REGIME SWITCHING VECTOR AUTOREGRESSIONS: A BAYESIAN MARKOV CHAIN MONTE CARLO APPROACH

Glen R. Harris

Australian Mutual Provident Society
33 Alfred Street
SYDNEY AUSTRALIA 2000

Telephone: 61 2 9257 6752
Facsimile : 61 2 9257 5278
Email: Glen_Harris@amp.com.au

ABSTRACT

Many financial time series processes appear subject to periodic structural changes in their dynamics. Regression relationships are often not robust to outliers nor stable over time, whilst the existence of changes in variance over time is well documented. This paper considers a vector autoregression subject to pseudocyclical structural changes. The parameters of a vector autoregression are modelled as the outcome of an unobserved discrete Markov process with unknown transition probabilities. The unobserved regimes, one for each time point, together with the regime transition probabilities, are to be determined in addition to the vector autoregression parameters within each regime.

A Bayesian Markov Chain Monte Carlo estimation procedure is developed which generates the joint posterior density of the parameters and the regimes, rather than the more common point estimates. The complete likelihood surface is generated at the same time. The procedure can readily be extended to produce joint prediction densities for the variables, incorporating parameter uncertainty.

Results using simulated and real data are provided. A clear separation of the variance between regimes is observed. Ignoring regime shifts is very likely to produce misleading volatility estimates, and is unlikely to be robust to outliers. A comparison with commonly used models suggests that the regime switching vector autoregression provides a particularly good description of the data.

Keywords: regime switching; joint parameter density; joint prediction densities; outliers; robust estimation; Gibbs sampler; Markov chains; Bayesian estimation.

1. INTRODUCTION

Stock & Watson (1996) examined the stability and predictive ability of 8 univariate models for each of 76 monthly U.S. times series, and 8 bivariate models for each of 5,700 bivariate relationships. They found evidence of substantial instability in a significant proportion of the univariate and bivariate autoregressive models considered.

Conditional heteroscedasticity, or changes in the level of volatility, has been found in financial series by numerous researchers, both actuarial and from the wider financial and econometric fields. Examples of the former include Praetz (1969), Becker (1991), Harris (1995b), Frees et al (1996) and Harris (1996). Examples of the latter include McNees (1979), Engle (1982), Akgiray (1989), Hamilton & Susmel (1994), Hamilton & Lin (1996) and Gray (1996).

The proposition underlying regime switching models is that, over time, changes in the financial environment may be closely associated with relatively discrete specific events. The process may have quite different characteristics in different regimes. A tractable mathematical model of structural changes and discrete market phases is the univariate Markov regime switching autoregressive process introduced by Hamilton (1989), and considered by Albert & Chib (1993) and Harris (1996).

Given that financial series appear interdependent, both in terms of their levels and their volatilities, e.g. Harris (1994, 1995b, 1995c) and Hamilton & Lin (1996), a vector regime switching process would seem to be an attractive description of the data. Hamilton (1990) proposed an EM maximum likelihood algorithm for estimating a Markov regime switching vector autoregression. The present paper develops an alternative Bayesian Markov Chain Monte Carlo (MCMC) estimation procedure which is more informative, flexible, and efficient than a maximum likelihood based approach.

In the process being considered, the various series are able to interact through regression relationships in the conditional means and through contemporaneous correlations in the residuals, as well as through joint regime switching in the conditional means, regressive correlations, variances and contemporaneous error correlations. Within each regime the process is assumed linear stationary. Joint regime switching produces nonlinear dependence between the series, and can account for discrete market phases and cycles, episodes of instability, and leptokurtic (i.e. fat-tailed) frequency distributions. In the univariate case, the model fitting results of Gray (1996) and Harris (1996) suggest that regime switching models compare more than favourably with common autoregressive and conditional heteroscedasticity models. The results in section 6.3 of the present paper suggest that the same is true of regime switching vector autoregressions.

The model is described in section 2. Markov Chain Monte Carlo methods, in the form of the Gibbs sampler and Metropolis-Hastings algorithm, are introduced in section 3, while the likelihood function is considered in section 4. The derivation of the Bayesian MCMC estimation procedure is presented in section 5, followed by model fitting results in section 6. Concluding remarks are made in section 7.

2. THE MODEL

Regime switching processes are characterised by multiple discrete regimes (states or phases), where each regime has different dynamics and is characterised by a different set of parameters. They are subject to probabilistic discrete shifts in the parameters. Within each regime the process is assumed stable a priori, and is hence linear stationary. The effect of the discrete regime shifting is to make the total process nonlinear stationary. The task, based on the observed data, is to make probabilistic inferences about when transitions between the various regimes occurred, the parameters characterising the different regimes, and the regime transition probabilities.

Define ρ_t to be a discrete-valued indicator variable, such that at any time t the process will be in regime $\rho_t \in \{1,..,K\}$. Define the transition probabilities, $p_{ij} = p(\rho_t = j \mid \rho_{t-1} = i)$, with $\sum_j p_{ij} = 1 \;\; \forall i$, and $\mathbf{P}^T \equiv \{p_{ij}\}$ $(K\times K)$.

Consider the following VAR(q) time series process with K discrete regimes, where each regime is characterised by a different set of parameters,

$$\mathbf{x}_t = \boldsymbol{\mu}_{(\rho_t)} + \sum_{h=1}^{q} \mathbf{A}_{(\rho_t)}^{(h)}\left(\mathbf{x}_{t-h} - \boldsymbol{\mu}_{(\rho_t)}\right) + \boldsymbol{\xi}_{t(\rho_t)},$$

$\boldsymbol{\xi}_{t(\rho_t)} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Omega}_{(\rho_t)}\right)$, $\mathrm{E}\boldsymbol{\xi}_{t(\rho_t)} = \mathbf{0}$ and $\mathrm{E}\boldsymbol{\xi}_{t(\rho_t)}\boldsymbol{\xi}_{t(\rho_t)}^T = \boldsymbol{\Omega}_{(\rho_t)} \;\; \forall t > q$. $\mathbf{x}$, $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$ are $m\times 1$ column vectors, while the $\mathbf{A}^{(h)}$ and the $\boldsymbol{\Omega}$ are $m\times m$ matrices. For convenience the above regime switching VAR(q) process will be denoted an RSVAR(q,K) process.

It will sometimes be more convenient to consider the equivalent VAR(1) form, namely

$$\mathbf{X}_t = \tilde{\boldsymbol{\mu}}_{(\rho_t)} + \mathbf{A}_{(\rho_t)}\left(\mathbf{X}_{t-1} - \tilde{\boldsymbol{\mu}}_{(\rho_t)}\right) + \tilde{\boldsymbol{\xi}}_{t(\rho_t)}$$

i.e.
$$\begin{pmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \\ \vdots \\ \mathbf{x}_{t-q+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_{(\rho_t)} \\ \boldsymbol{\mu}_{(\rho_t)} \\ \boldsymbol{\mu}_{(\rho_t)} \\ \vdots \\ \boldsymbol{\mu}_{(\rho_t)} \end{pmatrix} + \begin{pmatrix} \mathbf{A}_{(\rho_t)}^{(1)} & \mathbf{A}_{(\rho_t)}^{(2)} & \cdots & \mathbf{A}_{(\rho_t)}^{(q)} \\ \mathbf{I}_m & \mathbf{0}_m & \cdots & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{I}_m & \cdots & \mathbf{0}_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_m & \mathbf{0}_m & \cdots & \mathbf{0}_m \end{pmatrix} \left(\begin{pmatrix} \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \\ \mathbf{x}_{t-3} \\ \vdots \\ \mathbf{x}_{t-q} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu}_{(\rho_t)} \\ \boldsymbol{\mu}_{(\rho_t)} \\ \boldsymbol{\mu}_{(\rho_t)} \\ \vdots \\ \boldsymbol{\mu}_{(\rho_t)} \end{pmatrix} \right) + \begin{pmatrix} \boldsymbol{\xi}_{t(\rho_t)} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

$E\tilde{\xi}_{t(\rho_t)} = \mathbf{0}$ and $E\tilde{\xi}_{t(\rho_t)}\tilde{\xi}^T_{t(\rho_t)} = \tilde{\Omega}_{(\rho_t)}$ $\forall t > q$. $\mathbf{X}, \tilde{\mu}$ and $\tilde{\xi}$ are $mq \times 1$ column vectors, while the $\mathbf{A}$ are $mq \times mq$ matrices.

The total parameter set to be estimated is $\lambda \equiv \left\{\mu_{(1)},..,\mu_{(K)},\mathbf{A}_{(1)},..,\mathbf{A}_{(K)},\Omega_{(1)},..,\Omega_{(K)},\mathbf{P}\right\}$, which can be partitioned as $\lambda \equiv \{\Theta,\mathbf{P}\}$. To ensure that the process is identifiable, it will sometimes be necessary to define the regimes by insisting upon prior restrictions on the parameters, such as ordering of the variances of at least one of the variables (components of the $\mathbf{x}_t$). If this is not done, it is possible that the regime associated with essentially the same set of data points could be labelled differently in different iterations of the estimation procedure.

The next 3 sections, sections 3 to 5, develop the procedure used to generate the joint parameter density and estimate the model. Those who do not wish to consider the mathematics of the estimation procedure at this stage may wish to skip the next 14 or so pages and go directly to section 6.2, which reports the results of fitting the model to real data.

3. Markov Chain Monte Carlo Methods

Draws from the joint posterior distribution of the regimes and the parameters, given the sample data, can be simulated using Markov Chain Monte Carlo methods, such as the Gibbs sampler and Metropolis-Hastings algorithm. Chib & Greenberg (1995) provide a useful and readable description of MCMC methods.

Markov Chain theory would usually start with a transition matrix in the case of discrete states, $\mathbf{P}^T = \{p_{ij}\}$, or a transition kernel density in the case of continuous states, $p(x,y)$. Since the process must end up somewhere at each transition, $\sum_j p_{ij} = 1$ or $\int p(x,y)dy = 1$. The probability of the process being in state j after n transitions, given that it was initially in state i, is given by $p_{ij}^{(n)} = \sum_k p_{ik}^{(n-1)} p_{kj}$ (discrete case). A limiting or invariant distribution is said to exist whenever $p_{ij}^{(n)} \to \pi_j$ as $n \to \infty$. It follows therefore that $\pi_j = \sum_i \pi_i p_{ij}$ or $\pi(y) = \int \pi(x)p(x,y)dx$.

A major concern of the theory is to determine conditions under which there exists an invariant distribution, and conditions under which iterations of the transition matrix or kernel converge to the invariant distribution.

MCMC methods look at the theory from a different perspective. The invariant distribution is the target distribution from which we wish to sample, generally a Bayesian posterior distribution. The transition matrix or kernel is unknown.

3.1 THE GIBBS SAMPLER

The Gibbs Sampler generates samples from a joint density $f(X) = f(X_1,..,X_N)$ via a sequence of random draws or samples from full conditional densities, as follows

$$X_1^{(r+1)} \leftarrow f\left(X_1 | X_2^{(r)},..,X_N^{(r)}\right)$$
$$\vdots$$
$$X_j^{(r+1)} \leftarrow f\left(X_j | X_{i<j}^{(r+1)}, X_{i>j}^{(r)}\right)$$
$$\vdots$$
$$X_N^{(r+1)} \leftarrow f\left(X_N | X_1^{(r+1)},..,X_{N-1}^{(r+1)}\right).$$

That completes a transition from $X^{(r)}$ to $X^{(r+1)}$. The sequence $\{X^{(r)}\}$ forms a realisation of a Markov chain which converges in distribution to a random sample from the joint distribution $f(X) = f(X_1,..,X_N)$.

3.2 THE METROPOLIS-HASTINGS ALGORITHM

Suppose $p(x,y)$ is unknown, but that a density $q(x,y)$ exists, $\int q(x,y)dy = 1$, from which candidate values of y can be generated for given x, to be accepted or rejected. The candidate generating density, $q(x,y)$, is a first approximation to the unknown transition kernel density, $p(x,y)$. $q(x,y)$ needs to be modified to ensure convergence to the desired target density. This is done by introducing a move probability, $\alpha(x,y) < 1$. If a move is not made, with probability $1 - \alpha(x,y)$, then the process remains at x and again returns a value of x as a value from the target distribution. The move probability is given by

$$\alpha(x,y) = \begin{cases} \min\left\{\dfrac{\pi(y)}{\pi(x)} \cdot \dfrac{q(y,x)}{q(x,y)}, 1\right\} & \pi(x) \cdot q(x,y) > 0 \\ 1 & otherwise \end{cases}.$$

An important feature of the algorithm is that the calculation of $\alpha(x,y)$ only requires knowledge of the target density $\pi(\cdot)$ up to proportionality (which in the case of a Bayesian posterior is given by the product of the likelihood and the prior), since $\pi(\cdot)$ only appears as a ratio.

