

MARKOV CHAIN MONTE CARLO ESTIMATION OF REGIME SWITCHING VECTOR AUTOREGRESSIONS

BY

GLEN R HARRIS

Lend Lease Investment Management, Sydney

ABSTRACT

Financial time series data are typically found to possess leptokurtic frequency distributions, time varying volatilities, outliers and correlation structures inconsistent with linear generating processes, nonlinear dependence, and dependencies between series that are not stable over time. Regime Switching Vector Autoregressions are of interest because they are capable of explaining the observed features of the data, can capture a variety of interactions between series, appear intuitively reasonable, are vector processes, and are now tractable.

This paper considers a vector autoregression subject to periodic 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 determined in addition to the vector autoregression parameters within each regime.

A Bayesian Markov Chain Monte Carlo estimation procedure is developed which efficiently generates the posterior joint density of the parameters and the regimes. The complete likelihood surface is generated at the same time, enabling estimation of posterior model probabilities for use in non-nested model selection. The procedure can readily be extended to produce joint prediction densities for the variables, incorporating both parameter and model uncertainty.

Results using simulated and real data are provided. A clear separation of the variance between a stable and an unstable regime was 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 Regime Switching Vector Autoregressions provide a particularly good description of the observed data.

KEYWORDS

Vector regime switching; joint parameter density; outliers, Gibbs sampler; Metropolis-Hastings algorithm, Markov Chain Monte Carlo; posterior model probabilities, model selection.

1 INTRODUCTION

An important feature of financial data is the presence of short periods of instability characterised by large magnitude changes. Such extreme values are not consistent with common linear time series processes in either their magnitude or frequency of occurrence, so that they are often referred to as data "outliers", and often effectively ignored.

The chances that the observed "outliers" were generated from linear models is overwhelmingly small, e.g. consider the quarterly and monthly Australian share price return series over the period 1960-96. The largest magnitude outlier in each case corresponds to the October 1987 sharemarket crash. The chances of observing a return as extreme as observed in the samples from an i.i.d. Normal generating process is less than 1-in-a-million for quarterly data, and indistinguishable from zero (5×10^{-20}) for monthly data. Similarly, the quarterly rate of change in bond yields over the same period produced twice as many "outliers" (residuals in excess of 2 standard deviations) as would be expected from the fitted AR(1) data generating process.

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, 1996) and Frees et al (1996). Examples of the latter include McNees (1979), Engle (1982), Akgray (1989), Hamilton & Susmel (1994), Hamilton & Lin (1996) and Gray (1996).

This paper presents an attempt to deal with the observed difficulties in financial time series, a Regime Switching Vector Autoregression (RSVAR), the parameters of which are subject to periodic discrete changes. The process may have quite different characteristics in different regimes. A tractable mathematical model of structural changes and discrete market regimes is the univariate Markov regime switching autoregressive process introduced by Hamilton (1989), and subsequently 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, 1995a, 1995b, 1995c) and Hamilton & Lin (1996), a vector joint regime switching process would seem to be an attractive description of the data.

Hamilton (1990, p40) observed that the usual numerical maximisation of regime switching likelihood functions is subject to computational difficulties associated with the often ill-behaved likelihood surface (multiple local maxima, essential singularities, and local increases as boundary conditions are approached). He suggested a numerically robust Expectation-Maximisation (EM) maximum likelihood algorithm to overcome the numerical

problems. The author has performed convergence comparisons of EM and Markov Chain Monte Carlo (MCMC) estimation procedures for regime switching processes, which demonstrate much faster convergence of the MCMC estimation, particularly for larger data sets.

The present paper extends regime switching to vector processes and develops a Bayesian Markov Chain Monte Carlo estimation procedure that is more informative, efficient, and flexible than a maximum likelihood based approach. The estimation procedure estimates regimes at each time point, regime transition probabilities, and vector process parameters within each regime. In addition it is numerically efficient, ensures stability of the data generating process within each regime, is able to assess the joint significance of the large number of potential parameters, is numerically stable, produces robust parameter estimates, and enables the estimation of posterior model probabilities for use in non-nested model selection.

The RSVAR process 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. Model fitting results are presented in section 4, while concluding remarks are made in section 5. The derivation of the Bayesian MCMC estimation procedure is contained in the Appendix.

2. THE MODEL

Vector regime switching processes are characterised by multiple discrete regimes, where each regime has different dynamics and is characterised by a different set of parameters. They are subject to probabilistic joint discrete shifts in the parameters of the vector process, so that a regime shift may alter the dynamics of all the variables at the same time. Within each regime the vector 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 of the vector process characterising the different regimes, and the regime transition probabilities.

Define ρ_t to be an unobserved discrete-valued indicator variable, such that at any time t the process will be in regime $\rho_t \in \{1, \dots, K\}$. Define the transition probabilities, $p_{ij} = P(\rho_t = j | \rho_{t-1} = i)$ with $\sum_j p_{ij} = 1 \quad \forall i$, 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,

$$\mathbf{x}_t = \mu_{(\rho_t)} + \sum_{h=1}^q \mathbf{A}_{(\rho_t)}^{(h)} (\mathbf{x}_{t-h} - \mu_{(\rho_t)}) + \xi_{t(\rho_t)},$$

$\xi_{t(\rho_i)} \sim N(\mathbf{0}, \Omega_{(\rho_i)})$, $E\xi_{t(\rho_i)} = 0$ and $E\xi_{t(\rho_i)}\xi_{t(\rho_i)}^T = \Omega_{(\rho_i)} \quad \forall t > q$. The variables \mathbf{x} , μ and ξ are $m \times 1$ column vectors, while the $\mathbf{A}^{(h)}$ and the Ω are $m \times m$ matrices. For convenience the above regime switching VAR(q) process will be denoted an RSVAR(q, K) process.

The total parameter set to be estimated is $\lambda \equiv \{\mu_{(1)}, \dots, \mu_{(K)}, \mathbf{A}_{(1)}, \dots, \mathbf{A}_{(K)}, \Omega_{(1)}, \dots, \Omega_{(K)}, \mathbf{P}\}$, which can be partitioned as $\lambda \equiv \{\Theta, \mathbf{P}\}$. To ensure that the process is identifiable, it will usually 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

3. MARKOV CHAIN MONTE CARLO SAMPLING 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. The resulting simulated sample from the parameter space can be used to make inferences about the distribution of the process parameters and regimes. Chib & Greenberg (1995) provide a useful and readable description of MCMC methods

Posterior simulators are particularly attractive in the case of complicated processes which are impossible or impractical to estimate using maximum likelihood or other methods. Advantage can often be taken of structure within a process to decompose the inference problem into manageable components. In the case of Regime Switching VAR, the process is non-Normal and path dependent, however, conditional on the regime, the process is a regular VAR process

Markov Chain theory would usually start with a transition kernel density, $\kappa(x, y)$, which describes how the Markov Chain moves between states. In the current application, the state space is the joint parameter and regime space of the RSVAR process, $\{\rho, \lambda\}$, and the Markov Chain is the sequence of joint parameter and regime estimates, $\{\rho^{(c)}, \lambda^{(c)}\}$

Since the process must end up somewhere at each transition, $\int \kappa(x, y) dy = 1$. The probability of the process being in state y after n transitions, given that it was initially in state x , is given by $\kappa_n(x, y) = \int \kappa_{n-1}(x, z) \kappa(z, y) dz$. A limiting or invariant distribution, $\pi(\cdot)$, is said to exist whenever $\kappa_n(x, y) \rightarrow \pi(y)$ as $n \rightarrow \infty$. It follows therefore that $\pi(y) = \int \pi(x) \kappa(x, y) dx$

A major concern of Markov Chain theory is to determine conditions under which there exists an invariant distribution, and conditions under which iterations of the transition kernel converge to the invariant distribution. MCMC sampling 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 kernel is unknown.

