THE ANALYSIS OF AUTOREGRESSIVE PROCESSES: THE IDENTIFICATION AND THE PRIOR, POSTERIOR, AND PREDICTIVE ANALYSIS

By

MOHAMMAD ABDULHADI AL-MAHMEED

Bachelor of Statistics
Kuwait University
Kuwait City, Kuwait
1975

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1979

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
December, 1982

THE ANALYSIS OF AUTOREGRESSIVE PROCESSES:

THE IDENTIFICATION AND THE PRIOR,

POSTERIOR, AND PREDICTIVE

ANALYSIS

Thesis Approved:

Lyle Broemeling
Thesis Adviser

W D Warde

William H. Stewart

Darryll E. Ray

Norman N Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

The author wishes to express his appreciation to his major adviser, Dr. Lyle D. Broemeling, for his guidance and assistance. Appreciation also is expressed to the other committee members, Dr. William D. Warde, Dr. William H. Stewart, and Dr. Daryll E. Ray, for their aid and encouragement. Ms. Barbara Newport gave valuable assistance in the excellent typing.

Special gratitude is reserved for my wife and son for their patience.

TABLE OF CONTENTS

LIST OF TABLES

Table		Page

Table | Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

The family of autoregressive (AR) models is a subclass of a rich class of models for time-series known as autoregressive integrated moving average (ARIMA) models. When analyzing a time series, we think of the observations as a finite segment of a possible realization of an underlying stochastic process.

Let $\{Y(t)\}_{t=1}^{n}$ denote n successive measurements of the process made at equidistant points in time. $\{Y(t)\}$ may consist of the differenced values of some original set of observations, $\{X(t)\}_{t=1}^{M}$. An AR(p) model for $\{Y(t)\}$ has the form

$$Y(t) = \phi_1 Y(t-1) + \ldots + \phi_p Y(t-p) + \varepsilon(t),$$

for $t = 1,2,\ldots,n$.

The autoregressive coefficients $\phi_1,\ldots,\phi_p$ are real unknown parameters, and $\{\varepsilon(t)\}$ is a series of identically distributed, uncorrelated, zero mean random variables (white noise), usually assumed i.i.d. $N(0,\tau)$, where τ is a positive unknown precision parameter. The vector of previous values $Y_0 = [Y(0),\ldots,Y(1-p)]$ is assumed to be known. Using the backshift operator B, where $BY(t) = Y(t-1)$, the AR(p) can be written

$$(1 - \phi_1 B - \ldots - \phi_p B^p)Y(t) = \varepsilon(t)$$

or, simply

$$\phi(B)Y(t) = \varepsilon(t),$$

and $\phi(B)$ is called an autoregressive operator of order p. Formally, we can write $Y(t) = \phi^{-1}(B)\varepsilon(t)$, so that $Y(t) = \psi(B)\varepsilon(t)$. The polynomial $\psi(B)$ for AR(p) will be a polynomial operator of infinite order.

Y(t) can be expressed as the output of a linear filter which acts on the white noise $\varepsilon(t)$, that is, as a weighted sum of present and past values of $\varepsilon(t)$: $Y(t) = \mu + \varepsilon(t) + \psi_1\varepsilon(t-1) + \psi_2\varepsilon(t-2) + \ldots$, or, $Y(t) = \mu + \psi(B)\varepsilon(t)$. $\{\varepsilon(t)\}$ is regarded as "series shocks" which drive the system. The parameter μ is the level of the process. Y(t) will be a random variable for each t.

If $\sum_{t=0}^{\infty} \psi_t$ converges, $\{Y(t)\}$ is said to be stationary. In a simple way, $\{Y(t)\}$ is stationary if the influence of $\varepsilon(t-p)$ on the value Y(t), as represented by the ψ_p weight, decreases sufficiently rapidly as p increases. If $\{Y(t)\}$ is stationary, μ is the mean of the process, that is, $E(Y(t)) = \mu$, for all t's.

One may also have $Y(t) = (\phi_1 B + \phi_2 B^2 + \ldots + \phi_p B^p)Y(t) + \varepsilon(t)$, expressing Y(t) as a weighted sum of past p values of Y(t) plus the current shock $\varepsilon(t)$. Y(t) can be viewed as a regression on the last p y-values.

If $\phi(B)$, the operator polynomial, converges for $|B| \leq 1$, the linear process $\{Y(t)\}$ is said to be invertible, and in simple terms this means that in the autoregression of Y(t) on the past p Y(t)'s, the weights must decrease quickly as p increases; i.e., the influence of the past is cut off after the p-th lag of the AR(p) model.

The stationarity and invertibility conditions for AR(p) can be specified by conditions on $\phi(B)$: the AR(p) model $\phi(B)Y(t) = \varepsilon(t)$ will be stationary if and only if the roots of $\phi(B) = 0$ lie outside the unit circle (in the complex plane), and is always invertable.

A Bayesian solution to any statistical problem requires two essentials. Determination of the likelihood function based on the structure of the problem and assumptions about the generating mechanism of the data. Secondly, a prior distribution for the parameters of the likelihood function must be selected and the values of the parameters chosen in such a way to reflect the prior information available about the parameters. The posterior distribution of the parameters is found by combining the likelihood function (using the data) and the prior distribution, via Bayes theorem. The posterior densities can be used to compute probabilities that a parameter or set of parameters lie in a particular region, or may be used to find regions of highest posterior density (HPD).

Forecasting is accomplished with the Bayesian predictive probability density function of k future observations $Y(n+1), Y(n+2), \ldots, Y(n+k)$. The predictive pdf is obtained as follows:

$$p(\text{future} \mid \text{data}) = \int p(\text{future} \mid \text{data, parameters}) \times p(\text{parameters} \mid \text{data})\, d(\text{parameters}).$$

From the predictive density, point forecasts can be generated from a measure of central tendency. Probabilities of intervals and regions and finding regions of specified probability can also be evaluated.

The objective of this study is to develop the posterior analysis and forecasting techniques using a conjugate prior, assuming stationarity of AR(p) within a probability sense. That is, instead of restricting the parameter space of the A(1) model to $\{\phi: \ |\phi| < 1\} =$

(-1,1), which is the usual way to assume stationarity, choose the prior parameters so that, say, 95% of the marginal prior probability of ϕ is concentrated on (-1,+1).

A numerical study will be devised to assess the effects of prior information on the posterior and predictive distributions, and to study the effect of using mixtures of conjugate prior distributions (which is thought to be more flexible in the prior analysis).

Also, a Bayesian method for identifying the order of an autoregressive process based on the posterior distribution of the order will be examined, and the sensitivity of this method to the choice of prior parameters. The sample size will be investigated and compared with other classical methods.

Finally, the mathematical derivation of a Bayesian method for identifying the order of a multivariate autoregressive process will be given.

CHAPTER II

REVIEW OF LITERATURE

The idea of time related observations has been known for centuries, but according to Kendall (1973), much of the study of time-series analysis has been carried out since the work of Udny Yule in 1927, which led toward the ideas of stochastic processes. Chatfield (1975) notes that around 1920 Yule introduced autoregressive-type processes. In fact, AR(2) processes are sometimes called Yule AR processes (Kendall, 1973).

Bayesian ideas were first formally expressed in the writings of Thomas Bayes (1958) more than two centuries ago, but were partially discredited and fell into disuse. The use of a conjugate family of distributions was introduced and formalized by Raiffa and Schlaifer (1961), in which they also studied many families of distributions. Discussions of Bayesian analysis can be found in many well-known books, among them, Lindley (1972), DeGroot (1970), and Box and Tiao (1973).

Classical analysis of AR processes has been known since the early 1920's, but in recent years, Box and Jenkins (1970) methods have become a popular technique for model building. Box and Jenkins discuss three stages in model building: identification, estimation, and diagnostic checking. The identification is made by using the plots of the estimated autocorrelations and partial autocorrelations from a sample. Given the form of the model, the parameters are then estimated by iterative least squares. Stationarity and invertibility are introduced by differencing the original process. This will force the determination of the process to be unique. Given the estimates of the parameters a model is then fitted and diagnostic checking is performed, using the residuals of the fitted model. The autocorrelations of the residuals can indicate inadequacies of the selected model, and can suggest alternative models.

Most published studies of AR processes from a Bayesian point of view are focused on parameter estimation. Vague prior information is often used to illustrate relationships to maximum likelihood estimators. For the most part, the priors used do not include the restrictions on the parameters which introduce stationarity and invertibility in the model.

Box and Jenkins outline some aspects of the Bayesian analysis of the AR model, where a Jeffrey's non-informative prior is used. For the AR(p) model, Box and Jenkins show the posterior distribution of ϕ is a multivariate t-distribution.

Zellner (1971) gives the posterior analysis of the AR(1) and AR(2) models using a vague prior density of ϕ and $\tau(=\sigma^{-2})$. The model he used for AR(1) is a two parameter model, namely,

$$Y(t) = \phi_1 + \phi_2 Y(t-1) + \varepsilon(t),\ t = 1,2,\ldots,n,\ Y(0) \text{ is fixed.}$$

The joint posterior distribution for ϕ_1 and ϕ_2 is a bivariate t-distribution, thus, the marginal posterior distributions for ϕ_1 and ϕ_2 are univariate t-distribution, and Y(n+1) is distributed as t with mean $\hat{\phi}_1 + \hat{\phi}_2 Y(n)$, where

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} = \begin{bmatrix} n & \sum_{t=1}^{n} Y(t-1) \\ \sum_{t=1}^{n} Y(t-1) & \sum_{t=1}^{n} Y^2(t-1) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{n} Y(t) \\ \sum_{t=1}^{n} Y(t-1)Y(t) \end{bmatrix} .$$

In his analysis, Zellner did not put the restriction $|\phi_2|<1$ on ϕ_2 to introduce stationarity. Following Thornber (1967), he analyzes the AR(1) model assuming that $|\phi_2|<1$, Y(0) is normally distributed, ϕ_1 and $\log \sigma$ are uniform, $p(\phi_2)$ (the prior distribution for ϕ_2) is chosen from the family of beta distribution defined on (-1,1). The posterior distribution of ϕ_2 must be analyzed numerically. For an AR(2) model

$$Y(t) = \phi_1 Y(t-1) + \phi_2 Y(t-2) + \varepsilon(t), \qquad t = 1,2,\ldots,n\ ,$$

Zellner assumes that Y(0) and Y(-1) are fixed and uses a vague prior for the parameters. The joint posterior of ϕ_1 and ϕ_2 is bivariate t-distribution.

Lahiff (1980) studies stationary AR(1) processes, where she assumes a beta-type prior for ϕ over (-1,1), namely,

$$P(\phi) = \frac{1}{B(p,q)} \frac{(\phi+1)^{p-1}\ (1-\phi)^{q-1}}{(2)^{p+q-1}}; \qquad p > 0,\ q > 0,\ -1 \le \phi \le 1.$$

However, since the prior does not combine well with the likelihood function of ϕ, she examined the posterior distribution of ϕ from a numerical point of view. Lahiff did not study the predictive density of future observations.

Land (1981) developed the prediction of k future observations using the Bayesian predictive density, however she did not assume a stationary model, and she used a vague prior density.

Litterman (1980) introduced forecasting in the Bayesian sense for vector autoregressions. More sophisticated procedures for estimating the parameters of AR processes are given by Harrison and Stevens (1976) and Bloomfield (1976).

The order of an AR process has been studied, and many different procedures were developed from a non-Bayesian point of view. Anderson

(1963) discussed the determination of the order of an AR process under the normality assumption explicitly as a multiple decision procedure, which takes the form of a sequence of tests of the model starting at the highest order and successively down to the lowest order. Akaike (1969) introduced a fitting procedure of univariate AR(p) models. In his procedure he uses the mean-square error of the one-step ahead prediction obtained by using the least squares estimates of the parameters of the model. The mean square error is called the final prediction error (FPE), and estimated from a sample by

$$FPE(p) = \frac{n+p}{n-p} [\hat{C}(0) - \sum_{j=1}^{p} \hat{a}_{pj}\hat{C}(j)],$$

where the mean of the sample is assumed to be 0, $\hat{C}(\ell) = \frac{1}{n} \sum_{i=1}^{n-\ell} Y(i+\ell)Y(i)$, and $\hat{a}_{pj}$'s are obtained by solving the Yule-Walker equations. The order of the model is estimated by the value $\hat{p}$ which gives a minimum for the FPE(p), p = 0,1,2,...,L, where L is some upper limit.

Akaike (1974) introduced a new estimation procedure to statistically identify the order of AR processes; the estimate is based on the minimum information theoretical criterion estimate (MAICE), which he introduced. The order is estimated by the value $\hat{p}$ which minimizes

$$AIC(p) = -2 \log \text{(maximum likelihood)} + 2p.$$

Parzen (1974) introduced another criterion for selecting the order of an AR process. Bhansali and Downham (1977) proposed a simultaneous use of modified AIC statistics

$$AIC_{\alpha}(p) = -2 \log \text{(maximum likelihood)} + \alpha p, \qquad \alpha > 0.$$

Akaike (1979) studied and checked the practicability of the Bhansali and

Downham suggestion, and he introduced a prior distribution for α, and suggested that α to be used for the definition of a minimum AIC_α procedure should be kept large only when the dominant values of the likelihood would be distributed over small p's.

Gary, et al. (1978), introduced another procedure for identification of AR processes, based on the D-statistic which they introduced.

Box and Jenkins (1970) identify the order of AR processes by comparing plots of the sampled autocorrelation and partial autocorrelation functions with theoretical ones.

A Bayesian procedure for identification of AR processes have been proposed by Diaz and Farah (1981) and will be investigated.

CHAPTER III

POSTERIOR AND PREDICTIVE ANALYSIS OF ALMOST STATIONARY AR(1) MODELS

Normal-Gamma Prior for (ϕ,τ)

Consider an autoregressive process of order p given by

$$\phi(B)Y(t) = \varepsilon(t), \qquad t = 1,2,\ldots,n,$$

where $\varepsilon(1), \varepsilon(2),\ldots, \varepsilon(n)$ are i.i.d. $N(0,\tau)$ and $\phi(B) = 1-\phi_1 B - \ldots - \phi_p B^p$, where $B^i Y(t) = Y(t-i)$ and $Y(0), Y(-1),\ldots, Y(-p+1)$ are known.

If $p = 1$, we have

$$Y(t) - \phi Y(t-1) = \varepsilon(t), \qquad t = 1,2,\ldots,n.$$

which is a two-parameter AR(1) process. If $|\phi| < 1$, the process is stationary; otherwise it is called non-stationary. The likelihood function of ϕ and τ is

$$\begin{aligned} L(S|\phi,\tau) &\propto \tau^{n/2} \operatorname{Exp} -\frac{\tau}{2}\{\sum_{t=1}^{n} [Y(t) - \phi Y(t-1)]^2\} \\ &\propto \tau^{n/2} \operatorname{Exp} -\frac{\tau}{2}[\sum_{t=1}^{n} Y^2(t) - 2\phi \sum_{t=1}^{n} Y(t)Y(t-1) + \phi^2 \sum_{t=1}^{n} Y^2(t-1)] \\ &\propto \tau^{n/2} e^{-\tau \sum_{t=1}^{n} Y^2(t)/2} e^{-\frac{\tau}{2}[\phi^2 \sum_{t=1}^{n} Y^2(t-1) - 2\phi \sum_{t=1}^{n} Y(t)Y(t-1)]}, \end{aligned} \tag{3.1}$$

for $\phi \in R$, $\tau > 0$, and $S = (Y(1),Y(2),\ldots,Y(n))$.

This suggests that the normal-gamma density

$$p(\phi,\tau|\mu,\xi,\alpha,\beta) \propto \tau^{\alpha-1}\, e^{-\beta\tau}\, \tau^{1/2}\, e^{-\frac{\tau\xi}{2}(\phi-\mu)^2}, \quad \phi\in R,\ \tau>0. \tag{3.2}$$

is conjugate to the likelihood function. From (3.2) the joint prior distribution of (ϕ,τ) is

$$p(\phi,\tau|\mu,\xi,\alpha,\beta) \propto \tau^{\frac{2\alpha+1}{2}-1}\, e^{-\frac{\tau}{2}[\xi(\phi-\mu)^2+2\beta]}, \quad \phi\in R,\ \tau>0.$$

The marginal prior distribution of ϕ is

$$p(\phi|\alpha,\beta,\xi,\mu) \propto [\xi(\phi-\mu)^2 + 2\beta]^{-\frac{2\alpha+1}{2}}$$

$$\propto \left[\frac{1}{1+\frac{\xi}{2\beta}(\phi-\mu)^2}\right]^{\frac{2\alpha+1}{2}}, \quad \phi\in R \tag{3.3}$$

which is a t-distribution with 2α d.f., location parameter μ, precision $\frac{\xi\alpha}{\beta}$, and variance $\frac{\beta}{\xi(\alpha-1)}$. The marginal prior distribution of τ is

$$p(\tau|\alpha,\beta,\xi,\mu) \propto \tau^{\frac{2\alpha+1}{2}-1} \int_{-\infty}^{\infty} e^{-\frac{\tau}{2}[\xi(\phi-\mu)^2+2\beta]}\, d\phi$$

$$\propto \tau^{\alpha-1}\, e^{-\tau\beta}, \quad \tau>0 \tag{3.4}$$

which is a gamma density with parameters α and β. Thus

$$E(\tau|\alpha,\beta) = \frac{\alpha}{\beta}, \quad \text{and} \quad Var(\tau|\alpha,\beta) = \frac{\alpha}{\beta^2}.$$

Combining (3.1) and (3.2) we have

$$p(\phi,\tau|S) \propto \tau^{\frac{2\alpha+n+1}{2}-1}\, e^{-\frac{\tau}{2}[\xi(\phi-\mu)^2 + \sum_{t=1}^{n}(Y(t)-\phi Y(t-1))^2 + 2\beta]}$$

$$\propto \tau^{\frac{2\alpha+n+1}{2}-1}\, e^{-\frac{\tau}{2}[\phi^2(\sum_{t=1}^{n}Y^2(t-1)+\xi) - 2\phi(\sum_{t=1}^{n}Y(t)Y(t-1)+\xi\mu)}$$

$$+ \xi\mu^2 + 2\beta + \sum_{t=1}^{n} Y^2(t)], \qquad \phi \in R, \ \tau > 0, \tag{3.5}$$

as the joint posterior distribution of (ϕ,τ). Consider the exponent in (3.5), and complete the square in ϕ:

$$\phi^2[\sum_{t=1}^{n} Y^2(t-1)+\xi] - 2\phi[\sum_{t=1}^{n} Y(t)Y(t-1)+\xi\mu] + \xi\mu^2 + 2\beta + \sum_{t=1}^{n} Y^2(t)$$

$$= A[\phi - A^{-1}B]^2 + C - B^2A^{-1};$$

where

$$A = [\sum_{t=1}^{n} Y^2(t-1)+\xi],$$

$$B = [\sum_{t=1}^{n} Y(t)Y(t-1)+\xi\mu],$$

and

$$C = \xi\mu^2 + 2\beta + \sum_{n=1}^{n} Y^2(t).$$

Thus (3.5) can be written as

$$p(\phi,\tau|S) \propto \tau^{\frac{n+2\alpha+1}{2}-1} \operatorname{Exp} -\frac{\tau}{2}[(\phi-A^{-1}B)^2A + C - B^2A^{-1}], \quad \phi \in R, \ \tau > 0. \tag{3.6}$$

The marginal posterior distribution of ϕ is

$$p(\phi|S) \propto [A(\phi - A^{-1}B)^2 + C - B^2A^{-1}]^{-\frac{(2\alpha+n+1)}{2}}$$

$$\propto \frac{1}{[1 + \frac{A}{C - B^2A^{-1}}(\phi - A^{-1}B)^2]^{\frac{2\alpha+n+1}{2}}}, \quad \phi \in R. \tag{3.7}$$

which is a t-distribution with $2\alpha+n$ d.f., location parameter $A^{-1}B$, precision $\frac{A(2\alpha+n)}{C-B^2A^{-1}}$, and variance $\frac{C-B^2A^{-1}}{A(2\alpha+n-2)}$. The marginal posterior distribution of τ is

$$p(\tau|S) \propto \tau^{\frac{2\alpha+n+1}{2}-1} e^{-\frac{\tau}{2}(C-B^2A^{-1})} \int_{-\infty}^{\infty} e^{-\frac{\tau}{2}A(\phi-BA^{-1})^2} d\phi$$

$$\propto \tau^{\frac{2\alpha+n}{2}-1} e^{-\frac{\tau}{2}(C-B^2A^{-1})}, \qquad \tau > 0 \tag{3.8}$$

which is a gamma density with parameters $\frac{2\alpha+n}{2}$ and $\frac{C-B^2A^{-1}}{2}$.

$$E(\tau|S) = \frac{2\alpha+n}{2(C-B^2A^{-1})},$$

and

$$V(\tau|S) = \frac{2\alpha+n}{2(C-B^2A^{-1})^2}.$$

The posterior analysis is straight forward, since we have standard distributions for the parameters.

From (3.3) ϕ is distributed a priori as t. For an AR(1) to be stationary, $|\phi|$ should be less than 1. We have two ways to impose a priori that $-1 < \phi < 1$. The first is to truncate the prior distribution of ϕ over the interval $(-1,1)$, i.e.,

$$P^*(\phi|\alpha,\beta,\xi,\mu) = k\,P(\phi|\alpha,\beta,\xi,\mu), \qquad \phi \in (-1,1)$$

where

$$k = [\int_{-1}^{1} P(\phi|\alpha,\beta,\xi,\mu)\, d\phi]^{-1},$$

in which case k will be a function of the prior parameters of ϕ, and

$P^*(\phi|\alpha,\beta,\xi,\mu)$ will not be in a standard form, so using P^* as prior for ϕ will result in posterior distributions that are not mathematically convenient, and numerical integration must be carried out, in other words using $P^*(\phi|\alpha,\beta,\xi,\mu)$ will result in a joint prior distribution of (ϕ,τ) that is not conjugate to the likelihood function of (ϕ,τ). The second, is to make $\phi\varepsilon(-1,1)$ with a certain probability, say, 95%, i.e.,

$$P\left[\left|\frac{\phi-\mu}{\sqrt{\frac{\beta}{\xi(\alpha-1)}}}\right| < t_{(.025,2\alpha)}\right] = .95 .$$

Thus,

$$P\left[\mu - t_{(.025,2\alpha)}\sqrt{\frac{\beta}{\xi(\alpha-1)}} < \phi < \mu + t_{(.025,2\alpha)}\sqrt{\frac{\beta}{\xi(\alpha-1)}}\right] = .95$$

so we can choose the prior parameters μ, β, ξ, μ, such that,

$$P[-1 < \phi < 1] = .95,$$

to achieve this we need to have

$$1 = \mu + t_{(.025,2\alpha)}\sqrt{\frac{\beta}{\xi(\alpha-1)}} .$$

Thus, if ξ is chosen as the parameter to be determined by the above relation, and the other prior parameters are given, then

$$\xi = \frac{\beta t^2_{(.025,2\alpha)}}{[1-\mu]^2(\alpha-1)} . \qquad (3.9)$$

To forecast using this framework, a one-step predictive probability density function is used and is defined by

$$P[\text{future} \mid \text{data}] = \int P(\text{future, parameters} \mid \text{data})\, d(\text{parameters})$$

$$= \quad P(\text{future} \mid \text{data, parameters})\, p(\text{parameters} \mid \text{data})\, d(\text{parameters}).$$

Then, the one-step ahead predictive probability density function of Y is

$$P[Y(n+1) \mid S] \propto \int p(Y(n+1) \mid \phi,\tau,S)\; p(\phi,\tau \mid S)\; d(\phi,\tau).$$

$$P[Y(n+1),\phi,\tau \mid S] \propto p(\phi,\tau \mid S)\; p(\phi,\tau \mid \alpha,\beta,\xi,\mu)\; p(Y(n+1) \mid \phi,\tau,S)$$

$$\propto \tau^{n/2}\, e^{-\frac{\tau}{2}\sum_{t=1}^{n}(Y(t)-\phi Y(t-1))^2}\; \tau^{\alpha-1}\, e^{-\tau\beta}\; \tau^{\frac{1}{2}}\, e^{-\frac{\tau\xi}{2}(\phi-\mu)^2}$$

$$\tau^{\frac{1}{2}}\, e^{-\frac{\tau}{2}(Y(n+1)-\phi Y(n))^2}$$

$$\propto \tau^{\frac{2\alpha+n+2}{2}-1}\, e^{-\frac{\tau}{2}\{\sum_{t=1}^{n}[Y(t)-\phi Y(t-1)]^2 + [Y(n+1)}$$

$$^{- \phi Y(n)]^2 + \xi(\phi-\mu)^2 + 2\beta\}}. \qquad (3.10)$$

Consider the exponent in (3.10), and complete the square in ϕ, i.e.,

$$\phi^2\Big[\sum_{t=1}^{n} Y^2(t-1) + Y^2(n) + \xi\Big] - 2\phi\Big[\sum_{t=1}^{n} Y(t)\cdot Y(t-1) + Y(n+1)\cdot Y(n) + \xi\mu\Big]$$

$$+ \sum_{t=1}^{n} Y^2(t) + Y^2(n+1) + \xi\mu^2 + 2\beta\,.$$

Letting

$$A^* = \sum_{t=1}^{n} Y^2(t-1) + Y^2(n) + \xi,$$

$$B^* = \sum_{t=1}^{n} Y(t)\; Y(t-1) + Y(n+1)\; Y(n) + \xi\mu$$

and

$$C^* = \sum_{t=1}^{n} Y^2(t) + Y^2(n+1) + \xi\mu^2 + 2\beta,$$

(3.10) can be written

$$P(Y(n+1),\phi,\tau|S) \propto \tau^{\frac{2\alpha+n+2}{2}-1} e^{-\frac{\tau}{2}[A^*\phi^2 - 2B^*\phi + C^*]}$$

$$\propto \tau^{\frac{2\alpha+n+2}{2}-1} e^{-\frac{\tau}{2}[A^*(\phi - \frac{B^*}{A^*})^2 + C^* - \frac{B^{*2}}{A^*}]}.$$

Thus

$$P(Y(n+1)|S) \propto \int_0^\infty \int_{-\infty}^\infty \tau^{\frac{2\alpha+n+2}{2}-1} e^{-\frac{\tau}{2}[A^*(\phi - \frac{B^*}{A^*})^2 + C^* - \frac{B^{*2}}{A^*}]} d\phi d\tau$$

$$\propto [C^* - \frac{B^{*2}}{A^*}]^{\frac{2\alpha+n+1}{2}}, \qquad Y(n+1) \in R. \tag{3.11}$$

But,

$$C^* - \frac{B^{*2}}{A^*} = \Sigma\, Y^2(t) + Y^2(n+1) + \xi\mu^2 + 2\beta - \frac{[\sum_{t=1}^{n} Y(t)\cdot Y(t-1) + Y(n+1)\cdot Y(n) + \xi\mu]^2}{[\sum_{t=1}^{n} Y^2(t-1) + Y^2(n) + \xi]}$$

$$= DY^2(n+1) - 2GY(n+1) + F$$

$$= D[Y(n+1) - \frac{G}{D}]^2 + F - \frac{G^2}{D},$$

where

$$D = 1 - \frac{Y^2(n)}{\sum_{t=1}^{n} Y^2(t-1) + Y^2(n) + \xi} ,$$

$$G = \frac{Y(n)[\sum_{t=1}^{n} Y(t)\cdot Y(t-1) + \xi\mu]}{\sum_{t=1}^{n} Y^2(t-1) + Y^2(n) + \xi} ,$$

and

$$F = 2\beta + \xi\mu^2 + \Sigma\, Y^2(t) - \frac{[\sum_{t=1}^{n} Y(t)\, Y(t-1) + \xi\mu]^2}{\sum_{t=1}^{n} Y^2(t-1) + Y^2(n) + \xi} .$$

Then (3.11) can be written as

$$P[Y(n+1)|S] \propto [1 + \left[\frac{D}{F - \frac{G}{D}}\right](Y(n+1) - \frac{G}{D})^2]^{-\frac{(2\alpha+n+1)}{2}} , \quad Y(n+1) \in R, \qquad (3.12)$$

which is t-distribution with $2\alpha+n$ d.f., location parameter G/D, precision $\frac{D(2\alpha+n)}{F - G^2/D}$, and variance $\frac{F - G^2 D}{D(2\alpha+n-2)}$.

