An alternative to the Inverted Gamma for the variances to modelling outliers and structural breaks in dynamic models

Jairo Fúquene[a]**, María-Eglée Pérez** [b,1] **and Luis Pericchi Guerra**[b,1]

[a]*University of California, Santa Cruz.*
[b]*University of Puerto Rico, Río Piedras Campus.*

Abstract. In this paper we propose a new wide class of hypergeometric heavy tailed priors that is given as the convolution of a Student-t density for the location parameter and a Scaled Beta2prior for the squared scale parameter. These priors may have heavier tails than Student-t priors, and the variances have a sensible behaviour both at the origin and at the tail, making it suitable for objective analysis. Since the representation of our proposal is a scale mixture, it is suitable to detect sudden changes in the model. Finally we propose a Gibbs sampler using this new family of priors for modelling outliers and structural breaks in Bayesian dynamic linear models. We demonstrate in a published example, that our proposal is more suitable than the Inverted Gamma's assumption for the variances, which makes very hard to detect structural changes.

1 Introduction

Strong criticisms against the almost universal use of "vague" Inverted Gamma prior distributions has appeared in Gelman (2006), who also propose half-Student priors for the scale parameter, τ, in hierarchical models. On the other hand, Pericchi (2010) propose to use the Beta Distribution of the Second Kind, (or Beta 2 distribution) as a sensible general replacement of Inverted-Gammas as priors for scale parameters, for hierarchical models. The Scaled Beta 2 distribution for the τ scale is defined as:

$$\pi(\tau) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)\beta} \frac{(\tau/\beta)^{p-1}}{(1+\tau/\beta)^{p+q}}; \quad \tau > 0, \ p > 0 \ q > 0 \ \beta > 0. \qquad (1.1)$$

1

Pérez and Pericchi (2009) use the theory of Regularly Varying (RV) functions as in Andrade and O'Hagan (2005) for checking the robustness of this prior. Pericchi and Pérez (2010) introduce the "Cauchy-Beta2 prior", on which the location conditional on scale is Cauchy and the scale is Beta2. For example, for the case on which $p = q = 1$, it is shown that the marginal prior for the location parameters, fulfils the desiderata that obeys a "horseshoe" density (Carvalho et al., 2010, Theorem 1): i) unbounded at the origin and ii) tails at least as heavy those of a Cauchy (in fact heavier). Furthermore it has an explicit form.

The scaled Beta2 distribution can be defined as a scale mixture of Gammas for the square of the scale as follows (see Pérez and Pericchi, 2009):

$$\tau^2 \sim \text{Gamma}(p, \beta/\rho)$$
$$\rho \sim \text{Gamma}(q, 1)$$

where $\text{Gamma}(a, b)$ denotes the Gamma distribution with pdf given by

$$p(x|\alpha, b) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\{-x/\beta\} \quad a > 0, b > 0,$$

with β the scale parameter. Therefore, the distribution for the squared scale is given by equation (1.1) evaluated in τ^2. For precisions $\lambda = 1/\tau^2$, this prior assignment corresponds to a Scaled Beta 2 with scale $1/\beta$.

Typically the hyper-parameters p, q are fairly small, for example $p = q = 1$, and β quite small, obtaining in this way a bounded density at the origin, flat tails and an vague prior distribution.

Here we model the square of the scale of a Student-t as a scaled Beta 2 prior, and show that the marginal for the location can be written in explicit form. For particular values of the hyper-parameters, the marginal is found analytically. This strategy has several advantages. Among them, it is a suitable heavy tailed prior which can be used in general for scale parameters in Bayesian analysis. Also, our scheme lends itself naturally to a simple Gibbs-Sampling procedure, not adding substantial complication to the Inverted Gamma prior analysis, but improving its performance.

This paper is organized as follows: in Section 2, we show the new family of heavy tailed priors and illustrate their qualities. In section 3 we show the Gibbs sampler proposed for dynamic linear models. In section 4 the potential of our proposal is illustrated in a popular example of the series of quarterly gas consumption in the UK from 1960 and 1986. Some closing concluding remarks are presented in Section 5.

2 A new class of hypergeometric heavy tailed priors

In this paper we consider a conditional Student-t density for a location parameter coupled with the Scaled Beta 2 prior for its squared scale parameter, in order to achieve robustness with respect to the prior, to get sensible prior inputs with heavy tailed distributions and to get closed analytical results for particular values. The original proposal by Pericchi and Pérez (2010) of modelling the scale (as opposed to the square of the scale) leads also to a sensible analysis, and for specific values of the parameters it yields an explicit "Horseshoe" prior, with a pole at zero (for the definition of a horseshoe prior see Carvalho et al., 2010). The proposal here is very similar in its properties (without a pole at zero but a sizeable finite peak at the origin) but it is simpler and easier to implement.

