Bayesian Analysis of Long Memory and Persistence using ARFIMA Models

Gary Koop
Dept. of Economics, University of Toronto
Eduardo Ley
Resources for the Future, Washington DC
Jacek Osiewalski
Dept. of Econometrics, Academy of Economics, Kraków
Mark F.J. Steel
CentER and Dept. of Econometrics, Tilburg University

May 1993
Version: May 24, 1995

Abstract. This paper provides a Bayesian analysis of Autoregressive Fractionally Integrated Moving Average (ARFIMA) models. We discuss in detail inference on impulse responses, and show how Bayesian methods can be used to (i) test ARFIMA models against ARIMA alternatives, and (ii) take model uncertainty into account when making inferences on quantities of interest. Our methods are then used to investigate the persistence properties of real U.S. GNP.

Keywords. Fractionally Integrated Models, Impulse Responses, Time Series, Trend Stationarity, Unit Root

Address. Eduardo Ley, Resources for the Future, 1616 P St NW, Washington DC 20036-1400. Email: G. Koop: gkoop@epas.utoronto.ca; E. Ley: ley@rff.org; J. Osiewalski: eeosiewa@cyf-kr.edu.pl; M.J.F. Steel: steel@kub.nl.

Bayesian Analysis of Long Memory and Persistence using ARFIMA Models

Gary Koop
Dept. of Economics, University of Toronto
Eduardo Ley
Resources for the Future, Washington DC
Jacek Osiewalski
Dept. of Econometrics, Academy of Economics, Kraków
Mark F.J. Steel
CentER and Dept. of Econometrics, Tilburg University

1. Introduction

Over the last decade, there has been an increasing interest in investigating the degree of integration of macroeconomic time series, as well as in measuring the persistence of shocks. Much work has been done within the class of ARIMA models (see, for example, almost all of the unit root literature). In recent years several econometricians have argued that ARIMA models are too restrictive. For example, Sowell (1992b) claims that ARIMA models tend to fit mainly the short-run properties of the data and hence can provide misleading estimates of long-run properties. Autoregressive Fractionally Integrated Moving Average (ARFIMA) models provide an alternative to ARIMA models. They allow for series to exhibit stationary ARMA behaviour after being fractionally differenced. Granger and Joyeux (1980) and Hosking (1981) proposed the use of ARFIMA processes to model long memory. Some theoretical properties of these stochastic processes can also be found in Beran (1994), Brockwell and Davis (1991) and Odaki (1993). In an applied econometrics context, Sowell (1992b) describes how the ARMA component could pick up short-run, and the fractionally differenced component long-run behaviour.

A (stationary and invertible) ARMA(p, q) process is formally a special case

We would like to thank Fallaw Sowell for providing FORTRAN code for evaluating the likelihood function for the ARFIMA model, and Arnold Zellner, Dale Poirier and three anonymous referees for their useful comments. E. Ley and M.F.J. Steel were affiliated to the Universidad Carlos III de Madrid during the early work on this paper. J. Osiewalski was visiting CORE, Université Catholique de Louvain, and was partly supported by the Belgian Programme on Interuniversity Poles of Attraction during the preparation of the present version.

of a (stationary and invertible) ARFIMA(p, δ, q) process, corresponding to the value $\delta = 0$ of the fractional differencing parameter $\delta \in (-1, 0.5)$. However, the memory properties under $\delta = 0$ and $\delta \neq 0$ are so very much different that we consider these two cases to define separate (competing) model classes. The autocorrelation function of an ARFIMA process can be shown to decay at a hyperbolic rate for non-zero δ, which is much slower than the usual geometric rate associated with stationary ARMA processes. Typically, processes with $\delta > 0$ are called long memory processes (autocorrelations are not summable), whereas negative values of δ lead to so-called intermediate memory (summable autocorrelations).

Long memory may be perceived as an intrinsic characteristic of some economic phenomena, but even if it is not, it may still appear on a macro level due to aggregation. Assuming a Beta distribution for the squares of AR coefficients, Granger (1980) showed that a sum of a large number of stationary AR(1) processes with random parameters can possess long memory. Thus, when analyzing aggregated data, we should keep the possibility of long memory open.

There has been a growing use of ARFIMA models by empirical researchers (see, among many others, Baillie, Chung and Tieslau (1992), Diebold and Rudebusch (1989, 1991), Cheung (1993), Cheung and Lai (1993)). Virtually all of this work has been carried out using non-Bayesian statistical techniques. Exceptions are Koop (1991) and Carlin and Dempster (1989). However, the former of these papers uses a very simple model while the second carries out a conditional Bayesian analysis (*i.e.*, the analysis proceeds conditionally on fitted parameter values). The most commonly used techniques break down into three categories:

(i) Maximum likelihood (Sowell (1992a));

(ii) Approximate maximum likelihood (Baillie and Chung (1992), Li and McLeod (1986) or Fox and Taqqu (1986)); and

(iii) Two-step procedures (Geweke and Porter-Hudak (1983) or Janacek (1982)).

For asymptotic sampling-theory properties of exact and approximate maximum likelihood estimators see Dahlhaus (1989) and Beran (1994).

Exact maximum likelihood techniques have been criticized as being too computationally demanding, while the other methods have been criticized as being inaccurate for finite samples (see Sowell (1992a,b)). Recent simulations in Beveridge and Oickle (1993) indicate that the Li and McLeod (1986) approximation is hazardous in the presence of short memory (ARMA) parameters. There has been some work on method of moments estimators (see Tieslau, Schmidt and Baillie (1992)), but these have not been widely used in practice yet. In addition, detection of long-memory properties through the sample autocorrelations seems almost impossible for positive values of the fractional differencing parameter in view of the results in Newbold and Agiakloglou (1993). A recent Monte Carlo study by Cheung and Diebold (1994) finds that the superiority of the exact maximum likelihood approach over the Whittle approximation (advocated by Fox and Taqqu (1986)) is less prominent when the mean is unknown than in the case with known mean; however, they conclude that Sowell's exact maximum likelihood technique still seems preferable, especially in small and medium sample sizes.

In this paper, a Bayesian analysis of ARFIMA models is presented which is roughly as computationally demanding as maximum likelihood. It is argued that the Bayesian approach has several advantages over classical techniques in that:

(i) It provides exact finite sample distributions for any feature of interest (*e.g.*, an impulse response or the fractional differencing parameter).

(ii) Instead of presenting just a point estimate and standard error associated with, say, an impulse response, we can plot the whole density of that quantity of interest. Our empirical results indicate that this is important since impulse responses can be multimodal, highly skewed and fat-tailed. Furthermore, moments may not exist (see Koop, Osiewalski and Steel (1994)).

(iii) By attaching prior mass to $\delta = 0$, we put ARIMA and ARFIMA specifications on equal footing and we can perform small sample tests of memory properties.

(iv) It allows us to average across models rather than choosing just one model (*e.g.*, for predictive purposes). In particular, we find that ARIMA models may be favoured by the data. Since ARFIMA and ARIMA models have very different consequences for inference on persistence, it is instrumental to formally take into account this model uncertainty.