For more than one-step ahead forecasts, say, two-steps ahead, we can find the joint predictive probability density function of Y(n+1) and Y(n+2), then our two-step estimate is the conditional mean of Y(n+2) given Y(n+1). To predict Y(n+2) we may either use the mean of the predictive distribution of Y(n+1) which is substituted for Y(n+1) in E[Y(n+2)|Y(n+1)] if we have observed only n observations, or we can update the forecast if Y(n+1) is observed, and use one-step ahead estimate for the n+1 observations as before.

The prior variance for ϕ depends on the prior parameters α, β, and

ξ, and from (3.9) ξ depends on α, β, μ, and $t^2_{(.025,2\alpha)}$. Then, as μ (prior mean of ϕ) becomes closer to 1 the prior variance of ϕ becomes smaller, which means that as μ approaches 1 we have to increase the precision of our prior information. To overcome this difficulty, one may perhaps use mixtures of conjugate prior distributions which will give one more flexibility in the prior analysis.

Mixtures of Normal-Gamma Prior Densities for (ϕ,τ)

When ϕ is a priori thought to be close to one of the endpoints of the interval (-1,1), and our confidence in this prior information is not as high as suggested by the prior parameters and the relation (3.9), then a mixture of two or more normal-gamma priors could be used to reflect the prior information about ϕ and τ and preserve the 95% belief that $\phi \varepsilon (-1,1)$.

Let the assumptions be the same as for the ordinary normal-gamma prior case, namely, that we are concerned with the AR(1) model

$$Y(t) = \phi Y(t-1) + \varepsilon(t), \qquad t = 1,2,\ldots,n,$$

where $\varepsilon(t)$ are i.i.d. $N(0,\tau)$. The likelihood function of ϕ and τ is the same as (3.1).

The prior used in this case is a mixture of conjugate prior distributions, that is

$$p(\phi,\tau) \propto \sum_{i=1}^{k} \pi_i \, \tau^{\frac{2\alpha+1}{2}-1} e^{-\frac{\tau}{2}[\xi_i(\phi-\mu_i)^2 + 2\beta]}, \tag{3.13}$$

where $\phi \in R$, $\tau > 0$, $\sum_{i=1}^{k} \pi_i = 1$, and k is the number of mixtures. The marginal prior distribution of ϕ is

$$p(\phi) \propto \sum_{i=1}^{k} \frac{\pi_i}{[1 + \frac{\xi_i}{2\beta}(\phi-\mu_i)^2]^{(2\alpha+1)/2}}, \qquad \phi \in R, \tag{3.14}$$

which is a mixture of k t-distribution, each with 2α d.f. Thus,

$$p(\phi) = \sum_{i=1}^{k} \frac{\pi_i}{\sum_{i=1}^{k} \pi_i k_i^{-1}} \frac{[\frac{\xi_i}{2\beta}]^{1/2}}{B(2\alpha,.5)} \frac{1}{[1 + \frac{\xi_i}{2\beta}(\phi-\mu_i)^2]^{(2\alpha+1)/2}}, \quad \phi \in R \tag{3.15}$$

where k_i is the normalizing factor for component i in the prior mixture. The prior mean of ϕ is

$$E(\phi) = \sum_{i=1}^{k} \frac{\pi_i k_i^{-1}}{[\sum_{i=1}^{k} \pi_i k_i^{-1}]} \mu_i,$$

and the prior variance of ϕ is

$$V(\phi) = E(\phi^2) - [E(\phi)]^2$$

$$= \sum_{i=1}^{k} \left[\frac{2\beta\pi_i k_i^{-1}}{\xi_i(2\alpha-2)(\sum_{i=1}^{k} \pi_i k_i^{-1})} + \frac{\pi_i k_i^{-1}}{\sum_{i=1}^{k} \pi_i k_i^{-1}} \right] - \left[\sum_{i=1}^{k} \frac{\pi_i k_i^{-1} \mu_i}{(\sum_{i=1}^{k} \pi_i k_i^{-1})} \right]^2 .$$

The marginal prior distribution of τ is

$$p(\tau) \propto \sum_{i=1}^{k} \int_{-\infty}^{\infty} \pi_i \, \tau^{\frac{2\alpha+1}{2} - 1} e^{-\frac{\tau}{2}[\xi_i(\phi-\mu_i)^2 + 2\beta]}$$

$$\propto \sum_{i=1}^{k} \pi_i \, \tau^{\alpha-1} e^{-\tau\beta}$$

$$\propto \tau^{\alpha-1} e^{-\tau\beta}$$

which is a gamma density with parameters α and β, and its moments are

$$E(\tau) = \frac{\alpha}{\beta}, \quad \text{and} \quad Var(\tau) = \frac{\alpha}{\beta^2}.$$

The joint posterior distribution of ϕ and τ is obtained by combining (3.1) and (3.13), thus

$$p(\phi,\tau|S) \propto \sum_{i=1}^{k} \pi_i \tau^{\frac{2\alpha+n+1}{2}-1} e^{-\frac{\tau}{2}[\xi_i(\phi-\mu_i)^2 + \sum_{t=1}^{n}(Y(t)-\phi Y(t-1))^2+2\beta]}$$

$$, \quad \phi \in R, \ \tau > 0,$$

consider the exponent, then completing the square with respect to ϕ we have

$$p(\phi,\tau|S) \propto \sum_{i=1}^{k} \pi_i \tau^{\frac{2\alpha+n+1}{2}-1} e^{-\frac{\tau}{2}[A_i(\phi-A_i^{-1}B_i)^2 + C_i - B_i^2 A_i^{-1}]}, \tag{3.16}$$

where, $\phi \in R$, $\tau > 0$,

$$A_i = \sum_{t=1}^{n} Y^2(t-1) + \xi_i,$$

$$B_i = \sum_{t=1}^{n} Y(t)Y(t-1) + \xi_i\mu_i,$$

and

$$C_i = \sum_{t=1}^{n} Y^2(t) + 2\beta + \xi_i\mu_i^2.$$

The marginal posterior distribution of ϕ is

$$p(\phi|S) \propto \sum_{i=1}^{k} \pi_i \left[\frac{1}{1 + \frac{A_i}{C_i - B_i^2 A_i^{-1}}(\phi - A_i^{-1}B_i)^2}\right]^{(2\alpha+n+1)/2}, \quad \phi \in R \tag{3.17}$$

which is a mixture of k t-distributions each with $2\alpha+n$ d.f. Thus,

$$p(\phi|S) = \sum_{i=1}^{k} \frac{\pi_i}{\sum_{i=1}^{k} \pi_i (k_i')^{-1}} \frac{\left[\frac{A_i}{C_i - B_i^2 A_i^{-1}}\right]^{\frac{1}{2}}}{B(2\alpha+n,\frac{1}{2})} \left[\frac{1}{1+\frac{A_i}{C_i - B_i^2 A_i^{-1}}(\phi - A_i^{-1} B_i)^2}\right]^{\frac{2\alpha+n+1}{2}} \tag{3.18}$$

where $\phi \in R$, k_i' is the normalizing factor of the i-th component of the posterior distribution mixture. The posterior mean of ϕ is

$$E(\phi|S) = \sum_{i=1}^{k} \frac{\pi_i (k_i')^{-1}}{\sum_{i=1}^{k} \pi_i (k_i')^{-1}} (A_i B_i^{-1}) ,$$

and the posterior variance of ϕ is

$$V(\phi|S) = \sum_{i=1}^{k} \left[\frac{\pi_i (k_i')^{-1}}{\sum_{i=1}^{k} \pi_i (k_i')^{-1}} \left\{ \frac{C_i - B_i^2 A_i^{-1}}{A_i (2\alpha+n-2)} + (A_i^{-1} B_i)^2 \right\} \right] - (E(\phi/S))^2 .$$

The marginal posterior distribution of τ is

$$p(\tau|S) \propto \sum_{i=1}^{k} \pi_i \tau^{\frac{2\alpha+n}{2} - 1} e^{-\frac{\tau}{2}[C_i - B_i^2 A_i^{-1}]} , \qquad \tau > 0, \tag{3.19}$$

which is a mixture of k gamma distributions with parameters $\frac{2\alpha+n}{2}$ and $(C_i - B_i^2 A_i^{-1})/2$, $i = 1,2,\ldots,k$. Thus,

$$p(\tau|S) = \sum_{i=1}^{k} \frac{\pi_i}{\sum_{i=1}^{k} \pi_i (k_i^*)^{-1}} \tau^{\frac{2\alpha+n}{2} - 1} e^{-\frac{\tau}{2}[C_i - B_i^2 A_i^{-1}]} , \tag{3.20}$$

where $\tau > 0$, k_i^* is the normalizing factor of the i-th component of the posterior distribution mixture. The posterior mean of τ is

$$E(\tau|S) = \sum_{i=1}^{k} \frac{\pi_i (k_i^*)^{-1}}{\sum_{i=1}^{k} \pi_i (k_i^*)^{-1}} \left(\frac{2\alpha+n}{C_i - B_i A_i^{-1}} \right),$$

and the posterior variance of τ is

$$V(\tau|S) = \sum_{i=1}^{k} \frac{\pi_i (k_i^*)^{-1}}{\sum_{i=1}^{k} \pi_i (k_i^*)^{-1}} \left[\frac{(2\alpha+n)(1+2\alpha+n)}{(C_i - B_i A_i^{-1})^2} \right] - (E(\tau/S))^2 .$$

As in the case of the conjugate prior, the posterior analysis is straight forward.

The one-step ahead predictive probability density function of Y is

$$p[Y(n+1)|S] = \int_0^{\infty} \int_{-\infty}^{\infty} p[Y(n+1),\phi,\tau/S]\,d\phi\,d\tau,$$

where

$$p[Y(n+1),\phi,\tau|S] \propto \sum_{i=1}^{k} \pi_i \, \tau^{\frac{2\alpha+n+2}{2}-1} e^{-\frac{\tau}{2}[\sum_{t=1}^{n} Y(t) - \phi Y(t-1)]^2}$$

$$e^{-\frac{\tau}{2}[(Y(n+1) - \phi Y(n))^2 + \xi_i(\phi-\mu_i)^2 + 2\beta]},$$

$$\text{for} \quad Y(n+1) \in R, \ \phi \in R, \ \tau > 0. \qquad (3.20a)$$

Consider the exponents and complete the square with respect to ϕ. Then (3.20a) may be expressed as

$$p[Y(n+1),\phi,\tau|S] \propto \sum_{i=1}^{k} \pi_i \, \tau^{\frac{2\alpha+n+2}{2}-1} e^{-\frac{\tau}{2}[A_i^*(\phi - \frac{B_i^*}{A_i^*})^2 + C_i^* - \frac{B_i^{*2}}{A_i^*}]},$$

where

$$A_i^* = \sum_{t=1}^{n} Y^2(t-1) + Y^2(n) + \xi_i,$$

$$B_i^* = \sum_{t=1}^{n} Y(t)Y(t-1) + Y(n+1)Y(n) + \xi_i\mu_i ,$$

and

$$C_i^* = \sum_{t=1}^{n} Y^2(t) + Y^2(n+1) + \xi_i\mu_i + 2\beta .$$

Thus

$$P[Y(n+1)|S] \propto \sum_{i=1}^{k} \pi_i \left[\frac{1}{C_i^* - B_i^{*2} A_i^{-1}} \right]^{\frac{2\alpha+n+1}{2}} , \qquad Y(n+1) \in R. \qquad (3.21)$$

Consider the value $C_i^*-B_i^{*2}A_i^{-1}$ and complete the square with respect to $Y(n+1)$, we can write (3.21) as

$$P[Y(n+1)|S] \propto \sum_{i=1}^{k} \pi_i \left[\frac{1}{1 + \left(\frac{A_i}{C_i - B_i^2 A_i^{-1}}\right)(Y(n+1) - B_i A_i^{-1})^2} \right]^{\frac{2\alpha+n+1}{2}} , \qquad (3.22)$$

$$Y(n+1) \in R.$$

where

$$A_i = 1 - \frac{Y^2(n)}{\sum_{t=1}^{n} Y^2(t-1) + Y^2(n) + \xi_i} ,$$

$$B_i = \frac{Y(n)\left[\sum_{t=1}^{n} Y(t)Y(t-1) + \xi_i\mu_i\right]}{\sum_{t=1}^{n} Y^2(t-1) + Y^2(n) + \xi_i}$$

and

$$C_i = 2\beta + \xi_i\mu_i^2 + \sum_{t=1}^{n} Y^2(t) - \frac{\left[\sum_{t=1}^{n} Y(t)Y(t-1) + \xi_i\mu_i\right]^2}{\sum_{t=1}^{n} Y^2(t-1) + Y^2(n) + \xi_i} .$$

From (3.22), $Y(n+1)$ is distributed as a mixture of k t-distributions each with $2\alpha+n$ d.f. Thus

$$p[Y(n+1)|S] = \sum_{i=1}^{k} \frac{\pi_i}{\sum_{i=1}^{k} \pi_i (k_i')^{-1}} \frac{\left(\frac{A_i}{C_i - B_i^2 A_i^{-1}}\right)^{\frac{1}{2}}}{B(2\alpha+n, \frac{1}{2})}$$

$$\left[\frac{1}{1 + \frac{A_i}{C_i - B_i^2 A_i^{-1}} (Y(n+1) - B_i A_i^{-1})^2}\right]^{\frac{2\alpha+n+1}{2}}, \qquad (3.23)$$

$$Y(n+1) \in R,$$

where k_i' is the normalizing factor for the i-th component in the mixture. The mean of $Y(n+1)$ is

$$E(Y(n+1)|S) = \sum_{i=1}^{k} \frac{\pi_i (k_i')^{-1}}{\sum_{i=1}^{k} \pi_i (k_i')^{-1}} (A_i^{-1} B_i),$$

and the variance of $Y(n+1)$ is

$$V(Y(n+1)|S) = \sum_{i=1}^{k} \left\{ \frac{\pi_i (k_i')^{-1}}{\sum_{i=1}^{k} \pi_i (k_i')^{-1}} \left[\frac{C_i - A_i^{-1} B_i^2}{A_i (2\alpha+n-2)} + (A_i^{-1} B_i)^2 \right] \right\} - [E(Y(n+1)|S]^2.$$

Experiments and Methodology of the Posterior Density for (ϕ,τ) and Predictive Density for $Y(n+1)$

Since the behavior for negative ϕ can be described from work with $\phi > 0$, the values of ϕ used to numerically study the posterior analysis of (ϕ,τ) and the one-step ahead predictive analysis, are all positive values.

For the ordinary normal-gamma conjugate prior for (ϕ,τ) series of length 25, 50, 100 and 750 were used. AR(1) series were generated with ϕ taking the values 0.0, .25, .50, .75 and .90. τ was taken to be one. The first observation Y(0) was always taken to be zero.

Normal-gamma distributions which concentrate most of the marginal prior probability of ϕ in (-1,1) were used as the prior distributions for ϕ and were specified by $\alpha = 10$, $\beta = 9$. μ was taken to be 0.0, .25, .5, .75 and .90 for each value of ϕ and the different series length, ξ for each prior was calculated using formula (3.9).

For every combination of n (length of AR(1) series), ϕ, and μ, the mean and variance for the prior and posterior distributions of ϕ and τ and the one-step ahead predictive density for Y(n+1) were calculated using the exact formula of the first section. The prior and the posterior distributions for ϕ and for τ were plotted on the same graph, for each of the above combinations. The predictive density for Y(n+1) was also plotted for each case. A summary of means and variances for each case, and for all the above combinations are tabulated in Tables I - V in Appendix A.

When using mixtures of normal-gamma densities of (ϕ,τ), series of length 25, 50, 100 and 200 were generated with ϕ = 0.0, 0.5 and 0.9, τ was one, and Y(0) was zero.

A mixture of three marginal prior distributions for ϕ was also used as the marginal prior distribution. The three prior means were -.5, 0.0, 0.5. The prior parameter α was taken to be 10 and $\beta = \alpha - 1$, and ξ_i for each component in the prior mixture was calculated using formula (3.9). Three cases were used to calculate ξ_i by (3.9); (1) for the first case ξ_i was fixed for all the three components by using $\mu = .5$ in (3.9) for μ_i, $i = 1,2,3$, i.e., the variance for each component in the marginal prior mixture for ϕ, was fixed for the smallest variance among the three, (2) for the second case, ξ_i was fixed with $\mu = 0$ for all μ_i in the mixture, i.e., the variance for each component in the mixture was fixed for the largest variance among the three, (3) and for the last case, ξ_i was calculated by (3.9) using μ_i for each component, i.e., the variance for each component is proportional to its mean. Note that our 95% prior belief that $\phi \varepsilon (-1,1)$ is preserved under the first and the third cases, but not under the second.

To evaluate the effect of increasing the prior variance for τ on the posterior analysis, α values of 2 and 1.1 were used respectively, with $\beta = \alpha - 1$, and ξ_i was fixed using $\mu_i = .5$, for $i = 1,2,3$.

As in the case of the ordinary normal-gamma density, for every combination of n and ϕ, the mean and variance for the prior and posterior distributions for ϕ, and for the one-step ahead predictive density for Y(n+1) were calculated using the exact formula of the second section, and the plots of the prior and posterior distributions for ϕ and for τ were provided on the same graph. The predictive density for Y(n+1) was also plotted for each combination and for each case. Tables V - X summarize the mean and variance for each of the above.

Computations were done on the IBM 370 at Oklahoma State University Computer Center. All computations were done with double precision. The main program and subroutines used to generate the random series, the AR(1) series, and to compute all the parameters and statistics for the priors, posteriors and predictive densities, were written in FORTRAN. Tables I - XXVIII are in Appendix A, and Figures 1 - 18 are in Appendix B.

Results and Conclusions

Parameter estimates for ϕ = 0.0, .25, .5, and .9; and for τ = 1 are given in Tables I - V, for the normal-gamma prior density for (ϕ,τ), and in Tables VI - X for the mixture of normal-gamma prior densities for (ϕ,τ) along with estimates for Y(n+1). The Tables show the anticipated reduction in the posterior variance for ϕ and τ, with large sample sizes. This effect is clearly shown in Figures 1 - 12 when the true value are ϕ = .5, τ = 1, n = 25 and 100, respectively. The Tables also show that for the same AR(1) series (i.e., for the same true value for ϕ) and the same series length, the posterior variance for ϕ decreases as the prior mean for ϕ increases in absolute value for the normal gamma prior cases. This indicates that as the prior mean of ϕ increases in absolute value, the prior information will have more influence on the posterior, and this is anticipated from (3.9), for as μ increases, ξ will increase and the prior variance for ϕ will be smaller.

The results in the tables also indicate that for the same prior mean and same series length, and for the ordinary and mixture cases of normal-gamma prior densities, the posterior variance for ϕ decreases as the true value of ϕ increases in absolute value. This is also anticipated since from time-series theory we know that the variance of the sampling-theory estimator for ϕ is a function of sample size and the true value of ϕ, and is approximately $\frac{1}{n}(1-\phi^2)$. For a discussion of this see Box and Jenkins (1970). In the Bayesian analysis, if the true value of ϕ is close to 1 in absolute value, the likelihood function will strongly influence the posterior distribution of ϕ, and there should be less effect from the prior distribution of ϕ. This is clearly shown in

Table V, where even with short series length the posterior estimates of the parameters were reasonably close to their true values.

Tables VI - X show that for the same true value for ϕ and the same series length, when the prior variance of τ increases the posterior variance of τ also increases.

Tables VI - VIII indicate that when the variance of each component in the prior mixture is fixed for the smallest variance among all components, the influence of the prior distribution on the posterior will be stronger than for the two other cases.

The posterior estimates for the parameters shifted toward the true value of these parameters as the sample size of the AR(1) series increases. It is also clearly demonstrated that when the true absolute value of ϕ is close to 1 the influence of the prior distribution is minimal, and a small sample size will be adequate, and yield reasonable estimate for the parameters. But, when the true value of ϕ is close to zero, then a very large sample is needed to obtain estimates with good qualities, especially when a marginal prior distribution of ϕ with an absolute mean close to 1 is used. To overcome this difficulty, a mixture of normal-gamma prior densities was used, and a comparison between results in Tables I and VI indicates that even though, with a mixture of prior distributions (which is not too informative), a better result is obtained.

Generally speaking, we can conclude that if the true absolute value of ϕ is close to one, then a normal-gamma prior distribution for (ϕ,τ) could be used with a reasonable sample size to get good posterior estimates for the parameters. If the absolute value of ϕ is close to zero, then the prior information will have a great influence on the properties of the posterior distributions and therefore, prior distributions must be carefully selected. A reasonable way to choose a prior is by taking the marginal prior distribution of to be a mixture of t-densities with suitable means and variances chosen to describe the prior information available. If strong prior belief indicates that ϕ may be one of few values with equal probability, then components with small equal variance and with means equal to these values, will be appropriate. Otherwise, the variances could be adjusted to reflect our prior belief.

Results of the predictive density for Y(n+1) indicate that its mean is highly influenced by the direction of the last observation Y(n), and its variance is affected by the sample size.

The graphs in Figures 5, 6, 11, 12, 17 and 18 are all for mixtures of three densities, and they should exhibit trimodal shapes. Due to small variances for the components of the posterior distribution of ϕ, the prior and posterior distribution of τ, and the predictive distribution of Y(n+1), and the inaccuracy in using a line printer for plotting, the plots appear to be of a unimodal shapes.

