Federal Reserve Bank of Minneapolis
Research Department

Prior Density Ratio Class Robustness in Econometrics

John Geweke and Lea Petrella*

Working Paper 553

August 1995

ABSTRACT

This paper provides a general and efficient method for computing density ratio class bounds on posterior moments, given the output of a posterior simulator. It shows how density ratio class bounds for posterior odds ratios may be formed in many situations, also on the basis of posterior simulator output. The computational method is used to provide density ratio class bounds in two econometric models. It is found that the exact bounds are approximated poorly by their asymptotic approximation, when the posterior distribution of the function of interest is skewed. It is also found that posterior odds ratios display substantial variation within the density ratio class, in ways that cannot be anticipated by the asymptotic approximation.

JEL classification numbers: C11, C63

*Geweke, University of Minnesota and Federal Reserve Bank of Minneapolis; Petrella, Universita di Firenze and University of Minnesota. Without implicating him for errors and omissions, we thank Michael Lavine for useful discussions. The first author wishes to acknowledge partial financial support from NSF grant SES-9210070, and the second author thanks Professor G. Salinetti and L. Piccinato for encouragement and guidance. This work was undertaken while the second author was visiting the University of Minnesota.

1. Introduction

Good Bayesian investigators seek to report results in a way that will be most useful to their clients, that is, to those who read their work and go on to modify their opinions on the basis of the reported results, or who incorporate them in public or private decision making. Generally the investigator does not know what his clients' priors will be, or the posterior moments that interest them. Indeed, the investigator often does not know exactly who his clients will be.

There are a number of approaches to this Bayesian communication problem. An important one is to provide an indication of the sensitivity of posterior moments to changes in the prior distribution. This may be accomplished by reporting posterior moments corresponding to alternative priors, but such an enumeration likely becomes tiresome long before reasonable possibilities for prior distributions are exhausted. A more efficient approach is to report a range for each posterior moment, corresponding to all possible prior distributions within a specified class of distributions. Several interesting classes of prior distributions have been proposed; a concise and informative review is provided by Wasserman (1992). A thorough, recent overview of robust Bayesian analysis is given by Berger (1994).

In this paper we consider the density ratio class of prior distributions. This class consists of all prior distributions over the parameter vector θ with probability density kernel $\mathrm{p}(\theta)$, satisfying the inequalities $\mathrm{a}(\theta) \leq \mathrm{p}(\theta) \leq \mathrm{b}(\theta)$. The class has at least three attractions. First, it may be motivated with regard to prior elicitation. Suppose that an elicitation process seeks to establish $\mathrm{P}(A)/\mathrm{P}(B)$ for all subsets A and B of the parameter space Θ. Clearly there are limits on the degree of accuracy with which this can be accomplished. If one represents these limits by

$$k^{-1} \int_A \mathrm{r}(\theta)d\theta \Big/ \int_B \mathrm{r}(\theta)d\theta \leq \mathrm{P}(A)/\mathrm{P}(B) \leq k \int_A \mathrm{r}(\theta)d\theta \Big/ \int_B \mathrm{r}(\theta)d\theta$$

for all r-measurable A and B and specified $k>1$, then the corresponding density ratio class is given by $\mathrm{a}(\theta) = \mathrm{r}(\theta)$ and $\mathrm{b}(\theta) = k\,\mathrm{r}(\theta)$.

A second advantage of the density ratio class is that it applies to prior density kernels, and therefore to improper priors. Given one of the conventional diffuse priors, e.g. a Jeffreys prior or a flat prior, one may examine the sensitivity of posterior moments to density ratio classes about this prior by taking $\mathrm{a}(\theta)$ and $\mathrm{b}(\theta)$ proportional to the diffuse prior kernel density.

A third attraction of the density ratio class is its practicality. This fact is widely recognized, e.g. Wasserman (1992). In fact the density ratio class is very useful because bounds for posterior moments over density ratio classes may be computed on the basis of output from posterior simulators, e.g. importance sampling (Geweke, 1989) and Gibbs sampling (Gelfand and Smith, 1990). The objectives of this paper are to show how these computations may be accomplished routinely and efficiently, and then illustrate their application to some substantial econometric models. In meeting the first objective we build on the work of Lavine (1991a, 1991b) and especially Wasserman and Kadane (1992). Meeting the second objective answers the call of Wasserman (1992) for applications of density ratio robustness analysis in high-dimensional problems where robustness concerns are likely to be greatest.

The paper is written to be as self-contained as possible. In doing this we recapitulate some results due to others, providing proofs that use a common notation and are new in some cases. We also introduce a new algorithm for the key computation. The paper is organized as follows. The next section defines the problem and characterizes the solution. The main result here has been reported previously DeRobertis and Hartigan (1981) and Lavine (1991a). In Section 3 we extend the methods of Wasserman and Kadane (1992) for computing bounds on posterior moments from posterior simulators: we allow density ratio classes that need not include the prior used in the analysis, do not require the simulator to produce i.i.d. draws, and cope with weighting functions of the kind employed in importance sampling. Section 4 motivates the basic asymptotic relationship between the bounds on posterior moments over the density ratio class, and the posterior mean and standard deviation, first derived by DeRobertis and Hartigan (1981). Section 5 shows that the asymptotic approximation is excellent in a leading situation, linear functions of coefficients in the normal linear regression model. Section 6 takes up a dynamic time series model that is nonlinear in parameters and features non-Gaussian disturbances. We carry out the exact computations and find that the asymptotic approximations do not work well. An analysis of the relation between posterior densities and the density ratio class bounds appears to explain why this is so. The paper's final section offers some conclusions and conjectures.

2. Bounds on posterior moments

For a data set $\mathbf{Y}_T = \{\mathbf{y}_t\}_{t=1}^T$ a model specifies the probability density function of $\mathbf{Y}_T$ up to a vector of unknown parameters, $f(\mathbf{Y}_T|\theta) = \prod_{t=1}^T f(\mathbf{y}_t|\mathbf{Y}_{t-1}, \theta)$. For brevity, denote the likelihood function $L(\theta) = f(\mathbf{Y}_T|\theta)$. Given the function of interest $g(\theta)$, the posterior expectation of $g(\theta)$ corresponding to the prior density kernel $p(\theta)$ is

$$E = \frac{\int_\Theta g(\theta)L(\theta)p(\theta)d\theta}{\int_\Theta L(\theta)p(\theta)d\theta} = E(g, L, p). \quad (1)$$

The prior robustness question in general is the sensitivity of E to choice of $p(\theta)$. A *density ratio class* of prior distributions is all prior distributions with kernel densities that satisfy inequalities of the form $a(\theta) \le p(\theta) \le b(\theta)$; for brevity, $a \le p \le b$. The formal problem is to determine the range of values of posterior moments over this class of prior distributions, i.e., to minimize and maximize E in (1) subject to these inequalities; i.e., to find

$$\underline{E}(g, L; a, b) = \inf_{p:a \le p \le b} E(g, L, p) \text{ and } \overline{E}(g, L; a, b) = \sup_{p:a \le p \le b} E(g, L, p).$$

Since the technical issues involved in minimization and maximization of E are the same, we shall confine treatment to the maximization problem. The following result is shown by DeRobertis and Hartigan (1981). We provide a proof here because it parallels the similar result based on posterior simulators presented in Section 3.

Proposition 1. The conditions

(i) $p(\theta) = b(\theta)$ if $g(\theta) > f$; $p(\theta) = a(\theta)$ if $g(\theta) < f$; $a(\theta) \le p(\theta) \le b(\theta)$ if $g(\theta) = f$;

(ii) $E = f$

uniquely determine $\overline{E}(g, L; a, b)$.

Proof. We first show that if any prior density kernel $p^*(\theta)$ violates the stated conditions it will yield a smaller value of E. Given any $p(\theta)$ satisfying the conditions (i), construct $p^*(\theta) = a(\theta) \; \forall \; \theta: g(\theta) < f$ and $p^*(\theta) < b(\theta)$ on a subset of $\{\theta: g(\theta) > f\}$ with positive Lebesgue measure. Then

$$\int_\Theta g(\theta)L(\theta)p^*(\theta)d\theta = \int_\Theta g(\theta)L(\theta)p(\theta)d\theta + \int_\Theta g(\theta)L(\theta)[p^*(\theta) - p(\theta)]d\theta$$
$$< \int_\Theta g(\theta)L(\theta)p(\theta)d\theta + f \cdot \int_\Theta L(\theta)[p^*(\theta) - p(\theta)]d\theta$$
$$= \int_\Theta g(\theta)L(\theta)p(\theta)d\theta + E \cdot \int_\Theta L(\theta)[p^*(\theta) - p(\theta)]d\theta = E\int_\Theta L(\theta)p^*(\theta)d\theta.$$

Hence

$$\int_\Theta g(\theta)L(\theta)p^*(\theta)d\theta \Big/ \int_\Theta L(\theta)p^*(\theta)d\theta \le E = \int_\Theta g(\theta)L(\theta)p(\theta)d\theta \Big/ \int_\Theta L(\theta)p(\theta)d\theta.$$

A similar argument applies to $p^*(\theta) = b(\theta) \; \forall \; \theta: g(\theta) > f$ and $p^*(\theta) > a(\theta)$ on a subset of $\{\theta: g(\theta) < f\}$ with positive Lebesgue measure.

We finally show that any function $p(\theta)$ meeting the stated conditions yields posterior moment E. Denote $\Theta^* = \{\theta: g(\theta) = E\}$, $A = \int_{\overline{\Theta}^*} g(\theta)L(\theta)p(\theta)d\theta$, $B = \int_{\overline{\Theta}^*} L(\theta)p(\theta)d\theta$,

$s=\int_{\Theta^*}\mathrm{L}(\theta)\mathrm{p}(\theta)d\theta$. Then the posterior moment is $\mathrm{E}(s)=(A+fs)/(B+s)$, which is invariant with respect to the choice of s. ##

Figure 1 provides a graphical interpretation of the result that links it to what follows. Let the function $\mathrm{E}^*(f)$ denote the posterior moment corresponding to the function $\mathrm{p}(\theta)$ defined in condition (i) of Proposition 1. Figure 1 shows two such functions, $\mathrm{E}_1^*(f)$ and $\mathrm{E}_2^*(f)$. These functions will have the orientation shown: exactly one intersection with the 45° line, at the maximum of $\mathrm{E}^*(f)$. The function $\mathrm{E}_1^*(f)$ corresponds to a continuous function $\mathrm{g}(\theta)$, and the function $\mathrm{E}_2^*(f)$ corresponds to a function $\mathrm{g}(\theta)$ with two discrete values.

Since $\underline{\mathrm{E}}(\mathrm{g},\mathrm{L};\mathrm{a},\mathrm{b})=-\overline{\mathrm{E}}(-\mathrm{g},\mathrm{L};\mathrm{a},\mathrm{b})$, the solution of the problem of minimization of E over the same class of prior densities is essentially the same. In general, $f\colon \mathrm{E}(f)=f$ will not be the same for the minimization problem as for the maximization problem.

3. Systematic numerical approximations

Suppose that a simulator for the posterior distribution with kernel density $\mathrm{L}(\theta)\tilde{\mathrm{p}}(\theta)$ defines the stochastic process $\{\theta_m,\mathrm{w}(\theta_m)\}_{m=1}^{\infty}$, where θ_m is the m'th realization of the parameter vector and $\mathrm{w}(\theta_m)$ is an associated weight such that $\bar{g}_m=\sum_{m=1}^{M}\mathrm{g}(\theta_m)\mathrm{w}(\theta_m)\big/\sum_{m=1}^{M}\mathrm{w}(\theta_m)$ is an M-consistent approximation of $\bar{g}=\int_{\Theta}\mathrm{g}(\theta)\mathrm{L}(\theta)\tilde{\mathrm{p}}(\theta)d\theta\big/\int_{\Theta}\mathrm{L}(\theta)\tilde{\mathrm{p}}(\theta)d\theta$. Examples of such simulators include importance sampling (Geweke, 1989) and Markov chain Monte Carlo methods (Gelfand and Smith, 1990; Tierney, 1994). Without loss of generality suppose the simulator output is ordered so that $g_m=\mathrm{g}(\theta_m)$ is monotone nondecreasing. We wish to determine the robustness of $\bar{g}$ with respect to a prior density with kernel $\mathrm{p}(\theta)$, where $\mathrm{a}(\theta)\le\mathrm{p}(\theta)\le\mathrm{b}(\theta)$. Consider approximating the solution of (1) by maximizing

$$\mathrm{Q}(\mathbf{r})=\mathbf{g}'\mathbf{r}/\iota'\mathbf{r}=\sum_{m=1}^{M}g_m r_m\Big/\sum_{m=1}^{M}r_m$$

subject to $\mathbf{u}\le\mathbf{r}\le\mathbf{v}$, where $u_m=\mathrm{w}(\theta_m)\mathrm{a}(\theta_m)/\tilde{\mathrm{p}}(\theta_m)$ and $v_m=\mathrm{w}(\theta_m)\mathrm{b}(\theta_m)/\tilde{\mathrm{p}}(\theta_m)$ $(m=1,\ldots,M)$. We first develop an efficient method for solution of this problem, and then show that the solution provides an M-consistent approximation of $\overline{\mathrm{E}}(\mathrm{g},\mathrm{L};\mathrm{a},\mathrm{b})$.