The paper is organized as follows: The second section describes ARFIMA models and discusses their properties, with particular emphasis on impulse response functions. The third section introduces a Bayesian approach. In particular, we derive the posterior density for the parameters and discuss computational methods for calculating posterior properties about the functions of the parameters such as impulse responses. In addition, we briefly describe some results on post-sample prediction. The fourth section applies the theory to an important empirical problem, viz. the measurement of persistence in real U.S. GNP. Section 5 concludes.

2. ARFIMA Models

Assume that $y_0, \ldots, y_T$ are observations on a time series $\{y_t\}$ and let

$$z_t = (1-L)(y_t - \mu t - \alpha) = \Delta y_t - \mu.$$

The model considered[1] is:

$$\varphi(L)(1-L)^{\delta} z_t = \vartheta(L)\varepsilon_t \tag{1}$$

where $\vartheta(L) = (1 + \theta_1 L + \ldots + \theta_q L^q)$ and $\varphi(L) = 1 + \phi_1 L + \ldots + \phi_p L^p$ are polynomials in the lag operator whose roots are restricted to lie outside the unit circle. The corresponding regions for the parameters of $\vartheta(L)$ and $\varphi(L)$ will be denoted by C_q and C_p, respectively; the ε_t's are i.i.d. $N(0,\sigma^2)$; and $\delta \in (-1, 0.5)$. The fractional differencing operator, $(1-L)^{\delta}$, is defined as

$$(1-L)^{\delta} = \sum_{j=0}^{\infty} c_j(\delta) L^j, \tag{2}$$

where $c_0(\cdot) = 1$, and for $j > 0$

$$c_j(a) = \prod_{k=1}^{j} \left(1 - \frac{1+a}{k}\right).$$

In this formulation, which also appeared in, *e.g.*, Brockwell and Davis (1991, Ch.13) or Odaki (1993), it can be clearly seen that the infinite past of the z_t's is taken into account. We will refer to this model for z_t as the ARFIMA(p, δ, q). Using Granger and Andersen's (1978) general definition of invertibility, Odaki

[1] In this paper we consider only linear univariate models. In recent years there has been some criticism of such models by researchers who advocate going to nonlinear or multivariate frameworks. For example, there has been a growing interest in fractional cointegration (see Cheung and Lai (1993)). We do not deny that there may be great value in embedding fractional ideas in multivariate or nonlinear models, however we feel it is important to develop basic Bayesian tools for univariate models before proceeding to more complicated models.

(1993) proved that z_t is invertible whenever $\delta > -1$. The special case $\delta = 0$ corresponds to the standard ARIMA$(p, 1, q)$ model for y_t used in Campbell and Mankiw (1987). Note that the restriction $\delta < 0.5$ implies that Δy_t is stationary, which is reasonable for many economic time series, like real U.S. GNP or disposable income. However, we could allow for trend-stationarity of y_t by defining

$$v_t = y_t - \mu t - \alpha,$$

and then using an ARFIMA model for v_t. Values of the fractional differencing parameter $d = 1+\delta$ for v_t in $(-1, 0.5)$ correspond to values of δ in $(-2, -0.5)$ for z_t. As long as we assume $d > 0$, *i.e.*, $\delta > -1$, we can work with the differenced series, Δy_t, without violating invertibility of z_t. Note that this enables us to leave the possibility of trend-stationarity for y_t open, without explicitly treating α, albeit at the cost of imposing long memory on v_t. This puts the trend-stationary $(d = 0)$ versus unit root $(d = 1)$ debate in a new light, as it alleviates the need to choose one of these very special cases, corresponding to integer degrees of differencing in an ARIMA context. As long as $d \in (0, 0.5)$, y_t will be trend-stationary (with long memory) and for $d \in (0.5, 1.5)$ the differenced series, Δy_t, will be stationary, with intermediate memory for $d < 1$ and long memory for $d > 1$. In our framework, we shall test for trend-stationarity of y_t through the posterior probability that $-1 < \delta < -0.5$. By assuming long memory for v_t, we avoid the issue of having to introduce α explicitly under one of the contending hypotheses [see, *e.g.*, Schotman and van Dijk (1991)]. Of course, our interest is in properties of the level of the series, but for prediction of y_{T+n} given the sample, and posterior inference on impulse responses, α does not intervene, as will be shown in the sequel.

An impulse response function, $I(n)$, can be thought of as the effect of a shock of size one at time t on y_{t+n}. Impulse responses for a stationary process are the coefficients of its (infinite) moving average representation. In our case, z_t is a stationary process with impulse responses given by the coefficients of an infinite order lag polynomial $A(L) = (1-L)^{-\delta}\varphi^{-1}(L)\vartheta(L)$. The nth-order partial sums of these coefficients, *i.e.*, the cumulative impulse responses for z_t, are the impulse responses $I(n)$ for the level processes v_t and y_t. Equivalently, $I(n)$ can be represented as the nth coefficient of $A^*(L) = (1-L)^{-1}A(L) = (1-L)^{-d}\varphi^{-1}(L)\vartheta(L)$.

The coefficients of $\varphi^{-1}(L)\vartheta(L)$, *i.e.*, the standard ARMA(p, q) impulse responses, can be presented as

$$J(i) = \sum_{j=0}^{q} \theta_j f_{i+1-j},$$

with $\theta_0 = 1$, $f_h = 0$ for $h \leq 0$, $f_1 = 1$ and

$$f_h = -(\phi_1 f_{h-1} + \ldots + \phi_p f_{h-p}), \qquad \text{for} \quad h \geq 2.$$

Therefore, the coefficients of $A^*(L)$ can easily be seen to take the form

$$I(n) = \sum_{i=0}^{n} c_i(-d) J(n-i), \tag{3}$$

which is particularly convenient for computations. When $d = 1$ (*i.e.*, $\delta = 0$), then $c_i = 1$ $(i \geq 0)$ and (3) reduces to the formula presented in Koop, Osiewalski and Steel (1994) for the ARIMA$(p, 1, q)$ case. Also note that for the limiting case $d = 0$ (*i.e.*, $\delta = -1$, which we have excluded in this paper), $c_i = 0$ $(i \geq 1)$ and equation (3) reduces to $I(n) = J(n)$, where $J(n)$ is the impulse response for the case of ARMA(p, q) deviations from a linear trend for y_t.

Note that the limit behaviour of $I(n)$ as n tends to ∞ corresponds to the properties of the infinite sum $A(1)$, which is zero if $\delta < 0$, is equal to $\vartheta(1)/\varphi(1)$ for $\delta = 0$, and is ∞ for $\delta > 0$. Based on this behaviour of $I(\infty)$, Hauser, Pötscher and Reschenhofer (1992) have criticized the use of ARFIMA models for the measurement of persistence. They argue that "the estimated persistence [$I(\infty)$] obtained from an estimated ARFIMA model for Δy_t will necessarily be 0 or ∞ whenever the estimated differencing parameter is different from zero. This will, however, almost exclusively be the case, since finding the estimator [of δ] to be exactly equal to zero is extremely unlikely" (page 8). On this basis, they conclude "fractionally integrated ARMA models are inappropriate for the purpose of estimating persistence" (Abstract).

We offer two responses to this criticism, one general and one Bayesian. The general response is that the behaviour of our model for impulse responses at ∞ is of little relevance. Most economic policy questions focus on the effect of shocks at much shorter horizons. In this paper, although results for $I(\infty)$ are discussed, more attention is given to $I(4)$, $I(12)$ and $I(40)$. Since quarterly data are used, we interpret $I(4)$, $I(12)$ and $I(40)$ as "short-run", "medium-run" and "long-run" effects of a shock, respectively.

