Bayesian Econometrics and How to Get Rid of Those Wrong Signs

William E. Griffiths*

In many instances, Bayesian Econometrics offers a more natural interpretation of the results of a statistical investigation than does the sampling theory approach. Furthermore, the Bayesian approach provides a formal framework for incorporating prior information which is frequently available from economic theory. Despite these advantages of the Bayesian approach, applied econometric work has generally been dominated by the sampling theory approach. A simple regression example with one coefficient is used to describe the Bayesian approach using three different priors: a natural conjugate informative prior, a noninformative prior, and a prior with inequality restrictions on the sign and possibly magnitude of the coefficient. The differences between the sampling theory and Bayesian approaches are highlighted. Some practical problems with the first two priors are suggested as possible reasons for the non adoption of the Bayesian approach; it is argued that the inequality restricted prior provides a practical and meaningful alternative which is likely to increase the appeal of the Bayesian approach. The implications are outlined of extending the simple one coefficient model to one where the error variance is unknown and then one where there is an unspecified number of coefficients. An example is provided of how to compute Bayesian inequality restricted estimates using the econometric computer program SHAZAM.

1. Introduction

Although the number of applications of Bayesian econometrics is increasing (see, for example, Zellner 1983, 1984, 1985), it would be fair to say that there are few such applications in many of the applied economic journals, such as this *Review* and the *Australian Journal of Agricultural Economics*. Bayesian decision theory is popular (Anderson, Dillon and Hardaker 1977), but more mundane problems such as estimating regression coefficients, and testing hypotheses about such coefficients, are almost universally handled using sampling theory procedures. This situation is unfortunate because Bayesian methodology is convenient for many types of problems. There are three probable reasons for the restricted use of Bayesian econometrics. The teaching of Bayesian econometrics has not been widespread (although it is growing), the specification of prior information can be difficult or unsettling, and Bayesian options have not been provided on most of the popular econometric software packages.

The purpose of this paper is to increase awareness of Bayesian methodology, some of its advantages, and some of its problems. Some simple examples will be used to illustrate the Bayesian approach and to contrast it with the more conventional sampling theory approach. It is hoped that these examples will show that the use of prior information is not always difficult, nor unsettling, and that software development is well advanced. No material in this paper is new, nor is there any attempt to give a comprehensive review of Bayesian econometrics. Such reviews can be found elsewhere (Zellner 1983, 1984, 1985; Judge *et al.* 1985, Ch. 4).

2. A Simple Example

To introduce the Bayesian approach to inference, it is convenient to begin with a simple artificial example, namely, a regression model with only one explanatory variable and no constant term. This model can be written as:

$$(1) \qquad y_t = x_t\beta + e_t \qquad\qquad t = 1,2,\dots,T,$$

where the usual variable interpretations hold, and where the e_t are assumed to be independent normally distributed random variables with zero mean and constant variance σ^2. Because it will prove useful to give the model some economic content, it will be assumed that equation (1) represents a long-run consumption function where y_t and x_t denote consumption and income in period t, respectively, and β is the long-run marginal propensity to consume. Furthermore, it is convenient to begin with the assumption that σ^2 is known and equal to 1, an assumption that will later be relaxed. Given these assumptions the statistical problem of concern is to learn about the parameter β, given the following hypothetical sample information:

$$(2) \qquad T = 6 \qquad \Sigma y_t^2 = 181.89$$
$$\Sigma x_t^2 = 196 \qquad \Sigma x_t y_t = 186.2.$$

* Department of Econometrics, University of New England.

2.1 Sampling theory approach

The conventional sampling theory approach to this problem is to first find the least squares estimate (which is also the maximum likelihood estimate because e_t is normally distributed):

$$(3) \qquad \hat{\beta} = \frac{\Sigma x_t y_t}{\Sigma x_t^2} = \frac{186.2}{196} \approx 0.95,$$

and the variance of $\hat{\beta}$:

$$(4) \qquad \sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\Sigma x_t^2} = \frac{1}{196}.$$

Then, from the result:

$$(5) \qquad \hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\Sigma x_t^2}\right)$$

it is possible to construct a confidence interval for β. For example, a 95 per cent confidence interval would be given by:

$$(6) \qquad \hat{\beta} - 1.96\sigma_{\hat{\beta}} < \beta < \hat{\beta} + 1.96\sigma_{\hat{\beta}}$$

Substituting sample values yields:

$$0.95 – 1.96/14 < \beta < 0.95 + 1.96/14$$

or

$$(7) \qquad 0.81 < \beta < 1.09$$

Results (3), (4) and (7) constitute the typical sampling theory approach to reporting information about β. Although rather irrelevant for this particular example, it is also customary to present the "significance" of a coefficient in terms of a "t" or "Z" value. In this case, such a value is:

$$(8) \qquad Z = \hat{\beta}/\sigma_{\hat{\beta}} = 0.95 \times 14 = 13.3$$

Alternatively, significance is sometimes reported in terms of "P-values". That is:

$$(9) \qquad \text{Prob}(|Z| > 13.3) = 0.2 \times 10^{-39}$$

For later comparison with the Bayesian approach to reporting results, it is useful to ask why the above sampling theory results are used and what interpretations can be placed on results. First, if just a point estimate is required, the least squares estimator is used because it is minimum variance unbiased. That is, in a large number of future hypothetical samples, the $\hat{\beta}$'s obtained from each of these samples would average out to β and, relative to the estimates provided by any other unbiased estimator, the $\hat{\beta}$'s would vary less around β (in the sense of a lower variance). The important point to note is that the least squares estimator is chosen on the basis of the probability distribution of the estimates it will provide in future hypothetical samples (equation (5)). Probability is defined in terms of the relative frequency of future estimates. The estimate 0.95 given in equation (3) is one drawing from the probability distribution.

A confidence interval is intended to be an indication of the precision of a point estimate, with a wide confidence interval indicating that the sample has conveyed little information about β, and a narrow confidence interval indicating the sample information is precise. It is tempting to interpret the 95 per cent confidence interval (0.81, 1.09) as an interval which contains β with 95 per cent confidence, or 0.95 probability. Such an interpretation is erroneous, however, because, in the sampling theory approach, probability statements cannot be made about the nonrandom parameter β. As Berger (1985, p. 119) points out, in elementary statistics courses it is common to spend a great deal of effort pointing out that a 95 per cent confidence interval is not to be interpreted as an interval that has probability 0.95 of containing β. The correct interpretation is that, if a large number of future hypothetical samples were taken, and in each case a 95 per cent confidence interval was calculated, then in 95 per cent of these cases the calculated interval would contain β. Before a sample is taken, there is a 0.95 probability of obtaining an interval that contains β; once the sample has been taken there is *not* a 0.95 probability that β lies in the interval. This distinction is lost on many students and, indeed, on many nonstatisticians who must use and interpret results provided to them by statisticians. Both these groups invariably find the incorrect interpretation, that there is a 0.95 probability that the interval contains β, more sensible and more natural. Given that the incorrect interpretation seems a sensible and natural one, from a pragmatic standpoint it is reasonable to ask whether there is an alternative methodology which permits such an interpretation. This question will be taken up in the discussion of the Bayesian approach. See Berger (1985, p. 119-120) for an argument along similar lines.

Another difficulty with the sampling theory confidence interval (0.81, 1.09) is that it suggests that β, the long-run marginal propensity to consume, could "reasonably" be as high as 1.09, when there is no doubt in most economists' minds that it could be no greater than 1. The sampling theory approach makes little provision for including such prior information, or for somehow "adjusting" the confidence interval. A possible sampling theory solution is to obtain, via quadratic programming, the least squares estimator subject to inequality restrictions (Judge and Takayama 1966). However, the distribution theory for this estimator has not been developed to the extent that satisfactory confidence intervals can be generally obtained (Geweke 1986a). The problem of confidence intervals including regions of infeasible parameter values is a common one; every time an estimated coefficient which should be positive (say) is "insignificant" the corresponding confidence interval will include some negative values.

The P-value in (9) is the final piece of sampling theory information to be considered. Typically, using a 5 per cent significance level, a P-value greater than 0.05 indicates an estimated coefficient is not significant, a P-value less than 0.05 indicates significance. Implicit in the procedure is a test of the null hypothesis H_0:β=0 against the alternative H_1:β≠0. There is a tendency to incorrectly use P-values as probability statements about the null hypothesis. For example, the P-value of 0.2×10^{-39} given in equation (9) would often be given the interpretation that there is a zero probability that β=0. If the P-value was (say) 0.5, then some might give this the interpretation Prob(β=0) = 0.5; others, suspecting that such statements are not correct, might say there is a "reasonably high probability" that β is zero. In the sampling theory approach none of these interpretations is correct, because each involves probability statements about the coefficient β. Probability statements can only be made about random outcomes of future "experiments", not about what are likely and unlikely values for β. The correct interpretation of equation (9) is that, if a large number of future samples were taken (a very large number!), and if β=0, then the proportion of samples where $|\hat{\beta}/\sigma_{\hat{\beta}}| > 13.3$ would be 0.2×10^{-39}. The natural tendency to use P-values as precise or imprecise probability statements about null hypotheses again raises the question of whether an alternative body of inference which permits probability statements about parameters would be preferable. In this particular example, Prob (β=0) would be of little interest, as would be the significance or otherwise of β. However, probability statements such as Prob (β < 0.9) could be quite useful since information about likely values of the marginal propensity to consume has implications for multiplier effects of government policy. The sampling theory and Bayesian approaches to hypotheses such as β < 0.9 are contrasted later in the paper.

2.2 Bayesian approach

The starting point for the Bayesian approach to analysing and presenting information about β is the construction of a prior probability density function, g(β). In a Bayesian framework, probability is defined in terms of a degree of belief, the probability of an event being given by an individual's belief in how likely the event is to occur. This belief may depend on quantitative and/or qualitative information, but it does not necessarily depend on the relative frequency of the event in a large number of future hypothetical experiments. A consequence of this subjective definition of probability and one of the main features of Bayesian analysis, is that uncertainty about the value of an unkown parameter can be expressed in terms of a probability distribution. It is assumed that, before a sample is taken, an investigator's ideas about what are likely and unlikely values for a parameter can be formalised by assigning to that parameter a prior probability density function. Thus, in the consumption function example, the prior density g(β) represents an investigator's prior knowledge about possible values for the long-run marginal propensity to consume. Such prior knowledge would typically come from economic theory, past studies, or both.