3.1 The Gibbs Sampler

Starting from an arbitrary point, the Gibbs Sampler generates samples from a joint density $\pi(\cdot)$ via a sequence of random draws or samples from full conditional densities, as follows

$$\begin{aligned}
 y_1 &\leftarrow \pi(y_1|x_2, \dots, x_m) \\
 &\vdots \\
 y_j &\leftarrow \pi(y_j|y_{1<j}, x_{j>j}) \\
 &\vdots \\
 y_m &\leftarrow \pi(y_m|y_1, \dots, y_{m-1}).
 \end{aligned}$$

The above completes a transition from x to y in the state space. The sequence forms a realisation of a Markov Chain which converges in distribution to a random sample from the target joint distribution $\pi(\cdot)$

The Gibbs sampler will often be useful where a complicated process can be built up from components with standard conditional distributions. The target joint regime and parameter distribution of the RSVAR process is too complicated to sample directly. The structure of the RSVAR process is however such that draws from the full conditional densities can be made tractable, since essentially standard densities arise for the VAR parameters once one conditions on the sequence of regimes

3.2 The Metropolis-Hastings Algorithm

Suppose the target density $\pi(\cdot)$ is unknown, but that a density $q(\cdot, \cdot)$ 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 target density. The candidate generating density 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

$$\alpha(x, y) = \begin{cases} \min\left\{\frac{\pi(y)}{\pi(x)} \frac{q(x, y)}{q(y, x)}, 1\right\} & \pi(x) > q(x, y) \\ 1 & \text{otherwise} \end{cases}$$

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. $\pi(x) \propto \psi(x) \phi(x)$, 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 is 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.

3.3 Outline of the Estimation Procedure

Samples 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 $p(\rho, \lambda | \mathbf{Y})$.

4. RESULTS

4.1 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 process was incorrect. 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, even for vector processes. This is in contrast to maximum likelihood based approaches, which are subject to computational difficulties and/or are relatively slow to converge.

When data was generated from a VAR process (without regimes), the MCMC algorithm very rapidly collapsed to a single regime with $p_{11} \rightarrow 1$ & $p_{12} \rightarrow 0$. 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. $x_{t,2} = 0.01 + 0.75(x_{t-1,2} - 0.01) - 0.25(x_{t-2,2} - 0.01) + 0.005z_t$, where $z_t \sim \text{iid } N(0,1)$.

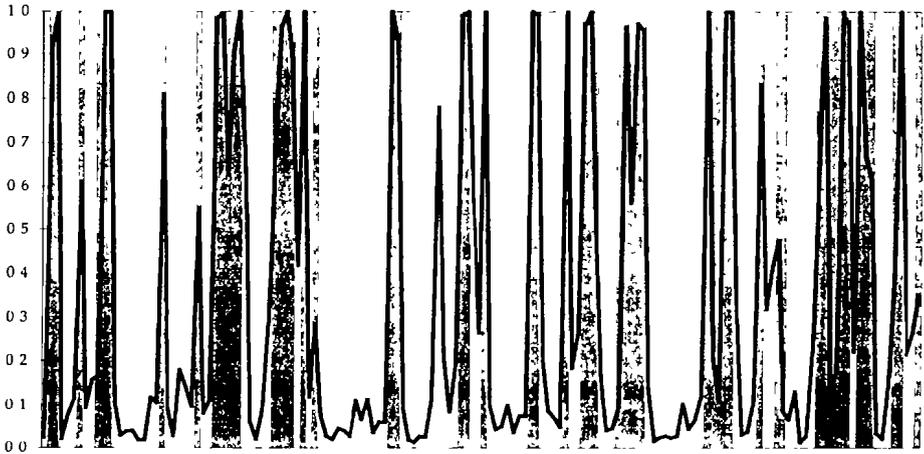
TABLE 1
SIMULATION TEST RESULTS

	True	1st 50 iterates/samples			2000 iterates/samples		
	Value	5% 'ile	Mean	95% 'ile	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%
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%

The MCMC estimation procedure described in the Appendix 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. 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 can be seen to have successfully differentiated between the low and high volatility regimes.



GRAPH 1 Probability of Being in Regime 2

4.2 Empirical Estimation Results

To further illustrate the estimation procedure, an RSVAR(1.2) process was fitted to a real financial data set ($m = 4$, $q = 1$, $K = 2$, $N = 147$). 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

- $\nabla \ln \text{GDP}_t$, where GDP_t is the real Gross Domestic Product for the quarter ending time t ;
- $\nabla^2 \ln \text{CPI}_t$, where CPI_t is the Consumer Price Index at time t ,
- $\nabla \ln \text{SPI}_t$, 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 Treasury bonds, where ∇ is the backward difference operator.

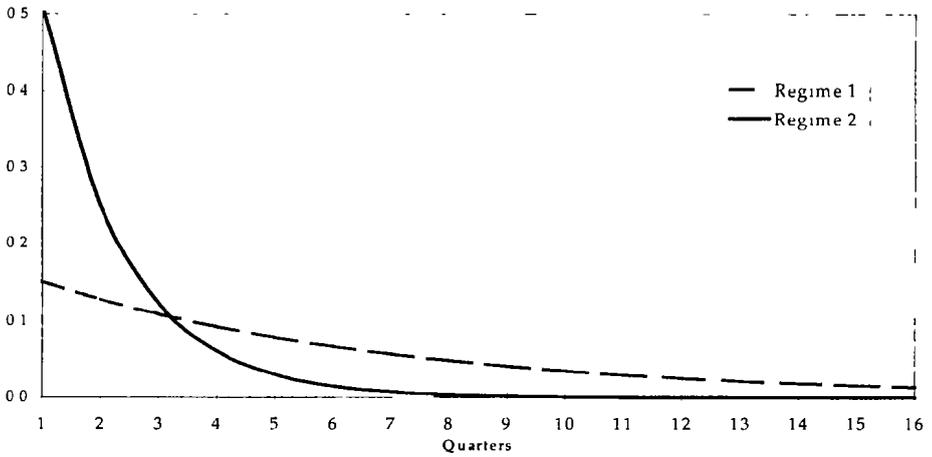
The data set was chosen on the basis of its general interest and convenience, to illustrate the MCMC estimation procedure developed in the paper. It is not suggested that the resulting model is appropriate for any other specific purpose, e.g. it would clearly be inappropriate for long term projections, given the degree of differencing.

5,000 iterations/samples were generated using the MCMC estimation procedure described in the Appendix. Rapid convergence was observed in the quantities measured. The first 50 samples were discarded and the remaining 4,950 samples used to describe the joint parameter density.

