*Working Paper Series*

WORKING PAPER NO. 833

# FURTHER RESULTS ON BAYESIAN METHOD OF MOMENTS ANALYSIS OF THE MULTIPLE REGRESSION MODEL

by

Justin Tobias and Arnold Zellner*

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS

BERKELEY

California Agricultural Experiment Station

*University of California*

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS
DIVISION OF AGRICULTURE AND NATURAL RESOURCES
UNIVERSITY OF CALIFORNIA AT BERKELEY

WORKING PAPER NO. 833

# FURTHER RESULTS ON BAYESIAN METHOD OF MOMENTS ANALYSIS OF THE MULTIPLE REGRESSION MODEL

by

**Justin Tobias and Arnold Zellner***



*University of Chicago; and Visiting Professor, Department of Agricultural and Resource Economics, University of California, Berkeley, respectively.

**California Agricultural Experiment Station**
**Giannini Foundation of Agricultural Economics**
**February 1998**

# Further Results on Bayesian Method of Moments Analysis of the Multiple Regression Model

Justin Tobias and Arnold Zellner*

April 13, 1997

**Abstract**

The Bayesian Method of Moments (BMOM) was introduced in 1994 to permit investigators to make inverse probability statements regarding parameters' possible values given the data when the form of the likelihood function is unknown. BMOM has been applied in analyses of several statistical and econometric models including location, multiple and multivariate regression, and simultaneous equation models. In Zellner (1996) and Zellner and Sacks (1996) some previous BMOM analyses of the multiple regression model have appeared that permit derivation of post-data densities for parameters and future observations to be calculated without use of a likelihood function, prior density, or Bayes' Theorem. In the present paper, we extend previous analyses by showing how information about a variance parameter and its relation to regression coefficients affects post-data densities. We also discuss estimation of functions of parameters and model selection techniques using BMOM and traditional Bayesian approaches. The developed theory is then applied using a model and data from a famous paper by Haavelmo (1947) .

*Research financed in part by the National Science Foundation and by income from the H.G.B. Alexander Endowment Trust Fund, Graduate School of Business, University of Chicago. We would also like to thank participants of the Student Econometrics Workshop at the University of Chicago for helpful comments and suggestions. This paper was presented at the Econometric Society meeting, Caltech, June 1997.

# Further Results on Bayesian Method of Moments Analysis of the Multiple Regression Model

Justin Tobias and Arnold Zellner*

April 13, 1997

**Abstract**

The Bayesian Method of Moments (BMOM) was introduced in 1994 to permit investigators to make inverse probability statements regarding parameters' possible values given the data when the form of the likelihood function is unknown. BMOM has been applied in analyses of several statistical and econometric models including location, multiple and multivariate regression, and simultaneous equation models. In Zellner (1996) and Zellner and Sacks (1996) some previous BMOM analyses of the multiple regression model have appeared that permit derivation of post-data densities for parameters and future observations to be calculated without use of a likelihood function, prior density, or Bayes' Theorem. In the present paper, we extend previous analyses by showing how information about a variance parameter and its relation to regression coefficients affects post-data densities. We also discuss estimation of functions of parameters and model selection techniques using BMOM and traditional Bayesian approaches. The developed theory is then applied using a model and data from a famous paper by Haavelmo (1947).

*Research financed in part by the National Science Foundation and by income from the H.G.B. Alexander Endowment Trust Fund, Graduate School of Business, University of Chicago. We would also like to thank participants of the Student Econometrics Workshop at the University of Chicago for helpful comments and suggestions. This paper was presented at the Econometric Society meeting, Caltech, June 1997.

## 1 Introduction

The Bayesian method of moments (BMOM) was introduced to permit investigators to compute post-data densities for parameters and future observations when not enough information is available to formulate a satisfactory likelihood function; see Green and Strawderman (1996) for a study in which the authors did not have enough information to formulate a likelihood function and use was made of the BMOM. The BMOM approach provides a solution to the famous inverse problem posed by Bayes (1763) and hence the name Bayesian method of moments. BMOM has been studied and applied to various models in Zellner (1995,1996,1997b), Green and Strawderman (1996), Zellner and Sacks (1996), Currie (1996), and Zellner Min, Dallaire and Currie (1994). In the BMOM approach, two basic assumptions are made that permit evaluation of post-data moments of parameters from a given set of data. Then, from among various methods of determining densities from given moments, the maximum entropy (maxent) approach is used to choose a proper density with the given moments that maximizes entropy. For discussion and applications of maxent, see e.g. Jaynes (1982,1988), Shore and Johnson (1980), Cover and Thomas (1991), and Zellner and Highfield (1988). Also, see Zellner (1997) for coherent procedures for updating BMOM maxent post-data densities for parameters and future observations.

The methods developed in this paper augment previous BMOM analyses of location, multiple regression, multivariate regression and simultaneous equations models with or without autocorrelation or heteroscedasticity. In the current paper we review and extend the existing theory of BMOM in analysis of the standard multiple regression model. In particular, we derive post-data densities for parameters and future, as yet unobserved observations using various moment side conditions and show how use of alternative moment side conditions affects the shapes and properties of maxent densities. For certain moment side conditions, post-data densities are very similar to those derived in a traditional Bayesian approach based on improper diffuse prior densities for parameters and a normal likelihood function. Further, as in Chaloner and Brant (1988), Zellner and Moulton (1985), and Zellner (1975), the BMOM approach yields moments and densities for realized error terms and functions of them that are very useful for diagnostic checking purposes. Last, it is shown how posterior odds can be computed to compare models produced by BMOM and those produced with a traditional Bayesian approach.

After presenting the above and pointing out key relations between BMOM results and traditional Bayesian results, we analyze generated data and data from a famous study by Haavelmo (1947) to illustrate BMOM empirical results. Also, we compute various measures, including traditional Bayes' factors to compare models produced by approaches based on different assumptions. As discussed in Min and Zellner (1993) and Palm and Zellner (1992), posterior odds can be utilized to compare and/or combine alternative predictive models. On this capability, Barnard (1997) has commented favorably on the value of being able to compare and select among BMOM and TB approaches.

The plan of the paper is as follows. In Section 2, we review the BMOM assumptions relating to the multiple regression model and demonstrate how various moment conditions are derived. Then these moments are used as side conditions in deriving proper maxent post-data densities for parameters and future observations. Included is a demonstration of the moment side conditions for the variance parameter that lead to a maxent density in the inverted gamma form, a form that is encountered in a traditional Bayes approach based on an improper diffuse prior and iid normal error terms. Also, the dependence of the variance of the variance parameter on the sample size is investigated. Finally, some comments on sampling properties of BMOM and traditional Bayes estimates are provided. Section 4 is devoted to presenting the results of analyses of generated data and data from Haavelmo's (1947) paper. Similarities and differences of results produced by various BMOM and traditional Bayes approaches are discussed. In Section 4, various measures including Bayes factors are employed to compare alternative models. The paper concludes with a summary in Section 5.

# 2 Review and Extension of BMOM

## 2.1 Post-Data Moments for Regression Parameters

Let $y$, an $n \times 1$ vector of given observations be assumed to be related to $X$, a given $n \times k$ matrix of rank $k$, as follows

$$y = X\beta + u, \tag{1}$$

where $\beta$ is a $k \times 1$ vector of regression coefficients with fixed but unknown values. and $u$ is an $n \times 1$ vector of realized error terms. As in earlier work on the analysis of realized error terms (see Chaloner and Brant (1988), Zellner (1975) and Zellner and Moulton (1985) ), we regard $\beta$ and $u$ to be subjectively random. We shall introduce assumptions that will enable us to obtain the moments of the elements of $\beta$ and $u$ and then use the principle of maximum entropy to obtain proper post-data densities.

From equation (1) we have

$$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u. \tag{2}$$

On taking the post-data expectation of both sides of (2) given the data $D = (y, X)$,

$$\hat{\beta} = E(\beta \mid D) + (X'X)^{-1}X'E(u \mid D), \tag{3}$$

where $E$ denotes the subjective, post-data expectation operator. We now introduce the following assumption,

**Assumption 1** $X'E(u \mid D) = 0$

namely that the columns of $X$ are orthogonal to the vector $E(u \mid D)$. This assumption would not be satisfied if relevant variables correlated with $X$ are omitted from (1), if the included independent variables are measured with error, or if other errors are made in formulating the form of (1). Given that assumption 1 is satisfied, we have from (3),

$$E(\beta \mid D) = \hat{\beta} = (X'X)^{-1}X'y. \tag{4}$$

That is, the post-data mean of $\beta$ is equal to the least squares estimate. Also, the mean of $\beta$ given in (4) is an optimal point estimate relative to a quadratic loss function, $L(\beta, \tilde{\beta}) = (\beta - \tilde{\beta})'Q(\beta - \tilde{\beta})$, where $\tilde{\beta}$ is some estimate and $Q$ is a given positive definite, symmetric matrix. The value of $\tilde{\beta}$ that minimizes expected loss is the mean given in (4).

Further, assumption 1 yields the following post-data mean for $u$,

$$E(u \mid D) = y - XE(\beta \mid D) = y - X\hat{\beta} = \hat{u} \tag{5}$$

where $\hat{u}$ is the least squares residual vector that satisfies $X'\hat{u} = 0$. Note that from (5) we can write

$$\begin{aligned} u - \hat{u} &= y - X\beta - \left(y - X\hat{\beta}\right) \\ &= X(X'X)^{-1}X'u \\ &= X(X'X)^{-1}X'(u - \hat{u}) \end{aligned}$$

where the last step follows from the orthogonality condition mentioned above, $X'\hat{u} = 0$. We can thus write

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

which defines a functional equation that the post-data covariance matrix for $u, V(u \mid D)$ must satisfy. Since there are only $k$ free elements of $u$ in the equations in (1), $V(u \mid D)$ must be of rank $k$. In view of these considerations, our second assumption is [1]

**Assumption 2** *$Var(u \mid \sigma^2, D) = \sigma^2 X(X'X)^{-1}X'$,*

where $\sigma^2$ is a variance parameter to be defined below. We use assumption 2 to evaluate the post-data covariance matrix of $\beta$ as follows

$$\begin{aligned} \text{Var}(\beta \mid \sigma^2.D) &= E\left[\left(\beta - \hat{\beta}\right)\left(\beta - \hat{\beta}\right)' \mid D\right] \\ &= (X'X)^{-1}X'E\left[(u-\hat{u})(u-\hat{u})' \mid D\right]X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned} \tag{6}$$

Assumption 2 will also enable us to evaluate the post-data moments of $\sigma^2$. In the arguments of this section, we assume that an intercept appears in the regression matrix. From assumption 1, observe that the presence of an intercept implies that the post-data mean of $\overline{u} = \sum_{i=1}^{n} u_i/n$ is zero, which simplifies the derivations to follow. We define the expectation of $\sigma^2$ given the data as follows

$$E[\sigma^2 \mid D] \equiv E\left[\sum_{i=1}^{n} \frac{(u_i - E(\overline{u} \mid D))^2}{n} \mid D\right] = E\left(\frac{u'u}{n} \mid D\right). \tag{7}$$

[1] Note that substituting assumption 2 in to the formula for $\text{Var}(u \mid \sigma^2, D)$ solves the fixed point problem.

