Bayesian Inference in Generalized Gamma Processes for Stochastic Volatility.

Roberto Leon-Gonzalez
National Graduate Institute for Policy Studies (GRIPS)
7-22-1, Roppongi, Minato-Ku, 106-8677 Tokyo, Japan
rlg@grips.ac.jp

February 2014

Preliminary Version

Abstract

This paper develops a novel and efficient algorithm for Bayesian inference in Gamma Stochastic Volatility models. It is shown that by conditioning on auxiliary variables, it is possible to sample all the volatilities jointly directly from their posterior conditional density, using simple and easy to draw from distributions. Furthermore, this paper develops a generalized Gamma process that allows for more flexible tails in the distribution of volatilities. Using several macroeconomic and financial datasets, it is shown that Gamma and Generalized Gamma processes can greatly outperform log normal volatility processes with student-t errors.

JEL: C11, C15

1 Introduction

There is overwhelming empirical evidence in favor of Stochastic Volatility models with both macroeconomic (e.g. Sims and Zha 2006) and financial data (e.g. Kim et al. (1998)). The first algorithms for posterior simulation where developed for the case in which the volatility σ_t follows an autoregressive log-normal process. The first algorithms used a single-move update for the volatilities (e.g. Jacquier, Polson and Rossi (1994)), which implies that σ_t is generated conditionally on the volatility values in other periods $(\sigma_1, ..., \sigma_{t-1}, \sigma_{t+1}, ..., \sigma_T)$. To improve the convergence speed, it was later proposed to sample several of the volatility values at a time using blocking strategies (e.g. Shephard and Pitt (1997), Watanabe and Omori (2004), Asai (2005)). In an influential paper, Kim et al. (1998) showed that by accurately approximating the likelihood with a mixture of normals, it is possible to draw jointly all the latent volatilities given some auxiliary variables. Furthermore, the log-volatilities can be integrated out when drawing the unknown parameters.

A more recent literature provides methods for Bayesian inference in models where σ_t follows some type of gamma process. In a multivariate stochastic volatility context, Philipov and Glickman (2006) proposed a single-move algorithm whereas Fox and West (2011) proposed to sample all the volatility matrices jointly in a Metropolis-step which conditions on auxiliary variables. Creal (2012), in the univariate context, proposed maximum likelihood estimation by accurately approximating the likelihood with a finite state Markov-switching model. There is also a recent literature that deals with Ornstein-Ulhlenbeck with marginal gamma laws (e.g. Barndorff-Nielsen and Shephard (2001), Roberts et al. (2004), Griffin and Steel (2006), Frühwirth-Schnatter and Sögner (2009)).

The purpose of this paper is to develop efficient posterior simulators for flexible gamma stochastic volatility models. We show that by conditioning on some auxiliary variables, it is possible to draw all the volatilities jointly using simple distributions such as the Poisson and Gamma. Furthermore, it is possible to draw the unknown parameters after integrating out all the volatilities. Because of these features, our algorithm mimicks the efficient algorithm that Kim et al (1998) developed for the lognormal model. Furthermore, this paper proposes a generalized gamma model that allows for a more flexible distribution for the volatility. The generalized gamma process allows for more abrupt jumps in volatility. In an empirical exercise we show that this feature makes the generalized gamma process especially suitable to model series with periods of great instability.

Section 2 describes the gamma and generalized gamma processes and Section 3 develops the posterior simulators. Section 4 presents evidence on the computational efficiency of the algorithms and Section 5 compares different models using several macroeconomic and financial time series. Section 6 concludes.

2 Models

2.1 Autoregressive Gamma Process (TARG)

We consider the following model of stochastic volatility:

$$\begin{aligned} y_t &= x_t\beta + \sigma_t\varepsilon_t \\ \varepsilon_t &\sim N(0,1) \end{aligned}$$

Although for simplicity in the exposition we are assuming normality for ε_t, in the empirical applications we will consider also models where ε_t follows a student-t. The student-t can be easily incorporated into this framework by writing it as a mixture of normals, as in Chib et al (2002). The stochastic process for the volatility σ_t can be described by defining $k_t = \sigma_t^{-2}$ and assuming that $k_t = z_t' z_t$, where z_t is a $n \times 1$ vector distributed as a Gaussian AR(1) process:

$$z_t = \rho z_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \theta^2 I_n) \tag{1}$$

Equation (1) implies that the conditional distribution of $(k_t/\theta^2)|k_{t-1}$ is a noncentral chi squared, which is also well defined for non-integer values of n, and therefore we will treat n as a continuous unknown parameter. The joint distribution of $(k_1, ..., k_T)$ is the multivariate gamma distribution analyzed by Krishnaiah and Rao (1961). It was proposed for observed volatility by Gourieroux and Jasiak (2006) and for unobserved volatility by Creal (2012). In our case we are using it for the inverse of the unobserved volatility, as this makes Bayesian computations simpler. This is in line with the Bayesian analysis of Fox and West (2011), who specify a Wishart distribution for the inverse volatility matrix.

The properties of $(k_1, ..., k_T)$ are well known (e.g. Krishnaiah and Rao (1961)) and the most important ones can be summarized as:

- $E(k_t) = \frac{n\theta^2}{1-\rho^2}$, $E(k_t^2) = \left(\frac{\theta^2}{1-\rho^2}\right)^2 n(n+2)$
- $corr(k_t, k_{t-h}) = \rho^{2h}$
- $E(k_t|k_{t-1}) = \rho^2 k_{t-1} + (1-\rho^2)E(k_t)$
- The conditional distribution $\frac{k_t}{\theta^2}|k_{t-1}$ is a noncentral chi squared.
- The stationary distribution of k_t is a $G(n/2, \frac{2\theta^2}{1-\rho^2})$, where $G(.)$ represents the gamma distribution (Bauwens et al. (1999), p. 290)
- A necessary and suficient condition for stationarity is $|\rho| < 1$

In the following it will be assumed that k_1 is drawn from the stationary distribution, that is $k_1 \sim G(n/2, 2\theta^2/(1-\rho^2))$.

2.2 Flexible Tail Autoregressive Gamma Process (FTARG)

The parameters (n, θ^2, ρ^2) control the unconditional mean, variance and the first order correlation of k_t. However, the degrees of freedom n also control the shape of the tails of the distribution of k and therefore also the tails of the distribution of y. Hence it might be desirable to consider models where the shape of the tails is not determined by the first two unconditional moments of the distribution. For this purpose we propose the Flexible Tail Autoregressive Gamma Process (FTARG). Recall that $k_t = z_t' z_t$. Instead of $z_t = \rho z_{t-1} + \varepsilon_t$ we now assume:

$$z_t = \sqrt{\widetilde{T}_t}(\rho z_{t-1} + \varepsilon_t)$$

where $(\widetilde{T}_2, ..., \widetilde{T}_T)$ are independent draws from a beta distrubution $B(\underline{\alpha}, \underline{\beta})$. If we write $\widetilde{\rho}_t = \sqrt{\widetilde{T}_t}\rho$ and $\widetilde{\theta}_t^2 = \widetilde{T}_t \theta^2$ it is clear that the FTARG process arises from (1) by writing $\widetilde{\rho}_t$ instead of ρ and $\widetilde{\theta}_t^2$ instead of θ^2. When the variance of $\widetilde{T}_t$ approaches 0, the variable $\widetilde{T}_t$ behaves similar to a constant, and therefore the FTARG becomes equivalent to a TARG. Thus, when the variance of $\widetilde{T}_t$ is close to 0, the mean of $\widetilde{T}_t$ is poorly identified. To avoid this local non-identification problem, we reparameterize $(\underline{\alpha}, \underline{\beta})$ as $A = A = \underline{\alpha}/(\underline{\alpha} + \underline{\beta})$ and $V = (\underline{\alpha} + \underline{\beta})$, and fix $A = 1/2$. Therefore with this normalization we have that $\underline{\alpha} = \underline{\beta} = V/2$. The parameter V controls the variance of $\widetilde{T}_t$ and will be estimated.