3.1 Solution of the discrete problem

The solution of the discrete problem has properties quite similar to the solution of the exact problem given in Proposition 1.

Proposition 2. Let

$$Q_\ell = \frac{N_\ell}{D_\ell} = \frac{\sum_{m=1}^{\ell} g_m u_m + \sum_{m=\ell+1}^{M} g_m v_m}{\sum_{m=1}^{\ell} u_m + \sum_{m=\ell+1}^{M} v_m} \quad (\ell = 0,\ldots,M), \tag{2}$$

and let

$$m^* = \max\{\ell : g_\ell \le Q_\ell \le g_{\ell+1}\}.$$

Then

$$Q_{m^*} = \max_{\{r_m\}}\left[\sum_{m=1}^{M} g_m r_m \Big/ \sum_{m=1}^{M} r_m\right] \text{ s.t. } u_m \le r_m \le v_m \quad (m = 1,\ldots,M).$$

Proof. Since the sign of

$$\frac{\partial Q}{\partial r_m} = \frac{\left(\sum_{j=1}^{M} r_j\right) g_m - \sum_{j=1}^{M} g_j r_j}{\left(\sum_{j=1}^{M} r_j\right)^2} = \frac{g_m - Q(r_1,\ldots,r_M)}{\sum_{j=1}^{M} r_j}$$

does not involve r_m, a necessary condition for solution of the problem is $r_m = u_m$ if $g_m < Q(\mathbf{r})$ and $r_m = v_m$ if $g_m > Q(\mathbf{r})$. Hence all solutions are characterized by (2) and

$$g_\ell \le Q_\ell \le g_{\ell+1}. \tag{3}$$

Since $\{g_m\}$ is monotone nondecreasing, the largest value of ℓ such that (3) is true will provide the largest possible value of Q_ℓ. ##

Direct implementation of Proposition 2 would require that all Q_ℓ be computed. However, if we can rule out the possibility depicted in Figure 2, then a much more efficient procedure is available: find *one* ℓ such that (3) is satisfied, and then search $\ell+1, \ell+2,\ldots$ until an m is found such that $Q_m < g_m$. This is indeed the case.

Proposition 3. If $Q_\ell < g_\ell$, then $Q_{\ell+1} < g_{\ell+1}$.

Proof. $Q_\ell < g_\ell \Rightarrow Q_\ell < g_{\ell+1} \Rightarrow \frac{N_\ell}{D_\ell} + \frac{(u_{\ell+1} - v_{\ell+1}) g_{\ell+1}}{D_\ell} < g_{\ell+1} + \frac{(u_{\ell+1} - v_{\ell+1}) g_{\ell+1}}{D_\ell} \Rightarrow$

$N_\ell + (u_{\ell+1} - v_{\ell+1}) g_{\ell+1} < g_{\ell+1}(D_\ell + u_{\ell+1} - v_{\ell+1}) \Rightarrow \frac{N_\ell + (u_{\ell+1} - v_{\ell+1}) g_{\ell+1}}{D_\ell + u_{\ell+1} - v_{\ell+1}} < g_{\ell+1} \Rightarrow Q_{\ell+1} < g_{\ell+1}.$

##

Since the situation in Figure 2 cannot occur, our search for the highest ℓ need not go beyond the point at which $g_\ell > Q_\ell$. Proposition 3 forms the basis of a computationally efficient solution of the discrete problem, as follows.

Corollary. The index m^* of Proposition 2 may be determined as follows.

(1) Sort $\{g(\theta_m)\}$ so that $g_m = g(\theta_m)$ is monotone nondecreasing in m.

(2) Using successive bisection find an index ℓ such that $g_\ell \le Q_\ell \le g_{\ell+1}$.

(3) Increment the index ℓ of step (2) until $g_\ell > Q_\ell$ and then set $m^* = \ell - 1$.

To minimize Q_ℓ over the same class of prior densities, apply the algorithm to $-g(\theta_m)$ and reverse the sign of the solution.

3.2 Consistency of the sequence of discrete problem solutions

For many problems E of Proposition 1 cannot be obtained analytically, but for many prior density kernels $p(\theta)$ E can be approximated consistently (in M) by a posterior simulator. If this is true for all kernel densities $p(\theta)$: $a(\theta) \le p(\theta) \le b(\theta)$ then the solutions we obtain to the sequence of discrete problems defined by the output of the posterior simulator will converge almost surely, as $M \to \infty$, to the maximum value of E defined in Proposition 1.

Proposition 4. Suppose that $\{\theta_m, w(\theta_m)\}_{m=1}^{\infty}$ is a posterior simulator for the posterior density kernel $L(\theta)\tilde{p}(\theta)$ with the property that for any kernel density $p(\theta)$: $a(\theta) \le p(\theta) \le b(\theta)$,

$$\bar{g}_M = \sum_{m=1}^{M} g(\theta_m) r_m \Big/ \sum_{m=1}^{M} r_m \xrightarrow{a.s.} \int_\Theta g(\theta) L(\theta) p(\theta) d\theta \Big/ \int_\Theta L(\theta) p(\theta) d\theta = \bar{g} \qquad (4)$$

if $r_m = w(\theta_m) p(\theta_m) / \tilde{p}(\theta_m)$. Define Q_{m^*} as in Proposition 2. Then

$$Q_{m^*} \xrightarrow{a.s.} \sup_{p:a \le p \le b} \int_\Theta g(\theta) L(\theta) p(\theta) d\theta \Big/ \int_\Theta L(\theta) p(\theta) d\theta .$$

Proof. Define $E(f) = \int_\Theta g(\theta) L(\theta) p(\theta) d\theta \Big/ \int_\Theta L(\theta) p(\theta) d\theta$, with $p(\theta) = a(\theta)$ if $g(\theta) < f$ and $p(\theta) = b(\theta)$ if $g(\theta) \ge f$. Let $\{\theta_m\}$ be the output of the posterior simulator ordered so that $g(\theta_m)$ is monotone nondecreasing. Take $f^0 = \sup\{f : E(f) = f\}$ and $m^0 = \sup\{m : g(\theta_m) \le f^0\}$.

Since $Q_{m^*} \ge Q_{m^0} \xrightarrow{a.s.} E(f^0)$,, $Q_{m^*} \ge E(f^0) - \varepsilon$ is an almost sure event for any $\varepsilon > 0$.

Take $f^1 > f^0$; $f^1 - f^0$ may be arbitrarily close to 0. Define $m^1 = \sup\{m : g(\theta_m) \le f^1\}$.

Proposition 1, $E(f^1) < E(f^0)$, and $Q_{m^1} \xrightarrow{a.s.} E(f^1)$ imply that $Q_{m^1} < g_{m^1}$ is an almost sure event. From Propositions 2 and 3, so is $m^* < m^1$. Hence

$$Q_{m^*} \le g_{m^*+1} \le g_{m^1} \le f^1$$

is an almost sure event. (The first weak inequality follows directly from Proposition 2 and the third follows from the definition of f^1.) Since f^1 can be taken arbitrarily close to $f^0 = \mathrm{E}(f^0)$, $Q_{m^*} \leq \mathrm{E}(f^0) + \varepsilon$ is an almost sure event for any $\varepsilon > 0$. ##

Condition (4) is weak. For example, the bound $\mathrm{b}(\theta)/\tilde{\mathrm{p}}(\theta) < \infty \ \forall\ \theta \in \Theta$ is sufficient for (4); see Geweke(1995).

4. Bounds on posterior moments in large samples

Given some standard regularity conditions, in large samples the distribution of $\mathrm{g}(\theta)|\mathbf{Y}_T$ is approximately normal with variance proportional to $T^{-1/2}$. A good summary discussion of these conditions is provided by Bernardo and Smith (1994, 285 - 297); for greater detail, see Heyde and Johnstone (1979), Hartigan (1983, Chapter 4), and Chen (1985). As the posterior variance becomes small, a continuous function of interest $\mathrm{g}(\theta)$ becomes locally linear and the upper and lower bounds $\mathrm{b}(\theta)$ and $\mathrm{a}(\theta)$ of the density ratio class robustness problem become locally constant. It is therefore of some interest to consider robustness of the posterior mean of

$$u = g_0 + \mathbf{g}_1'\theta,\ u|\mathbf{Y}_T \sim \mathrm{N}(\bar{\mathrm{g}}, \sigma^2)$$

over the class of prior density kernels $\mathrm{p}(\theta), a = \mathrm{a}(\theta) < \mathrm{p}(\theta) < \mathrm{b}(\theta) = b$.

From Result 1,

$$\overline{\mathrm{E}}(\mathrm{g}, \mathrm{L}; \mathrm{a}, \mathrm{b}) = \sup_f \frac{a\int_{-\infty}^{f} u\phi\left[(u-\bar{g})/\sigma\right]du + b\int_f^{\infty} u\phi\left[(u-\bar{g})/\sigma\right]du}{a\int_{-\infty}^{f} \phi\left[(u-\bar{g})/\sigma\right]du + b\int_f^{\infty} \phi\left[(u-\bar{g})/\sigma\right]du} = \bar{g} + \sigma \cdot \sup_f \beta(f,k)$$

with

$$\beta(f,k) = \left[\int_{-\infty}^{f} x\phi(x)dx + k\int_f^{\infty} x\phi(x)dx\right] \Bigg/ \left[\int_{-\infty}^{f} \phi(x)dx + k\int_f^{\infty} \phi(x)dx\right] = \frac{-\phi(f) + k\phi(f)}{\Phi(f) + k\left[1 - \Phi(f)\right]}$$

and

$$k = b/a,$$

where $\phi(z) = (2\pi)^{-1/2}\exp(-z^2/2)$ and $\Phi(x) = \int_{-\infty}^{x} \phi(z)dz$. Standard first-order conditions show that $\mathrm{f}^*(k) = \arg\max_f \beta(f,k)$ is the unique root of

$$(k-1)\left[\phi(f) + f\Phi(f)\right] - kf = 0.$$

Defining $\gamma(k) = \beta\left[\mathrm{f}^*(k), k\right]$, we have

$$\sup_{\mathrm{p}:\mathrm{a} \leq \mathrm{p} \leq \mathrm{b}} \int_{\Theta} \mathrm{g}(\theta)\mathrm{p}(\theta|\mathbf{Y}_T)d\theta = \bar{g} + \sigma\gamma(b/a).$$

The same analysis shows

$$\inf_{p:a\le p\le b}\int_{\Theta} g(\theta)\,p(\theta|\mathbf{Y}_T)d\theta = \bar{g} - \sigma\gamma(b/a).$$

Figure 3 shows the function $\gamma(k)$ which is monotone increasing, strictly concave, and unbounded as $k \to \infty$.

This leading case provides the behavior of the upper and lower bounds on $\bar{g}_T = \mathrm{E}[g(\theta)|\mathbf{Y}_T]$ as $T \to \infty$, under standard regularity conditions. We recapitulate a result due to DeRobertis and Hartigan (1981). Let $\sigma_T^2 = \mathrm{var}[g(\theta)|\mathbf{Y}_T]$ and let $\overline{\mathrm{E}}_T(\mathrm{g,L;a,b})$ denote $\overline{\mathrm{E}}(\mathrm{g,L;a,b})$ in a sample of size T. The function of interest $g(\theta)$ is *asymptotically normal under the prior* $p(\theta)$ if for all bounded functions $f:\mathrm{R}^1 \to \mathrm{R}^1$,

$$\mathrm{E}\left\{f\left[\left(g(\theta)-\bar{g}_T\right)/\sigma_T\right]\middle|\mathbf{Y}_T\right\} \xrightarrow{a.s.} \int_{-\infty}^{\infty} f(t)\phi(t)dt.$$

Proposition 5. Suppose

(i) $\theta|\mathbf{Y}_T \xrightarrow{a.s.} \theta^*$;

(ii) The functions $a(\theta)$ and $b(\theta)$ are continuous at $\theta = \theta^*$;

(ii) $g(\theta)$ is asymptotically normal under the prior with kernel $b(\theta)$.

