

Working Paper 92-10
March 1992

División de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Madrid)
Fax (341) 624-9849

BAYESIAN LONG-RUN PREDICTION IN TIME SERIES MODELS

Gary Koop, Jacek Osiewalski and Mark F.J. Steel

Abstract

This paper considers Bayesian long-run prediction in time series models. We allow time series to exhibit stationary or non-stationary behavior and show how differences between prior structures which have little effect on posterior inferences can have a large effect in a prediction exercise. In particular, the Jeffreys' prior given in Phillips (1991) is seen to prevent the existence of one-period ahead predictive moments. A Bayesian counterpart is provided to Sampson (1991) who takes parameter uncertainty into account in a classical framework. An empirical example illustrates our results.

Key words

forecasting, predictive moments, unit root, parameter uncertainty.

Acknowledgements: The first author acknowledges the hospitality of the Center for Economic Research, Tilburg University. The second author received support from the Polish State Committee for Scientific Research (KBN Grant number 1-0894-91-02) and benefitted from the hospitality of Tilburg University and Universidad Carlos III de Madrid. The third author held a research fellowship from the Royal Netherlands Academy of Arts and Sciences (KNAW).

* Koop, Department of Economics, Boston University, Boston, U.S.A.; Osiewalski, Institute of Statistics, Econometrics and Informatics, Academy of Economics, Kraków, Poland; Steel, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain and Department of Econometrics, Tilburg University, The Netherlands.

Section 1: Introduction

Prediction is an important activity of econometricians and statisticians, but one that has received relatively little attention in the Bayesian time series literature. This paper represents an attempt to fill this gap. We present several new results and link these to existing results, paying special attention to the role of prior information in long-term forecasting.

Chow (1973) considers an autoregressive process with exogenous variables and Normal i.i.d. errors and discusses Bayesian multi-period prediction from a decision-theoretic point of view. Assuming a diffuse prior, he derives the Bayesian predictor under quadratic loss and demonstrates that it requires calculation of high order moments of the Student-t posterior of the coefficients of the autoregressive process. Since not all moments of Student-t densities are finite, the Bayesian predictor may not exist when the researcher wants to forecast far ahead. Differences between Chow's predictor and what he calls a "naively constructed predictor" are due to his Bayesian predictor being the mean of the predictive density rather than an estimated conditional expectation of future values given the sample, as in the sampling theory approach (see, e.g., Fuller and Hasza (1981)).

Bayesian forecasting, whether performed in a decision context or not, is based on the predictive density. For an AR(p) process with Normal i.i.d errors, Broemeling and Land (1984) show that the multi-period predictive density corresponding to a Normal-gamma or diffuse prior can be written as a product of conditional univariate Student-t densities. This representation facilitates, and gives a Bayesian interpretation to, sequential forecasting. In other words, the researcher can predict the (k+1)th future value given forecasts of the preceding k future values. Stressing that the joint multi-period predictive density is not a multivariate-t density, Broemeling and Land (1984) give explicit formulae for the two period ahead marginal predictive mean and variance.

The computational aspects of Bayesian multi-period

forecasting are considered by Thompson and Miller (1986), who propose: i) generating parameter values from the joint posterior; ii) simulating future paths of the time series; and iii) presenting certain characteristics of the resulting bundle of future paths. Schnatter (1988) proposes using Gaussian sum approximations to the n period ahead marginal predictive distribution.

Since all these authors assume a natural conjugate or diffuse prior, they do not examine the sensitivity of the predictive distribution with respect to the prior. In this paper we show the crucial role played by prior assumptions, especially with respect to long-term forecasting.

In Section 2 we consider Bayesian multi-period prediction in a regression model with a general lag structure for the endogenous variable and possibly correlated or heteroscedastic errors. No specific prior for the autoregressive coefficients is assumed; rather we attempt to establish relationships between the form of the prior and the existence of first and second order predictive moments. Section 3 examines the behavior of the predictive variance for a simple AR(1) model as the forecast horizon, n , tends to infinity. Contrary to the case where the autoregressive coefficient α is known (and in the stationary region), we demonstrate that imposing stationarity through the prior for α need not lead to the stabilization of the predictive variance as n increases. In addition, we show how assuming a prior which strongly favors explosiveness (as in the case of Phillips' (1991) "objective prior") can prevent the existence of predictive moments even for forecasting one period ahead. Section 4 demonstrates that the predictive variance in a simple trend-stationary model with $0 < \alpha < 1$ grows at least at the same rate as in the corresponding unit root model (which imposes not only $\alpha=1$, but also an additional parametric restriction). This is a Bayesian counterpart of the result obtained by Sampson (1991) whose sampling theory points to the crucial role played by a researcher's lack of knowledge of trend or drift parameters. We also stress that neglecting the additional restriction of the unit root model has severe predictive consequences. Finally,

Section 5 gives an empirical illustration of some of the theoretical results described in the other sections and Section 6 concludes.

Section 2: General Formulae for Predictive Densities and Moments

Consider the following dynamic regression model:

$$y_t = \alpha_1 y_{t-1} + \dots + \alpha_q y_{t-q} + x_t' \beta + \epsilon_t \quad (2.1)$$

where x_t ($t=1, 2, \dots$) is a $k \times 1$ vector of exogenous variables treated as known constants¹, $\alpha = (\alpha_1 \dots \alpha_q)' \in \Gamma \subseteq \mathbb{R}^q$ and $\beta \in \mathbb{R}^k$ are vectors of unknown coefficients, and ϵ_t is an error term. We are interested in forecasting y_t for $t=T+1, \dots, T+n$, given that we observe y_t ($t=1, \dots, T$) and all the x_t s ($t=1, \dots, T+n$) and condition on initial values $Y_{(0)} = (Y_{1,q}, \dots, Y_0)'$. Defining y and y^* to be the vector of T observations and n future values respectively, we can rewrite (2.1) in matrix form as:

$$\begin{bmatrix} y \\ y^* \end{bmatrix} = \begin{bmatrix} Y_- \\ Y_* \end{bmatrix} \alpha + \begin{bmatrix} X \\ X^* \end{bmatrix} \beta + \begin{bmatrix} e \\ e^* \end{bmatrix} \quad (2.2)$$

