Noncausal autoregressions for economic time series

Lanne, Markku and Saikkonen, Pentti

University of Helsinki

August 2010

Noncausal Autoregressions for Economic Time Series[†]

Markku Lanne[*] Pentti Saikkonen[**]
University of Helsinki University of Helsinki

August 2010

Abstract

This paper is concerned with univariate noncausal autoregressive models and their potential usefulness in economic applications. In these models, future errors are predictable, indicating that they can be used to empirically approach rational expectations models with nonfundamental solutions. In the previous theoretical literature, nonfundamental solutions have typically been represented by noninvertible moving average models. However, noncausal autoregressive and noninvertible moving average models closely approximate each other, and therefore,the former provide a viable and practically convenient alternative. We show how the parameters of a noncausal autoregressive model can be estimated by the method of maximum likelihood and derive related test procedures. Because noncausal autoregressive models cannot be distinguished from conventional causal autoregressive models by second order properties or Gaussian likelihood, a model selection procedure is proposed. As an empirical application, we consider modeling the U.S. inflation which, according to our results, exhibits purely forward-looking dynamics.

[†] This is a compeletely revised version of our previous working paper entitled "Modeling Expectations with Noncausal Autoregressions". We thank Michael McAleer, Efrem Castelnuovo, Martin Ellison and Antti Ripatti for useful comments. Financial support from the Academy of Finland and the OP-Pohjola Group Research Foundation is gratefully acknowledged. This research was completed while the second author was a Fernand Braudel Fellow in the Economics Department of the European University Institute.

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

1 Introduction

Univariate autoregressive models are commonly employed in characterizing the dynamics of economic time series. Typical empirical applications include forecasting and the measurement of persistence (Andrews and Chen (1994)), but also in theoretical macroeconomic (see, e.g., Canova (2007)) and financial (see, e.g., Campbell et al. (1997)) models, the dynamics of the variables is often described by an autoregressive structure. However, to the best of our knowledge, 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 or on the present and past values of the errors of the model. The noncausal autoregressive model proposed in this paper, in contrast, also contains leads in addition to lags, and thus, allows for dependence on the future. We argue that this is a particularly useful feature in economic applications where expectations play a central role. Another interesting feature of the noncausal autoregressive model is that its errors are pedictable by past values of the series which may be made use of in improving forecast accuracy if noncausality is detected. An interpretation of such predictability is that the errors contain effects of omitted variables that are predictable by the considered series.

In this paper, we concentrate on introducing a new noncausal autoregressive model formulation and illustrating its potential usefulness in economic applications while leaving its motivation in terms of economic theory mostly for future research. However, at least two cases leading to noncausality have already been discussed in the economics literature. First, a noncausal autoregressive model may arise as a nonfundamental solution of a rational expectations model, when the agents' information set is greater than that of the econometrician, who is estimating only a univariate model (see, e.g., Hansen and Sargent (1991)). The presence of noncausality indicates that the agents are able to forecast a part of the future values of the economic variable in question by information unknown to the econometrician, and this results in a noncausal autoregressive representation with predictable errors. Hence, noncausal

autoregressive models allow for taking the effect of the agents' true information set into account without explicitly specifying it. In this context, the setup has virtually always been formulated in such a way that a noninvertible moving average process arises as a solution instead of a noncausal autoregression, but as pointed out in Section 2 below, these models closely approximate each other. Second, besides this kind of discrepancy of information sets, heterogeneous information has been shown to be a potential cause of nonfundamental solutions with nonrevealing equilibria (see, e.g. Kasa et al. (2007)).

Although economic applications of noncausal time series models are virtually nonexistent, in the statistics literature, noncausal autoregressive and autoregressive moving average models have been studied, inter alia, by Breidt et al. (1991), Lii and Rosenblatt (1996), Huang and Pawitan (2000), Rosenblatt (2000), Breidt et al. (2001), Andrews et al. (2006, 2009), and Wu and Davis (2010). However, this literature is not voluminous, and typical applications have been confined to natural sciences and engineering.[1]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. In contrast to the models in the previous literature, our formulation achieves depencence on future errors by explicitly including both leads and lags of the variable in question. A useful implication of this is that, unlike in the previously introduced formulations, statistical inference on autoregressive parameters is facilitated, and it becomes, for example, straightforward to obtain likelihood based diagnostic tests for

[1]As far as we know, the only empirical examples of noncausal autoregressive (moving average) models with economic data are provided by Breidt et el. (2001) who demonstrate that a noncausal first order autoregressive model is appropriate for modeling a daily time series of Microsoft trading volume, and Andrews et al. (2009) and Wu and Davis (2010) who model the daily Wal-Mart trading volume. Empirical economic examples of related models with a noninvertible moving average part are given in Huang and Pawitan (2000) and Breidt et al. (2001). In the former paper a noninvertible moving average model is applied to U.S. unemployment rate, whereas the latter fits the so-called all-pass model to New Zealand/U.S. exchange rate.

the specified number of leads and lags. A further advantage is that 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.

Once allowance for noncausality is made, model selection becomes a more complicated empirical issue than in conventional causal autoregressions. In particular, in addition to the order of the autoregression, the number of leads and lags to include, must be decided upon. Which model is selected is also of great economic interest, as it yields information on the extent to which the variable in question depends on the past and future. Dependence on the future (or equivalently, predictability of future errors) suggests nonfundamentalness, as discussed above. It is well-known that in noncausal autoregressions a non-Gaussian error term is required to achieve identification. In previous economic applications, causal autoregressive processes with Gaussian error terms have typically been assumed. However, this approach has usually been justified by quasi maximum likelihood (ML) arguments as significant departures from Gaussianity, especially excess kurtosis, have been detected by diagnostic checks. In this paper, an error term with a t-distribution is found to provide an adequate fit, but other leptokurtic distributions could also be considered. Once the distribution of the error term has been specified, we follow Breidt et al. (1991) and consider a model selection algorithm based on the maximized log-likelihood function, augmented by diagnostic tests.

The proposed model is applied to study the U.S. inflation dynamics. A large part of the related voluminous previous literature based on univariate methods concentrates on the finding that inflation seems to be highly persistent which is in contrast with typical New Keynesian models assuming inflation to be forward-looking such that it depends on expected future but not past inflation. Previous empirical results are based on conventional causal autoregressive models where high persistence indeed necessarily implies backward-looking behavior in the basic New Keynesian model framework. However, our results suggest that a purely noncausal autoregressive model far better captures the U.S. inflation process, and this finding is recon-

firmed by the superior forecast accuracy of the peferred specification. Hence, the persistence previously found with univariate methods does not seem to be caused by agents' relying on past inflation. Instead, it is caused by predictability inherent in the noncausal autoregressive nature of the process, which, in turn, may be explained by nonfundamentalness due to omitting relevant variables in the univariate model. It should be pointed out that although a large part of the literature on inflation persistence, including this paper, is based on univariate models, typical New Keynesian models incorporate also other drivers of inflation, such as a measure of marginal costs. We argue that it is indeed the omission of such variables that makes the univariate autoregressive model too simple, which shows up as noncausality. However, if lagged inflation is not a significant predictor of current inflation in the univariate model, it should not be significant in a fully specified New Keynesian Phillips curve equation either. These findings indicate that analyses based on univariate conventional autoregressions are not useful in this context as they in no way take other relevant variables into account and may, therefore, yield misleading conclusions.

The rest of the paper is organized as follows. In Section 2, the noncausal autoregressive model is introduced and its properties are discussed. Section 3 considers (approximate) ML estimation and statistical inference in noncausal autoregressive models. In Section 4, a small-scale simulation study is conducted to examine the practical relevance of the asymptotic results of Section 3 as well as the aforementioned model selection procedure. Section 5 presents an empirical application to U.S. inflation. Finally, Section 6 concludes.

2 Model

Let y_t $(t = 0, \pm1, \pm2, ...)$ be a stochastic process 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, $\epsilon_t \sim i.i.d.(0, \sigma^2)$. Moreover, B is the usual backward shift operator, that is, $B^k y_t = y_{t-k}$ $(k = 0, \pm 1, ...)$, and the polynomials $\phi(z)$ and $\varphi(z)$ have their zeros outside the unit circle so that

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

If $\varphi_j \neq 0$ for some $j \in \{1, .., s\}$, equation (1) defines a noncausal autoregression referred to as purely noncausal when $\phi_1 = \cdots = \phi_r = 0$. The conventional causal autoregression is obtained when $\varphi_1 = \cdots = \varphi_s = 0$. Then the former condition in (2) guarantees the stationarity of the model. In the general set up of equation (1) the same is true for the process $u_t = \varphi(B^{-1}) y_t$ which has the backward moving average representation

$$u_t = \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j}, \tag{3}$$

where $\alpha_0 = 1$ and the coefficients α_j decay to zero at a geometric rate as $j \to \infty$. Similarly, the latter condition in (2) guarantees the stationarity of the purely noncausal process $v_t = \phi(B) y_t$ and the validity of its forward moving average representation

$$v_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t+j}, \tag{4}$$

where $\beta_0 = 1$ and the coefficients β_j decay to zero at a geometric rate as $j \to \infty$. The process y_t itself has the two-sided moving average representation

$$y_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}, \tag{5}$$

where ψ_j is the coefficient of z^j in the Laurent series expansion of $\phi(z)^{-1} \varphi(z^{-1})^{-1} \overset{def}{=} \psi(z)$. Specifically, by condition (2),

$$\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$$

exists in some annulus $b < |z| < b^{-1}$ with $b < 1$ and reduces to the one-sided special cases obtained from (3) and (4) when y_t is causal and purely noncausal, respectively.

The representation (5) implies that y_t is a stationary and ergodic process with finite second moments. We use the abbreviation AR(r, s) for the model defined by (1). In the causal case $s = 0$, the conventional abbreviation AR(r) is also used.