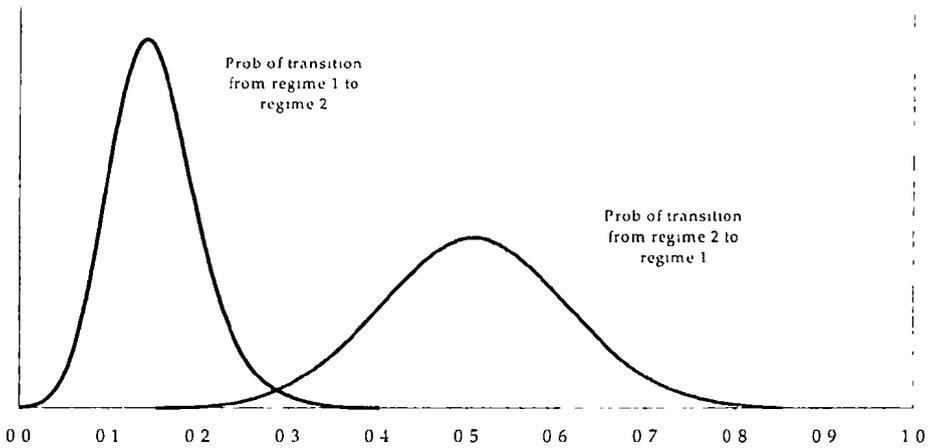
The estimation procedure identified two clearly distinct regimes. One regime (regime 1) was characterised by strong economic growth, stable inflation and interest rates, and relatively stable share price growth. The other regime (regime 2) was characterised by weak economic growth, volatile inflation and interest rates, and volatile and generally falling share prices. The low volatility regime was relatively stable in the sense that it was the more persistent of the two regimes.

The duration of a regime i episode is a discrete random variable, with expected value given by $\sum_{k=1}^{\infty} k \times p_{ii}^{k-1} \times (1 - p_{ii}) = 1/(1 - p_{ii})$. Given the estimated mean transition probability of 0.15, the expected duration of a regime 1 episode is about $6\frac{1}{2}$ quarters. The high volatility regime was unstable in the sense that it was not expected to persist for long. Given its estimated mean transition probability of 0.51, the expected duration of a regime 2 episode is 2 quarters. The identified regimes seem highly intuitive.

The theoretical density of the duration of an episode of each regime, given the mean transition probabilities, $\text{Prob}(\text{duration of regime } i = k) = p_{ii}^{k-1}(1 - p_{ii})$, is shown in graph 2. The duration of an episode of either regime can be seen to vary over a considerable range, even if the transition probabilities were known with certainty. The uncertainty in the estimated transition probabilities themselves is illustrated by graph 3.

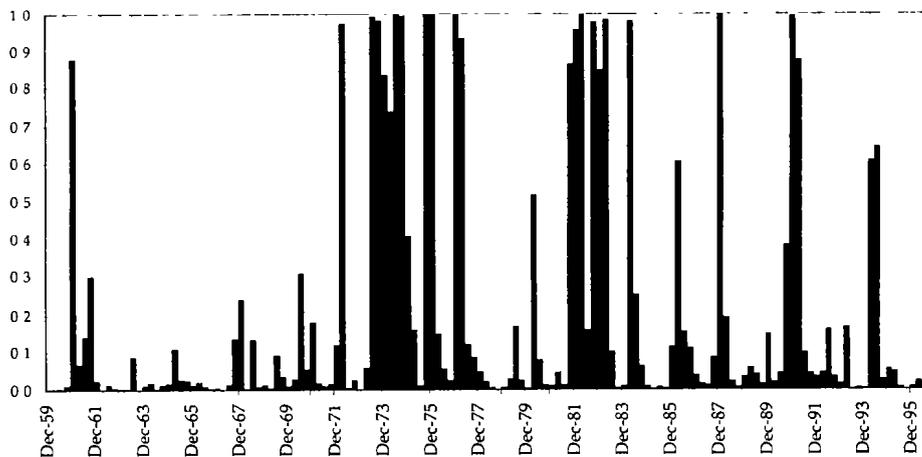


GRAPH 2 Density of Duration of Episodes



GRAPH 3 Density of Transition Probabilities

Graph 4 shows the mean regime at each time point. 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

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.

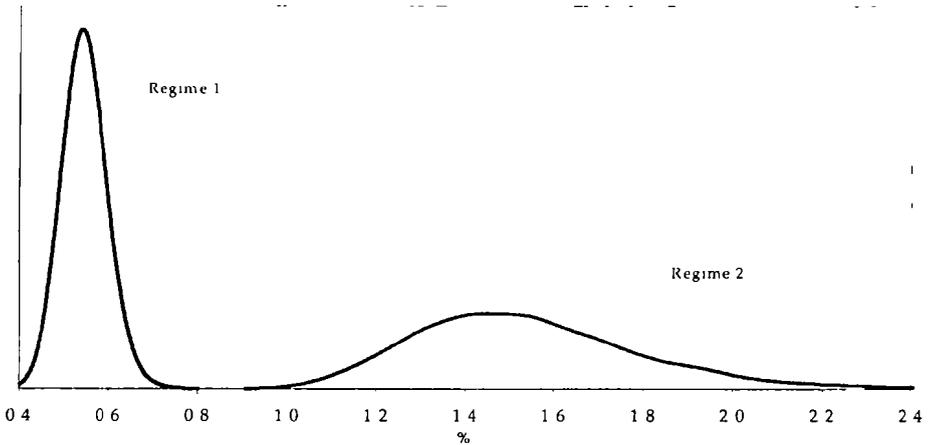
TABLE 2
PARAMETER ESTIMATES

	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%

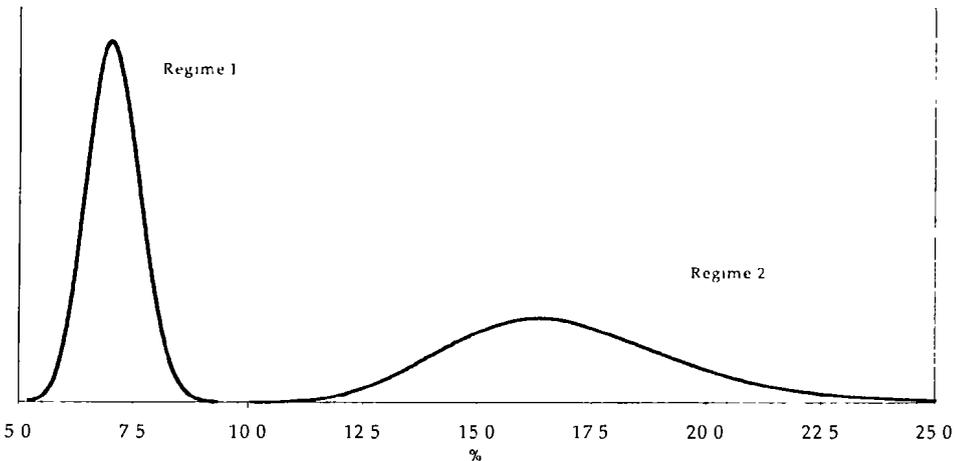
variable 1 = $\nabla \ln GDP_t$, variable 2 = $\nabla^2 \ln CPI_t$, variable 3 = $\nabla \ln SPI_t$, variable 4 = $\nabla \ln B_t$

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.



