NONINFORMATIVE PRIORS AND BAYESIAN TESTING FOR THE AR(1) MODEL

by

James O. Berger and Ruo-yong Yang
Purdue University

Technical Report #92-45C

Department of Statistics
Purdue University

October 1992

Noninformative Priors and Bayesian Testing for the AR(1) Model *

James O. Berger
Department of Statistics
Purdue University

Ruoyong Yang
Department of Statistics
Purdue University

June, 1992

Abstract

Various approaches to development of a noninformative prior for the AR(1) model are considered and compared. Particular attention is given to the *reference prior* approach, which seems to work well for the stationary case but encounters difficulties in the explosive case. A symmetrized (proper) version of the stationary reference prior is ultimately recommended for the problem.

Bayesian testing of the unit root, stationary, and explosive hypotheses is also considered. Bounds on the Bayes factors are developed, and shown to yield answers that appear to conflict with classical tests.

Keywords: Jeffreys prior, reference prior, Bayes factors, unit root, mean-squared error.

*This research was supported in part by NSF grant DMS-8923071 at Purdue University.

1 Introduction

The AR(1) model, in which the data $\underset{\sim}{X} = (X_1, ..., X_T)$ follow the model

$$X_t = \rho X_{t-1} + \epsilon_t \,, \tag{1}$$

where the ϵ_t are i.i.d. $\mathcal{N}(0, \sigma^2)$, is surprisingly challenging to "objective" Bayesians, even in the comparatively simple case of known σ^2 (which is the case studied here). Phillips [21] and the discussants therein highlight the issues and controversies in developing a noninformative prior for this model. A brief review of some of these issues is given in Section 2.1.

A recent approach to development of noninformative priors, the *reference prior* approach, has had considerable success in developing noninformative priors for "difficult" problems (cf., Berger and Bernardo [7]). In Section 2.2, we apply this approach to the AR(1) model. Somewhat surprisingly, the AR(1) model proves resistant to a definitive reference prior analysis in the explosive case. In Section 2.3, therefore, a noninformative prior is proposed that is based on a symmetrization of the reference prior for the stationary case.

Section 3 compares the various noninformative priors developed in Section 2, in terms of both mean squared error of the resulting Bayes estimators and frequentist coverage of Bayesian credible sets. The "constant" prior and the new prior perform quite well.

In Section 4, Bayesian testing in the AR(1) model is considered, with special emphasis placed on testing for unit roots. The new noninformative prior developed in Section 2.3 proves to be particularly suitable for this testing. Lower bounds on the Bayes factors for the unit root hypothesis are also presented, bounds which are substantially larger than corresponding classical P-values. The necessity for approaching the unit root problem from a Bayesian perspective is thus reinforced.

2 Possible Noninformative Priors

2.1 Background

Use of noninformative priors has an extensive tradition in statistics, starting with Bayes [2] and Laplace [16] [17] who used the "uniform" prior

$$\pi_U(\theta) = 1 \,. \tag{2}$$

In developing Bayesian methodology, use of π_U is generally very successful, although there are concerns about its lack of invariance to transformation (since one cannot, for instance, be simultaneously "uniform" in θ and $\eta = \log(\theta)$). Also, a number of counterexamples to its use have been encountered (see, e.g., Monette, Fraser, and Ng [20] and Ye and Berger[28]).

Jeffreys [15] sought to overcome the lack of invariance of π_U through development of the now famous Jeffreys prior

$$\pi_J(\theta) = \sqrt{\det(I(\theta))} \,, \tag{3}$$

where $I(\theta)$ is the Fisher information matrix with (i, j) entry

$$I_{ij}(\theta) = -E_\theta[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log f(X \mid \theta)] \,, \tag{4}$$

where E_θ stands for expectation over X, given θ. Not only is this method of deriving a noninformative prior invariant to reparameterization of the problem, but it seems to correct a number of the counterexamples to use of $\pi_U(\theta) = 1$, especially those arising from nonintegrability of the posterior distribution, $\pi_U(\theta \,|\text{data})$. For the AR(1) model, with $\theta = (\rho, \sigma)$,

$$\pi_J(\theta) = \sqrt{\det(I(\theta))} = \left[\frac{2T}{\sigma^2}(\frac{T}{1-\rho^2} + \frac{1-\rho^{2T}}{1-\rho^2}\{E[\frac{X_0^2}{\sigma^2}] - \frac{1}{1-\rho^2}\})\right]^{1/2} \tag{5}$$

(cf., Jeffreys [15], Zellner [29], and Box and Jenkins [13], for the $|\rho| < 1$ case, and Phillips [21] for the general case). At $\rho = 1$, $\pi_J(\theta)$ is defined by continuity (see Phillips [21]).

There are a number of interesting features of $\pi_J(\theta)$. One is that it depends on $E[X_0^2/\sigma^2]$, where X_0 is the initialization of the AR(1) process. Often it is simply assumed that $X_0 = 0$, but more reasonable would be placing a prior distribution on X_0, in which case $E[X_0^2/\sigma^2]$ would refer to expectation with respect to this prior distribution. Thus even Jeffreys prior depends on the possibly subjective prior distribution on X_0.

Another striking feature of $\pi_J(\theta)$ is that, for large $| \rho |$ and $X_0 = 0$,

$$\pi_J(\theta) \approx \frac{\sqrt{2}}{\sigma}\rho^{(T-2)} , \tag{6}$$

which grows exponentially fast if $T > 2$. For $| \rho |< 1$, $\pi_J(\theta)$ is an integrable function of ρ, so π_J gives enormously greater prior mass to the explosive case ($| \rho |> 1$) than to the stationary case ($| \rho |< 1$). Many econometricians feel that this is unreasonable. Phillips [21] and the ensuing discussants vigorously discuss this issue.

2.2 The Reference Prior Approach

2.2.1 Motivation

Bernardo [12] initiated an information based approach to development of noninformative priors, called the *reference prior* approach. A review and discussion of the current status of the approach can be found in Berger and Bernardo [7].

The motivation for developing the approach was the acknowledged problems of the Jeffreys prior in higher dimensions. Even Jeffreys would often alter $\pi_J(\theta)$ in multiparameter problems to remove perceived inadequacies. The reference prior approach sought to overcome these difficulties by breaking up multiparameter problems into a series of conditional one-parameter problems, for which reasonable noninformative priors could be determined. The approach has proven to be remarkably successful in overcoming the inadequacies of Jeffreys prior in multiparameter problems.

Unfortunately, the motivation for the reference prior method was primarily based on i.i.d. asymptotics. An attempt to generalize this to the dependent-data AR(1) model met with only partial success: the reference prior exists for the stationary case ($| \rho |< 1$), but not for the explosive case ($| \rho |> 1$). Thus a symmetrized version of the stationary case reference is ultimately recommended.

2.2.2 The Reference Prior Algorithm

The reference prior algorithm consists of four components: (i)information maximization; (ii) maximizing asymptotic missing information; (iii)finding limits of reference priors on compact sets and (iv) dealing with multiparameter problems by conditional decompositions. Although step (iv) is the original motivation for the algorithm, we focus here on the difficulties of steps(i) through(iii). Indeed in our applications to the AR(1) model we will avoid the multiparameter issue entirely by considering only the case $\sigma^2 = 1$. Undoubtedly, the only change in the reference prior analysis that would result from having σ^2 unknown, would be to introduce a multiplicative factor of $1/\sigma$ in the prior. For the rest of this section, therefore, θ will be assumed to be a real valued parameter.

(i) *Information Maximization*

Denote the Kullback-Liebler divergence between two density functions $f(\theta)$ and $g(\theta)$ by:

$$D(f(\theta), g(\theta)) = \int f(\theta) \log \frac{f(\theta)}{g(\theta)} d\theta \, . \tag{7}$$

Following Lindley [19], the expected information about θ in the data X, when the prior is $\pi(\theta)$, is defined to be the expected Kullback-Liebler divergence between the prior $\pi(\theta)$ and the posterior $\pi(\theta \mid X)$, where the expectation is taken w.r.t. the marginal distribution of X, to be denoted by m(x):

$$I_X^\theta(\theta) = E^X D(\pi(\theta \mid x), \pi(\theta)) = \int m(x) D(\pi(\theta \mid x), \pi(\theta)) dx \, , \tag{8}$$

where

$$m(x) = \int f(x \mid \theta) \pi(\theta) d\theta \, . \tag{9}$$

The idea behind the reference prior algorithm is to choose a prior π to maximize $I_X^\theta(\pi)$ (i.e., to maximize the information provided by the data, relative to π). By variational arguments, this maximizing π satisfies the (implicit) equation:

$$\pi(\theta) \propto \exp\{\int f(x \mid \theta) \log \pi(\theta \mid x) dx\} \, . \tag{10}$$

Solution of this implicit equation is, in general, difficult; indeed, the solution is typically a discrete prior π (see Berger, Bernardo, and Mendoza [8]). But interesting approximate solutions can be obtained in the "large sample" case, as in the following example.