where $X = (x_1, \dots, x_T)'$ and $X^* = (x_{T+1}, \dots, x_{T+n})'$, and the matrices Y and Y^* group the appropriate lagged values of y_t . We assume that the errors are jointly Normal with a positive definite symmetric covariance matrix dependent on a precision factor, $\tau \in \mathbb{R}_+$, and possibly on a parameter vector, $\eta \in H$. In other words,

$$p(e, e^* | y_{(0)}, \omega) = f_N \left(\begin{bmatrix} e \\ e^* \end{bmatrix} \mid 0, \frac{1}{\tau} \begin{bmatrix} V(\eta) W(\eta) \\ W'(\eta) V^*(\eta) \end{bmatrix} \right), \quad (2.3)$$

where $\omega' = (\beta' \ \alpha' \ \eta' \ \tau) = (\beta' \ \theta' \ \tau)$ with $\theta' = (\alpha' \ \eta') \in \Theta \subseteq \Gamma \times H$. For example, η is the vector of MA coefficients if the errors in (2.2) follow a moving average process.

Since we are mainly interested in the effect of prior assumptions about α on long-term Bayesian forecasts, we simplify our presentation by assuming improper uniform priors for β and

¹Our notation does not make explicit the conditioning on the x_t s. In cases of particular interest, like the trend-stationary model of Section 4, the x_t s are known constants.

$\log(\tau)$. The same qualitative results could be derived by adopting a Normal-gamma prior for (β, τ) .

It is well known that posterior inference about the parameters is not affected by the presence of lagged endogenous variables. Under the following prior structure² (possibly depending on initial values),

$$p(\omega | y_{(0)}) = p(\beta) p(\tau) p(\theta | y_{(0)}) \propto \tau^{-1} p(\theta | y_{(0)}) ; \quad (2.4)$$

the joint posterior density of the parameters of (2.2) and (2.3) is:

$$p(\omega | y_{(0)}, y) = p(\theta | y_{(0)}, y) f_G\left(\tau \mid \frac{T-k}{2}, \frac{SSE_\theta}{2}\right) f_N^k(\beta | \beta_\theta, (\tau X'V^{-1}X)^{-1}) , \quad (2.5)$$

where $V=V(\eta)$,

$$\beta_\theta = (X'V^{-1}X)^{-1}X'V^{-1}(y - Y_\alpha) ,$$

$$SSE_\theta = (y - Y_\alpha)' M_\eta (y - Y_\alpha) ,$$

$$M_\eta = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} ,$$

and

$$p(\theta | y_{(0)}, y) \propto (|V| |X'V^{-1}X|)^{-\frac{1}{2}} p(\theta | y_{(0)}) SSE_\theta^{-\frac{(T-k)}{2}} ; \quad (2.6)$$

provided that the right hand side of (2.6) is integrable over θ . However, as documented by Chow (1973) and Broemeling and Land (1984), the presence of lagged y_t s has implications for Bayesian forecasting if the focus is on forecasting more than one period

²The improper prior on τ , $p(\tau) \propto \tau^{-1}$, used in (2.4) leads to perfect robustness of our predictive results with respect to departures from Normality of the errors within the class of jointly elliptical distributions. Following Osiewalski and Steel (1992), we can prove that if $(\epsilon' \epsilon'')$ ' given $y_{(0)}$ and ω has a $(T+n)$ -variate continuous elliptical distribution with the same location and scale as in (2.3), then integrating out τ leaves the same marginal density for (y, y', β, θ) , given $y_{(0)}$, as follows from (2.3), i.e. under Normality of errors. In particular, we can prove that the form of the predictive density $p(y' | y, y_{(0)})$ would remain the same.

ahead.

In order to obtain relatively simple and easily programmable formulae for marginal predictive densities and moments, we adopt the following notation:

$$y^* - Y^* \alpha = [-C \ A] \begin{bmatrix} y \\ y^* \end{bmatrix} = Ay^* - Cy \quad , \quad (2.7)$$

where C is $n \times T$ with non-zero elements $c_{i,Tj} = \alpha_{i+j}$ for $i=1, \dots, \min\{q, n\}$ and $j=0, \dots, q-i$. Furthermore, A is an $n \times n$ lower triangular matrix with 1's on the diagonal and $a_{i+j,j} = -\alpha_i$ ($j=1, \dots, n-1$ and $i=1, \dots, \min\{q, n-j\}$) its remaining non-zero elements. Note that the determinant of A is one and thus A is nonsingular. For example if $q=3$,

$$[-C \ A] = \left[\begin{array}{cccc|cccc} 0 & \dots & 0 & -\alpha_3 & -\alpha_2 & -\alpha_1 & 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & -\alpha_3 & -\alpha_2 & -\alpha_1 & 1 & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & 0 & -\alpha_3 & -\alpha_2 & -\alpha_1 & 1 & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 0 & \dots & -\alpha_3 & -\alpha_2 & -\alpha_1 & 1 & \dots & \dots & \dots \end{array} \right] .$$

From (2.2) and (2.7) we obtain:

$$y^* = A^{-1} (Cy + X^* \beta + e^*) \quad , \quad (2.8)$$

and in particular,

$$y_n^* = y_{T+n} = a_n' Cy + a_n' X^* \beta + a_n' e^* \quad , \quad (2.9)$$

where a_n' denotes the last (i.e. n th) row of A^{-1} . Since the marginal predictive density requires numerical integration over α , it is important to find a simple analytical expression for A^{-1} . The first row of A^{-1} , a_1' , consists of a one followed by $n-1$ zeros. It is easy to show that, if a_r' is the r th row of A^{-1} ($r=1, 2, \dots, n-1$), then the elements of the $(r+1)$ th row, a_{r+1}' , can be obtained as follows:

$$a^{r+1,1} = \alpha_1 a^{r,1} + \dots + \alpha_h a^{r,h} \quad , \quad h = \min\{q, n\} \quad , \quad (2.10)$$

$$a^{r+1,j} = a^{r,j-1} \quad , \quad j = 2, \dots, n.$$

This simple recursive formula demonstrates that the non-zero elements of A^{-1} are polynomials in the α_i s, and that the maximal power in the $(r+1)$ th row is α_1^r . Therefore, the maximal power in the last row, a_n' , is α_1^{n-1} and $a_n'Cy$ is an n th order polynomial in the α_i s.

