SHORT COURSE IN BAYESIAN INFERENCE: THEORY AND METHODS: PART III EXAMPLES: NORMAL MODELS, REGRESSION MODELS THE KALMAN FILTER AND BAYESIAN STATE SPACE MODELS

DAVID A. STEPHENS

July 2002

Contents

Chapter 1 EXAMPLE 1. Bayesian Testing for Normal Data

1.1 Classical Two Sample Tests

The majority of statistical analyses of microarray data that concentrate on detecting differences in gene expression utilize a parametric hypothesis testing framework, and test only for a difference in mean level in the two samples. In this section we begin by describing the standard tests and some generalizations, but then look at non-parametric alternatives. For notational simplicity, we suppress dependence on gene index j and consider the two samples $(y_{11}, ..., y_{1n_1})$ and $(y_{21}, ..., y_{2n_2})$ with summary statistics the sample means $\overline{y}_1$ and $\overline{y}_2$ and sample variances s_1^2 and s_2^2 where

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2$$

1.1.1 Parametric Tests

The simplest classical two sample test is the **two sample t-test** under the assumption of **equal population variances** and two normally distributed populations, in which the test statistic, $t^{(1)}$, is defined as

$$t^{(1)} = \frac{\overline{y}_1 - \overline{y}_2}{s_P \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \tag{1.1}$$

where s_P^2 is the pooled sample variance

$$s_P^2 = \frac{1}{n_1 + n_2 - 2} \left[(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2 \right].$$

Under the null hypothesis of equal means in the two populations, the test statistic has a Student t distribution with $n_1 + n_2 - 2$ degrees of freedom. Naturally, much attention has been paid to this test as it captures the idea of a standardized difference in means measure, and is very straightforward to compute..

A natural extension to the simple two sample t-test is to allow for **differences in population variance** but retains the normality assumption. For this test, the test statistic $t^{(2)}$ is used, where

$$t^{(2)} = \frac{\overline{y}_1 - \overline{y}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \tag{1.2}$$

which is again a standardized difference in means but with a difference approach to standardization compared to (1.1). Under the null distribution of no difference in means in the two populations, the distribution of the test statistic is not available in closed form, but is, via Satterthwaite's approximation, approximately Student-t distributed with

$$\frac{\left(\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}\right)^2}{\dfrac{s_1^2/n_1}{n_1-1}+\dfrac{s_2^2/n_2}{n_2-1}}$$

degrees of freedom . If this quantity is large enough, a normal approximation may be used; this approach can also be justified from a likelihood-ratio test perspective Alternately, the Cochran and Cox approximation computes an approximate p-value for the test via the quantity

$$\frac{\dfrac{s_1^2}{n_1}c_1+\dfrac{s_2^2}{n_2}c_2}{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}$$

where c_1 and c_2 are the critical values at significance level p of the t-tests assuming n_1-1 and n_2-1 degrees of freedom respectively. In addition, for completeness, an F-test for equality of population variances may be carried out.

1.2 Bayesian Two Sample Tests

The Bayesian approach to two sample testing centres on the posterior probability of the two distribution hypothesis, in some manifestation, in comparison to the single distribution alternative. The canonical Bayesian calculation in a two sample testing situation compares the posterior probabilities as follows; if M_0 and M_1 represent the null and alternative models (that is, one population and two population models) respectively, then the posterior probability of M_0 in light of the data is given by

$$\pi_{M|Y}\left(M_0|y_1,y_2\right)=\frac{f_{Y|M}\left(y_1,y_2|M_0\right)\pi\left(M_0\right)}{f_{Y|M}\left(y_1,y_2|M_0\right)\pi\left(M_0\right)+f_{Y|M}\left(y_1,y_2|M_1\right)\pi\left(M_1\right)} \tag{1.3}$$

Variations on (1.3) that are often used are the posterior odds ratio

$$\frac{\pi\left(M_0|y_1,y_2\right)}{1-\pi\left(M_0|y_1,y_2\right)} \tag{1.4}$$

and the Bayes Factor or marginal likelihood ratio

$$\frac{f_{Y|M}\left(y_1,y_2|M_0\right)}{f_{Y|M}\left(y_1,y_2|M_1\right)}. \tag{1.5}$$

In a parametric context, where M_0 and M_1 are parameterized via parameters θ_0 and θ_1 respectively, it is also possible to formulate the Bayesian significance tests in terms of posterior quantities of interest (moments, quantiles, credible intervals). In fact, it transpires that we can mimic the types of test described in the section 1.1 within a fully Bayesian context. We again study both parametric and non-parametric tests.

1.2.1 Bayesian Parametric Tests

The most common parametric tests are based on the normal distribution. In the above notation, if we let $\theta_0=\left(\mu,\sigma^2\right)$ and $\theta_1=\left(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2\right)$ then the likelihoods for the two models (conditional on θ_0 and θ_1) are

$$\begin{aligned}
f_{Y|\theta,M}\left(y_1,y_2;\theta_0,M_0\right) &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(y_{ij}-\mu\right)^2\right\}\\
f_{Y|\theta,M}\left(y_1,y_2;\theta_1,M_1\right) &= \left(\frac{1}{2\pi\sigma_1^2}\right)^{n_1/2}\left(\frac{1}{2\pi\sigma_2^2}\right)^{n_2/2}\exp\left\{-\frac{1}{2\sigma_1^2}\sum_{j=1}^{n_i}\left(y_{1j}-\mu_1\right)^2-\frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}\left(y_{2j}-\mu_2\right)^2\right\}
\end{aligned} \tag{1.6}$$

The next stage in the Bayes computation is to specify prior distributions for θ_0 and θ_1; on the surface it would be useful to be able to make vague prior specifications, but care must be taken in this case, otherwise, for example the Bayes factor in (1.5) is indeterminate. For the moment we specify conjugate Normal/Inverse Gamma priors

$$\pi\left(\mu,\sigma^2\right)=\pi\left(\mu|\sigma^2\right)\pi\left(\sigma^2\right)=\left(\frac{1}{2\pi\lambda\sigma^2}\right)^{1/2}\exp\left\{-\frac{1}{2\lambda\sigma^2}\left(\mu-\phi\right)^2\right\}\frac{\left(\frac{\beta}{2}\right)^{\alpha/2}}{\Gamma\left(\frac{\alpha}{2}\right)}\left(\frac{1}{\sigma^2}\right)^{\frac{\alpha}{2}+1}\exp\left\{-\frac{\beta}{2\sigma^2}\right\}$$

with independent priors for the two population situation

$$\begin{aligned}\pi\left(\mu_i,\sigma_i\right) &= \pi\left(\mu_i|\sigma_i^2\right)\pi\left(\sigma_i^2\right)\\ &= \left(\frac{1}{2\pi\lambda_i\sigma_i^2}\right)^{1/2}\exp\left\{-\frac{1}{2\lambda_i\sigma_i^2}\left(\mu-\phi_i\right)^2\right\}\frac{\left(\frac{\beta_i}{2}\right)^{\alpha/2}}{\Gamma\left(\frac{\alpha_i}{2}\right)}\left(\frac{1}{\sigma_i^2}\right)^{\frac{\alpha_i}{2}+1}\exp\left\{-\frac{\beta_i}{2\sigma_i^2}\right\} \qquad i=1,2\end{aligned}$$

respectively, with hyperparameters ψ to be specified. We describe methods for appropriate choice of the hyperparameters in Section **??**.

With this model specification, routine calculations give that the posterior distributions for the unknown parameters are available in closed form

$$\pi_{\theta|Y,M}\left(\mu|y_1,y_2,\sigma^2,\varphi\right)\equiv N\left(\frac{n\lambda\overline{y}+\phi}{n\lambda+1},\frac{\lambda\sigma^2}{n\lambda+1}\right)$$

$$\pi_{\theta|Y,M}\left(\sigma^2|y_1,y_2,\varphi\right)\equiv IGamma\left(\frac{n+\alpha}{2},\frac{1}{2}\left[\left(n-1\right)s^2+\frac{n}{n\lambda+1}\left(\overline{y}-\phi\right)^2+\beta\right]\right)$$

giving, marginally, a Student-t distribution for μ, and with appropriate subscripts added for the parameters in the two population situation. Further, the marginal likelihoods, obtained as

$$f_{Y|M}\left(y_1,y_2|M_0\right)=\int f_{Y|\theta,M}\left(y_1,y_2;\mu,\sigma^2,M_0\right)\pi\left(\mu|\sigma^2\right)\pi\left(\sigma^2\right)d\mu d\sigma^2 \tag{1.7}$$

for M_0 and

$$f_{Y|M}\left(y_1,y_2|M_1\right)=f_{Y|M}\left(y_1|M_1\right)f_{Y|M}\left(y_2|M_1\right)$$

for M_1 are given by

$$f_{Y|M}\left(y_1,y_2|M_0\right)=\left(\frac{1}{\pi}\right)^{n/2}\left(\frac{1}{n\lambda+1}\right)^{1/2}\beta^{\alpha/2}\frac{\Gamma\left(\frac{n+\alpha}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)}\left\{\left(n-1\right)s^2+\frac{n}{n\lambda+1}\left(\overline{y}-\phi\right)^2+\beta\right\}^{-(n+\alpha)/2} \tag{1.8}$$

and hence the marginal likelihood ratio is

$$\frac{f_{Y|M}\left(y_1,y_2|M_0\right)}{f_{Y|M}\left(y_1,y_2|M_1\right)}=\frac{c_1c_2}{c}\frac{g_1(y_1)g_2(y_2)}{g\left(y_1,y_2\right)} \tag{1.9}$$

where

$$c=\left(\frac{1}{n\lambda+1}\right)^{1/2}\beta^{\alpha/2}\frac{\Gamma\left(\frac{n+\alpha}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)} \qquad g(y)=\left\{\left(n-1\right)s^2+\frac{n}{n\lambda+1}\left(\overline{y}-\phi\right)^2+\beta\right\}^{(n+\alpha)/2}$$

with equivalent definitions for c_i and g_i for $i = 1, 2$.

As in Section 1.1, it is possible to make a simplifying assumption of a common variance in the two populations. In this case, the posterior distributions of μ_1 and μ_2 are independent normal distributions conditional on σ^2, and

$$\pi_{\theta|Y,M}\left(\sigma^2|y_1, y_2, \varphi, M_1\right) \equiv IGamma\left(\frac{n_1+n_2+\alpha}{2}, \frac{1}{2}\left[(n_1+n_2-2)\,s_P^2 + \sum_{i=1}^{2}\frac{n_i}{n_i\lambda_i+1}\left(\overline{y}_i-\phi_i\right)^2+\beta\right]\right)$$

with a corresponding modified marginal likelihood ratio.

1.2.2 Bayesian Simulation-based Posterior Inference

For the comparison of gene expression in two samples, posterior summaries for the two sets of population parameters, or the marginal likelihood ratio in (1.7) can be used. For a direct comparison with two sample tests, consider, for example, the marginal posterior distribution of parameter $\delta = \mu_1 - \mu_2$ that is

$$\pi_{\theta|Y,M}\left(\delta|y_1, y_2, \varphi, M_1\right). \tag{1.10}$$

Now, this marginal posterior is not available readily in closed form, but can be sampled routinely because the joint posterior distribution $\pi_{\theta|Y,M}\left(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2|y_1, y_2, \varphi, M_1\right)$ and $\pi_{\theta|Y,M}\left(\mu_1, \mu_2, \sigma^2|y_1, y_2, \varphi, M_1\right)$ can be sampled directly by standard methods. Thus we may achieve full, simulation-based Bayesian inference.

Note that, in a posterior inference rather than model comparison problem, a non-informative and improper prior specification, such as Jeffreys' prior

$$\pi\left(\mu_i, \sigma_i\right) \propto \frac{1}{\sigma_i} \qquad \text{or} \qquad \pi\left(\mu_1, \mu_2, \sigma\right) \propto \frac{1}{\sigma}$$

can be used, leading to a joint posterior where

$$\begin{aligned} \pi_{\theta|Y,M}\left(\mu_i|y_i, \sigma_i^2\right) &\equiv N\left(\overline{y}_i, \frac{\sigma_i^2}{n_i}\right) \qquad i = 1, 2 \\ \pi_{\theta|Y,M}\left(\sigma^2|y_1, y_2\right) &\equiv IGamma\left(\frac{n+1}{2}, \frac{(n_i-1)\,s_i^2}{2}\right) \end{aligned} \tag{1.11}$$

for different population variances and

$$\begin{aligned} \pi_{\theta|Y,M}\left(\mu_i|y_i, \sigma^2\right) &\equiv N\left(\overline{y}_i, \frac{\sigma^2}{n_1+n_2}\right) \qquad i = 1, 2 \\ \pi_{\theta|Y,M}\left(\sigma^2|y_1, y_2\right) &\equiv IGamma\left(\frac{n_1+n_2+1}{2}, \frac{(n_1+n_2-2)\,s_P^2}{2}\right) \end{aligned} \tag{1.12}$$

for a common population variance. Hence, for a simulation based two-sample test, we may proceed by simulating a large number of variates via either (1.11) or (1.12), forming a corresponding sample from $\pi_{\theta|Y,M}\left(\delta|y_1, y_2, \varphi, M_1\right)$, and assessing a the support for the model $\delta = \mu_1 - \mu_2 = 0$ by means of a "Bayesian p-value" or sample posterior credible interval calculation.

Chapter 2 EXAMPLE 2: A simple generalized-linear model

When the data are not normally distributed, the analytical results found elsewhere in this report are not available. This example illusterates how the Bayesian inferential framework proceeds via simulation based computation.

