CORE DISCUSSION PAPER

2015/XX

SPARSE CHANGE-POINT TIME SERIES MODELS

Arnaud Dufays[1] and Jeroen V.K. Rombouts[2]

July 10, 2015

Abstract

Change-point time series specifications constitute flexible models that capture unknown structural changes by allowing for switches in the model parameters. Nevertheless most models suffer from an over-parametrization issue since typically only one latent state variable drives the breaks in all parameters. This implies that all parameters have to change when a break happens. We introduce sparse change-point processes, a new approach for detecting which parameters change over time. We propose shrinkage prior distributions allowing to control model parsimony by limiting the number of parameters which evolve from one structural break to another. We also give clear rules with respect to the choice of the hyper parameters of the new prior distributions. Well-known applications are revisited to emphasize that many popular breaks are, in fact, due to a change in only a subset of the model parameters. It also turns out that sizeable forecasting improvements are made over recent change-point models.

Keywords: Time series, Shrinkage prior, Change-point model, Online forecasting

JEL Classification: C11, C15, C22, C51.

[1]Université catholique de Louvain, CORE, B-1348 Louvain-La-Neuve.
Website : https://sites.google.com/site/websiteofarnauddufays/
[2]ESSEC Business School, CREST

Arnaud Dufays is also an associate member of the Centre de Recherche en Economie et Statistique (CREST) and a Belgian FSR fellow. The scientific responsibility is assumed by the authors.

1 Introduction

It is now well known that time series observed over a long period are subject to structural breaks. Evidence can be found in Stock and Watson (1996) and in Bauwens, Koop, Koribilis, and Rombouts (2015) for macroeconomic series, and in Pastor and Stambaugh (2001) for financial series. Since break dates are unknown, models allowing for the possibility of changing their structure or parameters have been developed over the last two decades. Nevertheless, to date these models forecast generally not much better than simple techniques. Indeed, Bauwens, Koop, Koribilis, and Rombouts (2015) find no significant improvements in forecasting a wide variety of US macroeconomic time series when comparing cutting edge structural break models to simple forecasting techniques like rolling least squares. A potential explanation is that the existing specifications have the unattractive feature of forcing all parameters to change after a break creating models with an artificial high number of parameters. For example, a structural break model fitted to US GDP data implies a change in mean and variance due to the great moderation.

This paper proposes a new approach to model time series subject to multiple structural breaks. We extend the literature in two important directions. First, we relax the assumption that all parameters need to change when there is a structural break. This is done by shrinking irrelevant break-parameters towards zero in our estimation procedure which is based on a sampling scheme tailored to change-point (CP) models. In addition, our shrinkage methodology alleviates the search of an optimal number of regimes and does only rely on one estimation output. We propose two new shrinkage priors based on a finite mixture of respectively uniform and Gaussian distributions. In contrast to existing shrinkage priors which are typically used for high dimensional regression problems, our new priors satisfy desirable properties when working with structural break models.

The second main contribution of our paper is that we allow for rich and parsimonious within regime dynamics thanks to an autoregressive moving average (ARMA) type structure in the model. The current literature has at most an autoregressive (AR) structure in the conditional mean of each regime. In fact, ARMA type dynamics in the conditional mean or variance introduce more complexity in the estimation and model selection procedure because of the path dependence problem. Therefore, apart from the standard ARMA model, our approach also allows to estimate break dates and parameter changes in time varying variance and correlation models, specifications which are intensively employed in empirical economics and finance.

The classical reference that initiated much of the research on regime models is Hamilton (1989). He considers a linear AR model the parameters of which are allowed to switch according to a Markov-switching discrete process. The latent states are recurrent as the process can move from one state to any other state at any date. This approach being proposed for regime switches in a stationary time series has been adapted to the change-point or structural break specification. As in this paper, this specification has non-recurrent states so that the process can only stay in the same state or move to the next one. This implies a restricted probability transition matrix and is therefore a particular case of the classical switching model.

Inference on CP models has been introduced by Chib (1998). Apart from parameter estimation, he also tackles the computation of the marginal likelihood which is required to estimate the number of breaks. More recently, new specifications and inference have been proposed by Pesaran, Pettenuzzo, and Timmermann (2006), Koop and Potter (2007), Giordani and Kohn (2008), Maheu and Gordon (2008) and Geweke and Jiang (2011). These models adopt a hierarchical prior for regime coefficients, which allows for the coefficients of one regime to be informative about coefficients of other regimes. This permits to compute forecasts that contain information on the size and frequency of past breaks instead of neglecting observations prior to the most recent break-point.

While the recently proposed papers are quite general in their model specification, they have the feature that a break triggers abrupt changes in all parameters indexed by the latent state variable. This holds also for another standard procedure to deal with structural breaks. This procedure consists of first detecting the break dates by testing, see for example Andrews (1993) and Bai and Perron (1998), or by implementing an efficient LASSO algorithm as for example proposed by Chan, Yip, and Zhang (2014). Then given the detected break dates, independent locally stationary models typically from the same model class are fitted. When subsets of parameters are invariant over regimes, such an approach looses efficiency.

Using new shrinkage priors suited for CP configurations, we allow for changes in subsets of model parameters that are virtually zero, implying that only relevant parameters break. We state clear rules with respect to the choice of the hyper parameters of the prior distributions.

In fact, a novelty of our approach is that we can state explicitly how much in terms of log-likelihood it has to take for triggering a parameter break. Shrinkage priors, proposed by Mitchell and Beauchamp (1988) and George and McCulloch (1993) in discrete form, are popular in high-dimensional regressions to select relevant explanatory variables (e.g. Inoue and Kilian (2008) and Stock and Watson (2012) for time series applications). Continuous shrinkage priors like the spike and slab of Ishwaran and Rao (2005) or the Normal-Gamma of Griffin and Brown (2010) consist of mixtures of parametric distributions that exhibit a high density around zero in order to shrink non-relevant parameters to this value. Although extendible beyond linear regressions, see for example Scheipl, Fahrmeir, and Kneib (2012), we show that those standard shrinkage priors are not performing well when applied to CP models.

The existing literature considers models with at most an AR structure in the conditional mean of each regime. Adding a moving average part is a major obstacle for inference because of the path dependence problem. This occurs because the conditional mean at time t depends on the entire sequence of regimes visited up to time t. When computing the likelihood function one needs to integrate over all possible regime paths which grow exponentially with t, see Billio, Monfort, and Robert (1999) for an example of a two regime ARMA model. In this paper, we do inference and compute the marginal likelihood using a Metropolis-type MCMC sampler. Our flexible approach is based on an approximate model of the CP-ARMA process which is used to generate a candidate that is accepted/rejected according to the Metropolis-Hastings ratio. The inference is adapted from Dufays (2012) and Billio, Casarin, and Osuntuyi (2014) who develop an efficient MCMC sampler to infer CP and MS-GARCH models, see also Bauwens, Carpantier, and Dufays (2015). We further innovate by incorporating the MCMC scheme into a Sequential Monte Carlo sampler (see Del Moral, Doucet, and Jasra (2006)). In fact, as shown by Jasra, Stephens, and Holmes (2007) and Herbst and Schorfheide (2012), a MCMC sampler based on one single Markov-chain are less appropriate than SMC methods to explore distribution supports that are multi-modal, which is typically the case when dealing with shrinkage priors.

We apply our model to two macroeconomic time series. First, for the quarterly US GDP growth, rate we find the typical change point highlighting the great moderation. This induces a change in the variance of the time series but not in its mean. However, the global financial crisis changes both the dynamics in the mean and uncertainty in the variance. The second application fits the monthly US 3 Month Treasury Bill rate. We find evidence of eight change points all caused by changes in the variance of the time series rather than its mean. In terms of forecasting performance, it turns out that for the two applications our new sparse CP time series model performs particularly well both in terms of predictive likelihood as well as root mean squared error.

The rest of the paper is organized as follows. Section 2 introduces the desirable properties of shrinkage priors dedicated to CP settings and discusses popular shrinkage priors. We present our new priors specially designed for CP time series models, and discuss in detail how hyper parameters can be chosen. Section 3 defines the base-line model on which the Sparse CP is applied. We sketch the algorithms used for estimation. Two empirical applications are documented in Section 4. A forecasting exercise is provided in Section 5. Section 6 concludes. Appendices A and B give precise details on the implementation of the estimation and online forecasting algorithms.