Due to the Normality of the conditional sampling distribution of ϵ^* given $\epsilon = y - Y\alpha - X\beta$, we obtain from (2.8),

$$p(y^* | y, y_{(0)}, \omega) = f_N^n(y^* | A^{-1}[Cy + Q\beta + W'V^{-1}(y - Y\alpha)], \frac{1}{\tau}A^{-1}S(A^{-1})'), \quad (2.11)$$

where $S = V^* - W'V^{-1}W$, $Q = X^* - W'V^{-1}X$. Integrating out β and τ with the use of the Normal-gamma conditional posterior density in (2.5), we obtain the following n -variate Student t post-sample predictive density given θ :

$$p(y^* | y, y_{(0)}, \theta) = f_S^n(y^* | T-k, A^{-1}[Cy + Q\beta_\theta + W'V^{-1}(y - Y\alpha)], \frac{T-k}{SSE_\theta}A'[S + Q(X'V^{-1}X)^{-1}Q']^{-1}A), \quad (2.12)$$

where the notation specifies degrees of freedom, location vector and precision matrix, respectively.

The unconditional predictive density, $p(y^* | y, y_{(0)})$, can be obtained by marginalizing (2.12) with respect to the posterior density for θ given in (2.6). No convenient analytical expression exists for this unconditional predictive density and numerical integration is required.³ Although $p(y^* | y, y_{(0)})$ is a proper density for any n , its unconditional predictive moments need not exist, especially for long-term forecasting. We will focus on the predictive distribution of y_{T+n} and assume that $T > \max\{k+2, k+q\}$. $p(y_{T+n} | y, y_{(0)}, \theta)$ takes the form of a univariate t density with $T-k$ degrees of freedom, mean

$$E(y_{T+n} | y, y_{(0)}, \theta) = a_n'[Cy + Q\beta_\theta + W'V^{-1}(y - Y\alpha)] \text{ if } T-k > 1, \quad (2.13)$$

and variance

³Note that our approach requires numerical integration over the space of θ only. The approaches of Thompson and Miller (1986) and Schnatter (1988) amount to integration over the entire parameter space.

$$\text{Var}(y_{T+n} | y, y_{(0)}, \theta) = \frac{SSE_{\theta}}{T-k-2} a_n' [S+Q(X'V^{-1}X)^{-1}Q'] a_n \quad (2.14)$$

if $T-k > 2$.

Note that (2.13) is a polynomial of degree n and that (2.14) is a polynomial of degree $2n$ in the α_i s. Hence, $E(y_{T+n}^r | y, y_{(0)}, \eta)$ exists iff the conditional posterior of α given η has finite (rn) th order moments ($r=1,2$). This conditional posterior takes the form:

$$p(\alpha | y, y_{(0)}, \eta) = K_{\eta}^{-1} p(\alpha | y_{(0)}, \eta) f_S^q(\alpha | T-k-q, \hat{\alpha}_{\eta}, \frac{T-k-q}{s_{\eta}^2} Y' M_{\eta} Y) \quad (2.15)$$

where

$$\hat{\alpha}_{\eta} = (Y' M_{\eta} Y)^{-1} Y' M_{\eta} y, \quad s_{\eta}^2 = (y - Y \hat{\alpha}_{\eta})' M_{\eta} (y - Y \hat{\alpha}_{\eta}) .$$

The existence of high order moments of y_{T+n} depends crucially on $p(\alpha | y_{(0)}, \eta)$. If we let $h_j(\alpha)$ denote a polynomial of order j in the α_i s and f_{\max} denote the maximum value for the Student t density given in (2.15). Then,

$$\int_{R^q} h_j(\alpha) p(\alpha | y, y_{(0)}, \eta) d\alpha \leq K_{\eta}^{-1} f_{\max} \int_{R^q} h_j(\alpha) p(\alpha | y_{(0)}, \eta) d\alpha ,$$

and, therefore, the existence of the (rn) th order conditional prior moment of α is sufficient for the existence of $E(y_{T+n}^r | y, y_{(0)}, \eta)$ ($r=1,2$).

To examine the dependence of predictive moments on the prior consider the following situations:

A. If $p(\alpha | y_{(0)}, \eta)$ possesses all moments, then the conditional predictive mean and variance (given η) exist at any finite forecast horizon. Examples of this case include Normality for the conditional prior or uniformity for the conditional prior over any bounded support (e.g. the stationary region).

B. If $p(\alpha | y_{(0)}, \eta)$ is a Student t density with ν degrees of freedom, then $p(\alpha | y, y_{(0)}, \eta)$ is a 2-0 poly-t density whose moments exist up to any order less than $\nu+T-k$ (see Dreze (1977)). Hence, the predictive mean of y_{T+n} given η exists iff $n < \nu+T-k$, and the predictive variance exists iff $n < \frac{1}{2}(\nu+T-k)$.

C. If $p(\alpha, \eta | Y_{(0)}) = p(\alpha)p(\eta | Y_{(0)})$ and $p(\alpha)$ is improper uniform over R^q , then

$$p(\alpha | y, y_{(0)}, \eta) = f_S^q(\alpha | T-k-q, \hat{\alpha}_\eta, \frac{T-k-q}{S_\eta^2} Y' M_\eta Y_-) ,$$

and the predictive mean and variance of y_{T+n} (given η) exist iff $n < T-k-q$ and $n < \frac{1}{2}(T-k-q)$, respectively.

As cases A, B, and C indicate, aspects of the prior for α (given η) which are practically irrelevant for posterior inference on the parameters and for short term forecasting, may be crucial for long term prediction. For example, if we use a uniform prior, it makes little difference for first and second posterior moments whether the prior is proper and defined over a sufficiently large but bounded subset of R^q , or whether it is improper and defined over all of R^q . However, predictive means and variances for the proper uniform prior will always exist and will become infinite at some horizon for the improper uniform prior.

