Federal Reserve Bank of Minneapolis
Research Department

Bayesian Inference for Linear Models Subject to Linear Inequality Constraints

John F. Geweke*

Working Paper 552

July 1995

ABSTRACT

The normal linear model, with sign or other linear inequality constraints on its coefficients, arises very commonly in many scientific applications. Given inequality constraints Bayesian inference is much simpler than classical inference, but standard Bayesian computational methods become impractical when the posterior probability of the inequality constraints (under a diffuse prior) is small. This paper shows how the Gibbs sampling algorithm can provide an alternative, attractive approach to inference subject to linear inequality constraints in this situation, and how the GHK probability simulator may be used to assess the posterior probability of the constraints.

*Federal Reserve Bank of Minneapolis and University of Minnesota. The editor and a referee, who bear no responsibility for any errors or omissions, have provided suggestions that improved the paper. Forthcoming in *Forecasting, Prediction and Modeling in Statistics and Econometrics: Bayesian and non-Bayesian Approaches,* ed. W. O Johnson, J. C. Lee, and A. Zellner. New York: Springer-Verlag. This work was supported in part by National Science Foundation Grant SES-9210070. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Introduction

The normal linear regression model subject to linear inequality constraints for the coefficients arises commonly in applied econometrics as well as other scientific applications. Typically the motivating economic model restricts the signs of certain coefficients or of known linear combinations of coefficients. A well-known pedagogical example is provided by Pindyck and Rubinfeld (1981, p. 44) who take up the demand for student housing near the University of Michigan. Rent paid per person is a linear function of the number of rooms per person, with a positive coefficient, and the distance from campus, with a negative coefficient. (We return to this example below.)

The subsequent empirical work in this and other linear regression models subject to linear inequality constraints for the coefficients then focuses on two related but distinct questions. First, how plausible are the linear inequality restrictions delivered by the economic model? Second, conditional on these restrictions what is to be inferred about the coefficients of the regression model? These turn out to be nontrivial questions, and historically investigators have taken different, usually informal, approaches to these tasks. Difficulties for classical inference are discussed in Judge and Takayama (1966) and Lovell and Prescott (1970); for classical testing in Gourieroux, Holly, and Monfort (1982) and Wolak (1987).

This work takes up a Bayesian approach to the problem of linear regression with linear inequality constraints on the coefficients. Extending earlier analytical work by Davis (1978), Chamberlain and Leamer (1976), and Leamer and Chamberlain (1976), it uses fast numerical methods for the determination of posterior moments and probabilities, advancing the methods reported in Geweke (1986). But whereas Geweke (1986) takes up any inequality constraints on the coefficients, this paper limits attention to inequality constraints that are linear. The more specialized algorithms provide faster, more accurate numerical approximations to posterior moments than do the more general ones. In particular, when the posterior probability of linear inequality constraints is low or the number of coefficients is large the methods in Geweke (1986) may be slow to the point of impracticality. In contrast computation time in the approach taken here does not increase systematically with the inverse of the posterior probability of the inequality constraints, and increases only linearly with the number of coefficients.

In standard notation the normal linear regression model is

$$y_{T\times 1} = X_{T\times k}\beta_{k\times 1} + \varepsilon_{T\times 1}, \quad \varepsilon \sim N(0, \sigma^2 I_T) \tag{1}$$

where y is the vector of dependent variables, X is the matrix of explanatory variables (regressors), and ε is the vector of disturbances. There are T observations and k explanatory variables. In the unconstrained case a standard diffuse reference prior for the parameter vector (β', σ) is

$$f(\beta, \sigma) \propto \sigma^{-1}. \tag{2}$$

The inequality constraints are expressed

$$a_{k\times 1} \leq D_{k\times k}\beta \leq w_{k\times 1}. \tag{3}$$

In this expression the inequalities are to read line by line: $a_i \leq \sum_{j=1}^{k} d_{ij}\beta_j \leq w_i (i = 1, ..., k)$. The matrix D is composed of real numbers and is nonsingular. The vectors a and w are composed of extended real numbers, with $-\infty$ and $+\infty$ explicitly permitted, thus allowing single-sided inequality constraints. Since a constraint in (3) has no effect if $a_i = -\infty$ and $w_i = +\infty$, fewer than k linear inequality restrictions—perhaps only one—may be involved. Inequality constraints on more than k linear combinations are precluded by (3), a point to which the concluding section of the paper returns briefly. Extending the standard reference prior of the unconstrained model the prior distribution employed in this model is

$$\begin{aligned} f(\beta,\sigma) &\propto \sigma^{-1} \text{ if } \beta \in Q, \ f(\beta,\sigma) = 0 \text{ if } \beta \notin Q; \\ Q &\equiv \{\beta : a \leq D\beta \leq w\}. \end{aligned} \tag{4}$$