2.1 Binary Regression: O-ring Failure data

The Space Shuttle Challenger mission failed catastrophically in 1986 when the craft exploded during the very early phase of its flight. The problem was traced and found to be the failure of all of the six O-rings that attached part of the propulsion mechanism to the body of the craft, due largely to the low air temperature at the time of launch. This example attempts a Bayesian anaysis of a simulated data set of a similar nature to the real O-ring failure data.

The following table contains the numbers of failures (out of six) that were observed during previous launches at a range of air temperatures.

Temperature (F) x_i	Failures (out of 6) y_i
26.57	4
34.94	2
43.67	4
47.63	2
52.00	1
58.48	0
62.03	0
66.66	0

The objective is to learn about the temperature dependence of the failure probability.

2.2 Likelihood assumptions

A natural model to use for the likelihood is the *Binomial distribution*, that is for the $n = 8$ data points,

$$f_{Y|\theta}(y_i;\theta_i) = \binom{n_i}{y_i}\theta_i^{y_i}\,(1-\theta_i)^{n_i-y_i} \qquad y_i = 0, 1, ..., n_i \tag{2.1}$$

where $n_i = 6$, and θ_i is the failure probability for each set of rings at temperature x_i. Note that θ is subscripted by i to acknowledge the dependence on temperature x_i, for $i = 1, ..., 8$. A common assumption is to capture this dependence via a logistic **link** function

$$\theta_i = \frac{\exp\{\beta_0 + \beta_1 x_i\}}{1+\exp\{\beta_0+\beta_1 x_i\}} \qquad \Leftrightarrow \qquad \log\left(\frac{\theta_i}{1-\theta_i}\right) = \beta_0 + \beta_1 x_i \qquad i = 1, ..., 8 \tag{2.2}$$

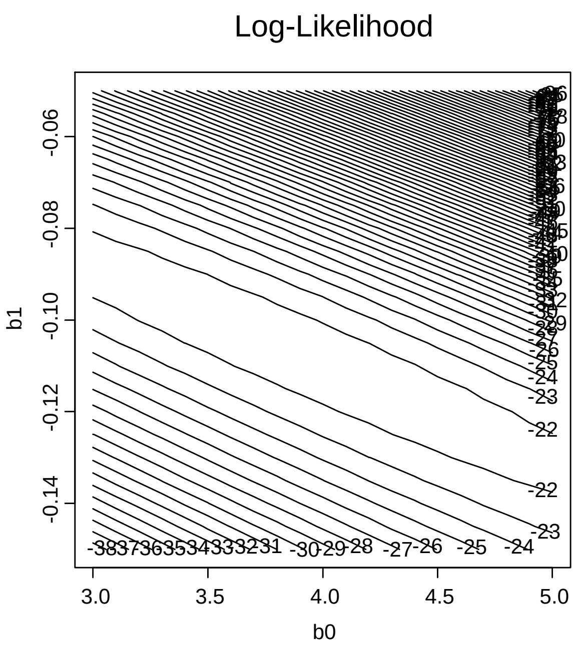

Figure 2.1: O-ring Data: Log-likelihood function

This is an example of a **Generalized Linear Model.** The two parameters (β_0, β_1) are the centre of inferential interest: specifically if we can deduce that β_1 is negative than zero then we conclude that there is a tendency for increases in temperature to decrease failure probability.

The resulting likelihood function from (2.1) and (2.2) is

$$f_{Y|\theta}(y; \beta_0, \beta_1) = \prod_{i=1}^{n} f_{Y|\theta}(y_i; \beta_0, \beta_1) = \prod_{i=1}^{n} \binom{n_i}{y_i} \left\{ \frac{\exp\{\beta_0 + \beta_1 x_i\}}{1 + \exp\{\beta_0 + \beta_1 x_i\}} \right\}^{y_i} \left\{ 1 - \frac{\exp\{\beta_0 + \beta_1 x_i\}}{1 + \exp\{\beta_0 + \beta_1 x_i\}} \right\}^{n_i - y_i} \tag{2.3}$$

For the data above, a contourplot of the (log) likelihood function for a grid of values of (β_0, β_1) can be obtained.

2.3 Posterior Analysis via Simulation

Under a non-informative prior

$$\pi(\beta_0, \beta_1) \propto \text{constant}$$

the joint posterior $\pi_{\theta|Y}(\beta_0, \beta_1|y)$ is proportional to (2.3). Analytic inference is not available, but simulation-based inference via Markov chain Monte Carlo (MCMC) can be obtained routinely.

The **Metropolis-Hastings** algorithm, using the acceptance probability in a proposed move from θ^{old} to θ^{new}

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(\theta^{new})}{\pi(\theta^{old})} \frac{q(\theta^{new}, \theta^{old})}{q(\theta^{old}, \theta^{new})} \right\}$$

for some **transition density** q, can be used, either on the joint posterior or within a Gibbs sampler context

$$\begin{array}{rcll} \pi_{\theta_0|Y,\theta_1}(\beta_0|y, \beta_1) & \propto & f_{Y|\theta}(y; \beta_0, \beta_1) & \quad \beta_0 \in \mathbb{R} \\ \pi_{\theta_1|Y,\theta_0}(\beta_1|y, \beta_0) & \propto & f_{Y|\theta}(y; \beta_0, \beta_1) & \quad \beta_1 \in \mathbb{R} \end{array}$$

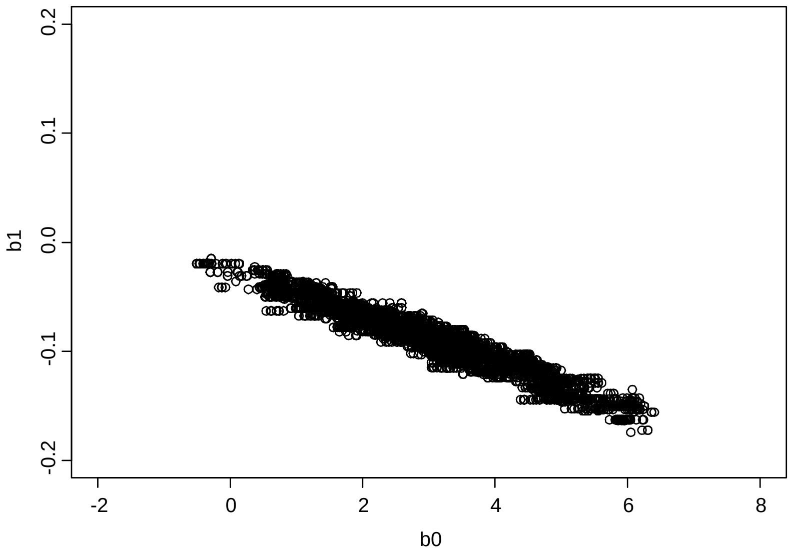

Figure 2.2: Metropolis-within-Gibbs algorithm for the O-ring data

Figure (**??**) depicts the joint posterior sample obtained via a Metropolis-within-Gibbs algorithm. The output from this chain can be further analyzed: figure (2.3) shows the trace and autocorrelation plots for the twoparameters. Clearly there is extreme dependence in the output.

This problem can be lessened by reparameterization that reducesthe correlation in the posterior between the two parameters. Setting

$$\begin{aligned}\varphi_0 &= m_0 + a_{00}\beta_0 + a_{01}\beta_1 \\ \varphi_1 &= m_1 + a_{10}\beta_0 + a_{11}\beta_1\end{aligned}$$

for appropriately chosen constants $(m_0, m_1, a_{00}, a_{10}, a_{01}, a_{11})$, the posterior can be framed in terms of the new parameters (φ_0, φ_1). After this re-parameterization, a joint **indepedence** Metropolis sampler centred at the origin can be used.

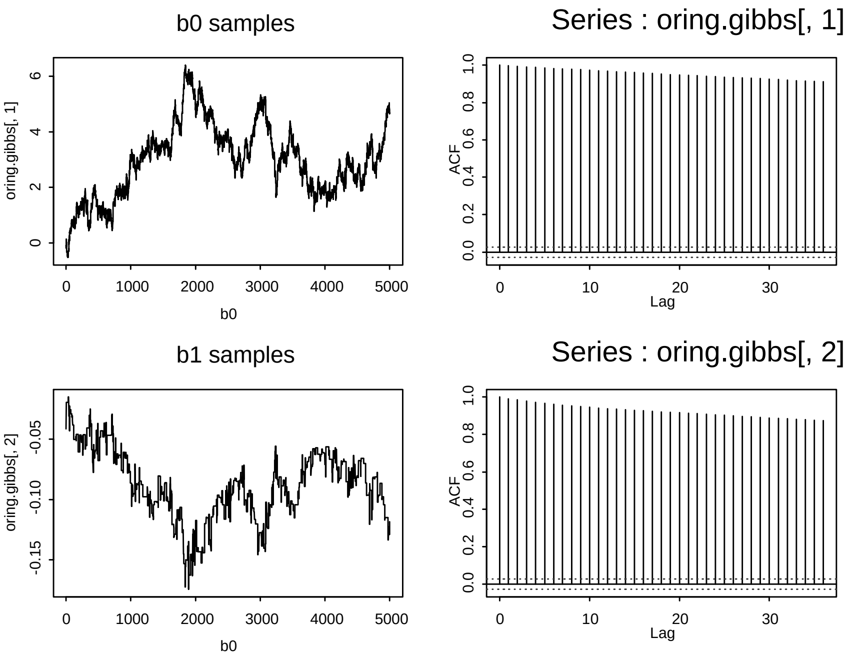

Figure 2.3: Output analysis from the Metropolis within Gibbs Markov chain.

Chapter 3 EXAMPLE 3: The Normal Linear Regression Model

A general **regression model** for data y can be written

$$y_t = f(\beta, t) + \varepsilon_t.$$

More specifically, we attempt to represent y using a linear model

$$y_t = X_t \beta + \varepsilon_t$$

where X_t is (in general) a $1 \times p$ vector of specified functions of t, and β is a $p \times 1$ parameter vector. In vector representation, where $y = (y_1, ..., y_T)$, we have

$$y = X\beta + \varepsilon \tag{3.1}$$

and a classical linear model. The precise form of **design matrix** X is at the moment left unspecified. Typically we take the random error terms $\{\varepsilon_t\}$ as independent and identically distributed Normal variables with variance σ^2, implying that the conditional distribution of the responses Y is multivariate normal

$$Y|X, \beta, \sigma^2 \sim N\left(X\beta, \sigma^2 I_T\right) \tag{3.2}$$

where now X is $T \times p$ where I_T is the $T \times T$ identity matrix.

For this model, the maximum likelihood/ordinary least squares estimates of β and σ^2 can be obtained straightforwardly: The likelihood is

$$\widehat{\beta}_{ML} = \left(X^T X\right)^{-1} X^T y \qquad \widehat{\sigma}^2 = \frac{1}{T-p}\left(y - \widehat{y}\right)^T \left(y - \widehat{y}\right)$$

for fitted values $\widehat{y} = X\widehat{\beta}_{ML} = X\left(X^T X\right)^{-1} X^T y$.

3.1 Bayesian Inference

In a Bayesian analysis of the model in (3.1) a joint prior distribution is specified for $\left(\beta, \sigma^2\right)$. Typically, a conjugate prior specification is used where

$$p\left(\beta|\sigma^2\right) \equiv N\left(v, \sigma^2 V\right) \qquad p\left(\sigma^2\right) \equiv InverseGamma\left(\frac{\alpha}{2}, \frac{\gamma}{2}\right) \tag{3.3}$$

(v is $p \times 1$, V is $p \times p$ positive definite and symmetric, all other parameters are scalars) and using this prior standard Bayesian calculations show that conditional on the data

$$p\left(\beta|y, \sigma^2\right) \equiv N\left(v^*, \sigma^2 V^*\right) \qquad p\left(\sigma^2|y\right) \equiv InverseGamma\left(\frac{T+\alpha}{2}, \frac{c+\gamma}{2}\right) \tag{3.4}$$

where

$$
\begin{aligned}
V^* &= \left(X^T X + V^{-1}\right)^{-1} \\
v^* &= \left(X^T X + V^{-1}\right)^{-1}\left(X^T y + V^{-1} v\right) \\
c &= y^T y + v^T V^{-1} v - \left(X^T y + V^{-1} v\right)^T \left(X^T X + V^{-1}\right)^{-1}\left(X^T y + V^{-1} v\right)
\end{aligned}
\tag{3.5}
$$

In regression modelling, it is usual to consider a centred parameterization for β so that $v = 0$, giving

$$
\begin{aligned}
v^* &= \left(X^T X + V^{-1}\right)^{-1} X^T y \\
c &= y^T y - y^T X\left(XX + V^{-1}\right)^{-1} X^T y = y^T\left(I_T - X\left(X^T X + V^{-1}\right)^{-1} X^T\right) y
\end{aligned}
$$

In a vague prior specification, $V^{-1} \to 0$ and $\alpha = \gamma = 0$ then

$$
c = y^T\left(I_T - X\left(X^T X\right)^{-1} X^T\right) y
$$

that is, the residual sum of squares, and

$$
p\left(\beta|y,\sigma^2\right) \equiv N\left(\left(X^T X\right)^{-1} X^T y, \sigma^2\left(X^T X\right)^{-1}\right) \qquad p\left(\sigma^2|y\right) \equiv InverseGamma\left(\frac{T}{2}, \frac{c}{2}\right) \tag{3.6}
$$

The critical quantity in our clustering procedure will be the marginal likelihood or prior predictive distribution

$$
p(y) = \int p\left(y|\beta,\sigma^2\right) p\left(\beta|\sigma^2\right) p\left(\sigma^2\right) d\beta d\sigma^2. \tag{3.7}
$$

Combining (3.2) and (3.3) gives that

$$
p(y) = \left(\frac{1}{\pi}\right)^{T/2} \frac{\gamma^{\alpha/2}\Gamma\left(\frac{T+\alpha}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)} \frac{|V^*|^{1/2}}{|V|^{1/2}} \frac{1}{\{c+\gamma\}^{(T+\alpha)/2}} \tag{3.8}
$$

which can be re-expressed

$$
p(y) = g\left(T,\alpha,\gamma\right)|V|^{-1/2} \frac{1}{|X^T X + V^{-1}|^{1/2}\{c+\gamma\}^{(T+\alpha)/2}} \tag{3.9}
$$

We note here that $V^{-1} \to 0$ in (3.9) leads to $p(y) \to 0$ and impropriety or indeterminacy.