A particularly useful application of the Metropolis-Hastings algorithm is where an intractable density arises within a Gibbs Sampler as the product of a standard density and another density, e.g. $\pi(x) \propto \psi(x)\cdot\phi(x)$, where $\phi(x)$ is a standard density that can be sampled. Then $q(x,y) = \phi(y)$ can be used to generate candidate y, which are accepted with probability $\alpha(x,y) = \min\{\psi(y)/\psi(x), 1\}$. The Metropolis-Hastings algorithm will be superior to direct acceptance/rejection methods since the move probability will be higher than $\psi(\cdot)$, the acceptance probability under the acceptance/rejection method, particularly where $\psi(\cdot)$ is small.

4. The Likelihood Function

The contribution of the t-th data vector to the likelihood conditional on the regime is

$$l(\mathbf{x}_t|\rho_t,\mathbf{Y}_{t-1},\boldsymbol{\lambda})=(2\pi)^{-\frac{m}{2}}\cdot\left|\boldsymbol{\Omega}_{(\rho_t)}^{-1}\right|^{\frac{1}{2}}\cdot\exp\left\{-\tfrac{1}{2}\boldsymbol{\xi}_{t(\rho_t)}^T\boldsymbol{\Omega}_{(\rho_t)}^{-1}\boldsymbol{\xi}_{t(\rho_t)}\right\}$$

$$\boldsymbol{\xi}_{t(\rho_t)}=\mathbf{x}_t-\boldsymbol{\mu}_{(\rho_t)}-\sum_{h=1}^{q}\mathbf{A}_{(\rho_t)}^{(h)}\left(\mathbf{x}_{t-h}-\boldsymbol{\mu}_{(\rho_t)}\right)$$

in the case of $t > q$, where $\mathbf{Y}_t \equiv (\mathbf{x}_1,.., \mathbf{x}_t)$.

The first q data vectors can be taken together. $l(\mathbf{X}_q|\rho_q,\boldsymbol{\lambda})$ can be obtained by exploiting stationarity, i.e.

$$\begin{aligned}\mathbf{X}_t&=\tilde{\boldsymbol{\mu}}+\mathbf{A}(\mathbf{X}_{t-1}-\tilde{\boldsymbol{\mu}})+\tilde{\boldsymbol{\xi}}_t\\&=\tilde{\boldsymbol{\mu}}+\tilde{\boldsymbol{\xi}}_t+\mathbf{A}\tilde{\boldsymbol{\xi}}_{t-1}+\mathbf{A}^2(\mathbf{X}_{t-2}-\tilde{\boldsymbol{\mu}})\\&=\tilde{\boldsymbol{\mu}}+\sum_{\tau=0}^{\infty}\mathbf{A}^{\tau}\tilde{\boldsymbol{\xi}}_{t-\tau},\end{aligned}$$

$\mathrm{E}\mathbf{x}_t = \boldsymbol{\mu}$, and assuming the process is stable[1] within each regime,

$$\begin{aligned}\mathrm{Var}\mathbf{X}_t&=\mathrm{E}(\mathbf{X}_t-\tilde{\boldsymbol{\mu}})(\mathbf{X}_t-\tilde{\boldsymbol{\mu}})^T\\&=\mathrm{E}\left\{\left(\sum_{\tau=0}^{\infty}\mathbf{A}^{\tau}\tilde{\boldsymbol{\xi}}_{t-\tau}\right)\left(\sum_{\tau=0}^{\infty}\tilde{\boldsymbol{\xi}}_{t-\tau}^T(\mathbf{A}^{\tau})^T\right)\right\}\\&=\mathrm{E}\left\{\sum_{\tau=0}^{\infty}\mathbf{A}^{\tau}\tilde{\boldsymbol{\xi}}_{t-\tau}\tilde{\boldsymbol{\xi}}_{t-\tau}^T(\mathbf{A}^{\tau})^T\right\}\\&=\sum_{\tau=0}^{\infty}\mathbf{A}^{\tau}\tilde{\boldsymbol{\Omega}}(\mathbf{A}^{\tau})^T\\&=\mathbf{V}(\boldsymbol{\Omega},\mathbf{A})\\&=\mathbf{V}_{(\rho_t)}.\end{aligned}$$

The unconditional or stationary $mq{\times}mq$ variance-covariance matrix, $\mathbf{V}$, can also be determined from

$$\mathrm{vec}\mathbf{V}=\left(\mathbf{I}_{m^2q^2}-\mathbf{A}\otimes\mathbf{A}\right)^{-1}\mathrm{vec}\tilde{\boldsymbol{\Omega}},$$

as described by Lütkepohl (1991, p21-22). vec is the column stacking operator[2], and $\otimes$ is the (right) kronecker product[3]. The contribution to the likelihood from the first q data vectors is therefore

$$l(\mathbf{X}_q|\rho_q,\boldsymbol{\lambda})=(2\pi)^{-\frac{mq}{2}}\cdot\left|\mathbf{V}\left(\boldsymbol{\Omega}_{(k)},\mathbf{A}_{(k)}\right)^{-1}\right|^{\frac{1}{2}}\cdot\exp\left\{-\tfrac{1}{2}\left(\mathbf{X}_q-\tilde{\boldsymbol{\mu}}_{(k)}\right)^T\mathbf{V}\left(\boldsymbol{\Omega}_{(k)},\mathbf{A}_{(k)}\right)^{-1}\left(\mathbf{X}_q-\tilde{\boldsymbol{\mu}}_{(k)}\right)\right\},$$

where $\rho_q = k$.

In practice, to avoid inversion of an $m^2 q^2 \times m^2 q^2$ matrix, $\mathbf{V}$ could be approximated as a finite sum of the form $\tilde{\Omega} + \sum_{\tau=1}^{\nu} \mathbf{A}^{\tau} \tilde{\Omega} (\mathbf{A}^{\tau})^T$, and then an approximation to $\mathbf{V}^{-1}$ obtained by inverting the approximation to $\mathbf{V}$ (an $mq \times mq$ matrix).

The full likelihood conditional on the regimes is therefore

$$L(\mathbf{Y}|\rho,\lambda) = l(\mathbf{X}_q|\rho_q,\lambda) \cdot \prod_{t=q+1}^{N} l(\mathbf{x}_t|\rho_t,\mathbf{Y}_{t-1},\lambda),$$

where $\mathbf{Y} = \mathbf{Y}_N$ and $\rho \equiv \{\rho_q,.., \rho_N\}$.

The exact or unconditional likelihood of λ ("<u>the</u> likelihood") is obtained by integrating over all possible regimes, i.e.

$$L = L(\mathbf{Y}|\lambda) = l(\mathbf{X}_q|\lambda) \cdot \prod_{t=q+1}^{N} l(\mathbf{x}_t|\mathbf{Y}_{t-1},\lambda).$$

The exact or unconditional maximum likelihood parameter estimate is given by the value of λ that maximises L.

5. Markov Chain Monte Carlo Estimation of the Model

Draws from the joint posterior distribution of the regimes and the parameters given the data, $p(\rho,\lambda|\mathbf{Y})$, can be simulated via the Gibbs Sampler and the Metropolis-Hastings algorithm. The algorithm will involve the repeated generation of variates from their full conditional densities, as follows:

$$\begin{aligned}
\rho^{(c+1)} &\leftarrow \mu^{(c)}, \mathbf{A}^{(c)}, \Omega^{(c)}, \mathbf{P}^{(c)} \\
\mu^{(c+1)} &\leftarrow \rho^{(c+1)}, \mathbf{A}^{(c)}, \Omega^{(c)}, \mathbf{P}^{(c)} \\
\mathbf{A}^{(c+1)} &\leftarrow \rho^{(c+1)}, \mu^{(c+1)}, \Omega^{(c)}, \mathbf{P}^{(c)} \\
\Omega^{(c+1)} &\leftarrow \rho^{(c+1)}, \mu^{(c+1)}, \mathbf{A}^{(c+1)}, \mathbf{P}^{(c)} \\
\mathbf{P}^{(c+1)} &\leftarrow \rho^{(c+1)}, \mu^{(c+1)}, \mathbf{A}^{(c+1)}, \Omega^{(c+1)}.
\end{aligned}$$

In each case, $\mathbf{V}$ will be a function of the $\mathbf{A}$ and the Ω on the right hand side. Under mild regularity conditions, the sequence $\{\rho^{(c+1)}, \lambda^{(c+1)}\} \equiv \{\rho^{(c+1)}, \mu^{(c+1)}, \mathbf{A}^{(c+1)}, \Omega^{(c+1)}, \mathbf{P}^{(c+1)}\}$ will form a Markov chain whose limiting distribution will be $p(\rho,\lambda|\mathbf{Y})$.

5.1 Generating the Regimes

$\rho^{(c+1)} \leftarrow \mu^{(c)}, \mathbf{A}^{(c)}, \Omega^{(c)}, \mathbf{P}^{(c)}$. The regimes can be generated jointly from.

$$p(\rho|\mathbf{Y},\lambda) = p(\rho_N|\mathbf{Y},\lambda) \cdot \prod_{t=q}^{N-1} p(\rho_t|\rho_{t+1},\mathbf{Y}_t,\lambda).$$

The filter probabilities, $p(\rho_N|\mathbf{Y},\lambda)$[4], can be calculated from

$$l(\mathbf{X}_q,\rho_q|\lambda)=l(\mathbf{X}_q|\rho_q,\lambda)\cdot p(\rho_q|\lambda)$$
$$l(\mathbf{X}_q|\lambda)=\sum_{\rho_q=1}^{K}l(\mathbf{X}_q,\rho_q|\lambda)$$
$$p(\rho_q|\mathbf{X}_q,\lambda)=l(\mathbf{X}_q,\rho_q|\lambda)/l(\mathbf{X}_q|\lambda).$$

For $t = q+1,..,N$,

$$l(\mathbf{x}_t,\rho_t,\rho_{t-1}|\mathbf{Y}_{t-1},\lambda)=l(\mathbf{x}_t|\rho_t,\mathbf{Y}_{t-1},\lambda)\cdot p(\rho_t|\rho_{t-1})\cdot p(\rho_{t-1}|\mathbf{Y}_{t-1},\lambda)$$
$$l(\mathbf{x}_t|\mathbf{Y}_{t-1},\lambda)=\sum_{\rho_t=1}^{K}\sum_{\rho_{t-1}=1}^{K}l(\mathbf{x}_t,\rho_t,\rho_{t-1}|\mathbf{Y}_{t-1},\lambda)$$
$$p(\rho_t,\rho_{t-1}|\mathbf{Y}_t,\lambda)=l(\mathbf{x}_t,\rho_t,\rho_{t-1}|\mathbf{Y}_{t-1},\lambda)/l(\mathbf{x}_t|\mathbf{Y}_{t-1},\lambda)$$
$$p(\rho_t|\mathbf{Y}_t,\lambda)=\sum_{\rho_{t-1}=1}^{K}p(\rho_t,\rho_{t-1}|\mathbf{Y}_t,\lambda)$$
$$\forall\rho_t,\ \rho_{t-1}\in\{1,..,K\}.$$

Once the filter probabilities, $p(\rho_N|\mathbf{Y},\lambda)$, have been calculated, a sample can easily be generated from $p(\rho_N|\mathbf{Y},\lambda)$, since it is a discrete density.

The above iterations require the evaluation of the contributions to the conditional likelihood, $l(\mathbf{x}_t|\rho_t,\mathbf{Y}_{t-1},\lambda)$. These will require evaluation of the $m\times m$ determinants of the K Ω^{-1} [5].

To initialise the previous iterations, the K $p(\rho_q|\lambda)$ will be required. They can be derived as the limiting distribution of the Markov chain[6]. Define the $K\times 1$ column vector $\pi \equiv \{p(\rho_t = i|\lambda),\ i = 1,..,K\}$, then $\pi = \mathbf{P}\pi$. π can be estimated by iterating on $\pi^{(n+1)} = \mathbf{P}\pi^{(n)}$ until convergence to the desired level of accuracy. The $p(\rho_q|\lambda)$ are given as the elements of π.