We also note that

$$\begin{aligned} E(u'u \mid D) &= nE(\sigma^2 \mid D) = \left(y - X\hat{\beta}\right)' \left(y - X\hat{\beta}\right) + E\left[\left(\beta - \hat{\beta}\right)' X'X \left(\beta - \hat{\beta}\right) \mid D\right] \\ &= \hat{u}'\hat{u} + \mathrm{tr}\left[X'XE\left\{\left(\beta - \hat{\beta}\right)\left(\beta - \hat{\beta}\right)' \mid D\right\}\right] \\ &= \hat{u}'\hat{u} + kE(\sigma^2 \mid D), \qquad (8) \end{aligned}$$

where the first line follows from (7). Solving (8) for $E(\sigma^2 \mid D)$, we obtain the result

$$E(\sigma^2 \mid D) = \frac{\hat{u}'\hat{u}}{n-k} = s^2. \qquad (9)$$

Thus relative to quadratic loss, $L(\sigma^2, \tilde{\sigma}^2) = (\sigma^2 - \tilde{\sigma}^2)^2$, the estimate $\tilde{\sigma}^2$ which minimizes post-data expected loss is $E(\sigma^2 \mid D) = s^2$, with $s^2$ defined in (9). Note that this post-data expectation differs from the post-data mean of $\sigma^2$ in a diffuse prior - normal likelihood traditional Bayesian approach, namely $E_{TB}(\sigma^2 \mid D) = vs^2/(v-2)$, with $v = n - k > 2$. For small values of $v$, the last expression is much larger than $s^2$.

The results in equations (6) and (9) yield

$$\mathrm{Var}(\beta \mid D) = s^2(X'X)^{-1} \qquad (10)$$

Note that the results in (4) and (10) are what one might use for moments of parameters in a large sample, approximate traditional Bayesian approach when there are difficulties in formulating a likelihood function and/or prior density. Here the results in (4) and (10) are exact results and are not large sample approximations.

We will now explain how maximum entropy is used to compute a density function from given moment conditions. We will show that using the first and second moments in (4) and (10) as constraints, the proper density that maximizes entropy is the following normal density for $\beta$ given $D$ [2]

$$\beta \sim N(\hat{\beta}, s^2(X'X)^{-1}).$$

[2] Note that if $\beta$ and $\sigma^2$ are assumed a priori independent given the data, $D$, then the entropy of their joint density, $p(\beta, \sigma^2 \mid D) = f(\beta \mid D)g(\sigma^2 \mid D)$ is $H(p) = H(fg) = H(f) + H(g)$. Thus, $f$ in (13) maximizes the first term of the entropy of the joint density and a proper $g$ that maximizes the second term, $H(g)$, subject to moment constraints can be derived. See below for examples.

The entropy, or negative information, of a density function relative to uniform measure is defined as

$$H(f) = -\int f(x) \log f(x) dx. \tag{11}$$

Thus, it is seen that entropy measures the average log height of a density function. Viewed as an objective function, maximizing the entropy of a given density is to minimize the amount of information in that density. We can then consider the problem

$$\begin{aligned} &\max_f - \int f(x) \log f(x) dx \\ &\text{subject to} \int x^i f(x) dx = \mu_i, \quad i = 0, 1, \cdots n \end{aligned} \tag{12}$$

with $\mu_0 = 1$. That is, we seek to find the most conservative, or least informative density that incorporates the information in the moment side conditions. Using the calculus of variations, we find that the maxent density solving (12) is of the form

$$f^*(x) = \exp\left(-(\lambda_0 + \lambda_1 x + \cdots + \lambda_n x^n)\right), \tag{13}$$

where $\lambda_i$ is the Lagrange multiplier associated with the moment side condition $E[x^i]$ in (12). Substituting this maxent density into the $n+1$ moment conditions in (12) defines $n+1$ nonlinear integral equations in $n+1$ unknowns. Zellner and Highfield (1988) describe an iterative solution to this problem by taking a first-order expansion of the system of moment equations about initial values for the $\lambda$'s. They show existence and uniqueness of a solution and provide applications of the procedures. We apply this algorithm in the application of section 4 and find that for all models used, convergence is achieved to a tolerance of $10^{-4}$ in under 50 iterations.

It is also interesting to note that if the moments $E(\log(x))$ and $E(1/x)$ are employed as side conditions, the proper maxent density is

$$f^*(x) = \exp\left(-(\lambda_0 + \lambda_1 \log(x) + \lambda_2 \frac{1}{x})\right) \propto x^{-\lambda_1} \exp\left(-\frac{\lambda_2}{x}\right),$$

which has the form of an inverted gamma density with parameters $\lambda_1$ and $\lambda_2$. The inverted gamma is also the form of the traditional Bayesian posterior density for $\sigma^2$ based on a diffuse prior and iid normal likelihood function. We shall return to this case when discussing post-data densities for the scale parameter.

Adding more moment conditions imposes more constraints on the problem, and hence, the entropy of the resulting maxent density will be reduced (unless the constraints are redundant). Given the post-data moment conditions we have described thus far, we can impose these conditions as constraints and find the maxent density which is as "flat" or uninformative as possible given these conditions. By minimizing the amount of information in the density, we are letting the shape of the post-data densities be determined only by the data via derived moment conditions. A natural question, then is to ask why not add as many moment conditions as possible? Some additional moment conditions may be redundant, and it may be that we are overfitting the model by adding too many constraints. We will address this issue in discussing model selection techniques and the computation of posterior odds in sections 3 and 4.

Given the above maxent results, the proper maxent density for $\beta$ given (4) and (10) is multivariate normal,

$$f(\beta \mid D) \sim N(\hat{\beta}, s^2(X'X)^{-1}). \tag{14}$$

This exact finite sample density can be employed to compute optimal point estimates, marginal densities, intervals, etc.. Further, inequality restrictions on the elements of $\beta$ can easily be imposed in a traditional Bayesian approach using the methods described in Geweke (1986). That is, by making draws from the normal density in (14) and accepting only those draws that satisfy the given inequality constraints, post-data densities can be obtained that incorporate information regarding the restricted values of the elements of the coefficient vector. We can also impose these inequality constraints in the BMOM approach by restricting the region of integration so that the inequality constraints are satisfied. We can then maximize the entropy of our density over the appropriate region subject to given moment side conditions. Last, as shown below in computed examples, for moderate sample sizes, the BMOM post-data density in (14) is very similar in form to traditional Bayesian posterior densities derived using a normal likelihood and a diffuse prior.

Next, using moment conditions in (4) and (6), the following is the proper maxent density for $\beta$ given $\sigma^2$ and $D$,

$$f(\beta \mid \sigma^2, D) \sim N(\hat{\beta}, \sigma^2(X'X)^{-1}) \tag{15}$$

This density can be employed when the value of $\sigma^2$ is known. When its value is unknown, it is interesting to note that use of (15) in conjunction with the first expression in (8) permits us to evaluate higher-order moments of $\sigma^2$. That is, from the definition of $\sigma^2$, we have

$$\sigma^2 = \frac{u'u}{n} = \frac{1}{n}\left(vs^2 + (\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta})\right) = \frac{1}{n}(vs^2 + \sigma^2 Q), \tag{16}$$

where $v = n - k$ and $Q \equiv (\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta})/\sigma^2$, which has a chi-square density with $k$ degrees of freedom provided that $\beta$ has the normal density in (15). This fact can be employed to evaluate the moments of $\sigma^2$ as illustrated below.

First. note that the powers of $\sigma^2$ are defined from (16) as

$$\sigma^{2j} = \frac{1}{n^j}\left(vs^2 + \sigma^2 Q\right)^j, \quad j = 1, 2, \cdots. \tag{17}$$

The right-hand side of this equation can be simplified using the binomial expansion. Taking expectations through the expanded equation produces an equation to be solved for $E(\sigma^{2j} \mid D)$. Using the moments of the chi-square variable and solving the resulting expression, we obtain the following recursive formula for obtaining the desired moments for all $j > 1$,

$$E(\sigma^{2j} \mid D) = \frac{(vs^2)^j + \sum_{i=1}^{j-1} \binom{j}{i} (vs^2)^{j-i} E(\sigma^{2i} \mid D)\left[k(k+2)\cdots\left(k + 2(i-1)\right)\right]}{n^j - \left[k(k+2)\cdots\left(k+2(j-1)\right)\right]}. \tag{18}$$

From this expression we find that the first two moments are given as

$$E(\sigma^2 \mid D) = s^2 \quad \text{and} \quad E(\sigma^4 \mid D) = \frac{s^4\,(v^2 + 2vk)}{n^2 - k(k+2)}. \tag{19}$$

Hence. the posterior variance of $\sigma^2$ is given by

$$Var(\sigma^2 \mid D) = s^4\left(\frac{2k}{n^2 - k(k+2)}\right) \tag{20}$$

It is seen that for given $s^4$ and $k$. the post-data variance of $\sigma^2$ declines with rate $n^2$ given the assumptions above. Further, the variance of the scale parameter is increasing with $k$, the dimension of the regression coefficient vector. To compare this to the TB result, we note that the traditional Bayes posterior variance for $\sigma^2$ when a normal likelihood function and diffuse prior are employed is $2s^4v^2/(v-2)^2(v-4)$. If $\sigma^2$ has an exponential density (a case which will be described in the following sections), then $\sigma^2$ will have a post-data variance of $s^4$. Thus we have a range of alternative densities which allow the sample size to have different effects on the post-data variance of $\sigma^2$. We shall describe in sections 3 and 4 how to compute posterior odds for these alternative models and thus select the model which is most supported by the data.