The properties of the FTARG can be derived using basic properties of the gamma and beta distributions. Defining $\widetilde{\rho}^2 = E(\widetilde{T}_t)\rho^2$ and $\widetilde{\theta}^2 = E(\widetilde{T}_t)\theta^2$, the main properties of $(k_1, ..., k_T)$ are:

- $E(k_t) = \frac{n\widetilde{\theta}^2}{1-\widetilde{\rho}^2}$, $E(k_t^2) = \left(\widetilde{\theta}^2\right)^2 n(n+2)\frac{E(v_c^2)}{\left[E(\widetilde{T}_t)\right]^2}$
- $corr(k_t, k_{t-1}) = \widetilde{\rho}^2$
- The sufficient and necessary condition for the process to be stationary is $\widetilde{\rho}^2 < 1$. When $E(\widetilde{T}_t)$ is normalized to be 1/2 the stationarity condition becomes $\rho^2 < 2$.
- The stationary distribution of k_t is that of the product of $\varepsilon_t' \varepsilon_t$ (i.e. a gamma distribution) and v_c, where $v_c = (1 + \rho^2 \widetilde{T}_t + \rho^4 \widetilde{T}_t \widetilde{T}_{t-1} + \rho^6 \widetilde{T}_t \widetilde{T}_{t-1} \widetilde{T}_{t-2} + ...)$. Note that $\varepsilon_t' \varepsilon_t$ and v_c are independent and therefore all the moments of k_t can be derived from basic properties of the gamma and beta distributions.

In the following it will be assumed that k_1 is drawn from a distribution whose first two moments coincide with the stationary distribution: $k_1 \sim G(n/2, 2\widetilde{\theta}^2/(1-\widetilde{\rho}^2))$.

3 Computation by Gibbs Sampling

3.1 Autoregressive Gamma Process (TARG)

In this section we will use the notation $\widetilde{\rho}_t = \sqrt{\widetilde{T}_t}\rho$ and $\widetilde{\theta}_t^2 = \widetilde{T}_t\theta^2$ for $t = 2, ..., T$ and $\widetilde{\rho}_1 = \widetilde{\rho} = \sqrt{E(\widetilde{T}_t)}\rho$, $\widetilde{\theta}_1^2 = \widetilde{\theta}^2 = E(\widetilde{T}_t)\theta^2$ with the understanding that in the TARG model $\widetilde{T}_t = 1$ and so $\widetilde{\rho}_t = \rho$ and $\widetilde{\theta}_t^2 = \theta^2$ for every t. In this way the conditional posterior densities derived in this section will be valid for both the TARG and the FTARG models when $\widetilde{T}$ is among the conditioning variables As noted before the prior of $\frac{k_t}{\widetilde{\theta}_t^2}|k_{t-1}$ is a noncentral chi squared. From Muirhead (1982, p. 23) it turns out that a noncentral chi squared can be written as a mixture of (central) chi-squared with degrees of freedom $n + 2h_t$, where h_t follows a Poisson. Using this representation, the model can be written as:

$$\begin{aligned} y_t &= x_t\beta + \sqrt{\frac{1}{k_t}}\varepsilon_t \qquad (2) \\ \varepsilon_t &\sim N(0,1) \\ k_t|k_{1:(t-1)}, h_{1:t}, \Theta, \beta &\sim G(n/2 + h_t, 2\widetilde{\theta}_t^2) \\ h_t|k_{1:(t-1)}, h_{1:(t-1)}, \Theta, \beta &\sim P(\lambda_t) \quad \text{with } \lambda_t = \frac{\widetilde{\rho}_t^2 k_{t-1}}{2\widetilde{\theta}_t^2} \end{aligned}$$

where $G(.)$ represents the gamma distribution (Bauwens et al. (1999), p. 290), $P(.)$ is the Poisson distribution (Koop (2003), p. 325) and $k_{1:(t-1)}$ is notation for $(k_1, ..., k_{(t-1)})$. Let $\Theta = (n, \theta^2, \rho)$, $k = (k_1, ..., k_T)$ and $h = (h_2, ..., h_T)$. The representation (2) suggests the first Gibbs sampling algorithm that we consider:

The h-Gibbs

- Generate $\Theta|h, \beta$ (Metropolis step)
- Generate $k|h, \Theta, \beta$ (draw from independent gamma).
- Generate $h|k, \Theta, \beta$ (draw from independent Bessel distributions).
- Generate $\beta|k, h, \Theta$ (draw from a multivariate normal).

Note that for greater efficiency Θ is drawn marginally on k. For this reason k needs to be drawn immediately after Θ, so that the algorithm converges to the joint posterior distribution. An advantage of this algorithm is that all the precisions in the vector k can be drawn jointly from the conditional posterior. Similarly, as noted by Creal (2012), the vector h can be drawn jointly from the posterior conditional using a discrete distribution known as Bessel distribution (Yuan and Kalbfleisch (2000)). Devroye (2002) and Iliopoulos and Karlis (2003) have developed efficient algorithms to draw from the Bessel distribution. The conditional distributions needed in the h-Gibbs algorithm are summarized in the following proposition, whose proof is in the appendix.

Proposition 1 *Consider the model defined by (2), and define:*

$$\begin{aligned} r_t^2 &= (y_t - x_t\beta)^2 \\ \widetilde{r}_t^2 &= \left(\frac{1+\widetilde{\rho}_t^2}{\widetilde{\theta}_t^2} + r_t^2\right)^{-1} \quad \textit{for } t = 2, ..., T-1 \\ \widetilde{r}_t^2 &= \left(\frac{1}{\widetilde{\theta}_t^2} + r_t^2\right)^{-1} \quad \textit{for } t = 1 \textit{ and } t = T \\ h_1 &= h_{T+1} = 0 \end{aligned}$$

The conditional posteriors are as follows:

$$\begin{aligned} k_t|h,\Theta,\beta,Y &\sim G((n+1)/2 + h_t + h_{t+1}, 2\widetilde{r}_t^2) \quad \textit{for } t = 1, ..., T \\ h_t|k,\Theta,\beta,Y &\sim Bessel\left(\frac{n-2}{2}, \widetilde{\rho}_t \frac{\sqrt{k_t k_{t-1}}}{\widetilde{\theta}_t^2}\right) \quad \textit{for } t = 2, ..., T \end{aligned}$$

and

$$\begin{aligned} p(\Theta|Y,h,\beta) \propto & \int p(\Theta)p(k,h|\Theta)L(Y|k,\beta)dk = \\ & \prod_{t=1}^{T}\left[\left(2\widetilde{r}_t^2\right)^{\frac{n+1}{2}+h_{t+1}+h_t}\Gamma\left(\frac{n+1}{2}+h_{t+1}+h_t\right)\right] \\ & \left[\prod_{t=2}^{T}\frac{1}{\left(2\widetilde{\theta}_t^2\right)^{n/2}}\frac{\left(\frac{\rho}{2\theta^2}\right)^{2h_t}}{h_t!}\frac{1}{[n/2]_{h_t}}\right](1-\widetilde{\rho}^2)^{n/2}\left(2\widetilde{\theta}^2\right)^{-\frac{n}{2}}\left(\Gamma\left(\frac{n}{2}\right)\right)^{-T}p(\Theta) \end{aligned} \tag{3}$$

where $L(Y|k,\beta)$ is the density function of the observed data Y given the volatilities k and $p(\Theta)$ is the prior.