Thus to generate a sample from the joint distribution of ρ we first generate ρ_N from $p(\rho_N|\mathbf{Y},\lambda)$. Then for $t = N-1$ to q, calculate $p(\rho_t|\rho_{t+1},\mathbf{Y}_t,\lambda)$ using the most recently generated value of ρ_{t+1} and the previously calculated filter probabilities, as follows

$$p(\rho_{t+1},\rho_t|\mathbf{Y}_t,\lambda)=p(\rho_{t+1}|\rho_t,\lambda)\cdot p(\rho_t|\mathbf{Y}_t,\lambda)\quad\text{for }\rho_t=1,..,K$$
$$p(\rho_{t+1}|\mathbf{Y}_t,\lambda)=\sum_{\rho_t=1}^{K}p(\rho_{t+1},\rho_t|\mathbf{Y}_t,\lambda)$$
$$p(\rho_t|\rho_{t+1},\mathbf{Y}_t,\lambda)=\frac{p(\rho_{t+1},\rho_t|\mathbf{Y}_t,\lambda)}{p(\rho_{t+1}|\mathbf{Y}_t,\lambda)}\quad\text{for }\rho_t=1,..,K.$$

Once the probabilities, $p(\rho_t|\rho_{t+1},\mathbf{Y}_t,\lambda)$, have been calculated, ρ_t can easily be generated from $p(\rho_t|\rho_{t+1},\mathbf{Y}_t,\lambda)$, since it is a discrete density. For the regime switching process to be defined, each of the K regimes needs to be visited.

5.2 Generating the Parameters

The conditional densities of the parameters are given by

$$p\left(\lambda_j | \rho, \lambda_{\neq j}, \mathbf{Y}\right) \propto L(\mathbf{Y}|\rho,\lambda) \cdot p(\rho|\lambda) \cdot p\left(\lambda_j\right),$$

$$\text{i.e. } p\left(\Theta_j | \rho, \Theta_{\neq j}, \mathbf{P}, \mathbf{Y}\right) \propto L(\mathbf{Y}|\rho,\lambda) \cdot p\left(\Theta_j\right)$$

$$\text{and } p(\mathbf{P}|\rho,\Theta,\mathbf{Y}) \propto p\left(\rho_q|\mathbf{P}\right) \cdot \prod_{t=q+1}^{N} p(\rho_t|\rho_{t-1},\mathbf{P}) \cdot p(\mathbf{P}).$$

5.3 Generating the Level Parameters

$\mu^{(c+1)} \leftarrow \rho^{(c+1)}, \mathbf{A}^{(c)}, \Omega^{(c)}, \mathbf{P}^{(c)}$. In this section, $\mu_{(r)}$ represents one of the possible discrete values of $\mu_{(\rho_t)}$, $\rho_t \in \{1,..,K\}$. Independent uniform priors can be used for the $\mu_{(r)}$, conveying no prior information. The prior would therefore be uniform where the identifiability restrictions (if any) are met, and zero everywhere else. The level parameter vectors can be generated jointly from

$$p\left(\mu_{(1)},..,\mu_{(K)} | \rho, \lambda_{\neq\mu}, \mathbf{Y}\right) \propto L(\mathbf{Y}|\rho,\lambda) \times p\left(\mu_{(1)},..,\mu_{(K)}\right),$$

which is the product of K independent multivariate Normal densities (and the identifiability prior), since the contribution at each time t involves only one of the $\mu_{(r)}$. The exponent in the above expression is

$$-\tfrac{1}{2}\tilde{\xi}_q^T \mathbf{V}_{(\rho_q)}^{-1} \tilde{\xi}_q - \tfrac{1}{2}\sum_{t=q+1}^{N} \xi_t^T \Omega_{(\rho_t)}^{-1} \xi_t$$

$$= -\tfrac{1}{2}\left(\mathbf{X}_q - \tilde{\mu}_{(\rho_q)}\right)^T \mathbf{V}_{(\rho_q)}^{-1}\left(\mathbf{X}_q - \tilde{\mu}_{(\rho_q)}\right)$$
$$-\tfrac{1}{2}\sum_{t=q+1}^{N}\left(\mathbf{x}_t - \sum_{h=1}^{q}\mathbf{A}_{(\rho_t)}^{(h)}\mathbf{x}_{t-h} - \left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(\rho_t)}^{(h)}\right)\mu_{(\rho_t)}\right)^T \Omega_{(\rho_t)}^{-1}\left(\mathbf{x}_t - \sum_{h=1}^{q}\mathbf{A}_{(\rho_t)}^{(h)}\mathbf{x}_{t-h} - \left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(\rho_t)}^{(h)}\right)\mu_{(\rho_t)}\right)$$

$$= -\tfrac{1}{2}\left(\tilde{\mu}_{(\rho_q)} - \mathbf{X}_q\right)^T \mathbf{V}_{(\rho_q)}^{-1}\left(\tilde{\mu}_{(\rho_q)} - \mathbf{X}_q\right)$$
$$-\tfrac{1}{2}\sum_{t=q+1}^{N}\left(\mu_{(\rho_t)} - \left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(\rho_t)}^{(h)}\right)^{-1}\left(\mathbf{x}_t - \sum_{h=1}^{q}\mathbf{A}_{(\rho_t)}^{(h)}\mathbf{x}_{t-h}\right)\right)^T \mathbf{W}_{(\rho_t)}^{-1}\left(\mu_{(\rho_t)} - \left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(\rho_t)}^{(h)}\right)^{-1}\left(\mathbf{x}_t - \sum_{h=1}^{q}\mathbf{A}_{(\rho_t)}^{(h)}\mathbf{x}_{t-h}\right)\right),$$

where $\mathbf{W}_{(\rho_t)}^{-1} = \left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(\rho_t)}^{(h)}\right)^T \Omega_{(\rho_t)}^{-1}\left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(\rho_t)}^{(h)}\right)$.

Suppose $\rho_q = k$ and that n_r of the $\rho_t = r$, then the exponent can be rewritten as

$$-\tfrac{1}{2}\tilde{\xi}_q^T \mathbf{V}_{(k)}^{-1}\tilde{\xi}_q$$

$$-\tfrac{1}{2}\left(\mu_{(k)} - \frac{1}{n_k - 1}\left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(k)}^{(h)}\right)^{-1}\sum_{\substack{\rho_t = k\\ t>q}}\left(\mathbf{x}_t - \sum_{h=1}^{q}\mathbf{A}_{(k)}^{(h)}\mathbf{x}_{t-h}\right)\right)^T (n_k - 1)\mathbf{W}_{(k)}^{-1}\left(\mu_{(k)} - \frac{1}{n_k - 1}\left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(k)}^{(h)}\right)^{-1}\sum_{\substack{\rho_t = k\\ t>q}}\left(\mathbf{x}_t - \sum_{h=1}^{q}\mathbf{A}_{(k)}^{(h)}\mathbf{x}_{t-h}\right)\right)$$

$$-\tfrac{1}{2}\sum_{r\neq k}\left(\mu_{(r)} - \frac{1}{n_r}\left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(r)}^{(h)}\right)^{-1}\sum_{\rho_t = r}\left(\mathbf{x}_t - \sum_{h=1}^{q}\mathbf{A}_{(r)}^{(h)}\mathbf{x}_{t-h}\right)\right)^T n_r\mathbf{W}_{(r)}^{-1}\left(\mu_{(r)} - \frac{1}{n_r}\left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(r)}^{(h)}\right)^{-1}\sum_{\rho_t = r}\left(\mathbf{x}_t - \sum_{h=1}^{q}\mathbf{A}_{(r)}^{(h)}\mathbf{x}_{t-h}\right)\right).$$

Ignoring the first term in $\mathbf{V}^{-1}$, which is a function of $\mu_{(k)}$, the above expression is in the form of K independent vector Normal densities in the $\mu_{(r)}$.

$$\begin{aligned}
&\therefore p\left(\mu_{(1)},\ldots,\mu_{(K)}\mid\rho,\lambda_{\neq\mu},\mathbf{Y}\right)\\
&\quad\propto \left|\mathbf{V}_{(k)}^{-1}\right|^{\frac{1}{2}}\exp\left\{-\tfrac{1}{2}\tilde{\xi}_q^T\mathbf{V}_{(k)}^{-1}\tilde{\xi}_q\right\}\\
&\quad\times \mathrm{N}\left(\frac{1}{n_k - 1}\left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(k)}^{(h)}\right)^{-1}\sum_{\substack{\rho_t = k\\ t>q}}\left(\mathbf{x}_t - \sum_{h=1}^{q}\mathbf{A}_{(k)}^{(h)}\mathbf{x}_{t-h}\right), \frac{1}{n_k - 1}\mathbf{W}_{(k)}\right)\\
&\quad\times \prod_{r\neq k}\mathrm{N}\left(\frac{1}{n_r}\left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(r)}^{(h)}\right)^{-1}\sum_{\rho_t = r}\left(\mathbf{x}_t - \sum_{h=1}^{q}\mathbf{A}_{(r)}^{(h)}\mathbf{x}_{t-h}\right), \frac{1}{n_r}\mathbf{W}_{(r)}\right)\\
&\quad\times p\left(\mu_{(1)},\ldots,\mu_{(K)}\right),
\end{aligned}$$

where N(.,.) is the multivariate or vector Normal density[7]. The K-1 $\mu_{(r)}$, $r \neq k$, can therefore be independently generated from the above multivariate Normal densities. Asymptotically, the means of the above densities, for each regime, are the average of the data vectors in each regime, as expected.

The terms in $\mu_{(k)}$ are not quite vector Normal, since $\mathbf{V}^{-1}$ is also a function of $\mu_{(k)}$. A Metropolis-Hastings step can be used to generate $\mu_{(k)}$. First, a candidate $\mu_{(k)}$ is generated, $\mu_{(k)}^{(*)} \sim \mathrm{N}\left(\frac{1}{n_k - 1}\left(\mathbf{I}_m - \sum_{h=1}^{q}\mathbf{A}_{(k)}^{(h)}\right)^{-1}\sum_{\substack{\rho_t = k\\ t>q}}\left(\mathbf{x}_t - \sum_{h=1}^{q}\mathbf{A}_{(k)}^{(h)}\mathbf{x}_{t-h}\right), \frac{1}{n_k - 1}\mathbf{W}_{(k)}\right)$, and accepted, i.e. $\mu_{(k)}^{(c+1)} = \mu_{(k)}^{(*)}$, with probability

$$\min\left\{\sqrt{\frac{\left|\mathbf{V}_{(k)}^{-1(*)}\right|}{\left|\mathbf{V}_{(k)}^{-1(c)}\right|}}\exp\left\{-\tfrac{1}{2}\tilde{\xi}_q^T\left(\mathbf{V}_{(k)}^{-1(*)} - \mathbf{V}_{(k)}^{-1(c)}\right)\tilde{\xi}_q\right\}, 1\right\},$$

otherwise the value from the previous Gibbs iteration is retained, i.e. $\mu_{(k)}^{(c+1)} = \mu_{(k)}^{(c)}$. Here, $\mathbf{V}_{(k)}^{-1(*)}$ is a function of $\mathbf{A}_{(k)}^{(c)}$ and $\Omega_{(k)}^{-1(c)}$, while $\mathbf{V}_{(k)}^{-1(c)}$ is a function of $\mathbf{A}_{(k)}^{(c-1)}$ and $\Omega_{(k)}^{-1(c-1)}$.

Generation of the K $\mu_{(r)}$ would continue until the prior conditions represented by $p\left(\mu_{(1)},..,\mu_{(K)}\right)$ are satisfied (i.e. by direct acceptance/rejection).

5.4 Generating the Regressive Correlation Parameters

$\mathbf{A}^{(c+1)} \leftarrow \rho^{(c+1)}, \mu^{(c+1)}, \Omega^{(c)}, \mathbf{P}^{(c)}$. In this section, $\mathbf{A}_{(r)}$ represents one of the possible discrete values of $\mathbf{A}_{(\rho_t)}$, $\rho_t \in \{1,..,K\}$. The regressive correlation parameter matrices can be generated jointly from

$$p\left(\mathbf{A}_{(1)},..,\mathbf{A}_{(K)}|\rho,\lambda_{\neq\mathbf{A}},\mathbf{Y}\right) \propto L(\mathbf{Y}|\rho,\lambda)\cdot p\left(\mathbf{A}_{(1)},..,\mathbf{A}_{(K)}\right).$$

Recall that

$$\mathbf{A}_{(\rho_t)} = \begin{pmatrix} \mathbf{A}^{(1)}_{(\rho_t)} & \mathbf{A}^{(2)}_{(\rho_t)} & \cdots & \mathbf{A}^{(q)}_{(\rho_t)} \\ \mathbf{I}_m & \mathbf{0}_m & \cdots & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{I}_m & \cdots & \mathbf{0}_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_m & \mathbf{0}_m & \cdots & \mathbf{0}_m \end{pmatrix},$$

so that only the first m rows need to be generated. Define the $m \times mq$ matrix operator $\vartheta \equiv (\mathbf{I}_m, \mathbf{0}_m, \mathbf{0}_m, .., \mathbf{0}_m)$, so that $\vartheta\mathbf{A} = (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, .., \mathbf{A}^{(q)})$. It is the K $\vartheta\mathbf{A}_{(r)}$ that need to be generated.

