DEPARTMENT OF STATISTICS
University of Wisconsin
Madison, Wisconsin

Technical Report No. 22

December, 1963

Bayesian Analysis of the Regression Model With Autocorrelated Errors *

by

Arnold Zellner and George C. Tiao

This research was supported in part by the National Science Foundation under Grant GS-151 and by the Wisconsin Alumni Research Foundation.

* Also issued as Systems Formulation and Methodology Workshop Paper No. 6313.

1. INTRODUCTION

In the present investigation we utilize Bayesian methods to analyze the regression model with errors generated by a first order auto-regressive scheme. For a simple regression model, we derive finite sample joint, conditional and marginal posterior distributions of the parameters of the model. With these distributions, an investigator can make inferences about parameters and investigate how departures from independence, very often encountered in economic data, affect his inferences about parameters. Further, this approach provides a unified treatment of non-explosive and explosive models and in fact yields results for deciding whether a process is or is not explosive. To illustrate application of the techniques, two sets of artificially generated data, one set from a non-explosive model and the other from an explosive model, are analyzed in detail. We then go on to develop techniques for a Bayesian analysis of the multiple regression model with autocorrelated errors.

2. SPECIFICATION OF MODEL AND DERIVATION OF POSTERIOR DISTRIBUTIONS

2.1 Specification of Model

Initially, we consider a simple regression model with an error term generated by a first order autoregressive process, that is,

(2.1a) $$y_t = \beta x_t + u_t$$

$$t = 1, 2, \ldots, T$$

(2.1b) $$u_t = \rho u_{t-1} + \epsilon_t ..$$

In (2.1a) β is a regression coefficient, y_t the t^{th} observation, x_t the t^{th} fixed element, and u_t the t^{th} error term.

Equation (2.1b) defines the autoregressive scheme generating the error term u_t which involves a parameter ρ and an error term ϵ_t. It is assumed that the ϵ_t are normally and independently distributed with zero means and common variance σ^2. From (2.1a-b), we obtain:

$$(2.2) \qquad y_t = \rho\, y_{t-1} + \beta(x_t - \rho\, x_{t-1}) + \epsilon_t \qquad t = 1, 2, \ldots, T.$$

We note that y_o appears in (2.2). Without assumptions regarding how y_o is generated, it is impossible to proceed with the analysis. Below we consider a range of assumptions appropriate for a variety of possible "real world" situations.

If we assume that the process represented by (2.1a-b) has been operative for $t = 0, -1, -2, \ldots, -T_o$, where T_o is unknown, we can write $y_o - \beta\, x_o = M + \epsilon_o$ where $M = \rho(y_{-1} - \beta\, x_{-1})$. M is regarded as a parameter since it depends on certain unobservable and unobserved quantities. Under these assumptions y_o is normally distributed with mean $\beta x_o + M$ and variance σ^2. These assumptions are broad enough to apply to explosive ($|\rho| \geq 1$) as well as non-explosive ($|\rho| < 1$) schemes and to situations in which the process commences at any unknown point in the past.

On the other hand, it may be that the situation being represented by the model in (2.1) is such that the initial value, y_o, is fixed and known. For example, if the observations relate to a price and if the period $t = 0$ is the last period during which this price was fixed by a governmental body, then it would be appropriate to take y_o as fixed and known. This situation can also be represented in the framework introduced in the preceding paragraph by assuming that ϵ_o has zero variance. Other assumptions which may be appropriate for other circumstances are that ϵ_o is normal with known variance, σ_o^2, or that y_o is distributed independently of $y_1, y_2, \ldots, y_T$ and has a distribution not involving any of the parameters of the model. As can readily be ascertained from what follows, any of these assumptions regarding y_o lead to the same joint posterior distribution for the parameters of the model.

Under the assumptions embedded in (2.1), the likelihood function for β, ρ, σ, and M is given by:

$$(2.2) \quad \ell(\beta,\rho,\sigma,M|y_o,y_1,\ldots,y_T) \propto \sigma^{-(T+1)} \exp\left\{-\frac{1}{2\sigma^2}(y_o-\beta x_o-M)^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}[y_t-\rho y_{t-1}-\beta(x_t-\rho x_{t-1})]^2\right\}$$

with $-\infty < \beta < \infty$, $-\infty < \rho < \infty$, $-\infty < M < \infty$, and $\sigma > 0$.

In the next section, this likelihood function is used in conjunction with Bayes' Theorem to derive posterior distributions for the parameters.

2.2 Derivation of Posterior Distributions

We assume that we are in a situation wherein our prior knowledge about the parameters β, ρ, M, and log σ, can be suitably represented by locally uniform and independent distributions [cf. Jeffreys (1948), Savage (1961) and Box and Tiao (1962)]; that is,

$$(2.3) \quad p(\beta) \propto k_1; \quad p(\rho) \propto k_2; \quad p(\sigma) \propto \frac{1}{\sigma} \quad \text{and } p(M) \propto k_3.$$

With these prior distributions and the likelihood function in (2.2), application of Bayes' Theorem leads to the following joint posterior distribution:

$$(2.4) \qquad p(\beta,\rho,\sigma,M|y_o,y_1,\ldots,y_T) = k\,\sigma^{-1}\,\ell(\beta,\rho,\sigma,M|y_o,y_1,\ldots,y_T,x)$$

where $\ell(\beta,\rho,\sigma,M|y_o,y_1,\ldots,y_T,x)$ is the likelihood function in (2.2) and k is a normalizing constant.

If one is interested in investigating M, the initial level of the process in (2.1b), it is possible to obtain the posterior distribution of M by integrating (2.4) over β, ρ, and σ. If interest does not center on M, the influence of this parameter can be eliminated by integration to yield:

$$(2.5) \qquad p(\beta,\rho,\sigma|y_o,y_1,\ldots y_T) = k\,\sigma^{-(T+1)}\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}[y_t-\rho y_{t-1}-\beta(x_t-\rho x_{t-1})]^2\right\}$$

which is the joint posterior distribution of β, ρ, and σ. We note that in obtaining (2.5), y_o was assumed normal with mean $M + \beta\, x_o$ and variance σ^2. It is straightforward to verify that employing the other assumptions about y_o discussed in the preceding section, one would, in each case, obtain the posterior distribution given in (2.5).