However, the convergence of this algorithm can be slow because of the high correlation between k and h. Indeed, once we condition upon h, the different components of k become independent of each other, even if unconditionally the serial correlation of k_t is tipically very high. This suggests that h contains too much information about k and so ideally we would like to draw k and h jointly. Thus we consider a second Gibbs algorithm that surpasses this problem, and that also has the advantage of drawing from distributions that are simpler than the Bessel. For this purpose we introduce two vectors of auxiliary variables, one of them continuous $m = (m_2, ..., m_T)$ and another discrete $d = (d_2, ..., d_T)$, such that we will be able to draw (k, h) jointly conditioning on (m, d) and viceversa. Using simple properties of the beta distribution the appendix shows that the transition equation for h in (2) can be equivalently written as:

$$\begin{aligned}
\Pr(h_t &= s|k_{1:(t-1)}, h_{1:(t-1)}, m_{1:t}) = \frac{\frac{\lambda_t^s}{s!} m_t^s \frac{[n/2]_s}{[(n-1)/2]_s}}{{}_1F_1(n/2; (n-1)/2; m_t \lambda_t)} && (4)\\
\lambda_t &= \frac{\widetilde{\rho}_t^2 k_{t-1}}{2\widetilde{\theta}_t^2} \\
p(m_t|m_{1:(t-1)}) &= \frac{\Gamma(\alpha_m + \beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)} \frac{({}_1F_1(n/2; (n-1)/2; m_t \lambda_t))}{\exp(\lambda_t)} m_t^{\alpha_m - 1}(1 - m_t)^{\beta_m - 1} \\
\alpha_m &= \frac{n-1}{2} \qquad \beta_m = \frac{1}{2} \qquad m_1 = 1 && (5)
\end{aligned}$$

where $[x]_s$ is notation for the rising factorial $[x]_s = (x)(x + 1)...(x + s - 1)$, with $[x]_0 = 1$, and ${}_1F_1(.)$ is a hypergeometric function (e.g. Muirhead (1982, p. 258)):

$${}_1F_1(n/2; (n-1)/2; \lambda_t) = \sum_{s=0}^{\infty} \frac{\lambda_t^s}{s!} m_t^s \frac{[n/2]_s}{[(n-1)/2]_s}$$

The advantage of this parameterization is that the posterior of $h_t|(k_{1:(t-1)}, h_{1:(t-1)}, m_{1:t})$ is a finite mixture of shifted Poissons, whereas the posterior of $k_t|k_{1:(t-1)}, h_{1:t}, m_{1:t}$ continues to be a Gamma. This is what makes possible the joint sampling of the two vectors k and h conditional on m. However, the calculation of the probabilities of each component of the mixture could be time consuming, especially when T is large. For this reason we have preferred to condition on a mixture indicator d_t, such that the conditional posterior of h_t becomes a shifted Poisson. This implies that conditional on (m, d), the two vectors k and h can be drawn jointly from the conditional posterior using simple gamma and Poisson distributions. In turn, $(m, d)|(k, h)$ can be drawn using independent beta distributions (for m) and a finite discrete distribution for d.

A shifted Poisson results from adding a fixed constant to a random variable with Poisson distribution (Winkelmann (2008, p.10)). We use the notation $h_t \sim SP(\lambda_t, d_t)$ to mean that $(h_t - d_t)$ follows a Poisson distribution (i.e. $(h_t - d_t) \sim P(\lambda_t)$). The probability density function of a shifted Poisson distribution is:

$$f_{SP}(h|d, \lambda) = \lambda^{h-d} \frac{1}{(h-d)!} \frac{1}{\exp(\lambda)} \qquad h = d, (d+1), ... \qquad (6)$$

The vector d is formally introduced in the model by using the following prior that depends on h:

$$\Pr\left(d_t = s | h_t, d_{t+1}\right) = \frac{\frac{[h_t]^s}{s!} \frac{[1+d_{t+1}]^s}{[(n-1)/2]_s}}{\sum_{s=0}^{(1+d_{t+1})} \left(\frac{[h_t]^s}{s!} \frac{[1+d_{t+1}]^s}{[(n-1)/2]_s} \right)} \qquad \begin{array}{c} t = 2, ..., T \\ 0 \leq s \leq (1 + d_{t+1}) \\ d_{T+1} = 0 \end{array} \quad (7)$$

where $[x]^s$ is notation for the falling factorial $[x]^s = (x)(x-1)...(x-s+1)$, with $[x]^0 = 1$. Note that d_T can take only two values, 0 and 1. The support of $d_{T-1}|d_T$ is from 0 up to $(1 + d_T)$, so d_{T-1} could at most take value 2. Similarly, the support of $d_t|d_{(t+1):T}$ is from 0 up to $(1 + d_{t+1})$, such that d_2 could take at most value $(T-1)$. However, in our applications to real data we have found d_t to be at most 20 even when $T = 10168$, and so each d_t was drawn from a discrete distribution defined on a relatively small set of values. Note also that the term $[h_t]^s$ in the probability implies that $d_t \leq h_t$.

Thus the Gibbs algorithm that uses (m, d) as auxiliary variables can be described as:

The m-Gibbs for TARG

- $\Theta|(m, d), \beta$ using a Metropolis step.
- $(k, h)|(m, d), \Theta, \beta$ using gammas and poisson.
- $(m, d)|(k, h), \Theta, \beta$ using beta and the finite discrete distribution in (7).
- Generate $\beta|k, h, \Theta$ (draw from a multivariate normal).

Note that for greater efficiency Θ is drawn marginally on (k, h). Therefore, the step to draw (k, h) needs to come just after drawing Θ, so that the joint posterior continues to be the stationary distribution. The following proposition describes the distributions that are used in the m-Gibbs.

Proposition 2 *Given the model described in equations (2), (4) - (5), and the following definitions:*

$$\begin{aligned} \widehat{r}_T^2 &= \widetilde{r}_T^2 \\ \widehat{r}_t^2 &= \left(\frac{1}{\widetilde{r}_t^2} - m_{t+1} \left(\frac{\widetilde{\rho}_{t+1}}{\widetilde{\theta}_{t+1}^2} \right)^2 \widehat{r}_{t+1}^2 \right)^{-1} \quad \textit{for } t = 1, ..., T-1 \\ m_1 &= 1,\ d_1 = d_{T+1} = 0,\ \lambda_t = \frac{\widetilde{\rho}_t^2 k_{t-1}}{2\widetilde{\theta}_t^2},\ \widehat{\lambda}_t = \lambda_t \frac{m_t \widehat{r}_t^2}{\widetilde{\theta}_t^2} \end{aligned}$$

the conditional posteriors are as follows:

$$\begin{aligned} m_t|Y, k, h &\sim B((n-1)/2 + h_t, 1/2), \\ k_t|Y, k_{1:(t-1)}, h_{1:t}, m, d &\sim G((n+1)/2 + h_t + d_{t+1}, 2\widehat{r}_t^2) \\ h_t|Y, k_{1:(t-1)}, h_{1:(t-1)}, m, d &\sim SP(\widehat{\lambda}_t, d_t) \end{aligned}$$

The conditional posterior $d|Y, k, h, m$ is the same as the conditional prior in (7). In addition:

$$
\begin{aligned}
p(\Theta|Y,m,d) \quad \propto \quad & \int p(\Theta)p(k,h,m,d|\Theta)L(Y|k)dkdh = \\
& \left[\prod_{t=1}^{T}\left(2\widehat{r}_t^2\right)^{\frac{n+1}{2}+d_{t+1}+d_t}\right]\left[\prod_{t=2}^{T}\left(m_t\left(\frac{\widetilde{\rho}_t}{2\widetilde{\theta}_t^2}\right)^2\right)^{d_t}\right]\left[\prod_{t=1}^{T}\Gamma\left(\frac{n+1}{2}+d_{t+1}\right)\right]\times \\
& \left[\prod_{t=2}^{T}\frac{1}{d_t!}\frac{[1+d_t]^{d_{t-1}}}{[(n-1)/2]_{d_t}}\right]\left[\prod_{t=2}^{T}m_t^{\alpha_m-1}(1-m_t)^{\beta_m-1}\right]\times \\
& \left(\Gamma\left(\frac{n}{2}\right)\right)^{-T} C_p C_L C_B p(\Theta)
\end{aligned}
$$

where

$$
\begin{aligned}
C_p &= \left(1-\widetilde{\rho}^2\right)^{n/2}\prod_{t=1}^{T}\left(2\widetilde{\theta}_t^2\right)^{-\frac{n}{2}} \\
C_L &= (2\pi)^{-T/2} \\
C_B &= \left(\frac{\Gamma(\alpha_m)\Gamma(\beta_m)}{\Gamma(\alpha_m+\beta_m)}\right)^{-(T-1)} \\
\alpha_m &= (n-1)/2, \quad \beta_m = 1/2
\end{aligned}
$$