A suitable prior for each of the $\vartheta\mathbf{A}_{(r)}$ is a matrix Normal density in the region of stability of the VAR process within each regime, and zero everywhere else. Thus it is assumed a priori that the process is stable in each regime. The prior for $\vartheta\mathbf{A}_{(r)}$ will be represented by $p(\vartheta\mathbf{A}_{(r)}) \propto \mathrm{N}(\mathbf{B}_{(r)}, 1/\nu_{(r)}\ \mathbf{I}_{m^2q}) \times g(\mathbf{A}_{(r)})$, where the first term is a matrix Normal[8] density, $\mathbf{B}_{(r)} = \left(\mathbf{B}^{(1)}_{(r)},\mathbf{B}^{(2)}_{(r)},..,\mathbf{B}^{(q)}_{(r)}\right)$ is the $m \times mq$ prior estimate of $\vartheta\mathbf{A}_{(r)}$, $\mathbf{I}_{m^2q}$ is an $m^2q \times m^2q$ identity matrix, and the last term is uniform in the stable region of $\mathbf{A}_{(r)}$ and zero everywhere else. In the absence of strong prior evidence, each of the $\mathbf{B}^{(h)}_{(r)}$ is likely to be zero everywhere except perhaps instances on the diagonal where serial correlation is clearly present, e.g. inflation rate series. The prior variance of each element of $\vartheta\mathbf{A}_{(r)}$ is $1/\nu_{(r)}$, where $\nu_{(r)}$ can be interpreted as the equivalent number of prior observations in regime r.

Stability requires that $\mathbf{A}^{\tau}$ converge rapidly to zero as $\tau \to \infty$, so that the sequence $\{\mathbf{A}^{\tau}, \tau = 0,1,2,..\}$ is absolutely summable[9], converging to $(\mathbf{I} - \mathbf{A})^{-1}$. This is equivalent to insisting that all eigenvalues of $\mathbf{A}$ have modulus less than one. The latter condition holds if and only if the determinant of $(\mathbf{I} - z\mathbf{A})$ is nonzero for $|z| \leq 1$, i.e. iff $\det(\mathbf{I} - z\mathbf{A}^{(1)} - .. - z^q\mathbf{A}^{(q)}) \neq 0$ on the interval $|z| \leq 1$. For practical implementation it is wise to insist that $\det(\mathbf{I} - z\mathbf{A})$ exceeds a fixed positive constant, say 0.15, for $z = \pm 1$ (noting that it equals 1 for $z = 0$), to control the occurrence of $(\mathbf{I} - \mathbf{A})^{-1}$ becoming large, which can lead to the level estimates visiting unlikely values (since the level parameters are not defined when $(\mathbf{I} - \mathbf{A})$ is not invertible).

Supposing $\rho_q = k$, consider regimes $\rho_t = r$ ($\neq k$), and define the $m \times n_r$ matrix of regime r residual vectors, $\varsigma_{(r)} \equiv (\xi_t : \rho_t = r)$, and matrices of deviation vectors, $\widetilde{\chi}_{(r)} \equiv \left(\left(\mathbf{X}_{t-1} - \widetilde{\mu}_{(r)}\right) : \rho_t = r\right)$ and $\chi_{(r)} \equiv \left(\left(\mathbf{X}_t - \widetilde{\mu}_{(r)}\right) : \rho_t = r\right)$. Noting that $-\widetilde{\xi}_t = \mathbf{A}_{(r)}\left(\mathbf{X}_{t-1} - \widetilde{\mu}_{(r)}\right) - \left(\mathbf{X}_t - \widetilde{\mu}_{(r)}\right)$ from the VAR(1) form, consider the following (dropping the references to regime r, for brevity).

$$
\begin{aligned}
-\varsigma &= \vartheta\mathbf{A}\widetilde{\chi} - \vartheta\chi \\
-\varsigma\widetilde{\chi}^T &= \vartheta\mathbf{A}\widetilde{\chi}\widetilde{\chi}^T - \vartheta\chi\widetilde{\chi}^T \\
-\varsigma\widetilde{\chi}^T &= \left[\vartheta\mathbf{A} - \vartheta\left(\chi\widetilde{\chi}^T\right)\left(\widetilde{\chi}\widetilde{\chi}^T\right)^{-1}\right]\left(\widetilde{\chi}\widetilde{\chi}^T\right) \\
-\varsigma &= \left[\vartheta\mathbf{A} - \vartheta\left(\chi\widetilde{\chi}^T\right)\left(\widetilde{\chi}\widetilde{\chi}^T\right)^{-1}\right]\widetilde{\chi} \\
-\text{vec}\varsigma &= \left(\widetilde{\chi}^T \otimes \mathbf{I}_m\right)\left[\text{vec}\vartheta\mathbf{A} - \text{vec}\left\{\vartheta\left(\chi\widetilde{\chi}^T\right)\left(\widetilde{\chi}\widetilde{\chi}^T\right)^{-1}\right\}\right] \\
&= \left(\widetilde{\chi}^T \otimes \mathbf{I}_m\right)\left[\text{vec}\vartheta\mathbf{A} - \text{vec}\mathbf{C}\right],
\end{aligned}
$$

where $\mathbf{C} = \vartheta\left(\chi\widetilde{\chi}^T\right)\left(\widetilde{\chi}\widetilde{\chi}^T\right)^{-1}$ is an $m \times mq$ sample correlation matrix at lag 1. Note that vec(**AB**) = ($\mathbf{B}^T \otimes \mathbf{I}$) vec(**A**). Lütkepohl (1991, Appendix A.11-A.12) provides a useful summary of the properties of the kronecker product and the vec and trace operators.

The contribution at each time t involves only one of the $\mathbf{A}_{(r)}$. The exponent of the likelihood term can be expressed as

$$
\begin{aligned}
&-\tfrac{1}{2}\widetilde{\xi}_q^T\mathbf{V}_{(\rho_q)}^{-1}\widetilde{\xi}_q - \tfrac{1}{2}\sum_{t=q+1}^{N}\xi_t^T\Omega_{(\rho_t)}^{-1}\xi_t \\
&= -\tfrac{1}{2}\widetilde{\xi}_q^T\mathbf{V}_{(k)}^{-1}\widetilde{\xi}_q - \tfrac{1}{2}\text{vec}\varsigma_{(k)}^{\ T}\left(\mathbf{I}_{n_k-1} \otimes \Omega_{(k)}^{-1}\right)\text{vec}\varsigma_{(k)} - \tfrac{1}{2}\sum_{r \neq k}\text{vec}\varsigma_{(r)}^{\ T}\left(\mathbf{I}_{n_r} \otimes \Omega_{(r)}^{-1}\right)\text{vec}\varsigma_{(r)},
\end{aligned}
$$

where $\varsigma_{(k)} \equiv (\xi_t : \rho_t = k, t > q)$ is the $m \times (n_k - 1)$ matrix of regime k residual vectors, excluding the first ($t = q$). The exponent, including the prior, can therefore be expressed as

$$
\begin{aligned}
&-\tfrac{1}{2}\widetilde{\xi}_q^T\mathbf{V}_{(k)}^{-1}\widetilde{\xi}_q - \tfrac{1}{2}\sum_r\left[\text{vec}\vartheta\mathbf{A}_{(r)} - \text{vec}\mathbf{C}_{(r)}\right]^T\left(\left(\widetilde{\chi}_{(r)}\widetilde{\chi}_{(r)}^T\right) \otimes \Omega_{(r)}^{-1}\right)\left[\text{vec}\vartheta\mathbf{A}_{(r)} - \text{vec}\mathbf{C}_{(r)}\right] \\
&\qquad - \tfrac{1}{2}\sum_r\left[\text{vec}\vartheta\mathbf{A}_{(r)} - \text{vec}\mathbf{B}_{(r)}\right]^T v_{(r)}\mathbf{I}_{m^2q}\left[\text{vec}\vartheta\mathbf{A}_{(r)} - \text{vec}\mathbf{B}_{(r)}\right] \\
&= -\tfrac{1}{2}\widetilde{\xi}_q^T\mathbf{V}_{(k)}^{-1}\widetilde{\xi}_q - \tfrac{1}{2}\sum_r\left[\text{vec}\vartheta\mathbf{A}_{(r)} - \alpha_{(r)}\right]^T\Psi_{(r)}^{-1}\left[\text{vec}\vartheta\mathbf{A}_{(r)} - \alpha_{(r)}\right],
\end{aligned}
$$

where $\widetilde{\chi}_{(k)}$ and $\mathbf{C}_{(k)}$ are defined to exclude the first vector ($t = q$), and

$$\Psi_{(r)}^{-1} = \left(\left(\tilde{\chi}_{(r)}\tilde{\chi}_{(r)}^{T}\right) \otimes \Omega_{(r)}^{-1}\right) + \nu_{(r)}\mathbf{I}_{m^2 q}$$

$$\alpha_{(r)} = \Psi_{(r)}\left[\left(\left(\tilde{\chi}_{(r)}\tilde{\chi}_{(r)}^{T}\right) \otimes \Omega_{(r)}^{-1}\right)\text{vec}\mathbf{C}_{(r)} + \nu_{(r)}\text{vec}\mathbf{B}_{(r)}\right].$$

Note that the term involving $\mathbf{V}^{-1}$ is also a function of $\mathbf{A}_{(k)}$. Excluding the term in $\mathbf{V}^{-1}$, the previous expression is in the form of independent matrix Normal densities in the $\mathbf{A}_{(r)}$.

$$\therefore p\left(\mathbf{A}_{(1)},..,\mathbf{A}_{(K)}|\rho,\lambda_{*\mathbf{A}},\mathbf{Y}\right) \propto \left|\mathbf{V}_{(k)}^{-1}\right|^{\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\tilde{\xi}_q^T\mathbf{V}_{(k)}^{-1}\tilde{\xi}_q\right\} \times \prod_r \mathrm{N}\left(\alpha_{(r)},\Psi_{(r)}\right) \times g\left(\mathbf{A}_{(r)}\right).$$

The means of the above densities for the $\mathbf{A}_{(r)}$ are weighted averages of the sample correlation matrices of the data vectors within each regime and the prior estimates of the $\mathbf{A}_{(r)}$, as expected. The prior variance provides a floor under the inverse of the variance matrix, and hence limits the variance of the $\mathbf{A}_{(r)}$.

Direct acceptance/rejection can be used to independently generate the regression matrices, $\mathbf{A}_{(r)}$, $r \neq \rho_q$. Candidate $\mathbf{A}_{(r)}$ are generated from the matrix Normal densities until they fall within the stable region.

The terms in $\mathbf{A}_{(k)}$ are not quite matrix Normal, since $\mathbf{V}^{-1}$ is also a function of $\mathbf{A}_{(k)}$. A Metropolis-Hastings step can be used to generate $\mathbf{A}_{(k)}$. First, a candidate $\mathbf{A}_{(k)}$ is generated, $\mathbf{A}_{(k)}^{(*)} \sim \mathrm{N}\left(\alpha_{(k)},\Psi_{(k)}\right)$, and if it is stable, accepted, i.e. $\mathbf{A}_{(k)}^{(c+1)} = \mathbf{A}_{(k)}^{(*)}$, with probability

$$\min\left\{\sqrt{\frac{\left|\mathbf{V}_{(k)}^{-1(*)}\right|}{\left|\mathbf{V}_{(k)}^{-1(c)}\right|}}\exp\left\{-\tfrac{1}{2}\tilde{\xi}_q^T\left(\mathbf{V}_{(k)}^{-1(*)} - \mathbf{V}_{(k)}^{-1(c)}\right)\tilde{\xi}_q\right\},1\right\},$$

otherwise the value from the previous Gibbs iteration is retained, i.e. $\mathbf{A}_{(k)}^{(c+1)} = \mathbf{A}_{(k)}^{(c)}$. Here, $\mathbf{V}_{(k)}^{-1(*)}$ is a function of $\mathbf{A}_{(k)}^{(*)}$ and $\Omega_{(k)}^{-1(c)}$, while $\mathbf{V}_{(k)}^{-1(c)}$ is a function of $\mathbf{A}_{(k)}^{(c)}$ and $\Omega_{(k)}^{-1(c-1)}$.