Proposition 2.1. *Let θ follow a Student t distribution with ν degrees of freedom, location μ and scale τ, with pdf*

$$\pi(\theta|\tau^2) = \frac{k_1}{\tau}\left(1 + \frac{1}{\nu}\left(\frac{\theta - \mu}{\tau}\right)^2\right)^{-(\nu+1)/2}, \quad -\infty < \theta < \infty, \nu > 0, -\infty < \mu < \infty, \tau > 0 \tag{2.1}$$

where $k_1 = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}}$, and assign τ^2 a Scaled-Beta2 distribution (1.1) with parameters p, q and β. Therefore

$$\pi(\theta) = \begin{cases} k\beta^q\nu/(\theta-\mu)^{q+1/2}2F1(p+q, q+1/2, (\upsilon+1)/2+p+q, 1-\beta\nu/(\theta-\mu)^2) & if \quad \theta \neq \mu, \\ \\ k_1 Be(p-1/2, q+1/2)/(\beta^{1/2}Be(p,q)) & if \quad \theta = \mu. \end{cases}$$

with $k = k_1 Be(q + 1/2, p + \upsilon/2)/Be(p, q)$, where $Be(a, b)$ denotes the beta function and $2F1(a, b, c, z)$ denotes the hypergeometric function (see 15.1.1, Abramowitz and Stegun, 1970).

Proof. See Appendix. $\qquad\qquad\square$

In the sequel, the marginal prior $\pi(\theta)$ will be called *Student t-Beta2(ν, μ, p, q, β)*. To the best of our knowledge, this prior represents a new and wider class of hypergeometric heavy tailed distributions.

In order to illustrate the properties of these priors, consider the following case: for $\nu = p = q = 1$, the pdf of the Student t-Beta2 is given by

$$\pi(\theta) = \frac{1}{2\sqrt{\beta}\left(1 + \frac{|\theta - \mu|}{\sqrt{\beta}}\right)^2}. \tag{2.2}$$

We can find (2.2) using identities 15.2.5 and 15.1.13 of Abramowitz and Stegun (1970). It is also easy to show that (2.2) is a proper prior. For this specific selection of parameters, this distribution is a particular case of the p-generalized Student-distribution with n degrees of freedom defined in Richter (2007), where $p = n = 1$.

In order to compare (2.2) with the Normal and the Cauchy distributions, we centered them at 0 and matched quartiles to be at ± 1, so that the scale of the Normal is 1.47 and the scale parameters for the Cauchy and the Student t-Beta2 with $\nu = p = q = 1$ are equal to 1. Figures 1 and 2 show that the tails of the Student t-Beta2 distribution are heavier than those of the Normal or the Cauchy.

3 Model Specification and modelling outliers and structural breaks

A Dynamic Linear Model (DLM) is specified (see West and Harrison, 1997; Prado and West, 2010) by the set of equations:

$$
\begin{aligned}
y_t &= F_t \theta_t + \nu_t \quad \nu_t \sim N(0, V_t), \\
\theta_t &= G_t \theta_{t-1} + \omega_t \quad \omega_t \sim N(0, W_t),
\end{aligned}
\tag{3.1}
$$

$t = 1, ...T$. The specification of (3.1) is given by the prior distribution for the initial state θ_0. This is assumed to be normally distributed with mean m_0 and variance C_0, y_t and θ_t are m and n-dimensional random vectors and F_t, G_t, V_t and W_t are real matrices of the appropriate dimension. In our applications y_t is the value of an univariate time series at time t, while θ_t is an unobservable state vector. The original proposal for using heavy tailed priors for modelling and detecting outliers is considered in West (1984). The idea put into action in Petris et al. (2010) is to represent the distributions as a scale mixture, and check when the latent variable is too big or too low, far from one. On the other hand, Petris et al. (2010) propose a Bayesian approach for modelling outliers in dynamic linear models replacing the normal distribution of each component ν_t and ω_t with a scale mixture of normal distributions, leading to a Student-t distribution to obtain a model that accounts for possible outliers and structural breaks (not only in the observation process but also in the state process). Petris et al. (2010) use a Gibbs sampler in their proposal and priors are specified for the degrees of freedom of the Student-t distribution. In our view, although a combination of Gibbs and Metropolis Sampling can be implemented, the clever model proposed by Petris et al. (2010) is overly complex, slow and difficult to analyze and elicit. Our proposal is to use