Example. AR(1) Model, assuming σ known.

Define $A = \sum_{i=1}^{T} X_{i-1}^2$, χ_1^2 to be the chi-square distribution with 1 degree of freedom, and denote the MLE estimate of ρ by $\hat{\rho}_{mle}$,

$$\hat{\rho}_{mle} = \sum_{i=1}^{T} X_{i-1} X_i \Big/ \sum_{i=1}^{T} X_{i-1}^2 \,. \tag{11}$$

Since, for large T, $(\sum_{i=1}^{T} X_{i-1}^2)(\hat{\rho}_{mle} - \rho)^2 \sim \chi_1^2$ (see Anderson [1] and White [27]), $\hat{\rho}_{mle} \stackrel{a.s.}{\to} \rho$ (see Rubin [22]), and $A \stackrel{a.s.}{\to} \infty$, the variational equation for large T heuristically becomes:

$$\begin{aligned}
\pi(\rho) &\propto \exp\{\int f(\underset{\sim}{x} \mid \rho) \log \pi(\rho \mid \underset{\sim}{x}) d\underset{\sim}{x}\} \\
&= \exp\{\int f(\underset{\sim}{x} \mid \rho) \log \left[\frac{\pi(\rho) e^{-\frac{A}{2}(\rho - \hat{\rho}_{mle})^2}}{\int \pi(\rho) e^{-\frac{A}{2}(\rho - \hat{\rho}_{mle})^2} d\rho} \right] d\underset{\sim}{x}\} \\
&\sim \exp\{\int f(\underset{\sim}{x} \mid \rho) \log \left[\frac{\pi(\rho) e^{-\frac{A}{2}(\rho - \hat{\rho}_{mle})^2}}{\pi(\hat{\rho}_{mle}) \sqrt{2\pi/A}} \right] d\underset{\sim}{x}\} \\
&\sim \exp\{\int f(\underset{\sim}{x} \mid \rho) [\frac{1}{2} \log(\frac{\sum X_{i-1}^2}{2\pi}) - \frac{1}{2} (\sum X_{i-1}^2)(\hat{\rho}_{mle} - \rho)^2] d\underset{\sim}{x}\} \\
&\propto \exp\{\frac{1}{2} E_\rho [\log \sum_{i=1}^{T} X_{i-1}^2]\} \equiv \pi_R(\rho) \,,
\end{aligned} \tag{12}$$

where "E_ρ" denotes expectation over $(X_1, ..., X_T)$, given ρ. This will be called the "Nonasymptotic reference" prior (though it's justification is a "large T" justification). Computation and hence use of $\pi_R(\rho)$ requires simulation (see Section 3).

(ii)Maximizing Asymptotic Missing Information

Because of the difficulties with exact solution to (10), it is natural to attempt to formalize the asymptotic idea implicit in the above example. There are, however, two possible types of asymptotics. In the usual reference prior approach, one considers n (imaginary) replications (i.i.d.) of $\underset{\sim}{X} : Z_n = (\underset{\sim}{X}^{(1)}, ..., \underset{\sim}{X}^{(n)})$, where $\underset{\sim}{X}^{(i)} = (X_1^{(i)}, ..., X_T^{(i)})$. Define:

$$I_{Z_n}^{\theta}(\pi) = E^{Z_n} D(\pi(\theta \mid Z_n), \pi(\theta)) \,. \tag{13}$$

As $n \to \infty$, $\pi(\theta \mid Z_n)$ converges to a point mass at the true value of θ, so the π that maximizes $I_{Z_n}^{\theta}(\pi)$ can be said to asymptotically maximize the missing information about θ. If $\pi(\theta \mid Z_n)$ is asymptotically normal, then it can be argued, as in the above example, that

$$\pi_n(\theta) \propto \exp\{\int f(Z_n \mid \theta) \log \pi_n(\theta \mid Z_n) dZ_n\} \overset{(n\to\infty)}{\longrightarrow} \sqrt{\det I(\theta)} \tag{14}$$

(after renormalization - see Berger and Bernardo [7]). In the AR(1) model, if ρ is restricted to a compact set, the above argument can be carried out yielding the Jeffreys prior (restricted to the compact set) as the reference prior.

The above asymptotic information argument was designed for the i.i.d. case, where independent replication of the error structure seems reasonable. In the AR(1) model, however, this corresponds to considering a large number of independent realizations of the series.

A move natural asymptotic information argument for the AR(1) model would be to let $T \to \infty$. Earlier heuristics suggested that the appropriate limit of (13) would define this asymptotic reference prior. Because of impropriety, it is typical to actually operate on a compact set C, so we define the *(asymptotic) reference prior on C* as

$$\pi_C(\rho) = \lim_{T\to\infty} \frac{\exp\{\frac{1}{2}E_\rho[\log \sum_{i=1}^{T} X_{i-1}^2]\}}{\int_C \exp\{\frac{1}{2}E_\rho[\log \sum_{i=1}^{T} X_{i-1}^2]\}d\rho} \, . \tag{15}$$

Lemma 1 **(i)** *In the stationary AR(1) case, i.e.,* $C = [-1, 1]$, $\pi_C(\rho) \propto (1-\rho^2)^{-1/2}$.

(ii) *If C=[a, b], where* $a < -1$ *or* $b > 1$, $\pi_C(\rho)$ *is a discrete prior giving mass only to the end points of* C.

Proof. See Appendix.

(iii) *Limits of Reference Priors on Compact Sets*

The asymptotic reference prior can typically only be defined on compact sets Θ_i, at which point it is necessary to pass to a limit as $\Theta_i \to \Theta$ to obtain a global reference prior. Typically, a sequence $\Theta_1 \subset \Theta_2 \subset \ldots \to \Theta$ is chosen and the reference priors on Θ_i, $\pi_i(\theta)$, are found. If $\pi^*(\theta)$ is such that:

$$E_i^X D(\pi_i(\theta \mid X), \pi^*(\theta \mid X)) = \int m_i(x) D(\pi_i(\theta \mid x), \pi^*(\theta \mid x)) dx \overset{i\to\infty}{\to} 0 \, , \tag{16}$$

so that $\pi_i \to \pi^*$ in an information sense, then, π^* is defined to be a reference prior on Θ. One may also require $\pi^*(\theta) = \lim_{i\to\infty} \pi_i(\theta)/\pi_i(\theta_0)$ (this is the usual way to compute π^*). Typically, the above definition will yield the Jeffreys prior(in one dimension and with i.i.d. asymptotics) with

$$\pi_i(\theta) \propto 1_{\Theta_i}(\theta)\sqrt{\det I(\theta)} \stackrel{i\to\infty}{\longrightarrow} \sqrt{\det I(\theta)} = \pi^*(\theta)\ ; \tag{17}$$

here $1_{\Theta_i}(\theta)$ denotes the indicator function on Θ_i.

Initially, we had thought that (16) would be violated for π^* equal to the Jeffreys prior. We now feel, however, that Jeffreys prior will satisfy the condition, although we have not carried out the technical verification.

For the $T \to \infty$ (asymptotic) reference prior, however, it is clear that no π^* can exist for which (16) is satisfied, since Lemma 1 shows that the mass of π_i will be concentrated at the endpoints of Θ_i; this mass will escape to infinity as $i \to \infty$. Hence no sensible asymptotic reference prior can be defined for the explosive case ($|\rho| > 1$).

Conclusion: The reference prior algorithm works in the AR(1) model only for $|\rho| < 1$, and yields as the ($T \to \infty$ asymptotic) reference prior on this set

$$\pi(\rho) = (1-\rho^2)^{-1/2}\ . \tag{18}$$

(If the i.i.d. asymptotic reference algorithm is used, the Jeffreys prior results.)