Secondly, although we agree with Hauser, Pötscher and Reschenhofer (1992) that the ultimate impact of a shock being either 0 or ∞ is a theoretical weakness of ARFIMA specifications for Δy_t, we see no reason to dismiss them altogether on that basis before seeing the data. We cannot claim that ARMA models are more appropriate for estimating persistence, unless they are favoured by the data evidence. Unlike classical methods, our Bayesian method allows for non-degenerate distributions for $I(\infty)$ even if we take ARFIMA models into account. By putting some prior mass at $\delta = 0$, we allow for finite positive values of $I(\infty)$. We can easily consider two models, M_1 and M_2, for Δy_t which correspond to δ in the interval $(-1, 0.5)$ excluding 0 and $\delta = 0$, respectively. By allocating prior weight to each model we obtain a positive posterior probability for both M_1 and M_2. The ARFIMA model M_1 will lead to point masses for $I(\infty)$ at 0 and ∞, given by the posterior probabilities that $\delta < 0$ and $\delta > 0$, respectively. The ARMA model M_2, however, results in a continuous distribution over $(0, \infty)$ for $I(\infty)$. Posterior inferences about $I(\infty)$ are then based on weighted averages of results from M_1 and M_2, where the weights are the posterior model probabilities.

Any Bayesian analysis which allocates non-zero prior probability to $\delta > 0$ automatically implies that the impulse response at infinity is ∞ with the corresponding posterior probability, which can, however, be much smaller than 1. In fact, it could be argued on the basis of economic plausibility that finite values of $I(\infty)$ are a priori more reasonable than infinity. Of course, if our subjective opinion excludes an infinite value for $I(\infty)$, we can express this by allocating zero prior weight to $\delta > 0$.[2] If our prior allows for $\delta > 0$ then the posterior mean of $I(\infty)$ will be infinite.[3]

To summarize: in this section we have briefly described the ARFIMA model and discussed some of its properties. In particular, we have discussed impulse response functions and argued that:

(i) In an applied context it makes more sense to focus on $I(n)$ for n finite (say, $n = 4, 12, 40$ for quarterly data) than for $n = \infty$; and

(ii) If one wishes $I(\infty)$ to have a continuous posterior on $(0, +\infty)$, then a prior

[2] Previous writers have found that this area of the parameter space does not seem empirically relevant for real U.S. GNP [see Diebold and Rudebusch (1989)]. However, the Bayesian methods used in this paper indicate that a non-negligible part of the posterior lies in this region. For disposable income, moreover, the results of Diebold and Rudebusch (1991) suggest that this area is important.

[3] The fact that the posterior mean of $I(\infty)$ is infinite need not bother us. For example, a uniform prior for δ over $(-1, 0.5)$ would yield a posterior for $I(\infty)$ composed of two point masses, one at 0 and one at ∞. If $p(\delta < 0|\text{Data}) > \frac{1}{2}$ then the posterior mode would be zero and this could be used as a point estimate. The picture becomes much more interesting when we put some prior mass at $\delta = 0$. In this case, the posterior for $I(\infty)$ still contains two point masses at 0 and ∞, but is continuous and nonzero between these two points. Even though no moments exist for such a posterior, reasonable point estimates (if required) can be based on posterior medians or other quantiles.

structure which puts some prior mass at $\delta = 0$ should be adopted.

Note that attaching a prior mass to $\delta = 0$ is essential for any formal comparison between ARFIMA and ARMA models done through posterior odds testing.

3. A Bayesian Approach

The basic ARFIMA model is given in (1). In addition to the assumptions presented there, we will also assume that the roots of $\varphi(L)$ are simple.[4] We begin by introducing some notation. Our parameter space is partitioned into μ, σ^2, and $\omega' = (\delta, \Theta', \Phi')'$, where $\Theta = (\theta_1, \ldots, \theta_q)' \in C_q$ and $\Phi = (\phi_1, \ldots, \phi_p)' \in C_p$.[5] T in-sample data points are observed for $w' = (\Delta y_1, \ldots, \Delta y_T)'$, and predictions are to be made for n out of sample points $w^{*\prime} = (\Delta y_{T+1}, \ldots, \Delta y_{T+n})'$. Our model can be written as:

$$\begin{pmatrix} w \\ w^* \end{pmatrix} = \mu \begin{pmatrix} \iota_T \\ \iota_n \end{pmatrix} + \begin{pmatrix} \xi \\ \xi^* \end{pmatrix}, \tag{4}$$

$$\begin{pmatrix} \xi \\ \xi^* \end{pmatrix} \sim N(0_{T+n}, \sigma^2 \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}) \sim N(0, \sigma^2 V), \tag{5}$$

where the elements of V are given by $v_{ij} = \sigma^{-2}\gamma(i-j)$ for $i, j = 1, \ldots, T+n$; $\gamma(s)$ is the autocovariance function given in Sowell (1992a) (Equation (8), page 173); and ι_i is an $i \times 1$ vector of ones. The sampling density of w is given by

$$p(w|\omega, \mu, \sigma^2) = f_N^T(w|\mu\iota_T, \sigma^2 V_{11}),$$

where f_N^T is the T-variate Normal density function.

We will assume the following prior structure:

$$p(\omega, \mu, \sigma^{-2}) = p(\omega)p(\mu)p(\sigma^{-2}) \propto \sigma^2 p(\omega),$$

where $\omega \in \Omega$, $\mu \in R$, $\sigma^{-2} \in R_+$, and $\Omega = (-1, 0.5) \times C_q \times C_p$.

---

[4] The reasons for making this assumption are computational and are given in detail in Sowell (1992a). Section 5.4 of Sowell (1992a) presents convincing evidence that this assumption is not restrictive in practice. From a Bayesian point of view, multiple roots have zero prior probability under any continuous prior.

[5] For posterior inference, the conditions on Θ and Φ can be enforced using the procedure in Monahan (1984).

As is shown in Osiewalski and Steel (1993), the improper prior on σ^{-2} adopted here leads to perfect robustness of posterior inference on (ω, μ) and predictive inference with respect to any deviation from joint Normality of the sampling distribution within the class of all $(T+n)$-variate elliptical densities with the same location and scale. That is, we could replace the Normal error vector in (5) by any other jointly elliptical error vector with zero mean and the same scale matrix $\sigma^2 V$, and the marginal posterior density of (ω, μ) and the predictive density would remain unchanged. The latter densities are also unaffected by rescaling the V matrix, as a result of the scale invariance of the assumed prior on σ^{-2}.

The posterior analysis can be simplified considerably by noting that, conditional on ω, we face a standard Bayesian exercise which can be solved analytically. In other words, we can integrate out μ and σ^{-2}, yielding[6]

$$p(\omega|\text{Data}) = K^{-1}|V_{11}|^{-\frac{1}{2}}(\iota_T' V_{11}^{-1}\iota_T)^{-\frac{1}{2}} SSE^{-\frac{T-1}{2}} p(\omega), \tag{6}$$

where

$$K = \int_\Omega |V_{11}|^{-\frac{1}{2}}(\iota_T' V_{11}^{-1}\iota_T)^{-\frac{1}{2}} SSE^{-\frac{T-1}{2}} p(\omega) d\omega. \tag{7}$$

In (6) and (7) we use the notation:

$$SSE = (w - \hat{\mu}\iota_T)' V_{11}^{-1}(w - \hat{\mu}\iota_T),$$

where

$$\hat{\mu} = (\iota_T' V_{11}^{-1}\iota_T)^{-1}\iota_T' V_{11}^{-1} w.$$

Equation (6) is the posterior density which forms the basis for our impulse response analysis.