Upon eliminating the scale parameter σ from (2.5), we obtain the following bivariate posterior distribution:

$$(2.6) \qquad p(\beta,\rho|y) = k\left\{\Sigma[y_t-\beta x_t-\rho(y_{t-1}-\beta x_{t-1})]^2\right\}^{-T/2}$$

$$= k\left\{\Sigma[y_t-\rho y_{t-1}-\beta(x_t-\rho x_{t-1})]^2\right\}^{-T/2}.$$

This distribution summarizes all the information about β and ρ. Although the normalizing constant cannot be expressed in terms of simple functions, for any given set of data the joint density function can always be evaluated numerically and the density contours plotted. Further, the marginal distributions of β and of ρ are respectively:

(2.7) $$p(\beta|y) = k[\Sigma(y_{t-1}-\beta x_{t-1})^2]^{-\frac{1}{2}} \left\{\Sigma(y_t - x_t\beta)^2 - \frac{[\Sigma(y_{t-1}-\beta x_{t-1})(y_t-\beta x_t)]^2}{\Sigma(y_{t-1}-\beta x_{t-1})^2}\right\}^{-\frac{T-1}{2}}$$

(2.8) $$p(\rho|y) = k[\Sigma(x_t-\rho x_{t-1})^2]^{-\frac{1}{2}} \left\{\Sigma(y_t-\rho y_{t-1})^2 - \frac{[\Sigma(x_t-\rho x_{t-1})(y_t-\rho y_{t-1})]^2}{\Sigma(x_t-\rho x_{t-1})^2}\right\}^{-\frac{T-1}{2}} .$$

In order for the distribution in (2.8) to be proper, the quantity'

(2.8a) $$\sum_{t=1}^{T} (x_t - \rho\, x_{t-1})^2$$

must be positive. This implies that we must assume $x_t \neq \rho\, x_{t-1}$ for all ρ.

For illustrative purposes, we have computed these density functions with data generated from the following model:

$$y_t = 3\, x_t + u_t$$

$$u_t = \rho\, u_{t-1} + \epsilon_t$$

where the ϵ's given in Table I were drawn from a table of standardized random normal deviates. The x's are rescaled investment expenditures taken from Haavelmo (1953). The first series of 15 observations was generated with $\rho = 0.5$ while the second set, $\rho = 1.25$. Hereafter we shall refer to the first set as the "non-explosive" series and the second set as the "explosive" series. While we distinguish these two cases, it is important to realize that the results given in (2.7) and in (2.8) are appropriate in the analysis of both.

Table I

t	ϵ_t	x_t	y_t (for $\rho = .5$)	y_t (for $\rho = 1.25$)
0	--	3.0	9.500	9.500
1	.699	3.9	12.649	13.024
2	.320	6.0	18.794	19.975
3	-.799	4.2	12.198	14.270
4	-.927	5.2	14.372	16.760
5	.373	4.7	13.909	15.923
6	-.648	5.1	14.556	16.931
7	1.572	4.5	14.700	17.111
8	-.319	6.0	18.281	22.195
9	2.049	3.9	13.890	18.992
10	-3.077	4.1	10.318	18.338
11	-.136	2.2	5.473	14.012
12	-.492	1.7	4.044	13.873
13	-1.211	2.7	6.361	17.855
14	-1.994	3.3	7.036	20.099
15	.400	4.8	13.368	27.549

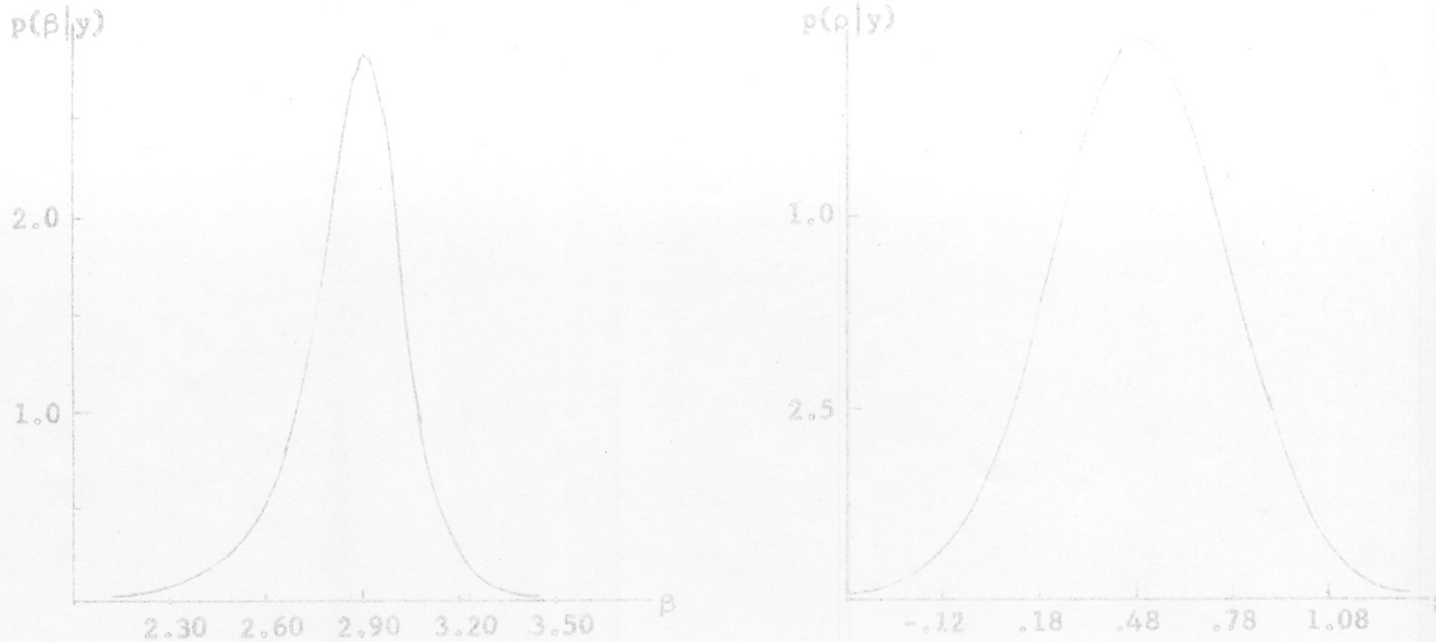
$u_o = 0.5$

The marginal distributions of β and of ρ for these data are shown in Fig. 1. It is seen that the posterior distribution of ρ derived from the explosive series is much sharper than that relating to the non-explosive case. As will be seen in the discussion in Section 3, one would indeed expect such a result.

