# Bayesian Analysis of the Stochastic Switching Regression Model Using Markov Chain Monte Carlo Methods

**Maria Ana E. Odejar**

**Abstract**

This study develops Bayesian methods of estimating the parameters of the stochastic switching regression model. Markov Chain Monte Carlo methods data augmentation and Gibbs sampling are used to facilitate estimation of the posterior means. The main feature of these two methods is that the posterior means are estimated by the ergodic averages of samples drawn from conditional distributions which are relatively simple and more feasible to sample from than the complex joint posterior distribution.

A simulation study is conducted to compare model estimates obtained using data augmentation, Gibbs sampling and maximum likelihood EM algorithm and to determine the effect of accuracy and bias of the researcher's prior distributions on parameter estimates.

## 1. INTRODUCTION

Economic systems are intrinsically dynamic. These dynamics are characterized by changing economic relationships. Shocks to the economy may be due to an economic crisis, a change in the society's economic behavior, improvement in technology, economic policy revisions or a major revamp in the political system. Whatever the source of change is, econometric models must be able to incorporate changing economic relationships.

An econometric model that allows for change is the stochastic switching regression model where it is assumed that an observation $y_j$ , may have been generated by one of s alternative regression models or regimes, i.e.

$$y_j = \mathbf{x}_{ij}'\boldsymbol{\beta}_i + \varepsilon_{ij} \qquad y_j \;\varepsilon\; \text{Regime } i \qquad i = 1, 2, \ldots, s \qquad j = 1, 2, \ldots, n \qquad (1.1)$$

where n is the sample size, $\mathbf{x}_{ij}$ is a vector of observed independent variables, $\boldsymbol{\beta}_i$ is a vector of regression coefficients and $\varepsilon_{ij}$ is an unobserved random error.

Information on the separation of the sample observations into the various regimes is often not available. If the switch is stochastic, some random process determines which regression model an observation is believed to have been generated from. The stochastic switching regression model is formulated as follows:

$$y_j = \mathbf{x}_{ij}' \boldsymbol{\beta}_i + \varepsilon_{ij} \quad \text{with probability } \lambda_i \quad 0 < \lambda_i < 1 \quad \sum_{i=1}^{s} \lambda_i = 1 \qquad (1.2)$$

$$\varepsilon_{ij} \sim \text{iid } N(0, \sigma_i^2) \qquad i = 1, 2, \ldots, s \quad j = 1, 2, \ldots, n$$

where the mixing parameter $\lambda_i$ is the probability that an observation $y_j$ is generated by the $i$th regime.

The stochastic switching regression model has numerous applications in the field of economics. Quandt (1972) used this model to estimate the demand and supply schedules of housing starts in disequilibrium markets. Quandt and Ramsey (1978) applied the stochastic switching regression model to estimate the Hamermesh wage bargains model. Kon and Jen (1978) also applied the stochastic switching regression model to characterize a mutual funds manager's decision making process. Beard, Caudill and Gropper (1991) estimated the multiproduct cost function of all U.S. banks using this model.

In the absence of information as to which observations follow which regression model, estimation of the regression parameters becomes complicated and cumbersome to implement. In the past, three methods have been introduced to estimate the stochastic switching regression model.

One method is by the method of moments discussed by Day (1969) and Cohen (1967). The sample moments are equated to the corresponding theoretical central moments about the mean. However, this method does not provide standard errors of the estimates.

Another method, introduced by Quandt and Ramsey (1978) is the moment generating function estimation method. Their method minimizes the sum of squared deviations of the sample from the theoretical moment generating function to derive parameter estimates. The moment generating function method produces estimates closer to the true values than estimates from the method of moments. However, there is a problem of choosing a satisfactory value of $\boldsymbol{\theta}$ in the moment generating function, $E[e^{\theta Y}]$.

Another strategy of estimating the stochastic switching regression model is to implement maximum likelihood estimation using iterative algorithms. According to Hartley (1978), the Expectation Maximization (EM) of Dempster and Rubin (1977) can provide maximum likelihood estimates of the stochastic switching regression parameters. However, the EM algorithm has some known drawbacks. When the components are not well separated or when initial values are far from the true values, the EM algorithm converges intolerably slowly (Celeux and Diebolt 1985). To remedy this problem, Celeux and Diebolt added a stochastic step to prevent the iteration from staying in an unstable stationary point of the likelihood function. This modified EM algorithm, is the Stochastic Expectation Maximization (SEM) algorithm. However, for small sample sizes, the EM and SEM algorithms are not reliable. Another well known criticism pertains to the maximum likelihood of the switching regression itself. According to Maddala and Nelson (1975), Kiefer (1978), Swamy and Mehta (1975) and Quandt and Ramsey (1978), the maximum likelihood function for the switching regression model is unbounded at the edges of the parameter space. Another criticism of the maximum likelihood method is that it does not provide parameter estimates accurate enough to be useful for small and moderately large samples. Hosmer (1973) conducted a Monte Carlo experiment which showed this disadvantage of using the maximum likelihood method.

This study will focus on estimating the stochastic switching regression model using Bayesian analysis. This method is particularly useful when there is some prior knowledge of the parameters as suggested by economic theory or previous research. Bayesian analysis combines prior information on the parameters (through the prior distribution) with information on the current sample (through the likelihood function) to provide and update estimates of the parameters using the posterior distribution. The complexity involved in evaluating the integral necessary to obtain the mean of the posterior distribution is a major drawback of Bayesian analysis. Even when a closed form expression for the posterior distribution exists, computation of the posterior mean is computationally very cumbersome. This is especially true for a complex model like the stochastic switching regression model. As a solution to this computational problem in implementing the Bayesian approach, Markov Chain Monte Carlo (MCMC) methods can be used to indirectly sample from the joint posterior distribution. This alleviates the computational difficulty of evaluating the posterior mean.

This research explores the usefulness of Markov Chain Monte Carlo methods for computing Bayesian estimates of the parameters of the stochastic switching regression model. A simulation study is conducted to compare the parameter estimates of a market disequilibrium model and a structural change model obtained by data augmentation, Gibbs sampling and the maximum likelihood EM algorithm. The simulation study focused on the effect of accuracy and bias in the prior distributions on the Bayesian estimates of the regression coefficients.

In Section 2, the posterior distribution of the stochastic switching regression model is presented. Section 3 discusses the Markov Chain Monte Carlo methods, Gibbs sampling and data augmentation and how they are implemented for the stochastic switching regression model. The simulation methodology and results are described in Section 4.

## 2. POSTERIOR DISTRIBUTION OF THE STOCHASTIC SWITCHING REGRESSION MODEL

Consider the stochastic switching regression model, for $n$ observations over $s$ regimes

$$y_j = \mathbf{x}_{ij}'\boldsymbol{\beta}_i + \varepsilon_{ij} \quad \text{with probability } \lambda_i \quad 0 < \lambda_i < 1 \quad \sum_{i=1}^{s} \lambda_i = 1 \qquad (2.1)$$

$$i = 1, 2, \ldots, s\,, \; j = 1, 2, \ldots, n$$

The errors $\varepsilon_{ij}$ are normally and independently distributed as normal with mean 0 and variance $\sigma_i^2$ i.e. $\varepsilon_{ij} \sim \text{iid } N(0, \sigma_i^2)$. Define the vector of parameters as $\boldsymbol{\theta} = \{\lambda_1, \ldots, \lambda_s, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s, \sigma_1^2, \ldots, \sigma_s^2\}$. The conditional distribution of the sample observation $y_j$ given a value of the parameter vector is

$$f(y_j|\boldsymbol{\theta}) = \sum_{i=1}^{s} \lambda_i \varphi_i(y_j|\boldsymbol{\theta}) \qquad (2.2)$$

where:

$$\varphi_i(y_j|\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma_i^2)^{1/2}} \exp\left( -\frac{1}{2\sigma_i^2} (y_j - \boldsymbol{x}_{ij}'\boldsymbol{\beta}_i)^2 \right)$$

Therefore, the likelihood for the switching regression model is

$$L(\boldsymbol{\theta}\,|\,\boldsymbol{Y}) = \prod_{j=1}^{n} [\sum_{i=1}^{s} \lambda_i \varphi_i (y_j|\boldsymbol{\theta})] \qquad (2.3)$$