For a collection of data sequences $y_1, ..., y_N$ (3.9) can be evaluated and used as the basis of a similarity measure. In fact, we shall use this marginal likelihood as the basis of a Bayesian hierarchical procedure that is readily computable for large data samples.

Chapter 4 State Space Modelling

Consider the model for observables $\{y_t\}$ specified via a *state-space representation*

$$\begin{array}{lll} y_t & = \theta_t + u_t & u_t \sim N(0, \sigma_u^2) \\ \theta_t & = \mu + \phi(\theta_{t-1} - \mu) + v_t & v_t \sim N(0, \sigma_v^2) \end{array} \quad t = 1, 2, ...T \tag{4.1}$$

$\{u_t\}, \{v_t\}$ are independent error terms, and where for completenes we may define $\theta_0 = \mu$. This model is the Gaussian AR(1) model; the first line is the *observation equation* and the second is the *state equation.* The θ_t are the *states*, which are unobserved parameters, and which, along with the model parameters $\mu, \phi, \sigma_u^2, \sigma_v^2$ are the objectives of our inference problem; we desire either maximum likelihood or Bayesian posterior inferences. Clearly, we may write the likelihood as

$$L\left(\theta_1, \theta_2, ..., \theta_T, \mu, \phi, \sigma_u^2, \sigma_v^2\right) = f(y_1, y_2, ..., y_T; \theta_1, \theta_2, ..., \theta_T, \mu, \phi, \sigma_u^2, \sigma_v^2)$$

which factorizes by the independence of the $\{u_t\}$ into

$$L\left(\theta_1, \theta_2, ..., \theta_T, \mu, \phi, \sigma_u^2, \sigma_v^2\right) = \left\{\prod_{t=1}^{T} f\left(y_t|\theta_t\right)\right\} f\left(\theta_1, \theta_2, ..., \theta_T|\mu, \phi, \sigma_u^2, \sigma_v^2\right)$$

and hence the posterior pdf for the parameters is

$$f\left(\theta_1, \theta_2, ..., \theta_T, \mu, \phi, \sigma_u^2, \sigma_v^2|y_1, y_2, ..., y_T\right) \propto \left\{\prod_{t=1}^{T} f\left(y_t|\theta_t\right)\right\} f\left(\theta_1, \theta_2, ..., \theta_T|\mu, \phi, \sigma_u^2, \sigma_v^2\right) p\left(\mu, \phi, \sigma_u^2, \sigma_v^2\right) \tag{4.2}$$

Note that, by the conditional independence structure imposed where y_t and y_{t+s} are conditionally independent, we may re-write the left-hand side of (2) as

$$f\left(\theta_1, \theta_2, ..., \theta_T, \mu, \phi, \sigma_u^2, \sigma_v^2|y_1, y_2, ..., y_T\right) = \left\{\prod_{t=1}^{T} f\left(\theta_t|\theta_1, ..., \theta_{t-1}, \mu, \phi, \sigma_u^2, \sigma_v^2, y_1, y_2, ..., y_t\right)\right\} f\left(\mu, \phi, \sigma_u^2, \sigma_v^2; y_1, y_2, ..., y_T\right) \tag{4.3}$$

4.0.1 Filtering

In general, the complex, high-dimensional nature of (2) has encouraged many sophisticated computational approaches. In the simple Gaussian case in (1), analytic results are straightforward to derive and have led to the technique of *Filtering*. This proceeds by formulating the inference problem sequentially as follows. For each time point t define z_t as the vector of observables up to and including time t, that is $z_t = (y_1, y_2, ..., y_t)$. Suppressing the dependence on the model parameters, we then consider pdfs

$$f(\theta_{t+1}|z_t) = \int f(\theta_{t+1}|\theta_t) f(\theta_t|z_t)\, d\theta_t \tag{4.4}$$

which is termed the *prediction density*, and

$$f(\theta_{t+1}|z_t, y_{t+1}) = \frac{f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}|z_t)}{f(y_{t+1}|z_t)} = \frac{f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}|z_t)}{\int f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}|z_t)\,d\theta_{t+1}} \tag{4.5}$$

which is termed the *filtering density*. Inspection of (3) reveals that

$$f\left(\theta_{t+1}|\theta_1, ..., \theta_t, \mu, \phi, \sigma_u^2, \sigma_v^2, y_1, y_2, ..., y_{t+1}\right) \equiv f(\theta_{t+1}|z_t, y_{t+1})$$

and thus recursive calculation via (5) will allow complete computation of (2). Note also that (4) plays an integral role in (5), so that, in fact recursive calculation via (4) and (5) give a method of inference for the state space model.

4.0.2 Gaussian Models: The Kalman Filter

In the case described in (1), analytic calculations via (4) and (5) are possible, revealing that in fact the pdfs in (4) and (5) are both Gaussian pdfs. We set up the recursion as follows; denote by μ_t and τ_t the mean and precision parameters in (4), that is

$$\theta_t|z_t \sim Normal\left(\mu_t, \tau_t^{-1}\right)$$

Then, recalling the identity

$$A(x-a)^2 + B(x-b)^2 = (A+B)\left(x - \frac{Aa+Bb}{A+B}\right)^2 + \frac{AB}{A+B}(a-b)^2$$

we may derive $f(\theta_{t+1}|z_t)$ as follows

$$\begin{aligned}
f(\theta_{t+1}|z_t) &= \int f(\theta_{t+1}|\theta_t)f(\theta_t|z_t)\,d\theta_t \\
&= \int \left(\frac{\tau_v}{2\pi}\right)^{1/2} \exp\left\{-\frac{\tau_v}{2}\left(\theta_{t+1} - \mu - \phi(\theta_t - \mu)\right)^2\right\} \left(\frac{\tau_t}{2\pi}\right)^{1/2} \exp\left\{-\frac{\tau_t}{2}\left(\theta_t - \mu_t\right)^2\right\} d\theta_t \\
&= \left(\frac{\tau_v}{2\pi}\right)^{1/2} \left(\frac{\tau_t}{2\pi}\right)^{1/2} \int \exp\left\{-\frac{1}{2}\left[\tau_v\left(\theta_{t+1} - \mu - \phi(\theta_t - \mu)\right)^2 + \tau_t\left(\theta_t - \mu_t\right)^2\right]\right\} d\theta_t \\
&= \left(\frac{\tau_v}{2\pi}\right)^{1/2} \left(\frac{\tau_t}{2\pi}\right)^{1/2} \int \exp\left\{-\frac{1}{2}\left[\tau_v\phi^2\left(\theta_t - \mu - (\theta_{t+1} - \mu)/\phi\right)^2 + \tau_t\left(\theta_t - \mu_t\right)^2\right]\right\} d\theta_t
\end{aligned} \tag{4.6}$$

Now, using the identity above, we have

$$\tau_v\phi^2\left(\theta_t - \mu - (\theta_{t+1} - \mu)/\phi\right)^2 + \tau_t\left(\theta_t - \mu_t\right)^2 = \left(\tau_v\phi^2 + \tau_t\right)\left(\theta_t - m_{t+1}\right)^2 + \tau\prime_t\left(\theta_{t+1} - \mu\prime_t\right)^2$$

where

$$m_{t+1} = \frac{\tau_v\phi\left(\theta_{t+1} - (1-\phi)\mu\right) + \tau_t\mu_t}{\tau_v\phi^2 + \tau_t} \qquad \mu\prime_t = \mu + \phi(\mu_t - \mu) \qquad \tau\prime_t = \frac{\tau_v\tau_t}{\tau_v\phi^2 + \tau_t} \tag{4.7}$$

and thus integrating out θ_t in (6) we have

$$\begin{aligned}
f(\theta_{t+1}|z_t) &= \left(\frac{\tau_v}{2\pi}\right)^{1/2} \left(\frac{\tau_t}{2\pi}\right)^{1/2} \int \exp\left\{-\frac{1}{2}\left[\tau_v\phi^2\left(\theta_t - \mu - (\theta_{t+1} - \mu)/\phi\right)^2 + \tau_t\left(\theta_t - \mu_t\right)^2\right]\right\} d\theta_t \\
&= \left(\frac{\tau_v}{2\pi}\right)^{1/2} \left(\frac{\tau_t}{2\pi}\right)^{1/2} \exp\left\{-\frac{\tau\prime_t}{2}\left(\theta_{t+1} - \mu\prime_t\right)^2\right\} \int \exp\left\{-\frac{\left(\tau_v\phi^2 + \tau_t\right)}{2}\left(\theta_t - m_{t+1}\right)^2\right\} d\theta_t \\
&= \left(\frac{\tau\prime_t}{2\pi}\right)^{1/2} \exp\left\{-\frac{\tau\prime_t}{2}\left(\theta_{t+1} - \mu\prime_t\right)^2\right\}
\end{aligned} \tag{4.8}$$

as the remaining integrand is merely a Normal pdf, and other constants cancel. Hence

$$\theta_{t+1}|z_t \sim Normal\left(\mu\prime_t, \tau\prime_t^{-1}\right) \tag{4.9}$$

Finally, for (5), we have

$$\begin{aligned} f(\theta_{t+1}|z_t, y_{t+1}) &\propto f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}|z_t) \\ &= \exp\left\{-\frac{\tau_u}{2}\left(y_{t+1}-\theta_{t+1}\right)^2\right\}\exp\left\{-\frac{\tau\prime_t}{2}\left(\theta_{t+1}-\mu\prime_t\right)^2\right\} \end{aligned} \tag{4.10}$$

from which we can deduce using the identity that

$$\theta_{t+1}|z_t, y_{t+1} \sim Normal\left(\mu_{t+1}, \tau_{t+1}^{-1}\right)$$

where

$$\mu_{t+1} = \frac{\tau_u y_{t+1} + \tau\prime_{t+1}\mu\prime_{t+1}}{\tau_u + \tau\prime_{t+1}} \qquad \tau_{t+1} = \frac{\tau_u \tau\prime_{t+1}}{\tau_u + \tau\prime_{t+1}} \tag{4.11}$$

Finally, we have the a data prediction density from (5)

$$f(y_{t+1}|z_t) = \int f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}|z_t)\, d\theta_{t+1}$$

which is the normalizing constant in (11), that is

$$y_{t+1}|z_t \sim Normal\left(\mu\prime_{t+1}, \tau_{t+1}^{-1}\right) \tag{4.12}$$

Chapter 5 Stochastic Volatility: Introduction and Model Formulation

The homogeneous stochastic volatility model in discrete time can be formulated as follows: for data $y_1, ..., y_n$

$$
\begin{aligned}
y_t &= \exp\left\{\frac{h_t}{2}\right\} z_t \qquad && z_t \sim N(0,1) \text{ i.i.d} \qquad && \text{(I)} \\
h_1 &= \mu + u_1 \qquad && u_1 \sim N\left(0, \frac{\sigma^2}{(1-\phi^2)}\right) \text{ i.i.d} && \\
h_t &= \mu + \phi\,(h_{t-1} - \mu) + u_t \qquad && u_t \sim N(0, \sigma^2) \text{ i.i.d} \qquad && \text{(II)}
\end{aligned}
$$

and where, for stationarity of the state equation (II) we have the constraint $|\phi| < 1$.

5.1 Likelihood calculation

The full likelihood under this model is given by

$$
\begin{aligned}
L(\mu, \phi, \sigma^2, h_1, ..., h_n) &= \left\{\prod_{t=1}^{n} \left(\frac{1}{2\pi}\right)^{1/2} \exp\{-h_t\} \exp\left\{-\frac{y_t^2}{2e^{h_t}}\right\}\right\} \\
&\times \left(\frac{(1-\phi^2)}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{(1-\phi^2)}{2\sigma^2}(h_1-\mu)^2\right\} \\
&\times \left\{\prod_{t=2}^{n} \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{1}{2\sigma^2}\left(h_t - \mu - \phi\,(h_{t-1}-\mu)\right)^2\right\}\right\}
\end{aligned} \tag{5.1}
$$

5.1.1 Prior specification

The joint prior distribution for the parameters $\left(\mu, \phi, \sigma^2\right)$ is to be expressed as follows:

$$p\left(\mu, \phi, \sigma^2\right) = p\left(\mu\right) p\left(\phi\right) p\left(\sigma^2\right)$$

where, *a priori*,

$$
\begin{array}{lll}
\mu & \sim N\left(0,\sigma_\mu^2\right) & p\left(\mu\right) = \left(\frac{1}{2\pi\sigma_\mu^2}\right)^{1/2}\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\} \\
\frac{\phi+1}{2} & \sim Beta\left(\alpha_\phi,\beta_\phi\right) & p\left(\phi\right) = \left(\frac{1}{2}\right)^{\alpha_\phi+\beta_\phi-1}\frac{\Gamma\left(\alpha_\phi+\beta_\phi\right)}{\Gamma\left(\alpha_\phi\right)\Gamma\left(\alpha_\phi\right)}\left(1+\phi\right)^{\alpha_\phi-1}\left(1-\phi\right)^{\beta_\phi-1} \\
\sigma^2 & \sim InverseGamma\left(\alpha_\sigma,\beta_\sigma\right) & p\left(\sigma^2\right) = \frac{\beta_\sigma^{\alpha_\sigma}}{\Gamma\left(\alpha_\sigma\right)}\left(\frac{1}{\sigma^2}\right)^{\alpha_\sigma+1}\exp\left\{-\frac{\beta_\sigma}{\sigma^2}\right\}
\end{array}
\tag{5.2}
$$

Typically, following Kim Shephard and Chib (199?) - denoted KSC - in part, we might choose $\sigma_\mu = 5$, $\alpha_\phi = 2.5, \beta_\phi = 0.25$, and $\alpha_\sigma = \beta_\sigma = 1$. Therefore, combining likelihood and prior, we can make inference via the full posterior distribution.