Predictive distributions are also of interest, and inferences about y_{T+n} can be made on the basis of $p(w^*|\text{Data})$. As $y_{T+n} = y_T + \iota_n' w^* = y_T + n\mu + \iota_n'\xi^*$, we obtain [see, *e.g.*, Osiewalski and Steel (1993)]:

$$\begin{aligned} p(y_{T+n}|\omega, \text{Data}) = f_s^1(y_{T+n}|T-1, y_T + n\hat{\mu} + \iota_n' V_{21} V_{11}^{-1}(w - \hat{\mu}\iota_T),& \\ \frac{T-1}{SSE}\left[\iota_n' V_{22\cdot 1}\iota_n + \frac{(n - \iota_n' V_{21} V_{11}^{-1}\iota_T)^2}{\iota_T' V_{11}^{-1}\iota_T}\right]^{-1}),& \end{aligned} \tag{8}$$

[6] In general, if we had other exogenous variables which entered linearly with uniform or natural-conjugate priors on their coefficients, we could integrate the latter out in the same fashion. However, for proper priors on σ^{-2} the robustness with respect to deviations from Normality in (5) will be lost.

where $V_{22\cdot 1} = V_{22} - V_{21}V_{11}^{-1}V_{12}$, and $f_s^k(\cdot|r, b, A)$ is the k-variate Student t density with r degrees of freedom, location vector b and precision matrix A. Inferences about μ can be made on the basis of

$$p(\mu|\omega, \text{Data}) = f_s^1(\mu|T-1, \hat{\mu}, \frac{T-1}{SSE}\iota_T' V_{11}^{-1}\iota_T). \tag{9}$$

Note that the densities in (8) and (9) are conditional on ω, but we can integrate out ω using (6) through a numerical procedure.

In the empirical section, posterior properties of the parameters and impulse responses are calculated using Monte Carlo integration with importance sampling on ω (see Geweke (1989)). Both the prior $p(\omega)$ and the importance function are taken to be proper uniform densities on Ω.

It should be noted that standard Gibbs sampling methods are not easy to use with ARFIMA models. There have been two recent papers (Chib and Greenberg (1994) and Marriott, Ravishanker, Gelfand and Pai (1995)) which use Gibbs sampling methods with ARMA models. These papers use data augmentation and treat the initial conditions as latent variables. Recently, Pai and Ravishanker (1994) proposed a similar approach for ARFIMA models. Note, however, that their full conditional densities have nonstandard forms and they employ a Metropolis-within-Gibbs algorithm. It is an open question which numerical method is better, but given that our simple Monte Carlo approach works well, we see little need to use such complicated Markov chain algorithms with their inherent difficulties in assessing convergence.

In our application we discuss the shape of impulse responses at different horizons for real U.S. GNP. There are a few general theoretical results that can be obtained. As noted previously, the posterior for $I(\infty)$ is a point mass at 0 or ∞ for $\delta < 0$ or $\delta > 0$, respectively. If $\delta = 0$ then the model collapses to a standard ARIMA$(p, 1, q)$ model, which Koop, Osiewalski and Steel (1994) discusses in detail. In the latter paper we found that the posterior for $I(\infty)$ usually has no moments, but that, if $\Phi \in C_p$, all posterior moments exist for $I(n)$ when n is finite. The empirical illustration in Koop, Osiewalski and Steel (1994) indicated that, in practice, impulse responses can have very non-Normal posterior densities and, hence, that point estimates accompanied by standard deviations could be misleading.

As yet, we have presented Bayesian inference for the general ARFIMA(p, δ, q) model in (1), assuming the orders p and q fixed. However, in the next Section we will consider several (say, m) specifications which will differ in the values for p and q and in the assumptions about δ. In view of the discussion in the previous Sections, the cases with $\delta = 0$ and with non-zero δ will be treated as different models. All our models are derived by putting simplifying parameter constraints (zero restrictions) on the unrestricted ARFIMA$(3, \delta, 3)$ specification. Thus, functions of the parameters, like impulse responses, are not model-specific quantities and it is formally possible to average them over models.[7] For the purpose of model comparison, we require proper, normalized prior densities for free parameters other than location μ and scale σ (which appear in all the specifications and retain the same improper prior). This is fulfilled by proper uniform priors on $C_p \times C_q$ for the free ARFIMA coefficients, provided that their integrating constants are taken into account when the marginal data density values $K_i (i = 1, \dots, m)$ are calculated using the generic formula (7). Moreover, the uniform joint prior density of the coefficients in the nesting model leads to uniform conditional densities given the zero restrictions. Thus, uniform priors form an overall coherent prior structure, as advocated by, *e.g.*, Poirier (1985), and a marginal posterior density in a restricted model can be treated as the appropriate conditional posterior density in the unrestricted specification. Finally, the posterior probability of M_i, model i, is given by

$$p(M_i|\text{Data}) = \frac{p(M_i)K_i}{\sum_{j=1}^{m} p(M_j)K_j},$$

where $p(M_j)$ is the prior model probability of M_j.

4. The Persistence of Real GNP

In this section, we investigate the persistence properties of post-war quarterly real U.S. GNP from 1947:1 to 1989:4 (CITIBASE series GNP82).[8] This series has been examined by many authors (*e.g.*, Campbell and Mankiw (1987), Diebold and Rudebusch (1989), Hauser, Pötscher and Reschenhofer (1992), Sowell (1992b)). There has been great debate over the persistence of this series. In general, though, most researchers agree that $I(n)$ can be rather different from zero for most values of n of practical relevance. However, the fractional methods of Diebold and Rudebusch provide more evidence that $I(n) < 1$ for $n > 16$,

[7] Min and Zellner (1993) use a decision-theory context to show that mixing over models is optimal for forecasting with squared error loss, provided the set of models under consideration is exhaustive. However, they also stress that if the latter condition does not hold, mixing need not be the optimal strategy. Consequences of having other vaguely specified models in the background are discussed by Poirier (1995) for model comparison and model selection.

[8] The series $\{y_t\}$ here is the natural logarithm of U.S. GNP.

than do the non-fractional methods of Campbell and Mankiw (1987) or Hauser, Pötscher and Reschenhofer (1992).