3.2 Flexible Tail Autoregressive Gamma Process (FTARG)

As described later in Proposition 3, the conditional posterior density of $\widetilde{T}_t|V, h, \Theta$ is proportional to:

$$
\left(\widetilde{T}_t\right)^{\overline{\alpha}_t-1}\left(1-\widetilde{T}_t\right)^{V/2-1}\left(\frac{1}{1+\widetilde{T}_t S_t}\right)^{v_t} \tag{8}
$$

with:

$$
\begin{aligned}
\overline{\alpha}_t &= \frac{V}{2}+h_{t+1}+\frac{1}{2} \\
v_t &= \frac{n+1}{2}+h_t+h_{t+1} \\
S_t &= \theta^2(r_t^2+\rho^2/\theta^2)
\end{aligned}
$$

It can be seen that this kernel is that of an infinite mixture of beta distributions if we write the last term as a series (e.g. Muirhead (1985, p. 259)):

$$
\left(\frac{1}{1+\widetilde{T}_t S_t}\right)^{v_t} = \frac{1}{(1+S_t)^{v_t}}\sum_{s=0}^{\infty}\left(\frac{S_t}{1+S_t}(1-\widetilde{T}_t)\right)^s\frac{[v_t]_s}{s!}
$$

Thus one possibility to draw $\widetilde{T}_t$ is to draw from a mixture of betas. However, calculating the probability of each component of the mixture requires evaluation of the hypergeometric function ${}_2F_1(.)$, which could be computationally demanding. Another possibility is to draw from (8) using a Metropolis-step. However, this would require calibrating the proposal density. In this paper we facilitate the drawing of $\widetilde{T}_t$ by introducing an auxiliary variable J_t such that $\widetilde{T}_t|J_t$ and $J_t|\widetilde{T}_t$ can be both drawn from simple distributions. The variable J_t is introduced as a discrete random variable with probability density function:

$$\Pr\left(J_t = s|\widetilde{T}_t, S_t\right) = \left(\frac{1+\widetilde{T}_t S_t}{1+S_t}\right)^{v_t}\left(\frac{S_t}{1+S_t}(1-\widetilde{T}_t)\right)^s \frac{[v_t]_s}{s!} \tag{9}$$

Since the probabilities in (9) can be easily calculated, it is possible to draw J_t using the inverse transform sampling method (e.g. Gamerman and Lopes (2006, p. 13)). Furthermore, $\widetilde{T}_t$ conditional on J_t becomes a simple beta distribution $B(\overline{\alpha}_t, V/2 + J_t)$.

Therefore, a sampling algorithm for the FTARG model can be obtained by adding the following three steps to sample $\widetilde{T} = (\widetilde{T}_2, ..., \widetilde{T}_T)$, $J = (J_2, ..., J_T)$ and V to any of the two algorithms described in the previous section:

Additional Steps for the FTARG

- $V|(k,h), \Theta, J, \beta$ using a Metropolis step.
- $\widetilde{T}|(k,h), \Theta, J, V, \beta$ using beta distributions.
- $J|(k,h), \Theta, \widetilde{T}, J, V, \beta$ using the discrete distribution in (9).

Note that V is sampled marginally on $\widetilde{T}$ to increase the efficiency of the algorithm. However, for this reason, the step to sample $\widetilde{T}$ needs to come immediately after sampling V, so as to ensure that the algorithm converges to the joint posterior distribution.

Proposition 2 in the previous section and the following proposition describe the distributions that are necessary in this algorithm.

Proposition 3 *The conditional posterior densities for $\widetilde{T}$, and V in the FTARG model are as follows:*

$$\begin{aligned} \widetilde{T}_t|J_t &\sim B(\overline{\alpha}_t, V/2 + J_t) \\ p(V|Y,J) &\propto p(V)\left(\frac{\Gamma(V)}{\Gamma(V/2)\Gamma(V/2)}\right)^{T-1}\prod_{t=2}^{T}\frac{\Gamma(\overline{\alpha}_t)\Gamma(V/2+J_t)}{\Gamma(\overline{\alpha}_t+V/2+J_t)} \end{aligned}$$

The conditional posterior density for J_t is the same as the conditional prior given in (9).

4 Evidence on the Efficiency of the Algorithms

First let us compare the efficiency of the h-Gibbs and the m-Gibbs algorithm using 2000 daily observations of the exchange rate Yen - US dollar (6th Aug

	ESS		ESS/TIME	
	h-Gibbs	m-Gibbs	h-Gibbs	m-Gibbs
$\sigma^2_{T/2}$	0.0224	0.0755	64.5	151.0
$\overline{\sigma}^2$	0.0471	0.4472	135.7	894.4
n	0.0022	0.0135	6.2	26.9
θ	0.0002	0.0024	0.6	4.8
ρ	0.0003	0.0030	0.8	6.1
β_0	0.2701	0.7507	777.8	1501.4
β_1	0.5709	0.7790	1644.1	1558.0

Table 1: ESS and ESS/TIME for the h-Gibbs and the m-Gibbs algorithm using 2000 observations of the US-Japan exchange rate.

2003 - 15th Jul. 2011). y_t is the first difference of the log exchange rate and x_{t-1} includes a constant and a lag, so that $\beta = (\beta_0, \beta_1)$. We compare the efficiency of these algorithms using the effective sample size (e.g. Brooks (1999)). The effective sample size measures the number of independent draws from the posterior that is equivalent to 1 draw from an MCMC algorithm. Thus, algorithms with larger values of ESS are more efficient. Since the m-Gibbs takes more time per iteration in our implementation, we present also the ESS adjusted for computation time (ESS/TIME). We can see that the m-Gibbs is 12 times more efficient in terms of ESS for sampling θ^2 and ρ. When we control for computation time it is still 8 times more efficient. For the parameter n the m-Gibbs is about 4 times more efficient when controlling for computation time. The m-Gibbs is also about 6.6 times more efficient to sample the volatility at the middle of the sample ($\sigma^2_{T/2}$) and 2.3 times more efficient to sample the average volatility ($\overline{\sigma}^2 = \sum_{t=1}^{T} \sigma_t^2$) after adjusting for computation time. The same pattern can be observed with the other datasets that are used in this paper. The acceptance rate of the Metropolis-step to sample Θ was about 55% in both algorithms, which lies within the recommended range

5 Empirical Application

The aim of this section is to compare the empirical performance of several models using real macroeconomic and financial data. In addition to the TARG and FTARG described in the previous section, we consider the model where σ_t follows a log-normal distribution (LNORM), as in Kim et al (1998). In addition, we consider 3 models where ε_t follows a student-t distribution: TARG-T, FTARG-T and LNORM-T. These 3 models are the same as TARG, FTARG and LNORM models, respectively, but assume a student-t distribution for ε_t instead of normal. We run the models separately on 5 datasets, 4 of which are exchange rates (2 daily exchange rates and 2 monthly) and one dataset corresponds to UK inflation (see Table 2 for more details on the data). The dependent variable y_t is either the level of inflation or the first difference of the log exchange rate. When y_t is the exchange rate, x_t contains a constant and a lag of y_t. When y_t is inflation, x_t contains a constant, two lags of inflation, the unemployment rate and two lags of the unemployment rate (as in the estimation of a Phillips curve, e.g. Staiger et al. (1997) or Sargent et al. (2006)). The exchange rate data was obtained from the Federal Reserve Bank of St. Louis, and the inflation and unemployment rate data from OECD (2010).

