

A BAYESIAN ANALYSIS OF SOME THRESHOLD SWITCHING MODELS

A.M. POLE and A.F.M. SMITH

University of Nottingham, Nottingham, NG7 2RD, England

This paper presents a Bayesian analysis of various threshold switching regression models, including simple time series models, where the change of regime is governed by a known function of exogenous variables. Some special features arising from the choice of a two-dimensional linear dichotomy function are then discussed and the concepts of concurrency and duality introduced. Within this framework, we compare the maximum likelihood and Bayesian methodologies for inference and prediction. In particular, we show that the Bayesian approach solves the non-uniqueness problem which affects maximum likelihood prediction in certain situations. These results are illustrated with two numerical examples.

1. Introduction

The term switching regression has been used to describe a number of different types of model discussed in the literature. A comprehensive list would include two-phase (or changepoint) models [Quandt (1960), Hinkley (1969, 1971)] from the maximum likelihood viewpoint, and Ferreira (1975), Chin-Choy and Broemeling (1980), Booth and Smith (1982), Holbert (1982) from the Bayesian viewpoint]; stochastic choice of regime, which has nature choosing between two competing regimes with probability λ [Goldfeld and Quandt (1973a, b), Kieffer (1978, 1979)]; the Markov model, which extends the λ model to include a probability transition matrix [Goldfeld and Quandt (1973a, b)]; time-trending regression [Farley and Hinich (1970), Farley, Hinich and McGuire (1975), Hackl (1980)]; variable parameter regression, more familiar to statisticians as the dynamic linear model [Garbade (1977), Griffiths, Dryman and Prakash (1979)]; disequilibrium models [Goldfeld and Quandt (1975), Fair and Jaffee (1972), Maddala and Nelson (1974), Hartley and Mallela (1977), Quandt (1982) from the maximum likelihood perspective, and Lubrano (1983) from the Bayesian perspective]; and the threshold switching model [Goldfeld and Quandt (1972, 1973a), Tishler and Zang (1979)].

Various applications of these models have been proposed in the literature. The Housing Starts model, first discussed by Fair and Jaffee (1972) in a disequilibrium context, was reformulated as a switching model with a single-variable dichotomy function by Tishler and Zang (1979). If the disturbance terms are autocorrelated, as assumed by Fair and Jaffee, then we cannot simply

reorder the data in terms of the dichotomy variable and proceed with an analysis of the standard changepoint model. The more general threshold switching formulation is required. Furthermore, it is often the case that decisions are based upon the cumulative evidence of more than one period; the specifications we shall consider will allow this situation to be investigated.

Another application was discussed by Ginsburgh et al. (1979) within the context of export price setting behaviour. They proposed a model whereby the export price of a commodity is determined either as a linear function of the domestic price or as a linear function of (foreign) competitors' price, with the decision as to which obtains depending on the value of some (unspecified) function of exogenous variables. For the purposes of their paper, they analyse the case where $g(x)$, the dichotomy function, is specified so as to give a disequilibrium type of model. However, their general argument certainly suggests that the switching models we describe later in this paper are applicable.

In this paper we shall confine attention to a Bayesian analysis of some threshold switching models, comparing and contrasting the Bayesian and maximum likelihood approaches. In section 2 we shall develop inference procedures for the standard linear threshold model and discuss some of the problems encountered in using these models for prediction. Section 3 presents the extensions to linear time series models, together with a simple numerical illustration.

2. Bayesian analysis of the linear threshold model

2.1. Introduction

The general linear threshold model with two regimes is

$$y_t = X_{1t}'\theta_1 + \epsilon_{1t} \quad \text{if } g(\pi, z_t) \leq 0 \quad \text{Regime I,}$$

$$= X_{2t}'\theta_2 + \epsilon_{2t} \quad \text{if } g(\pi, z_t) > 0 \quad \text{Regime II,}$$

where

- X_{1t} (X_{2t}) = k_1 (k_2)-dimensional vector of explanatory variables,
- θ_1 (θ_2) = k_1 (k_2)-dimensional vector of parameters,
- $g(\cdot)$ = some function of a set of r parameters π and exogenous variables z_t ,
- $\epsilon_{it} \sim \text{iid } N(0, \sigma_i^2), i = 1, 2,$
- $\text{cov}(\epsilon_{1t}, \epsilon_{2t}) = 0, \forall i, j.$

Thus, according to the value of some dichotomy function $g(\cdot)$, a given

observation is generated by either regime I or regime II. In the most general case, no restrictions are placed on the components of X_{1t} , X_{2t} , z or the functional form of $g(\cdot)$. That is, the two regimes may have some common independent variables; the dichotomy variables, z , may be a subset of or include a subset of the independent variables of either or both regimes; $g(\cdot)$ may be linear or non-linear in π .

However, the dichotomy function parameters, π , must be distinct from the regime parameters, θ_1, θ_2 . Also, θ_1 and θ_2 must not contain any common parameters.

2.2. Basic notation and results

We first introduce some simplifying notation. Given a data set $y = (y_1, \dots, y_T)'$ with independent variables $X_1^* = [X_{11}, X_{12}, \dots, X_{1T}]$, $X_2^* = [X_{21}, X_{22}, \dots, X_{2T}]$, $z = [z_1', z_2', \dots, z_T']$, then, for any given π , we may write the model as

$$\begin{bmatrix} y_1^* \\ y_2^* \end{bmatrix} = \begin{bmatrix} X_1^* & 0 \\ 0 & X_2^* \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1^* \\ \epsilon_2^* \end{bmatrix},$$

where y_1^*, \dots, y_T^* is the time ordering of the observations generated by regime I, and similarly y_{11}^*, \dots, y_T^* for regime II; $X_1^{*'} = [X_{11}^*, \dots, X_{1T}^*]$, $X_2^{*'} = [X_{21}^*, \dots, X_{2T}^*]$. The interpretation of the asterisked variables is analogous for X_{1t} and X_{2t} ($t = 1, \dots, T$), $j = 1, \dots, T$ and so on. What we have done, for notational convenience, is simply to put in order of occurrence regime I observations followed by regime II observations – the dichotomy conditional on the values of the parameters π .

With the model in this form, we adopt the usual non-informative prior forms [see Box and Tiao (1970), pp. 116–118] and consider first the case of constant regime variances, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, so that

$$p(\theta_1, \theta_2, \sigma^2) = p(\theta_1 | \sigma^2) p(\theta_2 | \sigma^2) p(\sigma^2),$$

where $p(\theta_i | \sigma^2) \propto \text{constant}$, $i = 1, 2$, and $p(\sigma^2) \propto \sigma^{-2}$. Combining this with the likelihood, we get the following posterior forms:

$$p(\theta | \pi, D) \propto \frac{|X'X|^{1/2} S^{-k}}{p^{k/2}} \left[1 + \frac{(\theta - \theta)' X' X (\theta - \theta)}{p S^2} \right]^{-\frac{(v+k)/2}{2}},$$

so that $\theta' = (\theta_1', \theta_2')$ has, a posteriori, a multivariate- t distribution, conditional

on π and D , where $D = \{y_1, \dots, y_T, z_1^1, \dots, z_T^1\} = \{y, z\}$, and

$$X'X = \begin{bmatrix} X_1^* X_1^* & 0 \\ 0 & X_2^* X_2^* \end{bmatrix},$$

$$\theta = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \begin{bmatrix} (X_1^* X_1^*)^{-1} X_1^* y_1^* \\ (X_2^* X_2^*)^{-1} X_2^* y_2^* \end{bmatrix},$$

$$vS^2 = v_1 S_1^2 + v_2 S_2^2, \quad k = k_1 + k_2,$$

$$S_i^2 = v_i^{-1} (y_i^* - \hat{y}_i)' (y_i^* - \hat{y}_i), \quad \hat{y}_i = X_i^* \hat{\theta}_i, \quad i = 1, 2,$$

$$v_1 = t_{\pi} - k_1, \quad v_2 = T - t_{\pi} - k_2.$$

Since $p(\theta|D) = \int p(\theta|\pi, D)p(\pi|D)d\pi$, it follows that $p(\theta|D)$ is a weighted average of multivariate- t densities, with weights given by the posterior distribution of π (which we shall examine in more detail below). By standard arguments, it follows that, conditional on π , σ^2/vS^2 is distributed as inverse chi-square with v degrees of freedom, where $v = v_1 + v_2$. Clearly, posterior moments are weighted averages of moments of the individual conditional (multivariate) t and χ^{-2} densities, so that computation of these quantities is straightforward.