5.2 Markov Chain Monte Carlo: full conditional distributions

Samples from the joint posterior distribution are to be obtained using a Gibbs-sampler type Markov chain Monte Carlo algorithm, where the parameters are sampled iteratively with updating. Three parameterizations are used for the state space equation; the required full conditionals are found to be of the following forms (as given by KSC - see appendix for full details).

The full conditionals of μ and states $h_1,...,h_n$ are identical under each parameterization, that is, if $m_t = h_t - \mu$ and $d_t = h_t - \phi h_{t-1}$ and

$$\bar{d} = \frac{1}{(n-1)}\sum_{t=2}^{n} d_t$$

then

$$\mu\,|\,- \equiv N\left(\frac{(1+\phi)\,h_1 + (n-1)\bar{d}}{(1+\phi)+(1-\phi)\,(n-1)}, \left(\frac{(1-\phi^2)+(1-\phi)^2\,(n-1)}{\sigma^2} + \frac{1}{\sigma_\mu^2}\right)^{-1}\right)$$

$$p(h_1\,|\,-) \propto \exp\left\{-\frac{h_1}{2}\right\}\exp\left\{-\frac{y_1^2}{2e^{h_1}}\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(h_1-\mu-\phi\left(h_2-\mu\right)\right)^2\right\}$$

$$p(h_t\,|\,-) \propto \exp\left\{-\frac{h_t}{2}\right\}\exp\left\{-\frac{y_t^2}{2e^{h_t}}\right\}\exp\left\{-\frac{(1+\phi^2)}{2\sigma^2}\left(h_t - \frac{(1-\phi)^2\mu+\phi\left(h_{t-1}+h_{t+1}\right)}{(1+\phi^2)}\right)^2\right\}$$

$$p(h_n\,|\,-) \propto \exp\left\{-\frac{h_n}{2}\right\}\exp\left\{-\frac{y_n^2}{2e^{h_n}}\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(h_n-\mu-\phi\left(h_{n-1}-\mu\right)\right)^2\right\}$$

where parameters $\left(\phi,\sigma^2\right)$ are either sampled directly, or formed as functions of the sampled parameters.

Parameterization 1: $\left(\mu,\phi,\sigma^2\right)$

$$\sigma^2\,|\,- \equiv InverseGamma\left(\frac{n}{2}+\alpha_\sigma, \frac{(1-\phi^2)}{2}\left(h_1-\mu\right)^2 + \frac{1}{2}\sum_{t=2}^{n}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)^2+\beta_\sigma\right)$$

$$p(\phi\,|\,-) \propto (1+\phi)^{\alpha_\phi+1/2-1}\left(1-\phi\right)^{\beta_\phi+1/2-1}\exp\left\{-\frac{m_1^2}{2\sigma^2}\left(1-\phi^2\right)\right\}\exp\left\{-\frac{M_1}{2\sigma^2}\left(\phi-M_2\right)^2\right\}$$

Parameterization 2: (μ, ψ, σ^2) Under reparameterization 2, where

$$\psi = \frac{1+\phi}{1-\phi}$$

the full conditional for ψ is

$$p(\psi \,|\, -) \propto \left(1 + \frac{\psi-1}{\psi+1}\right)^{\alpha_\phi+1/2-1} \left(1 - \frac{\psi-1}{\psi+1}\right)^{\beta_\phi+1/2-1}$$
$$\times \exp\left\{-\frac{m_1^2}{2\sigma^2}\left(1 - \left(\frac{\psi-1}{\psi+1}\right)^2\right)\right\} \exp\left\{-\frac{M_1}{2\sigma^2}\left(\frac{\psi-1}{\psi+1} - M_2\right)^2\right\} \frac{1}{(\psi+1)^2}$$

Parameterization 3: (μ, θ, τ^2) Under reparameterization 3, where

$$\theta = \left(\frac{1+\phi}{1-\phi}\right)\tau^2 \qquad \tau^2 = \sigma^2$$

the full conditionals

$$p(\tau^2 \,|\, -) \propto \left(\frac{1}{\tau^2}\right)^{n/2+\alpha_\sigma+1} \left(1 + \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)\right)^{\alpha_\phi-1} \left(1 - \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)\right)^{\beta_\phi-1} \left(1 - \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)^2\right)^{1/2}$$
$$\times \exp\left\{-\frac{1}{2\tau^2}\left[\left(1 - \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)^2\right) m_1^2 + \textstyle\sum_{t=2}^n \left(m_t - \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right) m_{t-1}\right)^2 + 2\beta_\sigma\right]\right\}$$
$$\times \frac{\tau^2}{(\theta+\tau^2)^2}$$

$$p(\theta \,|\, -) \propto \left(1 - \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)^2\right)^{1/2} \exp\left\{-\frac{1}{2\tau^2}\left[\left(1 - \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)^2\right) m_1^2 + \sum_{t=2}^n \left(m_t - \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right) m_{t-1}\right)^2\right]\right\}$$
$$\times \theta^{\alpha_\phi-1}\left(\frac{1}{\theta+\tau^2}\right)^{\alpha_\phi+\beta_\phi}$$

For each parameterization, the full conditionals will either be sampled directly where possible, or using a local or independent MH step. Details are given in Appendix 3, but broadly the local step will be based on a symmetric (normal) proposal for parameters on $\mathbb{R}$, and using a Gamma proposal for parameters on $\mathbb{R}^+$. Parameters in the MH proposal are chosen to optimize mixing of the chain; typical values used are $\lambda = 0.1$ and $\alpha = 500$.

5.3 Discrete time Changepoint SV Models

To relax the assumption of homogeneity of the SV process, we allow for an unknown changepoints. Assume that the entire data sequence can be segmented into $K+1$ homogeneous regions, that is, for integers $0 = r_0 < r_1 < r_2 < ... < r_K < n = r_{K+1}$, defining segments labelled$1, 2, ..., K+1$ and running from $r_{k-1} + 1 \le t \le r_k$ respectively, the observed data in segment k follow an SV model with parameters

$(\mu_k, \phi_k, \sigma_k^2)$ or equivalent. Conditional on the (integer-valued) changepoint positions, the full conditional distribution for the $K+1$ sets of SV parameters are of the same form as described above, but with data indices restricted to the segment range.

For the multiple changepoint model, it is also necessary to update the changepoint position parameters during the MCMC cycle; we can update each changepoint position in turn, conditional on the other $K-1$ parameters. Given all other changepoint positions, because of the changepoint model assumptions, the full conditional posterior of r_k depends only on r_{k-1} and r_{k+1}, segment parameters $(\mu_k, \phi_k, \sigma_k^2)$ and $(\mu_{k+1}, \phi_{k+1}, \sigma_{k+1}^2)$, and states with indices in the range $r_{k-1}+1, ..., r_{k+1}$; all other terms are independent of r_k Specifically

$$p(r_k \mid -) \propto l\left(h_{r_{k-1}+1}, ..., h_{r_k}; \mu_k, \phi_k, \sigma_k^2\right) l\left(h_{r_k+1}, ..., h_{r_{k+1}}; \mu_{k+1}, \phi_{k+1}, \sigma_{k+1}^2\right) p(r_k \mid r_{k-1}, r_{k+1})$$

where $p(r_k \mid r_{k-1}, r_{k+1})$ is the (conditional) prior which we take to be uniform

$$p(r_k \mid r_{k-1}, r_{k+1}) = \frac{1}{(r_{k+1} - r_{k-1} - 1)} \qquad r_{k-1}+1 \le r_k < r_{k+1},$$

although it may be advantageous to restrict the conditional range for r_k so that no segment is shorter than some minimum length. Note that the full conditional for r_k does not depend on the data y in any way, but does depend on some of the (unobserved) states.

The full conditional for r_k therefore takes the form

$$\begin{aligned} p(r_k \mid -) \quad \propto & \left(\frac{1}{\sigma_k^2}\right)^{\frac{(r_k - r_{k-1})}{2}+\alpha+1} \exp\left\{-\frac{\left(1-\phi_k^2\right)}{2\sigma_k^2}\left(h_{r_{k-1}+1} - \mu_k\right)^2\right\} \\ & \times \left\{\prod_{t=r_{k-1}+2}^{r_k} \exp\left\{-\frac{1}{2\sigma_k^2}\left(h_t - \mu_k - \phi_k\left(h_{t-1} - \mu_k\right)\right)^2\right\}\right\} \\ & \times \left(\frac{1}{\sigma_{k+1}^2}\right)^{\frac{(r_{k+1} - r_k)}{2}+\alpha+1} \exp\left\{-\frac{\left(1-\phi_{k+1}^2\right)}{2\sigma_{k+1}^2}\left(h_{r_k+1} - \mu_{k+1}\right)^2\right\} \\ & \times \left\{\prod_{t=r_k+1}^{r_{k+1}} \exp\left\{-\frac{1}{2\sigma_{k+1}^2}\left(h_t - \mu_{k+1} - \phi_k\left(h_{t-1} - \mu_{k+1}\right)\right)^2\right\}\right\} \end{aligned}$$

for $r_{k-1}+1 \le r_k < r_{k+1}$. This is a discrete distribution on this set of indices, and hence can be evaluated exactly and sampled using cdf inversion, or sampled using a Metropolis step with candidate changepoint sampled uniformly on (r_{k-1}, r_{k+1}).

5.4 Variable Dimension Models

In section 2 the K changepoint model was described, assuming that there were precisely $K+1$ homogeneous segments in the hidden process. In reality, K is unknown (typically a Poisson prior distribution with some parameter λ_K truncated to $[0, K_{\max}]$ is used to express a priori opinion about K) and hence variable dimension (reversible jump) MCMC algorithms must be utilized.

Essentially, variable dimension MCMC is identical to fixed dimension MCMC with additional moves to allow changes of dimension. In changepoint problems, the simplest dimension changing moves (that make proposals in models with one more or one fewer changepoint) - *Birth* and *Death* moves - are easy to formulate. Suppose that, after a number of iterations, the current model has K changepoints. A Birth or Death move is selected and a candidate point (in the $K+1$ or $K-1$) dimensional model is proposed, and accepted or rejected with some probability, detailed below. The birth and death moves are formulated in complimentary pairs.

For a generic acceptance probability calculation for moves between two real spaces, consider current point $x \in \mathbb{R}^{d_1}$ and a proposed move to point $y \in \mathbb{R}^{d_1+d_2}$. Suppose that y is generated as a function of x and random variables $u \sim q(.)$ where $u \in \mathbb{R}^{d_2}$ as

$$y = g(x, u)$$

where g is 1-1, and q is some proposal density. Then move to y is accepted with probability

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \frac{1}{q(u)} \frac{p(d_1 + d_2 \to d_1)}{p(d_1 \to d_1 + d_2)} \left| \frac{\partial y}{\partial(x, u)} \right| \right\}$$

where $p(d_1 \to d_1 + d_2)$ and $p(d_1 + d_2 \to d_1)$ are the probabilities of choosing moves from $d_1$to $d_1 + d_2$ and from $d_1 + d_2$ to d_1 dimensions respectively, and

$$\left| \frac{\partial y}{\partial(x, u)} \right|$$

is the Jacobian of the transformation from (x, u) to y. If $y = (y_1, ..., y_{d_i+d_2})$ then the (i, j)th element of the Jacobian matrix is

$$\left| \frac{\partial y_i}{\partial x_j} \right| \qquad 1 \leq j \leq d_1 \qquad \text{and} \qquad \left| \frac{\partial y_i}{\partial u_{j-d_1}} \right| \qquad d_1 + 1 \leq j \leq d_2$$

For computational convenience, we use the (μ, ψ, σ^2) parameterization in the variable dimension changepoint formulation.

Birth Move: For birth moves, the number of changepoints is to change from K to $K+1$, and the number of changepoint segments from $K+1$ to $K+2$. The move involves the generation of a new changepoint position, and a new set of changepoint parameters. To ensure reversibility of the birth move, we must consider the complimentary death move, and propose the move to the large model as follows

(1) sample a segment uniformly from the $K+1$ available; suppose segment k is selected.