The nonasymptotic version of the reference prior algorithm suggested consideration of $\pi_R(\rho)$ in (12), but there are two concerns here. First, the nonasymptotic version has typically been viewed as inferior (admittedly without much justification). More importantly, the derivation of $\pi_R(\rho)$ suggests that it might be very similar to $\pi_J(\rho)$; further evidence of this will be seen in Section 3.

For $|\rho| < 1$, the prior (18) had previously been given in Zellner [29] as an approximate Jeffreys prior. Interestingly, Zellner also suggests the inverse of this prior, namely $(1-\rho^2)^{1/2}$, based on his MDIP approach (see also Zellner [30]).

2.3 The Symmetrized Asymptotic Reference Prior

In our attempts to apply the reference prior algorithm, we encountered a very interesting prior that has considerable motivation as a noninformative prior, namely

$$\pi_{SR}(\rho) = \begin{cases} 1/[2\pi\sqrt{1-\rho^2}] & \text{if } \mid \rho \mid < 1 \\ 1/[2\pi \mid \rho \mid \sqrt{\rho^2 - 1}] & \text{if } \mid \rho \mid > 1\,. \end{cases} \tag{19}$$

This prior has several appealing properties. First, it is actually a proper prior, integrating to one. Indeed it gives equal probability (of 1/2) to $\mid \rho \mid < 1$ and $\mid \rho \mid > 1$. Since few econometricians would actually view the explosive case as being more plausible a priori than the stationary case, this is an attractive property in comparison with $\pi_U(\rho) = 1$ or $\pi_J(\rho)$, which assign infinite mass to the explosive case and only finite mass to $\mid \rho \mid < 1$. The propriety of $\pi_{SR}(\rho)$ and its assignment of equal probability to $\mid \rho \mid < 1$ and $\mid \rho \mid > 1$ also make it eminently suitable for use in testing. (Recall that improper priors cannot typically be used for testing.)

A second motivation for $\pi_{SR}(\rho)$ is that it is the reference prior for $\mid \rho \mid < 1$. And it deals with $\mid \rho \mid > 1$ by imposing invariance via the mapping $\rho \to \frac{1}{\rho}$. (Transforming to $\eta = \frac{1}{\rho}$ results in the same prior for η.). This is particularly appealing in light of the fact that $\eta = \frac{1}{\rho}$ can be considered the parameter of the "backwards" process

$$X_{t-1} = \eta X_t + \epsilon^*_{t-1} \tag{20}$$

and $\mid \eta \mid < 1$ corresponds to $\mid \rho \mid > 1$. Unfortunately, we were unable to make this argument more than a heuristic argument.

2.4 Perspective

The "holy grail" of objective Bayesians has been the search for noninformative priors that will be perceived as truly objective representations of ignorance. Many no longer believe in the existence of this grail; indeed, the AR(1) model has been argued to be an example of the failure of objective Bayesian methods, the argument being that the information-based Jeffreys prior is unreasonable. See Phillips [21] for a defense of the Jeffreys prior, and the discussants therein for a variety of views. We have avoided involvement in this type of debate, perhaps because we are subjective

Bayesians at heart and do not believe any noninformative prior can be more than a convenient approximation, or perhaps because we have seen too many counterexamples to specific "objective" criteria to believe true objectivity is possible. We do, however, feel that information-based criteria have by far the best track record in suggesting noninformative priors that turn out to have nice operational properties.

This leads to the second school of thought concerning noninformative priors, the "operationalist" school: a noninformative prior is good if it yields answers that are deemed to be reasonably objective by most statisticians. Often there is even a frequentist component to this validation of the noninformative prior, showing that the resulting Bayes procedures have good frequentist properties.

Unfortunately, not all problems are amenable to an operationalist objective analysis. A common example is testing of a precise hypothesis (or, more generally, testing between two models of differing dimension), where it is impossible to produce answers that will generally be perceived as objective (cf., Berger and Berry [11] and Berger and Delampady [10]). It appears that the AR(1) model will be added to this list of problems.

If an operationally objective answer is not available, Bayesians have three choices. First of course, is performing a subjective Bayesian analysis. If one is working in a situation where this is feasible, great. A second option is to perform a robust Bayesian analysis over a wide range of subjective priors; this is the option discussed in Section 4. The third possibility is to professionally "agree" that a certain prior will be used as the "conventional" or "default" prior. The Jeffreys [15] priors for precise hypothesis testing and the Zellner and Siow [31] priors for model comparison can be considered to be of this type. Because of its attractive properties (see Section 2.3) and performance (see Section 3), the prior $\pi_{SR}(\rho)$ in (19) deserves to be seriously considered as the default prior for the AR(1) model. (If σ is unknown, the suggested default prior is $\sigma^{-1}\pi_{SR}(\rho)$.)

3 Comparison of Noninformative Priors

3.1 The Four Candidate Priors

As candidate noninformative priors, we have encountered the following:

Uniform Prior: $\pi_U(\rho) = 1$;

Jeffreys Prior: $\pi_J(\rho) = \left[\frac{T}{1-\rho^2} + \frac{1-\rho^{2T}}{1-\rho^2}\{E[\frac{X_0^2}{\sigma^2}] - \frac{1}{1-\rho^2}\}\right]^{1/2}$;

Nonasymptotic Reference Prior: $\pi_R(\rho) = \exp\{\frac{1}{2}E[\log(\sum_{i=1}^T X_{i-1}^2)]\}$;

Symmetrized Reference Prior: $\pi_{SR}(\rho) = \begin{cases} 1/[2\pi\sqrt{1-\rho^2}] & \text{if } \mid\rho\mid < 1 \\ 1/[2\pi \mid\rho\mid \sqrt{\rho^2-1}] & \text{if } \mid\rho\mid > 1. \end{cases}$

In the remainder of Section 3, we compare these four priors.

3.2 Frequentist Comparison of Noninformative Priors

There is no unique way to compare noninformative priors, but various frequentist criteria have proved helpful in such evaluations. The basic idea is to use the prior to generate a statistical procedure, and investigate the frequentist properties of the procedure. If the procedure resulting from one prior has substantially better properties than that resulting from another prior, then the latter prior is suspect. Note, however, that one cannot typically expect one prior to completely dominate another according to a given criterion.

The most common frequentist comparison of noninformative priors is via admissibility or risk dominance of resulting estimators (see Berger, [3]). Another common method is to compare confidence properties of sets arising from the posteriors. Here we follow these traditions of comparing noninformative priors, utilizing the following simulation for the AR(1) model with $x_0 = 0$ and $\sigma^2 = 1$:

Simulation: Set T = 5, 10, 20, 50, 100 and ρ = .25, .5, .75, 1.0, 1.25, 1.5. For fixed T and ρ, we do the following:

(i) Generate 1600 groups of observations $\underset{\sim}{X} = (X_1, X_2, ..., X_T)$ from the AR(1) process.

(ii) For each specific observation set $\underset{\sim}{X}$, compute the MLE estimate $\hat{\rho}_{mle}$ (see (11))and calculate the posterior means: $\hat{\rho}_U, \hat{\rho}_J, \hat{\rho}_R, \hat{\rho}_{SR}$ w.r.t the above four candidate priors, via importance sampling. It is easy to see that $\hat{\rho}_U = \hat{\rho}_{mle}$, so that case need not be separately analyzed. The importance sampling used sample sizes of 1000 with a $Cauchy(\hat{\rho}_{mle}, 1/\sqrt{\sum_{i=1}^T x_{i-1}^2})$ importance function, and was accurate to within 4%.

(iii) Use of $\pi_R(\rho)$ requires an additional simulation to compute $\pi_R(\rho)$ itself (see (12)) at each of the thousand ρ_i generated from the above importance function; the $\pi_R(\rho_i)$ are needed to compute the importance sampling weights. The simulation to compute $\pi_R(\rho_i)$ was carried out by the simple device of generating samples from the AR(1) process with the given ρ_i, and then computing the expectation in (12) by simple averaging. This second simulation greatly increased computational cost, however. Since the resulting answers appeared to closely mimic those for the Jeffreys prior (interestingly, they were slightly better for the cases we computed), we dropped consideration of $\pi_R(\rho)$.

(iv) Record $(\hat{\rho}_{mle} - \rho)^2$ and $(\hat{\rho}_i - \rho)^2$ $(i = J, R, SR)$ for each observation, and then average over the 1600 observations. This average is regarded as an estimate of the mean-squared error for the corresponding estimator. The prior which produces posterior means having the smallest mean-squared error is preferred.

