A Robust Bayesian Dynamic Linear Model to Detect Abrupt Changes in an Economic Time Series: The Case of Puerto Rico

Jairo Fúquene. *
Marta Álvarez.†
Luis Pericchi.‡

March 26, 2013

Abstract

Economic indicators time series are usually complex with high frequency data. The traditional time series methodology requires at least a preliminary transformation of the data to get stationarity. On the other hand, the Robust Bayesian Dynamic Model (RBDM) does not assume a regular pattern and stability of the underlying system but can include points of statement breaks. In this paper, we estimate the Consumer Price Index and the Economic Activity Index of Puerto Rico using a RBDM with observational and states variances to model the outliers and structural breaks in the time series. The results show the model detects structural changes in both series, some of them just before the beginning of a recession period in Puerto Rico's economy.

*Department of Applied Mathematics and Statistics Jack Baskin School of Engineering University of California, Santa Cruz, USA. jfuquene@soe.ucsc.edu.

†Institute of Statistics, School of Business Administration, University of Puerto Rico, Río Piedras Campus. PO Box 23332, San Juan, Puerto Rico 00931-3332.

‡Department of Mathematics, University of Puerto Rico, Río Piedras Campus. PO Box 23332, San Juan, Puerto Rico 00931-3332. luis.pericchi@upr.edu.

Keywords: *Dynamic Models, Consumer Price Index, Bayesian Robustness.*

1 Introduction

In this paper we present a Robust Bayesian Dynamic Model (RBDM) to model the outliers and structural breaks of a time series. This model permits to take into account not only the outliers and structural breaks of the historical series, but also allow us to do posterior inference using a Bayesian approach. A linear trend model, also called linear growth model, with robust priors for the distributions of the state and observational precisions errors was fitted to the historical series of two economic indexes widely used to describe the economic situation of a country: the Consumer Price Index (CPI) and the Economic Activity Index (EAI). The former measures inflation through the fluctuations of prices, and the latter measures the real economic activity.

In Puerto Rico, the CPI is produced and published by the Department of Labor and Human Resources (DTRH) and the EAI is produced and published by the Government Development Bank (GDB). The CPI represents the average monthly change in prices of the items and services that urban families usually buy, and it is used to study the behavior of inflation. The CPI began as an index of cost of living in Puerto Rico, published as the Index of Cost of Life for the Families of Workers (see Department of labor and human resources (2008)), and was developed in the early years of the 1940's as an index of cost of living for working families. This index was reviewed approximately every ten years. For the first time, in the Study of Income and Expenses of 1977, information was collected about the urban families in Puerto Rico, not only of working families, including the self-employed and the pensioners, among others. The use of the original index was discontinued in 1980. With the new information collected in 1977, a new index was born, the index of prices to urban consumers, widely known as the Consumer Price Index (CPI). After the Study of 1977, a Study of Income and Expenses is done again in the years 1999-2003. Since the elapsed time between studies was more than twenty years, in 1990 a few adjustments to the items and services included in the basket that makes up the CPI were made, including the addition of new products. The continuing revision of the index is extremely important so that it reflects accurately the changes in tastes and preferences of the population, as well as the inclusion of new products in the market that are not represented in the previous basket. A major change in the CPI calculation methodology is implemented in March 2010, incorporating the use of the geometric mean for the aggregation of prices, a measure used by the majority of countries, including the United States. Currently, the base year for the basket of goods for the CPI is the year 2006 (the value of the CPI for 2006 is 100). This basket is composed of the following major groups: food and beverages, housing, apparel, transportation, medical care, entertainment, education and communication, and other items and services. These groups are similar to the ones in the United States basket (see Department of labor and human resources (2008)).

The Economic Activity Index (EAI) of the Government Development Bank, on the other hand, is a monthly index that includes the behavior of four economic indicators: total of salaried employment (thousands), sale of cement (million bags), fuel consumption (millions of gallons) and electricity generation (million KWH). Until March 2012, one of the indicators was electrical energy consumption, but it was replaced since then by electricity generation. The Government Development Bank for Puerto Rico (2012) attributed this change to “the electricity generation data is obtained in a timelier manner, is more reliable because they have fewer revisions, and there is a slightly higher correlation with the Gross National Product (GNP) of Puerto Rico". The Government Development Bank for Puerto Rico (2012) reported that the correlation of this new index with the GNP is 0.97, which means that they are highly correlated.

We consider the monthly series of the CPI from January 1980 to December 2012; the monthly data for the AEI goes from January 1980 to December 2011. The CPI and the AEI have had different abrupt changes in the last years. Therefore, it is interesting to have a model that allows the detection of these changes in both indexes. As the CPI and AEI are two important measures of the economy, we will also investigate the relationship between the changes the two series have had in Puerto Rico.

The paper is organized as follows. Section 2 shows the RBMD prior variances specification for the RBDM. Section 3 presents the RBDM to modeling the CPI and the AEI. Finally we have the conclusions and remarks in Section 4.