5.5 Generating the Variance and Covariance Parameters

$\Omega^{(c+1)} \leftarrow \rho^{(c+1)}, \mu^{(c+1)}, \mathbf{A}^{(c+1)}, \mathbf{P}^{(c)}$. In this section, $\Omega_{(r)}$ represents one of the possible discrete values of $\Omega_{(\rho_t)}$, $\rho_t \in \{1,..,K\}$. It is more convenient to generate the inverse of the variance-covariance matrices (i.e. to generate the precision matrices), rather than the $\Omega_{(r)}$ directly.

Suitable priors for the $m \times m$ precision matrices would be Wishart densities with parameters $\eta_{(r)}$ and $\mathbf{F}^{-1}$, where $\mathbf{F}$ is diagonal with i-th diagonal element equal to $\eta_{(r)}s_i^2$, where s_i^2 is the prior error variance for the i-th series. The $\eta_{(r)}$ can be interpreted as the equivalent number of prior observations in each regime. The generation of the precision matrices presents no stability problems, so that the prior need only be diffuse, hence $\eta_{(r)}$ is likely to be smaller than $\nu_{(r)}$. The complete prior would therefore be of the form $p\left(\Omega_{(1)}^{-1},..,\Omega_{(K)}^{-1}\right)=\prod_{r=1}^{K}\mathrm{W}_m\left(\eta_{(r)},\mathbf{F}_{(r)}^{-1}\right)\times h\left(\Omega_{(1)},..,\Omega_{(K)}\right)$, where $h\left(\Omega_{(1)},..,\Omega_{(K)}\right)$ captures the identifiability prior restrictions (if any). An example of an identifiability restriction might be that the variance of the second series increases with the regime, i.e. $\omega_{22(1)} < .. < \omega_{22(K)}$, where $\omega_{ii(r)}$ is the i-th diagonal element of $\Omega_{(r)}$. As before, define $\varsigma_{(r)} \equiv \left(\xi_t : \rho_t = r\right)$, $r \neq \rho_q$ and $\varsigma_{(k)} \equiv \left(\xi_t : \rho_t = k, t > q\right)$, $k = \rho_q$ (noting that the ξ_t are functions of the most recently generated $\mathbf{A}_{(r)}$).

The precision matrices can be generated jointly from

$$p\left(\Omega_{(1)}^{-1},..,\Omega_{(K)}^{-1}|\rho,\lambda_{\neq\Omega},\mathbf{Y}\right)\propto L(\mathbf{Y}|\rho,\lambda)\cdot p\left(\Omega_{(1)}^{-1},..,\Omega_{(K)}^{-1}\right)$$

$$\propto \left|\mathbf{V}_{(k)}^{-1}\right|^{\frac{1}{2}}\exp\left\{-\tfrac{1}{2}\tilde{\xi}_q^T\mathbf{V}_{(k)}^{-1}\tilde{\xi}_q\right\} \times \left|\Omega_{(k)}^{-1}\right|^{\frac{n_k-1}{2}}\exp\left\{-\tfrac{1}{2}\operatorname{vec}\varsigma_{(k)}^T\left(\mathbf{I}_{n_k-1}\otimes\Omega_{(k)}^{-1}\right)\operatorname{vec}\varsigma_{(k)}\right\}$$
$$\times\prod_{r\neq k}\left|\Omega_{(r)}^{-1}\right|^{\frac{n_r}{2}}\exp\left\{-\tfrac{1}{2}\operatorname{vec}\varsigma_{(r)}^T\left(\mathbf{I}_{n_r}\otimes\Omega_{(r)}^{-1}\right)\operatorname{vec}\varsigma_{(r)}\right\}\times p\left(\Omega_{(1)}^{-1},..,\Omega_{(K)}^{-1}\right)$$

$$\propto \left|\mathbf{V}_{(k)}^{-1}\right|^{\frac{1}{2}}\exp\left\{-\tfrac{1}{2}\tilde{\xi}_q^T\mathbf{V}_{(k)}^{-1}\tilde{\xi}_q\right\}\times\left|\Omega_{(k)}^{-1}\right|^{\frac{n_k+\eta_{(k)}-m-2}{2}}\exp\left\{-\tfrac{1}{2}\operatorname{tr}\left(\left(\varsigma_{(k)}\varsigma_{(k)}^T+\mathbf{F}_{(k)}\right)\Omega_{(k)}^{-1}\right)\right\}$$
$$\times\prod_{r\neq k}\left|\Omega_{(r)}^{-1}\right|^{\frac{n_r+\eta_{(r)}-m-1}{2}}\exp\left\{-\tfrac{1}{2}\operatorname{tr}\left(\left(\varsigma_{(r)}\varsigma_{(r)}^T+\mathbf{F}_{(r)}\right)\Omega_{(r)}^{-1}\right)\right\}\times h\left(\Omega_{(1)},..,\Omega_{(K)}\right)$$

$$\therefore p\left(\Omega_{(1)}^{-1},..,\Omega_{(K)}^{-1}|\rho,\lambda_{\neq\Omega},\mathbf{Y}\right)$$
$$\propto \left|\mathbf{V}_{(k)}^{-1}\right|^{\frac{1}{2}}\exp\left\{-\tfrac{1}{2}\tilde{\xi}_q^T\mathbf{V}_{(k)}^{-1}\tilde{\xi}_q\right\}\times\mathrm{W}_m\left(n_k+\eta_{(k)}-1,\left(\varsigma_{(k)}\varsigma_{(k)}^T+\mathbf{F}_{(k)}\right)^{-1}\right)$$
$$\times\prod_{r\neq k}\mathrm{W}_m\left(n_r+\eta_{(r)},\left(\varsigma_{(r)}\varsigma_{(r)}^T+\mathbf{F}_{(r)}\right)^{-1}\right)\times h\left(\Omega_{(1)},..,\Omega_{(K)}\right).$$

Therefore the precision matrices, other than $\Omega_{(k)}^{-1}$, can be generated independently from Wishart densities. $\Omega_{(k)}^{-1}$ can be generated via a Metropolis-Hastings step with a Wishart candidate generating density, i.e. generate a candidate $\Omega_{(k)}^{-1}$, $\Omega_{(k)}^{-1(*)}\sim\mathrm{W}_m\left(n_k+\eta_{(k)}-1,\left(\varsigma_{(k)}\varsigma_{(k)}^T+\mathbf{F}_{(k)}\right)^{-1}\right)$, and accept it, i.e. set $\Omega_{(k)}^{-1(c+1)}=\Omega_{(k)}^{-1(*)}$, with probability

$$\min\left\{\sqrt{\frac{\left|\mathbf{V}_{(k)}^{-1(*)}\right|}{\left|\mathbf{V}_{(k)}^{-1(c)}\right|}}\exp\left\{-\tfrac{1}{2}\tilde{\xi}_q^T\left(\mathbf{V}_{(k)}^{-1(*)}-\mathbf{V}_{(k)}^{-1(c)}\right)\tilde{\xi}_q\right\},1\right\},$$

otherwise retain the value from the previous Gibbs iteration, i.e. set $\Omega_{(k)}^{-1(c+1)} = \Omega_{(k)}^{-1(c)}$. Here, $\mathbf{V}_{(k)}^{-1(*)}$ is a function of $\Omega_{(k)}^{-1(*)}$ and $\mathbf{A}_{(k)}^{(c+1)}$, while $\mathbf{V}_{(k)}^{-1(c)}$ is a function of $\Omega_{(k)}^{-1(c)}$ and $\mathbf{A}_{(k)}^{(c)}$.

A variate from the Wishart density, $\mathbf{W} \sim W_m(\eta,\Sigma)$, can be generated as $\mathbf{W} = \mathbf{Q}\mathbf{Q}^T$, where $\mathbf{Q} = \mathbf{L}\mathbf{U}$, $\mathbf{L}$ is lower triangular given by the Choleski decomposition $\Sigma = \mathbf{L}\mathbf{L}^T$, and $\mathbf{U}$ is upper triangular given by the Bartlett decomposition, $u_{ij} = 0$ for $i > j$, $u_{ii}^2 \sim \chi_\eta^2$ $(i = j)$ and $u_{ij} \sim N(0,1)$ for $i < j$ (so that $\mathbf{U}^T\mathbf{U} \sim W_m(\eta,\mathbf{I}_m)$).

5.6 Generating the Transition Probabilities

$\mathbf{P}^{(c+1)} \leftarrow \rho^{(c+1)}$, $\mu^{(c+1)}$, $\mathbf{A}^{(c+1)}$, $\Omega^{(c+1)}$. The transition probability matrix can be generated from

$$p(\mathbf{P}|\rho,\Theta,\mathbf{Y}) \propto p(\rho_q|\mathbf{P})\cdot\prod_{t=q+1}^{N}p(\rho_t|\rho_{t-1},\mathbf{P})\cdot p(\mathbf{P}).$$

Suppose that ρ represents n_{ij} transitions from regime i to regime j. Define the prior for the p_{ij} to be Beta(m_{ij} + 1, m_{ii} + 1), where m_{ij} has the interpretation as the equivalent number of prior transitions, then

$$\begin{aligned}p(\mathbf{P}|\rho,\Theta,\mathbf{Y}) &\propto p(\rho_q|\mathbf{P})\times\prod_{i\neq j}p_{ij}^{n_{ij}+m_{ij}}\times\prod_{i=1}^{K}\left(1-\sum_{i\neq j}p_{ij}\right)^{n_{ii}+m_{ii}}\\ &\propto p(\rho_q|\mathbf{P})\times\prod_{i=1}^{K}\left\{\left(\prod_{i\neq j}p_{ij}^{n_{ij}+m_{ij}}\right)\left(1-\sum_{i\neq j}p_{ij}\right)^{n_{ii}+m_{ii}}\right\}.\end{aligned}$$

In the above expression, $p(\rho_q|\mathbf{P})$ is a function of each of the p_{ij}. Draws from the above joint density can be generated using a Metropolis-Hastings step, using independent Beta densities as the candidate generating densities. Generate candidate $\mathbf{P}$, $\mathbf{P}^{(*)}$, from $p_{ij}^{(*)} \sim$ Beta(n_{ij} + m_{ij} + 1, n_{ii} + m_{ii} + 1) for $i \neq j$, $p_{ii}^{(*)} = 1-\sum_{i\neq j}p_{ij}^{(*)}$, until $p_{ii}^{(*)} > 0$, which is then accepted, i.e. $\mathbf{P}^{(c+1)}$ set equal to $\mathbf{P}^{(*)}$, with probability $\min\left\{\frac{\pi^{(*)}/q^{(*)}}{\pi^{(c)}/q^{(c)}},1\right\}$, where

$$\pi/q = p(\rho_q|\mathbf{P})\left(\frac{p_{ii}}{\prod_{i\neq j}(1-p_{ij})}\right)^{n_{ii}+m_{ii}}$$

otherwise the value from the previous Gibbs iteration is retained, i.e. set $\mathbf{P}^{(c+1)} = \mathbf{P}^{(c)}$. Recall that $p(\rho_q|\mathbf{P})$ is given by iterating on $\mathbf{P}$.

The acceptance rate is high for stable regimes where the p_{ij} are small. The acceptance rate can become very low when a p_{ii} becomes small, since then $p_{ii} = 1 - \Sigma p_{ij} < \Pi(1 - p_{ij})$, and hence their ratio can become very small when raised to the power $n_{ii} + m_{ii}$. However if a p_{ii} is small, perhaps the appropriateness of modelling the regime at all should be questioned.

6. RESULTS

6.1 VALIDATION AGAINST SIMULATED DATA

The estimation procedure was tested against a number of simulated data sets. The mean parameter estimates were found to converge extremely rapidly, even when the initial parameter estimates were very poor and the order of the fitted model was incorrect. The effects of the initial parameter estimates appeared to dissipate after only several iterations/samples. The MCMC procedure can therefore be expected to supply a good estimate of the mean parameter values within seconds, regardless of the initial parameter estimates, in contrast to an EM maximum likelihood approach. The results of one of the simulation tests are briefly reported below.