Table 3 shows the value of the log-likelihood at the posterior median of parameters, calculated using a bootstrap particle filter (e.g. Gordon et al. 1993). In Table 3 we can see that the TARG model has a much higher value of the log likelihood than the LNORM and LNORM-T models for the monthly India-US and Brazil - US exchange rates. Furthermore, for these two exchange rates the FTARG model is much superior than all the other simpler models (by more than 20 points or 36 points increase in the log likelihood with respect to the TARG). The extension to student-t errors does not bring any noticeable improvement in the value of the log-likelihood of the TARG or FTARG models, although it does increase the log likelihood of the LNORM model. In summary, the FTARG is a clear winner in the case of the monthly India-US and Brazil - US exchange rates.

In the case of the Japan-US daily exchange rate, although the LNORM and LNORM-T are clearly superior to the TARG and TARG-T, the FTARG-T model seems to be the best as it gains 20 points in the log-likelihood over the second best model (LNORM-T) for just one extra parameter. For this dataset the assumption of student-t errors greatly improves the performance of all models.

Regarding the EU-US exchange rate, the LNORM-T and TARG-T are substantially better than LNORM and TARG, again indicating that it is important to allow for student-t errors. Both the LNORM-T and the TARG-T seem to perform equally well, whereas the FTARG and FTARG-T models do not bring any noticeable increase in the log likelihood. Hence, the LNORM-T and TARG-T could be said to joint winners for the EU-US exchange rate.

Finally, regarding the estimation of the Phillips curve for UK inflation, all models have very similar values for the log likelihood, indicating that the simpler models (LNORM and TARG) might be more adequate to model this series.

6 Conclusions

This paper has developed efficient posterior simulators for gamma and generalized gamma process for stochastic volatility. By conditioning on some auxiliary variables, it is shown that it is possible to draw all the volatilities jointly using simple distributions such as Poisson and Gamma. Furthermore, the unknown parameters can be drawn after integrating out the volatilities.

IND-US	Exchange rate Indian Rupee - US dollar, monthly average: March 1973 - June 2013, 484 observations
BRA-US	Exchange rate Brazilian Real - US dollar, monthly average: March 1995 - June 2013, 220 observations
JP-US	Exchange rate Japanese Yen - US dollar, daily: 6 Jan 1971 - 15 Jul 2011, 10168 observations
US-EU	Exchange rate Euro - US dollar, daily: 6 Jan 1999 - 17 May 2013, 3616 observations
UK-INFL	Quarterly Inflation based on GDP deflator, seasonally adjusted, 1971Q1 - 2011Q4, 162 observations.
UK-UR	Harmonized Unemployment Rate: All Persons for United Kingdom, seasonally adjusted, 1971Q1 - 2011Q4, 162 observations.

Table 2: Description of variables used in empirical analysis

	IND-US	BRA-US	JP-US	US-EU	UK-INFL
LNORM	1052.3	413.5	-7734.0	13275.9	-198.7
LNORM-T	1401.3	446.2	-7563.7	13284.8	-197.7
TARG	1427.1	489.3	-8228.1	13271.3	-197.5
TARG-T	1426.9	489.6	-7914.3	13283.4	-197.6
FTARG	1446.1	526.0	-7593.9	13273.1	-196.8
FTARG-T	1445.9	525.5	-7543.1	13280.5	-197.3

Table 3: Value of Log-Likelihood at th posterior median, calculated with particle filter for different models and datasets.

The empirical exercise shows that gamma and generalized gamma models outperform the lognormal volatility model with student-t errors specially in the datasets that exhibit greater instability, such as the exchange rate of Brazil-US or India-US.

References:

Asai, M. (2005) "Comparison of MCMC Methods for Estimating Stochastic Volatility Models," *Computational Economics*, 25, 281-301.

Barndorff-Neilsen, O. E. and N. Shephard (2001) "Non-Gaussian Ornstein-Ulhlenbeck-based models and some of their uses in financial economics" *Journal of the Royal Statistical Society*, Series B, 63, 167–241.

Abraham, B., N. Balakrishna and R. Sivakumar (2006) "Gamma Stochastic Volatility Models" *Journal of Forecasting*, 25, 153–171.

Casarin, R. and D. Sartore, (2007), "Matrix-state particle filters for Wishart stochastic volatility processes," in *Proceedings SIS, 2007 Intermediate Conference, Risk and Prediction*, 399-409, CLEUP Padova.

Chib, S., F. Nardarib and N. Shephard (2002) "Markov Chain Monte Carlo Methods for Stochastic Volatility Models," *Journal of Econometrics*, 108, 281 – 316.

Creal, D. (2013) "A Class of Non-Gaussian State Space Models with Exact Likelihood Inference," Available at SSRN: http://ssrn.com/abstract=2310256 or http://dx.doi.org/10.2139/ssrn.2310256

Devroye, L. (2002) "Simulating Bessel Random Variables," *Statistics & Probability Letters*, 57, 249–257.

Fox, E.B. and M. West (2011) "Autoregressive Models for Variance Matrices: Stationary Inverse Wishart Processes" arXiv:1107.5239.

Frühwirth-Schnatter, S. and L. Sögner (2009) "Bayesian estimation of stochastic volatility models based on OU processes with marginal Gamma law," *Annals of the Institute of Statistical Mathematics*, 61, 159–179.

Gamerman, D. and H.F. Lopes (2006) Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Chapman & Hall.

Geweke, J., 1996, Bayesian reduced rank regression in econometrics, *Journal of Econometrics*, 75, 121-146.

Gordon, N., D. Salmond and A.F.M. Smith, (1993) "Novel approach to nonlinear and non-Gaussian Bayesian state estimation," *Proc. Inst. Elect. Eng., F.*, 140, 107-113.

Gourieroux, C. and J. Jasiak (2006) "Autoregressive Gamma Processes" *Journal of Forecasting*, 129–152.

Gourieroux, C., J. Jasiak and R. Sufana (2009) "The Wishart Autoregressive process of multivariate stochastic volatility," *Journal of Econometrics*, 150, 167 - 181.

Griffin, J. E. and M.F.J. Steel, (2006) "Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility," *Journal of Econometrics*, 134, 605–644.

Iliopoulos, G. and D. Karlis (2003) "Simulation from the bessel distribution with applications," *Journal of Statistical Computation and Simulation*, 73:7, 491-506.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). "Bayesian analysis of stochastic volatility models (with discussion)" *Journal of Business and Economic Statistics,* 12, 371 - 389.

Johnson, N.L., S. Kotz and N. Balakrishnan (1995), *Continuous Univariate Distributions, Volume 2*, John Wiley & Sons, Inc.

Kim, S., N. Shephard, and S. Chib (1998) "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65, 361-393.

Krishnaiah, P. and M. Rao (1961), "Remarks on a multivariate Gamma distribution," *The American Mathematical Monthly*, 68, 342-346.

Mathai, A.M. and R.K. Saxena (1973), *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences*, Lecture Notes in Mathematics, Springer.

OECD (2010), "Main Economic Indicators - complete database", Main Economic Indicators (database), http://dx.doi.org/10.1787/data-00052-en (Accessed on Feb 2014)

Philipov, A. and M. E. Glickman (2006) "Multivariate stochastic volatility via Wishart processes," *Journal of Business and Economic Statistics* 24, 313–328.

Pitt, M. K. and N. Shephard (1999) "Time varying covariances: a factor stochastic volatility approach" In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), Bayesian Statistics, Volume 6, pp. 547–570. Oxford: Oxford University Press.

Roberts, G. O., O. Papaspiliopoulos and P. Dellaportas, (2004) "Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes" *Journal of Royal Statistical Society*, Series B, 66, 369–393.

Sargent, T.,N. Williams, and T. Zha, (2006), "Shocks and Government Beliefs: The Rise and Fall of American Inflation," *American Economic Review*, 96, 1193–1224.

Shephard, N. and M. K. Pitt (1997). "Likelihood analysis of non-Gaussian measurement time series" *Biometrika* 84, 653–667.

Sims, C. and Zha, T., 2006. Were there regime switches in macroeconomic policy? *American Economic Review* 96, 54-81.

Slater, L.J. (1966), *Generalized Hypergeometric Functions*, Cambridge University Press.

Staiger, D., J. Stock and M. Watson (1997), “The NAIRU, Unemployment and Monetary Policy,” *Journal of Economic Perspectives*, 11, 33–49.