Higher-order moments of $\sigma^2$ can be obtained recursively by evaluating the above expression. Further, moments of functions of $\sigma^2$, say $g(\sigma^2)$ can be evaluated numerically by making draws from the chi-square density with $k$ degrees of freedom and noting from (16) that $g(\sigma^2) = g\left(vs^2/(n-Q)\right)$. Hence, for each draw from the chi-square distribution, we can compute $g(\cdot)$. Repeating this process and averaging over the resulting values will give an approximation to the mean of $g(\sigma^2)$. These can be used in connection with imposing $E(\log(\sigma^2) \mid D)$, and $E\left(1/\sigma^2 \mid D\right)$ as side conditions and deriving the proper maxent post-data density. We can evaluate the values of these moments numerically and then proceed to solve for the Lagrange multipliers as discussed previously. In Table 1 below, we consider the assumptions used and the resulting post-data densities for $\sigma^2$ in selected BMOM and TB models. These models will then be used in the application of section 4.

**Table 1** [3]

**Restrictions and Post-Data Densities for $\sigma^2$**

| Model | Restrictions | Density Function |
|---|---|---|
| BMOM(1) | $E(\sigma^2 \mid D) = s^2$ | $\frac{1}{s^2}\exp\left(-\frac{\sigma^2}{s^2}\right)$ |
| BMOM(2) | $E(\sigma^2 \mid D) = s^2, \quad E(\sigma^4 \mid D) = \frac{s^4(v^2+2vk)}{n^2-k(k+2)}$ | $\exp\left(-(\lambda_0 + \lambda_1\sigma^2 + \lambda_2\sigma^4)\right)$ |
| BMOM(IG) | $E(\log\sigma^2 \mid D) = \mu_1, \quad E\left(\frac{1}{\sigma^2} \mid D\right) = \mu_2$ | $\exp\left(-(\lambda_0 + \lambda_1\log(\sigma^2) + \lambda_2\frac{1}{\sigma^2})\right)$ |
| TB(IG) | $p(\beta,\sigma^2 \mid D) \propto p_N(y \mid \beta,\sigma^2)p(\beta,\sigma^2)$ | $\propto \frac{1}{\sigma^{v+2}}\exp\left(-\frac{vs^2}{2\sigma^2}\right)$ |

As seen from Table 1, when just the first moment of $\sigma^2$ is employed, the proper maxent density for $\sigma^2$ is in the exponential form, and when two moments of $\sigma^2$ are employed, the maxent post-data density is in a truncated normal form. Finally, when means of $\log \sigma^2$ and of $1/\sigma^2$ are used as side conditions, the maxent post-data density for $\sigma^2$ is in the inverted gamma form, just as in the case in a diffuse prior, normal likelihood traditional Bayesian analysis. From the table, it is seen that the values of the moments of $\sigma^2$ are different in small samples. For example, the mean of $\sigma^2$ under BMOM(1) and BMOM(2) is $s^2$, whereas in the diffuse prior, normal likelihood function approach, the posterior mean of $\sigma^2$ is $vs^2/(v-2)$ which is larger than $s^2$. Under BMOM(IG) the mean will be functions of the Lagrange multipliers $\lambda_1$ and $\lambda_2$. Although the densities are both of the inverted gamma form, the Lagrange multipliers may depart from the corresponding parameters of the TB(IG) density, producing different values for moments of $\sigma^2$ (see, for example Table 4 and Figure 1 in the appendix). Of course, for large $n$, the difference between the values of the means is negligible.

Since use of alternative assumptions leads to different densities for $\sigma^2$, a question as to which assumptions to employ naturally arises. In some applications, higher order moments may be redundant and in others their use may lead to a better model. In sections 3 and 4 we describe and implement model selection techniques using predictive densities and standard Bayesian procedures. In this way, we can choose the model which incorporates the appropriate amount of information regarding the variance parameter $\sigma^2$.

Shown below in Table 2 are alternative marginal post-data densities for an element, $\beta_i$, of the coefficient vector $\beta$. For BMOM(N) and TB, we know that the vector $\beta$ is distributed multivariate normal and multivariate Student-t, respectively. [4]

[4] See Zellner (1971) for a discussion of TB results.

**Table 2** [5]

**Restrictions and Post-Data Densities for $\beta_i$**

| Model | Restrictions | Density Function |
|---|---|---|
| BMOM(N) | $E(\beta) = \hat{\beta}, \quad \text{Var}(\beta) = s^2(X'X)^{-1}$ | $\text{N}(\hat{\beta}_i, s_i^2)$ |
| BMOM(1) | $\int_0^\infty p_N(\beta \mid \sigma^2, D) p_{EXP}(\sigma^2 \mid D) d\sigma^2$ | $\propto \frac{1}{s_i} \exp\left(-\frac{\sqrt{2}\lvert\beta_i - \hat{\beta}_i\rvert}{s_i}\right)$ |
| BMOM(2) | $\int_0^\infty p_N(\beta \mid \sigma^2, D) p_{TN}(\sigma^2 \mid D) d\sigma^2$ | Evaluate Numerically |
| TB | $p(\beta, \sigma^2 \mid D) \propto p_N(y \mid \beta, \sigma^2) p(\beta, \sigma^2)$ | Student-t $(\hat{\beta}_i, \frac{v}{v-2} s_i^2, v)$ |

If we do not impose an independence assumption with respect to $\beta$ and $\sigma^2$, we can write their joint post-data density as $p(\beta, \sigma^2 \mid D) = f(\beta \mid \sigma^2, D) g(\sigma^2 \mid D)$ and maximize the entropy of the joint density with respect to the choice of $f$ and $g$ subject to their being proper and to moment side conditions such as considered above. When this is done, the maxent density for $\beta$ given $\sigma^2$, the data and the first two conditional moment restrictions in (4) and (6) is $f(\beta \mid \sigma^2, D) \sim N(\hat{\beta}, \sigma^2(X'X)^{-1})$. Also, the maxent post-data density for $\sigma^2$ using just the first moment condition, $E(\sigma^2 \mid D) = s^2$ is in the gamma form. Other moment conditions can be imposed to provide a range of possible forms for the marginal post-data density for $\sigma^2$. As mentioned earlier, a draw can be made from the post-data density for $\sigma^2$ and inserted into the conditional normal post-data density for $\beta$. Repeating this process and taking a draw from the conditional normal density for each $\sigma^2$ draw enables numerical estimation of the density function, moments, intervals, and other statistics of interest for the coefficient vector $\beta$. Also, by first integrating over all elements of $\beta$ but one, say $\beta_i$, in the conditional normal density for $\beta$ given $\sigma^2$ and $D$, the joint post-data density for $\beta_i$ and $\sigma^2$ is obtained that can be analyzed by bivariate integration techniques as an

[5] In the table, BMOM(i) $i = 1, 2$ denotes that $i$ moments were employed in deriving the marginal post-data density for $\sigma^2$. BMOM(N) is the normal maxent density using the post-data moments in (4) and (10). In the table, we present the forms of density functions for an element of the $\beta$ vector. Further, we let $s_i^2$ be the $(i, i)$ element of the matrix $s^2(X'X)^{-1}$. BMOM(IG) is not carried along in this table, but the results will be similar in functional form to TB, with the Lagrange multipliers entering as parameters of the density function. For TB, the arguments of the density are the mean, variance, and degrees of freedom, respectively.

alternative to the Monte Carlo method mentioned above. In some cases, these integrations can be performed analytically.

Finally, we note that the post-data densities for $\sigma^2$ can be employed to compute post-data densities for the realized error terms and functions of the realized errors that are often useful for diagnostic purposes as has been recognized in traditional Bayesian analyses: see e.g. Chaloner and Brant (1988), Zellner and Moulton (1985) and Zellner (1975). Since $u_i = y_i - x_i'\beta$, given $y_i, x_i$, and the post-data density for $\beta$, it is possible to compute numerically or perhaps analytically the density function, moments and intervals for the $u_i's$ or functions of them. These can be employed to analyze outlier problems or to obtain the distributions of interesting and useful functions of the $u_i's$, say $\rho_1 = \sum_{i=2}^{n} u_i u_{i-1} / \sum_{i=1}^{n} u_i^2$. [6] That similar analyses can be carried forward when the form of the likelihood function is unknown is noteworthy.

Having derived a range of post-data densities for parameters and indicating how BMOM realized error term analysis can be performed, we now turn to derive post-data predictive densities for future observations.

## 2.2 BMOM Predictive Densities

Here we assume that a $q \times 1$ vector of as yet unobserved future values of the dependent variable, denoted $y_f$, satisfies the following $q$ equations,

$$y_f = X_f\beta + u_f, \tag{21}$$

where $X_f$ is a $q \times k$ matrix with elements having known values, $\beta$ is the $k \times 1$ vector of regression coefficients considered in sections 2.1 and 2.2, and $u_f$ is a $q \times 1$ vector of as yet unrealized error terms. We shall make the same assumptions regarding the properties of $u_f$ as made in previous BMOM work and from these assumptions deduce the moments of and maxent densities for $y_f$ given the past data, $(y, X)$, $X_f$, and assumptions. First we assume that $E(u_f \mid D') = 0$ which expresses the belief that there is no systematic element in the future error vector. Second, we shall assume that the future, as yet unobserved error

[6] See Zellner and Hong (1989) and Hong (1989) for analysis using a traditional Bayesian approach.

terms each have the same variance $\sigma^2$ and are mutually uncorrelated and also uncorrelated with the elements of $\beta$.[7] With these assumptions, the first two moments of $y_f$ given $D' = (y, X, X_f)$ are

$$(22) \qquad E(y_f \mid D') = X_f\hat{\beta}$$

and

$$(23) \qquad \mathrm{Var}(y_f \mid \sigma^2, D') = Ms^2;$$

where $M \equiv I_q + X_f(X'X)^{-1}X_f'$. The proper predictive maxent post-data density for $y_f$ subject to (22) and (23) is

$$(24) \qquad f(y_f \mid D') \sim N(X_f\hat{\beta}, Ms^2),$$

which can be used to compute marginal densities, predictive intervals, and other quantities of interest.