Then except possibly for θ^* on a set of b-measure zero,

$$\sigma_T^{-1}\left\langle \overline{\mathrm{E}}_T(\mathrm{g,L;a,b}) - \left\{\bar{g}_T + \sigma_T\gamma\left[b(\theta^*)/a(\theta^*)\right]\right\}\right\rangle \xrightarrow{a.s.} 0$$

5. Linear regression

Perhaps the most commonly applied model in econometrics is the normal linear regression model,

$$\underset{T\times 1}{\mathbf{y}} = \underset{T\times k}{\mathbf{X}}\beta + \underset{T\times 1}{\varepsilon}, \quad \varepsilon \sim \mathrm{N}\left(\mathbf{0}, \sigma^2\mathbf{I}_T\right)$$

in which each row corresponds to one of T observations, $\mathbf{X}$ is a matrix of fixed or strictly exogenous regressors, $\mathbf{y}$ is a vector of corresponding dependent variables, and ε is a vector of disturbances. Perhaps the most commonly employed prior distribution in Bayesian approaches to the normal linear regression model is

$$p(\beta, \sigma^2) \propto \sigma^{-1},$$

which is of course improper. This prior distribution can be motivated by appeal to Jeffreys' invariance principle (e.g. Zellner, 1971, 41-53).

It is well known that the posterior distribution for β in this model is

$$\beta|(\mathbf{y},\mathbf{X}) \sim t\left(\mathbf{b}, s^2(\mathbf{X}'\mathbf{X})^{-1}; \nu\right).$$

That is, the posterior distribution of β is multivariate Student-t with location parameter $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, scale matrix $s^2(\mathbf{X}'\mathbf{X})^{-1}$ where $s^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})/\nu$, and $\nu = T - k$ degrees of freedom. (For derivation, see Zellner, 1971, 66-70.) In many applications, linear functions of β are of interest. Since the posterior distribution of β is multivariate Student-t, for any vector $\mathbf{c} \neq \mathbf{0}$,

$$\mathbf{c}'\beta|(\mathbf{y},\mathbf{X}) \sim \mathrm{t}\left(\mathbf{c}'\mathbf{b}, s^2\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c};\ \nu\right).$$

That is, the posterior distribution of $\mathbf{c}'\beta$ is univariate Student-t with location parameter $\mathbf{c}'\mathbf{b}$, scale parameter $s^2\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$, and ν degrees of freedom.

Consider the density ratio class of prior density kernels,

$$\mathrm{p}(\beta, \sigma^2):\ a\sigma^{-1} \leq \mathrm{p}(\beta, \sigma^2) \leq b\sigma^{-1}.$$

Proceeding as in Section 4, we have from Result 1 that for $\mathrm{g}(\beta) = \mathbf{c}'\beta$,

$$\overline{\mathrm{E}}(\mathrm{g}, \mathrm{L}; \mathrm{a}, \mathrm{b}) = \mathbf{c}'\mathbf{b} + \left[s^2\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}\right]^{1/2}\tilde{\gamma}(k;\ \nu)$$

where $k = b/a$. The function $\tilde{\gamma}(k;\ \nu)$ can be derived in the same way as the function $\gamma(k)$ in Section 4, replacing the standardized normal density and distribution $\phi(\cdot)$ and $\Phi(\cdot)$ with $\mathrm{t}(\cdot;\ \nu)$ and $\mathrm{T}(\cdot;\ \nu)$, respectively. Thus,

$$\tilde{\beta}(f,k;\ \nu) = \left[\int_{-\infty}^{f} x\,\mathrm{t}(x;\ \nu)dx + k\int_{f}^{\infty} x\,\mathrm{t}(x;\ \nu)dx\right] \Big/ \left\{\mathrm{T}(f;\ \nu) + k[1 - \mathrm{T}(f;\ \nu)]\right\},$$

$\tilde{\mathrm{f}}^*(k;\ \nu) = \arg\max_f \tilde{\beta}(f,k;\ \nu)$ is the unique root of

$$(k-1)\left[\int_{-\infty}^{f} x\,\mathrm{t}(x;\ \nu)dx + f\int_{f}^{\infty} x\,\mathrm{t}(x;\ \nu)dx\right] - kf = 0,$$

and

$$\tilde{\gamma}(k;\ \nu) = \tilde{\beta}\left[\tilde{\mathrm{f}}^*(k;\ \nu), k;\ \nu\right].$$

Figure 4 shows $\tilde{\gamma}(k;\ \nu)$, for $\nu \to \infty$, $\nu = 15$, $\nu = 6$, and $\nu = 3$. The case $\nu \to \infty$ is of course exactly the one portrayed in Figure 3. All functions are strictly concave to the horizontal axis. For all ν, $\tilde{\gamma}(k;\ \nu)/\gamma(k)$ is an increasing function of k, with $\lim_{k\to 1}\tilde{\gamma}(k;\ \nu)/\gamma(k) > 1$ and $\lim_{k\to\infty}\tilde{\gamma}(k;\ \nu)/\gamma(k) = \infty$. That is, the bound $\gamma(k)$ yielded by the asymptotic analysis is a better approximation the smaller is k, but it does not become perfect as $k \to 1$. And as k becomes very large, the approximation becomes very poor.

Figure 4 shows that at least for modest degrees of freedom ($\nu \geq 15$) and small values of k ($k \leq 70$) the asymptotic approximation of Section 4 is quite good; in the worst case, $\nu = 15$ and $k = 70$, the approximation is correct to within about 10%. For linear functions of interest of the coefficient vector in the normal linear model, the asymptotic approximation will be excellent in most cases.

6. An example: The Markov chain model

The first-order Markov chain model is widely used to describe transitions of an entity between a finite number of discrete states over time. For example, it is often used in epidemiology to model transitions of individuals between infected and noninfected states (e.g., Singer and Cohen, 1980). In economics it provides a simple way to describe the labor market status of individuals as employed, unemployed, or not in the labor force (Abowd and Zellner, 1985). It has also been a useful tool in studying the mobility of individuals over different earnings quantiles (Shorrocks, 1978). Classical methods of inference in the Markov chain model are described by Bartholomew (1973). A precursor to Bayesian inference is provided by Zellner's treatment of the multinomial distribution (Zellner, 1971) and an early Bayesian treatment is that of Geweke, Marshall and Zarkin (1986).

We use this model to illustrate numerical methods of studying density ratio class robustness for two reasons. First, there is a posterior simulator for this model that is simple, yet to our knowledge has not been presented in the literature. Second, typical posterior moments of interest in this model do not have closed-form analytic expressions, thus requiring the use of numerical methods.

6.1 The model and data

Index individuals by $\ell = 1,\ldots,L$, time by $t = 1,\ldots,T$, and states by $i = 1,\ldots,m$. Let $s_{\ell t}$ denote the state of individual ℓ at time t. In the first-order Markov chain model the probability that an individual occupies state j at time t, conditioned on the history of all individuals from time 1 to time $t-1$ depends only on the individual's own state at time $t-1$. Hence if the probability that an individual is in state i at time t is denoted π_{it} we may write

$$\pi_{jt} = \sum_{i=1}^{m} \pi_{i,t-1} p_{ij},$$

where p_{ij} is the probability that an individual who was in state i at time $t-1$ is in state j at time t. Defining $\pi_t = (\pi_{1t},\ldots \pi_{mt})'$ and $\underset{m\times m}{\mathbf{P}} = [p_{ij}]$ we have

$$\pi_t' = \pi_{t-1}'\mathbf{P}.$$

If the distribution of individuals over states at time 1 is fixed exogenously, and the states of the L individuals are observed from time 1 to time T, then the likelihood function is

$$\mathrm{L}(\mathbf{P}) = \prod_{i=1}^{m}\prod_{j=1}^{m} p_{ij}^{n_{ij}} \qquad \left[p_{ij} \geq 0\,(i,j=1,\ldots m);\ \sum_{j=1}^{m} p_{ij} = 1\,(i=1,\ldots m)\right],$$

where n_{ij} is the total number of observed transitions from state i at time $t-1$ to state j at time t over all L individuals at times $t=2,\ldots,T$.

The likelihood function is the product of m kernels of the multinomial distribution indexed by i: $\prod_{j=1}^{m} p_{ij}^{n_{ij}}$ $(i=1,\ldots,m)$. A natural conjugate prior distribution for the multinomial distribution is the Dirichlet (also known as the multivariate beta) distribution,

$$\mathrm{p}_i\left(p_{ij}\right) \propto \prod_{j=1}^{m} p_{ij}^{a_{ij}} \quad \left[p_{ij} \geq 0\,(j=1,\ldots,m); \sum_{j=1}^{m} p_{ij}=1; a_{ij} \geq -1\,(j=1,\ldots,m)\right]. \qquad (5)$$

(For further discussion see Zellner, 1971, 38-39.) Hence a natural conjugate prior distribution for the first-order Markov chain model is

$$\mathrm{p}(\mathbf{P}) \propto \prod_{i=1}^{m} \prod_{j=1}^{m} p_{ij}^{a_{ij}}$$

$$\left[p_{ij} \geq 0\,(i,j=1,\ldots,m); \sum_{j=1}^{m} p_{ij}=1\,(i=1,\ldots,m); a_{ij} \geq -1\,(i,j=1,\ldots,m)\right]$$

and the corresponding posterior density kernel is

$$\mathrm{p}\left[\mathbf{P} \middle| s_{\ell t}\,(\ell=1,\ldots,L; t=1,\ldots,T)\right] \propto \prod_{i=1}^{m} \prod_{j=1}^{m} p_{ij}^{n_{ij}+a_{ij}}$$

$$\left[p_{ij} \geq 0\,(i,j=1,\ldots,m); \sum_{j=1}^{m} p_{ij}=1\,(i=1,\ldots,m)\right].$$

In the illustration here we use a flat prior distribution ($a_{ij}=0\,(i,j=1,\ldots,m)$). The data are constructed from the Penn World Tables, version 5.6 (Summers and Heston, 1991). The raw data is real GDP per worker in 1985 international prices, for each of 125 countries, in 1960 and 1985. The raw data is then converted to relative income levels by dividing each observation by a population weighted average of GDP per worker for that year. Finally, relative income levels are partitioned into six states: less than 1/8; 1/8 to 1/4; 1/4 to 1/2; 1/2 to 1; 1 to 2; and greater than 2. The resulting values of the n_{ij} are shown in Table 1; e.g., there were 8 countries that were in state 3 (1/4 to 1/2) in 1960 and then state 2 (1/8 to 1/4) in 1985. The data and model are similar to those of Chari, Kehoe and McGrattan (1995) in their study of the dynamics of international differences in real growth rates.

6.2 The posterior simulator

The posterior density kernel is the product of m kernels of the multivariate beta density,

$$\prod_{j=1}^{m} p_{ij}^{n_{ij}+a_{ij}} \quad \left[p_{ij} \geq 0\,(j=1,\ldots,m); \sum_{j=1}^{m} p_{ij}=1\right] \quad (i=1,\ldots,m). \qquad (6)$$

There is a convenient genesis for this density (Johnson and Kotz, 1972, 232-233). Construct the m^2 independent random variables

$$d_{ij} \sim \chi^2\left[2\left(n_{ij}+a_{ij}+1\right)\right] \quad (i,j=1,\ldots,m). \qquad (7)$$

Then

$$p_{ij} = d_{ij} \Big/ \sum_{k=1}^{m} d_{ik} \quad (i, j = 1, \ldots, m) \tag{8}$$

has probability density kernel (6). It is thus simple to construct i.i.d. drawings from the posterior distribution. On a Sun 20 Sparcstation, 10,000 draws for a six-state model ($m = 6$) requires 8.2 seconds. Computation is independent of sample size and proportional to m^2.

6.3 Posterior moments and their robustness

Under conditions that are not very restrictive, the first-order Markov model implies a unique, equilibrium value of π_t. The vector π^* is an equilibrium value of π_t if and only if $\pi^{*\prime} = \pi^{*\prime}\mathbf{P}$, i.e., π^* is a left eigenvector of $\mathbf{P}$ corresponding to the eigenvalue 1. (Since the entries of each row of $\mathbf{P}$ sum to 1, there must always be such an eigenvalue.) If there is only one eigenvalue of 1, then π^* is unique. An eigenvalue of $\mathbf{P}$ cannot exceed 1, and if all eigenvalues but one have modulus less than unity, then π_t converges to π^* from any initial value. For any Dirichlet prior (5), the probability that π^* is unique and that π_t therefore converges to π^* from any initial value, is 1.

The posterior density and moments of π^* can be computed by evaluating the appropriate left eigenvector of $\mathbf{P}$ for each $\mathbf{P}$ drawn from the posterior distribution by means of (7) and (8). (This evaluation requires 32.1 seconds for 10,000 draws of $\mathbf{P}$, on a Sun 20 Sparcstation. Computation time is proportional to m^3.) Posterior densities for π_1^* and π_5^* are provided in Figure 5. We focus on these two values because of the contrast in their posterior densities: while both are skewed to the right, the skewness for π_1^* is considerably stronger than that for π_5^*.