2 Shrinkage priors for CP models

We provide new shrinkage priors tailored to time series CP models. In the applications below, we estimate models with several parameters in each regime. However, to simplify the exposition here, we focus only one of its parameters that we denote μ. A CP model always starts in the first regime with parameter μ_1. When a break happens, the parameter changes to μ_2 and the break size is given by $\Delta\mu_2 = \mu_2 - \mu_1$ and subsequent breaks are measured analogously by $\Delta\mu_3, \Delta\mu_4$, etc. For ease of notation, we work with $\Delta\mu$ in this section. Shrinkage means that a substantial part of the prior density on $\Delta\mu$ is concentrated around zero implying no structural break.

In the next section, we discuss desirable features of CP shrinkage priors that are not necessarily shared with existing shrinkage priors. In Section 2.2 we propose two new shrinkage priors suited for CP models. Section 2.3 compares in more detail with existing priors and Section 2.4 proposes some ways for choosing the hyper parameters.

2.1 Desirable features of CP shrinkage priors

Existing CP models assume that all the parameters of the model change when a structural break occurs. Although investigating which parameters are affected by a break seems a natural idea, this has rarely been done in practice. Considering all the switching possibilities for each number of regimes and determining the best specification following a criterion such as the marginal likelihood (e.g. Eo (2012)) is burdensome in many cases. This is because the possibilities grow exponentially with the number of regimes, and the marginal likelihood is sensitive to the prior distributions. For instance, considering our US Treasury bond illustration below, this approach would have required 4^{10} model estimations if we want to consider an upper bound of 10 regimes. Based on shrinkage priors, our method achieves the same goal in one estimation.

Two popular choices of continuous shrinkage priors are the spike-and-slab of Mitchell and Beauchamp (1988) and the Normal-Gamma of Griffin and Brown (2010) extending the double exponential prior of Park and Casella (2008), see also the scale mixture of normals in West (1987). The normal-Gamma prior is defined in hierarchical form as follows

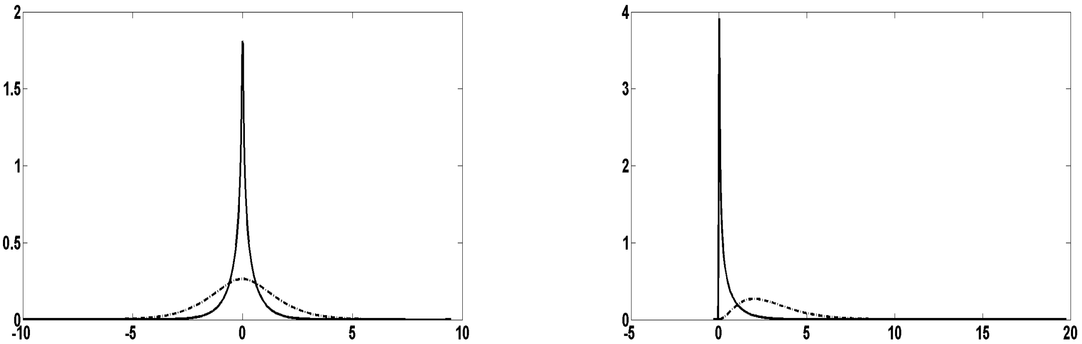
$$\begin{aligned} \Delta\mu|\Psi_\mu &\sim N(0, \Psi_\mu) \\ \Psi_\mu &\sim G \end{aligned}$$

where G is a Gamma distribution $G(\lambda, \frac{1}{2\nu^2})$. The marginal density of $\Delta\mu$ has variance and kurtosis respectively equal to $2\lambda\nu^2$ and $\frac{3}{\lambda}$. The hyper-parameters λ, ν^2 allow to generate sparsity when most of the density mass is close around zero as shown in the left panel Figure 1. The right panel shows the variance distribution Ψ_μ.

The spike and slab prior is hierarchically defined as

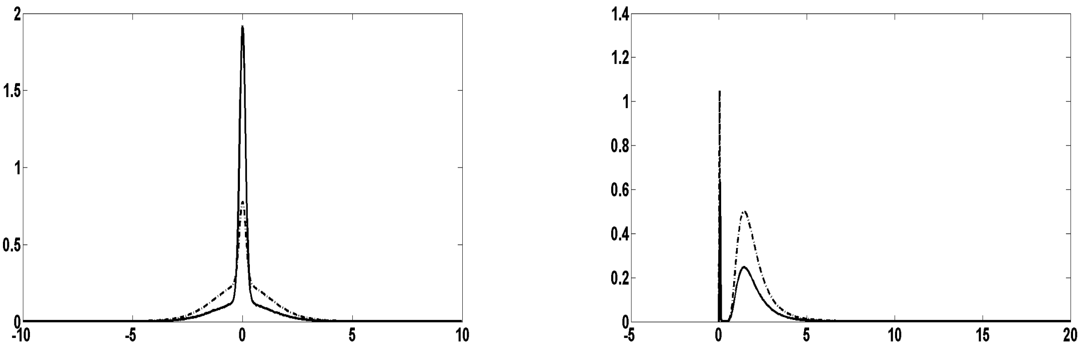
$$\begin{aligned} \Delta\mu|\Psi_\mu &\sim N(0, \Psi_\mu\sigma_\mu^2) \\ \Psi_\mu &\sim (1-\omega)\delta_{\Psi_\mu=c} + \omega\delta_{\Psi_\mu=1} \\ \sigma_\mu^2 &\sim IG(a, b) \end{aligned}$$

with mixing probability ω. The point mass c is chosen close to but different from zero with appropriate hyper-parameters to generate sparsity. Figure 2 presents an illustration of the marginal distribution as well as the distribution of the variance $(\Psi_\mu\sigma_\mu^2)$.



(a) Marginal distribution of $\Delta\mu$ (b) Variance distribution Ψ_μ

Figure 1: Normal-Gamma shrinkage prior. Solid line $\lambda = 0.5$, dotted line $\lambda = 3$ ($\nu = 1$)



(a) Marginal distribution of $\Delta\mu$ (b) Variance distribution $\Psi_\mu \sigma_\mu^2$

Figure 2: Spike and slab prior. Solid line $\omega = 0.4$, dotted line $\omega = 0.8$ ($c = 0.01, a = 6b = 10$)

The sparsity implied by those priors highly depends on the hyper-parameters and therefore some guidance have been proposed to select them, see Ishwaran and Rao (2005) and Griffin and Brown (2010). So far, these shrinkage distributions have typically been used to generate sparsity in high-dimensional regression models (see also Kalli and Griffin (2014) for an example for models time varying sparsity). However, we will argue that these priors are not suited for CP models. We state next three desirable properties of shrinkage priors in sparse CP configurations:

P1 The prior should be able to distinguish a clear change from a small variation in the model parameter.

P2 When there is a break, the prior should not distinguish between a small or a large break.

P3 The hyper-parameters of the prior distributions should depend on the user's likeness of having breaks and should be derived from a simple rule.

The first property is not shared by the existing shrinkage priors since they are not able to distinguish sharp changes from small deviations of the previous parameter value because the density functions are continuous and the mixture components are centered at zero. More details are given in Section 2.3. The second property can be validated by the existing priors as explained below. For the third desirable feature, we work with a penalty in terms of log-likelihood. For example, a penalty of -3 implies that any new regime has to improve the log-likelihood function of at least 3 to be detected. The smaller the penalty, the lesser the number of detected breaks and the more parsimonious is the model. To implement this simple rule, we choose a threshold value which defines a significant change in the parameter value. A deviation above this threshold is evidence of a new regime. The hyper-parameters of the prior distributions are derived such that the difference of the prior log-density evaluated at the threshold minus the prior log-density evaluated at zero is exactly equal to the user-defined penalty. Such an approach is inconvenient for the existing shrinkage priors exhibiting hierarchical layers.

2.2 Finite mixture shrinkage priors

We present two priors which comply with all the desirable conditions. They are both characterized by a finite mixture instead of the existing continuous mixtures priors defined above.