The next section takes up evaluation of the posterior probability of the hypothesis (4) relative to (2). An algorithm for the evaluation of this probability based on the Geweke-Hajivassiliou-Keane (GHK) probability simulator is constructed. (The simulator was named by Hajivassiliou, McFadden, and Ruud (1995) in a paper that first appeared in 1992. See also Keane (1990, 1993, 1994) and Geweke, Keane and Runkle (1995).) Section 3 turns to the problem of inference for β—construction of posterior means, evaluation of the posterior probabilities of regions, etc.—in the constrained model (4). A Gibbs sampling algorithm for drawing β and σ from the posterior distribution that builds on work in Geweke (1991, 1995) is presented. The algorithm described in Section 3 is superficially similar to that in Geweke (1995), but that paper treats mixed equality and inequality constraints on individual coefficients, whereas the approach here takes up inequality constraints for arbitrary sets of linear combinations of coefficients. Section 4 employs both methods in the two empirical examples used in Geweke (1986). A concluding section suggests extensions of this work.

2 Evaluating the Hypothesis of Linear Inequality Constraints

One may take two approaches to the evaluation of the inequality constraints (3) as a formal hypothesis. The first is to regard (1) and (2) as the maintained hypothesis, and evaluate the posterior probability that (3) is true. The advantage of this method is conceptual simplicity: no problems arise from the fact that (3) is not a proper prior distribution. In particular suppose that (2) is regarded as the limit of a sequence of proper prior distributions, e.g. $\beta \sim N(\beta, \alpha V)$, where β is a fixed vector, V is a fixed matrix, and $\alpha \to \infty$. Then the limiting posterior probability of the region defined by (3) is correctly given as the posterior probability of (3) when the posterior density kernel is formed as the product of (2) and the likelihood function

$$L(\beta,\sigma) = \sigma^{-T} \exp\{-[(T-k)s^2 + (\beta - b')'(X'X)(\beta - b)]/2\sigma^2\} \tag{5}$$

corresponding to (1). (This expression employs standard notation for the least squares estimator $b \equiv (X'X)^{-1}X'y$; $s^2 \equiv (y - Xb)'(y - Xb)/(T-k)$. For details on the derivation of (5) see Zellner (1971, pp. 65-66).) Denote this probability $p_{2|1}$. The disadvantage of this approach is that (1) through (4) is frequently a hypothesis competing with (1) and (2) rather than a region of interest of the parameter space. For example, if the posterior distribution of β is centered at $\beta = 0$ in the unconstrained model, $p_{2|1}$ for $\beta \geq 0$ will generally become smaller the larger is k; yet in this case (1) through (4) is clearly competitive with (1) and (2).

The second approach is to regard (1) and (2) and (1) through (4) as competing hypotheses for which a posterior odds ratio is to be formed. Without loss of generality, assume that the prior odds ratio is 1:1. If all elements of a and w are finite then the prior distribution (4) is proper for β and the limit of a posterior odds ratio in favor of (1) and (2) over (1) through (4) with a sequence of increasingly diffuse priors is 0. (This phenomenon is sometimes called the Lindley paradox and is well treated in the literature, e.g. Lindley (1957), Press (1989, pp. 36–37) and Bernardo and Smith (1994, p. 394).) If some elements of a are $-\infty$ and/or some elements of w are $+\infty$, then in general the limiting ratio formed from a sequence of increasingly diffuse priors depends on the particular sequence chosen. Certain specific cases are of some general interest, however. In particular if the linear inequality constraints may be brought into the form $\delta_i\beta_j \geq a_i (i = 1, ..., r)$ with $\delta_i = \pm 1$, and if the sequence of prior distributions is $\beta \sim N(a, \alpha I_k)$ for the unconstrained case and $\beta \sim N(a, \alpha I_k)$ constrained to $\delta_i\beta_j \geq a_i(i = 1, \ldots, r)$ for the constrained case, then the limiting posterior odds ratio (as $\alpha \to \infty$) is $p_{2:1} \equiv 2^r p_{2|1}$.

With these considerations in mind, focus on computation of the event probability $p_{2|1}$. Geweke (1986) proposed a *crude frequency simulator* as follows. In the context of (2) and (5), the marginal posterior probability distribution of σ is provided by

$$(y - Xb)'(y - Xb)/\sigma^2 \sim \chi^2(T-k). \tag{6}$$

Conditional on σ,

$$\beta \sim N[b, \sigma^2(X'X)^{-1}].$$

Hence random draws $\{\sigma^{(i)}, B^{(i)}\}_{i=1}^{m}$ may be taken easily. Let $d^{(i)} = 1$ if $\beta^{(i)} \in Q$, $d^{(i)} = 0$ if $B^{(i)} \notin Q$. Then $\breve{p}_{2|1} \equiv \sum_{i=1}^{m} d^{(i)}/m \xrightarrow{a.s} p_{2|1}$ as $m \to \infty$. Moreover, for large m the standard error of approximation of $p_{2|1}$ by $\breve{p}_{2|1}$ is given by $[p_{2|1}(1 - p_{2|1})/m]^{1/2}$. The advantage of this method is its simplicity. Its disadvantage lies in the need to make many draws, m, if $p_{2|1}$ is small; and if k is large, $p_{2:1}$may be large enough that the constraint hypothesis is competitive with the unconstrained model even though $p_{2|1}$ is small.