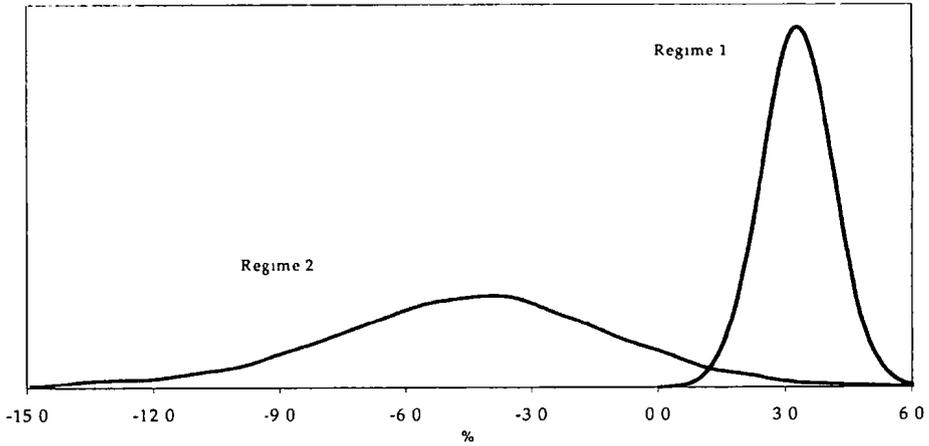
GRAPH 5a Density of Std Dev of Change in Inflation Rate Parameter



GRAPH 5b Density of Std Deviation of Share Price Return 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% versus 0.55%). In the case of the quarterly share price return, the mean

standard deviation parameter estimate in regime 2 is $2\frac{1}{2}$ times as large as in regime 1 (17% versus 7%). The mean share price return level parameter estimate in regime 1 is 3.3% compared with -4.6% in regime 2.



GRAPH 5c. Density of Mean Share Price Return Parameter

Examination of the marginal parameter densities revealed that relatively few regressive cross-correlation parameters were significantly nonzero once joint regime switching was allowed for. 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.

<i>mean A₍₁₎</i>				<i>mean A₍₂₎</i>			
-0.178	0.078	0.021	-0.009	0.137	-0.021	-0.009	0.084
0.049	-0.332	0.000	0.006	-0.108	-0.449	0.019	0.080
-0.156	0.078	0.072	-0.049	0.033	-0.030	-0.111	-0.125
-0.052	0.086	0.048	0.222	0.054	0.008	0.098	0.160

variable 1 = $\nabla \ln GDP_t$, variable 2 = $\nabla^2 \ln CPI_t$, variable 3 = $\nabla \ln SPI_t$, variable 4 = $\nabla \ln B_t$. Numbers in bold are significantly nonzero at the 5% one-sided level under the posterior density, i.e. zero falls outside the region bounded by the 5% tile and the 95% tile of the posterior parameter density

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 therefore interesting to compare the mean regression parameters of the RSVAR(1,2) process with the corresponding VAR(1) parameters, which reflect the usual sample correlations/regression relationships (the stability of which was previously questioned).

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

variable 1 = $\nabla \ln \text{GDP}_t$, variable 2 = $\nabla^2 \ln \text{CPI}_t$, variable 3 = $\nabla \ln \text{SPI}_t$, variable 4 = $\nabla \ln B_t$
 Numbers in bold are simply large in magnitude

Notable differences are the large feedback of lagged changes in inflation into share price returns (-0.74) and changes in interest rates (0.56), and the feedback of lagged changes in interest rates into share price returns (-0.28). 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 was observed empirically (refer section 4.3).

A number of contemporaneous error correlations were found to be significantly nonzero. In the stable regime changes in bond yields were found to be contemporaneously negatively correlated with share price returns (-0.195), so that unanticipated shocks that caused bond yields to rise were more likely to be associated with a downward (than an upward) shock to share prices. Similarly, increases in inflation were found to be contemporaneously correlated with real economic growth (0.15). Both contemporaneous relationships appear consistent with intuition.

No significant relationship was revealed between inflation and share price returns, at least in the short term. In the unstable regime the only significant contemporaneous error correlation detected was between share price returns and real economic growth (-0.34), the sign of which is somewhat counter-intuitive.

<i>Mean contemporaneous error correlations implied by $\Omega_{(1)}$</i>				<i>Mean contemporaneous error correlations implied by $\Omega_{(2)}$</i>			
1				1			
0.148	1			-0.098	1		
-0.044	-0.053	1		-0.337	0.004	1	
0.087	-0.084	-0.195	1	0.132	-0.120	-0.181	1

variable 1 = $\nabla \ln \text{GDP}_t$, variable 2 = $\nabla^2 \ln \text{CPI}_t$, variable 3 = $\nabla \ln \text{SPI}_t$, variable 4 = $\nabla \ln B_t$
 Numbers in bold are significantly nonzero at approximately the 5% one-sided level under the posterior density, i.e. zero falls outside the region bounded by approximately the 5% 'ile and the 95% 'ile of the posterior parameter density

4.3 Empirical Comparison With Common Models

In this section the statistical goodness-of-fit of the Regime Switching VAR model is compared with commonly used models. The models considered were independent random/noise, independent autoregressive (i.e. diagonal VAR), independent non-Gaussian autoregressive, independent GARCH, 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. The Student t density, standardised to have zero mean and unit variance, was considered as an alternative to the standard Normal error distribution.

The Generalised ARCH model was 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 $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

The models were compared in terms of their maximum log-likelihood, posterior model probabilities, prediction errors, ability to predict volatility, and their ability to explain the observed excess kurtosis (a measure of non-Normality). For the purposes of the comparison, the MCMC iteration/sample which produced the largest log-likelihood value was used to calculate the measures for the regime switching model. The regime switching measures were therefore based on a single realisation from the joint parameter and regime density.

The maximum log-likelihoods, both unconditional and conditioned on the first data point, are reported in table 3. The standard likelihood ratio test was used to assess the significance of the respective maximum log-likelihoods. Where one model is completely nested within another, twice the increase in the maximum log-likelihood is asymptotically distributed as χ_k^2 where k is the number of additional parameters fitted in the more general of the two models (equal to the number of parameter restrictions imposed by the null hypothesis). Simulation experiments confirmed the appropriateness of the asymptotic χ_k^2 distribution for samples of the size considered. Thus the data suggests the AR(1) model is significantly more likely than the Random model and both the Student t 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 (Student t) produced a substantial increase in the maximum likelihood with the addition of only 4 parameters. The VAR models are *not* significantly more likely than the independent AR(1) model, suggesting spurious regressive correlations between the series based on sample cross-correlations.

The usual asymptotic statistical distribution theory fails to apply in the regime switching case, since the transition probabilities are not defined under the null hypothesis that the regime switching model is inappropriate. If standard distribution theory did apply, the RSVAR(1,2) model would be overwhelmingly more likely than the independent AR(1) or VAR models.

Though not a statistical test, it is at least reassuring that there is a large increase in the maximum log-likelihood after allowing for the larger number of potential parameters. The addition of the second lag in the RSVAR(2,2) model produced only a modest increase in the maximum log-likelihood. 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 standard LR test assesses the significance of the evidence *against* a nested null hypothesis, based on the maximum likelihood of model j , i.e. maximum over λ_j of $p(\mathbf{Y}|\lambda_j, M_j)$. An alternate, and arguably more sensible model selection approach, which does not require the models to be nested, is to consider the evidence in the observed data in *favour* of the alternative models. The probability of model j given the observed data is given by $p(M_j|\mathbf{Y}) \propto p(\mathbf{Y}|M_j) \times p(M_j)$. A priori convictions regarding null hypotheses are reflected in the prior probabilities assigned to the alternative models, $p(M_j)$, which are then modified by the likelihood of the observed data given the models, $p(\mathbf{Y}|M_j)$. MCMC estimation enables $p(\mathbf{Y}|M_j) = \int p(\mathbf{Y}|\lambda_j, M_j)p(\lambda_j|M_j)d\lambda_j$ to be readily estimated as the harmonic average of the likelihood over the MCMC samples/iterations, e.g. refer to Kass & Raftery (1995, 4.3). The harmonic average estimate was found to have stabilised after 1,000 samples or so in the present case.