2 Model Specification and modelling outliers and structural breaks

A univariate Dynamic Linear Model (DLM) is specified (see West & Harrison (1997)) by the set of equations:

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

where $t = 1 : T$. The specification 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. y_t is the value of an univariate time series at time t, while θ_t is an unobservable state vector. On the other hand, the scaled Beta2 prior for the precision $\lambda = 1/\tau^2$ is the following:

$$\pi(\lambda) = \frac{\Gamma(q+p)}{\Gamma(q)\Gamma(p)}\beta\frac{(\beta\lambda)^{q-1}}{(1+\beta\lambda)^{p+q}}; \quad \lambda > 0 \tag{2}$$

where small values of β are considered in order to have heavy tails priors for robust inference. This paper consider the Student-t density coupled with the scaled Beta2 (see Appendix B) for modelling the observational and states variances in RBDM (see Fúquene, Pérez & Pericchi (2013)). Therefore let $\theta \sim$ Student-t(μ, τ, υ) where υ are the degrees of freedom, μ the location and τ the scale of the Student-t density:

$$\pi(\theta|\tau^2) = \frac{k_1}{\tau}\left(1+\frac{1}{\upsilon}\left(\frac{\theta-\mu}{\tau}\right)^2\right)^{-(\upsilon+1)/2}, \quad \upsilon > 0, -\infty < \mu < \infty, -\infty < \theta < \infty, \tag{3}$$

where $k_1 = \dfrac{\Gamma((\upsilon+1)/2)}{\Gamma(\upsilon/2)\sqrt{\upsilon\pi}}$. We have that $\pi(\theta) = \int_0^\infty \pi(\theta|\tau^2)\pi(\tau^2)d\tau^2$ therefore

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

with $2F1(a, b, c, z)$ the hypergeometric function (see 15.1.1 of Abramowitz & Stegun (1970)) and we have that $\pi(\theta)$ is the Student-t-Beta(υ,p,q,β) (see Fúquene et al. (2013) for the proof of this result). For example if $\upsilon = p = q = 1$ we have the Student-t-Beta2(1,1,1,β)) such as:

$$\pi(\theta) = \frac{1}{2\sqrt{\beta}\left(1 + \frac{|\theta - \mu|}{\sqrt{\beta}}\right)^2} \tag{4}$$

We can see in Figures 1 and 2 the student-t-Beta(1,1,1,β) prior has heavier tails than the Cauchy prior. That means these priors can detect outliers even far from to these obtained with student priors. Also, using these priors a simple Gibbs sampler can be used because all full conditional densities in the gamma hierarchical parameters have gamma distributions.

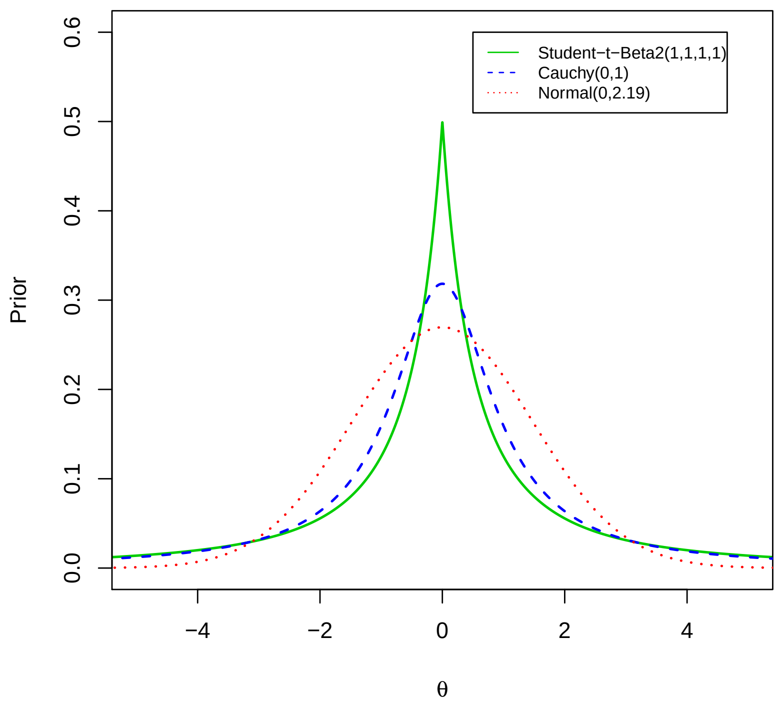

Figure 1: Comparison of the Student-t-Beta2(1,1,1,1) ,Cauchy(0,1), Normal(0,2.19) priors.