2000 observations were generated from a bivariate RSVAR(2,2) process. The data generating process was a random noise process within each regime, apart from variable 2 in regime 1, which was generated from an AR(2) process with autoregressive parameters of 0.75 and -0.25, i.e. $x_{t2} = 0.01 + 0.75(x_{t-1,2} - 0.01) - 0.25(x_{t-2,2} - 0.01) + 0.005z_t$, where $z_t \sim$ iid N(0,1).

The MCMC estimation procedure described in section 5 was used to generate 2000 samples from the joint parameter density of the model. The mean parameter estimates are summarised in table 1. The procedure successfully identified the data generating process with very tight densities centred over the true parameter values. The significance or otherwise of the various parameter estimates is beyond doubt. Maximum likelihood estimated (MLE) parameters were taken as the set of parameters corresponding to the sample/iteration with the highest log-likelihood of the 2000 samples/iterations. The mean and MLE parameters were very close, which is to be expected, given the large data set (2000 observations). In the test shown the initial parameter estimates were reasonably close to the true parameters. Other tests demonstrated the robustness of the estimation procedure to various starting values.

Graph 1 compares the mean regime (line) with the true regime (shaded bands) for the first 150 time points. The procedure successfully differentiated between the low and high volatility regimes.

Graph 1: Probability of Being in Regime 2

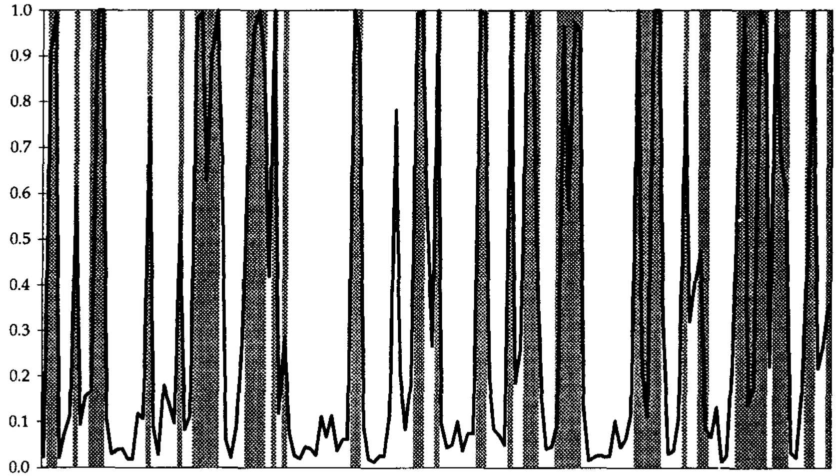


Table 1: Simulation Test Results

	True Value	1st 50 iterates/samples 5%'ile	Mean	95%'ile	2000 iterates/samples 5%'ile	**Mean**	95%'ile
Number in regime 2	**462**	404	451.3	502	424	**463.9**	508
Transition Prob p_{12}	**0.161**	0.131	0.159	0.189	0.138	**0.161**	0.185
Transition Prob p_{21}	**0.500**	0.497	0.548	0.605	0.466	**0.527**	0.589
Regime 1:							
μ_1	**1.00%**	0.91%	0.97%	1.04%	0.93%	**0.98%**	1.03%
$a_{11}^{(1)}$	**0.000**	-0.027	0.005	0.031	-0.039	**0.002**	0.042
$a_{11}^{(2)}$	**0.000**	-0.049	-0.009	0.024	-0.048	**-0.010**	0.029
$\sqrt{\omega_{11}}$	**1.00%**	0.95%	0.98%	1.02%	0.95%	**0.98%**	1.02%
μ_2	**1.00%**	0.93%	0.98%	1.03%	0.93%	**0.98%**	1.03%
$a_{22}^{(1)}$	**0.750**	0.747	0.770	0.798	0.740	**0.767**	0.795
$a_{22}^{(2)}$	**-0.250**	-0.281	-0.261	-0.238	-0.287	**-0.261**	-0.236
$\sqrt{\omega_{22}}$	**0.50%**	0.48%	0.51%	0.53%	0.48%	**0.50%**	0.52%

Table 1: Simulation Test Results (continued)							
	True	1st 50 iterates/samples			2000 iterates/samples		
	Value	5%'ile	Mean	95%'ile	5%'ile	**Mean**	95%'ile
Regime 2:							
μ_1	**0.00%**	-0.08%	0.07%	0.20%	-0.19%	**0.03%**	0.24%
$a_{11}^{(1)}$	**0.000**	-0.073	0.015	0.104	-0.074	**0.025**	0.120
$a_{11}^{(2)}$	**0.000**	-0.075	0.034	0.131	-0.072	**0.035**	0.145
$\sqrt{\omega_{11}}$	**1.50%**	1.47%	1.58%	1.65%	1.48%	**1.57%**	1.67%
μ_2	**2.50%**	2.27%	2.43%	2.54%	2.26%	**2.46%**	2.69%
$a_{22}^{(1)}$	**0.000**	-0.068	0.066	0.171	-0.034	**0.077**	0.187
$a_{22}^{(2)}$	**0.000**	-0.112	0.033	0.143	-0.071	**0.045**	0.157
$\sqrt{\omega_{22}}$	**1.50%**	1.49%	1.57%	1.67%	1.45%	**1.54%**	1.63%

6.2 Empirical Estimation Results

The data set considered, derived from the Reserve Bank of Australia database, consisted of 147 quarterly observations, for the quarters ending December 1959 through to June 1996, of the continuously compounded rates of

- real economic growth;
- change in the rate of price inflation;
- share price return; and
- change in the 10 year bond yield.

More precisely, the data series examined were

- $\nabla \ln GDP_t$ where GDP_t is the real Gross Domestic Product for the quarter ending time t;
- $\nabla^2 \ln CPI_t$ where CPI_t is the Consumer Price Index at time t;
- $\nabla \ln SPI_t$ where SPI_t is the All Ordinaries Share Price Index at time t; and
- $\nabla \ln B_t$ where B_t is the yield to maturity on 10 year Commonwealth Government bonds,

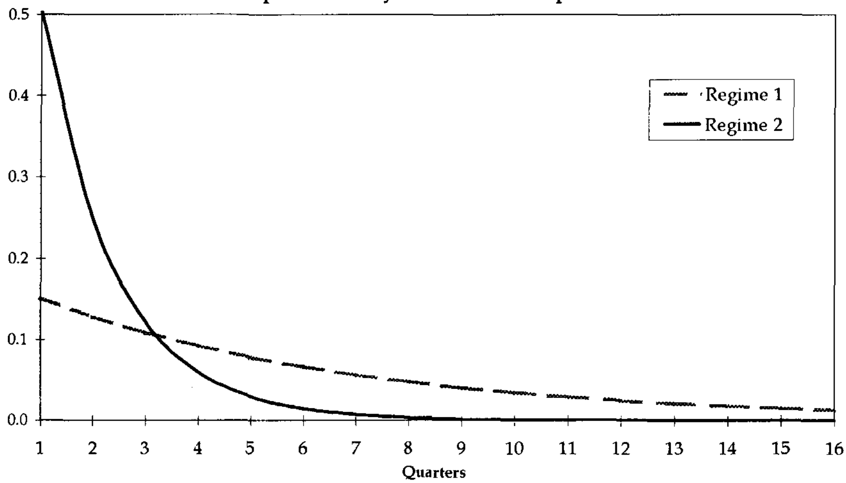
where ∇ is the backward difference operator.

The results reported in this section relate to the fitting of an RSVAR(1,2) process to the above Australian financial data ($m = 4$, $q = 1$, $K = 2$, $N = 147$). 5,000 iterations/samples were generated using the MCMC estimation procedure described in section 5. The first 50 samples were discarded and the remaining 4,950 samples used to describe the joint parameter density.

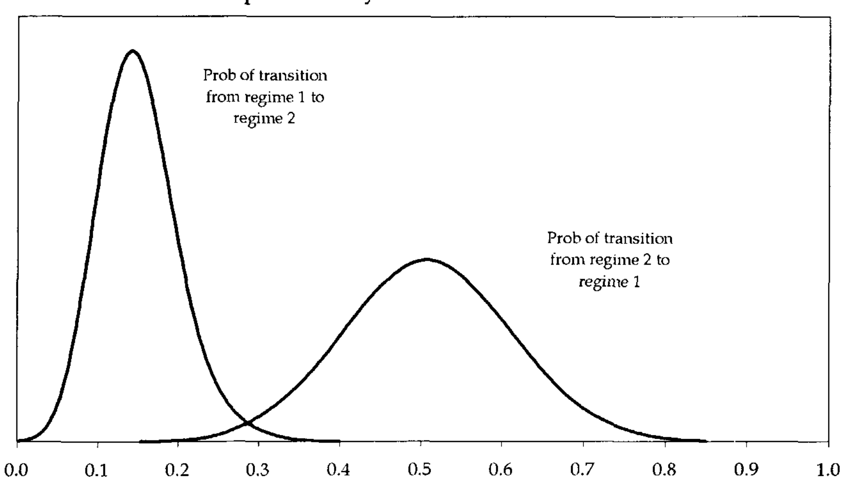
The estimation procedure identified two clearly distinct regimes. One regime (regime 1) was characterised by stable inflation and interest rates, and relatively stable share price growth.

The other regime (regime 2) was characterised by volatile inflation and interest rates, and volatile and generally falling share prices. The low volatility regime is relatively stable in the sense that it is the more persistent of the two regimes. Given the estimated mean transition probability of 0.15, the expected duration of a regime 1 episode is about 6½ quarters[10]. The high volatility regime is unstable in the sense that it is not expected to persist for long. Given its estimated mean transition probability of 0.51, the expected duration of a regime 2 episode is only 2 quarters. The identified regimes seem highly intuitive.

Graph 2: Density of Duration of Episodes



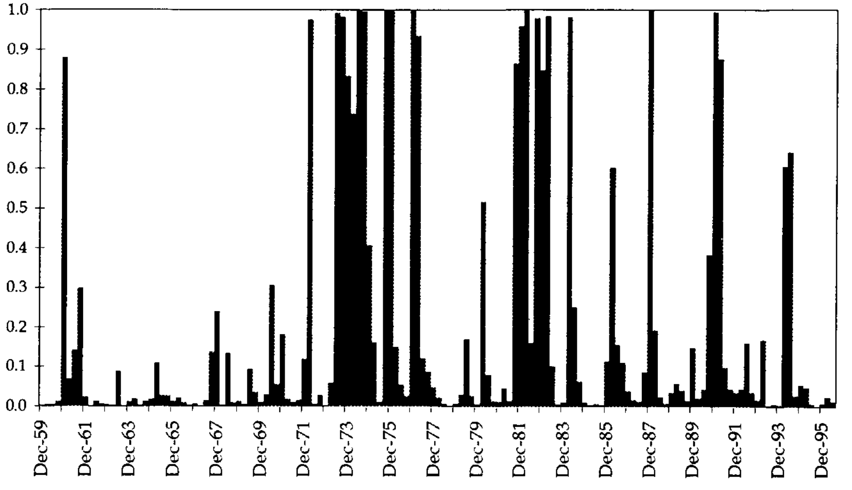
Graph 3: Density of Transition Probabilities



The theoretical density of the duration of an episode of each regime, given the mean transition probabilities[11], is shown in graph 2. The uncertainty in the estimated transition probabilities is illustrated by graph 3.

The mean regime at each time point is shown in graph 4. The economic environment was identified as almost certainly being in the unstable regime during the turbulence of the mid-1970s and early 1980s, the last quarter of 1987 stock market crash, and briefly during the early 1990s. There is a slightly better than average chance that mid-1994 also witnessed a regime shift. Again, the identified regimes seem highly intuitive.

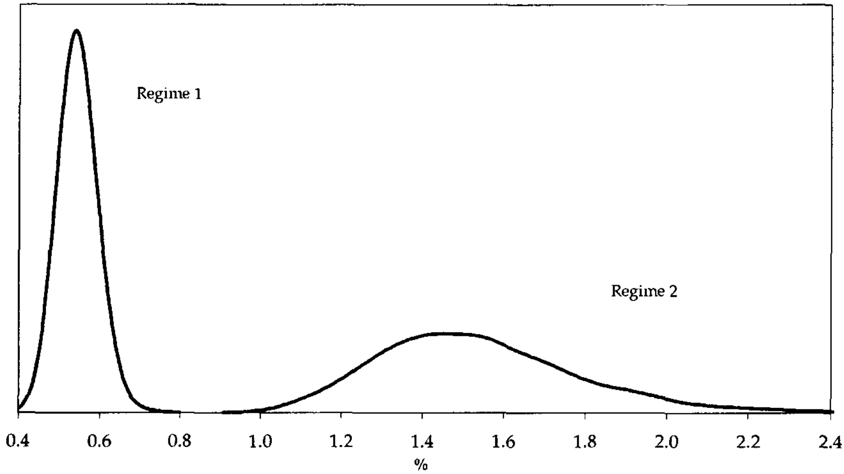
Graph 4: Probability of Being in Regime 2



It will be noticed that the unstable high volatility regime captures the extreme events that might otherwise be termed outliers. Outliers have the potential to seriously distort the estimation of process dynamics. Regime shifting can therefore be viewed as providing a robust data driven treatment of outliers in this case, which should enable more robust parameter estimates.