A more efficient method for approximating $p_{2|1}$ is based on the *GHK probability simulator,* an algorithm proposed independently by Hajivassiliou and McFadden (1990) and Keane (1990); more accessible references are Keane (1993), Keane (1994), and Geweke, Keane and Runkle (1995). Let $z \equiv D(\beta - b)$, $a^* \equiv D(a - b)$, and $w^* \equiv D(w-b)$. Since $\beta|\sigma \sim N[b, \sigma^2(X'X)^{-1}]$, $z|\sigma \sim N[0, \sigma^2 D(X'X)^{-1}D']$and $P[a \leq D\beta \leq w|\sigma] = P[a^* \leq z \leq w^*|\sigma]$. Let FF' denote the Choleski decomposition of $D(X'X)^{-1}$ D' : F is the unique lower triangular matrix with positive diagonal elements such that $FF' = D(X'X)^{-1}D'$. A conventional construction for $z \mid \sigma$ is then given by

$$z|\sigma = \sigma F\varepsilon, \quad \varepsilon \sim N(0, I_k). \tag{7}$$

(Most software for generation of synthetic random multivariate normal vectors is based on this construction.) Denote a typical row of (7) $z_i = \sigma \sum_{j=1}^{i} f_{ij}\varepsilon_i$. The probability $P[a^* \leq z \leq w^*|\sigma]$ may be decomposed

$$P[a_1^* \leq z_1 \leq w_1^*|\sigma] \cdot P[a_2^* \leq z_2 \leq w_2^*|\sigma, a_1^* \leq z_1 \leq w_1^*]\cdot$$

$$\ldots \cdot P[a_j^* \leq z_j \leq w_j^*|\sigma, a_i^* \leq z_i \leq w_i^*(i < j)]\cdot$$

$$\ldots \cdot P[a_k^* \leq z_k \leq w_k^*|\sigma, a_i^* \leq z_i \leq w_i^*(i < k)].$$

The GHK probability simulator provides independent, unbiased simulations of each conditional probability in this product. It does so by drawing $\tilde{z}_1, \ldots, \tilde{z}_{j-1}$, from the $N(0, \sigma^2 D(X'X)^{-1}D']$ distribution subject to the constraints $a_i^* \leq z_i \leq w_i^*(i = 1, \ldots, j-1)$ and then computing $P[a_j^* \leq z_j \leq c_j^*|\tilde{z}_1, \ldots \tilde{z}_{j-1}]$. The draws are accomplished by generating

$$\begin{aligned}
\tilde{\varepsilon}_1 &\sim N(0,1) s.t. \quad a_1^*/\sigma f_{11} \leq \tilde{\varepsilon}_1 \leq w_1^*/\sigma f_{11} \\
\tilde{\varepsilon}_2 &\sim N(0,1) s.t. \quad (a_2^* - \sigma f_{21}\tilde{\varepsilon}_1)/\sigma f_{22} \leq \tilde{\varepsilon}_2 \leq (w_2^* - \sigma f_{21}\tilde{\varepsilon}_1)/\sigma f_{22} \\
&\vdots \\
\tilde{\varepsilon}_j &\sim N(0,1) s.t. \quad \left(a_j^* - \sum_{i=1}^{j-1} \sigma f_{ji}\tilde{\varepsilon}_i\right)/\sigma f_{jj} \leq \tilde{\varepsilon}_j \\
&\leq \left(w_j^* - \sum_{i=1}^{j-1} \sigma f_{ji}\tilde{\varepsilon}_i\right)/\sigma f_{jj} \qquad (8) \\
&\vdots \\
\tilde{\varepsilon}_k &\sim N(0,1) s.t. \quad \left(a_k^* - \sum_{i=1}^{k-1} \sigma f_{ki}\tilde{\varepsilon}_i\right)/\sigma f_{kk} \leq \tilde{\varepsilon}_k \\
&\leq \left(w_k^* - \sum_{i=1}^{k-1} \sigma f_{ki}\tilde{\varepsilon}_i\right)/\sigma f_{kk}
\end{aligned}$$

Let $\tilde{p}_j = P[(a_j^* - \sum_{i=1}^{j-1} \sigma f_{ji}\tilde{\varepsilon}_i)/\sigma f_{jj} \leq \tilde{\varepsilon}_j \leq (w_j^* - \sum_{i=1}^{j-1} \sigma f_{ji}\tilde{\varepsilon}_i)/\sigma f_{jj} | \tilde{\varepsilon}_1, ..., \tilde{\varepsilon}_{j-1}]$. Then $\tilde{p}_1 = P[a_1^* \leq z_1 \leq w_1^* | \sigma]$ and $E(\tilde{p}_j) = P[a_j^* \leq z_j \leq w_j^* | \sigma] (j > 1)$. The probabilities $\tilde{p}_j$ may be computed by direct evaluation of the univariate standard normal c.d.f. Because the $\tilde{p}_j$ are mutually independent,

$$E\left[\prod_{i=1}^{k} \tilde{p}_i\right] = P(a^* \leq z \leq w^* | \sigma).$$

(Note that in fact it is not necessary to take the last draw in (8).)