Section 3: Predictive Variances in the AR(1) Model

In this section we focus on the influence of prior uncertainty about α on the long term predictive variance in the model:

$$y_t = \alpha y_{t-1} + \epsilon_t, \quad \epsilon_t \sim i.i.N(0, \tau^{-1}) , \quad (3.1)$$

with $\alpha \in R$, which is the simplest case of (2.1). Assuming the prior structure:

$$p(\alpha, \tau | y_0) = p(\tau)p(\alpha | y_0) \propto \tau^{-1} p(\alpha | y_0) , \quad (3.2)$$

we obtain the joint posterior density:

$$p(\alpha, \tau | y, y_0) \propto p(\alpha | y_0) f_S^1(\alpha | T-1, \hat{\alpha}, \frac{T-1}{S^2} w) f_G(\tau | \frac{T}{2}, \frac{SSE_\alpha}{2}) , \quad (3.3)$$

where

$$w = \sum_{t=0}^{T-1} y_t^2, \quad \hat{\alpha} = \frac{1}{w} \sum_{t=1}^T y_t y_{t-1}, \quad s^2 = \sum_{t=1}^T (y_t - \hat{\alpha} y_{t-1})^2,$$

and

$$SSE_{\alpha} = s^2 + w(\alpha - \hat{\alpha})^2.$$

In the case of (3.1), (2.12) simplifies and leads to

$$P(y_{T+n} | y, y_0, \alpha) = f_S^1(y_{T+n} | T, \alpha^n y_T, \frac{T}{SSE_{\alpha}} (\sum_{i=0}^{n-1} \alpha^{2i})^{-1}). \quad (3.4)$$

The conditional variance from (3.4), treated as a function of the forecast horizon, is bounded from above when $|\alpha| < 1$, is proportional to n when $|\alpha| = 1$, and grows exponentially with n when $|\alpha| > 1$. However, the unconditional predictive variance takes the form:

$$(3.5) \quad \text{Var}(y_{T+n} | y, y_0) = \frac{1}{T-2} E([s^2 + w(\alpha - \hat{\alpha})^2] \sum_{i=0}^{n-1} \alpha^{2i} | y, y_0) + y_T^2 \text{Var}(\alpha^n | y, y_0),$$

which requires the existence of posterior moments of α up to order $2n$. To examine the limiting behavior of predictive variance (3.5) as n increases, note that

$$\text{Var}(y_{T+n} | y, y_0) > \frac{s^2}{T-2} \sum_{i=0}^{n-1} E(\alpha^{2i} | y, y_0), \quad (3.6)$$

which gives a restriction that bounds the variance from below. By way of illustration of the role of prior uncertainty, consider the following priors for α .

1. If $p(\alpha | y_0)$ is uniform over $(0, 1)$, then, for any r ,

$$0 < c_*(r+1)^{-1} < E(\alpha^r | y, y_0) < c^*(r+1)^{-1} < \infty, \quad (3.7)$$

where c_* and c^* are, respectively, the inf and sup of the marginal posterior density of α , a truncated Student t density. Note also that $c_* > 0$ due to truncation. Using (3.6) and (3.7) we deduce that

$$\text{Var}(y_{T+n} | y, y_0) > \frac{s^2 c_*}{T-2} \sum_{i=0}^{n-1} (2i+1)^{-1} > \frac{s^2}{T-2} \frac{c_*}{2} \sum_{i=1}^n \frac{1}{i},$$

and, since the harmonic series $\sum i^{-1}$ diverges,

$$\lim_{n \rightarrow \infty} \text{Var}(y_{T+n} | y, y_0) = \infty. \quad (3.8)$$

Contrary to the case where $\alpha \in (0,1)$ and is known, the unconditional predictive variance tends to infinity as the forecast horizon increases. Therefore, imposing stationarity on (3.1) does not guarantee that the predictive variance will stabilize at some finite level as n grows.⁴

2. Suppose $p(\alpha | y_0)$ is uniform over $(0, 1+d)$, $d > 0$. For any r ,

$$0 < \frac{c_*}{r+1} (1+d)^{r+1} < E(\alpha^r | y, y_0) < \frac{c_*}{r+1} (1+d)^{r+1} < \infty, \quad (3.9)$$

and thus,

$$\text{Var}(y_{T+n} | y, y_0) > \frac{s^2 c_*}{T-2} \sum_{i=0}^{n-1} \frac{(1+d)^{2i+1}}{2i+1} \rightarrow \infty \text{ as } n \rightarrow \infty.$$

Allowing for explosiveness magnifies the effect found in the previous case, even if d is very small. In the present case, the predictive variance is bounded from below by a series growing faster than the harmonic one. Note, however, that the variance is finite for any finite n .

3. Suppose $p(\alpha | y_0)$ is uniform over the real line and thus is improper. Then

$$p(\alpha | y, y_0) = f_s^1(\alpha | T-1, \hat{\alpha}, \frac{T-1}{s^2} w),$$

and the predictive variance exists iff $n < \frac{1}{2}(T-1)$. In the present case there is too much prior uncertainty about α in the sense that the explosiveness of (3.1) is infinitely many times more

⁴If there is uncertainty about α in the stationary AR(1) model and if $p(\alpha | y_0) \geq c > 0$ for $\alpha \in (d, 1)$ or for $\alpha \in (-1, d)$, where $d \in (-1, 1)$, then (3.8) holds. However, if $p(\alpha | y_0)$ is, e.g., uniform over $(0, d]$ where $d \in (0, 1)$, then $\text{Var}(y_{T+n} | y, y_0)$ has a finite limit as $n \rightarrow \infty$.

probable a priori than is stationarity, and the predictive variance exists in the case of short and medium term forecasting only.

4. Suppose that Phillips' "objective" prior is assumed (Phillips (1991)). In this case, the marginal posterior of α is a proper distribution with Cauchy tails and thus lacks integer moments. In this case, predictive moments do not exist at all, and consequently, Chow's predictor does not exist, even for one period forecasting.

This section has illustrated for a very simple case some implications of the more general results derived in Section 2. For typical macroeconomic time series, the priors described in 1, 2, and 3 of this section would yield very similar posterior inferences; however the implications for long-term forecasting would differ greatly. Phillips' prior, which has been advocated as one which yields sensible posterior inferences in the context of unit root testing, is shown to yield surprising predictive results.