As already discussed in the Introduction, we can think the autoregressive representation of an economic variable as a solution of a rational expectations model. The solution may be fundamental or nonfundamental, the latter case being characterized by the process of the economic variable depending on future (nonfundamental) shocks. This property is shared by the noncausal autoregressive model (see (4) and (5)). As Hansen and Sargent (1991), among others, have pointed out, one case where an estimated model may turn out to be nonfundamental, arises if the econometrician's information set is smaller than that of the agents. Therefore, finding noncausality may be interpreted as the causal univariate autoregressive model being inadequate, despite the causal and noncausal models having the same autocorrelation function (see the discussion below). In this case, the noncausal autoregressive model captures effects of missing variables (the discrepancy between the agents' and the econometrician's information sets, with the latter consisting only of $\mathcal{F}_t$, the history of the variable in question up to time point t), and allows for explicitly modeling the dependence of realized values on future errors.[2]

It is easy to see that in the model, the current value of the process, y_t, is indeed affected by expected future errors. Using the definition of the process v_t and taking conditional expectation with respect to $\mathcal{F}_t$ on both sides of equation (4) yields

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_r y_{t-r} + \sum_{j=0}^{\infty} \beta_j E_t(\epsilon_{t+j}). \tag{6}$$

[2]In the previous literature, including Hansen and Sargent (1991), the nonfundamental solution has all but been equated with the nonivertible moving average (MA) representation. However, it is easy to see that the noncausal AR and noninvertible MA representations closely approximate each other, as in both y_t depends on future errors with the difference that in the latter the lead polynomial is of finite order. To see this, for simplicity, consider the noninvertible first order MA model $y_t = (1 - \gamma B)\epsilon_t$, where $|\gamma| > 1$. The aforementioned fact is seen by defining $\gamma_* = 1/\gamma$ and $\varepsilon_t = \epsilon_{t-1}/\gamma$ and observing that we can write $y_t = (1 - \gamma_* B^{-1})\varepsilon_t = -\sum_{j=1}^{\infty} \gamma_*^j y_{t+j} + \varepsilon_t$.

In a causal model, $\beta_j = 0$, $j > 0$, and the last term is just ϵ_t implying that expected future errors have no effect on y_t. However, because, from (5), ϵ_{t+j} $(j = 0, 1, ...)$ are not independent of $\mathcal{F}_t$, the last term is generally nonzero in a noncausal model, indicating the potential dependence of y_t on (an infinite number of) expected future errors. This, of course, implies that future errors are predictable by past values of y_t. Note, however, that in noncausal autoregressive models the prediction problem is, in general, nonlinear (see Rosenblatt (2000, Corollary 5.4.2, and Lanne et al. (2010)) and, therefore, $E_t(\epsilon_{t+j})$ cannot be computed in a simple fashion (see Lanne et al. (2010)). In particular, in a noncausal model $E_t(\epsilon_t) \neq \epsilon_t$ because ϵ_t depends on y_{t+j} $(0 < j \leq s)$ (see (1)).

A practical complication of noncausal autoregressive processes that probably underlies their unpopularity in the empirical economic literature, is that they cannot be identified by second order properties or Gaussian likelihood. This can be seen as follows. First, conclude from well-known results on linear filters that the spectral density function of the process y_t defined by (1) is given by $\sigma^2/2\pi \left|\phi\left(e^{-i\omega}\right)\varphi\left(e^{-i\omega}\right)\right|^2$. This also applies to the alternative formulation defined by (8) and (9) discussed in Section 3 below. The same spectral density is obtained from a causal autoregressive process with lag polynomial $\varphi(B)\,\phi(B)$ having its zeros outside the unit circle. These observations explain that y_t also has the causal representation

$$\varphi(B)\,\phi(B)\,y_t = \xi_t, \tag{7}$$

where the (stationary) innovation sequence ξ_t is uncorrelated but, in general, not independent with mean zero and variance σ^2 (cf. Brockwell and Davis (1987, p. 124–125)). Thus, even if y_t is noncausal, its spectral density and, hence, autocovariance function cannot be distinguished from those of a causal autoregressive process. Before applying a noncausal model it is therefore advisable in practice to first fit an (adequate) causal autoregression to the observed series by standard least squares or Gaussian ML and check whether the residuals look non-Gaussian.

Unless otherwise stated, we shall henceforth assume that 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 introduced earlier.

3 Parameter estimation and statistical inference

In this section, we derive the approximate likelihood function of the noncausal autoregression (1) and related asymptotic tests. It is useful to start by highlighting the benefits of our formulation over that used in the previous literature from the viewpoint of statistical inference. In the previous literature on noncausal autoregressions, it has been common to specify the model as

$$a(B)\, y_t = \varepsilon_t, \tag{8}$$

where $a(B) = 1 - a_1 B - \cdots - a_p B^p$ with $a_p \neq 0$ and ε_t is an $i.i.d.$ sequence with zero mean and finite variance (see, e.g., Breidt et al. (1991), Rosenblatt (2000) and the references therein). In this set up the relevant stationarity condition is $a(z) \neq 0$, $|z| = 1$. When it holds, y_t has a two-sided moving average representation similar to that in (5) (see Brockwell and Davis (1987, p. 88)). Moreover, when $p = r + s$ and the number of zeros of $a(z)$ outside (inside) the unit circle is r (s), one can factor the polynomial $a(z)$ as

$$a(z) = \varphi^*(z)\,\phi(z), \tag{9}$$

where $\phi(z)$ is as in (1) and $\varphi^*(z) = 1 - \varphi_1^* z - \cdots - \varphi_s^* z^s$ has its zeros inside the unit circle, that is, $\varphi^*(z) \neq 0$ for $|z| \geq 1$. Note that this particularly means that in the noncausal case $s > 0$, the condition $|\varphi_s^*| > 1$ holds.

The polynomial $\varphi^*(z)$ can be expressed as

$$\begin{aligned} \varphi^*(z) &= -\varphi_s^* z^s \left(1 + \frac{\varphi_{s-1}^*}{\varphi_s^*} z^{-1} + \cdots + \frac{\varphi_1^*}{\varphi_s^*} z^{1-s} - \frac{1}{\varphi_s^*} z^{-s}\right) \\ &= -\varphi_s^* z^s \varphi\left(z^{-1}\right), \end{aligned}$$

where $\varphi(z^{-1})$ is as in (1) so that $\varphi_{s-j}^*/\varphi_s^* = -\varphi_j$ for $j = 1, ..., s-1$ and $1/\varphi_s^* = \varphi_s$. Because the zeros of $\varphi^*(z)$ lie inside the unit circle, those of $\varphi(z)$ lie outside the unit

circle, as can be readily checked. Thus, the latter condition in (2) holds and model (1) can be obtained from (8) by defining $\epsilon_t = -(1/\varphi_s^*)\varepsilon_{t+s}$. Similarly, if $\varphi_s \neq 0$ is assumed in (1), the preceding reasoning can be reversed to obtain the specification (8) with $\varepsilon_t = -(1/\varphi_s)\,\epsilon_{t-s}$ and the coefficients of the polynomial $\varphi^*(z)$ in (9) given by $\varphi_j^* = -\varphi_j/\varphi_s$, $j = 1, ..., s-1$, and $\varphi_s^* = 1/\varphi_s$. Thus, when $\varphi_s \neq 0$ there is a one-to-one correspondence between the parameters in (1) and (8).[3]

The formulation (1) appears more convenient than (8) and (9) when one needs to specify the (usually) unknown model orders r and s. Indeed, it turns out to be quite feasible to construct conventional likelihood based tests for hypotheses such as $\phi_{r_0+1} = \cdots = \phi_r = 0$ $(r_0 < r)$ and $\varphi_{s_0+1} = \cdots = \varphi_s = 0$ $(s_0 < s)$. For the latter hypothesis similar test procedures seem to be more difficult to obtain if the model is formulated as in (8) and (9) because $|\varphi_s^*| > 1$ by assumption and because the logarithm of $|\varphi_s^*|$ appears in the likelihood function (see Breidt et al. (1991)). A further statistical convenience of the specification (1) is that the autoregressive parameters $\phi = (\phi_1, ..., \phi_r)$ and $\varphi = (\varphi_1, ..., \varphi_s)$ turn out to be orthogonal to the parameters σ^2 and λ implying asymptotic independence of the corresponding ML estimators.[4]

3.1 Approximate likelihood function

ML estimation of the parameters of a noncausal autoregression was studied by Breidt et al. (1991) by using the formulation based on equation (8). Even in this set up our model is slightly more general than theirs because we allow the distribution of the error term to depend on the additional parameter vector λ. This generalization

[3] This kind of reparameterization of model (8) is mentioned in Lii and Rosenblatt (1996, p. 17) in the context of a noncausal and noninvertible autoregressive moving average model. However, in that paper the model is not explicitly written as in (1) and the case $\varphi_s = 0$ allowed in (1) is not discussed.

[4] We use the notation $x = (x_1, ..., x_n)$ to introduce the n-dimensional vector x and its components. The same convention is also used when the components are vectors. In matrix calculations all vectors are interpreted as column vectors and a prime is used to signify the transpose of a vector or a matrix.

has been considered by Andrews et al. (2006) in a related context and, following the arguments used in their paper, it can also be straightforwardly handled in our case. Thus, we shall assume that the density function $f(x;\lambda)$ satisfies the regularity conditions of Andrews et al. (2006) which, among other things, require 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. The permissible parameter space of λ, denoted by Λ, is some subset of $\mathbb{R}^d$ whereas the permissible space of the parameters ϕ, φ and σ is defined by the conditions in (2) and by $\sigma > 0$. For convenience, the regularity conditions of Andrews et al. (2006) are also presented in the appendix and, unless otherwise stated, they will henceforth be assumed. Densities that satisfy these conditions include a rescaled t-density and a weighted average of Gaussian densities.

If the model is defined as in (8) and (9), ML estimators of the parameters in (1) can be derived by a smooth one-to-one transformation from ML estimators of the parameters in (8), and hence their limiting distribution can also be easily obtained. However, because this reasoning is not directly applicable if the degree of the polynomial $\varphi(z)$ is overspecified (i.e., $\varphi_s = 0$), we shall provide details based directly on the specification (1). We start by deriving the likelihood function.