(v) The posterior 0.05 and 0.95 quantiles w.r.t. the four candidate priors for each specific observation set $\underset{\sim}{X}$ are also recorded. Then the proportion of quantiles that exceed ρ, in the 1600 observation sets, is an estimate of the frequentist coverage of posterior 0.05 and 0.95 quantiles. Typically one prefers the prior that gives proportions close to 0.05 and 0.95, respectively. Also relevant in this comparison is the expected difference between the 0.95 and 0.05 quantiles, with small values of this difference being more desirable (implying more concentrated knowledge).

The simulation results for mean-squared error are given in Table 1, with the standard errors in parentheses. For T=50 and 100, the various estimators were numerically virtually indistinguishable for larger ρ, so only $\rho = 1.1$ is reported. Table 2 presents the estimated coverage of the posterior quantiles. The standard errors of these estimated coverage probabilities is about 0.005.

Table 1. Mean-squared error of posterior means and the MLE.

Parameter		Prior		
T	ρ	Symmetrized Reference	Constant (and MLE)	Jeffreys
5	.25	.21(.0060)	.24(.011)	.68(.060)
	.50	.20(.0073)	.25(.012)	.67(.072)
	.75	.21(.0091)	.27(.014)	.65(.080)
	1.0	.22(.011)	.29(.015)	.63(.092)
	1.25	.27(.013)	.30(.017)	.58(.11)
	1.5	.30(.015)	.29(.019)	.51(.12)
10	.25	.129(.0041)	.093(.0032)	.266(.013)
	.50	.104(.0043)	.090(.0037)	.205(.011)
	.75	.073(.0043)	.085(.0041)	.136(.0088)
	1.0	.054(.0043)	.077(.0044)	.089(.0080)
	1.25	.048(.0034)	.054(.0040)	.039(.0038)
	1.5	.038(.0044)	.039(.0049)	.026(.0080)
20	.25	.058(.0020)	.046(.0016)	.096(.0059)
	.50	.048(.0018)	.041(.0017)	.086(.0047)
	.75	.031(.0015)	.034(.0017)	.057(.0035)
	1.0	.016(.0014)	.025(.0017)	.019(.0013)
	1.25	.0057(.00069)	.0063(.00090)	.0023(.00036)
	1.5	.00045(.00030)	.00047(.00037)	.00022(.00017)
50	.25	.019(.00067)	.018(.00064)	.021(.00092)
	.50	.017(.00058)	.016(.00057)	.024(.0013)
	.75	.012(.00045)	.011(.00048)	.022(.00097)
	1.0	.0028(.00026)	.0046(.00032)	.0026(.00023)
	1.1	.0014(.00028)	.0017(.00033)	.00080(.00020)
100	.25	.0091(.00032)	.0089(.00032)	.0091(.00032)
	.50	.0075(.00027)	.0074(.00027)	.0080(.00040)
	.75	.0049(.00019)	.0048(.00020)	.0091(.00053)
	1.0	.00077(.000086)	.0012(.00010)	.00061(.000047)
	1.1	1.1E-6(7.7E-7)	9.6E-7(6.9E-7)	3.2E-7(1.8E-7)

Table 2. Frequentist probabilities that the 0.05, 0.95 posterior quantiles exceed ρ, and (in parentheses) the frequentist expected difference between the 0.95 and 0.05 quantiles.

Parameter		Prior		
T	ρ	Symmetrized Reference	Constant	Jeffreys
5	.25	.05, .96(1.77)	.03, .97(2.08)	.06, .96(2.61)
	.50	.05, .97(1.66)	.03, .97(2.01)	.07, .96(2.50)
	.75	.04, .98(1.49)	.03, .96(1.89)	.07, .97(2.31)
	1.0	.002,.97(1.31)	.04, .95(1.72)	.07, .98(2.06)
	1.25	.01, .81(1.16)	.04, .94(1.53)	.07, .97(1.78)
	1.5	.01, .75(1.04)	.04, .93(1.31)	.07, .97(1.52)
10	.25	.06, .96(1.23)	.03, .96(1.17)	.07, .95(1.49)
	.50	.06, .97(1.09)	.03, .96(1.09)	.08, .96(1.38)
	.75	.06, .98(.87)	.03, .95(.95)	.07, .97(1.18)
	1.0	.006,.97(.60)	.03, .94(.72)	.08, .98(.85)
	1.25	.03, .78(.38)	.03, .89(.43)	.07, .98(.48)
	1.5	.03, .86(.21)	.04, .90(.22)	.06, .95(.24)
20	.25	.06, .96(.82)	.04, .96(.76)	.07, .96(.90)
	.50	.06, .96(.74)	.04, .94(.70)	.07, .96(.87)
	.75	.07, .97(.56)	.03, .94(.57)	.08, .97(.74)
	1.0	.005, .97(.29)	.02, .93(.35)	.07, .99(.42)
	1.25	.03, .90(.067)	.03, .91(.072)	.05, .95(.077)
	1.5	.05, .94(.0067)	.05, .94(.0067)	.05, .95(.0068)
50	.25	.06, .96(.47)	.05, .96(.46)	.06, .96(.47)
	.50	.05, .96(.43)	.04, .95(.41)	.06, .96(.47)
	.75	.06, .96(.35)	.03, .94(.33)	.09, .96(.45)
	1.0	.007,.97(.12)	.03, .91(.14)	.08, .99(.18)
	1.1	.04, .90(.023)	.04, .90(.023)	.05, .95(.027)
100	.25	.05, .95(.32)	.05, .95(.32)	.05, .95(.32)
	.50	.05, .95(.29)	.05, .95(.29)	.05, .95(.30)
	.75	.06, .96(.23)	.04, .95(.22)	.07, .96(.28)
	1.0	.008,.97(.059)	.03, .90(.068)	.07, 1.0(.090)
	1.1	.05,.95(.00035)	.05,.95(.00035)	.05,.95(.00035)

3.3 Conclusions

1. The mean-squared errors for $\pi_{SR}(\rho)$ seem generally superior, except for larger T and $\rho > 1$. The constant prior posterior mean (and MLE) have quite satisfactory mean-squared error. The good performance of the Jeffreys prior for larger T and $\rho > 1$ is noteworthy, because the criticisms of the Jeffreys prior often focus on its supposedly extreme form for $\rho > 1$; apparently this form is not really "extreme" operationally (see, also, Phillips [21]).

2. The coverage results are not clear cut. The Jeffreys (and nonasymptotic reference) posterior seem to be systematically shifted too far right, but their coverage performance is not egregiously bad. The coverages for the symmetrized reference and constant priors are generally more attractive, with the exception of the "small T - large ρ" behavior for π_{SR}. The generally fine coverages for the symmetrized reference prior are achieved even though it typically yields a substantially narrower posterior, as is evidenced by the smaller expected posterior quantile differences (in parentheses in Table 2).

3. In conjunction with the intuitive arguments from Section 2.3, a fairly strong case can be made for adopting $\pi_{SR}(\rho)$ as the recommended "default" prior for the AR(1) model.

4 Bayesian Testing for the AR(1) Model

4.1 Introduction

In this section, we consider testing from the robust Bayesian perspective, developing the lower bounds on Bayes factors over a large range of priors. The Bayes factor based on the symmetrized reference prior developed in Section 2.3 is also investigated. Since there are substantial differences between $|\rho| < 1$ and $|\rho| > 1$ in the AR(1) model, and since $\rho = 1$ has possibly special significance, we consider testing

$$H_0 : \rho = 1 \;\; vs. \;\; H_1 : \rho < 1 \;\; vs. \;\; H_2 : \rho > 1 \, . \tag{21}$$

For simplicity, we will assume that $\rho > 0$, though similar results could be developed for the more general case. Note that $H_0 : \rho = 1$ is typically a surrogate for $H_0 : |\rho - 1| < \epsilon$. In Bayesian analysis, it can be used as such a surrogate if $\epsilon < \frac{1}{2}\sigma_{\hat{\rho}_{mle}} \approx 1/[2\sqrt{\sum x_i^2}]$ (cf., Berger and Delampady, [10]). Define

$g_1 =$ prior density on $\rho < 1$, conditional on H_1 being true;

$g_2 =$ prior density on $\rho > 1$, conditional on H_2 being true.