The likelihood of the data given model j can be seen to be related to a likelihood where the parameters are eliminated by *integration* rather than *maximisation*. The logarithm of $p(\mathbf{Y}|M_j)$ is shown in table 3 for each of the models which could be estimated based on the MCMC estimation scheme described in the paper. The model probabilities are also shown, assuming the alternative models considered were assumed equally likely a priori. The data overwhelmingly supports the RSVAR model over the AR and VAR models. Of the 4 models considered, the probability that the data was generated by the AR(1) or VAR alternatives is virtually zero, while the probability that the data was generated by the RSVAR alternative is virtually 1.

The average prediction or forecast errors for each model were assessed using the root-mean-square error, which for series i was defined as $rmse_i = \sqrt{\frac{1}{N-q} \sum \varepsilon_i^2}$, where ε_i 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. $wrms\ error = \sqrt{w_i \times rmse_i^2}$. Despite their high likelihood, both the Student t AR(1) and GARCH(1,1)-AR(1) models produced forecasts no better than the simpler AR(1) model on average.

To assess the explanatory information contained in the regime, the regime switching model residuals in each period were calculated conditional on the realised regime in that period, i.e. the $\xi_{t(\rho_i)} \sim N(0, \Omega_{(\rho_i)})$ were tested.

Allowing for the regime, the RSVAR model produced the smallest errors on average (it is noted that the RSVAR model does have the largest number of parameters).

Two measures were used to assess the ability to predict volatility. The first measure used was the root-mean-square squared error, defined for series i as $rmsse_i = \sqrt{\frac{1}{N-q} \sum (\epsilon_i^2 - \sigma_i^2)^2}$, where σ_i is the one-period-ahead predicted error standard deviation according to the model. The $rmsse$ measure directly compares the sample volatility with the volatility predicted by the model, since $E(\epsilon_i^2) = \sigma_i^2$. While intuitive, the $rmsse$ measure is highly skewed, since $\epsilon_i^2 \sim \sigma_i^2 \chi_1^2$, and hence is susceptible to outliers. The rms squared errors were also combined into a single *weighted rms squared error* using the same weights as used for the *wrms error*.

The second measure used was the root-mean-square normalised absolute error, defined for series i as $rmsnae_i = \sqrt{\frac{1}{N-q} \sum \left(|\epsilon_i|^{2/3} - \frac{7}{9} \sigma_i^{2/3} \right)^2}$. While less

intuitive than the previous measure, the $rmsnae$ should be more robust, since the deviations should be less skew and more Normal. The measure was motivated by the approximately Normal transformation of the chi-square density, $(\chi_\nu^2/\nu)^{1/3} \sim N(1 - \frac{2}{9\nu}, \frac{2}{9\nu})$, so that $|\epsilon_i|^{2/3} \sim \sigma_i^{2/3} N(\frac{7}{9}, \frac{2}{9})$.

Allowing for the regime, the RSVAR model produced better predictions of volatility than the other models on average. Discrete regime switching would appear to be a better explanation of conditional heteroscedasticity than the commonly used GARCH and ARCH processes, which, despite their high log-likelihood, generally impute too much persistency in the volatility (see, for example, Hamilton & Susmel (1994)).

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 model was able to successfully account for the excess kurtosis in terms of discrete regime switching in the variance, i.e. conditional heteroscedasticity. The non-Normal process explicitly models excess kurtosis by assuming the residuals are drawn from a leptokurtic non-Normal distribution. The excess kurtosis of a standardised Student t density, when finite, is $6/(\nu - 4)$, where ν is the degrees of freedom parameter. Since the fitted ν values of 3 out of the 4 series were less than 4, the kurtoses of the fitted non-Normal processes are in general not finite, which appears inconsistent with the observed residuals, which had an average excess kurtosis of 4.4.

TABLE 3
 MODEL COMPARISON

	<i>Random</i>	<i>AR(1)</i>	<i>Student t AR(1)</i>	<i>GARCH (1,1) -AR(1)</i>	<i>VAR(1)</i>	<i>VAR(2)</i>	<i>RSVAR (1,2)</i>
uncond max lnL	1231.4	1258.9			1271.1	1284.6	1326.5
cond max lnL	1221.5	1249.0	1296.4	1287.3	1261.2		1315.5
$\Delta \ln L$ over AR(1)		0.0	47.4	38.3	12.2		66.5
significance of std χ^2			$< 10^{-6}$	$< 10^{-6}$	0.44		$< 10^{-6}$
$\ln p(Y M)$		1221.0			1253.6	1261.1	1296.7
$p(M Y)$		$< 10^{-30}$			$< 10^{-15}$	$< 10^{-15}$	1.000
<i>wrms error</i>	1.48%	1.40%	1.41%	1.41%	1.39%	1.36%	1.30%
<i>wrmse</i> as % Random	100%	95.0%	95.6%	95.6%	94.0%	92.0%	88.1%
<i>wrmse</i> as % Random	100%	99.5%	101.0%	103.4%	92.7%	93.5%	70.8%
<i>wrmsnac</i> as % Random	100%	89.6%	93.2%	84.7%	87.9%	84.1%	78.1%
ave excess kurtosis	4.6	4.0	4.4	2.1	3.5	3.2	0.8

5. CONCLUSIONS

A Bayesian Markov Chain Monte Carlo (MCMC) procedure was developed for estimating the joint parameter and regime density of Regime Switching Vector Autoregressions (RSVAR). 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 process was incorrect. 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, even for vector processes (in contrast to maximum likelihood based approaches).

The estimation procedure identified two clearly distinct regimes in quarterly Australian financial data. One regime was characterised by strong economic growth, stable inflation and interest rates, and relatively stable share price growth. The other regime was characterised by weak economic growth, 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.

Regression relationships are often not robust to outliers nor stable over time. Relatively few regressive cross-correlations appeared important in the dynamics once joint regime switching was allowed for, in contrast to the

large lagged 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

MCMC estimation enabled the calculation of posterior model probabilities, i.e. the probabilities of the various models given the observed data. The data overwhelmingly supported the RSVAR model. The RSVAR model also produced the lowest average prediction errors and better predictions of volatility on average. Discrete regime switching would appear to be a better explanation of conditional heteroscedasticity than the commonly used GARCH and ARCH processes.

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 periodic discrete shifts. The Regime Switching Vector Autoregression process was found to provide a particularly good description of an Australian quarterly financial data set.

ACKNOWLEDGEMENTS

I am indebted to Professor Robert Kohn of the Australian Graduate School of Management (UNSW) for introducing me to Bayesian MCMC estimation. I am also indebted to my then employer, AMP Life, for enabling me to present the basis of this paper at the 7th AFIR International Colloquium in Cairns, Australia, in August 1997.