Suppose we have an observed time series $y_1, ..., y_T$. Using the definitions $u_t = \varphi(B^{-1})\, y_t$ and $v_t = \phi(B)\, y_t$ we can write

$$\begin{bmatrix} u_1 \\ \vdots \\ u_{T-s} \\ v_{T-s+1} \\ \vdots \\ v_T \end{bmatrix} = \begin{bmatrix} y_1 - \varphi_1 y_2 - \cdots - \varphi_s y_{s+1} \\ \vdots \\ y_{T-s} - \varphi_1 y_{T-s+1} - \cdots - \varphi_s y_T \\ y_{T-s+1} - \phi_1 y_{T-s} - \cdots - \phi_r y_{T-s+1-r} \\ \vdots \\ y_T - \phi_1 y_{T-1} - \cdots - \phi_r y_{T-r} \end{bmatrix} = A \begin{bmatrix} y_1 \\ \vdots \\ y_{T-s} \\ y_{T-s+1} \\ \vdots \\ y_T \end{bmatrix}$$

or briefly

$$x = Ay.$$

Similarly,

$$\begin{bmatrix} u_1 \\ \vdots \\ u_r \\ \epsilon_{r+1} \\ \vdots \\ \epsilon_{T-s} \\ v_{T-s+1} \\ \vdots \\ v_T \end{bmatrix} = \begin{bmatrix} u_1 \\ \vdots \\ u_r \\ u_{r+1} - \phi_1 u_r - \cdots - \phi_r u_1 \\ \vdots \\ u_{T-s} - \phi_1 u_{T-s-1} - \cdots - \phi_r u_{T-s-r} \\ v_{T-s+1} \\ \vdots \\ v_T \end{bmatrix} = C \begin{bmatrix} u_1 \\ \vdots \\ u_r \\ u_{r+1} \\ \vdots \\ u_{T-s} \\ v_{T-s+1} \\ \vdots \\ v_T \end{bmatrix}$$

or

$$z = Cx.$$

Hence, the vectors z and y are related by

$$z = CAy.$$

Note that from (3) and (4) it can be seen that the components of z given by $(u_1, ..., u_r)$, $(\epsilon_{r+1}, ..., \epsilon_{T-s})$, and $(v_{T-s+1}, ..., v_T)$ are independent. The joint density function of z under true parameter values can thus be expressed as

$$h_U(u_1, ..., u_r) \left(\prod_{t=r+1}^{T-s} f_\sigma(\epsilon_t; \lambda) \right) h_V(v_{T-s+1}, ..., v_T),$$

where h_U and h_V signify the joint density functions of $(u_1, ..., u_r)$ and $(v_{T-s+1}, ..., v_T)$, respectively. It is easy to see that the (nonstochastic) matrices A and C are non-singular and the determinant of C is unity so that we can express the joint density function of the data vector y as

$$h_U\left(\varphi\left(B^{-1}\right) y_1, ..., \varphi\left(B^{-1}\right) y_r\right) \left(\prod_{t=r+1}^{T-s} f_\sigma\left(\varphi\left(B^{-1}\right) \phi(B)\, y_t; \lambda\right) \right) \times h_V(\phi(B)\, y_{T-s+1}, ..., \phi(B)\, y_T) \left|\det(A)\right|.$$

It is also easy to check that the determinant of the $(T-s)\times(T-s)$ block in the upper left hand corner of A is unity and, using the well-known formula for the determinant of a partitioned matrix, it can furthermore be seen that the determinant of A is independent of the sample size T. This suggests approximating the joint density of y by the second factor in the preceding expression, giving rise to the approximate log-likelihood function

$$l_T(\theta) = \sum_{t=r+1}^{T-s} g_t(\theta), \tag{10}$$

where $\theta = (\phi, \varphi, \sigma, \lambda)$ and

$$\begin{aligned} g_t(\theta) &= \log f\left(\sigma^{-1}\left(u_t(\varphi) - \phi_1 u_{t-1}(\varphi) - \cdots - \phi_r u_{t-r}(\varphi)\right); \lambda\right) - \log\sigma \\ &= \log f\left(\sigma^{-1}\left(v_t(\phi) - \varphi_1 v_{t+1}(\phi) - \cdots - \varphi_s v_{t+s}(\phi)\right); \lambda\right) - \log\sigma. \end{aligned}$$

Here $u_t(\varphi)$ and $v_t(\phi)$ signify the series u_t and v_t treated as functions of the parameters φ and ϕ, respectively. Maximizing $l_T(\theta)$ over permissible values of θ gives an approximate ML estimator of θ. Note that here, as well as in the next section, the orders r and s are assumed known. Procedures to specify these quantities will be discussed in later sections of the paper.

3.2 Asymptotic properties of the approximate ML estimator

In what follows, it will be convenient to use the notation θ_0 for the true value of θ and similarly for its components. It is assumed that λ_0, the true value λ, is an interior point of Λ.

We shall first consider the score of θ evaluated at true parameter values. Define the vectors $U_{t-1} = (u_{t-1}, ..., u_{t-r})$ and $V_{t+1} = (v_{t+1}, ..., v_{t+s})$ where u_t and v_t are defined in terms of true parameter values so that $u_t = \sum_{j=0}^{\infty} \alpha_{0j}\epsilon_{t-j}$ and $v_t = \sum_{j=0}^{\infty} \beta_{0j}\epsilon_{t+j}$. By straightforward differentiation (cf. Breidt et al. (1991)) we find from (10) that

$$\frac{\partial}{\partial\phi} g_t(\theta_0) = -\frac{f'\left(\sigma_0^{-1}\epsilon_t; \lambda_0\right)}{\sigma_0 f\left(\sigma_0^{-1}\epsilon_t; \lambda_0\right)} U_{t-1} \quad (r \times 1)$$

and

$$\frac{\partial}{\partial\varphi}g_t(\theta_0) = -\frac{f'\left(\sigma_0^{-1}\epsilon_t;\lambda_0\right)}{\sigma_0 f\left(\sigma_0^{-1}\epsilon_t;\lambda_0\right)}V_{t+1} \quad (s\times 1),$$

where $f'(x,\lambda) = \partial f(x,\lambda)/\partial x$ and use has also been made of the fact that $\phi_0(B)u_t = \epsilon_t = \varphi_0(B)v_t$ with $\phi_0(B)$ and $\varphi_0(B)$ defined in terms of true parameter values (e.g. $\phi_0(B) = 1-\phi_{01}B-\cdots-\phi_{0r}B^r$). Similarly,

$$\frac{\partial}{\partial\sigma}g_t(\theta_0) = -\sigma_0^{-2}\left(\frac{f'\left(\sigma_0^{-1}\epsilon_t;\lambda_0\right)}{f\left(\sigma_0^{-1}\epsilon_t;\lambda_0\right)}\epsilon_t + \sigma_0\right)$$

and

$$\frac{\partial}{\partial\lambda}g_t(\theta_0) = \frac{1}{f\left(\sigma_0^{-1}\epsilon_t;\lambda_0\right)}\frac{\partial}{\partial\lambda}f\left(\sigma_0^{-1}\epsilon_t;\lambda_0\right) \quad (d\times 1).$$

The following lemma presents the asymptotic distribution of the score vector. For the presentation of this lemma we need some notation. Let $\eta_t \sim i.i.d.(0,1)$ and define the AR(r) process u_t^* by $\phi_0(B)u_t^* = \eta_t$ and the AR(s) process v_t^* by $\varphi_0(B)v_t^* = \eta_t$. Note that u_t^* and v_t^* are jointly stationary and causal with finite second moments. Next form the vectors $U_{t-1}^* = \left(u_{t-1}^*,...,u_{t-r}^*\right)$ and $V_{t-1}^* = \left(v_{t-1}^*,...,v_{t-s}^*\right)$ and the associated covariance matrices $\Gamma_{U^*} = Cov\left(U_{t-1}^*\right)$, $\Gamma_{V^*} = Cov\left(V_{t-1}^*\right)$, and $\Gamma_{U^*V^*} = Cov\left(U_{t-1}^*, V_{t-1}^*\right) = \Gamma_{V^*U^*}'$. We also define

$$\mathcal{J} = \int \frac{\left(f'(x;\lambda_0)\right)^2}{f(x;\lambda_0)}dx$$

and set

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} \mathcal{J}\Gamma_{U^*} & \Gamma_{U^*V^*} \\ \Gamma_{V^*U^*} & \mathcal{J}\Gamma_{V^*} \end{bmatrix}.$$

Note that $\Gamma_{U^*} = \sigma_0^{-2}Cov(U_{t-1})$, $\Gamma_{V^*} = \sigma_0^{-2}Cov(V_{t+1})$, and $\mathcal{J} > 1$ (see condition (A5) of Andrews et al. (2006)). Finally, define the $(d+1)\times(d+1)$ matrix

$$\Omega = \begin{bmatrix} \omega_\sigma^2 & \omega_{\sigma\lambda} \\ \omega_{\lambda\sigma} & \Omega_{\lambda\lambda} \end{bmatrix}, \tag{11}$$

where

$$\Omega_{\lambda\lambda} = \int \frac{1}{f(x;\lambda_0)}\left(\frac{\partial}{\partial\lambda}f(x;\lambda_0)\right)\left(\frac{\partial}{\partial\lambda}f(x;\lambda_0)\right)' dx,$$

$$\omega_{\lambda\sigma} = -\sigma_0^{-1} \int x \frac{f'(x;\lambda_0)}{f(x;\lambda_0)} \frac{\partial}{\partial\lambda} f(x;\lambda_0)\, dx = \omega'_{\sigma\lambda},$$

and

$$\omega_\sigma^2 = \sigma_0^{-2} \left(\int x^2 \frac{(f'(x;\lambda_0))^2}{f(x;\lambda_0)} dx - 1 \right).$$

Now we can present the limiting distribution of the score vector.[5]

Lemma 1 *If conditions (A1)–(A7) of Andrews et al. (2006) hold, then*

$$(T-p)^{-1/2} \sum_{t=r+1}^{T-s} \frac{\partial}{\partial\theta} g_t(\theta_0) \xrightarrow{d} N(0, diag(\Sigma, \Omega)).$$

Moreover, the matrices Σ and Ω are positive definite.