CHAPTER IV

THE IDENTIFICATION OF UNIVARIATE AR PROCESSES

Most published discussions of AR models from a Bayesian point of view are focused on parameter estimation rather than on order identification. Different "classical" approaches to AR identification are available, among those which are used is the Box and Jenkins (1970) method and the minimum information theoretic criterion estimate (MAICE) presented by Akaike (1974).

The Box and Jenkins method identifies the AR model by comparing graphs of the observed autocorrelation and partial autocorrelation functions with their known theoretical counter parts. This method requires a minimum sample size of 50 to identify the appropriate model satisfactorily, and it is quite subjective since one needs experience in matching observed and theoretical patterns.

The MAICE method identifies the order p of AR models by minimizing AIC. Under the assumption of Gaussian model the AIC statistics for the p-th order autoregressive model is defined by

$$AIC(p) = n \log [C(0) - \sum_{j=1}^{p} \phi_{pj} C(p)] + 2p,$$

where $C(\ell)$ is defined by

$$C(\ell) = \begin{cases} \frac{1}{n} \sum_{n=1}^{n-\ell} Y(n+\ell)Y(n), & (\ell \geq 0), \\ C(-\ell) & (\ell < 0). \end{cases}$$

The procedure chooses the order to be that p which minimizes AIC(p) for $p = 0,1,\ldots,L$, where L is a preassigned upper limit to the order.

The objective of this Chapter is to examine a Bayesian procedure proposed by Diaz and Farah (1981) for identifying the order of an AR model; the procedure is based on the posterior distribution of the order.

Posterior Analysis of the Order of an AR Processes

Let $Y(t)$, $t = 1,2,\ldots,n$ be n observations from an AR(p) process

$$\phi(B)Y(t) = \varepsilon(t), \qquad t = 1,2,\ldots,n \tag{4.1}$$

where $Y(t)$ is the observation at time t, the $\varepsilon(t)$ are i.i.d. $N(0,\tau)$ with $\tau > 0$ and

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p , \tag{4.2}$$

where $\phi_i \in R$, and $B^s Y(t) = Y(t-s)$. The ϕ_i are unknown real autoregressive parameters, and the process is stationary when the roots (in the complex plane) of $\phi(B) = 0$, lie outside the unit circle.

If $\{Y(t): t = 1,2,\ldots,n\}$ are available, then to estimate the order p of the process is known as the identification problem.

Once p has been identified, one may do the usual Bayesian analysis of an AR(p) model.

Let $\phi^{(p)} = (\phi_{p1}\phi_{p2},\ldots,\phi_{pp})'$ be the vector of autoregressive parameters associated with a p-th order model, for $p = 1,2,\ldots,k$, where k is the maximum order of the model, and let $\mu^{(p)}$ and $Q^{(p)}$ be the prior mean and precision matrix for $\phi^{(p)}$ given p. For a given p, assume that $\phi^{(p)}$ and τ have a normal-gamma prior distribution, namely

$$\xi[\phi^{(p)},\tau|\mu^{(p)},Q^{(p)},\alpha,\beta,p] \propto (2\pi)^{-p/2}\tau^{p/2}\, e^{-\frac{\tau}{2}(\phi^{(p)}-\mu^{(p)})'Q^{(p)}(\phi^{(p)}-\mu^{(p)})}$$

$$|Q^{(p)}|^{\frac{1}{2}}\ \tau^{\alpha-1}\ e^{-\tau\beta}, \qquad (4.3)$$

where $\phi^{(p)} \in R^p$, $\tau > 0$ and the $\mu^{(p)}$, $Q^{(p)}$, α and β are known parameters. Let $\xi(p)$ be the marginal prior mass function of p, $p = 1,2,\ldots,k$.

The likelihood function for $\phi^{(p)}$, τ, and p, based on n observations $Y(t)$, $t = 1,2,\ldots,n$ is

$$L[S|\phi^{(p)},\tau,p] \propto \tau^{n/2}\ e^{-\frac{\tau}{2}\sum_{t=1}^{n}(Y(t)\ -\ \sum_{j=1}^{p}\phi_{pj}Y(t-j))^2}, \qquad (4.4)$$

where $\phi^{(p)} \in R^p$, $\tau > 0$ and $p = 1,2,\ldots,k$, and $S = (Y(1),Y(2),\ldots,Y(N))$, thus the posterior density of $\phi^{(p)}$, τ and p is

$$f[\phi^{(p)},\tau,p|S] \propto (2\pi)^{-\frac{p}{2}}\ \tau^{\frac{n+2\alpha+p}{2}-1} \exp[\ -\tfrac{\tau}{2}\{\ (\phi^{(p)}-\mu^{(p)})'Q^{(p)}$$

$$(\phi^{(p)}-\mu^{(p)}) + \sum_{t=1}^{n}[Y(t)\ -\ \sum_{j=1}^{p}\phi_{pj}Y(t-j)]^2 + 2\beta\}]$$

$$|Q^{(p)}|^{\frac{1}{2}}\xi(p) \qquad (4.5)$$

for $\phi^{(p)} \in R^p$, $\tau > 0$ and $p = 1,2,\ldots,k$.

The likelihood function (4.4) can be written as

$$L[S|\phi^{(p)},\tau,p] \propto \tau^{n/2}\ \exp -\tfrac{\tau}{2}\{\ \phi'^{(p)}A(p,S)\phi^{(p)} - 2\phi'^{(p)}B(p,S) + C(S)\}, \qquad (4.6)$$

where A(p,S)

$$\begin{bmatrix} \sum_{t=1}^{n} Y^2(t-1) & \sum_{t=1}^{n} Y(t-1)Y(t-2) & \cdots & \sum_{t=1}^{n} Y(t-1)Y(t-p) \\ \sum_{t=1}^{n} Y(t-2)Y(t-1) & \sum_{t=1}^{n} Y^2(t-2) & \cdots & \sum_{t=1}^{n} Y(t-2)Y(t-p) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^{n} Y(t-p)Y(t-1) & \sum_{t=1}^{n} Y(t-p)Y(t-2) & \cdots & \sum_{t=1}^{n} Y^2(t-p) \end{bmatrix}_{p\times p},$$

$$B(p,S) = \begin{bmatrix} \sum_{t=1}^{n} Y(t)Y(t-1) \\ \sum_{t=1}^{n} Y(t)Y(t-2) \\ \vdots \\ \sum_{t=1}^{n} Y(t)Y(t-p) \end{bmatrix}_{p\times 1},$$

and

$$C(S) = \sum_{t=1}^{n} Y^2(t).$$

This implies that (4.5) may be written as

$$f(\phi^{(p)},p,\tau|S) \propto (2\pi)^{-p/2}\, \tau^{\frac{n+2\alpha+p}{2}-1} \exp[\,-\frac{\tau}{2}\{\,\phi'^{(p)}A^*(p,S)\phi^{(p)} - 2\phi'^{(p)}B^*(p,S) + C^*(p,S)\}]\xi(p)\;|Q^{(p)}|^{\frac{1}{2}}, \qquad (4.6)$$

where

$$A^*(p,S) = Q^{(p)} + A(p,S),$$

$$B^*(p,S) = Q^{(p)}\mu^{(p)} + B(p,S),$$

$$C^*(p,S) = \mu'^{(p)}Q^{(p)}\mu^{(p)} + C(S) + 2\beta.$$

Consider the exponent of (4.6), and complete the square with respect to $\phi^{(p)}$, we can write (4.6) as

$$f(\phi^{(p)},p,\tau|S) \propto (2\pi)^{-\frac{p}{2}}\tau^{\frac{n+2\alpha+p}{2}-1}\xi(p)\cdot|Q^{(p)}|^{\frac{1}{2}}\,\mathrm{Exp}$$

$$-\frac{\tau}{2}\{[\phi^{(p)}-A^{*-1}(p,S)B^*(p,S)]'A^*(p,S)[\phi^{(p)}-A^{*-1}(p,S)B^*(p,S)]$$

$$+ C^*(p,S)-B^{*\prime}(p,S)A^{*-1}(p,S)B^*(p,S)\}. \tag{4.7}$$

Integrating (4.7) with respect to $\phi^{(p)}$ and τ gives the posterior distribution of p as

$$\xi(p|S) \propto \frac{\xi(p)|Q^{(p)}|^{\frac{1}{2}}}{|A^*(p,S)|^{\frac{1}{2}}[C^*(p,S)-B^{*\prime}(p,S)A^{*-1}(p,S)B^*(p,S)]^{\frac{2\alpha+n}{2}}}, \tag{4.8}$$

$p = 1,2,\ldots,k.$

To compare the performance of this method with other well-known "classical" methods for identifying the order of an AR(p) process, it is more reasonable to use a vague prior density for $\phi^{(p)}$ and τ. If we take the vague prior density for $\phi^{(p)}$ and τ given p as

$$\xi(\phi^{(p)},\tau|p) \propto \frac{1}{\tau}, \quad \tau > 0, \tag{4.9}$$

thus the joint posterior distribution for $\phi^{(p)}$, τ, and p is

$$\xi(\phi^{(p)},\tau,p|S) \propto \tau^{\frac{n-2}{2}} \text{Exp}[-\frac{\tau}{2} \sum_{t=1}^{n} [Y(t) - \sum_{i=1}^{p} \phi_{p_i} Y(t-i)]^2]\xi(p), \qquad (4.10)$$

for $\phi^{(p)} \in R^p$, $\tau > 0$, and $p = 1,2,\ldots,k$. We can write (4.10) as follows

$$\xi(\phi^{(p)},\tau,p|S) \propto \tau^{\frac{n-2}{2}} \text{Exp}[-\frac{\tau}{2}\{ [\phi^{(p)}-A^{-1}(p,S)B(p,S)]'A(p,S)$$

$$[\phi^{(p)}-A^{-1}(p,S)B(p,S)] + C(S) - B'(p,S)A^{-1}(p,S)B(p,S)\}], \qquad (4.11)$$

$\phi^{(p)} \in R^p$, $\tau > 0$, and $p = 1,2,\ldots,k$.

Eliminating $\phi^{(p)}$, and τ from (4.11) will give the following posterior distribution for p

$$\xi(p|S) \propto \frac{\xi(p)(2\pi)^{\frac{p}{2}}}{|A(p,S)|^{\frac{1}{2}}[C(S)-B'(p,S)A^{-1}(p,S)B(p,S)]^{\frac{n-p}{2}}}, \qquad (4.12)$$

for $p = 1,2,\ldots,k$.

Numerical results for identifying the order of several different generated series of order 2 using the vague prior (4.9), and a uniform $\xi(p)$ with $k = 10$, and different sample sizes ($N = 200$ the largest) shows that $\hat{p}$ (the estimated order of the AR(p)) is always 1, where $\hat{p}$ is the mode of (4.12) for $p = 1,2,\ldots,10$. A preliminary investigation indicated that the dominant factor here is the power of the component in brackets in the denominator of (4.12), namely, $\frac{n-p}{2}$.

If the prior of $\phi^{(p)}$ and τ given p is taken to be

$$\xi(\phi^{(p)},\tau|p) \propto \tau^{\frac{p}{2}-1}, \qquad \tau > 0,\ \phi^{(p)} \in R^p, \tag{4.13}$$

then the posterior distribution of p is

$$\xi(p|S) \propto \frac{\xi(p)(2\pi)^{\frac{p}{2}}}{|A(p,S)|^{\frac{1}{2}}[C(S)-B'(p,S)A^{-1}(p,S)B(p,S)]^{\frac{n}{2}}}, \tag{4.14}$$

$p = 1,2,\ldots,k.$

Because of the $(2\pi)^{\frac{p}{2}}$, (4.14) may tend to favor a higher value for p, especially when the sample size is small. So, to remove this effect the prior of $\phi^{(p)}$ and τ given p is taken as

$$\xi(\phi^{(p)},\tau|p) \propto \tau^{\frac{p}{2}-1}(2\pi)^{-\frac{p}{2}}, \qquad \tau > 0,\ \phi^{(p)} \in R^p, \tag{4.15}$$

and the posterior distribution for p is

$$\xi(p|S) \propto \frac{\xi(p)}{|A(p,S)|^{\frac{1}{2}}[C(S)-B'(p,S)A^{-1}(p,S)B(p,S)]^{\frac{n}{2}}}, \tag{4.16}$$

$p = 1,2,\ldots,k.$

In either case the posterior distribution of the order p is not a known form, and must be normalized numerically. If $\xi(p)$ is uniform the formula simplifies.

To estimate p (the order of AR(p)) using $\xi(p|S)$ a measure of central tendency could be used, we will use the mode and the nearest integer to the mean of $\xi(p|S)$, and refer to the first method as Bayesian posterior mode and the second as Bayesian posterior mean.

Numerical Study for Sensitivity Analysis and Method Assessment

Our main concern in this part is to study the effect of changing the values of the hyperparameters, namely, $\mu^{(p)}$, $Q^{(p)}$, α, and β, and the sample size N on the identification of the true order of an AR(p) process, in the case where conjugate prior density is used. We are also concerned with how well these Bayesian procedures perform when compared to some well-known "classical" procedures, such as Akaike's identification procedures. Finally, we will try the Bayesian procedure on some real data.

For the sensitivity of the Bayesian identification method on the choice of the values of the hyper-parameters, the following three AR models are used

(1) $Y(t) = .65Y(t-1) + .3Y(t-2) + \varepsilon(t)$,

(2) $Y(t) = .35Y(t-1) - .35Y(t-2) + \varepsilon(t)$,

(3) $Y(t) = .55Y(t-1) + .05Y(t-2) + \varepsilon(t)$,

where model number 3 is taken from Akaike's (1979) paper, and $\varepsilon(t)$, $t=1, 2,\ldots,n$ are i.i.d. $N(0,\tau)$, and τ is taken to be one. Each of the above models is explored with series of length 30, 50, 60, 75, 100 and 200, and the maximum order tried (k) is 10.

Six prior μ vectors are used as follows:

(1) $\mu' = (.65, .3, .352, .282, .225, .18, .144, .115, .092, .074)$,

(2) $\mu' = (.35, -.35, .352, .282, .225, .18, .144, .115, .092, .074)$,

(3) $\mu' = (.55, .05, .352, .282, .225, .18, .144, .115, .092, .074)$,

(4) $\mu' = (.65, .3, -.352, -.282, -.225, -.18, -.144, -.115, -.092, -.074)$,

(5) $\mu' = (.35, -.35, -.352, -.282, -.225, -.18, -.144, -.115, -.092, -.074)$,

(6) $\mu' = (.55, .05, -.352, -.282, -.225, -.18, -.144, -.115, -.092, -.074)$.

We will refer to these vectors by their numbers in Table XI - Table XVI. The prior precision matrix Q is taken as $d*I$, where I is (10×10) identity matrix and d takes the values .01, 1 and 100. The prior mean and variance for τ are also changed through the choice of the values of α and β, where α takes the value 2 and β equals 1 and 10.

Table XI - Table XVI summarize all information about the estimated order by the Bayesian posterior mode method for each of the above 648 cases.

To compare the Bayesian procedures with some well-known "classical" procedures, the following four procedures by Akaike (1969, 1974, 1979) are chosen: the final prediction error (FPE) method, the information criterion (AIC) method, the information criterion 4 (AIC4) method, and the Akaike's mode method with prior density $\pi(\mu) = 1$, $0 < \mu < 1$, and $\mu = e^{-\frac{1}{2}(\alpha-2)}$.

Five-hundred simulation runs each with a sample size of 50 generated observations from the following models were made, using the Bayesian and "classical" procedures with k = 10.

(1) $Y(t) = .65Y(t-1) + .3Y(t-2) + \varepsilon(t)$,

(2) $Y(t) = .5Y(t-1) - .06Y(t-2) + .45Y(t-8) + \varepsilon(t)$,

(3) $Y(t) = .55Y(t-1) + .05Y(t-2) + \varepsilon(t)$,

(4) $Y(t) = .35Y(t-1) - .35Y(t-2) + \varepsilon(t)$

where model (3) and (2) are taken from Akaike (1979) paper, with an eighth order for model (2) instead of the original order of fifteen, and $\varepsilon(t)$, $t = 1,2,\ldots,n$ are i.i.d. $N(0,\tau)$, and τ is taken to be 1, with $\xi(p) \propto \tau^{\frac{p}{2}-1}$.

Five-hundred simulation runs were also made for the same prior for p but, for sample size of 60 and 70, using model (1) and model (2).

Table XVII - Table XXIV summarize the mean, variance, M.S.E., and the number of correct identifications out of the 500 runs, for each of the different procedures. The Bayesian posterior mean method was used in some of the runs to see how well it compares to the Bayesian posterior mode method, which was chosen initially as the Bayesian identification method.

The same number of simulation runs were also made using model (1) and model (2) for $\xi(p) \propto \tau^{(p/2)-1}(2\pi)^{-p/2}$, and with sample sizes of 50 and 70. Table XXV - Table XXVIII summarize the statistics of the above cases for the Bayesian and "classical procedures.

Finally, the Bayesian posterior mode procedure was applied to identify the classical series of Wolfer's sunspot numbers with N = 176 (Anderson, 1971), and to the series E and F given in Box and Jenkins (1970).

Results and Conclusions

Table XI - Table XVI show that the correct identification of the order of an AR model depends on the true model itself. If the last non-zero parameter in the process is close to zero, such as in the model $Y(t) = .55Y(t-1) + .05Y(t-2) + \varepsilon(t)$, then most likely it will be considered as zero in the data information, i.e., the true order is not likely to be correctly estimated, especially when the data information dominates the prior information. The use of prior mean vector for $\phi^{(p)}$ with correct or values close to the true values of these parameters, and with high prior precision, will yield a correct estimate of the true order of the process, if the prior information dominates, i.e., if N is small (N = 30 in Table XV and Table XVI).

For the other two models used in Table XI - Table XVI, namely, $Y(t) = .65Y(t-1) + .3Y(t-2) + \varepsilon(t)$, and $Y(Y) = .35Y(t-1) - .35Y(t-2) + \varepsilon(t)$, the results show that depending on the model used, a sample size of at least 50 observations is needed for the data to begin to overcome prior information (which is not that strong), but if the prior information is wrong, then a very large sample size may be required if the prior precision is high, and vice versa.

It is also clear from the tables that the choice of the prior vector μ will influence the estimated order of the model, especially for small sample sizes, and that the same could be said about the prior precision matrix.

By comparing the tables it seems that the estimated order of an

AR process is not very sensitive to the choice of values of α and β, but actually when β increased to 10 the probability of the posterior mode decreases, and a big change in the value of α or β may have greater influence on our estimate.

Table XXIII shows that the same problem is encountered in estimating the true order of an AR process correctly, when the last non-zero AR parameter is close to zero, as encountered in the conjugate prior for $\phi^{(p)}$ and τ case, when a non-informative prior density is used. The same is true for all the "classical" procedures used.

Tables XVII - Table XXIV show that when the vague prior $\xi(\phi^{(p)},\tau/p) \propto \tau^{(p/2)-1}$ is used, with n = 50 the two Bayesian procedures did not do a good job in terms of the mean, variance and M.S.E. of the estimated orders of the 500 simulation runs, when the true model is of low order compared to the maximum order tried (k), even though the Bayesian mode method did as well as Akaike's F.P.E. and AIC methods in terms of the number of the correct identifications out of the 500 simulation runs. This problem was anticipated because of the appearance of $(2\pi)^{p/2}$ in the posterior distribution of p, when the above prior is used. But when the sample size increased to 60 and 70 the Bayesian procedures improved considerably, especially the Bayesian posterior mode method.

When the true order of the AR process was high, and using the above prior, Table XX shows that the Bayesian mode procedure did considerably better than Akaike's procedures, and the Bayesian posterior mean procedure did a good job in identifying the true order of the process compared to the Akaike's methods, when the number of correct identifications is considered, even though its mean and variance were lower than Akaike's F.P.E. and AIC methods, and when the sample size was increased to 60 and then to 70 Akaike's F.P.E. and AIC methods improved considerably and were very close to the Bayesian posterior methods, which also improved as the sample size became larger.

Tables XXV - Table XVIII show that when the prior density for $\phi^{(p)}$ and τ was $\xi(\phi^{(p)},\tau/p) \propto \tau^{(p/2)-1}(2\pi)^{-(p/2)}$, then for n = 50 and the model Y(t) = .65Y(t-1) + .3Y(t-2) + ε(t), the Bayesian posterior mean did better than all other procedures, and the Bayesian posterior mode was also good. When n increased to 70 the two Bayesian procedures preserve their position. But when we used the model Y(t) = .5Y(t-1) - .06Y(t-2) + .45Y(t-8) + ε(t), with the above prior, the two Bayesian procedures did very poor job in estimating the true value of the order.

The results of the simulation studies shown in Table XVII - XXVIII also indicate that Akaike's F.P.E. and AIC procedures were almost the same in the ability of estimating the order of an AR process, and Akaike's AIC4 and mode procedures were also close to each other when the true order of an AR process was low. When the order was high then Akaike's AIC4 procedure was the poorest among the Akaike's procedures, even though the Akaike's mode procedure was also considerably weaker than the F.P.E. and the AIC procedures.

The simulation results for the Bayesian procedures also show the difficulty in choosing one vague prior density, to best serve in comparing a Bayesian procedure with "classical" ones, where the Bayesian posterior mode did very well with one prior and the Bayesian posterior mean did better in the other. The results also show that the Bayesian posterior mode's ability to correctly estimate the true order of an AR process when $\xi(\phi^{(p)},\tau|p) \propto \tau^{(p/2)-1}$ was used, was very close to

Akaike's F.P.E. and AIC procedures if we consider the number of correct identifications, even though it tends to have higher means and variances when the true order was low and the sample size was small. When the sample size get larger these methods will be very close to each other in their ability to estimate the correct order. The Bayesian posterior mean was the worst among the methods tried when the above prior is used and the true order was low, but when the true order was high this method was comparable with the Bayesian mode, F.P.E. and AIC methods. This property was reversed when $\xi(\phi^{(p)},\tau/p) \propto \tau^{(p/2)-1}(2\pi)^{-(p/2)}$ was used.

From all the results we conclude that if informative prior information about the order and the values of the AR parameters were available, then expressing them in a conjugate prior density, and using them in the posterior analysis of the order of an AR process would be of benefit to us, especially if the sample size is not large enough to yield reliable estimates. If the prior information is not informative or vague, then either a conjugate prior with a very low prior precision could be used, or one of the above priors, and based on the simulations results the Bayesian posterior mode seem to be preferable, especially with the prior $\xi(\phi^{(p)},\tau|p) \propto \tau^{(p/2)-1}$, and a sample size which is not considerably small for the identification problem.

Using the vague prior density $\xi(\phi^{(p)},\tau|p) \propto \tau^{(p/2)-1}$, the Bayesian posterior mode estimate for series E was the second order model, and this agrees with the second-or-third order AR model suggested by Box and Jenkins for this series, which is a part of the Wolfer's sunspot numbers series with $n = 100$. For series F with $n = 70$, the Bayesian posterior mode estimate for the order was 2, which also agrees with the suggestion made by the authors.

For the whole classical series of Wolfer's sunspot numbers, with $n = 176$, the Bayesian posterior mode estimate was the third-order model.

CHAPTER V

THE IDENTIFICATION OF MULTIVARIATE AR PROCESSES

Let $Y(t)$, $t = 1,2,\ldots,n$ be n observations from an $AR_m(p)$ process.

$$Y'(t) = \sum_{j=1}^{p} Y'(t-j)\theta_{pj} + \varepsilon'(t), \qquad t = 1,2,\ldots,n, \tag{5.1}$$

where $Y(t)$ is the $(m\times 1)$ observed vector at time t, $Y(t-j)$, $j = 1,2,\ldots,p$ are previous vectors of observation each of dimension $(m\times 1)$, θ_{pj} is $(m\times m)$ matrix of real unknown autoregressive parameters, $j = 1,2,\ldots,p$, and $\varepsilon(t)$ are i.i.d. $N_m(0,T)$, where T is positive definite symmetric matrix of precision. Model (5.1) is known as a vector or multivariate autoregressive process of order p ($AR_m(p)$) if θ_{pp} is not the zero $(m\times m)$ matrix.

Suppose $\{Y(t)\colon t = 1,2,\ldots,n\}$ is a set of observations, and if $Y(0)$, $Y(-1),\ldots,Y(1-p)$ are known vectors, and if we let

$$Y = \begin{bmatrix} Y'(1) \\ Y'(2) \\ \vdots \\ Y'(n) \end{bmatrix}, \quad X = \begin{bmatrix} Y'(0) & Y'(-1) & \cdots & Y'(1-p) \\ Y'(1) & Y'(0) & \cdots & Y'(2-p) \\ & & & \\ Y'(n-1) & Y'(n-2) & \cdots & Y'(n-p) \end{bmatrix},$$

$$\theta = \begin{bmatrix} \theta_{p1} \\ \theta_{p2} \\ \cdot \\ \cdot \\ \cdot \\ \theta_{pp} \end{bmatrix}, \quad E = \begin{bmatrix} \varepsilon'(1) \\ \varepsilon'(2) \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon'(n) \end{bmatrix},$$

where Y is an (n×m) matrix, X is an (n×mp) matrix, θ is an (mp×m) matrix, and E is an (n×m) matrix, then we can write our model as

$$Y = X\theta + E, \tag{5.2}$$

which is a special case of a multivariate regression model.