We now consider the case of different regime variances, $\sigma_1^2 \neq \sigma_2^2$, so that, conditional on π , we have the two models

$$y_1^* = X_1^* \theta_1 + e_1^*, \quad y_2^* = X_2^* \theta_2 + e_2^*,$$

with $v(e_1^*) = \sigma_1^2$ and $v(e_2^*) = \sigma_2^2$. The diffuse prior is now given by $p(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2) = p(\theta_1|\sigma_1^2)p(\theta_2|\sigma_2^2)p(\sigma_1^2)p(\sigma_2^2)$, where $p(\theta_i|\sigma_i^2) \propto \text{constant}$ and $p(\sigma_i^2) \propto \sigma_i^{-2}$, $i = 1, 2$. Again, straightforward Bayesian manipulation establishes that the joint and marginal distributions of regime parameters are, conditional on π , multivariate- and univariate- t respectively, and also that $\sigma_i^2 | v_i S_i^2 \sim \chi_{v_i}^{-2}$, $i = 1, 2$.

We now examine the posterior distribution of π , beginning with the case where $\sigma_1^2 = \sigma_2^2 = \sigma^2$. The likelihood in this case is given by

$$\begin{aligned} \ell(y|\theta_1, \theta_2, \sigma^2, \pi, z) &\propto \sigma^{-T} \exp \left\{ -\frac{1}{2\sigma^2} \left[(y_1^* - X_1^* \theta_1)' (y_1^* - X_1^* \theta_1) \right. \right. \\ &\quad \left. \left. + (y_2^* - X_2^* \theta_2)' (y_2^* - X_2^* \theta_2) \right] \right\}, \end{aligned}$$

which, combined with the diffuse prior forms and the prior density $p(\pi)$ for π , gives the joint posterior density of $\theta_1, \theta_2, \sigma^2, \pi$,

$$p(\theta_1, \theta_2, \sigma^2, \pi|D) \propto \sigma^{-2} \ell(y|\theta_1, \theta_2, \sigma^2, \pi, z) p(\pi).$$

Integrating out θ_1, θ_2 and σ^2 then yields

$$p(\pi|D) \propto \left| (X_1^* X_1^*)^{-1} \right| \left| (X_2^* X_2^*)^{-1} \right|^{1/2} [vS^2]^{-(T-k_1)/2} p(\pi). \quad (1)$$

In fact, in all our later numerical illustrations we shall take $p(\pi) \propto \text{constant}$.

In the case where $\sigma_1^2 \neq \sigma_2^2$, we have the likelihood

$$\begin{aligned} \ell(y|\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \pi, z) &\propto \sigma_1^{-t_{\pi}} \exp \left\{ -\frac{1}{2\sigma_1^2} (y_1^* - X_1^* \theta_1)' (y_1^* - X_1^* \theta_1) \right\} \sigma_2^{-(T-t_{\pi})} \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma_2^2} (y_2^* - X_2^* \theta_2)' (y_2^* - X_2^* \theta_2) \right\}, \end{aligned}$$

and hence the joint posterior

$$p(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \pi|D) \propto \sigma_1^{-2} \sigma_2^{-2} \ell(y|\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \pi, z) p(\pi),$$

which, on integrating out $\theta_1, \theta_2, \sigma_1^2$ and σ_2^2 , yields

$$\begin{aligned} p(\pi|D) &\propto \frac{\Gamma[(t_{\pi} - k_1)/2] \Gamma[(T - t_{\pi} - k_2)/2]}{\left| (X_1^* X_1^*)^{-1} \right| \left| (X_2^* X_2^*)^{-1} \right|^{1/2}} \\ &\quad \frac{(v_1 S_1^2)^{(t_{\pi} - k_1)/2} (v_2 S_2^2)^{(T - t_{\pi} - k_2)/2}}{(v_1 S_1^2)^{(t_{\pi} - k_1)/2} (v_2 S_2^2)^{(T - t_{\pi} - k_2)/2}}. \end{aligned} \quad (2)$$

There are a number of points to note. First, we must be careful when computing $p(\pi|D)$ to avoid infeasible values of π (i.e., with $t_{\pi}, T - t_{\pi}$ less than required for identifiability). It may happen, of course, that for some samples there are no such regions. In general, we can always avoid the problem by specifying a proper prior form for any unidentified parameters. Secondly, the analytic forms for $p(\pi|D)$ in (1) and (2) are only available up to proportionally. For any given application, the normalising factor must be found (approximately) by numerical computation. Thirdly, the forms of (1)

and (2) reveal that the value of $p(\pi|D)$ is inversely proportional to a measure of the overall model variance based upon the dichotomy of the data given by π . Thus, combinations of π which determine sample separations into the two regimes yielding a high estimated residual variance (i.e., which are poorly fitting) acquire low posterior probability.

2.3. Special features arising from (π, z) combinations

In what follows, we assume a (two-dimensional) linear dichotomy function, $g(\pi, z) = \pi_1 z_1 + \pi_2 z_2$, and examine some peculiar features of regime classification inference which do not appear to have been fully discussed by previous authors.

The first feature arises from what we shall call *concurrency*. For a given data set, using the observed $z = (z_1, z_2)'$, we construct a diagram in π space (i.e., the plane with orthogonal π_1, π_2 -axes) which consists of rays through the origin. It is the configuration of these rays which can lead to special problems. For n observations with distinct z_1/z_2 ratios there are $2n$ possible groupings [so that, assuming a uniform prior distribution for π , $p(\pi|D)$ will comprise at most $2n$ distinct values]. However, concurrent rays (given by coincident z_1/z_2 ratios) lead to complications in the number of possible groupings of a given data set as we show in the following example (corresponding to $n = 3$).

