

Efficient Bayesian Inference in Generalized Inverse Gamma Processes for Stochastic Volatility

Roberto León González (GRIPS)

15 Sept 2016

Departamento de Métodos Cuantitativos en Economía y Gestión
(DMC)

Universidad de Las Palmas de Gran Canaria (ULPGC)

LITERATURA SOBRE MODELOS DE VOLATILIDAD ESTOCASTICA

- Hay mucha evidencia empírica a favor de los modelos de volatilidad estocástica con datos macro (e.g. Sims and Zha 2006) y datos de finanzas (e.g. Kim , Shephard and Chib (1998)).
- Analizamos el siguiente modelo de volatilidad estocástica:

$$y_t = \beta x_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1)$$

- Una amplia literatura ha propuesto mejoras en los algoritmos para el caso en que σ_t sigue un proceso autoregresivo log-normal:
 - Los primeros algoritmos movían solo una volatilidad cada vez, condicionando en todas las demás (*single-move*, e.g. Jacquier, Polson and Rossi 1994).
 - Estrategias de formación de bloques: e.g. Shephard and Pitt (1997), Watanabe and Omori (2004), Asai (2005).
 - Kim , Shephard and Chib (1998) muestran que es posible mover todas las volatilidades latentes al mismo tiempo cuando se condiciona con respecto a algunas variables auxiliares. Además, se puede calcular la integral con respecto a las volatilidades de forma analítica

PROCESOS DE TIPO GAMMA/WISHART

- Una literatura más reciente propone algoritmos para el caso en que σ_t sigue algún tipo de proceso gamma:
 - Philipov and Glickman (2006): volatilidad estocástica multivariante, la distribución condicional de la matriz de volatilidad es una Wishart invertida, mueve una cada vez.
 - Procesos OU con distribución marginal Gamma: Barndorff-Nielsen and Shephard (2001), Roberts et al. (2004), Griffin and Steel (2006), Frühwirth-Schnatter and Sögner (2009).
 - Casarin and Sartore (2007) utiliza Monte Carlo secuencial y filtro de partículas para modelos de volatilidad estocástica multivariante con distribución marginal de Wishart invertida.
 - Fox and West (2011): Procesos estacionarios de Wishart invertida.
 - Triantafyllopoulos (2011) propone un método en dos etapas para estimar una versión restringida de un proceso Wishart.

ESTE TRABAJO

- Este trabajo propone un nuevo algoritmo eficiente y sencillo para el caso en que σ_t^{-1} sigue un proceso Gamma autoregresivo de orden 1 (ARG(1)):
 - Condicionando en unas variables auxiliares, todas las volatilidades se pueden generar conjuntamente. Además, cuando se condiciona en estas variables auxiliares, la integral con respecto a todas las volatilidades se puede realizar analíticamente.
 - Las volatilidades y variables auxiliares se generan con distribuciones sencillas y bien conocidas.
- Este trabajo generaliza el proceso ARG(1) definiendo un proceso donde las colas de la distribución de la volatilidad son más flexibles:
 - En tiempos de turbulencia económica, la volatilidad puede repentinamente tomar valores extremos.
 - Para la misma media, varianza y correlación de primer orden, podemos controlar la probabilidad en las colas de la distribución.
- El modelo propuesto se ajusta mejor que los modelos log-normales para datos de algunas tasas de cambio.

PROCESO GAMMA AUTOREGRESIVO PARA VOLATILIDAD ESTOCASTICA

- Definamos $k_t = \sigma_t^{-1}$ y asumamos que $k_t = z_t' z_t$, donde z_t es un vector $n \times 1$ distribuido como un proceso Gaussiano AR(1):

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

- La distribución conjunta de (k_1, \dots, k_T) es la distribución gamma multivariante, que ha sido analizada por Krishnaiah and Rao (1961). Fue propuesta para analizar la volatilidad observada por Gouriéroux and Jasiak (2006).
- Las propiedades de este proceso son bien conocidas, concretamente:
 - $E(k_t) = \frac{n\theta^2}{1-\rho^2}$, $E(k_t^2) = \left(\frac{\theta^2}{1-\rho^2}\right)^2 n(n+2)$
 - $\text{corr}(k_t, k_{t-h}) = \rho^{2h}$
 - $E(k_t | k_{t-1}) = \rho^2 k_{t-1} + (1-\rho^2)E(k_t)$
 - Distribución condicionada: $\frac{k_t}{\theta^2} | k_{t-1}$ es una chi cuadrado no central.
 - La distribución estacionaria de k_t es una gamma. La condición para que el proceso sea estacionario es: $|\rho| < 1$