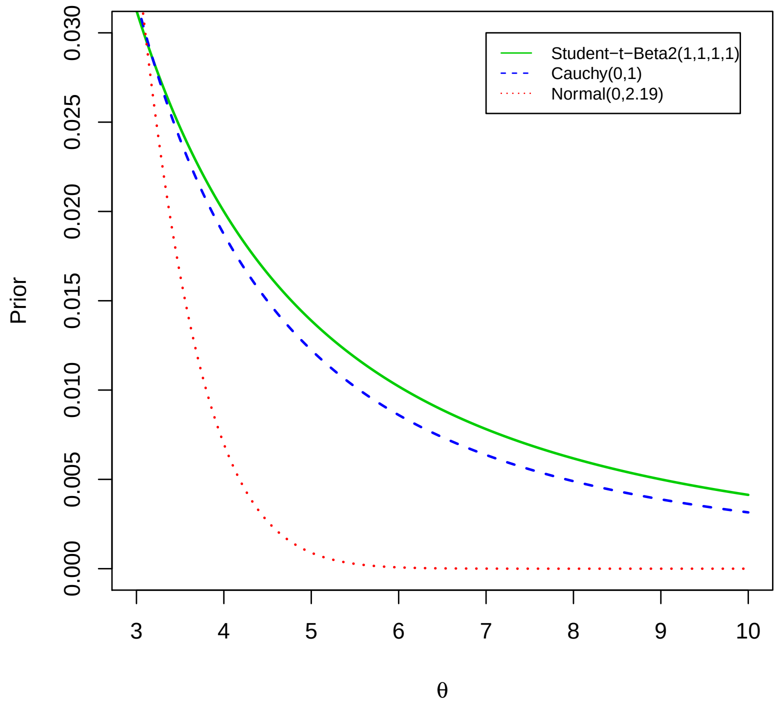

Figure 2: Comparison of the tails of the Student-t-Beta2(1,1,1,1) ,Cauchy(0,1), Normal(0,2.19) priors.

In order to model the CPI we use the Student-t-Beta(υ,q,p,$\frac{1}{\beta}$) (using the Beta2 prior for the precision $\lambda = 1/\tau^2$). $W_{t,i}$ 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 such as:

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

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, Petrone & Campagnoli (2010)). A Gibbs sampler is implemented using the posterior distribution of parameter and states of the model specified above.

We have the annual Consumer Price Index in Puerto Rico on a log-scale in Figure 3.

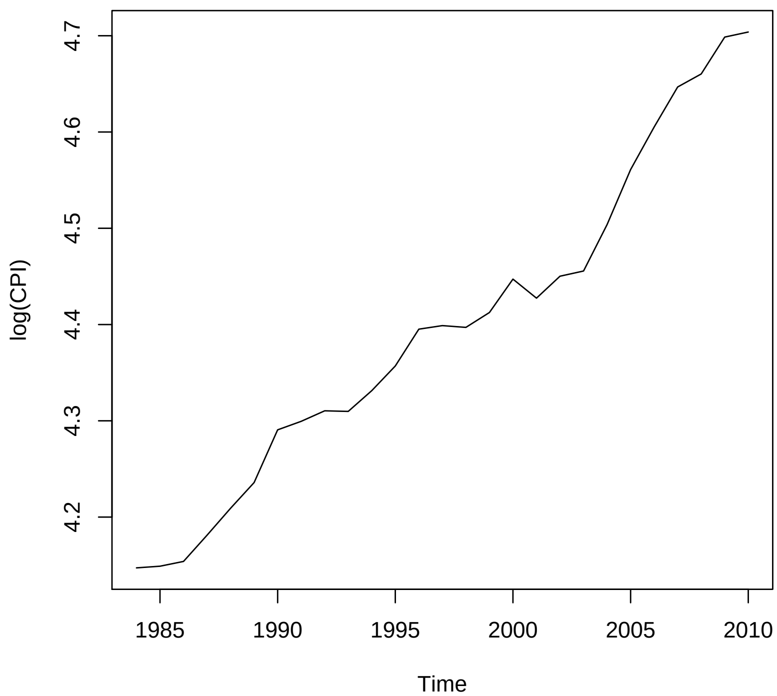

Figure 3: Annual Consumer Price Index in Puerto Rico on a log-scale, 1984-2010.

The natural choice for modelling the logarithm of the CPI using a DLM is a local linear trend model or also called linear growth model which fit the trend and slope of the CPI logarithm. The linear growth model is the following:

$$
\begin{aligned}
y_t &= \mu_t + \nu_t, & \nu_t &\sim N(0, V_t), \\
\mu_t &= \mu_{t-1} + \xi_{t-1} + \omega_{t,1}, & \omega_{t,1} &\sim N(0, \sigma^2_{t,\mu}), \\
\xi_t &= \xi_{t-1} + \omega_{t,2}, & \omega_{t,2} &\sim N(0, \sigma^2_{t,\xi}),
\end{aligned} \tag{5}
$$

with uncorrelated errors ν_t, $\omega_{t,1}$ and $\omega_{t,2}$. This is a DLM with:

$$
\theta_t = \begin{bmatrix} \mu_t \\ \xi_t \end{bmatrix}, \qquad G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad W_t = \begin{bmatrix} \sigma^2_{\mu,t} & 0 \\ 0 & \sigma^2_{\xi,t} \end{bmatrix}, \qquad F = \begin{bmatrix} 1 & 0 \end{bmatrix}.
$$

3 The Robust Bayesian Dynamic model for the CPI

In this section we implement the proposed model. First a motivating example is showed where the proposed model is illustrate using the annual logarithm of the CPI. Next we apply our approach to the CPI and AEI monthly indexes.

3.1 A Robust Bayesian Dynamic model for the annual logarithm of the CPI

We implement the proposed model for modelling the annual logarithm of the CPI. We use a Student-t with four degrees of freedom and a non-informative Gamma for modelling the outliers and changes in the variance states. The choice $\nu = 4$ degrees of freedom is not new. For example Gelman (2004) recommend the Student-t prior with four degrees of freedom in order to obtain robustness in Bayesian applications.