The six groupings are the distinct infinite cones numbered 1-6 in fig. 1. If there is no concurrency, each boundary is associated with two groupings – one on either side – and the boundary itself belongs to only one of them. In the case of concurrency, however, a boundary can be associated with three different groupings: one either side and the boundary itself. This will be the case if the concurrency is what we shall call an ‘awkward’ concurrency, whose meaning we can illustrate as follows with the aid of fig. 2. Consider two observations (a) and (b) of the z variable taking the following values:

$$\begin{array}{ll} \text{‘Nice’ concurrency} & \begin{array}{ll} \text{(a)} & z_1 = 1 & \text{(b)} & z_1 = 2 \\ & z_2 = -1 & & z_2 = 2 & z_1/z_2 = -1, \end{array} \\ \text{‘Awkward’ concurrency} & \begin{array}{ll} \text{(a)} & z_1 = 1 & \text{(b)} & z_1 = -1 & z_1/z_2 = -1, \\ & z_2 = -1 & & z_2 = 1 & \end{array} \end{array}$$

In both cases, the two sets of z values define only a single ray or boundary in π space. In fig. 2 the boundary is labelled (a), (b) and α , β , γ are three representative points in the π plane. If we calculate the value of the dichotomy function $\pi_1 z_1 + \pi_2 z_2$ for both observations at each of the points $\alpha = (\frac{1}{2}, 1)$, $\beta = (1, 1)$, $\gamma = (1, \frac{1}{2})$ for the two cases we obtain the following: (1) either both

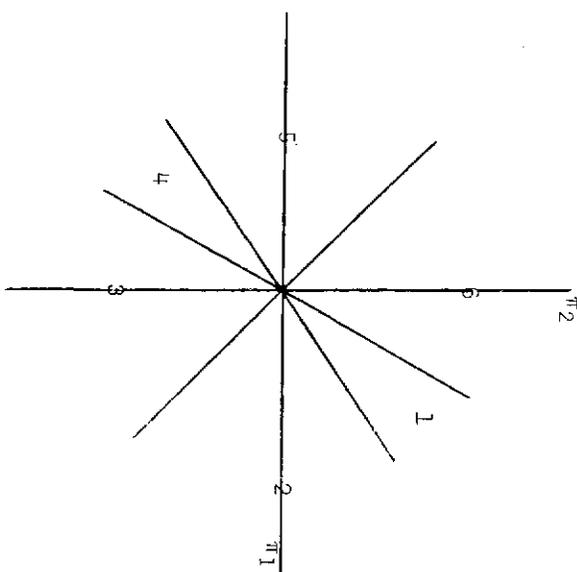


Fig. 1. π -cones for three distinct z ratios.

observations are classified to regime I or both to regime II, and (2) $\alpha \Rightarrow (a) \in I$, (b) $\in II$, $\beta \Rightarrow (a)$, (b) $\in I$, $\gamma \Rightarrow (a) \in II$, (b) $\in I$.

‘Nice’ concurrencies – i.e., those where z_1 values are of the same sign for all concurrent observations, and likewise for z_2 , but z_1 and z_2 may have any sign combination for a particular observation – pose no further problem. However, ‘awkward’ concurrencies – where at least one pair in any set of concurrent observations violates the above conditions – do cause problems, whose nature will be examined as and when appropriate in the following sections.

Another phenomenon, which we call *distality*, occurs when, in parameter switching models, including models where only a subset of parameters change – i.e., those involving identical independent variables in both regimes ($X_1 = X_2$) – the prior specifications are non-informative within each model.

For example, suppose we have four observations from (say) two competing straight lines and π_1 gives the dichotomy $1, 2 \in I$, $3, 4 \in II$. Suppose that the corresponding likelihood function is l_1 . Then $\pi_2 = -\pi_1$ will give the opposite dichotomy $1, 2 \in II$, $3, 4 \in I$ (assuming π does not lie on a boundary) and a likelihood function l_2 . If we maximise such l_1 and l_2 , we find they are equivalent, and yield the following pattern of maximum likelihood estimates:

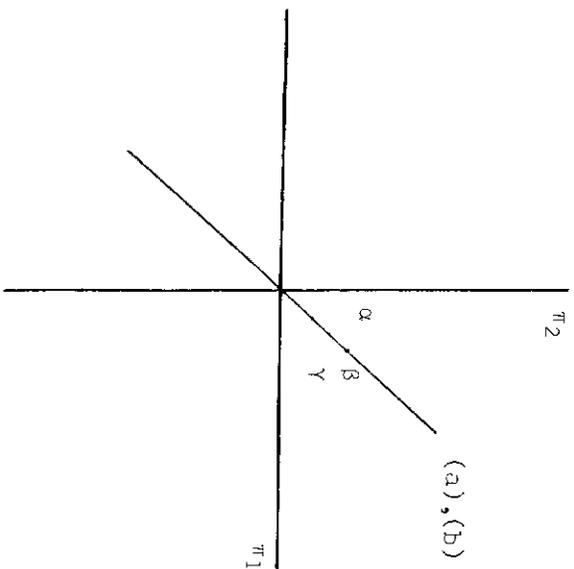


Fig. 2. The effect of concurrency on the number of possible data groupings.

from $l_1, \hat{\alpha}_1 = a, \hat{\beta}_1 = b, \hat{\alpha}_2 = a', \hat{\beta}_2 = b'$, and from $l_2, \hat{\alpha}_1 = a', \hat{\beta}_1 = b', \hat{\alpha}_2 = a, \hat{\beta}_2 = b$. In this case, we have the same underlying information in both cases, and so, if there are no awkward concurrencies in the data, we need only look at a half-plane in π space. Awkward concurrencies cause problems because they do not give opposite classifications either side of the origin (see fig. 3).

From a Bayesian point of view, the consequence of duality is that if we discretize π over the whole π -plane (by which we mean a suitable subset covering all possible groupings of the data – for example, uniformly around the unit circle), then the posterior distributions of the regime parameters are identical. For a moderate sized data set, these posterior distributions will be bimodal with a mode corresponding to each regime. Similarly, the posterior distribution of π will consist of a twice repeated pattern of plateaux (see, for example, the time series illustration given in section 3.3).

Awkward concurrencies introduce perturbations from symmetry, but these will be negligible if the data set is large relative to the number of such awkward concurrencies.

Maximum likelihood analysis does not reflect these features. If we use the whole π -plane and the selected optimal region is one of the infinite cones (awkward concurrencies notwithstanding) then the opposite grouping is also optimal (see above example).

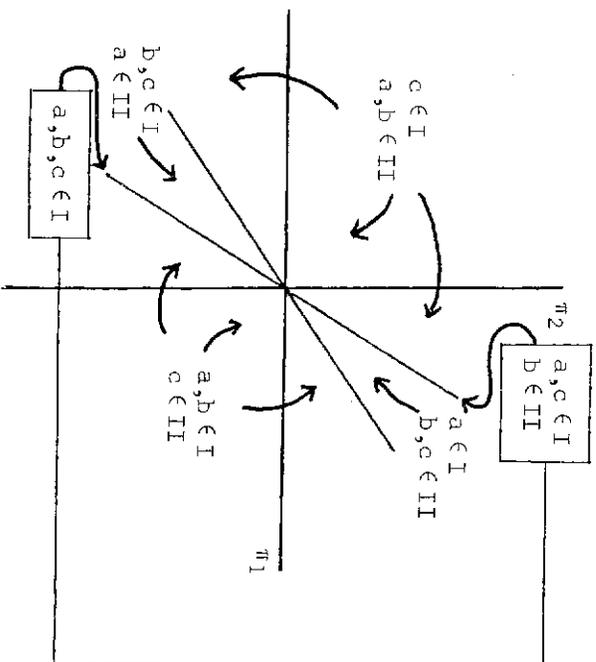


Fig. 3. Awkward type of concurrency.

2.4. Comparison of the Bayesian and maximum likelihood approaches

In general, maximum likelihood analysis will not yield a unique estimate of π . This is easiest to see from a diagram of two-dimensional π space using a linear dichotomy function, as shown in fig. 4. All π values in the shaded region in fig. 4 [ignoring (a) and (b), whose significance will be discussed in the following section] give the same sample separation and hence the same likelihood. Even if we impose some normalisation rule on π (such as $\pi_1^2 + \pi_2^2 = 1$) we still have an infinite set of optimal values. There are two exceptions to this general case: if either the number of observations tends to infinity, so that the optimal cone reduces to a single ray, or if the maximum likelihood dichotomy occurs at an awkward concurrency; then, in both cases, a normalisation rule leads to a unique result.