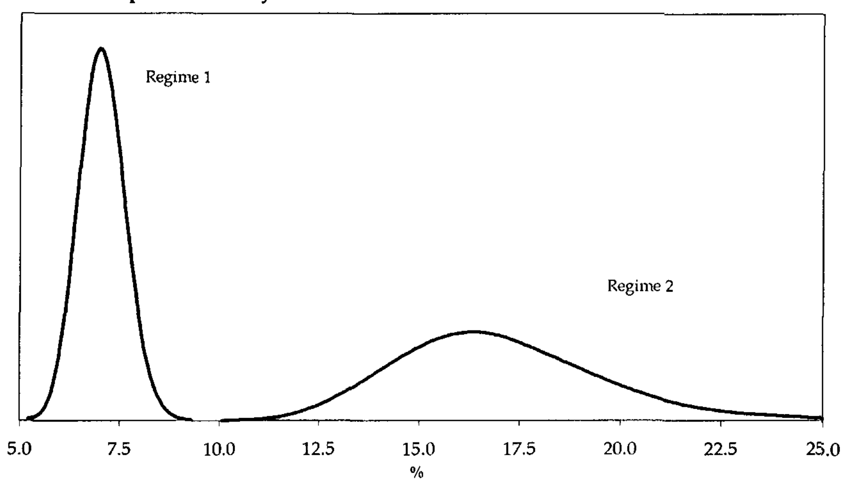
The clear distinction between the parameters in each regime is illustrated by the clear separation of a number of the parameter densities, particularly those relating to the volatility of share price returns and changes in inflation and interest rates, and to the level of share price returns (refer to table 2 and graphs 5a, 5b & 5c). Given the clear separation of the variance of the variables in the two regimes, any procedure that ignores the regime shifts is very likely to produce misleading volatility estimates, and is unlikely to be robust to outliers.

Graph 5a: Density of Std Dev of Change in Inflation Rate Parameter



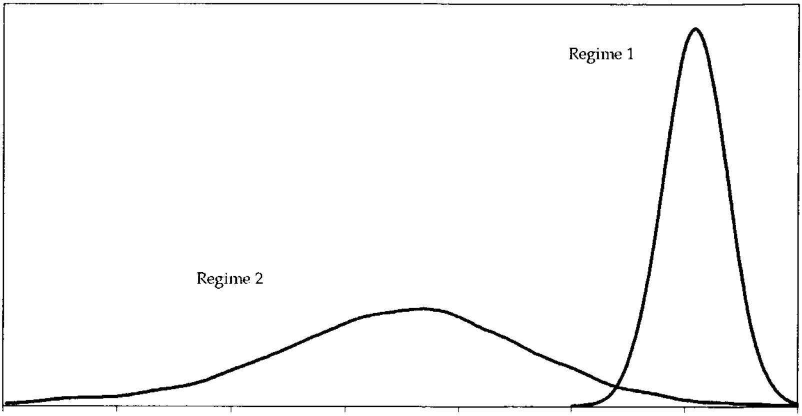
In the case of the quarterly change in inflation, the mean standard deviation parameter estimate in regime 2 is 3 times as large as in regime 1 (1.55% verses 0.55%). The MLE regime 2 parameter is slightly lower at 1.3%.

Graph 5b: Density of Std Deviation of Share Price Return Parameter



In the case of the quarterly share price return, the mean standard deviation parameter estimate in regime 2 is 2½ times as large as in regime 1 (17% verses 7%). The MLE regime 2 parameter is somewhat lower at 13½%.

Graph 5c: Density of Mean Share Price Return Parameter



In the case of the quarterly share price return, the mean level parameter estimate in regime 1 is 3.8% compared with -4.6% in regime 2. The corresponding MLE parameters are 3.7% and -5.1%.

Table 2: Parameter Estimates

	Regime 1			Regime 2		
	5%'ile	**Mean**	95%'ile	5%'ile	**Mean**	95%'ile
Transition Prob p_{ij}	0.089	**0.151**	0.231	0.351	**0.509**	0.668
Mean Parameters:						
μ_1	0.98%	**1.14%**	1.31%	-0.82%	**0.13%**	1.05%
μ_2	-0.07%	**0.00%**	0.07%	-0.39%	**0.11%**	0.65%
μ_3	1.99%	**3.28%**	4.55%	-10.82%	**-4.63%**	0.86%
μ_4	-1.18%	**0.22%**	0.81%	-1.69%	**2.67%**	7.51%
Std Dev Parameters:						
$\sqrt{\omega_{11}}$	1.03%	**1.17%**	1.32%	1.18%	**1.50%**	1.91%
$\sqrt{\omega_{22}}$	0.48%	**0.55%**	0.62%	1.19%	**1.55%**	2.01%
$\sqrt{\omega_{33}}$	6.24%	**7.06%**	7.96%	13.45%	**17.05%**	21.63%
$\sqrt{\omega_{44}}$	3.71%	**4.53%**	5.37%	7.82%	**10.02%**	12.91%

variable 1 = $\nabla \ln GDP_t$, variable 2 = $\nabla^2 \ln CPI_t$, variable 3 = $\nabla \ln SPI_t$, variable 4 = $\nabla \ln B_t$

mean $A_{(1)}$			
-0.178	0.078	0.021	-0.009
0.049	**-0.332**	0.000	0.006
-0.156	0.078	0.072	-0.049
-0.052	0.086	0.048	**0.222**

mean $A_{(2)}$			
0.137	-0.021	-0.009	0.084
-0.108	**-0.449**	0.019	0.080
0.033	-0.030	-0.111	-0.125
0.054	0.008	0.098	0.160

mean contemporaneous error correlations implied by $\Omega_{(1)}$			
1			
0.148	1		
-0.044	-0.053	1	
0.087	-0.084	**-0.195**	1

mean contemporaneous error correlations implied by $\Omega_{(2)}$			
1			
-0.098	1		
-0.337	0.004	1	
0.132	-0.120	-0.181	1

variable 1 = $\nabla \ln GDP_t$, variable 2 = $\nabla^2 \ln CPI_t$, variable 3 = $\nabla \ln SPI_t$, variable 4 = $\nabla \ln B_t$

Only a few regressive and contemporaneous error correlation parameters were found to be significant. The notable regression parameters were serial correlation in the inflation rate in both regimes, and serial correlation in real GDP and the change in interest rates in the stable regime. Interestingly, no cross-correlations appear important in the dynamics once joint regime switching is allowed for.

Ignoring regime shifts would expose estimates of regression parameters to the effects of "outliers" generated during episodes of the high volatility regime, the effects of which would then be assumed to operate at all times. It is interesting to compare the estimated regression parameters of the RSVAR(1,2) process (above) with the corresponding VAR(1) parameters (below).

A from VAR(1)			
-0.10	0.06	0.02	0.03
0.01	**-0.51**	0.00	0.03
0.15	**-0.74**	-0.03	**-0.28**
0.06	**0.56**	0.08	**0.27**

Notable differences are the large feedback of lagged changes in inflation into share price returns and changes in interest rates, and the feedback of lagged changes in interest rates into share price returns. If the feedback were as strong as indicated by the VAR model, the VAR model should have a significantly higher likelihood than an independent AR(1) model (where the off-diagonal elements of **A** are all zero) and one ought to be able to make better predictions than models without the feedback. Neither of these features is observed empirically (refer section 6.3).

It was also noticable that the MLE regression parameters could diverge considerably from the mean of the parameter density when the density was diffuse, so that parameter uncertainty was poorly represented by the usual asymptotic MLE errors.

The notable contemporaneous error correlations were between real GDP growth and changes in inflation in the stable regime (+ve), changes in interest rates and share price returns in the stable regime (-ve), and between real GDP growth and share price returns in the unstable regime (-ve).

6.3 Empirical Comparison With Common Models

In this section the statistical goodness-of-fit of the vector regime switching model is compared with commonly used models. The models considered were independent random/noise, independent autoregressive, independent non-Gaussian autoregressive, independent GARCH[12], Vector Autoregression and RSVAR. The results are summarised in table 3.

Non-Gaussian error distributions are sometimes used in an attempt to directly capture the leptokurtosis observed in the frequency distribution of many series. Two densities were tried as alternatives to the standard Normal density, the Student *t* density and the Generalised Error or Exponential Power Density (GED)[13] (both standardised). As the GED provided the better fit to the data, only the GED results are reported.

The models were compared in terms of their likelihood and prediction errors, their ability to predict volatility, and their ability to explain the observed excess kurtosis (a measure of non-Normality).

The log-likelihoods, both unconditional and conditional on the first data point, are reported in table 3. Where one model is completely nested within another, the increase in the log-likelihood is asymptotically distributed as $\frac{1}{2}\chi_p^2$ where p is the number of additional parameters fitted in the more general of the two models. Thus the AR(1) model is significantly more likely than the Random model and both the GED-AR(1) and the GARCH-AR(1) models are significantly more likely than the AR(1) model. The introduction of the non-Normal error density (GED) produced a substantial increase in the likelihood with the addition of only 4 parameters. It is important to note however that neither the GED nor GARCH models were able to produce lower prediction errors than the AR(1) model. The VAR model is *not* significantly more likely than the AR(1) model.

Since the transition probabilities are not defined under the null hypothesis that the regime switching model is inappropriate, the usual asymptotic statistical distribution theory fails to apply in this case. If it did, the RSVAR(1,2) model would be extremely more likely than either the AR(1) or VAR(1) models. Though not a statistical test, it is at least reassuring that there is a large increase in the log-likelihood, even after allowing for the larger number of parameters. The addition of the second lag in the RSVAR(2,2) model produced only a very modest increase in the log-likelihood (p-value of 0.38). The addition of a further regime (K = 3) proved problematic, due to the degree of instability of the third regime in iterations where $p_{33} \approx 0$. A third regime would appear to be superfluous given its virtual unidentifiability.

The average prediction or forecast errors for each model were assessed using the root-mean-square, or standard, error, which for series *i* was defined as $rmse_i = \sqrt{\frac{1}{N-q}\sum \varepsilon_t^2}$, where ε_t is the residual or one-period-ahead prediction error at time *t*. The *rms errors* for each series were combined into a single *weighted rms error* for each model for ease of comparison. The weights used were proportional to the reciprocals of the corresponding AR(1) residual variances, i.e. $wrms\ error = \sqrt{w_i \times rmse_i^2}$. As noted earlier, both the GED-AR(1) and GARCH(1,1)-AR(1) models produced forecasts no better than the simpler AR(1) model on average. The RSVAR models however produced the smallest errors on average by a considerable margin.

Two measures were used to assess the ability to predict volatility. The first measure used was the root-mean-square absolute error, defined for series *i* as $rmsae_i = \sqrt{\frac{1}{N-q}\sum \left(\left|\varepsilon_t\right| - \sigma_t\right)^2}$, where σ_t is the one-period-ahead predicted error standard deviation according to the model. The *rms absolute errors* were also combined into a single *weighted rms absolute error* using the same weights as used for the *wrms error*. The second measure used was the root-mean-square log-absolute error, defined for series *i* as $rmslae_i = \sqrt{\frac{1}{N-q}\sum \left(\ln\left|z_t\right|\right)^2} = \sqrt{\frac{1}{N-q}\sum \left(\ln\left|\varepsilon_t\right| - \ln\sigma_t\right)^2}$, where $z_t = \varepsilon_t / \sigma_t$.

The GED model produced particularly poor forecasts of volatility. The RSVAR models produced considerably better predictions of volatility on the whole, and notably better than the GARCH model, which also explicitly models conditional heteroscedasticity. Regime switching would appear to be a better explanation of conditional heteroscedasticity than the commonly used GARCH and ARCH processes, which generally impute too much persistency in the volatility (see, for example, Hamilton & Susmel (1994)).