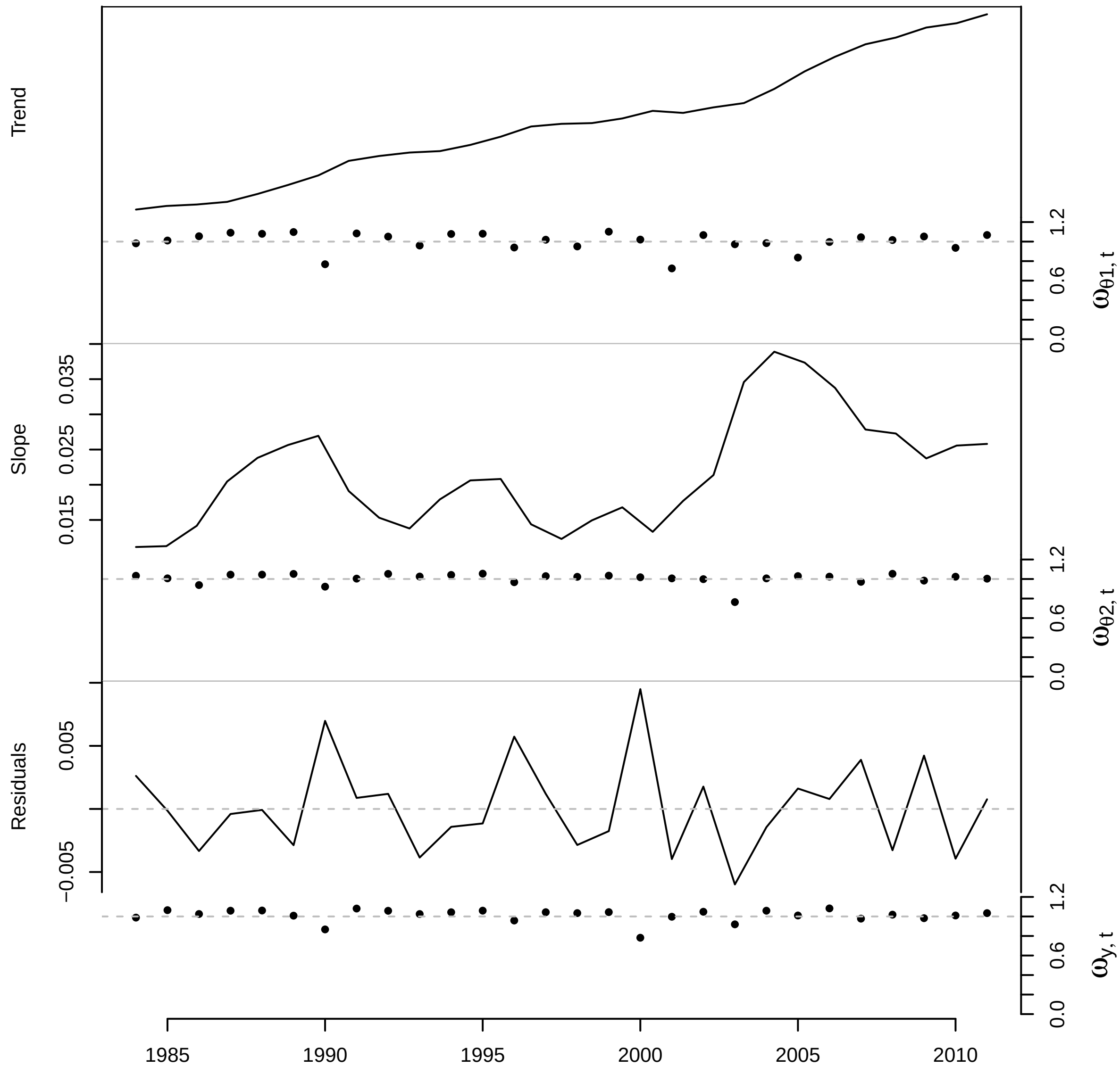

Figure 4: Outliers and structural breaks in the annual logarithm Consumer Price Index in Puerto Rico using the robust approach

In the proposed model we use a Student-t-Beta2 where $p = q = 1$, and $1/\beta = 10000$ for the posterior robustness inference. We obtained convergence of all parameters using 30000 iterations after a burn-in period of 10000 iterations. The Figure 4 displays the proposed model we can see there is an outlier at 2000. The trend shows changes in 1990 and 2001 and the slope has a change in 2003. The proposed model has taken into account not only the outliers but also the structural breaks for the annual logarithm of the CPI.

3.2 Bayesian Dynamic Robust Model for the CPI and EAI monthly indexes

The motivating example of the previous section shows how the proposed model works. However, the most interesting case is when the model is fitted to the monthly series of the two indexes studied. Since the CPI and AEI indexes could be related, we apply the proposed model to both series, in order to have a broader analysis of the changes during the period studied.

The EAI is one of the indicators used to detect a recession period in the economy, since it is highly correlated to the Gross National Product (GDP). When the GDP decreases for more than six months consecutively, the economy could be going through a recession period. In Puerto Rico, the recession periods identified during the last 35 years are: 1980-82, 1990-91, 2001-02 and since 2006.