The first situation is irrelevant in finite-sample practice, and the second situation is highly unlikely if the z 's are continuous variates. Of course, this non-uniqueness extends to the posterior distribution (corresponding to a uniform prior) if we are looking for modal values. However, the implications for prediction are very different in the two cases as we shall see in section 2.5.

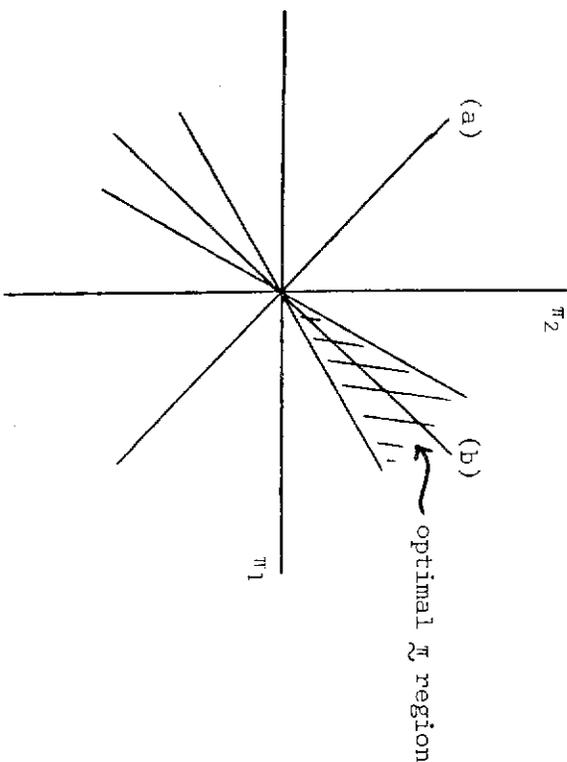


Fig. 4. Non-uniqueness of ML estimate for π .

A further major difference between the Bayesian and maximum likelihood methodologies for these threshold problems is that for the models under study maximum likelihood analysis does not appear capable of producing measures of uncertainty for π . Thus, we can get no feel for the properties of π , even locally. Bayesian analysis, on the other hand, readily provides posterior moments, and the complete posterior distribution provides a rich global overview of the inferences to be drawn in these rather non-standard situations as we see in the following example.

We consider the example of a straight line versus a constant:

$$y_t = \theta + \epsilon_t, \quad \text{Regime I,}$$

$$y_t = \alpha + \beta x_t + \epsilon_t, \quad \text{Regime II,}$$

$$g(\pi, z_t) = \pi_1 z_{1t} + \pi_2 z_{2t},$$

where $\theta = 0$, $\alpha = 2$, $\beta = 0.5$, $\epsilon_t \sim \text{iid } N(0, 1)$, $\pi_1 = \pi_2 = 1$.

A simulation was performed which generated twenty observations, ten from each regime. Comparative summary inferences are presented in table 1 and complete Bayesian posterior densities (based on non-informative priors) are

Table 1
ML estimates and posterior moments

	ML	Mean	SD
θ	0.3216	0.5432	0.679
α	2.3755	1.9175	0.691
β	0.4343	0.3736	0.178
σ^2	1.0998	1.8442	0.808

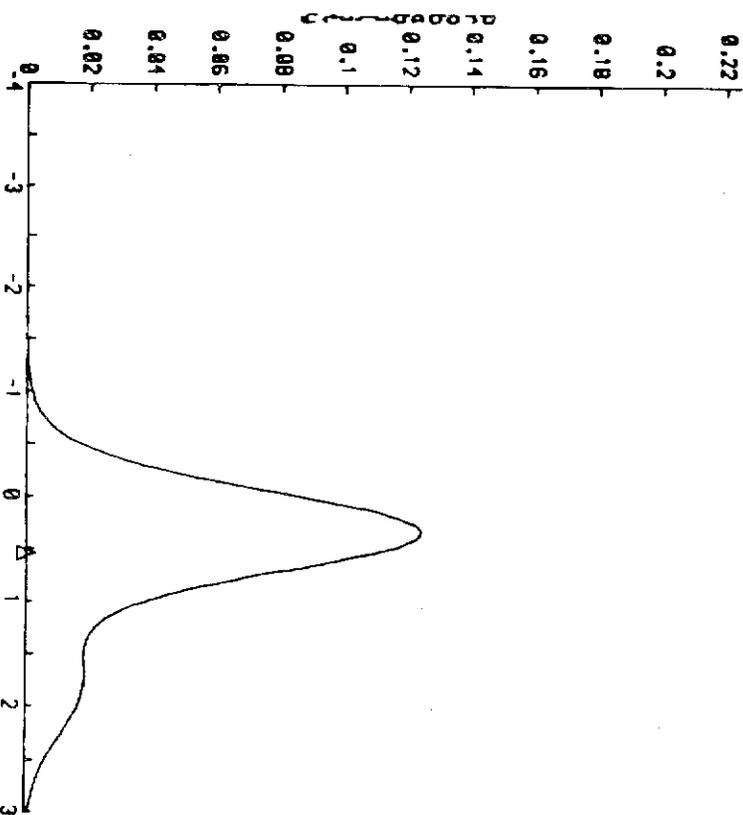
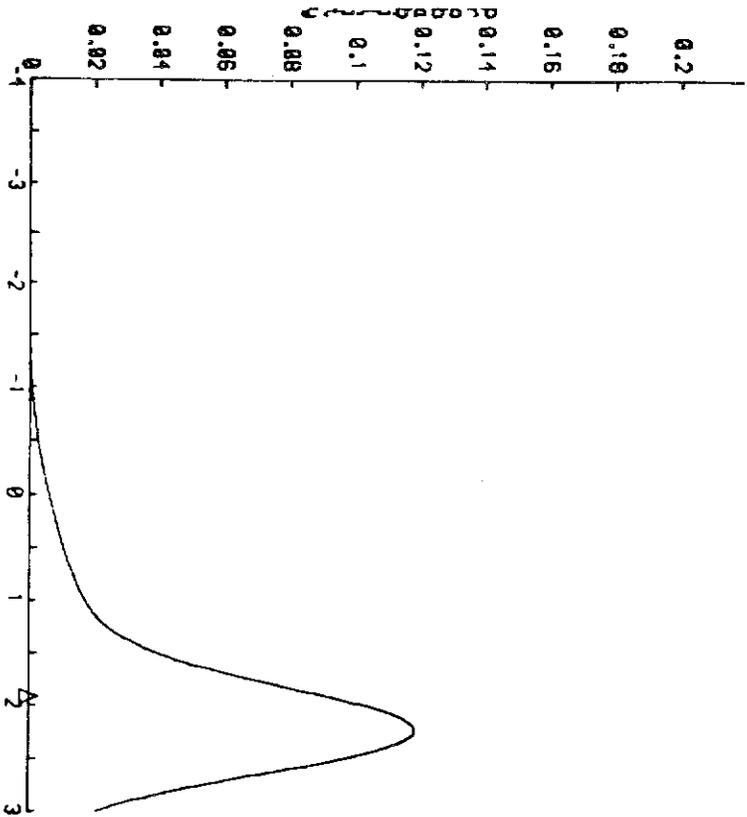