Then B_{ij}, the *Bayes factor* of H_i to H_j, is given as follows:

$$B_{01} = f(\underset{\sim}{x} \mid 1) / \int_0^1 f(\underset{\sim}{x} \mid \rho) g_1(\rho) d\rho \,, \tag{22}$$

$$B_{02} = f(\underset{\sim}{x} \mid 1) / \int_1^\infty f(\underset{\sim}{x} \mid \rho) g_2(\rho) d\rho \,, \tag{23}$$

$$B_{12} = \frac{\int_0^1 f(\underset{\sim}{x} \mid \rho) g_1(\rho) d\rho}{\int_1^\infty f(\underset{\sim}{x} \mid \rho) g_2(\rho) d\rho} \,. \tag{24}$$

The choices of g_1 and g_2 that will be considered are:

(I) *Default*: Choose g_1 and g_2 corresponding to $\pi_{SR}(\rho)$, i.e.,

$$g_1 = 1/\left[\pi\sqrt{1-\rho^2}\right] \,, \quad g_2 = 1/\left[\pi\rho\sqrt{\rho^2 - 1}\right] \,. \tag{25}$$

(II) *"Objective" Classes*: For many econometric problems, reasonable subjective prior opinions will reside in the following classes:

$$g_1 \in \mathcal{G}_1 = \{nondecreasing \;\; densities \;\; on \;\; (0,1)\} \,,$$

$$g_2 \in \mathcal{G}_2 = \{nonincreasing \;\; densities \;\; on \;\; (1,\infty)\} \,. \tag{26}$$

4.2 Bayes Factors for the Default Prior

To reiterate, one of the attractive properties of $\pi_{SR}(\rho)$ (in (19)) is that it is proper and gives equal probability to $\rho < 1$ and $\rho > 1$. Computation of the Bayes factors B_{01}, B_{02} and B_{12} using (19) (or (25)) is straightforward numerically.

Example 1. Suppose T=10. The Bayes factors depend on both $A = \sum_{i=1}^{T} x_{i-1}^2$ and on $\hat{\rho}_{mle}$. Classical testing depends on the test statistic

$$W = \sqrt{A}(\hat{\rho}_{mle} - 1) \,. \tag{27}$$

It is of interest to graph the Bayes factors as, say, a function of A, when W is fixed at the standard classical 0.05 critical value $W_{0.05} = -1.75$ for B_{01}, or $W_{0.05} = 1.38$ for B_{02} and B_{12}. The graph for

B_{01}, B_{02}, and B_{12} are presented in Figure 1 over a reasonably large range of A.

That B_{01} B_{02} are substantially larger than 0.05 is not a great surprise; in testing a point null, classical P-values are well known to seriously overstate the evidence against the null (cf., Berger and Sellke [9] and Berger and Delampady [10]). It is somewhat interesting that B_{02} seems to be almost twice as large as B_{01}.

More surprising is the graph for B_{12} over this reasonable range of A, since B_{12} is still substantially bigger than the classical P-value 0.05. This is surprising because of the perception (cf., Casella and Berger [14]) that in one-sided testing (here $H_0 : \rho < 1$ versus $H_1 : \rho > 1$), P-values and Bayes factors tend to be similar.

4.3 Bayes Factors Over "Objective" Classes of Priors

If indeed it is felt that reasonable prior opinions reside in $\mathcal{G}_1$ and $\mathcal{G}_2$ in (26), two robust Bayesian techniques can be exploited. Both are based on observing that the "extreme" points of $\mathcal{G}_1$ and $\mathcal{G}_2$ are, respectively, the following densities:

$$g_{1,r} = Uniform(1-r, 1), \;\; and \;\; g_{2,r} = Uniform(1, 1+r)\,. \tag{28}$$

Robust Bayesian Technique 1: Graph the Bayes factors as functions of the extreme points, thus indicating the range of reasonable opinions. Indeed, by specifying the range of r he/she considers reasonable, a statistician can determine from the graph the relevant personal range of Bayes factors. Using (28), the expressions for B_{01} and B_{02} are

$$B_{01}(r) = \frac{f(\underset{\sim}{x}\,|1)}{\frac{1}{r}\int_{1-r}^{1} f(\underset{\sim}{x}\,|\rho)d\rho}, \;\; B_{02}(r) = \frac{f(\underset{\sim}{x}\,|1)}{\frac{1}{r}\int_{1}^{1+r} f(\underset{\sim}{x}\,|\rho)d\rho}\,. \tag{29}$$

(These had earlier been considered, with a different motivation, by Schotman and Van Dijk [23].)

Although (28) could also be used to define $B_{12}(r)$, there is a concern: $g_{1,r}$ and $g_{2,r}$ might reflect quite different levels of information. Thus we, instead, follow the "invariance" idea of Section 2.3 and define $g_{2,r}^*$ to be the transformation of $g_{1,r}(\rho)$ induced by the mapping $\rho \to 1/\rho$. The result is

$$g_{2,r}^*(\rho) = \frac{1}{\rho^2} g_1(\frac{1}{\rho}) \;\; on \;\; \rho > 1\,. \tag{30}$$

This ensures, in a sense, that the priors assigned to H_1 and H_2 carry similar levels of information, which is important in the effort to be as "objective" as possible. The resulting Bayes factor is

$$B_{12}(r) = \int_{1-r}^{1} f(\underset{\sim}{x} \mid \rho) d\rho \Big/ \int_{1}^{(1-r)^{-1}} f(\underset{\sim}{x} \mid \rho) \frac{1}{\rho^2} d\rho \, . \tag{31}$$

(Note that the class of priors for which $B_{12}(r)$ represents the "extreme" Bayes factors is the class specified by $\mathcal{G}_1$, with the invariance map to $\rho > 1$.)

Example 1 (continued). For the situation of Example 1, $B_{01}(r), B_{02}(r)$, and $B_{12}(r)$ are graphed in Figure 2 (as functions of r) for rather different pairs of (A, $\hat{\rho}_{mle}$). In all cases, a classical test for the same data would have P-value=0.05.

As an example of the use of such graphs, suppose a user felt that values of ρ less than 0.3 were very unlikely apriori, while values of ρ around 0.8 were quite plausible. This would suggest that the range $0.2 \le r \le 0.7$ should be considered. For the six graphs in Figure 2, the corresponding range of Bayes factors are (.29, .46), (.30, .57), (.47, .59), (.47, 1.04), (.15, .27) and (.12, .15). There is clearly a reasonable degree of robustness here.

Robust Bayesian Technique 2: It is interesting to consider the minimum values of $B_{01}(r), B_{02}(r)$, and $B_{12}(r)$, namely $\underline{B}_{01} = \inf_r B_{01}(r)$, $\underline{B}_{02} = \inf_r B_{02}(r)$, $\underline{B}_{12} = \inf_r B_{12}(r)$. These have the interpretation that the evidence for the precise unit root hypothesis $H_0 : \rho = 1$ is *at least* $\underline{B}_{01}$ or $\underline{B}_{02}$ (depending on the alternative being considered), while the evidence for stationarity as opposed to explosive behavior is *at least* $\underline{B}_{12}$; and these are lower bounded on the Bayes factor over *any* prior in the relevant "objective" class. Thus $\underline{B}_{01}$ gives the lower bound on the Bayes factor for testing $H_0 : \rho = 1$ versus $H_1 : \rho < 1$ over *all* priors whose density on $\rho < 1$ is nondecreasing.

It can be shown that $\underline{B}_{01}$ and $\underline{B}_{02}$ depend only on W, and are given by

$$\underline{B}_{01} = \inf_r \frac{\exp(-W^2/2)}{\frac{1}{r}\int_{-r}^{0} \exp[-\frac{1}{2}(W-\xi)^2] d\xi} \, , \tag{32}$$

$$\underline{B}_{02} = \inf_r \frac{\exp(-W^2/2)}{\frac{1}{r}\int_{0}^{r} \exp[-\frac{1}{2}(W-\xi)^2] d\xi} \, . \tag{33}$$

Since the classical test level, α, also depends only on W, it is possible to graph $\underline{B}_{01}$ and $\underline{B}_{02}$ as functions of α. This is done in Figure 3 for the situation of Example 1. Thus, when $\alpha = 0.05$ and $T = 10$, $\underline{B}_{01} = 0.29$ and $\underline{B}_{02} = 0.47$. These comparatively large lower bounds, which indicate that the evidence for H_0 to H_1 or H_2 is *at least* 3/10 or 1/2, respectively, demonstrate the truly misleading nature here of the classical P-value, α. Classical P-values are far too ready to reject the "precise" unit root hypothesis $H_0 : \rho = 1$.