The frequency distribution of financial series typically display excess kurtosis, i.e. are fat-tailed and peaked at the mean (e.g. Mandelbrot (1963), Praetz (1969), Akgiray (1989), Peters (1991), Becker (1991) and Harris (1994, 1995a)). The excess kurtosis of the residuals of each series was calculated, and the average reported in table 3. Autoregressive, VAR and GARCH models failed to explain the observed excess kurtosis. The RSVAR models were able to successfully account for the excess kurtosis in terms of regime switching in the variance, i.e. conditional heteroscedasticity. The only other model to account for the excess kurtosis was the GED model, which explicitly models excess kurtosis by assuming the residuals are drawn from a non-Normal distribution, without explaining the mechanism leading to the non-Normality.

Table 3: Model Comparison

	Random	AR(1)	AR(4)	GED-AR(1)	GARCH (1,1)-AR(1)	VAR (1)	**RS VAR (1,2)**	RS VAR (2,2)
unconditional lnL	1231.5						**1326.5**	1343.4
lnL cond on t=1	1221.5	1249.0		1306.2	1287.4	1258.4	**1315.3**	
ΔlnL over AR(1)		0.0		57.2	38.3	9.4	**66.3**	
standard χ^2 p-value				$<10^{-6}$	$<10^{-6}$	0.1	$\mathbf{<10^{-6}}$	
wrms error	1.48%	1.40%	1.37%	1.42%	1.41%	1.38%	**1.32%**	1.26%
wrmse as % Random	100%	95.0%	92.5%	96.2%	95.6%	93.5%	**89.4%**	85.4%
wrms absolute error	1.18%	1.08%	1.04%	1.15%	1.05%	1.06%	**0.88%**	0.83%
wrmsae as % Random	100%	91.4%	88.1%	97.3%	88.7%	89.4%	**74.4%**	69.9%
av rms log abs error	1.98	1.75	1.78	2.36	1.63	1.55	**1.42**	1.29
av rmslae as % Random	100%	88.2%	80.6%	119.2%	82.3%	77.9%	**71.5%**	65.2%
ave excess kurtosis	4.6	4.0	3.9	4.5	2.1	3.9	**0.3**	0.4

7. CONCLUSIONS

A Bayesian Markov Chain Monte Carlo procedure was developed for estimating the joint parameter and regime density for a regime switching vector autoregression given the observed data. The mean parameter estimates were found to converge extremely rapidly, even when the initial parameter estimates were very poor. The MCMC procedure can therefore be expected to supply a good estimate of the mean parameter values within seconds, regardless of the initial parameter estimates, in contrast to an EM maximum likelihood approach.

The estimation procedure identified two clearly distinct regimes in quarterly Australian financial data. One regime was characterised by stable inflation and interest rates, and relatively stable share price growth. The other regime was characterised by volatile inflation and interest rates, and volatile and generally falling share prices. The high volatility regime was found to be unstable, with an expected duration of only 6 months.

The unstable high volatility regime captured extreme events that might otherwise be termed outliers. Outliers have the potential to seriously distort the estimation of process dynamics. Regime shifting can therefore be viewed as providing a robust data driven treatment of outliers in this case, which should enable more robust parameter estimates.

No cross-correlations appeared important in the dynamics once joint regime switching was allowed for, in contrast to the large cross-correlation terms observed when a standard vector autoregression was fitted to the data. If the feedback were as strong as indicated by the VAR model, the VAR model should have a significantly higher likelihood than an independent autoregressive model (where there are no cross-correlation terms) and one ought to be able to make better predictions than models without the feedback. Neither of these features was observed empirically.

Whilst the GARCH and non-Normal Generalised Error Distribution models were able to produce significant increases in the likelihood over the simpler autoregressive model, it was noted that neither model was able to produce lower average prediction errors than the simple AR model. The RSVAR models produced the lowest average prediction errors by a considerable margin, however.

The RSVAR models produced considerably better predictions of volatility on the whole, and notably better than the GED and GARCH models. Regime switching would appear to be a better explanation of conditional heteroscedasticity than the commonly used GARCH and ARCH processes.

Autoregressive, VAR and GARCH models failed to explain the fat tails observed in the frequency distribution of the data. The RSVAR models were able to successfully account for the excess kurtosis in terms of regime switching in the variance, i.e. conditional heteroscedasticity. The only other model to account for the excess kurtosis was the GED model, which explicitly models excess kurtosis by assuming the residuals are drawn from a non-Normal distribution, without explaining the mechanism leading to the non-Normality.

In conclusion, many financial time series processes appear subject to periodic structural changes in their dynamics. Regression relationships are often not robust to outliers nor stable over time, whilst the existence of changes in variance over time is well known. This paper presented an attempt to deal with such difficulties in financial time series, a Regime Switching Vector Autoregression, the parameters of which are subject to pseudocyclical discrete changes. The regime switching vector autoregression model was found to provide a particularly good description of an Australian quarterly financial data set.

END NOTES

[1] Stability requires that $\mathbf{A}^\tau$ converge rapidly to zero as $\tau \to \infty$ so that the partial sum $\sum_{\tau=0}^{n} \mathbf{A}^\tau$ converges rapidly to $(\mathbf{I} - \mathbf{A})^{-1}$ as $\tau \to \infty$. A stable process is also a stationary process.

[2] If $\mathbf{A} \equiv (a_1,..,a_n)$ is an $m \times n$ matrix with $m \times 1$ columns a_i, then vec$\mathbf{A}$ is the $mn \times 1$ column vector $(a_1^T,.., a_n^T)^T$.

[3] If $\mathbf{A}$ and $\mathbf{B}$ are two matrices, $m \times n$ and $p \times q$ respectively, then $\mathbf{A} \otimes \mathbf{B}$ is the $mp \times nq$ matrix $(a_{ij}\mathbf{B})$.

[4] $p(\rho_t|\mathbf{Y}_t,\lambda)$ is a K-tuple of probabilities, representing $p(\rho_t = i\,|\mathbf{Y}_t,\lambda)$ for $i = 1,..,K$.

[5] using the Choleski decomposition, $|\Omega^{-1}| = |\mathbf{L}||\mathbf{L}^T| = \Pi l_{ii}^2$.

[6] $\mathbf{P}$ is a stochastic matrix since all column sums equal one and all elements are non-negative, hence 1 is an eigenvalue of $\mathbf{P}$, and all eigenvalues of $\mathbf{P}$ have magnitude no greater than one. A necessary and sufficient condition for the existence of a limiting distribution is that $\mathbf{P}$ has a distinct non-repeated unit eigenvalue.

[7] A variate from the multivariate or vector Normal density, $\mathbf{x} \sim N(\mu,\Omega)$, can be generated as $\mathbf{x} = \mu + \mathbf{Lz}$, where $\mathbf{z}$ is a vector of i.i.d. N(0,1) variates, and $\mathbf{L}$ is a lower triangular matrix obtained from the Choleski decomposition of Ω, such that $\mathbf{LL}^T = \Omega$.

[8] Generation from a matrix Normal density is the same as from a vector Normal density, since $\mathbf{A} \sim N(\Theta,\Sigma) \equiv$ vec$\mathbf{A} \sim N(\text{vec}\Theta,\Sigma)$. If $\mathbf{A}$ is $m \times m$, then Θ is also $m \times m$, while Σ is $m^2 \times m^2$, and vec$\mathbf{A}$ and vecΘ are $m^2 \times 1$.

[9] since $\mathbf{X}_t = \mu + \mathbf{A}(\mathbf{X}_{t-1} - \mu) + \xi_t = (\mathbf{I} - \mathbf{A})\mu + \mathbf{AX}_{t-1} + \xi_t = (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + .. + \mathbf{A}^j)(\mathbf{I} - \mathbf{A})\mu + \mathbf{A}^{j+1}\mathbf{X}_{t-j-1} + \sum_{i=0}^{j} \mathbf{A}^i \xi_{t-i}$.

[10] The duration of a regime i episode is a discrete random variable, with expected value given by $\sum_{k=1}^{\infty} k \times p_{ii}^{k-1} \times (1 - p_{ii}) = 1/(1 - p_{ii})$.

[11] p(duration of regime $i = k$) = $p_{ii}^{k-1}(1 - p_{ii})$

[12] The Generalised ARCH model introduced by Bollerslev (1986). The conditional variance is modelled as a linear combination of lagged squared residuals and variances. For example, the commonly used GARCH(1,1) conditional variance is such that $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

[13] The Generalised Error or Exponential Power density is a generalisation of the Normal density where the exponential is raised to a general power. The density function is $f(x) = \frac{\nu}{2} \cdot \sqrt{\frac{\Gamma(\frac{3}{\nu})}{\Gamma(\frac{1}{\nu})^3}} \cdot \frac{1}{\sigma_t} \cdot e^{-\left(\sqrt{\frac{\Gamma(\frac{3}{\nu})}{\Gamma(\frac{1}{\nu})}}|z_t|\right)^\nu}$, where $z = (x - \mu)/\sigma$. When $\nu = 2$ the density is Normal, when $\nu < 2$ it is leptokurtic, and when $\nu = 1$ it is Double-Exponential.

REFERENCES

AKGIRAY, V. (1989). Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts. *Journal of Business* 62, 55-80.

ALBERT, J. H. AND CHIB, S. (1993). Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts. *Journal of Business and Economic Statistics* 11, 1, 1-15.

BARNETT G., KOHN R. AND SHEATHER, S. (1996). Bayesian Estimation of an Autoregressive Model Using Markov Chain Monte Carlo. *Journal of Econometrics* 74 (October 1996), 237-254.

BECKER, D. (1991). Statistical Tests of the Lognormal Distribution as a Basis for Interest Rate Changes. *Transactions of the Society of Actuaries* XLIII.

BOLLERSLEV, T (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31 (June 1986), 307-327.

CASELLA, G. AND GEORGE, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician* 46, No.3, 167-174.

CHIB, S. AND GREENBERG, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician* 49, No.4, 327-335.

ENGLE, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica* 50, 987-1008.

FREES, E. W., KUNG, Y-C., ROSENBERG, M. A., YOUNG, V. R. AND LAI, S-W. (1996). Forecasting Social Security Actuarial Assumptions. Submitted to the *North American Actuarial Journal*.

GELFAND, A. AND SMITH, A. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* 85, 398-409.

GRAY, S. F. (1996). Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process. *Journal of Financial Economics* 42, 27-62.

HAMILTON, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57, 357-384.

HAMILTON, J. D. (1990). Analysis of Time Series Subject to Changes in Regime. *Journal of Econometrics*, 45, 39-70.

HAMILTON, J. D. AND SUSMEL, R. (1994). Autoregressive Conditional Heteroskedasticity and Changes in Regime. *Journal of Econometrics*, 64, 307-333.

HAMILTON, J. D. AND LIN, G. (1996). Stock Market Volatility and the Business Cycle. *Journal of Applied Econonmetrics*, 11, 573-593.

HARRIS, G. R. (1994). On Australian Stochastic Share Return Models for Actuarial Use. *The Institute of Actuaries of Australia Quarterly Journal*, September 1994, 34-54.

HARRIS, G. R. (1995a). Statistical Data Analysis and Stochastic Asset Model Validation. *Transactions of the 25th International Congress of Actuaries* 3, 313-331.

HARRIS, G. R. (1995b). Low Frequency Statistical Interest Rate Models. *Proceedings of the 5th AFIR International Colloquium* 2, 799-831.

HARRIS, G. R. (1995c). A Comparison of Stochastic Asset Models for Long Term Studies. *The Institute of Actuaries of Australia Quarterly Journal*, September 1995, 43-75.

HARRIS, G. R. (1996). Market Phases and Cycles? A Regime Switching Approach. *The Institute of Actuaries of Australia Quarterly Journal*, December 1996, Part 2, 28-44.

LÜTKEPOHL, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag.

MANDELBROT, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business* 36 (October 1963), 394-419.

MCNEES, S. S. (1979). The Forecasting Record for the 1970's. *New England Economic Review* September/October 1979, 33-53.

PRAETZ, P. D. (1969). Australian Share Prices and the Random Walk Hypothesis. *Australian Journal of Statistics* 11, no. 3 (1969), 123-139.

SHERRIS, M., TEDESCO, L. AND ZEHNWIRTH, B. (1996). Stochastic Investment Models: Unit Roots, Cointegration, State Space and GARCH Models for Australian Data. Working Paper, Macquarie University and Melbourne University.

SMITH, A. AND ROBERTS, G. (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society* B 55, 3-23.

STOCK, J. H. AND WATSON, M. W. (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business and Economic Statistics* 14, 1, 11-30.