The need to specify a prior distribution is one reason many investigators baulk at the thought of applying Bayesian techniques. It is useful, therefore, to illustrate some difficulties, and solutions, associated with using various prior specifications. Before dealing with each of three priors in turn, Bayesian analysis with the general specification g(β) will be examined.

After formulating the prior g(β) and collecting some sample observations, the next step in a Bayesian analysis is to use Bayes' Theorem to combine the prior information with the sample information to form what is known as the posterior information. The prior information is represented by the prior density function g(β). The sample information is repesented by the joint density function $f(\underline{y}|\beta)$ where $y = (y_1, y_2, ..., y_T)'$ is the vector of sample observations; this density is conditioned on β, because, in the Bayesian subjective probability sense, β is a random variable. The function f(y|β) when viewed as a function of β, given the sample observations, is known as the likelihood function. The posterior information is represented by the posterior density function g(β|y). This function is conditioned on the sample y because it summarises all the information about β, after y has been observed. It can be viewed as a prior distribution, updated to include the information provided by the sample. Because it contains all the information about β, and it is the source of all inferences about β, the attainment and reporting of the posterior density function can be viewed as the endpoint of any statistical investigation. However, it is also customary to report some summary statistics from the posterior density function, such as its mean and standard deviation, and the probability of particular intervals containing β. The reporting of such quantities is similar to the sampling theory approach of reporting joint estimates and confidence intervals, although the meaning is quite different.

Bayes' Theorem is given by the following relationship between conditional and marginal probability density functions

$$(10) \qquad g(\beta|\underset{\sim}{y}) = \frac{f(\underset{\sim}{y}|\beta)g(\beta)}{f(\underset{\sim}{y})}$$

The marginal density function f(y) can be viewed as the average of the conditional functions f(y|β) over all possible values of β, with the prior density function g(β) used as a weighting function. Since f(y) depends only on y, once the sample has been observed f(y) is simply a single number. Consequently, it is common to write Bayes' Theorem as:

$$(11) \qquad g(\beta|\underset{\sim}{y}) \propto f(\underset{\sim}{y}|\beta)g(\beta)$$

or

$$(12) \qquad \text{posterior information} \propto \text{sample information} \times \text{prior information}$$

The additional factor which is required to make the product f(y|β)g(β) *exactly* equal to the posterior density g(β|y) is usually found by computing the constant necessary to make the area under the curve g(β|y) equal to 1.

To summarise, the Bayesian approach to inference consists of (a) formulating a prior density function g(β); (b) using relation (11) to multiply the prior density g(β) by the likelihood function f(y|β) to form the posterior density function g(β|y); and (c) computing, from g(β|y), any summary quantities of particular interest.

For the consumption function in equation (1), the joint density function (or likelihood function) f(y|β) is given by:

$$(13) \qquad f(\underset{\sim}{y}|\beta) = (2\pi)^{-T/2} \exp\{-\tfrac{1}{2}\Sigma(y_t - x_t\beta)^2\},$$

After some algebra and after discarding terms which do not contain β and can be viewed as part of the proportionality constant, equation (13) can be written as:

$$(14) \qquad f(\underset{\sim}{y}|\beta) \propto \exp\{-\tfrac{1}{2}\Sigma x_t^2(\beta - \hat{\beta})^2\}$$

where $\hat{\beta}$ is the least squares estimator. Thus, from relations (11) and (14), the posterior density function for the marginal propensity to consume is:

$$(15) \qquad g(\beta|\underset{\sim}{y}) \propto g(\beta) \exp\{-\tfrac{1}{2}\Sigma x_t^2(\beta - \hat{\beta})^2\}$$

In the next three subsections, three different posterior density functions, each one corresponding to a different specification for the prior density function g(β), will be considered.

2.2.1 Bayesian analysis with an informative prior

Suppose that an investigator's prior information about β is such that he believes there is a 0.9 probability that β lies between 0.75 and 0.95, there is a 50-50 chance that β is above (or below) 0.85, and his prior views can be adequately expressed in terms of a normal distribution. That is, Prob(0.75 < β < 0.95) =0.9, Prob(β < 0.85) = Prob(β > 0.85) = 0.5, and β is normally distributed. Because this prior conveys some definite prior information about β, it is known as an informative prior. Using properties of the normal distribution, and standard normal probabilities, it can be shown that the normal prior distribution g(β) which has these properties has mean and standard deviation given respectively by:

$$(16) \qquad \bar{\beta} = 0.85 \text{ and } \bar{\sigma}_\beta = 0.06079$$

Its density function is:

$$(17) \qquad g(\beta) = (2\pi)^{-\frac{1}{2}}(0.06079)^{-1} \exp\{-(\beta - 0.85)^2/(2 \times 0.06079^2)\}$$

Let $h_0 = 1/\bar{\sigma}^2$ be the reciprocal of the prior variance; h_0 is known as the precision of the prior density. Also, let $h_s = 1/\sigma_{\hat{\beta}}^2 = \Sigma x_t^2$ be the reciprocal of the sampling variance of the least squares estimator; h_s is known as the precision of the sample information. Then, after substituting equation (17) into relation (15), and carrying out some algebra, it can be shown that:

$$(18) \qquad g(\beta|\underset{\sim}{y}) = (2\pi)^{-\frac{1}{2}}\,\bar{\bar{\sigma}}_\beta^{-1} \exp\{-(\beta - \bar{\bar{\beta}})^2/2\bar{\bar{\sigma}}_\beta^2\}$$

where

$$(19) \qquad \bar{\bar{\beta}} = \frac{h_s\hat{\beta} + h_0\bar{\beta}}{h_s + h_0}$$

and

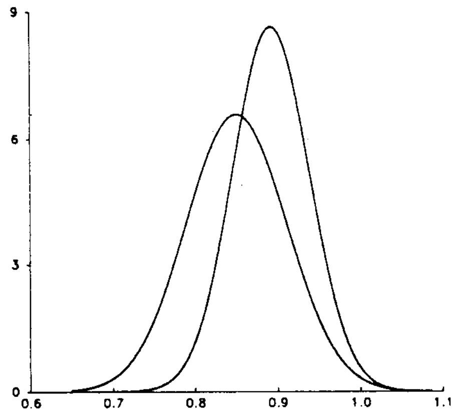
$$(20) \qquad \bar{\bar{\sigma}}_\beta^2 = \frac{1}{h_s+h_0} = \frac{1}{h_1}$$

The posterior density function in equation (18) is a normal distribution with mean $\bar{\bar{\beta}}$ and standard deviation $\bar{\bar{\sigma}}_\beta$. The posterior mean $\bar{\bar{\beta}}$ is a weighted average of the prior mean $\bar{\beta}$ and the least squares (sample) estimator $\hat{\beta}$ with the weights attached to each being given by the precision of each source of information. The posterior variance $\bar{\bar{\sigma}}_\beta^2$ is given by the reciprocal of the precision of the posterior information, this precision being defined as the sum of the prior and sample precisions ($h_1 = h_0+h_s$). Making the calculations for the consumption example yields:

$$(21) \qquad \bar{\bar{\beta}} = 0.892 \quad \text{and} \quad \bar{\bar{\sigma}}_\beta = 0.04629$$

This posterior density function, along with the prior from which it was derived, is graphed in Figure 1. Relative to the prior, the posterior density function is centred more to the right, and has lower variance. These characteristics reflect the fact that the sample information suggests a relatively high marginal propensity to consume, and that the addition of sample information has made knowledge about β more precise.

Figure 1: Informative Prior and Normal Posterior.



The posterior density in Figure 1 represents the current state of knowledge about β, and, as such, it could be regarded as the final step in a Bayesian investigation into the marginal propensity to consume. However, it is often useful to provide some summary quantities derived from the posterior distribution. If a point estimate for β is required, then the Bayesian approach to providing such an estimate is to set up a loss function and to find that value of β which minimises expected loss, where the expectation is taken with respect to the posterior density g(β|y). If the loss function is quadratic (an implicit assumption of sampling theory searches for minimum variance unbiased estimators), then the optimal point estimate is the posterior mean. In this case, $\bar{\bar{\beta}} = 0.892$.

The Bayesian alternative to the sampling theory confidence interval is the highest posterior density (HPD) interval. A 90 per cent (say) HPD interval for β is the shortest interval which, according to the posterior g(β|$\underset{\sim}{y}$), has a 0.9 probability of containing β. It is given by:

$$(22) \qquad \bar{\bar{\beta}} - 1.645\,\bar{\bar{\sigma}}_\beta < \beta < \bar{\bar{\beta}} + 1.645\,\bar{\bar{\sigma}}_\beta$$

or

$$(23) \qquad 0.816 < \beta < 0.968$$

In contrast to the confidence interval approach, the correct interpretation of relations (23) is that there is a 0.9 probability that β lies in the interval (0.816, 0.968). A comparison between this interval, and the corresponding interval from the prior density g(β), (0.75, 0.95), is one indication of the effect of the sample on the prior information.