To avoid problems with noninformative priors only conjugate priors are considered in this study. The joint prior distribution of the parameters is

$$\begin{aligned} g(\boldsymbol{\theta}) &= g(\lambda_1, \ldots, \lambda_s, \boldsymbol{\beta_1}, \ldots, \boldsymbol{\beta_s}, \sigma_1^2, \ldots, \sigma_s^2) \\ &= g(\lambda_1, \ldots, \lambda_s) \prod_{i=1}^{s} [g(\boldsymbol{\beta_i}|\sigma_i^2)\ g(\sigma_i^2)] \end{aligned} \qquad (2.4)$$

assuming that the mixing parameters $(\lambda_1,\ldots,\lambda_s)$ are independent of the regime parameters $(\boldsymbol{\beta_1},\ldots,\boldsymbol{\beta_s},\sigma_1^2,\ldots,\sigma_s^2)$ and that $(\boldsymbol{\beta_i}, \sigma_i^2)$ is independent of $(\boldsymbol{\beta_{i'}}, \sigma_{i'}^2)$, $i \neq i'$. The assumptions regarding the form of the prior distributions are that the conjugate prior $g(\lambda_1,\ldots,\lambda_s)$ of $\lambda_i$ is the Dirichlet distribution with parameters $(\alpha_1,\ldots,\alpha_s)$ and the conjugate prior $g(\boldsymbol{\beta_i}, \sigma_i^2)$ is the normal-gamma distribution. The prior distribution $g(\boldsymbol{\beta_i}|\sigma_i^2)$ is then a multivariate normal with mean $\boldsymbol{A_i}$ and covariance $\sigma_i^2\boldsymbol{Q_i}$ and the prior distribution $g(\sigma_i^2)$ of $\sigma_i^2$ is then an inverse-gamma with parameters $\gamma_i/2$ and $v_i/2$. Combining the likelihood (2.3) and the prior distribution (2.4), the joint posterior distribution becomes,

$$g(\boldsymbol{\theta}\,|\,\boldsymbol{Y}) \propto \prod_{j=1}^{n}\left[\sum_{i=1}^{s}\lambda_i\varphi_i(y_j|\boldsymbol{\theta})\right] g(\lambda_1,\ldots,\lambda_s)\prod_{i=1}^{s}[g(\boldsymbol{\beta_i}|\sigma_i^2)\ g(\sigma_i^2)] \quad . \qquad (2.5)$$

Since the posterior distribution (2.5) takes into account all the possible partitions of the sample $\{y_1, y_2, \ldots, y_n\}$ into at most $s$ groups, implementing the Bayesian approach of the stochastic switching regression model becomes intractable as the sample size becomes reasonably large, like n = 50. As a solution to this computational problem in implementing the Bayesian approach, Markov Chain Monte Carlo methods can be used to sample indirectly from the joint posterior distribution.

## 3. MARKOV CHAIN MONTE CARLO METHODS FOR MIXED REGRESSION

Markov Chain Monte Carlo methods provide a way of estimating the features of a distribution by using samples drawn indirectly from the distribution. These methods are most valuable for complicated distributions, such as high-dimensional joint distributions, from which it is generally infeasible to sample directly. To circumvent the difficulty of sampling from a complicated distribution, the MCMC method constructs a Markov chain with equilibrium distribution identical to the desired distribution. Suppose $\theta^{(1)}, \theta^{(2)}, \ldots \theta^{(l)} \ldots, \theta^{(m)}$ is a realization of a Markov chain where $\theta^{(l)}$ converges to $\theta$ in distribution almost surely as m approaches infinity, then, the ergodic average of a function $h(\theta^{l})$ of θ will approach $E[h(\theta)]$ almost surely. In order to make this approach useful, the Markov chain must be easy to simulate from.

### 3.1 Gibbs Sampling

A form of the MCMC method that is especially suited to the problem at hand is Gibbs sampling. In Gibbs sampling, the parameter vector $\boldsymbol{\theta}$ is partitioned into r arbitrary groups, $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_r\}$. Instead of directly sampling from the complex joint conditional distribution a Markov chain is generated by sampling from the full conditional distributions

$$f\,(\theta_k \mid \mathbf{Y},\ \theta_1, \ldots \theta_{k-1},\ \theta_{k+1}, \ldots \theta_r)\ ,\ k = 1, \ldots, r$$

corresponding to the partitioning. Under regularity conditions (Tierney 1991) as t approaches infinity $\theta_k^{(m)}$ converges in distribution to the $\theta_k$ and $(\theta_1^{(m)}, \ldots \theta_k^{(m)} \ldots, \theta_r^{(m)})$ converges in distribution to $(\theta_1, \ldots \theta_k \ldots, \theta_r)$. In contrast to the maximum likelihood method EM and SEM algorithms and the bootstrap resampling method, the MCMC Bayesian methods are useful even for finite sample size since convergence results depend only on the number of iterations.

The key to estimating the parameters of the stochastic switching regression model, is to express the model in terms of missing (or latent) data $\mathbf{Z} = \{z_{ij}\}$ , i=1, 2,..., s , j=1, 2,..., n. The latent variable assumes the value 1 if the j th observation $y_j$ is generated by the i th regression model and 0 otherwise. If the data $\mathbf{Z} = \{z_{ij}\}$ were observed, estimation would be straightforward. However, since $\mathbf{Z}$ is not observed, the elements of $\mathbf{Z}$ can be viewed as unknowns that must be estimated along with the other model parameters. Consequently, the parameter vector is $\boldsymbol{\theta} = \{\theta_k\} = \{\mathbf{Z},\ \lambda_1 \ldots \lambda_s, \boldsymbol{\beta}_1 \ldots \boldsymbol{\beta}_s,\ \sigma_1^2 \ldots \sigma_s^2\}$ for the stochastic switching regression model. In Gibbs sampling the parameter vector $\boldsymbol{\theta}$ is broken down into subvectors, $\{\mathbf{Z}\}$, $\boldsymbol{\lambda} = \{\lambda_1,...,\lambda_s\}$, $\{\boldsymbol{\beta}_1,..., \boldsymbol{\beta}_s\}$, and $\{\sigma_1^2,...,\sigma_s^2\}$. The Gibbs sampling algorithm is then to generate parameter values from the full conditional distributions with starting values $\mathbf{Z}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$, $\sigma_i^{2\,(0)}$ and $\boldsymbol{\beta}_i^{(0)}$

(1) $\boldsymbol{Z}^{(m+1)} \sim f(\boldsymbol{Z} \mid \boldsymbol{Y}, \lambda^{(m)}, \boldsymbol{\beta}_i^{(m)}, \sigma_i^{2\,(m)})$

(2) $\boldsymbol{\lambda}^{(m+1)} \sim f(\boldsymbol{\lambda} \mid \boldsymbol{Y}, \boldsymbol{Z}^{(m+1)}, \boldsymbol{\beta}_i^{(m)}, \sigma_i^{2\,(m)})$

(3) $\sigma_i^{2\,(m+1)} \sim f(\sigma_i^2 \mid \boldsymbol{Y}, \boldsymbol{Z}^{(m+1)}, \lambda_i^{(m+1)}, \boldsymbol{\beta}_i^{(m)})$

(4) $\boldsymbol{\beta}_i^{(m+1)} \sim f(\boldsymbol{\beta}_i \mid \boldsymbol{Y}, \boldsymbol{Z}^{(m+1)}, \lambda_i^{(m+1)}, \sigma_i^{2\,(m+1)})$

After equilibrium is reached at the a th iteration, sample values are averaged to provide consistent estimates of the parameters or their function,

$$\hat{E}[h(\theta_k)] = \frac{\sum_{m=a+1}^{t} h(\theta_k)^{(m)}}{t - a}$$

The marginal posterior distribution is estimated as

$$f(\theta_k) \approx \frac{\sum_{m=a+1}^{t} f(\theta_k | \theta_1^{(m)},...\theta_{k-1}^{(m)},\theta_{k+1}^{(m)}...\theta_r^{(m)})}{t - a}$$

and the estimate of the predictive density is

$$\int f(\boldsymbol{Y}^f|\boldsymbol{\theta})\, g(\boldsymbol{\theta}|\boldsymbol{Y})\, d\boldsymbol{\theta} \approx \frac{\sum_{m=a+1}^{t} f(\boldsymbol{Y}^f|\boldsymbol{\theta}^{(m)}, \boldsymbol{X}^{(f)})}{t - a}$$