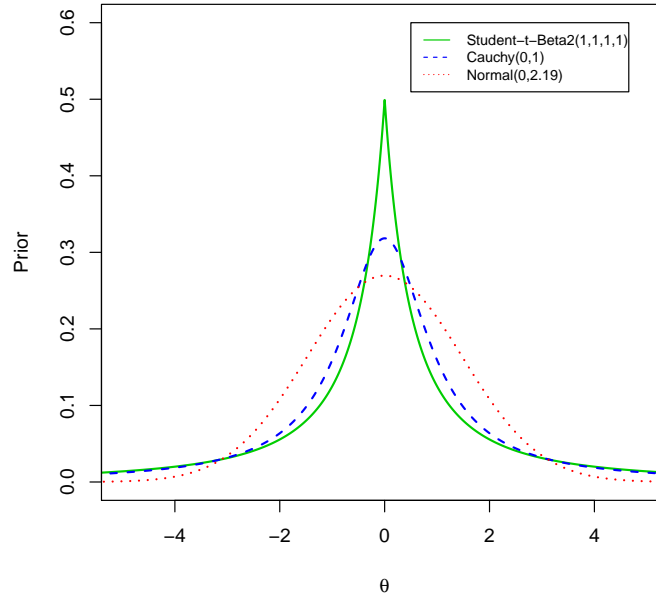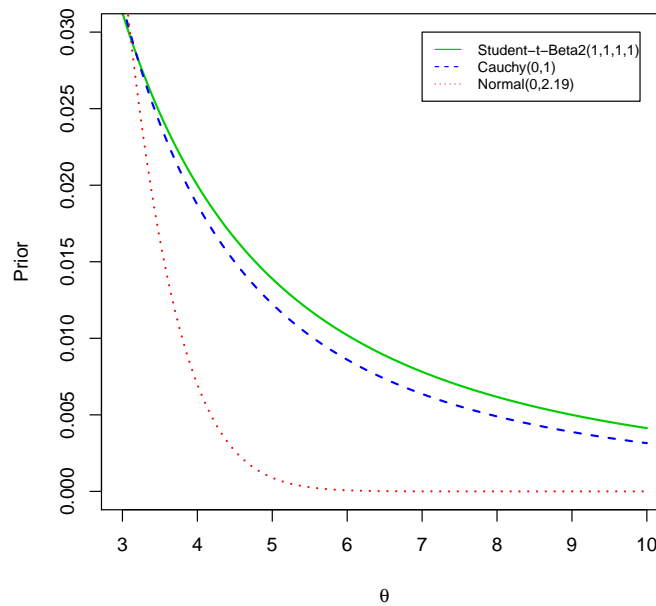**Figure 1** *Comparison of the Cauchy-Beta2,Cauchy(0,1), Normal(0,2.19) priors.*

Figure 2 *Comparison of the tails of the Cauchy-Beta2 ,Cauchy(0,1), Normal(0,2.19) priors.*

the Student-t-Beta(υ,q,p,$\frac{1}{\beta}$) (using the Beta2 prior for the precision $\lambda = 1/\tau^2$) prior for modelling outliers in DLM in order to account outliers in the observations (i.e. abrupt changes in the state vector) of the specify model. $W_{t,i}$ denotes the ith diagonal element of $W_{t,i}$, $i = 1, ..., n$ the hierarchical Student-t-Beta(υ,q,p,$\frac{1}{\beta}$) prior can be summarized in the following display:

$$
\begin{aligned}
V_t^{-1} &= \lambda_y \omega_{y,t}, & W_{t,i}^{-1} &= \lambda_{\theta,i} \omega_{\theta,t_i}, \\
\lambda_y | q &\sim \text{Gamma}(q, (\beta \rho_y)^{-1}), & \lambda_{\theta,i} | q &\sim \text{Gamma}(q, (\beta \rho_{\theta,t_i})^{-1}), \\
\omega_{y,t} &\sim \text{Gamma}(\upsilon/2, 2/\upsilon), & \omega_{\theta,t_i} &\sim \text{Gamma}(\upsilon/2, 2/\upsilon), \\
\rho_y &\sim \text{Gamma}(p, 1), & \rho_{\theta,t_i} &\sim \text{Gamma}(p, 1),
\end{aligned}
$$

We follow here the approach of West (1984) for detection of outliers. Small values of $\omega_{s,t}$ indicate changes in seasonality, small values of $\omega_{\psi,t}$ indicate changes in slope, and similarly for $\omega_{y,t}$ (the observation may be an outlier) and $\omega_{\mu,t}$ (changes in level). For each t, the posterior distribution of $\omega_{y,t}$ (i.e. ω_{θ,t_i}) contains the information of outliers and abrupt changes in the states. Values of $\omega_{y,t}$ (i.e. ω_{θ,t_i}) smaller than one indicate possible outliers or abrupt changes in the states (see Petris et al., 2010).