Lemma 1 can be proved in the same way as Propositions 1 and 2 of Breidt et al. (1991). An outline of the needed arguments is provided in the appendix. Here we note that the positive definiteness of the matrix Σ follows from the above mentioned inequality $\mathcal{J} > 1$ which holds when ϵ_t is non-Gaussian (see Remark 2 of Andrews et al. (2006)). The matrix Σ is positive definite even if the model order r or s is overspecified or both are overspecified. For instance, suppose that $r = s$ and consider the extreme case where $\phi = \varphi = 0$. Then, $\Sigma_{11} = \Sigma_{22} = \mathcal{J} I_r$ and $\Sigma_{12} = I_r$ so that the matrix Σ is clearly positive definite when $\mathcal{J} > 1$. In the general case of Lemma 1 the positive definiteness of the matrix Ω must be assumed (cf. condition (A6) of Andrews et al. (2006)). The block diagonality of the covariance matrix of the limiting distribution implies that the scores of (ϕ, φ) and (σ, λ) are asymptotically independent. This property, commonly referred to as orthogonality of the parameters (ϕ, φ) and (σ, λ), is convenient because it means that statistical inference on the autoregressive parameters ϕ and φ, which is typically of primary interest, is asymptotically independent of the estimation of the parameters σ and λ describing the distribution of the error term ϵ_t. It may be noted that similar orthogonality does not hold if the formulation given by (8) and (9) is used because then the score of the autoregressive parameter φ_s^* is

[5] The notation $diag(A_1, A_2)$ signifies a block diagonal matrix with diagonal blocks A_1 and A_2.

asymptotically correlated with the score of the scale parameter of the error term ε_t (see Proposition 2 of Breidt et al. (1991)).

Using a conventional Taylor series expansion of the score in conjunction with Lemma 1 and the assumed regularity conditions one can show the existence of a consistent and asymptotically normal (local) maximizer of the approximate likelihood function. Specifically, the following theorem can be established. Its proof makes use of arguments similar to those in Breidt et al. (1991) and Andrews et al. (2006) and is outlined in the appendix.

Theorem 2 *If conditions (A1)–(A7) of Andrews et al. (2006) hold, there exists a sequence of (local) maximizers* $\hat{\theta} = (\hat{\phi}, \hat{\varphi}, \hat{\sigma}, \hat{\lambda})$ *of* $l_T(\theta)$ *in (10) such that*

$$(T-p)^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, diag\left(\Sigma^{-1}, \Omega^{-1}\right)\right).$$

Due to the block diagonality of the covariance matrix of the limiting distribution, the (approximate) ML estimators $(\hat{\phi}, \hat{\varphi})$ and $(\hat{\sigma}, \hat{\lambda})$ are asymptotically independent. This means that if a consistent initial estimator $(\tilde{\phi}, \tilde{\varphi})$ of (ϕ, φ) is available an estimator of (σ, λ) with the same asymptotic distribution as the ML estimator $(\hat{\sigma}, \hat{\lambda})$ can be obtained by maximizing the function $l_T(\tilde{\phi}, \tilde{\varphi}, \sigma, \lambda)$. As the initial estimator $(\tilde{\phi}, \tilde{\varphi})$ one may consider the least absolute deviation (LAD) estimator based on the (possibly incorrect) assumption that ϵ_t has a Laplace (or double exponential) distribution. In the case of the specification (8) Huang and Pawitan (2000) establish the consistency of the LAD estimator when, in a certain sense, the true distribution of ϵ_t has tails heavier than the normal distribution. Their result applies to a variety of known distributions including the rescaled t-distribution and weighted averages of Gaussian densities. An inspection of the residuals based on a LAD estimation may also help to specify an appropriate distribution for the error term ϵ_t.

3.3 Statistical inference

To be able to compute approximate standard errors for the components of the estimator $\hat{\theta}$ and construct confidence intervals and conventional Wald tests we need

consistent estimators of the covariance matrices Σ and Ω. We use the conventional estimator based on the Hessian of the approximate log-likelihood function which yields a consistent estimator, as discussed in the proof of Theorem 2. Specifically, we have

$$\hat{Q} \stackrel{def}{=} -(T-p)^{-1} \sum_{t=r+1}^{T-s} \frac{\partial^2}{\partial\theta\partial\theta'} g_t(\hat{\theta}) \xrightarrow{p} diag\,(\Sigma, \Omega)\,. \tag{12}$$

Approximate standard errors of the components of $\hat{\theta}$ can be obtained by computing the square roots of the diagonal elements of the matrix $(T-p)^{-1}\hat{Q}^{-1}$. Conventional Wald tests are also readily obtained. For instance, one can consider testing the null hypotheses

$$H_{r_0,s_0} : \phi_{0,r_0+1} = \cdots = \phi_{0r} = 0 \quad \text{and} \quad \varphi_{0,s_0+1} = \cdots = \varphi_{0s} = 0,$$

where $r_0 < r$ and $s_0 < s$ with the case $r_0 = r$ or $s_0 = s$ obtained in with an obvious modification. Under this null hypothesis at least one of the model orders can be reduced. To generalize slightly, consider the null hypothesis $H : R\theta_0 = 0$ where the (known) $m \times (r+s+d+1)$ matrix R is of full row rank. The conventional Wald test statistic can be written as

$$\mathcal{W} = (T-p)\,\hat{\theta}' R'(R\hat{Q}^{-1}R')^{-1}R\hat{\theta} \xrightarrow{d} \chi_m,$$

where the convergence assumes the null hypothesis and is an immediate consequence of Theorem 2.

One may also use the likelihood ratio (LR) test. Let $\tilde{\theta}$ signify the ML estimator of the parameter θ constrained by the null hypothesis H so that in the case of the hypothesis H_{r_0,s_0} the estimator $\tilde{\theta}$ is obtained by applying ML in the model with orders r_0 and s_0. The LR test statistic is

$$\mathcal{LR} = 2[l_T(\hat{\theta}) - l_T(\tilde{\theta})] \xrightarrow{d} \chi_m,$$

where the null hypothesis is again assumed. The limiting distribution can be justified by a standard application of the results given in the appendix which can also be

used to obtain the corresponding score (Lagrange multiplier) test. To the best of our knowledge, test procedures of this kind have not been explicitly considered in the previous literature of noncausal autoregressive models where the model is formulated as in (8) and (9). In this formulation treating the null hypothesis which specifies $s_0 < s$ is hampered by the condition $|\varphi_s^*| > 1$.

4 Simulation study

To study the finite-sample properties of the estimators and tests proposed in Section 3, we conducted a small simulation study. Following Breidt et al. (1991), we concentrate on the second-order process as the data-generating process (DGP) because it is the simplest model that allows for a versatile analysis of various aspects of estimation and testing. Throughout, the results are based on 10,000 realizations. We generate each realization in two steps. First, a series from the causal AR(r) model $\phi(B)\, v_t = \epsilon_t$ $(t = r+1, ..., T)$ is generated. Then y_t is computed recursively from $\varphi(B^{-1})\, y_t = v_t$ for $t = T-s, ..., 1$. The r and s initial observations, respectively, are set to zero, and to eliminate initialization effects, 100 observations at the beginning and end of each realization are discarded. In all experiments, the error term ϵ_t is assumed to follow the t-distribution with 3 degrees of freedom and σ is set equal to 0.1. We consider three different combinations of parameter values, $(\phi_1, \varphi_1) = \{(0.9, 0.9), (0.9, 0.1), (0.1, 0, 9)\}$. In the first case, the roots of the lag polynomials are equal and close to the unit circle, in the two other cases the roots of the "causal" and "noncausal" polynomials are clearly different. Three sample sizes, 100, 200 and 500 are considered.

The mean and standard deviation of the ML estimators of ϕ_1 and φ_1 are presented in Table 1. Even with as few as 100 observations the parameters are relatively accurately estimated in each case, and the biases as well as the standard deviations clearly diminish as the sample size increases. In the case $(\phi_1, \varphi_1) = (0.9, 0.9)$, ϕ_1 is more accurately estimated in terms of both criteria, whereas in the other two cases it

is the parameter taking the smaller value that is estimated with a somewhat smaller bias. The differences are, however, minor.

The results concerning the Wald and LR tests of hypotheses involving a single parameter in Table 2 indicate that both tests tend to overreject to some extent, but the problem is mitigated as the sample size increases. For the Wald test, the case $(\phi_1, \varphi_1) = (0.9, 0.9)$ seems to be the most difficult, while the differences between the rejection rates of the Wald and LR tests are minor in the other cases. In general, the LR tests on the parameter with the smaller value have somewhat better size properties, in accordance with the properties of the ML estimator above. In contrast, this pattern does not carry over to the Wald test.

As the Wald test tends to overreject in the $(\phi_1, \varphi_1) = (0.9, 0.9)$ case, we only present simulation results on power for the LR test. Because the size properties do not differ much between the different DGP's, only the rejection rates of the LR test (at the nominal 5% level of significance) for the first DGP ($(\phi_1, \varphi_1) = (0.9, 0.9)$) are presented in Figure 1. Moreover, we concentrate on tests concerning ϕ_1 because there is no reason to expect the power properties to greatly depend on the particular parameter. The values of ϕ_1 in the alternative DGP's that are used to generate the data are given by $0.9 - c/\sqrt{T}$ ($c = 0.0, 0.2, 0.4, \ldots, 2.0$), and the null hypothesis in the test is $\phi_1 = 0.9$. The rejection rates for alternatives very close to the null are moderate for all sample sizes considered ($T = 100, 200, 500$), but they rapidly increase with c, especially with the greater sample sizes. Hence, the LR test seems to have reasonable power. These results, however, suggest that in small samples, one should not rely on this test alone in model selection.

Breidt et al. (1991) suggested a model selection procedure based on maximizing the likelihood function. In other words, all purely causal, noncausal and mixed models of a given order (p) are estimated, and the model yielding the greatest value of the likelihood function is selected. Their simulation results lend support to this procedure, and in Table 4, we present similar results when the DGP is the mixed second-order model. The procedure seems to work relatively well even with 100 observations, and

the performance greatly improves with the sample size. However, there seem to be some differences depending on the parameter values. When $(\phi_1, \varphi_1) = (0.9, 0.9)$, the correct model is selected in 95% of the realizations with 200 observations, and the corresponding figure is 99.9% with 500 observations. In contrast, in the cases with different parameter values, the noncausal (causal) model is selected too often when ϕ_1 is greater (smaller), even with 500 observations. In these cases model selection is presumably complicated by the fact that the considered processes are rather close to first-order processes. Although the proposed procedure works fairly well even in these difficult cases, additional simulation experiments involving greater values of the other parameter (not reported) indicated improvements, with the correct model sometimes being selected even more frequently than in the $(\phi_1, \varphi_1) = (0.9, 0.9)$ case. Despite the quite satisfactory performance of this procedure, the results suggest that model selection should not be based on this criterion alone, but, in addition, diagnostic tests should be employed.

