Bayesian Method of Moments Analysis of Time Series Models with an Application to Forecasting Turning Points in Output Growth Rates

A. Zellner, J. Tobias and H. Ryu, *

October 15, 1998

Abstract

Bayesian method of moments (BMOM) analyses of central time series models are presented. These include derivations of post data densities for parameters, predictive densities for future observations and relative expected losses associated with alternative model specifications, *e.g.* a unit root versus a non-unit root AR(1) process or an AR(1) versus higher order AR processes. BMOM results are compared with those provided by traditional Bayesian and non-Bayesian approaches. An application to forecasting turning points in 18 countries' annual output growth rates, 1980-1995 is provided using several variants of an autoregressive leading indicator model. Optimal forests include not only forecasts of dichotomous outcomes, *e.g.* downturn or no downturn, as in previous work, but also trichotomous outcomes, *e.g.*, minor downturn, major downturn or no downturn or minor upturn, major upturn or no upturn. Empircal results indicate that about 70 percent of dichotomous outcomes are forecasted correctly, in line with previous results obtained using earlier data for the period 1974-1986 for the same 18 countries. A summary of results and some comments on future research are provided.

Key words: Time series analysis; Bayesian analysis; Bayesian method of moments; turning point forecasts; model selection.

*This paper will appear in *Estadistica*, 1999. The authors are affiliated with the University of Chicago, University of Chicago, and Chung Ang University, respectively. Research financed in part by the National Science Foundation and by income from the H.G.B. Alexander Endowment Fund, Graduate School of Business, University of Chicago.

1 Introduction

In this paper, we present Bayesian Method of Moments (BMOM) analyses of central time series models and then use variants of an autoregressive leading indicator (ARLI) model to make forecasts of output growth rates for 18 industrialized economies. See Currie (1996), Green and Strawderman (1996), Tobias and Zellner (1997), Zellner (1994,1995,1997a,b), Zellner and Sacks (1996), Zellner, Min, Dallaire and Currie(1994) and Zellner, Tobias and Ryu (1997) for previous work on the theory and application of the BMOM.

The BMOM is particularly useful for obtaining post data moments and densities for parameters and future observations when the form of the likelihood function is unknown and thus a traditional Bayesian (TB) approach can not be used. Also, even when the likelihood function's form is assumed known, in time series problems there is sometimes difficulty in formulating an appropriate prior density; see, e.g. Phillips (1991), Min and Zellner(1993) and Zellner (1977, 1997c) for a review and discussion of alternative priors for parameters of AR(1) and AR(2) processes. Here, we shall show how the BMOM approach in connection with a conceptual sample yields a prior density for parameters of AR processes. As in previous work with binomial, multiple and multivariate regression, semiparametric regression and simultaneous equations models, it will be shown how the BMOM approach can be employed to analyze time series models when the form of the likelihood function and/or prior density is unknown. In addition, BMOM procedures are developed for unit root and other testing problems as well as for general model selection problems. Also, AR models containing lagged leading indicator variables or ARLI models, employed in previous forecasting work, *e.g.* Garcia-Ferrer et al (1987), Le Sage (1996), Min and Zellner (1993, 1997), Zellner, Hong and Gulati (1990) and Zellner, Hong and Min (1991), will be utilized to provide optimal turning point forecasts for 18 industrialized countries' annual output growth rates, 1981-1995 using several loss structures. The results, based on a broader, revised data set, are compared with earlier results obtained in the works cited above.

The plan of the paper is as follows. In Section 2, we derive BMOM results for AR and ARLI models and compare them with those provided by traditional Bayesian and non-Bayesian approaches. Loss function and post data odds approaches for model selection are employed to test for unit roots and to select among alternative AR and ARLI models. Computed examples are provided. Then in Section 3, turning point forecasting techniques are described and it is shown how to incorporate "add factors" in a BMOM analysis as well as how to utilize an elaborated loss structure that allows for choices of various types of turing points, *e.g.* minor downturn, major downturn, no downturn, etc. Section 4 is devoted to reporting the results of applications of the BMOM procedures discussed above while in Section 5, a summary and some concluding remarks are presented.

2 BMOM Analysis of AR and ARLI Models

Herein, we consider models of the following forms:

$$y_t = \rho_1 y_{t-1} + u_t \tag{1}$$

$$y_t = \alpha + \rho_1 y_{t-1} + u_t \tag{2}$$

$$y_t = \alpha + \gamma t + \rho_1 y_{t-1} + u_t \tag{3}$$

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + u_t \tag{4}$$

$$y_t = \Phi(L) y_{t-1} + z_t' \theta + u_t, \tag{5}$$

where $t = 1, 2, \cdots T$, y_t is the observed value of an output variable, u_t is a realized error term, z_t is a vector of observed input variables, α, γ, θ, ρ_1 and ρ_2 are parameters unknown in value, and $\Phi(L)$ is a finite polynomial lag operator with parameters that are fixed and unknown in value. As in other traditional Bayesian analyses, *e.g.* Chaloner and Brant (1988) and Zellner and Moulton (1985), the parameters and realized errors are considered subjectively random. A "generic" representation of the above models is given by

$$y = X\beta + u, \tag{6}$$

where y is a $T \times 1$ vector of given values, X is a $T \times k$ full column rank matrix composed of given values of lagged $y's$ and other input vairables, β is a $k \times 1$ vector of parameters with unknown values and u is a $T \times 1$ vector of realized error terms. By appropriate definition of X and β, (6) provides a representation of each of the relations in (1-5).

2.1 Post Data Moments of Parameters and Future Observations

Given our assumption that the form of the likelihood function is unknown, it is the case that maximum likelihood and traditional Bayesian procedures can not be used to analyze the relations for the data $D = (y, X)$ in (6). However, (6) can be analyzed in a BMOM approach to yield moments of parameters and future values of observations as explained in the references above by use of two assumptions, the first of which is:

Assumption I : $X'E(u \mid D) = 0$,

where E denotes the subjective expectation operator and $D = (y, X)$ is the given data. Note that Assumption I indicates that the columns of X are orthogonal to the post data expectation of u, $E(u \mid D)$, that is that there is nothing "systematic" in $E(u \mid D)$ that is correlated with the variables in X. Also note from (6), $y = XE(\beta \mid D) + E(u \mid D)$ and Assumption I implies that y is represented by the sum of two orthogonal vectors, $XE(\beta \mid D)$ and $E(u \mid D)$. Finally, if there is an intercept in the relation (2), that is X has one column that consists of just ones, denoted by e, then $e'E(u \mid D)/T = E(\overline{u} \mid D) = 0$, where $\overline{u} = 1/T \sum_{t=1}^{T} u_t$.

We can utilize Assumption I to obtain the first post data moments of β and u as follows. From (6), with $\hat{\beta} = (X'X)^{-1}X'y$, we have $\hat{\beta} = \beta + (X'X)^{-1}X'u$, and on taking the post data expectation of both sides of this last equation and using Assumption I, we have

$$E(\beta \mid D) \quad = \quad \hat{\beta} = (X'X)^{-1}X'y. \tag{7}$$

That is, the first post data moment of β is $\hat{\beta}$, the least squares estimate. Also, from (6), using (7),

$$E(u \mid D) \quad = \quad y - X\hat{\beta} = \hat{u}. \tag{8}$$

Thus, the post data mean of the realized error vector is $\hat{u}$, the least squares residual vector. Further, the post data mean of a future, as yet unobserved value of the dependent variable, say $y_{T+1} = x'_{T+1}\beta + u_{T+1}$ is given by $E(y_{T+1} \mid D) = x'_{T+1}E(\beta \mid D) + E(u_{T+1} \mid D)$. If it is appropriate to assume that $E(u_{T+1} \mid D) = 0$, then

$$E(y_{T+1} \mid D) = x'_{T+1}E(\beta \mid D) = x'_{T+1}\hat{\beta}, \tag{9}$$

using (7), which is the least squares forecast. For y_{T+2}, we have $y_{T+2} = x'_{T+2}\beta + u_{T+2}$, where x_{T+2} contains the as yet unobserved quantity y_{T+1} as well as other variables. Then the conditional post data predictive mean of y_{T+2} given y_{T+1} and the other variables in x_{T+2} (which we include in D), is

$$E\left(y_{T+2} \mid y_{T+1}, D\right) \quad = \quad x'_{T+2}E(\beta \mid y_{T+1}, D) + E(u_{T+2} \mid y_{T+1}, D).$$

Given the assumption that this last expectation is zero, the conditional mean of y_{T+2} given y_{T+1} is available. Note that $E(\beta \mid y_{T+1}, D)$ is the least squares quantity based on $T+1$ observations where y_{T+1} has not as yet been observed. Similarly, the first moment of y_{T+3} given y_{T+2}, y_{T+1} and D can be derived given a zero mean assumption for the future error term, u_{T+3} and similarly for other future values of y. These conditional means along with conditional second moments of future values of y, derived below, can be utilized as side conditions in deriving maxent conditional predictive densities. Draws can be made from these marginal and conditional densities for future y's to obtain their marginal densities and marginal moments.[1] Of course, in this approach the future values of any input variables are required.