The likelihood function for θ, T, and p, based on the observations $\{Y(t),\ t = 1,2,\ldots,n\}$ is

$$L(Y|\theta,T,p) \propto |T|^{\frac{n}{2}} \operatorname{Exp}\{-\frac{1}{2}\operatorname{tr}[(Y-X\theta)'(Y-X\theta)]\}$$

$$\propto |T|^{\frac{n}{2}} \operatorname{Exp}\{-\frac{1}{2}\operatorname{tr}[S + (\theta-\hat{\theta})'X'X(\theta-\hat{\theta})T]\}, \tag{5.3}$$

where X is a full-rank matrix,

$$\hat{\theta} = (X'X)^{-1}X'Y ,$$

and

$$S = (Y-X\hat{\theta})'(Y-X\hat{\theta}).$$

To use a conjugate prior density, the form of the likelihood function suggests a normal-wishart distribution, namely

$$f(\theta,T) \propto (2\pi)^{-\frac{mp}{2}} |D|^{\frac{1}{2}} |T|^{\frac{mp+\alpha}{2}-1} \operatorname{Exp}\{-\frac{1}{2}\operatorname{tr}[(\theta-A)'D(\theta-A)T]$$

$$+ \frac{1}{2}\operatorname{tr}(\beta T)\}, \tag{5.4}$$

where

A is (mp×m) prior mean matrix,

D is (mp×mp) positive definite symmetric (PDS) matrix,

β is (m×m) PDS matrix,

and $\alpha > 0$.

Assuming that the order p has a prior density $\xi(p)$, $p = 1,2,\ldots,k$, the joint posterior distribution of θ, T, and p is

$$\xi(\theta,T,p|Y) \propto \xi(p)(2\pi)^{-\frac{mp}{2}} |D|^{\frac{1}{2}} |T|^{\frac{n+mp+\alpha}{2}-1} \text{Exp}\{-\frac{1}{2}\text{tr}[(\theta-A)'D(\theta-A)T$$

$$+ (\theta-\hat{\theta})'X'X(\theta-\hat{\theta})T + ST + \beta T]\}$$

$$\propto \xi(p)(2\pi)^{-\frac{mp}{2}} |D|^{\frac{1}{2}} |T|^{\frac{n+mp+\alpha}{2}-1} \text{Exp}\{-\frac{1}{2}\text{tr}[\theta'D\theta-\theta'DA-A'D\theta$$

$$+ A'DA+\theta'X'X\theta-\theta'X'X\hat{\theta}-\hat{\theta}'X'X\theta+\hat{\theta}'X'X\hat{\theta}+S+\beta]T\}$$

$$\propto \xi(p)(2\pi)^{-\frac{mp}{2}} |D|^{\frac{1}{2}} |T|^{\frac{n+mp+\alpha}{2}-1} \text{Exp}\{-\frac{1}{2}\text{tr}[\theta'(D+X'X)\theta$$

$$-\theta'(DA+X'X\hat{\theta})-(A'D+\hat{\theta}X'X)\theta+\hat{\theta}'X'X\hat{\theta}+S+\beta]T\}$$

$$\propto \xi(p)(2\pi)^{-\frac{mp}{2}} |D|^{\frac{1}{2}} |T|^{\frac{n+mp+\alpha}{2}-1} \text{Exp}\{-\frac{1}{2}\text{tr}[\theta'(D+X'X)\theta$$

$$-\theta'(DA+X'Y)-(DA+X'Y)'\theta+Y'Y+\beta]T\}. \tag{5.5}$$

Let

$$E_1 = D + X'X,$$

and

$$E_2 = X'Y + DA.$$

Then we can write (5.5) as

$$\xi(\theta,T,p|Y) \propto \xi(p)(2\pi)^{-\frac{mp}{2}} |D|^{\frac{1}{2}} |T|^{\frac{n+mp+\alpha}{2}-1} \text{Exp}\{-\frac{1}{2}\text{tr}[\theta'E_1\theta-\theta'E_2$$

$$-E_2'\theta+Y'Y+\beta]T\}. \qquad (5.6)$$

Consider the exponent of (5.6), and complete the square with respect to θ, we can write (5.6) as

$$\xi(\theta,T,p|Y) \propto \xi(p)(2\pi)^{-\frac{mp}{2}} |D|^{\frac{1}{2}} |T|^{\frac{n+mp+\alpha}{2}-1} \text{Exp}\{-\frac{1}{2}\text{tr}[(\theta-E_1^{-1}E_2)'\cdot$$

$$E_1(\theta-E_1^{-1}E_2)+E_2'E_1^{-1}E_2+Y'Y+\beta]T\}. \qquad (5.7)$$

Let

$$E_3 = Y'Y + \beta + E_2'E_1^{-1}E_2 = Y'Y + \beta + (X'Y+DA)'(D+X'X)^{-1}(X'Y+DA),$$

and

$$\tilde{\theta} = E_1^{-1}E_2 = (D+X'X)^{-1}(X'Y+DA).$$

Then (5.7) is simplified to

$$\xi(\theta,T,p|y) \propto \xi(p)(2\pi)^{-\frac{mp}{2}} |D|^{\frac{1}{2}} |T|^{\frac{n+mp+\alpha}{2}-1} \operatorname{Exp}\{-\tfrac{1}{2}\operatorname{tr}[(\theta-\tilde{\theta})'E_1 \cdot$$

$$(\theta-\tilde{\theta})+E_3]T\}. \tag{5.8}$$

Now, (5.8) can be written as follows

$$\xi(\theta,T,p|Y) \propto \xi(p)(2\pi)^{-\frac{mp}{2}} |D|^{\frac{1}{2}} |T|^{\frac{n+mp+\alpha}{2}-1} \operatorname{Exp}\{-\tfrac{1}{2}(\theta^*-\tilde{\theta}^*)'T \otimes$$

$$E_1(\theta^*-\tilde{\theta}^*)\}, \tag{5.9}$$

where

$$\theta^{*\prime} = (\theta'_{p1}, \theta'_{p2}, \ldots, \theta'_{pp}),$$

$$\tilde{\theta}^{*\prime} = (\tilde{\theta}'_{p1}, \tilde{\theta}'_{p2}, \ldots, \tilde{\theta}'_{pp}),$$

and $\otimes$ denotes Kronecker or direct matrix multiplication.

Integrating out θ^* from (5.9), using the properties of the multivariate normal distribution, we have the joint posterior distribution of T, and p as

$$\xi(T,p|Y) \propto \xi(p)\ |D|^{\frac{1}{2}} |T|^{\frac{n+mp+\alpha}{2}-1} |E_1 \otimes T|^{-\frac{1}{2}} \operatorname{Exp}[-\tfrac{1}{2}\operatorname{tr}(E_3T)]$$

$$\propto \xi(p)\ |D|^{\frac{1}{2}} |E_1|^{-\frac{m}{2}} |T|^{\frac{n+mp+\alpha}{2}-1} |T|^{-\frac{mp}{2}} \operatorname{Exp}[-\tfrac{1}{2}\operatorname{tr}(E_3T)]$$

$$\propto \xi(p)\ |D|^{\frac{1}{2}} |E_1|^{-\frac{m}{2}} |T|^{\frac{n+\alpha}{2}-1} \operatorname{Exp}[-\tfrac{1}{2}\operatorname{tr}(E_3T)]. \tag{5.10}$$

Integrating T from (5.10), using the properties of a Wishart distribution, we have the posterior distribution of the order as

$$\xi(p|Y) \propto \frac{\xi(p)|D|^{\frac{1}{2}}}{|E_1|^{\frac{m}{2}} [E_3]^{\frac{n+\alpha+(m+1)}{2}}}, \qquad p = 1,2,\ldots,k,$$

$$\xi(p|Y) \propto \frac{\xi(p)|D|^{\frac{1}{2}}}{|D+X'X|^{\frac{m}{2}} [Y'Y+\beta+(X'Y+DA)'(D+X'X)^{-1}(X'Y+DA)]^{\frac{n+\alpha+m-1}{2}}}, \quad (5.11)$$

$p = 1,2,\ldots,k,$

where k is the maximum order of the model. If m = 1, then (5.11) will reduce to the posterior distribution of p for the univariate case obtained in Chapter IV, namely (4.8), with $D = Q^{(p)}$, $D+X'X = A^*(p,S)$, and α and β will be 2α, and 2β for the univariate case.

If we use a vague prior density, then depending on this density the posterior distribution of p will change. If the vague prior of θ, T, given p is

$$f_{\tilde{}}(\theta,T|p) \propto |T|^{-\frac{(m+1)}{2}}, \quad (5.12)$$

then the posterior distribution of p is

$$\xi(p|Y) \propto \frac{\xi(p)(2\pi)^{\frac{mp}{2}}}{|X'X|^{\frac{m}{2}} [Y'Y-Y'X(X'X)^{-1}X'Y]^{\frac{n-mp}{2}}}, \quad (5.13)$$

$p = 1,2,\ldots,k.$

We may encounter the same problem in using this vague prior as was encountered for the univariate AR(p), namely, the estimated order will always be 1.

If we use the general form for the vague prior we used for the univariate AR(p) case, i.e.,

$$f_2(\theta,T|p) \propto |T|^{\frac{mp}{2}-1}, \tag{5.14}$$

then the posterior distribution of p will have the following form

$$\xi(p|Y) \propto \frac{\xi(p)(2\pi)^{\frac{mp}{2}}}{|X'X|^{\frac{m}{2}}\,[Y'Y-Y'X(X'X)^{-1}X'Y]^{\frac{n+m-1}{2}}}, \tag{5.15}$$

$p = 1,2,\ldots,k,$

and this will favor higher order because of the appearance of $(2\pi)^{\frac{mp}{2}}$, and may be biased upward, especially for small sample size, as was the situation for the univariate case.

If the following vague prior density is used

$$f_3(\theta,T|p) \propto |T|^{\frac{mp}{2}-1}(2\pi)^{-\frac{mp}{2}}, \tag{5.16}$$

then the posterior distribution of p is

$$\xi(p|Y) \propto \frac{\xi(p)}{|X'X|^{\frac{m}{2}}\,[Y'Y-Y'X(X'X)^{-1}X'Y]^{\frac{n+m-1}{2}}}, \tag{5.17}$$

$p = 1,2,\ldots,k.$

In either case the posterior distribution of the order p must be normalized numerically, and if $\xi(p)$ is uniform, then the formula simplifies, and as for the identification of the univariate AR(p) process, the posterior estimate of p, i.e., $\hat{p}$ could be taken as the mode or the nearest integer to the mean of the posterior distribution of the order.

Further Work and Assessment

Further study will be done later in a separate work to assess the performance of these methods for estimating the order of a multivariate autoregressive process, and the following objectives are included:

1. Design a numerical study to assess the performance of the posterior distribution of the order of an AR process in selecting the order of a multivariate autoregressive process.
2. Study the sensitivity of these methods to the choice of the prior parameters.
3. Compare these methods to well-known identification procedures such as the AIC(p) and FPE(p) of Akaike (1969, 1974).
4. Examine other statistical models, such as linear dynamic systems, where identification is an important problem.
5. Derive an identification technique for MA and ARMA models.

A SELECTED BIBLIOGRAPHY

Akaike, H. "Fitting Autoregressive Models for Prediction." *Ann. Inst. Statist. Math.*, Vol. 21 (1969), 243-247.

Akaike, H. "A New Look at the Statistical Model Identification." *IEEE Trans. Autom. Control* AC-19 (1974), 716-723.

Akaike, H. "A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting." *Biometrika*, Vol. 66, 2 (1979), 237-242.

Anderson, T. W. "Determination of the Order of Dependence in Normally Distributed Time Series." *Time Series Analysis*, Rosenblatt, M. (Ed.). New York: Wiley, 1963, 425-446.

Anderson, T. W. *The Statistical Analysis of Time Series*. New York: Wiley, 1971.

Bayes, Rev. R. "Studies in the History of Probability and Statistics". IX. Thomas Bayes Essay "Toward Solving a Problem in the Doctrine of Chances". G. A. Barnard (Ed.). *Biometrika*, Vol. 45 (1958), 293-315.

Bhansali, R. J. and D. Y. Downham. "Some properties of the order of an autoregressive model selected by a generalization of Akaike's EPE Criterion." *Biometrika*, Vol. 64 (1977), 547-551.

Bloomfield, P. *Forier Analysis of Time Series: An Introduction*. New York: John Wiley and Sons, 1976.

Box, G. E. P. and G. M. Jenkins. *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day, 1970.

Box, G. E. P. and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. Reading, Ma.: Addison-Wesley, 1973.

Chatfield, Christopher. *The Analysis of Time Series: Theory and Practice*. London: Chapman and Hall, 1975.

DeGroot, M. H. *Optimal Statistical Decisions*. New York: McGraw-Hill Book Company, 1970.

Diaz, Jaoquin and Jose L. Farah. "On the identification of autoregressive processes." Paper presented at the NBER-NSF Seminar on Bayesian Inference in Econometrics, May, 1981, Cornell University, Ithaca, New York.

Gray, H. L., G. D. Kelley and D. D. McIntire. "A New Approach to ARMA Modeling." Commun. Statist.-Simula. Computa, Vol. B7(1) (1978) 1-77.

Harrison, P. J. and C. F. Stevens. "Bayesian Forecasting (with Discussion)." Journal of the Royal Statistical Society, Series B, Vol. 38 (1976), 205-247.

Kendall, M. G. Time Series. London: Charles Griffin and Company, Limited, 1973.

Lahiff, Maureen. "Time-Series Forecasting with informative prior distributions." Technical Report No. 111, Department of Statistics, The University of Chicago, May, 1980.

Land, Margaret Foster. "Bayesian Forecasting for Switching Regression and Autoregressive Processes." (Ph.D. Dissertation, Oklahoma State University, Stillwater, Oklahoma, 1981.)

Lindley, D. V. Bayesian Statistics, A Review. Philadelphia, Pa: Society for Industrial and Applied Mathematics, 1972.

Litterman, Robert B. "A Bayesian Procedure for Forecasting with Vector Autoregressive." (Unpublished paper, Massachusetts Institute of Technology, Boston, Massachusetts, 1980), 36 pp.

Parzen, E. Manuel. "Some Recent Advances in Time Series Modeling." IEEE Trans. Autom. Control AC-19 (1974), 723-730.

Raiffa, H. and R. Schlaifer. Applied Statistical Decision Theory. Cambridge: Harvard University Press, 1961.

Zellner, Arnold. An Introduction to Bayesian Inference in Econometrics. New York, New York: John Wiley and Sons, 1971.

APPENDICES

APPENDIX A

TABLES

TABLE I

AR(1), $Y_t=\varepsilon_t, \varepsilon_t \sim$ i.i.d. N(0,1), $y_0=0$, N-G Prior $\alpha=10, \beta=9$

μ	n	Prior Information				Posterior Information				Predictive Information	
		E(ϕ)	V(ϕ)	E(τ)	V(τ)	E(ϕ/S)	V(ϕ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
0.00	25	0	.2298	1.1111	.1235	.1020	.0326	.8548	.0325	.1373	1.2834
	50	0	.2298	1.1111	.1235	.1805	.0180	.9211	.0242	-.2759	1.1596
	100	0	.2298	1.1111	.1235	.1425	.0094	.9948	.0165	.7584	1.0250
	750	0	.2298	1.1111	.1235	.0418	.0013	1.0216	.0027	.8453	.9868
0.25	25	.25	.1293	1.1111	.1235	.1408	.0299	.8538	.0324	.1895	1.2800
	50	.25	.1293	1.1111	.1235	.2008	.0171	.9227	.0243	-.3068	1.1554
	100	.25	.1293	1.1111	.1235	.1554	.0091	.9949	.0165	.8273	1.0247
	750	.25	.1293	1.1111	.1235	.0442	.0013	1.0212	.0027	.0894	.9872
0.50	25	.5	.0575	1.1111	.1235	.2476	.0249	.8290	.0305	.3334	1.3076
	50	.5	.0575	1.1111	.1235	.2651	.0151	.9080	.0236	-.4051	1.1690
	100	.5	.0575	1.1111	.1235	.1986	.0085	.9805	.0160	.1057	1.0396
	750	.5	.0575	1.1111	.1235	.0527	.0013	1.0168	.0027	.1065	.9914
0.75	25	.75	.0144	1.1111	.1235	.5450	.0139	.7290	.0237	.7339	1.4591
	50	.75	.0144	1.1111	.1235	.4982	.0099	.8179	.0191	-.7613	1.2817
	100	.75	.0144	1.1111	.1235	.3889	.0066	.8845	.0130	.2070	1.1516
	750	.75	.0144	1.1111	.1235	.1034	.0013	.9802	.0025	.2093	1.0281
0.90	25	.9	.0023	1.1111	.1235	.8443	.0036	.6285	.0176	1.1370	1.6715
	50	.9	.0023	1.1111	.1235	.8174	.0030	.6916	.0137	-1.2492	1.4955
	100	.9	.0023	1.1111	.1235	.7543	.0026	.7136	.0085	.4015	1.4258
	750	.9	.0023	1.1111	.1235	.3617	.0011	.8065	.0017	.7318	1.2474

TABLE II

AR(1), $Y_t = .25Y_{t-1} + \varepsilon_t ; \varepsilon_t \sim$ i.i.d. N(0,1), $y_0 = 0$,
N-G Prior $\alpha = 10, \beta = 9$

μ	n	Prior Information				Posterior Information				Predictive Information	
		E(ϕ)	V(ϕ)	E(τ)	V(τ)	E(ϕ/S)	V(ϕ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
0.00	25	0	.2298	1.1111	.1235	.3004	.0299	.8458	.0318	.4189	1.2954
	50	0	.2298	1.1111	.1235	.3892	.0156	.9104	.0237	-.5178	1.1583
	100	0	.2298	1.1111	.1235	.3684	.0083	.9887	.0163	.2017	1.0311
	750	0	.2298	1.1111	.1235	.2803	.0012	1.0205	.0027	.4964	.9863
0.25	25	.25	.1293	1.1111	.1235	.3209	.0275	.8521	.0323	.4475	1.2815
	50	.25	.1293	1.1111	.1235	.3973	.0148	.9166	.0240	-.5286	1.1493
	100	.25	.1293	1.1111	.1235	.3738	.0080	.9927	.0164	.2047	1.0269
	750	.25	.1293	1.1111	.1235	.2815	.0012	1.0210	.0027	.4986	.9859
0.50	25	.5	.0575	1.1111	.1235	.3883	.0227	.8477	.0319	.5414	1.2786
	50	.5	.0575	1.1111	.1235	.4315	.0131	.9175	.0241	-.5741	1.1451
	100	.5	.0575	1.1111	.1235	.3968	.0075	.9920	.0164	.2172	1.0274
	750	.5	.0575	1.1111	.1235	.2865	.0012	1.0199	.0027	.5074	.9869
0.75	25	.75	.0144	1.1111	.1235	.6063	.0124	.7908	.0278	.8453	1.3476
	50	.75	.0144	1.1111	.1235	.5839	.0086	.8743	.0218	-.7768	1.1926
	100	.75	.0144	1.1111	.1235	.5170	.0057	.9469	.0149	.2831	1.1057
	750	.75	.0144	1.1111	.1235	.3194	.0012	1.0023	.0026	.5658	1.0039
0.90	25	.9	.0023	1.1111	.1235	.8558	.0031	.7074	.0222	1.1932	1.4855
	50	.9	.0023	1.1111	.1235	.8343	.0026	.7770	.0173	-1.1099	1.3295
	100	.9	.0023	1.1111	.1235	.7880	.0023	.8219	.0113	.4314	1.2380
	750	.9	.0023	1.1111	.1235	.5002	.0009	.8936	.0021	.8851	1.1249

TABLE III

$AR(1), Y_t = .5Y_{t-1} + \varepsilon_t, \varepsilon_t \sim$ i.i.d. $N(0,1), y_0 = 0,$
N-G Prior $\alpha=10, \beta=9$

μ	n	Prior Information				Posterior Information				Predictive Information	
		E(φ)	V(φ)	E(τ)	V(τ)	E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
0.00	25	0	.2298	1.1111	.1235	.4854	.0254	8316	.0307	.7905	1.3259
	50	0	.2298	1.1111	.1235	.5846	.0121	8926	.0228	-.8737	1.1783
	100	0	.2298	1.1111	.1235	.5896	.0062	9788	.0160	.3234	1.0408
	750	0	.2298	1.1111	.1235	.5208	.0010	1.0190	.0027	.8017	.9862
0.25	25	.25	.1293	1.1111	.1235	.4909	.0235	.8409	.0314	.7995	1.3070
	50	.25	.1293	1.1111	.1235	.5842	.0116	.8998	.0231	-.8399	1.1681
	100	.25	.1293	1.1111	.1235	.5893	.0061	.9838	.0161	.3232	1.0355
	750	.25	.1293	1.1111	.1235	.5209	.0010	1.0198	.0027	.8019	.9854
0.50	25	.5	.0575	1.1111	.1235	.5232	.0197	.8490	.0320	.8522	1.2849
	50	.5	.0575	1.1111	.1235	.5945	.0105	.9085	.0236	-.8547	1.1547
	100	.5	.0575	1.1111	.1235	.5952	.0057	.9900	.0163	.3265	1.0290
	750	.5	.0575	1.1111	.1235	.5226	.0010	1.0206	.0027	.8045	.9846
0.75	25	.75	.0144	1.1111	.1235	.6644	.0110	.8289	.0305	1.0821	1.2937
	50	.75	.0144	1.1111	.1235	.6722	.0071	.9020	.0232	-.9664	1.1560
	100	.75	.0144	1.1111	.1235	.6489	.0045	.9831	.0161	.3559	1.0357
	750	.75	.0144	1.1111	.1235	.5376	.0009	1.0161	.0027	.8277	.9889
0.90	25	.9	.0023	1.1111	.1235	.8654	.0028	.7690	.0263	1.4096	1.3684
	50	.9	.0023	1.1111	.1235	.8503	.0023	.8426	.0203	-1.2225	1.2265
	100	.9	.0023	1.1111	.1235	.8200	.0019	.9140	.0139	.4498	1.1132
	750	.9	.0023	1.1111	.1235	.6365	.0007	.9653	.0024	.9798	1.0403

TABLE IV

AR(1), $Y_t = .75Y_{t-1} + \varepsilon_t, \varepsilon_t$ i.i.d. N(0,1), $y_0 = 0$,
N-G Prior $\alpha=10, \beta=9$

μ	n	Prior Information				Posterior Information				Predictive Information	
		E(ϕ)	V(ϕ)	E(τ)	V(τ)	E(ϕ/S)	V(ϕ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
0.00	25	0	.2298	1.1111	.1235	.7231	.0168	.8119	.0293	1.6301	1.3742
	50	0	.2298	1.1111	.1235	.7987	.0069	.8743	.0218	-1.7879	1-2121
	100	0	.2298	1.1111	.1235	.8100	.0033	.9667	.0156	.2774	1.0524
	750	0	.2298	1.1111	.1235	.7591	.0006	1.0168	.0027	1.0167	.9871
0.25	25	.25	.1293	1.1111	.1235	.7167	.0159	.8200	.0299	1.6157	1.3570
	50	.25	.1293	1.1111	.1235	.7943	.0068	.8793	.0221	-1.7779	1.2046
	100	.25	.1293	1.1111	.1235	.8075	.0032	.9700	.0157	.2765	1.0487
	750	.25	.1293	1.1111	.1235	.7587	.0006	1.0175	.0027	1.0162	.9864
0.50	25	.5	.0575	1.1111	.1235	.7149	.0140	.8331	.0309	1.6116	1.3270
	50	.5	.0575	1.1111	.1235	.7893	.0063	.8881	.0225	-1.7668	1.1909
	100	.5	.0575	1.1111	.1235	.8044	.0031	.9760	.0159	.2755	1.0423
	750	.5	.0575	1.1111	.1235	.7584	.0006	1.0186	.0027	1.0158	.9853
0.75	25	.75	.0144	1.1111	.1235	.7584	.0087	.8487	.0320	1.7097	1.2772
	50	.75	.0144	1.1111	.1235	.7991	.0048	.9038	.0233	-1.7887	1.1632
	100	.75	.0144	1.1111	.1235	.8083	.0027	.9874	.0162	.2768	1.0302
	750	.75	.0144	1.1111	.1235	.7605	.0005	1.0201	.0027	1.0187	.9838
0.90	25	.9	.0023	1.1111	.1235	.8809	.0025	.8314	.0307	1.9858	1.2713
	50	.9	.0023	1.1111	.1235	.8779	.0019	.8976	.0230	-1.9651	1.1564
	100	.9	.0023	1.1111	.1235	.8668	.0014	.9805	.0160	.2969	1.0373
	750	.9	.0023	1.1111	.1235	.7887	.0005	1.011	.0027	1.0504	.9924