5 Empirical application

In this section, we apply the models and methods discussed above to modeling U.S. inflation dynamics. Our focus is on examining the nature of inflation persistence that has given rise to a voluminous literature in the past few decades. The central question in this line of research is whether inflation is a purely forward-looking variable as required by the basic New Keynesian model. This assumption has been tested by checking for serial correlation in inflation, and typically measures based on univariate autoregressive models, such as the cumulative impulse response (CIR) (Andrews and Chen (1994)), have indicated quite high persistence of inflation in industrialized countries (for a survey of the recent empirical literature, see Cecchetti and Debelle (2006)). The presence of high autocorrelation has been interpreted as evidence in favor of the dependence of inflation on its past values, and, hence, against the forward-looking inflation expectations assumed in the basic New Keynesian model. This, in turn, has

led to modifications of existing theory that try to explain the apparently backward-looking behavior (see, e.g., Gali and Gertler (1999)).

This paper contributes to the large empirical literature that studies inflation persistence in the univariate framework only. This approach excludes potential drivers of the inflation process included in macroeconomic theories of price determination, such as the marginal costs and output gap. However, if lagged inflation turns out not to enter the univariate model, it should not be significant in a model augmented with any of these additional drivers either. As already discussed, the noncausal AR model is to some extent able to take these additional variables into account, while the conventional AR model with potentially predictable errors fails to do so, and, therefore, may yield misleading conclusions.

To the best of our knowledge, only causal autoregressive models have been entertained in the previous literature on inflation. As a consequence, high persistence has automatically been interpreted as evidence of the dependence of inflation expectations on past inflation (see Cecchetti and Debelle (2006), and the references therein). However, as discussed in Section 2, high autocorrelation and, hence, strong persistence do not, per se, indicate such dependence. Even if current inflation only depends on expected future inflation (or equivalently expected future errors to inflation, incorporating factors that drive inflation), the process may be persistent if autocorrelation is used as a measure of persistence. The same is true if the CIR based on a causal autoregressive model is used to measure persistence. Indeed, as seen in Section 2, for any purely noncausal autoregressive process there is a corresponding causal process with the same lag polynomial and, hence, the same autocorrelation function and impulse response function. Thus, causality or noncausality and, hence, dependence on past or expected future errors, cannot be distinguished by examining the autocorrelation function or the impulse response function of a causal autoregressive model fitted to the series.

In what follows, we will use the procedures proposed earlier in the paper to argue that the U.S. inflation series is purely noncausal despite its strong persistence. This

can be interpreted as evidence in favor of inflation being dependent on expectations of future inflation and not on past inflation. In view of the discussion in Section 2, finding noncausality would suggest the presence of nonfundamentalness in the univariate inflation process, with the likeliest explanation being that agents have other information relevant for predicting inflation besides the past and present of the inflation series alone. Moreover, if pure noncausality is found, past inflation is not useful in predicting current inflation over and above this other information.

The inflation series that we model, is the annualized quarterly inflation rate computed from the seasonally adjusted U.S. consumer price index (for all urban consumers) published by the Bureau of Labor Statistics. The sample period comprises 155 observations, from 1970:1 to 2008:3. There is positive autocorrelation even at high lags as shown by the autocorrelation function depicted in Figure 2. The Ljung-Box test indicates that autocorrelation is also significant at all reasonable significance levels. However, by visual inspection and unit root tests, the series can be considered stationary. Further evidence of persistence is provided by the CIR based on the causal Gaussian AR(3) model that turned out to adequately capture the linear dependence in the inflation series (see model AR(3,0)-N inTable 5). The CIR of this model equals 6.98 which is comparable to the values obtained by Cecchetti and Debelle (2006) for the OECD countries, indicating high persistence.

In Table 5, we present the estimation results of a number of autoregressive models for the demeaned inflation, along with some diagnostic tests.[6] Of Gaussian autoregressive models up to order 4, the AR(3) model (AR(3,0)-N) was selected by both the Akaike (AIC) and Bayesian (BIC) information criteria. However, the diagnostic tests suggest that this model is misspecified. Although the Ljung-Box test does not indicate the presence of unmodeled autocorrelation, there is evidence of conditional heteroskedasticity, as the p-value of the McLeod-Li test is only 0.003.[7] Moreover, the

[6] Estimation is done using the BHHH algorithm in the GAUSS CMLMT library.

[7] Note that, when the orders of the model are misspecified, the Ljung-Box and McLeod-Li tests are not exactly valid as they do not take estimation errors correctly into account. The reason is that

quantile-quantile plot of the residuals in the upper panel of Figure 3 indicates that the normal distribution fails to capture the tails of the error distribution. Also, normality of the residuals of the AR(3,0)-N model is rejected by the Jarque-Bera test at the 5% level (p-value is 0.031). These findings suggest that a more leptokurtic distribution, such as the t-distribution with a relatively small degrees-of-freedom parameter might provide a more satisfactory fit.

Because a Gaussian AR(3) model is deemed adequate in describing the autocorrelation structure of the inflation series, we proceed by estimating all alternative causal and noncausal AR(r, s) models with $r + s = 3$, following the procedure proposed in Section 4. The error term is assumed to have a t-distribution with λ degrees of freedom.[8] Of the four models, the purely noncausal model (AR(0,3)-t) maximizes the log-likelihood function by a clear margin to the other specifications. With the exception of the AR(2,1)-t model, all specifications with t-distributed errors exhibit little evidence of remaining autocorrelation or conditional heteroskedasticity. The adequacy of the AR(0,3)-t model was also checked by testing it against higher-order specifications, and the coefficients of the additional terms turned out to be insignificant in the LR test. The p-values of the extra parameter in the AR(1,3)-t and AR(0,4)-t models, are 0.339 and 0.395, respectively. Hence, the results attest to purely noncausal inflation dynamics, indicating that it is the expectations of future errors that drive the inflation process (see (6)).

---

a misspecification of the model orders makes the errors dependent, as pointed out in the case of the causal specification (7). Nevertheless, p-values of these tests can be seen as convenient summary measures of the autocorrelation remaining in residuals and their squares.

[8] The log-likelihood function equals

$$l_T(\theta) = \sum_{t=r+1}^{T-s} g_t(\theta),$$

where

$$g_t(\theta) = \log\left\{\frac{\Gamma\left[(\lambda+1)/2\right]}{\pi^{1/2}\Gamma(\lambda/2)}(\lambda-2)^{-1/2}\left[1+\frac{\sigma^{-2}\epsilon_t^2}{\lambda-2}\right]^{-(\lambda+1)/2}\right\} - \log\sigma,$$

and $\Gamma(\cdot)$ is the gamma function.

In all cases, the degrees-of-feedom parameter λ is estimated small, indicating fat-tailed error distributions. This is not surprising given the bad fit of the Gaussian AR(3) model. The quantile-quantile plot of the AR(0,3)-t model depicted in the lower panel of Figure 3 lends support to the adequacy of the t-distribution. As a matter of fact, all models with t-distributed errors generated a similar quantile-quantile plot, indicating that great improvements in fit are brought about by merely appropriately selecting the error distribution.

Further evidence in favor of the selected purely noncausal specification is provided by the pseudo out-of-sample forecast comparisons in Table 6. The forecasts are computed by the simulation-based method introduced in Lanne et al. (2010). The out-of-sample period starts in 1982:1, and thus comprises 107 quarters. The forecasts are based on an expansive estimation window, with the first estimates based on the first 48 observations. At all forecast horizons considered, employing the t-distribution brings about slight improvements in the root mean square error (RMSE) compared to the Gaussian AR model. The noncausal AR(0,3)-t model, in general, leads to further diminution of the RMSE, which is also statistically significant at the three- and four-quarter horizons compared to the AR(3,0)-N model at the 1% level according to the test of Diebold and Mariano (1995) and West (1996).

In summary, the results strongly indicate purely noncausal inflation dynamics. Hence, the apparent persistence in inflation observed in univariate analyses does not seem to be caused by dependence on past inflation but by the predictability of nonfundamental shocks to inflation. These findings lend little support to the hybrid Phillips curve specification incorporating lagged inflation (see, e.g., Gali and Gertler (1999)).

6 Conclusion

In this paper, we have considered univariate noncausal autoregressive models that, to the best of our knowledge, have so far not attracted attention in the economics and finance literatures. Their economic motivation is mostly left for future research, but

generally they can be seen as representations of nonfundamental solutions to rational expectations models. In the previous theoretical literature, economic models have virtually invariably been formulated in such a way that these solutions have a noninvertible moving average instead of a noncausal autoregressive representation. However, these two representations are close approximations of each other, and, therefore, the noncausal AR model provides a viable and useful alternative to the noninvertible moving average model. In particular, the outlined model selection procedure facilitates distinguishing between fundamental and nonfundamental representations, and if noncausality (or equivalently nonfundamentalness) is detected, estimating the resulting noncausal AR model. Corresponding results for noninvertible moving average models appear less straightforward.

We discuss ML estimation and develop related tests for noncausal autoregressive models. Furthermore, based on a number of simulation experiments and our experience with actual economic data, we propose a three-step procedure for specifying a potentially noncausal autoregressive model. The first step is to fit a conventional causal autoregressive model by least squares or Gaussian ML and determine its order by using conventional procedures such as diagnostic checks and model selection criteria. Once an adequate causal model is found, its error term should be tested for Gaussianity. Because identification requires the error term to be non-Gaussian, we can proceed only if deviations from Gaussianity are detected. A variety of error distributions can be considered; in our empirical application we successfully employed the t-distribution. With the chosen error distribution, all causal and noncausal autoregressive models of the selected order are then estimated and the model maximizing the log-likelihood function is selected. Finally, through diagnostic tests the adequacy of this model is confirmed. These diagnostic checks should give information on directions in which the model potentially fails.

The model is applied to U.S. inflation dynamics that are shown to be purely noncausal. In the previous macroeconomic literature, the strong autocorrelation in inflation series has been seen as evidence in favor of the dependence of current on past

inflation, invalidating the basic New Keynesian model. Within the purely noncausal autoregression suggested by our results, the persistence can instead be interpreted as arising from the predictability of future errors to inflation that does not contradict the basic model. This predictability, in turn, indicates the nonfundamentalness of the simple univariate model without additional driving variables. However, since lagged inflation does not enter the selected univariate model for inflation, it should not enter a (fundamental) model augmented with additional variables.