To obtain second order moments of β and y_{T+1}, the following assumption regarding the second moment matrix of the realized vector u has been utilized in past work:

Assumption II: $\text{Var}(u \mid \sigma^2, D) = E\left[(u - E(u \mid D))(u - E(u \mid D))'\right] = \sigma^2 X(X'X)^{-1}X'$,

where σ^2 is a positive parameter to be defined below. Note that since the elements of u satisfy the T equations in (6) and X is of rank k, there are only k free elements of u and thus $\text{Var}(u \mid D)$ should be of rank k. Further $\text{Var}(u \mid D)$ has to satisfy the following functional equation: $\text{Var}(u \mid D) = X(X'X)^{-1}X'\text{Var}(u \mid D)X(X'X)^{-1}X$. The assumed form for $\text{Var}(u \mid D)$ in Assumption II satisfies this functional equation. Since

$$E\left((\beta - \hat{\beta})(\beta - \hat{\beta})' \mid D\right) = (X'X)^{-1}X'E\left[(u - \hat{u})(u - \hat{u})' \mid D\right]X(X'X)^{-1},$$

where $E(u \mid D) = \hat{u}$ and $X'\hat{u} = 0$ have been utilized, use of Assumption II yields:

$$V(\beta \mid \sigma^2, D) \quad = \quad E\left[(\beta - \hat{\beta})(\beta - \hat{\beta})' \mid \sigma^2, D\right] = \sigma^2(X'X)^{-1} \tag{10}$$

which is the post data variance-covariance matrix for β given D and the positive scale parameter, σ^2. As in previous work, we define the parameter σ^2 as follows: $\sigma^2 \equiv 1/T\sum_{t=1}^{T} u_t^2 = u'u/T$. We note

$$E(u'u \mid D) \quad = \quad TE(\sigma^2 \mid D) = (y - X\hat{\beta})'(y - X\hat{\beta}) + E\left[(\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) \mid D\right]$$

[1] To see this, suppose we are interested in the marginal density for y_{T+2} and assume that we have the marginal density for y_{T+1}. Observe that $p(y_{T+2}) = \int p(y_{T+2} \mid y_{T+1})p(y_{T+1})dy_{T+1}$. Thus, by taking a draw from the marginal density of y_{T+1}, say y_{T+1}^0 and using this value to draw from the conditional density, $p(y_{T+2} \mid y_{T+1}^0)$, we obtain a draw from the marginal for y_{T+2}. Given these draws, we can obtain estimates of the marginal density of y_{T+2} or compute moments from the draws. This procedure can also be employed to derive the marginal density for a j-period ahead, as yet unrealized value of the dependent variable, y_{T+j}.

$$
\begin{aligned}
&= \hat{u}'\hat{u} + \mathrm{tr}\left[X'XE\left\{(\beta-\hat{\beta})(\beta-\hat{\beta})' \mid D\right\}\right] \\
&= \hat{u}'\hat{u} + kE(\sigma^2 \mid D),
\end{aligned}
$$

where the first line is obtained by noting $u = y - X\beta$ and $X'\hat{u} = 0$, and the third line is obtained after substituting the variance of β given in (10). We can solve the above expression for $E(\sigma^2 \mid D)$ to obtain

$$
E(\sigma^2 \mid D) \quad = \quad \hat{u}'\hat{u}/(T-k) = s^2. \tag{11}
$$

Then using (10) and (11) gives

$$
\mathrm{Var}(\beta \mid D) = s^2(X'X)^{-1} \tag{12}
$$

as the unconditional post data covariance matrix for β .

Assumption II permits us to derive second moments for future, as yet unobserved observations given that the future, as yet unobserved error term, u_f, has mean zero, variance σ^2, and is uncorrelated with β. Then,

$$
\mathrm{Var}(y_{T+1} \mid \sigma^2, x_{T+1}, D) \quad = \quad x'_{T+1}\mathrm{Var}(\beta \mid D)x_{T+1} + \sigma^2 = \sigma^2[1 + x'_{T+1}(X'X)^{-1}x_{T+1}] \tag{13}
$$

and using (14) in conjunction with (11),

$$
\mathrm{Var}(y_{T+1} \mid x_{T+1}, D) = s^2[1 + x'_{T+1}(X'X)^{-1}x_{T+1}]. \tag{14}
$$

2.2 Post Data Densities for Parameters and Future Observations

As noted in previous work, *e.g.* Soofi (1996) and Zellner (1994, 1997a-c), the above moments can be used as side conditions in deriving proper maxent densities for regression parameters as well as post data predictive densities that incorporate the information in side conditions as conservatively as possible. For example, using the first and second moments of y_{T+1} in (9) and (14), the proper maxent density which satisfies these moment side conditions is a normal density, which we denote as BMOM(N). As shown in Tobias and Zellner (1997), these predictive densities can be utilized along with new data to calculate Bayes' factors that are used with prior odds for model comparison and selection purposes. In some cases, we may wish to use additional information to supplement the information in the data. In this connection, as in Zellner (1994), a conceptual sample, say $y_c = X_c\beta + u_c$, (where y_c and u_c are $T_c \times 1$ vectors, and X_c is a $T_c \times k$ matrix), can be introduced and analyzed using the above methods and assumptions to yield, for example, a maxent normal prior density, $N(\hat{\beta}_c, s_c^2(X_c'X_c)^{-1})$, where $\hat{\beta}_c = (X_c'X_c)^{-1}X_c'y_c$, $X_c'X_c$, T_c and $s_c^2 = (y_c - X_c\hat{\beta}_c)'(y_c - X_c\hat{\beta}_c)/(T_c - k)$ are assigned values by the investigator. For example, such a prior can be employed to introduce information that a parameter's value is probably between -1 and +1. Further the above conceptual sample can be combined with the actual sample as follows:

$$\begin{bmatrix} y_c \\ y \end{bmatrix} = \begin{bmatrix} X_c \\ X \end{bmatrix}\beta + \begin{bmatrix} u_c \\ u \end{bmatrix}, \tag{15}$$

or,

$$w = W\beta + v, \tag{16}$$

where $w' = (y_c'\ y')$, $W' = (X_c'\ X')$, and $v = (u_c'\ u')$.

When the analysis, based on Assumptions I and II is applied to (16), the results are

$$\begin{aligned} E(\beta \mid D_c) &= (W'W)^{-1}W'w \\ &= (X_c'X_c + X'X)^{-1}(X_c'y_c + X'y) \\ &\equiv \hat{\beta}_*, \end{aligned} \tag{17}$$

where D_c denotes the given and conceptual samples and the c subscript denotes the use of an informative conceptual sample. Then,

$$\mathrm{Var}(\beta \mid D) = s_*^2(W'W)^{-1}, \tag{18}$$

with $s_*^2 = (w - W\hat{\beta}_*)'(w - W\hat{\beta}_*)/(T_c + T - k)$. Then, using (17) and (18) as side conditions, the proper maxent density for β is $N(\hat{\beta}_*, s_*^2(W'W)^{-1})$.

2.3 Hypothesis Testing

In the previous section, we described how BMOM can be used to obtain finite sample post data densities for parameters as well as predictive densities without the use of an assumed likelihood function, prior density, or Bayes' Theorem. We will now discuss how to perform tests of hypotheses.

In this section we consider a time series model of the form

$$y = X_1\beta_1 + X_2\beta_2 + u, \tag{19}$$

where y is a $T \times 1$ vector, X_1 is a $T \times k_1$ matrix, X_2 is a $T \times k_2$ matrix, and β_1 and β_2 are $k_1 \times 1$ and $k_2 \times 1$ vectors, respectively. Here, lagged $y's$ and other variables can be included in the $X's$ as in (5). We consider two hypotheses, $H_0 : \beta_2 = 0$ and $H_1 : \beta_2 \neq 0$. For convenience, we reparameterize the model as follows

$$y = X_1\phi + \hat{V}\beta_2 + u, \tag{20}$$

where $\hat{V} = X_2 - X_1\hat{\Pi}$, with $\hat{\Pi} = (X_1'X_1)^{-1}X_1'X_2$ and $\phi = \beta_1 + \hat{\Pi}\beta_2$. Note that by construction, $X_1'\hat{V} = 0$. Further, using the BMOM regression approach applied to (20), we have $E(\beta_2 \mid D) = (\hat{V}'\hat{V})^{-1}\hat{V}'y$ and $\mathrm{V}(\beta_2 \mid D) = s^2(\hat{V}'\hat{V})^{-1}$, where $s^2 = \hat{u}'\hat{u}/(T - k_1 - k_2)$ and $\hat{u}$ is the least squares residual vector.