The GHK probability simulator for $p_{2|1}$ is therefore an iterative process with two steps in each iteration. In the first step, $(y - Xb)'(y - Xb)/(\sigma^{(i)})^2 \sim \chi^2(T - k)$. In the second step, one or more values of $\prod_{j=1}^{k} \tilde{p}_j$ are drawn as described in the previous paragraph, with $\sigma = \sigma^{(i)}$; let the average of these values be denoted $p^{*(i)}$. After m iterations of the two steps, $p_{2|1}$ is approximated by $\hat{p}_{2|1} \equiv \sum_{i=1}^{m} p^{*(i)}/m$. Since $E[p^{*(i)}|\sigma] = E[d^{(i)}|\sigma]$ but $0 < p^{*(i)} < 1$ whereas $d(i) = 0$ or $d^{(i)} = 1$, $var(\hat{p}_{2|1}) < var(\check{p}_{2|1})$. Hence the GHK probability simulator always provides a more accurate approximation to $p_{2|1}$ than does the crude frequency simulator, given the same number of iterations.

3 Inference Subject to Linear Constraints

A related but distinct task is to find posterior moments and probabilities corresponding to the restricted model (1) through (4). The posterior density for this model is

$$\begin{aligned}
f(\beta, \sigma | y, X) &\propto \sigma^{-(T+1)} \exp\{-[(T-k)s^2 + (\beta - b)'(X'X)(B - b)]/2\sigma^2\} \\
\text{if } \beta &\in Q; \\
f(\beta, \sigma | y, X) &= 0 \quad \text{if } \beta \notin Q; \\
Q &\equiv \{\beta : a \leq D\beta \leq w\}.
\end{aligned} \tag{9}$$

The crude frequency simulator described in the previous section may also be used to produce draws from the posterior distribution whose kernel is given by (9). One draws from the unconstrained posterior distribution, and accepts the draw if and only if $a \leq D\beta \leq w$. This algorithm again has the advantage of simplicity and the disadvantage of inefficiency. In this section we summarize a Gibbs sampling algorithm for generating a sequence $\{\beta^{(i)}, \sigma^{(i)}\}$ that converges in distribution to the posterior distribution whose probability density kernel is given by (9). Supporting technical details may be found in Geweke (1991).

As in the previous section define $z \equiv D(\beta - b)$, $a^* \equiv D(a - b)$, $w^* \equiv D(w - b)$. The posterior distribution of (β, σ) is then a simple transformation of the posterior distribution of (z, σ). The posterior distribution of σ conditional on z (equivalently, on $\beta = b + D^{-1}z$) is given by

$$(y - X\beta)'(y - X\beta)/\sigma^2 | (\beta, y, X) \sim \chi^2(T), \tag{10}$$

which may be derived easily from (9) (Geweke 1992). The posterior distribution of z conditional on σ is

$$z|(\sigma, y, X) \sim N[0, \sigma^2 D(X'X)^{-1}D')] \; s.t. \; a^* \leq z \leq w^*;$$

let $R \equiv \sigma^2 D(X'X)^{-1}D'$. The distribution of

$$z_j|(\sigma, z_i(i \neq j), y, X)$$

is univariate normal, truncated below by a_j^* and above by w_j^*. Following Geweke (1991) this normal distribution has mean $b_j + \sum_{i \neq j} c_{ji} z_i$ and variance h_j^2. The vector

$$c^j \equiv (c_{j1}, ..., c_{j,j-1}, c_{j,j+1}, ..., c_{jk})'$$

is given by $c^j = -(R^{jj})^{-1} T^{j,<j}$, where R^{jj} is the element in row j and column j of R^{-1}, and $T^{j,<j}$ is row j of R^{-1} with R^{jj} deleted. The variance is $h_j^2 = (R^{jj})^{-1}$. (These expressions follow from the conventional theory for the multivariate normal distribution (Rao, 1965, p. 441) and expressions for the inverse of a partitioned symmetric matrix (Rao, 1965, p. 29). Consequently,

$$\begin{aligned} z_j|(\sigma, z_i(i \;\; \neq \;\; j), y, X) &= \sum_{i \neq j} c_{ij} z_j + h_j \varepsilon_j, \\ \varepsilon_j \;\; \sim \;\; N(0,1) \quad & s.t. (a_j^* - \sum_{i \neq j} c_{ij} z_j / h_j \leq \varepsilon_j \leq w_j^* - \sum_{i \neq j} c_{ij} z_j / h_j. \end{aligned} \tag{11}$$