Various hypotheses may also be of interest. Suppose that it is important to know whether or not β is less than 0.9. The sampling theory solution to this question is to test an appropriate null hypothesis against the corresponding alternative. For example, to test:

$$(24) \qquad H_0: \ \beta \le 0.9 \quad \text{against} \quad H_1: \ \beta > 0.9$$

at the 5 per cent level of significance, we compute:

$$(25) \qquad Z = (\hat{\beta} - 0.9)/\sigma_{\hat{\beta}} = 0.7$$

Then, since 0.7 is less than the critical value of 1.645, we accept H_0, or at least we say there is insufficient evidence to reject H_0. The Bayesian approach to hypothesis testing does not necessarily lead to the rejection or acceptance of either hypothesis. It is more correctly referred to as an approach for comparing hypotheses, although, if suitable loss functions are set up, that hypothesis which minimizes expected loss can be chosen. The Bayesian approach compares two hypotheses by calculating the posterior odds ratio in favour of one hypothesis relative to another. The posterior odds ratio is the ratio of the posterior probability that one hypothesis is true to the posterior probability that the alternative is true. Using standard normal probabilities and the results in equations (18)-(21), the posterior probabilities for each of the hypotheses in relations (24) can be computed as:

$$\text{Prob}(H_0) = \text{Prob}(\beta \leq 0.9) = 0.569 \tag{26}$$

$$\text{Prob}(H_1) = \text{Prob}(\beta > 0.9) = 0.431 \tag{27}$$

The posterior odds ratio in favour of H_0 relative to H_1 is:

$$K = \frac{\text{Prob}(\beta \leq 0.9)}{\text{Prob}(\beta > 0.9)} = 1.32 \tag{28}$$

That is, H_0 is 1.32 times more likely to be true than is H_1. In a Bayesian approach, reporting of the posterior odds ratio can be viewed as the bottom line in the comparison of two hypotheses. Knowing the probability of one hypothesis relative to another conveys much more information than does knowledge of the sampling theory result that a hypothesis has simply been accepted or rejected. The sampling theory result is heavily dependent on which hypothesis is chosen as the null, the sample size, and the level of significance.

The normal prior density function is a mathematically convenient one because it combines nicely with the likelihood function to yield a posterior density function which is also a normal distribution. In general, priors which lead to posteriors belonging to the same family of density functions are known as natural conjugate priors. Because of their mathematical convenience, natural conjugate priors have been popular in much of the Bayesian literature (Raiffa and Schlaifer 1961). However, the apparent need for an investigator to specify an informative prior in general, or a natural conjugate prior in particular, is one of the major reasons why researchers shy away from Bayesian analysis. In the consumption function example, many investigators would feel unhappy specifying their prior information about the marginal propensity to consume in terms of the symmetric, infinite-range normal distribution. It is likely that many more others would feel uncomfortable specifying any kind of informative prior, despite the fact that the marginal propensity to consume is a parameter about which economic theory has a great deal to say. The main areas of concern are likely to be the "accuracy" of the prior information, and the sensitivity of the posterior density function to the prior specification. There is a lack of generality in reporting in the sense that the posterior density function and quantities derived from it, such as the posterior mean and standard deviation in equation (21), and the posterior odds ratio in equation (28), are heavily dependent on the prior specification; different investigators may possess different prior information. The two prior density functions outlined in the next two subsections overcome these problems.

2.2.2 Bayesian analysis with a noninformative prior

The purpose of Bayesian analysis with a non-informative prior is to provide a posterior density function in which the sample information is dominant. If the prior density function has little or no bearing on the shape of the posterior density function, then questions such as the dependence of the results on the subjective opinions of a single researcher do not arise. Zellner (1971) and Box and Tiao (1973) make extensive use of non-informative priors.

A completely noninformative prior for β would allow for values of β anywhere from $-\infty$ to $+\infty$ and, roughly speaking, would treat all values as equally likely. A prior density function with these properties is the uniform density:

$$g(\beta) = 1 \qquad -\infty < \beta < \infty \tag{29}$$

This density function is an improper one. That is, the total area under the density is infinite rather than unity. Under these circumstances, calculation of prior probabilities for regions of the parameter space for β is meaningless. However, as we shall see, the posterior density function derived from this prior is a proper one and, furthermore, possesses the desired characteristic of being dominated by the sample information.

Substituting equation (29) into relation (15) and inserting the appropriate proportionality constant yields the posterior density function:

$$(30)\qquad g(\beta|\underset{\sim}{y}) = (2\pi)^{-\frac{1}{2}}\ (\Sigma x_t^2)^{\frac{1}{2}}\ \exp\{-\tfrac{1}{2}\Sigma x_t^2(\beta - \hat{\beta})\}$$

This density function is a normal one, with mean $\hat{\beta}$, the least squares estimator, and variance $(1/\Sigma x_t^2)$. That is:

$$(31)\qquad (\beta|\underset{\sim}{y}) \sim N(\hat{\beta},\ 1/\Sigma x_t^2)$$

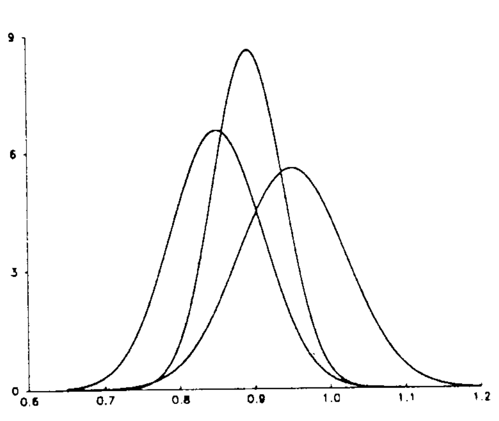
The similarity between this result, and the sampling theory result:

$$(32)\qquad \hat{\beta} \sim N(\beta, 1/\Sigma x_t^2)$$

indicates that the posterior density function does, indeed, contain only sample information. Although relations (31) and (32) look similar, their interpretations differ, and the advantages of the interpretation in relation (31) should not be overlooked. With the sampling theory result in relation (32), probability statements are made about possible values of $\hat{\beta}$ *before* a sample is taken. Once the sample has been observed, it is no longer possible to make probability statements. In particular, it is not possible to make probability statements about the values of β which are likely to have generated the sample. In contrast, the Bayesian result in relation (31) makes it possible to make probability statements about possible values of β *after* a sample has been taken; what is important is the information contained in the data collected so far, not what information might be contained in future hypothetical samples.

Figure 2 contains the posterior density described by relation (31), superimposed on Figure 1, the informative prior and corresponding posterior. Drawing the three densities on one diagram clearly illustrates the difference between the posterior density derived from an informative prior and that derived from a noninformative prior. It also shows how both the sample information (illustrated by the posterior density from a noninformative prior), and the prior information, contribute to the posterior density from the informative prior.

Figure 2: Informative Prior, Corresponding Normal Posterior, and Posterior from Noninformative Prior.



As before, the posterior density can be used to provide summary quantities of interest. With quadratic loss, the Bayesian point estimate is $\hat{\beta} = 0.95$, the same point estimate provided by the sampling theory least squares estimator. This correspondence between the sampling theory and Bayesian results holds as long as the assumption that the equation errors are independent, identically distributed normal random variables is maintained. When heteroscedasticity or autocorrelation exists, the mean of the Bayesian posterior density is not equal to a feasible generalised least squares or maximum likelihood estimator.

The 95 per cent HPD interval from relation (31) is (0.81, 1.09), an interval identical to the sampling theory 95 per cent confidence interval. However, the interpretation of the HPD interval is a pragmatically more reasonable one. It implies there is a 0.95 probability that β lies in the interval (0.81, 1.09). If the probability content of other intervals is of interest, then such probabilities can be calculated using standard normal probability tables. For example, $\text{Prob}(\beta \leq 0.9) = 0.242$ and $\text{Prob}(0 < \beta < 1) = 0.758$. The first of these probabilities can be used, as we did with the posterior derived from the informative prior, to compute the posterior odds in favour of H_0: $\beta \leq 0.9$ relative to H_1: $\beta > 0.9$. In this case the posterior odds ratio in favour of H_0 is:

$$(33)\qquad K = \frac{0.242}{0.758} = 0.319$$

Alternatively, we can say that the posterior odds ratio in favour of H_1: $\beta > 0.9$ is $(0.319)^{-1} = 3.13$. It is interesting to compare this sample dominated Bayesian result with the sampling theory result for testing H_0 against H_1. We note that the sampling theory approach led to acceptance of H_0 despite the fact that it is over 3 times more likely for H_1 to be true. Thus, implicit in the sampling theory approach, there is a high cost associated with accepting an incorrect H_1.

The probability that β lies in the interval (0, 1) is likely to be of particular interest to a researcher. Most investigators would insist that β cannot lie outside the interval (0, 1). That is, they would hold the prior view $\text{Prob}(0 < \beta < 1) = 1$. Because the completely noninformative prior allows for β to be in the range $(-\infty, +\infty)$, it makes no allowance for the stronger prior view, and it is possible for the posterior probability that β lies in the interval (0, 1) to be less than 1. The statement $\text{Prob}(0 < \beta < 1) = 0.758$ is based only on sample information; it gives the probability that β lies in the feasible range when no prior information has been provided. Alternatively, it can be viewed as a sample-based probability of the prior information being "correct", when the prior information is that β lies somewhere in the interval (0, 1). This discussion suggests that it would be good to have a prior density function which includes the information that β lies between 0 and 1, but which is not so informative that it leaves many investigators uneasy about its specification. Before considering such a prior, however, it is convenient to use the example in this subsection to introduce the concepts of numerical integration and Monte Carlo numerical integration.

Numerical and Monte Carlo numerical integration

Consider the statement $\text{Prob}(0 < \beta < 1)$ which has just been discussed. This probability is given by the area under the posterior density function $g(\beta|y)$ between the points 0 and 1, or by the following integral:

$$(34) \qquad \text{Prob}(0 < \beta < 1) = \int_0^1 g(\beta|\underline{y})d\beta$$

Fortunately, when β is normally distributed, there is no need to evaluate this integral to compute the required probability. Others have already evaluated corresponding integrals associated with the standard normal distribution; the resulting probabilities have been tabulated, and are found in all statistics textbooks. However, it is not possible to evaluate integrals like the one in equation (34) analytically. Tabulated probabilities associated with the standard normal distribution have been obtained using numerical integration. Numerical integration is simply a numerical computational technique for finding the area under a curve between any two points. It is not a technique which is usually required when the sampling theory approach is adopted because commonly used normal probabilities, t-values, χ^2-values, F-values, *etc* have all been tabulated. However, it is common for the Bayesian approach to require numerical integration, particularly as models become more complicated or when informative priors which are not natural conjugate priors are employed. The need for numerical integration is another reason practitioners have tended to shy away from Bayesian analysis. It is worth emphasizing, however, that numerical integration is not a difficult computational problem; certainly, it is much less difficult than solving a linear programming problem or finding a nonlinear least squares estimate.

