Efficient Bayesian Inference in Generalized Inverse Gamma Processes for Stochastic Volatility

Roberto Leon-Gonzalez

September 2014

GRIPS

NATIONAL GRADUATE INSTITUTE
FOR POLICY STUDIES

Efficient Bayesian Inference in Generalized Inverse Gamma Processes for Stochastic Volatility*

Roberto Leon-Gonzalez

National Graduate Institute for Policy Studies (GRIPS)

September 2014

Abstract

This paper develops a novel and efficient algorithm for Bayesian inference in inverse 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 inverse Gamma process with more flexible tails in the distribution of volatilities, which still allows for simple and efficient calculations. Using several macroeconomic and financial datasets, it is shown that the inverse Gamma and Generalized inverse Gamma processes can greatly outperform the commonly used log normal volatility processes with student-t errors.

Keywords: Markov Chain Monte Carlo, Gibbs Sampling, Flexible Parametric Models, Particle Filters.

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^2 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^2 is generated conditionally on the volatility values in other periods $(\sigma_1^2, ..., \sigma_{t-1}^2, \sigma_{t+1}^2, ..., \sigma_T^2)$. 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 log-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^2 follows some type of gamma or inverse 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. In the multivariate context Casarin and Sartore (2007) proposed sequential monte carlo and particle filters for estimation of the states and parameters and Triantafyllopoulos (2010) proposed a simplified Wishart stochastic volatility model which allows for fast and simple computations. Abraham et al. (2006) proposed method of moments estimators for gamma type univariate stochastic volatility models and Gourieroux et al. (2009) develop maximum likelihood inference for a Wishart autoregressive process for observed volatility. There is also a recent literature that deals with Ornstein-Ulhlenbeck processes with marginal gamma laws (e.g. Barndorff-Nielsen and Shephard (2001), Roberts et al. (2004), Griffin and Steel (2006a), Frühwirth-Schnatter and Sögner (2009)).

A related strand of literature proposes flexible models for stochastic volatility. Although there are many papers that provide alternative methods to model flexibly the distribution of the observed dependent variable (e.g. Steel (1998), Durham (2007), Jensen and Maheu (2010), Delatola and Griffin (2011), Griffin and Steel (2011)), there are few that model flexibly the distribution of the unobserved volatility. As argued by Janssen and Drees (2013), the latter approach is more appropriate in datasets that exhibit persistence of volatility outliers. In this line Griffin and Steel (2006b) and Jensen and Maheu (2014) provide semiparametric methods of inference based on infinite mixtures for the volatility distribution. However, there is a lack of models that specify the volatility process in a flexible yet parametric manner. Flexible parametric models could potentially perform better than semiparametric ones in some datasets, while taking advantage of simpler and more efficient computational methods.

The purpose of this paper is to develop efficient posterior simulators for flexible inverse 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 generate the unknown parameters after integrating out all the volatilities. Because of these features, our algorithm mimics the efficient algorithm that Kim et al (1998) developed for the lognormal model, without

requiring the use of an approximation to the likelihood. Moreover, this paper proposes a generalized inverse gamma time-series model that specifies a more flexible distribution for the volatility, allows for more abrupt jumps in volatility, and can be estimated using simple and efficient methods. In an empirical exercise we show that the generalized inverse gamma process is especially suitable to model series with greater volatility jumps and persistence in outliers. Furthermore, we use real and simulated data to illustrate the efficiency of the new algorithm and show that it is much more efficient than the recently proposed Particle Markov Chain Monte Carlo methods (Andrieu et al. 2010) which sample the volatilities and parameters in a joint move using a particle filter.

This paper differs from previous work on gamma type stochastic volatility models in two main aspects. Firstly, we find a method to sample all the volatilities jointly from the posterior using well-known distributions such as the Poisson and Gamma, whereas previous work mostly used single-move or blocking strategies in a Metropolis-step to sample the volatilities. As mentioned before, sampling the volatilities jointly from the posterior is an important characteristic of efficient algorithms. Secondly, we develop and study the properties of a flexible inverse gamma time series model that can be estimated with simple and efficient computations. Thus this paper provides a new class of flexible stochastic volatility models that can be estimated with simple and efficient MCMC methods.

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

2   Models

2.1   The Autoregressive Gamma Process (ARG)

We consider the following model of stochastic volatility:

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

Although for simplicity in the exposition we are assuming normality for e_t, in the empirical applications we will consider also models where e_t follows a student-t. The student-t can be easily incorporated into this framework by writing it as a scale mixture of normals, as in Chib et al (2002). The stochastic process for the volatility σ_t^2 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 (or intertrade durations) 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), Gourieroux and Jasiak (2006)) 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 $k_1 \sim G(n/2, 2\theta^2/(1-\rho^2))$. Note finally that the autocorrelations are defined by ρ^2, so that they cannot be negative. In fact ρ enters the likelihood always in the form of ρ^2, so that the sign of ρ is not identified. For this reason in our empirical section we will specify the prior not on ρ but directly on ρ^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 it also controls 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 k_t. There is previous literature that develops more flexible gamma-type distributions, such as the generalized gamma distribution of Stacy (1962) or the compound gamma of Dubey (1970) (see also Johnson et al. (1994, section 17.8) for a review). However, here we propose a different type of distribution that lends itself better to the context of time-series and the use of MCMC methods for computation. For this purpose we define 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:

$$z_t = \sqrt{\widetilde{T}_t}(\rho z_{t-1} + \varepsilon_t) \tag{2}$$

where $(\widetilde{T}_2, ..., \widetilde{T}_T)$ are independent draws from a Beta distribution $B(\underline{\alpha}, \underline{\beta})$. Given that we are more concerned with modelling the left tail of k_t (which corresponds to the right tail of σ_t^2) and given that the stationarity of the process requires $E(\widetilde{T}_t) < 1/\rho^2$, it seems appropriate to specify a distribution with bounded support for $\widetilde{T}_t$.

4

If we write $\widetilde{\rho}_t = \rho\sqrt{\widetilde{T}_t}$ 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, and therefore the FTARG is equivalent to the ARG with time-varying parameters. Furthermore, the FTARG can be also compared to the ARG process by defining $\widetilde{\rho} = \sqrt{E(\widetilde{T}_t)}\rho$, $\widetilde{\theta}^2 = E(\widetilde{T}_t)\theta^2$ and $\widetilde{\varepsilon}_t \sim N(0, \widetilde{\theta}^2)$, such that (2) can be equivalently written as:

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

so that $k_t = z_t'z_t$ becomes:

$$k_t = \frac{\widetilde{T}_t}{E(\widetilde{T}_t)}(\widetilde{\rho}z_{t-1} + \widetilde{\varepsilon}_t)'(\widetilde{\rho}z_{t-1} + \widetilde{\varepsilon}_t) \tag{3}$$

From this expression it is clear that when $\widetilde{T}_t > E(\widetilde{T}_t)$ ($\widetilde{T}_t < E(\widetilde{T}_t)$), the value of k_t is higher (lower) than in the ARG model, which adds flexibility to the model. Furthermore, when the variance of $\widetilde{T}_t$ approaches 0, the ratio $\widetilde{T}_t/E(\widetilde{T}_t)$ behaves as a constant of value 1, and therefore the FTARG becomes equivalent to the ARG. However this implies that 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 fix $E(\widetilde{T}_t) = 1/2$. For this purpose, we reparameterize $(\underline{\alpha}, \underline{\beta})$ as $A = E(\widetilde{T}_t) = \underline{\alpha}/(\underline{\alpha} + \underline{\beta})$ 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 and are summarized in the following proposition whose proof is in the appendix.

Proposition 1 *Define* $\widetilde{\rho}^2 = E(\widetilde{T}_t)\rho^2$ *and* $\widetilde{\theta}^2 = E(\widetilde{T}_t)\theta^2$. *The main properties of* $(k_1, ..., k_T)$ *implied by (2) are:*

$$E(k_t|k_{t-1}) = \widetilde{\rho}^2 k_{t-1} + (1 - \widetilde{\rho}^2)E(k_t) \qquad if\ \widetilde{\rho}^2 < 1 \tag{4}$$

$$corr(k_t, k_{t-h}) = \widetilde{\rho}^{2h} \qquad if\ \widetilde{\rho}^2 < 1 \tag{5}$$

$$E(k_t|k_{t-1}, \widetilde{T}_t) = \frac{\widetilde{T}_t}{E(\widetilde{T}_t)}(\widetilde{\rho}^2 k_{t-1} + (1 - \widetilde{\rho}^2)E(k_t)) \tag{6}$$

$$E(k_t) = \frac{n\widetilde{\theta}^2}{1 - \widetilde{\rho}^2} \quad if\ \widetilde{\rho}^2 < 1 \tag{7}$$

$$E(k_t^2) = (\theta^2)^2 n(n+2)E(v_{c,t}^2) \quad if\ \rho^4 E(\widetilde{T}_t^2) < 1 \tag{8}$$

where $v_{c,t} = \widetilde{T}_t(1 + \rho^2\widetilde{T}_{t-1} + \rho^4\widetilde{T}_{t-1}\widetilde{T}_{t-2} + \rho^6\widetilde{T}_{t-1}\widetilde{T}_{t-2}\widetilde{T}_{t-3} + ...)$ *and:*

$$E(v_{c,t}^2) = \frac{E(T^2)(1 + \widetilde{\rho}^2)}{(1 - \widetilde{\rho}^2)(1 - \rho^4 E(T^2))} \quad if\ \rho^4 E(\widetilde{T}_t^2) < 1$$