If our interest centers on the value of β_2, namely is it equal to zero or not, we can consider use of the following quadratic loss function,

$$L = (\beta_2 - \hat{\beta}_2)'\hat{V}'\hat{V}(\beta_2 - \hat{\beta}_2). \tag{21}$$

Now we evaluate expected loss given that $\beta_2 = 0$ to obtain

$$E(L \mid \beta_2 = 0) = P_0\hat{\beta}_2'\hat{V}'\hat{V}\hat{\beta}_2, \tag{22}$$

and similarly, expected loss under the alternative hypothesis that $\beta_2 \neq 0$ is given as

$$\begin{aligned} E(L \mid \beta_2 \neq 0) &= P_A\left[E\left[(\beta_2 - \hat{\beta}_2)'\hat{V}'\hat{V}(\beta_2 - \hat{\beta}_2) \mid D\right] + L_\alpha\right] \\ &= P_A\left[k_2 s^2 + L_\alpha\right]. \end{aligned} \tag{23}$$

where L_α is an added loss term, assigned by the investigator, which serves as a penalty for model complexity, P_0 is the probability associated with the hypothesis $\beta_2 = 0$, P_A is the probability associated with the hypothesis $\beta_2 \neq 0$, and $\beta_2 \sim N(\hat{\beta}_2, s^2(\hat{V}'\hat{V})^{-1})$ has been used to evaluate the expectation in (23). Thus, the ratio of expected losses, denoted by C_{12}, is just

$$C_{12} = \frac{P_0\hat{\beta}_2'\hat{V}'\hat{V}\hat{\beta}_2}{P_A(k_2 s^2 + L_\alpha)} = \frac{P_0 F}{P_A(1+z)}, \tag{24}$$

where $z \equiv L_\alpha/k_2 s^2$, and F is the traditional sampling theory $F-$statistic for testing $H_0 : \beta_2 = 0$. It is interesting to note that the value for L_α can be chosen so that the decisions under the rule of minimizing expected losses can lead to decisions that are identical to those provided by certain sampling theory approaches. For time series problems this method can be used to determine the appropriate order of an autoregression, or to test other hypotheses such as the presence of a unit root. Finally, other methods for comparing models such as post data odds and comparison of expected predictive losses can also be employed.[2]

To illustrate the above procedure, we consider the hypothesis $H_1 : \rho_1 = 1$ and $H_2 : \rho_1 \neq 1$ in connection with the model in (2), with $t = 1, 2, \cdots, T$. Using quadratic loss, $L = (\rho_1 - \hat{\rho}_1)^2$, where $\hat{\rho}_1$ is the postdata mean of ρ_1, the least squares estimate, and the BMOM(N) model, we have for the ratio of expected losses $C_{12} = (P_1/P_2)[(\hat{\rho}_1 - 1)^2/s^2_{\rho_1}][1/(1+z)]$, where $s^2_{\rho_1} = s^2(y'_{-1}y_{-1})^{-1}$ and $z = L_\alpha/s^2_{\rho_1}$. Note that $(\hat{\rho}_1 - 1)^2/s^2_{\rho_1}$ is just the square of the usual "t-statistic" that is in widespread use in sampling theory tests for unit roots and also appears in traditional Bayesian posterior odds expressions relating to the hypothesis that $\rho = 1$ versus $\rho \neq 1$ in a normal AR(1) process; see *e.g.* Zellner and Plosser (1976) and Manas-Anton (1986). It is thus straightforward to compare expected losses in this case and to choose the hypothesis with the lower or lowest expected loss.

In the next section we make use of the BMOM theory developed in this section and apply the techniques to the problem of forecasting turning points in output growth rates. We also employ the expected loss criterion described above to determine the variables to include in a model of output growth rate determination.

3 BMOM and Traditional Bayes Methods for Forecasting Turning Points

In the previous sections we have discussed how BMOM can be applied to analyze time series problems. In this section, we apply both BMOM as well as traditional Bayesian techniques to the problem of forecasting turning points in output growth rates. In previous work, Zellner and Hong (1989), and Zellner, Hong and Min (1991) have developed a Bayesian decision theoretic procedure and have used it with several variants of an autoregressive-leading indicator model to forecast correctly about 70 percent of 158 turning points in 18 countries' annual output growth rates for the period 1974-1986, using data for 1954-1973 to fit the models which were then sequentially updated year by year in the "holdout" sample, 1974-86. In our current application, we apply such procedures and models using revised annual data, 1948-80, to fit our models and 1981-95 as a "holdout sample" that is used to forecast 192 turning points in 18 countries' annual output growth rates. In addition,

[2]See Tobias and Zellner (1997) and Zellner, Tobias and Ryu (1997).

we develop procedures that permit forecasts of various kinds of turning points, *e.g.* a minor or a major downturn or a minor or a major upturn.

We begin by introducing our definition of a turning point.[3] We call y_{iT+1}, the growth rate in output for country i in year $T+1$, a **downturn (DT)** if

$$y_{iT-1}, y_{iT-2} \leq y_{iT} \ \text{ and } \ y_{iT} > y_{iT+1},$$

and y_{iT+1} is **not a downturn (NDT)** if

$$y_{iT-1}, y_{iT-2} \leq y_{iT} \ \text{ and } \ y_{iT} \leq y_{iT+1}.$$

Similarly, y_{iT+1} is an **upturn (UT)** if

$$y_{iT-1}, y_{iT-2} \geq y_{iT} \ \text{ and } \ y_{iT} < y_{iT+1},$$

and is **not an upturn (NUT)** if

$$y_{iT-1}, y_{iT-2} \geq y_{iT} \ \text{ and } \ y_{iT} \geq y_{iT+1}.$$

With these definitions in hand, we can proceed to forecast outcomes for future, as yet unobserved growth rates, provided that the conditions for a turning point described above are satisfied. In past work, Zellner and Hong (1989), and Zellner, Hong and Min (1991) described how turning points can be forecasted using a traditional Bayesian approach. We now briefly review this forecasting procedure and also discuss how to produce turning point forecasts using the Bayesian Method of Moments.

Given the definitions of turning points above, the problem of forecasting a turning point arises when we have a sequence of output growth rates satisfying one of the definitions above and have to forecast whether an as yet unobserved growth rate, y_{iT+1}, is below or above a most recently observed growth rate, y_{iT}. Given a predictive density for y_{iT+1}, the needed probabilities for obtaining an optimal turning point forecast are readily available. For example, using traditional Bayesian techniques, we can obtain the predictive density by noting the following:

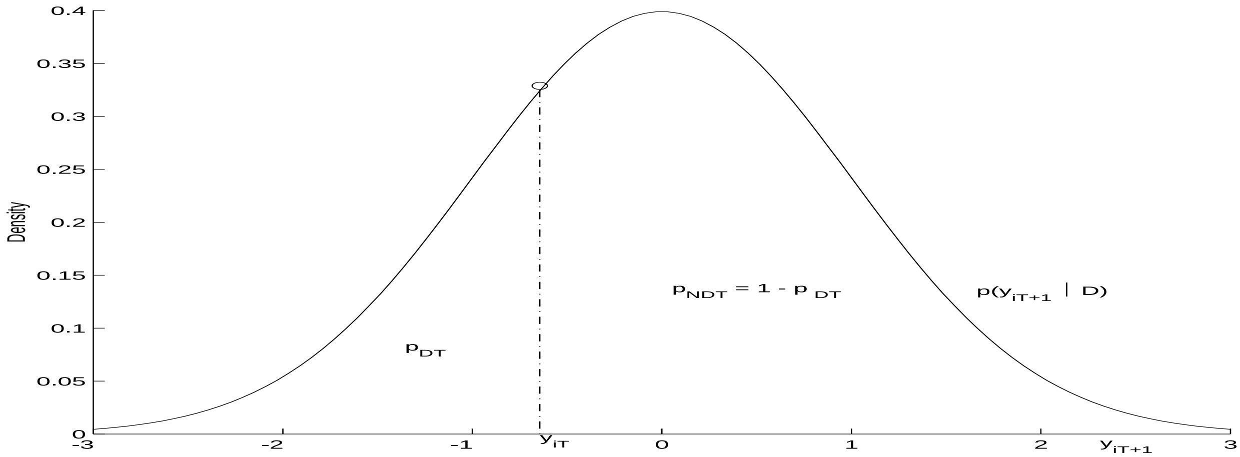
$$p(y_{iT+1} \mid D) = \int_{\Theta} p(y_{iT+1} \mid \theta, D)\pi(\theta \mid D)d\theta,$$

where $p(y_{iT+1} \mid \theta, D)$ is the conditional pdf for y_{iT+1} given the parameter vector $\theta \in \Theta$, and $\pi(\theta \mid D)$ is the posterior pdf for the parameter vector θ. Given this result, the probability of say, a downturn

[3]Note that other definitions of turning points could also be entertained.

(for a sequence of growth rates that satisfy the definition) is given as

$$p_{DT} = \Pr(y_{iT+1} < y_{iT} \mid D) = \int_{-\infty}^{y_{iT}} p(y_{iT+1} \mid D) dy_{iT+1}.$$



Calculation of DT and NDT Probabilities in year $T+1$

With the probabilities p_{DT} and $p_{NDT} = 1 - p_{DT}$ computed, we can use 2×2 loss structure to obtain an optimal turning point forecast, namely one which minimizes expected loss. With a symmetric loss structure, which has been employed in past work, we shall forecast a DT if p_{DT} exceeds 1/2, and otherwise forecast NDT. Similarly, we will forecast UT if p_{UT} exceeds 1/2, and otherwise forecast NUT.