Figure 5 displays the results using the proposed model for the CPI for the period of January 1980 to December 2012. Important remarks can be made. The residuals in the bottom of Figure 6 are given by $\hat{\epsilon}_t = y_t - E(F\theta_t|y_{1:T})$. By looking at the residuals it can be seen there is only one outlier in January 2009. The slope is dynamically changing with only a sudden jump in September 1981, within the first recession period.

The trend has different jumps, the most dramatic one in September 2005 with $E(\omega_{\theta,t_1}|y_{1:T}) =$

Table 1: Posterior mean of ω_{θ,t_1} for the monthly logarithm of the CPI.

month/year	$E(\omega_{\theta,t_1}\|y_{1:T})$ - CPI
May/1980	0.26512911
July/1989	0.15719286
September/1990	0.25577646
April/2005	0.20028956
September/2005	**0.07095877**
December/2005	0.26580040
April/2006	0.25441340
October/2006	0.13192695
December/2006	0.12591796
June/2008	0.26553021
November/2008	0.11320069
December/2008	0.13743595

0.07 and some other abrupt changes in the precedent years. Even though the Consumer Price Index is an indicator of inflation and not recession, this dramatic change could have been an "alarm" for the economic recession Puerto Rico is facing since 2006. Other dramatic changes are found in May 1980, July 1989 and September 1990, all of them just before the beginning of a recession period in Puerto Rico, as shown in the previous paragraph. The structural break of 1980 could be also tied to the changes in the methodology to compute the CPI, as discussed in the Introduction. The change in the trend of the CPI at the years 1989 and 1990 could also be related to the implementation of the "Joint Committee on Taxation" in the United States, and its effect on the island.

Table 2: Posterior mean of ω_{θ,t_1} for the monthly logarithm of the EAI.

month/year	$E(\omega_{\theta,t_1}\|y_{1:T})$ - EAI
July/1980	0.4879160
March/1983	0.3089463
December/1987	0.4912573
September/1989	0.2574237
December/1989	0.2864106
January/1990	0.4640351
July/1996	0.2484716
December/1996	0.3511382
September/1998	0.1563733
October/1998	0.3810037
November/1998	0.2929802
December/1998	0.2276613
December/2001	0.4530024
September/2005	0.3313795

Figure 6 shows that the abrupt changes in the trend for the EAI are much related with those found in the CPI. In particular, the most dramatic ones are presented in September 1989 and December 1989, just before the beginning of the recession period of 1990, July 1996, September 1998, December 1998, and September 2005, coinciding with the break in the CPI series, just before the beginning of the 2006 recession. Figure 7 displays the relationship between the indexes for the periods 2000-2005 and 2005-2010. In particular, the changes in the period 2000-2005 for the two time series are very similar. The abrupt changes in both indexes during the second period may be the consequence of the economic crisis that Puerto Rico has been suffering since 2006.

In conclusion, the RBD models fitted to the series of the Economic Activity Index and the Consumer Price Index of Puerto Rico, during the period 1980 - 2012, detect abrupt changes in both series that may be tied to the beginning of a recession period in the economy.