Steel, M. (1998) "Bayesian Analysis of Stochastic Volatility Models with Flexible Tails," *Econometric Reviews*, 17, 109-143.

Triantafyllopoulos, K. (2010) "Multi-variate stochastic volatility modelling using Wishart autoregressive processes" *Journal of Time Series Analysis*, 33, 48-60.
Watanabe, T. and Y. Omori (2004) "A multi-move sampler for estimating non-Gaussian times series models: Comments on Shephard and Pitt (1997)" *Biometrika* 91, 246–248.

Weisstein, E. W. "Chu-Vandermonde Identity." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Chu-VandermondeIdentity.html

Winkelmann, R. (2008) Econometric Analysis of Count Data, Springer.

7 Appendix

Proof of the equivalence between (2) and (4)

First let us verify that integrating out m_t from (4) gives a Poisson distribution with parameter λ_t:

$$\begin{aligned}
\int \Pr(h_t &= s|k_{1:(t-1)}, h_{1:(t-1)}, m_{1:t})p(m_t|m_{1:(t-1)})dm_t \\
&= \int \frac{\frac{\lambda_t^s}{s!}m_t^s\frac{[n/2]_s}{[(n-1)/2]_s}}{\exp(\lambda_t)}\frac{\Gamma(\alpha_m+\beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)}m_t^{\alpha_m-1}(1-m_t)^{\beta_m-1}dm \\
&= \frac{\frac{\lambda_t^s}{s!}}{\exp(\lambda_t)}
\end{aligned}$$

which is the probability density function for a $P(\lambda_t)$ and where we have used that if $m_t \sim B(\alpha_m, \beta_m)$ then $E(m_t^s) = \frac{[\alpha_m]_s}{[\alpha_m+\beta_m]_s}$ (Johnson et al. 1995, p. 217).

It can also be verified that $\int p(m_t|m_{1:(t-1)}) = 1$ by the following derivation:

$$\begin{aligned}
\int p(m_t|m_{1:(t-1)})dm &= \frac{\Gamma(\alpha_m+\beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)}\frac{1}{\exp(\lambda_t)}\sum_{s=0}^{\infty}\frac{\lambda_t^s}{s!}\frac{[n/2]_s}{[(n-1)/2]_s}\int m_t^{\alpha_m+s-1}(1-m_t)^{\beta_m-1}dm_t \\
&= \frac{\Gamma(\alpha_m+\beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)}\frac{1}{\exp(\lambda_t)}\sum_{s=0}^{\infty}\frac{\lambda_t^s}{s!}\frac{[\alpha_m+\beta_m]_s}{[\alpha_m]_s}\frac{\Gamma(\alpha_m+s)\Gamma(\beta_m)}{\Gamma(\alpha_m+s+\beta_m)}
\end{aligned}$$

Using the following properties of the gamma function:

$$\begin{aligned}
\Gamma(\alpha_m+\beta_m+s) &= \Gamma(\alpha_m+\beta_m)[\alpha_m+\beta_m]_s \\
\Gamma(\alpha_m+s) &= \Gamma(\alpha_m)[\alpha_m]_s
\end{aligned}$$

we obtain that $\int p(m_t|m_{1:(t-1)}) = 1$.

Proof of Proposition 1
The likelihood is:

$$L(Y|k) = (2\pi)^{-T/2}\left[\prod_{t=1}^{T}(k_t)^{1/2}\right]\exp\left(-\frac{1}{2}\sum_{t=1}^{T} r_t^2 k_t\right) \qquad r_t^2 = (y_t - x_t\beta)^2$$

The prior $p(k, h|\Theta)$ is equal to:

$$\begin{aligned} & p(k_1|\Theta)\prod_{t=2}^{T}\left(p(k_t|h_t,\Theta)p(h_t|k_{t-1},\Theta)\right) \\ = \; & p(k_1|\Theta)\prod_{t=2}^{T}\left(p(k_t|h_t,\Theta)\frac{\frac{\lambda_t^{h_t}}{h_t!}}{\exp(\lambda_t)}\right) \end{aligned}$$

The densities $p(k_1|\Theta)$ and $p(k_t|h_t,\Theta)$ are Gamma densities:

$$\begin{aligned} p(k_1|\Theta) &= \frac{|k_1|^{\frac{n-2}{2}}}{c_1}\exp\left(-\frac{1-\widetilde{\rho}^2}{2\widetilde{\theta}^2}k_1\right) \qquad c_1 = \Gamma\left(\frac{n}{2}\right)\left(\frac{2\widetilde{\theta}^2}{1-\widetilde{\rho}^2}\right)^{n/2} \qquad (10) \\ p(k_t|h_t,\Theta) &= \frac{|k_t|^{\frac{n+2h_t-2}{2}}}{c_t}\exp\left(-\frac{1}{2\widetilde{\theta}_t^2}k_t\right) \qquad c_t = \Gamma\left(\frac{n}{2}+h_t\right)\left(2\widetilde{\theta}_t^2\right)^{n/2+h_t} \qquad t = 2,...,T \end{aligned}$$

Thus, the product of the prior and the likelihood, $p(\Theta)p(k,h|\Theta)L(Y|k)$, can be written as:

$$\begin{aligned} & (2\pi)^{-T/2}\left[\prod_{t=1}^{T}(k_t)^{\frac{n+2h_t-2}{2}+\frac{1}{2}}\right]\exp\left(-\frac{1}{2}\sum_{t=2}^{T}k_t\left(\frac{1}{\widetilde{\theta}_t^2}+r_t^2\right)\right)\times \qquad (11) \\ & \exp\left(-\frac{1}{2}k_1\left(\frac{1-\widetilde{\rho}^2}{\widetilde{\theta}^2}+r_t^2\right)\right)\prod_{t=2}^{T}\left(\frac{\frac{\lambda_t^{h_t}}{h_t!}}{\exp(\lambda_t)}\right)\left(\prod_{t=1}^{T}c_t\right)^{-1}p(\Theta) \end{aligned}$$

Recalling that $\lambda_t = \widetilde{\rho}_t^2 k_{t-1}/(2\widetilde{\theta}_t^2)$, it is clear that $k_t|h, Y \sim G((n+1)/2 + h_t + h_{t+1}, 2\widetilde{\theta}_t^2)$. To find the conditional distribution of h given k note that c_t depends on h_t and also that standard properties of the gamma function (e.g. Slater (1966, p.3)) imply that:

$$\Gamma\left(\frac{n}{2}+h_t\right) = [n/2]_{h_t}\Gamma\left(\frac{n}{2}\right)$$

Thus, putting together the terms in (11) that depend on h_t we get:

$$\prod_{t=2}^{T}\left(\frac{1}{h_t!}\frac{1}{[n/2]_{h_t}}\left(\left(\frac{\widetilde{\rho}_t}{2\widetilde{\theta}_t^2}\right)^2 k_t k_{t-1}\right)^{h_t}\right)$$

which shows that $h_t|k,Y \sim Bessel(\frac{n-2}{2}, \widetilde{\rho}_t \frac{\sqrt{k_t k_{t-1}}}{\widetilde{\theta}_t^2})$ for $t = 2...T$. The expression for $p(\theta^2, n, \rho^2|Y, h)$ can be obtained by integrating (11) with respect to k using basic properties of the Gamma function.