TABLE V

$AR(1), Y_t = .9Y_{t-1} + \varepsilon_t, \varepsilon_t \sim i.i.d.\ N(0,1), y_0 = 0,$
N-G Prior $\alpha=10, \beta=9$

μ	n	Prior Information				Posterior Information				Predictive Information	
		E(φ)	V(φ)	E(τ)	V(τ)	E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
0.00	25	0	.2298	1.1111	.1235	.9144	.0082	.7961	.0282	3.7187	1.4498
	50	0	.2298	1.1111	.1235	.9332	.0025	.8641	.0213	-2.0811	1.2035
	100	0	.2298	1.1111	.1235	.9292	.0013	.9549	.0152	.2090	1.0651
	750	0	.2298	1.1111	.1235	.9036	.0002	1.0152	.0027	.9144	.9878
0.25	25	.25	.1293	1.1111	.1235	.9073	.0080	.7994	.0284	3.6900	1.4411
	50	.25	.1293	1.1111	.1235	.9307	.0024	.8661	.0214	-2.0755	1.2007
	100	.25	.1293	1.1111	.1235	.9277	.0013	.9563	.0152	.2087	1.0635
	750	.25	.1293	1.1111	.1235	.9033	.0002	1.0155	.0027	.9142	.9875
0.50	25	.5	.0575	1.1111	.1235	.8958	.0075	.8062	.0289	3.0431	1.4216
	50	.5	.0575	1.1111	.1235	.9262	.0024	.8700	.0216	-2.0655	1.1951
	100	.5	.0575	1.1111	.1235	.9251	.0013	.9592	.0153	.2081	1.0603
	750	.5	.0575	1.1111	.1235	.9028	.0002	1.0162	.0027	.9137	.9869
0.75	25	.75	.0144	1.1111	.1235	.8814	.0056	.8253	.0303	3.5844	1.3508
	50	.75	.0144	1.1111	.1235	.9174	.0021	.8814	.0222	-2.0458	1.1785
	100	.75	.0144	1.1111	.1235	.9197	.0012	.9672	.0156	.2069	1.0515
	750	.75	.0144	1.1111	.1235	.9019	.0002	1.0178	.0027	.9128	.9853
0.90	25	.9	.0023	1.1111	.1235	.9105	.0021	.8495	.0321	3.7030	1.2663
	50	.9	.0023	1.1111	.1235	.9218	.0012	.9023	.0233	-2.0558	1.1473
	100	.9	.0023	1.1111	.1235	.9223	.0008	.9818	.0161	.2075	1.0358
	750	.9	.0023	1.1111	.1235	.9041	.0002	1.0199	.0027	.9150	.9832

TABLE VI

AR(1), $Y_t=\varepsilon_t$, $\varepsilon_t \sim$ i.i.d. N(0,1), $y_0=0$, Mixture of N-G Prior $\alpha=10, \beta=9$

Condition on ξ_i for each prior comp.	i	μ_i	n	Prior Information				Posterior Information				Predictive Information	
				E(φ)	V(φ)	E(τ)	V(τ)	E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
Fixed for all prior components for μ=.5 i.e ξ_i=17.4056 i=1,2,3	1	-.5	25	0	.2241	1.1111	.1235	.0743	.0484	.8391	.0317	.1001	1.3531
	2	0	50	0	.2241	1.1111	.1235	.1479	.0244	.9107	.0239	-.2260	1.2132
	3	.5	100	0	.2241	1.1111	.1235	.1268	.0120	.9873	.0163	.6750	1.0577
			200	0	.2241	1.1111	.1235	.0917	.0057	.9725	.0086	-.1299	1.0490
Fixed for all prior components for μ=0 i.e ξ_i=4.3514 i=1,2,3	1	-.5	25	0	.3965	1.1111	.1235	.1018	.0352	.8467	.0319	.1371	1.3031
	2	0	50	0	.3965	1.1111	.1235	.1804	.0190	.9161	.0240	-.2757	1.1719
	3	.5	100	0	.3965	1.1111	.1235	.1425	.0097	.9909	.0164	.0758	1.0468
			200	0	.3965	1.1111	.1235	.0975	.0050	.9741	.0086	-1.1380	1.031
depends on μ_i of each prior component ξ_1=17.4056 ξ_2=4.3514 ξ_3=17.4056	1	-.5	25	0	.2686	1.1111	.1235	.0839	.0472	.8404	.0318	.1117	1.3575
	2	0	50	0	.2686	1.1111	.1235	.1589	.0260	.9138	.0241	-.2418	1.2143
	3	.5	100	0	.2686	1.1111	.1235	.1321	.0123	.9890	.0164	.7018	1.0578
			200	0	.2686	1.1111	.1235	.0936	.0058	.9729	.0086	-.1335	1.0485

TABLE VII

AR(1), $Y_t=.5Y_{t-1}+\varepsilon_t$, $\varepsilon_t \sim$ i.i.d. N(0,1), $y_0=0$, Mixture of N-G Prior $\alpha=10, \beta=9$

Condition on ξ_i for each prior comp.	i	μ_i	n	Prior Information				Posterior Information				Predictive Information	
				E(φ)	V(φ)	E(τ)	V(τ)	E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
Fixed for all prior components for μ=.5 i.e ξ_i=17.4056 i=1,2,3	1	-.5	25	0	.2241	1.1111	.1235	.3182	.0450	1.2084	.0662	-.2450	1.0273
	2	0	50	0	.2241	1.1111	.1235	.4615	.0226	1.3904	.0558	.1040	1.0231
	3	.5	100	0	.2241	1.1111	.1235	.5426	.0080	1.0689	.0192	-.3116	.9749
			200	0	.2241	1.1111	.1235	.5543	.0037	1.0199	.0095	-.2152	.8486
Fixed for all prior components for μ=0 i.e ξ_i=4.3514 i=1,2,3	1	-.5	25	0	.3965	1.1111	.1235	.4640	.0323	1.2042	.0654	-.0357	.9980
	2	0	50	0	.3965	1.1111	.1235	.5747	.0146	1.3852	.0555	.1295	.9690
	3	.5	100	0	.3965	1.1111	.1235	.5904	.0065	1.0665	.0191	-.3390	.8968
			200	0	.3965	1.1111	.1235	.5784	.0033	1.0186	.0095	-.2245	.7691
depends on μ_i of each prior component ξ_1=17.4056 ξ_2=4.3514 ξ_3=17.4056	1	-.5	25	0	.2686	1.1111	.1235	.3685	.0550	1.2099	.0655	-.2791	1.0210
	2	0	50	0	.2686	1.1111	.1235	.4987	.0254	1.3847	.0551	.1118	1.0119
	3	.5	100	0	.2686	1.1111	.1235	.5582	.0086	1.0673	.0190	-.3203	.9603
			200	0	.2686	1.1111	.1235	.5622	.0039	1.0193	.0095	-.2182	.8327

TABLE VIII

$AR(1), Y_t = .9Y_{t-1} + \varepsilon_t, \varepsilon_t \sim$ i.i.d. $N(0,1), y_0 = 0$, Mixture of N-G Prior $\alpha=10, \beta=9$

Condition on ξ_i for each prior comp.	i	μ_i	n	Prior Information				Posterior Information				Predictive Information	
				E(φ)	V(φ)	E(τ)	V(τ)	E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
Fixed for all prior components	1	-.5	25	0	.2241	1.1111	.1235	.7661	.0172	1.2428	.0610	1.9526	1.2683
for μ=.5 i.e ξ_i=17.4056	2	0	50	0	.2241	1.1111	.1235	.7780	.0080	1.2588	.0454	-.6240	1.0192
i=1,2,3	3	.5	100	0	.2241	1.1111	.1235	.8228	.0035	1.2397	.02556	1.7716	1.0171
			200	0	.2241	1.1111	.1235	.8700	.0012	1.0599	.0102	1.2994	.9515
Fixed for all prior	1	-.5	25	0	.3965	1.1111	.1235	.8934	.0109	1.3002	.0769	2.2769	.9617
components for μ=0 i.e	2	0	50	0	.3965	1.1111	.1235	.8485	.0057	1.2855	.0479	-.6806	.9512
ξ_i=4.3514 i=1,2,3	3	.5	100	0	.3965	1.1111	.1235	.8592	.0028	1.2556	.0265	1.8502	.8480
			200	0	.3965	1.1111	.1235	.8832	.0012	1.0668	.0104	1.3099	.8479
depends on μ_i of each	1	-.5	25	0	.2686	1.1111	.1235	.8043	.0208	1.2331	.0678	2.0449	1.2307
prior component ξ_1=17.4056	2	0	50	0	.2686	1.1111	.1235	.7997	.0091	1.2515	.0449	-.6408	1.0044
ξ_2=4.3514	3	.5	100	0	.2686	1.1111	.1235	.8342	.0037	1.2355	.0255	1.7960	.9792
ξ_3=17.4056			200	0	.2686	1.1111	.1235	.8743	.0012	1.0584	.0102	1.2967	.9281

TABLE IX

AR(1), $Y_t=\phi Y_{t-1}+\varepsilon_t$, $\varepsilon_t \simeq$ i.i.d. N(0,1), $y_0=0$, Mixture of N-G Prior $\alpha=2, \beta=1$

Condition on ξ_i for each prior comp.	i	μ_i	n	Prior Information E(φ)	V(φ)	E(τ)	V(τ)	Posterior Information E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	Predictive Information E(Yn+1)	V(Yn+1)
Fixed for all prior components for μ=.5	1	-.5	25	0	.1991	2	2	.0561	.0665	.7675	.0417	.0756	1.5772
ξ_i=30.8247	2	0	50	0	.1991	2	2	.1230	.0344	.8834	.0291	-.1880	1.3124
i=1,2,3	3	.5	100	0	.1991	2	2	.1134	.0166	.9840	.0188	.0604	1.0744
			200	0	.1991	2	2	.0864	.0073	.9709	.0093	-.1223	1.0741

φ = 0.0

Condition on ξ_i for each prior comp.	i	μ_i	n	E(φ)	V(φ)	E(τ)	V(τ)	E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
Fixed for all prior components	1	-.5	25	0	.1991	2	2	.6561	.0273	1.2117	.1080	1.6720	1.8304
for μ=.5 ξ_i=30.8247	2	0	50	0	.1991	2	2	.7120	.0118	1.2509	-0599	-.5711	1.2366
i=1,2,3	3	.5	100	0	.1991	2	2	.7868	.0046	1.2190	.0291	1.6943	1.0839
			200	0	.1991	2	2	.8566	.0014	1.0321	.0105	1.2705	1.0646

φ = .5

Condition on ξ_i for each prior comp.	i	μ_i	n	E(φ)	V(φ)	E(τ)	V(τ)	E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	V(Yn+1)
Fixed for all prior components	1	-.5	25	0	.1991	2	2	.2255	.0621	1.3771	.1345	-.0174	1.0850
for μ=.5 ξ_i=30.8247	2	0	50	0	.1991	2	2	.3752	.03367	1.5740	.0922	.0846	1.0601
i=1,2,3	3	.5	100	0	.1991	2	2	.4991	.0110	1.0814	.0226	-.2866	1.0422
			200	0	.1991	2	2	.5310	.0046	1.0223	.0103	-.2061	0.8839

φ = .9

TABLE X

AR(1), $Y_t=\phi Y_{t-1}+\varepsilon_t$, $\varepsilon_t \simeq$ i.i.d. N(0,1), $y_0=0$, Mixture of N-G Prior $\alpha=1.1$, $\beta=.1$

Condition on ξ_i for each prior comp.	prof comp #	μ_i	n	Prior Information E(φ)	V(φ)	E(τ)	V(τ)	Posterior Information E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	Predictive Information E(Yn+1)	Y (Yn+1)
Fixed for all prior components	1	-.5	25	0	.1827	11	110	.3320	.0949	.7546	.0433	.4470	1.1463
for μ=.5 ξ_i=62.3784	2	0	50	0	.1827	11	110	.0856	.0580	.8738	.0295	-.1308	1.1337
i=1,2,3	3	.5	100	0	.1827	11	110	.0897	.0320	.9817	.0190	.4776	1.1094
			200	0	.1827	11	110	.0757	.0129	.9713	.0094	-.1071	1.1052

φ = 0.0

Condition on ξ_i for each prior comp.	prof comp #	μ_i	n	E(φ)	V(φ)	E(τ)	V(τ)	E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	Y (Yn+1)
Fixed for all prior components	1	-.5	25	0	.1827	11	110	.1190	.0887	1.4192	.1515	-.0092	1.7018
for μ=.5 ξ_i=62.3784	2	0	50	0	.1827	11	110	.2503	.0565	1.6004	.0982	.0564	1.4155
i=1,2,3	3	.5	100	0	.1827	11	110	.4166	.0204	1.0814	.0229	-.2393	1.2008
			200	0	.1827	11	110	.4821	.0081	1.0219	.0103	-.1871	1.1266

φ = .5

Condition on ξ_i for each prior comp.	prof comp #	μ_i	n	E(φ)	V(φ)	E(τ)	V(τ)	E(φ/S)	V(φ/S)	E(τ/S)	V(τ/S)	E(Yn+1)	Y (Yn+1)
Fixed for all prior components	1	-.5	25	0	.1827	11	110	.4811	.0467	1.3339	.1228	1.2261	2.6844
for μ=.5 ξ_i=62.3784	2	0	50	0	.1827	11	110	.5882	.0222	1.3181	.0644	-.4718	1.6690
i=1,2,3	3	.5	100	0	.1827	11	110	.7111	.0085	1.2626	.0307	1.5312	1.3381
			200	0	.1827	11	110	.8260	.0021	1.0546	.0109	1.2251	1.2236

φ = .9

TABLE XI

ESTIMATED ORDER BY THE POSTERIOR MODE WITH CONJUGATE PRIOR, TRUE MODEL: $Y(t)=.65Y(t-1)+.3Y(t-2)+\varepsilon(t), \tau=1, \alpha=2, \beta=1$

n	Diagonal Elements of Q	The Number of Prior Vector μ Used					
		1	2	3	4	5	6
30	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	2	4	3	2	1	2
50	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	2	4	2	2	1	2
60	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	2	3	2	2	1	2
75	.01	2	2	2	2	2	2
	1	2	2	2	2	2	2
	100	2	3	2	2	1	2
100	.01	2	2	2	2	2	2
	1	2	2	2	2	2	2
	100	2	3	2	2	1	2
200	.01	2	2	2	2	2	2
	1	2	2	2	2	2	2
	100	2	3	2	2	2	2

TABLE XII

ESTIMATED ORDER BY THE POSTERIOR MODE WITH CONJUGATE PRIOR, TRUE MODEL: $Y(t)=.65Y(t-1)+.3Y(t-2)+\varepsilon(t), \tau=1, \alpha=2, \beta=10$

n	Diagonal Elements of Q	The Number of Prior Vector μ Used					
		1	2	3	4	5	6
30	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	2	4	3	2	1	2
50	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	2	4	2	2	1	2
60	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	2	3	2	2	1	2
75	.01	1	1	1	1	1	1
	1	2	2	2	2	2	2
	100	2	3	2	2	1	2
100	.01	2	2	2	2	2	2
	1	2	2	2	2	2	2
	100	2	3	2	2	1	2
200	.01	2	2	2	2	2	2
	1	2	2	2	2	2	2
	100	2	3	2	2	2	2

TABLE XIII

ESTIMATED ORDER BY THE POSTERIOR MODE WITH CONJUGATE PRIOR, TRUE MODEL: $Y(t)=.35Y(t-1)-.35Y(t-2)+\varepsilon(t),\tau=1,\alpha=2,\beta=1$

n	Diagonal Elements of Q	The Number of Prior Vector μ Used					
		1	2	3	4	5	6
30	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	2	1	1	2	1
50	.01	1	1	1	1	1	1
	1	2	2	2	2	2	2
	100	1	2	2	1	2	3
60	.01	1	1	1	1	1	1
	1	2	2	2	2	2	2
	100	1	2	2	1	2	2
75	.01	1	1	1	1	1	1
	1	2	2	2	2	2	2
	100	1	2	2	1	2	2
100	.01	1	1	1	1	1	1
	1	2	2	2	2	2	2
	100	1	2	2	1	2	2
200	.01	2	2	2	2	2	2
	1	2	2	2	2	2	2
	100	1	2	2	1	2	2

TABLE XIV

ESTIMATED ORDER BY THE POSTERIOR MODE WITH CONJUGATE PRIOR, TRUE MODEL: $Y(t)=.35Y(t-1)-.35Y(t-2)+\varepsilon(t), \tau=1, \alpha=2, \beta=10$

n	Diagonal Elements of Q	The Number of Prior Vector μ Used					
		1	2	3	4	5	6
30	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	1	1	1	1	1
50	.01	1	1	1	1	1	1
	1	2	2	2	2	2	2
	100	1	2	2	1	2	3
60	.01	1	1	1	1	1	1
	1	2	2	2	2	2	2
	100	1	2	2	1	2	2
75	.01	1	1	1	1	1	1
	1	2	2	2	2	2	2
	100	1	2	2	1	2	2
100	.01	1	1	1	1	1	1
	1	2	2	2	2	2	2
	100	1	2	2	1	2	2
200	.01	2	2	2	2	2	2
	1	2	2	2	2	2	2
	100	1	2	2	1	2	2

TABLE XV

ESTIMATED ORDER BY THE POSTERIOR MODE WITH CONJUGATE PRIOR, TRUE MODEL: $Y(t)=.55Y(t-1)+.05Y(t-2)+\varepsilon(t),\tau=1,\alpha=2,\beta=1$

n	Diagonal Elements of Q	The Number of Prior Vector μ Used					
		1	2	3	4	5	6
30	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	2	1	2	4	1	4
50	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	1	1	4	1	3
60	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	1	1	3	1	3
75	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	1	1	1	1	1
100	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	1	1	1	1	1
200	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	1	1	1	1	1

TABLE XVI

ESTIMATED ORDER BY THE POSTERIOR MODE WITH CONJUGATE PRIOR, TRUE MODEL: $Y(t)=.55Y(t-1)+.05Y(t-2)+\varepsilon(t), \tau=1, \alpha=2, \beta=10$

n	Diagonal Elements of Q	The Number of Prior Vector μ Used					
		1	2	3	4	5	6
30	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	2	1	2	4	1	4
50	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	1	1	4	1	3
60	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	1	1	1	1	3
75	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	1	1	1	1	1
100	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	1	1	1	1	1
200	.01	1	1	1	1	1	1
	1	1	1	1	1	1	1
	100	1	1	1	1	1	1

TABLE XVII

STATISTICS OF ESTIMATED ORDERS OF 500 SIMULATIONS, TRUE MODEL:

$Y(t)=.65Y(t-1)+.3Y(t-2)+\varepsilon(t),\xi(\phi^{(p)},\tau|p)\propto\tau^{p/2-1},n=50,K=10$

Procedure \ Statistics			Mean	Variance	MSE	# of Correct Identification
Bayesian	Posterior	Mode	2.9400	6.0285	6.9121	194
Bayesian	Posterior	Mean	3.8960	2.7507	6.3455	92
Akaike	FPE		1.8940	1.4376	1.4489	195
Akaike	AIC		1.8980	1.4545	1.4649	195
Akaike	AIC4		1.3020	0.2473	0.7345	133
Akaike	Mode		1.4180	0.5323	0.8711	138

TABLE XVIII

STATISTICS OF ESTIMATED ORDERS OF 500 SIMULATIONS, TRUE MODEL:
$Y(t)=.65Y(t-1)+.3Y(t-2)+\varepsilon(t), \xi(\phi^{(p)}, \tau|p) \propto \tau^{p/2-1}, n=60, K=10$

Procedure \ Statistics	Mean	Variance	MSE	# of Correct Identification
Bayesian Posterior Mode	2.6240	3.9024	4.2918	253
Akaike FPE	1.9980	1.4810	1.4810	248
Akaike AIC	2.0140	1.5289	1.5291	248
Akaike AIC4	1.4400	0.3271	0.6407	186
Akaike Mode	1.5780	0.5330	0.7111	207

TABLE XIX

STATISTICS OF ESTIMATED ORDERS OF 500 SIMULATIONS, TRUE MODEL:
$Y(t)=.65Y(t-1)+.3Y(t-2)+\varepsilon(t), \xi(\phi^{(p)},\tau|p)\propto\tau^{p/2-1}, n=70, K=10$

Procedure \ Statistics	Mean	Variance	MSE	# of Correct Identification
Bayesian Posterior Mode	2.3200	2.4906	2.5930	280
Bayesian Posterior Mean	3.1360	1.5727	2.8631	162
Akaike FPE	2.1080	1.4432	1.4549	276
Akaike AIC	2.1100	1.4408	1.4529	277
Akaike AIC4	1.5200	0.3062	0.5366	235
Akaike Mode	1.7040	0.7539	0.8415	240

TABLE XX

STATISTICS OF ESTIMATED ORDERS OF 500 SIMULATIONS, TRUE MODEL:

$Y(t)=.5Y(t-1)-.06Y(t-2)+.45Y(t-8)+\varepsilon(t), \xi(\phi^{(p)},\tau|p)\propto\tau^{p/2-1}$,

$n=50, K=10$

Procedure \ Statistics	Mean	Variance	MSE	# of Correct Identification
Bayesian Posterior Mode	6.5860	9.9625	11.962	215
Bayesian Posterior Mean	3.1360	5.5727	6.8631	162
Akaike FPE	4.4800	10.382	22.773	155
Akaike AIC	4.6140	10.374	21.839	164
Akaike AIC4	1.4120	1.2888	44.691	8
Akaike Mode	2.6160	7.3032	36.291	65

TABLE XXI

STATISTICS OF ESTIMATED ORDERS OF 500 SIMULATIONS, TRUE MODEL:
$Y(t)=.5Y(t-1)-.06Y(t-2)+.45Y(t-8)+\varepsilon(t), \xi(\phi^{(p)},\tau|p)\propto\tau^{p/2-1}$,
n=60,K=10

Procedure \ Statistics	Mean	Variance	MSE	# of Correct Identification
Bayesian Posterior Mode	6.7940	8.8300	10.536	263
Akaike FPE	5.7120	9.7722	15.560	241
Akaike AIC	5.7600	9.7880	15.368	240
Akaike AIC4	1.8100	3.3566	41.673	29
Akaike Mode	3.6980	11.624	29.131	129

TABLE XXII

STATISTICS OF ESTIMATED ORDERS OF 500 SIMULATIONS, TRUE MODEL:
$Y(t)=.5Y(t-1)-.06Y(t-2)+.45Y(t-8)+\varepsilon(t), \xi(\phi^{(p)},\tau|p)\propto\tau^{p/2-1}$,
$n=70, K=10$

Procedure \ Statistics			Mean	Variance	MSE	# of Correct Identification
Bayesian	Posterior	Mode	6.9360	7.8365	9.1999	297
Bayesian	Posterior	Mean	7.1640	4.4500	5.1489	229
Akaike	FPE		6.7260	8.4506	10.8289	287
Akaike	AIC		6.7660	8.3507	10.6367	288
Akaike	AIC4		2.3820	6.2566	37.818	67
Akaike	Mode		4.8040	11.673	21.887	210

TABLE XXIII

STATISTICS OF ESTIMATED ORDER OF 500 SIMULATIONS, TRUE MODEL:

$Y(t)=.55Y(t-1)+.05Y(t-2)+\varepsilon(t), \xi(\phi^{(p)},\tau|p)\propto\tau^{p/2-1}, n=50, K=10$

Statistics / Procedure	Mean	Variance	MSE	# of Correct Identification
Bayesian Posterior Mode	2.4380	6.1986	6.3904	45
Akaike FPE	1.4280	1.1591	1.4863	43
Akaike AIC	1.4360	1.1763	1.4944	42
Akaike AIC4	1.0560	0.0810	0.9722	17
Akaike Mode	1.1120	0.3281	1.1167	23

TABLE XXIV

STATISTICS OF ESTIMATED ORDERS OF 500 SIMULATIONS, TRUE MODEL:
$Y(t)=.35Y(t-1)-.35Y(t-2)+\varepsilon(t), \xi(\phi^{(p)},\tau|p)\propto\tau^{p/2-1}, n=50, K=10$

Procedure \ Statistics			Mean	Variance	MSE	# of Correct Identification
Bayesian	Posterior	Mode	2.0800	5.4044	6.5708	273
Akaike	FPE		2.2400	0.9884	1.0460	348
Akaike	AIC		2.2600	1.1627	1.2303	348
Akaike	AIC4		1.7420	02840	0.3506	334
Akaike	Mode		1.8900	0.4548	0.46691	345

TABLE XXV

STATISTICS OF ESTIMATED ORDERS OF 500 SIMULATIONS, TRUE MODEL:
$Y(t)=.65Y(t-1)+.3Y(t-2)+\varepsilon(t), \xi(\phi^{(p)},\tau|p)\propto\tau^{p/2-1}(2\pi)^{-p/2}, n=50,$
$K=10$