A difficulty does arise if numerical integration is required for an integral of dimension greater than 3. For such integrals, computational time can be prohibitive. In these circumstances it is preferable to use Monte Carlo numerical integration. This technique can be illustrated using the integral in equation (34) although, in practice, because the integral is one-dimensional, Monte Carlo numerical integration would not be required. To apply Monte Carlo numerical integration to equation (34), a random sample is artificially drawn from the density $g(\beta|y)$. That is, a sample of observations is drawn (using a random number generator) from a normal distribution with mean $\hat{\beta} = 0.95$ and standard deviation $(\Sigma x_t^2)^{-1/2} = 0.07143$. The proportion of these observations which lie between 0 and 1 is an estimate of $\text{Prob}(0 < \beta < 1)$. Furthermore, the accuracy of this estimate can be controlled by an appropriate choice of the size of the artificial sample. Monte Carlo numerical integration is useful for more than just the estimation of probabilities. Any functions of parameters, such as means, variances, and marginal posterior density functions, can be estimated.

2.2.3 Bayesian analysis for a prior with inequality restrictions

One of the advantages of the Bayesian approach is its ability to formally include prior information. The first prior that was considered was criticised on the grounds that many researchers would feel uneasy about the need to specify their prior views in terms of a normal distribution, and that they would worry about the sensitivity of the posterior density function to changes in the prior. On the other hand, the second prior can be criticised on the grounds that everybody knows the marginal propensity to consume lies between 0 and 1, and this prior information has not been included. The prior considered in this section is designed to overcome both these criticisms. It is given by:

$$(35) \qquad g(\beta) = \begin{cases} 1 & \text{if } 0 < \beta < 1 \\ 0, & \text{otherwise} \end{cases}$$

This prior suggests that only values of β between 0 and 1 are feasible, and that all values within this range are equally likely. Few would argue that a value for β of (say) 0.2 is as likely as a value of 0.8. However, as soon as judgements such as these are quantified, this prior becomes subject to the same criticisms which were levelled at the first prior.

Substituting equations (35) into relation (15) yields the posterior density function:

$$(36)\qquad g(\beta|\underset{\sim}{y}) = \begin{cases} (0.758)^{-1}(2\pi)^{-\frac{1}{2}}(\Sigma x_t^2)^{\frac{1}{2}} \\ \qquad \exp\{-\tfrac{1}{2}\Sigma x_t^2(\beta - \hat{\beta})^2\} \\ \qquad\qquad \text{if } 0 < \beta < 1 \\ 0, \qquad\qquad \text{otherwise} \end{cases}$$

This posterior density function is almost identical to the normal posterior density function given in equation (30). The difference is that the density in equations (36) cannot take values outside the interval (0, 1); it is a truncated normal distribution, truncated at the points 0 and 1. The constant $(0.758)^{-1}$ is included in equations (36) to make the area under the density between 0 and 1 equal to 1; that is, Prob(0 < β < 1) = 1. Recall that, with the posterior density in equation (30), it was found that Prob(0 < β < 1) = 0.758. In equations (36) this value "cancels" with $(0.758)^{-1}$ leaving a probability of 1.

The truncated normal posterior density function is graphed in Figure 3, alongside the non-truncated version of the previous subsection. The truncated normal follows the same shape as the normal, but is higher, so as to include the additional area (probability) which is lost when the distribution cuts off at β = 1. The calculation of summary statistics for this distribution requires somewhat more effort. Probability statements can be obtained by suitable adjustment of tabulated standard normal probabilities, or by numerical integration. In this case Prob(β≤ 0.9) = 0.319, a value slightly greater than that of 0.242 from the non-truncated version. The 95 per cent HPD interval is (0.82, 1.00), relative to (0.81, 1.09) before truncation. There is no corresponding way of adjusting a sampling theory 95 per cent confidence interval to incorporate the prior information that β lies in the interval (0, 1).

When the normal distribution has been truncated, the mode and the mean are no longer identical; for computation of the mean, more weight is placed on values of β less than 1 and values of β greater than 1 do not contribute at all. Thus, if the Bayesian point estimate under quadratic loss is required, special attention has to be devoted to such computation. Likewise, if the standard deviation is of interest, then it too needs to be computed. Formulae for the mean and variance of a truncated normal distribution are given in Johnson and Kotz (1970, pp. 81-83) or, alternatively, both could be obtained by numerical integration. Monte Carlo numerical integration could also be used, although its use would not be seriously entertained in this simple case where numerical integration yields exact results and is not computationally expensive. To see how numerical integration is relevant, note that the posterior mean $\bar{\bar{\beta}}$ and posterior variance $\bar{\bar{\sigma}}_\beta^2$ are defined, respectively, by:

$$(37)\qquad \bar{\bar{\beta}} = \int_0^1 \beta g(\beta|\underset{\sim}{y})d\beta$$

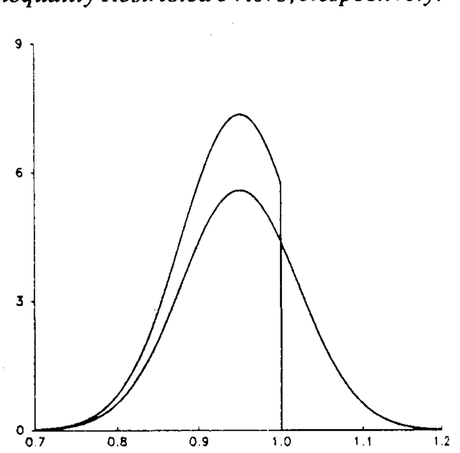
and

$$(38)\qquad \bar{\bar{\sigma}}_\beta^2 = \int_0^1 \beta^2 g(\beta|\underset{\sim}{y})d\beta - \bar{\bar{\beta}}^2$$

In equation (37) the numerical integration process finds the area under βg (β|y) and, for the first term in equation (38), it finds the area under β^2g (β|y). For the truncated normal distribution in Figure 3 these computations yield $\bar{\bar{\beta}}$ = 0.921 and $\bar{\bar{\sigma}}_\beta$ = 0.05192. To see how Monte Carlo numerical integration could be carried out, recall the usual sampling procedure used to estimate the mean μ of some distribution f(z), the mean being defined by:

$$(39)\qquad \mu = \int_{-\infty}^{\infty} zf(z)dz$$

Figure 3: Normal and Truncated Normal Posteriors from Noninformative and Inequality Restricted Priors, Respectively.



In this case a random sample $z_1, z_2, \ldots, z_n$ from the distribution f(z) is observed and the sample mean $\hat{\mu} = \bar{z} = n^{-1} \Sigma^n_{i=1} z_i$ is used as an estimator for μ. A similar procedure is followed for Monte Carlo numerical integration evaluation of $\bar{\bar{\beta}}$ and $\bar{\bar{\sigma}}^2$, except that a random number generator is used to artificially generate a sample. Since random number generators are not generally available for truncated normal random variables, the sample would be generated using the complete normal distribution g(β|y) given in equation (30). Those values falling outside the interval (0, 1) would be discarded, and the Monte Carlo based estimate of β is given by the sample average of those values falling inside (0, 1). For $\bar{\bar{\sigma}}^2_\beta$ in equation (38) the sample average of β^2's is calculated for those β's falling in (0, 1), and $\bar{\bar{\beta}}^2$ is subtracted from the result.

It is instructive to consider a further example of a truncated normal distribution, one which arises when the least squares estimate for the marginal propensity to consume is greater than 1. In general it is common for least squares estimates to fall into what are regarded as infeasible regions, another example being estimates with the wrong signs. Suppose, therefore, the least squares estimate for β is given by:

$$(40) \qquad \hat{\beta} = \frac{\Sigma x_t y_t}{\Sigma x_t^2} = \frac{205.8}{196} = 1.05$$

The posterior density function derived from this sample information, and a completely noninformative prior density, is:

$$(41) \qquad (\beta|\underline{y}) \sim N[1.05, (0.07143)^2]$$

This posterior density has exactly the same variance (and hence exactly the same shape) as the earlier one given in equation (30), but it is now centred at 1.05 instead of 0.95. If the inequality restricted prior which insists that β must lie between 0 and 1 is employed, then the resulting posterior density is a truncated version of relation (41). Both posterior densities are shown in Figure 4. Note that the scale used in this Figure is completely different from that used for the other figures.

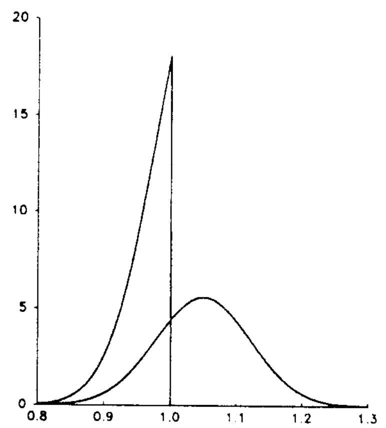
Some summary quantities describing the information about β provided by the truncated posterior density function are:

$$(42) \qquad \bar{\bar{\beta}} = 0.958, \qquad \bar{\bar{\sigma}}_\beta \quad 0.03492,$$

$$\text{Prob}(\beta \leq 0.9) = 0.074$$

$$95\% \text{ HPD interval is } (0.89, 1.00)$$

Figure 4: Normal and Truncated Normal Posteriors when Least Squares Estimate Greater Than 1.



Clearly, the Bayesian results which yield a point estimate of 0.958, and a 95 per cent HPD interval of (0.89, 1.00), are much more reasonable than the sampling theory results which yield a point estimate of 1.05 and a 95 per cent confidence interval of (0.91, 1.19). In fact, it is doubtful whether an investigator would bother reporting the sampling theory results. A possible criticism of the Bayesian approach is that the prior information may not be "correct". Some investigators may lack strength in their convictions about the prior information. If such is the case, then the Bayesian approach can be used to assess the probability that the prior information is correct. Based on the posterior density function from the noninformative prior, equation (41), this probability is Prob(0 < β < 1) = 0.242.