Expressions (10) and (11) provide an algorithm for producing draws for any element of (z, σ) conditional on all the other elements. Inference for the full posterior distribution whose density is given by (9) may therefore be accomplished using the Gibbs sampler described by Gelfand and Smith (1990) and Tierney (1994). Beginning with any values of σ and $z_1, ..., z_k$ in the support of (9), successively draw and replace $\sigma, z_1, ..., z_k$ using (10) and (11). Call this new draw $(z^{(1)}, \sigma^{(1)})$ and construct $\beta^{(1)} = b + D^{-1} z^{(1)}$. Then repeat the process, obtaining $\beta^{(2)}$ and $\sigma^{(2)}$, and continue on in this way. Since there is a positive probability of moving from any given $(\beta^{(j)}, \sigma^{(j)})$ to any region of the parameter space with positive posterior probability in one iteration, the sequence $\{\beta^{(j)}, \sigma^{(j)}\}$ converges in distribution to the posterior distribution whose kernel density is given by (9). If the posterior expectation $E[g(\beta, \sigma)]$ of $g(\beta, \sigma)$exists, then $g_m - m^{-1} \sum_{j=1}^{m} g(\beta^{(j)}, \sigma^{(j)}) \xrightarrow{a.s} E[g(\beta, \sigma)] \equiv g$. If the posterior variance $\text{var}[g(\beta, \sigma)]$ also exists, then the accuracy of the approximation of g by g_m may be assessed by computing the numerical standard error of g_m, NSE(g_m), as described in Geweke (1992). This variance provides an asymptotic (in m) approximation to the sampling standard deviation of the numerical approximation based on m iterations. Observe that the posterior mean and variance exist for $g(\beta, \sigma) = \beta_j$, $g(\beta, \sigma) = \sigma$, and $g(\beta, \sigma) = \sigma^2$.

4 Two Examples

Two illustrations provide an indication of the absolute and relative performance of the crude frequency simulator, GHK probability simulator, and Gibbs sampling algorithm. These examples were also used in Geweke (1986), which provides more elaborate discussion of substantive aspects.

Pindyck and Rubinfeld (1981, p. 44) provide 32 observations on rent paid, number of rooms rented, number of occupants, sex, and distance from campus in blocks for undergraduates at the University of Michigan. These data are used by the authors in developing the linear regression model at several points in their text. Denote rent paid per person by y_i, rooms per person by r_i, and distance from campus in blocks by d_i, and let s_i be a sex dummy, one for male and zero for female. The equation estimated is

$$y_i = \beta_1 + \beta_2 s_i r_i + \beta_3(1 - s_i)r_i + \beta_4 s_i d_i + \beta_5(1 - s_i)d_i + \varepsilon_i.$$

The inequality constraints are $\beta_2 \geq 0,\ \beta_3 \geq 0,\ \beta_4 \leq 0,\ \beta_5 \leq 0.$

Table I

Constraint Probabilities $p_{2|1}$, rent data set[a]

	Probability	Numerical standard error	Time[b]
Crude frequency simulator (1 evaluation per iteration)	.05070	.00200	8.37
GHK probability simulator (1 evaluation per iteration)	.04810	.00126	22.83
GHK probability simulator (25 evaluations per iteration)	.04921	.00020	369.88

[a]All results are based on $m = 10,000$ iterations.
[b]In seconds. All computations were carried out on a Sun Sparc IIPC 4/40, using compiled Fortran code with extensive calls to the IMSL mathematical and statistical libraries.

Alternative numerical approximations to the constraint probabilities $p_{2|1}$ are given in Table I. All approximations are based on m = 10,000 iterations for the algorithms described in Section 2. In each iteration one or more evaluations of $P(\beta \in Q|\sigma)$ may be made. In Table I results are provided for one evaluation in the crude frequency simulator, one evaluation in the GHK probability simulator, and 25 evaluations in the GHK probability simulator. Note that the numerical standard error of the GHK probability simulator with one evaluation is less than that of the crude frequency simulator (by over one-third), and the GHK probability simulator with 25 evaluations produces a probability approximation whose numerical standard error is about one-sixth of that with one evaluation. In terms of required computation time, however, the GHK probability simulator is preferred: the crude frequency simulator requires about $(8.37/369.88) \times (0.00200/0.00020)^2 = 2.26$ times as much computation time to achieve a numerical approximation of the same accuracy, while the GHK probability simulator with one evaluation per iteration requires about $(22.83/369.88) \times (0.00126/0.00020)^2 = 2.45$ times as much computation time. Finally, the substantive results show that the constrained model is competitive with the unconstrained model, using the set of reference priors described in Section 2: with a prior odds ratio of 1:1 the posterior odds ratio in favor of the constraints is $2^4 \times 0.049{:}1 = 0.78{:}1$.