### 3.2 Data Augmentation

One form of Gibbs sampling that is well suited to the stochastic switching regression model is data augmentation, introduced by Tanner and Wong (1987). In data augmentation, the natural hierarchical structure of the model is incorporated in partitioning. For the stochastic switching regression model, the parameter vector $\mathbf{w} = \{\boldsymbol{Z}, \lambda_1, ..., \lambda_s, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_s, \sigma_1^2, ..., \sigma_s^2\}$ is partitioned into two groups, $\{\mathbf{Z}\}$ and $\boldsymbol{\theta} = \{\lambda_1, ..., \lambda_s, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_s, \sigma_1^2, ..., \sigma_s^2\}$ where each group corresponds to a hierarchical level. In data augmentation r independent values of of $\boldsymbol{Z}^{(m+1)}$ are drawn to reduce variability. Although r=1 is enough for point estimation, a larger r is required for small sample size (Diebolt and Robert 1994). The data augmentation algorithm is

(1) Generate r independent values of $\boldsymbol{Z}^{(m+1)}$ from the conditional distribution

$$\boldsymbol{Z}^{(m+1)} \sim f[\boldsymbol{Z} \mid (\boldsymbol{Y}, \boldsymbol{\theta}^{(m)})]$$

(2) Generate $\boldsymbol{\theta}^{(m+1)}$ from the conditional distribution

$$\boldsymbol{\theta}^{(m+1)} \sim f[\boldsymbol{\theta} \mid (\boldsymbol{Y}, \boldsymbol{Z}^{(m+1)})] \ .$$

Since $\boldsymbol{\theta}$ is a vector, values of the parameters are generated sequentially.

(2.1) $\boldsymbol{\lambda}^{(m+1)} \sim f[\boldsymbol{\lambda} \mid (\boldsymbol{Y}, \boldsymbol{Z}^{(m+1)})]$

(2.2) $\sigma_i^{2\ (m+1)} \sim f[\sigma_i^2 \mid (\boldsymbol{Y}, \boldsymbol{\lambda}^{(m+1)}, \boldsymbol{Z}^{(m+1)})]$

(2.3) $\boldsymbol{\beta}_i^{(m+1)} \sim f[\boldsymbol{\beta}_i \mid (\boldsymbol{Y}, \sigma_i^{2\ (m+1)}, \boldsymbol{\lambda}^{(m+1)}, \boldsymbol{Z}^{(m+1)})]$

where starting values $\boldsymbol{Z}^{(0)}$, $\boldsymbol{\theta}^{(0)}$ are specified. The expected value of the parameters or their function, the marginal as well as the predictive distributions are estimated similarly as in Gibbs sampling.

**3.3 Conditional Distributions**

The conditional distributions necessary to implement the MCMC methods Gibbs sampling and data augmentation are identical except for the conditional distributions of the error variances which differ in parameters. The conditional distribution of the latent variable $f[\boldsymbol{Z} \mid (\boldsymbol{Y}, \boldsymbol{\theta}^{(m)})]$ is a multinomial distribution

$$z_{ij}^{\ (m+1)} \mid \boldsymbol{Y}, \boldsymbol{\theta}^{(m)} \sim \mathit{Multinomial}(w_j^{\ (m+1)})$$

where:

$$w_{ij}^{(m+1)} = \frac{\lambda_i^{\ (m)}\, \varphi_i(y_j \mid \boldsymbol{\beta}_i^{(m)}, \sigma_i^{2\ (m)})}{\sum_{i=1}^{s} \lambda_i^{\ (m)}\, \varphi_i(y_j \mid \boldsymbol{\beta}_i^{(m)}, \sigma_i^{2\ (m)})}$$

$$\varphi_i(y_j \mid \boldsymbol{\beta}_i^{(m)}, \sigma_i^{2\ (m)}) = \frac{1}{(2\pi\sigma_i^{2\ (m)})^{1/2}} \exp\left( -\frac{1}{2\sigma_i^{2\ (m)}} (y_j - \boldsymbol{x}_{ij}{}'\boldsymbol{\beta}_i^{(m)})^2 \right)$$

The conditional distribution of the mixing parameters is a Dirichlet distribution

$$\boldsymbol{\lambda}^{(m+1)} \mid (\boldsymbol{Y}, \boldsymbol{Z}^{(m+1)}) \sim \mathit{Dirichlet}(\alpha_1 + n_1^{(m+1)}, \alpha_2 + n_2^{(m+1)}, \ldots, \alpha_s + n_s^{(m+1)})$$

where:

$$n_i^{\ (m+1)} = \sum_{j=1}^{n} z_{ij}^{\ (m+1)} \ .$$

is the number of observations in each regime according to the latent data $\mathbf{Z}$. Conditional upon the latent data $\mathbf{Z}$ which classifies each observation into a regime, the Bayesian posterior distributions for the standard regression model apply. Thus, the conditional distribution of the error variance $f[\sigma_i^{\,2} \,|\, (\boldsymbol{Y}, \boldsymbol{\lambda}^{(m+1)}, \boldsymbol{Z}^{(m+1)})]$ is an inverse-gamma

$$\sigma_i^{2\,(m+1)} \sim IG[(\gamma_i + n_i^{(m+1)}) / 2 \,,\, (\nu_i + \boldsymbol{S}_i^{(m+1)}) / 2]$$

where:

$$\boldsymbol{S}_i^{(m+1)} = \boldsymbol{Y}_i^{(m+1)\,\prime}\boldsymbol{Y}_i^{(m+1)} + \boldsymbol{A}_i^{\prime}\boldsymbol{Q}_i^{-1}\boldsymbol{A}_i - \boldsymbol{C}_i^{(m+1)\,\prime}(\boldsymbol{X}_i^{(m+1)\,\prime}\boldsymbol{X}_i^{(m+1)} + \boldsymbol{Q}_i^{-1})\boldsymbol{C}_i^{(m+1)}$$

$$\boldsymbol{C}_i^{(m+1)} = (\boldsymbol{X}_i^{(m+1)\,\prime}\boldsymbol{X}_i^{(m+1)} + \boldsymbol{Q}_i^{-1})^{-1}\,(\boldsymbol{X}_i^{(m+1)\,\prime}\boldsymbol{Y}_i^{(m+1)} + \boldsymbol{Q}_i^{-1}\boldsymbol{A}_i) \;.$$

The conditional distribution of the regression coefficients $f[\boldsymbol{\beta}_i \,|\, (\boldsymbol{Y}, \sigma_i^{2\,(m+1)}, \lambda^{(m+1)}, \boldsymbol{Z}^{(m+1)})]$ is a multivariate normal with mean $\boldsymbol{C}_i^{(m+1)}$ and variance-covariance matrix $\sigma_i^{2\,(m+1)}(\boldsymbol{X}_i^{(m+1)\,\prime}\boldsymbol{X}_i^{(m+1)} + \boldsymbol{Q}_i^{-1})^{-1}$

$$\boldsymbol{\beta}_i^{(m+1)} \sim N[\boldsymbol{C}_i^{(m+1)}, \sigma_i^{2\,(m+1)}(\boldsymbol{X}_i^{(m+1)\,\prime}\boldsymbol{X}_i^{(m+1)} + \boldsymbol{Q}_i^{-1})^{-1}]] \;.$$