Section 4: A Bayesian Counterpart to Sampson's (1991) Result

Assume that the deviations from the linear trend of y_t follow a simple AR(1) process. That is,

$$y_t = \gamma + \mu t + u_t, \quad t=1, 2, \dots \quad (4.1)$$

$$u_t = \alpha u_{t-1} + \epsilon_t, \quad \epsilon_t \sim i.i.d. N(0, \tau^{-1}).$$

The structural model in (4.1) has the following reduced form:

$$y_t = \alpha y_{t-1} + x_t' \beta + \epsilon_t, \quad (4.2)$$

which is a special case of (2.1), where $q=1$, $k=2$, $x_t' = (1 \ t)$, $\beta = [\gamma(1-\alpha) + \alpha\mu \ \mu(1-\alpha)]'$, and $\omega' = (\alpha \ \beta' \ \tau)$. For this case, (2.11) leads to:

$$P(y_{T+n} | y, y_0, \omega) = f_N^1(y_{T+n} | \alpha^n y_T + \alpha^n \beta, \frac{1}{\tau} \sum_{i=0}^{n-1} \alpha^{2i}), \quad (4.3)$$

where

$$\alpha'_n = \left[\sum_{i=1}^n \alpha^{n-i} \sum_{j=1}^n (T+i) \alpha^{n-i} \right] .$$

We will see that taking the structural form (4.1) as a starting point will have serious consequences for our predictive analysis.

Note that if we consider the conditional variance of Y_{T+n} given the parameters as a function of the forecast horizon, then this function is bounded above by $[\tau(1-\alpha^2)]^{-1}$ when $|\alpha| < 1$, is equal to n/τ when $|\alpha| = 1$, and grows exponentially when $|\alpha| > 1$, since then it equals $\tau^{-1}(\alpha^{2n}-1)/(\alpha^2-1)$. This behavior of the conditional variance of Y_{T+n} has been used to justify taking the unit root model as the null hypothesis in classical unit root testing (see Dickey, Bell and Miller (1986)) or even as an argument against trend-stationary models (see, e.g., Phillips (1991)). As Sampson (1991, p. 67) writes: "The idea that uncertainty is bounded in the long run would appear to be implausible, since it seems unlikely that one could be as confident of a prediction for GNP in the year 2000 as the year 2200. Thus there would seem to be a strong a priori argument for unit root models". However, as Sampson shows, this argument relies crucially on the assumption that the parameters are known. In this section, we will show that the same result holds from a Bayesian viewpoint.

In the previous section we have discussed the role of the form of the prior distribution of α . Here we will focus on prior uncertainty about γ and μ . This corresponds to the purely classical analysis of Sampson (1991) who shows for both the unit root and the trend-stationary models, that the sampling theoretic forecast variances grow with n^2 (i.e. at a much faster rate than without parameter uncertainty) and that "the key parameter is the trend or drift coefficient".

Assume the following prior structure

$$p(\alpha, \beta, \tau | y_0) = p(\alpha | y_0) p(\beta) p(\tau) \propto \tau^{-1} p(\alpha | y_0) , \quad (4.4)$$

where $\beta \in \mathbb{R}^2$, $\alpha \in \mathbb{R}$, and $\tau \in \mathbb{R}_+$. Note that (4.4) corresponds to

$$p(\alpha, \mu, \gamma, \tau | y_0) \propto \tau^{-1} (1-\alpha)^2 p(\alpha | y_0) ,$$

where $(1-\alpha)^2$ is the Jacobian of the transformation $(\beta_1, \beta_2) \rightarrow (\gamma, \mu)$.

The factor $(1-\alpha)^2$ ensures integrability of the posterior, $p(\alpha, \mu, \gamma, \tau | y, y_0)$, in the case where $p(\alpha | y_0)$ is flat near $\alpha=1$. A flat prior on the original parameters, $p(\alpha, \mu, \gamma, \tau) \propto \tau^{-1}$, leads to an improper posterior (see Schotman and van Dijk (1991)).

Under (4.2) and (4.4), the joint density (2.5) can be rewritten as

$$p(\alpha, \beta, \tau | y, y_0) \propto p(\alpha | y_0) f_S^1(\alpha | T-3, \hat{\alpha}, \frac{T-3}{S^2} w) f_G(\tau | \frac{T-2}{2}, \frac{SSE_\alpha}{2}) f_N^2(\beta | \beta_\alpha, (\tau X'X)^{-1}), \quad (4.5)$$

where

$$w = y' My_-, \quad \hat{\alpha} = \frac{1}{w} y' My, \quad s^2 = (y - \hat{\alpha} y_-)' M (y - \hat{\alpha} y_-),$$

$$SSE_\alpha = (y - \alpha y_-)' M (y - \alpha y_-) = s^2 + w(\alpha - \hat{\alpha})^2,$$

$$M = I_T - X(X'X)^{-1}X', \quad y_- = (y_0 \ y_1 \ \dots \ y_{T-1})', \quad \beta_\alpha = (X'X)^{-1}X'(y - \alpha y_-).$$

Marginalizing (4.3) with respect to the Normal-gamma posterior of (β, τ) given α in (4.5) leads to

$$p(y_{T+n} | y, y_0, \alpha) = f_S^1(y_{T+n} | T-2, \alpha^n y_T + Q_\alpha' \beta_\alpha, \frac{T-2}{SSE_\alpha} [Q_\alpha' (X'X)^{-1} Q_\alpha + \sum_{i=0}^{n-1} \alpha^{2i}]^{-1}). \quad (4.6)$$

The conditional variance given α (when $T > 4$) is

$$\text{Var}(y_{T+n} | y, y_0, \alpha) = \frac{SSE_\alpha}{T-4} \left[\sum_{i=0}^{n-1} \alpha^{2i} + \frac{2}{T(T^2-1)} \sum_{j=1}^n \sum_{i=1}^n h(i, j) \alpha^{2n-i-j} \right], \quad (4.7)$$

where

$$h(i, j) = 6ij + 3(T-1)(i+j) + 2T^2 - 3T + 1.$$

Since, for $\alpha \in (0, 1)$,

$$h(n, n) \leq \sum_{j=1}^n \sum_{i=1}^n h(i, j) \alpha^{2n-i-j} \leq h(n, n) \sum_{j=1}^n \sum_{i=1}^n \alpha^{2n-i-j} < \frac{h(n, n)}{(1-\alpha)^2},$$

it follows that $\text{Var}(y_{T+n} | y, y_0, \alpha) = O(n^2)$. That is, for the trend-

stationary model, the conditional predictive variance given α grows with the square of the forecast horizon.