Unfortunately, $\underset{\sim}{B}_{12}$ is not simply a function of W. Thus, for a specified classical level α, we can present $\underset{\sim}{B}_{12}$ only as a function of, say, A. This is done in Figure 3 for $\alpha = 0.01$ and $\alpha = 0.05$. While $\underset{\sim}{B}_{12}$ is still substantially larger than α, indicating that α overstates the evidence against stationarity, the difference for this one- sided testing case is not as dramatic as for $\underset{\sim}{B}_{01}$ or $\underset{\sim}{B}_{02}$.

Appendix: Proof of Lemma 1

For the Gaussian AR(1) model with $\sigma^2 = 1$ and $x_0 = 0$, Anderson [1] and White [27] established that

$$\frac{1}{T}\sum_{i=1}^{T} X_{i-1}^2 \ \underset{T\to\infty}{\overset{d}{\longrightarrow}} \ \frac{1}{1-\rho^2} \qquad |\rho| < 1\,, \tag{34}$$

$$\rho^{-2T}\sum_{i=1}^{T} X_{i-1}^2 \ \underset{T\to\infty}{\overset{d}{\longrightarrow}} \ \frac{[N(0,1)]^2}{(\rho^2-1)^2} \qquad |\rho| > 1\,. \tag{35}$$

To compute the reference prior, we first establish that

$$E_\rho[\log(\frac{1}{T}\sum_{i=1}^{T} X_{i-1}^2)] \ \to \ \log\frac{1}{1-\rho^2} \qquad |\rho| < 1\,, \tag{36}$$

$$E_\rho[\log(\rho^{-2T}\sum_{i=1}^{T} X_{i-1}^2)] \ \to \ E\log\frac{[N(0,1)]^2}{(\rho^2-1)^2} \qquad |\rho| > 1\,. \tag{37}$$

These are immediate from (34) and (35) if uniform integrability of $\log(\sum_{i=1}^{T} X_{i-1}^2)$ can be demonstrated. Uniform integrability is guaranteed by uniform boundness of the square integral. This will be established in the following, using the inequality $|\log\sqrt{z}\,|^2 < z + 1/z$:

(i) *For* $|\rho| < 1$:

$$\begin{aligned} E_\rho\left|\log\sqrt{\tfrac{1}{T}\textstyle\sum_{i=1}^{T} X_{i-1}^2}\right|^2 &\le E_\rho\left[\frac{1}{T}\sum_{i=1}^{T} X_{i-1}^2\right] + E_\rho\left[T/\sum_{i=1}^{T} X_{i-1}^2\right] \\ &\overset{say}{=} C + D\,. \end{aligned} \tag{38}$$

Clearly,

$$C = \frac{1}{T}\sum_{i=1}^{T-1}\frac{1-\rho^{2i}}{1-\rho^2} < \frac{1}{1-\rho^2} . \tag{39}$$

In order to bound $E_\rho\left[T/\sum_{i=1}^T X_{i-1}^2\right]$, it is convenient to use the notation: $\xi \succ \eta$ means that ξ is stochastically larger than η, in the sense that $\Pr(\xi > t) > \Pr(\eta > t)$, $\forall t$. Note that if $\xi \succ \eta$ and they are both nonnegative, then since $E\xi = \int_0^\infty \Pr(\xi > t)dt$, $E(1/\xi) < E(1/\eta)$. Also, one can see that if $\xi_1 \succ \eta_1$, $\xi_2 \succ \eta_2$, then $\xi_1 + \xi_2 \succ \eta_1 + \eta_2$. Since $X_i^2 \succ \chi_1^2$ for $i = 1, ..., T-1$, it follows that $\sum_{i=1}^T X_{i-1}^2 \succ \chi_{T-1}^2$ and $T/\sum_{i=1}^T X_{i-1}^2 \prec T/\chi_{T-1}^2$. Thus

$$D = E_\rho(T/\sum_{i=1}^T X_{i-1}^2) \leq E(T/\chi_{T-1}^2) = \frac{T}{T-3} < 1 . \tag{40}$$

(ii) *For* $|\rho| > 1$: by the same methods as above

$$\begin{aligned} E_\rho\left|\log\sqrt{\rho^{-2T}\textstyle\sum_{i=1}^T X_{i-1}^2}\right|^2 &\leq E_\rho[\rho^{-2T}\sum_{i=1}^T X_{i-1}^2] + E_\rho[\rho^{2T}/\sum_{i=1}^T X_{i-1}^2] \\ &\overset{say}{=} C + D . \end{aligned} \tag{41}$$

Clearly

$$\begin{aligned} C &= \frac{1}{\rho^{2T}}\sum_{i=1}^{T-1}\frac{1-\rho^{2i}}{1-\rho^2} \\ &= \frac{1-\rho^{2(1-T)}}{(\rho^2-1)^2} - \frac{T-1}{\rho^{2T}(\rho^2-1)} < \frac{1}{(\rho^2-1)^2} . \end{aligned} \tag{42}$$

In order to bound $E_\rho[\rho^{2T}/\sum_{i=1}^T X_{i-1}^2]$, notice that $X_i^2 \succ \rho^{2(T-4)}\chi_1^2$, for $i = T-4, ..., T-1$, so that $\sum_{i=1}^T X_{i-1}^2 \succ \rho^{2(T-4)}\chi_4^2$. Therefore:

$$D = E_\rho[\rho^{2T}/\sum_{i=1}^T X_{i-1}^2] \leq E[\rho^{2T}/(\rho^{2(T-4)}\chi_4^2)] = \rho^8/2 . \tag{43}$$

Therefore, we have established the uniform integrability of $\frac{1}{2}[\log(\sum_{i=1}^T X_{i-1}^2/T)]$ for $\rho < 1$ and $\frac{1}{2}[\log(\sum_{i=1}^T X_{i-1}^2/\rho^{2T})]$ for $\rho > 1$. Thus (36) and (37) have been established.

To complete the proof, we take the limit as $T \to \infty$ inside the integral over ρ in the denominator of (15), noticing that, if $C = [-1, 1]$, then $\exp\{\frac{1}{2}E_\rho \log[\frac{1}{T}\sum_{i=1}^T X_{i-1}^2]\}$ is uniformly bounded over

T by $1/\sqrt{1-\rho^2}$ (Jensen's inequality applies here) which is integrable in $[-1,1]$. Thus, by the Dominated Convergence Theorem and (36), $\pi_C(\rho) \propto (1-\rho^2)^{-1/2}$ for $C=[-1,1]$.

For $C=[a,b]$ with $a<-1$ or $b>1$, w.l.o.g., consider $|b|>|a|$. For $0<\epsilon<(|b|-|a|)/3$, write:

$$\begin{aligned} C &= [a,b] = [a, b-3\epsilon] \cup [b-3\epsilon, b-\epsilon] \cup [b-\epsilon, b] \\ &\overset{say}{=} C_1 \cup C_2 \cup C_3 \,. \end{aligned} \tag{44}$$

If $\rho \in C_1 \setminus [-1,1]$, then by (37):

$$E_\rho \log\left[(b-2\epsilon)^{-2T}\sum_{i=1}^{T} X_{i-1}^2\right] = E_\rho \log(\rho^{-2T}\sum_{i=1}^{T} X_{i-1}^2) + 2T[\log|\rho| - \log(b-2\epsilon)] \overset{T\to\infty}{\longrightarrow} -\infty \,. \tag{45}$$

The limit of $-\infty$ can similarly be established for $\rho \in [-1,1]$. For $\rho \in C_3$, also by (37):

$$E_\rho \log\left[(b-2\epsilon)^{-2T}\sum_{i=1}^{T} X_{i-1}^2\right] = E_\rho \log(\rho^{-2T}\sum_{i=1}^{T} X_{i-1}^2) + 2T[\log\rho - \log(b-2\epsilon)] \overset{T\to\infty}{\longrightarrow} +\infty \,. \tag{46}$$

Thus, as T grows,

$$\pi(\rho) = \frac{\exp\{\frac{1}{2}E_\rho[\log \sum_{i=1}^{T} X_{i-1}^2]\}}{\int_C \exp\{\frac{1}{2}E_\rho[\log \sum_{i=1}^{T} X_{i-1}^2]\} d\rho} \tag{47}$$

will concentrate its mass on $C_2 \cup C_3$. As ϵ can be arbitrarily small, the conclusion of the lemma follows. ■

Acknowledgment

We are grateful to J.K. Ghosh for helpful discussions, and to Peter Phillips and a referee for valuable comments and suggestions.