Higher moments of k_t *are given by:*

$$E(k_t^s) = E(v_{c,t}^s)(\theta^2)^s \prod_{i=0}^{s-1}(n + 2i)\ if\ \rho^{2s}E(\widetilde{T}_t^s) < 1$$

where $E(v_{c,t}^s)$ can be calculated recursively as:

$$E(v_{c,t}^s) = \frac{E(\widetilde{T}_t^s)}{1 - \rho^{2s}E(\widetilde{T}_t^s)} \sum_{i=0}^{s-1} \binom{s}{i} \rho^{2i} E(v_{c,t}^i) \quad if \ \rho^{2s}E(\widetilde{T}_t^s) < 1 \tag{9}$$

and where the properties of the Beta distribution imply that:

$$E(\widetilde{T}_t^s) = \prod_{i=0}^{s-1} \frac{\underline{\alpha} + i}{\underline{\alpha} + \underline{\beta} + i}$$

The stationary distribution of k_t is that of the product of $\varepsilon_t'\varepsilon_t$ (i.e. a gamma distribution) and $v_{c,t}$, where $\varepsilon_t'\varepsilon$ and $v_{c,t}$ are independent of each other. The sth moment $E((\sigma_t^2)^s) = E(k_t^{-s})$ is finite if and only if $\underline{\alpha} > s$ and $n > s$.

Since $E(\widetilde{T}_t)$ is normalized to be $1/2$, the condition for the first order moment of the stationary distribution of k_t to be finite is $\rho^2 < 2$ However, the existence of higher moments of k_t requires a tighter restriction on ρ^2. In the empirical analysis of Section 5 we only impose the restriction $\rho^2 < 2$, implying that the first order correlation coefficient $\widetilde{\rho}^2$ is allowed to vary on the whole range of the interval $(0,1)$. Note also that the restriction $\rho^2 < 2$ is sufficient for σ_t^2 (i.e. the inverse of k_t) to have finite moments up to the order $\min(\underline{\alpha}, n)$.

Equation (4) indicates that the conditional expectation of k_t given k_{t-1} is a weighted average of k_{t-1} and the unconditional mean $E(k_t)$, as in a standard AR(1) model. Furthermore, equation (5) indicates that the autocorrelation structures of the ARG and the FTARG are the same.

The expression for $E(k_t|k_{t-1}, \widetilde{T}_t)$ in equation (6) indicates that when $\widetilde{T}_t > E(\widetilde{T}_t)$ ($\widetilde{T}_t < E(\widetilde{T}_t)$) the expected value of $k_t|k_{t-1}$ is above (below) what would be expected in the ARG model, making the tails more flexible. In particular, very small values of $\widetilde{T}_t$ will imply low values for k_t and consequently very large values for the volatility σ_t^2. As we will see in the empirical section, this feature makes the FTARG model specially useful for data with periods of greater instability.

Using the Poisson representation of the non-central chi-squared distribution (Muirhead (1982, p. 23)), the conditional distribution of $k_t|k_{t-1}$ can be written as a Gamma $G(n/2 + h_t, 2\theta^2\widetilde{T}_t)$, where h_t follows a Poisson distribution $P(\lambda_t)$ with $\lambda_t = \rho^2 k_{t-1}/2\theta^2$ and $\widetilde{T}_t$ follows a beta distribution (as described in Section 3.1). Therefore we are generalizing the conditional distribution of $k_t|k_{t-1}$ by using a scale mixture of gammas, in which the mixing distribution is a beta distribution. Similarly, the stationary distribution of k_t is a scale mixture of Gammas, where the mixing distribution is that of $v_{c,t}$. Note that restricting the support of $\widetilde{T}_t$ to $(0,1)$ does not restrict the support of $v_{c,t}$, which is unbounded. This approach to generalize the distribution is somehow analogous to the compound gamma distribution of Dubey (1970), which is also derived as a scale mixture of gammas, but with a gamma as the mixing distribution. Our framework could be further generalized by assuming that $\widetilde{T}_t$ follows a discrete mixture of Beta distributions, as a mixture of beta distributions can accurately approximate any distribution on the $(0,1)$ interval (e.g. Petrone, 1999).

Tables 1 and 2 show how V affects the percentiles of the stationary distribution of k_t while keeping $E(k_t)$, $E(k_t^2)$ and $cov(k_t, k_{t-1})$ constant. Even if the parameter for the degrees of freedom n increases from 1 to 100, by decreasing V and θ in a suitable manner, the moments can be kept constant while the tail of the distribution

6

V	n	θ^2	ρ^2	1%	5%	95%	99%	$var(\widetilde{T}_t)$
∞	1.28	0.078	1.96	0.003	0.031	8.78	14.53	0
15000	1.29	0.078	1.96	0.003	0.032	8.85	14.61	1.7E-05
7500	1.29	0.078	1.96	0.003	0.032	8.76	14.54	3.3E-05
1000	1.34	0.075	1.96	0.003	0.036	8.77	14.50	2.5E-04
500	1.40	0.072	1.96	0.004	0.041	8.65	14.58	5.0E-04
200	1.60	0.062	1.96	0.009	0.064	8.46	14.70	1.2E-03
150	1.76	0.057	1.96	0.012	0.080	8.32	14.92	1.7E-03
100	2.16	0.046	1.96	0.028	0.125	8.02	14.74	2.5E-03
50	6.31	0.016	1.96	0.203	0.396	7.12	13.85	4.9E-03
45	10.84	0.009	1.96	0.295	0.498	6.84	13.48	5.4E-03
42	22.04	0.005	1.96	0.380	0.581	6.71	13.37	5.8E-03
41	35.34	0.003	1.96	0.412	0.613	6.63	13.24	6.0E-03
40	96.20	0.001	1.96	0.451	0.645	6.54	13.24	6.1E-03

Table 1: Percentiles of k_t for different values of V. The value of $E(k_t)$, $E(k_t^2)$ and $cov(k_t, k_{t-1})$ are kept equal in all cases to 2.5, 16 and 0.98, respectively. The percentiles are calculated using 150000 independent draws. The table does not show values of V smaller than 40 because it is not possible to maintain the same values of $(E(k_t), E(k_t^2), cov(k_t, k_{t-1}))$ when $V < 40$.

varies considerably. In particular Table 1 shows that the 1% percentile varies from 0.003 to 0.45 as V varies from ∞ to 40. In Table 2 the 1% percentile varies from 3.5E-12 to 0.2157 as V varies from ∞ to 27.4. Thus, when V is large and n is small, the tail of k_t towards 0 is fatter, whereas decreasing the value of V allows n to be larger and in this way reduce the probability of values near 0. This implies that the right tail of the volatility σ_t^2 is fatter when V is large and n is small. To see the impact on the distribution of the volatility σ_t^2, Figures 1 - 3 plot one random realization of $(\sigma_1^2, ..., \sigma_{1000}^2)$ for 3 of the processes in Table 2 (those corresponding to $V = 33$, $V = 29$ and $V = 27.5$). Even though the 3 processes imply the same values for $E(k_t)$, $E(k_t^2)$ and $cov(k_t, k_{t-1})$, we can see that σ_t^2 takes occasionally very large values (larger than 800 in Figure 1) when $V = 33$, but when $V = 27.5$ the values for σ_t^2 in Figure 3 are all below 11.

For simplicity, instead of assuming that k_1 is drawn from the stationary distribution, it will be assumed that k_1 is drawn from a distribution which has the same mean as 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 (ARG)

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 ARG 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 ARG 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

V	n	θ^2	ρ^2	1%	5%	95%	99%	$var(\widetilde{T}_t)$
∞	0.327	0.183	1.96	4.0E-12	6.2E-08	8.12	18.39	0
15000	0.328	0.183	1.96	3.4E-12	6.2E-08	8.11	17.91	1.67E-05
40	1.09	0.055	1.96	0.0003	0.006	5.97	14.19	6.10E-03
35	1.59	0.038	1.96	0.0031	0.024	5.47	13.13	6.94E-03
33	2.04	0.029	1.96	0.0092	0.045	5.30	12.47	7.35E-03
30	4.00	0.015	1.96	0.0501	0.125	4.76	11.12	8.06E-03
29	6.28	0.010	1.96	0.0884	0.179	4.62	10.54	8.33E-03
28	16.0	0.004	1.96	0.1574	0.254	4.41	10.11	8.62E-03
27.6	45.0	0.001	1.96	0.1944	0.293	4.27	10.00	8.74E-03
27.5	83.8	0.001	1.96	0.2020	0.304	4.24	9.63	8.77E-03
27.4	631.2	9.5E-05	1.96	0.2157	0.314	4.19	9.53	8.80E-03

Table 2: Percentiles of k_t for different values of V. The value of $E(k_t)$, $E(k_t^2)$ and $cov(k_t, k_{t-1})$ are kept equal in all cases to 1.5, 16 and 0.98, respectively. The percentiles are calculated using 150000 independent draws. The table does not show values of V smaller than 27.4 because it is not possible to maintain the same values of $(E(k_t), E(k_t^2), cov(k_t, k_{t-1}))$ when $V < 27.4$.
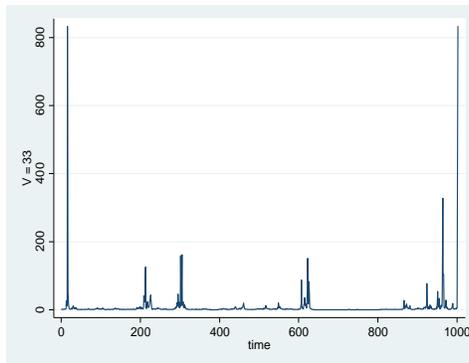

Figure 1: One random draw of $(\sigma_1^2, ..., \sigma_{1000}^2)$ with $V = 33$, $n = 2.04$, $\theta^2 = 0.029$ and $\rho^2 = 1.96$. These values imply that $E(k_t) = 1.5$, $E(k_t^2) = 16$, and $cov(k_t, k_{t-1}) = 0.98$.