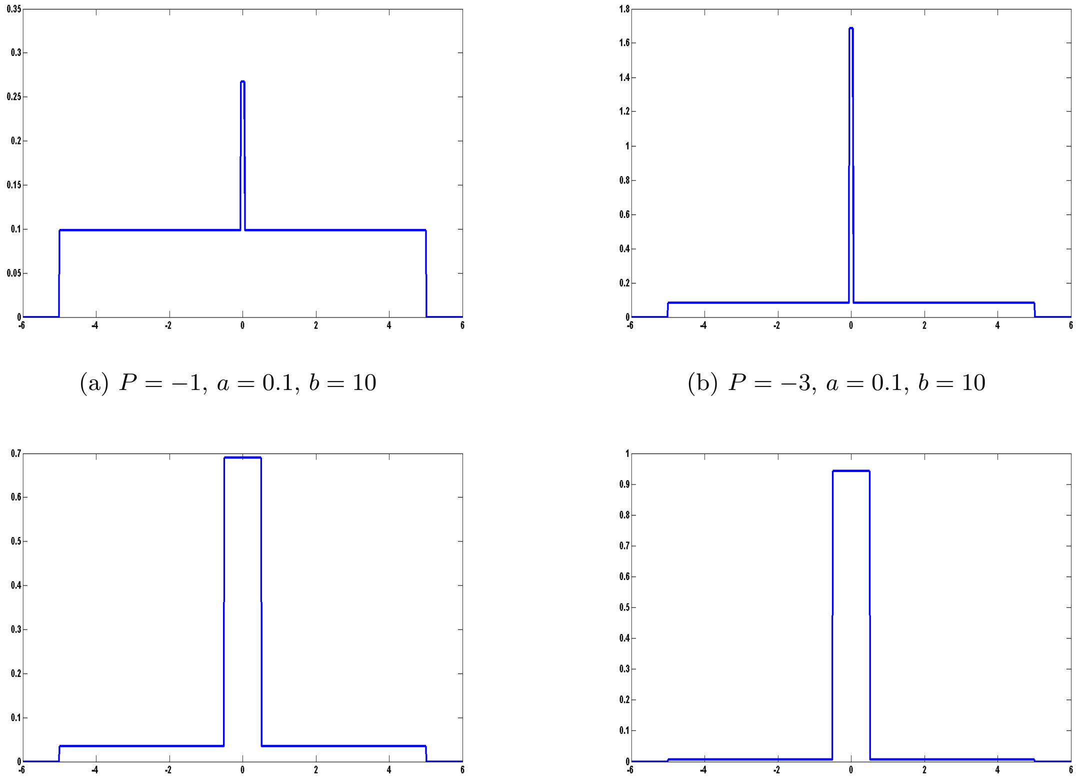
Mixture of two uniform distributions (2MU)

$$\begin{aligned} \Delta\mu &\sim \omega \;\; U[\frac{-a}{2},\frac{a}{2}] + (1-\omega) \;\; U[\frac{-b}{2},\frac{b}{2}] \\ f(\Delta\mu) &= \frac{\omega}{a}\delta_{\Delta\mu\in[\frac{-a}{2},\frac{a}{2}]} + \frac{1-\omega}{b}\delta_{\Delta\mu\in[\frac{-b}{2},\frac{b}{2}]} \end{aligned}$$

where ω is the mixture probability and the bounds a and b verify $a < b$. If we denote P the penalty on the log-likelihood function for detecting a new regime and by x, any point in the wider uniform component (i.e. $|x| > \frac{a}{2}$), then $\log f(x) - \log f(0) = P$ yields

$$\omega = \frac{a(1-e^P)}{be^P + a(1-e^P)} \tag{1}$$

which belongs to [0,1] for any positive values of a,b and any negative value of P. In practice, we define a very small and b very large. This is in line with the first desirable property. Furthermore, in accordance with the second property, a and b do not depend on P. From Equation (1), the distribution verifies the third property. In the following, the distribution is denoted by $2MU(a, b, P)$. Figure 3 displays examples of the density function for different values of P, a and b. Note also that when the penalty is set to 0, the 2MU reduces to a Uniform distribution with support $[-\frac{b}{2}, \frac{b}{2}]$ which is equivalent to the standard CP model with diffuse priors.



(a) $P = -1$, $a = 0.1$, $b = 10$

(b) $P = -3$, $a = 0.1$, $b = 10$

(c) $P = -3$, $a = 1$, $b = 10$

(d) $P = -5$, $a = 1$, $b = 10$

Figure 3: Mixture of two uniform distributions with different values of P, a and b.

Mixture of three Normal distributions (3MN)

The 2MU prior fulfills all the desirable properties but displays a compact support which might not always be adequate. We also propose a distribution with support on the real line based on a Normal mixture with three components specified by

$$\begin{aligned}\Delta\mu &\sim \frac{\omega}{2}\, N(-\bar{x},\bar{\sigma}^2)\delta_{\Delta\mu<-\bar{x}} + \frac{\omega}{2}\, N(\bar{x},\bar{\sigma}^2)\delta_{\Delta\mu>\bar{x}} + (1-\omega)\; N(0,\underline{\sigma}^2),\\ f(\Delta\mu) &= \frac{\omega}{\bar{\sigma}}\phi(\frac{|\Delta\mu|-\bar{x}}{\bar{\sigma}})\delta_{|\Delta\mu|\geq\bar{x}} + \frac{1-\omega}{\underline{\sigma}}\phi(\frac{\Delta\mu}{\underline{\sigma}}),\end{aligned}$$

where the parameter ω is the mixture probability, $\phi(\;)$ denotes the standard Normal density function and the parameters $\underline{\sigma}$ and $\bar{\sigma}$ define the standard deviations of the mixture components and verify the condition $\underline{\sigma} < \bar{\sigma}$. The interpretation of the variable P remains the minimum penalization on the log-likelihood function of adding a new regime and the value $\bar{x}$ stands for a threshold from which (in absolute value) the change in the parameter value is significant. Then from $\log f(\bar{x}) - \log f(0) = P$, the mixture probability ω is given by

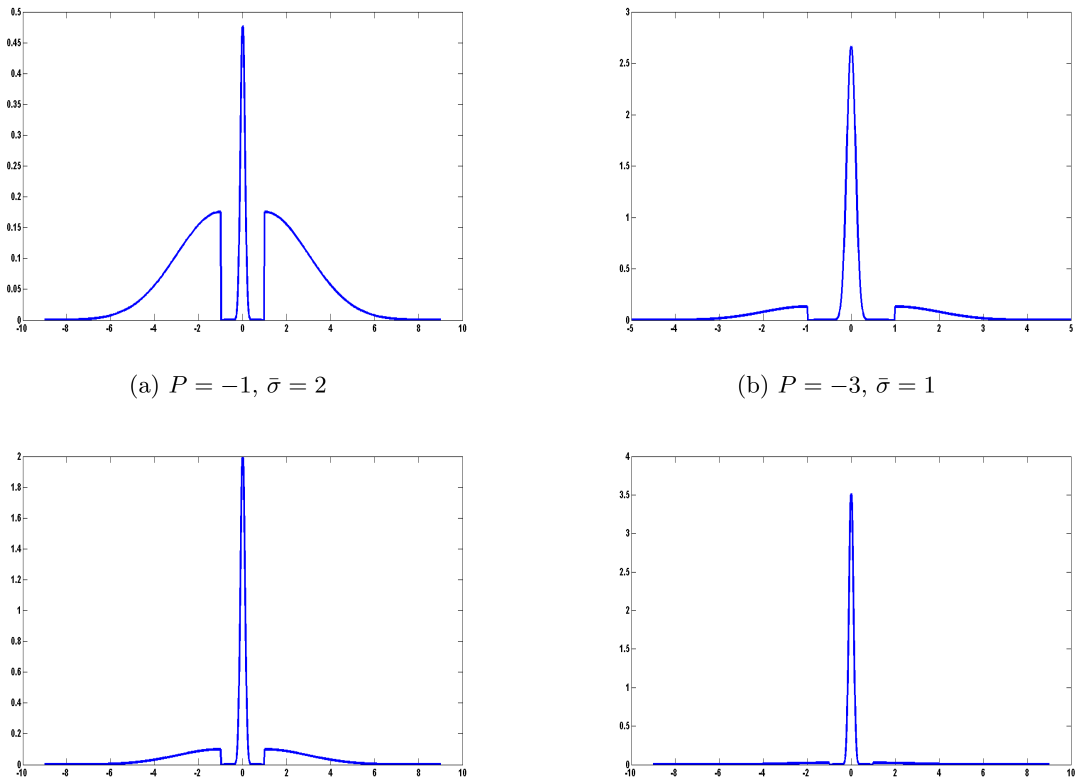
$$\omega = \frac{\bar{\sigma}(e^P - \frac{\phi(\bar{x}/\underline{\sigma})}{\phi(0)})}{\underline{\sigma} + \bar{\sigma}(e^P - \frac{\phi(\bar{x}/\underline{\sigma})}{\phi(0)})}, \tag{2}$$

implying $\omega \in [0,1]$ if $P \geq \log\phi(\bar{x}/\underline{\sigma}) - \log\phi(0)$. Note that this condition does not depend on $\bar{\sigma}$. In the empirical section, $\underline{\sigma}$ is fixed to $0.1\bar{x}$ implying that P should be larger than -50 which is a sufficiently high penalization for our purpose. The distribution is denoted by $3MN(\bar{x},\underline{\sigma},\bar{\sigma},P)$ hereafter. Note that in contrast to the 2MU prior, when $P = 0$, the 3MN prior does not collapse to a diffuse Normal prior.