Proof of Proposition 2:

For the proof it will be convenient to rewrite the denominator of (7) more compactly as a ratio of two hypergeometric coefficients. For this purpose we can use the Chu-Vandermonde identity (e.g. Slater (1966, p.2), Mathai and Saxena (1973, p. 110), or Weisstein) which states that:

$$\frac{[x+d]_h}{[x]_h} = {}_2F_1(-h, -d; x; 1) \tag{12}$$

where ${}_2F_1()$ is a hypergeometric function. If d is an integer and $(0 \leq d \leq h)$, this function can be written as:

$${}_2F_1(-h, -d; x; 1) = \sum_{s=0}^{d} \frac{1}{s!} \frac{[h]^s [d]^s}{[x]_s} \tag{13}$$

Therefore, (7) can be rewritten as:

$$\Pr(d_t = s|h_t, d_{t+1}) = \frac{[(n-1)/2]_{h_t}}{[(n+1)/2 + d_{t+1}]_{h_t}} \frac{[h_t]^s}{s!} \frac{[1+d_{t+1}]^s}{[(n-1)/2]_s}$$

Thus, the joint prior of $(d = (d_2, ..., d_T))$ given (h, k, m), denoted as $\pi(d|h, k, m)$, can be written as:

$$\prod_{t=T}^{2} p(d_t|h_t, d_{t+1}) = \prod_{t=T}^{2} \left(\frac{[(n-1)/2]_{h_t}}{[(n+1)/2 + d_{t+1}]_{h_t}} \frac{[h_t]^{d_t}}{d_t!} \frac{[1+d_{t+1}]^{d_t}}{[(n-1)/2]_{d_t}} \right) \quad \text{with} \quad d_{T+1} = 0$$

and we will also use the notation $p(d_{2:T-l}|d_{T-l+1}, h, k, m)$ for:

$$p(d_{2:T-l}|d_{T-l+1}, h, k, m) = \prod_{t=T-l}^{2} p(d_t|h_t, d_{t+1})$$

The prior $p(k, h, m|\Theta)$ is equal to:

$$\begin{aligned} & p(k_1) \prod_{t=2}^{T} (p(k_t|h_t) p(h_t|m_t, k_{t-1}) p(m_t|k_{t-1})) \\ = \quad & p(k_1) \prod_{t=2}^{T} \left(p(k_t|h_t) \frac{\frac{\lambda_t^{h_t}}{h_t!} m_t^{h_t} \frac{[n/2]_{h_t}}{[(n-1)/2]_{h_t}}}{\exp(\lambda_t)} \frac{\Gamma(\alpha_m + \beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)} m_t^{\alpha_m - 1} (1-m_t)^{\beta_m - 1} \right) \end{aligned}$$

where $p(k_1)$ and $p(k_t|h_t)$ have been defined in (10).

Thus, the product of the prior and the likelihood, $p(k,h,m,d|\Theta)L(Y|k)$, can be written as:

$$(2\pi)^{-T/2}\left[\prod_{t=1}^{T}(k_t)^{\frac{n+2h_t-2}{2}+\frac{1}{2}}\right]\exp\left(-\frac{1}{2}\sum_{t=1}^{T}k_t\left(\frac{1}{\widetilde{\theta}^2}+r_t^2\right)\right)\times$$
$$\prod_{t=2}^{T}\left(\frac{\frac{\lambda_t^{h_t}}{h_t!}m_t^{h_t}\frac{[n/2]_{h_t}}{[(n-1)/2]_{h_t}}}{\exp(\lambda_t)}\frac{\Gamma(\alpha_m+\beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)}m_t^{\alpha_m-1}(1-m_t)^{\beta_m-1}\right)\times$$
$$\pi(d|h,k,m)\left(\prod_{t=1}^{T}c_t\right)^{-1}$$

where h_1 is fixed as $h_1=0$. It is clear that the conditional posterior of $k_T|h_T,m,d$ is a $G((n+1)/2+h_T,2\widehat{r}_T^2)$. Integrating out k_T we find:

$$\Gamma\left(\frac{n+1+2h_T}{2}\right)\left(2\widehat{r}_T^2\right)^{\frac{n+1+2h_T}{2}}(2\pi)^{-T/2}\left[\prod_{t=1}^{T-1}(k_t)^{\frac{n+2h_t-2}{2}+\frac{1}{2}}\right]\times \quad (14)$$
$$\exp\left(-\frac{1}{2}\sum_{t=1}^{T-1}k_t\left(\frac{1}{\theta^2}+r_t\right)\right)\prod_{t=2}^{T}\left(\frac{\frac{\lambda_t^{h_t}}{h_t!}m_t^{h_t}\frac{[n/2]_{h_t}}{[(n-1)/2]_{h_t}}}{\exp(\lambda_t)}\frac{\Gamma(\alpha_m+\beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)}m_t^{\alpha_m-1}(1-m_t)^{\beta_m-1}\right)\times$$
$$p(d|h,k,m)\left(\prod_{t=1}^{T}c_t\right)^{-1}$$

Again using the properties of the gamma function (e.g. Slater (1966, p.3)), it can be showed that:

$$\begin{aligned}\Gamma\left(\frac{n+1}{2}+h_T\right) &= [(n+1)/2]_{h_T}\Gamma\left(\frac{n+1}{2}\right)\\ \Gamma\left(\frac{n}{2}+h_T\right) &= [n/2]_{h_T}\Gamma\left(\frac{n}{2}\right)\end{aligned}$$

In order to find out the posterior conditional of h_T, note that c_T depends on h_T and so the terms that contain h_T in expression (14) can be written as:

$$\begin{aligned}&[(n+1)/2]_{h_T}\left(2\widehat{\theta}_T^2\right)^{h_T}\frac{\lambda_T^{h_T}}{h_T!}m_T^{h_T}\frac{[n/2]_{h_T}\left(2\theta^2\right)^{-h_T}}{[(n-1)/2]_{h_T}}\frac{[h_T]^{d_T}}{[n/2]_{h_T}}\times \quad (15)\\ &\frac{[(n-1)/2]_{h_T}}{[(n+1)/2+d_{T+1}]_{h_T}}\\ =\;&\left(\widehat{\lambda}_T\right)^{h_T}\frac{1}{h_T!}[h_T]^{d_T}\end{aligned}$$

where we have implicitly defined:

$$\widehat{\lambda}_T=\left(\lambda_T\frac{m_T\widehat{r}_T^2}{\widetilde{\theta}_T^2}\right)\text{ and }d_{T+1}=0$$

Note that (15) can be written as:

$$\left(\widehat{\lambda}_T\right)^{h_T} \frac{1}{h_T!}[h_T]^{d_T} = \left(\widehat{\lambda}_T\right)^{h_T} \frac{1}{(h_T - d_T)!} \quad \text{for} \;\; h_T \in [d_T, \infty) \tag{16}$$

From which it is clear that $h_T|k_{T-1}, m, d$ is a $SP(\widehat{\lambda}_T, d_T)$. Summing up expression (16) over all values of h_T gives $\left(\widehat{\lambda}_T\right)^{d_T} \exp(\widehat{\lambda}_T)$. Thus, integrating out h_T from (14) we obtain:

$$\begin{aligned}
&\Gamma\left(\frac{n+1}{2}\right)\left(2\widehat{r}_T^2\right)^{\frac{n+1}{2}}(2\pi)^{-T/2}\left[\prod_{t=1}^{T-1}(k_t)^{\frac{n+2h_t-2}{2}+\frac{1}{2}}\right]\times \\
&\left(\widehat{\lambda}_T\right)^{d_T}\exp\left(-(\lambda_T-\widehat{\lambda}_T)\right)\frac{1}{d_T!}\frac{1}{[(n-1)/2]_{d_T}}\times \\
&\exp\left(-\frac{1}{2}\sum_{t=1}^{T-1}k_t\left(\frac{1}{\widetilde{\theta}_t^2}+r_t\right)\right)\prod_{t=2}^{T-1}\left(\frac{\frac{\lambda_t^{h_t}}{h_t!}m_t^{h_t}\frac{[n/2]_{h_t}}{[(n-1)/2]_{h_t}}}{\exp(\lambda_t)}\right)\times \\
&\prod_{t=2}^{T}\left(\frac{\Gamma(\alpha_m+\beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)}m_t^{\alpha_m-1}(1-m_t)^{\beta_m-1}\right)\times \\
&\left(\Gamma\left(\frac{n}{2}\right)\left(2\widetilde{\theta}_T^2\right)^{n/2}\right)^{-1}p(d_{2:T-1}|d_T,h,k,m)\left(\prod_{t=1}^{T-1}c_t\right)^{-1}
\end{aligned} \tag{17}$$