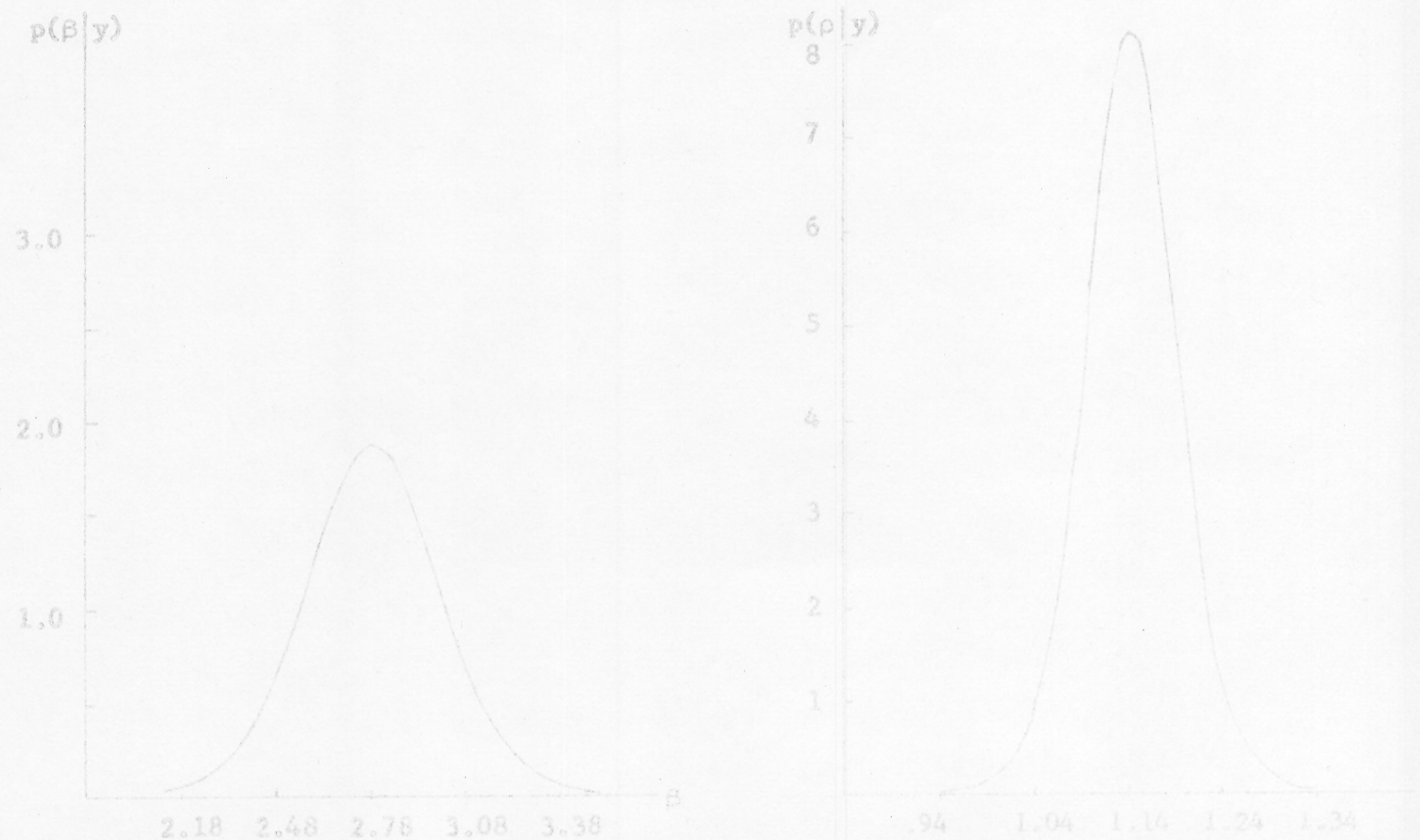
The posterior distributions of β in Fig. 1 enable an investigator to make inferences about this parameter which incorporate an allowance for the departure from independence postulated in the model. That allowance be made for such a departure is extremely important because

Fig. 1: MARGINAL DISTRIBUTIONS OF β and ρ

a. <u>Non-Explosive Series (T = 15)</u>



b. <u>Explosive Series (T = 15)</u>



inferences will be markedly different if one analyzed these data under the assumption of independence. It is well known that in the case of independence the posterior distribution of β can be expressed in terms of a Student-t distribution, that is,

$$(2.9) \qquad p\left(\frac{\beta - \hat{\beta}}{s} \mid y\right) = p(t_{T-1})$$

where $\hat{\beta} = \Sigma x_t y_t / \Sigma x_t^2$ and $s^2 = \Sigma (y_t - \hat{\beta} x_t)^2 / (T-1) \Sigma x_t^2$

as shown in Jeffreys (1948). It is to be remembered that in (2.9) $\hat{\beta}$ and s are regarded as known quantities calculated from the data. For our two sets of data the posterior distributions of β under the independence assumption are shown in Fig. 2 by the curves labelled $\rho = 0$. These distributions are far different from those shown in Fig. 1.

In order to appreciate the situation fully, it is instructive to write the marginal distribution of β as:

$$(2.10) \qquad p(\beta|y) = \int p(\beta|\rho, y)\, p(\rho|y)\, d\rho .$$

The integrand in (2.10) contains two factors, the conditional distribution of β, $p(\beta|\rho, y)$, and the marginal distribution of ρ, $p(\rho|y)$, given in expression (2.8). The conditional distribution of β for fixed values of ρ, $p(\beta|\rho, y)$, is obtained directly from (2.6) and is given by

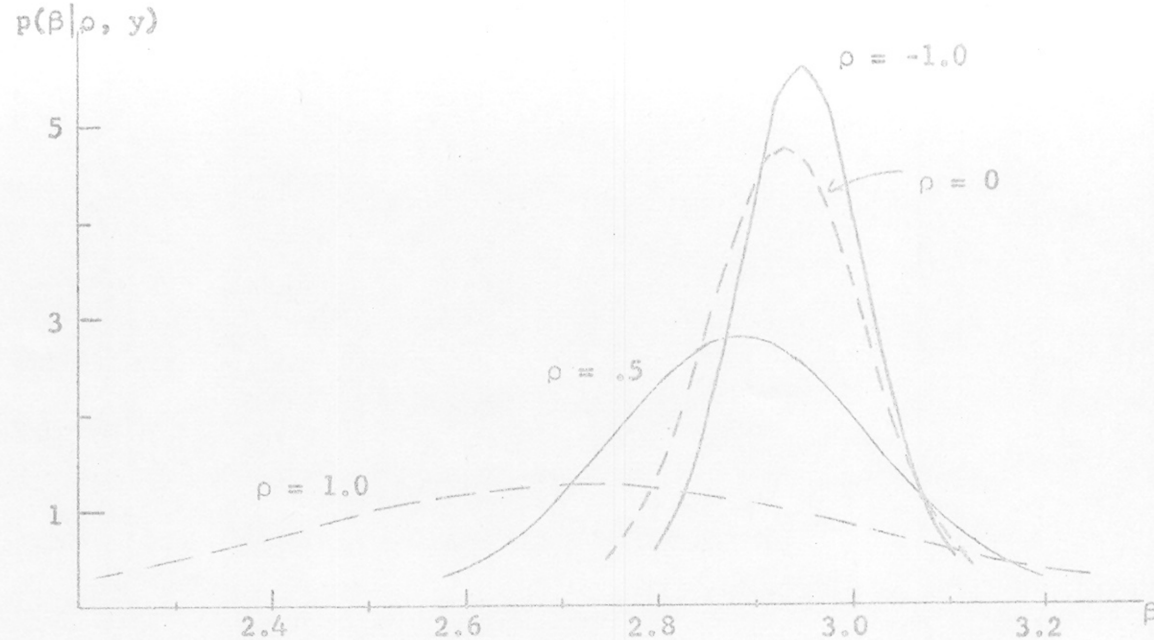
$$(2.11) \qquad p(\beta|\rho, y) = \frac{\Gamma\left(\frac{T}{2}\right)}{\Gamma\left(\frac{T-1}{2}\right)\sqrt{\Pi(T-1)}} \left\{s^2(\rho)\right\}^{-\frac{1}{2}} \left\{1 + \frac{[\beta - \hat{\beta}(\rho)]^2}{s^2(\rho)(T-1)}\right\}^{-\frac{T}{2}}$$

where (2.11a) $\hat{\beta}(\rho) = \Sigma(x_t - \rho x_{t-1})(y_t - \rho y_{t-1}) / \Sigma(x_t - \rho x_{t-1})^2$