Otras propiedades de $(\sigma_1, \sigma_2, \dots, \sigma_T)$:

- La distribución estacionaria de σ_t^2 es una $IG_2(\frac{1-\rho^2}{\theta^2}, n)$.
- $E(\sigma_t^2) = \frac{1-\rho^2}{\theta^2(n-2)}$ para $n > 2$, $var(\sigma_t^2) = (E(\sigma_t^2))^2 \frac{2}{n-4}$ para $n > 4$
- $corr(\sigma_t^2, \sigma_{t-h}^2) = (n/2 - 2)[{}_2F_1(1, 1; n/2; \rho^{2h}) - 1]$, para $n > 4$, donde ${}_2F_1(\cdot)$ es una serie hipergeométrica.
- $E(\sigma_t^2 | \sigma_{t-1}^2) = \frac{1}{\theta^2(n-2)} [{}_1F_1(1; n/2; -\frac{\rho^2}{2\theta^2\sigma_{t-1}^2})]$, para $n > 2$.

PROCESO AUTOREGRESIVO DE GAMMA CON COLAS FLEXIBLES (FTARG)

- Modificamos el proceso ARG para obtener una distribución con colas más flexibles. Recordando que $k_t = z_t' z_t$, en vez de asumir que $z_t = \rho z_{t-1} + \varepsilon_t$ ahora asumimos que:

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

donde $(\tilde{T}_2, \dots, \tilde{T}_T)$ son v.a. independientes beta $B(\alpha_T, \beta_T)$.

- Definiendo $\tilde{\rho}^2 = E(\tilde{T}_t)\rho^2$, $\tilde{\theta}^2 = E(\tilde{T}_t)\theta^2$ y $\tilde{\varepsilon}_t \sim N(0, \tilde{\theta}^2)$, podemos reescribir el modelo como:

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

de tal forma que $k_t = z_t' z_t$ es igual a:

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

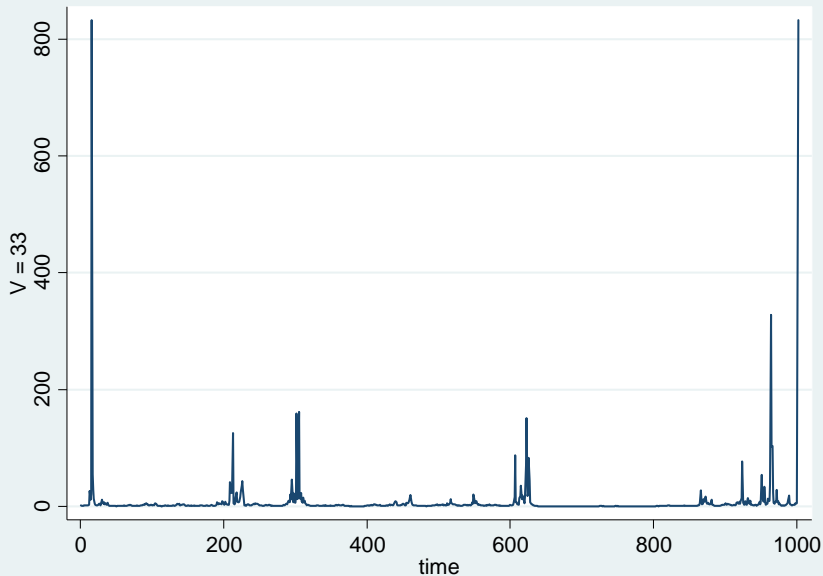
Las propiedades principales de (k_1, \dots, k_T) son:

$$\begin{aligned} E(k_t | k_{t-1}) &= \tilde{\rho}^2 k_{t-1} + (1 - \tilde{\rho}^2) E(k_t) \quad \text{si } \tilde{\rho}^2 < 1 \\ \text{corr}(k_t, k_{t-h}) &= \tilde{\rho}^{2h} \quad \text{si } \tilde{\rho}^2 < 1 \\ E(k_t | k_{t-1}, \tilde{T}_t) &= \frac{\tilde{T}_t}{E(\tilde{T}_t)} (\tilde{\rho}^2 k_{t-1} + (1 - \tilde{\rho}^2) E(k_t)) \end{aligned}$$