Figure 6 presents upper and lower bounds for the posterior expectation of π_1^*, for prior density kernels $\mathrm{p}(\theta)$ in the class $1 \le \mathrm{p}(\theta) \le k$, Values of k are plotted on the horizontal axis. The vertical axis indicates the bounds; it is centered at the original ($k = 1$) posterior mean of .082. The posterior mean of π_1^* -- the equilibrium probability for the lowest output state -- is quite sensitive to changes in the prior within the density ratio class. When k=4, it ranges from .065 to .103. At all values of k, variations in the prior can increase the posterior mean of π_1^*, more than they can lower it: e.g., at k=50 the range extends from .043 to .152. The dashed line indicates the asymptotic normal approximation to the bounds, taken from the original posterior mean of .082 and standard deviation of .036 and the function $\gamma(k)$. The approximation is poor, understating the upward sensitivity and understating the downward sensitivity.

Lower and upper bounds for the posterior mean of π_5^* are indicated in Figure 7. Here, the asymptotic approximation is excellent. Since the posterior mean of π_5^* is .211 and its posterior standard deviation is .058, variations in the prior within the density ratio class have a smaller effect, relatively, on π_5^*, than they do on π_1^*. The asymptotic approximation understates upward sensitivity and overstates downward sensitivity but the effect is mild, amounting to less than 10%.

7. An example: Trend and difference stationarity

Beginning with the investigation of Nelson and Plosser (1982), the propositions that most macroeconomic aggregates are trend stationary or, alternatively, that they are difference stationary, have captured the attention of applied and theoretical econometricians. These hypotheses have been approached using Bayesian methods by a number of investigators, including Zellner and Tiao (1964), Sims (1988), DeJong and Whiteman (1991), Phillips (1991), Sims and Uhlig (1991), Schotman and van Dijk (1991a, 1991b, 1992) and Geweke (1994).

Consideration of the sensitivity of inference and conclusions about trend and difference stationarity in these models to different prior distributions within the density ratio class is an interesting task for two reasons. First, because questions about trend and difference stationarity involve the long run, conclusions are likely to be sensitive to prior distributions; arguments to this effect are made in Sims (1988) and Geweke (1994). Second, the parameters of interest in models best suited to study of these questions enter nonlinearly (Schotman and van Dijk, 1991a) and there is strong evidence that disturbances are leptokurtic (Geweke, 1993). One may therefore anticipate that posterior expectations of functions of interest might be sensitive to prior distributions within the density ratio class, and that this sensitivity might not be well approximated by the asymptotic analysis of DeRobertis and Hartigan presented in Section 4.

7.1 The model and data

The essentials of the model are presented here. For further discussion the reader is referred to Geweke (1994). The trend stationary model is

$$y_t = \gamma + \delta t + u_t, \tag{9}$$

$$A^*(L)u_t = \varepsilon_t; \quad A^*(L) = (1-\rho L) - A(L)(1-L); \quad 0 \le \rho < 1, \tag{10}$$

$$\varepsilon_t \overset{IID}{\sim} \mathrm{t}(0, \sigma^2; \nu).$$

A little manipulation of (9) and (10) yields

$$y_t = \gamma(1-\rho) + \delta\left(\rho - \sum_{j=1}^4 a_j\right) + \delta(1-\rho)t + \rho y_{t-1} + \sum_{j=1}^4 a_j\left(y_{t-j} - y_{t-j-1}\right) + \varepsilon_t$$

and the likelihood function may be expressed

$$\sigma^{-T}\prod_{t=1}^{T}\left[1+\varepsilon_t^2/\nu\sigma^2\right]^{-(\nu+1)/2}.$$

For computation it is convenient to adopt the hierarchical representation

$$\nu_t^{-1} \sim \chi^2(\nu)/\nu, \quad \varepsilon_t|\nu_t \sim \mathrm{N}(0, \sigma^2\nu_t).$$

This yields the equivalent likelihood function

$$\prod_{t=1}^{T}\nu_t^{-(\nu+3)/2}\exp\left[\sum_{t=1}^{T}\left(\sigma^{-2}\varepsilon_t^2+\nu\right)/2\nu_t\right].$$

The prior distribution consists of several independent components. An important one is

$$\mathrm{p}_1(\rho)=(s+1)\rho^s\chi_{(0,1)}(\rho). \tag{11}$$

In the analysis below in Section 7.3 we focus on $s=9$ to keep the presentation compact. Geweke (1994) compares results for different values of s. In Section 7.4 we focus on $s \to \infty$, which yields the difference stationary model with $\rho=1$,

$$y_t-y_{t-1}=\delta+u_t, \quad u_t=\sum_{j=1}^{4}a_ju_{t-j}+\varepsilon_t,$$

and on the posterior odds ratio for the models distinguished by $s=9$ and $s\to\infty$.

The prior specification for $a_1,\ldots,a_4$ is

$$a_j \sim \mathrm{N}\left(0,\pi_0\pi_1^j\right) \quad \left(\pi_0=.731, \pi_1=.342\right).$$

This is a "Minnesota prior" (Doan, Litterman and Sims, 1984) in which the prior standard deviation of a_1 is .5 and the prior standard deviation of a_4 is .1. The prior distribution of γ is diffuse:

$$\gamma|y_0 \sim \mathrm{N}\left(y_0,\sigma_\gamma^2\right) \quad \left(\sigma_\gamma=10\right).$$

Finally,

$$\delta \sim \mathrm{N}\left(\overline{\delta},\sigma_\delta^2\right) \quad \left(\overline{\delta}=0, \sigma_\delta=.05\right),$$

$$\mathrm{p}_2(\sigma)\propto\sigma^{-1},$$

and

$$\mathrm{p}_3(\nu)=\omega\exp(-\omega\nu) \quad (\omega=.25).$$

The example presented here uses for $\{y_t\}_{t=1}^{T}$ the real GNP time series of Nelson and Plosser (1982). Data were furnished by Charles Nelson, and the least-squares estimates reported in Nelson and Plosser (1982, Table 5) were reproduced to all reported places. The sample period for the results here is the same as that used by Nelson and Plosser (1982), except for a few early observations that had to be dropped because more lagged values were used here.

7.2 The posterior simulator

The full posterior density for all parameters, including $\{\nu_t\}_{t=1}^{T}$, is intractable, but conditional distributions are relatively simple. In particular, the joint conditional distribution of (γ, δ) is bivariate normal, and that of (a_1, a_2, a_3, a_4) is multivariate normal. The conditional distribution of ρ has kernel density

$$\rho^s \exp\left[-(\rho-\hat{\rho})^2/2\lambda^2\right]\chi_{(0,1)}(\rho)$$

where $\hat{\rho}$ and λ are known functions of the other parameters and the data. The conditional distribution of ν has kernel density

$$(\nu/2)^{(T\nu/2)}\Gamma(\nu/2)^{-T}\exp(-\eta\nu)$$

where $\eta = \frac{1}{2}\sum_{t=1}^{T}\left[\log(v_t)+v_t^{-1}\right]+\omega$. The $v_1,\ldots,v_T$ are conditionally independent with

$$\left(\sigma^{-2}\varepsilon_t^2+\nu\right)/v_t \sim \chi^2(\nu+1).$$

Finally, the conditional distribution of σ^2 is

$$\sum_{t=1}^{T}\left(\varepsilon_t^2/v_t\right)/\sigma^2 \sim \chi^2(T).$$

The posterior density satisfies the conditions of Roberts and Smith (1994) for convergence of a Gibbs sampler to the posterior distribution. A Gibbs sampler may therefore be used to construct a posterior simulator, using six blocks corresponding to the foregoing six conditional distributions. The conditional distributions of ρ and ν are unconventional; efficient algorithms for drawing from these distributions are developed in Geweke (1992, Appendices A and B).

The output of the simulator provides the sequence of parameter vectors $\{\theta_m\}$, which may be used to approximate the upper and lower bounds for any function of interest for which the posterior expectation exists, as described in Section 3. Results presented here are based on $M = 10{,}000$ iterations of the Gibbs sampler, which required 90 seconds on a Sun 20 Sparcstation.

7.3 Posterior moments and their robustness.

Figure 8 provides posterior densities for three parameters of particular interest in the trend-stationary model: δ, the trend growth rate; ν, the degrees-of-freedom parameter of the Student-t distribution of the shocks; and ρ, the leading root of the autoregressive process for y_t. In each case the posterior density is plotted as a heavy line, and the prior density as a light line. Notice that the posterior distribution of δ is approximately Gaussian whereas those of ν and ρ are not.

Figure 9 presents upper and lower bounds for the posterior expectation of δ, for prior density kernels $\mathrm{p}(\theta)$ in the class

$$\tilde{\mathrm{p}}(\theta) \le \mathrm{p}(\theta) \le k\,\tilde{\mathrm{p}}(\theta),$$

where $\tilde{\mathrm{p}}(\theta)$ is the prior density described in Section 7.1. Values of k are plotted on the horizontal axis. The vertical axis indicates the bounds; it is centered at the posterior mean of 3.05%. The solid line indicates the actual upper and lower bounds, computed as described in Section 3. The bounds are nearly symmetric about the posterior mean. The mean displays some sensitivity to variation within the density ratio class: e.g., variations by a factor of 40 lead to an upper bound on the trend growth rate half again as great as the lower bound. The dashed line indicates the asymptotic normal approximation to the bounds, taken from the posterior mean and standard deviation and the function $\gamma(k)$. The approximation is good, though not nearly as good as would be the case in the linear model discussed in Section 5 with the same degrees of freedom. In view of the near-normality of the posterior density of δ in Figure 8, the quality of the approximation is perhaps not surprising.

Figure 10 presents upper and lower bounds for the posterior expectation of ν. The horizontal axis is the same as in Figure 9, the vertical axis is again centered at the posterior mean (6.0, for ν), and this axis has been scaled so that one posterior standard deviation is the same distance in Figures 9 and 10. Reflecting the imprecision of the posterior density for ν displayed in Figure 8, the bounds encompass a substantial range for this parameter, e.g. from $\nu = 3$ to $\nu = 13$ when $k = 40$. The asymmetry of the bounds is a reflection of the asymmetry in the posterior density, which in turn is due to the characteristics of the Student-*t* distribution itself: a change of one degree of freedom implies a more substantial change in the distribution when ν is lower than when ν is greater. The asymptotic approximation is quite poor for this posterior moment. It systematically overstates the ability of priors within the density ratio class to lower the posterior mean of ν, and understates their ability to increase it. This effect may be understood from the skewed posterior distribution of ν indicated in Figure 8. The data speak strongly against values of ν much below the posterior mean, relative to values much above it. Hence the intervals in Figure 10 are centered above the horizontal axis.

Figure 11 presents upper and lower bounds for the posterior expectation of ρ. The horizontal and vertical axes are set up the same as in Figures 9 and 10, with the vertical scaling again set so that one posterior standard distribution is the same distance in all three figures. Variation in the posterior mean of ρ, over the density ratio class of priors, is quite substantial. For example, when $k = 50$ the lower bound is less than .80 and the upper bound is about .97. The asymmetry in the bounds is due to the closeness of the original posterior mean to the upper bound of the support of ρ, and to the asymmetric shape of the posterior density exhibited in Figure 8. The asymptotic normal approximation to the bounds is better than is the approximation for ν, but not as good as the approximation for δ. As was the case with ν, the difference between the actual and approximate bounds may be understood in terms of the original posterior density in Figure 8. The posterior distribution for ρ is skewed to the left, in part because of the upper truncation of the support at $\rho = 1$. This limits the ability of prior distributions within the density ratio class to raise the posterior mean much above its original value of .90. The potential to lower the mean is greater. Once again, skewness in the posterior distribution is key in the discrepancy between the exact and asymptotic bounds.

7.4 Posterior odds ratios

Since much of the literature has focused on whether macroeconomic time series are trend or difference stationary, it is of some interest to compute the posterior odds ratio for the hypotheses

$$H_A: \quad \rho = 1$$

and

$$H_B: \quad 0 < \rho < 1.$$

Given H_B the prior distribution for ρ is (11). Under both hypotheses, the prior distributions for all other parameters are the ones indicated above in Section 7.1. (For further discussion of these points see Schotman and van Dijk (1991a, 1991b) and Geweke (1994).) For conveneince we take the prior odds ratio to be 1:1 in what follows.