Before going to our small-sample, parametric Bayesian approach, let us present the results of the asymptotic, nonparametric sampling-theory test of short-range dependence developed in Lo (1991). The latter modifies the well-known rescaled range (R/S) statistic such that its asymptotic behaviour is invariant over a general class of short memory processes, but deviates for long or intermediate memory. For Δy_t, our differenced series of length $T = 171$, the value of the normalized modified R/S statistic, $Q_T(k)/\sqrt{T}$, decreases from 1.262 for $k = 0$ to 0.928 for $k = 4$ and increases very slowly for higher k. Here k is the order of sample autocovariances taken into account. For the OLS detrended level series, the values of Lo's statistic are much higher, starting from 4.373 for $k = 0$ and for $k = 6$ we obtain 1.780. Since, under the null hypothesis of short memory[9] $Q_T(k)/\sqrt{T}$ is distributed, for T going to ∞ and k of the order $T^{\frac{1}{4}}$, as the range of a Brownian bridge on the unit interval, with mean 1.25, standard deviation 0.27 and tail probabilities of 0.025 below 0.809 and above 1.862, Lo's test would suggest lack of long memory for the differenced y_t, but its possible presence for the detrended level series. However, Lo (1991, p. 1308) concludes "Direct estimation of particular parametric models may provide more positive evidence of long-term memory" and adds (p. 1296) "Of course, if one is interested exclusively in fractionally-differenced alternatives, a more efficient means of detecting long-range dependence might be to estimate the fractional differencing parameter directly."

We consider 32 different models for $z_t = \Delta y_t - \mu$ corresponding to all possible ARMA(p, q) and ARFIMA(p, δ, q) for $p, q \leq 3$. We consider two sets of prior probabilities $p(M_i)$ attached to each model. First of all, we use a flat model prior with equal $p(M_i)$ for all $i = 1, \ldots, m$ as an "agnostic" or reference case. Secondly, we will present results with prior model weights based on both economic and statistical reasoning: in particular, one could argue that m-dependence is very unlikely for this economic time series, so that we should downweigh the pure MA(q) models; in addition, ARFIMA models will require AR parameters to adequately model short run behaviour, and to avoid the inference on δ to be largely driven by the short run properties of the data. The latter fact would lead to lower prior probabilities for ARFIMA$(0, \delta, q)$ models. Thus, the second "informative" prior downweighs the MA(q) models by a factor 10 with respect to the other ARMA models, and assigns five times less prior

[9] The null hypothesis restricts the degree of dependence and heterogeneity and essentially imposes strong mixing.

Table 1. Posterior Model Probabilities for ARFIMA(p, δ, q) and ARMA(p, q)

	ARFIMA(p,δ,q)		ARMA(p,q)	
p,q	flat prior	informative prior	flat prior	informative prior
0,0	0.045	0.011	0.000	0.000
0,1	0.005	0.001	0.004	0.001
0,2	0.020	0.005	0.167	0.022
0,3	0.005	0.001	0.018	0.002
1,0	0.166	0.212	0.249	0.327
1,1	0.018	0.023	0.043	0.057
1,2	0.018	0.022	0.046	0.060
1,3	0.002	0.003	0.005	0.006
2,0	0.019	0.024	0.056	0.073
2,1	0.010	0.012	0.021	0.028
2,2	0.006	0.007	0.019	0.025
2,3	0.001	0.001	0.001	0.002
3,0	0.010	0.012	0.027	0.036
3,1	0.005	0.007	0.009	0.011
3,2	0.002	0.003	0.003	0.004
3,3	0.000	0.000	0.001	0.001
Total	0.331	0.345	0.669	0.655

mass to the ARFIMA$(0, \delta, q)$ models than to the ARFIMA models with AR terms, while retaining a total prior mass of $\frac{1}{2}$ for both ARFIMA and ARMA alternatives. We shall discuss posterior model probabilities under both priors, but for the rest of the discussion we will focus on the informative prior, which is more in accordance with our prior beliefs.

The flat prior for the parameters is described in Section 3. Note that the prior for the parameters is proper, so that meaningful posterior odds can be defined.[10] Table 1 presents posterior model probabilities for these 32 models. A few points are worth emphasizing:

(i) ARFIMA models receive one third of the posterior model probability as opposed to two thirds for the ARMA models. In other words, there is some (but not overwhelming) evidence for the simple unit root. This holds for

[10] In practice, the uniform prior typically puts a lot of prior weight in regions with negligible likelihood. This will tend to increase posterior odds in favour of more parsimonious models. If this is found bothersome, more informative priors could easily be accommodated in the present framework.

both sets of prior model probabilities.

(ii) The posterior model probability is scattered widely across models, indicating the hazard of choosing just one. The informative prior leads to a somewhat higher concentration of posterior model probability, but still leads to attaching high posterior credibility to models with very different long-run properties.

(iii) The top two models are the ARMA(1,0) and ARFIMA$(1, \delta, 0)$. The ARMA$(0, 2)$ also receives a lot of posterior probability under the flat model prior, but gets downweighted by the informative prior. The overall ranking roughly corresponds to that found in Sowell (1992b) on the basis of the Schwarz criterion. Notable exceptions are the ARMA$(0, 2)$, which ranks first in Sowell's paper, but is only the 12^{th} best model under our informative prior, and the ARMA$(0, 3)$ which is 5^{th} for Sowell and 24^{th} for us. Sowell also reports values for the AIC criterion, which favours the ARFIMA$(3, \delta, 2)$ specification, a model that always receives less than 0.003 posterior probability in our framework.

(iv) The posterior odds of trend-stationarity with long memory (*i.e.*, $-1 < \delta < -.5$) versus difference-stationarity with intermediate memory ($-.5 \leq \delta < 0$) are 0.252 to one. Posterior odds in favour of $-1 < \delta < -.5$ (trend-stationarity) decrease to 0.065 and 0.061 to one if we consider the more general alternatives $-.5 \leq \delta \leq 0$ and $\delta \in [-.5, .5)$, respectively. The two latter alternatives take into account the point mass at $\delta = 0$, which imposes the simple unit root.

For the sake of brevity, we do not report information on all parameters. However, Table 2 provides posterior means and standard deviations for δ for the 16 ARFIMA models. Figure 1 plots the posterior density of δ mixed over the ARFIMA models. This plot indicates that the posterior for δ is highly non-Normal, so that posterior means and standard deviations will not be adequate summary measures. The strongest impression given by an examination of Table 2 is that δ is not estimated very precisely. Previous work (*e.g.*, Diebold and Rudebusch (1989)) using classical econometric techniques has tended to find estimates of δ less than zero. Here, too, we find that posterior means of δ are less than zero for most models. As expected, the models without autoregressive terms will tend to use δ to aid in capturing the short-run behaviour of the series (which displays positive first-order autocorrelation), thus leading to posterior mass on the positive real line for δ. However, such models were downweighted by the informative prior and if we average across all ARFIMA models, $p(\delta <$

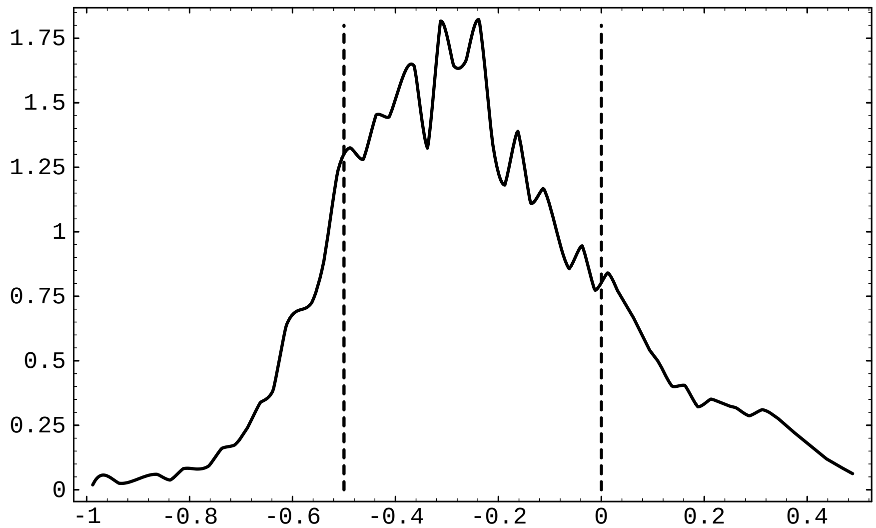

Fig. 1. Posterior of δ.
(The dashed lines delimit the regions where $\delta < -0.5$, $-0.5 < \delta < 0$ and $\delta > 0$.)