GLEN R. HARRIS

Lend Lease Investment Management

Level 43, Australia Square

Sydney NSW 2000

Australia

Email Glen_Harris@lendlease.com.au

Telephone 61 2 9237 5364

Facsimile 61 2 9236 6345

APPENDIX

The derivation of the Bayesian MCMC estimation procedure is described in this Appendix. At times it will be more convenient to consider the equivalent VAR(1) form of the RSVAR(q, K) process, namely

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

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

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

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

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

$$\xi_{t(\rho_t)} = x_t - \mu_{(\rho_t)} - \sum_{h=1}^q \mathbf{A}_{(\rho_t)}^{(h)}(x_{t-h} - \mu_{(\rho_t)})$$

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

The first q data vectors can be taken together. $l(\mathbf{X}_q | \rho_q, \lambda)$ can be approximated by exploiting stationarity within each regime (effectively ignoring regime shifts prior to time q), so that the contribution to the likelihood from the first q data vectors can be approximated by

$$l(\mathbf{X}_q | \rho_q, \lambda) = (2\pi)^{-\frac{mq}{2}} \cdot |\mathbf{V}(\Omega_{(k)}, \mathbf{A}_{(k)})^{-1}|^{\frac{1}{2}}$$

$$\exp\left\{-\frac{1}{2}(\mathbf{X}_q - \tilde{\mu}_{(k)})^T \mathbf{V}(\Omega_{(k)}, \mathbf{A}_{(k)})^{-1} (\mathbf{X}_q - \tilde{\mu}_{(k)})\right\}.$$

The within regime unconditional or stationary $mq \times mq$ variance-covariance matrix, \mathbf{V} , could be determined from $\text{vec} \mathbf{V} = (\mathbf{I}_{m^2 q^2} - \mathbf{A} \otimes \mathbf{A})^{-1} \text{vec} \Omega$, as described by Lütkepohl (1991, p21-22). vec is the column stacking operator, such that if $\mathbf{A} \equiv (a_1, \dots, a_n)$ is an $m \times n$ matrix with $m \times 1$ columns a_i , then $\text{vec} \mathbf{A}$ is the $mn \times 1$ column vector $(a_1^T, \dots, a_n^T)^T$, and \otimes is the (right) kronecker product, such that 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 $(a_{ij} \mathbf{B})$

Assuming within regime stationarity,

$$\begin{aligned} \mathbf{X}_q &= \tilde{\mu} + \mathbf{A}(\mathbf{X}_{q-1} - \tilde{\mu}) + \tilde{\xi}_q \\ &= \tilde{\mu} + \tilde{\xi}_q + \mathbf{A}\tilde{\xi}_{q-1} + \mathbf{A}^2(\mathbf{X}_{q-2} - \tilde{\mu}) \\ &= \tilde{\mu} + \sum_{j=0}^{\infty} \mathbf{A}^j \tilde{\xi}_{q-j} \end{aligned}$$

The unconditional mean is therefore $E\mathbf{X}_q = \tilde{\mu}_{(k)}$, where $\rho_{ij} = k$. Assuming the process is stable within each regime,

$$\begin{aligned} \text{Var}\mathbf{X}_q &= E(\mathbf{X}_q - \tilde{\mu})(\mathbf{X}_q - \tilde{\mu})^T \\ &= E\left\{ \left(\sum_{j=0}^{\infty} \mathbf{A}^j \tilde{\xi}_{q-j} \right) \left(\sum_{j=0}^{\infty} \tilde{\xi}_{q-j}^T (\mathbf{A}^j)^T \right) \right\} \\ &= E\left\{ \sum_{j=0}^{\infty} \mathbf{A}^j \tilde{\xi}_{q-j} \tilde{\xi}_{q-j}^T (\mathbf{A}^j)^T \right\} \\ &= \sum_{j=0}^{\infty} \mathbf{A}^j \tilde{\Omega} (\mathbf{A}^j)^T \\ &= V(\Omega_{(k)}, \mathbf{A}_{(k)}) \\ &= V_{(k)} \end{aligned}$$

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

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 $\tilde{\Omega} + \sum_{j=1}^n \mathbf{A}^j \tilde{\Omega} (\mathbf{A}^j)^T$, and then an approximation to \mathbf{V}^{-1} obtained by inverting the approximation to \mathbf{V} (an $m q \times m q$ matrix)

The full likelihood conditional on the regimes is

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

where $\mathbf{Y} = \mathbf{Y}_N$ and $\rho \equiv \{\rho_q, \dots, \rho_N\}$. The exact or unconditional likelihood of λ 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 .

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 $p(\rho, \lambda|\mathbf{Y})$.

Generating the Regimes

$\rho^{(c+1)} \leftarrow \mu^{(c)}, \mathbf{A}^{(c)}, \Omega^{(c)}, \mathbf{P}^{(c)}$. Carter & Kohn (1994) suggested that it is generally much more efficient to generate the regime variables simultaneously from the joint distribution of the ρ_t rather than one-at-a-time from the full conditional densities. Generating the regimes one-at-a-time can lead to little movement at each iteration due to conditioning on the neighbouring regimes from the previous iteration. Their empirical results are supported theoretically by the results of Liu, Wong & Kong (1994). Following Carter & Kohn (1994), the regimes can be generated jointly from

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

Estimation of the regimes at each time t is based on the probability filters of Hamilton (1989, 1990) and Kim (1994). The first step in calculating the smoothed regime probability estimates is to calculate the joint regime filter probabilities and conditional likelihoods, which can be determined recursively. The filter probabilities, $p(\rho_N|\mathbf{Y}, \lambda)$, 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 $t = q + 1, \dots, N$,

$$l(\mathbf{x}_t, \rho_t, \rho_{t-1}|\mathbf{Y}_{t-1}, \lambda) = l(\mathbf{x}_t|\rho_t, \mathbf{Y}_{t-1}, \lambda) \cdot p(\rho_t|\rho_{t-1}) \cdot p(\rho_{t-1}|\mathbf{Y}_{t-1}, \lambda)$$

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

$$p(\rho_t, \rho_{t-1}|\mathbf{Y}_t, \lambda) = l(\mathbf{x}_t, \rho_t, \rho_{t-1}|\mathbf{Y}_{t-1}, \lambda)/l(\mathbf{x}_t|\mathbf{Y}_{t-1}, \lambda)$$

$$p(\rho_t|\mathbf{Y}_t, \lambda) = \sum_{\rho_{t-1}=1}^K p(\rho_t, \rho_{t-1}|\mathbf{Y}_t, \lambda)$$

$$\forall \rho_t, \rho_{t-1} \in \{1, \dots, K\}$$

Note the filter probabilities $p(\rho_t|\mathbf{Y}_t, \lambda)$ are each a K -tuple of probabilities, representing $p(\rho_t = i|\mathbf{Y}_t, \lambda)$ for $i = 1, \dots, K$. 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, $l(\mathbf{x}_t|\rho_t, \mathbf{Y}_{t-1}, \lambda)$, which will require evaluation of the $m \times m$ determinants of the $K \Omega^{-1}$. Using the Choleski decomposition, $|\Omega^{-1}| = |\mathbf{L}||\mathbf{L}^T| = \prod \sigma_{ii}^2$.

To initialise the previous iterations, the K $p(\rho_q|\lambda)$ will be required. They can be derived as the limiting distribution of the regime Markov chain. Define the $K \times 1$ column vector $\pi \equiv \{p(\rho_t = i|\lambda), i = 1, \dots, K\}$, then $\pi = \mathbf{P}\pi$. The limiting distribution 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 . Observe that \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.