Table II

Posterior moments, rent data set[a]

	Crude frequency simulator			Gibbs sampler (No skips)			Gibbs sampler (Every 10th)		
Parameter	Mean	s.d.	NSE	Mean	s.d.	NSE	Mean	s.d.	NSE
β_1	37.0	35.6	.4	36.5	34.6	2.7	37.0	34.7	.7
β_2	137.6	39.3	.4	138.3	38.4	2.8	137.6	38.5	.8
β_3	124.6	40.9	.4	125.0	40.1	3.1	124.7	40.1	.7
β_4	−.917	.857	.009	−.927	.843	.011	−.922	.858	.011
β_5	−1.194	.581	.006	−1.192	.580	.010	−1.211	.578	.004
σ	40.87	5.80	.058	40.75	5.79	.098	40.79	5.84	.059
σ^2	1703.9	499.4	5.0	1693.9	497.8	8.5	1698.0	507.4	5.1
Time[b]		131.1			12.63			118.8	

[a]All results are based on m = 10,000 iterations.
[b]In seconds. All computations were carried out on a Sun Sparc I IPC 4/40, using complied Fortran code with extensive calls to the IMSL mathematical and statistical libraries.

Posterior means and standard deviations for the coefficients β_j, σ, and σ^2 are provided in Table II. Three methods were used to approximate these moments. In the first, the crude frequency simulator was run until $m = 10{,}000$ draws of (β, σ) had been made for which the value of β satisfied the constraints. This produces i.i.d. drawings from the posterior distribution, and consequently the numerical standard error is $(10{,}000)^{-1/2} = 0.01$ times the posterior standard deviation. In the second method, the Gibbs sampler was used to make 10,000 successive draws as described in Section 3. In general the Gibbs sampler produces a serially correlated sequence, and that is evident here in the numerical standard errors that are substantially higher than are obtained from a sequence of i.i.d. drawings from the posterior distribution, for many coefficients. This problem can be alleviated by increasing the number of iterations, or by recording only every nth draw: as $n \to \infty$ the Gibbs sampled parameters become serially uncorrelated and the numerical standard error approaches that of an i.i.d. sequence. The last panel of Table II shows the results corresponding to this procedure with $n = 10$. At a cost of a 10-fold increase in computation time numerical standard errors are reduced to values closer to those of the crude frequency simulator than to those of the Gibbs sampler with no skips. Execution time for this modified Gibbs sampler is still less than that for the crude frequency simulator. (Note that the crude frequency simulator stochastically records about every 20th draw, whereas the Gibbs sampler for the third panel of Table II records every 10th draw.)

The second example is taken from Bails and Peppers (1982). In Appendix G they provide 60 quarterly observations on unit sales of automobiles in the U.S., and 10 explanatory variables. They consider the normal linear regression model $y_t = \sum_{j=1}^{11} \beta_j x_{tj} + \varepsilon_t$, with y_t denoting unit sales of automobiles at time t; x_{1t}, an intercept term; x_{2t}, personal income less transfer payments; x_{3t}, index of consumer sentiment; x_{4t}, unemployment rate; x_{5t}, index of cost of car ownership; x_{6t}, average miles per gallon of current model-year cars; x_{7t}, dummy variable for automobile strikes; x_{8t}, depreciation rate of the stock of cars; x_{9t}, average price of a new car; x_{10t}, stock of automobiles; and x_{11t}, interest rate on automobile loans. (A discussion of these variables is provided by Bails and Peppers on pp. 246–247.) We use the data exactly as presented, except that x_{7t} is scaled by 10^3. The coefficients β_2, β_3, β_6, and β_8 are anticipated to be nonnegative; all the rest except the intercept are anticipated to be nonpositive. In the Bails and Peppers text the model and data are used as an instructive example of how algorithmic addition and deletion of variables can be used in conjunction with the informal imposition of sign constraints to yield a satisfactory final equation. The numbers of observations and regressors here seem to be typical of the fairly common situation in which sign constraints are imposed in an informal, descriptive regression equation.

Table III

Constraint Probabilities $p_{2|1}$, auto sales data set[a]

	Probability	Numerical standard error	Time[b]
Crude frequency simulator (1 evaluation per iteration)	2.0×10^{-4}	1.3×10^{-4}	15.96
GHK probability simulator (1 evaluation per iteration)	8.02×10^{-5}	$.38 \times 10^{-5}$	70.52
GHK probability simulator (25 evaluations per iteration)	9.06×10^{-5}	$.15 \times 10^{-5}$	1379.93

[a]All results are based on $m = 10,000$ iterations.
[b]In seconds. All computations were carried out on a Sun Sparc IIPC 4/40, using compiled Fortran code with extensive calls to the IMSL mathematical and statistical libraries.

Table III provides alternative numerical approximations to the constraint probabilities in the same style of presentation as Table I. Because the constraint probability $p_{2|1}$ is so small, the crude frequency simulator with m = 10,000 iterations provides an unsatisfactory approximation. The GHK probability simulator provides this small probability to one significant figure and clearly could provide it for two figures. The crude frequency simulator requires about $[(1.3\text{x}10^{-4})/(0.38\text{x}10^{-5})]^2 \times (15.96/70.53)$ = 265 times as much computation time to achieve the same numerical standard error as the GHK probability simulator. Increasing the number of evaluations by a factor of 25 leads to less than a five-fold reduction in numerical standard error. This indicates that variation in σ plays an important role in evaluation of the constraint probability. Employing the reference prior described in Section 2, the posterior odds ratio in favor of the constrained model is about $2^{10} \times 9 \times 10^{-5}$:1 = 0.092:1 when the prior odds ratio is 1:1.