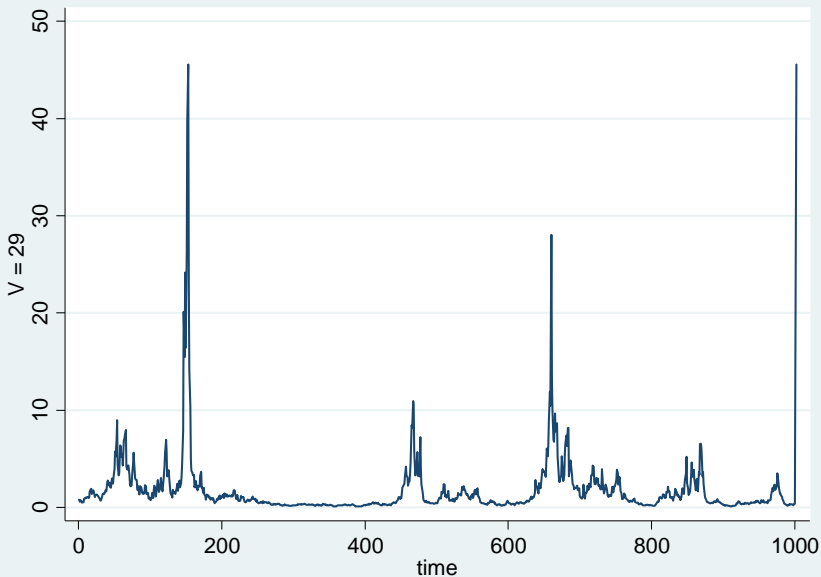
V	n	θ^2	ρ^2	1%	5%	95%	99%	$var(\tilde{T}_t)$
∞	1.28	0.078	1.96	0.003	0.031	8.78	14.53	0
15000	1.29	0.078	1.96	0.003	0.032	8.85	14.61	1.7E-05
7500	1.29	0.078	1.96	0.003	0.032	8.76	14.54	3.3E-05
1000	1.34	0.075	1.96	0.003	0.036	8.77	14.50	2.5E-04
500	1.40	0.072	1.96	0.004	0.041	8.65	14.58	5.0E-04
200	1.60	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: 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.

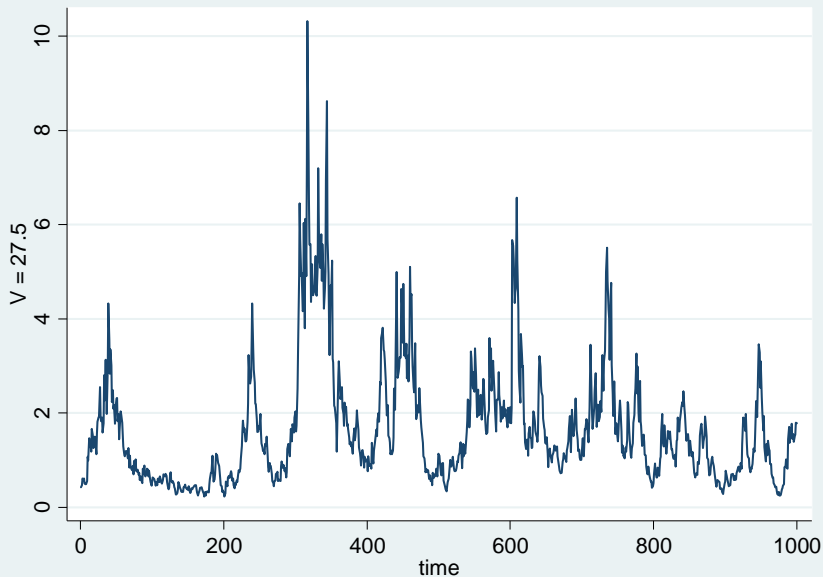
UN VALOR ALEATORIO DE $(\sigma_1^2, \dots, \sigma_T^2)$ (I)



UN VALOR ALEATORIO DE $(\sigma_1^2, \dots, \sigma_T^2)$ (II)



UN VALOR ALEATORIO DE $(\sigma_1^2, \dots, \sigma_T^2)$ (III)



REPRESENTACION DE POISSON

Utilizando la representación de Poisson el modelo puede ser escrito de forma equivalente de esta forma:

$$y_t = x_t \beta + \sqrt{\frac{1}{k_t}} \varepsilon_t$$

$$\varepsilon_t \sim N(0, 1)$$

$$k_t | k_{1:(t-1)}, h_{1:t}, \Theta, \beta \sim G(n/2 + h_t, 2\tilde{\theta}_t^2)$$

$$h_t | k_{1:(t-1)}, h_{1:(t-1)}, \Theta, \beta \sim P(\lambda_t) \quad \text{donde } \lambda_t = \frac{\tilde{\rho}_t^2 k_{t-1}}{2\tilde{\theta}_t^2}$$

donde $G(\cdot)$ representa la distribución gamma, $P(\cdot)$ es la distribución de Poisson, y $k_{1:(t-1)}$ denota $(k_1, \dots, k_{(t-1)})$.