References

[1] Anderson, T.W. On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30 (1959): 676-687.

[2] Bayes, T.R. Essay towards solving a problem in the doctrine of chances, 1763. Reprinted in *Biometrika*, 45 (1958): 243-315.

[3] Berger, J.O. *Statistical Decision Theory and Bayesian Analysis.* New York: Springer-Verlag, 1985.

[4] Berger, J. and Bernardo, J.M. Estimating a product of means: Bayesian analysis with reference priors. *Journal of American Statistical Association* 84 (1989): 200-207.

[5] Berger, J. and Bernardo, J.M. Reference priors in a variance components problem. In *Bayesian Analysis in Statistics and Econometrics*, P. Goel and N.S. Iyengar(eds.). New York: Springer Verlag, 1992.

[6] Berger, J. and Bernardo, J.M. Ordered group reference priors with application to a multinomial problem. *Biometrika* 79 (1992): 25-37.

[7] Berger, J. and Bernardo, J.M. On the development of the reference prior method. In *Bayesian Statistics* 4, J.M.Bernardo, J.O.Berger, D.V.Lindley, and A.F.M.Smith (eds.). London: Oxford University Press, 1992.

[8] Berger, J., Bernardo, J.M. and Mendoza, M. On priors that maximize expected information. In *Recent Developments in Statistics and Their Applications*, J.P.Klein and J.C.Lee (eds.). Seoul: Freedom Academy Publishing, 1989.

[9] Berger, J., Sellke, T. Testing of a point null hypothesis: the irreconcilability of significance levels and evidence (with discussion). *Journal of American Statistical Association* 82 (1987): 112-139.

[10] Berger, J. and Delampady, M. Testing precise hypotheses, *Statistical Science* 2 (1987): 317-352.

[11] Berger, J., Berry, D. Analyzing data: Its objectivity possible. *American Scientist* 76 (1988): 159-165.

[12] Bernardo, J.M. Reference posterior distributions for Bayes inference. *Journal of Royal Statistical Society* B 41 (1979): 113-147 (with discussion).

[13] Box, G.E.P. and Jenkens, G.M. *Time Series Analysis: Forecasting and Control.* San Francisco: Holden Day, 1976.

[14] Casella, G. and Berger, R. Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *Journal of American Statistical Association* 82 (1987): 106-111.

[15] Jeffreys, H. *Theory of Probability.* London: Oxford University Press, 1937, 1961.

[16] Laplace, P. Mémoire sur la probabilityé des causes par les évenemens. *Mem. Acad. R. Sci. Presentés par Divers Savans* 6 (1774): 621-656 (translated in Statistical Science 1,359-378).

[17] Laplace, P. *Theorie Analytique des Probabilites.* Paris: Courcier, 1812.

[18] Mitchell, A. Discussion of paper by I.J. Good. *Journal of Royal Statistical Society* B 29 (1967): 423-424.

[19] Lindley, D.V. On a measure of the information provided by an experiment. *Annals of Mathematical Statistics* 27 (1956): 986-1005.

[20] Monette, G., Fraser, D.A.S., and Ng, K.W. Marginalization, likelihood, and structural models. In *Multivariate Analysis VI*, P.R. Krishnaiah (ed.). Amsterdam North-Holland, 1984.

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

[22] Rubin, H. Consistency of maximum likelihood estimates in the explosive case. In *Statistical Inference in Dynamic Economic Models* (T.C. Koopmans, ed.), 356-364. New York: Wiley, 1950.

[23] Schotman, P.C. and Van Dijk, H.K. On Bayesian routes to unit roots. *Journal of Applied Econometrics* 6 (1991): 387-401.

[24] Stein, C. Approximation of improper prior measures by prior probability measures. In *Bernoulli-Bayes-Laplace Festschrift*, 217-240. New York: Springer-Verlag, 1965.

[25] Uhlig, H. What macroeconomists should know about unit roots as well: the Bayesian perspective. Presented at the NSF-Yale Symposium on Bayesian Methods and Unit Roots (1992).

[26] Welch, B.L. and Peers, H.W. On formulae for confidence points based on intervals of weighted likelihoods. *Journal of Royal Statistical Society* B 25 (1963): 318-329.

[27] White, J.S. The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* 29 (1958): 1188-1197.

[28] Ye, K.Y. and Berger, J. Noninformative priors for inference in exponential regression models. *Biometrika* 78 (1991): 645-656.

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

[30] Zellner, A. Maximal data information prior distributions. In *New Methods in the application of Bayesian Methods*, A. Aykac (ed.). Amsterdam: North-Holland, 1977.

[31] Zellner, A. and Siow, A. Posterior odds for selected regression hypotheses. In *Bayesian Statistics*, J.M.Bernardo, M.H.DeGroot, D.V.Lindley, and A.F.M.Smith (eds.). Valencia: University Press, 1980.