Procedure \ Statistics			Mean	Variance	MSE	# of Correct Identification
Bayesian	Posterior	Mode	1.5280	0.6265	0.8493	184
Bayesian	Posterior	Mean	1.8020	0.5158	0.5550	274
Akaike	FPE		1.8940	1.4376	1.4489	195
Akaike	AIC		1.8980	1.4545	1.4649	195
Akaike	AIC4		1.3020	0.2473	0.7345	133
Akaike	Mode		1.4180	0.5323	0.8711	138

TABLE XXVI

STATISTICS OF ESTIMATED ORDERS OF 500 SIMULATIONS, TRUE MODEL:
$Y(t)=.65Y(t-1)+.3Y(t-2)+\varepsilon(t), \xi(\phi^{(p)},\tau|p)\propto\tau^{p/2-1}(2\pi)^{-p/2}, n=70,$
$K=10$

Procedure \ Statistics	Mean	Variance	MSE	# of Correct Identification
Bayesian Posterior Mode	1.6260	0.4671	0.6069	243
Bayesian Posterior Mean	1.7920	0.4056	0.4488	310
Akaike FPE	2.1080	1.4432	1.4549	276
Akaike AIC	2.1100	1.4408	1.4529	277
Akaike AIC4	1.5200	0.3062	0.5366	235
Akaike Mode	1.7040	0.7539	0.8415	240

TABLE XXVII

STATISTICS OF ESTIMATED ORDERS OF 500 SIMULATIONS, TRUE MODEL:
$Y(t)=.5Y(t-1)-.06Y(t-2)+.45Y(t-8)+\varepsilon(t), \xi(\phi^{(p)},\tau|p)\propto\tau^{p/2-1}$
$(2\pi)^{-p/2}, n=50, K=10$

Procedure \ Statistics	Mean	Variance	MSE	# of Correct Identification
Bayesian Posterior Mode	2.2240	5.7694	39.132	48
Bayesian Posterior Mean	2.6960	4.1799	32.312	21
Akaike FPE	4.4800	10.382	22.773	155
Akaike AIC	4.6140	10.374	21.839	164
Akaike AIC4	1.4120	1.2888	44.691	8
Akaike Mode	2.6160	7.3032	36.291	65

TABLE XXVIII

STATISTICS OF ESTIMATED ORDERS OF 500 SIMULATIONS, TRUE MODEL: $Y(t)=.5Y(t-1)-.06Y(t-2)+.45Y(t-8)+\varepsilon(t), \xi(\phi^{(p)},\tau|p)\propto\tau^{p/2-1}(2\pi)^{-p/2}, n=70, K=10$

Procedure \ Statistics	Mean	Variance	MSE	# of Correct Identification
Bayesian Posterior Mode	3.3220	10.0990	31.982	128
Bayesian Posterior Mean	3.5420	7.1706	27.044	69
Akaike FPE	6.7260	8.4506	10.8289	287
Akaike AIC	6.7660	8.3507	10.6367	288
Akaike AIC4	2.3820	6.2566	37.818	67
Akaike Mode	4.8040	11.6730	21.887	210

APPENDIX B

FIGURES

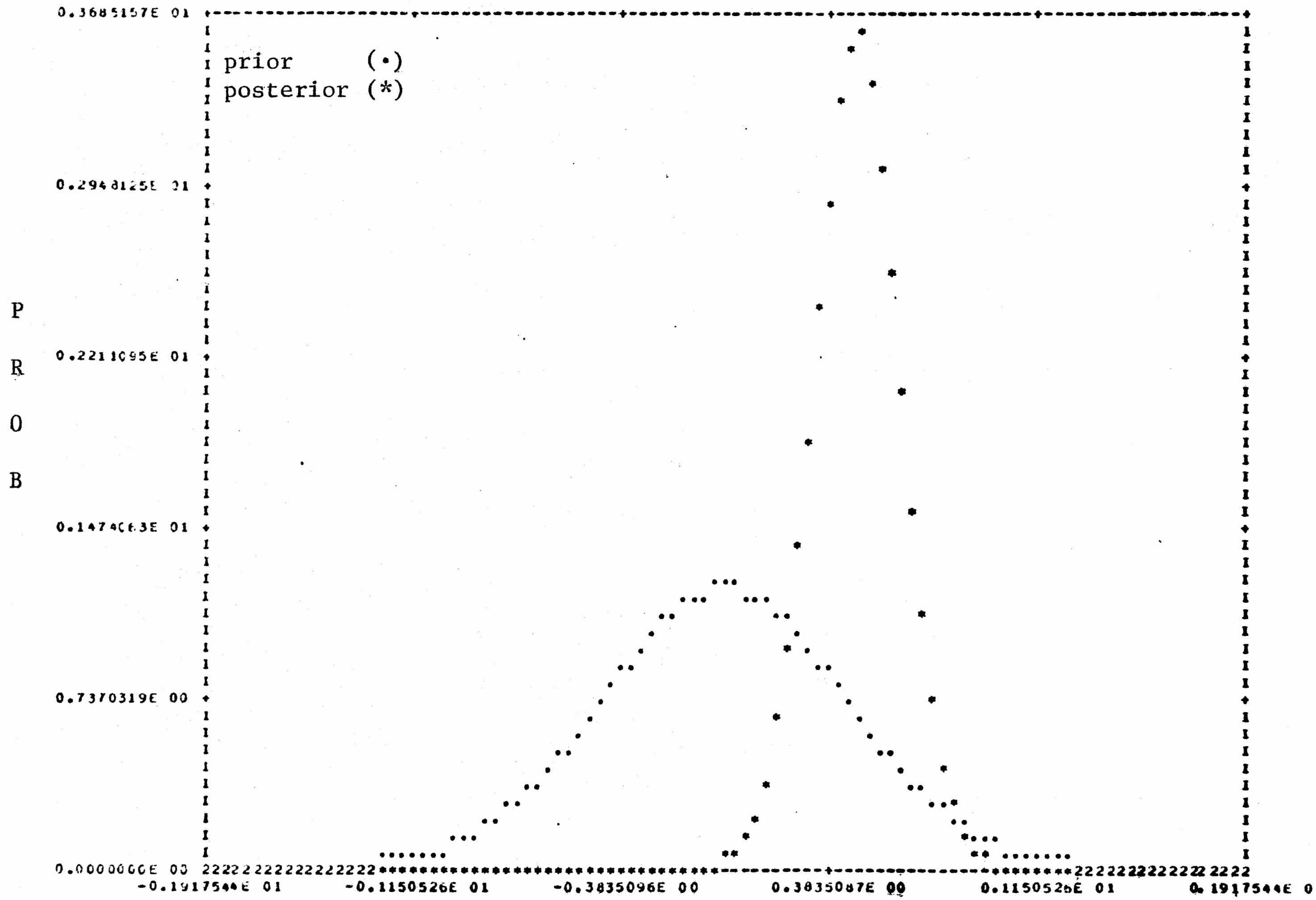

Figure 1. Prior and Posterior Distribution for ϕ: True ϕ=.5, n=25, Prior Mean = 0, N-G Prior for (ϕ,τ)

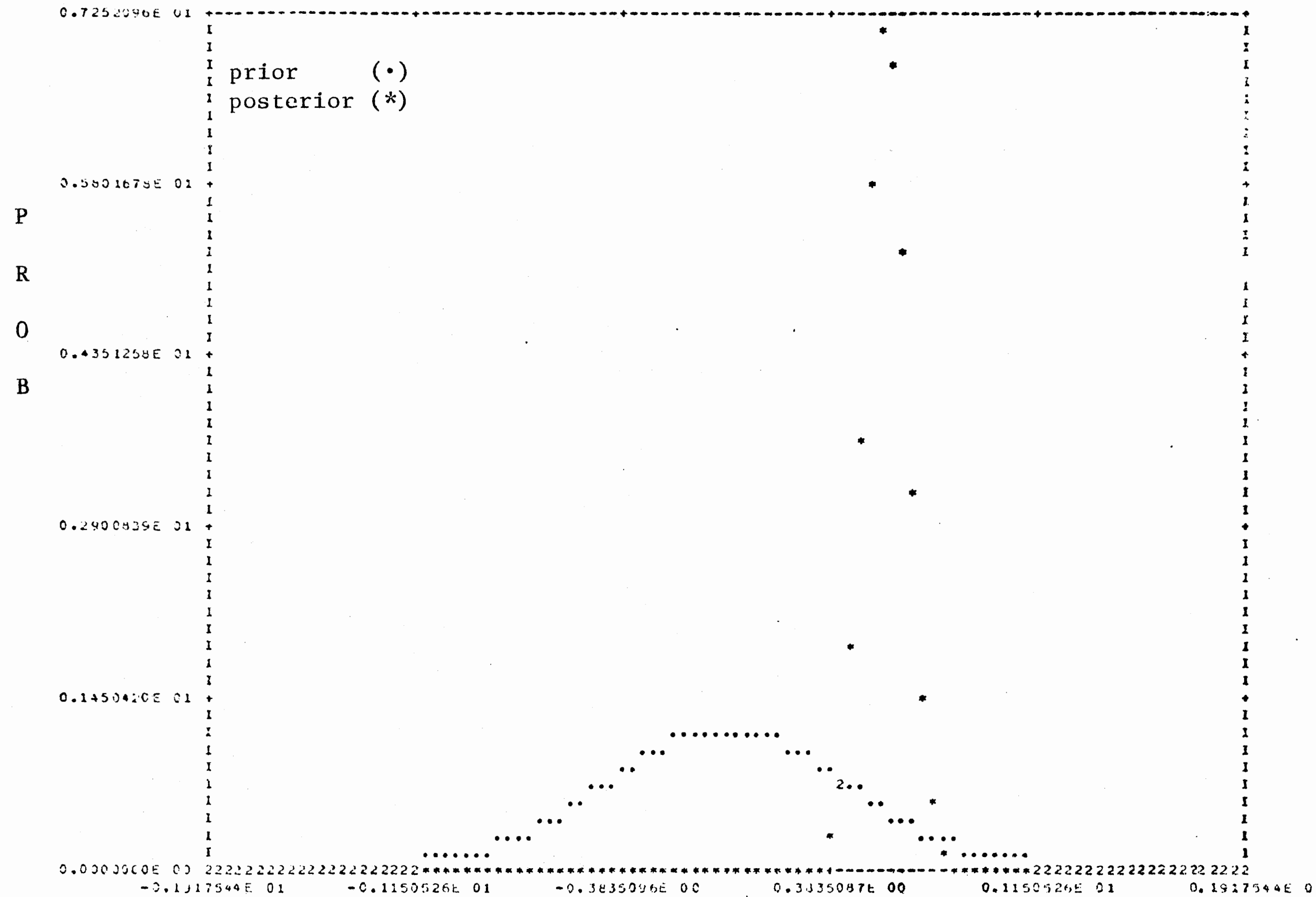

Figure 2. Prior and Posterior Distribution for ϕ: True ϕ=.5, n=100, Prior Mean = 0, N-G Prior for (ϕ,τ)

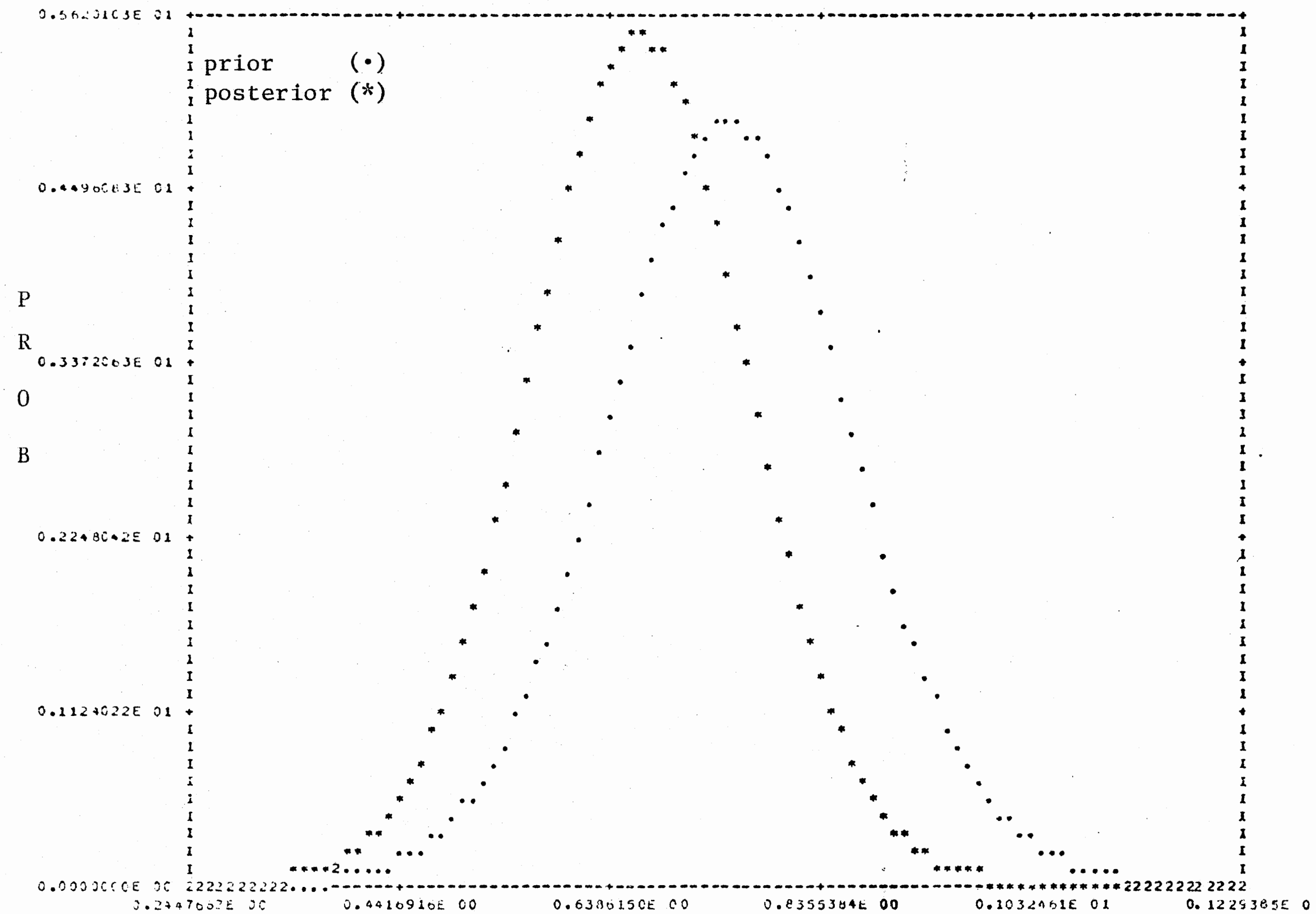

Figure 3. Prior and Posterior Distribution for ϕ: True ϕ=.5, n=25, Prior Mean = .75, N-G Prior for (ϕ,τ)

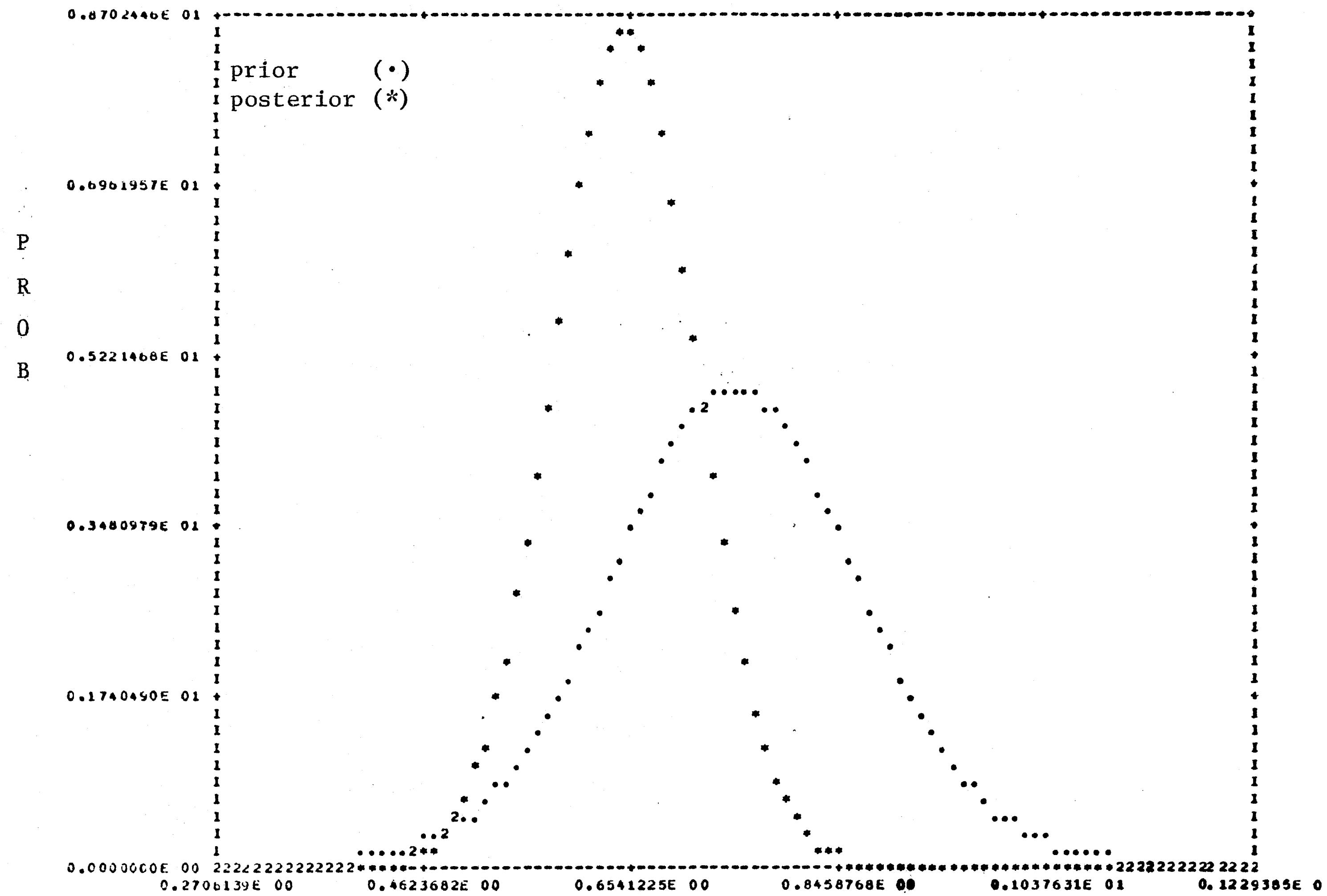

Figure 4. Prior and Posterior Distribution for ϕ: True ϕ=.5, n=100, Prior Mean = .75, N-G Prior for (ϕ,τ)

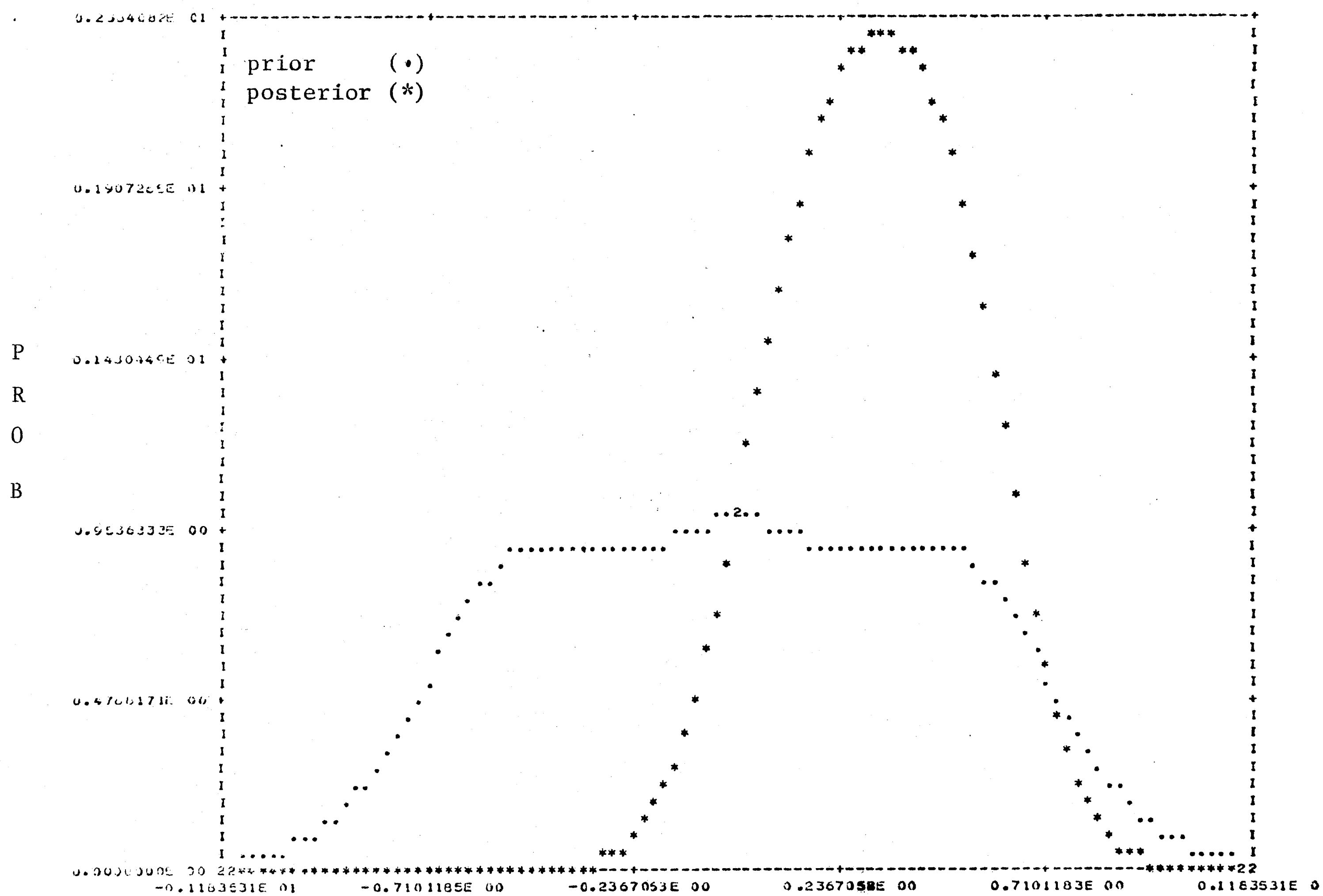

Figure 5. Prior and Posterior Distribution for ϕ: True ϕ=.5, n=25, Prior Mean = 0.
Mixture of N-G Prior for (ϕ,τ), Prior ξ_i is fixed for μ=.5 for All i

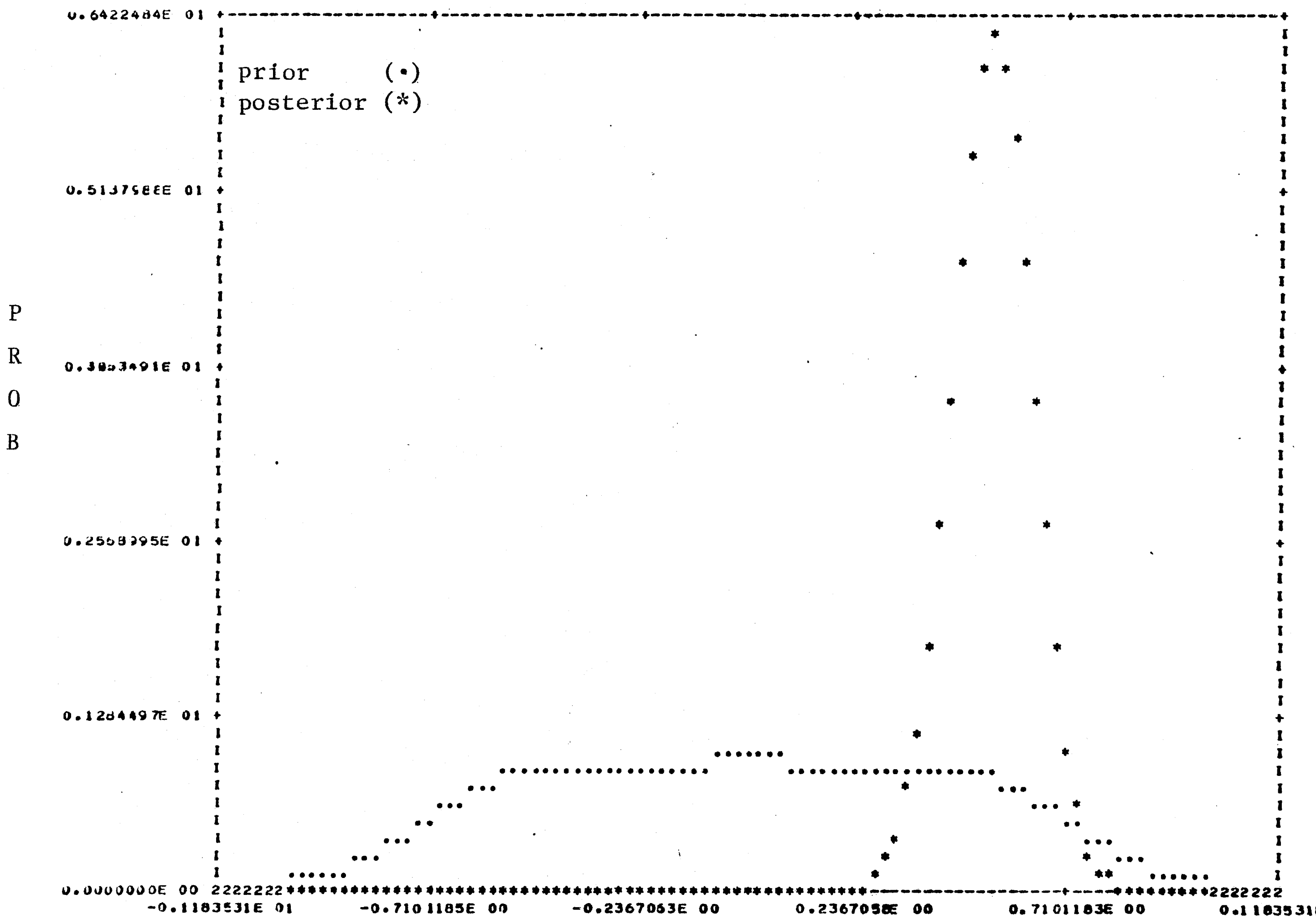

Figure 6. Prior and Posterior Distribution for ϕ: True ϕ=.5, n=100, Prior Mean = 0. Mixture of N-G Prior for (ϕ,τ), Prior ξ_i is Fixed for μ=.5 for All i