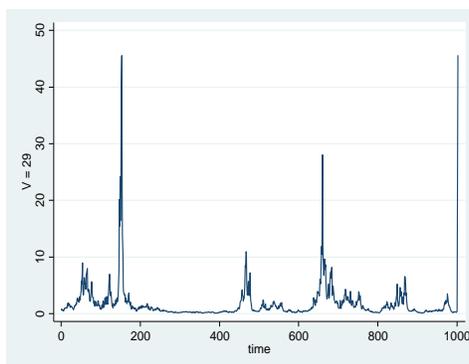Figure 2: One random draw of $(\sigma_1^2, ..., \sigma_{1000}^2)$ with $V = 29$, $n = 6.28$, $\theta^2 = 0.0095$ and $\rho^2 = 1.96$. The values for $E(k_t)$, $E(k_t^2)$ and $cov(k_t, k_{t-1})$ are the same as in Figure 1.
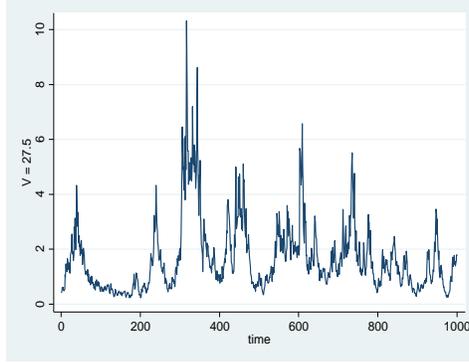
Figure 3: One random draw of $(\sigma_1^2, ..., \sigma_{1000}^2)$ with $V = 27.5$, $n = 83.77$, $\theta^2 = 0.00072$ and $\rho^2 = 1.96$. The values for $E(k_t)$, $E(k_t^2)$ and $cov(k_t, k_{t-1})$ are the same as in Figure 1.

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:

$$
\begin{aligned}
y_t &= x_t\beta + \sqrt{\frac{1}{k_t}}e_t & (10) \\
e_t &\sim N(0,1) \\
k_t|k_{1:(t-1)}, h_{1:t}, \Theta, \beta &\sim G(n/2 + h_t, 2\widetilde{\theta}_t^2) \\
h_t|k_{1:(t-1)}, h_{1:(t-1)}, \Theta, \beta &\sim P(\lambda_t) \quad \text{with } \lambda_t = \frac{\widetilde{\rho}_t^2 k_{t-1}}{2\widetilde{\theta}_t^2}
\end{aligned}
$$

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 $\Theta = (n, \theta^2, \rho^2)$, $k = (k_1, ..., k_T)$ and $h = (h_2, ..., h_T)$. The representation (10) 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

9

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 2 *Consider the model defined by (10), and define:*

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

The conditional posteriors are as follows:

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

and

$$
\begin{aligned}
p(\Theta|Y, h, \beta) &\propto \int p(\Theta)p(k, h|\Theta, \beta)L(Y|k, \beta)dk = \qquad (11) \\
&(2\pi)^{-T/2} \prod_{t=1}^{T} \left[(2\widetilde{r}_t^2)^{\frac{n+1}{2} + h_{t+1} + h_t} \Gamma\left(\frac{n+1}{2} + h_{t+1} + h_t\right) \right] \\
&\left[\prod_{t=2}^{T} \frac{1}{\left(2\widetilde{\theta}_t^2\right)^{n/2}} \frac{\left(\frac{\widetilde{\rho}_t}{2\widetilde{\theta}_t^2}\right)^{2h_t}}{h_t!} \frac{1}{\Gamma(n/2 + h_t)} \right] (1 - \widetilde{\rho}^2)^{n/2} \left(2\widetilde{\theta}^2\right)^{-\frac{n}{2}} \left(\Gamma\left(\frac{n}{2}\right)\right)^{-1} p(\Theta)
\end{aligned}
$$

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. Let us introduce m_t by assuming that m_t conditional on h_t has a beta distribution:

$$m_t|h_t \sim B(\alpha_m + h_t, \beta_m), \qquad \alpha_m = (n-1)/2, \qquad \beta_m = 1/2 \tag{12}$$

Note that this requires $n > 1$. This is a restriction that is also needed to ensure that $E(\sigma_t^2)$ is finite. 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 it seems preferable to condition on a mixture indicator d_t, such that the conditional posterior of h_t becomes simply 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 shifted Poisson distributions. In turn, $(m, d)|(k, h)$ can be drawn using independent beta distributions (for m) and the hypergeometric 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:

$$f_{SP}(h|\lambda, d) = \lambda^{h-d} \frac{1}{(h-d)!} \frac{1}{\exp(\lambda)} \qquad h = d, (d+1), ... \tag{13}$$

Note that a draw from a shifted Poisson $h_t \sim SP(\lambda_t, d_t)$ can be obtained by first obtaining a draw x from the Poisson distribution $P(\lambda_t)$ and then calculating $h_t = x + d_t$. The vector d is formally introduced in the model by using a hypergeometric distribution (e.g. Monahan (2001, p. 305)) as a prior for each of the components of d given h:

$$\Pr(d_t = s|h_t, d_{t+1}) = \frac{\binom{M_{d_t}}{s}\binom{N_{d_t} - M_{d_t}}{n_{d_t} - s}}{\binom{N_{d_t}}{n_{d_t}}} \qquad \begin{array}{c} t = 2, ..., T \\ d_{T+1} = 0 \\ 0 \leq s \leq \min((1 + d_{t+1}), h_t) \end{array} \tag{14}$$

$$M_{d_t} = h_t, \quad n_{d_t} = 1 + d_{t+1}, \quad N_{d_t} = (n-1)/2 + h_t + d_{t+1}$$

Because in our case N_{d_t} is not an integer, the corresponding binomial coefficient should be written using the gamma function instead of the factorial, based on the relationship $\Gamma(x+1) = x!$ (see proof of Proposition 3 in the appendix for more details). There are several algorithms that efficiently draw from the hypergeometric distribution, are available in some standard statistical packages and are applicable in the case that N_{d_t} is not an integer (e.g. Stadlober, (1989), Kachitvichyanukul and Schmeiser (1988) or see Monahan (2001, p. 306) for a review). 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 $d_t \leq h_t$, so if $h_t = 0$ then d_t should also be fixed to be 0.

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

The m-Gibbs for the ARG model.

- Generate $\Theta|(m,d),\beta$ using a Metropolis step.

- Generate $(k,h)|(m,d),\Theta,\beta$ using gammas and poisson.

- Generate $(m,d)|(k,h),\Theta,\beta$ using beta and the hypergeometric distribution in (14).

- 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 3 *Given the model described in equations (10), (12), (14), and the following definitions:*

$$
\begin{aligned}
\widehat{r}_T^2 &= \widetilde{r}_T^2 \\
\widehat{r}_t^2 &= \left(\frac{1}{\widetilde{r}_t^2} - m_{t+1} \left(\frac{\widetilde{\rho}_{t+1}}{\widetilde{\theta}_{t+1}^2} \right)^2 \widehat{r}_{t+1}^2 \right)^{-1} \quad for \ t = 1, ..., T-1 \\
m_1 &= 1, \ d_1 = d_{T+1} = h_1 = 0, \ \lambda_t = \frac{\widetilde{\rho}_t^2 k_{t-1}}{2\widetilde{\theta}_t^2}, \ \widehat{\lambda}_t = \lambda_t \frac{m_t \widehat{r}_t^2}{\widetilde{\theta}_t^2},
\end{aligned}
$$

with $\widetilde{r}_t^2$ defined in Proposition 2, the conditional posteriors are as follows:

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

The conditional posterior $d|k,h,m$ is the same as the conditional prior in (14). In addition:

$$
\begin{aligned}
p(\Theta|Y,m,d,\beta) \ \propto \ & \int p(\Theta) p(k,h,m,d|\Theta) L(Y|k,\beta) dk dh = \\
& \left[\prod_{t=1}^{T} \left(2\widehat{r}_t^2 \right)^{\frac{n+1}{2} + d_{t+1} + d_t} \right] \left[\prod_{t=2}^{T} \left(m_t \left(\frac{\widetilde{\rho}_t}{2\widetilde{\theta}_t^2} \right)^2 \right)^{d_t} \right] \times \\
& \left[\prod_{t=2}^{T} \frac{1}{d_t!} \frac{\Gamma((n+1)/2 + d_{t+1})}{\Gamma((n-1)/2 + d_t)} \frac{\Gamma(2 + d_{t+1})}{\Gamma(2 + d_{t+1} - d_t)} \right] \times \\
& \left[\prod_{t=2}^{T} m_t^{\alpha_m - 1} (1 - m_t)^{\beta_m - 1} \right] \frac{\Gamma((n+1)/2 + d_2)}{\Gamma(n/2)} C_p C_L C_B p(\Theta)
\end{aligned}
$$

(15)

where

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

and $p(\Theta)$ is the prior of Θ.