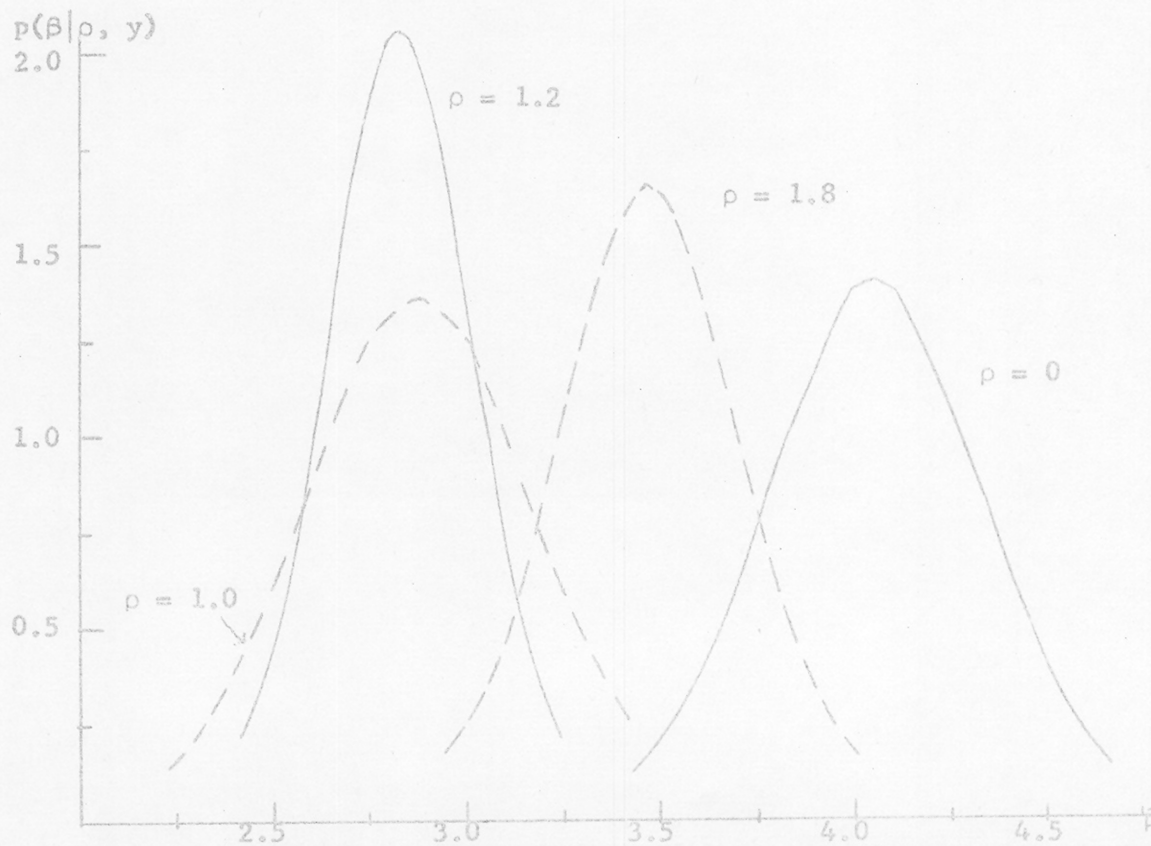
and (2.11b) $s^2(\rho) = \Sigma[y_t - \rho y_{t-1} - \hat{\beta}(\rho)(x_t - \rho x_{t-1})]^2 / (T-1)\, \Sigma(x_t - \rho x_{t-1})^2$.

Fig. 2: CONDITIONAL POSTERIOR DISTRIBUTION OF β FOR VARIOUS ρ

a. Non-Explosive Series (T = 15)



b. Explosive Series (T = 15)



It is clear that

$$(2.12) \qquad p\left(\frac{\beta - \hat{\beta}(\rho)}{s(\rho)} \mid y\right) = p(t_{T-1})$$

where t_{T-1} is a Student-t variable with (T-1) degrees of freedom. In particular, when $\rho = 0$, (2.12) reduces to (2.9).

The conditional distribution $p(\beta|\rho, y)$ provides inferences about β for an assumed value of ρ. On the other hand, the marginal density function $p(\rho|y)$ which appears as the other factor in the integrand of (2.10) reflects the plausibility of assertions about the value of ρ in the light of the data and our original assumptions. Thus the marginal distribution $p(\beta|y)$ in (2.10) can be regarded as a suitably weighted average of the conditional distributions $p(\beta|\rho, y)$ with $p(\rho|y)$ serving as the weight function. Unless the conditional distribution is insensitive to changes in ρ, it is clear that assuming ρ equals some fixed value, say $\rho = 0$ (corresponding to assuming the observations to be independent) or $\rho = 1$ (corresponding to the assumption that the first differences of the observations are independent), could lead to a posterior distribution of β far different from that given in (2.7). To illustrate this point, we have computed conditional distributions of β for various values of ρ which are plotted in Fig. 2. This figure shows that for the non-explosive series the center of the conditional distribution is relatively insensitive to changes in ρ whereas the spread of the distribution is quite sensitive to such changes. On the other hand, both the center and the spread in the case of the explosive series change markedly as ρ is varied. Thus an inappropriate assumption about ρ can vitally affect an analysis. This fact underlines the importance of working with the marginal posterior distribution of β which incorporates a proper allowance for the role of ρ in the model.

3. SOME SAMPLING THEORY CONSIDERATIONS

It is of interest to compare the above Bayesian analysis with analyses in the sampling theory framework. In the latter approach, one may investigate the sensitivity of the distribution of a specific estimator of β to departures from independence. For example, in our model, Wold (1949) shows that while the least squares estimator $\hat{\beta} = \Sigma x_t y_t / \Sigma x_t^2$ is unbiased for β, its variance is in general quite sensitive to the value which ρ assumes.

Alternatively, given a particular departure from independence, say $\rho = \rho_o$, an estimator with optimal properties is readily obtained. For we may then write the model in (2.2) as:

$$(3.1) \qquad y_t - \rho_o y_{t-1} = \beta(x_t - \rho_o x_{t-1}) + \epsilon_t$$

which is in the usual least squares form. It follows that the quantity

$$(3.2) \qquad t_{T-1} = \frac{\hat{\beta}(\rho_o) - \beta}{s(\rho_o)}$$

has a Student-t distribution with T-1 degrees of freedom where $\hat{\beta}(\rho_o)$ and $s(\rho_o)$, now regarded as random variables, are given in (2.11a-b). The properties of optimal testing and estimation procedures, which are seen to depend critically on the value assigned to ρ_o in (3.2), can be studied as a function of ρ_o. This latter analysis may be regarded as a direct analogue of our use of conditional posterior distributions in the Bayesian framework.