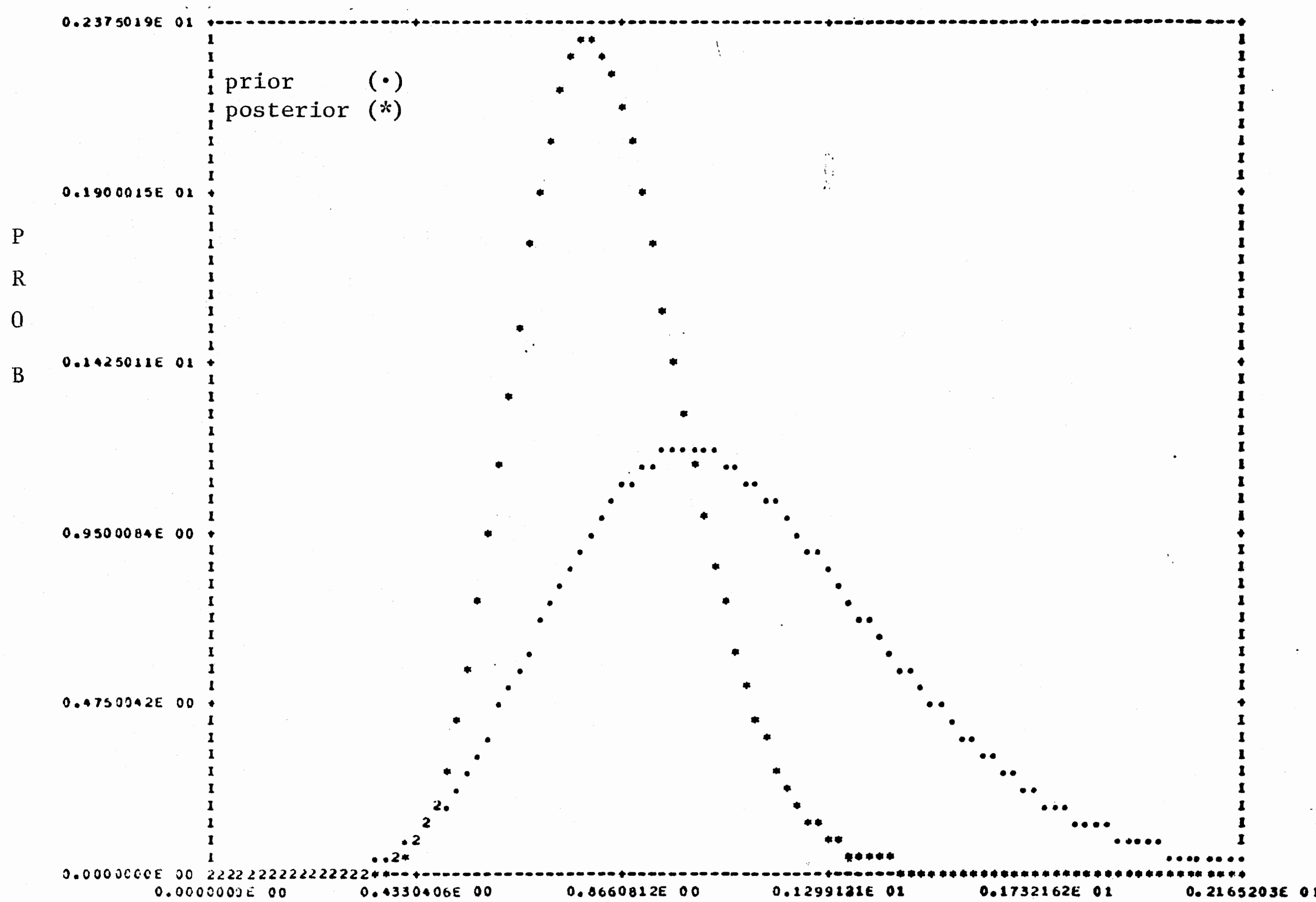

Figure 7. Prior and Posterior Distribution for τ: True τ=1, True ϕ=.5, n=25, $E(\tau)$=1.1111, $E(\phi)$=0, N-G Prior for (ϕ,τ)

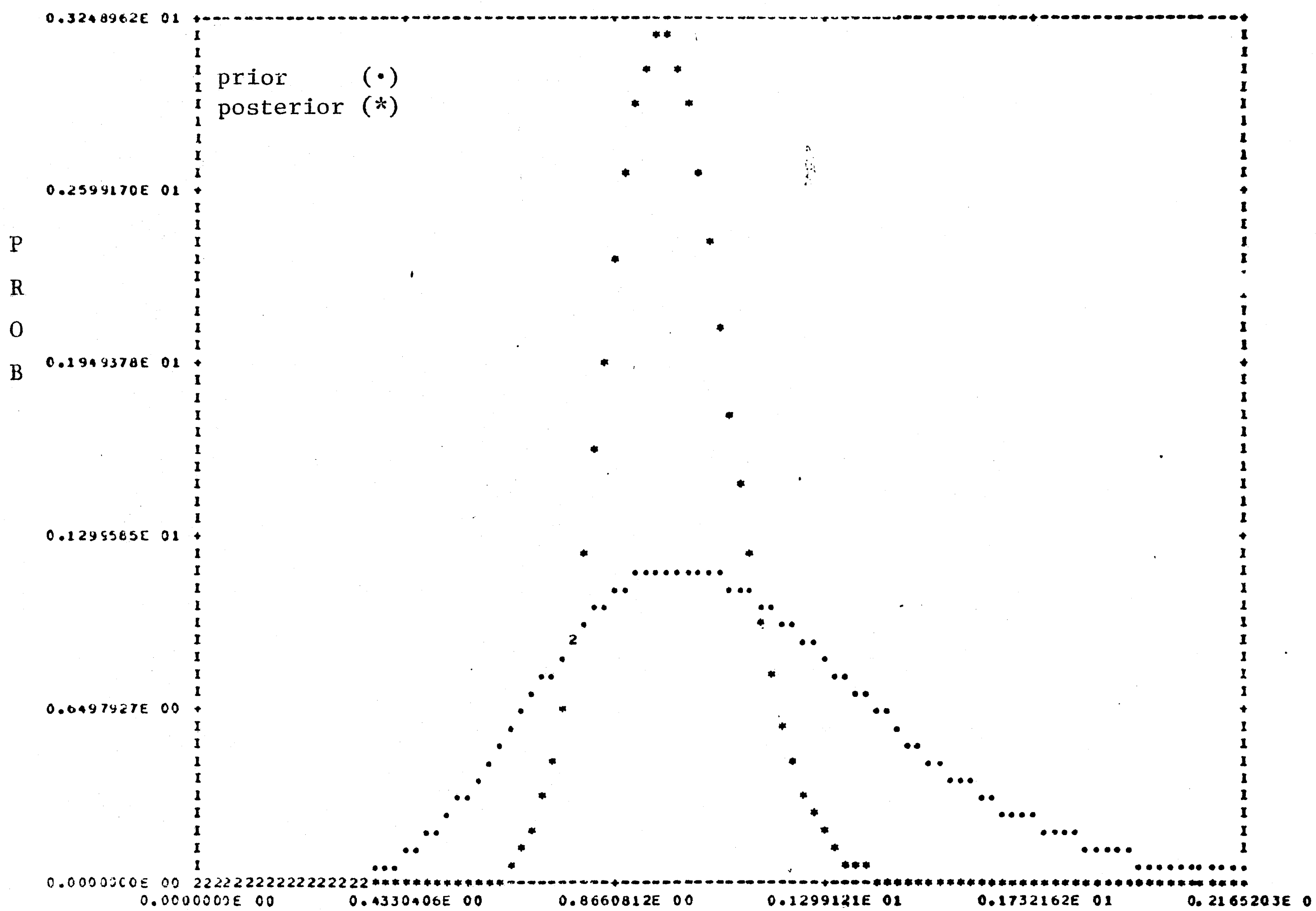

Figure 8. Prior and Posterior Distribution for τ: True τ=1, True ϕ=.5, n=100, $E(\tau)$=1.1111, $E(\phi)$=0, N-G Prior for (ϕ,τ)

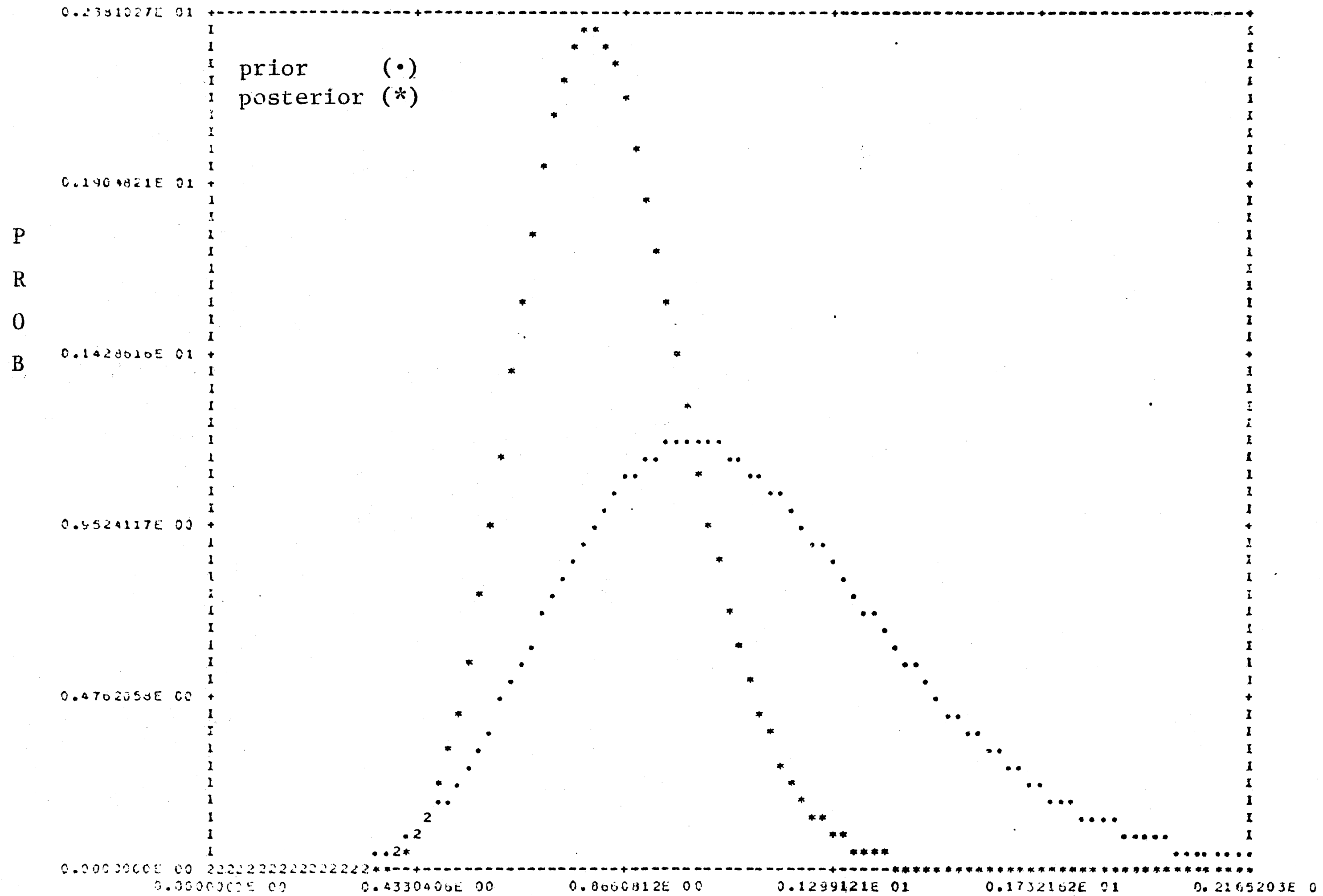

Figure 9. Prior and Posterior Distribution for τ: True $\tau=1$, True $\phi=.5$, $n=25$, $E(\tau)=1.1111$, $E(\phi)=.75$, N-G Prior for (ϕ,τ)

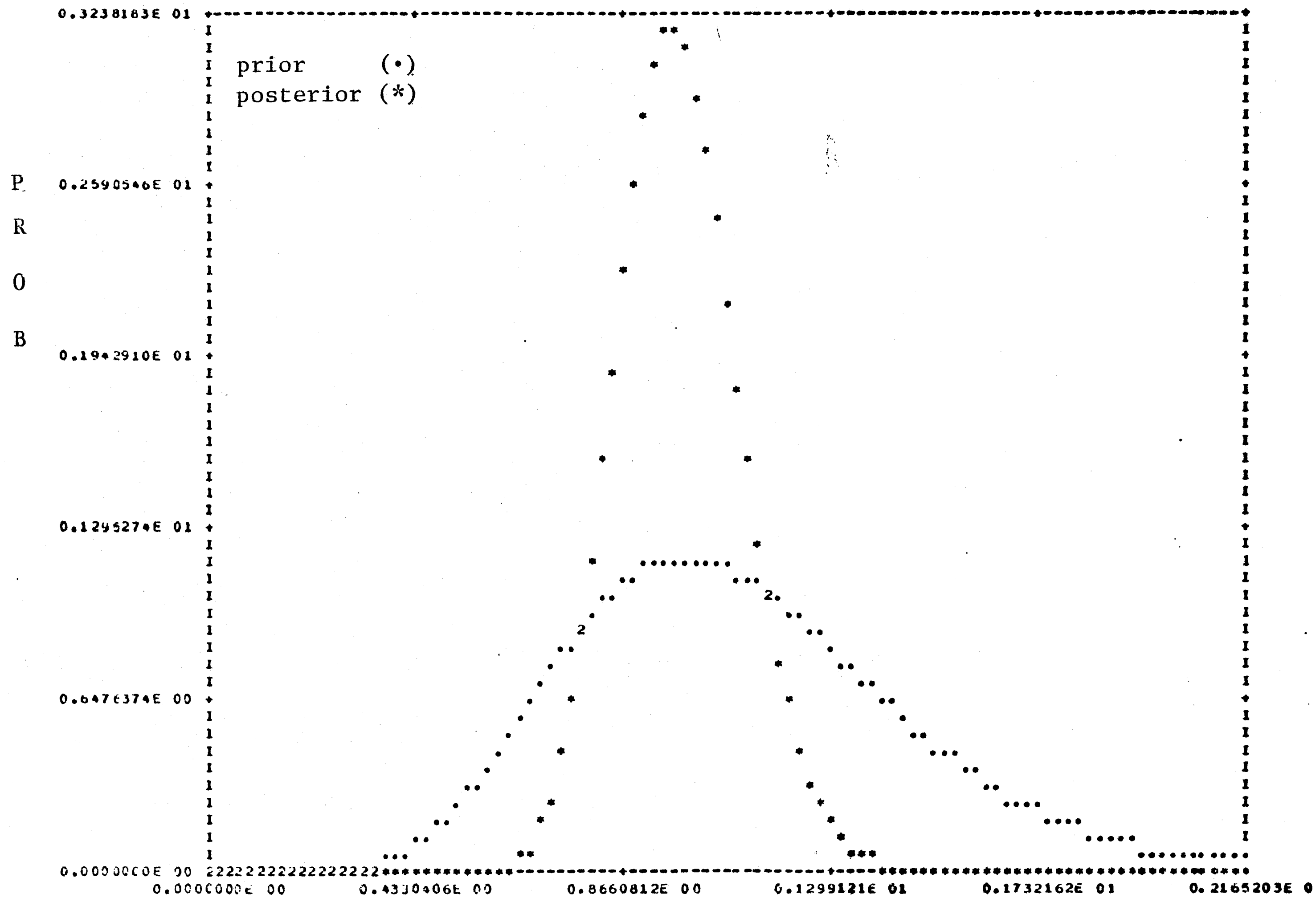

Figure 10. Prior and Posterior Distribution for τ: True $\tau=1$, True $\phi=.5$, $n=100$, $E(\tau)=1.1111$, $E(\phi)=.75$, N-G Prior for (ϕ,τ)

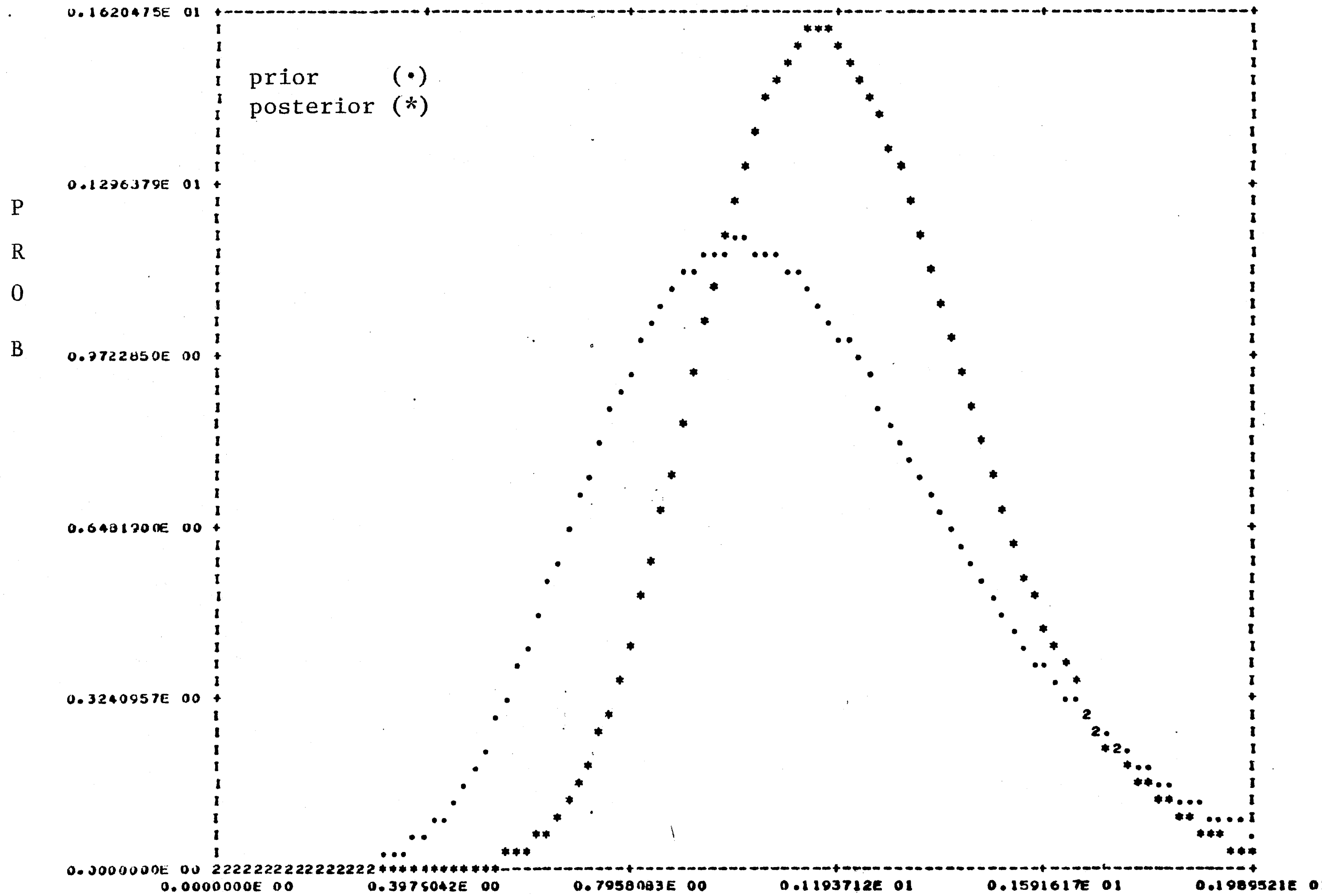

Figure 11. Prior and Posterior Distribution for τ: True τ=1, True ϕ=.5, n=25, $E(\tau)$=1.1111, $E(\phi)$=0. Mixture of N-G for (ϕ,τ), Prior ξ_i Fixed for μ=.5 for All i

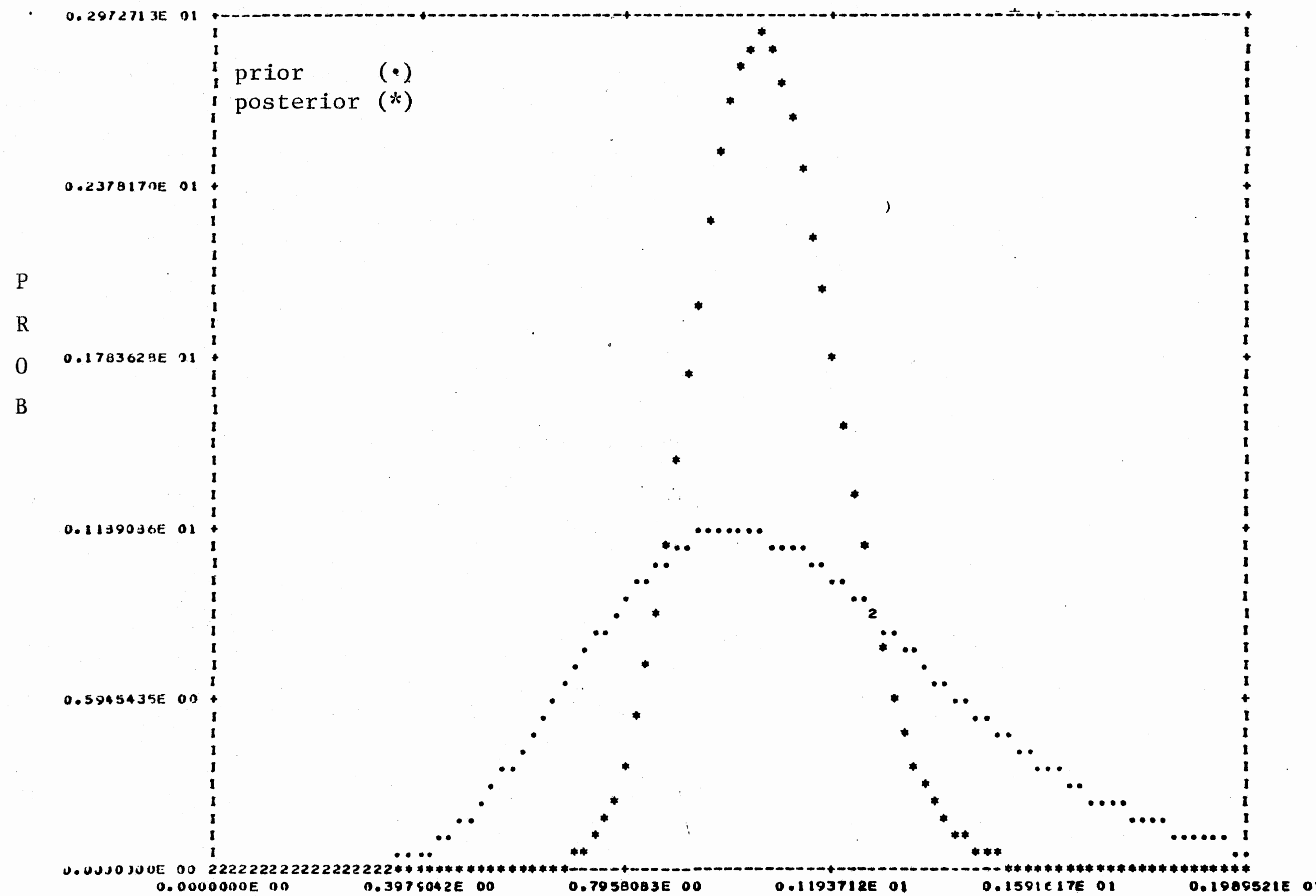

Figure 12. Prior and Posterior Distribution for τ: True τ=1, True ϕ=.5, n=100, $E(\tau)$=1.1111, $E(\phi)$=0. Mixture of N-G for (ϕ,τ), Prior ξ_i Fixed for μ=.5 for All i