Using Proposition 3, a draw of $(k,h)|(m,d)$ can be obtained by first drawing k_1 from a Gamma (recall that $h_1 = 0$), then $h_2|k_1$ from a shifted Poisson, then $k_2|h_2$ again from a Gamma and so on until we finally draw $h_T|k_{T-1}$ and $k_T|h_T$. Conversely, a draw from the conditional posterior of (m,d) is obtained by using the prior distributions (12) and (14). Thus, m_t is drawn using independent beta distributions, and d_t is drawn recursively using the hypergeometric distribution, starting with d_T, and then $d_{T-1}|d_T$ and so on until we finally draw $d_2|d_3$. The vector of unknown parameters Θ is generated by targeting the kernel in (15) using a Metropolis step. It seems recommendable to repeat the Metropolis step several times (between 5 and 15) since this could reduce the autocorrelations while not having much impact on computation time.

3.2 Flexible Tail Autoregressive Gamma Process (FTARG)

As shown in the proof of Proposition 4 in the appendix, the conditional posterior density of $\widetilde{T}_t|V, h, \Theta$ is proportional to:

$$
\left(\widetilde{T}_t\right)^{\overline{\alpha}_t - 1} \left(1 - \widetilde{T}_t\right)^{V/2 - 1} \left(\frac{1}{1 + \widetilde{T}_t S_t}\right)^{v_t} \qquad t = 2, ..., T \tag{16}
$$

with:

$$
\begin{aligned}
\overline{\alpha}_t &= \frac{V}{2} + h_{t+1} + \frac{1}{2} & v_t &= \frac{n+1}{2} + h_t + h_{t+1} \\
S_t &= \theta^2(r_t^2 + \rho^2/\theta^2) \text{ for } t = 2, ..., T-1 & S_T &= \theta^2 r_T^2
\end{aligned}
$$

This kernel can be written as that of an infinite mixture of beta distributions if we write the last term of this density 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. An easier method is to draw from (16) using a Metropolis-step with a random walk proposal density. A third possibility is to introduce 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. This variable J_t can be introduced as a negative binomial (e.g. Johnson et al. (2005, p. 208)) discrete random variable with probability of success p_t and number of failures v_t (denoted as $J_t \sim NB(v_t, p_t)$):

13

$$\Pr\left(J_t = s | \widetilde{T}_t, S_t\right) \quad = \quad (1 - p_t)^{v_t} (p_t)^s \binom{v_t + s - 1}{v_t - 1} \tag{17}$$

$$p_t \quad = \quad \frac{S_t}{1 + S_t}(1 - \widetilde{T}_t) \qquad t = 2, ..., T$$

Draws from the negative binomial distribution can be obtained using efficient algorithms which are implemented in a wide range of statistical software. Alternatively, J_t can be drawn from a Poisson $P(c_t)$ where c_t is a draw from a Gamma $G(v_t, p_t/(1 - p_t))$ (e.g. Johnson et al. (2005, p.p. 212-213)). 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

- $J|(k, h), \Theta, \widetilde{T}, V, \beta$ using the negative binomial distribution in (17).

- $\widetilde{T}|(k, h), \Theta, J, V, \beta$ using beta distributions.

- $V|(k, h), \Theta, \widetilde{T}, \beta$ using a Metropolis step.

Proposition 3 in the previous section and the following proposition describe the distributions that are necessary in this algorithm.

Proposition 4 *The conditional posterior densities for $\widetilde{T}$, and V in the FTARG model are as follows:*

$$\widetilde{T}_t | J_t \quad \sim \quad B(\overline{\alpha}_t, V/2 + J_t)$$

$$p(V|Y, \widetilde{T}) \quad \propto \quad p(V) \left(\frac{\Gamma(V)}{\Gamma(V/2)\Gamma(V/2)}\right)^{T-1} \prod_{t=2}^{T} \left(\widetilde{T}_t\right)^{V/2-1} \left(1 - \widetilde{T}_t\right)^{V/2-1}$$

where $p(V)$ is the prior for V. The conditional posterior density for J_t is the same as the conditional prior given in (17).

4 Evidence on the Efficiency of the Algorithms

We use real and simulated data to compare the computational efficiency of the two algorithms developed in this paper (the h-Gibbs and the m-Gibbs) with the recently developed Particle marginal Metropolis - Hastings sampler (PMMH, Andrieu et al. 2010) that updates jointly the unknown parameters Θ and the volatilities k. The PMMH is a general purpose algorithm and it uses a particle filter to evaluate the conditional posterior of Θ marginally on the volatilities. The efficiency of the algorithm depends on the number of particles used, and as the number of particles increases, the performance of the PMMH (in terms of autocorrelations) approaches that of an ideal algorithm that generates Θ marginally on the volatilities. To be able to set optimally the proposal

density for Θ in the PMMH algorithm, we simplify the estimation by keeping β equal to the OLS estimate, so that (n, θ^2, ρ^2) remain as the only parameters to be estimated. In all algorithms we use a random walk proposal density for Θ and for optimality we fix the variance-covariance matrix of the proposal density proportional to the posterior variance-covariance matrix of Θ (Gelman et al. 1996), which is obtained in a previous estimation. For simplicity in the PMMH algorithm we use the bootstrap filter (Gordon et al. 1993). In the h-Gibbs and m-Gibbs algorithms, we repeat the Metropolis step 10 times to obtain a single value for Θ. This reduces significantly the autocorrelation for the parameter n (not much for θ^2 and ρ^2) while increasing computation time by 60% or 70% (when T=100) and 70% or 100% (when T=2000), respectively. In terms of comparing the efficiency among the algorithms, results would be very similar if we did not repeat the Metropolis-step.

We use the prior described in the appendix and in the Metropolis step we use a transformation of the parameters that maps them into an unbounded space. In particular, we target the conditional posterior of $\delta = (\delta_1, \delta_2, \delta_3)$ defined as: $\delta_1 = \ln(n) + \ln(\theta^2) - \ln(1 - \rho^2)$, $\delta_2 = \ln(\theta^2)$ and $\delta_3 = -\ln(1 - \rho^2)$. By this transformation the only restriction on δ is $\delta_3 > 0$, which is likely to be satisfied provided that ρ^2 is not close to 0. To be more precise, we are not using a proposal density for Θ but for δ, calibrated using the posterior var-cov of δ.

First we simulate a short time series of $T = 100$ using parameter values $n = 2$, $\theta^2 = 0.15$, $\rho = 0.95$ with $y_t = 2 + \sigma_t e_t$, and $x_t = (1, y_{t-1})$, so that the true value of β is $\beta = (2, 0)$. We compare the efficiency of the 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 computation time per iteration differs for different algorithms, we present also the ESS adjusted for computation time (ESS/TIME), which is the number of independent draws from the posterior obtained in one minute (using GAUSS software and Intel Xeon CPU with 2.9 GHz).

The ESS of the PMMH depends on the number of particles used in the bootstrap filter. Table 3 shows that when considering computation time choosing 50 particles gives better results (although for n choosing 25 particles gives slightly better results). However, the m-Gibbs sampler is 4.7 times better than the best PMMH in terms of ESS/TIME to sample n, whereas the improvements for ρ and θ^2 are 62% and 9.4%, respectively. When we compare the m-Gibbs with the h-Gibbs, we can see that the m-Gibbs is between 4.9 and 6.2 times more efficient.

In Table 4 we can see that choosing 500 or 1000 particles gives roughly the same ESS for the PMMH, indicating that there is not much further gain in increasing the number of particles. Thus we can expect that the PMMH algorithm with 1000 particles has practically the same ESS as the ideal algorithm that samples Θ marginally on the volatilities (Andrieu et al. (2010)). Thus it is interesting to compare the ESS sample size of the m-Gibbs and the h-Gibbs with the ESS of such ideal algorithm. In Table 4 we can see that the m-Gibbs has roughly the same ESS for n as the ideal algorithm, but the ESS for θ^2 and ρ is 15% and 22.6%, respectively, of the ideal algorithm. Because the number of observations is relatively small and the prior for ρ is quite spread, the 95% posterior credible interval for ρ is wide and equal to $(0.76, 0.98)$. Although not shown in the tables, all algorithms produced the same summary of the posterior distribution, indicating the absence of programming errors. Overall Tables 3 and 4 show that the m-Gibbs algorithm is much more efficient than the best PMMH even when T is as small as 100 and much more efficient also than the h-Gibbs.

	h-Gibbs	m-Gibbs	P_25	P_50	P_100	P_500	P_1000
n	0.0037	0.0302	0.0113	0.015	0.021	0.042	0.028
	[150.3]	[730.6]	[155.9]	[122.9]	[83.5]	[14.9]	[2.8]
θ^2	0.0011	0.0105	0.015	0.029	0.034	0.063	0.069
	[46.1]	[253.2]	[201.9]	[231.3]	[134.5]	[22.4]	[6.8]
ρ	0.0015	0.015	0.016	0.029	0.034	0.06	0.07
	[60.1]	[372.8]	[219.3]	[230.1]	[137.5]	[21.9]	[6.8]
Accept R.	93%	93%	21%	34%	45%	50%	52%

Table 3: Effective Sample Size (ESS) and ESS over time (ESS/TIME) for the h-Gibbs, the m-Gibbs and PMMH algorithms using 100 artificial observations. ESS/TIME is in squared brackets and represents the number of independent samples per minute. The column P_25 refers to the PMMH algorithm that uses 25 particles. The row Accept R. gives the acceptance rate in the Metropolis step. Note that in the h-Gibbs and m-Gibbs the Metropolis step is repeated 10 times, and Accept R. is the probability of accepting a new value in the sequence of 10 draws.

	h-Gibbs	m-Gibbs	P_25	P_50	P_100	P_500	P_1000
n	13.1	107.7	40.3	54.9	74.2	149.2	100
θ^2	1.6	15.3	21.4	42.3	49.0	91.9	100
ρ	2.2	22.6	23.3	42.3	50.3	90.2	100

Table 4: Effective Sample Size (ESS) as a proportion of the ESS of the PMMH with 1000 particles.