The mixture of Normal distributions prior displays the three qualities of a CP shrinkage prior. Indeed, the first property is fullfilled because the Normal components are truncated at $\bar{x}$. Regarding the second desirable feature, the arbitrary large density mass at the left of $|\bar{x}|$ is almost uniform for a wide range of values and for any penalty P. When a break is detected, a large deviation from the previous parameter value is therefore not more penalized than a small one. Finally, given equation (2), the penalty value determines the weights of the components which verifies the last property. Figure 4 illustrates the prior distribution for different values of the parameters.

2.3 Why not using the existing shrinkage priors?

The 2MU and 3MN priors being introduced, we can further motivate why the existing priors are inappropriate for CP models. As emphasized earlier, the hierarchy implied by the Normal-Gamma and the spike and slab priors make their dependence on the penalty parameter difficult to interpret. However, for the particular case of the spike and slab prior with $\sigma_\mu^2 \sim$



(a) $P = -1$, $\bar{\sigma} = 2$

(b) $P = -3$, $\bar{\sigma} = 1$

(c) $P = -3$, $\bar{\sigma} = 2$

(d) $P = -5$, $\bar{\sigma} = 2$

Figure 4: Mixture of three Normal distributions with different values of penalization and $\bar{\sigma}$. The other hyper-parameters are fixed to $\bar{x} = 1$ and $\underline{\sigma} = 0.1\bar{x}$.

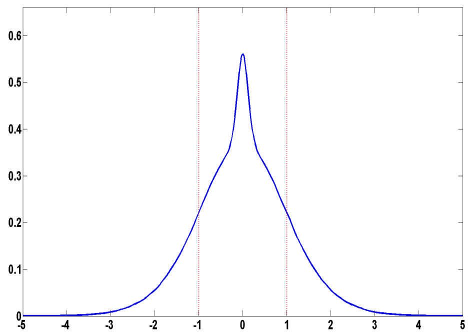
$IG(\frac{v}{2}, \frac{v}{2})$ the dependence on the penalty parameter can be made explicit. Indeed, under this specification, the marginal density of $\Delta\mu$ is given by

$$\Delta\mu \quad \sim \quad (1-\omega)X\sqrt{c} + \omega X$$

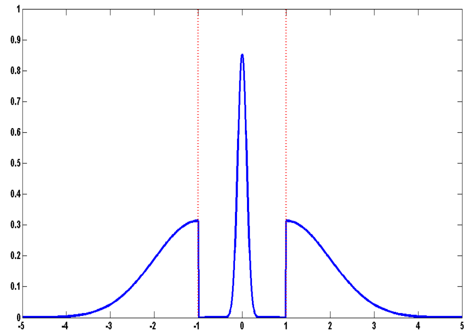
where $X \sim t_v$ denotes a random variable following a Student distribution with v degrees of freedom and density denoted by $f_t(\ |v)$. To set the weight according to a chosen penalty P, one can solve the equation $\log f(\bar{x}) - \log f(0) = P$ which leads to

$$\omega \quad = \quad \frac{e^P f_t(0|v) - f_t(\bar{x}/\sqrt{c}|v)}{f_t(0|v)e^P(1-\sqrt{c}) + \sqrt{c} f_t(\bar{x}|v) - f_t(\bar{x}/\sqrt{c}|v)}. \tag{3}$$

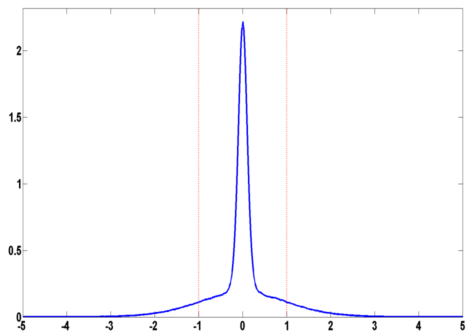
Nevertheless, the relation (3) only verifies property P3. Instead, property P1 implies a discontinuous density function which is not satisfied by the spike and slab prior. More specifically, the density function of the existing continuous shrinkage priors evaluated at any point between the threshold value and zero may drastically depend on the user's break sensitivity. Figure 5 makes clear that different penalty values affect the density function between zero and the threshold value for the spike and slab prior whereas they do not impact the 3MN priors. Consequently, the spike and slab prior implies that the model parameters are allowed to smoothly change from one regime to another for some penalties. In such a case, the detection of abrupt switches in their values is more delicate.



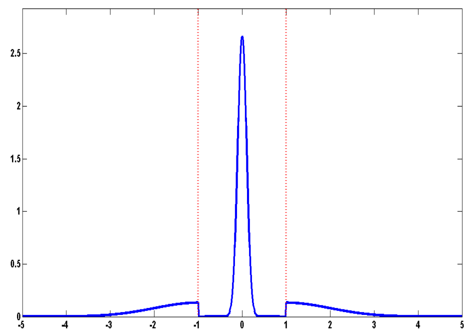
(a) Spike and slab prior, $P = -1$



(b) 3MN prior, $P = -1$



(c) Spike and slab prior, $P = -3$



(d) 3MN prior, $P = -3$

Figure 5: Mixture of two Student distributions centered at zero with $\bar{\sigma} = 1$, $v = 50$, $c = 0.1$ and different values of penalizations on the left. 3MN priors with $\bar{\sigma} = 1, \underline{\sigma} = 0.1$ and different values of penalizations on the right. The threshold value is fixed to $\bar{x} = 1$.

2.4 Choosing the threshold and the penalty values

So far, we have developed two prior distributions for shrinking non-relevant parameters toward zero in the CP context. However, these distributions depend strongly on the penalty P

and on the threshold value $\bar{x}$ in the case of the 3MN prior. We propose next some ways for choosing these parameters.

2.4.1 The penalty value P

Consider a model M_1 without abrupt switches in the parameters and M_2 the same model but with one break in one of its parameters. The latter model therefore contains at least two additional parameters, i.e. the parameter modifying the process dynamics and the one controlling when this change occurred. As suggested by Kass and Raftery (1995), model M_2 is strongly preferred when its marginal probability $p(M_2|y_{1:T})$ is higher or equal to 95%. Ideally, the penalty value P should be chosen such that a break is detected only if it implies such a marginal posterior probability. Assuming equal model prior probabilities, the ratio of the posterior odds simplifies into the ratio of marginal likelihoods, also known as Bayes factor. Instead of using the marginal likelihood, we use as approximation the easier to compute Bayesian information criterion (BIC) to obtain a rule to select P. More precisely, we have that

$$\begin{aligned}
\frac{P(M_2|y_{1:T})}{P(M_1|y_{1:T})} &\geq \frac{0.95}{1-0.95} \\
\approx e^{-0.5(BIC_{M_2}-BIC_{M_1})} &\geq \frac{0.95}{1-0.95} \\
\log f(y_{1:T}|\{\hat{\Theta},\hat{\Sigma}\}_{M_2}) - \log f(y_{1:T}|\{\hat{\Theta},\hat{\Sigma}\}_{M_1}) - 0.5(k_{M_2}-k_{M_1})\log T &\geq \log\frac{0.95}{1-0.95} \\
\log f(y_{1:T}|\{\hat{\Theta},\hat{\Sigma}\}_{M_2}) - \log f(y_{1:T}|\{\hat{\Theta},\hat{\Sigma}\}_{M_1}) &\geq \log\frac{0.95}{1-0.95} + \log T
\end{aligned}$$

where $y_{1:T} = \{y_1, ..., y_T\}$ is a time series of T observations, k_{M_i} stands for the number of parameters in model i and $\hat{\Theta}$ and $\hat{\Sigma}$ denote the maximum likelihood estimates of the corresponding model. In fact, the maximum log likelihood of the model exhibiting a break in one parameter should be at least higher than $\log\frac{0.95}{1-0.95} + \log T$ to have a posterior probability higher than 95% if the BIC approximation holds. A natural value for the penalty parameter is therefore $P = -(\log\frac{0.95}{1-0.95} + \log T)$.

The previous rule to select P assumes that the BIC provides a good approximation of the marginal likelihood. Because CP models are non-stationary by nature this assumption might be questionable. Hence, an alternative strategy would consist of estimating P by considering it as a random variable with a non-informative Uniform prior. In such a case, the shape of the posterior distribution of P is illustrated in the following example.

Example

We compute the posterior distribution for P under the following conditions: (i) only one parameter $\Delta\mu$ can be shrunk, (ii) the shrinkage prior is 2MU with upper bound b ten times the lower bound a, (iii) the prior distribution of P is Uniform on $[u_1,u_2]$ ($u_1 < u_2 \leq 0$). Under these conditions, we can easily compute the posterior distribution of P in the case when $|\Delta\mu| < \frac{a}{2}$ (no break) and when $|\Delta\mu| \geq \frac{a}{2}$ (break).