Table IV

Posterior moments, auto sales data set[a]

	Crude frequency simulator			Gibbs sampler (No skips)			Gibbs sampler (Every 10th)		
Parameter	Mean	s.d.	NSE	Mean	s.d.	NSE	Mean	s.d.	NSE
β_1	−7.59	2.25	.06	−7.75	2.17	.25	−7.70	2.23	.10
β_2	.0242	.0025	.0001	.0240	.0023	.0003	.0242	.0024	.0001
β_3	.0405	.0125	.0003	.0403	.0125	.0008	.0410	.0120	.0003
β_4	−.0232	.0212	.0005	−.0236	.0221	.0004	−.0237	.0227	.0002
β_5	−3.212	1.110	.042	−3.458	1.093	.080	−3.190	1.122	.024
β_6	.1281	.0928	.0027	.1375	.1036	.0099	.1297	.0969	.0030
β_7	−.1282	.0296	.0009	−.1277	.0306	.0004	−.1287	.0307	.0002
β_8	33.04	26.08	.90	32.36	26.30	1.39	32.45	25.98	.53
β_9	−.421	.322	.012	−.372	.287	.027	−.432	.329	.012
β_{10}	−.0160	.0143	.0004	−.0146	.0128	.0010	−.0154	.0140	.0003
β_{11}	−.1285	.0953	.0020	−.1175	.0906	.0064	−.1270	.0957	.0020
σ	.5439	.0537	.0015	.5431	.0557	.0016	.5454	.0562	.0005
σ^2	.2987	.0599	.0016	.2981	.0621	.0017	.3006	.0629	.0006
Time[b]		13846.24			300.15			2939.66	

[a]All results are based on m = 10,000 iterations.
[b]In seconds. All computations were carried out on a Sun Sparc I IPC 4/40, using complied Fortran code with extensive calls to the IMSL mathematical and statistical libraries.

Posterior means and standard deviations for the model parameters are provided in Table IV. Results for the crude frequency simulator are based on m = 1,000 draws of (β, σ) from the support of the constrained posterior distribution. This required over four hours of execution time. The Gibbs sampler produces drawings with very high positive serial correlation for most parameters, and results for m = 10,000 with no skips yield very poor approximations and these results are not shown in Table IV. When every 10th draw is used until m = 10,000 draws have been obtained, numerical standard errors are as high as 10 percent of the posterior standard deviation for several coefficients. Increasing to every 100th draw lowers this error to about 4 percent. The Gibbs sampler requires about 49 minutes to achieve this accuracy, whereas the crude frequency simulator demands almost four hours. For this problem, the Gibbs sampler is clearly the method of choice.

5 Extensions

The treatment of linear inequality constraints developed in this paper may be extended and applied in models other than the normal linear regression model. The presentation here has limited treatment to inequality constraints on no more than k linear combinations of the k coefficients. More than k linear combinations can be handled by combining the GHK probability simulator and Gibbs sampler with an accept-reject algorithm. Given $s > k$ constraints, apply the Gibbs sampler using k of the linear constraints, but accept only those draws that satisfy the other $s - k$ constraints, and retain the acceptance rate. The retained sample converges in distribution to the posterior distribution. Use the GHK probability simulator to approximate the probability of the same k constraints, and then scale it by the acceptance rate from the application of the Gibbs sampler to obtain an approximation to $p_{2|1}$.

The semi-informative prior distribution with kernel density $\sigma^{-1}\exp\{-R\beta - r)'\Psi^{-1}(R\beta - r)\}$ can be used in place of (4). Conditional on σ the distribution of β is still truncated multivariate normal, and consequently the GHK probability simulator and Gibbs sampler may be applied in the same way. The only modification of substance is that since the variance of this multivariate normal distribution is $[\sigma^{-2}X'X + R'\Psi^{-1}R]^{-1}$, the Choleski decomposition of the $k \times k$ variance matrix must be recomputed each iteration in both the crude frequency simulator and the GHK probability simulator. In the Gibbs sampler the vectors c^j must be recomputed, requiring the inversion of a $k \times k$ matrix, each iteration. The times required to compute a Choleski decomposition, and to invert a matrix, are similar and proportional to the cube of the dimension of the matrix. Consequently when the number of regressors, k, is large (say, $k \geq 8$) computation time with a semi-informative prior will be roughly proportional to the number of iterations. In the automobile sales example, the crude frequency simulator requires more than 100 times as many iterations as the Gibbs sampler. For this case, the Gibbs sampler would clearly have been the method of choice had a semi-informative prior been employed.