There is a convenient generic approach to computing the posterior odds ratio (POR) that may be applied in situations like this. In the general case, suppose that two models (A and B) have the same likelihood function $L(\theta)$ and differ only in their prior distributions, which have probability density kernels $p_A(\theta) = \pi_A(\theta)p(\theta)$ and $p_B(\theta) = \pi_B(\theta)p(\theta)$ respectively. The functions $\pi_A(\theta)$ and $\pi_B(\theta)$ are normalized proper p.d.f.'s, but $p(\theta)$ need not be proper. In this case it is of interest to examine the robustness of the posterior odds ratio over the density ratio class $p(\theta)$: $a(\theta) \le p(\theta) \le b(\theta)$. This class leaves the prior density ratio $p_A(\theta)/p_B(\theta) = \pi_A(\theta)/\pi_B(\theta)$ unaffected, while varying the absolute levels of the densities within the bounds specified by the density ratio class. Since

$$\text{POR} = \frac{\int_\Theta \pi_A(\theta)p(\theta)L(\theta)d\theta}{\int_\Theta \pi_B(\theta)p(\theta)L(\theta)d\theta} = \frac{\int_\Theta [\pi_A(\theta)/\pi_B(\theta)]\pi_B(\theta)p(\theta)L(\theta)d\theta}{\int_\Theta \pi_B(\theta)p(\theta)L(\theta)d\theta},$$

one may compute the posterior odds ratio by taking $g(\theta) = \pi_A(\theta)/\pi_B(\theta)$ as the function of interest in Model B. All the methods of Sections 1 - 4 apply to the analysis of this situation, with $g(\theta) = \pi_A(\theta)/\pi_B(\theta)$ the function of interest, and $p(\theta)$: $a(\theta) \le p(\theta) \le b(\theta)$ the density ratio class.

Next, consider a specialization of this approach to the case in which $\theta' = (\theta_1', \theta_2')$, $p_A(\theta) \propto \pi_A^*(\theta_1)p(\theta_2)$ and $p_B(\theta) \propto \pi_B^*(\theta_1)p(\theta_2)$. The models differ only in the prior distribution for the subvector θ_1. The prior densities $\pi_A^*(\theta_1)$ and $\pi_B^*(\theta_1)$ are proper and normalized. The models have a common, independent prior density kernel $p(\theta_2)$ for θ_2. This prior need not be proper and if proper $p(\theta_2)$ need not be normalized. Then

$$\mathrm{POR} = \frac{\int_{\Theta_2}\left[\int_{\Theta_1}\pi_A^*(\theta_1)\mathrm{L}(\theta)d\theta_1 \Big/ \int_{\Theta_1}\pi_B^*(\theta_1)\mathrm{L}(\theta)d\theta_1\right]\int_{\Theta_1}\pi_B^*(\theta_1)\mathrm{L}(\theta)\mathrm{p}(\theta_2)d\theta_1 d\theta_2}{\int_{\Theta_2}\int_{\Theta_1}\pi_B^*(\theta_1)\mathrm{L}(\theta)\mathrm{p}(\theta_2)d\theta_1 d\theta_2}. \quad (12)$$

If the expressions $\int_{\Theta_1}\pi_A^*(\theta_1)\mathrm{L}(\theta)d\theta_1$ and $\int_{\Theta_1}\pi_B^*(\theta_1)\mathrm{L}(\theta)d\theta_1$ can be evaluated quickly, then one may approximate the posterior odds ratio by drawing θ from the posterior distribution under Model B, and treating the term in brackets in (12) as the function of interest.

This method applies in the limit as $\pi_A^*(\theta_1)$ places all its mass on a single point θ_1^*. The numerator in the term in brackets in (12) then becomes $\mathrm{L}(\theta_1^*, \theta_2)$. This situation is precisely the one that occurs here. Model A is the model of Section 7.1 with $\rho = 1$ imposed; model B is the model of Section 7.1 with the prior $\pi_B^*(\cdot)$ indicated by (11); θ_1 consists of ρ; and θ_2 consists of all the other parameters. Geweke (1994) shows that the bracketed term in (12) is

$$\frac{\exp\left[-(1-\hat{\rho})^2/2\lambda^2\right]}{(s+1)\int_0^1 \rho^s \exp\left[-(\rho-\hat{\rho})^2/2\lambda^2\right]d\rho}, \quad (13)$$

where $\hat{\rho}$ and λ are functions of the data and θ_2.

Suppose now that we examine robustness of the posterior odds ratio to changes in $p(\theta_2)$ within a density ratio class $p(\theta_2)$: $a(\theta_2) \le p(\theta_2) \le b(\theta_2)$. The bracketed term in (12) is invariant with respect to changes in $p(\theta_2)$. The methods of Sections 1 - 3 therefore apply to this situation, taking as the function of interest $\int_{\Theta_1}\pi_A^*(\theta_1)\mathrm{L}(\theta)d\theta_1 \Big/ \int_{\Theta_1}\pi_B^*(\theta_1)\mathrm{L}(\theta)d\theta_1$. This expression for our H_A and H_B is (13). The methods of Section 4, however, do not apply because the function of interest changes with sample size.

We examine the sensitivity of the posterior odds ratio in favor of H_A over the density ratio class $p(\theta_2)$: $\tilde{p}(\theta_2) \le p(\theta_2) \le k\tilde{p}(\theta_2)$ where $\tilde{p}(\theta_2)$ is the prior density kernel for all the parameters except ρ described in Section 7.1. The results of this analysis are presented in Figure 12, in the solid curve. This figure is set up similar to Figures 9-11, vertically centered at the posterior odds ratio of .53 computed with the original prior, and scaled according to the posterior standard deviation of (13). The posterior odds ratio is quite sensitive to priors within the density ratio class. When $k = 3$ the posterior odds ratio can exceed 1 or drop below .30; when $k = 30$, it can exceed 3 or fall below 0.05. Clearly the limits are not at all well described by the function $\gamma(k)$ of Section 4. Moreover, the assumptions of Proposition 5 are not satisfied here. (The function of interest changes with sample size.) Whether there is a variant of Proposition 5 that would apply to the posterior odds ratio is an interesting and open question.

Figure 13 provides bounds for the log posterior odds ratio. The shape of the bounds here is closer to the one suggested by $\gamma(k)$ but the resemblance is still not close. Moreover, there is no formal reason to expect this to be a useful approximation.

8. Conclusion

The density ratio class provides a range of prior distributions that reflects uncertainty about the prior distribution of the kind that is likely to arise from uncertainty in elicitation. It is a very convenient class to use. Given the output of any posterior simulator, one can form upper and lower bounds for a posterior moment. The computations involved are very fast and are not specific to the problem at hand, i.e., the same algorithm and code is applicable for all problems.

The examples taken up here are nontrivial applications that cannot be approached by analytical methods. It proved straightforward to obtain upper and lower bounds over several posterior moments for many density ratio classes. These bounds often differed substantially from their asymptotic (in sample size) values. In particular, the more skewed the posterior probability density the less accurate was the asymptotic approximation. It is likely that this feature will be found in most nontrivial applications. A modest extension of these methods makes it possible to examine the robustness of posterior odds ratios within the density ratio class, but in this extension there is no applicable asymptotic approximation. The posterior odds ratio studied in one of our examples proved sensitive to variations in the prior, These observations, combined with the ease of obtaining good numerical approximations to exact results, suggest that investigators should work with the exact results and not asymptotic approximations.

A little additional work substantially broadens the applicability of the method. Suppose that a client has access to the output of the investigator's posterior simulator, $\left\{\theta_m, \mathrm{w}(\theta_m)\right\}_{m=1}^{\infty}$ for the posterior density kernel $\mathrm{L}(\theta)\tilde{\mathrm{p}}(\theta)$, and the corresponding evaluations of the function of interest, $\left\{\mathrm{g}(\theta_m)\right\}_{m=1}^{M}$. The client's density ratio class of prior distributions is $\mathrm{p}(\theta)\colon \mathrm{a}(\theta) \le \mathrm{p}(\theta) \le \mathrm{b}(\theta)$. The client may proceed exactly as indicated in Section 3, so long as the assumptions of Proposition 4 are satisfied. The practical limitations of this procedure are currently being explored (Geweke, 1995). This analysis may be undertaken for any function of interest $\mathrm{g}(\theta_m)$ that the client can compute, and need not be limited to those whose posterior expectation has been reported by the investigator.

References

Abowd, J., and A. Zellner, 1985, "Estimating Gross Labor-Force Flows," *Journal of Business and Economic Statistics* **3**: 254-283.

Bartholomew, J., 1973, *Stochastic Models for Social Processes*. London: Wiley.

Berger, J.O., 1994, "An Overview of Robust Bayesian Analysis" (with discussion), *Test* **3**: 5-124.

Bernardo, J. M., and A. F. M. Smith, 1994, *Bayesian Theory*. Chichester: Wiley.

Chen, C.F., 1985, "On Asymptotic Normality of Limiting Density Functions with Bayesian Implications," *Journal of the Royal Statistical Society* Series B **47**: 540-546.

Chari, V.V., P. Kehoe, and E. McGrattan, 1995, "The Poverty of Nations: A Quantitative Exploration," Federal Reserve Bank of Minneapolis working paper.

DeJong, D.N., and C.H. Whiteman, 1991, "Reconsidering 'Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics* **28**: 221-254.

DeRobertis, L., and J. A. Hartigan, 1981, "Bayesian Inference Using Intervals of Measures," *The Annals of Statistics* **9:** 235-244.

Doan, T., R. Litterman, and C.A. Sims, 1984, "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews* **3**: 1-56.

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., 1989, "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica* **57**: 1317-1340.

Geweke, J., 1992, "Priors for Macroeconomic Time Series and Their Application," Institute for Empirical Macroeconomics Discussion Paper No. 62. Minneapolis: Federal Reserve Bank of Minneapolis.

Geweke, J., 1993, "Bayesian Treatment of the Independent Student-t Linear Model," *Journal of Applied Econometrics* **8**: S19-S40.

Geweke, J., 1994, "Priors for Macroeconomic Time Series and Their Application," *Econometric Theory* **10**: 609-632.

Geweke, J., 1995, "Bayesian Communication," working paper in preparation.

Geweke, J., R.C. Marshall, and G.A. Zarkin, 1986, "Exact Inference for Continuous Time Markov Chain Models," *Review of Economic Studies* **53**: 653-699.

Hartigan, J.A., 1983, *Bayes Theory*. Berlin: Springer.

Heyde, C.C., and I.M. Johnstone, 1979, "On Asymptotic Posterior Normality for Stochastic Processes," *Journal of the Royal Statistical Society* Series B **41**: 184-189.

Johnson, N.L., and S. Kotz, 1972, *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.

Lavine, M., 1991a, "Sensitivity in Bayesian Statistics: The Prior and the Likelihood," *Journal of the American Statistical Association* **86**: 396-399.

Lavine, M., 1991b, "An Approach to Robust Bayesian Analysis for Multidimensional Parameter Spaces," *Journal of the American Statistical Association* **86**: 400-403.

Nelson, C.R., and C.I. Plosser, 1982, "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics* **10**: 139-162.

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

Roberts, G.O., and A.F.M. Smith, 1994, "Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithms," *Stochastic Processes and Their Applications* **49**: 207-216.

Schotman, P.C. and H.K. van Dijk, 1991a, "A Bayesian Analysis of the Unit Root in Real Exchange Rates," *Journal of Econometrics* **49**: 195-238.

Schotman, P.C. and H.K. van Dijk, 1991b, "On Bayesian Routes to Unit Roots," *Journal of Applied Econometrics* **6**: 387-401.

Schotman, P.C., and H.K. van Dijk, 1992, "Posterior Analysis of Possibly Integrated Time Series with an Application to Real GNP," in D. Brillinger et al. (eds.), *New Directions in Time Series Analysis, Part II*, 341-362. Berlin: Springer-Verlag.

Shorrocks, A.F., 1978, "The Measurement of Mobility," *Econometrica* **46**: 1013-1024.

Sims, C.A., 1988, "Bayesian Skepticism on Unit Root Econometrics," *Journal of Economic Dynamics and Control* **12**: 436-474.

Sims, C..A., and H. Uhlig, 1991, "Understanding Unit Rooters: A Helicopter Tour," *Econometrica* **59**: 1591-1599.

Singer, B., and J. Cohen, 1980, "Estimating Malaria Incidence and Recovery Rates from Panel Surveys," *Mathematical Biosciences* **49**: 273-305.

Summers, R., and A. Heston, 1991, "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-88," *Quarterly Journal of Economics* **106**: 327-368.

Tierney, L., 1994, "Markov Chains for Exploring Posterior Distributions" (with discussion and rejoinder), *Annals of Statistics* **22**: 1701-1762. (Also Technical Report No. 560, University of Minnesota School of Statistics.)

Wasserman, L., 1992, "Recent Methodological Advances in Robust Bayesian Inference," in J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds.), *Bayesian Statistics 4*, 483-502. Oxford: Oxford University Press.

Wasserman, L., and J. B. Kadane, 1992, "Computing Bounds on Expectations," *Journal of the American Statistical Association* **87**: 516-522.