Thus, drawing on Carter & Kohn (1994), 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

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

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

Generating the Parameters

The conditional densities of the parameters are given by

$$\begin{aligned} p(\lambda_j|\rho, \lambda_{\neq j}, \mathbf{Y}) &\propto L(\mathbf{Y}|\rho, \lambda) \cdot p(\rho|\lambda) \cdot p(\lambda_j), \\ \text{i.e. } p(\Theta_j|\rho, \Theta_{\neq j}, \mathbf{P}, \mathbf{Y}) &\propto L(\mathbf{Y}|\rho, \lambda) \cdot p(\Theta_j) \\ \text{and } p(\mathbf{P}|\rho, \Theta, \mathbf{Y}) &\propto p(\rho_q|\mathbf{P}) \prod_{t=q+1}^N p(\rho_t|\rho_{t-1}, \mathbf{P}) \cdot p(\mathbf{P}) \end{aligned}$$

Generating the Level Parameters

$\mu^{(c+1)} \leftarrow \rho^{(c+1)}, \mathbf{A}^{(c)}, \Omega^{(c)}, \mathbf{P}^{(c)}$. In this section, $\mu_{(t)}$ represents one of the possible discrete values of $\mu_{(\rho_t)}$, $\rho_t \in \{1, \dots, K\}$. Independent uniform priors can be used for the $\mu_{(t)}$, 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

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

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

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

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

$$\begin{aligned}
 & -\frac{1}{2}\tilde{\xi}_q^T \mathbf{V}_{(k)}^{-1} \tilde{\xi}_q \\
 & - \frac{1}{2} \left(\mu_{(k)} - \frac{1}{n_k - 1} \left(\mathbf{I}_m - \sum_{h=1}^q \mathbf{A}_{(k)}^{(h)} \right)^{-1} \sum_{\substack{\rho_t=k \\ t>q}} \left(\mathbf{x}_t - \sum_{h=1}^q \mathbf{A}_{(k)}^{(h)} \mathbf{x}_{t-h} \right) \right)^T \\
 & (n_k - 1) \mathbf{W}_{(k)}^{-1} \left(\mu_{(k)} - \frac{1}{n_k - 1} \left(\mathbf{I}_m - \sum_{h=1}^q \mathbf{A}_{(k)}^{(h)} \right)^{-1} \sum_{\substack{\rho_t=k \\ t>q}} \left(\mathbf{x}_t - \sum_{h=1}^q \mathbf{A}_{(k)}^{(h)} \mathbf{x}_{t-h} \right) \right) \\
 & - \frac{1}{2} \sum_{r \neq k} \left(\mu_{(r)} - \frac{1}{n_r} \left(\mathbf{I}_m - \sum_{h=1}^q \mathbf{A}_{(r)}^{(h)} \right)^{-1} \sum_{\rho_t=r} \left(\mathbf{x}_t - \sum_{h=1}^q \mathbf{A}_{(r)}^{(h)} \mathbf{x}_{t-h} \right) \right)^T \\
 & n_r \mathbf{W}_{(r)}^{-1} \left(\mu_{(r)} - \frac{1}{n_r} \left(\mathbf{I}_m - \sum_{h=1}^q \mathbf{A}_{(r)}^{(h)} \right)^{-1} \sum_{\rho_t=r} \left(\mathbf{x}_t - \sum_{h=1}^q \mathbf{A}_{(r)}^{(h)} \mathbf{x}_{t-h} \right) \right)
 \end{aligned}$$

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 $\mu_{(r)}$.

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

where $\mathbf{N}(\cdot, \cdot)$ is the multivariate or vector Normal density. 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.

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

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,

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

and accepted, i.e. $\mu_{(k)}^{(c+1)} = \mu_{(k)}^{(*)}$, with probability

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

otherwise the previous value 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 $p(\mu_{(1)}, \dots, \mu_{(K)})$ are satisfied (i.e. by direct acceptance/rejection).

Generating the Regressive Correlation Parameters

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

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

Recall that

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

so that only the first m rows need to be generated. Define the $m \times mq$ matrix operator $\vartheta \equiv (\mathbf{I}_m, \mathbf{0}_m, \mathbf{0}_m, \dots, \mathbf{0}_m)$, so that $\vartheta \mathbf{A} = (\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(q)})$. 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 $p(\vartheta \mathbf{A}_{(r)}) \propto \mathbf{N}(\mathbf{B}_{(r)}, 1/\nu_{(r)} \mathbf{I}_{m^2q}) \times g(\mathbf{A}_{(r)})$, where the first term is a matrix Normal density, $\mathbf{B}_{(r)} = (\mathbf{B}_{(r)}^{(1)}, \mathbf{B}_{(r)}^{(2)}, \dots, \mathbf{B}_{(r)}^{(q)})$ 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 $\mathbf{B}_{(r)}^{(i)}$ 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}^j converge rapidly to zero as $j \rightarrow \infty$, so that the sequence $\{\mathbf{A}^j, j = 0, 1, 2, \dots\}$ is absolutely summable, converging to $(\mathbf{I} - \mathbf{A})^{-1}$, since $\mathbf{X}_t = \mu + \mathbf{A}(\mathbf{X}_{t-1} - \mu) + \xi_t = (\mathbf{I} - \mathbf{A})\mu + \mathbf{A}\mathbf{X}_{t-1} + \xi_t = (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \dots + \mathbf{A}^j)(\mathbf{I} - \mathbf{A})\mu + \mathbf{A}^{j+1}\mathbf{X}_{t-j-1} + \sum_{i=0}^j \mathbf{A}^i \xi_{t-i}$. 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 $\det(\mathbf{I} - z\mathbf{A}^{(1)} - \dots - z^q \mathbf{A}^{(q)}) \neq 0$ 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 (depending on the dimension of the problem) 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, $\varsigma_{(r)} \equiv (\xi_t \mid \rho_t = r)$, and matrices of deviation vectors, $\tilde{\chi}_{(r)} \equiv ((\mathbf{X}_{t-1} - \tilde{\mu}_{(r)}) \mid \rho_t = r)$ and $\chi_{(r)} \equiv ((\mathbf{X}_t - \tilde{\mu}_{(r)}) \mid \rho_t = r)$. Noting that $-\tilde{\xi}_t = \mathbf{A}_{(r)}(\mathbf{X}_{t-1} - \tilde{\mu}_{(r)}) - (\mathbf{X}_t - \tilde{\mu}_{(r)})$ from the VAR(1) form of the process, consider the following (dropping the references to regime r , for brevity).

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

where $\mathbf{C} = \vartheta (\chi \tilde{\chi}^T) (\tilde{\chi} \tilde{\chi}^T)^{-1}$ is an $m \times m$ sample regression matrix. Note that $\text{vec}(\mathbf{AB}) = (\mathbf{B}^T \otimes \mathbf{I}) \text{vec}(\mathbf{A})$. 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 $\mathbf{A}_{(t)}$. The exponent of the likelihood term can be expressed as

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

where $\varsigma_{(k)} \equiv (\xi_t \mid \rho_t = k, t > q)$ 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

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

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

$$\begin{aligned} \Psi_{(t)}^{-1} &= \left((\tilde{\chi}_{(t)} \tilde{\chi}_{(t)}^T) \otimes \Omega_{(t)}^{-1} \right) + \nu_{(t)} \mathbf{I}_{m^2 q} \\ \alpha_{(t)} &= \Psi_{(t)} \left[\left((\tilde{\chi}_{(t)} \tilde{\chi}_{(t)}^T) \otimes \Omega_{(t)}^{-1} \right) \text{vec} \mathbf{C}_{(t)} + \nu_{(t)} \text{vec} \mathbf{B}_{(t)} \right] \end{aligned}$$

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 $\mathbf{A}_{(t)}$

$$\begin{aligned} \therefore p(\mathbf{A}_{(1)}, \dots, \mathbf{A}_{(k)} | \rho, \lambda_{\neq A}, \mathbf{Y}) &\propto \left| \mathbf{V}_{(k)}^{-1} \right|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \tilde{\xi}^T \mathbf{V}_{(k)}^{-1} \tilde{\xi} \right\} \\ &\times \prod_r \mathbf{N}(\alpha_{(t)}, \Psi_{(t)}) \times g(\mathbf{A}_{(t)}). \end{aligned}$$

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

Direct acceptance/rejection can be used to independently generate the regression matrices, $\mathbf{A}_{(t)}$, $r \neq \rho_q$. Candidate $\mathbf{A}_{(t)}$ are generated from the matrix Normal densities until they fall within the stable region. The acceptance rate can however fall to very low levels as the dominant eigenvalue of \mathbf{A} approaches 1 and as the dimension of \mathbf{A} increases, causing the procedure to get “stuck” in the \mathbf{A} dimension. This is an area of current research. A possible approach would be to use a Metropolis-Hastings step with a suitably well constructed candidate density to enforce the stability constraint.