We can also obtain probabilities such as p_{DT} above using the BMOM approach. For example, in Section 2.1 we derived a normal maxent density for a future growth rate using the moments shown in equations (9) and (14). This normal maxent predictive density can be used to compute the probabilities of UT/NUT, or DT/NDT, just as in the TB case described above. With a symmetric 2×2 loss structure, the forecasts using the BMOM normal predictive density will be exactly the same as those provided by a traditional Bayesian approach using a diffuse prior and usual normal likelihood function derived from an iid normal sampling assumption for the error terms. In this case, both predictive densities will be symmetric about the same mean $\hat{y}_{T+1} = x'_{T+1}\hat{\beta}$, and thus the turning point forecasts in this case will be identical.[4] Of course, other BMOM predictive densities can be utilized to yield different turning point forecasts. For example, if we utilize higher-order

[4] If the loss function is asymmetric, BMOM and TB turning point forecasts will not be identical.

moments of y_{T+1}, we can obtain, say an exponential quartic predictive density for y_{T+1}. The use of an exponential quartic density will allow for possible bimodality, skewness, or kurtosis, and computed probabilities based on the exponential quartic density may produce forecasts that are quite different from traditional Bayesian forecasts. This issue is taken up in section 3.3.

3.1 Forecasting Major and Minor Turning Points in Output Growth Rates

In the previous section, we considered only two possibilities associated with each turning point - either DT or NDT can occur or either UT or NUT can occur. It is of interest to expand these alternatives to include a wider range of possible outcomes, namely, a minor DT, major DT, or NDT and, similarly, a minor UT, major UT, or NUT. We can differentiate between major and minor turning points with a slight modification of our definitions given in the previous section. Without loss of generality, we focus on prediction of major and minor downturns - the theory developed for this case easily generalizes to the case of forecasting major and minor upturns.

We call y_{iT+1} a **major downturn (MADT)** if the sequence of output growth rates for country i satisfies the following inequalities: [5]

$$y_{iT-2}, y_{iT-1} \leq y_{iT} \;\text{ and }\; y_{iT}\left(1-(1/2)z_{iT}\right) > y_{iT+1},$$

where $z_{iT} = 1$ if $y_{iT} \geq 0$, and $z_{iT} = -1$ if $y_{iT} < 0$. Similarly, we call y_{iT+1} a **minor downturn (MIDT)** if the growth rates satisfy the following inequalities:

$$y_{iT-2}, y_{iT-1} \leq y_{iT} \;\text{ and }\; y_{iT} > y_{iT+1} \geq y_{iT}\left(1-(1/2)z_{iT}\right).$$

Finally, y_{iT+1} is **not a downturn (NDT)** if the output growth rates satisfy the following inequalities:

$$y_{iT-2}, y_{iT-1} \leq y_{iT} \;\text{ and }\; y_{iT} \leq y_{iT+1}.$$

In the above definitions, the choice of 1/2 in defining major and minor downturns is somewhat arbitrary. If other values are thought to be more appropriate, they can be used without difficulty in what follows. Given a predictive density for y_{iT+1}, we can compute the probabilities of a major DT

[5]Note that other definitions of major and minor DT's can also be entertained. In this approach, we are concerned about the size of the future growth rate relative to the size of the current growth rate. With the definition employed here, if y_t is near zero, then the length of the interval defining a minor DT will be "small", and thus the probability associated with this outcome will also be "small." An alternate approach would define a major DT if the future growth rate is smaller than some fixed distance from the current growth rate, i.e., $y_{it+1} \leq y_{it} - c, \;\; c > 0$.

(denoted p_1), the probability of a minor downturn (denoted p_2), and the probability of no downturn, $1 - p_1 - p_2$ for country i in year $T+1$. We can then choose a loss structure for the problem to determine our forecast. We introduce the following general 3×3 loss structure,[6]

Table 1
Loss Structure for Forecasting Downturns #1

Predicted Outcome	Actual Outcome		
	Major DT	Minor DT	NDT
Major DT	0	c_{12}	c_{13}
Minor DT	c_{21}	0	c_{23}
NDT	c_{31}	c_{32}	0
Probability	p_1	p_2	$1 - p_1 - p_2$

with $c_{ij} > 0 \ \forall \ i, j$. In the next section we consider two alternative specifications for this loss structure and examine the decision rules under each structure.

3.1.1 Alternative Loss Structures and Decision Rules

We first take up the case of equal loss, where $c_{ij} = c$ in Table 1. In this case, the expected loss associated with forecasting a major DT is $c(1 - p_1)$, with forecasting a minor DT is $c(1 - p_2)$, and with forecasting NDT is $c(p_1 + p_2)$. Thus, use of the following decision rules leads to minimum expected loss:

Table 2
Decision Rules Under Loss Structure #1

Forecast Major DT if	$p_1 > p_2$ and $2p_1 + p_2 > 1$
Forecast Minor DT if	$p_2 > p_1$ and $2p_2 + p_1 > 1$
Forecast NDT if	$2\max\{p_1, p_2\} + \min\{p_1, p_2\} < 1$

We note that with the above decision rules, we might not forecast any type of DT even if $\text{Prob}(y_{iT+1} < y_{iT}) > .5$. For example, if $p_1 = .31$ and $p_2 = .29$, we would still forecast NDT. Recall that in the two decision case, with forecasts DT and NDT, use of a 2×2 symmetric loss structure leads us

[6] In Tables 1 and 3, the (i, j) entry gives the loss when the event in row i is predicted and the event in column j actually occurrs.

to forecast a DT if Prob($y_{iT+1} < y_{iT}$) > .5. Hence, the decisions in the three action case do not necessarily preserve the decisions under the 2×2 symmetric loss structure.

We now consider a second symmetric loss structure which does preserve the basic forecasts of the two-action case.[7].

Table 3
Loss Structure for Forecasting Downturns #2

Predicted Outcome	Actual Outcome		
	Major DT	Minor DT	NDT
Major DT	0	c	$2c$
Minor DT	c	0	c
NDT	$2c$	c	0

With this loss function, we penalize "large" misses more severely than small misses. That is, if we forecast NDT when the economy experiences a Major DT, the loss incurred, $2c$, is greater than if we forecasted a Minor DT. For this reason, we expect more Minor DT's to be forecasted than with the equal, symmetric loss function, since the relative cost of forecasting a minor downturn has decreased. We derive the following decision rules under this loss structure

Table 4
Decision Rules Under Loss Structure #2

Forecast Major DT if	$p_1 > 1/2$
Forecast Minor DT if	$p_1 < 1/2$ and $p_1 + p_2 > 1/2$
Forecast NDT if	$p_1 + p_2 < 1/2$

The decision rules under this loss function are clear - we will only forecast the "extreme" outcomes (*i.e.* NDT or Major DT) if the probability of those actions are greater than 1/2. Otherwise we forecast the "safe" outcome, Minor DT. Thus, compared to the equal loss structure, we shall see fewer Major DT and NDT forecasts, offset by more Minor DT forecasts. Finally, note that if Prob($y_{iT+1} < y_{iT}$) > .5, then some type of downturn (either major or minor) will be forecasted, and similarly, if Prob($y_{iT+1} < y_{iT}$) < .5, NDT will be forecasted, consistent with the results in the

[7]It may not always be the case that $c_{13} = c_{31} = 2c$. In some cases, forecasting NDT when the outcome is Major DT may be more serious than forecasting Major DT when the outcome is NDT. For example, if we specify $c_{31} = 4c$ and $c_{13} = 2c$, then we forecast Major DT if $p_1 > 1/2$, Minor DT if $p_1 < 1/2$ and $4p_1 + 2p_2 > 1$, and NDT if $4p_1 + 2p_2 < 1$. Of course, loss structures that are deemed more appropriate for a given situation can also be employed.

two-action case. The computed examples in the following section present the results of forecasting minor and major upturns and downturns for a variety of autoregressive leading indicator (ARLI) models. Both loss structures described above are used. As mentioned earlier, other loss structures or definitions of turning points can also be employed.

3.2 The Model and Results

The basic model that we employ is one of those described and used in Garcia-Ferrer et al. (1987), Zellner and Hong (1989), and Zellner, Hong and Min (1991). The model is given below

$$y_{it} = \alpha_0 + \alpha_1 y_{it-1} + \alpha_2 y_{it-2} + \alpha_3 y_{it-3} + \beta GM_{it-1} + \gamma_1 SR_{it-1} + \gamma_2 SR_{it-2} + \delta WR_{t-1} + u_{it}, \quad (25)$$

where y_{it} is the real growth rate in output, GM_{it} is the growth rate of the real money supply, SR_{it} is the growth rate in real stock prices for country i in year t, and WR_t is a world return variable, a proxy for common "world" effects, which we define as the median of the 18 countries' stock prices (SR_{it}) in year t. Here, we employ a third-order AR to allow for the possibility of having 2 complex roots associated with a cycle and one real root associated with a trend, as found empirically in Hong (1989). The addition of the world return variable has been shown in past work to reduce contemporaneous correlation of the error terms.

In addition to the base model in (25), we also consider models with a time trend and country indicator variables. Another new feature that we employ in the application of the following section is the inclusion of "add factors." After observing the past data, new information may arrive which shapes our expectations about the future output growth rates. We can modify our assumptions about future error terms' properties, *e.g.* their means, to take account of such information which will result in shifts of predictive densities. In the section to follow, we apply the techniques of this section and present our forecasting results.