Let us now compare the efficiency of the algorithms using 2000 daily observations of the exchange rate Yen - US dollar (6th Aug 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 $\beta = (\beta_0, \beta_1)$. In Table 5 we can see that it is best to choose 500 particles for the PMMH and that the m-Gibbs is 710 times more efficient than the best PMMH to sample n, 15 times more efficient to sample ρ and 12 times more efficient to sample θ^2. With respect to the h-Gibbs algorithm, the m-Gibbs is about 36 times more efficient to sample θ^2 or ρ and 4.6 times more efficient to sample n. The posterior 95% credible interval for ρ is $(0.956, 0.99)$, which is quite close to 1. That is one reason why the relative performance of the h-Gibbs is particularly bad in this case.

	h-Gibbs	m-Gibbs	P_300	P_500	P_750	P_1000
n	0.004	0.041	0.002	0.004	0.007	0.007
	[9.68]	[44.26]	[0.061]	[0.062]	[0.060]	[0.032]
θ^2	0.00001	0.0012	0.0016	0.0065	0.0079	0.0078
	[0.033]	[1.23]	[0.058]	[0.101]	[0.064]	[0.037]
ρ	0.00002	0.0014	0.0019	0.0062	0.0082	0.0091
	[0.040]	[1.450]	[0.071]	[0.096]	[0.066]	[0.044]
Accept R.	93%	96%	9%	20%	28%	34%

Table 5: Effective Sample Size (ESS) and ESS over time (ESS/TIME) for the h-Gibbs, the m-Gibbs and PMMH algorithms using 2000 observations of the US-Japan exchange rate. ESS/TIME is in squared brackets and represents the number of independent samples per minute. See explanation in Table 3 for other definitions.

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 ARG and FTARG described in Section 2, we consider the model where σ_t^2 follows a log-normal distribution (LNORM) (using the SvPack in Ox provided by Kim et al (1998)). In addition, we consider 3 models where e_t follows a student-t distribution: ARG-T, FTARG-T and LNORM-T. These 3 models are the same as the ARG, FTARG and LNORM models, respectively, but assume a student-t distribution for e_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 6 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 return of 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 7 shows the value of the log-likelihood at the posterior median of parameters, calculated using the bootstrap particle filter (e.g. Gordon et al. (1993)), and using the prior specification shown in the appendix. Marginal likelihood values (calculated with the method of Chib and Jeliazkov, (2001)), show a similar patter and are given in Table 8. We can see that the ARG 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. The improvement in the log-likelihood is as much as 30 (India-US) or 40 (Brazil-US) points over the LNORM-T. 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 ARG). The extension to student-t errors does not bring any noticeable improvement in the value of the log-likelihood of the ARG 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 ARG and ARG-T, the FTARG-T model seems to be the best as it gains more than 30 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 ARG-T are substantially better than the LNORM and ARG, again indicating that it is important to allow for student-t errors. Both the LNORM-T and the ARG-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 ARG-T could be said to be joint winners for the EU-US exchange rate, as confirmed by the marginal likelihood values in Table 8.

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 ARG) might be more adequate in the estimation of the Phillips curve with UK data.

Figure 4 shows the OLS residuals for each of the 5 datasets. We can observe larger jumps in volatility in the exchange rates of India and Brazil, which might be one of the reasons why the inverse gamma models perform

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				
EU-US	Exchange rate Euro - US dollar, daily: 6 Jan 1999 - 17 May 2013, 3615 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 6: Description of variables used in empirical analysis

	IND-US	BRA-US	JP-US	EU-US	UK-INFL
LNORM	1258.5	413.6	n.a.	13275.1	-198.7
	(0.09)	(0.21)		(0.16)	(0.32)
LNORM-T	1398.4	448.0	-7568.9	13285.8	-197.8
	(0.23)	(0.26)	(0.64)	(0.28)	(0.16)
ARG	1427.2	490.5	n.a.	13276.5	-195.8
	(0.08)	(0.67)		(0.05)	(0.09)
ARG-T	1427.6	490.4	-7913.8	13287.0	-195.9
	(0.09)	(0.19)	(0.71)	(0.25)	(0.08)
FTARG	1447.8	526.0	-7634.4	13276.2	-196.5
	(0.09)	(0.15)	(0.82)	(0.05)	(0.07)
FTARG-T	1446.4	526.1	-7531.2	13286.1	-195.5
	(0.09)	(0.30)	(0.62)	(0.05)	(0.12)

Table 7: Value of Log-Likelihood at the posterior median, calculated with a particle filter for different models and datasets. Numerical standard error in brackets (obtained using independent estimates of the likelihood). The particle filter failed to give estimates of the likelihood in the case of normal errors with the Yen-Dollar dataset due to the presence of data points which were too far on the tails of the distribution, causing underflow.

much better than the log-normal models in these datasets. Another reason might be that inverse gamma models allow for greater correlation in the volatility outliers than the LNORM-T model. To see this recall that the LNORM-T model can be written as a mixture of normals: $y_t = x_t \beta + e_t$, where $e_t \sim N(0, \chi_t^{-1} \sigma_t^2)$ and χ_t are i.i.d. draws from a gamma distribution. Therefore the volatility of e_t has two components, one determined by σ_t^2 and another by χ_t. Because χ_t has no serial correlation, the LNORM-T model does not allow for persistence in the volatility outliers. This is not so in the inverse gamma and generalized inverse gamma models, where the volatility of e_t has only one component σ_t^2, which has positive correlation with σ_{t-1}^2 regardless of whether σ_{t-1}^2 was on the tail of the distribution or not.

	IND-US	BRA-US	JP-US	EU-US	UK-INFL
LNORM	1259.1	390.7	n.a.	13236.1	-236.7
	(0.45)	(0.21)		(0.16)	(0.32)
LNORM-T	1365.2	421.5	-7607.2	13247.2	-231.9
	(0.23)	(0.26)	(0.64)	(0.28)	(0.16)
ARG	1400.9	467.2	n.a.	13241.0	-259.0
	(0.08)	(0.67)		(0.05)	(0.09)
ARG-T	1400.9	466.7	-7976.0	13249.8	-258.4
	(0.09)	(0.19)	(0.71)	(0.25)	(0.07)
FTARG	1425.9	496.9	-7688.7	13234.5	-263.1
	(0.12)	(0.19)	(0.82)	(0.05)	(0.09)
FTARG-T	1419.4	493.9	-7592.0	13242.2	-266.5
	(0.09)	(0.33)	(0.70)	(0.05)	(0.12)

Table 8: Value of Marginal Likelihood calculated using the method of Chib and Jeliazkov (2001), but the posterior ordinate for $(n, \theta^2, \rho, V, \varpi)$ was calculated using a normal approximation. Numerical standard error in brackets.
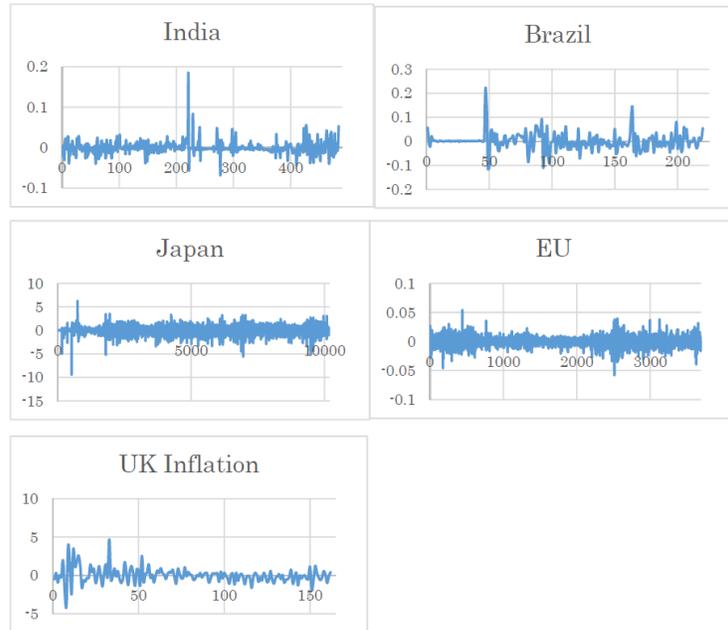

Figure 4: OLS residuals for 5 different datasets: 4 exchange rates versus the US dollar and a Phillips Curve for UK inflation.

6 Conclusions

This paper has developed efficient posterior simulators for inverse gamma and generalized inverse gamma processes 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. Estimations with real and simulated data show that the new algorithm is much more efficient than the recently developed Particle MCMC algorithms that generate the volatilities and unknown parameters in a joint move.