(2) sample new changepoint position uniformly on the set of integers $(r_{k-1} + 1, ..., r_k - 1)$; suppose r is selected

(3) sample random variates u, v, w independently according to the following distributions

$$u \sim N(0, \lambda^2) \qquad v, w \sim Gamma(\alpha, \alpha)$$

(4) transform the original parameters from segment k, that is $(\mu_k, \psi_k, \sigma_k^2)$ and u, v, w to obtain the parameters for the two new segments as follows

$$\left.\begin{aligned} \mu_{(1)} &= \mu_k \\ \mu_{(2)} &= \mu_k + u \\ \\ \psi_{(1)} &= \psi_k \\ \psi_{(2)} &= v\psi_k \\ \\ \sigma^2_{(1)} &= \sigma^2_k \\ \sigma^2_{(2)} &= w\sigma^2_k \end{aligned}\right\} \iff \left\{\begin{aligned} \mu_k &= \mu_{(2)} - \mu_{(1)} \\ \\ \psi_k &= \psi_{(2)}/\psi_{(1)} \\ \\ \sigma^2_k &= \sigma^2_{(2)}/\sigma^2_{(1)} \end{aligned}\right. .$$

Each of these steps makes a contribution to the transition acceptance probability; generically we need to compute $q(x, y)$ (and its compliment $q(y, x)$). Term by term, we have contributions from (1) to (3) as follows

$$q(x, y) = \frac{1}{K+1} \times \frac{1}{r_k - r_{k-1} - 1} \times q(u)q(v)q(w)$$

that appear in the denominator of the transition probability. For the birth move, the Jacobian term is given by

$$\begin{vmatrix} \frac{\partial \mu_{(1)}}{\partial \mu_k} & \frac{\partial \mu_{(1)}}{\partial u} & \frac{\partial \mu_{(1)}}{\partial \psi_k} & \frac{\partial \mu_{(1)}}{\partial v} & \frac{\partial \mu_{(1)}}{\partial \sigma_k^2} & \frac{\partial \mu_{(1)}}{\partial w} \\ \frac{\partial \mu_{(2)}}{\partial \mu_k} & \frac{\partial \mu_{(2)}}{\partial u} & \frac{\partial \mu_{(2)}}{\partial \psi_k} & \frac{\partial \mu_{(2)}}{\partial v} & \frac{\partial \mu_{(2)}}{\partial \sigma_k^2} & \frac{\partial \mu_{(2)}}{\partial w} \\ \frac{\partial \psi_{(1)}}{\partial \mu_k} & \frac{\partial \psi_{(1)}}{\partial u} & \frac{\partial \psi_{(1)}}{\partial \psi_k} & \frac{\partial \psi_{(1)}}{\partial v} & \frac{\partial \psi_{(1)}}{\partial \sigma_k^2} & \frac{\partial \psi_{(1)}}{\partial w} \\ \frac{\partial \psi_{(2)}}{\partial \mu_k} & \frac{\partial \psi_{(2)}}{\partial u} & \frac{\partial \psi_{(2)}}{\partial \psi_k} & \frac{\partial \psi_{(2)}}{\partial v} & \frac{\partial \psi_{(2)}}{\partial \sigma_k^2} & \frac{\partial \psi_{(2)}}{\partial w} \\ \frac{\partial \sigma_{(1)}^2}{\partial \mu_k} & \frac{\partial \sigma_{(1)}^2}{\partial u} & \frac{\partial \sigma_{(1)}^2}{\partial \psi_k} & \frac{\partial \sigma_{(1)}^2}{\partial v} & \frac{\partial \sigma_{(1)}^2}{\partial \sigma_k^2} & \frac{\partial \sigma_{(1)}^2}{\partial w} \\ \frac{\partial \sigma_{(2)}^2}{\partial \mu_k} & \frac{\partial \sigma_{(2)}^2}{\partial u} & \frac{\partial \sigma_{(2)}^2}{\partial \psi_k} & \frac{\partial \sigma_{(2)}^2}{\partial v} & \frac{\partial \sigma_{(2)}^2}{\partial \sigma_k^2} & \frac{\partial \sigma_{(2)}^2}{\partial w} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \psi_k & v & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \sigma_k^2 & w \end{vmatrix} = vw$$

Death Move: For death moves, the number of changepoints is to change from $K+1$ to K, and the number of changepoint segments from $K+2$ to $K+2$. The move involves the removal of an existing changepoint position, and the combination of the two sets of parameters. To ensure reversibility of the birth move, we must consider the complimentary death move, and propose the move to the smaller model as follows

(1) sample a changepoint uniformly from the $K+1$ available; suppose changepoint k is selected.

(2) reverse transform the two sets of original parameters from segment k (denoted by subscript (1)) and $k+1$ (denoted by subscript (2)) that is

$$\mu_k = \mu_{(2)} - \mu_{(1)} \qquad \psi_k = \psi_{(2)}/\psi_{(1)} \qquad \sigma_k^2 = \sigma_{(1)}^2/\sigma_{(2)}^2.$$

in accordance with the birth move above

These two steps make a contribution to the transition acceptance probability The contributions from (1) is

$$q(x, y) = \frac{1}{K+1}$$

that appear in the denominator of the transition probability. For the death move, the Jacobian term is given by

$$\frac{1}{\psi_{(1)}\sigma_{(1)}^2}$$

Therefore, the acceptance probability for a birth move

$$\left(\mu_k, \psi_k, \sigma_k^2; r_{k-1}, r_k, K\right) \quad \rightarrow \quad \left(\mu_{(1)}, \psi_{(1)}, \sigma_{(1)}^2, \mu_{(2)}, \psi_{(2)}, \sigma_{(2)}^2; r_{k-1}, r, r_k, K+1\right)$$

is given by the minimum of 1 and

$$\frac{\pi(\mu_{(1)}, \psi_{(1)}, \sigma_{(1)}^2, \mu_{(2)}, \psi_{(2)}, \sigma_{(2)}^2; r_{k-1}, r, r_k, K+1)}{\pi(\mu_k, \psi_k, \sigma_k^2; r_{k-1}, r_k, K)} \times \frac{\frac{1}{K+1}}{\frac{1}{K+1} \times \frac{1}{r_k - r_{k-1} - 1} \times q(u)q(v)q(w)} \times vw$$

where

$$\pi(\mu_k, \psi_k, \sigma_k^2; r_{k-1}, r_k, K) \propto l\left(h_{r_{k-1}+1}, ..., h_{r_k}; \mu_k, \phi_k, \sigma_k^2\right) p(\mu_k, \psi_k, \sigma_k^2) p(k)$$

and

$$\begin{aligned}\pi(\mu_{(1)}, \psi_{(1)}, \sigma_{(1)}^2, \mu_{(2)}, \psi_{(2)}, \sigma_{(2)}^2; r_{k-1}, r, r_k, K+1) \quad &\propto l\left(h_{r_{k-1}+1}, ..., h_r; \mu_{(1)}, \psi_{(1)}, \sigma_{(1)}^2\right) l\left(h_{r+1}, ..., h_{r_{k+1}}; \mu_{(2)}, \psi_{(2)}, \sigma_{(2)}^2\right) \\ &\times p(\mu_{(1)}, \psi_{(1)}, \sigma_{(1)}^2)\, p(\mu_{(2)}, \psi_{(2)}, \sigma_{(2)}^2)\, p(r|\, r_{k-1}, r_k) p(k+1)\end{aligned}$$

and cancelling terms, we have that $\alpha(x, y)$ is the minimum of 1 and

$$\begin{aligned}&\frac{l\left(h_{r_{k-1}+1}, ..., h_r; \mu_{(1)}, \psi_{(1)}, \sigma_{(1)}^2\right) l\left(h_{r+1}, ..., h_{r_{k+1}}; \mu_{(2)}, \psi_{(2)}, \sigma_{(2)}^2\right)}{l\left(h_{r_{k-1}+1}, ..., h_{r_k}; \mu_k, \phi_k, \sigma_k^2\right)} \times \frac{p(\mu_{(1)}, \psi_{(1)}, \sigma_{(1)}^2)\, p(\mu_{(2)}, \psi_{(2)}, \sigma_{(2)}^2)}{p(\mu_k, \psi_k, \sigma_k^2)} \\ &\qquad \times \frac{1}{q(u)q(v)q(w)} \times \frac{p(k+1)}{p(k)} \times vw\end{aligned}$$

The acceptance probability for the reverse (death) move is given essentially the reciprocal of this expression. Note that care must be taken to ensure that all constants that appear in this expression (via the likelihood, prior, or Metropolis proposal density) must be included.

An algorithm based on this variable dimension MCMC approach appears to work well in practice. We have also experimented with other algorithms using other proposal mechanisms (see Appendix 5) which also appear to work adequately.

5.5 GARCH(1,1): Introduction and Model Formulation

The GARCH(1,1) model is formulated as follows: for data $y_1, ..., y_n$

$$\begin{aligned} y_t &= h_t z_t & z_t &\sim N(0,1) \text{ i.i.d} & \text{(I)} \\ h_t^2 &= \mu + \alpha h_{t-1}^2 + \beta y_{t-1}^2 & & & \text{(II)}\end{aligned}$$

where $h_1^2 = \mu$, and where, for stationarity of the state equation (II) we have the constraint $0 < \alpha + \beta < 1$.

5.5.1 Likelihood calculation

The full likelihood under this model is given by

$$L(\mu, \alpha, \beta) \quad = \prod_{t=1}^{n} \left(\frac{1}{2\pi h_t^2}\right)^{1/2} \exp\left\{-\frac{y_t^2}{2h_t^2}\right\} \tag{5.3}$$

This three parameter likelihood is straightforward, but computationally quite laborious, to evaluate.

5.5.2 Prior specification

The joint prior distribution for the parameters (μ, α, β) is to be expressed as follows:

$$p(\mu, \alpha, \beta) = p(\mu)\, p(\alpha, \beta)$$

where, *a priori*,

$$\begin{array}{llll} \log\mu & \sim N\left(0, \sigma_\mu^2\right) & p(\mu) & = \left(\dfrac{1}{2\pi\sigma_\mu^2}\right)^{1/2} \mu^{-1} \exp\left\{-\dfrac{(\log\mu)^2}{2\sigma_\mu^2}\right\} \\ \alpha, \beta & \sim Dirichlet\,(\gamma_1, \gamma_2, \gamma_3) & p(\alpha, \beta) & = \dfrac{\Gamma(\gamma_1+\gamma_2+\gamma_3)}{\Gamma(\gamma_1)\Gamma(\gamma_2)\Gamma(\gamma_3)} \alpha^{\gamma_1-1}\beta^{\gamma_2-1}(1-\alpha-\beta)^{\gamma_3-1} \end{array}. \quad (5.4)$$

We might choose $\sigma_\mu = 5$, $\gamma_1 = \gamma_2 = \gamma_3 = 1$ Therefore, combining likelihood and prior, we can make inference via the full posterior distribution.

5.5.3 Markov Chain Monte Carlo: full conditional distributions

Samples from the joint posterior distribution are to be obtained using a Gibbs-sampler type Markov chain Monte Carlo algorithm. Three parameterizations are used for the state space equation; the required full conditionals are found to be of the following forms (as given by KSC - see appendix for full details).

Parameterization 1:	(μ, α, β)
Parameterization 2:	(μ, ϕ_1, ϕ_2)
Parameterization 3:	(μ, ψ_1, ψ_2)

where

$$\left.\begin{array}{ll} \phi_1 & = \dfrac{\alpha}{(1-\alpha-\beta)} \\ \phi_2 & = \dfrac{\beta}{(1-\alpha-\beta)} \end{array}\right\} \iff \left\{\begin{array}{ll} \alpha & = \dfrac{\phi_1}{(1+\phi_1+\phi_2)} \\ \beta & = \dfrac{\phi_2}{(1+\phi_1+\phi_2)} \end{array}\right. \qquad \text{Jacobian} : \frac{1}{(1+\phi_1+\phi_2)^3}$$

for parameterization 2, and

$$\left.\begin{array}{ll} \psi_1 & = \log\left(\dfrac{\alpha}{1-\alpha-\beta}\right) \\ \psi_2 & = \log\left(\dfrac{\beta}{1-\alpha-\beta}\right) \end{array}\right\} \iff \left\{\begin{array}{ll} \alpha & = \dfrac{e^{\psi_1}}{(1+e^{\psi_1}+e^{\psi_2})} \\ \beta & = \dfrac{e^{\psi_2}}{(1+e^{\psi_1}+e^{\psi_2})} \end{array}\right. \qquad \text{Jacobian} : \frac{e^{\psi_1+\psi_2}}{(1+e^{\psi_1}+e^{\psi_2})^3}$$

For each parameterization, the full conditionals are sampled using a local MH step, as per the details are given in Appendix 3 based on a symmetric (normal) proposal for parameters on $\mathbb{R}$, and using a Gamma proposal for parameters on $\mathbb{R}^+$. Parameters in the MH proposal are again chosen to optimize mixing of the chain; typical values used are $\lambda = 0.1$ and $\alpha = 500$.

5.6 T-GARCH(1,1): Introduction and Model Formulation

The T-GARCH(1,1) model can be formulated as follows: for data $y_1, ..., y_n$

$$\begin{aligned} y_t &= h_t z_t \qquad z_t \sim N(0, \lambda_t) \text{ i.i.d} && \text{(I)} \\ \lambda_t &\sim IGamma\left(\frac{\upsilon}{2}, \frac{\upsilon}{2}\right) && t = 1, ..., n \\ h_t^2 &= \mu + \alpha h_{t-1}^2 + \beta y_{t-1}^2 && \text{(II)} \end{aligned}$$

where $h_1^2 = \mu$, and where, for stationarity of the state equation (II) we have the constraint $0 < \alpha + \beta < 1$. The parameters $\lambda_t, t = 1, ..., n$ modify the model so that in stage (I) the observed data follow a Student-t distribution with υ degrees of freedom. Typically, we let υ take some small positive integr value ($\upsilon = 3-7$), or possibly regard υ as further unknown parameter. Note that setting each $\lambda_t = 1$ recovers the original GARCH(1,1) model.

5.6.1 Likelihood calculation

The full likelihood under this model is given by

$$L(\mu, \alpha, \beta) = \prod_{t=1}^{n} \left(\frac{1}{2\pi\lambda_t h_t^2}\right)^{1/2} \exp\left\{-\frac{y_t^2}{2\lambda_t h_t^2}\right\} \tag{5.5}$$

This three parameter likelihood is straightforward, but computationally quite laborious, to evaluate.