EL ALGORITMO h

Algorithm

El Gibbs h

- *Generar $\Theta|h, \beta$ (método de Metropolis)*
- *Generar $k|h, \Theta, \beta$ (variables independientes con distribución gamma).*
- *Generar $h|k, \Theta, \beta$ (variables independientes con distribución de Bessel).*
- *Generar $\beta|k, h, \Theta$ (sigue una normal multivariante).*

DESVENTAJAS DEL ALGORITMO h

- Tiene dos desventajas principales:
 - Puede resultar en valores altos de autocorrelación porque al fijar h las variables k_t son independientes. Aunque todas las volatilidades se generan de la distribución conjunta $(k_1, \dots, k_T)|h$, en realidad cada una de las volatilidades k_t es independiente de las demás.
 - Para generar la distribución de Bessel correspondiente a $h|k$ es necesario calcular una serie infinita convergente. Sería más sencillo si solo fuese necesario generar h utilizando una Poisson.

EL ALGORITMO m PARA ARG(1)

- Asumamos que m_t condicionado a h_t tiene una distribución beta:

$$m_t | h_t \sim B(\alpha_m + h_t, \beta_m), \quad \alpha_m = (n - 1)/2, \quad \beta_m = 1/2 \quad (2)$$

Observamos que esto requiere que $n > 1$, pero $E(\sigma_t^2)$ es finito solo si $n > 2$.

- La a posteriori de $h_t | (k_{1:(t-1)}, h_{1:(t-1)}, m_{1:t})$ es una mixtura finita de Poissons desplazadas.
- La a posteriori de $k_t | k_{1:(t-1)}, h_{1:t}, m_{1:t}$ continúa siendo una Gamma.

INTRODUCIENDO UNA VARIABLE AUXILIAR MAS

El vector d sigue una distribución hipergeométrica a priori:

$$\Pr(d_t = s | h_t, d_{t+1}) = \frac{\binom{M_{d_t}}{s} \binom{N_{d_t} - M_{d_t}}{n_{d_t} - s}}{\binom{N_{d_t}}{n_{d_t}}} \quad \begin{array}{l} t = 2, \dots, T \\ 0 \leq s \leq \min((1 + d_{t+1}), h_t) \\ d_{T+1} = 0 \end{array}$$
$$\underline{M}_{d_t} = h_t, \quad \underline{n}_{d_t} = 1 + d_{t+1},$$
$$\underline{N}_{d_t} = (n - 1)/2 + h_t + d_{t+1}$$

EL ALGORITMO m

Algorithm

El Gibbs m

- $\Theta | (m, d), \beta$ utilizando el método de Metropolis
- $(k, h) | (m, d), \Theta, \beta$ utilizando distribución Gamma y Poisson.
- $(m, d) | (k, h), \Theta, \beta$ utilizando distribución Beta e Hipergeométrica.
- $\beta | k, h, \Theta$ utilizando una normal multivariante.

FORMULAS SENCILLAS

$$r_t^2 = (y_t - x_t\beta)^2$$

$$\tilde{r}_t^2 = \left(\frac{1 + \tilde{\rho}_t^2}{\tilde{\theta}_t^2} + r_t^2 \right)^{-1} \quad \text{para } t = 2, \dots, T - 1$$

$$\tilde{r}_t^2 = \left(\frac{1}{\tilde{\theta}_t^2} + r_t^2 \right)^{-1} \quad \text{para } t = 1 \text{ y } t = T$$

$$h_1 = h_{T+1} = 0$$

FORMULAS SENCILLAS

$$\hat{r}_T^2 = \tilde{r}_T^2$$

$$\hat{r}_t^2 = \left(\frac{1}{\tilde{r}_t^2} - m_{t+1} \left(\frac{\tilde{\rho}_{t+1}}{\tilde{\theta}_{t+1}^2} \right)^2 \hat{r}_{t+1}^2 \right)^{-1} \quad \text{para } t = 1, \dots, T-1$$

$$m_1 = 1, d_1 = d_{T+1} = h_1 = 0, \lambda_t = \frac{\tilde{\rho}_t^2 k_{t-1}}{2\tilde{\theta}_t^2}, \hat{\lambda}_t = \lambda_t \frac{m_t \hat{r}_t^2}{\tilde{\theta}_t^2},$$