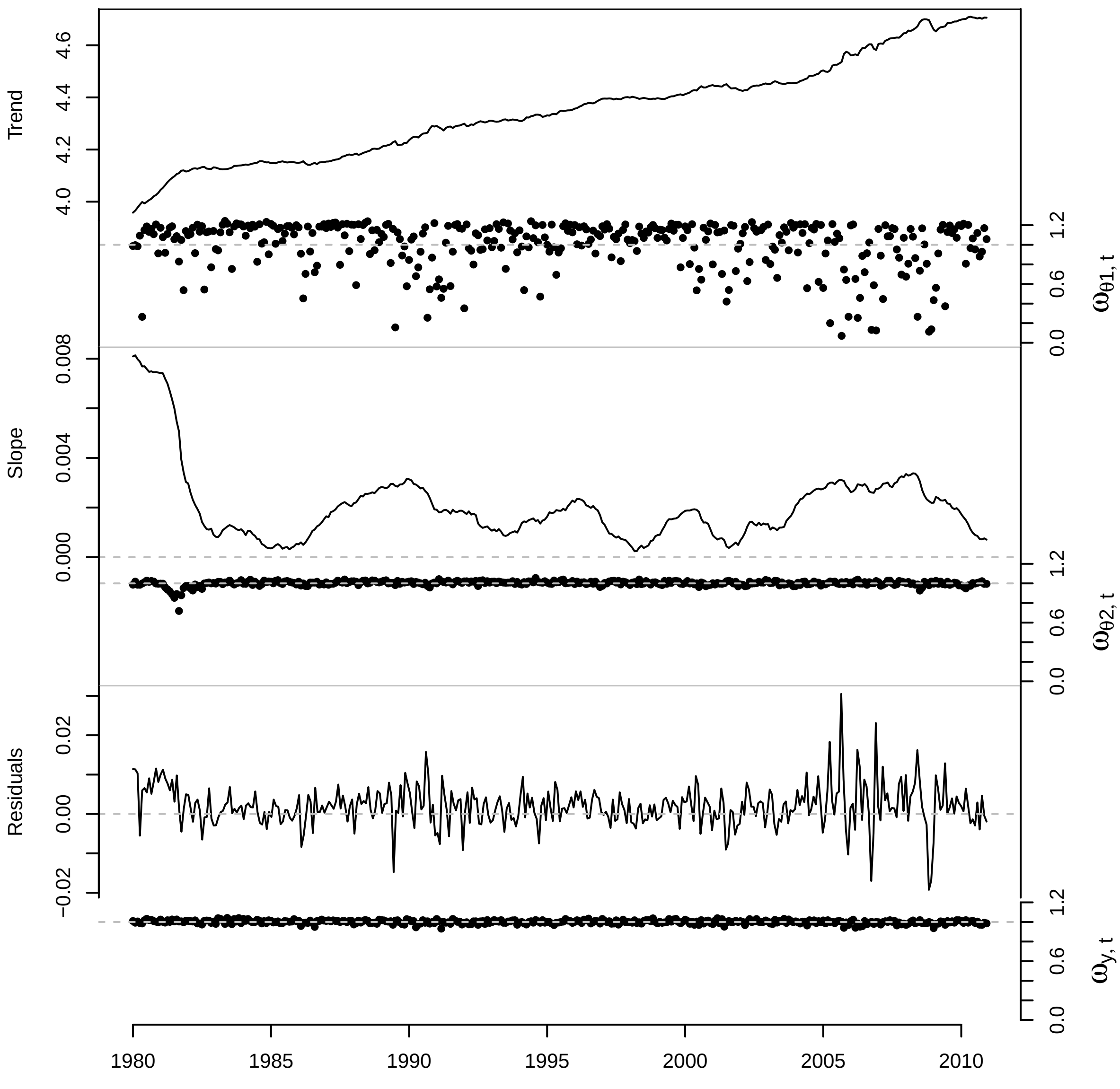

Figure 5: Outliers and structural breaks in the logarithm monthly Consumer Price Index in Puerto Rico

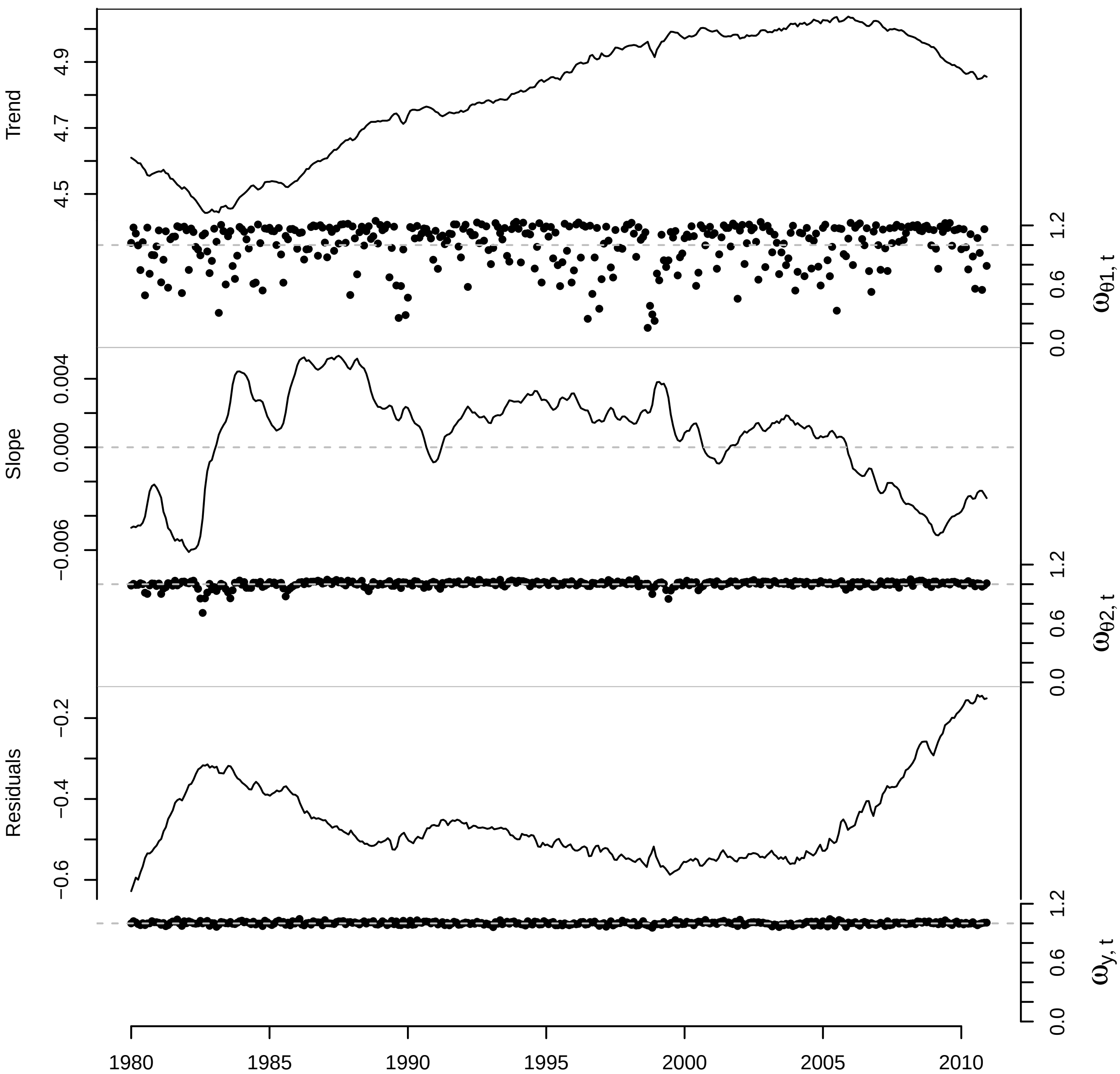

Figure 6: Outliers and structural breaks in the logarithm monthly Economic Activity Index in Puerto Rico using the usual approach