Noting that:

$$\begin{aligned}
\exp\left(-(\lambda_T-\widehat{\lambda}_T)\right) &= \exp\left(-\frac{1}{2}\left(\frac{\widetilde{\rho}_T^2}{\widetilde{\theta}_T^2}-m_T\left(\frac{\widetilde{\rho}_T}{\widetilde{\theta}_T^2}\right)^2\widehat{r}_T^2\right)k_{T-1}\right) \\
\exp\left(-\frac{1}{2}k_{T-1}\left(\frac{1}{\widetilde{\theta}_t^2}+r_t^2\right)\right)\exp\left(-(\lambda_T-\widehat{\lambda}_T)\right) &= \exp\left(-\frac{1}{2\widehat{r}_{T-1}^2}k_{T-1}\right)
\end{aligned}$$

we can see that the conditional posterior $k_{T-1}|h_{T-1}, m, d$ is a $G((n+1)/2+h_{T-1}+d_T, 2\widehat{r}_{T-1}^2)$. Thus, integrating out k_{T-1} from (17) we obtain:

$$\begin{aligned}
&\Gamma\left(\frac{n+1}{2}\right)\left(2\widehat{r}_T^2\right)^{\frac{n+1}{2}}\Gamma\left(\frac{n+1}{2}+h_{T-1}+d_T\right)\times \\
&\left(2\widehat{r}_{T-1}^2\right)^{\frac{n+1}{2}+h_{T-1}+d_T}\left(\frac{m_T}{2}\left(\frac{\widetilde{\rho}_T}{\widetilde{\theta}_T^2}\right)^2\widehat{r}_T^2\right)^{d_T}(2\pi)^{-T/2}\left[\prod_{t=1}^{T-2}(k_t)^{\frac{n+2h_t-2}{2}+\frac{1}{2}}\right]\times \\
&\exp\left(-\frac{1}{2}\sum_{t=1}^{T-2}k_t\left(\frac{1}{\widetilde{\theta}_t^2}+r_t\right)\right)\prod_{t=2}^{T-1}\left(\frac{\frac{\lambda_t^{h_t}}{h_t!}m_t^{h_t}\frac{[n/2]_{h_t}}{[(n-1)/2]_{h_t}}}{\exp(\lambda_t)}\right)\times \\
&\prod_{t=2}^{T}\left(\frac{\Gamma(\alpha_m+\beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)}m_t^{\alpha_m-1}(1-m_t)^{\beta_m-1}\right)\times \\
&\left(\Gamma\left(\frac{n}{2}\right)\left(2\widetilde{\theta}_T^2\right)^{n/2}\right)^{-1}p(d_{2:T-1}|d_T,h,k,m)\left(\prod_{t=1}^{T-1}c_t\right)^{-1}
\end{aligned} \tag{18}$$

The terms that depend on h_{T-1} are:

$$\Gamma\left(\frac{n+1}{2}+h_{T-1}+d_T\right)\left(2\widehat{r}_{T-1}^2\right)^{h_{T-1}}\left(\Gamma\left(\frac{n}{2}+h_{T-1}\right)\right)^{-1}\times \tag{19}$$

$$\left(\left(2\widetilde{\theta}_{T-1}^2\right)^{h_{T-1}}\right)^{-1}\frac{\lambda_{T-1}^{h_{T-1}}}{h_{T-1}!}m_{T-1}^{h_{T-1}}\frac{[n/2]_{h_{T-1}}}{[(n-1)/2]_{h_{T-1}}}[h_{T-1}]^{d_{T-1}}\frac{[(n-1)/2]_{h_{T-1}}}{[(n+1)/2+d_T]_{h_{T-1}}}$$

Again noting that from the properties of the gamma function we have that:

$$\Gamma\left(\frac{n+1}{2}+d_T+h_{T-1}\right)=[(n+1)/2+d_T]_{h_{T-1}}\Gamma\left(\frac{n+1}{2}+d_T\right)$$

So (19) can be written as:

$$\frac{\Gamma\left((n+1)/2+d_T\right)}{\Gamma\left(n/2\right)}\left(\lambda_{T-1}\frac{m_{T-1}\widehat{r}_{T-1}^2}{\widetilde{\theta}_{T-1}^2}\right)^{h_{T-1}}\frac{1}{h_{T-1}!}[h_{T-1}]^{d_{T-1}} \tag{20}$$

From which it is clear that h_{T-1} follows a poisson with parameter $\widehat{\lambda}_{T-1}$:

$$\widehat{\lambda}_{T-1}=\lambda_{T-1}\frac{m_{T-1}\widehat{r}_{T-1}^2}{\theta^2}$$

Therefore, if we integrate out h_{T-1} from (18) we get:

$$\left(\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{n+1}{2}+d_T\right)\right)\left(2\widehat{r}_T^2\right)^{\frac{n+1}{2}}\left(2\widehat{r}_{T-1}^2\right)^{\frac{n+1}{2}+d_T}$$

$$\left(\frac{m_T}{2}\left(\frac{\widetilde{\rho}_T}{\widetilde{\theta}_T^2}\right)^2\widehat{r}_T^2\right)^{d_T}\left(\frac{m_{T-1}}{2}\left(\frac{\widetilde{\rho}_{T-1}}{\widetilde{\theta}_{T-1}^2}\right)^2\widehat{r}_{T-1}^2\right)^{d_{T-1}}(2\pi)^{-T/2}\left[\prod_{t=1}^{T-2}(k_t)^{\frac{n+2h_t-2}{2}+\frac{1}{2}}\right]$$

$$(k_{t-2})^{d_{T-1}}\exp\left(-(\lambda_{T-1}-\widehat{\lambda}_{T-1})\right)\frac{1}{d_T!}\frac{1}{[(n-1)/2]_{d_T}}\frac{1}{d_{T-1}!}\frac{[1+d_T]^{d_{T-1}}}{[(n-1)/2]_{d_{T-1}}}$$

$$\exp\left(-\frac{1}{2}\sum_{t=1}^{T-2}k_t\left(\frac{1}{\widetilde{\theta}_t^2}+r_t\right)\right)\prod_{t=2}^{T-2}\left(\frac{\frac{\lambda_t^{h_t}}{h_t!}m_t^{h_t}\frac{[n/2]_{h_t}}{[(n-1)/2]_{h_t}}}{\exp(\lambda_t)}\right)$$

$$\prod_{t=2}^{T}\left(\frac{\Gamma(\alpha_m+\beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)}m_t^{\alpha_m-1}(1-m_t)^{\beta_m-1}\right)$$

$$\left(\Gamma\left(\frac{n}{2}\right)\right)^{-2}\left(2\widetilde{\theta}_T^2\right)^{n/2}\left(2\widetilde{\theta}_{T-1}^2\right)^{n/2}p(d_{2:T-2}|d_{T-1},h,k,m)\left(\prod_{t=1}^{T-2}c_t\right)^{-1}$$

The other results in Proposition 2 can be obtained by using similar operations to recursively integrate out $(k_{t-2},h_{t-2},...,k_2,h_2,k_1)$.

Proof of Proposition 3

The conditional posterior of $\widetilde{T}$, which is given in (8), comes simply from finding the terms that depend on $\widetilde{T}$ in expression (3) in Proposition 1. Multiplying expression (8) times the conditional prior of J (9) gives $\widetilde{T}|J$, which is clearly a

Beta distribution. Similarly, the conditional posterior of $V|\widetilde{T}, J$ is proportional to expression (3) times the prior of V and times the prior of J:

$$p(V)\left(\frac{\Gamma(\underline{\alpha}+\underline{\beta})}{\Gamma(\underline{\alpha})\Gamma(\underline{\beta})}\right)^{T-1}\prod_{t=2}^{T}\left[\left(\widetilde{T}_t\right)^{\overline{\alpha}_t-1}\left(1-\widetilde{T}_t\right)^{\underline{\beta}+J_t-1}\right]$$

Using the properties of the Beta distribution we can integrate $\widetilde{T}$ from this expression to obtain the desired result.