If we set $\alpha=1$ in (4.7), the predictive variances grow at a much faster rate ($O(n^4)$). Remember, however, that the unit root model derived from the structural form, (4.1), also imposes a zero coefficient on the time trend. In the reduced form parameterization, (4.2), this unit root model will thus correspond to $\alpha=1$, $\beta_1=\mu$ and $\beta_2=0$. In this case (indexed here by a *),

$$P_*(Y_{T+n} | y, y_0, \mu, \tau) = f_N^1(Y_{T+n} | Y_{T+n}\mu, \frac{n}{\tau}) ,$$

and under the improper prior:⁵

$$P_*(\mu, \tau) \propto \tau^{-1} , \quad (4.8)$$

we obtain

$$P_*(\mu, \tau | y, y_0) = f_N^1(\mu | \hat{\mu}_*, \frac{1}{T\tau}) f_G(\tau | \frac{T-1}{2}, \frac{z}{2}) ,$$

and

$$P_*(Y_{T+n} | y, y_0) = f_S^1(Y_{T+n} | T-1, Y_{T+n}\hat{\mu}_*, \frac{T-1}{z} (n + \frac{n^2}{T})^{-1}) , \quad (4.9)$$

where

$$\hat{\mu}_* = \frac{1}{T} \sum_{t=1}^T (y_t - y_{t-1}) , \quad z = \sum_{t=1}^T (y_t - y_{t-1} - \hat{\mu}_*)^2 .$$

For the simple unit root model considered here, the predictive variance is

$$\text{Var.}(y_{T+n} | y, y_0) = \frac{z}{T-3} (\frac{n^2}{T} + n) , \quad (4.10)$$

which grows exactly at the same rate as the conditional predictive variance in the case of trend-stationarity. Therefore,

⁵Alternatively, the same results can be derived by using (4.2) with the prior structure given in (4.4) and conditioning on $\alpha=1$ and $\beta_2=0$. Hence, the prior structure in (4.8) is fully consistent with the improper prior given in (4.4).

from the viewpoint of a Bayesian forecaster, the unit root model is no longer more plausible than the trend-stationary model when the parameters of (4.1) are unknown and prior (4.4) is assumed.⁶ Finally, note that the unconditional predictive variance for the trend-stationary model grows at least at the same rate as (4.10) since, from (4.7),

$$\text{Var}(y_{T+n} | y, y_0) > \frac{s^2}{T-4} \left[\sum_{i=0}^{n-1} E(\alpha^{2i} | y, y_0) + \frac{2h(n, n)}{T(T^2-1)} \right] .$$

and the argument based on (3.7) can be applied.

In this section, we contrasted trend-stationary and unit root models. If, in addition, we wish to consider models that allow for explosiveness, we can apply an argument similar to that given in (3.9) to the formula in (4.7) to show that marginal predictive variances grow explosively with n .⁷ This exponential growth is faster than occurs for the trend-stationary or unit root models.

Section 5: Empirical Illustration

In the Bayesian unit root literature, it has been common to specify truncated uniform priors for α (see, e.g. DeJong and Whiteman (1991) and Koop and Steel (1991)). Our application investigates the sensitivity of predictive distributions to the choice of truncation points. In addition, we examine the predictive consequences of imposing a unit root both in the structural parameterization (4.1) and in the reduced form (4.2). Our likelihood function is based on (4.2) and we use a prior of the form (4.4), where $p(\alpha|y_0)$ is uniform on the interval $(0, 1+d)$.

⁶The restriction that $\beta_2=0$ implicit in the structural parameterization (4.1) is crucial for this result. If this restriction is neglected and we only impose $\alpha=1$ on the reduced form parameterization (4.2), then the predictive variance ((4.7) with $\alpha=1$ imposed) grows much faster than it would for trend-stationarity ((4.7) with $\alpha \in (0, 1)$).

⁷Such reasoning holds, not only for a uniform prior over $(0, 1+d)$ for $d>0$, but also for any prior defined on $(0, 1+d)$ with $p(\alpha|y_0) \geq c > 0$.

We use as data the natural logarithm of U.S. nominal wages.⁸

Table 1 reports posterior means and standard deviations as well as predictive means for $n=1,10,25,35,42,50,60,70,80,85,90,100$ and predictive standard deviations for $n=1,10,25,35,42,50$ for various values of d . The last columns report results for a flat, untruncated prior⁹ and for the unit root cases: $\alpha=1$ and $(\alpha=1, \beta_2=0)$. A graphical display of the behavior of predictive means and standard deviations as functions of the forecast horizon, n , is provided in Figures 1 and 2. Finally, Figure 3 depicts the predictive density of y_{T+42} marginalized with respect to α .

For the truncated priors, the posterior of α takes the form of a truncated Student t density (see (4.5)). The quantities in Table 1 are calculated using Monte Carlo integration with importance sampling and antithetic acceleration. As an importance function we choose a Student t with mean and variance taken as the usual ordinary least squares quantities. Standard Monte Carlo diagnostics indicate that this importance function yields results which are approximately as good as those obtained by drawing directly from the posterior.

For the flat prior without truncation, the posterior for α is a Student t density with moments existing up to $T-3$ and finite predictive variances for $n < (T-3)/2$. Entries in Table 1 for this prior are calculated analytically using results in Zellner (1971, p. 336) and Chow (1973).

Several things are worth noting about the results. i) The predictive variances increase rapidly with n and do not stabilize since parameter uncertainty about β is taken into account.¹⁰ ii)

⁸This series, which has 89 annual observations from 1900-88 on U.S. data, is taken from the extended Nelson-Plosser data set and is described in Koop and Steel (1991). The series was chosen since there is little evidence that it contains a moving average component.

⁹It is worth emphasizing that, for finite d , all predictive moments exist, but for the untruncated uniform prior predictive variances only exist up to and including $n=42$ whereas predictive means exist up to and including $n=85$.