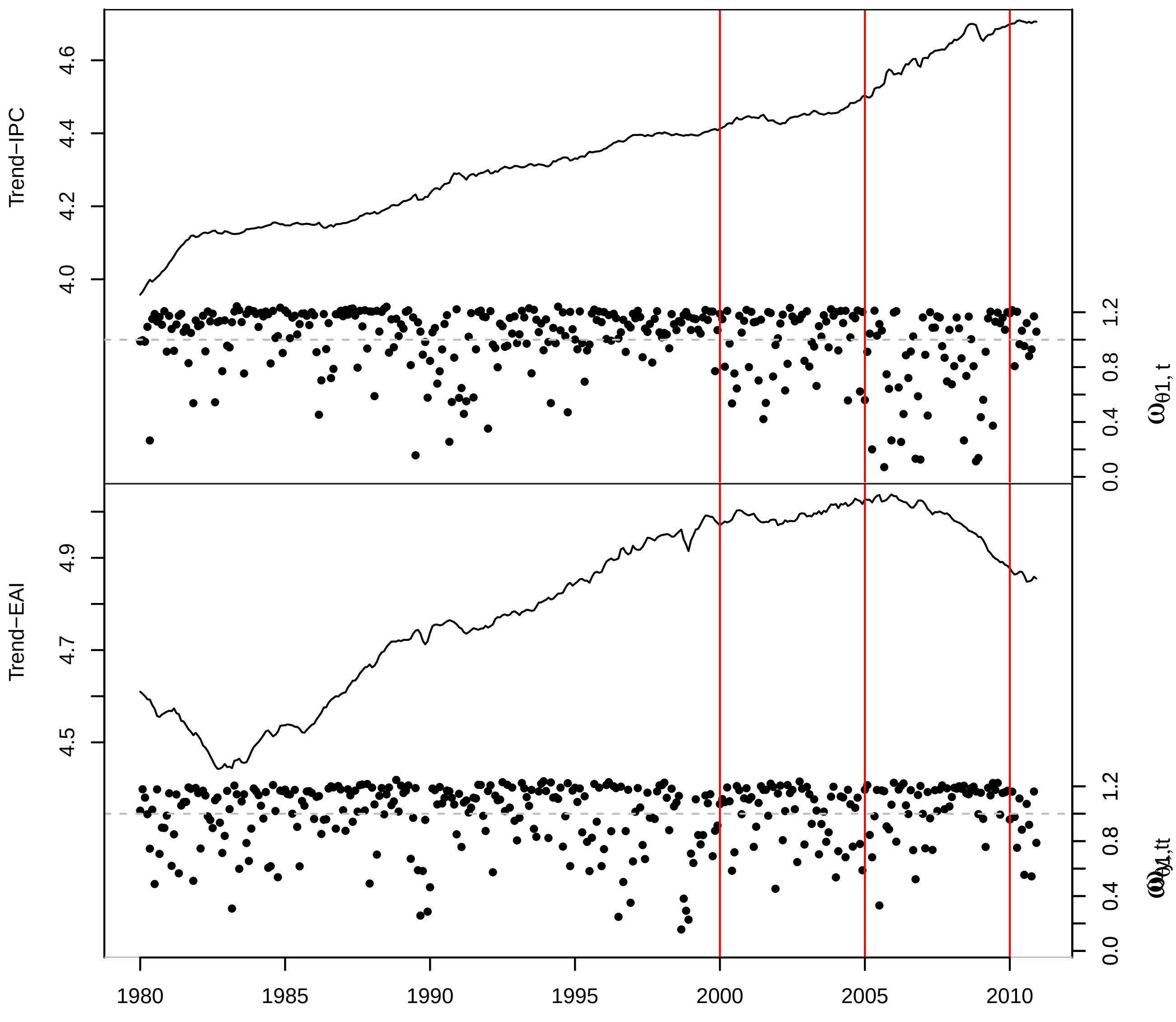

Figure 7: Comparison structural breaks for the monthly Consumer Price and Economic Indexes in Puerto Rico

4 Conclusions

In this paper we present a robust Bayesian dynamic model for the CPI and the EAI. We fitted a linear trend model with robust priors for the distributions of the state and observational precisions errors. The Bayesian dynamic robust model has the quality to detect outliers and historical structural breaks. The changes of the CPI trends are associated with those changes obtained for the EAI series. In fact, the structural changes in both series have a contextual historical and economical meaning. Some of these structural changes in both series coincide with the beginning of a recession period in the economy of Puerto Rico. Finally, the proposed model has the feature that it produces credible intervals that are not constant over time.