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

Zellner, A., and G.C. Tiao, 1964, "Bayesian Analysis of the Regression Model with Autocorrelated Errors," *Journal of the American Statistical Association* **59**: 763-778.

Table 1

Data for Markov chain example, n_{ij}

j:	1	2	3	4	5	6
i:						
1	2	1	0	0	0	0
2	3	12	1	1	0	0
3	0	8	15	3	0	0
4	0	0	3	15	13	2
5	0	0	1	4	14	3
6	0	0	0	0	2	22

Figure 1: Some hypothetical E (f) functions

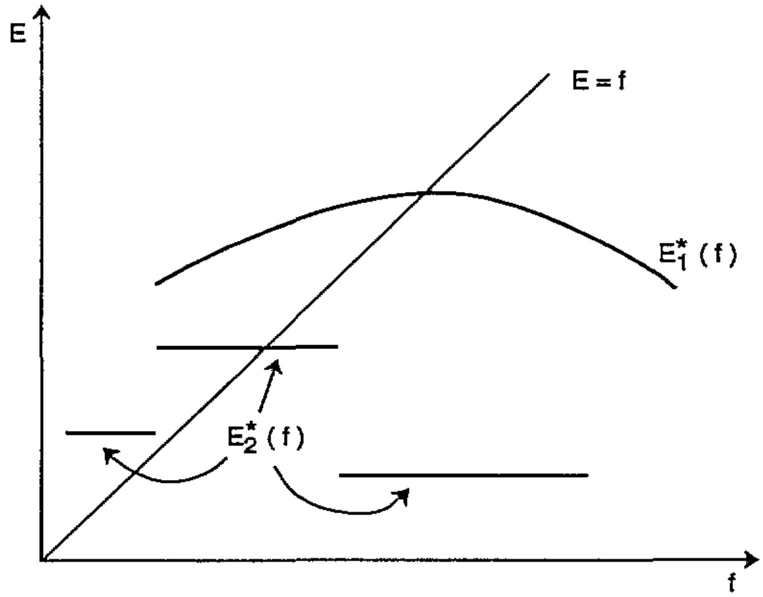


Figure 2: An impossible configuration of g_ℓ and Q_ℓ

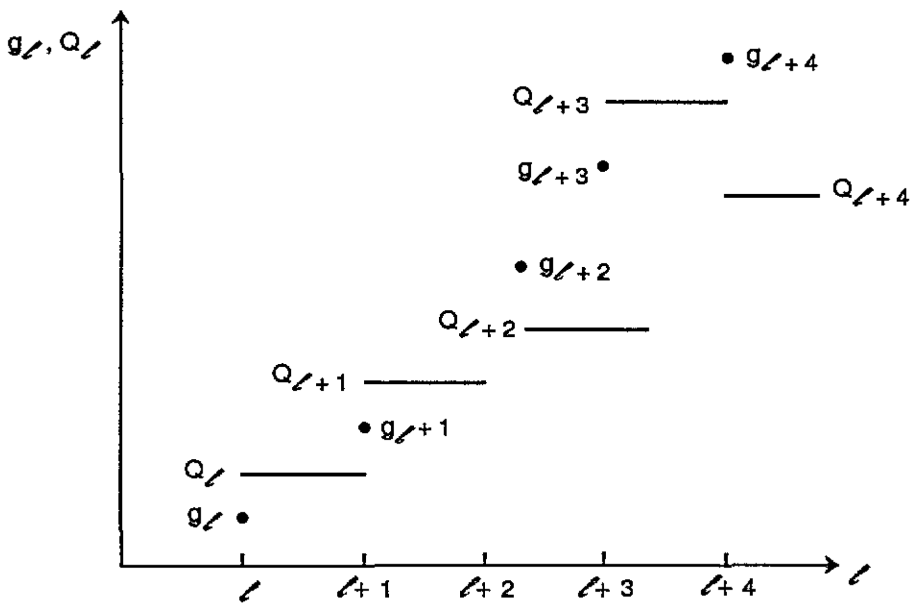


Figure 3: $\gamma(k)$

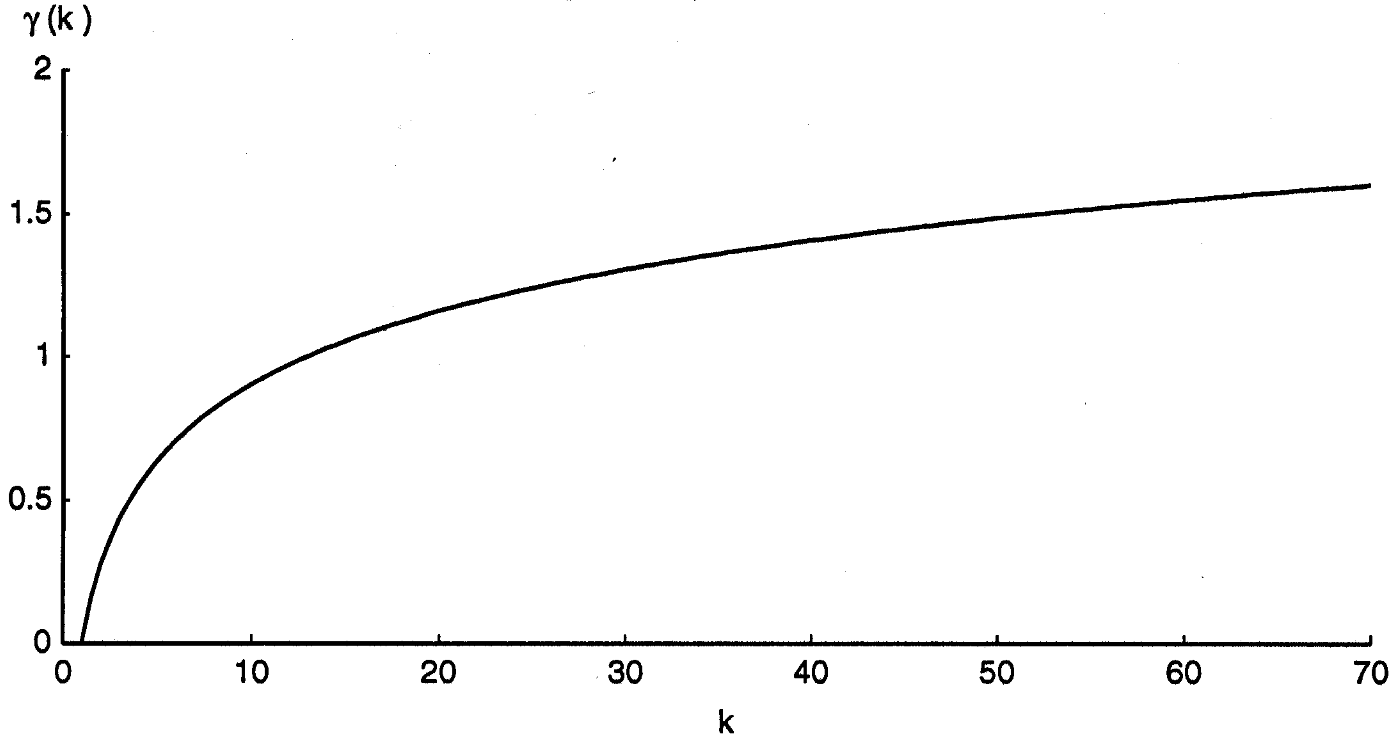


Figure 4: $\tilde{\gamma}(k;\nu)$

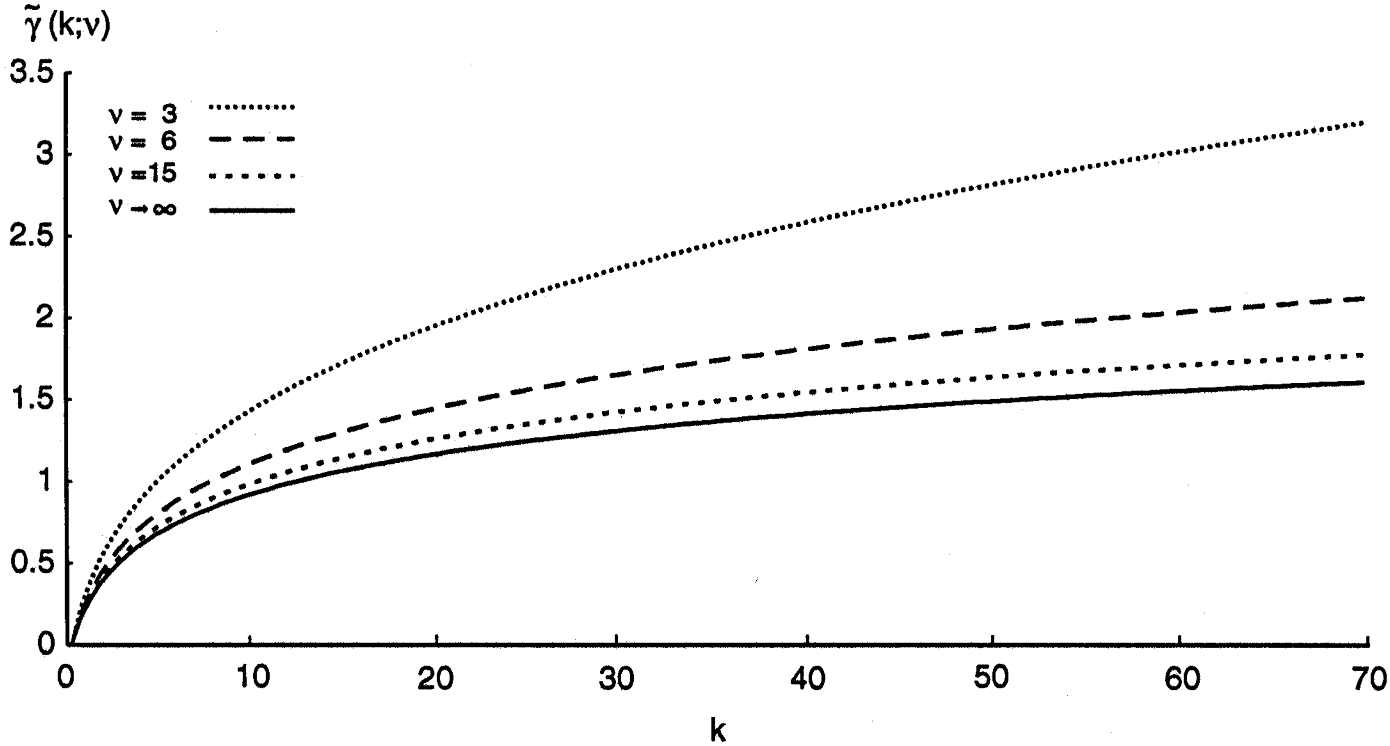
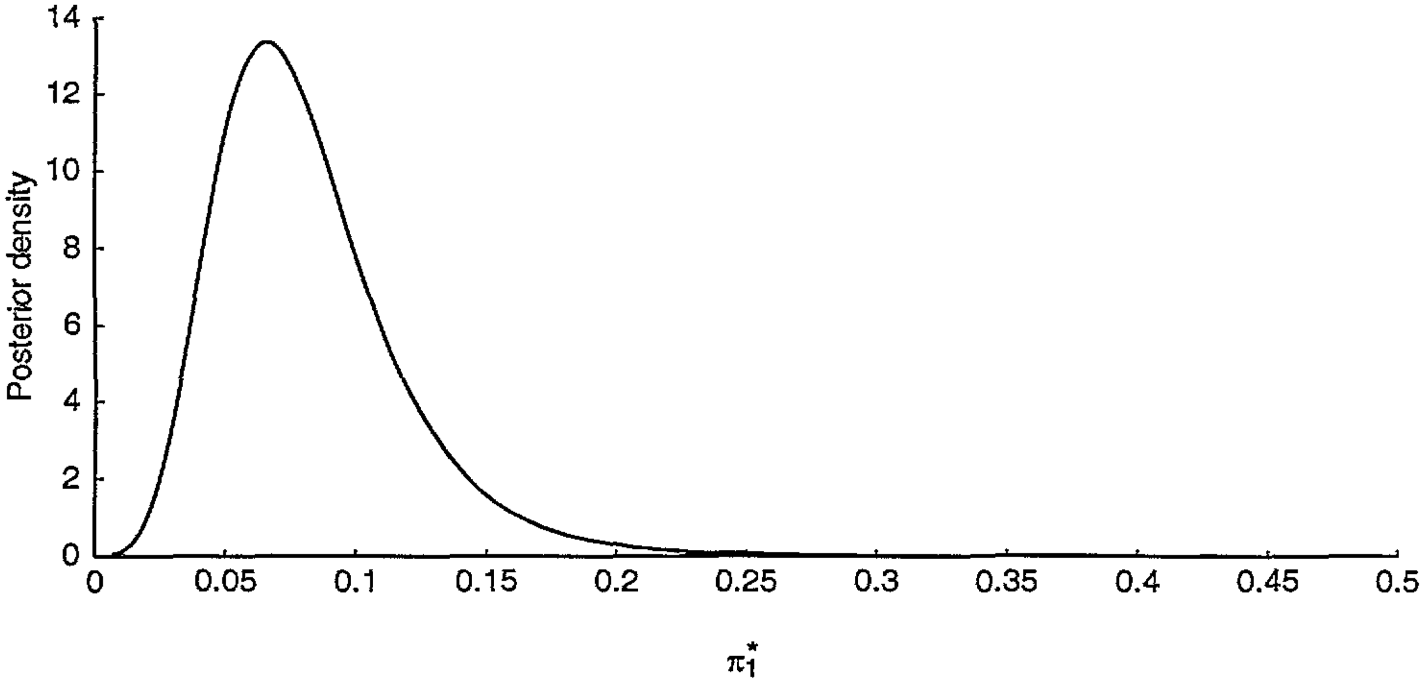