- When no break exists ($|\Delta\mu| < \frac{a}{2}$), the posterior density of P is given by

$$\begin{aligned} f(P|\ |\Delta\mu| < \frac{a}{2}) &= \frac{1}{(1+9e^P)(u_2 - u_1 + \log(9e^{u_1}+1) - \log(9e^{u_2}+1))}\delta_{P\in[u_1,u_2]} \\ &\approx \frac{1}{(1+9e^P)(|u_1| - \log(10))}\delta_{P\in[u_1,u_2]} \text{ if } u_1 << 0 \text{ and } u_2 = 0. \end{aligned}$$

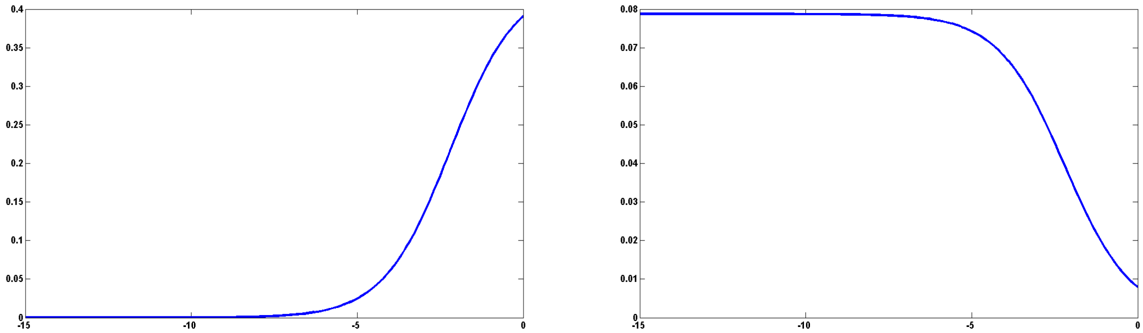
- When a break exists ($|\Delta\mu| > \frac{a}{2}$), the posterior density of P is given by

$$\begin{aligned} f(P|\ |\Delta\mu| > \frac{a}{2}) &= \frac{9e^P}{(1+9e^P)(\log(e^{u_1}+\frac{1}{9}) - \log(e^{u_2}+\frac{1}{9}))}\delta_{P\in[u_1,u_2]} \\ &\approx 3.8989\frac{e^P}{1+9e^P}\delta_{P\in[u_1,u_2]} \text{ if } u_1 \to -\infty \text{ and } u_2 = 0. \end{aligned}$$

Figure 6 displays both posterior densities when $u_1 = -15$ and $u_2 = 0$. To ensure that the break is detected, the log-likelihood has to increase by 5 since P has most density mass in the [-5,0] range. On the other hand, when there is no break in the parameter, the posterior has the undesirable feature that P is still likely to be in the [-5,0] range. Moreover, the problem becomes even more severe when u_1 increases.

Since a Uniform prior on P implies an undesirable posterior density, in our empirical applications, we use an informative prior centered at the BIC (i.e. $P \sim N(P_{BIC}, 0.5)$). By so-doing, the uncertainty on the parameter P is reflected in the posterior distribution of the model parameters whereas its fluctuation is under control.

Finally, we emphasize that one cannot choose the same parameter P for all the parameters of the model. To illustrate the issue, Figure 7 displays posterior densities of P for a wide range



(a) Break in the parameter

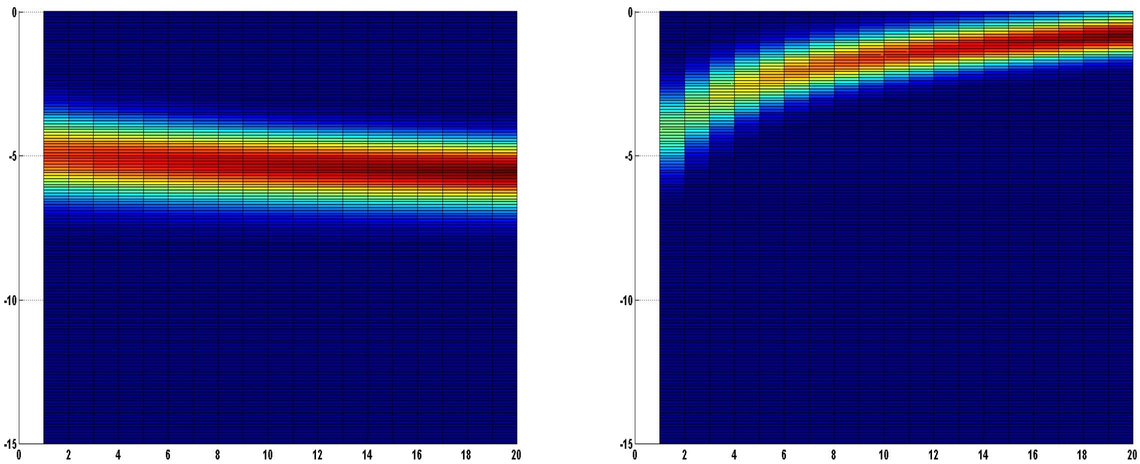
(b) No break in the parameter

Figure 6: Posterior densities of P under two different cases of a break in one model parameter.

of model parameters that either breaks or not. The upper left panel assumes an increasing number of parameters (from 1 to 20) that is not significantly different from zero. We see that whatever the number of parameters, the posterior density of P remains more or less identical and centered around the BIC value. The right panel of Figure 7 shows posterior densities of P for an increasing number of parameters which are all deviating significantly from zero. We observe that the distributions shift gradually towards zero, i.e. the required change in the log-likelihood to detect a new break decreases. This feature is not desirable as we loose the interpretation of the penalty parameter. Moreover, for the 3MN shrinkage prior, the implied penalty for an increasing number of breaks will also depend on the threshold $\bar{x}$. By including one penalty variable per model parameter, we avoid these issues as the marginal posterior distribution of each penalty is less sensitive to breaks in the parameters. This is a common solution in the literature on the standard shrinkage priors including the normal-gamma and the discrete or continuous spike and slab priors. Note that in the application below, we also estimate the models with nonrandom P and find only minor differences.

2.4.2 The threshold value in the 3MN prior

The threshold value $\bar{x}$ is the required deviation of a parameter from its previous value to be considered as a new regime. Above this value, the parameter is no longer shrunk toward zero and a new regime is allowed. Since the scale of the model parameters as well as their standard deviations are data dependent, a different threshold value is used per parameter. For instance, in the case of an ARMA(1,1) process, four thresholds values are needed. Choosing



No break in the parameters

Break in the parameters

Figure 7: Posterior densities of P when the prior is a Normal distribution with $P_{BIC} = -5$ and with the shrinkage distribution $2MU(0.2, 2, P)$. The horizontal axis indicates the number of parameters in the model. The vertical axis indicates the $[-15, 0]$ support of P. The upper left panel assumes an increasing number of parameters (from 1 to 20) that is insignificantly different from zero. The right panel shows the densities for an increasing number of parameters that deviating from zero.

an appropriate value requires knowledge of the possible variation of the model parameters within a regime. To this purpose, we first perform inference on a model without a break. Then we use the parameter draws to define the threshold value by the difference between the median and the 2.5%-th quantile. Therefore, if a new regime appears, the value of a model parameter in this regime will be more or less significantly different from its previous value at 5% level (similar to a two-sided test). Other quantiles are tested in the application to assess the sensitivity with respect to this parameter.