The GHK probability simulator and the Gibbs sampler can generally be employed in any Gibbs sampling algorithm in which the linear inequality restrictions apply to elements of a vector whose conditional distribution is normal. This extension applies to a wide range of models, for example the seemingly unrelated regressions model, censored regression model, and the multinomial probit model. A particularly interesting class of cases is formed by scale-mixture normal models. For example, the models

$$y_t = x_t'\beta + \varepsilon_t, \quad \varepsilon_t \overset{IID}{\sim} t(0, \sigma^2; \nu)$$

and

$$y_t = x_t'\beta + \varepsilon_t, \quad \varepsilon_t \overset{ID}{\sim} N(0, \sigma^2 v_t)$$

are equivalent if in the prior distribution the v_t are mutually independent, $\nu v_t^{-1} \sim \chi^2(v)$ (Geweke 1993). Conditional on σ and the v_t, the distribution of β is truncated normal, and the GHK probability simulator and Gibbs sampler may be applied to treat linear inequality constraints in the way described in this paper. Other prior distributions for the v_t produce other unconditional distributions for the disturbance term. The methods for treatment of linear inequality constraints described in this paper may therefore be extended to a variety of nonnormal linear regression models.

References

Bails, D. G., and L. C. Peppers (1982), Business fluctuations. Englewood Cliffs: Prentice-Hall.

Bernardo, J. M., and A. F. M. Smith (1994), Bayesian Theory. New York: Wiley.

Chamberlain, G., and E. Leamer (1976), Matrix weighted averages and posterior bounds, Journal of the Royal Statistical Society, Series B, 38, 73–84.

Davis, W. W. (1978), Bayesian analysis of the linear model subject to linear inequality constraints, Journal of the American Statistical Association 73, 573–79.

Gelfand, A. E., and A. F. M. Smith (1990), Sampling based approaches to calculating marginal densities, Journal of the American Statistical Association 85, 398–409.

Geweke, J. (1986), Exact inference in the inequality constrained normal linear regulation model, Journal of Applied Econometrics 1, 127–41.

Geweke, J. (1991), Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. In E. M. Keramidas (ed.), Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface, pp. 571–78. Fairfax, VA: Interface Foundation of North America.

Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. O. Berger et al. (eds.), Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics. Oxford: Oxford University Press.

Geweke, J. (1993), Bayesian treatment of the student-t linear model, Journal of Applied Econometrics 8, S19–S40.

Geweke, J. (1995), Variable selection and model comparison in regression. Forthcoming in J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith (eds.), Proceedings of the Fifth Valencia International Meeting on Bayesian Statistics. Also, Research Department Working Paper 539, Federal Reserve Bank of Minneapolis, (1994).

Geweke, J., M. Keane, and D. Runkle (1995), Recursively simulating multinomial multiperiod probit probabilities, American Statistical Association 1994 Proceedings of the Business and Economic Statistics Section, forthcoming.

Gourieroux, C., A. Holly, and A. Monfort (1982), Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters, Econometrica 50, 63–80.

Hajivassiliou, V., and D. McFadden (1990), The method of simulated scores for the estimation of LDV models with an application to external debt crises. Cowles Foundation Discussion Paper 967, Yale University.

Hajivassiliou, V., D. McFadden, and P. Ruud (1995), Simulation of multivariate normal orthant probabilities: Methods and programs, Journal of Econometrics, forthcoming.

Judge, G. G., and T. Takayama (1966), Inequality restrictions in regression analysis, Journal of the American Statistical Association 61, 166–81.

Keane, M. (1990), Four essays in empirical macro and labor economics. Ph.D. dissertation, Brown University.

Keane, M. (1993), A computationally practical simulation estimator for panel data, Econometrica 62, 95–116.

Keane, M. (1994), Simulation estimation for panel data models with limited dependent variables. In G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.), Handbook of Statistics vol. 11, pp. 545–71. Amsterdam: Elsevier Science Publishers.

Leamer, E., and G. Chamberlain (1976), A Bayesian interpretation of pretesting, Journal of the Royal Statistical Society, Series B, 38, 85–94.

Lindley, D. V. (1957), A statistical paradox, Biometrika 44, 187–92.

Lovell, M. C., and E. Prescott (1970), Multiple regression with inequality constraints: pretesting bias, hypothesis testing, and efficiency, Journal of the American Statistical Association 65, 913–25.

Pindyck, R. S., and D. L. Rubinfeld (1981), Econometric models and economic forecasts. New York: McGraw-Hill.

Press, S. J. (1989), Bayesian statistics: Principles, models, and applications. New York: John Wiley.

Rao, C.R. (1965), Linear statistical inference and its applications. New York: Wiley.

Tierney, L., (1994), Markov chains for exploring posterior distributions, (with discussion and rejoinder), Annals of Statistics, 22: 1701–1762. (Also, School of Statistics Technical Report 560, University of Minnesota, (1991).

Wolak, F. A. (1987), An exact test for multiple inequality and equality constraints in the linear regression model, Journal of the American Statistical Association 82, 782–93.

Zellner, A. (1971), An introduction to Bayesian analysis in econometrics. New York: Wiley.