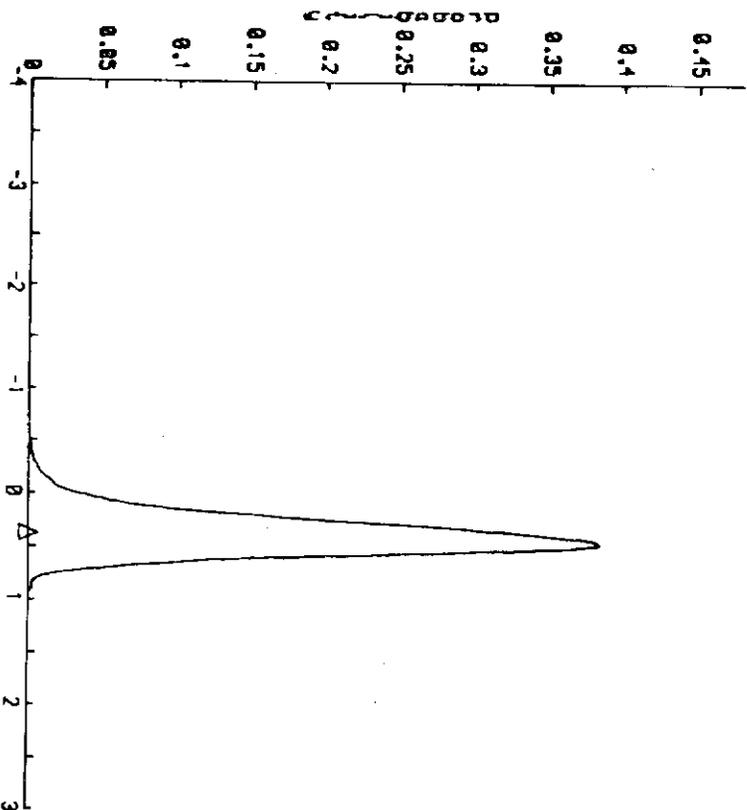
Fig. 5. Marginal posterior distribution of θ .

Fig. 6. Marginal posterior distribution of α .

shown in figs. 5–9. To obtain a representative picture of the posterior density for π (which actually consists of a step function defined over infinite wedges in the plane), we calculate values on a grid of 160 points, 40 equally spread along each side of the unit square [proceeding anti-clockwise, starting at $(-0.95, 1)$]. The modal spike gives a maximum likelihood estimate (recall that a uniform prior for π has been used). In what follows, we shall continue to use such a summation in place of the integral over π .

2.5. Prediction

The problem of uniqueness discussed in the previous section leads to situations in which the maximum likelihood approach runs into fundamental

Fig. 7. Marginal posterior distribution of β .

difficulties. Again, we shall illustrate this in the two-dimensional case. Suppose we want to forecast a new observation given a previous sample in which we observe all the independent and dichotomy variables. If our new z is such that it defines a boundary of the type depicted by (a) in fig. 4, then we have no problem. Whichever maximum likelihood estimate we select, $\pi_1 z_1 + \pi_2 z_2$ uniquely assigns the observation to one or other of the regimes. However, if the new boundary is similar to that depicted by (b), we have a problem, since different maximum likelihood estimates of π would classify the new observation to different regimes. The choice of regime then appears totally arbitrary and the maximum likelihood approach gives no guidance. We know of no satisfactory solution to this problem. There are circumstances in which this inability of maximum likelihood to select a predictive regime is not a

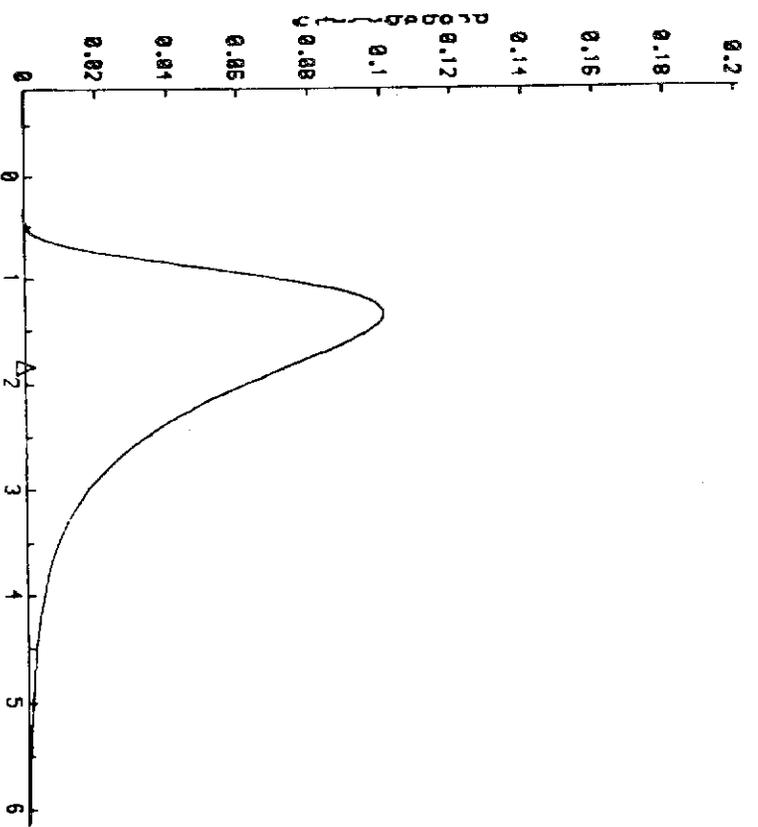


Fig. 8. Marginal posterior distribution of σ^2 .

serious problem, but this occurs basically when the predicted values for each regime are very similar – not, perhaps, a typical situation.

If we adopt the Bayesian approach, we encounter no such problem. There are various ways in which to approach the choice of predictive regime. One possibility is simply to select that regime which has the greatest posterior probability for the new z . Denoting the new value of z by z_n , and with

$$p = p(I|z_n) = \sum_{\pi: \pi z_n < 0} p(\pi|D),$$

we choose regime I if $p > \frac{1}{2}$. The only arbitrariness in this case is what happens when $p = \frac{1}{2}$, and clearly this will happen, if at all, far less often than case (b)

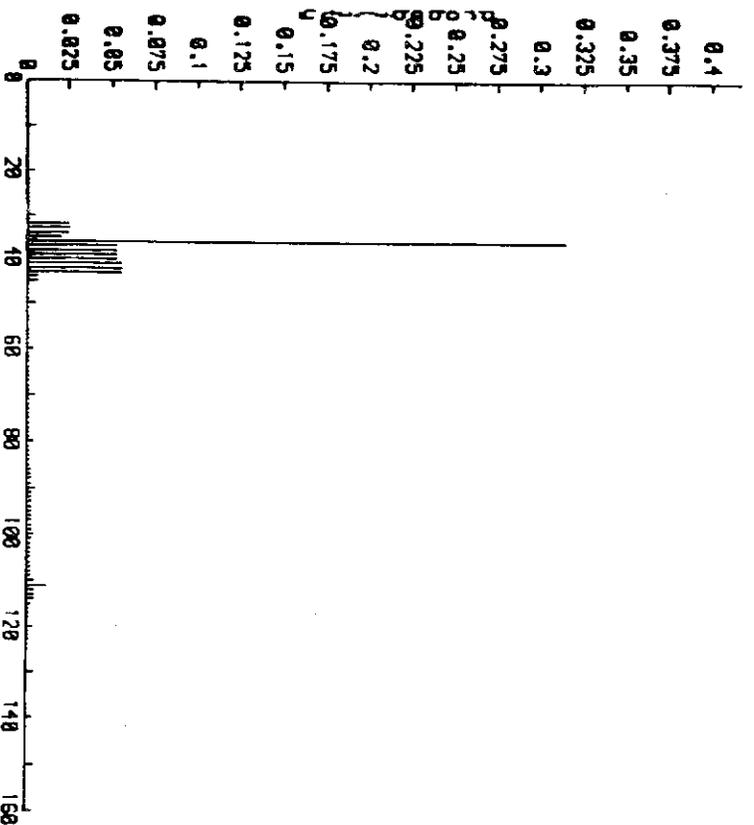


Fig. 9. Marginal posterior distribution of π .