In addition to the result in (24) that parallels that for estimation in (14) we can also derive the following maxent conditional predictive density for $y_f$ given $D'$ that incorporates the conditional moments $E(y_f \mid \sigma^2, D') = X_f\hat{\beta}$, and $\mathrm{Var}(y_f \mid \sigma^2, D') = M\sigma^2$:

$$(25) \qquad f(y_f \mid \sigma^2, D') \sim N(X_f\hat{\beta}, M\sigma^2).$$

This conditional normal density can be multiplied by any of the marginal post-data maxent densities for $\sigma^2$ that are shown in Table 1. The marginal predictive density of $y_f$ can be computed numerically by drawing a value of $\sigma^2$ from its marginal density, inserting this drawn value into (25), and drawing a vector $y_f$ from the conditional normal density. Repeating this procedure will provide draws from the marginal density and thus enable calculation of predictive intervals, moments, etc. of $y_f$. Also, we can compute traditional Bayesian predictive densities based on, say, normal sampling assumptions and diffuse priors. As will be shown in the next section, posterior odds can be employed to evaluate alternative models and to provide a means for combining alternative models and their predictions as discussed and applied in a traditional Bayesian framework by Min and Zellner (1993). The BMOM predictive densities that we shall consider are shown in Table 3. We present density

[7] If the future errors were assumed to have non-zero means and were correlated, we could incorporate this information into our analysis.

functions (when available) for the case in which $y_f$ is a scalar. As discussed in Table 2. BMOM(N) and TB results generalize to the multivariate case.

**Table 3** [8]

**Restrictions and Post-Data Predictive Densities for a scalar $y_f$**

| Model | Restrictions | Density Function |
|---|---|---|
| BMOM(N) | $E(y_f \mid D) = \hat{y}_f, \quad \mathrm{Var}(y_f \mid D) = s_e^2$ | $\mathrm{N}(\hat{y}_f, s_e^2)$ |
| BMOM(1) | $\int_0^\infty p_N(y_f \mid \sigma^2, D) p_{EXP}(\sigma^2 \mid D) d\sigma^2$ | $\propto \frac{1}{s_e} \exp\left(-\frac{\sqrt{2}\lvert y_f - \hat{y}_f \rvert}{s_e}\right)$ |
| BMOM(2) | $\int_0^\infty p_N(y_f \mid \sigma^2, D) p_{TN}(\sigma^2 \mid D) d\sigma^2$ | Evaluate Numerically |
| TB | $p(y_f \mid D') = \int p_N(y_f \mid \beta, \sigma^2, D') p(\beta, \sigma^2 \mid D) d\beta d\sigma^2$ | Student-t $(\hat{y}_f, \frac{v}{v-2} s_e^2, v)$ |

Given the alternative BMOM models for a scalar $y_f$ in Table 3, a vector $y_f$, or predictive densities based on other moment side conditions, there is a need for model comparison and selection methods, a problem which we take up in the following section.

# 3 Model Comparison and Selection Techniques

If we have observed the values of the elements of $y_f$, the predictive post-data densities discussed above can be utilized to compute Bayes factors. That is, for two alternative

[8] Here, as in Table 2, BMOM(i), $i = 1, 2$ denotes the use of a maxent density for $\sigma^2$ with the use of just a first moment constraint and first and second moment constraints, respectively. BMOM(N) is the normal maxent post-data density using conditions (22) and (23). TB utilizes a normal likelihood function, diffuse prior and Bayes rule to obtain the joint posterior, $p(\beta, \sigma^2 \mid D)$. Multiplying this density by the conditional normal for $y_f$ given $\beta$ and $\sigma^2$ and integrating out the nuisance parameters $\sigma^2$ and $\beta$ leads to a marginal predictive density for $y_f$ in the univariate Student-t form. The BMOM(IG) density will also be of the Student-t form with the Lagrange multipliers entering as parameters of the density. We have also defined $s_e^2 = s^2(1 + x_f(X'X)^{-1}x_f')$.

predictive densities $f_1(y_f \mid D')$ and $f_2(y_f \mid D')$, the TB posterior odds, $K_{12}$ is given by

$$(26) \qquad K_{12} = \frac{\pi_1 f_1(y_f \mid D')}{\pi_2 f_2(y_f \mid D')},$$

namely the product of the prior odds $\pi_1/\pi_2$ times the Bayes factor. On inserting $y_f = y_f^0$, the observed value of $y_f$, and a value for the prior odds, a numerical value for $K_{12}$ is obtained; see section 4 for computed examples. Further as Good (1950) and Kullback (1959) have noted, on taking logs of both sides of (26), and averaging with respect to $f_1(y_f \mid D')$, the following result is obtained

$$(27) \quad W_{12} = \int \log K_{12} f_1(y_f \mid D')dy_f - \log\frac{\pi_1}{\pi_2} = \int f_1(y_f \mid D')\log\frac{f_1(y_f \mid D')}{f_2(y_f \mid D')}dy_f.$$

Thus, the averaged log posterior odds minus the log prior odds, or "the weight of the evidence", is equal to the cross entropy of $f_1$ with respect to $f_2$, and is denoted by $CE(f_1, f_2)$, a non-negative measure of the distance between $f_1$ and $f_2$. Further, from (27) we have

$$(28) \qquad W = W_{12} + W_{21} = CE(f_1, f_2) + CE(f_2, f_1).$$

The quantity $W$ in (28) above is symmetric with respect to $f_1$ and $f_2$ and is well-known as the Jeffreys-Kullback-Leibler distance measure. This derivation illustrates the connection between TB model comparison methods and the concept of entropy and cross entropy. We make use of the Jeffreys-Kullback-Leibler metric in computing distances between alternative post-data densities of the scale parameter of section 4.

From (26), $K_{12} = P_1/P_2$, where $P_i$ is the posterior probability associated with model $f_i(y_f \mid D')$, $i = 1, 2$. Since these two models are not exhaustive, $P_1 + P_2 < 1$. Even so, in the standard two-action, two-state model selection problem, (see *e.g* DeGroot (1970)), if the loss structure is symmetric and $K_{12} = P_1/P_2 > 1$, then it is optimal in terms of minimizing expected loss to choose $f_1(y_f \mid D')$. If $K_{12} = P_1/P_2 < 1$, $f_2(y_f \mid D')$ is the optimal choice. See Palm and Zellner (1992) and Min and Zellner (1993) for application of these techniques in choosing between or combining alternative forecasting models. Of course, loss functions such as those described in Dehling et al (1996) can also be employed in evaluating expected losses associated with choices of alternative densities. Other model selection criteria can also be employed; see Judge et al (1985) for a discussion of alternative criteria.

We now turn to presenting applications of the techniques described in this and the previous sections to illustrate how BMOM can be implemented in practice.

# 4 Computed Examples

We apply the techniques of the previous section to the Haavelmo (1947) model of estimating the marginal propensity to consume [9]. Haavelmo estimates the following reduced form "investment multiplier" equation that is in the form of a simple regression model,

$$y_t = \beta_1 + \beta_2 x_t + u_t,$$

where $y_t$ is real per-capita income and $x_t$ is real per-capita investment. The original study includes data for the twenty year period 1922 - 1941. These data are presented in Haavelmo (1947) or Zellner (1971).

We begin with a discussion of the post-data densities for the scale parameter, $\sigma^2$. As discussed in section 2, we derive post-data moment conditions for the scale parameter and then use the principle of maximum entropy to obtain a most conservative density function. The Lagrange multipliers associated with the moment side conditions enter this density function and numerical values for them are determined following the algorithm of Zellner and Highfield (1988). Given these results, we compare the traditional Bayes posterior to post-data densities obtained via BMOM. These include the exponential (BMOM(1)), truncated normal (BMOM(2)), and the BMOM inverted gamma (BMOM(IG))[10]. These results are provided in Figure 1 of the appendix.

[9] We also perform analysis of the model using generated data which satisfy the iid normal error term condition. That is, we estimate the model using least squares and then generate a vector of $N(0, s^2 I_n)$ error terms. We then generate the per-capita consumption vector using the estimated regression coefficients and the drawn normal errors. This case is carried along with the analysis of the original data in order to abstract from the possibility of serial correlation in the error terms. The problem of serial correlation is taken up in Zellner and Sacks (1996) and Zellner (1997b), and is not addressed here. We find that the results of the generated data and the original data do not depart significantly. In what follows, we present only results for the original data set.

[10] When integrations which do not have a known analytical solution were required, the integrations were estimated numerically in MATLAB using the quadrature routine, "quad8".

We observe that by imposing an additional side condition and thus reducing entropy, the truncated normal distribution is more informative or "spiked" than the exponential post-data density. BMOM(IG) also appears more informative than the traditional Bayesian inverted-gamma counterpart, which results from the departure of the Lagrange multipliers $\lambda_1$ and $\lambda_2$ from parameters of the TB(IG), $v+2$ and $vs^2/2$, respectively. This rich class of density functions for the scale parameter will then translate into a variety of possible post-data densities for regression coefficients as well as predictive densities. We will use posterior odds to select the model which is best supported by the data and thus choose the model with the appropriate amount of information regarding the scale parameter, $\sigma^2$. We now present several moments for the post-data densities for $\sigma^2$ in Table 4. We also compute distance measures for these densities using integrated absolute distance, integrated square distance, and the Jeffreys-Kullback-Leibler distance measure as alternative metrics. These tables are provided in the appendix.

**Table 4 [11]**
**Moments of Alternative Post-Data Densities**
**for $\sigma^2$ in Haavelmo Model**

| | Trad. Bayes | BMOM(1) | BMOM(2) | BMOM(IG) |
|---|---|---|---|---|
| $E(\sigma^2 \mid D)$ | 745 | 663 | 663 | 674 |
| $\text{STD}(\sigma^2 \mid D)$ | 281 | 663 | 66.6 | 87.9 |
| $E(\sigma^6 \mid D)$ | $6.28 \times 10^8$ | $1.75 \times 10^9$ | $3.00 \times 10^8$ | $3.21 \times 10^8$ |
| $E(\sigma^8 \mid D)$ | $7.39 \times 10^{11}$ | $4.62 \times 10^{12}$ | $2.04 \times 10^{11}$ | $2.27 \times 10^{11}$ |
| $E(\log \sigma^2 \mid D)$ | 6.55 | 5.90 | 6.49 | 6.51 |
| $E\left(\frac{1}{\sigma^2} \mid D\right)$ | $1.51 \times 10^{-3}$ | $8.87 \times 10^{-2}$ | $1.53 \times 10^{-3}$ | $1.51 \times 10^{-3}$ |

From Table 4 we see that $E(\sigma^2 \mid D)$ is the same under BMOM(1) and BMOM(2) which must be the case since the mean enters as a moment side condition in both maxent densities. BMOM(IG) may not possess the same mean as BMOM(1) and BMOM(2) since BMOM(IG) is the proper maxent density satisfying $E(\log \sigma^2 \mid D) = \mu_1$ and $E(1/\sigma^2 \mid D) = \mu_2$ and does not conform to a mean restriction. For this particular application with 20 observations and

[11] "STD" denotes standard deviation. The log and reciprocal moments for all models and all moments for BMOM(2) were computed numerically.

two regression coefficients, $v = 18$, and thus the mean of the traditional Bayes posterior density for $\sigma^2$, $vs^2/(v-2)$, is much larger than $s^2$, the post-data mean of BMOM(1) and BMOM(2). By imposing an additional side condition and thus adding more information to the post-data density for $\sigma^2$, the post-data variance for $\sigma^2$ under BMOM(2) is much smaller than the post-data variance under BMOM(1). It is also interesting to note from Table 4 the similarity in log and reciprocal moments between BMOM(IG) and the traditional Bayes inverted gamma. Graphs of these four densities are provided in Figure 1 of the appendix.