¹⁰For our priors, even with β fixed, the predictive variances would grow indefinitely with n , albeit at a slower rate (see footnote 4).

The different priors have very little effect on predictive means and variances for small n . However, as n increases differences between the priors grow. As we expect, differences across priors are especially large in the case of predictive variances. iii) The explosion of the predictive variance at $n=43$ under the untruncated prior is found to be quite sudden, as this predictive variance is only slightly above its $d=.1$ counterpart when $n \leq 42$. The same behavior is witnessed for the predictive mean at $n=86$. iv) Rates of growth of predictive means are linear for the structural unit root case (see (4.9)), at least of order n for the trend-stationary case (from (4.6) and (3.7)), of order n^2 for the reduced form unit root model (from (4.6) with $\alpha=1$),¹¹ and are exponential when explosive behavior is not excluded. These differences are illustrated in Figure 1. v) It can be seen from Figure 3 that the marginal predictive densities for $n=42$ are highly skewed to the right, especially if we do not exclude the explosive region. Note that for $n=1$ the predictive density is Student-t if an untruncated uniform prior is assumed. Furthermore, conditional on α the predictive density is Student-t. Hence, the considerable deviation from the Student-t that is illustrated in Figure 3 is due to uncertainty about α and a prediction horizon greater than one. vi) Imposing a unit root in the structural model (4.1) is seen to have very different consequences for prediction than imposing it in the reduced form (4.2). Koop and Steel (1991) use a decision framework with a loss function based on avoiding errors in predictive variances. They contrast trend-stationary, reduced form unit root, and explosive models and find that the reduced form unit root model is favored as $n \rightarrow \infty$. Figure 2 confirms these results for one of the series they used, but also suggests that their decision analysis would favor trend-stationarity for moderate or large n if a structural unit root model was considered.

Section 6: Conclusions

This paper considers long-run prediction in time series

¹¹The reduced form unit root model implicitly imposes a quadratic trend.

models and emphasizes the role played by prior information. It is shown how aspects of the prior, which have little effect on posterior inferences or short-run predictions, can have large effects on medium- or long-run forecasts. In particular, the existence of predictive means and variances depends crucially on the existence of prior moments. We also prove that imposing stationarity through the prior on the autoregressive coefficient, α , in the simple AR(1) model need not lead to stabilization of the predictive variance as the forecast horizon increases, contrary to the case with known α in the stationary region. For models with a linear deterministic trend, we derive a Bayesian analogue to Sampson's (1991) result, which shows how the forecast variance of trend-stationary and unit root models both become infinite at infinite forecast horizons when parameter uncertainty is allowed for. Furthermore, the rate of growth of predictive variances in the trend-stationary model is at least as fast as that associated with the structural unit root model. Finally, we examine the predictive consequences of imposing a unit root in the reduced form model as opposed to the structural unit root model. Our results are illustrated in an empirical example.

References

- Broemeling, L. and M. Land (1984). "On forecasting with univariate autoregressive processes: A Bayesian approach," **Communications in Statistics: Theory and Methods**, 13, 1305-1320.
- Chow, G. C. (1973). "Multiperiod predictions from stochastic difference equations by Bayesian methods," **Econometrica**, 41, 109-118 and 796 (Erratum).
- DeJong, D. and C. Whiteman (1991). "The temporal stability of dividends and stock prices: Evidence from the likelihood function," **American Economic Review**, 81, 600-617.
- Dickey, D. A., W. R. Bell and R. Miller (1986). "Unit roots in time series models: Tests and implications," **The American Statistician**, 40, 12-26.
- Dreze, J. H. (1977). "Bayesian regression analysis using poly-t densities," **Journal of Econometrics**, 6, 329-354.
- Fuller, W. A. and D. P. Hasza (1981). "Properties of predictors for autoregressive time series," **Journal of the American Statistical Association**, 76, 155-161.
- Koop, G. and M. F. J. Steel (1991). "A decision theoretic analysis of the unit root hypothesis using mixtures of general elliptical models," Center for Economic Research Discussion Paper No. 9150, Tilburg University.
- Osiewalski, J. and M. F. J. Steel (1992). "Robust Bayesian inference in elliptical regression models," forthcoming in **Journal of Econometrics**
- Phillips, P. C. B. (1991). "To criticize the critics: An objective Bayesian analysis of stochastic trends," **Journal of Applied Econometrics**, 6, 333-364.
- Sampson, M. (1991). "The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend-stationary time series," **Journal of Applied Econometrics**, 6, 67-76.
- Schnatter, S. (1988). "Bayesian forecasting of time series by Gaussian sum approximation," in J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds., **Bayesian Statistics 3** (Clarendon Press, Oxford), 757-764.
- Schotman, P. C. and H. K. van Dijk (1991). "On Bayesian Routes to Unit Roots," **Journal of Applied Econometrics**, 6, 387-401.
- Thompson, P. and R. Miller (1986). "Sampling the future: A Bayesian approach to forecasting from univariate models," **Journal of Business and Economic Statistics**, 4, 427-436.

Zellner, A. (1971). **An Introduction to Bayesian Inference in Econometrics**, J. Wiley, New York.