A Gibbs sampler is implemented using the posterior distribution of parameter and states of the model specified above. For example the full conditional[1] for λ_y is given by:

$$
\pi(\lambda_y | ...) \propto \prod_{t=1}^{T} \lambda_y^{1/2} \exp\left\{ -\frac{\lambda_y \omega_{y,t}}{2}(y_t - F_t \theta_t)^2 \right\} \cdot \lambda_y^{q-1} \exp\left\{ -\beta \rho_y \lambda_y \right\}, \quad (3.2)
$$

hence,

$$
\lambda_y | ... \sim \text{Gamma}\left(q + \frac{T}{2}, \frac{1}{2} SSy^* + \beta \rho_y \right) \tag{3.3}
$$

where $SSy^* = \sum_{t=1}^{T} \omega_{y,t}(y_t - F_t \theta_t)^2$. Now, we make a summary of all the full conditional distributions.

$$
\lambda_y | ... \sim \text{Gamma}\left(q + \frac{T}{2}, \frac{1}{2} SSy^* + \beta \rho_y \right), \quad \lambda_{\theta,i} | ... \sim \text{Gamma}\left(q + \frac{T}{2}, \frac{1}{2} SS_{\theta,i}^* + \beta \rho_{\theta,t_i} \right)
$$

---

[1]The dots on the right-hand side of the conditional vertical bar in $\pi(\lambda_y | ...)$ denote that the distribution is conditioned on every other random variable in the model except λ_y

where $SS_{\theta,i}^* = \sum_{t=1}^{T} \omega_{\theta,t_i}(\theta_{t_i} - (G_t\theta_{t-1})_i)^2$ for $i = 1, 2, ..., n;$

$$\omega_{y,t}|... \sim \text{Gamma}\left(\frac{\upsilon+1}{2}, \frac{\upsilon + \lambda_y(y_t - F_t\theta_t)^2}{2}\right),$$

$$\omega_{\theta,t_i}|... \sim \text{Gamma}\left(\frac{\upsilon+1}{2}, \frac{\upsilon + \lambda_y(\theta_{t_i} - \lambda_{\theta,i}(G_t\theta_{t-1})_i)^2}{2}\right)$$

$$\rho_y|... \sim \text{Gamma}\left(p + q, \beta\lambda_y + 1\right), \quad \rho_{\theta,t_i}|... \sim \text{Gamma}\left(p + q, \beta\lambda_{\theta,i} + 1\right),$$

for $i = 1, ..., n$ and $t = 1, ...T$. Given all the unknown parameters, the states of the DLM are generated using the forward filtering backward sampling (FFBS) given in Fruwirth-Schnatter (1994) which is practically a simulation of the smoothing recursions.

We now show an application of this model. The DLM was fitted using the R software package dlm recently developed by Giovanni Petris (see Petris, 2010). Also, this package implement the FFBS algorithm.

3.1 Application: Quarterly gas consumption in the UK

In this section we consider the series of quarterly gas consumption in the UK from 1960 and 1986 analyzed in Fruwirth-Schnatter (1994), West and Harrison (1997) and Petris et al. (2010) to mention some. In the latest reference an interesting detection of outliers is presented, but with an extremely complicated model, which however assumes that the scales are modeled trough a Inverted Gamma priors. We first show that with a natural model to detect outliers but that use Inverted-Gammas is unable to detect the obvious change in the series.
On the other hand, here we show that a far simpler and easier to understand and implement model is able to detect the changes, when the ScaledBeta2 is assumed, instead of Inverted Gammas.
Plotting the series on the log scale shows some changes in the seasonal factor in the third quarter of 1970 and a DLM obtained as a quarterly seasonal factor model plus a local linear trend model could fit this data reasonably well. The observations (F) and system matrices of the model are:

$$F = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \end{bmatrix} \qquad G = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

The unknown parameters are the observations variance V_t and three elements for W_t:

$$W_t = \left[\begin{array}{ccccc} \sigma^2_{\mu,t}, & \sigma^2_{\xi,t}, & \sigma^2_{s,t}, & 0, & 0 \end{array} \right]$$

where $\sigma^2_{\mu,t}$, $\sigma^2_{\xi,t}$ and $\sigma^2_{s,t}$ are the unknown variances of the level of the series, the slope of the linear trend and the seasonal respectively. We implemented our Gibbs sampler proposed and it is compared with the objective Bayesian strategy:

$$
\begin{aligned}
V_t^{-1} &= \lambda_y \omega_{y,t}, & W_{t,i}^{-1} &= \lambda_{\theta,i} \omega_{\theta,t_i}, \\
\lambda_y &\sim \mathrm{Gamma}(10000, 10000), & \lambda_{\theta,i} &\sim \mathrm{Gamma}(10000, 10000), \\
\omega_{y,t} &\sim \mathrm{Gamma}(2, 1/2), & \omega_{\theta,t_i} &\sim \mathrm{Gamma}(2, 1/2),
\end{aligned}
$$

Note that in summary this approach is to use a Student-t with four degrees of freedom and a non-informative Gamma for modelling the outliers and changes in the states. Selecting 4 degrees of freedom allows having heavy tails without being as peaked as the Cauchy around the origin.