The above approach is a powerful one, it gives a solution to the age-old problem of obtaining least squares regression coefficients with the wrong signs. However, so far the discussion has been restricted to the case of one coefficient and a known disturbance variance σ^2. In the next section the methodology is extended to allow for an unknown σ^2; the general case of several unknown coefficients is considered in the subsequent section.

3. Relaxing the Known Variance Assumption

3.1 Sampling theory approach

Consider again the consumption function example, but assume that the variance σ^2 is unknown. In the sampling theory approach the point estimate for β is the same as before, namely, the least squares estimate $\hat\beta = 0.95$. For the construction of a confidence interval or the testing of hypotheses, an estimate of σ^2 is required, and is given by:

$$(43) \qquad \hat\sigma^2 = \frac{1}{T-1}\left(\Sigma y_t^2 - \hat\beta \Sigma x_t y_t\right)$$

$$= \frac{1}{5}\left(181.89 - 0.95 \times 186.2\right)$$

$$= 1.00$$

The example has been rigged so that this estimate is identical to the assumed value of $\sigma^2 = 1$ of the previous section. As will become clear, rigging the example in this way ensures that the different results in this section can be attributed to use of the t-distribution rather than a different assessment of σ^2. An estimate of the variance of $\hat\beta$ is also the same as before:

$$(44) \qquad \hat\sigma^2_{\hat\beta} = \frac{\hat\sigma^2}{\Sigma x_t^2} = \frac{1}{196}$$

and confidence intervals and hypothesis tests are based on the fact that $(\hat\beta - \beta) / \hat\sigma_{\hat\beta}$ follows a t-distribution with 5 degrees of freedom. A 95 per cent confidence interval is given by:

$$(45) \qquad \hat\beta - 2.571\,\hat\sigma_{\hat\beta} < \beta < \hat\beta + 2.571\,\hat\sigma_{\hat\beta}$$

or

$$(46) \qquad 0.766 < \beta < 1.134$$

This interval is slightly wider than the one given in inequalities (7), reflecting the additional uncertainty created by an unknown σ^2. As before, the interval contains a region of infeasible parameter values.

3.2 Bayesian approach

When there are two unknown parameters, β and σ, the Bayesian approach begins by specifying a joint prior density function g(β, σ) on both these parameters. The likelihood function is written as f(y | β, σ), and a joint posterior density function is obtained using Bayes' Theorem:

$$(47) \qquad g(\beta,\sigma|\underset{\sim}{y}) \propto f(\underset{\sim}{y}|\beta,\sigma)g(\beta,\sigma)$$

Because σ is seldom of interest, the joint posterior density function is not a convenient way of summarising the combined prior and sample information. It is preferable to use the marginal posterior density function for β which is obtained by integrating σ out of the joint density function. That is:

$$(48) \qquad g(\beta|\underset{\sim}{y}) = \int_0^\infty g(\beta,\sigma|\underset{\sim}{y})d\sigma$$

Whether this integral is evaluated analytically or numerically depends on the prior specification and how nicely it combines with the likelihood function. In all the cases considered here analytical integration is possible.

The first step towards specifying a joint prior density function g(β, σ) is to specify a marginal prior density for σ, which is denoted by g(σ). Then, if β and σ are regarded as *a priori* independent—the investigator's views about β do not depend on prior knowledge about σ—the joint prior density is given by:

$$(49) \qquad g(\beta,\sigma) = g(\beta)g(\sigma)$$

On the other hand, if knowledge of σ influences the investigator's views about β, a conditional prior density on β is required, and the joint prior density is given by:

$$(50) \qquad g(\beta,\sigma) = g(\beta|\sigma)g(\sigma)$$

As was the case for β, a prior density for σ can be informative or noninformative. The conventional noninformative one, obtained by treating the distribution of log σ as uniform over (–∞, +∞), is given by:

$$(51) \qquad g(\sigma) \propto \sigma^{-1} \qquad 0 < \sigma < \infty$$

Like the noninformative prior density for β given in equation (29), this prior density is improper. Informative priors for σ will not be considered since it is difficult to imagine too many situations where prior information about σ exists.

The next three subsections will deal with the marginal posterior density functions for β which arise from the three different types of prior information about the marginal propensity to consume, each used in conjunction with equation (51). Before turning to these cases it is useful to give a convenient expression for the likelihood function, namely:

$$(52) \qquad f(\underset{\sim}{y}|\beta,\sigma) = (2\pi)^{-T/2}\sigma^{-T}$$

$$\exp\{-[\Sigma x_t^2(\beta - \hat\beta)^2 + (T-1)\hat\sigma^2]/2\sigma^2\}$$

This expression is obtained by expressing the usual joint normal distribution in terms of the sampling theory estimators $\hat\beta$ and $\hat\sigma^2$. Using Bayes' Theorem, the joint posterior density function is given by:

$$g(\beta,\sigma|\underset{\sim}{y}) \propto g(\beta,\sigma)\sigma^{-T} \exp\{-[\Sigma x_t^2(\beta - \hat{\beta})^2 + (T - 1)\hat{\sigma}^2]/2\sigma^2\} \tag{53}$$

Different versions of this posterior density or, more particularly, different versions of the marginal posterior density g (β | y), obtained by integrating σ out of relation (53), are considered in the next three subsections.

3.2.1 Bayesian analysis with an informative prior

When σ was known, the informative prior for β which combined nicely with the likelihood function because it was a natural conjugate prior was $\beta \sim N(\bar{\beta}, \bar{\sigma}_\beta^2)$. When σ is unknown, the natural conjugate prior is such that the *conditional* prior density for β, given σ, must be a normal distribution. That is, the investigator's prior views about β depend on knowledge about σ. This dependence is made explicit by writing $\bar{\sigma}_\beta^2 = \sigma^2/\tau$ where τ is a prior parameter which controls the prior variance of β for a given σ². Thus, the conditional prior density for β is given by:

$$(\beta|\sigma) \sim N(\bar{\beta},\sigma^2/\tau) \tag{54}$$

Returning to the consumption function example, when σ was known to be equal to 1, the prior density was β ~ N [0.85, (0.06079)²]. Assuming that when σ is unknown these prior views still hold conditional on σ = 1, the value of τ can be found from:

$$\frac{1}{\tau} = (0.06079)^2 \quad \text{or} \quad \tau = 270.6 \tag{55}$$

Thus, equation (54) becomes (β | σ) ~ N (0.85, σ²/270.6). The implications of the prior dependence of β on σ can be seen more clearly by considering another value of σ. For example, if the investigator knew that σ = 2, then he would be of the opinion Prob(0.65 < β < 1.05) = 0.9, compared to Prob(0.75 < β < 0.95) = 0.9 if he knew σ = 1. These required implications of a natural conjugate prior are likely to be very unsettling to many potential Bayesian investigators. The doubts associated with attempts to formulate prior information within these guidelines are likely to be even greater than they were when σ² was known.

Proceeding with the analysis anyway, the joint prior density function is given by:

$$\begin{aligned} g(\beta,\sigma) &= g(\beta|\sigma)g(\sigma) \\ &\propto \left(\frac{\sigma^2}{\tau}\right)^{-\frac{1}{2}} \exp\left\{-\frac{\tau}{2\sigma^2}(\beta - \bar{\beta})^2\right\}\cdot\frac{1}{\sigma} \\ &\propto \frac{1}{\sigma^2} \exp\left\{-\frac{\tau}{2\sigma^2}(\beta - \bar{\beta})^2\right\} \end{aligned} \tag{56}$$

The next steps towards obtaining the marginal posterior density function g (β | y) are to use Bayes' Theorem of equation (47) to multiply equation (56) by equation (52), and then to integrate σ out of the result, as indicated in equation (48). Omitting the proportionality constant, this process yields:

$$g(\beta|\underset{\sim}{y}) \propto \left[\frac{(T-1)\bar{\bar{s}}^2}{\tau + \Sigma x_t^2} + (\beta - \bar{\bar{\beta}})^2\right]^{-T/2} \tag{57}$$

where

$$\bar{\bar{\beta}} = \frac{\bar{\beta}\tau + \hat{\beta}\Sigma x_t^2}{\tau + \Sigma x_t^2} = 0.892 \tag{58}$$

and

$$\begin{aligned} \bar{\bar{s}}^2 &= [\Sigma y_t^2 - (\tau + \Sigma x_t^2)\bar{\bar{\beta}}^2 + \tau\bar{\beta}^2]/(T-1) \\ &= 1.227 \end{aligned} \tag{59}$$

The density function in equation (57) is that of a t-distribution. Unlike the normal distribution, the t-distribution is not a distribution whose density function is dealt with in most courses in economic statistics. Such courses usually concern themselves with percentage points from the t-distribution without worrying about the actual t density function. Thus, equations such as equation (57) may be unfamiliar to many. The first point to note is that the t-distribution is a distribution described by three parameters, namely, the degrees of freedom, the mean (providing the degrees of freedom is greater than one), and the precision. If υ denotes the degrees of freedom parameter, and h the precision, then, providing υ > 2, the variance of a t-distribution is given by υ/h(υ–2). The parameters of the t-distribution in relation (57) are degrees of freedom (T–1), mean β, and precision $(\tau+\Sigma x_t^2)/\bar{\bar{s}}^2$. The t-distribution which arises as the distribution of $(\hat{\beta}-\beta)/\hat{\sigma}_{\hat{\beta}}$ in the sampling theory approach is a t-distribution with degrees of freedom (T-1), mean 0, and precision 1. It is a standardised t-distribution, just like a normal distribution with zero mean and unit variance is known as a standard normal distribution. In the sampling theory approach, where $\hat{\beta}$ and $\hat{\sigma}_{\hat{\beta}}$ are random variables, it is necessary to consider the standardised version of the t-distribution; in the Bayesian approach, where quantities such as $\hat{\beta}$ and $\hat{\sigma}_{\hat{\beta}}$ (or $\bar{\bar{\beta}}$ and $\bar{\bar{\sigma}}^2$) are *a posteriori* fixed, it is convenient for them to be treated as parameters in a more general formulation of the t-distribution.