For Gibbs sampling the conditional distribution of the error variance $f(\sigma_i^{\,2} \,|\, \boldsymbol{Y}, \boldsymbol{Z}^{(m+1)}, \lambda_i^{(m+1)}, \boldsymbol{\beta}_i^{(m)})$ is also an inverse gamma but with parameters $(\gamma_i + k_i + n_i^{(m+1)}) / 2$ and $(\nu_i + \boldsymbol{E}_i^{(m+1)}) / 2$ ,

$$\sigma_i^{2\,(m+1)} \,|\, \boldsymbol{\beta}_i^{(m)} \sim IG\left( \frac{\gamma_i + k_i + n_i^{(m+1)}}{2} , \frac{\nu_i + \boldsymbol{E}_i^{(m+1)}}{2} \right)$$

where :

$$\begin{aligned} \boldsymbol{E}_i^{(m+1)} &= (\boldsymbol{Y}_i^{(m+1)} - \boldsymbol{X}_i^{(m+1)}\boldsymbol{\beta}_i^{(m)})^{\prime}(\boldsymbol{Y}_i^{(m+1)} - \boldsymbol{X}_i^{(m+1)}\boldsymbol{\beta}_i^{(m)}) \\ &\quad + (\boldsymbol{\beta}_i^{(m)} - \boldsymbol{A}_i)^{\prime}\boldsymbol{Q}_i^{-1}(\boldsymbol{\beta}_i^{(m)} - \boldsymbol{A}_i) \;. \end{aligned}$$

**3.4 Specification of Hyperparameters**

In the following discussion a two component stochastic switching regression model in log-form will be considered. Bayesian estimation requires the prior distribution of the mixing parameter $\lambda \sim \text{Beta}\,(\alpha_1 , \alpha_2)$ . It is assumed that there is little prior information about $\lambda$ so $\alpha_1$ and $\alpha_2$ are set so that the mean of $\lambda$ was 0.5 and the standard deviation is 0.15. These values are $\alpha_1 = [1 - 4\,Var(\lambda)] / [8\,Var(\lambda)]$ and $\alpha_2 = \alpha_1$ .

Bayesian estimation also requires the prior distributions $\sigma_i^2 \sim IG(\gamma_i/2, \nu_i/2)$, $i = 1, 2$ , $\beta \sim N(\boldsymbol{A_1}, \sigma_1^{\,2}\boldsymbol{Q_1})$ and $\alpha \sim N(\boldsymbol{A_2}, \sigma_2^{\,2}\boldsymbol{Q_2})$ . When a regression model is fit in log-form, the model parameters are interpreted as elasticities. Economists often have prior beliefs about elasticities. Let the prior values for the means of the regression parameters be $\boldsymbol{A_1} = [\hat{\beta}_l]$ and $\boldsymbol{A_2} = [\hat{\alpha}_l]$. The actual values used for $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$ will be discussed in the simulation section. Given the specification of $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$, the parameters $\gamma_i$ and $v_i$ i (i=1, 2) were set so that the mean of $\sigma_1^2$ was $\sum_{l=1}^{k} \hat{\beta}_l^2$ and the mean of $\sigma_2^2$ was $\sum_{l=1}^{k} \hat{\alpha}_l^2$. These means give a coefficient of determination $R^2$ of 0.50 , which is a typical value for economic data. It is assumed that there is weak prior information about the error variance thus, $\gamma_i$ and $v_i$ were also chosen so that the standard deviation of $\sigma_1^2$ was $\sum_{l=1}^{k} \hat{\beta}_l^2 / 3$ and $\sigma_2^2$ was $\sum_{l=1}^{k} \hat{\alpha}_l^2 / 3$. These choices set to 0, the lower bound for $\sigma_i^2$ three standard deviations from the mean. Finally, values for $\boldsymbol{Q_1}$ and $\boldsymbol{Q_2}$ are required. Let the economist 's prior values for the standard deviation of $\beta_l$ be $\sigma_{\beta_l}$ and for the standard deviation of $\alpha_l$ be $\sigma_{\alpha_l}$. Small values of $\sigma_{\beta_l}$ and $\sigma_{\alpha_l}$ reflect high levels of confidence in the prior means $\hat{\beta}_l$ and $\hat{\alpha}_l$. For simplicity it is assumed that $\boldsymbol{Q_1}$ and $\boldsymbol{Q_2}$ are diagonal. The diagonal elements of $\boldsymbol{Q_1}$ are set at $\sigma_{\beta_l}^2 / \mu_{\sigma_1^2}$ and the diagonal elements of $\boldsymbol{Q_2}$ are set at $\sigma_{\alpha_l}^2 / \mu_{\sigma_2^2}$ where $\mu_{\sigma_1^2}$ and $\mu_{\sigma_2^2}$ are the expected values of $\sigma_1^2$ and $\sigma_2^2$ respectively.

## 4. SIMULATION

### 4.1 Methodology

Bayesian estimation of two different switching regression models is investigated. The first model is a market disequilibrium model

$$y_{1j} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \varepsilon_{1j}, \quad \varepsilon_{1j} \sim N(0,\sigma_1^2) \tag{4.3}$$

$$y_{2j} = \alpha_0 + \alpha_1 x_{1j} + \alpha_2 x_{4j} + \varepsilon_{2j}, \quad \varepsilon_{2j} \sim N(0,\sigma_2^2) \tag{4.4}$$

$$y_j = y_{1j} \quad \text{with probability} \quad \lambda$$

$$y_j = y_{2j} \quad \text{with probability} \quad (1-\lambda)$$

where $x_1$ represents price, $x_2$ is a random supply shock and $x_4$ is a random demand shock. The prior parameter values are values are $\hat{\beta}_0 = 0$, $\hat{\beta}_1 = 1$, $\hat{\beta}_2 = 1$, $\hat{\alpha}_0 = 0, \hat{\alpha}_1 = -1, \hat{\alpha}_2 = 1$, which are typical elasticities.

The second model represents a structural change model

$$y_{1j} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \varepsilon_{1j} , \quad \varepsilon_{1j} \sim N(0,\sigma_1^2) \tag{4.5}$$

$$y_{2j} = \alpha_0 + \alpha_1 x_{1j} + \alpha_2 x_{2j} + \varepsilon_{2j} , \quad \varepsilon_{2j} \sim N(0,\sigma_2^2) \tag{4.6}$$

$$y_j = y_{1j} \qquad \text{with probability} \quad \lambda$$

$$y_j = y_{2j} \qquad \text{with probability} \quad (1-\lambda)$$

The prior parameter values are $\hat{\beta}_0 = 0$ , $\hat{\beta}_1 = 1$ , $\hat{\beta}_2 = 1$ , $\hat{\alpha}_0 = 0$ , $\hat{\alpha}_1 = 0.5$ , $\hat{\alpha}_2 = 0.5$. Again, these are typical elasticity values.

Two factors were investigated in the simulation. The first factor is the economist's degree of belief in the prior means on the regression coefficients **β** and **α**. This was investigated by performing one complete simulation for $\sigma_{\beta_l} = \sigma_{\alpha_l}$ , $l > 0$, set at each value 0.1, 0.2,..., 2.0 . The standard deviation of $\sigma_{\beta_0} = \sigma_{\alpha_0}$ was fixed at 5.0 . The second factor is whether the economist's prior means on **β** and **α** were correct. This was investigated by setting the true regression coefficients equal to a certain fraction of the prior means. The fractions used are m = 1.00, 0.75 , 0.50 , 0. 25 . The prior distributions on the regression parameters are $N(\boldsymbol{A_i}, \sigma_i^2 \boldsymbol{Q_i})$ while the true regression parameters are $N(m\boldsymbol{A_i}, \sigma_i^2 \boldsymbol{Q_i})$. The constant m represents "bias" in the prior distributions. When m = 1 the true and prior distributions are identical. For m less than 1 , the prior distribution is "biased" upward.

There are a total of 2 x 20 x 4 = 160 distinct sets of prior values, one for each of the two model at the twenty levels of $\sigma_{\beta_l} = \sigma_{\alpha_l}$ and four levels of m . The sample size was set at n = 50 which is typical of economic data. For each of the 160 distinct sets of initial values, 400 data sets where used. Occasionally, the methods gave extremely poor estimates. Hence, for each set of prior values, the simulation results are summarized by computing a 5% trimmed mean of the absolute deviations from the true parameter values .