3.3 BMOM Exponential Quartic Predictive Densities

As mentioned earlier in the paper, the forecasting results using BMOM(N) and TB based on a diffuse prior and normal likelihood will be practically identical given the relatively large sample of observations used in this application. Thus, it would be interesting to introduce a model produced using the Bayesian Method of Moments that will give different, and possibly improved forecasts. To this end, we derive some higher-order moments for a future value of the dependent variable, y_{iT+1}, which can be incorporated in the predictive density to introduce departures from the symmetric normal density. Specifically, if we assign values for skewness and kurtosis, then we can solve for the associated third and fourth moments of y_{iT+1}.

To see how the third and fourth moments can be obtained, note that

$$\text{Skewness} \equiv s_k \equiv \frac{E\left[(y_{iT+1} - \mu_{iT+1})^3\right]}{\sigma^3_{iT+1}} \quad \text{and} \quad \text{Kurtosis} \equiv \kappa \equiv \frac{E\left[(y_{iT+1} - \mu_{iT+1})^4\right]}{\sigma^4_{iT+1}}.$$

If $s_k > 0$ the future value is more likely to lie above the mean, and if $s_k < 0$, the future value is more likely to lie below the mean[8]. The value for kurtosis generally controls the thickness of the tails. If kurtosis exceeds three, then the predictive density will be more heavy-tailed than the normal, and similarly, if kurtosis is less than three, than the predictive density will usually come down more quickly in the tails than the normal. Some intuition may guide our choice of values for skewness and kurtosis. If the expected growth rate is very large and not far from the "full employment growth rate," then a growth rate far below the mean may be much more probable than a growth rate far above the mean, and thus the predictive densities are negatively skewed. Conversely, if the expected future growth rate is very small, then a growth rate greater than the mean may be more probable than a growth rate below the mean, implying that the predictive density is positively skewed. Use of a symmetric BMOM or TB predictive density precludes these possibilities. In our application, we choose values for skewness and kurtosis as follows:

Condition on Predictive Mean, μ_{iT+1}	Assigned Skewness and Kurtosis
$\mu_{iT+1} > 5$	s_k = -1.5 and κ = 5.2
$2 < \mu_{iT+1} < 5$	s_k = -.5 and κ = 4
$0 < \mu_{iT+1} > 2$	s_k = 0 and κ = 3

A similar definition is applied for $\mu_{iT+1} < 0$, only the sign on the assigned skewness value is positive. Thus, as shown above, when the mean growth rate is between 0 and 2 percentage points, the predictive density has a symmetric normal shape. In the cases that the mean is between 2 and 5 or greater than 5, the predictive densities are negatively skewed with kurtosis different from 3, the value for the normal predictive density.[9] In the figure below we present a graph of alternative

[8] That is, more likely relative to the normal case in which skewness and all odd moments about the mean are zero.
[9] Of course, other values for s_k and κ can be utilized if they are thought to be more appropriate.

BMOM densities, each with mean = 3 and variance = 2. The values of skewness and kurtosis in this figure are the same as those employed in our empirical analysis in the following section.

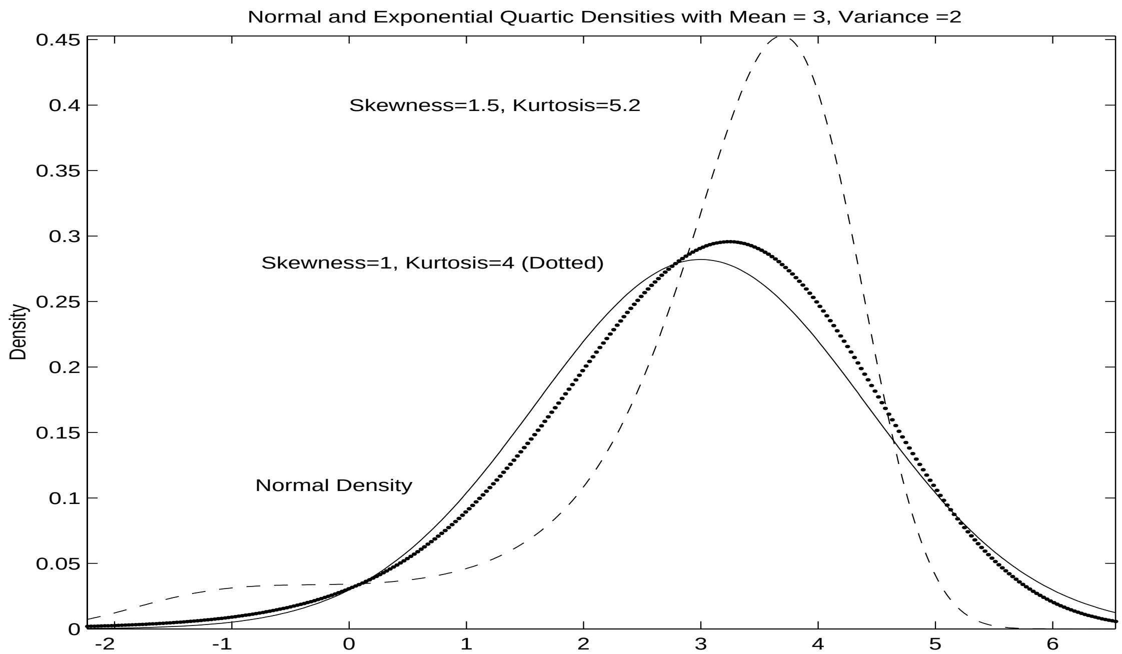


Figure1: Normal and Exponential Quartic Predictive Densities Each With Mean = 3 and Variance = 2

Given values for skewness and kurtosis measures above, we can solve for the third and fourth moments of y_{iT+1} as follows

$$E(y_{iT+1}^3 \mid \sigma_{iT+1}, s_k) = s_k \sigma_{iT+1}^3 + 3\sigma_{iT+1}^2 \mu_{iT+1} + \mu_{iT+1}^3,$$

and

$$E(y_{iT+1}^4 \mid \sigma_{iT+1}, \kappa) = \kappa \sigma_{iT+1}^4 + 4E(y_{iT+1}^3)\mu_{iT+1} - 6\sigma_{iT+1}^2 \mu_{iT+1}^2 - 3\mu_{iT+1}^4.$$

The values μ_{iT+1} and σ_{iT+1}^2 are the mean and variance for the future output growth rate, and can be obtained as in section 2.1. We can take expectations over σ_{iT+1} to get the unconditional third and fourth moments, using the methods discussed in Tobias and Zellner (1994).[10] With the first four moments in hand, the proper maxent density which satisfies these four moment restrictions is

[10] We use the approximate result $E(\sigma^j) = s^j$ in our applications. For a relatively large sample, this approximation will be quite accurate.

an exponential quartic density. The Lagrange multipliers associated with the five side conditions must be determined to obtain a proper predictive density. To do this, we employ the linearization method described in Zellner and Highfield (1988). Thus, for each value in the hold-out portion of our sample, we determine the appropriate moment conditions and solve for the exponential quartic maxent density. This density is then used to obtain the probability of DT, MADT, *etc.* as discussed in the previous sections.

4 Application to Forecasting Turning Points

The data employed in our calculations come from the IMF International Financial Statistics Database and cover the years 1948 - 1995.[11]. We use annual data for the following 18 countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Netherlands, Norway, Spain, Sweden, Switzerland, UK and the US. Real GDP, used as the output measure for each country, was logged and first-differenced to yield annual growth rates. Similarly, nominal money and stock prices were deflated by an annual price index, and then logged and first-differenced to create real growth rates for each country and each year. Data from 1948-1980 were used to estimate the models, and data for the remaining years were used to forecast the turning points. After forecasting turning points in a particular year, we employ the new observation and those preceding the new turning point epsiode to re-estimate our model. For the TB approach, we used a normal likelihood function and a diffuse prior for the parameters. In the BMOM approach, just assumptions I and II, explained above were employed and no sampling assumptions needed to formulate a likelihood function were introduced. Notice that we are considering the coefficients to be the same for all countries, a case of complete "shrinkage." See Zellner and Hong (1989) and Zellner, Hong and Min (1991) for earlier turning point forecasting results.

We begin with the results in Table 5 in which post data coefficient means and standard deviations associated with the BMOM(N) model are presented for the model in (25) and also for models that include a time trend and both a time trend and country indicator variables. Note that the posterior coviariance matrix for the coefficient vector using a TB approach based on a diffuse prior and normal

[11]We have found some differences between the revised IMF data used in this analysis and the earlier IMF data employed in previous analyses. For example, using our definition of a DT we do not find a downturn in the U.S. growth rate in 1989 whereas in the earlier IMF data, there is a downturn in 1989. In the revised data set, a downturn is almost experienced in 1989, but the growth rate in 1986 is slightly larger than the growth rate in 1988. However, most of the turning points found in the previous data set also appear in the current, revised IMF data. In addition, we do not have data for Germany for the years before 1970 and thus use data for 1971-1995 for Germany.

likelihood function is related to the BMOM post data coefficient covariance matrix as follows: Var (TB) $= v/(v-2)$Var (BMOM(N)) , where $v = T - k$ and Var (BMOM(N)) $= s^2(X'X)^{-1}$. These results indicate the relation between post data standard deviations for coefficients obtained in TB and BMOM(N) approaches.