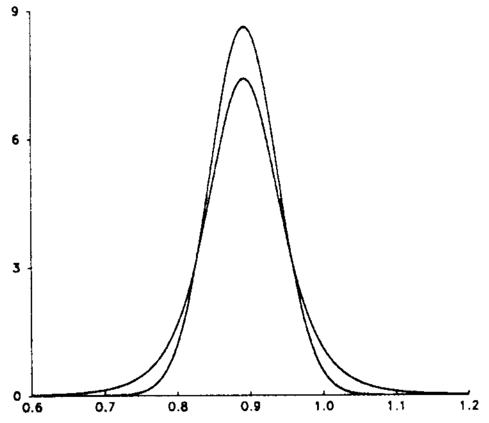
The posterior mean in equation (58) is again a weighted average of the prior mean $\bar{\beta}$ and the sample estimate $\hat{\beta}$, with weights given by the respective precisions. The formulation is slightly different from that in equation (19) because, in the present case, the precisions are given by (τ/σ^2) and $(\Sigma x^2/\sigma^2)$ and σ^2 cancels out. The value of $\bar{\bar{\beta}} = 0.892$ is identical to that obtained earlier and is the Bayesian point estimate under quadratic loss. The parameter value $\bar{\bar{s}}^2 = 1.227$ given in equation (59) can be regarded as an estimate of σ^2, although it is not exactly equal to the mean of the posterior density $g(\sigma^2 \mid y)$, the value which would be the Bayesian point estimate under quadratic loss. The posterior mean for σ^2 is $(T-1)\bar{\bar{s}}^2/(T-3) = 2.045$. Both values are greater than the sampling theory estimate $\hat{\sigma}^2 = 1$, reflecting the additional uncertainty created by the divergence of prior and sample information. If the results were such that $\bar{\bar{\beta}} = \tilde{\beta} = \hat{\beta}$, then the value $\bar{\bar{s}}^2 = 1$ would have been obtained.

When σ^2 is unknown, both the Bayesian and sampling theory approaches use the t-distribution to make inferences about β. However, the sampling theory approach makes inferences in terms of possible values for $\hat{\beta}$ and $\hat{\sigma}_{\hat{\beta}}$ in future hypothetical samples. The Bayesian approach uses the posterior density function to make inferences, and makes probability statements about β based on this posterior density function. The tabulated t-values found in all statistics books can be used to construct (say) 90 per cent or 95 per cent HPD intervals for β. The intervals are created by recognizing that the standardised quantity:

$$(60)\qquad t = \frac{\beta - \bar{\bar{\beta}}}{\bar{\bar{s}}/(\tau + \Sigma x_t^2)^{\frac{1}{2}}}$$

has a t-distribution with degrees of freedom (T–1), mean 0, and precision 1. Most statistics books do not contain sufficient information on distribution functions, or other probabilities associated with the t-distribution, to permit the computation of the probability of β lying within various intervals. Thus, to obtain such probabilities, scarcer more extensive tables have to be found, or numerical integration needs to be employed. Monte Carlo numerical integration could also be used by drawing a random sample from a t-distribution, but such a technique is not required for single-parameter problems.

Figure 5: Normal and t Posteriors from Informative Priors.



The posterior density function in equation (57) is illustrated in Figure 5 alongside its normal density counterpart which was relevant when σ^2 was known. Also, various summary quantities describing the posterior density are given in Table 1. When compared with the results for σ^2 known, the figure and the table values both reflect the fact that the t-distribution has fatter tails, a characteristic which in turn reflects the additional uncertainty created by an unknown σ^2.

3.2.2 Bayesian analysis with a non-informative prior

The conventional noninformative joint posterior density function for (β, σ) is given by:

$$(61)\qquad g(\beta,\sigma) = g(\beta)g(\sigma) = 1 \times \sigma^{-1} = \sigma^{-1}$$

It is assumed that β and σ are *a priori* independent, and that the noninformative marginal priors discussed earlier are relevant. This prior is noninformative in the sense that the posterior density derived from it is dominated by sample information. Its use avoids the difficult and sometimes controversial task of specifying a subjective informative prior.

Multiplying equation (61) by equation (52), as prescribed by the golden rule of Bayes' Theorem in equation (47), and integrating σ out of the result, as indicated in equation (48), yields the marginal posterior density function:

$$(62)\qquad g(\beta|\underset{\sim}{y}) \propto [(T-1)\hat{\sigma}^2_{\hat{\beta}} + (\beta - \hat{\beta})^2]^{-T/2}$$

Table 1: Summary Quantities from Posterior Density Functions

Prior	σ	Posterior	Mean	St.Dev.	95% HPD	$P(\beta \leq 0.9)$	$P(0 < \beta < 1)$
informative	known	normal	0.892	0.0463	(0.80, 0.98)	0.569	0.990
informative	unknown	t	0.892	0.0662	(0.76, 1.02)	0.558	0.954
noninformative	known	normal	0.95	0.0714	(0.81, 1.09)	0.242	0.758
noninformative	unknown	t	0.95	0.0922	(0.77, 1.13)	0.255	0.742
inequality (1)	known	truncated normal	0.921	0.0519	(0.82, 1.00)	0.319	1.000
inequality (1)	unknown	truncated t	0.913	0.0633	(0.79, 1.00)	0.345	1.000
inequality (2)	known	truncated normal	0.958	0.0349	(0.89, 1.00)	0.074	1.000
inequality (2)	unknown	truncated t	0.942	0.0562	(0.83, 1.00)	0.172	1.000

This posterior density function is a t-distribution with degrees of freedom (T–1), mean $\hat{\beta}$, and precision $\hat{\sigma}^{2}_{\hat{\beta}}$. Thus, the Bayesian results with a noninformative prior are very similar to the sampling theory results. The point estimate $\hat{\beta} = 0.95$ is identical to the least-squares estimate; a 95 per cent HPD interval for β is (0.766, 1.134), an interval identical to the sampling theory 95 per cent confidence interval. However, the Bayesian approach brings with it the advantage of being able to make probability statements about β. Some such probability statements, and other summary quantities, are given in Table 1. The complete posterior density, and the corresponding normal one for known σ, are shown in Figure 6. Of particular interest for the next subsection is $\text{Prob}(0 < \beta < 1) = 0.742$. Given their knowledge about the marginal propensity to consume, most researchers would claim $\text{Prob}(0 < \beta < 1) = 1$. A prior which yields a posterior density with this property is considered next.

3.2.3 Bayesian analysis with a prior with inequality restrictions

The analysis for a prior with inequality restrictions and unknown σ^2 parallels that followed when σ^2 was known. However, instead of the analysis yielding a truncated normal posterior distribution for β, a truncated t-distribution is obtained. If c_1 denotes the constant required to make the density in relation (62) integrate to unity, then the truncated t-distribution for this example is:

$$(63) \qquad g(\beta|\underset{\sim}{y}) = \begin{cases} (0.742)^{-1}\, c_1 [(T-1)\hat{\sigma}^2_{\hat{\beta}} + (\beta - \hat{\beta})^2]^{-T/2}, & \text{if } 0 < \beta < 1, \\ 0, & \text{otherwise} \end{cases}$$

The mean and standard deviation of this density, as well as the probability content of particular intervals, and any HPD intervals, all need to be obtained using numerical integration. Values obtained in this way are reported in Table 1, and the truncated t posterior is compared with the corresponding truncated normal posterior in Figure 7. Truncation has the effect of moving the mean to the left, and reducing the standard deviation. It also leads to HPD intervals which do not include values outside the interval (0, 1). The difference between the results of the truncated normal and the truncated t can be explained by the fatter tails of the t-distribution.

The results from a second example of a truncated t posterior, the one based on the sample where $\hat{\beta} = 1.05$, are given in Figure 8 and in the last row of Table 1. As was the case when σ^2 was known, the Bayesian approach to this problem has the advantage of yielding both a point estimate and a HPD interval which are feasible. If there is some doubt concerning whether or not the prior information is correct, then the probability of it being correct can be calculated from the posterior density derived from a completely noninformative prior. In this case this probability is $\text{Prob}(0 < \beta < 1) = 0.258$.

4. A General Regression Model

The previous two sections contain the essential ingredients of both the sampling theory and Bayesian approaches to estimating a regression model. However, the discussion was restricted to a model with just one coefficient, and so it is necessary to consider the implications of extending the analysis to a model where interest centres on a general (K × 1) coefficient vector β in the general regression model:

$$(64) \qquad \underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{e}$$

In equation (64) it is assumed that the usual notational definitions hold, that $\underset{\sim}{e} \sim N(\underset{\sim}{0}, \sigma^2 I)$, and that σ^2 is unknown. This set of assumptions is the same as before, except the problem is now one of finding information about the whole vector $\underset{\sim}{\beta}$.

4.1 Sampling theory approach

The first step in the sampling theory approach is to compute a value for the least squares estimator, $\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'y$. This estimator has a multivariate normal distribution, namely:

$$(65) \qquad \hat{\underset{\sim}{\beta}} \sim N[\underset{\sim}{\beta}, \sigma^2(X'X)^{-1}]$$

Single elements from $\hat{\underset{\sim}{\beta}}$ have univariate normal distributions which are easily derived from relation (65). Confidence intervals or hypothesis tests about single elements from $\underset{\sim}{\beta}$ are based on corresponding univariate t-distributions which arise when the unknown σ^2 is replaced by its sampling theory estimator $\hat{\sigma}^2 = (y'y - \hat{\underset{\sim}{\beta}}\,'\,X'\,y) / (T-K)$. The univariate t-distributions are used in exactly the same way as the t-distribution was used in the previous section. Hypothesis tests on linear functions of the coefficients, such as the equality of two or more coefficients or the sum of a set of coefficients, are based on the F-distribution. Likewise, although they do not seem to be commonly used, confidence regions for two or more parameters can be derived from the F-distribution.