References

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

Department of labor and human resources (2008), *Study of revenue and expenditure of the urban consumer of Puerto Rico 1999-2003*, Puerto Rico.

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

Fúquene, J. A., Pérez, M. & Pericchi, L. R. (2013), 'An alternative to the inverted gamma for the variances to modelling outliers and structural breaks in dynamic models', *Brazilian Journal of Probability and Statistics* **In press**, –.

Government Development Bank for Puerto Rico (2012), *Economic Activity Index for the month of June 2012*, Economy, Puerto Rico.

URL: *http://www.gdb-pur.com/economy/documents/2012-Jun-GDB-EAIvFINAL-GS.pdf*

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

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

A Details for the Markov Chain Monte Carlo algorithm

This appendix presents the MCMC scheme we use in the paper.

- ## FFBS algorithm for the states.

  In order to obtain posterior inference on the state parameters θ_t in the model (5), we use the forward filtering backward sampling (FFBS) given in Fruwirth-Schnatter (1994) which is practically a simulation of the smoothing recursions. The FFBS works for the model (5) as follow:

  1. Use the Kalman Filter equations for (5). Let m_0 and C_0 (known) with $(\theta_0|D_0) \sim N(m_0, C_0)$ and

  $$\theta_t|y_{1:t-1} \sim N(m_{t-1}, C_{t-1}) \tag{6}$$

     - The one step predictive distribution of θ_t given $y_{1:t-1}$ is Gaussian $(\theta_t|D_{t-1}) \sim N(a_t, R_t)$ with parameters:

  $$a_t = G_t m_{t-1}; \qquad R_t = G_t C_{t-1} G_t'. \tag{7}$$

– The one step predictive distribution of yt given $y_{1:t-1}$ is Gaussian $(y_t|D_{t-1}) \sim N(f_t, Q_t)$ with parameters:

$$f_t = F_t^{'} \boldsymbol{a}_t; \qquad\qquad Q_t = F_t^{'} R_t F_t + V_t.$$

– The filtering distribution of θ_t given $y_{1:t-1}$ is Gaussian $(\theta_t|D_t) \sim N(m_t, C_t)$ with parameters:

$$m_t = a_t + A_t e_t; \qquad\qquad C_t = R_t - A_t Q_t A_t^{'} \tag{8}$$

where $A_t = R_t F_t Q_t^{-1}$, and $e_t = y_t - f_t$.

2. At time $t = T$ sample θ_T from $N(\theta_T|m_t, C_t)$.

3. For $t = (T-1) : 0$ sample θ_t from $N(\theta_t|m_t^*, C_t^*)$with

$$m_t^* = m_t + B_t(\theta_{t+1} - a_{t+1}) \qquad\qquad C_t^* = C_t - B_t R_{t+1} B_t^{'}$$

where $B_t = C_t G_{t+1}^{'} R_{t+1}^{-1}$.

- **Full conditionals for the observational and state variances in the DLM.**

$$\begin{aligned}
\lambda_y|... \sim &\text{Gamma}\left(q + \frac{T}{2}, \frac{1}{2} SSy^* + \beta\rho_y\right), & \lambda_{\theta,i}|... \sim &\text{Gamma}\left(q + \frac{T}{2}, \frac{1}{2} SS_{\theta,i}^* + \beta\rho_{\theta,t_i}\right) \\
\omega_{y,t}|... \sim &\text{Gamma}\left(\frac{\upsilon+1}{2}, \frac{\upsilon + \lambda_y(y_t - F_t\theta_t)^2}{2}\right) & \omega_{\theta,t_i}|... \sim &\text{Gamma}\left(\frac{\upsilon+1}{2}, \frac{\upsilon + \lambda_y(\theta_{t_i} - \lambda_{\theta,i}(G_t\theta_{t-1})_i)^2}{2}\right) \\
\rho_y|... \sim &\text{Gamma}\left(p + q, \beta\lambda_y + 1\right), & \rho_{\theta,t_i}|... \sim &\text{Gamma}\left(p + q, \beta\lambda_{\theta,i} + 1\right).
\end{aligned}$$

where $SSy^* = \sum_{t=1}^{T} \omega_{y,t}(y_t - F_t\theta_t)^2$ and $SS_{\theta,i}^* = \sum_{t=1}^{T} \omega_{\theta,t_i}(\theta_{t_i} - (G_t\theta_{t-1})_i)^2$ for $i = 1, 2$ and $t = 1, ...T$.

B Scaled Beta 2 density

A Beta2 random variable is equal in distribution to the ratio of two gamma-distributed random variables having shape parameters p and q and common scale parameter β:

$$\tau^2 \sim \text{Gamma}(p, \beta/\rho) \tag{9}$$

$$\rho \sim \text{Gamma}(q, 1) \tag{10}$$

where Gamma(a, b) denotes the Gamma distribution:

$$p(x|\alpha, b) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} \exp\{-x/\beta\} \quad a > 0, b > 0, \tag{11}$$

with β the scale parameter. The scaled Beta2 prior can be defined such as:

$$\pi(\tau^2) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{1}{\beta} \frac{\left(\frac{\tau^2}{\beta}\right)^{p-1}}{\left(1 + \frac{\tau^2}{\beta}\right)^{p+q}}; \quad \tau > 0. \tag{12}$$