Table 5: BMOM Estimation of ARLI Models Using Annual Data:[12] : 1952-1995 for 18 Countries

	ARLI		ARLI + Time Trend (TT)		ARLI + Time Trend + Country Indicators	
Variable	Coeff. Mean	Std. Dev	Coeff. Mean	Std. Dev	Coeff. Mean	Std. Dev
Constant	1.54	.203	151	20.5	170	21.2
y_{-1}	.367	.039	.286	.039	.249	.040
y_{-2}	-.043	.042	-.088	.041	-.112	.041
y_{-3}	.160	.036	.108	.035	.078	.036
SR_{-1}	.025	.008	.028	.008	.029	.008
SR_{-2}	-.021	.006	-.016	.006	-.013	.006
GM_{-1}	.076	.014	.085	.014	.089	.014
WR_{-1}	.017	.012	.011	.011	.009	.011
t	—	—	-.076	.010	-.085	.011
Australia	—	—	—	—	.893	.604
Austria	—	—	—	—	.910	.633
Belgium	—	—	—	—	.319	.610
Canada	—	—	—	—	.628	.625
Denmark	—	—	—	—	-.124	.597
Finland	—	—	—	—	.238	.601
France	—	—	—	—	.536	.599
Germany	—	—	—	—	-.389	.736
Ireland	—	—	—	—	.683	.597
Italy	—	—	—	—	1.14	.612
Japan	—	—	—	—	1.74	.632
Netherlands	—	—	—	—	.557	.599
Norway	—	—	—	—	.189	.594
Spain	—	—	—	—	.934	.623
Sweden	—	—	—	—	-.232	.596
Switzerland	—	—	—	—	.207	.609
UK	—	—	—	—	-.384	.636
R^2	.29		.34		.36	
F			53.3		4.14	

[12]The means and standard deviations are based on BMOM(N). Also, 4.14 in the last cell of the table is the value of the F statistic associated with the null hypothesis that both the time trend and country indicator variables' coefficients are all equal to zero. The value of the F statistic associated with the null hypothesis that the country indicator variables' coefficients are all equal to zero is .63, a low value.

We see that for the basic ARLI model, most coefficients' post data means are several times larger in magnitude than their associated post data standard deviations. For example, the post data mean of the coefficient of the real money growth rate variable is .076 with a standard deviation of .014. Use of a time trend variable improved the fit of the relations somewhat as measured by R^2, but the addition of the country indicators did not significantly improve the fit of the model. Also, with prior odds = 1, the condition in (24) for rejecting the hypothesis that the coefficient of the trend is zero and the coefficients of the trend and country indicators are all zero is $F > 1 + z$. With the large value of F-statistic for testing whether the coefficient on the time trend is zero, it appears that this hypothesis can be rejected for many reasonable values of z. Further, the model with country indicators leads to higher expected loss for all choices of z, since the associated F statistic is .63.

The forecasting results for the two-action case are given in Tables 6 and 7 of the appendix. We see from the results in Table 6 that use of our methods and alternative models has led to correct turning point forecasts in 60 to 71 percent of 192 turning point episodes in annual output growth rates for 18 countries in the period 1981-95. The highest percentage correct, 71, is encountered for the ARLI model incorporating a time trend, country indicator variables and a trend add factor. For this model and add factor, 69 percent of the 96 DT/NDT forecasts are correct and 73 percent of the 96 UT/NUT forecasts are correct. The trend add factor involves adding one percentage point to the predictive mean in the DT/NDT situation and subtracting one percentage point in the UT/NUT situation, a correction that may reflect inertia or persistence in direction of movement. However, as Table 6 indicates, the performance of the various models with and without add factors is remarkably similar. Further, allowing for skewness and kurtosis in deriving predictive densities did not result in any major change in performance as shown in the bottom part of Table 6 where the percentage of correct turning point forecasts ranges from 63 to 69. Alternate rules for assigning values for skewness and kurtosis were also investigated, but variations in the rules did not significantly impact the forecasting results. Finally, considering DT/NDT and UT/NUT forecasts separately, the percentage of correct forecasts in the former case ranged from 61 to 70 and from 57 to 73 in the latter for the models in the upper part of Table 6. We see that for all models considered, we obtain correct forecasts for 60 to 71 percent of the 192 turning point episodes encountered in the IMF annual data, 1981-1995. In earlier work, Zellner, Hong and Min (1991) correctly forecasted about 70 percent of 158 turning points for the period 1974-1986 using the ARLI model described above and variants of it.

In Table 7, results are shown separately for different forecasts. The ARLI models including a time trend and both a time trend and country indicator variables along with use of the trend add factor produced 67 to 74 percent correct DT, NDT,UT and NUT forecasts. Other models with and without the use of add factors produced much more variable results. For example, the ARLI model without a trend term, country indicator variables and add factors, produced 61, 81, 98 and 16 percent correct DT, NDT, UT and NUT forecasts, respectively. The poor NUT forecasts have been noted in previous work and it is satisfying to see that use of the broadened model mentioned below has resulted in improved NUT forecasts. With a symmetric 2×2 loss structure, realized loss using the elaborated ARLI model with the trend add factor is 56, the lowest shown in the last column of Table 7, whereas use of the ARLI model with no trend and country indicator variables nor add factors has an associated loss of 64, 14 percent higher.[13]

Tables 8-9 give the results for forecasting major and minor turning points for the ARLI model with a time trend, while Tables 10-11 report results using the BMOM exponential quartic predictive density discussed in section 3.3. From inspection of the tables we see that the exponential quartic predictive density performs at least as well as the BMOM(N) model in Tables 8-9 in the forecast of major, minor or no upturns. However, the quartic predictive density fails to forecast correctly as many NDT's as the BMOM(N) model. This stems from the fact that in the case of a DT forecast, the current growth rate is often large, and thus, by construction, our exponential quartic predictive density places more mass on the left tail of the distribution. Consequently, the model will associate larger probabilities with Major DT and Minor DT, but a smaller probability with NDT. From the tables, we see that this is indeed the case - more Major DT's and Minor DT's are correctly forecasted in Tables 10 and 11 than are forecasted in Tables 8 and 9. For each loss structure, the total percentage of correct forecasts is slightly higher using the symmetric BMOM(N) predictive density than the exponential quartic density.

When comparing the results obtained by the use of loss structures #1 and #2, we see that loss structure #2 results in fewer "large" misses. We define a large miss if NDT was forecasted and a Major DT occurred or if Major DT was forecasted and NDT occurred. Recall that under loss structure #2, the loss associated with a large miss was twice as large as the loss associated with a small miss, and thus we would expect to see more "safe" forecasts (*i.e.* Minor DT / Minor UT) under loss structure #2. From inspection of the tables, we see that this is indeed the case. In particular, we note that under loss structure #1 we fail to forecast any Minor UT events,[14] while

[13]The realized losses are evaluated by assigning a loss of 1 for each incorrect forecast and summing the number of incorrect forecasts for each model. These numbers are 64,76,60,64,72,60,57,66,61,72,62,70,63,56,77,72, 59, and 61 for the rows of Table 7.

[14]Recall that we have defined a minor UT relative to the size of the current growth rate, $y_{iT}(1+.5z) \geq y_{iT+1} > y_{iT}$, so that if the current growth rate is small, the interval defining a minor UT will be small, and thus the probability

in Table 9 we forecast 5 of 14 (36 %) and in Table 11, 6 of 14 (43 %) Minor UT's correctly.

Finally, we note that the total number of correct forecasts for this three-action case ranges from 54-60 percent. Of course, we expect that the percentage of correct forecasts in the three-outcome case will be smaller than the two outcome case in which 60-71 percent of the forecasts were correct. Correctly forecasting 54-60 percent of the 192 turning point episodes in the case of three outcomes is certainly encouraging in view of the fact that it is more difficult forecasting in this situation as compared to forecasting in the two outcome case. The procedures presented here for forecasting Major and Minor turning points are quite operational; other turning point definitions or loss functions deemed to be appropriate for a given problem can also be employed.

The results for the two and three action case are broken down by year and country in tables 12-15. Tables 12 and 13 give the results for the two action case while tables 14 and 15 give the results for the three action case under loss structure #2. Note that under this loss structure, the event which is forecasted is not necessarily the event with the highest computed probability. The decision rules used to obtain these turning point forecasts are given in Table 4.