Table 2. Posterior Properties of δ for ARFIMA(p, δ, q)

p, q	Mean	Standard Deviation
0,0	0.326	0.074
0,1	0.235	0.112
0,2	0.038	0.118
0,3	−0.108	0.140
1,0	−0.290	0.217
1,1	−0.162	0.273
1,2	−0.272	0.262
1,3	−0.250	0.300
2,0	−0.208	0.254
2,1	−0.214	0.228
2,2	−0.181	0.274
2,3	−0.162	0.277
3,0	−0.246	0.252
3,1	−0.377	0.300
3,2	−0.376	0.302
3,3	0.059	0.295

$0|\text{Data}) = 0.831$ and $p(\delta > 0|\text{Data}) = 0.169$. That is, there is a small but non-negligible posterior probability that δ lies in a region which implies $I(\infty) = \infty$.

Note, however, that the prior probability of $\delta > 0$ ($\frac{1}{3}$ in the set of ARFIMA models) is halved by the data.

Table 3. Posterior Means of Impulse Responses for $n = 4$, 12 and 40

	ARFIMA$(1, \delta, 0)$	ARMA$(1, 0)$	Overall Model
$n = 4$	1.548	1.604	1.637
	(0.211)	(0.176)	(0.204)
$n = 12$	1.323	1.623	1.600
	(0.406)	(0.196)	(0.310)
$n = 40$	0.996	1.623	1.523
	(0.674)	(0.196)	(0.416)

(Posterior Standard Deviations in Parentheses)

Table 3 and Figures 2 through 4 provide information on the impulse responses for $n = 4$, 12 and 40. For the sake of brevity, three models are discussed: the ARMA$(1, 0)$, ARFIMA$(1, \delta, 0)$ and what we call the "overall" model which averages over all 32 individual ARMA and ARFIMA models. Table 3 presents posterior means and standard deviations of these impulse responses. The former are roughly the same for all three models, but the ARFIMA and overall models have larger standard deviations attached, especially at long horizons. This latter finding is consistent with the high posterior standard deviation of δ discussed above. The figures, however, give a warning against trusting too much in means and standard deviations. Especially at long horizons, the impulse responses are highly skewed and, for the fractional models, have very fat tails. The overall model, which is a mixture of models, yields multimodal impulse responses.

Impulse response $I(\infty)$, averaged over all models, has a point mass at zero equal to 0.287 and a point mass of 0.058 at infinity. Figure 5 presents the continuous part of the posterior distribution for $I(\infty)$, corresponding to the ARMA models. It can be seen that the ARMA models are more or less in agreement about $I(\infty)$ in that the bulk of the posterior probability lies between 1 and 2. However, the fractional models put point masses at zero and infinity as described above. Even if we ignore these point masses, no moments exist for the continuous part of the distribution (see Koop, Osiewalski and Steel (1994)). In order to evaluate the prior to posterior mapping in terms of $I(\infty)$, we also plot in Figure 5 prior and posterior densities of $I(\infty)$ corresponding to the dominant ARMA model, the ARMA$(1, 0)$. In this simple case, we can easily see that the prior on $I(\infty)$ induced by a flat prior on $(-1, 1)$ for ϕ_1 will be $p(I(\infty)) = \frac{1}{2}I(\infty)^{-2}$ on the interval $(\frac{1}{2}, \infty)$. Clearly, the data information changes the prior

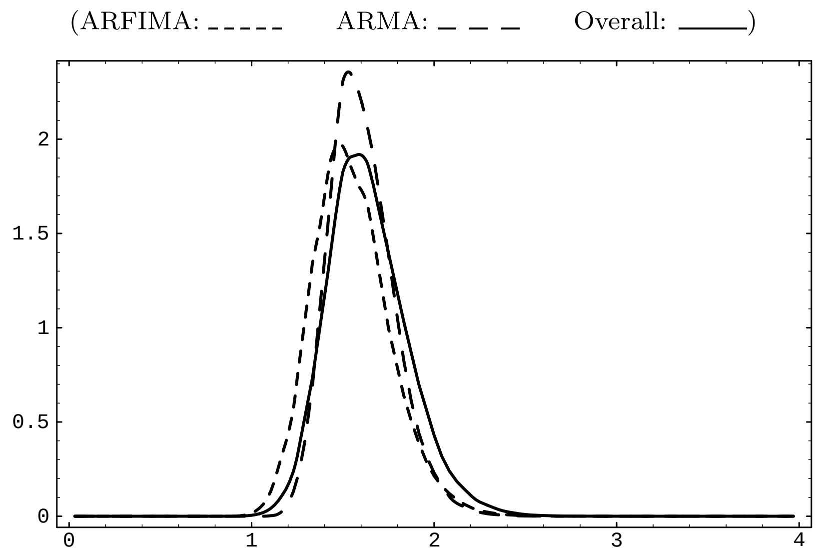

Fig. 2. Posterior of $I(4)$.

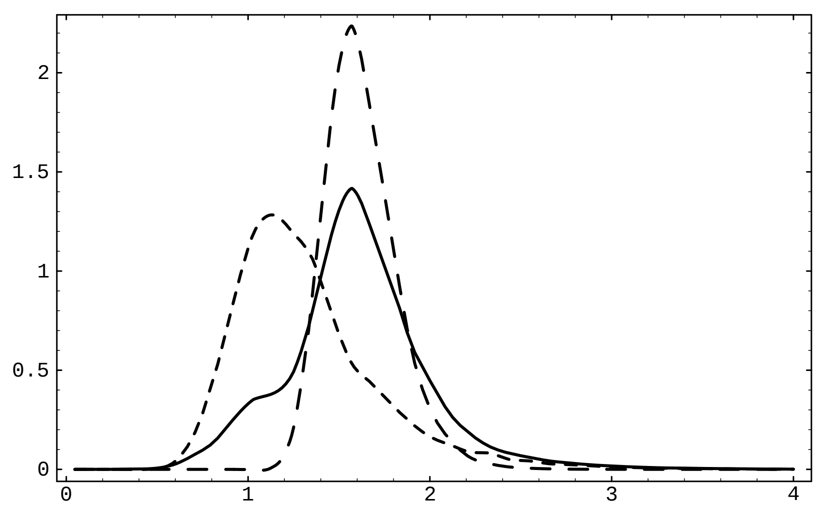

Fig. 3. Posterior of $I(12)$.

substantially in this case.

A direct comparison of the $\mathrm{ARMA}(1,0)$ and $\mathrm{ARFIMA}(1,\delta,0)$ models shows how, at short horizons, they yield similar impulse responses. However, at medium- and long-run horizons the two models yield rather different inferences. This is not surprising, since the fractional differencing parameter, which drives the long-run dynamics, is not concentrated in a small neighborhood of zero.