### 4.2. Results

Simulation results in Figures 1.1-1.4 and Figures 2.1-2.4 of the appendix show that in all cases considered, the Bayesian methods data augmentation and Gibbs sampling performed similarly. Also, the underlying model structure has relatively little impact on estimator performance. The general relationship between the error and the prior bias and the prior standard deviation appears to be the same for both the market disequilibrium and the structural change model, When there is no (m=1.00) or little bias (m=0.75) in the means of the prior distributions, the Bayesian sampling methods provided estimates with mean absolute errors smaller than the maximum likelihood method estimates, for prior parameter standard deviations less than 1.0. For prior parameter standard deviations between 1.0 and 2.0, the Bayesian and the maximum likelihood methods provided similar estimates. Since the regression coefficients for these models in log - form, may be interpreted as elasticities, the economist's prior information may well have standard deviations less than 1.0. This suggests that the Bayesian methods will provide a practical improvement over the maximum likelihood method, when the economist has some considerably accurate prior beliefs about the parameter means. With moderate bias (m=0.50) in the prior mean, the Bayesian and maximum likelihood estimates performed similarly. When bias was large ( m = 0.25 ), the maximum likelihood method performed better than the Bayesian sampling methods for small parameter standard deviations. The reason is that a considerable degree of confidence is attached to an incorrect prior value.

## 5. CONCLUSION

A Bayesian method of estimating the parameters of the stochastic switching regression model is developed. MCMC methods data augmentation and Gibbs sampling are implemented to facilitate computation of posterior means. Bayesian sampling estimates are compared with those of the maximum likelihood EM algorithm. The results of this study suggest that when there is some reliable prior information about the parameters from previous research or as suggested by economic theory, the Bayesian sampling methods are worth implementing for estimating and updating parameters since they provide more accurate estimates than the maximum likelihood method. When there is no reliable information about the parameters, maximum likelihood method is preferable to Bayesian estimation for small prior standard deviations. The sample size used for this study is relatively large n=50 to show that MCMC methods are easily implemented even for large samples. Results may differ when sample size is small because of the well documented poor performance of the EM and SEM algorithms and maximum likelihood method in general for small or even moderately large sample size.


## ACKNOWLEDGMENTS

I would like to thank my adviser Dr. Mark McNulty of the Los Alamos National Laboratory and my professors Dr. Paul Nelson and Dr. Shieh-Shiehn Yang of the Kansas State University Statistics Department. I am also grateful to my parents Benita and Juanito and my sister Lucila.


## REFERENCES


Bernardo, J. and F. Giron (1988) “A Bayesian Analysis of Simple Mixture Problems”, *Bayesian Statistics 3* (Bernardo, J., M. H. Degroot, D. V. Lindley and A. F. M. Smith, eds) Oxford University Press: New York.

Broemeling, L. (1985) *Bayesian Analysis of Linear Models*, Marcel Dekker : New York.

Beard, R., Caudill, S., and D. Gropper (1991) ”Finite Mixture Estimation of Multiproduct Cost Functions,” *Review of Economics and Statistics*, 73, 654-664.

Celeux, G and J. Diebolt (1985) “The SEM Algorithm: A Probabilistic Teacher Algorithm Derived from the EM Algorithm for the Mixture Problem”, *Computational Statistics Quarterly*, 2, 73-82.

Celeux, G. and J. Diebolt (1990) “Une versio de type recruit simule de l’algorithme EM”, *C. R. Acad. Sci. Paris*, 310, 119-124.

Dempster, A. P. , N. M. Laird and D. B. Rubin (1977) “Maximum Likelihood from Incomplete Data Via the EM Algorithm,” *Journal of the Royal Statistical Society (B)*, 39, 1-38.

Diebolt, J. and C. Robert (1994) "Estimation of Finite Mixture Distributions through Bayesian Sampling", *Journal of the Royal Statistical Society B*, 56, 363-375.

Geman, S. and D. J. Geman (1984) “Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis Economy*, 82, 1119-1143.

Hartley, M. J. (1978) Comment on "Estimating mixtures of Normal Distributions and Switching Regressions", *Journal of the American Statistical Association*, 73,738-741.

Kon, S. and F. Jen (1978) "Estimation of Time -Varying Systematic Risk and Performance for Mutual Fund Portfolios: An Application of Switching Regression", *Journal of Finance*, 33, 457-475.

Makov, U. E. (1988) Discussion on "A Bayesian Analysis of Simple Mixture Problems", by Bernardo, J. and F. Giron *Bayesian Statistics 3* (Bernardo, J., M. H. Degroot, D. V. Lindley and A. F. M. Smith, eds) Oxford University Press: New York, 76-78.

Quandt, R. (1972) "A New Approach to Switching Regressions", Journal of the American Statistical Association, 67, 306-311.

Quandt, R., J. Ramsey, B. James (1978) "Estimating Mixtures of Normal Distributions and Switching Regressions", *Journal of the American Statistical Association*, 73, 730-738.

Robert, C. (1994) *The Bayesian Choice*, Springer-Verlag New York Inc. : New York.

Smith, A. F. M. and G. O. Robert (1993) "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods", *Journal of the Royal Statistical Society (B)*, 55, 3-23.

Tanner, M. and W. Wong (1987) "The Calculation of Posterior Distributions by Data Augmentation (With Discussion)", *Journal of the American Statistical Association*, 82, 528-550.

Tierney, L. (1991) "Exploring Posterior Distributions Using Markov Chains", *Computing Science and Statistics: 23 Symposium on the Interface* (Keramidas, E., ed) A. S. A. : Alexandria, Virginia.

Titterington, D., A. Smith and U. Makov (1985) *Statistical Analysis of Finite Mixture Distributions*, J. Wiley : New York.