5 Summary and Concluding Remarks

In this paper we have shown how the BMOM approach can be applied in the analysis of various time series models and problems. Use of it permits derivation of post data moments and densities for parameters and future observations without use of data sampling assumptions, likelihood functions and prior densities for parameters. These BMOM results are operational and can easily be applied in analyses of time series models and in forecasting. Further, we showed how a BMOM loss function approach can be employed to choose between or among alternative models and hypotheses, for example, an AR(1) model with a unit root versus one without a unit root or an AR(1) model versus an AR(3) model. Such choices are made in a Bayesian decision theoretic manner by choosing the model or hypothesis that yields lower or lowest expected loss. While we have just employed quadratic loss functions in this paper, it is the case that such decisions can be made relative to a wide range of symmetric and asymmetric loss functions. In addition, as shown in previous work, Tobias and Zellner (1997), with new data, show how post data odds relating alternative models and

associated with a minor UT will also be small.

hypotheses can be evaluated and used to select and/or combine models. Traditional Bayesian and BMOM predictive densities have been employed in the current paper to forecast turning points in 18 countries' annual real GDP growth rates using revised IMF data and an extended time period relative to those used in earlier studies. It was found that our Bayesian decision theoretic procedure for forecasting turning points produced about 60 to 70 percent correct forecasts in 192 turning point episodes. Further, it was shown how to introduce "add factors" in the BMOM approach and their use was evaluated empirically. Also, several variants of a basic ARLI model were employed to forecast turning points and their performance relative to the basic ARLI model was evaluated. An extended Bayesian decision theoretic procedure for forecasting minor and major turning points was formulated and applied. This ability to distinguish minor downturns and upturns from major downturns and upturns is important and thus it is fortunate that probabilities associated with such possible outcomes are readily computed and can be employed to derive optimal turning point forecasts.

As regards future research, we are planning to study the effects of aggregation on the quality of point and turning point forecasts. By disaggregating total output, we can forecast the components and add them up to get a forecast of the total and compare it to that yielded by a model for the aggregate data. Some preliminary research by de Alba and Zellner (1991) indicates conditions under which disaggregation will lead to better forecasts, and some report actual improvement in practice; see, *e.g.* Espassa (1994). In calculations to be reported in a future paper, we have found that utilizing the information in the 18 countries' data and forecasting equations leads to better point forecasts of the median of the 18 countries' annual output growth rates than are obtained for an aggregative model of the median growth rates such as employed in Zellner and Hong (1989). Whether this result carries over to forecasting turning points will be investigated in future work.

6 Appendix

Table 6
Results of Forecasting Turning Points in Output Growth Rates, 1981-1995. Percentages of Correct Forecasts for Alternative ARLI Models[15]

		Percentage Correct		
Model	Add Factors	DT/NDT (96 Forecasts)	UT/NUT (96 Forecasts)	Total (192 Forecasts)
ARLI	None	70	61	66
	Optimist	64	57	60
	Pessimist	68	70	69
	Trend	64	70	67
	Change	68	57	63
ARLI with Time Trend	None	67	71	69
	Optimist	68	63	65
	Pessimist	63	69	66
	Trend	68	69	68
	Change	63	63	63
ARLI with Time Trend and Country Indicators	None	68	68	68
	Optimist	69	58	64
	Pessimist	61	73	67
	Trend	69	73	71
	Change	61	58	60
Models using Measures of Skewness and Kurtosis To Derive Exponential Quartic Predictive Densities				
ARLI	None	66	59	63
ARLI with Time Trend	None	68	71	69
ARLI with Time Trend and Country Indicators	None	69	68	68

[15]The models with skewness and kurtosis use four moments in the predictive density and are described in detail in the text. The "Optimist" always tends to favor NDT and UT, and shifts all predictive means by adding one percentage point of growth. The "Pessimist" favors DT and NUT and subtracts one percentage point from all predictive means. The "Trend" add factors capture "inertia" - the economy will continue in its current state, and thus adds one percentage point in the DT/NDT case and subtracts one point in the UT/NUT case. Finally, the "Change" add factors capture the opposite of inertia - by subtracting one point in the DT/NDT case, and adding one point in the UT/NUT case.

Table 7
Results of Forecasting Turning Points in Output Growth Rates for 18 Countries with Add Factors Equal to One Percentage Point Forecasting Period: 1981-1995[16]

Model	Add Factors	Number Correct				Percentage Correct			
		DT	NDT	UT	NUT	DT	NDT	UT	NUT
Actual TP's		54	42	53	43	100	100	100	100
ARLI	None	33	34	52	7	61	81	98	16
	Optimist	20	41	53	2	37	98	100	5
	Pessimist	50	15	49	18	93	36	92	42
	Trend	20	41	49	18	37	98	92	42
	Change	50	15	53	2	93	36	100	5
ARLI.with Time Trend	None	49	15	46	22	91	36	87	51
	Optimist	36	29	51	9	67	69	96	21
	Pessimist	53	7	37	29	98	17	70	67
	Trend	36	29	37	29	67	69	70	67
	Change	53	7	51	9	98	17	96	21
ARLI with Time Trend and Country Indicators	None	51	14	47	18	94	33	89	42
	Optimist	37	29	49	7	69	69	92	16
	Pessimist	53	6	39	31	98	14	74	72
	Trend	37	29	39	31	69	69	74	72
	Change	53	6	49	7	98	14	92	16
Models using Measures of Skewness and Kurtosis To Derive Exponential Quartic Predictive Densities									
ARLI	None	29	34	52	5	54	81	98	12
ARLI with Time Trend	None	49	16	46	22	91	38	87	51
ARLI with Time Trend and Country Indicators	None	49	17	47	18	91	40	89	42

[16]The models with exponential quartic predictive densities are asymmetric with heavy tails relative to the normal density and are described in the text. The "Optimist" always tends to favor NDT and UT, and shifts all predictive means by adding 1 percentage point of growth. The "Pessimist" favors DT and NUT and subtracts one percentage point from all predictive means. The "Trend" add factors capture "inertia" - the economy will continue in its current state, and thus adds one percentage point in the DT/NDT case and subtracts one point in the UT/NUT case. Finally, the "Change" add factors capture the opposite of inertia - by subtracting one point in the DT/NDT case, and adding one point in the UT/NUT case.

Table 8
Forecasting Results Using ARLI Model With Time Trend:
Loss Structure #1[17]

Forecasted Outcome	Actual Outcome						% Correctly Forecasted	
	Major DT	Minor DT	NDT	Major UT	Minor UT	NUT		
Major DT	5	8	4	···	···	···	Major DT	24
Minor DT	3	8	0	···	···	···	Minor DT	24
NDT	13	17	38	···	···	···	NDT	90
Major UT	···	···	···	36	8	20	Major UT	92
Minor UT	···	···	···	0	0	0	Minor UT	0
NUT	···	···	···	3	6	23	NUT	53
Total Outcomes	21	33	42	39	14	43		
Percent of Correct TP Forecasts							57	
Percent of DT's / NDT Forecasted Correctly							55	
Percent of UT's /NUT Forecasted Correctly							61	
Percent of DT's /NDT "Large" Misses							18	
Percent of UT's/NUT "Large" Misses							24	

Table 9
Forecasting Results Using ARLI Model With Time Trend:
Loss Structure #2.

Forecasted Outcome	Actual Outcome						% Correctly Forecasted	
	Major DT	Minor DT	NDT	Major UT	Minor UT	NUT		
Major DT	4	2	0	···	···	···	Major DT	19
Minor DT	7	23	13	···	···	···	Minor DT	70
NDT	10	8	29	···	···	···	NDT	69
Major UT	···	···	···	34	5	17	Major UT	87
Minor UT	···	···	···	2	5	6	Minor UT	36
NUT	···	···	···	3	4	20	NUT	47
Total Outcomes	21	33	42	39	14	43		
Percent of Correct TP Forecasts							60	
Percent of DT's / NDT Forecasted Correctly							58	
Percent of UT's /NUT Forecasted Correctly							61	
Percent of DT's /NDT "Large" Misses							10	
Percent of UT's/NUT "Large" Misses							21	

[17] The (i, j) entry of these tables gives the number of times the outcome in row i was forecasted when the event in column j actually occurred.

Table 10
Forecasting Results Using ARLI Model With Time Trend
Incorporating Measures of Skewness and Kurtosis: Loss Structure #1[18]

Forecasted Outcome	Actual Outcome						% Correctly Forecasted	
	Major DT	Minor DT	NDT	Major UT	Minor UT	NUT		
Major DT	9	14	17	⋯	⋯	⋯	Major DT	43
Minor DT	3	12	1	⋯	⋯	⋯	Minor DT	36
NDT	9	7	24	⋯	⋯	⋯	NDT	57
Major UT	⋯	⋯	⋯	36	8	20	Major UT	92
Minor UT	⋯	⋯	⋯	0	0	0	Minor UT	0
NUT	⋯	⋯	⋯	3	6	23	NUT	53
Total Outcomes	21	33	42	39	14	43		
Percent of Correct TP Forecasts							54	
Percent of DT's / NDT Forecasted Correctly							47	
Percent of UT's /NUT Forecasted Correctly							61	
Percent of DT's /NDT "Large" Misses							27	
Percent of UT's/NUT "Large" Misses							24	

Table 11
Forecasting Results Using ARLI Model With Time Trend
Incorporating Measures of Skewness and Kurtosis: Loss Structure #2.

Forecasted Outcome	Actual Outcome						% Correctly Forecasted	
	Major DT	Minor DT	NDT	Major UT	Minor UT	NUT		
Major DT	5	4	5	⋯	⋯	⋯	Major DT	24
Minor DT	14	25	20	⋯	⋯	⋯	Minor DT	76
NDT	2	4	17	⋯	⋯	⋯	NDT	40
Major UT	⋯	⋯	⋯	36	5	17	Major UT	92
Minor UT	⋯	⋯	⋯	1	6	6	Minor UT	43
NUT	⋯	⋯	⋯	2	3	20	NUT	47
Total Outcomes	21	33	42	39	14	43		
Percent of Correct TP Forecasts							57	
Percent of DT's / NDT Forecasted Correctly							49	
Percent of UT's /NUT Forecasted Correctly							65	
Percent of DT's /NDT "Large" Misses							7	
Percent of UT's/NUT "Large" Misses							20	

[18] The models with skewness and kurtosis use four moments in the predictive density and are described in detail in the text. The (i, j) entry of these tables gives the number of times the outcome in row i was forecasted when the event in column j actually occurred.