In the common situation in which <u>both</u> β and ρ are unknown, they must of course be estimated from the data. In the sampling theory framework it seems difficult to derive the distributions of optimal estimators of β and ρ even in the non-explosive case and only asymptotic results appear to be available; cf. e.g., Hurwicz (1950), Durbin (1960), Malinvaud (1961), White (1957, 1958) and the references in Anderson (1949). This contrasts with the Bayesian approach which provides a unified treatment of the explosive and non-explosive models and leads to finite sample results.

While we emphasize the positive contribution of the Bayesian approach, we recognize that many fruitful insights can be obtained from classical analyses. For example, it is of interest to evaluate the information matrix, I_θ, for $\theta = (M, \rho, \beta)$. We have for the joint density of $y_o, y_1, \ldots, y_T$:

$$P(y_o, y_1, \ldots, y_T | x, \beta, \rho, \sigma, M)$$

$$\propto \sigma^{-(T+1)} \exp\left\{-\frac{1}{2\sigma^2}(y_o - \beta x_o - M)^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}[y_t - \rho y_{t-1} - \beta(x_t - \rho x_{t-1})]^2\right\} = \sigma^{-(T+1)} \exp\left\{-\frac{1}{2\sigma^2} Q\right\}$$

and by definition:

$$I_\theta = -E\left\{\frac{\partial^2 \log P}{\partial \theta_i \, \partial \theta_j}\right\}$$

$$= \frac{1}{2\sigma^2} E\left\{\frac{\partial^2 Q}{\partial \theta_i \, \partial \theta_j}\right\}.$$

On performing the indicated differentiations, we find:

$$I_\theta = \frac{1}{2\sigma^2} E \begin{bmatrix} 1 & 0 & x_o \\ 0 & \sum_{t=1}^{T}(y_t - x_t\beta)^2 & \sum_{t=1}^{T}[x_{t-1}\epsilon_t + (x_t - \rho x_{t-1})(y_{t-1} - \beta x_{t-1})] \\ x_o & \sum_{t=1}^{T}[x_{t-1}\epsilon_t + (x_t - \rho x_{t-1})(y_{t-1} - \beta x_{t-1})] & x_o^2 + \sum_{t=1}^{T}(x_t - \rho x_{t-1})^2 \end{bmatrix}$$

To evaluate the expectations in this last expression, we utilize the following results:

$$y_t - \beta x_t = \rho^t M + \sum_{j=0}^{t} \rho^j \epsilon_{t-j}$$

$$E(y_t - \beta x_t) = \rho^t M \qquad t = 1, 2, \ldots, T$$

$$E(y_t - \beta x_t)^2 = \rho^{2t} M^2 + \sigma^2 \frac{1-\rho^{2(t+1)}}{1-\rho^2}$$

That part of the information matrix relating to ρ and β is then:

$$(3.3) \quad E \begin{bmatrix} \frac{\partial^2 Q}{\partial \rho^2} & \frac{\partial^2 Q}{\partial\rho\, \partial\beta} \\ \frac{\partial^2 Q}{\partial\rho\, \partial\beta} & \frac{\partial^2 Q}{\partial\beta^2} \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \frac{\sigma^2 (T+1)}{1-\rho^2} + \frac{M(1-\rho^{2T})}{1-\rho^2} - \frac{1-\rho^{2(T+1)}}{(1-\rho^2)^2} & M \sum_{t=1}^{T} (x_t - \rho x_{t-1})\rho^{t-1} \\ M \sum_{t=1}^{T} (x_t - \rho x_{t-1})\, \rho^{t-1} & x_o^2 + \sum_{t=1}^{T} (x_t - \rho x_{t-1})^2 \end{bmatrix}$$

In the case $|\rho| > 1$, information about ρ is extensive even in moderate-sized samples since the upper left-hand element of the information matrix in (3.3) is of order ρ^{2T}. This suggests that ρ can be estimated quite precisely in the explosive case. In the Bayesian approach, this phenomenon seems to be reflected by a posterior distribution for ρ which is sharply concentrated (see, for example, Figure 1b).

On the other hand, for $|\rho| < 1$ and relatively large T, the off-diagonal elements in (3.3) are small relative to the diagonal elements and thus the information matrix is approximately,

$$\frac{1}{\sigma^2} \begin{bmatrix} \frac{\sigma^2 T}{1-\rho^2} & 0 \\ 0 & \sum_{t=1}^{T} (x_t - \rho x_{t-1})^2 \end{bmatrix} .$$

This implies that the maximum likelihood estimates of β and ρ will be asymptotically uncorrelated. Also note that information about ρ is asymptotically independent of σ^2.

The sampling theory approach can also be utilized to provide an intuitive explanation of the fact that in our two computed examples, the conditional posterior mean of β is insensitive to ρ in the non-explosive case but very sensitive in the explosive case. That is, with

$$\hat{\beta}(\rho) = \sum_{t=1}^{T} (x_t-\rho x_{t-1})(y_t-\rho y_{t-1}) \Big/ \sum_{t=1}^{T} (x_t-\rho x_{t-1})^2$$

we have that

$$(3.4) \quad E\,\frac{\partial\hat{\beta}(\rho)}{\partial\rho} = -M \sum_{t=1}^{T}(x_t-\rho x_{t-1})\rho^{t-1} \Big/ \sum_{t=1}^{T}(x_t-\rho x_{t-1})^2 \;.$$

As T becomes large, we see that (3.4) approaches zero if $|\rho| < 1$ but grows without limit if $|\rho| > 1$. This suggests that in samples of moderate size, the conditional posterior mean of β will be insensitive to ρ if the data are generated from an non-explosive model, but sensitive if otherwise.

With these observations made, we now turn to discuss the multiple regression model with autocorrelated errors.

4. GENERALIZATION TO THE MULTIPLE REGRESSION MODEL

In this section we generalize the results in Section 2 to the multiple regression model with errors generated by a first order autoregressive process. Our model is:

(4.1a) $$y = X\beta + u$$

(4.1b) $$u = \rho\, u_{-1} + \epsilon$$

or alternatively,

(4.2) $$y = \rho\, y_{-1} + (X - \rho\, X_{-1})\, \beta + \epsilon$$

where $y' = (y_1, \ldots, y_T)$ and $y'_{-1} = (y_o, \ldots, y_{T-1})$ are $(1\times T)$ vectors of observations; $u' = (u_1, \ldots, u_T)$ and $u'_{-1} = (u_o, \ldots, u_{T-1})$ are $(1\times T)$ vectors of autocorrelated errors; $\beta' = (\beta_1, \ldots, \beta_K)$ is a $(1\times K)$ vector of regression coefficients; ρ is a scalar;

(4.3) $$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ \vdots & \vdots & & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{TK} \end{pmatrix} \text{ and } X_{-1} = \begin{pmatrix} x_{01} & x_{02} & \cdots & x_{0K} \\ \vdots & \vdots & & \vdots \\ x_{(T-1)1} & x_{(T-1)2} & \cdots & x_{(T-1)K} \end{pmatrix}$$

are $(T\times K)$ matrices of fixed elements; and $\epsilon' = (\epsilon_1, \ldots, \epsilon_T)$ is a $(1\times T)$ vector of random errors.