4.2 Bayesian approach

The same basic principles of the Bayesian approach are just as relevant for the general model as they were for the single coefficient model. However, the implementation of these principles can be more difficult. It is straightforward to set up a completely noninformative prior or a prior with inequality restrictions. The former is given by g (β, σ) = σ^{-1}, while the latter is given by g (β, σ) = σ^{-1} over the feasible parameter region, and 0 outside this region. It is more difficult to specify a natural conjugate informative prior for β. The natural conjugate prior for β is a multivariate normal distribution, conditional on σ. Its specification requires values for a prior mean vector and a prior covariance matrix. It is relatively easy to visualise how percentiles of the normal distribution can be used to specify a mean and variance for each of the elements in β, although many investigators may object to describing their prior information in this way. This was the approach taken with the marginal propensity to consume. What is difficult, even for the trained statistician, is to conceptualise prior information in terms of the covariances of the elements in β. Thus, when faced with the need to force their prior information about the complete vector β into a multivariate normal distribution, and to specify a complete prior covariance matrix, it is not surprising that few researchers opt for Bayesian analysis with a natural conjugate prior. There has been considerable research into methods for specifying informative priors; see, for example, Winkler (1980) and several papers in Goel and Zellner (1986). However, the desire to produce results which are not prior specific but relevant for a wide range of researchers is likely to be a hurdle which prevents the wide acceptance of these methods.

Whichever prior is employed, the natural conjugate prior, the non-informative prior, or the prior with inequality restrictions, the next steps are to use Bayes' Theorem to find the joint posterior density function g (β, σ | y), and to find the marginal posterior density function g (β | y) by integrating out σ from g (β, σ | y). For the first two cases, a natural conjugate or a noninformative prior, the resulting marginal posterior density for β is what is known as a multivariate t-distribution. This distribution is described by a degrees of freedom parameter, a mean vector, and a precision matrix; its relationship to the univariate t-distribution is like the relationship between the univariate and multivariate normal distributions. All marginal and conditional distributions derived from the multivariate t-distribution are also t-distributions. When a noninformative prior is employed, the resulting multivariate t posterior density function for β has degrees of freedom (T–K), mean vector $\hat{\beta}$ and precision matrix $\hat{\sigma}^{-2}$ X ' X. The sampling theory

Figure 6: Normal and t Posteriors from Noninformative Priors.

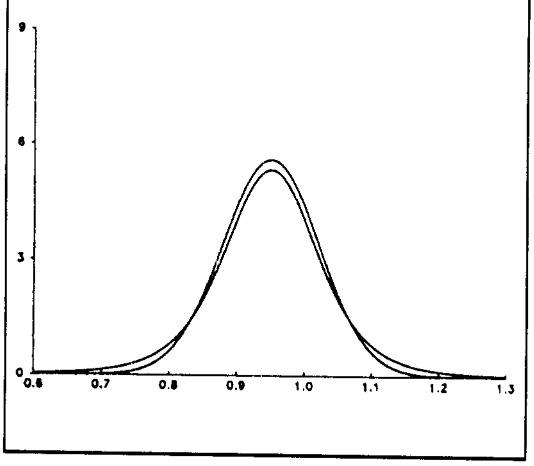
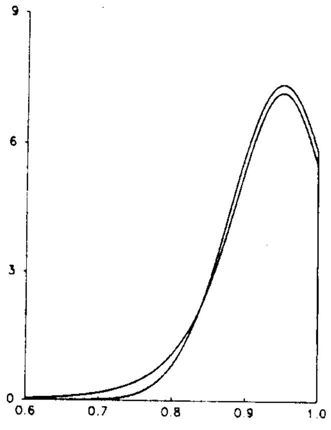


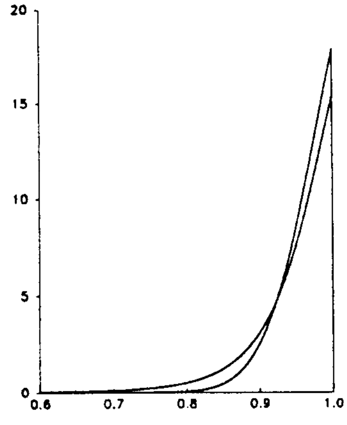
Figure 7: Truncated Normal and Truncated t Posteriors.



estimate $\hat{\beta}$ is identical to the Bayesian quadratic loss estimate, and confidence intervals for single elements in β are identical (numerically) to HPD intervals derived from corresponding univariate t-distributions. For probability statements to be made about one or more of the elements in β, numerical integration is required. Such integration will, of course, be more than unidimensional if the probability statement is for a region which involves more than one of the coefficients in β. If the probability statement involves 4 or more of the coefficients, then Monte Carlo numerical integration would be required to estimate the probability. In this case a large number of "observations" on β would be artificially drawn from the multivariate t posterior density function; the proportion of those values which fall in the required region is an estimate of the probability.

The third prior, that which allows for inequality restrictions, leads to a posterior density function for β which is a truncated multivariate t-distribution. The most common application of this prior is likely to be a regression model where a researcher has *a priori* knowledge about the signs of one or more of the coefficients. When this prior knowledge is included using a prior with inequality restrictions, the resulting multivariate t posterior attaches zero probability to regions of the parameter space which contain "wrong signs". Thus, Bayesian point estimates, such as the posterior mean, will always be of the correct sign. The requirement for numerical integration, or Monte Carlo numerical integration, is much greater for the truncated multivariate t posterior, than it is for the nontruncated version. All means, standard deviations and probabilities need to be evaluated numerically. Also, Monte Carlo numerical integration is required if the dimension of β is 4 or more, even if interest centres only on a quantity (such as the mean) associated with a single coefficient. The truncation of the multivariate t also means that marginal posterior densities for single coefficients can no longer be obtained analytically, but need to be evaluated numerically.

Figure 8: Truncated Normal and Truncated t Posteriors when Least Squares Estimate Greater Than 1.



The problem of obtaining least squares estimates with wrong signs is a common one in applied econometrics. Researchers faced with this problem typically embark on fishing expeditions until they find an alternative equation with the correct signs, or they reluctantly report the results with the incorrect signs. If standard econometric software packages included sampling from a multivariate t-distribution, and the associated computations for Monte Carlo numerical integration, then it is easy to see how Bayesian estimates, derived from the truncated multivariate t-distribution, could become very popular. Some progress in this direction has been made. Version 6 of the popular package SHAZAM (White 1978) has automated the Monte Carlo numerical integration procedure so that it is now straightforward to obtain estimates of posterior means and standard deviations from the truncated multivariate t-distribution. Posterior probability estimates can also be found for any specified region of the parameter space. In the next section, an illustrative example taken from Geweke (1986a) is reworked using SHAZAM. In addition to Geweke, authors who have examined inequality restrictions within a Bayesian framework are O'Hagan (1973) and Davis (1978).

5. An Example

The example taken from Geweke (1986a) is an attempt to explain apartment rentals paid by students at the University of Michigan using sample data provided by Pindyck and Rubinfeld (1981, p. 44). The estimated equation relates rent paid per person y_t to rooms per person r_t, and distance from campus in blocks d_t; a sex dummy s_t, one for male and zero for female, is included to examine whether the influence of number of rooms and distance on rent depends on sex. The equation estimated is:

$$y_t = \beta_1 + \beta_2 s_t r_t + \beta_3(1-s_t)r_t + \beta_4 s_t d_t + \beta_5(1-s_t)d_t + e_t \qquad (66)$$

The expected signs on the coefficients are $\beta_2 \geq 0$, $\beta_3 \geq 0$, $\beta_4 \leq 0$ and $\beta_5 \leq 0$.

Three sets of estimates of the equation are presented in Table 2. The first set, the least squares estimates, can be interpreted as the usual sampling theory results, or they can be viewed as the posterior means of the multivariate t posterior density function which arises when a completely noninformative prior is used. The numbers in parentheses are standard errors or, from the Bayesian viewpoint, the posterior standard deviations of the coefficients. (Strictly speaking, the numbers are not *exactly* the standard deviations because of the constant factor $\upsilon/(\upsilon-2)$ used to derive the variance of a t-distribution. They are, however, the numbers which would be used for the calculation of HPD intervals.) One of the least squares estimates, β_4, has the wrong sign. One possible solution to this problem is to use quadratic programming to compute the constrained least squares estimates. As pointed out by Geweke (1986a), this solution is equivalent to running regressions with all possible subsets of the variables included in equation (66), and choosing from those regressions which have the correct signs, that one with the highest R^2. This is a strategy which is informally followed by many researchers, although they may not consider all possible subsets. The required number of regressions is 2^r where r is the number of inequality restrictions. For equation (66), $2^r = 16$. The quadratic programming set of estimates is given in the second column of Table 2. It suffers from two defects. First, the standard errors, based on results from Liew (1976), are not strictly correct because they are simply those obtained using least squares when the variable corresponding to β_4 is omitted. They are conditional on *a priori* knowledge of which variable(s) was to be omitted. The required distribution theory has not been sufficiently well developed to provide unconditional standard errors. (For some progress in this direction, see Judge and Yancey 1986.) The second defect is that the estimate of zero for β_4 is not a very interesting one. Both defects can be overcome using the Bayesian approach with a prior density with inequality restrictions. The third set of estimates in Table 2 are the posterior means, estimated using SHAZAM's option for Monte Carlo numerical integration from the truncated multivariate t posterior density. The numbers in parentheses are corresponding posterior standard deviations. All were estimated using 100 000 replications. Note that all the estimates have the correct signs. Given that the inequality restrictions are true, these estimates are optimal Bayesian point estimates for quadratic loss, and the standard deviations are an accurate reflection of the precision of the information on each of the coefficients. The values are slightly different from those given by Geweke (1986a), but the differences can be explained by the random number generating process and by the difference in the number of replications.