3 Sparse change-point models

3.1 A baseline change-point model

Let $y_{1:T} = \{y_1, ..., y_T\}$ be a time series of T observations and let $s_{1:T} = \{s_1, ..., s_T\}$ denote discrete random variables taking values in $[1, K+1]$. Our baseline CP model with $K+1$ regimes is defined (for $t > 1$) by

$$y_t \;=\; \mu_{s_t} + \beta_{s_t} y_{t-1} + \phi_{s_t}\varepsilon_{t-1} + \epsilon_t \;\text{ with }\; \varepsilon_t \;\sim\; N(0, \sigma^2_{s_t}) \;\text{ for }\; s_t = 1, \ldots, K+1.$$

A more involved conditional mean or non Gaussian innovations (ε_t) can be introduced. Also, the variance in each regime is kept constant, though time varying variance models like GARCH specifications can be readily incorporated. The structural breaks or change-points are modeled by discrete variables $s_{1:T}$ driven by a Markov-chain with $(K+1)\times(K+1)$ probability transition matrix P given by

$$P = \begin{pmatrix} p_1 & 1-p_1 & 0 & ... & 0 \\ 0 & p_2 & 1-p_2 & ... & 0 \\ & & ... & & \\ 0 & 0 & ... & 0 & 1 \end{pmatrix}.$$

In order to apply our new shrinkage priors, we write the CP model relative to the first regime as follows

$$\begin{aligned} y_t \;&=\; (\mu_1 + \sum_{i=2}^{s_t} \Delta\mu_i) + (\beta_1 + \sum_{i=2}^{s_t} \Delta\beta_i) y_{t-1} + (\phi_1 + \sum_{i=2}^{s_t} \Delta\phi_i)\varepsilon_{t-1} \\ &\quad +\varepsilon_t \text{ with } \varepsilon_t \;\sim\; N(0, \sigma_1^2 + \sum_{i=2}^{s_t} \Delta\sigma_i^2) \;\text{ for }\; s_t = 2, \ldots, K+1, \end{aligned} \tag{4}$$

where the operator Δ stands for the difference of two consecutive regime parameters, e.g. $\Delta\mu_2 = \mu_2 - \mu_1$. Therefore, the parameters in levels can be simply obtained by $\mu_{s_t} = (\mu_1 + \sum_{i=2}^{s_t} \Delta\mu_i)$ for $s_t = 2, \ldots, K+1$ and similarly for the other parameters. Working with differences of consecutive regime parameters allows applying the shrinkage priors that encourage sparsity in our specification. To alleviate the notation in the next section, we define the continuous mean parameter set as $\Theta = (\mu_1, \beta_1, \phi_1, \Delta\mu_2, \Delta\beta_2, \Delta\phi_2, ..., \Delta\mu_{K+1}, \Delta\beta_{K+1}, \Delta\phi_{K+1})$, the variance parameter set as $\Sigma = \{\sigma_1^2, \Delta\sigma_2^2, ..., \Delta\sigma_{K+1}^2\}$ and $p_{1:K} = \{p_1, ..., p_K\}$, the set of transition probabilities.

3.2 Estimation of the Sparse CP ARMA model

For a given number of breaks K, inference is done by drawing sequentially from the following posterior distributions $\pi(\Theta|y_{1:T}, s_{1:T}, p_{1:K}, \Sigma)$,$\pi(\Sigma|y_{1:T}, s_{1:T}, p_{1:K}, \Theta)$, $\pi(p_{1:K}|y_{1:T}, s_{1:T}, \Theta, \Sigma)$ and $\pi(s_{1:T}|y_{1:T}, p_{1:K}, \Theta, \Sigma)$. Chib (1998) develops efficient Bayesian inference of CP models that relies on the forward-backward algorithm, see also Rabiner (1989). This latter algorithm is applicable for models without path dependence since it assumes that the likelihood of an observation y_t given the model parameters and the *current* state can directly be evaluated. The baseline CP model in (4) does not belong to this class since the likelihood at time t requires the computation of the current error term which, in turn, depends on the states *prior to* the current one. To solve the problem, Bauwens, Dufays, and Rombouts (2013) substitute the forward-backward algorithm by a Sequential Monte Carlo (SMC) algorithm. Since such an approach is time-consuming, we follow Dufays (2012) by adopting a Metropolis-Hastings method.

Regarding the unknown number of breaks K, our approach is to estimate a CP model exhibiting a high number of regimes and shrink the parameters of non-relevant regimes to zero. By so-doing, we save computational resources and coding efforts since only one estimation is required and computation of the marginal likelihood is avoided. In fact, the standard procedure is to maximize the marginal likelihood by estimating CP models with different numbers of regimes.

Algorithm 1 documents the MCMC scheme that is applied to infer the Sparse CP ARMA model. More details are given in Appendix A. The scheme exposed in Algorithm 1 is theoretically sufficient to draw from the posterior distribution. Nevertheless, the complexity of our model conjugated with the shrinkage priors that typically lead to multi-modal posterior distributions and the Metropolis-Hastings updates used in the MCMC approach make that the algorithm is bound to badly mix. As a result, the realizations of the Markov chain could be trapped in a single mode of the posterior distribution or simply not converge to the targeted distribution in a reasonable amount of time.

To solve these issues, we simulate the posterior distribution of the parameters by combining Markov-Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) methods, a technique called SMC sampler (see Del Moral, Doucet, and Jasra (2006)). This method exhibits several advantages compared to the standard MCMC approach. First, as the realizations of the SMC (termed 'particles', hereafter) evolve independently, the algorithm can be easily parallelized (e.g. Durham and Geweke (2012)). Secondly, as shown by Jasra, Stephens, and Holmes (2007) or Herbst and Schorfheide (2012), MCMC samplers based on one single Markov-chain usually mix slower than SMC methods to simulate multi-modal distributions. Furthermore, the SMC sampler delivers an estimate of the marginal likelihood while additional computations are required with MCMC algorithms. The sampler is also less sensitive to the initial values and rules out the difficult choice of the burn-in sample size. A final advantage of the SMC sampler is that it updates its particles in the light of new observations without requiring a new estimation as the MCMC would, a feature that refers to the *on-line* property of the SMC method with respect to the *off-line* MCMC approach.

Naturally, the SMC sampler also comes with its own disadvantages, the most important being the number of user-specified parameters to tune in order to run the algorithm. Therefore, inference is carried out by the Time and Tempered (TNT) algorithm (see Dufays (2014)), i.e. a variant of the Sequential Monte Carlo sampler (Del Moral, Doucet, and Jasra (2006)), which automates the choice of the SMC parameters. The algorithm sequentially iterates by producing realizations from a prior distribution without shrinkage (denoted $\bar{f}$ in the algorithm 2) to the posterior distribution by combining many importance sampling and MCMCs. The auxiliary distribution $\bar{f}$ is important as it allows to sequentially introduce the shrinkage feature (and thus the multimodality of the posterior distribution). The SMC sketch, when the number of observations is fixed, is detailed in Algorithm 2. More information is given in Appendix B which details how to estimate the model with a newly arriving of time series observations. The Matlab code is available on Arnaud Dufays' website.

Algorithm 1 MCMC scheme

Set initial values to $s_{1:T}, p_{1:K}, \Sigma, \Theta$

for $i = 1$ to N where N denotes the number of MCMC iterations **do**

[1] - Sample $\Theta \sim f(\Theta|y_{1:T}, s_{1:T}, p_{1:K}, \Sigma)$ by the DREAM algorithm.

[2] - Sample $\Sigma \sim f(\Sigma|y_{1:T}, s_{1:T}, p_{1:K}, \Theta)$ by the DREAM algorithm.

[3] - Sample $p_{1:K} \sim \prod_{i=1}^{K} Beta(\alpha_p + \sum_{t=1}^{T} \delta_{s_t=i}, \beta_p + 1)$ with α_p and β_p hyper-parameters.

[4] - Sample $s_{1:T} \sim f(s_{1:T}|y_{1:T}, p_{1:K}, \Theta, \Sigma)$ by the Metropolis-Hastings algorithm (Appendix A)

end for

3.3 Prior elicitation

We display the prior distributions of our CP-ARMA parameters in Table 1. Besides the ARMA parameters that are driven by one of the two new shrinkage priors, the other parameters follow standard distributions. In particular, the transition probabilities and the state vector are specified as in Chib (1998) and Pesaran, Pettenuzzo, and Timmermann (2006). The shrinkage priors are characterized by the penalty parameter P and a fixed threshold $\bar{x}$. The penalty parameter is taken as random and driven by a Normal distribution centered at the penalty value implied by the BIC, see the discussion in Section 2.4. While in principle no threshold is needed to specify the 2MU distribution, the support a of the inner Uniform component a is made dependent on the threshold $\bar{x}$ to avoid numerical problems that could occur if the bound a was chosen too small to be handled by the computer. Finally, the maximum number of breaks K is chosen to be large (in the application equal to five for quarterly macroeconomic time series). In fact, this number is not relevant as long as it is above the observed number of breaks in the time series.

4 Applications

In this section, we present two applications of our approach, namely the quarterly US GDP growth rate and the monthly US Treasury Bill rate.