We now focus upon estimation of $\beta_2$ in the regression model. Following the results of section 2 we obtain the traditional Bayes. BMOM(1), BMOM(2), and BMOM(N) densities and plot them in Figure 2. The BMOM(IG) density is not carried along in this analysis. The resulting posterior density will be similar in functional form to the traditional Bayes result. and a description of how to compute this density was provided in section 2. With respect to the BMOM(2) post-data density for $\beta_2$. we have previously demonstrated how this density must be evaluated numerically. We draw from the truncated normal for $\sigma^2$, substitute these draws into the conditional normal for $\beta_2$ and then draw from this density to obtain draws from the marginal density for $\beta_2$. Given these draws, we estimate the density nonparametrically via kernel density estimation. We choose a Gaussian kernel and use a fixed bandwith $h_n = .18$. For a discussion of kernel density estimation and bandwith selection techniques. see Silverman (1986) and Devroye and Györfi (1985).

The object of interest in this analysis may not be the regression coefficient $\beta_2$, but rather a transformation of the regression coefficient which defines the marginal propensity to consume or save. We note that the marginal propensity to consume is related to the regression coefficient $\beta_2$ by the transformation $\beta_2 = 1/(1 - MPC)$, and similarly, the marginal propensity to save is defined by $\beta_2 = 1/(MPS)$. Provided that the functional form of the post-data density for $\beta_2$ is known. we can derive the post-data density for $MPC$ by a simple change of variable. Since the analytical form of the BMOM(2) density is unknown, the post-data density for $MPC$ was evaluated numerically. That is, for each draw from the post-data density for $\beta_2$, we compute the corresponding value from $MPC$. Given these draws. the $MPC$ post-data density was then obtained via kernel density estimation using a Gaussian kernel with fixed bandwith $h_n = .022$. We present the alternative post-data densities for the marginal propensity to consume in Figure 3. [12]

Last, we compute alternative predictive densities for the dependent variable given that the future value of the regressor is equal to the sample mean. The graphs are provided in Figure 4 of the appendix. Again, the BMOM(2) density was obtained numerically via Monte Carlo integration and kernel density estimation using a Gaussian kernel with fixed bandwith $h_n = 8.3$.

As described in previous sections, use of BMOM and Traditional Bayesian techniques provides a rich class of predictive densities. We can then choose among the alternative densities which incorporate varying amounts of information in the data by computing posterior odds as described in section 3. For this particular problem we split the sample and estimated the models using the first fifteen observations. Given the estimation results, past data $D$ and future regressors $X_f$, we derived the multivariate predictive density for each model considered. The ordinate of the predictive density is then evaluated using the remaining five values from the sample [13]. We compute posterior odds by placing equal prior probability on all models considered and evaluating the Bayes factors. We present the computed posterior odds in the Table 5 below.

---

[12]. It is interesting to discuss point estimation in the context of this problem. Let us consider the problem of estimating the marginal propensity to save, which is given as $\beta_2 = 1/MPS$. The MLE estimate for this problem. $1/\hat{\beta}_2$ does not posses finite moments, and thus has infinite risk relative to quadratic and other loss functions. To avoid this problem, we can introduce the relative squared error loss function, $L\left((\theta - \hat{\theta})/\theta\right)^2$, where $\theta = 1/\beta_2$, and $\hat{\theta}$ is our estimate. Zellner and Park (1979) discuss this problem extensively and show that the optimal point estimate is $\hat{\theta} = z^2/\hat{\beta}_2\left(1 + z^2\right)$, where $z$ is the sampling theory test statistic for testing $\beta_2 = 0$. For our problem, the t-statistic is large, and hence both methods of estimation provide estimates for the marginal propensity to save around .3. See also Diebold and Lamb (1997) for an interesting analysis of the problem of estimating ratios of parameters.

[13]We do not focus attention here on alternative procedures for splitting the sample.

**Table 5** [14]

**Posterior Odds Calculations Using Last 5 Observations to Compute Bayes Factors**

| | BMOM(N) | Trad. Bayes | BMOM(1) | BMOM(2) |
|---|---|---|---|---|
| BMOM(N) | 1.00 | 1.14 | .760 | 1.62 |
| Trad. Bayes | .881 | 1.00 | .670 | 1.42 |
| BMOM(1) | 1.32 | 1.49 | 1.00 | 2.12 |
| BMOM(2) | .619 | .703 | .471 | 1.00 |

From the table above we conclude that BMOM(2) is the model most favored by the data when the last five observations were used to compute the ordinate of the predictive density[15]. Also, we see that TB and BMOM(N) are very close, with TB being slightly favored over BMOM(N) with posterior odds 1.14. Finally, we observe that BMOM(1) does not appear to be supported by the data relative to BMOM(2), BMOM(N) and TB. Recall that the BMOM(1) predictive density results from averaging the conditional normal for $y_f$ given $\sigma^2$ over the relatively uninformative exponential post-data density for $\sigma^2$. Thus it appears that the incorporation of additional information regarding the variance parameter is supported by the data. These illustrative results indicate that empirical comparisons of alternative models by use of traditional posterior odds are quite operational.

Finally, to investigate how BMOM adjusts to departures from normality, we re-compute the model with generated iid Student-t errors with 4 degrees of freedom. The last two figures of the appendix present the posterior distribution for $\beta_2$ and the predictive density for income using these generated data. The functional forms of the densities are relatively unchanged, but the post-data densities with generated Student-t errors are more spread out than the results presented in figures 2 and 4.

[14]The model in the column of the table is placed in the numerator of the posterior odds calculation and the row is placed in the denominator. For example, the first column computes posterior odds with BMOM(N) in the numerator and alternative models in the denominator. Note that the $(i,j)$ entry of the table is the reciprocal of the $(j,i)$ entry.

[15]Note that the computation of posterior odds for BMOM(2) does not involve any type of kernel estimation. To compute the Bayes factor we substitute the observed vector $y_f^0$ into the conditional normal predictive density, $p_N(y_f \mid \sigma^2, D')$. The ordinate of the marginal density for $y_f$ is then obtained numerically by computing $\int p(y_f \mid \sigma^2, D')p_{TN}(\sigma^2 \mid D)d\sigma^2$ given that $y_f = y_f^0$.

# 5 Conclusion

In this paper we have indicated how to apply the Bayesian Method of Moments in analysis of the standard multiple regression model when information is not available to formulate a likelihood function. On introducing simple assumptions relating to the moments of the realized error terms and the future, as yet unobserved error terms, we derived post-data moments of parameters and future values of the dependent variable. Using these moments as side conditions, proper maxent densities for parameters and future values of the dependent variable were derived that can easily be computed from the data. Further, the methods developed in this paper can be used to extend previous BMOM analyses of models with autocorrelated or heteroscedastic error terms, see e.g Currie (1996), Zellner and Sacks (1996), and Zellner (1997b), where use is made of the Gibbs sampler to compute post-data densities. Last, it was shown how alternative BMOM and TB predictive densities can be compared by use of posterior odds. As shown in computed examples, some BMOM maxent densities are very similar to TB densities, while others are not. Thus it is fortunate that it is possible to use posterior odds to ascertain the extent to which alternative assumptions and models are supported by information in the data.

With respect to future research, we, along with H. Ryu, are applying BMOM techniques to various semi-parametric models such as considered in Ryu (1993). Further attention is being given to understanding how alternative assumptions regarding the form of the dependence between regression coefficients and the variance parameter affects results statistically and economically. It has been recognized that the form of the mean-variance dependency is a critical factor in explaining decision-making under uncertainty in financial economics and other areas as well as in formulating likelihood functions. Last, work relating to multivariate regression similar to that contained in the present paper will extend the BMOM multivariate results in Zellner (1995).

## References

Barnard, G. Personal Communication, 1997.

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

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, University of Chicago, 1996.

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

Devroye, L. and L. Györfi, *Nonparametric Density Estimation : The $L_1$ View* (New York: Wiley, 1985).

Diebold, F. and R. Lamb, "Why are Estimates of Agricultural Supply Response so Variable?," *Journal of Econometrics* 76 (1997), 357–373.

G. Judge, W.E. Griffiths, R. H. Carter, H. L. Lütkepohl and T.-C. Lee, *The Theory and Practice of Econometrics* (New York: John Wiley & Sons, Inc., 1985).

Geweke, J., "Exact Inference in the Inequality Constrained Normal Linear Regression Model," *Journal of Applied Econometrics* 1 (1986), 127–141.

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

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

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

Haavelmo, T., "Methods of Measuring the Marginal Propensity to consume," *Journal of the American Statistical Association* 1 (1947), 105–122.

H.G. Dehling, T.K. Dijkstra, H.J. Guichelarr, W. Schaafsma, A.G.M. Steerneman, T.J. Wansbeek and J.T. van der Zee, "Structuring the Inferential Contest," in D. Berry, K. Chaloner & J. Geweke, eds, *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner* (New York: John Wiley & Sons, 1996).

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

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

Jaynes, E., "Comment on 'Optimal Information Processing and Bayes' Theorem'," *American Statistician* 42 (1988), 280–281.

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

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

Mead, L. and N. Papanicolaou, "Maximum Entropy in the Problem of Moments," *Journal of Mathematical Physics* 25 (1984), 2404–2417.

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.

Silverman, B., *Density Estimation for Statistics and Data Analysis* (New York: John Wiley & Sons, Inc., 1986).

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).

Zellner, A., *An Introduction to Bayesian Inference in Econometrics*, (New York: Wiley, 1971) reprinted in 1987 by Krieger Publishing Co., Malabar, FL., and by Wiley, 1996

——, "Bayesian Analysis of Regression Errors," *Journal of the American Statistical Association* 70 (1975), 138–144.