Table 1: Posterior and Predictive Means and Standard Deviations

	Prior		for		α	
	d=0	d=.05	d=.1	Untrun Unif.*	Unit Root in (4.2)	Unit Root in (4.1)
	MC**	MC	MC	Ex***	Ex	Ex
E(α)	.95330	.95877	.95898	.95898	1.0	1.0
SD(α)	.02639	.03074	.03090	.03093	---	---
E(Y_{T+1})	10.180	10.181	10.181	10.181	10.191	10.175
SD(Y_{T+1})	.06401	.06421	.06424	.06424	.06405	.06330
E(Y_{T+10})	10.636	10.651	10.652	10.652	10.755	10.578
SD(Y_{T+10})	.21081	.22190	.22298	.22311	.24544	.21005
E(Y_{T+25})	11.425	11.474	11.477	11.478	11.760	11.250
SD(Y_{T+25})	.36575	.43071	.44154	.44262	.51345	.35663
E(Y_{T+35})	11.962	12.044	12.052	12.054	12.476	11.698
SD(Y_{T+35})	.46445	.60461	.64009	.64411	.71535	.44024
E(Y_{T+42})	12.341	12.454	12.464	12.469	12.998	12.012
SD(Y_{T+42})	.53537	.75764	.83790	.84571	.87056	.49579
E(Y_{T+50})	12.777	12.934	12.915	12.960	13.616	12.370
SD(Y_{T+50})	.61927	.97813	1.1430	∞	1.0627	.55735
E(Y_{T+60})	13.326	13.551	13.589	13.597	14.422	12.818
E(Y_{T+70})	13.874	14.195	14.270	14.285	15.263	13.266
E(Y_{T+80})	14.427	14.872	15.021	15.052	16.141	13.714
E(Y_{T+85})	14.705	15.226	15.438	15.481	16.594	13.938
E(Y_{T+90})	14.983	15.593	15.894	∞	17.055	14.162
E(Y_{T+100})	15.540	16.374	17.002	∞	18.005	14.610

* Untruncated uniform prior.

** MC means calculated using Monte Carlo integration.

*** Ex means calculated using analytical formulae.

The fact that means and standard deviations are conditional on the observed data has not been made explicit in column 1.

FIGURE 1:
PREDICTIVE MEANS

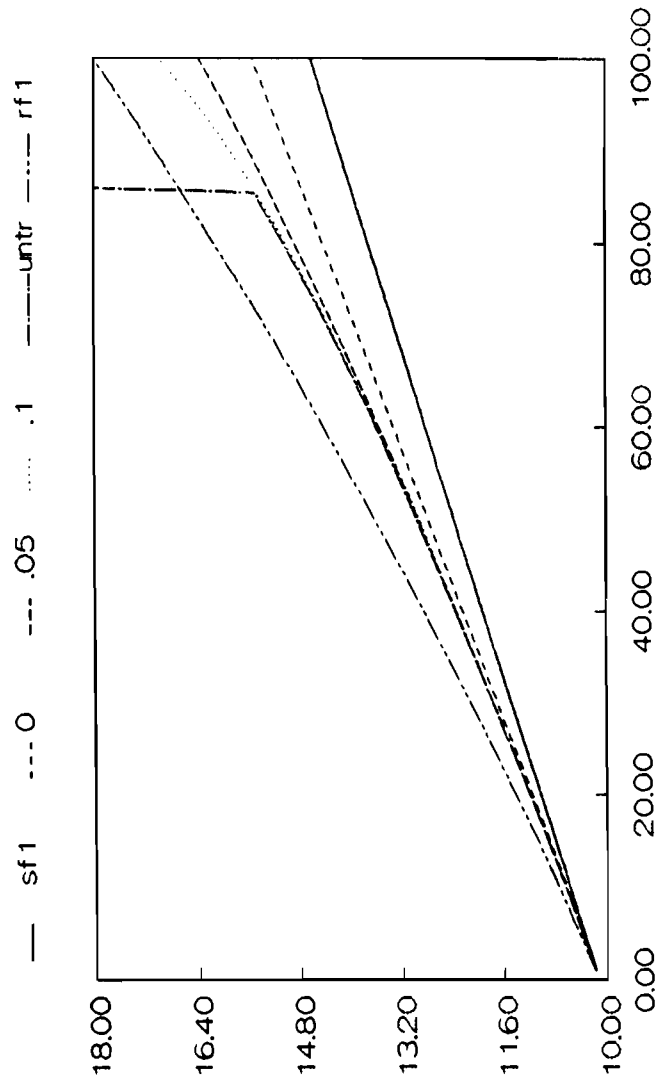


FIGURE 2:
PREDICTIVE STANDARD DEVIATIONS

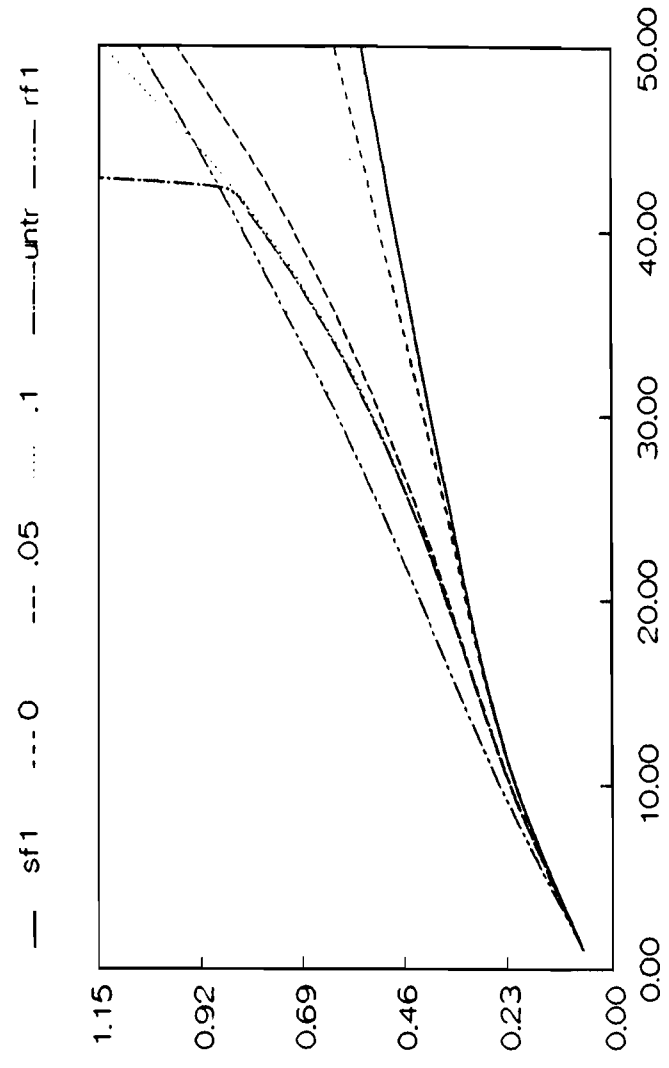


FIGURE 3:
PREDICTIVE DENSITIES $n=42$
..... $d=0$ $d=0.05$ -----untrunc.
uniform