Algorithm 2 SMC algorithm with fixed number of observation (off-line)

Sample M particles from the prior distribution without shrinkage : $\{\Theta^i\}_{i=1}^M, \{\Sigma^i\}_{i=1}^M$

Sample M particles from the prior distributions : $\{p^i_{1:K}\}_{i=1}^M, \{s^i_{1:T}\}_{i=1}^M$

Set the tempered function $\phi = 0$, the normalized weights $W_i = 1/M \ \forall i \in [1, M]$ and $ESS = M$

while $\phi < 1$ **do**

A - Correction step :

Find $\tilde{\phi} > \phi$ such that $M/(\sum_{i=1}^N \tilde{w}_i^2) = 0.95ESS$

where $\tilde{w}_i \propto W_i[f(y_{1:T}|s^i_{1:T}, p^i_{1:K}, \Theta^i, \Sigma^i)f(\Theta^i, \Sigma^i)]^{\tilde{\phi}-\phi}\bar{f}(\Theta^i, \Sigma^i)^{\phi-\tilde{\phi}}$

$\forall i \in [1, M]$; Set $w_i = W_i[f(y_{1:T}|s^i_{1:T}, p^i_{1:K}, \Theta^i, \Sigma^i)f(\Theta^i, \Sigma^i)]^{\tilde{\phi}-\phi}\bar{f}(\Theta^i, \Sigma^i)^{\phi-\tilde{\phi}}$

$\forall i \in [1, M]$; Compute the normalized weights $W_i = w_i / \sum_{j=1}^N w_j$ and $ESS = M/(\sum_{i=1}^N W_i^2)$

Set $\phi = \min(\tilde{\phi}, 1)$

B - Re-sample step if $ESS < 0.75M$

Re-sample the particles by stratified sampling (Carpenter, Clifford, and Fearnhead (1999))

C - MCMC step with targeted distribution :

$f_\phi(s_{1:T}, p_{1:K}, \Theta, \Sigma|y_{1:T}) \propto [f(y_{1:T}|s_{1:T}, p_{1:K}, \Theta, \Sigma)f(\Theta, \Sigma)]^\phi \bar{f}(\Theta, \Sigma)^{1-\phi} f(s_{1:T}, p_{1:K}, \Theta, \Sigma)$

for $i = 1$ to M **do**

for $j = 1$ to N **do**

[1] - Sample $\Theta^i \sim f_\phi(\Theta|y_{1:T}, s^i_{1:T}, p^i_{1:K}, \Sigma^i)$ by the DREAM algorithm.

[2] - Sample $\Sigma^i \sim f_\phi(\Sigma|y_{1:T}, s^i_{1:T}, p^i_{1:K}, \Theta^i)$ by the DREAM algorithm.

[3] - Sample $p^i_{1:K} \sim \prod_{r=1}^K Beta(\alpha_p + \sum_{t=1}^T \delta_{s^i_t = r}, \beta_p + 1)$ with α_p and β_p hyper-parameters.

[4] - Sample $s^i_{1:T} \sim f_\phi(s_{1:T}|y_{1:T}, p^i_{1:K}, \Theta^i, \Sigma^i)$ by the Metropolis-Hastings algorithm

end for

end for

end while

Table 1: Prior distributions and hyper-parameters.

Prior Distributions of the mean parameters

$$\{\mu_1, \beta_1, \phi_1\}' \sim N(\begin{pmatrix}0 & 0 & 0\end{pmatrix}', I_3)$$

2MU: For each regime $k \in [2, K+1]$, $\{\Delta\mu_k, \Delta\beta_k, \Delta\phi_k\}' \sim \prod_{j=1}^{3} 2MU(\bar{x}_j, 5, P_{\text{Mean}}^{k,j})$

3MN: For each regime $k \in [2, K+1]$, $\{\Delta\mu_k, \Delta\beta_k, \Delta\phi_k\}' \sim \prod_{j=1}^{3} 3MN(\bar{x}_j, 0.1\bar{x}_j, 0.5, P_{\text{Mean}}^{k,j})$

Prior Distributions of the variances

$$\sigma_1^{-2} \sim \text{G(1,1)}$$

2MU: For each regime $k \in [2, K+1]$, $\Delta\sigma_k^2 \sim 2MU(\bar{x}_\sigma, 10, P_\sigma^k)1_{\sigma_k^2>0}$

3MN: For each regime $k \in [2, K+1]$, $\Delta\sigma_k^2 \sim 3MN(\bar{x}_\sigma, 0.1\bar{x}_\sigma, 2, P_\sigma^k)1_{\sigma_k^2>0}$

Prior Distribution of transition probabilities

$$p_{1:K} \sim \prod_{i=1}^{K} \text{Beta}(100, 1)$$

Prior Distribution of the latent state vector

$s_1 = 1$, for $t > 1$, $s_t = s_{t-1} | s_{t-1}, p_{1:K} \sim p_{s_{t-1}}$ and $s_t = s_{t-1} + 1$ otherwise

Prior Distribution of the Penalty parameters

For each parameter $j \in [1, 3]$ in each regime $k \in [2, K+1]$ $P_{\text{Mean}}^{k,j} \sim N(P_{BIC}, 0.5)$
For each regime $k \in [2, K+1]$ $P_\sigma^k \sim N(P_{BIC}, 0.5)$

The d-dimensional identity matrix is denoted by I_d, $(\bar{x}_1, \bar{x}_2, \bar{x}_3) = (\bar{x}_\mu, \bar{x}_\beta, \bar{x}_\phi)$. The uniform mixture and Normal shrinkage priors are respectively denoted by $2MU()$ and $3MN()$. The Gamma distribution is written $G()$ and $Geo()$ means the Geometric distribution.

4.1 US GDP growth rate

We study quarterly US GDP growth from 1959Q2 to 2011Q3 (210 observations). Researchers have highlighted the presence of one structural break in the mid-1980s followed by the 'great moderation' era. The volatility substantially drops during this episode while GDP growth remains unchanged. As can been seen from Table 2, fitting a standard ARMA model over the full sample gives an autoregressive parameter of 0.46, a non-significant MA parameter, and a variance of 0.70.

We first estimate a CP-ARMA model without shrinkage prior, meaning that all parameters

Table 2: GDP growth: preferred CP-ARMA model with 3 regimes

	Sparse CP-ARMA-2MU			**Sparse CP-ARMA-3MN**		
	Regime 1	Regime 2	Regime 3	Regime 1	Regime 2	Regime 3
Dates	→1983Q2	→2007Q3	→ 2011Q3	→ 1983Q2	→ 2007Q4	→ 2011Q3
μ_k	0.26	no-change	no-change	0.28	no-change	no-change
	(0.10)			(0.09)		
β_k	0.59	no-change	no-change	0.60	no-change	no-change
	(0.14)			(0.11)		
ϕ_k	-0.09	no-change	no-change	-0.21	no-change	no-change
	(0.15)			(0.13)		
σ_k^2	1.13	0.32	no-change	1.16	0.32	no-change
	(0.16)	(0.05)		(0.17)	(0.04)	
	CP-ARMA (no shrinkage)			**Standard ARMA**		
	Regime 1	Regime 2	Regime 3	Regime 1	Regime 2	Regime 3
Dates	→1983Q2	→2007Q3	→ 2011Q3	→2011Q3		
μ_k	0.58	0.33	-0.01	0.40		
	(0.20)	(0.15)	(0.29)	(0.14)		
β_k	0.30	0.57	0.46	0.46		
	(0.21)	(0.19)	(0.31)	(0.17)		
ϕ_k	-0.06	-0.32	0.67	-0.15		
	(0.20)	(0.19)	(0.48)	(0.17)		
σ_k^2	1.13	0.27	0.57	0.70		
	(0.17)	(0.10)	(0.49)	(0.07)		

Posterior means and standard deviations in brackets for the CP-ARMA model estimated on GDP growth data between 1959Q2 and 2011Q3. Break dates are defined by the posterior modes of the state variables.

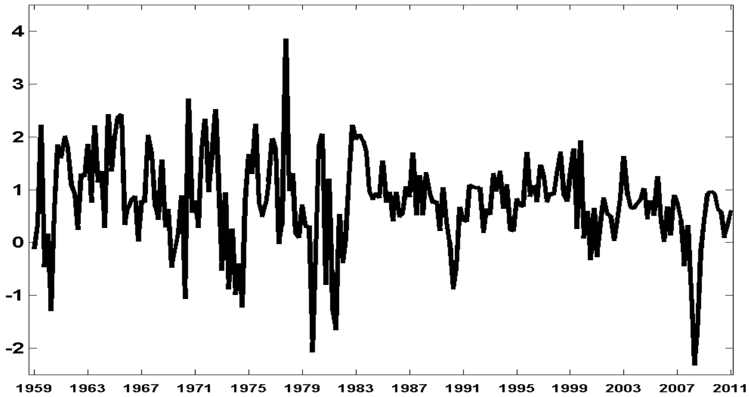
change after a structural break. This model is the natural extension of the CP-AR model proposed by Pesaran, Pettenuzzo, and Timmermann (2006). The optimal number of breaks is inferred by maximizing the marginal likelihood. This requires separately estimating the model with zero, one, two,... breaks and computing each time the marginal likelihood. For the GDP growth data we find two breaks, in 1983Q2 and 2007Q3. The estimation results in Table 2 confirm the sharp decline from 1.13 to 0.27 in the variance parameter after the first break. Since by construction all model parameters are required to change, the overall level and dynamics of GDP growth are modified as well. For example, the autoregressive parameter before and after the break has posterior means respectively equal to 0.30 and 0.57. However, given their posterior standard deviations of respectively 0.21 and 0.19 these parameters are not that different. A second break is detected around 2007Q3, indicating the end of the great moderation period, starting a zero growth rate new regime with again a substantially higher volatility. Note that the MA parameters are estimated with little precision in each of the three regimes.