The estimates given in the third column of Table 2 are those relevant for an investigator, say investigator A, who has a prior with the specified inequality restrictions. Suppose there is a second investigator, investigator B, who has a completely noninformative prior over the whole parameter space. The relevent estimates for investigator B are those given in the first column of Table 2. Also of interest might be B's assessment of whether A's prior information is correct; that is, the probability of getting the "correct signs", given a noninformative prior. This probability is estimated as:

$$\text{Prob}(\beta_2 \geqq 0,\ \beta_3 \geqq 0,\ \beta_4 \leqq 0,\ \beta_5 \leqq 0) = 0.05 \qquad (67)$$

This low probability would undoubtedly lead B to question the validity of A's prior information. It is interesting to note that there is no sampling theory measure of how unlikely prior information may be for the researcher who runs all $2^4 = 16$ possible regressions, and picks that feasible one with the highest R^2.

A possible reason for estimating the rent relationship of equation (66) might be to examine whether sex has any bearing on rent paid. In other words, the differences $\beta_2-\beta_3$ and $\beta_4-\beta_5$ might be of interest. Consider first the sampling theory approach to testing:

$$H_0: \beta_2 = \beta_3 \quad \text{against} \quad H_1: \beta_2 \neq \beta_3 \qquad (68)$$

This test is carried out by computing the value

Table 2: Rent Equation Estimates

	Least Squares	QP	Bayes
β_1	38.56 (32.22)	37.63 (33.27)	37.69 (35.34)
β_2	103.5 (38.37)	130.0 (36.29)	137.3 (39.28)
β_3	122.0 (37.36)	123.0 (38.57)	123.6 (40.52)
β_4	3.315 (1.961)	0.0	-0.9383 (0.8807)
β_5	-1.154 (0.5714)	-1.153 (0.5901)	-1.192 (0.5869)

[a] This table is taken from Geweke (1986a, p.132)

of the appropriate t (or F) statistic and comparing it with a critical value for a given significance level. In the example the t-value is –0.780 which falls well short of the 5 per cent critical values ± 2.052. Thus, the null hypothesis $\beta_2 = \beta_3$ is accepted. Following a similar procedure for testing H_0: $\beta_4 = \beta_5$ against the alternative H_1: $\beta_4 \neq \beta_5$ leads to a t-value of 2.187 and rejection of the null hypothesis.

With the Bayesian approach, the posterior probability for each hypothesis is computed. However, it is not informative to consider the posterior probability that $\beta_2 = \beta_3$. This probability is zero because $\beta_2 - \beta_3 = 0$ is just a single point from the continuous posterior density function for $\beta_2 - \beta_3$. One way to overcome this problem is to use a prior which assigns a positive probability to the point $\beta_2 - \beta_3 = 0$. This procedure of assigning an arbitrary positive prior probability to a single point is analogous to the sampling theory procedure of choosing an arbitrary significance level. An alternative approach, which does not require the setting of what could be an arbitrary positive prior probability, is to compare $\text{Prob}(\beta_2 < \beta_3)$ with $\text{Prob}(\beta_2 > \beta_3)$, where these are posterior probabilities resulting from either a noninformative prior or a prior with inequality restrictions. If the posterior odds ratio:

$$K = \frac{\text{Prob}(\beta_2 < \beta_3)}{\text{Prob}(\beta_2 > \beta_3)} \tag{69}$$

is close to unity, then there would be little evidence to suggest that sex makes a difference. If it is much greater than one, or much less than one, then there would be some evidence to suggest sex has some bearing on rent paid. Geweke (1986a) used Monte Carlo numerical integration to compute the required posterior probabilities for both the noninformative prior and the prior with inequality restrictions. Similar computations were carried out using SHAZAM; the estimates provided by SHAZAM, again from 100 000 replications, are given in Table 3.

When the noninformative prior is employed, the posterior odds ratio in favour of $(\beta_2 > \beta_3)$ is 0.287 or, using the inverse $[(0.287)^{-1} = 3.48]$, it is 3.48 times more likely that $\beta_3 > \beta_2$ than it is for $\beta_2 > \beta_3$. Note that the sampling theory procedure accepts the hypothesis that $\beta_2 = \beta_3$ despite this fact. When β_4 and β_5 are considered the sampling theory procedure rejects H_0: $\beta_4 = \beta_5$, and the odds ratio in favour of $\beta_4 > \beta_5$ is 51.63. The posterior probabilities and odds ratios are vastly different when the prior with inequality restrictions is employed. This result is not surprising given the values of the posterior means under each of the priors and given that the probability of the inequality restrictions holding is only 0.050.

The SHAZAM instructions used to obtain the entries in Tables 2 and 3 are given in an Appendix.

6. Conclusions

The purpose of this paper has been to introduce Bayesian econometric methodology in general, and Bayesian estimation of the regression model with inequality restrictions in particular. Building heavily on the work of Geweke (1986a), an attempt has been made to show that Bayesian estimation of the inequality constrained regression model is practically feasible. Two difficulties commonly associated with Bayesian analysis, the need to formulate prior information in terms of a prior density function, and the computational problems of numerical integration, need not be limiting factors. It has been argued that the Bayesian way of reporting results is both more pragmatic and more informative.

The vehicle used to introduce the Bayesian approach was the classical normal linear regression model. More complicated models such as models which involve heteroscedasticity or autocorrelation, distributed lag models, simultaneous equation models, *etc.* have not been considered. The principles involved in the analysis of such models are identical to those introduced in this paper, although the degree of difficulty can be greater. In particular, the choice of distribution from which to sample when using Monte Carlo numerical integration can be difficult. However, recent theoretical results derived by Geweke (1986b) suggest that, even for more complicated models, automated Bayesian computing is getting closer.

References

ANDERSON, J. R., DILLON, J. L. and HARDAKER, J. B. (1977), *Agricultural Decision Analysis*, Iowa State University Press, Ames, Iowa.

BAILS, D. G. and PEPPERS, L. C. (1982), *Business Fluctuations*, Prentice-Hall, Englewood Cliffs.

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

BOX, G. E. P. and TIAO, G. C. (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, Mass.

DAVIS, W. W. (1978), "Bayesian analysis of the linear model subject to linear inequality constraints", *Journal of the American Statistical Association*, 73 (363), 573–9.

GEWEKE, J. (1986a), "Exact inference in the inequality constrained normal linear regression model", *Journal of Applied Econometrics* 1(2), 127–41.

GEWEKE, J. (1986b), "Bayesian inference in econometric models using Monte Carlo numerical integration", working paper, Duke University.

GOEL, P. K. and ZELLNER, A. (eds) (1986), *Bayesian Inference and Decision Techniques*, North-Holland, Amsterdam.

JOHNSON, N. L. and KOTZ, S. (1970), *Continuous Univariate Distributions—1*, Wiley, New York.

JUDGE, G. G., GRIFFITHS, W. E., HILL, R. C., LUTKEPOHL, H. and LEE, T. C. (1985), *The Theory and Practice of Econometrics*, 2nd edition, Wiley, New York.

JUDGE, G. G. and TAKAYAMA, T. (1966), "Inequality restrictions in regression analysis", *Journal of the American Statistical Association* 61 (313), 166–81.

JUDGE, G. G. and YANCEY, T. A. (1968), *Improved Methods of Inference in Econometrics*, North-Holland, Amsterdam.

LIEW, C. K. (1976), "Inequality constrained least-squares estimation", *Journal of the American Statistical Association* 71(355), 746–51.

O'HAGAN, A. (1973), "Bayes estimation of a convex quadratic", *Biometrics* 60(3), 565–71.

PINDYCK, R. S. and RUBINFELD, D. L. (1981), *Econometric Models and Economic Forecasts*, 2nd edition, McGraw-Hill, New York.

RAIFFA, H. and SCHLAIFER, R. (1961), *Applied Statistical Decision Theory*, Harvard University Press, Boston.

WHITE, K. J. (1978), "A general computer program for econometric methods—SHAZAM", *Econometrica* 46(1), 239–40.

WINKLER, R. L. (1980), "Prior information, predicitive distributions, and Bayesian model-building", in A. Zellner (ed), *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, North-Holland, Amsterdam.

ZELLNER, A. (1971), *An Introduction to Bayesian Analysis in Econometrics*, Wiley, New York.

ZELLNER, A. (1983), "Applications of Bayesian analysis in econometrics", presented at the Institute of Statisticians International Conference on Practical Bayesian Statistics, St Johns College, University of Cambridge, July 21–24, 1982, and published in *The Statistician*, 32, 23–34.

ZELLNER, A. (1984), *Basic Issues in Econometrics*, University of Chicago Press, Chicago.

ZELLNER, A. (1985), "Bayesian econometrics", *Econometrica* 53(2), 253–70.

Table 3: Posterior Probabilities and Posterior Odds Ratios

	Noninformative Prior	Inequality Prior
$P(\beta_2 > \beta_3)$	0.223	0.723
Posterior Odds in favour of $\beta_2 > \beta_3$	0.287	2.610
$P(\beta_4 > \beta_5)$	0.981	0.679
Posterior Odds in favour of $\beta_4 > \beta_5$	51.632	2.115

Appendix
SHAZAM Instructions for Example Problem

```
file 6 aework.out
file 4 aework.dta
smpl 1 32
read(4) rent no room s d
genr y=rent/no
genr r=room/no
genr sr=s*r
genr osr=(1-s)*r
genr sd=s*d
genr osd=(1-s)*d
genr one=1
ols y one sr osr sd osd/noconstant
```

```
bayes/nsamp=50000 psigma
rest sr.gt.0
rest osr.gt.0
rest sd.lt.0
rest osd.lt.0
end
```

commands for Bayesian restricted estimates

```
bayes/nsamp=50000
rest sr.gt.osr
end
```

yields Prob($\beta_2 > \beta_3$)

```
bayes/nsamp=50000
rest sd.gt.osd
end
```

yields Prob($\beta_4 > \beta_5$)

```
bayes/nsamp=50000
rest sr.gt.0
rest osr.gt.0
rest sd.lt.0
rest osd.lt.0
rest sr.gt.osr
end
```

yields Prob($\beta_2 > \beta_3$ and β_i's have correct signs

```
bayes/nsamp=50000
rest sr.gt.0
rest osr.gt.0
rest sd.lt.0
rest osd.lt.0
rest sd.gt.osd
end
```

yields Prob($\beta_4 > \beta_5$ and β_i's have correct signs

```
stop
```