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

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

——, "The Finite Sample Properties of Simultaneous Equations' Estimates and Estimators : Bayesian and Non-Bayesian Approaches," in L. R. Klein, ed., *Journal of Econometrics: Special edition in honor of Carl Christ* 1995(to appear, 1997 and in Zellner (1997c)).

——, "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, and 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)[info@e-elgar.co.uk]

—— 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.

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

—— and Z.-B. Park, "Minimum Expected Loss (MELO) Estimators for Functions of Parameters and Structural Coefficients of Econometric Models," *Journal of the American Statistical Association* 74 (1979), 185–193.

—— and B. Sacks, "Bayesian Method of Moments (BMOM) Analysis of the Multiple Regression Model with Autocorrelated Errors," H.G.B. Alexander Research Foundation, University of Chicago, 1996.

# 6 Appendix

### Appendix A [16]

**Table 6**
**Distance Between Post-Data Densities for $\sigma^2$. Integrated Absolute Distance**
**Metric :** $d(f_1, f_2) = \int_0^\infty \mid f_1(\sigma^2) - f_2(\sigma^2) \mid d\sigma^2$

| | BMOM(1) | BMOM(2) | TB(IG) | BMOM(IG) |
|---|---|---|---|---|
| BMOM(1) | 0 | 1.45 | .926 | 1.38 |
| BMOM(2) | 1.45 | 0 | .975 | .208 |
| TB(IG) | .926 | .975 | 0 | .821 |
| BMOM(IG) | 1.38 | .208 | .821 | 0 |

**Table 7**
**Distance Between Post-Data Densities for $\sigma^2$. Integrated Square Distance**
**Metric :** $d(f_1, f_2) = \int_0^\infty (f_1(\sigma^2) - f_2(\sigma^2))^2 d\sigma^2$

| | BMOM(1) | BMOM(2) | TB(IG) | BMOM(IG) |
|---|---|---|---|---|
| BMOM(1) | 0 | .0038 | .0010 | .0030 |
| BMOM(2) | .0038 | 0 | .0020 | .0001 |
| TB(IG) | .0010 | .0020 | 0 | .0013 |
| BMOM(IG) | .0030 | .0001 | .0013 | 0 |

[16] It is interesting to note that the distance ordering is not invariant to the choice of metric. All distance measures have BMOM(IG) and BMOM(2) as the "closest" and BMOM(2) and BMOM(1) as the "farthest" apart. Graphs of these densities are provided in Appendix B

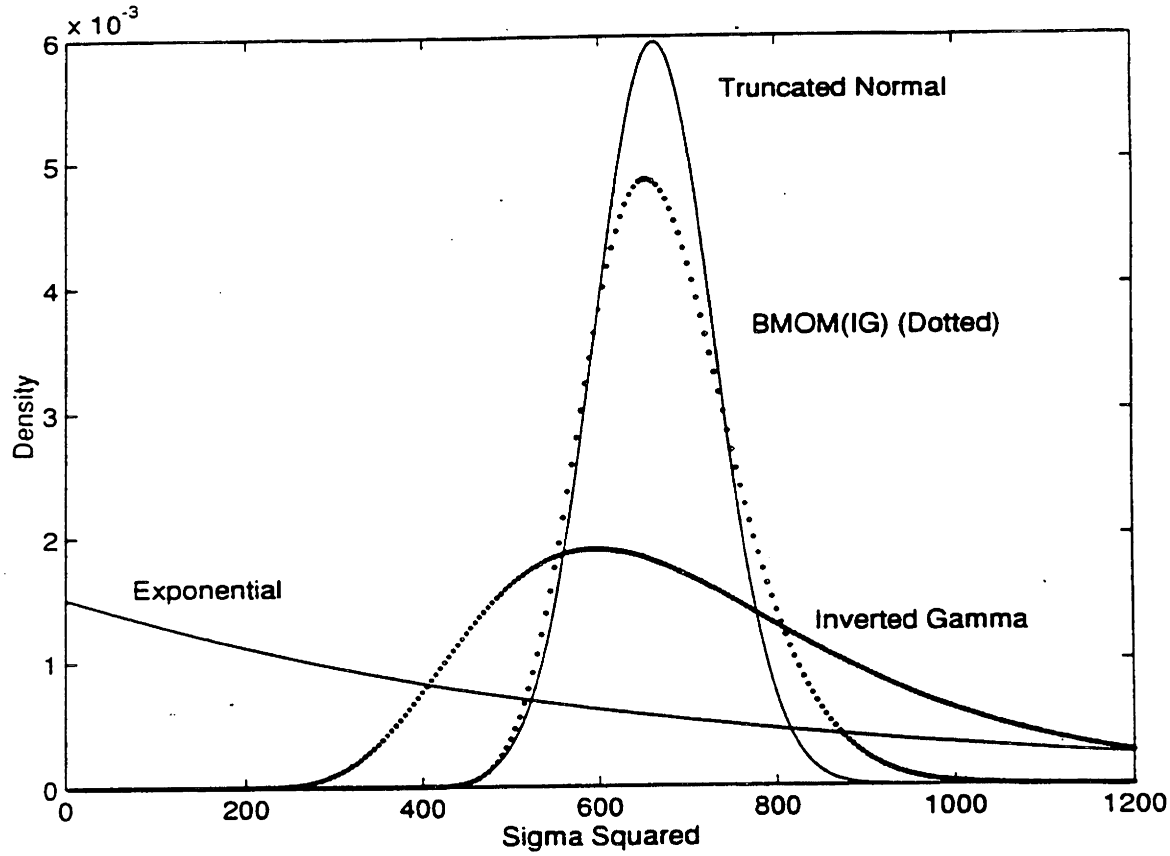
**Table 8**
**Distance Between Post-Data Densities**
**for $\sigma^2$. Jeffreys-Kullback-Leibler Distance Measure**

**Metric** : $d(f_1, f_2) = \int_0^\infty \left[ f_1(\sigma^2) \log\left(\frac{f_1(\sigma^2)}{f_2(\sigma^2)}\right) + f_2(\sigma^2) \log\left(\frac{f_2(\sigma^2)}{f_1(\sigma^2)}\right)\right] d\sigma^2$

| | BMOM(1) | BMOM(2) | TB(IG) | BMOM(IG) |
|---|---|---|---|---|
| BMOM(1) | 0 | 4.02 | .567 | 3.32 |
| BMOM(2) | 4.02 | 0 | 3.43 | .159 |
| TB(IG) | .567 | 3.43 | 0 | 1.81 |
| BMOM(IG) | 3.32 | .159 | 1.81 | 0 |

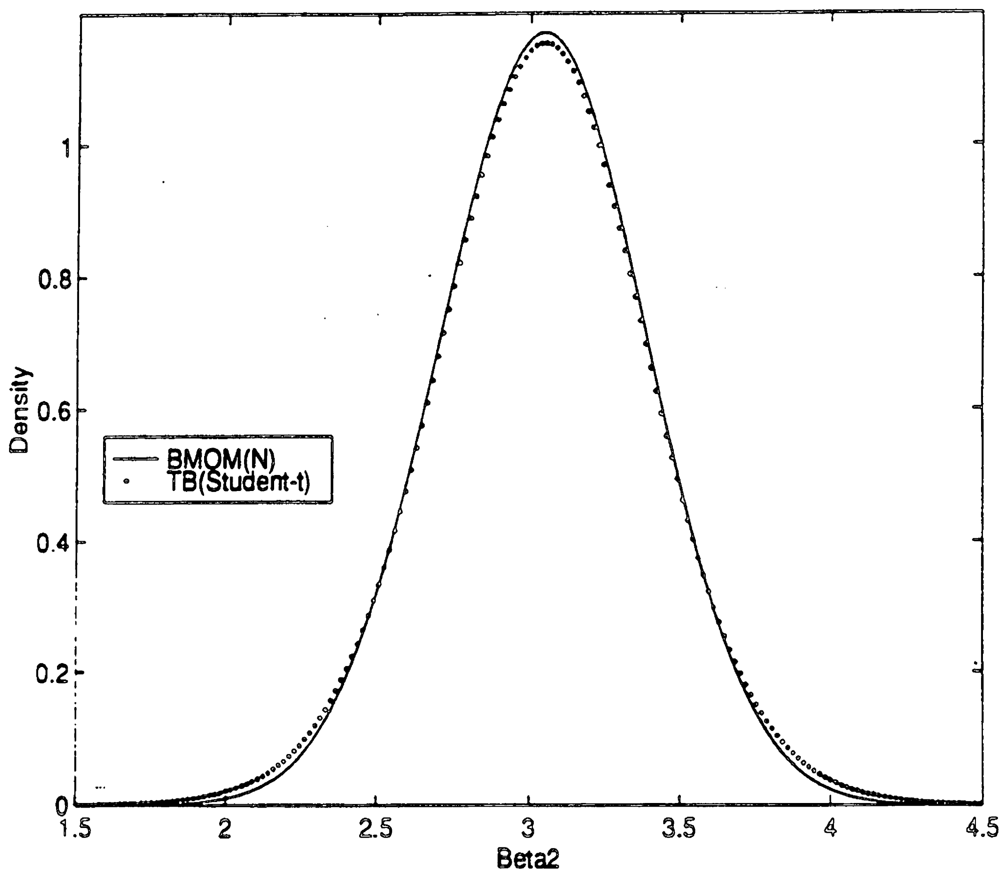
## Appendix B

**Figure 1**
**Traditional Bayes (Inverted Gamma) , BMOM(1) (Exponential), BMOM(IG), and BMOM(2) (Truncated Normal) Post-Data Densities for $\sigma^2$ Using Haavelmo Data [17]**