Figure 5: Some posterior densities in the first order Markov chain model



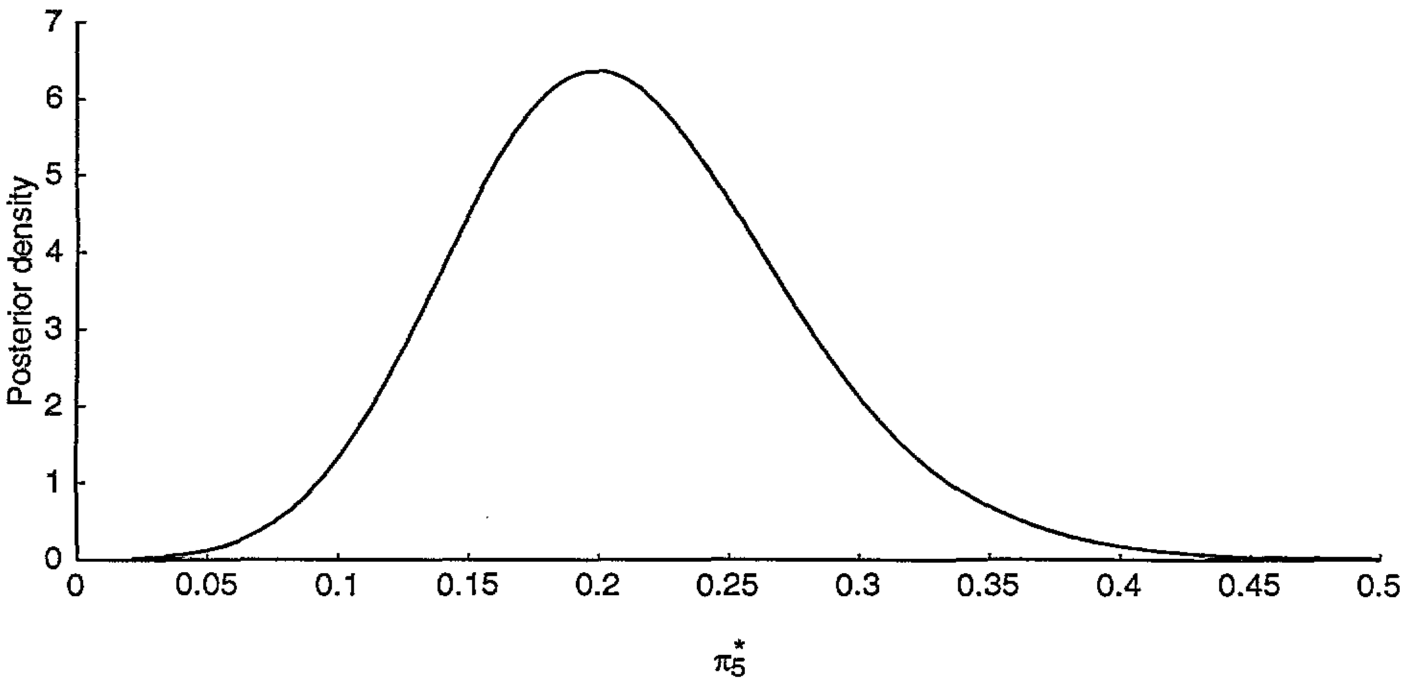


Figure 6: Upper and lower bounds for $E[\pi_1^* \ln_{ij} (i,j = 1,...,6)]$

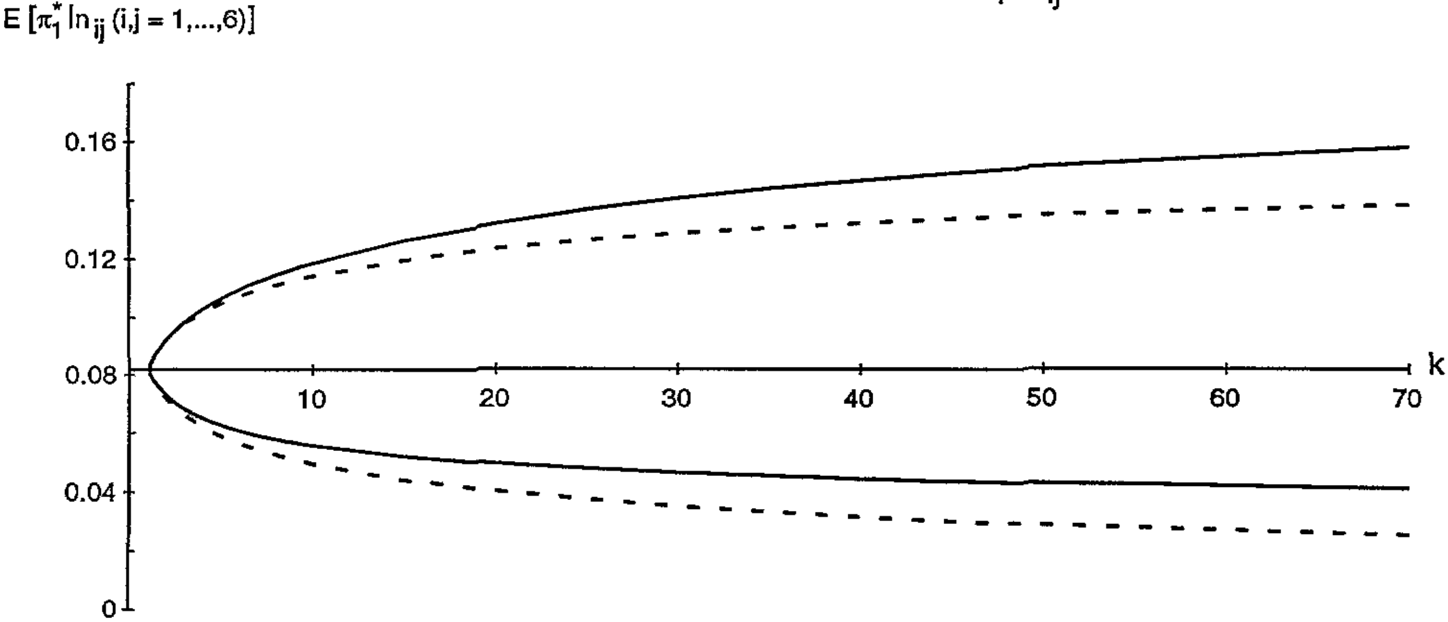


Figure 7: Upper and lower bounds for $E[\pi_5^* \ln_{ij} (i,j = 1,...,6)]$

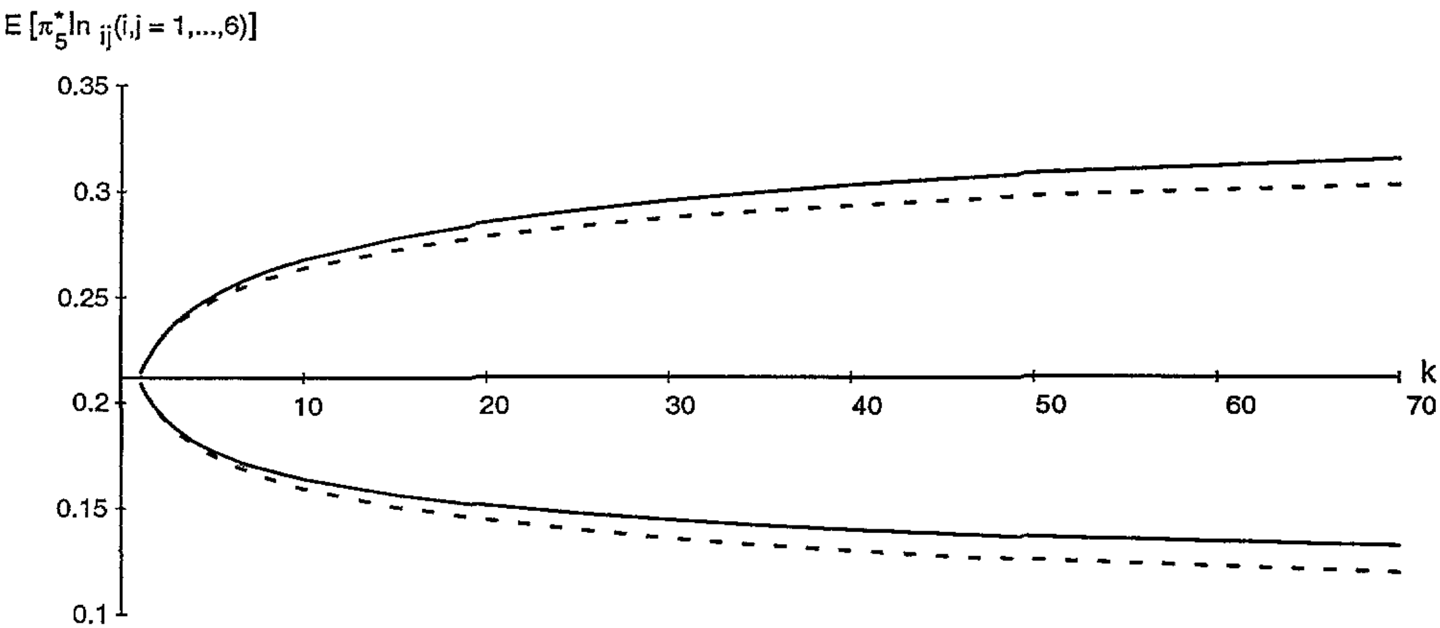


Figure 8: Some posterior densities in the trend stationary model

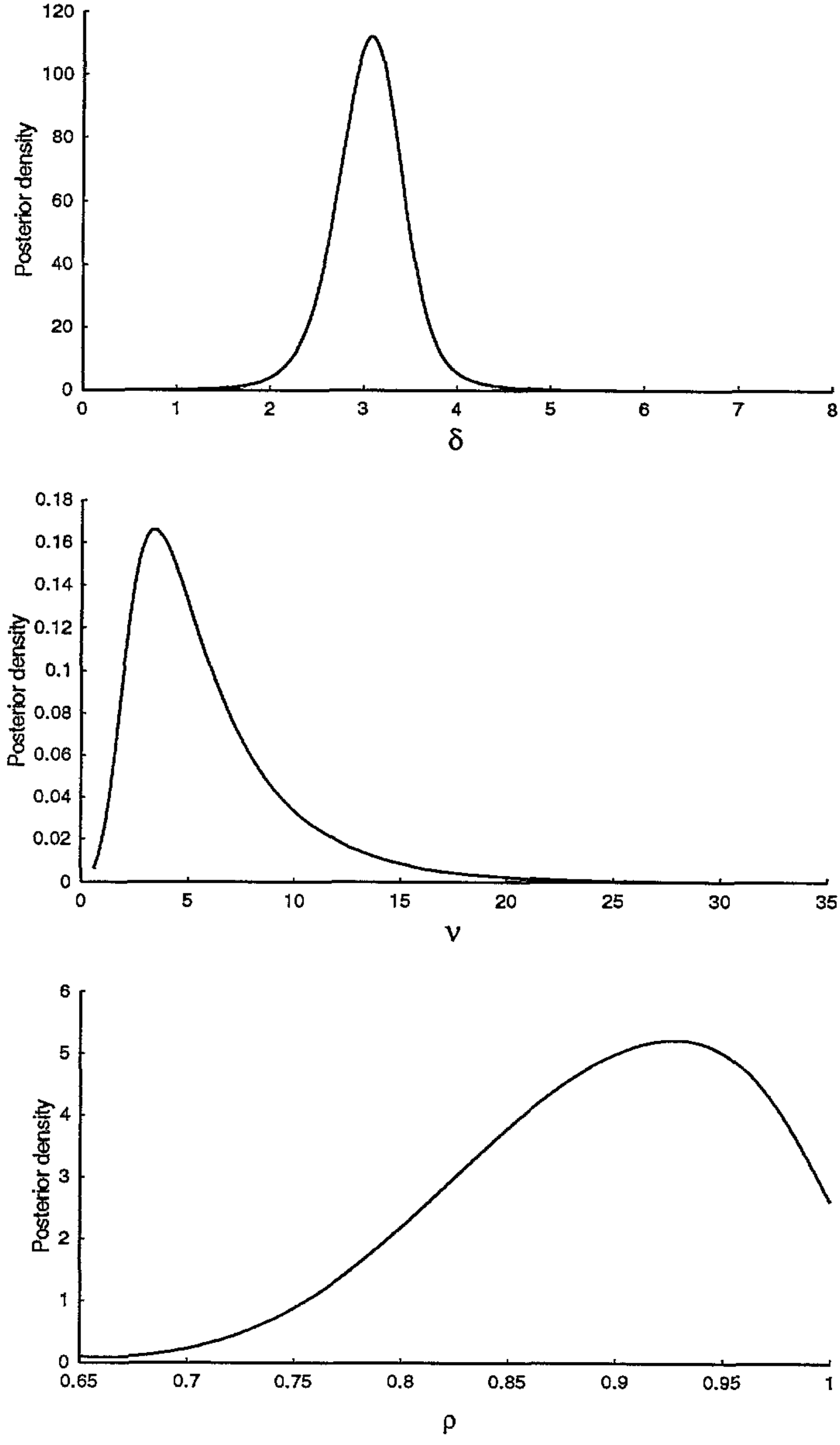


Figure 9: Upper and lower bounds for E (δ | y)