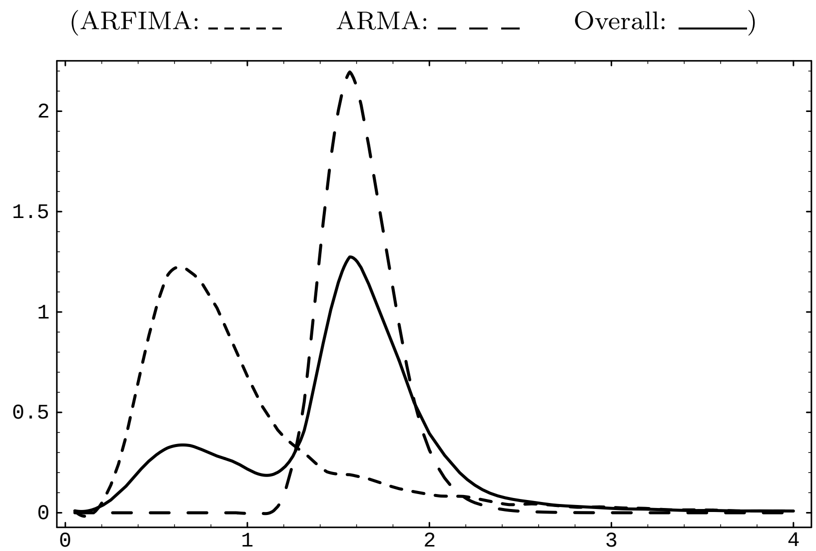

Fig. 4. Posterior of $I(40)$.

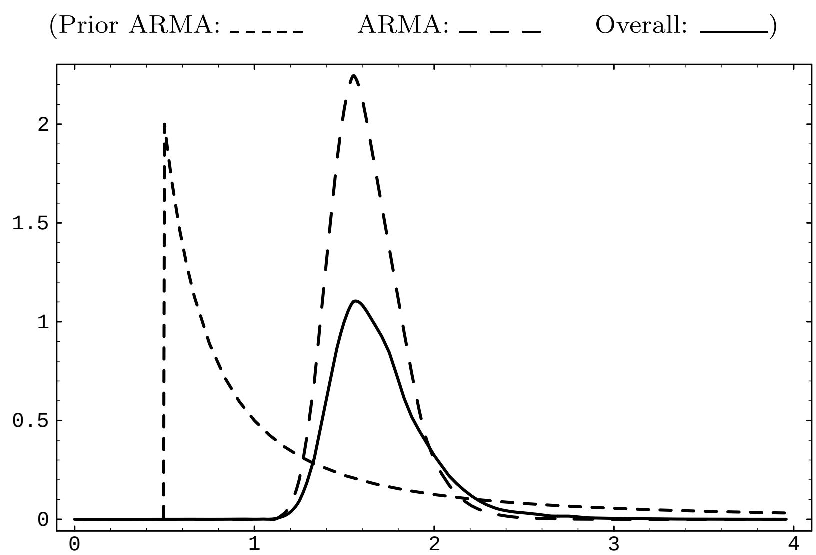

Fig. 5. Prior and Posterior of $I(\infty)$.

(Note: The overall posterior of $I(\infty)$ has point masses at 0 and ∞ of 0.287 and 0.058, respectively, which are not shown on this graph.)

Our Bayesian paradigm leads us to advocate the use of the overall model for making inferences. The figures make clear that this model indicates a large amount of uncertainty about the degree of persistence in real U.S. GNP. For instance, in Figure 4 there is substantial probability that $I(40)$ is less than 1. This stands in contrast to the model with highest posterior probability (*i.e.*, the ARMA$(1,0)$) which implies $I(40) > 1$ with virtual certainty. If we selected only one model we would make very different inferences than using the overall model.

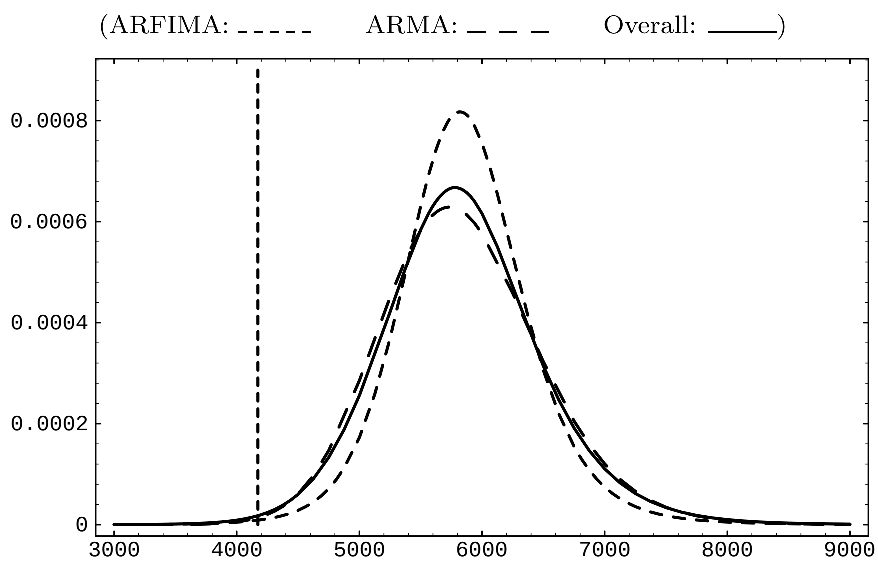

Fig. 6. Predictive Distribution of $\exp(y_{T+40})$.
(The dashed line shows the location of $\exp(y_T) = 4172.4$.)

Predictive distributions can be calculated using (8). Figure 6 plots the predictive distribution of $\exp(y_{T+40})$, *i.e.*, real U.S. GNP in the fourth quarter of 1999, for the three models discussed previously: the ARMA$(1,0)$, ARFIMA$(1,\delta,0)$ and the overall model. It can be observed that the three predictives are quite similar.

If out-of-sample prediction in a changing economic environment, rather than persistence issues, were the focus of this paper, we might consider models with some additional structural aspects, such as the leading indicator ARLI models used in e.g. Zellner and Min (1993).[11] Indeed, our pure time series models are not very successful in predicting the slowdown of 1990 and 1991: predictive means overestimate the actual y_{T+n} by about one predictive standard deviation for $4 \leq n < 12$. As regards the within-sample fit of the classes

[11] As mentioned in footnote 5, the introduction of additional explanatory variables into model (1) can trivially be handled in our theoretical framework

of models we consider here, we can report that posterior means of σ^2 are about 85% of the sample variance (of Δy_t) for most models, ranging from 83% for the ARFIMA$(1, \delta, 3)$ model to 100% (with certainty) for the ARMA$(0, 0)$ specification. The importance of σ^2 in explaining the sample variance sheds some light on the very different robustness properties of the models under consideration as regards forecasting on the one hand and measuring persistence on the other hand. While the predictive uncertainty, driven mainly by the sampling variance, is more or less the same across models, inference on long-run persistence is based solely on the parameters in ω, which leads to very different conclusions depending on the model.