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

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 $\mathbf{A}_{(k)}$ using $\mathbf{A}_{(k)}^{(*)} \sim \mathbf{N}(\alpha_{(k)}, \Psi_{(k)})$ as the candidate generating density. If the candidate is stable, it is accepted, i.e. $\mathbf{A}_{(k)}^{(c+1)} = \mathbf{A}_{(k)}^{(*)}$, with probability

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

otherwise the previous value 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 $\Omega_{(k)}^{-1(c-1)}$

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 $\Omega_{(\rho)}$, $\rho_t \in \{1, \dots, K\}$. 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

$$p\left(\Omega_{(1)}^{-1}, \dots, \Omega_{(K)}^{-1}\right) = \prod_{r=1}^K \mathbf{W}_m\left(\eta_{(r)}, \mathbf{F}_r^{-1}\right) \times h\left(\Omega_{(1)}, \dots, \Omega_{(K)}\right),$$

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. $\omega_{22(1)} < \dots < \omega_{22(K)}$, where $\omega_{ii(r)}$ is the i -th diagonal element of $\Omega_{(r)}$. As before, define $\varsigma_{(r)} \equiv (\xi_t : \rho_t = r)$, $r \neq \rho_q$ and $\varsigma_{(k)} \equiv (\xi_t : \rho_t = k, t > q)$, $k = \rho_q$ (noting that the ξ_t are functions of the most recently generated $\mathbf{A}_{(r)}$).

The precision matrices can be generated jointly from

$$\begin{aligned} p\left(\Omega_{(1)}^{-1}, \dots, \Omega_{(K)}^{-1} \mid \rho, \lambda \neq \Omega, \mathbf{Y}\right) &\propto L(\mathbf{Y} \mid \rho, \lambda) \cdot p\left(\Omega_{(1)}^{-1}, \dots, \Omega_{(K)}^{-1}\right) \\ &\propto \left|\mathbf{V}_{(k)}^{-1}\right|^{\frac{1}{2}} \exp\left\{-\frac{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\{-\frac{1}{2}\text{vec}\varsigma_{(k)}^T \left(\mathbf{I}_{n_k-1} \otimes \Omega_{(k)}^{-1}\right) \text{vec}\varsigma_{(k)}\right\} \\ &\times \prod_{r \neq k} \left|\Omega_{(r)}^{-1}\right|^{\frac{n_r}{2}} \exp\left\{-\frac{1}{2}\text{vec}\varsigma_{(r)}^T \left(\mathbf{I}_{n_r} \otimes \Omega_{(r)}^{-1}\right) \text{vec}\varsigma_{(r)}\right\} \times p\left(\Omega_{(1)}^{-1}, \dots, \Omega_{(K)}^{-1}\right) \\ &\propto \left|\mathbf{V}_{(k)}^{-1}\right|^{\frac{1}{2}} \exp\left\{-\frac{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\{-\frac{1}{2}\text{tr}\left(\left(\varsigma_{(k)}\varsigma_{(k)}^T + \mathbf{F}_{(k)}\right)\Omega_{(k)}^{-1}\right)\right\} \\ &\times \prod_{i \neq k} \left|\Omega_{(i)}^{-1}\right|^{\frac{n_i+\eta_{(i)}-m-1}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left(\left(\varsigma_{(i)}\varsigma_{(i)}^T + \mathbf{F}_{(i)}\right)\Omega_{(i)}^{-1}\right)\right\} \times h\left(\Omega_{(1)}, \dots, \Omega_{(K)}\right) \\ \therefore p\left(\Omega_{(1)}^{-1}, \dots, \Omega_{(K)}^{-1} \mid \rho, \lambda \neq \Omega, \mathbf{Y}\right) & \\ &\propto \left|\mathbf{V}_{(k)}^{-1}\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\tilde{\xi}_q^T \mathbf{V}_{(k)}^{-1} \tilde{\xi}_q\right\} \times \mathbf{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} \mathbf{W}_m\left(n_r + \eta_{(r)}, \left(\varsigma_{(r)}\varsigma_{(r)}^T + \mathbf{F}_{(r)}\right)^{-1}\right) \times h\left(\Omega_{(1)}, \dots, \Omega_{(K)}\right). \end{aligned}$$

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 the Wishart candidate generating density $\Omega_{(k)}^{-1(*)} \sim W_m \left(n_k + \eta_{(k)} - 1, \left(\zeta_{(k)} \zeta_{(k)}^T + \mathbf{F}_{(k)} \right)^{-1} \right)$. The candidate would be accepted, i.e. $\Omega_{(k)}^{-1(c+1)} = \Omega_{(k)}^{-1(*)}$, with probability

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

otherwise the previous value would be retained, i.e. $\Omega_{(k)}^{-1(c+1)} = \Omega_{(k)}^{-1(c)}$. 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 $i > j$, $u_{ii}^2 \sim \chi_n^2$ ($i = j$) and $u_{ij} \sim N(0, 1)$ for $i < j$ (so that $\mathbf{U}^T\mathbf{U} \sim W_m(\eta, \mathbf{I}_m)$)

Generating the Transition Probabilities

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

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

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 $p_{ij}^{(*)} \sim \text{Beta}(n_{ij} + m_{ij} + 1, n_{ii} + m_{ii} + 1)$ 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 previous value 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 $p_{ii} = 1 - \sum p_{ij} < \prod(1 - p_{ij})$, 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 corresponding regime at all should be questioned

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

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

CARTER, C K and KOHN, R (1994) On Gibbs Sampling for State Space Models *Biometrika* 81, 3, 541-553

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

ENGLER, 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 *North American Actuarial Journal* 1, No 4, 49-75

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 Econometrics*, 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 (Brussels, Belgium)
- HARRIS G R (1995b) Low Frequency Statistical Interest Rate Models *Proceedings of the 5th AFIR International Colloquium* 2, 799-831 (Brussels, Belgium)
- 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
- KASS R E and RAFTERY, A E (1995) Bayes Factors *Journal of the American Statistical Association* 90, 430, 773-795
- KIM, C (1994) Dynamic Linear Models with Markov-Switching *Journal of Econometrics* 60, 1-22
- LUTKEPOHL H (1991) *Introduction to Multiple Time Series Analysis* Springer-Verlag
- LIU, J, WONG, W H and KONG, A (1994) Covariance Structure of the Gibbs Sampler with Applications to the Comparison of Estimators and Augmentation Schemes *Biometrika* 81, 27-40
- 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
- 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

GLEN HARRIS
33 Struen Marie St
Kareela NSW 2232
Australia