In future work, we plan to look at extensions of the univariate model considered in this paper. Being able to handle multiple times series would be of interest in most economic applications, and a first attempt in this direction was recently put forth by Lanne and Saikkonen (2010). Using noncausal autoregressions to model financial returns is another obvious field of application. To be able to adequately capture the erratic behavior of these time series probably calls for extensions of the basic model proposed in this paper. In particular, allowing for noncausality is, per se, hardly sufficient to model the conditional heteroskedasticity prevalent in financial returns. Finally, examining the connections of noncausal time series models and economic theory, in particular finding theoretical economic models giving rise to noncausal solutions would be of great interest.

Mathematical appendix

We shall first present the regularity conditions (A1)–(A7) of Andrews et al. (2006). We use $\Lambda_0 \subset \Lambda$ to signify some neighborhood of λ_0.

(A1) For all $x \in \mathbb{R}$ and all $\lambda \in \Lambda$, $f(x;\lambda) > 0$ and $f(x;\lambda)$ is twice continuously differentiable with respect to (x, λ).

(A2) For all $\lambda \in \Lambda_0$, $\int x f'(x;\lambda)\, dx = x f(x;\lambda)\,|_{-\infty}^{\infty} - \int f(x;\lambda)\, dx = -1$.

(A3) $\int f''(x;\lambda_0)\, dx = f'(x;\lambda_0)\,|_{-\infty}^{\infty} = 0$.

(A4) $\int x^2 f''(x;\lambda_0)\, dx = x^2 f'(x;\lambda_0)\,|_{-\infty}^{\infty} - 2\int x f'(x;\lambda_0)\, dx = 2$.

(A5) $1 < \int \left(f'(x;\lambda_0)\right)^2 / f(x;\lambda_0)\, dx$.

(A6) The matrix Ω defined in (11) is positive definite.

(A7) For $j, k = 1, ..., d$ and all $\lambda \in \Lambda_0$,

- $f(x;\lambda)$ is dominated by a function $f_1(x)$ such that $\int x^2 f_1(x)\, dx < \infty$, and
- $x^2 \dfrac{\left(f'(x;\lambda)\right)^2}{f(x;\lambda)^2}$, $x^2 \left|\dfrac{f''(x;\lambda)}{f(x;\lambda)}\right|$, $|x| \left|\dfrac{\partial f'(x;\lambda)/\partial\lambda_j}{f(x;\lambda)}\right|$, $\dfrac{\left(\partial f'(x;\lambda)/\partial\lambda_j\right)^2}{f^2(x;\lambda)}$, and $\dfrac{|\partial^2 f(x;\lambda)/\partial\lambda_j\partial\lambda_k|}{f(x;\lambda)}$ are dominated by $a_1 + a_2|x|^{c_1}$, where a_1, a_2, and c_1 are nonnegative constants and $\int |x|^{c_1} f_1(x)\, dx < \infty$.

Proof of Lemma 1. First consider the covariance matrix of the score. For simplicity, denote $e_t = f'\left(\sigma_0^{-1}\epsilon_t;\lambda_0\right) / \left[f\left(\sigma_0^{-1}\epsilon_t;\lambda_0\right)\sigma_0\right] = f'_{\sigma_0}\left(\sigma_0^{-1}\epsilon_t;\lambda_0\right) / f_{\sigma_0}\left(\sigma_0^{-1}\epsilon_t;\lambda_0\right)$ and notice that

$$\begin{aligned} E(e_t^2) &= E\left[\left(f'_{\sigma_0}(\epsilon_t;\lambda_0)/f_{\sigma_0}(\epsilon_t;\lambda_0)\right)^2\right] \\ &= \sigma_0^{-2}\int \left(f'(x;\lambda_0)^2/f(x;\lambda_0)\right) dx \\ &= \sigma_0^{-2}\mathcal{J}, \end{aligned}$$

where the second equality is based on the fact that $f_{\sigma_0}(x;\lambda_0) = \sigma_0^{-1} f\left(\sigma_0^{-1}x;\lambda_0\right)$ is the density function of ϵ_t (cf. equation (2.13) of Breidt et al. (1991)). Thus, because e_t and U_{t-1} are independent and $\Gamma_{U^*} = \sigma_0^{-2} Cov\left(U_{t-1}\right)$,

$$\begin{aligned} Cov\left(\frac{\partial}{\partial\phi} g_t\left(\theta_0\right)\right) &= Cov\left(-U_{t-1}e_t\right) \\ &= E(e_t^2) Cov\left(U_{t-1}\right) \\ &= \mathcal{J}\Gamma_{U^*}. \end{aligned}$$

Because the sequence $U_{t-1}e_t$ is uncorrelated we have

$$\lim_{T\to\infty} (T-p)^{-1} Cov\left(\sum_{t=r+1}^{T-s} \frac{\partial}{\partial\phi} g_t\left(\theta_0\right), \sum_{t=r+1}^{T-s} \frac{\partial}{\partial\phi} g_t\left(\theta_0\right)\right) = \mathcal{J}\Gamma_{U^*}.$$

Similarly, the independence of e_t and V_{t+1} and the equality $\Gamma_{V^*} = \sigma_0^{-2} Cov\left(V_{t+1}\right)$ give

$$Cov\left(\frac{\partial}{\partial\varphi} g_t\left(\theta_0\right)\right) = \mathcal{J}\Gamma_{V^*}$$

and, by the uncorrelatedness of the sequence $V_{t+1}e_t$,

$$\lim_{T\to\infty} (T-p)^{-1} Cov\left(\sum_{t=r+1}^{T-s} \frac{\partial}{\partial\varphi} g_t\left(\theta_0\right), \sum_{t=r+1}^{T-s} \frac{\partial}{\partial\varphi} g_t\left(\theta_0\right)\right) = \mathcal{J}\Gamma_{V^*}.$$

As for the covariance matrix between $\partial g_t\left(\theta_0\right)/\partial\phi$ and $\partial g_t\left(\theta_0\right)/\partial\varphi$, first consider

$$\begin{aligned} Cov\left(-u_{t-i}e_t, -v_{k+j}e_k\right) &= \sum_{a=0}^{\infty}\sum_{b=0}^{\infty} \alpha_{0a}\beta_{0b} Cov\left(\epsilon_{t-i-a}e_t, \epsilon_{k+j+b}e_k\right) \\ &= \begin{cases} \alpha_{0,t-k-i}\beta_{0,t-k-j}, & t > k,\ 1 \le i \le r,\ 1 \le j \le s \\ 0, & t \le k,\ 1 \le i \le r,\ 1 \le j \le s \end{cases}, \end{aligned}$$

where the first equality follows from (3) and (4) and the second one is based on condition (A2) (see also Breidt et al. (1991, p. 181)). Hence, as in Breidt et al. (1991, p. 182), the element in position (i,j) of the matrix $(T-p)^{-1} Cov\left(\partial l_T\left(\theta_0\right)/\partial\phi, \partial l_T\left(\theta_0\right)/\partial\varphi\right)$ is

$$\begin{aligned} (T-p)^{-1} \sum_{k=r+1}^{T-s-1}\sum_{t=k+1}^{T-s} \alpha_{0,t-k-i}\beta_{0,t-k-j} &= (T-p)^{-1} \sum_{k=r+1}^{T-s-1}\sum_{t=0}^{T-s-k-i} \alpha_{0t}\beta_{0,t+i-j} \\ &\to \sum_{k=0}^{\infty} \alpha_{0k}\beta_{0,k+i-j}, \end{aligned}$$

where $\beta_{0l} = 0$ for $l < 0$. Note that the limit equals $\psi_{0,j-i}$, as can be easily checked.

Next recall that $u_t^* = \sum_{k=0}^{\infty} \alpha_{0k}\eta_{t-k}$ and $v_t^* = \sum_{l=0}^{\infty} \beta_{0l}\eta_{t-l}$ with $\eta_t \sim i.i.d.\,(0,1)$. Thus,

$$\begin{aligned} Cov\left(u_{t-i}^*, v_{t-j}^*\right) &= \sum_{k=0}^{\infty} \alpha_{0k} \sum_{l=0}^{\infty} \beta_{0l} E(\eta_{t-i-k}\eta_{t-j-l}) \\ &= \sum_{k=0}^{\infty} \alpha_{0k}\beta_{0,k+i-j,} \end{aligned}$$

and we can conclude that

$$\lim_{T\to\infty} (T-p)^{-1} Cov\left(\sum_{t=r+1}^{T-s} \frac{\partial}{\partial\phi} g_t(\theta_0), \sum_{t=r+1}^{T-s} \frac{\partial}{\partial\varphi} g_t(\theta_0)\right) = \Gamma_{U^*V^*}.$$

We have thus shown that the covariance matrix of the score of (ϕ, φ) evaluated at the true parameter value and divided by $(T-p)$ converges to Σ.

The score of (σ, λ) is $i.i.d.$ and, by condition (A7), has zero mean and finite second moments. The definitions show that its covariance matrix equals that of the score of the parameter (α_{p+1}, θ) in Andrews et al. (2006). Thus, if $\theta_2 = (\sigma, \lambda)$

$$(T-p)^{-1} Cov\left(\sum_{t=r+1}^{T-s} \frac{\partial}{\partial\theta_2} g_t(\theta_0), \sum_{t=r+1}^{T-s} \frac{\partial}{\partial\theta_2} g_t(\theta_0)\right) = \Omega.$$

Using the definitions it is also straightforward to check that, at true parameter values, the scores of (ϕ, φ) and (σ, λ) are uncorrelated so that we can conclude that

$$\lim_{T\to\infty} (T-p)^{-1} Cov\left(\sum_{t=r+1}^{T-s} \frac{\partial}{\partial\theta} g_t(\theta_0), \sum_{t=r+1}^{T-s} \frac{\partial}{\partial\theta} g_t(\theta_0)\right) = diag(\Sigma, \Omega).$$

The matrix Ω is positive definite by the assumed condition (A6). Because $\mathcal{J} > 1$ (see condition (A5)) the positive definiteness of Σ can be established in the same way as Proposition 1 of Breidt et al. (1991).