We also developed a new type of generalized inverse gamma time-series model and analytically derived its properties. Using simulation we calculated the percentiles of the distribution and illustrated that the generalized inverse gamma process has much greater flexibility in the right tail. In this way we provide a new class of flexible stochastic volatility models that can be estimated with simple and efficient MCMC algorithms. Furthermore, the FTARG process can be further generalized by specifying $\widetilde{T}_t$ to be a mixture of beta distributions, since such a mixture can approximate any distribution in the interval $(0,1)$. Finally, the empirical exercise shows that inverse gamma and generalized inverse gamma models outperform the lognormal volatility model with student-t errors specially in the datasets that exhibit greater jumps and correlation of volatility outliers, such as the exchange rates of Brazil-US or India-US.

References:

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

Andrieu, C., A. Doucet, and R. Holenstein, (2010), "Particle Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, 269–342.

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.

Bauwens, L., M. Lubrano, and J.F. Richard (1999) *Bayesian Inference in Dynamic Econometric Models*. Oxford: Oxford University Press.

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., I. Jeliazkov (2001) "Marginal Likelihood From the Metropolis–Hastings Output," *Journal of the American Statistical Association,* 96, 270-281.

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

Delatola, E.I. and J.E. Griffin (2011) "Bayesian nonparametric modelling of the return distribution with stochastic volatilty" *Bayesian Analysis* 6, 1–26.

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

Dubey, S. D. (1970) "Compound Gamma, beta and F distributions," *Metrika*, 16, 27-31.

Durham, G.B. (2007) "SV mixture models with application to S&P 500 index return," *Journal of Financial Economics,* 85, 822–856.

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.

Gelman, A., G.O. Roberts and W.R. Gilks (1996) "Efficient Metropolis Jumping Rules," in *Bayesian Statistics* 5, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp. 599–608.

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, (2006a) "Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility," *Journal of Econometrics*, 134, 605–644.

Griffin, J.E. and M.F.J. Steel (2006b) "Ordered-based dependent Dirichlet processes," *Journal of the American Statistical Association,* 101, 179–194.

Griffin, J.E. and M.F.J. Steel (2011) "Stick-breaking autoregressive processes," *Journal of Econometrics,* 162, 383–396.

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.

Janssen, A and H. Drees (2013) "A stochastic volatility model with flexible extremal dependence structure," *arXiv.org*, arXiv:1310.4621v1.

Jensen, M.J. and J.M. Maheu (2010) "Bayesian semiparametric stochastic volatility modeling", *Journal of Econometrics,* 157, 306–316.

Jensen, M.J. and J.M. Maheu (2014)."Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture," *Journal of Econometrics*, 178, 523-538.

Johnson, N.L., S. Kotz and N. Balakrishnan (1994), *Continuous Univariate Distributions, Volume 1*, 2nd Edition, John Wiley & Sons, Inc.

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

Johnson, N.L., A.W. Kemp and S. Kotz (2005), *Univariate Discrete Distributions*, 3rd Edition, John Wiley & Sons, Inc.

Kachitvichyanukul, V. and B.M. Schmeiser (1988) "Algorithm 668: H2PEC: sampling from the hypergeometric distribution," *ACM Transactions on.Mathematical Software*, 14, 397-398.

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.

Koop, G. (2003) *Bayesian Econometrics*, Wiley.

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

Monahan, J.F. (2001), *Numerical Methods of Statistics*, Volume 1, Cambridge University Press.

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

Petrone, S. (1999) "Random Bernstein polynomials," *Scandinavian Journal of Statistics*, 26, 373-393.

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

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.

Stacy, E. W. (1962) "A Generalization of the Gamma Distribution," *Annals of Mathematical Statistics*, 33, 1187-1192.

Stadlober, E (1989) "The Ratio of Uniforms Approach for Generating Discrete Random Variables," *Journal of Computational and Applied Mathematics*, 31, 181-189.

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.

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

Yuan, L. and J.D. Kalbfleisch (2000) "On the Bessel distribution and related problems" *Ann. Inst. Statist. Math.*, 52, 438–447.

Appendix

Prior Specification in the Empirical Application

For the Gamma type models we specified the prior as: $ln(n) \sim N(\ln(40), 1.5)$, $\rho^2 \sim B(8, 1)$, $\theta^2 \sim G(1, 200)$, $\ln(V) \sim N(\ln(20), 1)$.

For the log-normal volatility model we use the same prior specification and the same notation as in Kim, Shephard and Chib (1998): $\mu \sim N(0, 10)$, $(\sigma_\eta^2)^{-1} \sim G(2.5, 40)$, $(\phi + 1)/2 \sim B(20, 1.5)$.

In all models the prior for β is $N(0, TI)$, where I is the identity matrix and T is the sample size.

For the models with student-t errors, we specify $ln(\varpi) \sim N(\ln(40), 1.5)$, where ϖ is the parameter for the degrees of freedom of the student-t.

For some datasets the log-normal volatility model did not converge with the baseline prior, and in those cases we used a tighter prior for σ_η^2 to ensure convergence: $(\sigma_\eta^2)^{-1} \sim G(17.5, 57.14)$ (Brazil), $(\sigma_\eta^2)^{-1} \sim G(22.5, 444.4)$ (India, normal errors), $(\sigma_\eta^2)^{-1} \sim G(17.5, 57.14)$ (India, student-t errors), $(\sigma_\eta^2)^{-1} \sim G(3.5, 28.57)$ (UK inflation).

As mentioned above, in the Metropolis step we target the conditional posterior of $\delta = (\delta_1, \delta_2, \delta_3)$, defined as: $\delta_1 = \ln(n) + \ln(\theta^2) - \ln(1 - \rho^2)$, $\delta_2 = \ln(\theta^2)$ and $\delta_3 = -\ln(1 - \rho^2)$. The inverse transformation is $\Theta(\delta) = (n(\delta), \theta^2(\delta), \rho^2(\delta)) = (\exp(\delta_1 - \delta_2 - \delta_3), \exp(\delta_2), 1 - \exp(-\delta_3))$. Since our prior is defined on $\Theta^* = (\ln(n), \theta^2, \rho^2)$, the prior of δ can be written using the Jacobian as: $p(\Theta^*)\theta^2(1 - \rho^2)$, where $p(\Theta^*)$ is the prior of Θ^* and $[\theta^2(1 - \rho^2)]$ is the Jacobian of the transformation.

In the FTARG model instead of specifying the prior on (ρ^2, θ^2) we specify it on $(\widetilde{\rho}^2, \widetilde{\theta}^2)$, and the Metropolis step targets the conditional posterior of $\delta_1 = \ln(n) + \ln(\widetilde{\theta}^2) - \ln(1 - \widetilde{\rho}^2)$, $\delta_2 = \ln(\widetilde{\theta}^2)$ and $\delta_3 = -\ln(1 - \widetilde{\rho}^2)$.

Proof of Proposition 1

From equation (2) we can write the process for z_t as:

$$z_t = \sqrt{\widetilde{T}_t}\varepsilon_t + \rho\sqrt{\widetilde{T}_t}\sqrt{\widetilde{T}_{t-1}}\varepsilon_{t-1} + \rho^2\sqrt{\widetilde{T}_t}\sqrt{\widetilde{T}_{t-1}}\sqrt{\widetilde{T}_{t-2}}\varepsilon_{t-2} + \rho^3\sqrt{\widetilde{T}_t}\sqrt{\widetilde{T}_{t-1}}\sqrt{\widetilde{T}_{t-2}}\sqrt{\widetilde{T}_{t-3}}\varepsilon_{t-3} + ...$$

which implies that the stationary distribution of $z_t|\widetilde{T}$ is $N(0, \theta^2 v_{c,t})$ and so the stationary distribution of k_t is that of the product of $v_{c,t}$ and $\varepsilon_t' \varepsilon_t$. Hence the moments of k_t can be calculated as $E(k_t^s) = E(v_{c,t}^s) E((\varepsilon_t' \varepsilon_t)^s)$. Because $(\varepsilon_t' \varepsilon_t)$ is distributed as a $G(n/2, 2\theta^2)$, its moments are given by (e.g. Johnson et al. (1994 p. 339)):

$$E((\varepsilon_t' \varepsilon_t)^s) = (\theta^2)^s \prod_{i=0}^{s-1} (n + 2i)$$

To calculate $E(v_{c,t}^s)$ note that we can write $v_{c,t}$ as $v_{c,t} = \widetilde{T}_t + \rho^2 \widetilde{T}_t v_{c,(t-1)}$. so that $E(v_{c,t}^s) = E((\widetilde{T}_t + \rho^2 \widetilde{T}_t v_{c,(t-1)})^s)$. Using the binomial theorem we can write:

$$E((\widetilde{T}_t + \rho^2 \widetilde{T}_t v_{c,(t-1)})^s) = E(\widetilde{T}_t^s) \sum_{i=0}^{s} \binom{s}{i} \rho^{2i} E(v_{c,(t-1)}^i) \tag{18}$$

Because $E(v_{c,t}^s) = E(v_{c,(t-1)}^s)$, (18) implies property (9) and the other unconditional moments stated in Proposition 1. To obtain the conditional moments, note that equation (3) can be written as:

$$k_t = \frac{\widetilde{T}_t}{E(\widetilde{T}_t)} (\widetilde{\rho}^2 k_{t-1} + \widetilde{\varepsilon}_t' \widetilde{\varepsilon}_t + 2\widetilde{\rho} \widetilde{\varepsilon}_t' z_{t-1}) \tag{19}$$

Because $\widetilde{\varepsilon}_t$ is independent of z_{t-1} and $E(\widetilde{\varepsilon}_t) = 0$ we obtain that $E(\widetilde{\varepsilon}_t' z_{t-1}) = 0$. Taking into account that $E(\widetilde{\varepsilon}_t' \widetilde{\varepsilon}_t) = n\widetilde{\theta}^2$ we can take conditional expectations on both sides of (19) to get equations (4) and (6).