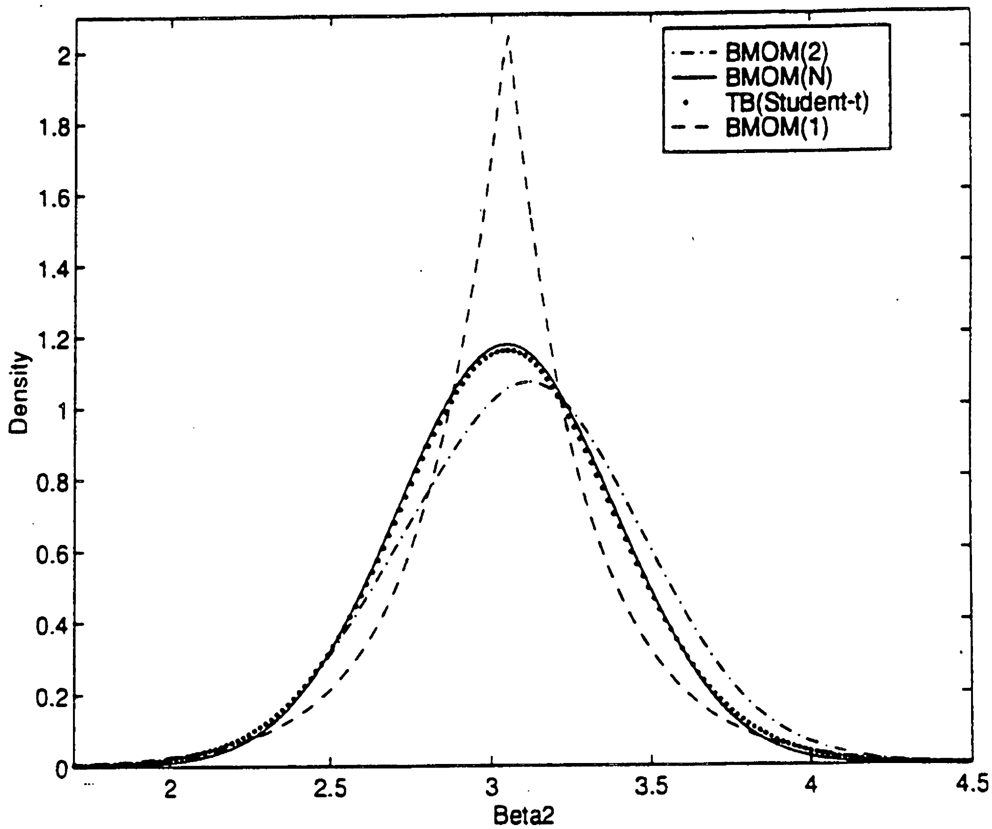

[17] The Traditional Bayes density assumes that the errors are iid normal with common variance $\sigma^2$ and uses the diffuse prior $p(\beta, \sigma) \propto \frac{1}{\sigma}$. BMOM(1) is a proper maxent density which has mean $E(\sigma^2) = s^2$, while BMOM(2) is a proper maxent density satisfying the conditions $E(\sigma^2) = s^2$ and $E(\sigma^4) = \frac{s^4(v^2+2vk)}{n^2-k(k+2)}$. Finally, BMOM(IG) is a proper maxent density that satisfies the moment conditions $E(\log(\sigma^2)) = \mu_1$ and $E(1/\sigma^2) = \mu_2$, where the $\mu_i$ are estimated numerically.

**Figure 2a**
**Traditional Bayes (TB(Student-t)) and BMOM(N)**
**Post-Data Densities for $\beta_2$ Using**
**Haavelmo Data**[18]




[18] The BMOM(N) post-data density imposes that the post-data mean for $\beta$ is $\hat{\beta}$ and the post-data variance-covariance matrix is $s^2(X'X)^{-1}$. The bivariate TB(Student-t) density results from the integration $\int p_N(\beta \mid \sigma) p_{IG}(\sigma) d\sigma$. The marginal densities for $\beta_2$ are obtained from the bivariate densities for $\beta$ and plotted in the figure.

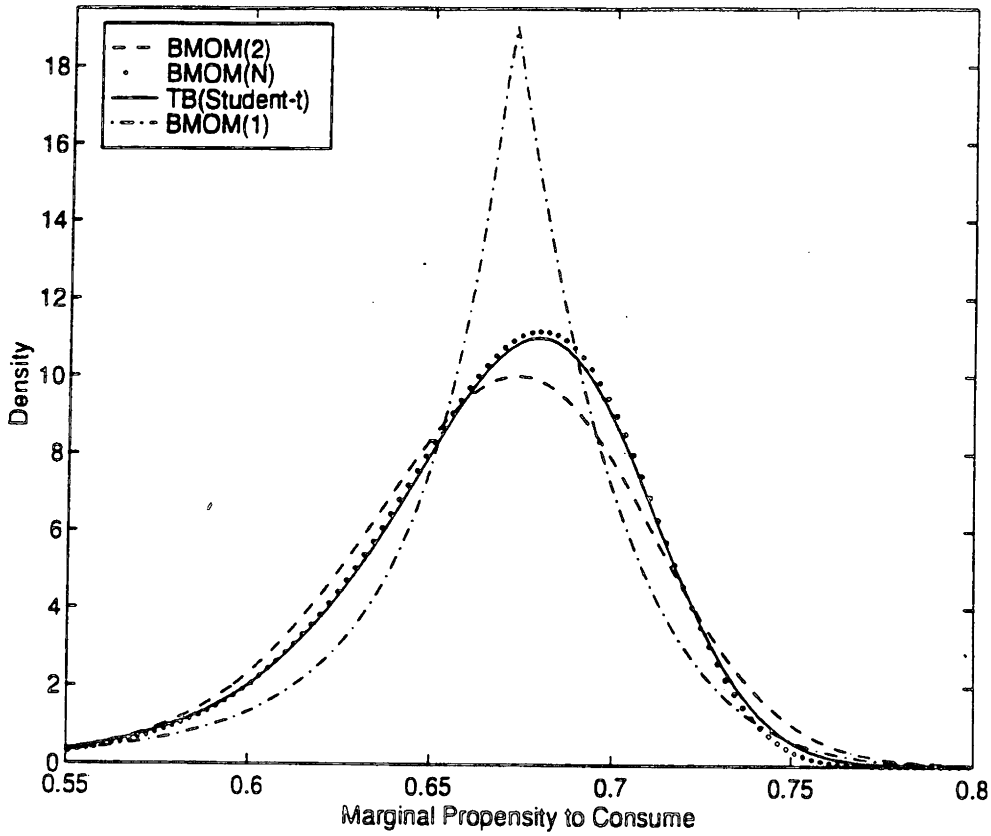
**Figure 2**
**BMOM(N), TB(Student-t) (Traditional Bayes), BMOM(1) (Double Exponential), and BMOM(2) Post-Data Densities for $\beta_2$ Using Haavelmo Data** [19]




---

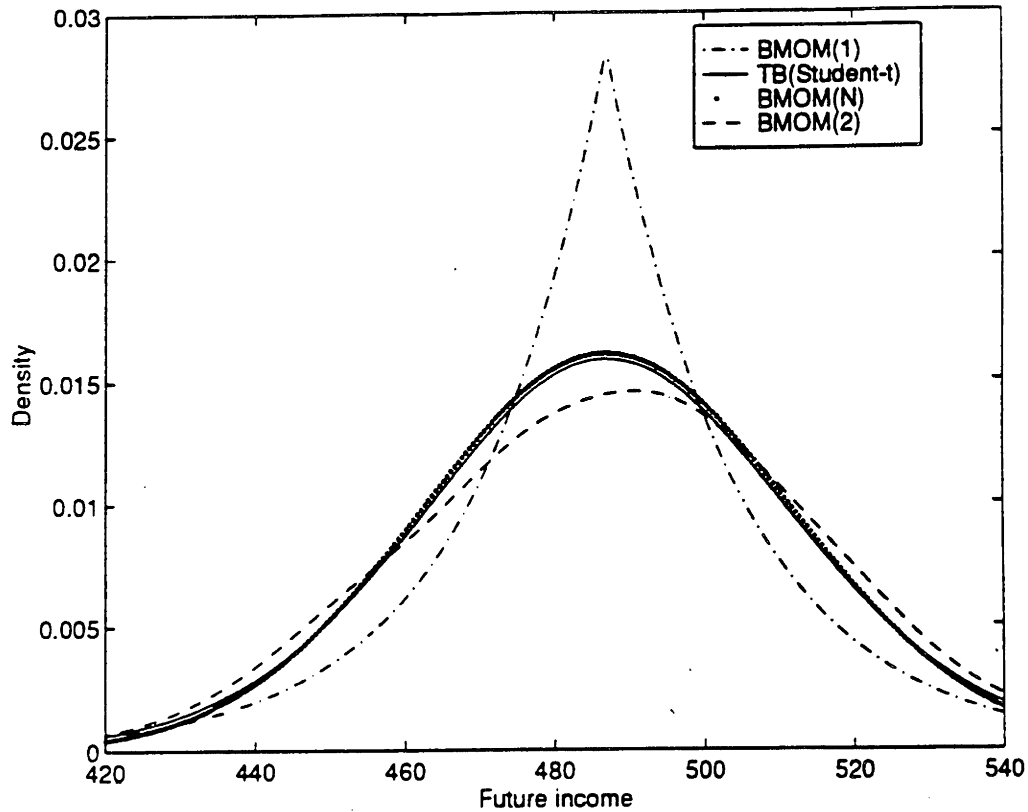
[19]The BMOM(N) post-data density imposes that the post-data mean for $\beta$ is $\hat{\beta}$ and the post-data variance-covariance matrix is $s^2(X'X)^{-1}$. The bivariate TB(Student-t) density results from the integration $\int p_N(\beta \mid \sigma, D) p_{IG}(\sigma \mid D) d\sigma$. From both of these bivariate densities the marginal density for $\beta_2$ is obtained and plotted in the figure. BMOM(1) results from $\int p_N(\beta_2 \mid \sigma^2, D) p_{EXP}(\sigma^2 \mid D) d\sigma^2$, producing the double exponential density. The BMOM(2) post-data density is estimated via direct Monte-Carlo integration and kernel density estimation. The marginal distribution for $\beta$ is obtained by solving the integral $\int_0^\infty p_N(\beta \mid \sigma^2, D) p_{TN}(\sigma^2 \mid D) d\sigma^2$. Under BMOM(2) we draw from the truncated normal, $p_{TN}(\sigma^2 \mid D)$, substitute these draws into the conditional normal for $\beta$, $p_N(\beta \mid \sigma^2, D)$, and obtain 2000 draws from the marginal distribution for $\beta_2$. Given these draws, we estimate post-data density for $\beta_2$ nonparametrically via kernel density estimation using a Gaussian kernel with fixed bandwith $h_n = .18$.