The computations were conducted in FORTRAN 77. All results are based on 25,000 antithetic drawings. On RISC-based workstations, 100 antithetic Monte Carlo replications took between 1 and 5 minutes of CPU time for a full analysis of an ARFIMA model (depending on the values of p and q).[12]

5. Conclusion

In this paper we have developed Bayesian techniques for the analysis of ARFIMA models and have shown how they can be implemented using Monte Carlo integration. With regards to the measurement of persistence, ARFIMA models have been criticized since they imply that impulse responses at infinity are either zero or infinity. We discuss how a Bayesian approach can surmount this problem by allocating prior weight to the standard ARMA model in first differences. This also enables us to test between ARFIMA and ARMA specifications through their posterior probabilities.

A post-war quarterly real U.S. GNP series is used to illustrate the Bayesian approach. Although some of our results for particular models are similar to those found in previous non-Bayesian studies, important differences remain. Since we do not choose one model, but rather average over 32 different specifications, our results formally reflect model uncertainty. Hence, posterior densities for impulse responses tend to be quite spread out. For instance, regions of the parameter space which imply $I(\infty) = \infty$ receive non-trivial posterior probability. It is worth stressing that we obtain finite sample results. Furthermore, the non-Normal shape of these finite sample distributions we find for impulse responses leads us to question the usefulness of standard asymptotic methods.

[12] We used DECstations 5000/200's, HP-Apollo 9000/710's and 9000/720's. The code is available from the authors upon request.

References

Baillie, R. and C. Chung (1992): "Estimation of fractionally integrated processes with ARCH innovations," manuscript.

Baillie, R., C. Chung, and M. Tieslau (1992): "The long memory and variability of inflation: A reappraisal of the Friedman hypothesis," CentER Discussion Paper 9246, Tilburg University.

Beran, J. (1994): *Statistics for Long-Memory Processes*. New York: Chapman and Hall.

Beveridge, S. and C. Oickle (1993): "Estimating fractionally integrated time series models," *Economics Letters*, **43**, 137–142.

Brockwell, P.J. and R.A. Davis (1991): *Time Series: Theory and Methods*. New York: Springer-Verlag.

Campbell, J. and N.G. Mankiw (1987): "Are output fluctuations transitory?," *Quarterly Journal of Economics*, **102**, 857–880.

Carlin, J. and A. Dempster (1989): "Sensitivity analysis of seasonal adjustments: Empirical case studies (with discussion)," *Journal of the American Statistical Association*, **84**, 6–32.

Cheung, Y.-W. (1993): "Long memory in foreign-exchange rates," *Journal of Business and Economic Statistics*, **11**, 93–102.

Cheung, Y.-W. and F.X. Diebold (1994): "On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean," *Journal of Econometrics*, **62**, 301–316.

Cheung, Y.-W. and K. Lai (1993): "A fractional cointegration analysis of purchasing power parity," *Journal of Business and Economic Statistics*, **11**, 103–112.

Chib, S. and E. Greenberg (1994): "Bayes inference in regression models with ARMA(p, q) errors," *Journal of Econometrics*, **64**, 183–203.

Dahlhaus, R. (1989): "Efficient parameter estimation for self-similar processes," *Annals of Statistics*, **17**, 1749–1766.

Diebold, F.X. and G.D. Rudebusch (1989): "Long memory and persistence in aggregate output," *Journal of Monetary Economics*, **24**, 189–209.

Diebold, F.X. and G.D. Rudebusch (1991): "Is consumption too smooth? Long memory and the Deaton paradox," *Review of Economics and Statistics*, **73**, 1–9.

Fox, R. and M. Taqqu (1986): "Large-sample properties of parameter estimates for strongly dependend stationary Gaussian time series," *Annals of Statistics*, **14**, 517–532.

Geweke, J. (1989): "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, **57**, 1317–1339.

Geweke, J. and S. Porter-Hudak (1983): "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, **4**, 221–238.

Granger, C.W.J. (1980): "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, **14**, 227–238.

Granger, C.W.J. and A. Andersen (1978): "On the invertibility of time series models," *Stochastic Processes and their Applications*, **8**, 87–92.

Granger, C.W.J. and R. Joyeux (1980): "An introduction to long-memory time series models and fractional differencing," *Journal of Time Series Analysis*, **1**, 15–39.

Hauser, M.A., B.M. Pötscher and E. Reschenhofer (1992): "Measuring persistence in aggregate output: ARMA models, fractionally integrated ARMA models and nonparametric procedures," manuscript.

Hosking, J.R.M. (1981): "Fractional differencing," *Biometrika*, **68**, 165–176.

Janacek, G. (1982): "Determining the degree of differencing for time series via the long spectrum," *Journal of Time Series Analysis*, **3**, 177–183.

Koop, G. (1991): "Intertemporal properties of real output: A Bayesian analysis," *Journal of Business and Economic Statistics*, **9**, 253-266.

Koop, G., J. Osiewalski and M.F.J. Steel (1994): "Posterior properties of long-run impulse responses," *Journal of Business and Economic Statistics*, **12**, 489–492.

Li, W. K. and A. I. McLeod (1986): "Fractional time series modelling," *Biometrika*, **73**, 217–221.

Lo, A.W (1991): "Long-term memory in stock market prices," *Econometrica*, **59**, 1279–1313.

Marriott, J., N. Ravishanker, A. Gelfand and J.S. Pai (1995): "Bayesian analysis of ARMA processes: Complete sampling based inference under full likelihoods," in *Bayesian Statistics and Econometrics: Essays in Honor of Arnold Zellner*, ed. D.A. Berry, K.M. Chaloner and J.K. Geweke. New York: Wiley.

Min, C. and A. Zellner (1993): "Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates," *Journal of Econometrics*, **56**, 89–118.

Monahan, J. (1984): "A note on enforcing stationarity in autoregressive-moving average models," *Biometrika*, **71**, 403–404.

Newbold, P. and C. Agiakloglou (1993): "Bias in the sample autocorrelations of fractional noise," *Biometrika*, **80**, 698–702.

Odaki, M. (1993): "On the invertibility of fractionally differenced ARIMA processes," *Biometrika*, **80**, 703–709.

Osiewalski, J. and M.F.J. Steel (1993): "Robust Bayesian inference in elliptical regression models," *Journal of Econometrics*, **57**, 345–363.

Pai, J.S. and N. Ravishanker (1994): "Bayesian analysis of autoregressive fractionally integrated moving average processes," manuscript.

Poirier, D. (1985): "Bayesian hypothesis testing in linear models with continuously induced conjugate priors across hypotheses," in *Bayesian Statistics 2*, ed. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith. Amsterdam: North-Holland.

Poirier, D. (1995): "Comparing and choosing between two models with a third model in the background," manuscript.

Schotman, P.C. and H.K. van Dijk (1991): "On Bayesian routes to unit roots," *Journal of Applied Econometrics*, **6**, 387–401.

Sowell, F. (1992a): "Maximum likelihood estimation of stationary univariate fractionally integrated models," *Journal of Econometrics*, **53**, 165–188.

Sowell, F. (1992b): "Modelling long-run behavior with the fractional ARIMA model," *Journal of Monetary Economics*, **29**, 277–302.

Tieslau, M., P. Schmidt and R. Baillie (1992): "A generalized method of moments estimator for long-memory processes," CentER Discussion Paper 9247, Tilburg University.