References

Bayes, T., "An Essay Towards Solving a Problem in the Doctrine of Chances," *Phil. Trans. Royal. Soc.* 53 (1763), 370–418.

Berger, J.O., *Statistical Decision Theory and Bayesian Analysis* (New York: Springer-Verlag, 2nd ed., 1985).

Chaloner, K. and R. Brant, "A Bayesian Approach to Outlier Detection and Residual Analysis," *Biometrika* 75 (1988), 651–659.

Cover, T. and J. Thomas, *Elements of Information Theory* (New York: John Wiley & Sons, 1991).

Currie, J., "The Geographic Extent of the Market: Theory and Application to the U.S. Petroleum Markets," Ph.D. Thesis, Department of Economics, University of Chicago, 1996.

de Alba, E. and A. Zellner, "Aggregation, Disaggregation, Predictive Precision and Modeling," working paper, H.G.B. Alexander Research Foundation (1991).

DeGroot, M., *Optimal Statistical Decisions* (New York: McGraw-Hill, 1970).

Espassa, A., "Comment on 'Time Series Analysis, Forecasting and Econometric Modeling: The Structural Econometric Modeling, Time Series Analysis (SEMSTA) Approach'," *Journal of Forecasting* 13 (1994).

Garcia-Ferrer, A., R.A. Highfield, F. Palm and A.Zellner, "Macro-Economic Forecasting Using Pooled International Data," *Journal of Business and Economic Statistics* 5, 53–67 (1987).

Golan, A., G. Judge, and D. Miller, *Maximum Entropy Econometrics* (Chichester: John Wiley & Sons, Ltd., 1996).

Good, I.J., *Probability and the Weighing of the Evidence* (New York: Hafner, 1950).

Green, E. and W. E. Strawderman, "A Bayesian Growth and Yield Model for Slash Pine Plantations," *Journal of Applied Statistics* 23 (1996).

Hong, C., "Forecasting Real Output Growth Rates and Cyclical Properties of Models: A Bayesian approach", Ph.D. thesis, Department of Economics, University of Chicago, 1989.

Jaynes, E.T., "On the Rationale of Maximum-Entropy Methods," *Proceedings of the IEEE* 70 (1982), 939–952.

Jaynes, E.T., "Comment on 'Optimal Information Processing and Bayes' Theorem'," *American Statistician* 42 (1988), 280–281.

Kass, R.E. and A.E. Raferty., "Bayes Factors" *Journal of the American Statistical Association* 90 (1995), 773-795.

Kullback, S., *Information Theory and Statistics* (New York: Wiley, 1959).

LeSage, J.P., "Forecasting Turning Points in Metropolitan Employment Growth Rates Using Bayesian Exponentially Weighted Regression, Time-Varying Parameter and Pooling Techniques" *Journal of Regional Science* 30(1990), 533–548.

Manas-Anton, L. P., "Empirical Regularities in Short-Run Exchange Rate Behavior," Ph.D. Thesis, Department of Economics, University of Chicago, 1986.

Min, C-K. and A. Zellner, "Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates," *Journal of Forecasting* 11 (1993), 687-701.

Orcutt, G.H., M. Greenberg, F.J. Korbel and A.M. Rivlin, *Microanalysis of Socioeconomic Systems: A Simulation Study* New York: Harper & Row (1961).

Palm, F.C. and A. Zellner, "To Combine or Not to Combine? Issues of Combining Forecasts," *Journal of Forecasting* 11 (1992), 687–701.

Phillips, P.C.B., "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics* 6 (1991), 333-364.

Ryu, H., "Orthogonal Basis and Maximum Entropy Estimation of Probability Density and Regression Functions," Ph.D. Thesis, Department of Economics, University of Chicago, 1990.

Ryu, H., "Maximum Entropy Estimation of Density and Regression Functions," *Journal of Econometrics* 56 (1993), 397–440.

Shore, J. and R. Johnson, "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," *IEEE Transactions* IT-26 (1980), 26–37.

Soofi, E., "Capturing the Intangible Concept of Information," *Journal of the American Statistical Association* 89 (1994), 1243–1254.

Soofi, E., "Information Theory and Bayesian Statistics," in D. Berry, K. Chaloner and J. Geweke, eds, *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner* (New York: John Wiley & Sons, 1996).

Tobias, J. and A. Zellner. "Further Results on the Bayesian Method of Moments Analysis of the Multiple Regression Model," H.G.B. Alexander Research Foundation, University of Chicago, 1997. Presented at Econometric Society meeting, June 1997, and the world meeting of the International Society for Bayesian Analysis, Istanbul, August 1997.

Zellner, A., *An Introduction to Bayesian Inference in Econometrics*, (New York: Wiley, 1971) reprinted in Wiley Calssics Library, 1996.

——, "Optimal Information Processing and Bayes' Theorem," *American Statistician* 42 (1988), 278–284, reprinted in Zellner 1997c.

——, "Guy H. Orcutt: Contributions to Economic Statistics," *Journal of Economic Behavior and Organization* (1990), 43–51.

——, "Bayesian Methods and Entropy in Economics and Econometrics," in T. Grandy and L. Schick, eds, *Maximum Entropy and Bayesian Methods* (Dordrecht, Netherlands: Kluwer, 1991), reprinted in Zellner 1997c.

——, "Time Series Analysis, Forecasting and Econometric Modeling: The Structural Econometric Modeling, Time Series Analysis (SEMSTA) Approach," *Journal of Forecasting* 13 (1994), 215-233.

——, "Bayesian Method of Moments/ Instrumental Variable (BMOM/IV) Analysis of Mean and Regression Models," in J. Lee, W. Johnson and A. Zellner, eds, *Modeling and Prediction: Honoring Seymour Geisser* Springer-Verlag, 1996, reprinted in Zellner 1997c.

——, "Comments on A 'Critique' of the "Bayesian" Method of Moments', " H.G.B. Alexander Research Foundation, University of Chicago, 1997a.

——, "The Bayesian Method of Moments (BMOM): Theory and Applications," in T. Fomby and R. Hill, eds, *Advances in Econometrics* 12 (1997b), 85-105.

——, *Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers* (Cheltenham, U.K.: Edward Elgar Publishing Company, 1997c).

—— and R. Highfield, R., "Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions," *Journal of Econometrics* 37 (1988), 195–210.

—— and C. Hong, "Forecasting International Growth Rates Using Bayesian Shrinkage and Other Procedures," *Journal of Econometrics* 40 (1989), 183–202, reprinted in Zellner 1997c.

——, C. Hong, and G.M Gulati, "Turning Points in Economic Time Series, Loss Structures, and Bayesian Forecasting," in S. Geisser, J. Hodges, S.J. Press, and A. Zellner, eds., *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard* Amsterdam: North-Holland 371–393.

——, C. Hong and C-k Min, "Forecasting Turning Points in International Output Growth Rates Using Bayesian Exponentially Weighted Autoregression, Time- Varying Parmaeter, and Pooling Techniques," *Journal of Econometrics* 49 (1991), 275–304, reprinted in Zellner 1997c.

——, and C-K. Min, "Bayesian Analysis, Model Selection and Prediction," *Physics and Probability: Essays in honor of Edwin T. Jaynes* (1993), 195-206, reprinted in Zellner 1997c.

——, C-k. Min, D. Dallaire and J. Currie, "Bayesian Analysis of Simultaneous Equation, Asset-Pricing and Related Models Using Markov Chain Monte Carlo Techniques and Convergence Checks," H.G.B. Alexander Research Foundation, University of Chicago, 1994.

—— and B. Moulton, "Bayesian Regression Diagnostics with Applications to International Consumption and Income Data," *Journal of Econometrics* 29 (1985), 187–211.

—— and C. Plosser, "Posterior Odds Ratios for Discriminating Random Walk Models from Stationary and Explosive Autoregressive Models," summary of presentations to the 11th meeting of the NBER-NSF Seminar on Bayesian Inference, U. of Rochester and to the Econometrics and Statistics Colloquium, U. of Chicago, 1976.

—— and B. Sacks, "Bayesian Method of Moments (BMOM) Analysis of the Multiple Regression Model with Autocorrelated Errors," presented at at the American Statistical Association meeting, August 1,1996.

—— and A. Siow, "Posterior Odds for Selected Regression Hypotheses," *Bayesian Statistical Regression Hypotheses* (Valencia, Spain: University Press) (1980), 585-603

——, J. Tobias and H. Ryu, "Bayesian Method of Moments (BMOM) Analysis of Parametric and Semiparametric Regression Models," presented at the world meeting of the International Society for Bayesian Analysis, Istanbul, August 1997, and to appear in *1998 Proceedings of the American Statistical Association Section on Bayesian Statistical Science.*