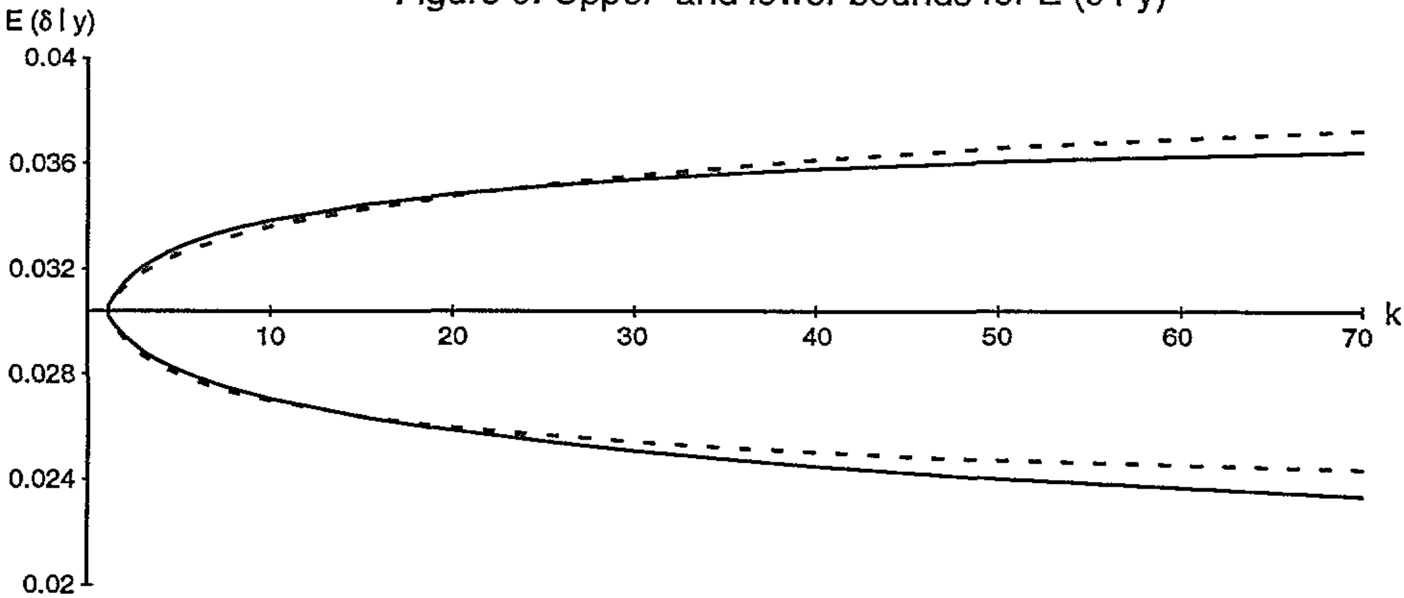


Figure 10: Upper and lower bounds for E (v | y)

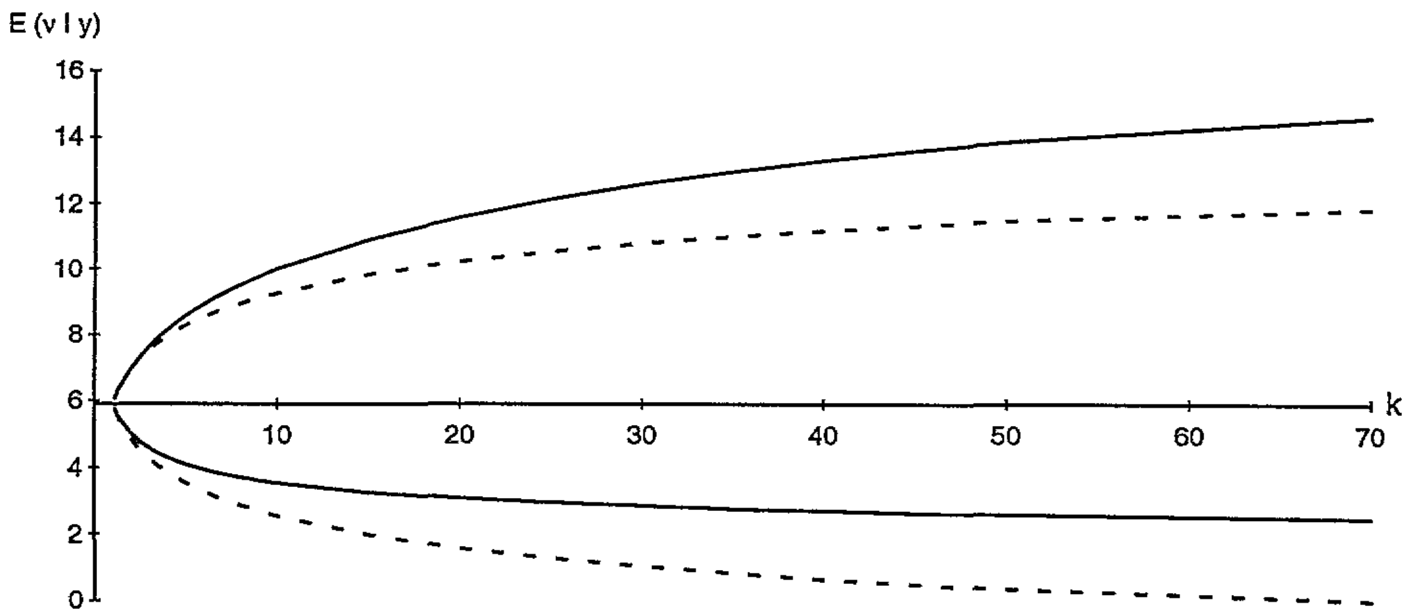


Figure 11: Upper and lower bounds for E (ρ l y)

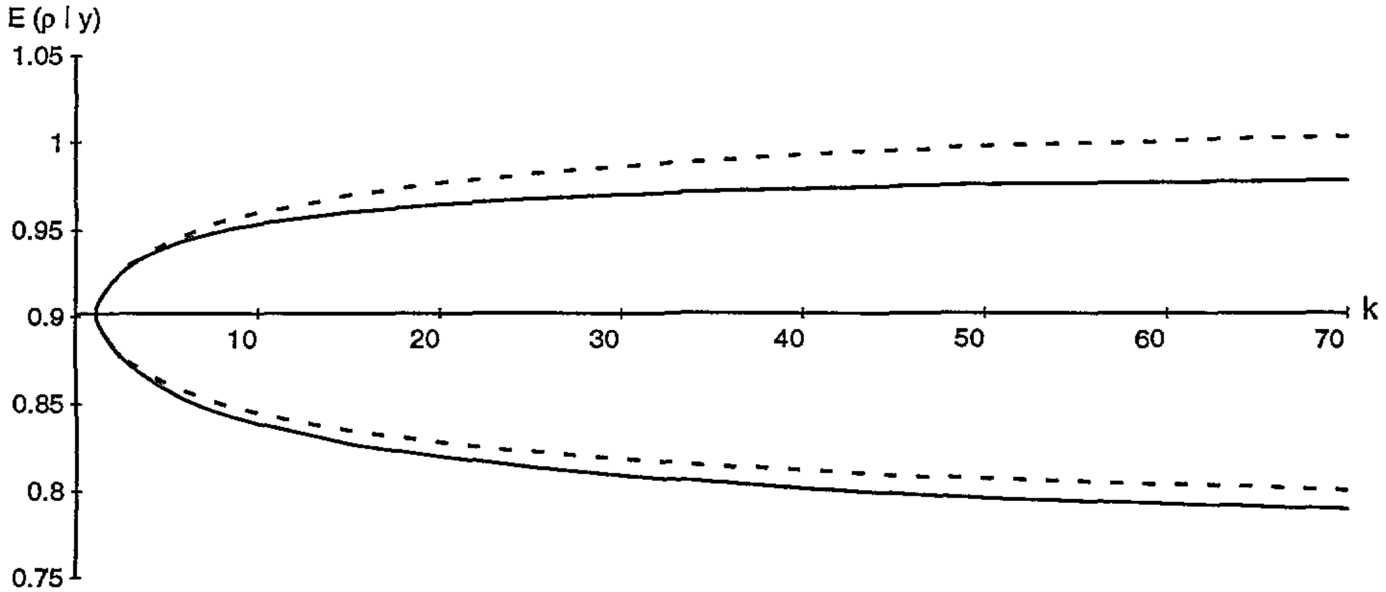


Figure 12: Upper and lower bounds for posterior odds ratio

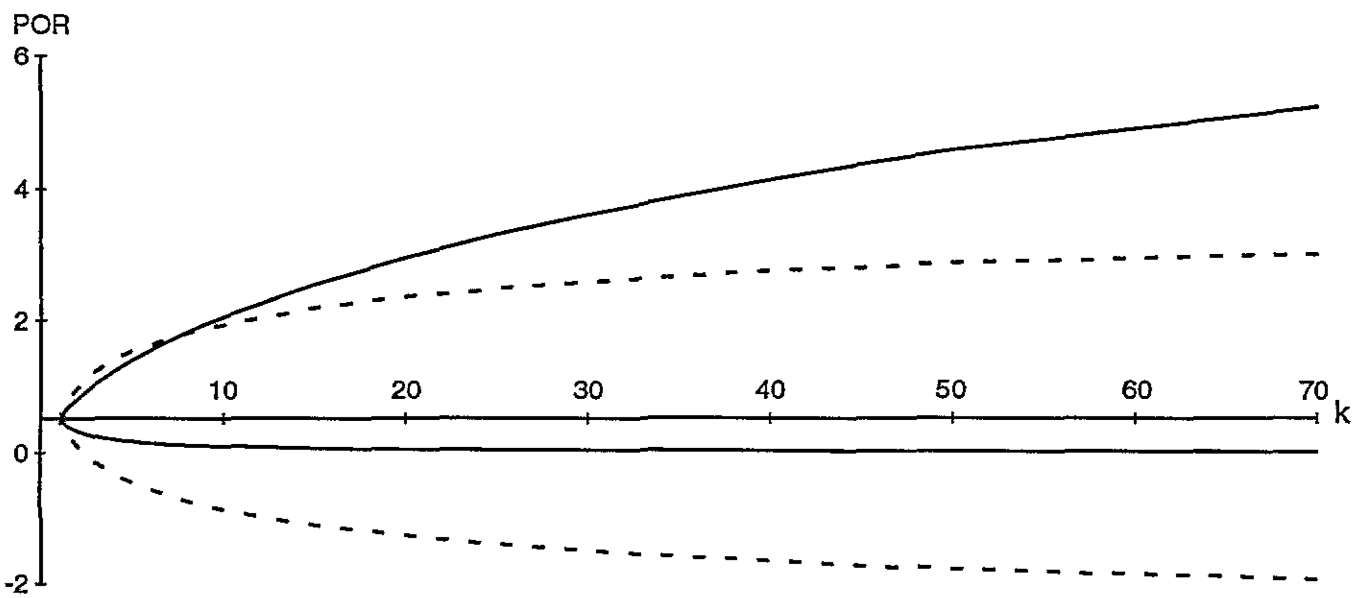


Figure 13: Upper and lower bounds for log posterior odds ratio