The asymptotic normality can be proved in the same way as Proposition 2 of Breidt et al. (1991) by approximating the processes U_{t-1} and V_{t+1} by long moving averages and using a standard central limit theorem for finitely dependent stationary processes.

Proof of Theorem 2. We shall first present the second partial derivatives of the function $g_t(\theta)$. To simplify notation, we write $\tilde{u}_t = u_t(\varphi)$ and $\tilde{v}_t = v_t(\phi)$ and, furthermore, $\tilde{U}_{t-1} = (\tilde{u}_{t-1}, ..., \tilde{u}_{t-r})$ and $\tilde{V}_{t+1} = (\tilde{v}_{t+1}, ..., \tilde{v}_{t+s})$. Similarly, $\tilde{\epsilon}_t = \tilde{u}_t - \phi_1\tilde{u}_{t-1} - \cdots - \phi_r\tilde{u}_{t-r} = \tilde{v}_t - \varphi_1\tilde{v}_{t+1} - \cdots - \varphi_s\tilde{v}_{t+s}$ will signify ϵ_t evaluated at an arbitrary point in the permissible parameter space, not the true parameter value. We also set $h(x;\lambda) = f'(x;\lambda)/f(x;\lambda)$, so that

$$h'(x;\lambda) = \frac{f''(x;\lambda)}{f(x;\lambda)} - \left(\frac{f'(x;\lambda)}{f(x;\lambda)}\right)^2,$$

and let Y_t stand for the $r \times s$ matrix with elements y_{t-i+j} $(i = 1, ..., r,\ j = 1, ..., s)$. By straightforward differentiation (cf. Breidt et al. (1991), p. 187),

$$\begin{aligned}
\partial^2 g_t(\theta)/\partial\phi\partial\phi' &= \sigma^{-2}h'\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)\tilde{U}_{t-1}\tilde{U}'_{t-1} \\
\partial^2 g_t(\theta)/\partial\varphi\partial\varphi' &= \sigma^{-2}h'\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)\tilde{V}_{t+1}\tilde{V}'_{t+1} \\
\partial^2 g_t(\theta)/\partial\sigma^2 &= 2\sigma^{-3}h\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)\tilde{\epsilon}_t + \sigma^{-4}h'\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)\tilde{\epsilon}_t^2 + \sigma^{-2} \\
\partial^2 g_t(\theta)/\partial\lambda\partial\lambda' &= \frac{1}{f\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)}\partial^2 f\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)/\partial\lambda\partial\lambda' \\
&\quad -\frac{1}{f^2\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)}\left(\partial f\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)/\partial\lambda\right)\left(\partial f\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)/\partial\lambda\right)' \\
\partial^2 g_t(\theta)/\partial\phi\partial\varphi' &= \sigma^{-2}h'\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)\tilde{U}_{t-1}\tilde{V}'_{t+1} + \sigma^{-1}h\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)Y_t \\
\partial^2 g_t(\theta)/\partial\phi\partial\sigma &= \sigma^{-3}h'\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)\tilde{\epsilon}_t\tilde{U}_{t-1} + \sigma^{-2}h\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)\tilde{U}_{t-1} \\
\partial^2 g_t(\theta)/\partial\phi\partial\lambda' &= -\sigma^{-1}\tilde{U}_{t-1}\partial h\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)/\partial\lambda' \\
\partial^2 g_t(\theta)/\partial\varphi\partial\sigma &= \sigma^{-3}h'\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)\tilde{\epsilon}_t\tilde{V}_{t+1} + \sigma^{-2}h\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)\tilde{V}_{t+1} \\
\partial^2 g_t(\theta)/\partial\varphi\partial\lambda' &= -\sigma^{-1}\tilde{V}_{t+1}\partial h\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)/\partial\lambda' \\
\partial^2 g_t(\theta)/\partial\sigma\partial\lambda' &= -\sigma^{-2}\tilde{\epsilon}_t\partial h\left(\sigma^{-1}\tilde{\epsilon}_t;\lambda\right)/\partial\lambda'.
\end{aligned}$$

Using conditions (A2)–(A4) and calculations similar to those in Breidt et al. (1991, p. 181) it is not difficult to check that $E\left[\partial^2 g_t(\theta_0)/\partial\theta\partial\theta'\right] = -diag(\Sigma, \Omega)$.

As in Andrews et al. (2006), we now use the Taylor series expansion

$$\sum_{t=r+1}^{T-s}\left[g_t\left(\theta_0+T^{-1/2}c\right)-g_t\left(\theta_0\right)\right]=T^{-1/2}\sum_{t=r+1}^{T-s}c'\frac{\partial g_t\left(\theta_0\right)}{\partial\theta}+\frac{1}{2}T^{-1}\sum_{t=r+1}^{T-s}c'\frac{\partial^2 g_t\left(\theta_0\right)}{\partial\theta\partial\theta'}c$$
$$+\frac{1}{2}T^{-1}\sum_{t=r+1}^{T-s}c'\left(\frac{\partial^2 g_t\left(\theta_T^*\left(c\right)\right)}{\partial\theta\partial\theta'}-\frac{\partial^2 g_t\left(\theta_0\right)}{\partial\theta\partial\theta'}\right)c,$$

where $c\in\mathbb{R}^{r+s+1+d}$ and the argument $\theta_T^*\left(c\right)$ in the matrix of second partial derivatives means that each row is evaluated at an intermediate point lying between θ_0 and $T^{-1/2}c$. Thus, if $\|\cdot\|$ signifies the Euclidean norm we have $\sup_{c\in K}\|\theta_T^*\left(c\right)-\theta_0\|\to 0$ for any compact set $K\subset\mathbb{R}^{r+s+1+d}$. Moreover, using the dominance conditions in (A7) and arguments similar to those in Breidt et al. (1991, p. 186-190) it can be shown that a uniform law of large numbers for stationary ergodic processes applies to $\partial^2 g_t(\theta)/\partial\theta\partial\theta'$ over any small enough compact neighborhood θ_0 (see Theorem A.2.2 in White (1994)). Thus, we can conclude that

$$T^{-1}\sum_{t=r+1}^{T-s}c'\left(\frac{\partial^2 g_t\left(\theta_T^*\left(c\right)\right)}{\partial\theta\partial\theta'}-\frac{\partial^2 g_t\left(\theta_0\right)}{\partial\theta\partial\theta'}\right)c\xrightarrow{p}0$$

for c belonging to any compact subset of $\mathbb{R}^{r+s+1+d}$. The proof can now be completed in the same way as the proof of Theorem 1 of Andrews et al. (2006).

Finally, note that the convergence (12) is an immediate consequence of the consistency of the estimator $\hat{\theta}$ obtained from Theorem 2 and the aforementioned fact that $\partial^2 g_t(\theta)/\partial\theta\partial\theta'$ obeys a uniform law of large numbers.

References

Andrews, B., M. Calder, and R.A. Davis (2009). Maximum likelihood estimation for α-stable autoregressive proceses. Annals of Statistics 37, 1946–1982.

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.

Andrews, D. and W. Chen (1994). Approximately median-unbiased estimation of autoregressive models. Journal of Business and Economic Statistics 12, 187–204.

Breidt, 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, 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.

Brockwell, P.J. and R.A. Davis (1987). Time Series: Theory and Methods. Springer-Verlag. New York.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay (1997). Econometrics of Financial Markets. Princeton University Press. Princeton.

Canova, F. (2007). Methods for Applied Macroeconomic Research. Princeton University Press. Princeton.

Diebold, F.X., and R.S. Mariano (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics 13, 134–144.

Cecchetti, S.G. and G. Debelle (2006). Has the inflation process changed? Economic Policy, April 2006, 311–352.

Gali, J., and M. Gertler (1999). Inflation dynamics: A structural econometric analysis. Journal of Monetary Economics 44, 195–222.

Hansen, L.P., and T.J. Sargent (1991). Two Difficulties in Interpreting Vector Autoregressions. In Rational Expectations Econometrics, ed. by L.P. Hansen, and T.J. Sargent, Westview Press, Inc., Boulder, CO, 77–119.

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

Kasa, K., T.B. Walker, and C.H. Whiteman (2007). Asset Prices in a Time Series Model with Perpetually Disparately Informed, Competitive Traders. Unpublished manuscript, Department of Economics, Simon Fraser University.

Lanne, M., and P. Saikkonen (2010). Noncausal vector autoregression. HECER Discussion Paper 293.

Lanne, M., J. Luoto, and P. Saikkonen (2010). Optimal forecasting of noncausal autoregressive time series. HECER Discussion Paper 286.

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.

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

White, H. (1994). Estimation, Inference and Specification Analysis. Cambridge University Press. New York.

Wu, R. and R. A. Davis (2010). Least absolute deviation estimation for general autoregressive moving average time-series models. Journal of Time Series Analysis 31, 98–112.

Figure 1: Rejection rates of the 5%-level LR test of $H_0 : \phi_1 = 0.9$ for $T = 100$ (solid line), $T = 200$ (long dashes) and $T = 500$ (dashes). The data are generated from a model with $\phi_1 = 0.9 - c/\sqrt{T}$.

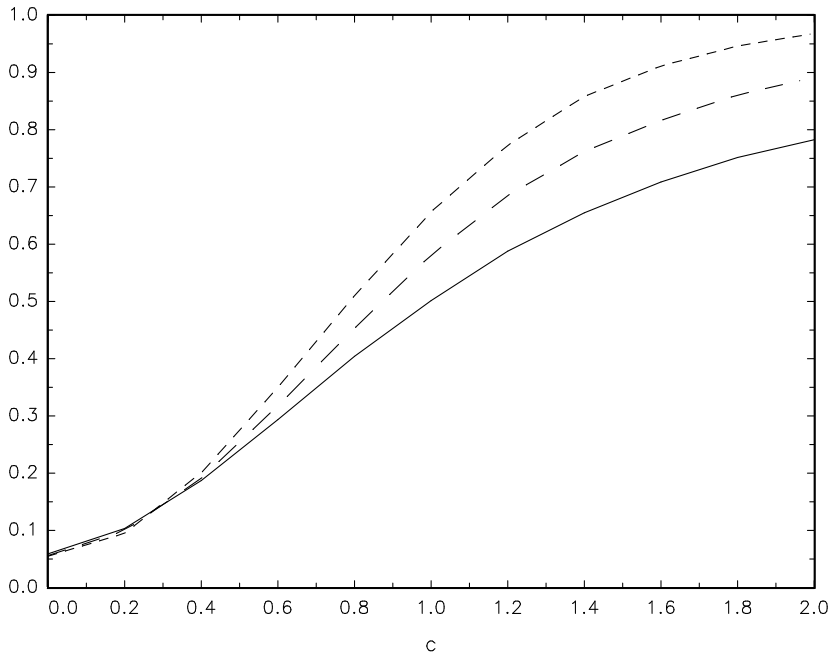

Figure 2: Autocorrelation function of the U.S. inflation. The dashed line depicts the upper bound of the two standard deviation band.