above. This simple approach corresponds to a decision-theoretic regime choice with zero-one loss function. A more philosophically sound approach might be to use decision theoretic ideas for the prediction itself. The first requirement is then to choose an appropriate utility function. Following Bernardo (1979), we adopt the form

$$U(p(\cdot), y) = A \log p(y) + B(y),$$

where A is a positive constant, B is a function of y , $p(y)$ is the predictive density, and y is the value to be predicted. Maximising expected utility, we select regime I if

$$\int \log \left[\frac{p_I(y|z_n)}{p_{II}(y|z_n)} \right] p^*(y|z_n) dy > 0,$$

where

$$p^*(y|z_n) = pp_I(y|z_n) + (1 - p) p_{II}(y|z_n),$$

and where $p_I(y|z_n)$ and $p_{II}(y|z_n)$ are the predictive densities resulting from the choice of regime I and II, respectively. We now require the forms of $p_I(y|z_n)$ and $p_{II}(y|z_n)$. We shall consider the case of constant regime variances; the extension to non-constant regime variances is straightforward. Consider, for example,

$$p_I(y|z_n) = \int \int p(y|\theta_1, \sigma^2) p(\theta_1, \sigma^2|D) d\theta_1 d\sigma^2.$$

Under our assumptions, the first term is a Normal density, but unfortunately the second term is not available in closed form, because of the nature of π . Thus, as it stands, the predictive density is only available numerically and seems to require heavy computation. However, we can reduce this to a single summation in the following way. Conditioning on π we have

$$p_I(y|z_n) = \int \int p(y|\theta_1, \sigma^2) \sum_{\pi} p(\theta_1, \sigma^2|\pi, D) p(\pi|D) d\theta_1 d\sigma^2.$$

If we now interchange integration and summation,

$$p_I(y|z_n) = \sum_{\pi} \int \int p(y|\theta_1, \sigma^2) p(\theta_1|\pi, D, \sigma^2) \times p(\sigma^2|\pi, D) d\theta_1 d\sigma^2 p(\pi|D),$$

from which we obtain

$$y|\theta_1, \sigma^2 \sim N(x'\theta_1, \sigma^2),$$

$$\theta_1|\sigma^2, \pi, D \sim N(\hat{\theta}_1, \sigma^2(X_1^*X_1^*)^{-1}),$$

$$\sigma^2/\nu S^2|\pi, D \sim \chi_{\nu-2}^2,$$

so that evaluation of the above integral yields

$$(y - x'_n\hat{\theta}_1) / \sqrt{S(1 + V)} | \pi \sim t_{\nu},$$

where $V = x'_n(X_1^*X_1^*)^{-1}x_n$, x_n is the new value of x , and all other quantities are previously defined. Thus, the predictive densities are weighted averages of t -densities with mixing parameters $p(\pi|D)$.

3. Extension to linear time series

3.1. Introduction

We confine attention to processes assumed to have started at some distant time in the past, and such that each regime may be modelled, for some choice of u, v by an ARMA(u, v) form

$$(1 - \phi_1 B - \dots - \phi_u B^u)(y_t - \mu_t) = (1 + \theta_1 B + \dots + \theta_v B^v) \epsilon_t.$$

We concentrate here on inference problems for this limited class of models. For an account of more general modelling possibilities, see Tong (1983).

The approach we shall use is an extension of that used by Booth and Smith (1982) whose change-point models are in fact special cases of the switching model we are considering here. What we seek is a transformation of y_1, \dots, y_T , conditional on π , to independent homoscedastic variables $\omega_1, \dots, \omega_T$, such that the models for each regime can be written in the standard linear model form $\omega = X\psi + \epsilon$. Once we have such a transformation, all our previous results apply, except that we are now conditioning on (ϕ, θ) as well as π [so that in the analysis we must not forget to multiply by the appropriate Jacobian $J(\phi, \theta)$]. For example, following such a transformation (1) becomes

$$p(\pi, \phi, \theta|D) \propto \left| (X'X_{(\phi, \theta)})^{-1} \right|^{1/2} \left[\nu S^2_{(\phi, \theta)} \right]^{-(T-k)/2} \times J(\phi, \theta) p(\pi) p(\phi, \theta).$$

Approximate marginal distributions can then be obtained by summing over a suitable grid for (ϕ, θ) (as long as u, v are small, which is usually the case in applications). Alternatively, more refined numerical integration procedures could be used.

Various special cases can be studied within this general class of models depending on the assumptions made about the model parameters. We shall look at (i) change in mean level and time series parameters across regimes, (ii) change in mean level only, (iii) change in time series parameters only.

3.2. AR(1) v AR(I)

We assume that the regimes correspond to the following models, with a linear threshold function as defined in earlier sections:

$$y_t - \mu_1 = \phi_1 (y_{t-1} - \mu_{t-1}) + \epsilon_t \quad \text{Regime I,}$$

$$y_t - \mu_2 = \phi_2 (y_{t-1} - \mu_{t-1}) + \epsilon_t \quad \text{Regime II,}$$

$$\mu_{t-1} = \mu_1 \quad \text{if } y_{t-1} \in I,$$

$$= \mu_2 \quad \text{if } y_{t-1} \in II.$$

Suppose without loss of generality that we begin with observations being generated by regime I. The required transformation to a standard linear model is then given by

$$\omega_1 = (1 - \phi_1^2)^{\frac{1}{2}} y_1,$$

$$\omega_i = y_i - \phi_j y_{i-1}, \quad i = 2, \dots, T, \quad j = \begin{cases} 1 & \text{if } y \in I, \\ 2 & \text{otherwise.} \end{cases}$$

with Jacobian $(1 - \phi_1^2)^{\frac{1}{2}}$.

3.2.1. Change in all parameters

Suppose we have switches occurring at r_1, r_2, \dots , so that $y_1, \dots, y_{r_1} \in I, y_{r_1+1}, \dots, y_{r_2} \in II$ and so on. Then the required form of X' is

$$X' = \begin{bmatrix} (r_1) & (r_2) \\ (1 - \phi_1^2)^{\frac{1}{2}} \cdot 1 - \phi_1, \dots, 1 - \phi_1, -\phi_2, & 0, \dots, 0, & 1, & 1 - \phi_1, \dots \\ 0, & 0, \dots, 0, & 1, & 1 - \phi_2, \dots, 1 - \phi_2, -\phi_1, & 0, \dots \end{bmatrix}$$

3.2.2. Change in mean level only

X has the same form as above except that now $\phi_1 = \phi_2$.