## APPENDIX

**Figure 1.1**
**Market Disequilibrium Model**
True Parameter = Prior x 1.00

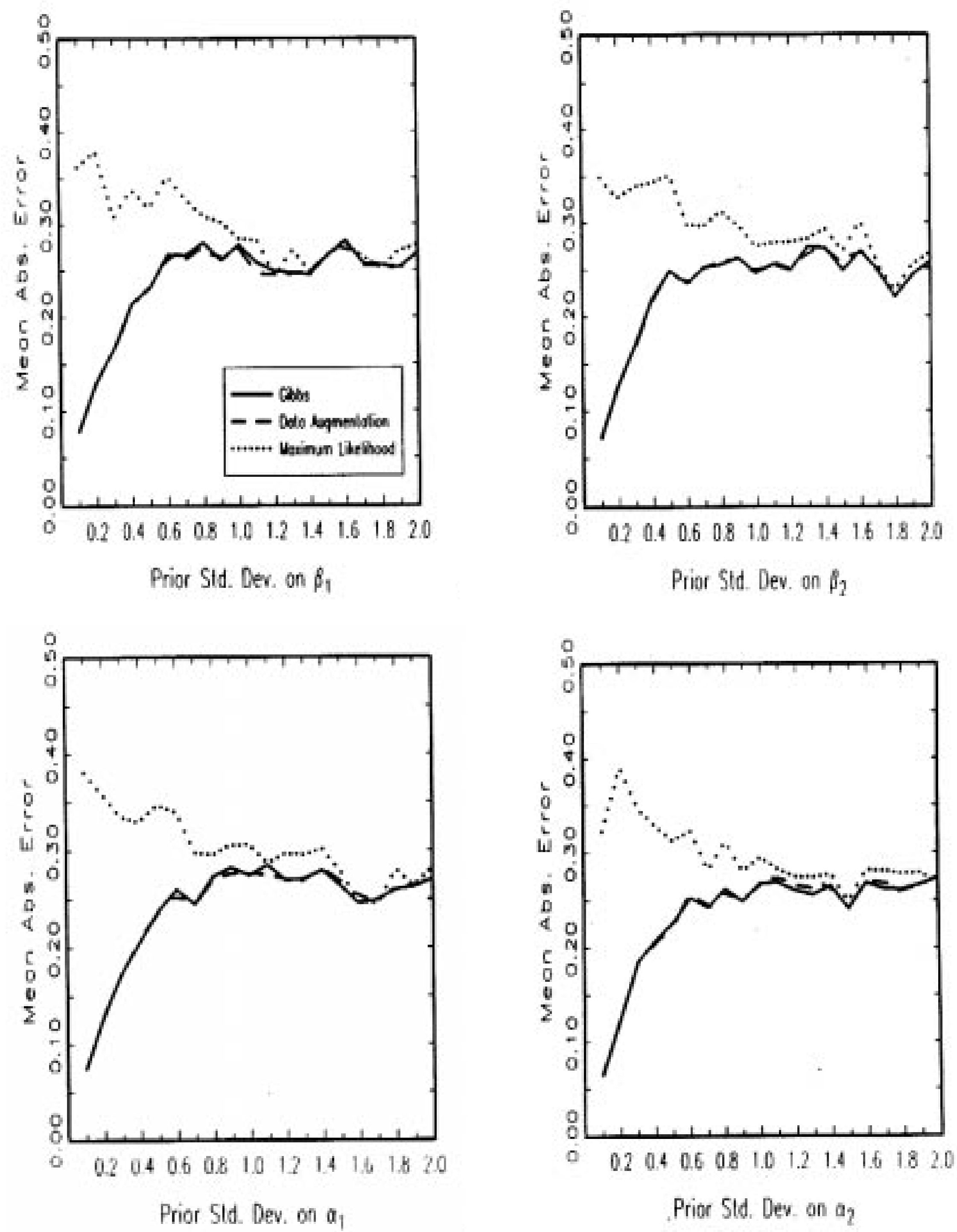

**Figure 1.2**
**Market Disequilibrium Model**
True Parameter = Prior x 0.75

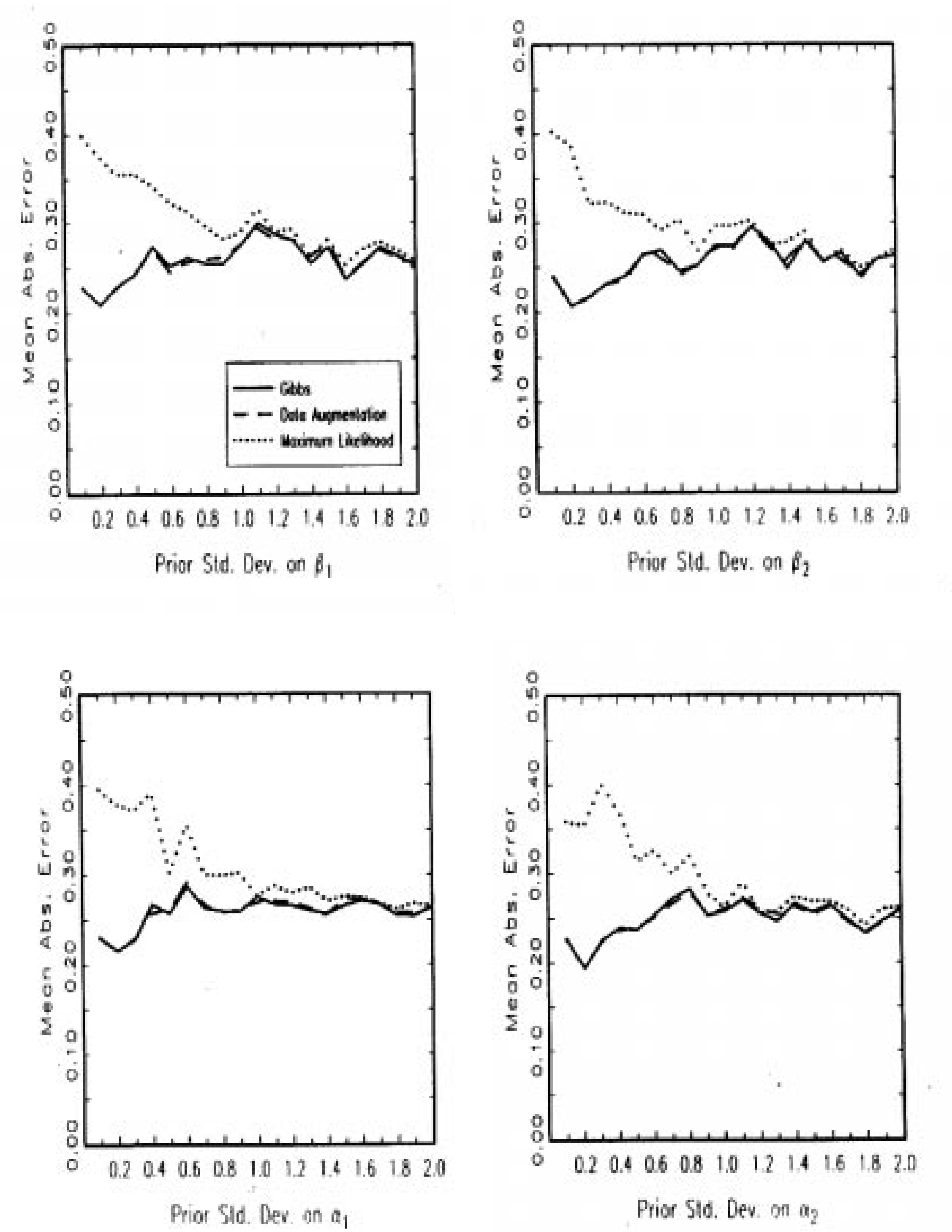

**Figure 1.3**
**Market Disequilibrium Model**
True Parameter = Prior x 0.50