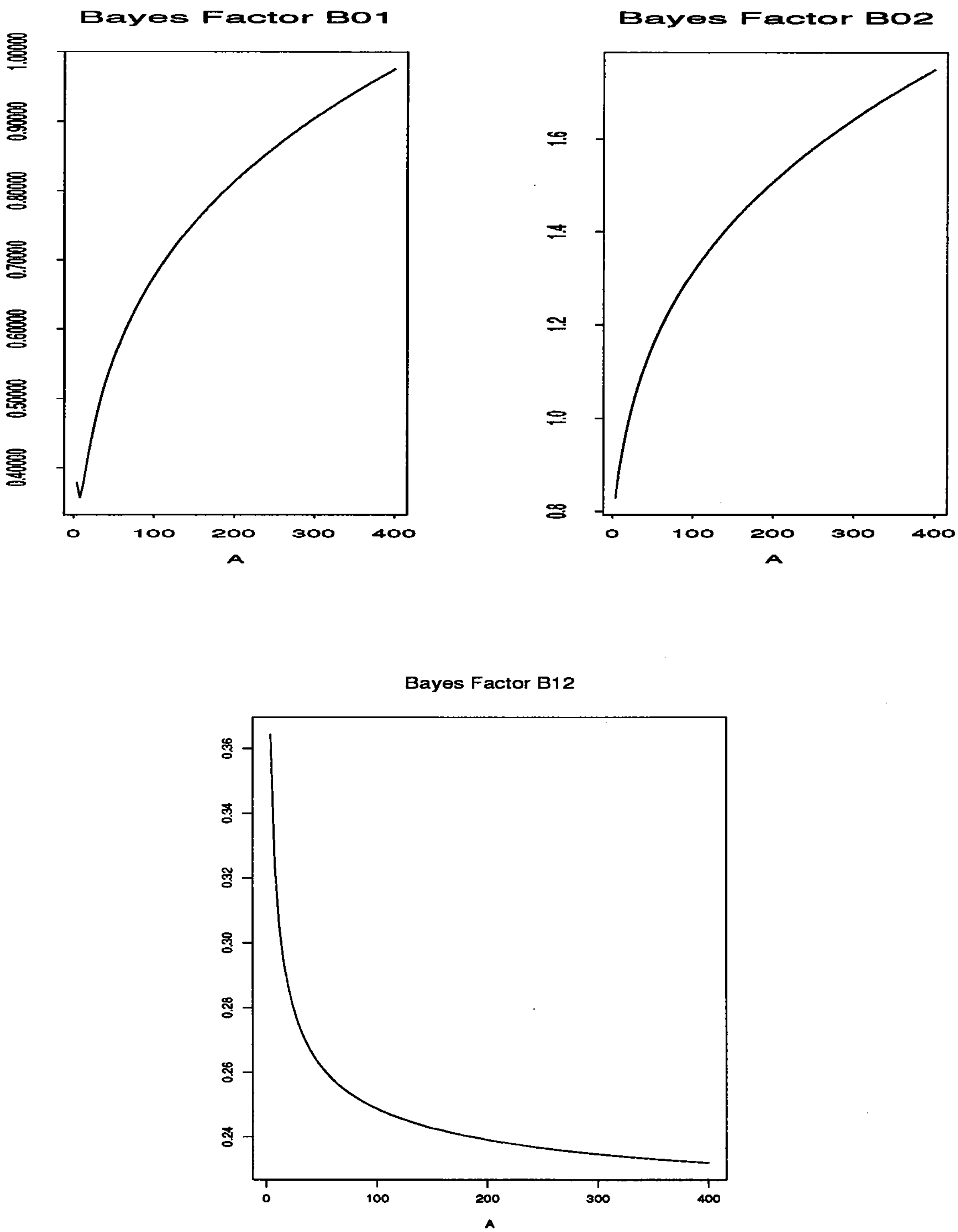

Figure 1: **Bayes Factors for the Default Prior**

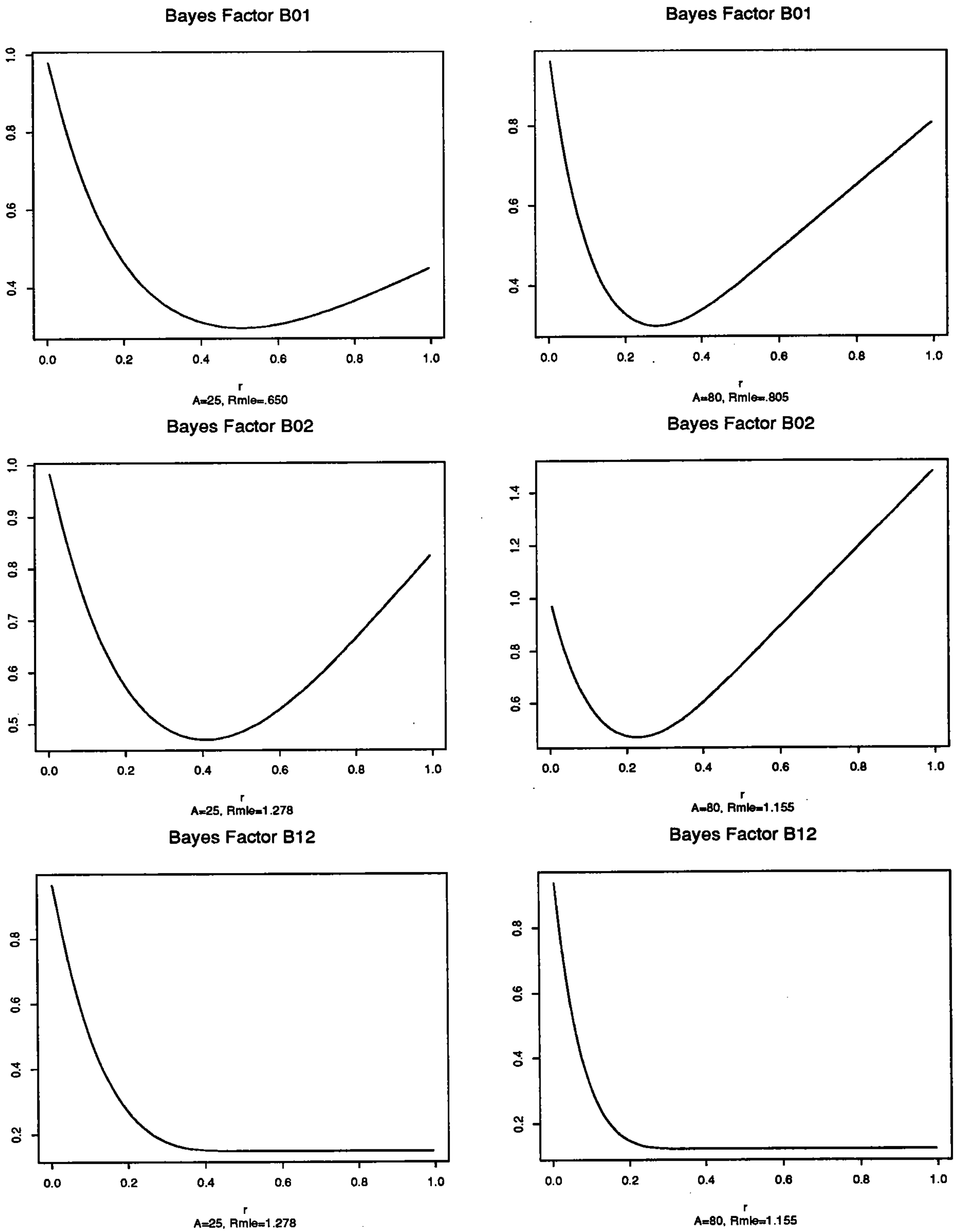

Figure 2: **Bayes Factors Over "Objective" Classes of Priors**

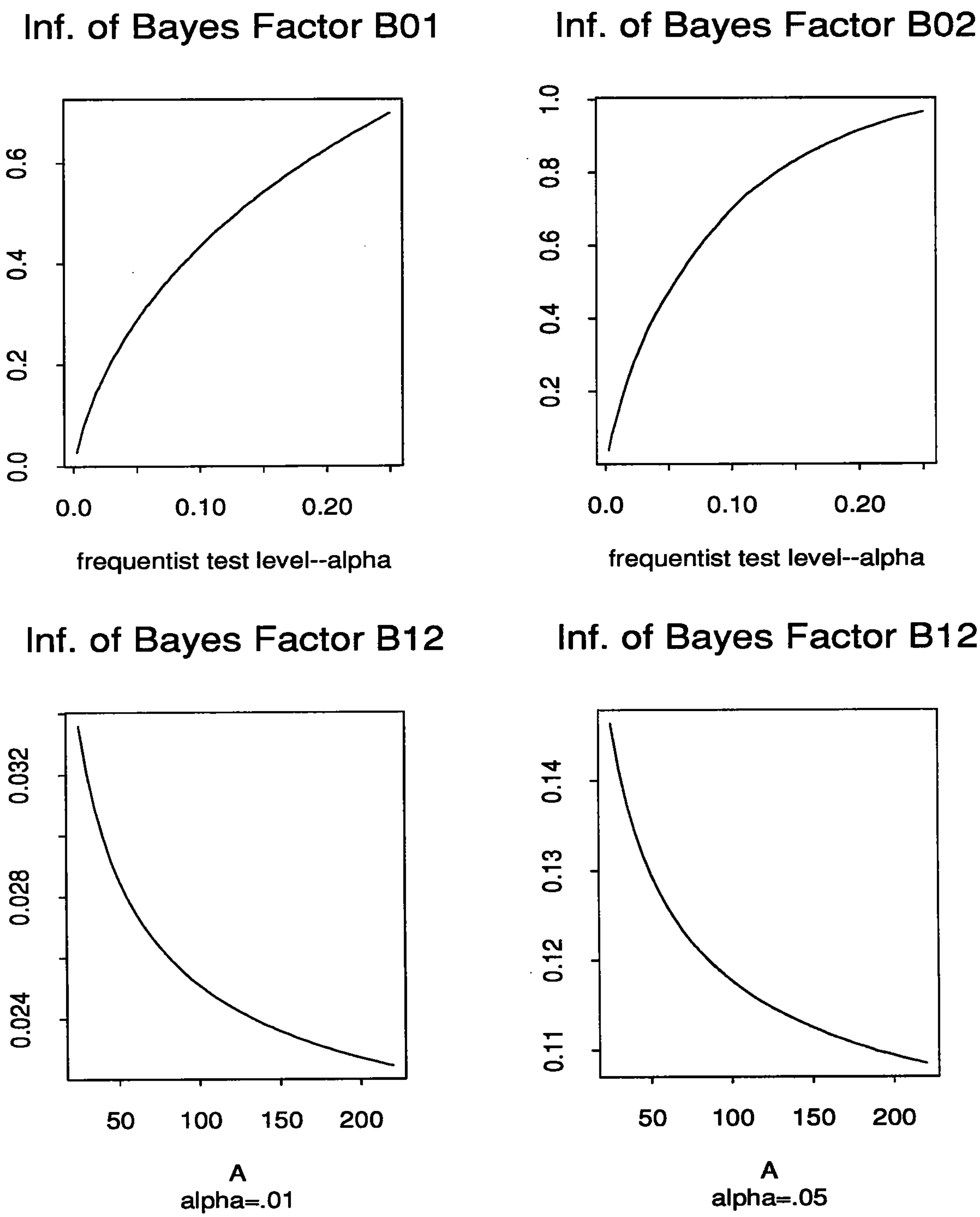

Figure 3: **Infimum of Bayes Factors Over "Objective" Classes of Priors**