3.2.3. Change in AR parameters only

The columns of the above matrix are now combined to form a single $(T \times 1)$ vector

$$X' = \left[(1 - \phi_1^2)^{\frac{1}{2}} \cdot 1 - \phi_1, \dots, 1 - \phi_1, 1 - \phi_2, \dots, 1 - \phi_2, 1 - \phi_1, \dots \right]$$

Table 2

Maximum likelihood estimates.			
μ_1	μ_2	ϕ^2	ϕ
0.138	1.97	0.899	0.6

Table 3

A discretized summary posterior distribution for ϕ .					
ϕ	-0.9	-0.8	-0.7	-0.6	-0.5
$p(\phi y)$	0.014	0.077	0.216	0.315	0.241
ϕ	-0.4	-0.3	0.2	-0.1	0.0..0.9
$p(\phi y)$	0.103	0.207	0.005	0.001	0

In cases (i) and (ii), if regime II generates the first observation then the form of X is obtained by interchanging the columns and the AR parameters. The Jacobian then becomes $(1 - \phi_2^2)^{\frac{1}{2}}$.

3.3. Example

We suppose that the model is AR(1) \vee AR(1) with a change in mean level only,

$$y_t = \mu_1 - \phi(y_{t-1} - \mu_{t-1}) + \epsilon_t, \quad \text{Regime I,}$$

$$y_t = \mu_2 = \phi(y_{t-1} - \mu_{t-1}) + \epsilon_t, \quad \text{Regime II,}$$

$$g(\pi, z_t) = \pi_1 z_{1t} + \pi_2 z_{2t},$$

$$\epsilon_t \sim \text{iid } N(0, \sigma^2),$$

where $\phi = -0.6, \mu_1 = 0, \mu_2 = 2.0, \sigma^2 = 1$.

Tables 2 and 3 above summarise maximum likelihood and Bayes' inferences based on a simulated sample of 40 observations, 17 from regime I and 23 from

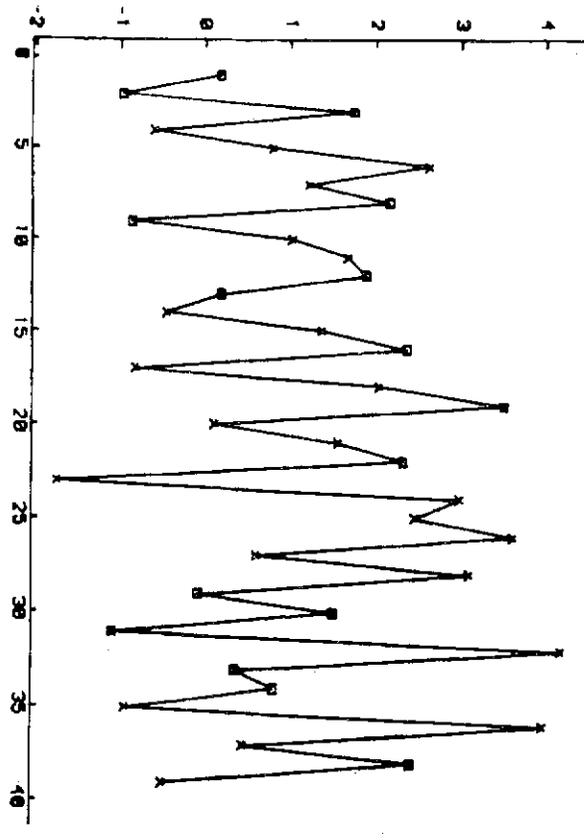
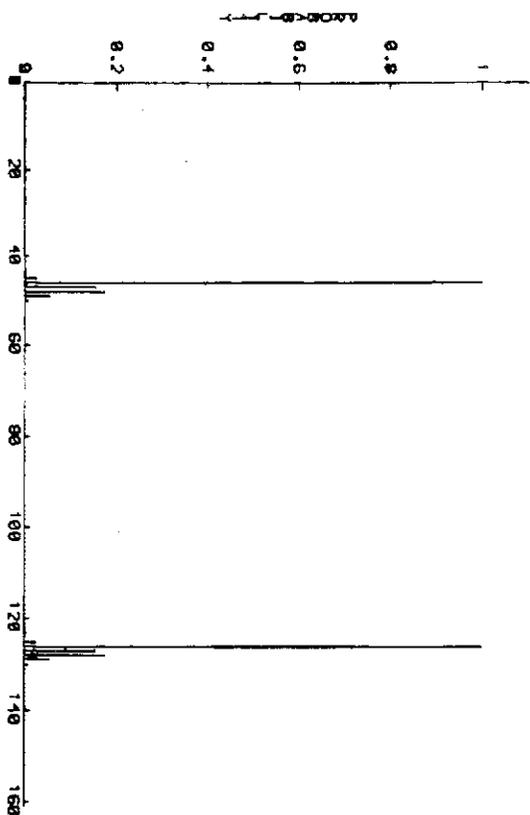
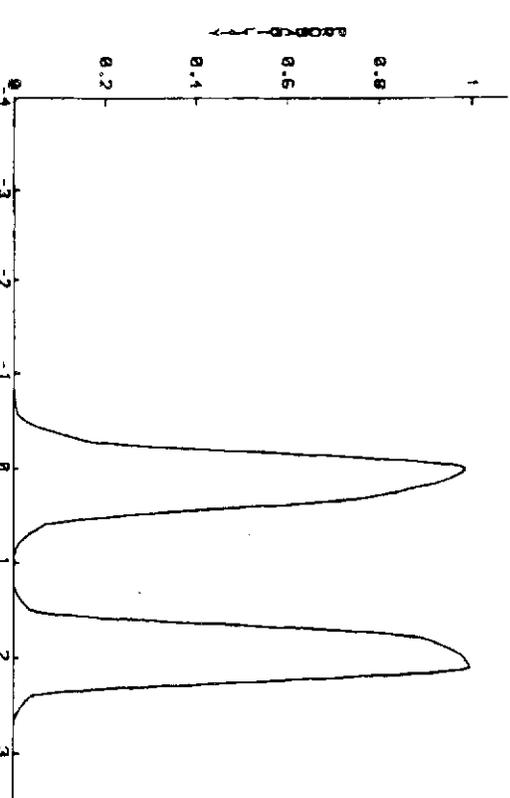


Fig. 10. Time plot of observations (x = regime I observation, o = regime II observation).

Fig. 11. Marginal posterior distribution of π .Fig. 12. Marginal posterior distribution of μ .

regime II, as shown in fig. 10. Table 3 clearly shows that inference on ϕ is accurate and quite sharp.

Although this is a time series example, the phenomenon of duality described earlier does occur when only the mean level changes. In order to illustrate its effects, we have used the full π plane in the calculations. In fig. 11 the repeated pattern of plateaux in $p(\pi|y)$ is clear; since $p(\mu_1|y) = p(\mu_2|y)$ we only illustrate one of these as shown in fig. 12.

3.4. A note on identification

Analysis of other simple time series examples (not reported here) shows that competing AR(1) regimes can give rise to sample autocorrelations and partial autocorrelations which depart radically from the anticipated patterns for a (single regime) AR(1) model. Thus, given a data set to which we wish to fit an ARMA model, indications of a complex model, for which we have no convincing theoretical foundations, may well be indicating the presence of a two-regime switching model in which each regime is a relatively simple form.