Figure 3 displays the posterior means of the $\omega_{y,t}$ and ω_{θ,t_i}, $t = 1, ..., 108$ and $i = 1, 2, 3$ using the Student-t(4) - non-informative Gamma approach. It is clear that using a Student-t(4)-Gamma(10000,10000) as prior, for modelling the series of quarterly gas consumption in the UK, we obtain that there are no outliers and structural breaks for this series.

Figure 4 displays the posterior means of the $\omega_{y,t}$ and ω_{θ,t_i}, $t = 1, ..., 108$ and $i = 1, 2, 3$ using the Student-t-Beta(4,1,1,$\frac{1}{\beta} = 10000$). We can see that there are different results with the two approaches. Using the Student-t-Beta(4,1,1,$\frac{1}{\beta} = 10000$) we have the expected results for modelling the changes in the dynamic linear models. According to the Bayesian approach proposed there are no observational outliers, excluding the mild outlier in the third quarter of 1983 with an estimated of $\omega_{y,t}$ of 0.83. We can see that this approach indicates that the trend and its slope are stable. There are a lot of structural changes in the seasonal component the most extreme one occurring in the third quarter of 1971 with $E(\omega_{\theta,t_3}|y_{1:t}) = 0.025$.
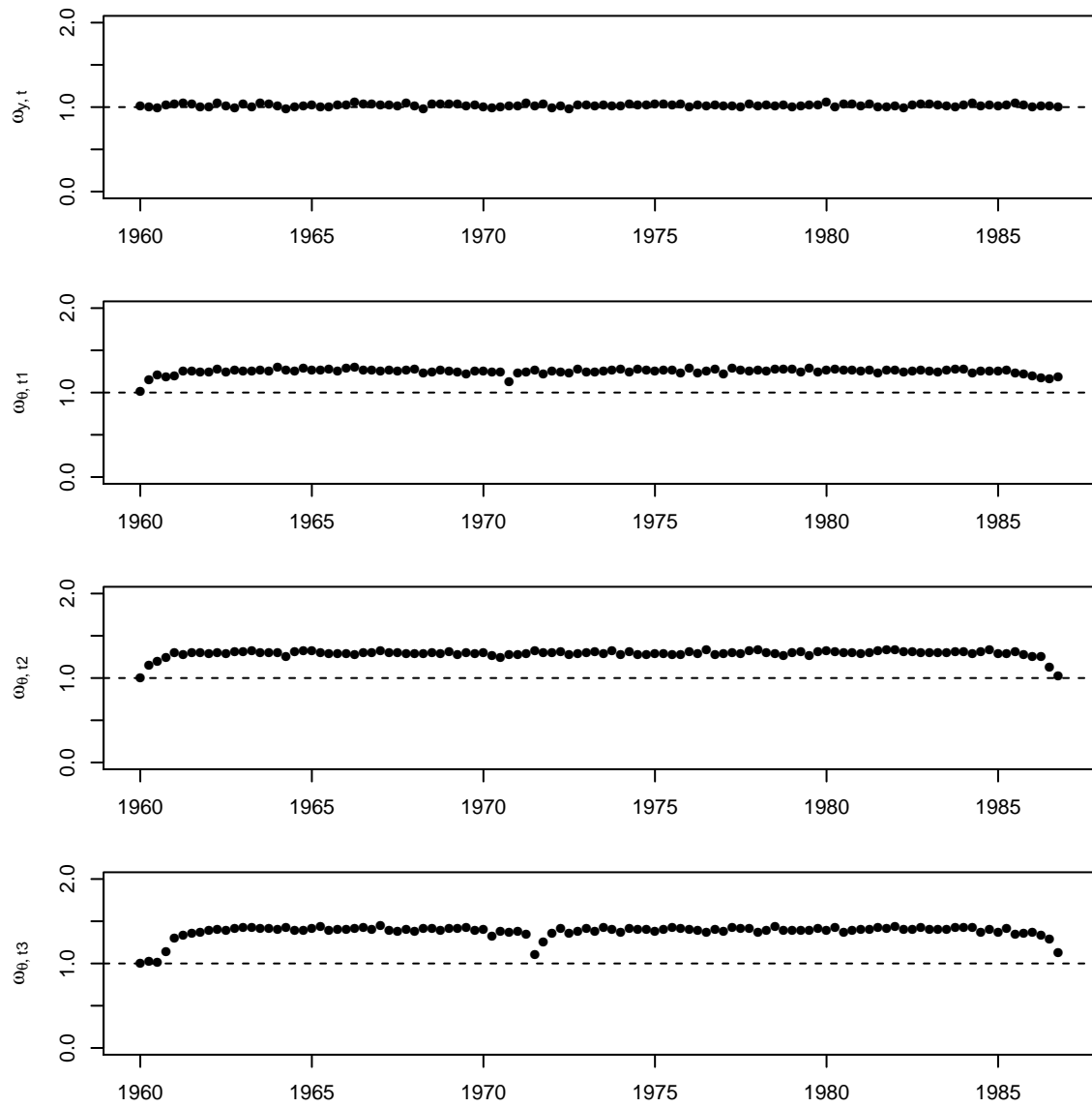
Figure 3 *UK gas consumption: posterior means of the ω_t's using the Student-t(4)-Gamma(10000,10000)*
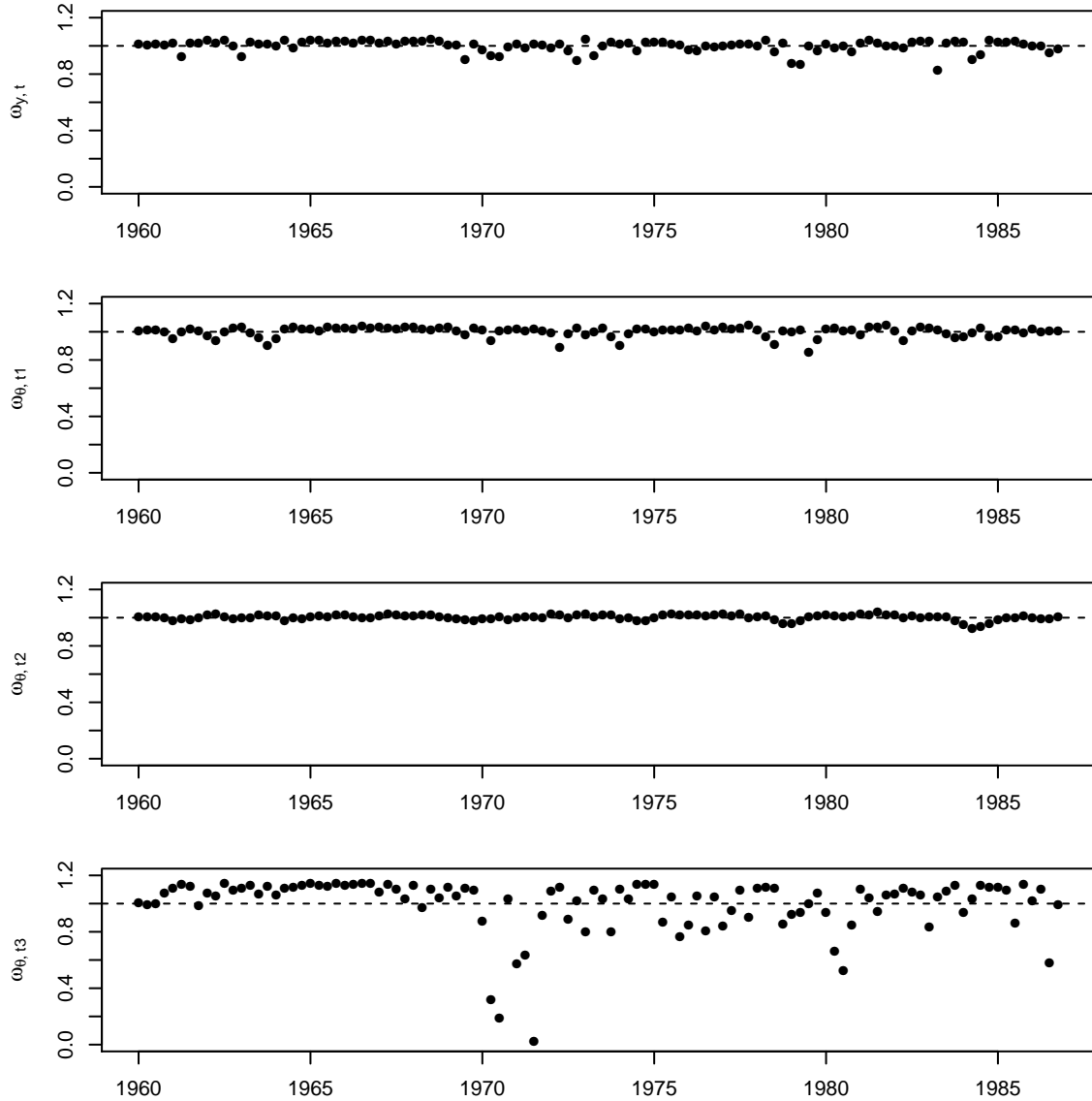
Figure 4 *UK gas consumption: posterior means of the ω_t's using the Student-t-Beta(4,1,1,$\frac{1}{\beta} = 10000$)*