5.6.2 Prior specification

The joint prior distribution for the parameters (μ, α, β) is to be expressed as follows:

$$p(\mu, \alpha, \beta) = p(\mu)\, p(\alpha, \beta)$$

where, *a priori*,

$$\begin{aligned} \log\mu &\sim N\left(0, \sigma_\mu^2\right) & p(\mu) &= \left(\frac{1}{2\pi\sigma_\mu^2}\right)^{1/2} \mu^{-1} \exp\left\{-\frac{(\log\mu)^2}{2\sigma_\mu^2}\right\} \\ \alpha, \beta &\sim Dirichlet\,(\gamma_1, \gamma_2, \gamma_3) & p(\alpha, \beta) &= \frac{\Gamma(\gamma_1 + \gamma_2 + \gamma_3)}{\Gamma(\gamma_1)\,\Gamma(\gamma_2)\,\Gamma(\gamma_3)} \alpha^{\gamma_1 - 1} \beta^{\gamma_2 - 1} (1 - \alpha - \beta)^{\gamma_3 - 1} \end{aligned} . \tag{5.6}$$

Again we choose $\sigma_\mu = 5$, $\gamma_1 = \gamma_2 = \gamma_3 = 1$ Therefore, combining likelihood and prior, we can make inference via the full posterior distribution. If υ is also to be included as an unknown parameter, we could use a discrete prior say, Geometric, Negative binomial or Truncated Poisson, or if continuous parameterization is used, perhaps a truncated Gamma or Pareto prior

$$p(\upsilon) = \frac{\gamma}{(1+\upsilon)^{\gamma+1}} \qquad \upsilon > 1$$

5.6.3 Markov Chain Monte Carlo: full conditional distributions

Samples from the joint posterior distribution are to be obtained using a Gibbs-sampler type Markov chain Monte Carlo algorithm. The required full conditionals are modified according to the changes in likelihood formulation. The λ_t parameters are also to be sampled, and elementary calculations show that the full conditional is

$$\lambda_t \sim IGamma\left(\frac{\upsilon+1}{2}, \frac{1}{2}\left[\frac{y_t^2}{h_t^2}+\upsilon\right]\right)$$

In the (joint) full conditional for the parameters in (II), we simply replace h_t^2 by $\lambda_t h_t^2$ in the relevant likelihood expressions; all other terms remain unaffected.

5.7 Appendix 1: Full conditionals in the standard SV model

Below we use the following "completing the square" identity:

$$A(x-a)^2 + B(x-b)^2 = (A+B)\left(\mu - \frac{Aa+Bb}{A+B}\right)^2 + \frac{AB}{A+B}(a-b)^2$$

Full conditional for μ : Collecting terms in μ from (1) and (2), we have

$$\begin{aligned}
p(\mu\,|\,-) &\propto \exp\left\{-\frac{(1-\phi^2)}{2\sigma^2}(h_1-\mu)^2\right\}\left\{\prod_{t=2}^{n}\exp\left\{-\frac{1}{2\sigma^2}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)^2\right\}\right\}\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\} \\
&= \exp\left\{-\frac{(1-\phi^2)}{2\sigma^2}(\mu-h_1)^2\right\}\left\{\prod_{t=2}^{n}\exp\left\{-\frac{1}{2\sigma^2}\left((1-\phi)\mu-(h_t-\phi h_{t-1})\right)^2\right\}\right\}\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\} \\
&= \exp\left\{-\frac{(1-\phi^2)}{2\sigma^2}(\mu-h_1)^2\right\}\left\{\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=2}^{n}\left(\mu-\frac{d_t}{(1-\phi)}\right)^2\right\}\right\}\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{where } d_t = h_t - \phi h_{t-1} \\
&= \exp\left\{-\frac{(1-\phi^2)}{2\sigma^2}(\mu-h_1)^2\right\}\left\{\exp\left\{-\frac{(1-\phi)^2(n-1)}{2\sigma^2}\left(\mu-\frac{1}{(1-\phi)(n-1)}\sum_{t=2}^{n}d_t\right)^2\right\}\right\}\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\left(1-\phi^2\right)(\mu-h_1)^2+(1-\phi)^2(n-1)\left(\mu-\frac{\bar d}{(1-\phi)}\right)^2\right]\right\}\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{where } \bar d = \frac{1}{(n-1)}\sum_{t=2}^{n}d_t \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\left(\left(1-\phi^2\right)+(1-\phi)^2(n-1)\right)\left(\mu-\frac{\left(1-\phi^2\right)h_1+(1-\phi)(n-1)\bar d}{\left(1-\phi^2\right)+(1-\phi)^2(n-1)}\right)^2\right]\right\}\exp\left\{-\frac{\mu^2}{2\sigma_\mu^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\frac{\left(1-\phi^2\right)+(1-\phi)^2(n-1)}{\sigma^2}\left(\mu-\frac{(1+\phi)h_1+(n-1)\bar d}{(1+\phi)+(1-\phi)(n-1)}\right)^2+\frac{\mu^2}{\sigma_\mu^2}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{\left(1-\phi^2\right)+(1-\phi)^2(n-1)}{\sigma^2}+\frac{1}{\sigma_\mu^2}\right)\left(\mu-\frac{(1+\phi)h_1+(n-1)\bar d}{(1+\phi)+(1-\phi)(n-1)}\right)^2\right]\right\}
\end{aligned}$$

and therefore

$$\mu\,|\,- \equiv N\left(\frac{(1+\phi)h_1+(n-1)\bar d}{(1+\phi)+(1-\phi)(n-1)},\left(\frac{\left(1-\phi^2\right)+(1-\phi)^2(n-1)}{\sigma^2}+\frac{1}{\sigma_\mu^2}\right)^{-1}\right)$$

Full conditional for σ^2 : Collecting terms in σ^2 from (1) and (2), we have

$$\begin{aligned}
p(\sigma^2 \,|\, -) \;&\propto \left(\frac{(1-\phi^2)}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{(1-\phi^2)}{2\sigma^2}(h_1-\mu)^2\right\}\left\{\prod_{t=2}^{n}\left(\frac{1}{2\pi\sigma^2}\right)^{1/2}\exp\left\{-\frac{1}{2\sigma^2}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)^2\right\}\right\} \\
&\times \left(\frac{1}{\sigma^2}\right)^{\alpha_\sigma+1}\exp\left\{-\frac{\beta}{\sigma^2}\right\} \\
&\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\alpha+1}\exp\left\{-\frac{1}{\sigma^2}\left[\frac{(1-\phi^2)}{2}m_1^2+\frac{1}{2}\textstyle\sum_{t=2}^{n}\left(m_t-\phi m_{t-1}\right)^2+\beta_\sigma\right]\right\}
\end{aligned}$$

and therefore

$$\sigma^2 \,|\, - \;\equiv InverseGamma\left(\frac{n}{2}+\alpha_\sigma,\, \frac{(1-\phi^2)}{2}m_1^2+\frac{1}{2}\sum_{t=2}^{n}\left(m_t-\phi m_{t-1}\right)^2+\beta_\sigma\right)$$

where $m_t = h_t - \mu$.

Full conditional for ϕ : Collecting terms in ϕ from (1) and (2), we have

$$\begin{aligned}
p(\phi \,|\, -) \;&\propto \left(\frac{(1-\phi^2)}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{(1-\phi^2)}{2\sigma^2}(h_1-\mu)^2\right\}\left\{\prod_{t=2}^{n}\left(\frac{1}{2\pi\sigma^2}\right)^{1/2}\exp\left\{-\frac{1}{2\sigma^2}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)^2\right\}\right\} \\
&\times (1+\phi)^{\alpha_\phi-1}(1-\phi)^{\beta_\phi-1} \\
&\propto (1+\phi)^{\alpha_\phi+1/2-1}(1-\phi)^{\beta_\phi+1/2-1}\exp\left\{-\frac{(1-\phi^2)}{2\sigma^2}m_1^2\right\}\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=2}^{n}\left(m_t-\phi m_{t-1}\right)^2\right\} \\
&= (1+\phi)^{\alpha_\phi+1/2-1}(1-\phi)^{\beta_\phi+1/2-1}\exp\left\{-\frac{m_1^2}{2\sigma^2}\left(1-\phi^2\right)\right\}\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=2}^{n}m_{t-1}^2\left(\phi-\frac{m_t}{m_{t-1}}\right)^2\right\} \\
&\propto (1+\phi)^{\alpha_\phi+1/2-1}(1-\phi)^{\beta_\phi+1/2-1}\exp\left\{-\frac{m_1^2}{2\sigma^2}\left(1-\phi^2\right)\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{t=2}^{n}m_{t-1}^2\right)\left(\phi-\frac{\sum_{t=2}^{n}m_{t-1}m_t}{\sum_{t=2}^{n}m_{t-1}^2}\right)^2\right\} \\
&= (1+\phi)^{\alpha_\phi+1/2-1}(1-\phi)^{\beta_\phi+1/2-1}\exp\left\{-\frac{m_1^2}{2\sigma^2}\left(1-\phi^2\right)\right\}\exp\left\{-\frac{M_1}{2\sigma^2}\left(\phi-M_2\right)^2\right\} \quad \text{, say}
\end{aligned}$$

where

$$M_1 = \sum_{t=2}^{n} m_{t-1}^2 \qquad \text{and} \qquad M_2 = \frac{\sum_{t=2}^{n} m_{t-1}m_t}{\sum_{t=2}^{n} m_{t-1}^2}$$

Hence the conditional density for ϕ is not available in closed form. However, we may sample from $p(\phi \,|\, -)$ using either rejection sampling (as in KSC) or merely by using a Metropolis step. Two types of Metropolis step may be used; either a local step conditional on the current point (recalling that the ϕ lies in $(-1,1)$) or an independence step; see Appendix 3 for a full discussion.

Full conditional for h_1 **:** Collecting terms in h_1 from (1) and (2), we have

$$\begin{aligned}
p(h_1\,|\,-) \quad &\propto \exp\left\{-\frac{h_1}{2}\right\}\exp\left\{-\frac{y_1^2}{2e^{h_1}}\right\}\exp\left\{-\frac{\left(1-\phi^2\right)}{2\sigma^2}\left(h_1-\mu\right)^2\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(h_2-\mu-\phi\left(h_1-\mu\right)\right)^2\right\} \\
&\propto \exp\left\{-\frac{h_1}{2}\right\}\exp\left\{-\frac{y_1^2}{2e^{h_1}}\right\}\exp\left\{-\frac{1}{2\sigma^2}\left[\left(1-\phi^2\right)\left(h_1-\mu\right)^2+\phi^2\left(h_1-\mu-\frac{\left(h_2-\mu\right)}{\phi}\right)^2\right]\right\} \\
&\propto \exp\left\{-\frac{h_1}{2}\right\}\exp\left\{-\frac{y_1^2}{2e^{h_1}}\right\}\exp\left\{-\frac{\left(\left(1-\phi^2\right)+\phi^2\right)}{2\sigma^2}\left(h_1-\frac{\left(1-\phi^2\right)\mu+\phi^2\mu+\phi\left(h_2-\mu\right)}{\left(1-\phi^2\right)+\phi^2}\right)^2\right\} \\
&= \exp\left\{-\frac{h_1}{2}\right\}\exp\left\{-\frac{y_1^2}{2e^{h_1}}\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(h_1-\mu-\phi\left(h_2-\mu\right)\right)^2\right\}
\end{aligned}$$

and therefore the conditional is again not available in closed form. We again use either a local or an independence Metropolis step.

Full conditional for $h_t, t = 2, ..., n-1$ **:** Collecting terms in h_t from (1) and (2), we have

$$\begin{aligned}
p(h_t\,|\,-) \quad &\propto \exp\left\{-\frac{h_t}{2}\right\}\exp\left\{-\frac{y_t^2}{2e^{h_t}}\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)^2\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(h_{t+1}-\mu-\phi\left(h_t-\mu\right)\right)^2\right\} \\
&\propto \exp\left\{-\frac{h_t}{2}\right\}\exp\left\{-\frac{y_t^2}{2e^{h_t}}\right\}\exp\left\{-\frac{1}{2\sigma^2}\left[\left(h_1-\mu-\phi\left(h_{t-1}-\mu\right)\right)^2+\phi^2\left(h_t-\mu-\frac{\left(h_{t+1}-\mu\right)}{\phi}\right)^2\right]\right\} \\
&\propto \exp\left\{-\frac{h_t}{2}\right\}\exp\left\{-\frac{y_t^2}{2e^{h_t}}\right\}\exp\left\{-\frac{\left(1+\phi^2\right)}{2\sigma^2}\left(h_1-\frac{\mu+\phi\left(h_{t-1}-\mu\right)+\phi^2\mu+\phi\left(h_{t+1}-\mu\right)}{\left(1+\phi^2\right)}\right)^2\right\} \\
&= \exp\left\{-\frac{h_t}{2}\right\}\exp\left\{-\frac{y_t^2}{2e^{h_t}}\right\}\exp\left\{-\frac{\left(1+\phi^2\right)}{2\sigma^2}\left(h_t-\frac{(1-\phi)^2\mu+\phi\left(h_{t-1}+h_{t+1}\right)}{\left(1+\phi^2\right)}\right)^2\right\}
\end{aligned}$$

Full conditional for h_n **:** Collecting terms in h_n from (1) and (2), we have

$$p(h_n\,|\,-) \propto \exp\left\{-\frac{h_n}{2}\right\}\exp\left\{-\frac{y_n^2}{2e^{h_n}}\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(h_n-\mu-\phi\left(h_{n-1}-\mu\right)\right)^2\right\}$$

and therefore the conditional is again not available in closed form. We again use either a local or an independence Metropolis step.