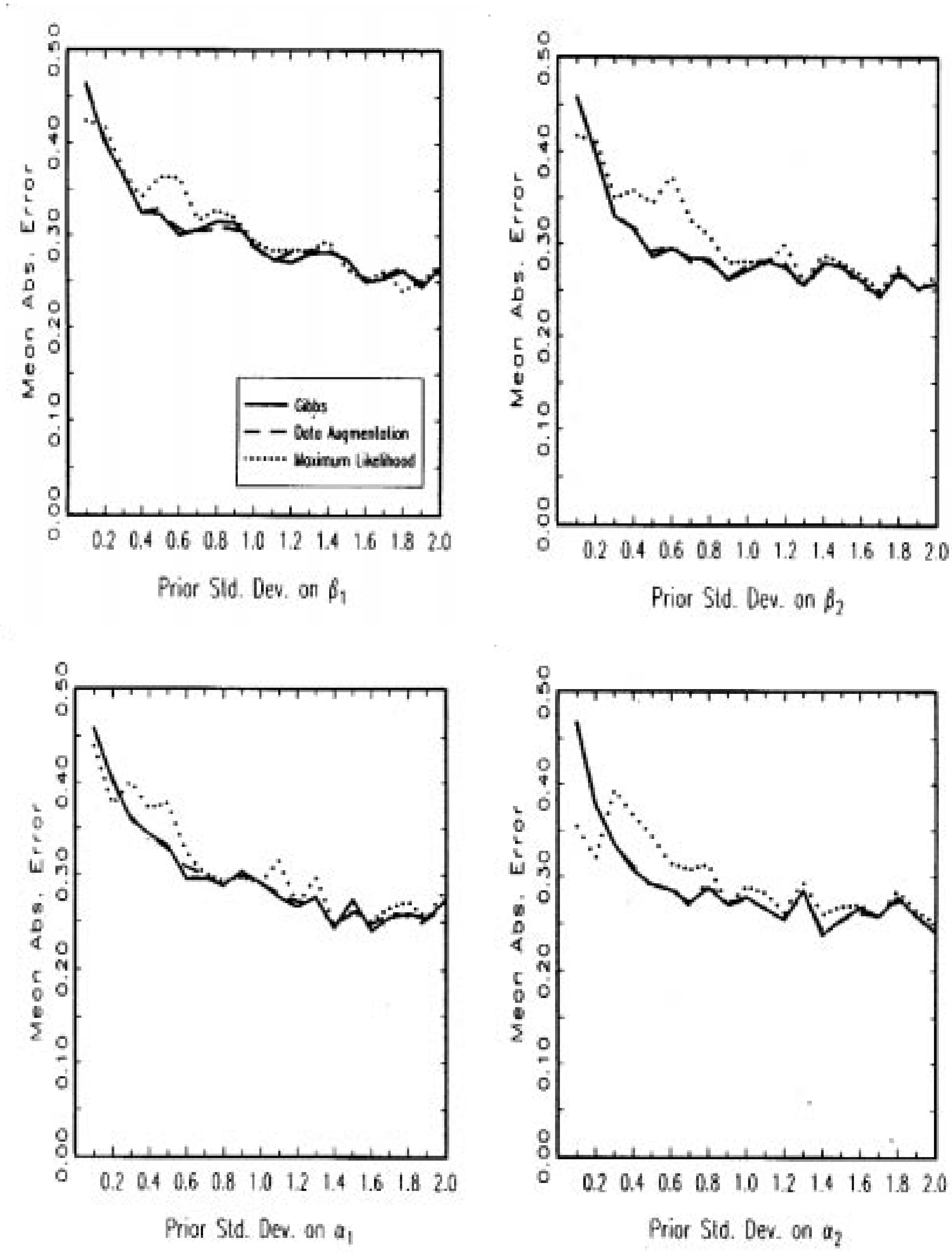

**Figure 1.4**
**Market Disequilibrium Model**
True Parameter = Prior x 0.50

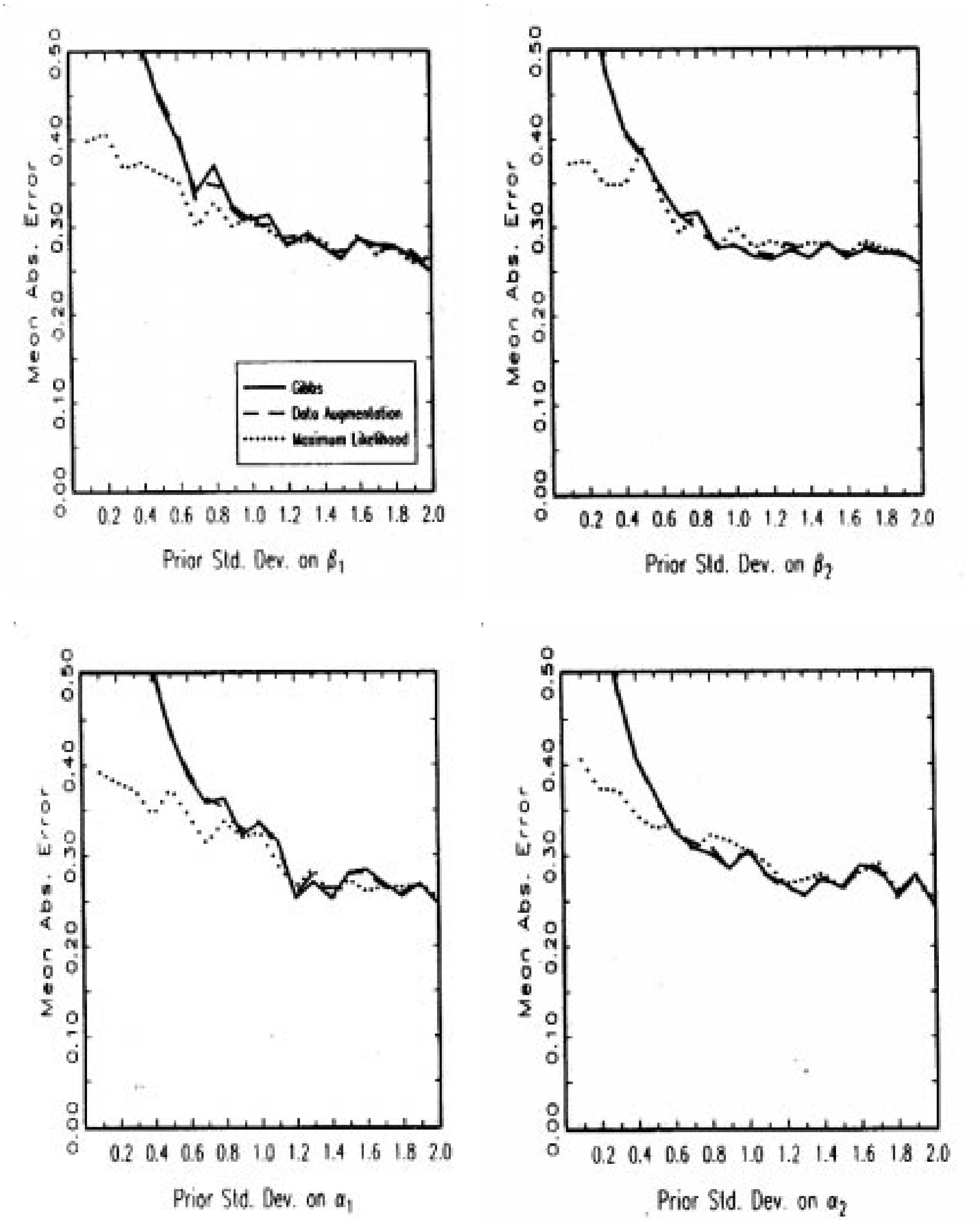

**Figure 2.1**
**Structural Change Model**
True Parameter = Prior x 1.00

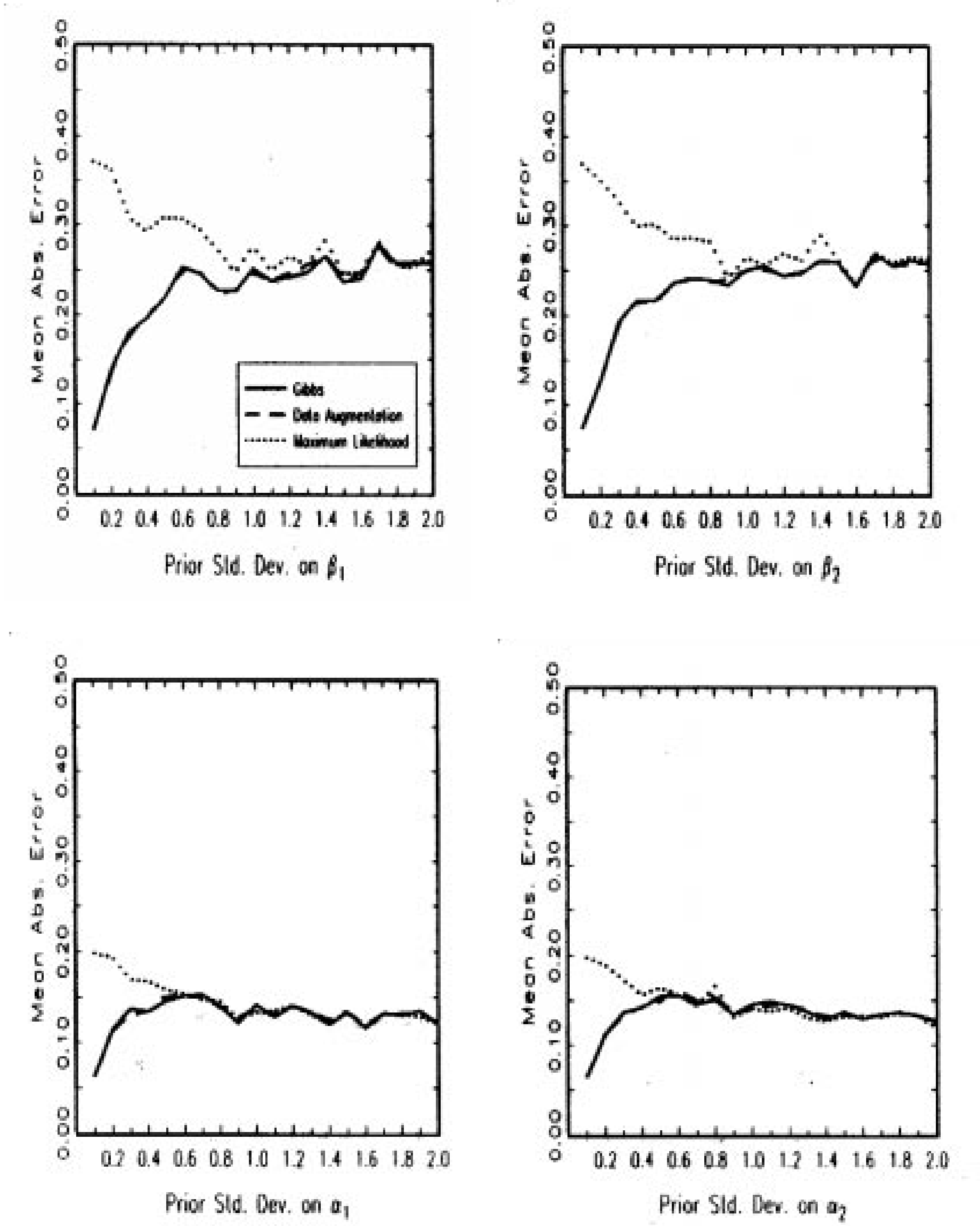