In Figure 5 we have the estimation of the 95% credible intervals for the unobservable seasonal and trend components. We can see that the credible interval for the seasonal component is wider beginning the seventies because it is a period of high variability. These results are very similar than the founded with a more complex model in Petris et al. (2010).
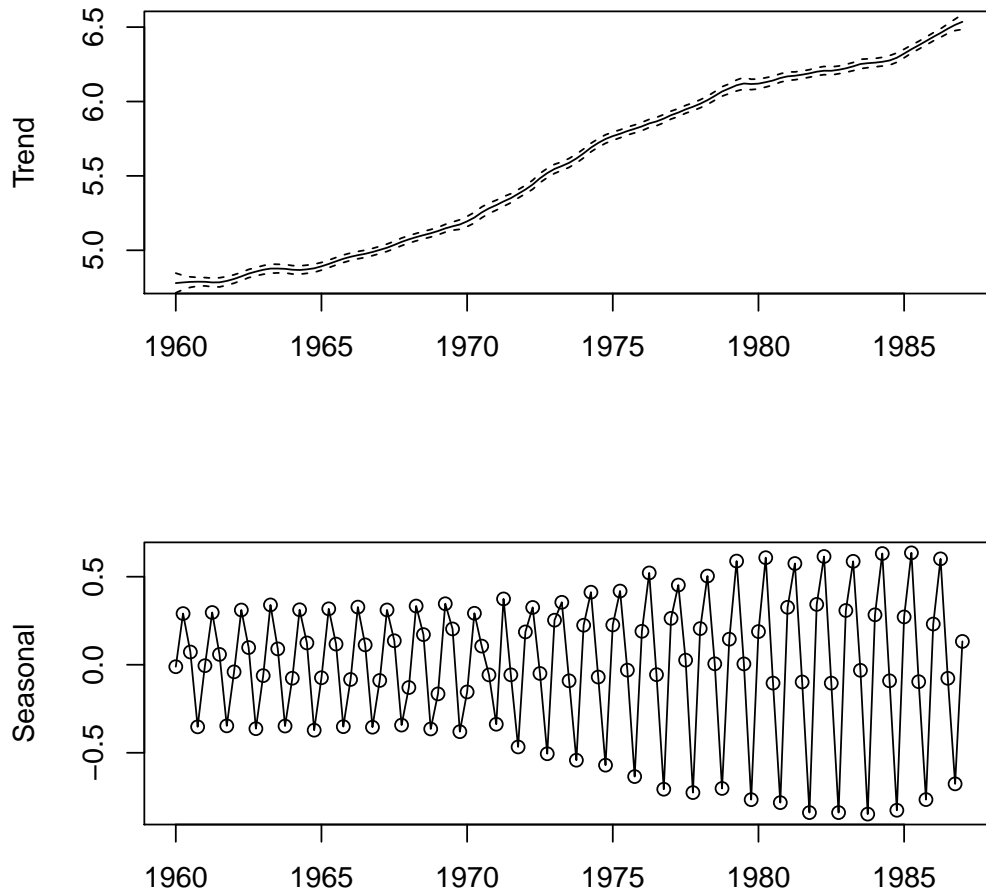

Figure 5 *UK gas consumption: trend and seasonal component, with 95% credible intervals using the Student-t-Beta(4,1,1, $\frac{1}{\beta} = 10000$)*

4 Conclusions

This paper follows up the proposal by Pericchi (2010) and Pérez and Pericchi (2009) to use the Scaled Beta 2 distribution as a sensible general replacement of Inverted-Gammas as priors for scale parameters, for hierarchical models. Here we show that if the **square** of the scales of a Cauchy (or more generally Student-t) are assumed to be distributed as a ScaleBeta2, a general result for the marginal of the location is obtained in terms of the beta and hypergeometric functions, with simple closed forms results for particular hyper-parameters. Furthermore, our scheme lends itself naturally to a simple Gibbs-Sampling procedure, not adding substantial complication to the Inverted Gamma prior analysis, but improving its performance. We suggest these priors as a suitable robust objective analysis for Dynamic Linear Models. The original proposal by Pericchi and Pérez (2010) of modelling the scale (as opposed to the square of the scales) leads also to a sensible analysis, and for particular values it yields an explicit "horseshoe" prior, with a pole at zero. The proposal here is very similar in its properties (without a pole at zero but a sizeable finite peak at the origin) but it is simpler and easier to implement.