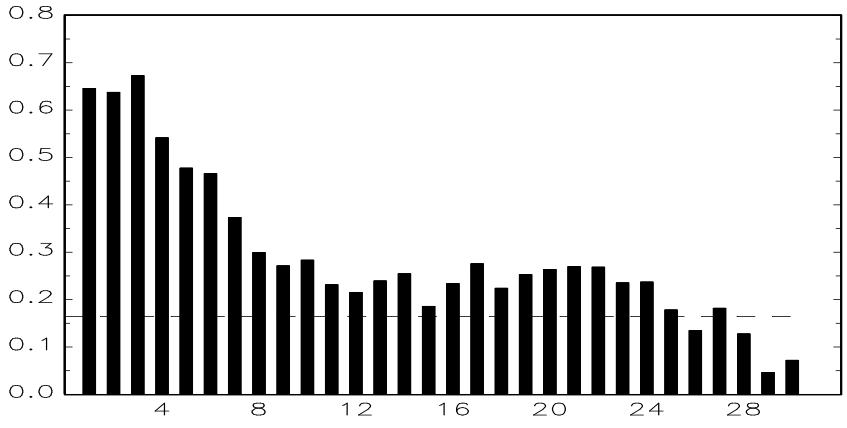

Figure 3: Quantile-quantile plots of the residuals of the AR(3,0)-N and AR(0,3)-t models for the U.S. inflation.

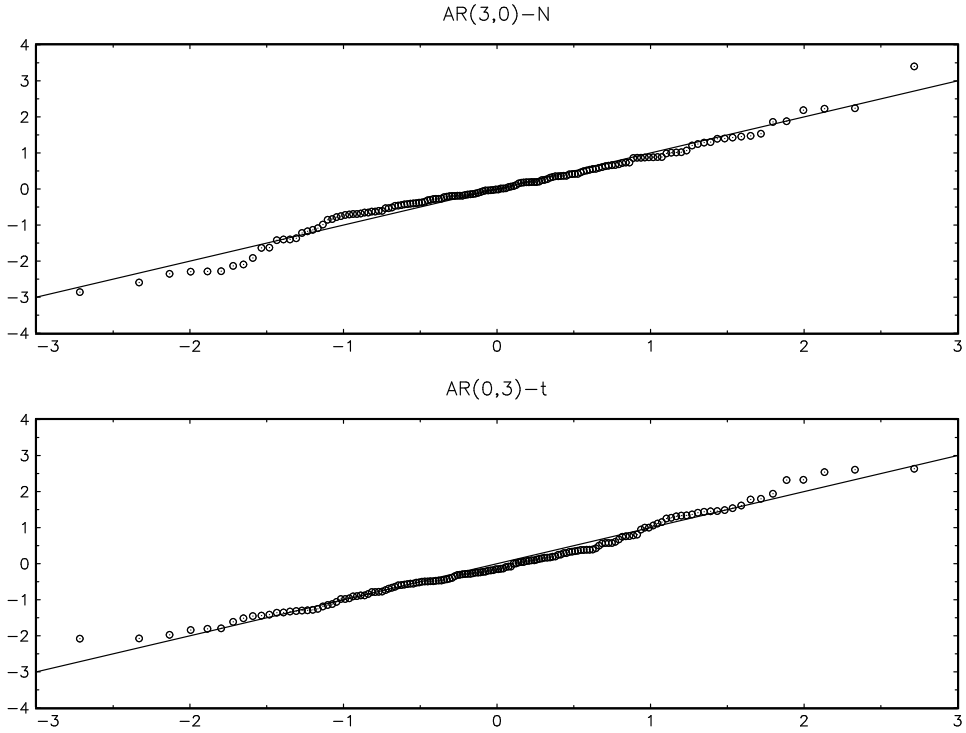

Table 1: Finite-sample properties of the ML estimator.

		DGP					
		$\phi_1 = 0.9, \varphi_1 = 0.9$		$\phi_1 = 0.9, \varphi_1 = 0.1$		$\phi_1 = 0.1, \varphi_1 = 0.9$	
T	Parameter	Mean	St.dev.	Mean	St.dev.	Mean	St.dev.
100	ϕ_1	0.882	0.048	0.871	0.063	0.108	0.097
	φ_1	0.874	0.051	0.107	0.095	0.869	0.064
200	ϕ_1	0.892	0.032	0.888	0.030	0.101	0.058
	φ_1	0.888	0.033	0.102	0.058	0.888	0.029
500	ϕ_1	0.897	0.019	0.896	0.016	0.100	0.035
	φ_1	0.896	0.019	0.100	0.035	0.896	0.016

The DGP is the AR(1,1) model where the error term follows the t-distribution with 3 degrees of freedom and $\sigma = 0.1$. The results are based on 10,000 realizations.

Table 2: Rejection rates of the Wald and likelihood ratio (LR) tests.

		DGP					
		$\phi_1 = 0.9, \varphi_1 = 0.9$		$\phi_1 = 0.9, \varphi_1 = 0.1$		$\phi_1 = 0.1, \varphi_1 = 0.9$	
T	Parameter	Wald test	LR test	Wald test	LR test	Wald test	LR test
100	ϕ_1	0.081	0.059	0.060	0.075	0.066	0.061
	φ_1	0.091	0.075	0.061	0.055	0.067	0.083
200	ϕ_1	0.074	0.056	0.059	0.062	0.060	0.054
	φ_1	0.075	0.061	0.057	0.052	0.057	0.062
500	ϕ_1	0.063	0.055	0.054	0.055	0.056	0.051
	φ_1	0.064	0.058	0.057	0.053	0.055	0.056

See notes to Table 1. The figures are rejection rates of Wald and LR tests of the null hypothesis that the parameter equals the true value. The nominal size of the tests is 5%.

Table 3: Simulation results on model selection by maximizing the likelihood function.

	DGP								
T	$\phi_1 = 0.9, \varphi_1 = 0.9$			$\phi_1 = 0.9, \varphi_1 = 0.1$			$\phi_1 = 0.1, \varphi_1 = 0.9$		
	AR(2,0)	AR(1,1)	AR(0,2)	AR(2,0)	AR(1,1)	AR(0,2)	AR(2,0)	AR(1,1)	AR(0,2)
100	1158	8077	765	3684	5472	844	990	5402	3608
200	332	9463	205	2975	6806	219	267	6699	3034
500	4	9991	5	1495	8501	4	4	8538	1458

See notes to Table 1. Each figure indicates the number of times the model in question maximizes the likelihood function out of 10,000 realizations.

Table 4: Simulation results on model selection by maximizing the likelihood function.

	DGP								
T	$\phi_1 = 0.9, \varphi_1 = 0.9$			$\phi_1 = 0.9, \varphi_1 = 0.1$			$\phi_1 = 0.1, \varphi_1 = 0.9$		
	AR(2,0)	AR(1,1)	AR(0,2)	AR(2,0)	AR(1,1)	AR(0,2)	AR(2,0)	AR(1,1)	AR(0,2)
100	765	8077	1158	844	5472	3684	3608	5402	990
200	205	9463	332	219	6806	2975	3034	6699	267
500	5	9991	4	4	8501	1495	1458	8538	4

See notes to Table 1. Each figure indicates the number of times the model in question maximizes the likelihood function out of 10,000 realizations.

Table 5: Estimation results of the autoregressive models for the demeaned U.S. inflation.

	Model				
	AR(3,0)-N	AR(3,0)-t	AR(2,1)-t	AR(1,2)-t	AR(0,3)-t
ϕ_1	0.270	0.294	0.674	0.947	
	(0.077)	(0.076)	(0.116)	(0.033)	
ϕ_2	0.240	0.272	0.204		
	(0.079)	(0.076)	(0.104)		
ϕ_3	0.347	0.329			
	(0.080)	(0.070)			
φ_1			–0.321	–0.643	0.259
			(0.109)	(0.076)	(0.081)
φ_2				–0.321	0.291
				(0.068)	(0.066)
φ_3					0.282
					(0.066)
σ		2.235	2.301	2.192	2.301
		(0.259)	(0.252)	(0.243)	(0.390)
λ		4.472	4.679	4.605	3.451
		(1.923)	(2.056)	(1.962)	(1.162)
Log-likelihood	–332.075	–328.431	–333.888	–326.208	–323.075
Ljung-Box (4)	0.243	0.342	< 0.001	0.195	0.073
McLeod-Li (4)	0.003	0.056	0.019	0.476	0.246

AR(r, s) denotes the autoregressive model with the rth and sth order polynomials $\phi(B)$ and $\varphi(B^{-1})$, respectively. N and t refer to Gaussian and t-distributed errors, respectively. The figures in parentheses are standard errors. Marginal significance levels of the Ljung-Box and McLeod-Li tests with 4 lags are reported.

Table 6: Root mean square errors of AR(3,0)-N, AR(3,0)-t and AR(0,3)-t models for the U.S. inflation.

	Forecast horizon (quarters)			
Model	$h = 1$	$h = 2$	$h = 3$	$h = 4$
AR(3,0)-N	2.121	1.978	2.093	2.270
AR(3,0)-t	2.108 (0.17)	1.973 (0.39)	2.088 (0.52)	2.263 (0.50)
AR(0,3)-t	2.052 (0.24)	1.979 (0.98)	2.042 (0.01)	2.183 (0.01)

The figures are based on an expansive estimation window. The first forecasts are based on the estimation period ending in 1981:4. The figures in parentheses are p-values of the Diebodl-Mariano test of equal predictive accuracy compared to the AR(3,0)-N model.