As in Section 2, we shall make the same assumptions about the distribution of the ϵ_t's, the prior distributions of ρ and σ, and the initial conditions. In addition, we shall assume that the regression coefficients are a priori locally independent and uniform, that is

(4.4) $$p(\beta) \propto \prod_{i=1}^{K} p(\beta_i) \propto c.$$

Under these assumptions, the joint posterior distribution of (β, ρ) is readily obtained as:

(4.5) $$p(\beta,\rho|y) \propto \left\{[(y-X\beta)-\rho(y_{-1}-X_{-1}\beta)]'\,[(y-X\beta)-\rho(y_{-1}-X_{-1}\beta)]\right\}^{-\frac{T}{2}}$$
$$\propto \left\{[(y-\rho y_{-1})-(X-\rho X_{-1})\beta]'\,[(y-\rho y_{-1})-(X-\rho X_{-1})\beta]\right\}^{-\frac{T}{2}}.$$

For any fixed value of ρ, the conditional distribution of β is

$$(4.6)\quad p(\beta|\rho,y) = \text{const.}\left\{1 + \frac{[\beta-\hat{\beta}(\rho)]'H[\beta-\hat{\beta}(\rho)]}{(T-K)\ s^2(\rho)}\right\}^{-\frac{T}{2}}$$

with

$$\text{const.} = \frac{\Gamma(\frac{T}{2})|H|^{\frac{1}{2}}\left\{s^2(\rho)\right\}^{-\frac{K}{2}}}{\Gamma(\frac{T-K}{2})\left\{\Pi(T-K)\right\}^{\frac{K}{2}}}$$

$$H = (X - \rho X_{-1})'(X - \rho X_{-1})$$

$$\hat{\beta}(\rho) = H^{-1}(X - \rho X_{-1})'(y - \rho y_{-1})$$

$$s^2(\rho) = \frac{1}{T-K}[y-\rho y_{-1}-\hat{\beta}(\rho)(X-\rho X_{-1})]'[y-\rho y_{-1}-\hat{\beta}(\rho)(X-\rho X_{-1})]\ .$$

The distribution in (4.6) is in the form of a multivariate Student-t distribution. This result is, of course, not surprising since for given ρ, (4.2) can be regarded as a usual regression model and it is well known [e.g., cf., Savage (1961) and Tiao and Zellner (1964)] that the posterior distribution of regression coefficients is of the Student-t form.

We note that in deriving the distribution in (4.6), it is implicitly assumed that the matrix H is positive definite for any fixed value of ρ. A sufficient condition for this to be so is given in the following lemma.

<u>Lemma:</u> Let X_*' be the K×(T+1) augmented matrix $X_*' = [x_o' \vdots X']$ where $x_o = (x_{01}, \ldots, x_{0K})$ and let $z' = (1, \rho, \rho^2, \ldots, \rho^T)$ be a 1×(T+1) vector. If Z and X_* are linearly independent, then H is positive definite.

<u>Proof:</u> It suffices to show that the matrix $X - \rho X_{-1}$ is of rank K. We can write

$$X - \rho X_{-1} = A X_*$$

where

$$A = \begin{bmatrix} -\rho & 1 & & & & \\ & -\rho & 1 & & & \\ & & \cdot & \vdots & \vdots & \vdots & \cdot \\ & & & & & -\rho & 1 \end{bmatrix}$$

is a T×(T+1) matrix with all elements not shown being zero. It is easily seen that A is of rank T and w = z is the only non-trivial solution of the system of equations Aw = 0. Since X_* and z are assumed linearly independent, there exists a (T+1)×(T-K) matrix C such that $B = [z \vdots X_* \vdots C]$ is a (T+1)×(T+1) non-singular matrix. Thus the rank of the product AB is T. But note that,

$$AB = [0 \vdots A\ X_* \vdots A\ C],$$

has only T non-zero columns. Hence the rank of A X_* must be K and the lemma follows.

One can easily establish that the above condition is in fact also necessary. This condition implies that any linear combination of the columns of X_*, the matrix of independent variables for periods 0, ..., T, must not satisfy an exact first order autoregressive scheme. This does not appear to us to be a very restrictive condition. For K = 1, it coincides with that given in connection with (2.8a).

To obtain the marginal posterior distribution of β, $p(\beta|y)$, and of ρ, $p(\rho|y)$, we simply perform the following integrations:

$$(4.7) \qquad p(\beta|y) = \int_{-\infty}^{\infty} p(\beta,\rho|y)d\rho = \int_{-\infty}^{\infty} p(\rho|y)\ p(\beta|\rho,y)\ d\rho$$

and

$$(4.8) \qquad p(\rho|y) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(\beta,\rho|y) \prod_{i=1}^{K} d\beta_i = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(\beta|y)\ p(\rho|\beta,y) \prod_{i=1}^{K} d\ \beta_i\ .$$

It is clear that each of these integrations can be interpreted as an averaging of conditional distributions with the weight function being a marginal posterior distribution. On performing the integrations in (4.7) and (4.8), we obtain:

$$(4.7a)\quad p(\beta|y) \propto [(y_{-1}-X_{-1}\beta)'(y_{-1}-X_{-1}\beta]^{-\frac{1}{2}}\left\{(y-X\beta)'(y-X\beta) - \frac{[(y-X\beta)'(y_{-1}-X_{-1}\beta)]^2}{(y_{-1}-X_{-1}\beta)'(y_{-1}X_{-1}\beta)}\right\}^{-\frac{T-1}{2}}$$

$$(4.8a)\quad p(\rho|y) \propto \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{(T-K)s^2(\rho)+[\beta-\hat{\beta}(\rho)]'\,H[\beta-\hat{\beta}(\rho)]\right\}^{-\frac{T}{2}}\prod_{i=1}^{K} d\,\beta_i$$

$$\propto |H|^{-\frac{1}{2}}[(T-K)\;s^2(\rho)]^{-\frac{T-K}{2}}$$

$$\propto |H|^{-\frac{1}{2}}\left\{(y-\rho y_{-1})'[I-(X-\rho X_{-1})H^{-1}(X-\rho X_{-1})'](y-\rho y_{-1})\right\}^{-\frac{T-K}{2}}.$$

Note that the conditions of our lemma insure that H is positive definite and thus the distribution in (4.8a) is proper.