Let us calculate $cov(k_t, k_{t-h})$ as $cov(k_t, k_{t-h}) = E(k_t k_{t-h}) - [E(k_t)]^2$. To derive $E(k_t k_{t-h})$ let us use iterative expectations to rewrite equation (4) as:

$$E(k_t | k_{t-h}) = \widetilde{\rho}^{2h} k_{t-h} + \sum_{i=0}^{h-1} \widetilde{\rho}^{2i} (1 - \widetilde{\rho}^2) E(k_t) \tag{20}$$

Multiplying both sides of (20) by k_{t-h} and then taking expectations with respect to k_{t-h} we obtain:

$$E(k_t k_{t-h}) = \widetilde{\rho}^{2h} E(k_{t-h}^2) + \sum_{i=0}^{h-1} \widetilde{\rho}^{2i} (1 - \widetilde{\rho}^2) [E(k_t)]^2 = \widetilde{\rho}^{2h} E(k_{t-h}^2) + (1 - \widetilde{\rho}^{2h}) [E(k_t)]^2$$

where we have used the formula for the sum of a geometric series. Thus $cov(k_t, k_{t-h}) = E(k_t k_{t-h}) - [E(k_t)]^2 = \widetilde{\rho}^{2h} (E(k_{t-h}^2) - [E(k_t)]^2) = \widetilde{\rho}^{2h} var(k_t)$. Thus, the correlation between k_t and k_{t-h} is $\widetilde{\rho}^{2h}$.

Because the stationary distribution of $\sigma_t^2 = 1/k_t$ is that of the product of $(v_{c,t})^{-1}$ and $(\varepsilon_t' \varepsilon_t)^{-1}$, with $(v_{c,t})^{-1}$ being independent of $(\varepsilon_t' \varepsilon_t)^{-1}$, the expectation $E(\sigma_t^{2s})$ is finite if and only if both $E((v_{c,t})^{-s})$ and $E((\varepsilon_t' \varepsilon_t)^{-s})$ are finite. Because $(\varepsilon_t' \varepsilon_t)^{-1}$ is an inverted gamma with n degrees of freedom, $E((\varepsilon_t' \varepsilon_t)^{-s})$ is finite only if $s < n$. In addition, from $v_{c,t} = \widetilde{T}_t(1 + \rho^2 v_{c,(t-1)})$ it follows that:

$$\frac{1}{v_{c,t}} = \frac{1}{\widetilde{T}_t} \frac{1}{1 + \rho^2 v_{c,(t-1)}}$$

Because $(1 + \rho^2 v_{c,(t-1)})^{-s} < 1$, it follows that $E((1 + \rho^2 v_{c,(t-1)})^{-s})$ is finite because the density function of $v_{c,(t-1)}$ integrates up to 1. Because $\widetilde{T}_t$ follows a $B(\underline{\alpha}, \underline{\beta})$, $E(\widetilde{T}_t^{-s})$ is finite if and only if $\underline{\alpha} > s$. Putting both conditions together, $E(\sigma_t^{2s})$ is finite when $\underline{\alpha} > s$ and $n > s$.

Proof of Proposition 2

The likelihood is:

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

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

$$p(k_1|\Theta, \beta) \prod_{t=2}^{T} \left(p(k_t|h_t, \Theta, \beta) p(h_t|k_{t-1}, \Theta, \beta) \right)$$

$$= p(k_1|\Theta, \beta) \prod_{t=2}^{T} \left(p(k_t|h_t, \Theta, \beta) \frac{\frac{\lambda_t^{h_t}}{h_t!}}{\exp(\lambda_t)} \right)$$

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

$$p(k_1|\Theta, \beta) = \frac{|k_1|^{\frac{n-2}{2}}}{c_1} \exp\left(-\frac{1-\widetilde{\rho}^2}{2\widetilde{\theta}^2} k_1 \right) \qquad c_1 = \Gamma\left(\frac{n}{2}\right) \left(\frac{2\widetilde{\theta}^2}{1-\widetilde{\rho}^2}\right)^{n/2} \tag{21}$$

$$p(k_t|h_t, \Theta, \beta) = \frac{|k_t|^{\frac{n+2h_t-2}{2}}}{c_t} \exp\left(-\frac{1}{2\widetilde{\theta}_t^2} k_t \right) \qquad c_t = \Gamma\left(\frac{n}{2} + h_t\right) \left(2\widetilde{\theta}_t^2\right)^{n/2+h_t} \qquad t = 2, ..., T$$

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

$$(2\pi)^{-T/2} \left[\prod_{t=1}^{T} (k_t)^{\frac{n+2h_t-2}{2}+\frac{1}{2}} \right] \exp\left(-\frac{1}{2} \sum_{t=2}^{T} k_t \left(\frac{1}{\widetilde{\theta}_t^2} + r_t^2 \right) \right) \times \tag{22}$$

$$\exp\left(-\frac{1}{2} k_1 \left(\frac{1-\widetilde{\rho}^2}{\widetilde{\theta}^2} + r_t^2 \right) \right) \prod_{t=2}^{T} \left(\frac{\frac{\lambda_t^{h_t}}{h_t!}}{\exp(\lambda_t)} \right) \left(\prod_{t=1}^{T} c_t \right)^{-1} p(\Theta)$$

Recalling that $\lambda_t = \widetilde{\rho}_t^2 k_{t-1}/(2\widetilde{\theta}_t^2)$ and also that $(\widetilde{\rho}_t^2/\widetilde{\theta}_t^2) = (\widetilde{\rho}_{t+1}^2/\widetilde{\theta}_{t+1}^2)$, it is clear that $k_t|h, \Theta, \beta \sim G((n+1)/2 + h_t + h_{t+1}, 2\widetilde{r}_t^2)$. To find the conditional distribution of h given k note that c_t depends on h_t and putting together the terms in (22) that depend on h_t we get:

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

which shows that $h_t|k, \Theta, \beta \sim Bessel(\frac{n-2}{2}, \widetilde{\rho}_t \frac{\sqrt{k_t k_{t-1}}}{\widetilde{\theta}_t^2})$ for $t = 2...T$. The expression for $p(n, \theta^2, \rho^2|Y, h, \beta)$ can be obtained by integrating (22) with respect to k using basic properties of the Gamma distribution.

Proof of Proposition 3:

For the proof let us write the hypergeometric distribution in (14) using the gamma function and the factorial

instead of the binomial coefficients, so that $\Pr\left(d_t = s | h_t, d_{t+1}\right)$ is equal to:

$$\frac{h_t!}{d_t!(h_t - d_t)!} \frac{\Gamma((n+1)/2 + d_{t+1})}{\Gamma((n-1)/2 + d_t)} \frac{\Gamma(2 + d_{t+1})}{\Gamma(2 + d_{t+1} - d_t)} \frac{\Gamma((n-1)/2 + h_t)}{\Gamma((n+1)/2 + d_{t+1} + h_t)}$$

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:

$$\pi(d|h, k, m) = \prod_{t=T}^{2} p\left(d_t | h_t, d_{t+1}\right) \quad \text{with} \quad d_{T+1} = 0$$

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

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

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

$$p(k_1|\Theta, \beta) \prod_{t=2}^{T} \left(p(k_t|h_t, \Theta, \beta) p(m_t|h_t) p(h_t|k_{t-1}, \Theta, \beta)\right)$$

$$= \quad p(k_1|\Theta, \beta) \prod_{t=2}^{T} \left(p(k_t|h_t, \Theta, \beta) \frac{\frac{\lambda_t^{h_t}}{h_t!} m_t^{h_t}}{\exp(\lambda_t)} \frac{\Gamma(\alpha_m + \beta_m + h_t)}{\Gamma(\alpha_m + h_t)\Gamma(\beta_m)} m_t^{\alpha_m - 1} (1 - m_t)^{\beta_m - 1}\right)$$

where $p(k_1|\Theta, \beta)$ and $p(k_t|h_t, \Theta, \beta)$ have been defined in (21), and where $\alpha_m = (n-1)/2$, $\beta_m = 1/2$, as defined before.