References

- Bacon, D.W. and D.G. Watts, 1971, Estimating transition between two intersecting straight lines, *Biometrika* 58, 525-534.
- Bernardo, J.M., 1979, Expected information as expected utility, *Annals of Statistics* 7, 686-690.
- Booth, N.B. and A.F.M. Smith, 1982, A Bayesian approach to retrospective identification of change-points, *Journal of Econometrics* 19, 7-22.
- Box, G.E.P. and G.C. Tiao, 1973, Bayesian inference in statistical analysis (Addison-Wesley, Reading, MA).
- Chin Choy, J.H. and L.D. Broemeling, 1980, Some Bayesian inferences for a changing linear model, *Technometrics* 22, 71-78.
- Diaz, J., 1982, Bayesian detection of a change of scale parameter in sequences of independent gamma random variables, *Journal of Econometrics* 19, 23-29.
- Fair, R.C. and D.M. Jaffe, 1972, Methods of estimation for markets in disequilibrium, *Econometrics* 40, 497-514.
- Farley, J.U. and M. Hinich, 1970, Testing for shifting slope coefficient in a linear model, *Journal of the American Statistical Association* 65, 1320-1329.
- Farley, J.U., M. Hinich and T.W. McGuire, 1975, Some comparisons of tests for a shift in the slopes of a multivariate linear time series model, *Journal of Econometrics* 3, 297-318.
- Ferreira, P.E., 1975, A Bayesian analysis of a switching regression model: Known number of regimes, *Journal of the American Statistical Association* 70, 370-374.
- Ginsburgh, V., A. Tishler and I. Zang, 1980, Alternative estimation methods for two regime models, *European Economic Review* 13, 207-228.
- Griffiths, W.E., Ross G. Drynan and Surekha Prakash, 1979, Bayesian estimation of a random coefficient model, *Journal of Econometrics* 10, 201-220.
- Goldfeld, S.M. and R.E. Quandt, 1972, Nonlinear methods in econometrics (North-Holland, Amsterdam).
- Goldfeld, S.M. and R.E. Quandt, 1973a, The estimation of structural shift by switching regressions, *Annals of Economic and Social Measurement* 2, 475-485.
- Goldfeld, S.M. and R.E. Quandt, 1973b, A Markov model for switching regressions, *Journal of Econometrics* 1, 3-16.
- Goldfeld, S.M. and R.E. Quandt, 1975, Estimation of a disequilibrium model and the value of information, *Journal of Econometrics* 3, 325-348.

- Gnutman, I. and U. Menzefricke, 1982, On the use of loss functions in the change-point problem. *Annals of the Institute of Statistics and Mathematics* 34, 319-326.
- Hackl, P., 1980, Testing the constancy of regression relationships over time (Vandenhoeck and Ruprecht, Göttingen).
- Hartley, M.J., 1978, Comment (on R.E. Quandt and J. Ramsay, 1978), *Journal of the American Statistical Association* 73, 738-741.
- Hartley, M.J. and P. Mallerla, 1977, The asymptotic properties of a maximum likelihood estimator for a model of markets in disequilibrium, *Econometrica* 45, 1205-1220.
- Harvey, A.C., 1981, Time series models (Philip Allan, Oxford).
- Hinkley, D.V., 1969, Inferences about the intersection in two-phase regression, *Journal of the American Statistical Association* 64, 495-504.
- Hinkley, D.V., 1971, Inferences in two phase regression, *Journal of the American Statistical Association* 66, 736-743.
- Holbert, D., 1982, A Bayesian analysis of a switching linear model, *Journal of Econometrics* 19, 77-87.
- Hsu, D.A., 1979, Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis, *Journal of the American Statistical Association* 74, 31-40.
- Hsu, D.A., 1982, Robust inferences for structural shift in regression models, *Journal of Econometrics* 19, 89-107.
- Kiefer, N.M., 1978, Discrete parameter variation: Efficient estimation of a switching regression model, *Econometrica* 46, 427-434.
- Kiefer, N.M., 1979, On the value of sample separation information, *Econometrica* 47, 997-1003.
- Kiefer, N.M., 1980, A note on switching regression and logistic discrimination, *Econometrica* 48, 1065-1069.
- Lee, A.J.S. and S.M. Heghinian, 1977, A shift in the mean level in a sequence of independent normal random variables: A Bayesian approach, *Technometrics* 19, 503-506.
- Lee, L. and R.H. Porter, 1984, Switching regression models with imperfect sample separation information: with an application to cartel stability, *Econometrica* 52, 391-418.
- Lubrano, M., 1983, Bayesian analysis of switching regression models, CORE discussion paper no. 8332 (CORE, Louvain-la-Neuve).
- Maddala, G.S. and F.D. Nelson, 1974, Maximum likelihood methods for model of markets in disequilibrium, *Econometrica* 42, 1013-1030.
- Menzefricke, U., 1981, A Bayesian analysis of a change in precision of a sequence of independent normal random variables at an unknown time point, *Applied Statistics* 30, 141-146.
- Ohtani, K., 1982, Bayesian estimation of the switching regression model with autocorrelated errors, *Journal of Econometrics* 18, 251-261.
- Painter, D.J., 1976, The economics of structural change (North-Holland, Amsterdam).
- Quandt, R.E., 1958, The estimation of parameters of a linear regression system obeying two separate regimes, *Journal of the American Statistical Association* 53, 873-880.
- Quandt, R.E., 1960, Tests of the hypothesis that a linear regression system obeys two separate regimes, *Journal of the American Statistical Association* 55, 324-330.
- Quandt, R.E., 1972, A new approach to estimating switching regressions, *Journal of the American Statistical Association* 67, 306-310.
- Quandt, R.E., 1982, Econometric disequilibrium models, *Econometric Reviews* 1, 1-63.
- Quandt, R.E. and J. Ramsay, 1978, Estimating mixtures of normal distributions and switching regressions, *Journal of the American Statistical Association* 73, 730-738.
- Salzard, D., 1982, Structural changes in time series models, *Journal of Econometrics* 19, 147-163.
- Schmidt, P., 1981, Further results on the value of sample separation information, *Econometrica* 49, 1339-1343.
- Schmidt, P., 1982, An improved version of the Quandt-Kamsey MGF estimator for mixtures of normal distributions and switching regressions, *Econometrica* 50, 501-516.
- Smith, A.F.M., 1975, A Bayesian approach to inference about a change-point in a sequence of random variables, *Biometrika* 62, 407-416.
- Smith, A.F.M., 1977, A Bayesian analysis of some time-varying models, in: J. Barra et al., *Recent developments in statistics* (North-Holland, Amsterdam).
- Smith, A.F.M. and D.G. Cook, 1980, Straight lines with a change-point: A Bayesian analysis of some renal transplant data, *Applied Statistics* 29, 180-189.

- Srivastava, M.S. and A. Sen, 1975, Some one-sided tests for change in level, *Technometrics* 17, 61-64.
- Swamy, P.A.V.B. and J.S. Mehta, 1975, Bayesian and non-Bayesian analysis of switching regressions and of random coefficient models, *Journal of the American Statistical Association* 70, 593-602.
- Tishler, A. and I. Zang, 1979, A switching regression method using inequality conditions, *Journal of Econometrics* 11, 259-274.
- Tishler, A. and I. Zang, 1981, A new maximum likelihood algorithm for piecewise regression, *Journal of the American Statistical Association* 76, 980-987.
- Tishler, A. and I. Zang, 1981, A maximum likelihood method for piecewise regression models with a continuous dependent variable, *Applied Statistics* 30, 116-124.
- Tong, H., 1982, Discontinuous decision approaches and threshold autoregressive time series modelling, *Biometrika* 69, 274-276.
- Tong, H., 1983, Threshold models in non-linear time series analysis (Springer-Verlag, New York).
- Tong, H. and K.S. Lim, 1980, Threshold autoregression, limit cycles and cyclical data (with discussion), *Journal of the Royal Statistical Society B* 42, 245-292.
- Tsurumi, H., 1977, A Bayesian tests of a parameter shift with an application, *Journal of Econometrics* 6, 371-380.