If interest centers on the marginal posterior distribution of a single element of β, say β_1, its posterior distribution can be obtained in principle from (4.7a) by integration. However, this integration, when viewed analytically or numerically, appears quite difficult to the present writers particularly when K is large. Therefore as an alternative, we suggest first obtaining $p(\beta_1, \rho|y)$ and then deriving $p(\beta_1|y)$ by integrating out ρ numerically. Note that

$$(4.9)\qquad p(\beta_1, \rho|y) = p(\rho|y)\; p(\beta_1|\rho, y)$$

with $p(\rho|y)$ given in (4.8a) and $p(\beta_1|\rho,y)$ is obtained from (4.6) by integration with respect to the elements of β other than β_1. It is well-known from the properties of the multivariate t-distribution that

$$(4.10)\qquad t = \frac{\beta_1 - \hat{\beta}_1(\rho)}{s(\rho)\sqrt{h^{11}}}$$

has a Student t-distribution with T-K degrees of freedom. In (4.10), h^{11} denotes the $(1,1)^{th}$ element of H^{-1}.

For greater computational simplicity, we can obtain $p(\beta_1, \rho|y)$ in a different form by integrating $p(\beta, \rho|y)$ in (4.5) with respect to $\beta_2, \ldots, \beta_K$. To perform this integration, we partition $\beta' = (\beta_1 \vdots \bar{\beta}')$, $X = (x \vdots \bar{X})$ and $X_{-1} = (x_{-1} \vdots \bar{X}_{-1})$, where x and x_{-1} denote the first column of X and X_{-1}, respectively. Then with

$$(4.11)\qquad W = y - \rho\, y_{-1} - (x - \rho\, x_{-1})\, \beta_1,$$

we have

$$p(\beta_1, \bar{\beta}, \rho|y) \propto \left\{[W - (\bar{X} - \rho\bar{X}_{-1})\bar{\beta}]'[W - (\bar{X} - \rho\bar{X}_{-1})\bar{\beta}]\right\}^{-\frac{T}{2}}.$$

Integration with respect to $\bar{\beta}$, performed as indicated above, yields:

$$(4.12)\qquad p(\beta_1, \rho|y) \propto |\bar{H}|^{-\frac{1}{2}} \left\{W'[I - (\bar{X} - \rho\bar{X}_{-1})\bar{H}^{-1}(\bar{X} - \rho\bar{X}_{-1})']\, W\right\}^{-\frac{T-K+1}{2}}$$

where

$$(4.13)\qquad \bar{H} = (\bar{X} - \rho\,\bar{X}_{-1})'\,(\bar{X} - \rho\,\bar{X}_{-1}).$$

The posterior distribution of β_1 can be obtained from (4.12) by numerical integration over ρ. The advantage of the form (4.12) is that its use involves inverting a $(K-1)\times(K-1)$ matrix, $\bar{H}$, whereas use of (4.9) would involve inverting a $K\times K$ matrix H. We note further that $\bar{H}$ is a λ-matrix of second degree in ρ. Thus the inverse can be expressed as a λ-matrix of degree $2(K-1)$ in ρ divided by a scalar polynomial of degree 2K in ρ. Putting the inverse of $\bar{H}$ in such a form is computationally convenient since this will avoid the necessity for inverting a matrix for each value of ρ in the integration.

5. A LARGE SAMPLE PROCEDURE

In this section we discuss a procedure for analyzing the multiple regression model with autocorrelated errors which can be conveniently applied when we work with large samples. Essentially, this procedure is the Bayesian analogue of a sampling theory approach suggested by Fuller (1962) and involves linearizing our model and applying linear theory to the linearized model. The goodness of this approximation can be checked within the Bayesian framework since we have the finite sample results of Section 2 and 4.

From (4.2), our model is:

$$(5.1) \qquad y = \rho\, y_{-1} + (X - \rho\, X_{-1})\,\beta + \epsilon$$

with $\rho\beta$ being our non-linearity. If we expand $\rho\beta$ about the maximum-likelihood estimates, say $\hat{\rho}$ and $\hat{\beta}$, we obtain:

$$(5.2) \qquad y \doteq \rho y_{-1} + X\beta - X_{-1}\,[\hat{\rho}\hat{\beta} + (\rho-\hat{\rho})\hat{\beta} + \hat{\rho}(\beta-\hat{\beta})] + \epsilon$$

or

$$(5.3) \qquad y - X_{-1}\hat{\rho}\hat{\beta} \doteq \rho(y_{-1} - X_{-1}\hat{\beta}) + (X - \hat{\rho}X_{-1})\beta + \epsilon$$

which is linear in the parameters ρ and β. With the locally uniform prior distributions with which we have been working, application of linear theory leads to a posterior distribution of ρ and β in the multivariate-t form.

To apply this approximation procedure, we require the maximum-likelihood estimates $\hat{\rho}$ and $\hat{\beta}$. These can be obtained using non-linear regression techniques in connection with (5.1) [cf. e.g., Fuller (1962), Box (1958)]. However, it appears computationally more efficient to utilize a step-wise procedure, suggested in Cochrane and Orcutt (1949), to minimize $\epsilon' \epsilon$ with respect to ρ and β. From our model in (5.1), it is seen that the conditional minimum of $\epsilon' \epsilon$ for a given ρ will be attained if we take

$$(5.4) \quad \beta(\rho) = [(X-\rho X_{-1})' (X-\rho X_{-1})]^{-1} (X-\rho X_{-1})' (y-\rho y_{-1})$$

whereas for a given β, the conditional minimizing value of ρ is given by:

$$(5.5) \quad \rho(\beta) = [(y_{-1}-X_{-1}\beta)' (y_{-1}-X_{-1}\beta)]^{-1} (y_{-1}-X_{-1}\beta)' (y-X\beta).$$

Thus, we can choose an initial ρ, say $\rho = \rho_o$, compute (5.4) to obtain a β_o. Substitute this value of β in (5.6) to obtain a new value of ρ, say ρ_1, and so on. When the computed values of β and ρ become stable, we have the minimizing values of ρ and β, namely $\hat{\rho}$ and $\hat{\beta}$, which are maximum-likelihood estimates if these values are associated with the <u>global</u> maximum of the likelihood function.

Since the problem of local maxima of the likelihood function may arise, it is suggested that the following procedure, described in Klein (1953), may be the safest to utilize. In

$$(5.6) \quad \epsilon' \epsilon = [y-\rho y_{-1}-(X-\rho X_{-1})\beta]' [y-\rho y_{-1}-(X-\rho X_{-1})\beta]$$

we substitute the conditional minimizing value of β given in (5.4) to obtain

$$(5.7) \quad \epsilon' \epsilon = [y-\rho y_{-1}-(X-\rho X_{-1})\beta(\rho)]' [y-\rho y_{-1}-(X-\rho X_{-1})\beta(\rho)]$$

$$= (y-\rho y_{-1})'(y-\rho y_{-1})-(y-\rho y_{-1})'(X-\rho X_{-1})\beta(\rho)$$

which we minimize with respect to ρ. The necessary condition on ρ will be in the form:

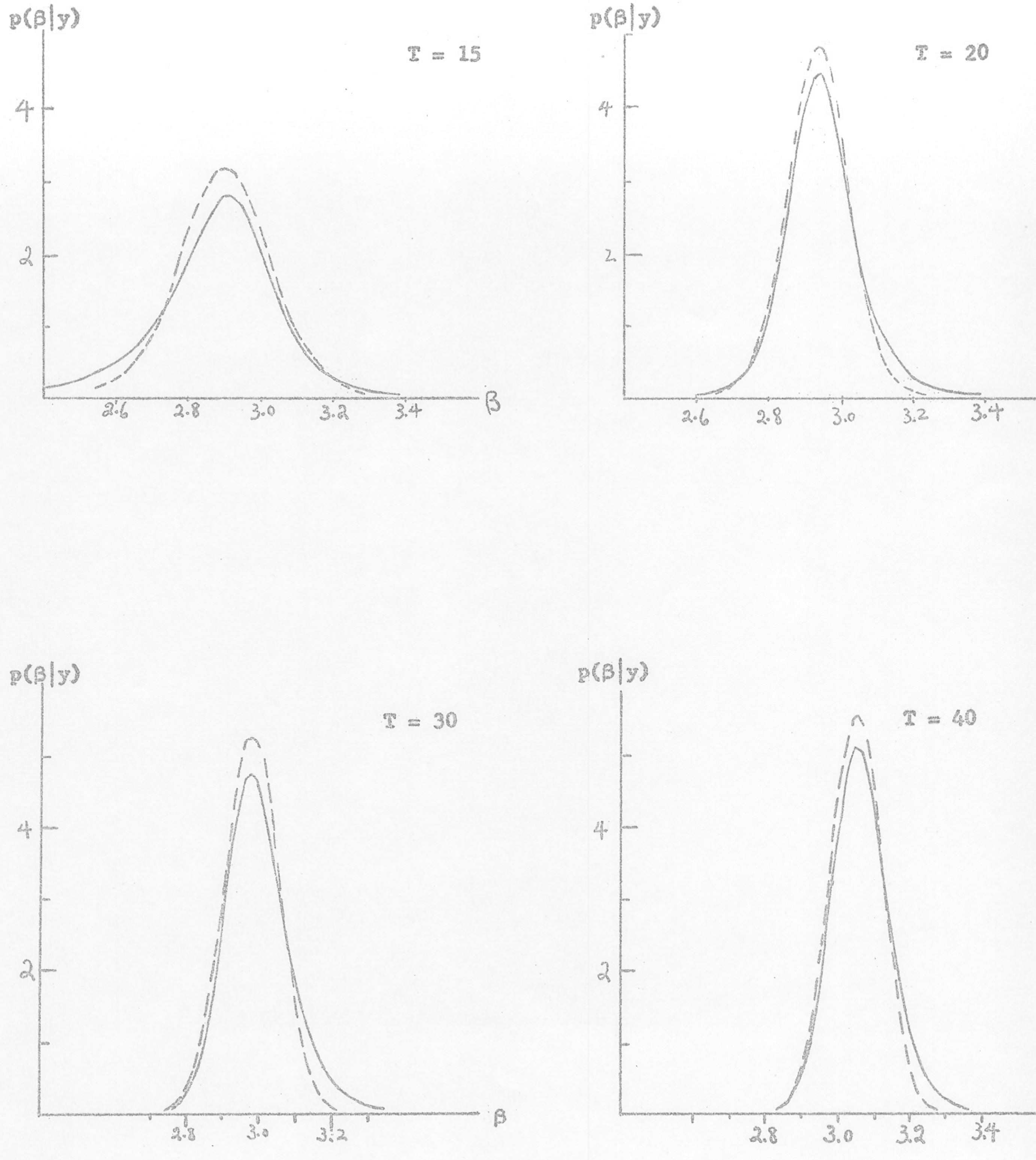
$$f(\rho) = 0, \tag{5.8}$$

where $f(\rho)$ is a polynomial of degree $4K + 1$ in ρ. The roots of the polynomial equation (5.9) can readily be obtained using standard numerical procedures. For each real root, evaluate (5.7) to determine which one is associated with the global minimum. Then use (5.4) to compute $\hat{\beta}$.

We have applied the linearization procedure to data generated from our non-explosive model described in Section 2 for samples of sizes 15, 20, 30 and 40. In Figure 3, the resulting approximate posterior distributions for our scalar β are compared with the exact distributions computed from (2.7). It is seen that for $T = 40$, the approximate and exact distributions are in fair agreement.

Fig. 3: EXACT AND APPROXIMATE MARGINAL DISTRIBUTION OF β FOR SEVERAL SAMPLE SIZES AND NON-EXPLOSIVE SERIES

(___ Exact; --- Approximate)



REFERENCES

Anderson, R. L. (1949). "The Problem of Autocorrelation in Regression Analysis," Journal of the American Statistical Association, Vol. XLIX, 113-129.

Box, G. E. P. (1958). "Use of Statistical Methods in the Elucidation of Basic Mechanisms," Bulletin de l'Institut International de Statistique, Vol. 36, 215-225.

Box, G. E. P. and Tiao, G. C. (1962). "A Further Look at Robustness via Bayes' Theorem," Biometrika, Vol. 49, 419-433.

Cochrane, D. and Orcutt, G. H. (1949). "Application of Least Squares Regressions to Relationships Containing Auto-correlated Error Terms," Journal of the American Statistical Association, Vol. 44, 32-61.

Durbin, J. (1960). "Estimation of Parameters in Time-series Regression Models," Journal of the Royal Statistical Society, (Series B), Vol. 22, 139-153.

Fuller, Wayne A. and Martin, James E. (1962). "The Effects of Autocorrelated Errors on the Statistical Estimation of Distributed Lag Models," Journal of Farm Economics, Vol. 44, p. 407.

Haavelmo, T. (1947). "Methods of Measuring the Marginal Propensity to Consume," Journal of the American Statistical Association, Vo. 42, 105-122.

Hurwicz, L. (1950). "Least Squares Bias in Time Series," in T. C. Koopmans (ed.), Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph No. 10, J. Wiley and Sons, New York, 365-383.

Jeffreys, H. (1948). Theory of Probability (2nd ed.), Clarendon Press, Oxford.

Klein, L. R. (1953). A Textbook of Econometrics, Row, Peterson and Company, Evanston, Illinois.

Malinvaud, E. (1961). "Estimation et Prevision dans les Modeles Economiques Autoregressifs," Rev. Inst. Int. de Statistique, Vol. 29.

Savage, L. J. (1961). The Subjective Basis of Statistical Practice, (manuscript), University of Michigan.

Tiao, G. C. and Zellner, A. (1964). "Bayes' Theorem and the Use of Prior Knowledge in Regression Analysis," to appear in Biometrika.

White, J. S. (1957). "Approximate Moments for the Serial Correlation Coefficient," Annals of Mathematical Statistics, Vol. 28, 798-802.

____ , (1958). "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," Annals of Mathematical Statistics, Vol. 29, 1188-1197.

Wold, H. O. A. (1949). "On Least Square Regression with Autocorrelated Variables and Residuals," Proceedings of the International Institute of Statistics, Berne, 277-289.