**Figure 3**
**BMOM(N), TB(Student-t) (Traditional Bayes), BMOM(1) (Double Exponential), and BMOM(2) Post-Data Densities for the Marginal Propensity to Consume Using Haavelmo Data** [20]



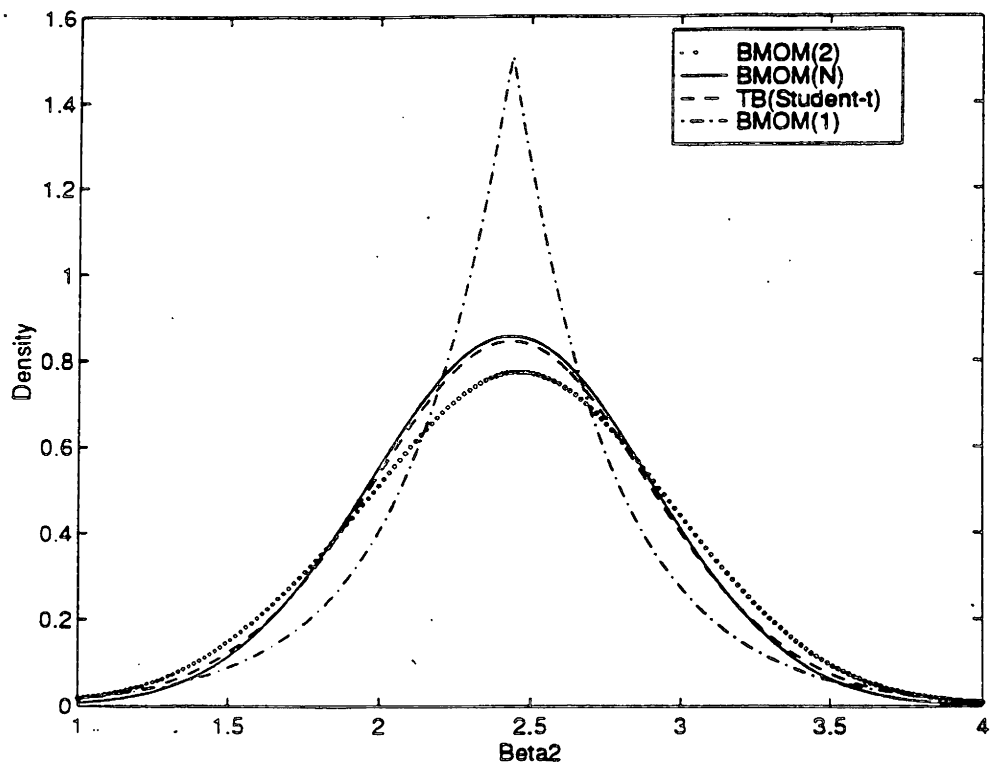

[20] For all posteriors except BMOM(2), an analytical solution for the post-data density for $\beta_2$ is available. Thus. BMOM(N), TB(Student-t), and BMOM(1) follow from the $\beta_2$ posterior densities of Figure 2 using the change of variable, $\beta_2 = \frac{1}{1-MPC}$. For BMOM(2), we compute draws from the MPC posterior given draws from the $\beta_2$ posterior. The post-data density for $MPC$ is then estimated nonparametrically via kernel density estimation using a Gaussian kernel with fixed bandwith $h_n = .022$.

**Figure 4**
**BMOM(N), TB(Student -t) (Traditional Bayes), BMOM(1) (Double Exponential), and BMOM(2) Predictive Densities for Income, $X_f$ = MEAN, Using Haavelmo Data [21]**



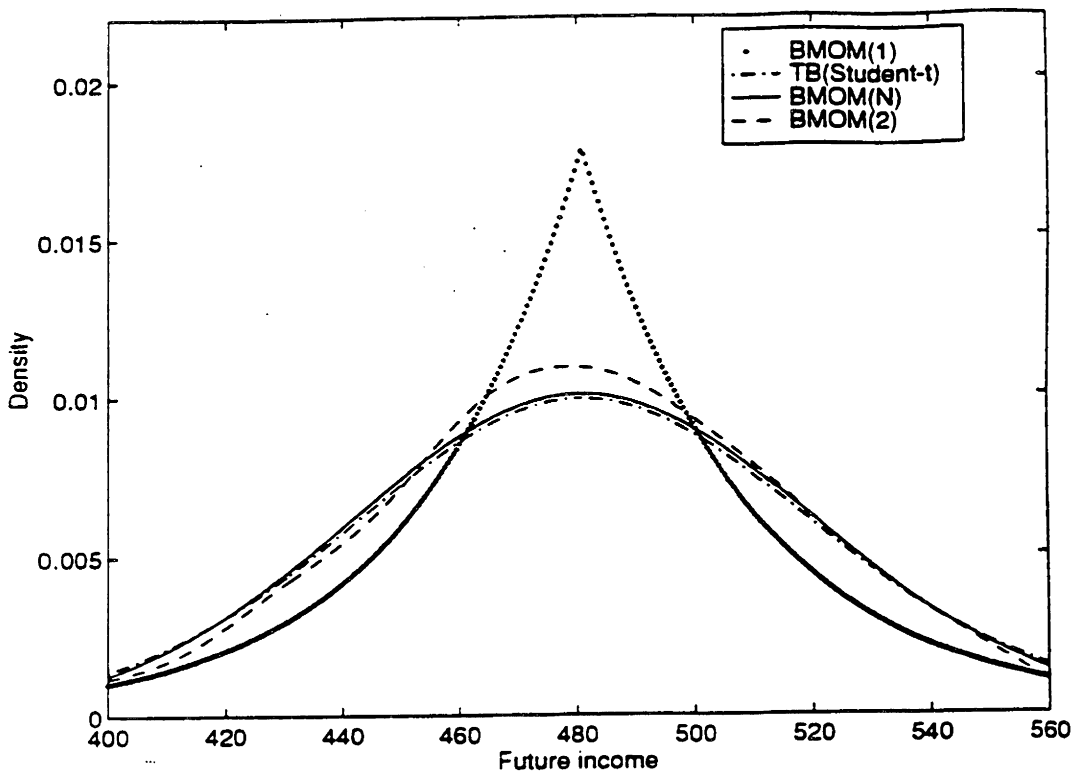

[21] The BMOM(N) post-data density imposes that the post-data mean for $y_f$ is $\hat{y}_f$ and the post-data variance-covariance matrix is $s^2(1+x_f(X'X)^{-1}x'_f)$. The TB(Student-t) density results from the integration $\int\int p_N(y_f \mid \beta, \sigma, D')p(\beta, \sigma \mid D)d\sigma d\beta$, where $p(\beta, \sigma \mid D)$ is the joint posterior density of the parameters which is obtained using a diffuse prior and iid normal likelihood function. Upon integrating $\int p_N(y_f \mid \sigma^2, D')p_{EXP}(\sigma^2 \mid D)d\sigma^2$, the BMOM(1) (double exponential) density is produced. BMOM(2) is estimated via direct Monte-Carlo integration and kernel density estimation. The marginal distribution for $y_f$ is obtained by solving the integral, $\int_0^\infty p_N(y_f \mid \sigma^2 D')p_{TN}(\sigma^2 \mid D)d\sigma^2$. Under BMOM(2), we draw from the truncated normal, $p_{TN}(\sigma^2 \mid D)$, substitute these draws into the conditional normal for $y_f$, $p_N(y_f \mid \sigma^2 D')$, and obtain 2000 draws from the predictive density for $y_f$. Given these draws we estimate the density nonparametrically via kernel density estimation using a Gaussian kernel with fixed bandwith $h_n = 8.3$.

**Figure 5**
**BMOM(N), TB(Student -t) (Traditional Bayes), BMOM(1) (Double Exponential), and BMOM(2) Densities for $\beta_2$ Using Data Generated with Student-t (4 d.f.) Error Terms** [22]




[22]The BMOM(N) post-data density imposes that the post-data mean for $\beta$ is $\hat{\beta}$ and the post-data variance-covariance matrix is $s^2(X'X)^{-1}$. The bivariate TB(Student-t) density results from the integration $\int p_N(\beta \mid \sigma, D) p_{IG}(\sigma \mid D) d\sigma$. From both of these bivariate densities, the marginal density for $\beta_2$ is obtained and plotted in the figure. BMOM(1) results from $\int p_N(\beta_2 \mid \sigma^2, D) p_{EXP}(\sigma^2 \mid D) d\sigma^2$, producing the double exponential density. The BMOM(2) post-data density is estimated via direct Monte-Carlo integration and kernel density estimation. The marginal distribution for $\beta$ is obtained by solving the integral $\int_0^\infty p_N(\beta \mid \sigma^2, D) p_{TN}(\sigma^2 \mid D) d\sigma^2$. Under BMOM(2), we draw from the truncated normal, $p_{TN}(\sigma^2 \mid D)$, substitute these draws into the conditional normal for $\beta$, $p_N(\beta \mid \sigma^2, D)$, and obtain 2000 draws from the marginal distribution for $\beta_2$. Given these draws we estimate post-data density for $\beta_2$ nonparametrically via kernel density estimation using a Gaussian kernel with fixed bandwith $h_n = .27$.

**Figure 6**
**BMOM(N), TB(Student -t) (Traditional Bayes), BMOM(1) (Double Exponential), and BMOM(2) Predictive Densities for Income, $x_f$ = MEAN, Using Data Generated with Student-t (4 d.f.) Error Terms [23]**




[23] The BMOM(N) post-data density imposes that the post-data mean for $y_f$ is $\hat{y}_f$ and the post-data variance-covariance matrix is $s^2(1+x_f(X'X)^{-1}x'_f)$. The TB(Student-t) density results from the integration $\int\int p_N(y_f \mid \beta, \sigma, D')p(\beta, \sigma \mid D)d\sigma d\beta$, where $p(\beta, \sigma \mid D)$ is the joint posterior density of the parameters which is obtained using a diffuse prior and iid normal likelihood function. Upon integrating $\int p_N(y_f \mid \sigma^2, D')p_{EXP}(\sigma^2 \mid D)d\sigma^2$, the BMOM(1) (double exponential) density is produced. BMOM(2) is estimated via direct Monte-Carlo integration and kernel density estimation. The marginal distribution for $y_f$ is obtained by solving the integral, $\int_0^\infty p_N(y_f \mid \sigma^2 D')p_{TN}(\sigma^2 \mid D)d\sigma^2$. Under BMOM(2), we draw from the truncated normal, $p_{TN}(\sigma^2 \mid D)$, substitute these draws into the conditional normal for $y_f$, $p_N(y_f \mid \sigma^2 D')$, and obtain 2000 draws from the predictive density for $y_f$. Given these draws we estimate the density nonparametrically via kernel density estimation using a Gaussian kernel with fixed bandwith $h_n = 8.8$.