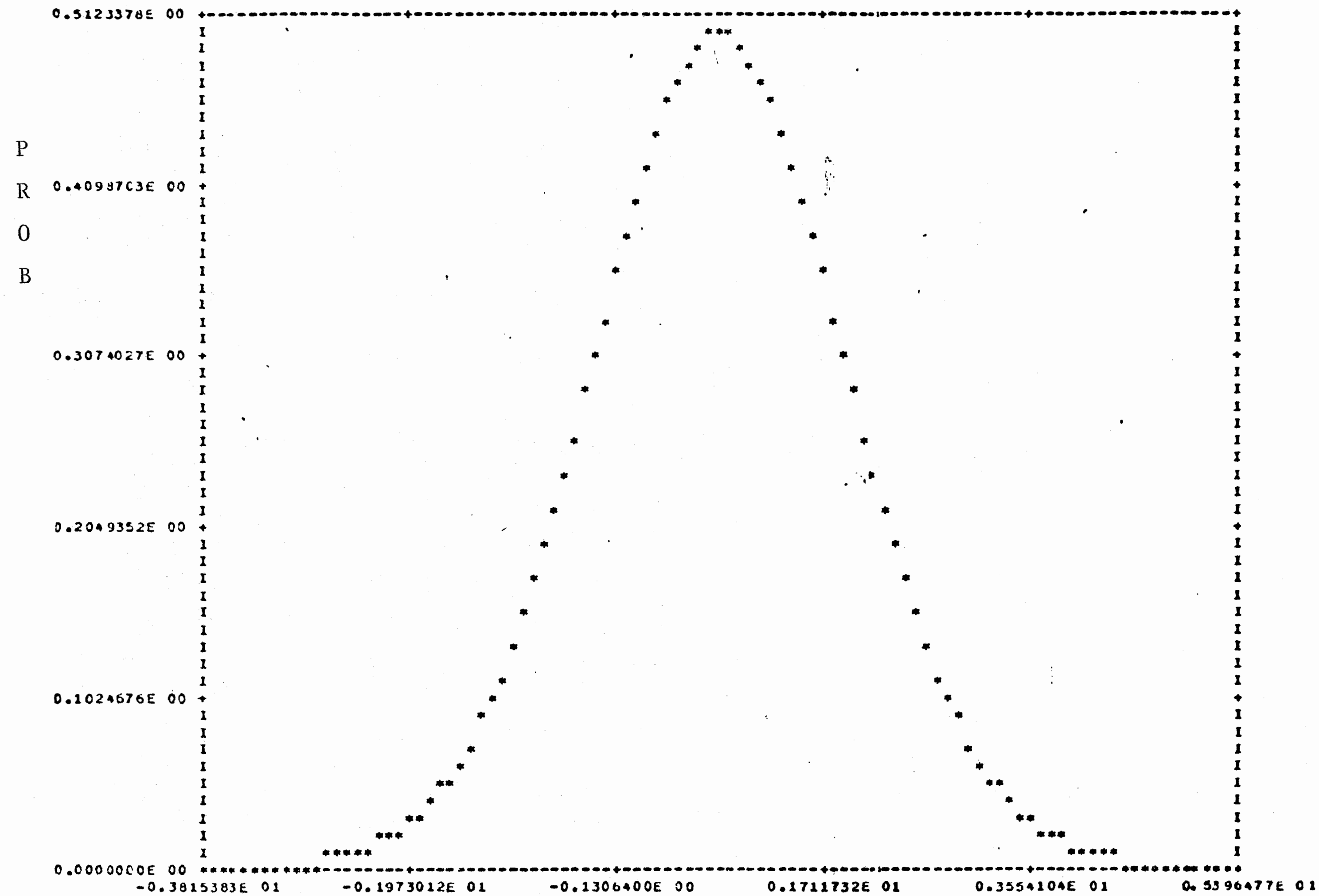

Figure 13. Predictive Prob. Density for Y(n+1), n=25, True τ=1, True ϕ=.5, E(ϕ)=0, N-G Prior for (ϕ,τ)

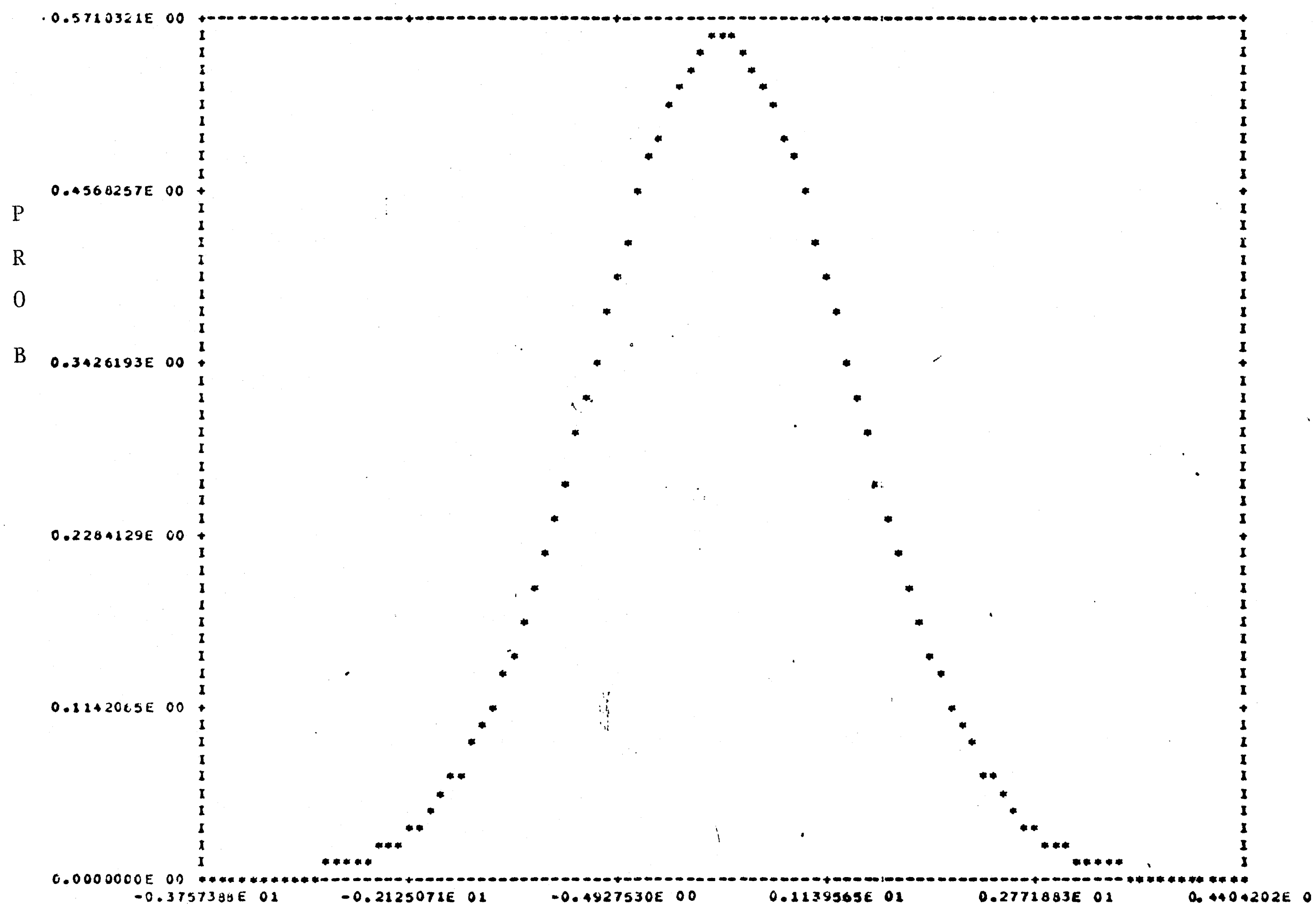

Figure 14. Predictive Prob. Density for Y(n+1), n=100, True τ=1, True ϕ=.5, E(ϕ)=0, N-G Prior for (ϕ,τ)

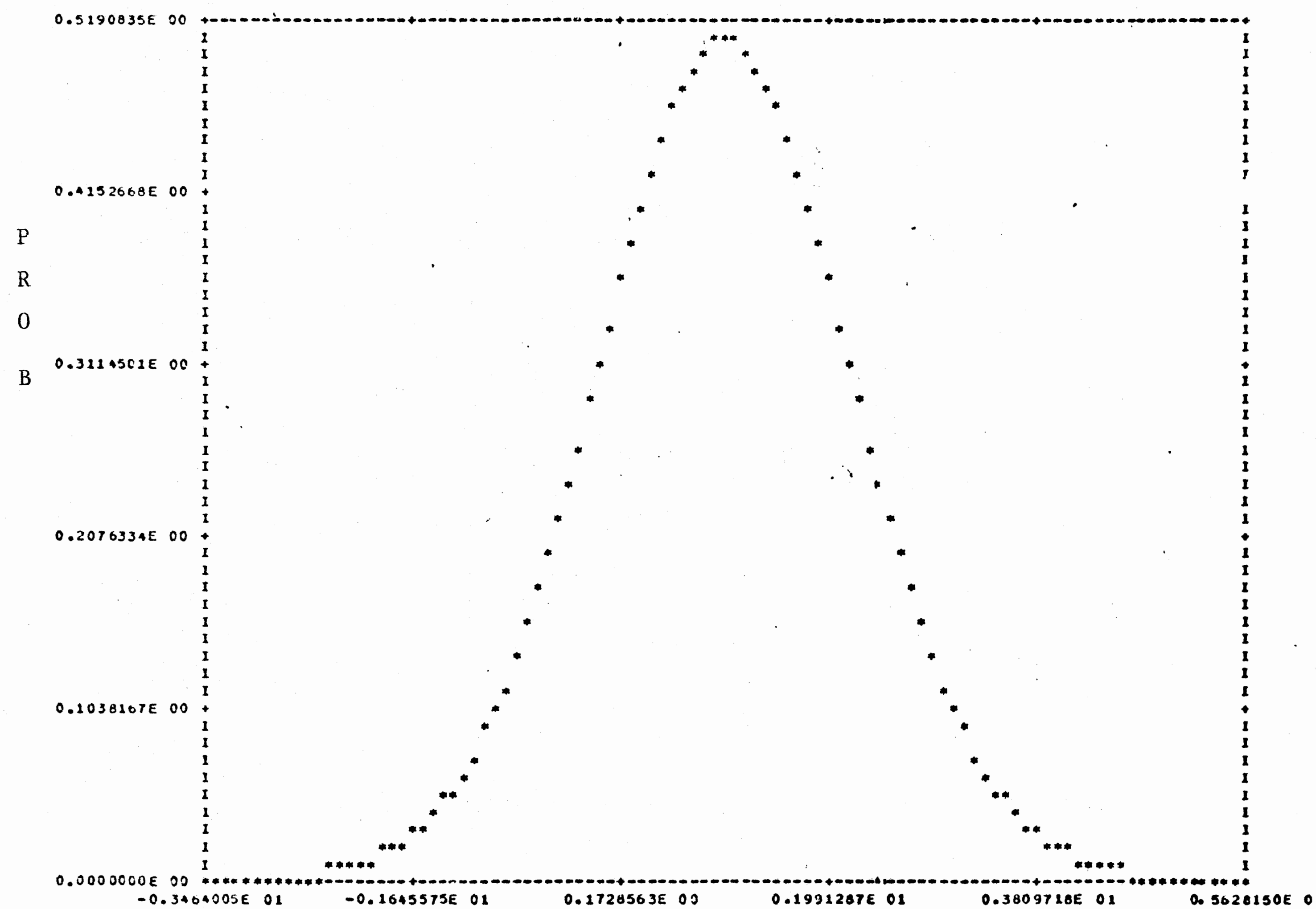

Figure 15. Predictive Prob. Density for Y(n+1), n=25, True τ=1, True ϕ=.5, E(ϕ)=.75, N-G Prior for (ϕ,τ)

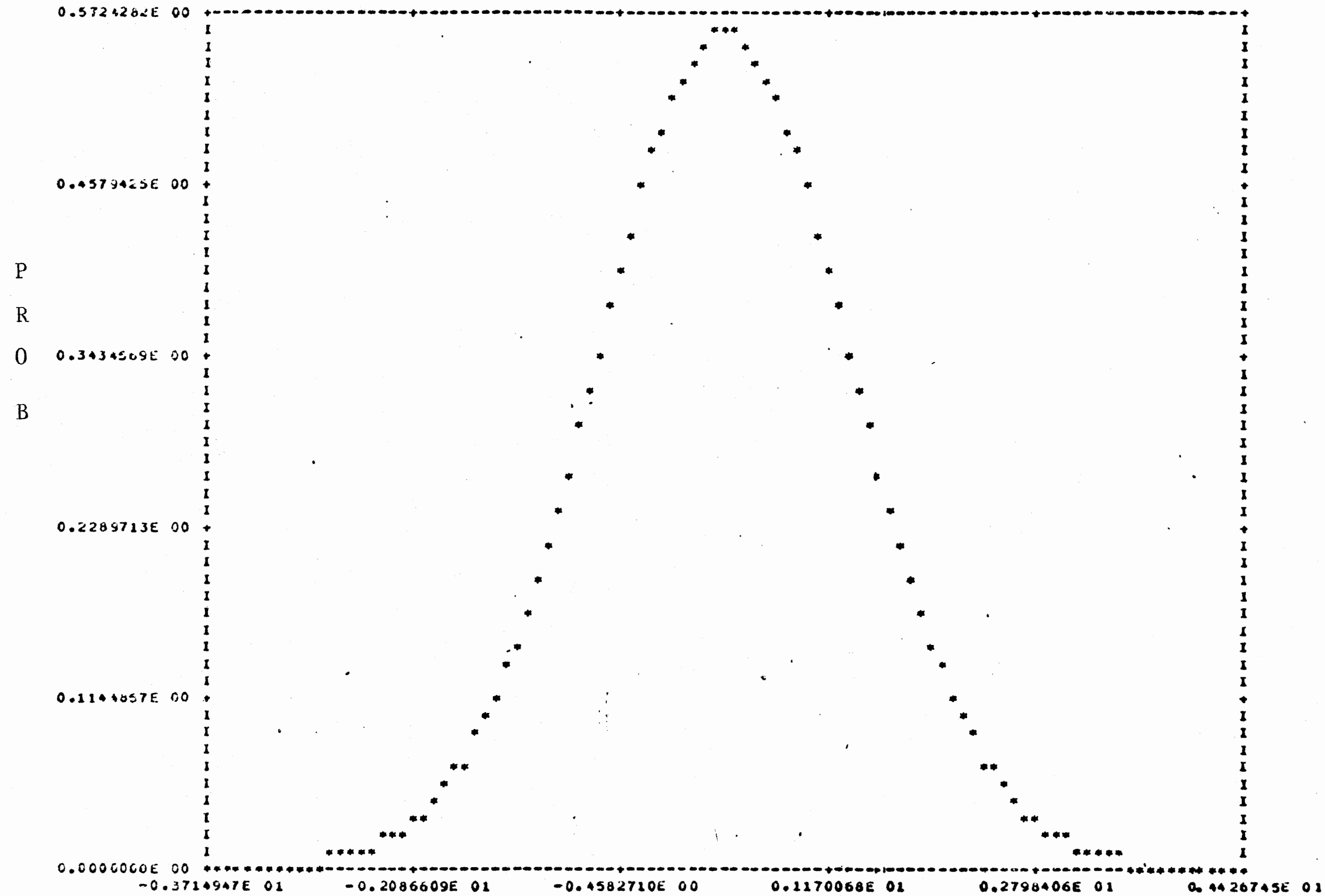

Figure 16. Predictive Prob. Density for Y(n+1), n=100, True τ=1, True ϕ=.5, E(ϕ)=.75, N-G Prior for (ϕ,τ)

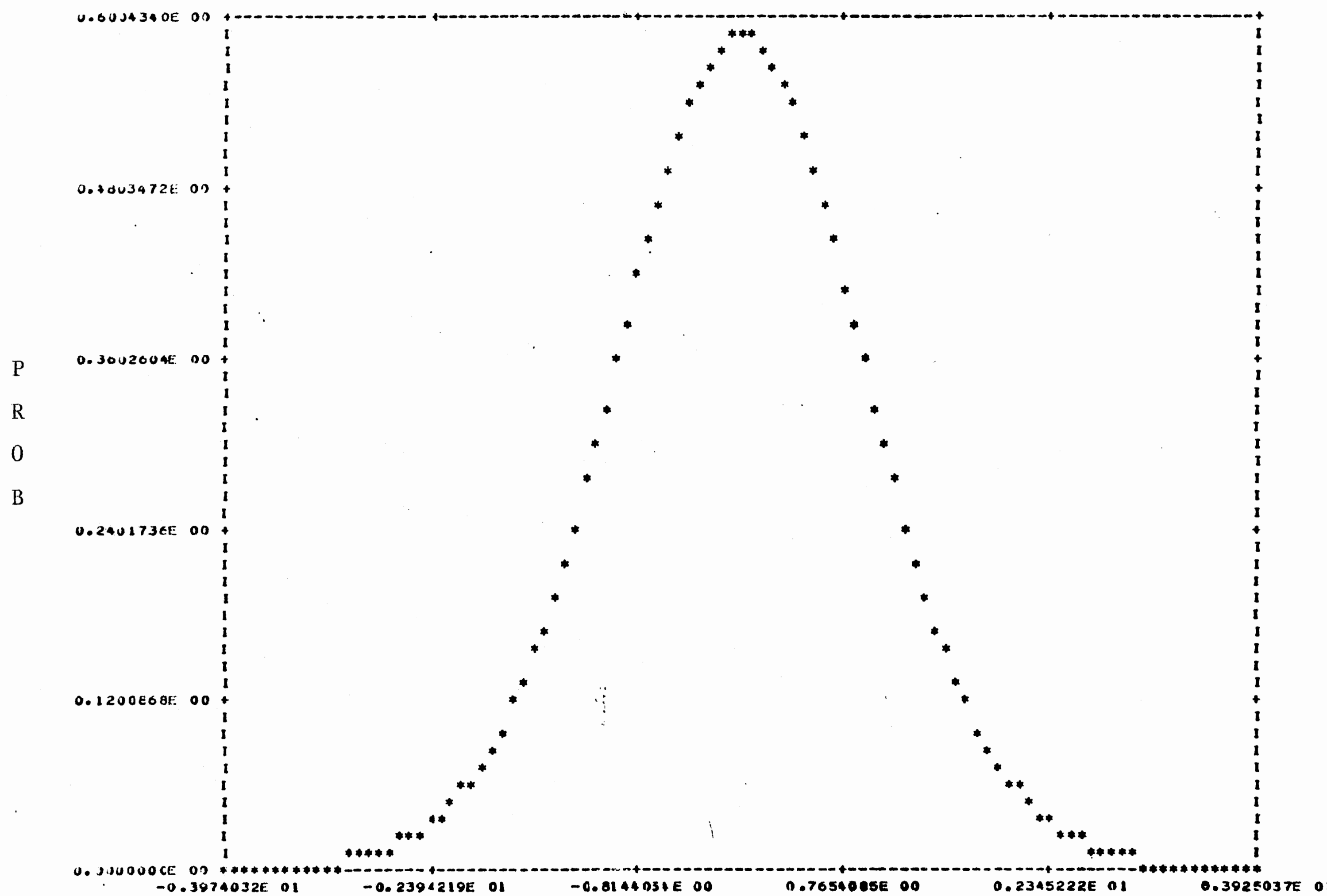

Figure 17. Predictive Prob. Density for Y(n+1), n=25, True τ=1, True ϕ=.5, E(ϕ)=0. Mixture of N-G Prior for (ϕ,τ)

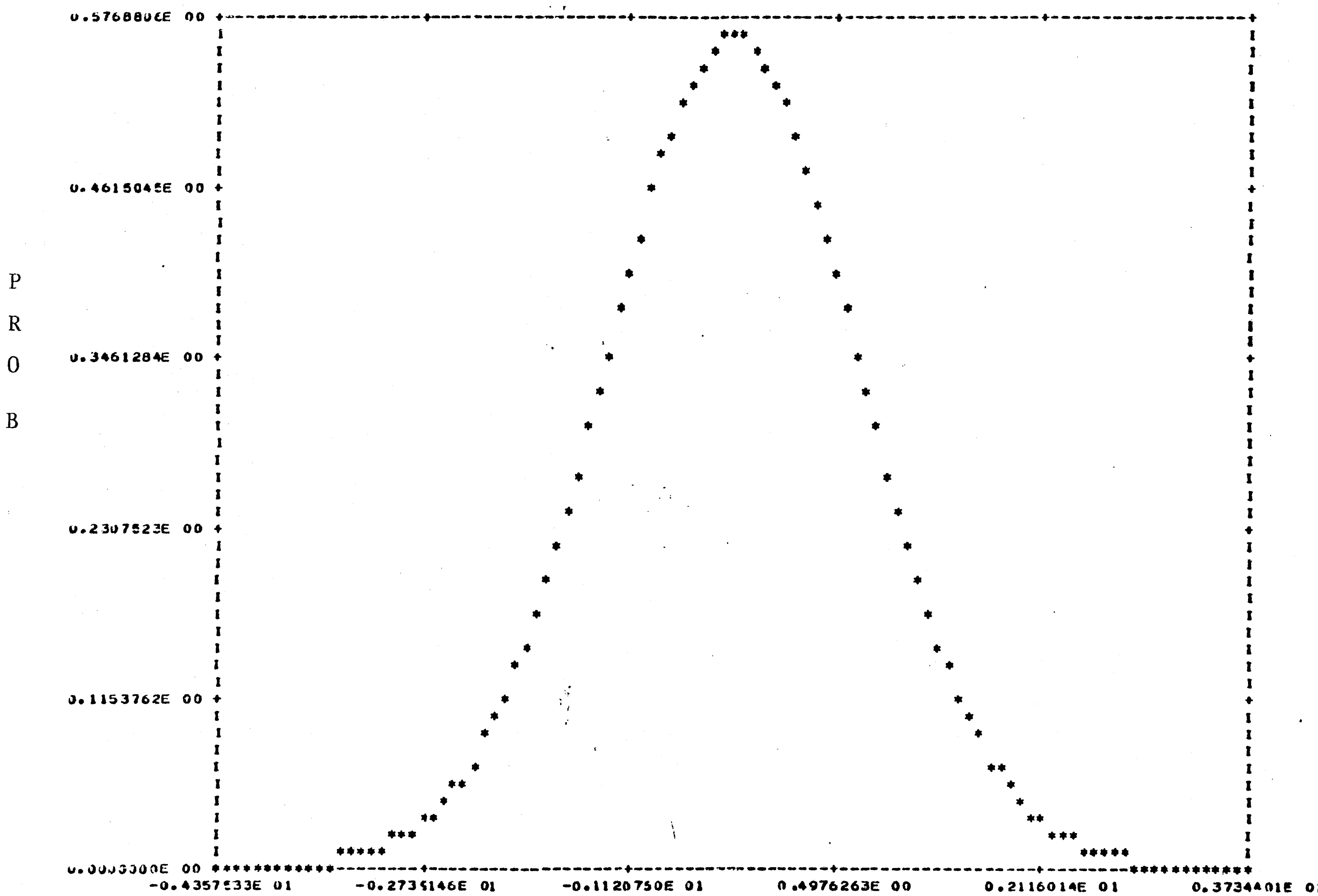

Figure 18. Predictive Prob. Density for Y(n+1), n=100, True τ=1, True ϕ=.5, E(ϕ)=0. Mixture of N-G Prior for (ϕ,τ)

VITA

Mohammad Abdulhadi Al-Mahmeed

Candidate for the Degree of

Doctor of Philosophy

Thesis: THE ANALYSIS OF AUTOREGRESSIVE PROCESSES: THE IDENTIFICATION AND THE PRIOR, POSTERIOR, AND PREDICTIVE ANALYSIS

Major Field: Statistics

Biographical:

Personal Data: Born in Kuwait City, Kuwait, June 3, 1953.

Education: Received the Bachelor of Statistics Degree from Kuwait University, Kuwait City, Kuwait, 1975; received Master of Science degree at Oklahoma State University in 1979; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in December, 1982.

Professional Experience: Demonstrator at Kuwait University, 1975-1976.

Professional Organizations: American Statistical Association.