5.8 Appendix 2: Reparameterization

We consider several reparameterizations that may be utilized either for computational convenience, or in an attempt to reduce the amount of autocorrelation in the Markov chain, or to remove posterior correlation in the sampled parameters; of course, only multivariate transformations will reduce posterior correlation. We shall consider univariate transformations of ϕ or joint transformations of $\left(\phi, \sigma^2\right)$

Reparameterization 1: $\psi = \dfrac{1+\phi}{1-\phi}$ Consider new parameter

$$\psi = \frac{1+\phi}{1-\phi} \qquad \Longleftrightarrow \qquad \phi = \frac{\psi-1}{\psi+1}$$

so that ψ takes values on $\mathbb{R}^+$. The Jacobian of the transformation is given by

$$J(\psi) = \left|\frac{d\phi}{d\psi}\right| = \frac{2}{(\psi+1)^2}$$

and hence the full conditional for ψ takes the form

$$\begin{aligned} p(\psi \,|\, -) \quad &\propto \left(1 + \frac{\psi-1}{\psi+1}\right)^{\alpha_\phi+1/2-1} \left(1 - \frac{\psi-1}{\psi+1}\right)^{\beta_\phi+1/2-1} \\ &\quad \times \exp\left\{-\frac{m_1^2}{2\sigma^2}\left(1 - \left(\frac{\psi-1}{\psi+1}\right)^2\right)\right\} \exp\left\{-\frac{M_1}{2\sigma^2}\left(\frac{\psi-1}{\psi+1} - M_2\right)^2\right\} \frac{2}{(\psi+1)^2} \end{aligned}$$

where

Reparameterization 2: $\xi = \log\left(\dfrac{1+\phi}{1-\phi}\right)$ Consider new parameter

$$\xi = \log\left(\frac{1+\phi}{1-\phi}\right) \qquad \Longleftrightarrow \qquad \phi = \frac{e^\xi-1}{e^\xi+1}$$

so that ξ takes values on $\mathbb{R}$. The Jacobian of the transformation is given by

$$J(\xi) = \left|\frac{d\phi}{d\xi}\right| = \left|\frac{\left(e^\xi+1\right)e^\xi - \left(e^\xi-1\right)e^\xi}{\left(e^\xi+1\right)^2}\right| = \frac{2e^\xi}{\left(e^\xi+1\right)^2}$$

and hence the full conditional for ξ takes the form

$$\begin{aligned} p(\xi \,|\, -) \quad &\propto \left(1 + \frac{e^\xi-1}{e^\xi+1}\right)^{\alpha_\phi+1/2-1} \left(1 - \frac{e^\xi-1}{e^\xi+1}\right)^{\beta_\phi+1/2-1} \\ &\quad \times \exp\left\{-\frac{m_1^2}{2\sigma^2}\left(1 - \left(\frac{e^\xi-1}{e^\xi+1}\right)^2\right)\right\} \exp\left\{-\frac{M_1}{2\sigma^2}\left(\frac{e^\xi-1}{e^\xi+1} - M_2\right)^2\right\} \frac{e^\xi}{\left(1+e^\xi\right)^2} \end{aligned}$$

Reparameterization 3: $\theta = \sigma^2 \left(\dfrac{1+\phi}{1-\phi}\right), \tau^2 = \sigma^2$ Consider a joint transformation with new parameters

$$\left.\begin{array}{l} \theta = \sigma^2 \left(\dfrac{1+\phi}{1-\phi}\right) \\ \\ \tau^2 = \sigma^2 \end{array}\right\} \quad \Longleftrightarrow \quad \left\{\begin{array}{l} \phi = \left(\dfrac{\theta-\tau^2}{\theta+\tau^2}\right) \\ \\ \tau^2 = \sigma^2 \end{array}\right.$$

so that (θ, τ^2) takes values on $\mathbb{R}^+ \times . \mathbb{R}^+$ The Jacobian of the transformation is given by

$$J(\theta, \tau^2) = \begin{vmatrix} \dfrac{d\phi}{d\theta} & \dfrac{d\phi}{d\tau^2} \\ \dfrac{d\sigma^2}{d\theta} & \dfrac{d\sigma^2}{d\tau^2} \end{vmatrix} = \left|\frac{d\phi}{d\theta}\right| = \left|\frac{(\theta+\tau^2)-(\theta-\tau^2)}{(\theta+\tau^2)^2}\right| = \frac{2\tau^2}{(\theta+\tau^2)^2}.$$

By inspection of (1) and (2), the joint (conditional) posterior for (θ, τ^2) takes the form

$$\begin{aligned} p(\theta, \tau^2 \,|\, -) \;&\propto\; \left(\frac{1}{\tau^2}\right)^{n/2} \left(1 - \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)^2\right)^{1/2} \exp\left\{-\frac{1}{2\tau^2}\left(1-\left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)^2\right) m_1^2\right\} \\ &\times \exp\left\{-\frac{1}{2\tau^2} \textstyle\sum_{t=2}^n \left(m_t - \left(\dfrac{\theta-\tau^2}{\theta+\tau^2}\right) m_{t-1}\right)^2\right\} \\ &\times \left(1 + \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)\right)^{\alpha_\phi - 1} \left(1 - \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)\right)^{\beta_\phi - 1} \times \left(\frac{1}{\tau^2}\right)^{\alpha_\sigma + 1} \exp\left\{-\frac{\beta_\sigma}{\tau^2}\right\} \times \frac{\tau^2}{(\theta+\tau^2)^2} \end{aligned}$$

which is proportional to the univariate conditional for τ^2, whereas the univariate conditional for θ is given up to proportionality by

$$\begin{aligned} p(\theta \,|\, -) \;&\propto\; \left(1 - \left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)^2\right)^{1/2} \exp\left\{-\frac{1}{2\tau^2}\left(1-\left(\frac{\theta-\tau^2}{\theta+\tau^2}\right)^2\right) m_1^2\right\} \exp\left\{-\frac{1}{2\tau^2} \textstyle\sum_{t=2}^n \left(m_t - \left(\dfrac{\theta-\tau^2}{\theta+\tau^2}\right) m_{t-1}\right)^2\right\} \\ &\times \theta^{\alpha_\phi - 1} \left(\frac{1}{\theta+\tau^2}\right)^{\alpha_\phi + \beta_\phi} \end{aligned}$$

These two conditionals are not available in closed form, but may be sampled using a Metropolis step.

5.9 Appendix 3: Local Metropolis moves

For parameters taking values on $\mathbb{R}$, the concept of a "local" Metropolis proposal is widely used. Let $q\left(x, y\right)$ be the Metropolis-Hastings (MH) proposal density in candidate point y for current point x; so that $q\left(x, y\right)$ can merely considered as a conditional density $q\left(x, y\right) \equiv q(y|x)$, where the MH acceptance probability for target density π is given by

$$\alpha(x, y) = \min\left\{1, \frac{\pi(y)}{\pi(x)}\frac{q(y, x)}{q(x, y)}\right\}$$

In a pure "local" Metropolis step with a symmetric proposal density, we have of course that

$$q(x, y) = q(y, x) \qquad \Longrightarrow \qquad \alpha(x, y) = \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\}$$

and such a local proposal is straightforward to achieve when the suppot of π is $\mathbb{R}$; set $q(y|x) \equiv N(x, \lambda^2)$ for some λ.

When the support of π is some sub-interval of $\mathbb{R}$, we consider the following moves:

Support $\mathbb{R}^+$ **:** If π has support $\mathbb{R}^+$, define $q\left(x, y\right) \equiv q(y|x) \equiv \mathit{Gamma}\left(\alpha, \frac{\alpha}{x}\right)$, so that

$$q(y|x) = \frac{\alpha^\alpha}{x^\alpha \Gamma(\alpha)} y^{\alpha-1} \exp\left\{-\frac{\alpha y}{x}\right\}$$

giving

$$\frac{q(y, x)}{q(x, y)} = \frac{q(x|y)}{q(y|x)} = \frac{\dfrac{\alpha^\alpha}{y^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left\{-\dfrac{\alpha x}{y}\right\}}{\dfrac{\alpha^\alpha}{x^\alpha \Gamma(\alpha)} y^{\alpha-1} \exp\left\{-\dfrac{\alpha y}{x}\right\}} = \left(\frac{x}{y}\right)^{2\alpha-1} \exp\left\{-\alpha\left[\frac{x}{y} - \frac{y}{x}\right]\right\}$$

where the "local" nature of q is determined by α; if α is large, then the proposed moves are small, and *vice versa.*

Support (a, b) **:** If π has bounded support (a, b), we consider a reparameterization of the move defined above: if π in variable X has support (a, b) then π_* for variable U

$$U = \frac{X - a}{b - X} \qquad \text{so that} \qquad X = \frac{U - b}{U - a}$$

has support $\mathbb{R}^+$, and we can define $q_*\left(u, v\right) \equiv q_*(v|u) \equiv \mathit{Gamma}\left(\alpha, \frac{\alpha}{u}\right)$, so that

$$q_*(v|u) = \frac{\alpha^\alpha}{u^\alpha \Gamma(\alpha)} v^{\alpha-1} \exp\left\{-\frac{\alpha v}{u}\right\}$$

giving, in the original parameterization,

$$q(y|x) = q_*\left(\frac{y-a}{b-y}|\frac{x-a}{b-x}\right)\left|\frac{dv}{dy}\right| = \frac{\alpha^\alpha}{\Gamma(\alpha)}\left(\frac{b-x}{x-a}\right)^\alpha \left(\frac{y-a}{b-y}\right)^{\alpha-1} \exp\left\{-\alpha\left(\frac{y-a}{b-y}\right)\left(\frac{b-x}{x-a}\right)\right\} \times \frac{(b-a)}{(b-y)^2}$$

and so

$$\frac{q(y,x)}{q(x,y)} = \frac{q(x|y)}{q(y|x)} = \frac{\frac{\alpha^\alpha}{\Gamma(\alpha)}\left(\frac{b-y}{y-a}\right)^\alpha\left(\frac{x-a}{b-x}\right)^{\alpha-1}\exp\left\{-\alpha\left(\frac{b-y}{y-a}\right)\left(\frac{x-a}{b-x}\right)\right\}\times\frac{(b-a)}{(b-x)^2}}{\frac{\alpha^\alpha}{\Gamma(\alpha)}\left(\frac{b-x}{x-a}\right)^\alpha\left(\frac{y-a}{b-y}\right)^{\alpha-1}\exp\left\{-\alpha\left(\frac{b-x}{x-a}\right)\left(\frac{y-a}{b-y}\right)\right\}\times\frac{(b-a)}{(b-y)^2}}$$

$$= \left(\frac{(x-a)(b-y)}{(b-x)(y-a)}\right)^{2\alpha-1}\exp\left\{-\alpha\left[\left(\frac{b-y}{y-a}\right)\left(\frac{x-a}{b-x}\right)-\left(\frac{b-x}{x-a}\right)\left(\frac{y-a}{b-y}\right)\right]\right\}\frac{(b-y)^2}{(b-x)^2}$$

Hence, if $a=-1, b=1$ we have

$$\frac{q(y,x)}{q(x,y)} = \left(\frac{(1+x)(1-y)}{(1-x)(1+y)}\right)^{2\alpha-1}\exp\left\{-\alpha\left[\left(\frac{1-y}{1+y}\right)\left(\frac{1+x}{1-x}\right)-\left(\frac{1-x}{1+x}\right)\left(\frac{1+y}{1-y}\right)\right]\right\}\frac{(1-y)^2}{(1-x)^2}$$

Mirror moves If π has support $\mathbb{R}^+$, consider the proposal mechanism that generates $z \sim N(0,\lambda^2)$, and proposes candidate point y as

$$y = |x+z|$$

Then, by elementary probability arguments we have

$$\mathrm{P}\left[Y \le y|X=x\right] = \mathrm{P}\left[|x+Z| \le y\right] = \mathrm{P}\left[-y \le x+Z \le y\right] = \mathrm{P}\left[-y-x \le Z \le y-x\right] = \Phi\left(\frac{y-x}{\lambda}\right) - \Phi\left(\frac{-y-x}{\lambda}\right)$$

where $\Phi(.)$ is the standard normal cdf and hence

$$q(y|x) = \frac{1}{\lambda}\left[\phi\left(\frac{y-x}{\lambda}\right)+\phi\left(\frac{-y-x}{\lambda}\right)\right] = \left(\frac{1}{2\pi\lambda^2}\right)^{1/2}\left[\exp\left\{-\frac{1}{2\lambda^2}(y-x)^2\right\}+\exp\left\{-\frac{1}{2\lambda^2}(y+x)^2\right\}\right]$$

and so

$$q(y|x) = q(x|y)$$

and the poposal is symmetric.

5.10 Appendix 4: Alternative Variable Dimension algorithm

Birth Move:

(1) sample a segment uniformly from the $K+1$ available; suppose segment k is selected.