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

$$(2\pi)^{-T/2} \left[\prod_{t=1}^{T} (k_t)^{\frac{n+2h_t - 2}{2} + \frac{1}{2}}\right] \exp\left(-\frac{1}{2}\sum_{t=2}^{T} k_t \left(\frac{1}{\theta_t^2} + r_t^2\right) - \frac{1}{2}k_1 \left(\frac{1 - \widetilde{\rho}^2}{\widetilde{\theta}^2} + r_1^2\right)\right) \times$$

$$\prod_{t=2}^{T} \left(\frac{\frac{\lambda_t^{h_t}}{h_t!} m_t^{h_t}}{\exp(\lambda_t)} \frac{\Gamma(n/2 + h_t)}{\Gamma((n-1)/2 + h_t)\Gamma(\beta_m)} m_t^{\alpha_m - 1} (1 - m_t)^{\beta_m - 1}\right) \times$$

$$\pi(d|h, k, m) \left(\prod_{t=1}^{T} c_t\right)^{-1}$$

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:

$$\Gamma\left((n+1)/2 + h_T\right) \left(2\widehat{r}_T^2\right)^{\frac{n+1+2h_T}{2}} (2\pi)^{-T/2} \left[\prod_{t=1}^{T-1} (k_t)^{\frac{n+2h_t - 2}{2} + \frac{1}{2}}\right] \times \tag{23}$$

$$\exp\left(-\frac{1}{2}\sum_{t=2}^{T-1} k_t \left(\frac{1}{\theta^2} + r_t^2\right) - \frac{1}{2}k_1 \left(\frac{1 - \widetilde{\rho}^2}{\widetilde{\theta}^2} + r_1^2\right)\right) \times$$

$$\prod_{t=2}^{T} \left(\frac{\frac{\lambda_t^{h_t}}{h_t!} m_t^{h_t}}{\exp(\lambda_t)} \frac{\Gamma(n/2 + h_t)}{\Gamma((n-1)/2 + h_t)\Gamma(\beta_m)} m_t^{\alpha_m - 1} (1 - m_t)^{\beta_m - 1}\right) p(d|h, k, m) \left(\prod_{t=1}^{T} c_t\right)^{-1}$$

26

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 (23) can be written as:

$$\Gamma\left((n+1)/2+h_T\right)\left(2\widehat{r}_T^2\right)^{h_T}\frac{\lambda_T^{h_T}}{h_T!}m_T^{h_T}\frac{\Gamma(n/2+h_T)}{\Gamma((n-1)/2+h_T)}\times \tag{24}$$

$$\frac{\left(2\widetilde{\theta}_T^2\right)^{-h_T}}{\Gamma(n/2+h_T)}\frac{h_T!}{(h_T-d_T)!}\frac{\Gamma((n-1)/2+h_T)}{\Gamma((n+1)/2+d_{T+1}+h_T)}$$

$$=\quad \left(\widehat{\lambda}_T\right)^{h_T}\frac{1}{(h_T-d_T)!}$$

where we have implicitly used that:

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

Recall that the restriction $d_T\le h_T$ comes from the prior of d_T. Therefore (24) implies that $h_T|k_{T-1},m,d$ is a $SP(\widehat{\lambda}_T,d_T)$. Summing up expression (24) over all values of $h_T\in[d_T,\infty)$ gives $\left(\widehat{\lambda}_T\right)^{d_T}\exp(\widehat{\lambda}_T)$. Thus, integrating out h_T from (23) we obtain:

$$\left(\widehat{\lambda}_T\right)^{d_T}\exp\left(-(\lambda_T-\widehat{\lambda}_T)\right)\times \tag{25}$$

$$\left(2\widehat{r}_T^2\right)^{\frac{n+1}{2}}(2\pi)^{-T/2}\left[\prod_{t=1}^{T-1}(k_t)^{\frac{n+2h_t-2}{2}+\frac{1}{2}}\right]\times$$

$$\exp\left(-\frac{1}{2}\sum_{t=2}^{T-1}k_t\left(\frac{1}{\widetilde{\theta}_t^2}+r_t^2\right)-\frac{1}{2}k_1\left(\frac{1-\widetilde{\rho}^2}{\widetilde{\theta}^2}+r_1^2\right)\right)\times$$

$$\prod_{t=2}^{T-1}\left(\frac{\lambda_t^{h_t}}{h_t!}m_t^{h_t}\frac{\Gamma(n/2+h_t)}{\exp(\lambda_t)}\frac{\Gamma(n/2+h_t)}{\Gamma((n-1)/2+h_t)}\right)\prod_{t=2}^{T}\left(\frac{1}{\Gamma(\beta_m)}m_t^{\alpha_m-1}(1-m_t)^{\beta_m-1}\right)\times$$

$$\frac{1}{d_T!}\frac{\Gamma((n+1)/2)}{\Gamma((n-1)/2+d_T)}\frac{\Gamma(2)}{\Gamma(2-d_T)}\times$$

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

Noting that:

$$\exp\left(-(\lambda_T-\widehat{\lambda}_T)\right)=\exp\left(-\frac{1}{2}\left(\frac{\widetilde{\rho}_T^2}{\widetilde{\theta}_T^2}-m_T\left(\frac{\widetilde{\rho}_T}{\widetilde{\theta}_T^2}\right)^2\widehat{r}_T^2\right)k_{T-1}\right)$$

we obtain that:

$$\exp\left(-\frac{1}{2}k_{T-1}\left(\frac{1}{\widetilde{\theta}_{T-1}^2}+r_{T-1}^2\right)\right)\exp\left(-(\lambda_T-\widehat{\lambda}_T)\right)=\exp\left(-\frac{1}{2\widehat{r}_{T-1}^2}k_{T-1}\right)$$

Therefore the conditional posterior $k_{T-1}|h_{T-1},m,d$ is a $G((n+1)/2+d_T+h_{T-1},2\widehat{r}_{T-1}^2)$. Thus, integrating out k_{T-1} from (25) we obtain:

$$\left(2\widehat{r}_T^2\right)^{\frac{n+1}{2}} \left(2\widehat{r}_{T-1}^2\right)^{\frac{n+1}{2}+h_{T-1}+d_T} \Gamma\left((n+1)/2+d_T+h_{T-1}\right) \times \tag{26}$$

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

$$\exp\left(-\frac{1}{2}\sum_{t=2}^{T-2}k_t\left(\frac{1}{\widetilde{\theta}_t^2}+r_t^2\right)-\frac{1}{2}k_1\left(\frac{1-\widetilde{\rho}^2}{\widetilde{\theta}^2}+r_1^2\right)\right) \times$$

$$\prod_{t=2}^{T-1}\left(\frac{\lambda_t^{h_t}}{h_t!}m_t^{h_t}\frac{\Gamma(n/2+h_t)}{\exp(\lambda_t)}\frac{1}{\Gamma((n-1)/2+h_t)}\right)\prod_{t=2}^{T}\left(\frac{1}{\Gamma(\beta_m)}m_t^{\alpha_m-1}(1-m_t)^{\beta_m-1}\right) \times$$

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

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

$$\Gamma\left((n+1)/2+d_T+h_{T-1}\right)\left(2\widehat{r}_{T-1}^2\right)^{h_{T-1}}\frac{\lambda_{T-1}^{h_{T-1}}}{h_{T-1}!}m_{T-1}^{h_{T-1}} \times \tag{27}$$

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

$$= \left(\widehat{\lambda}_{T-1}\right)^{h_{T-1}}\frac{1}{(h_{T-1}-d_{T-1})!}$$

where:

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

This shows that $h_{T-1}|k_{T-2},m,d$ is a $SP(\widehat{\lambda}_{T-1},d_{T-1})$. Therefore, if we integrate out h_{T-1} from (26) we get:

$$\left(\widehat{\lambda}_{T-1}\right)^{d_{T-1}}\exp\left(-(\lambda_{T-1}-\widehat{\lambda}_{T-1})\right)\left(2\widehat{r}_T^2\right)^{\frac{n+1}{2}}\left(2\widehat{r}_{T-1}^2\right)^{\frac{n+1}{2}+d_T} \times$$

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

$$\exp\left(-\frac{1}{2}\sum_{t=2}^{T-2}k_t\left(\frac{1}{\widetilde{\theta}_t^2}+r_t^2\right)-\frac{1}{2}k_1\left(\frac{1-\widetilde{\rho}^2}{\widetilde{\theta}^2}+r_1^2\right)\right) \times$$

$$\prod_{t=2}^{T-2}\left(\frac{\lambda_t^{h_t}}{h_t!}m_t^{h_t}\frac{\Gamma(n/2+h_t)}{\exp(\lambda_t)}\frac{1}{\Gamma((n-1)/2+h_t)}\right)\prod_{t=2}^{T}\left(\frac{1}{\Gamma(\beta_m)}m_t^{\alpha_m-1}(1-m_t)^{\beta_m-1}\right) \times$$

$$\frac{1}{d_T!}\frac{1}{(d_{T-1})!}\frac{\Gamma((n+1)/2)}{\Gamma((n-1)/2+d_T)}\frac{\Gamma(2)}{\Gamma(2-d_T)}\frac{\Gamma((n+1)/2+d_T)}{\Gamma((n-1)/2+d_{T-1})}\frac{\Gamma(2+d_T)}{\Gamma(2+d_T-d_{T-1})} \times$$

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

This shows that $k_{T-2}|h_{T-2},m,d$ is a $G((n+1)/2+h_{T-2}+d_{T-1},2\widehat{r}_{T-2}^2)$. The other results in Proposition 3

28

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 4

The conditional posterior of $\widetilde{T}$, which is given in (16), comes simply from finding the terms that depend on $\widetilde{T}$ in the product of expression (11) times the prior for $\widetilde{T}$. Multiplying expression (16) times the conditional prior of J (17) gives $\widetilde{T}|J$, which is clearly a Beta distribution. Similarly, the conditional posterior of $V|\widetilde{T}$ is proportional to expression (11) times the prior of V and the term $(\Gamma(V)/(\Gamma(V/2)\Gamma(V/2)))^{T-1}$, which comes from the prior for $\widetilde{T}$.