Appendix A: Appendix

We have that $\pi(\theta) = \int_0^\infty \pi(\theta|\tau^2)\pi(\tau^2)d\tau^2$, clearly

$$\pi(\theta) = \int_0^\infty \frac{k_1}{\beta \mathrm{Be}(p,q)} \left(\tau + \frac{(\theta-\mu)^2}{\upsilon}\right)^{-(\upsilon+1)/2} \tau^{\upsilon/2} \left(\frac{\tau}{\beta}\right)^{p-1} \left(\frac{\tau}{\beta}+1\right)^{-(p+q)} d\tau$$
$$(\text{A.1})$$

making a change of variable $z = 1/((\upsilon\tau)/(\theta-\mu)^2 + 1)$. Then for $\theta \neq \mu$

$$\pi(\theta) = \frac{\beta^q k_1}{\mathrm{Be}(p,q)} \int_0^1 (1-z)^{\upsilon/2+p-1}(1-z(1-\beta\nu/(\theta-\mu)^2))^{-(p+q)}dz \quad (\text{A.2})$$

therefore

$$\pi(\theta) = k\beta^q\nu/(\theta-\mu)^{q+1/2}2F1(p+q, q+1/2, (\upsilon+1)/2+p+q, 1-\beta\nu/(\theta-\mu)^2)$$
$$(\text{A.3})$$

(see 9.111 of Gradshteyn and Ryzhink, 1965). For $\theta = \mu$ we have that

$$\pi(\theta) = \frac{k_1}{\beta^p \mathrm{Be}(p,q)} \int_0^\infty \tau^{p-3/2} \left(\frac{\tau}{\beta}+1\right)^{-(p+q)} d\tau \qquad (\text{A.4})$$

making a change of variable $z = 1/(\tau/\beta + 1)$ then

$$\pi(\theta) = \frac{k_1}{\beta^{1/2}\mathrm{Be}(p,q)} \int_0^1 (1-z)^{p-3/2} z^{q-1/2} dz \qquad (A.5)$$

therefore

$$\pi(\theta) = k_1 \mathrm{Be}(p-1/2, q+1/2)/(\beta^{1/2}\mathrm{Be}(p,q)) \qquad (A.6)$$

(see 8.380 of Gradshteyn and Ryzhink, 1965).

References

Abramowitz, M. and Stegun, I. (1970), *Handbook of Mathematical Functions. National Bureau of Standards*, Vol. 46, Applied Mathematics Series.

Andrade, J. and O'Hagan, A. (2005), "Bayesian robustness modelling using regularly varying distributions", *Bayesian Analysis* , Vol. 1, pp. 169–188.

Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010), "The horseshoe estimator for sparse signals", *Biometrika* , Vol. 97, pp. 465–480.

Fruwirth-Schnatter, S. (1994), "Data augmentation and dynamic linear models", *Journal of Time Series Analysis* , Vol. 15, pp. 183–202.

Gelman, A. (2006), "Prior distributions for variance parameters in hierarchical models", *Bayesian Analysis* , Vol. 1, pp. 515–534.

Gradshteyn, I. and Ryzhink, I. (1965), *Table of Integrals, Series, and Products*, Academic Press.

Pérez, M. E. and Pericchi, L. R. (2009), The case for a fully robust hierarchical bayesian analysis of clinical trials, *in* 'The 2009 International Workshop on Objective Bayesian Methodology,Wharton School of Business'.

Pericchi, L. R. (2010), "Discussion of Polson, N, and Scott, J", *9th Valencia Meeting on Bayesian Statistics - Oxford University Press* .

Pericchi, L. R. and Pérez, M. E. (2010), "Limiting the shrinkage for the exceptional by objective robust bayesian analysis", *Technical Report Centro de Bioestadística y Bioinformática* , pp. 1–17.
   URL: *http://repositorio.upr.edu:8080/jspui/handle/10586%20/114*

Petris, G. (2010), "An R package for dynamic linear models", *Journal of Statistical Software* , Vol. 36-12, pp. 1–16.

Petris, G., Petrone, S. and Campagnoli, P. (2010), *Dynamic linear models with R*, Springer-Verlag.

Prado, R. and West, M. (2010), *Times Series: Modelling, Computation, and Inference*, Chapman and Hall.

Richter, W.-D. (2007), "Generalized spherical and simplicial coordinates", *Journal of Mathematical Analysis and Applications*, Vol. 336, pp. 1187–1202.

West, M. (1984), "Outliers models and prior distributions in bayesian linear regression", *Journal of the Royal Statistics Society. Series B.* , Vol. 46-3, pp. 431–439.

West, M. and Harrison, P. J. (1997), *Bayesian Forecasting and Dynamic Models*, Springer-Verlag.

Department of Applied Mathematics and Statistics   Department of Mathematics
Jack Baskin School of Engineering   College of Natural Sciences
University of California, Santa Cruz   University of Puerto Rico, Río Piedras Campus
1156 High Street   Box 70377
Mail Stop SOE2   San Juan, PR 00936-8377
Santa Cruz, CA 95064   E-mail: maria.perez34@upr.edu; luis.pericchi@upr.edu
E-mail: jfuquene@ucsc.edu