(2) sample new changepoint position uniformly on the set of integers $(r_{k-1}+1,...,r_k-1)$; suppose r is selected

(3) sample random variates u, v, w independently according to the following distributions

$$u \sim N\left(0,\lambda^2\right) \qquad\qquad v,w \sim Gamma(\alpha,\alpha)$$

(4) transform the original parameters from segment k, that is $\left(\mu_k, \psi_k, \sigma_k^2\right)$ and u, v, w to obtain the parameters for the two new segments as follows

$$\left.\begin{array}{ll} \mu_{(1)} & = \mu_k + u \\ \mu_{(2)} & = \mu_k + u \\ & \\ \psi_{(1)} & = \psi_k v \\ \psi_{(2)} & = \psi_k / v \\ & \\ \sigma^2_{(1)} & = w\sigma_k^2 \\ \sigma^2_{(2)} & = \sigma_k^2 / w \end{array}\right\} \iff \left\{\begin{array}{ll} \mu_k & = \frac{1}{2}\left(\mu_{(1)} + \mu_{(2)}\right) \\ & \\ \psi_k & = \sqrt{\psi_{(1)}\psi_{(2)}} \\ & \\ \sigma_k^2 & = \sqrt{\sigma^2_{(1)}\sigma^2_{(2)}} \end{array}\right. .$$

Each of these steps makes a contribution to the transition acceptance probability; generically we need to compute $q(x, y)$ (and its compliment $q(y, x)$). Term by term, we have contributions from (1) to (3) as follows

$$q(x, y) = \frac{1}{K+1} \times \frac{1}{r_k - r_{k-1} - 1} \times q(u)q(v)q(w)$$

that appear in the denominator of the transition probability. For the birth move, the Jacobian term is given by

$$\left|\begin{array}{cccccc} \frac{\partial \mu_{(1)}}{\partial \mu_k} & \frac{\partial \mu_{(1)}}{\partial u} & \frac{\partial \mu_{(1)}}{\partial \psi_k} & \frac{\partial \mu_{(1)}}{\partial v} & \frac{\partial \mu_{(1)}}{\partial \sigma_k^2} & \frac{\partial \mu_{(1)}}{\partial w} \\ \frac{\partial \mu_{(2)}}{\partial \mu_k} & \frac{\partial \mu_{(2)}}{\partial u} & \frac{\partial \mu_{(2)}}{\partial \psi_k} & \frac{\partial \mu_{(2)}}{\partial v} & \frac{\partial \mu_{(2)}}{\partial \sigma_k^2} & \frac{\partial \mu_{(2)}}{\partial w} \\ \frac{\partial \psi_{(1)}}{\partial \mu_k} & \frac{\partial \psi_{(1)}}{\partial u} & \frac{\partial \psi_{(1)}}{\partial \psi_k} & \frac{\partial \psi_{(1)}}{\partial v} & \frac{\partial \psi_{(1)}}{\partial \sigma_k^2} & \frac{\partial \psi_{(1)}}{\partial w} \\ \frac{\partial \psi_{(2)}}{\partial \mu_k} & \frac{\partial \psi_{(2)}}{\partial u} & \frac{\partial \psi_{(2)}}{\partial \psi_k} & \frac{\partial \psi_{(2)}}{\partial v} & \frac{\partial \psi_{(2)}}{\partial \sigma_k^2} & \frac{\partial \psi_{(2)}}{\partial w} \\ \frac{\partial \sigma^2_{(1)}}{\partial \mu_k} & \frac{\partial \sigma^2_{(1)}}{\partial u} & \frac{\partial \sigma^2_{(1)}}{\partial \psi_k} & \frac{\partial \sigma^2_{(1)}}{\partial v} & \frac{\partial \sigma^2_{(1)}}{\partial \sigma_k^2} & \frac{\partial \sigma^2_{(1)}}{\partial w} \\ \frac{\partial \sigma^2_{(2)}}{\partial \mu_k} & \frac{\partial \sigma^2_{(2)}}{\partial u} & \frac{\partial \sigma^2_{(2)}}{\partial \psi_k} & \frac{\partial \sigma^2_{(2)}}{\partial v} & \frac{\partial \sigma^2_{(2)}}{\partial \sigma_k^2} & \frac{\partial \sigma^2_{(2)}}{\partial w} \end{array}\right| = \left|\begin{array}{cccccc} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & v & \psi_k & 0 & 0 \\ 0 & 0 & \frac{1}{v} & -\frac{\psi_k}{v^2} & 0 & 0 \\ 0 & 0 & 0 & 0 & w & \sigma_k^2 \\ 0 & 0 & 0 & 0 & \frac{1}{w} & -\frac{\sigma_k^2}{w^2} \end{array}\right| = \frac{8\psi_k\sigma_k^2}{vw}$$

Death Move: For death moves, the number of changepoints is to change from $K+1$ to K, and the number of changepoint segments from $K+2$ to $K+2$. The move involves the removal of an existing changepoint position, and the combination of the two sets of parameters. To ensure reversibility of the birth move, we must consider the complimentary death move, and propose the move to the smaller model as follows

(1) sample a changepoint uniformly from the $K+1$ available; suppose changepoint k is selected.

(2) reverse transform the two sets of original parameters from segment k (denoted by subscript (1)) and $k+1$ (denoted by subscript (2)) that is

$$\mu_k = \frac{1}{2}\left(\mu_{(1)} + \mu_{(2)}\right) \qquad \psi_k = \sqrt{\psi_{(1)}\psi_{(2)}} \qquad \sigma_k^2 = \sqrt{\sigma^2_{(1)}\sigma^2_{(2)}}.$$

in accordance with the birth move above

These two steps make a contribution to the transition acceptance probability The contributions from (1) is

$$q(x,y)=\frac{1}{K+1}$$

that appear in the denominator of the transition probability. For the death move, the Jacobian term is given by

$$\frac{1}{8\psi_{(2)}\sigma^2_{(2)}}$$

Therefore, the acceptance probability for a birth move

$$\left(\mu_k,\psi_k,\sigma^2_k;r_{k-1},r_k,K\right)\qquad\rightarrow\qquad\left(\mu_{(1)},\psi_{(1)},\sigma^2_{(1)},\mu_{(2)},\psi_{(2)},\sigma^2_{(2)};r_{k-1},r,r_k,K+1\right)$$

is given by the minimum of 1 and

$$\frac{\pi(\mu_{(1)},\psi_{(1)},\sigma^2_{(1)},\mu_{(2)},\psi_{(2)},\sigma^2_{(2)};r_{k-1},r,r_k,K+1)}{\pi(\mu_k,\psi_k,\sigma^2_k;r_{k-1},r_k,K)}\times\frac{\frac{1}{K+1}}{\frac{1}{K+1}\times\frac{1}{r_k-r_{k-1}-1}\times q(u)q(v)q(w)}\times\frac{8\psi_k\sigma^2_k}{vw}$$

where

$$\pi(\mu_k,\psi_k,\sigma^2_k;r_{k-1},r_k,K)\propto l\left(h_{r_{k-1}+1},...,h_{r_k};\mu_k,\phi_k,\sigma^2_k\right)p(\mu_k,\psi_k,\sigma^2_k)$$

and

$$\begin{aligned}\pi(\mu_{(1)},\psi_{(1)},\sigma^2_{(1)},\mu_{(2)},\psi_{(2)},\sigma^2_{(2)};r_{k-1},r,r_k,K+1)\quad&\propto l\left(h_{r_{k-1}+1},...,h_r;\mu_{(1)},\psi_{(1)},\sigma^2_{(1)}\right)l\left(h_{r+1},...,h_{r_{k+1}};\mu_{(2)},\psi_{(2)},\sigma^2_{(2)}\right)\\&\times p(\mu_{(1)},\psi_{(1)},\sigma^2_{(1)})\,p(\mu_{(2)},\psi_{(2)},\sigma^2_{(2)})\,p(r|\,r_{k-1},r_k)\end{aligned}$$

and cancelling terms, we have that $\alpha\left(x,y\right)$ is the minimum of 1 and

$$\frac{l\left(h_{r_{k-1}+1},...,h_r;\mu_{(1)},\psi_{(1)},\sigma^2_{(1)}\right)l\left(h_{r+1},...,h_{r_{k+1}};\mu_{(2)},\psi_{(2)},\sigma^2_{(2)}\right)p(\mu_{(1)},\psi_{(1)},\sigma^2_{(1)})\,p(\mu_{(2)},\psi_{(2)},\sigma^2_{(2)})}{l\left(h_{r_{k-1}+1},...,h_{r_k};\mu_k,\phi_k,\sigma^2_k\right)p(\mu_k,\psi_k,\sigma^2_k)q(u)q(v)q(w)}\frac{8\psi_k\sigma^2_k}{vw}$$

5.11 Appendix 5: Full conditionals for GARCH and T-GARCH

The general GARCH(1,1) likelihood

$$L(\mu,\alpha,\beta)\quad=\prod_{t=1}^{n}\left(\frac{1}{2\pi\lambda_t h^2_t}\right)^{1/2}\exp\left\{-\frac{y^2_t}{2\lambda_t h^2_t}\right\}$$

where

$$\begin{aligned}y_t\quad&=h_t z_t\qquad\qquad z_t\sim N(0,\lambda_t)\text{ i.i.d}\qquad\qquad(\text{I})\\\lambda_t\quad&\sim IGamma\left(\frac{\upsilon}{2},\frac{\upsilon}{2}\right)\qquad\qquad\qquad t=1,...,n\\h^2_t\quad&=\mu+\alpha h^2_{t-1}+\beta y^2_{t-1}\qquad\qquad\qquad(\text{II})\end{aligned}$$

with $\lambda_t \equiv 1$ if GARCH rather than T-GARCH is required. Retaining this original parameterization, the joint prior distribution for the parameters (μ, α, β) is to be expressed as follows:

$$p(\mu, \alpha, \beta) = p(\mu)\, p(\alpha, \beta)$$

where, *a priori*,

$$\begin{array}{llll} \log \mu & \sim N\left(0, \sigma_\mu^2\right) & p(\mu) & = \left(\dfrac{1}{2\pi\sigma_\mu^2}\right)^{1/2} \mu^{-1} \exp\left\{-\dfrac{(\log \mu)^2}{2\sigma_\mu^2}\right\} \\ \alpha, \beta & \sim Dirichlet\left(\gamma_1, \gamma_2, \gamma_3\right) & p(\alpha, \beta) & = \dfrac{\Gamma(\gamma_1 + \gamma_2 + \gamma_3)}{\Gamma(\gamma_1)\Gamma(\gamma_2)\Gamma(\gamma_3)} \alpha^{\gamma_1 - 1} \beta^{\gamma_2 - 1} (1 - \alpha - \beta)^{\gamma_3 - 1} \end{array}.$$

Hence, the full conditional for μ, α, β given $\lambda_t, t = 1, ..., n$ is

$$\begin{aligned} p(\mu, \alpha, \beta \,|\, -) &= \left\{\prod_{t=1}^n \left(\frac{1}{2\pi\lambda_t h_t^2}\right)^{1/2} \exp\left\{-\frac{y_t^2}{2\lambda_t h_t^2}\right\}\right\} \\ &\quad \times \left(\frac{1}{2\pi\sigma_\mu^2}\right)^{1/2} \mu^{-1} \exp\left\{-\frac{(\log \mu)^2}{2\sigma_\mu^2}\right\} \times \frac{\Gamma(\gamma_1 + \gamma_2 + \gamma_3)}{\Gamma(\gamma_1)\Gamma(\gamma_2)\Gamma(\gamma_3)} \alpha^{\gamma_1 - 1} \beta^{\gamma_2 - 1} (1 - \alpha - \beta)^{\gamma_3 - 1} \\ &\propto \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^n \log h_t^2 + \frac{y_t^2}{\lambda_t h_t^2}\right]\right\} \times \mu^{-1} \exp\left\{-\frac{(\log \mu)^2}{2\sigma_\mu^2}\right\} \times \alpha^{\gamma_1 - 1} \beta^{\gamma_2 - 1} (1 - \alpha - \beta)^{\gamma_3 - 1} \end{aligned}$$

If required, the full conditional for $\lambda_t, t = 1, ..., n$ is calculated as follows

$$\begin{aligned} p(\lambda_t \,|\, -) &= \left\{\left(\frac{1}{2\pi\lambda_t h_t^2}\right)^{1/2} \exp\left\{-\frac{y_t^2}{2\lambda_t h_t^2}\right\}\right\} \times \frac{\left(\frac{\upsilon}{2}\right)^{\upsilon/2}}{\Gamma\left(\frac{\upsilon}{2}\right)} \left(\frac{1}{\lambda_t}\right)^{\upsilon/2 + 1} \exp\left\{-\frac{\upsilon}{2\lambda_t}\right\} \\ &\propto \left(\frac{1}{\lambda_t}\right)^{(\upsilon+1)/2 + 1} \exp\left\{-\frac{1}{2}\left[\frac{y_t^2}{h_t^2} + \upsilon\right]\frac{1}{\lambda_t}\right\} \\ &\equiv IGamma\left(\frac{\upsilon + 1}{2}, \frac{1}{2}\left[\frac{y_t^2}{h_t^2} + \upsilon\right]\right) \end{aligned}$$

If υ is also regarded as an unknown parameter, its full conditional is given by

$$\begin{aligned} p(\upsilon \,|\, -) &= \left\{\prod_{t=1}^n \frac{\left(\frac{\upsilon}{2}\right)^{\upsilon/2}}{\Gamma\left(\frac{\upsilon}{2}\right)} \left(\frac{1}{\lambda_t}\right)^{\upsilon/2 + 1} \exp\left\{-\frac{\upsilon}{2\lambda_t}\right\}\right\} \times p(\upsilon) \\ &\propto \left\{\frac{\left(\frac{\upsilon}{2}\right)^{\upsilon/2}}{\Gamma\left(\frac{\upsilon}{2}\right)}\right\}^n \times \frac{1}{\left\{\prod_{t=1}^n \lambda_t\right\}^{\upsilon/2 + 1}} \exp\left\{-\frac{\upsilon}{2}\sum_{t=1}^n \frac{1}{\lambda_t}\right\} \times p(\upsilon) \end{aligned}$$

A discrete or continuous parameterization for υ gives either a discrete or continuous full conditional; both can be sampled using a Metropolis step.