**Figure 2.2**
**Structural Change Model**
True Parameter = Prior x 0.75

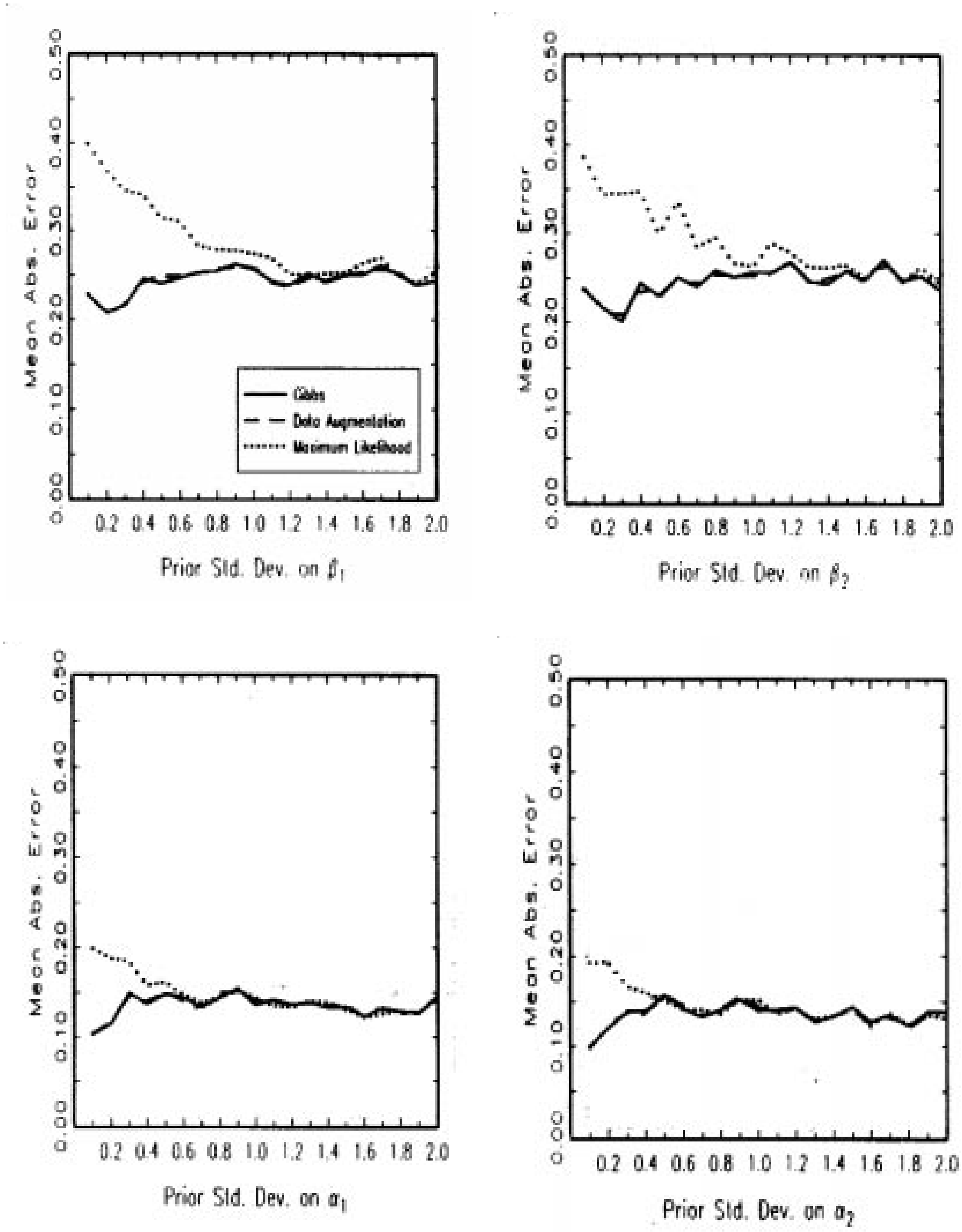

**Figure 2.3**
**Structural Change Model**
True Parameter = Prior x 0.50

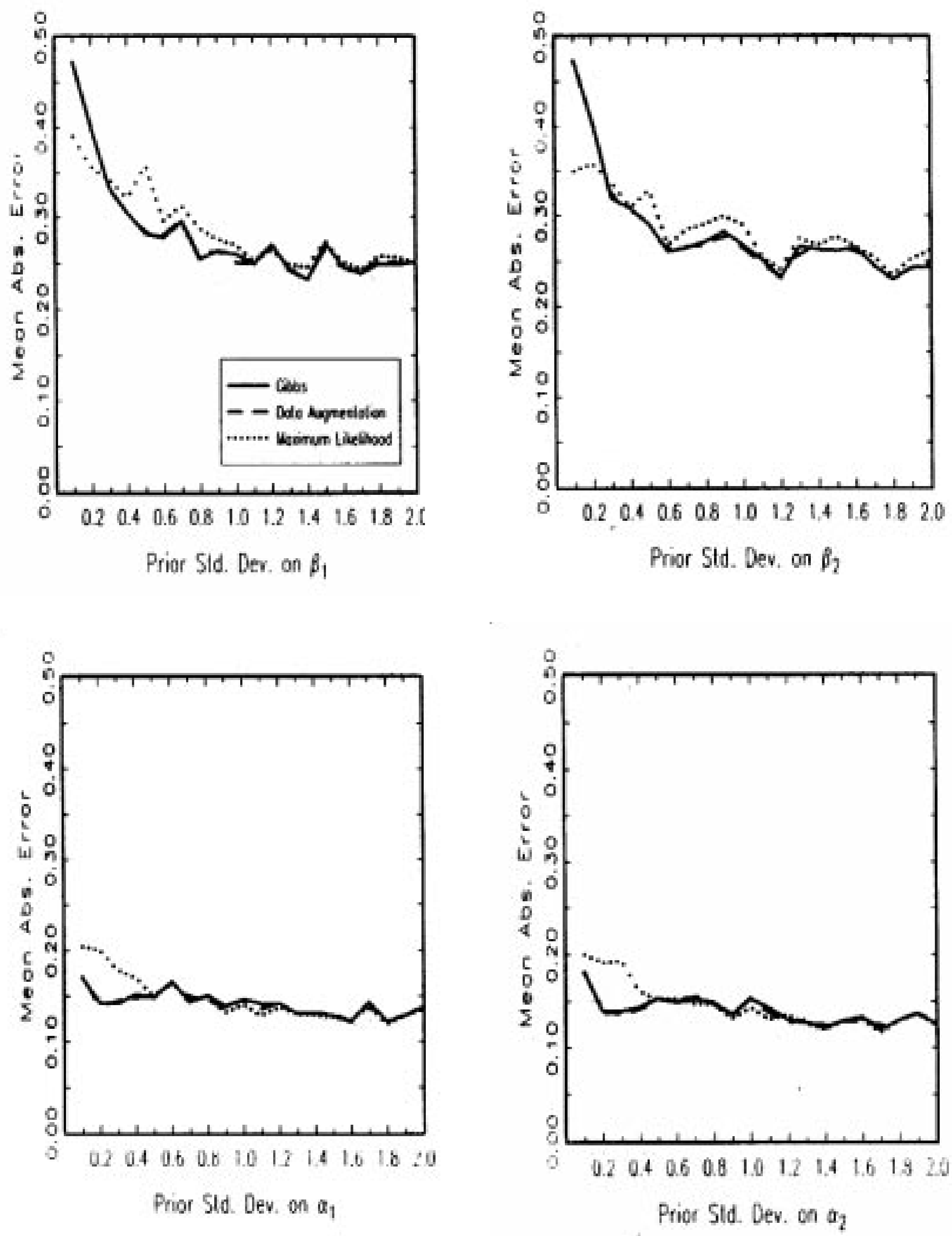

**Figure 2.4**
**Structural Change Model**
True Parameter = Prior x 0.25