GENERACION RECURSIVA

$$m_t | k, h, d, \Theta, \beta \sim B((n-1)/2 + h_t, 1/2),$$

$$k_t | k_{1:(t-1)}, h_{1:t}, m, d, \Theta, \beta \sim G((n+1)/2 + h_t + d_{t+1}, 2\hat{r}_t^2)$$

$$h_t | k_{1:(t-1)}, h_{1:(t-1)}, m, d, \Theta, \beta \sim SP(\hat{\lambda}_t, d_t)$$

EL ALGORITMO PARA EL MODELO FTARG

- Es igual que para el modelo ARG(1) pero añade los siguientes tres pasos:

Algorithm

Pasos adicionales para el modelo FTARG:

- $V|(k, h), \Theta, J, \beta$ utilizando el método de Metropolis.
- $\tilde{T}|(k, h), \Theta, J, V, \beta$ utilizando distribuciones beta.
- $J|(k, h), \Theta, \tilde{T}, J, V, \beta$ utilizando la distribución binomial negativa.

EFICIENCIA DE LOS ALGORITMOS

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

Table: 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 A. 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.

EFICIENCIA DE LOS ALGORITMOS

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

EFICIENCIA DE LOS ALGORITMOS

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

Table: 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.

ILUSTRACION EMPIRICA

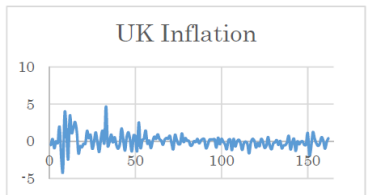
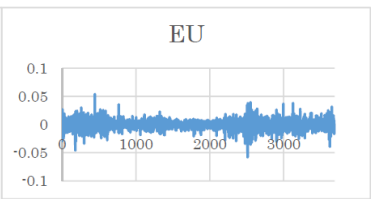
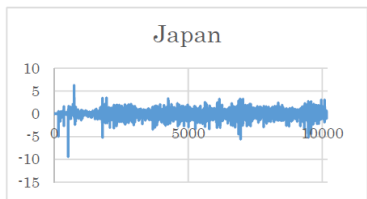
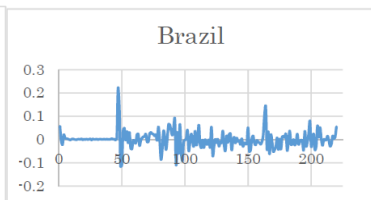
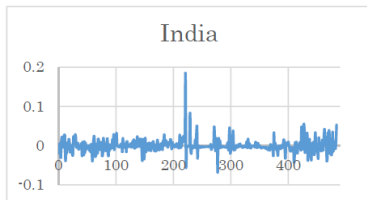
- y_t es la primera diferencia del log de la tasa de cambio:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \sigma_t \varepsilon_t$$

IND-US	Indian Rupee - US \$, monthly : 3/1973 - 6/2013, T=484
BRA-US	Brazilian Real - US \$, monthly: 3/1995 - 6/2013, T=220
JP-US	Yen - US \$, daily: 6 Jan 1971 - 15 Jul 2011, T=10168
EU-US	€ - US \$, daily: 6 Jan 1999 - 17 May 2013, T=3615
UK-INFL	Quarterly Inflation U.K., 1971Q1 - 2011Q4, T=162
UK-UR	Unemployment Rate U.K., 1971Q1 - 2011Q4, T=162

Table: Description of variables used in empirical analysis

RESIDUOS DEL METODO DE MCO



VALOR DE LA LOG-VEROSIMILITUD

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

Table: 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).

VALOR DE LA LOG VEROSIMILITUD MARGINAL

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

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

CONCLUSIONES

- Este trabajo desarrolla un nuevo algoritmo eficiente y sencillo para inferencia Bayesiana en un modelo de volatilidad estocástica de tipo Gamma.
- Al comparar varias estrategias algorítmicas, el algoritmo más eficiente es el algoritmo m:
 - $(k, h)|(m, d, n, \theta^2, \rho)$ utilizando Gamma y Poisson.
 - $(m, d,)|(k, h, n, \theta^2, \rho)$ utilizando beta y binomial negativa.
 - $(n, \theta^2, \rho)|(m, d)$ con el método de Metropolis.
- El algoritmo preferido es al menos 8 veces más eficiente que las otras alternativas consideradas en este trabajo, y es más sencillo de utilizar.
- Este trabajo propone una generalización del modelo ARG introduciendo un parámetro adicional que controla el comportamiento de las colas de la distribución. Este parámetro determina la intensidad del aumento de la volatilidad en periodos de turbulencia.
- Cuando los datos presentan volatilidad extrema, los procesos propuestos de Gamma generalizada se ajustan mejor que los modelos de volatilidad log-normal y términos de error con distribución t.