Table 3: GDP growth: Parameter configurations of CP-ARMA models

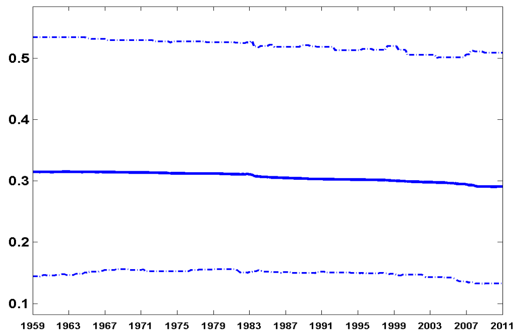
	CP-ARMA-2MU	CP-ARMA-3MN	CP-ARMA
$\{\mu, \beta, \phi, \sigma^2\}$ configuration	{1,1,1,2}-89%	{1,1,1,2}-44.4%	{3,3,3,3}-100%
	{1,1,2,2}-7.5%	{1,1,2,2}-43.2%	—
	{1,1,1,3}- 2.5%	{1,1,1,3}-8%	—

Posterior probabilities for the most likely break date configurations. The probabilities do not add up to one since there are several configurations associated with very small probabilities.

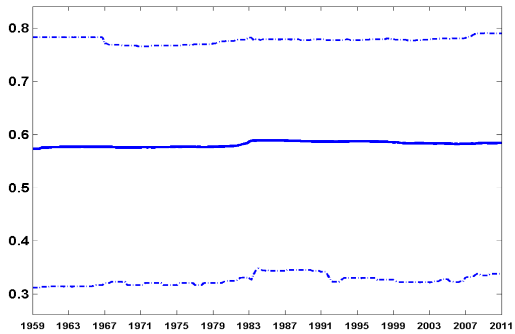
We next estimate sparse CP-ARMA models by using the 2MU and 3MN shrinkage priors. We fix the maximum number of regimes K equal to 5. Since irrelevant regimes are shrunk to zero, this number is not that important as long as it exceeds the actual number of breaks in the sample. Sparse CP-ARMA models can comprise different configurations of structural breaks. In fact, for both shrinkage priors the "active" number of regimes is either two or three as documented by the most likely break configurations in Table 3. Interestingly, the model with changes only in the variance dominates for the two priors while the configurations with no break in the parameters or breaks only in the mean parameters do not appear at all in Table 3 since they are associated with very low probabilities. Table 2 reports the posterior means and standard deviations under the preferred configuration (i.e. two regimes). For



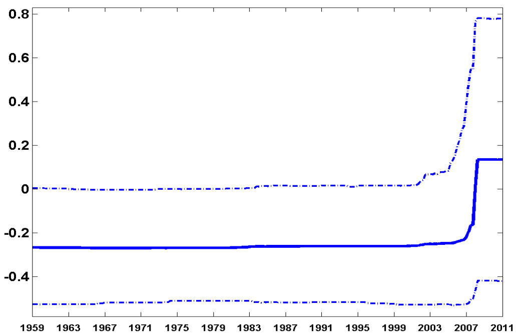
GDP growth rate



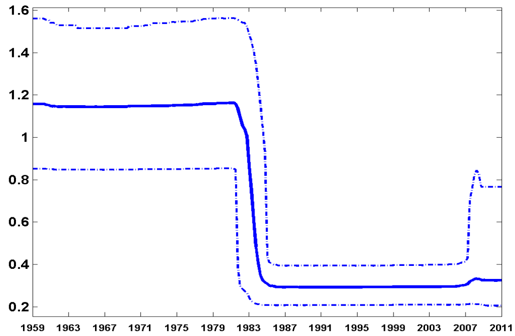
Intercept time variation



AR time variation



MA time variation



Variance time variation

Figure 8: US GDP growth rate with the 3MN shrinkage prior - Posterior means of the parameters over time and their corresponding 95% confidence intervals (in dashed line).

that specific case, only the variance parameter changes once at 1983Q2 from 1.16 to 0.32 for the 3MN shrinkage prior and from 1.13 to 0.32 for the other prior. When three breaks are detected (56% of the time for the 3MN and 10% for the 2MU), their dates coincide with those of the CP-ARMA model. In terms of model parameters, the sparse CP-ARMA model yields quite different results than the standard change-point one. Indeed, the intercept and the AR parameters remain constant over the full sample while the variance parameter unquestionably changes at the beginning of the great moderation era. Furthermore, depending on the break configuration, the MA parameter as well as the variance starts evolving at the beginning of the global financial crisis. The time variation of model parameters are displayed in Figure 8 for the 3MN shrinkage prior. We observe the break in the variance parameter at the beginning of the great moderation era. A sharp increase of the 95% posterior regions at the end of the sample indicates the uncertainties on the break of the MA and the variance parameters.

Since the sparse CP-ARMA model has less parameters than the CP-ARMA model (5 or 6 depending on the configuration instead of 12), they can be estimated with more precision. Obviously, by taking multiple break specifications into account, our approach allows for direct model combination without resorting to Bayesian model averaging. Not surprisingly, this more efficient use of data to estimate the model parameters also improves forecasting as demonstrated in Section 5 below.

The 3MN shrinkage priors necessitates the choice of threshold values. We reported results of a threshold value given by the difference between the median and the 5%-th quantile. In order to assess the sensitivity with respect to this choice, we also estimate the model with thresholds computed with the 10%-quantile and the 1%-quantile. These choices exactly deliver the same breaks and parameter values. The results are available on request.

Table 4: GDP growth: Parameter configurations of CP-ARMA models

	CP-ARMA-2MU	CP-ARMA-3MN
$\{\mu, \beta, \phi, \sigma^2\}$ configuration	{1,1,2,2}-36%	{1,1,2,4}-15%
	{1,1,1,2}-18%	{1,1,2,3}-12%
	{1,1,1,3}- 11%	{1,1,1,4}-9%

Posterior probabilities for the most likely break date configurations when the penalty parameter is fixed to -5. The probabilities do not add up to one since there are several configurations associated with very small probabilities.

The penalty parameter controls the number of breaks detected in the process. As long as its value increases, more regimes are identified and therefore more parameters have to be estimated. To illustrate its effect, Table 4 documents the best break configurations when the penalty parameter is fixed to -5 (instead of being random around the BIC value equal to -8.28). As expected, more breaks are detected and no configuration is clearly preferred as the model probabilities have sharply dropped. One can go a step further by estimating the model given several values of the penalty in order to assess its sensitivity with respect to each parameter. Figure 9 shows the posterior mode of the number of regimes observed for each parameter for a range of penalty values for both the 2MU and 3MN priors. As expected, when the penalty decreases, the breaks disappear in the parameters. Interestingly, the two priors provide very similar results. One can notice that the break of the variance is present even when the penalty is set to -20. The presence of a break in the MA term is also likely as it disappears only when the penalty amounts to -9. The full probabilities of the number of breaks per parameter for several penalties are displayed in Figure 10. Besides the variance and the MA term, the probability of having only one regime rapidly goes to one as the penalty decreases.

4.2 Monthly US 3-Month Tbill - 08/1947 – 12/2002

We revisit the empirical application of Pesaran, Pettenuzzo, and Timmermann (2006) (PPT). They estimate a CP-AR model on the Monthly US 3M Treasury Bill from August 1947 to December 2002 (we update the data later in the forecasting exercise). In their setting, all the parameters have to change after a break. Estimating several break models, PPT find that the Marginal Log-likelihood (MLL), is highest for a model with 7 regimes ($K = 6$). This leads to a process exhibiting 21 parameters. Table 5 documents the posterior means and standard deviations of the parameters. We observe that the intercept parameters μ_k are insignificant and that the AR parameter β_k indicates a unit root in each regime. Therefore, the principal regime changes are due to the variance which moves from values as small as 0.015 to levels up to 2.558.

We next estimate the Sparse CP-ARMA model with a potential number of breaks amounting to 13. As opposed to the standard CP model, the Sparse ARMA model has the advantage of encompassing multiple numbers of breaks during the estimation process. Table 6 documents the most frequent configurations in the posterior draws. In line with PPT, we observe

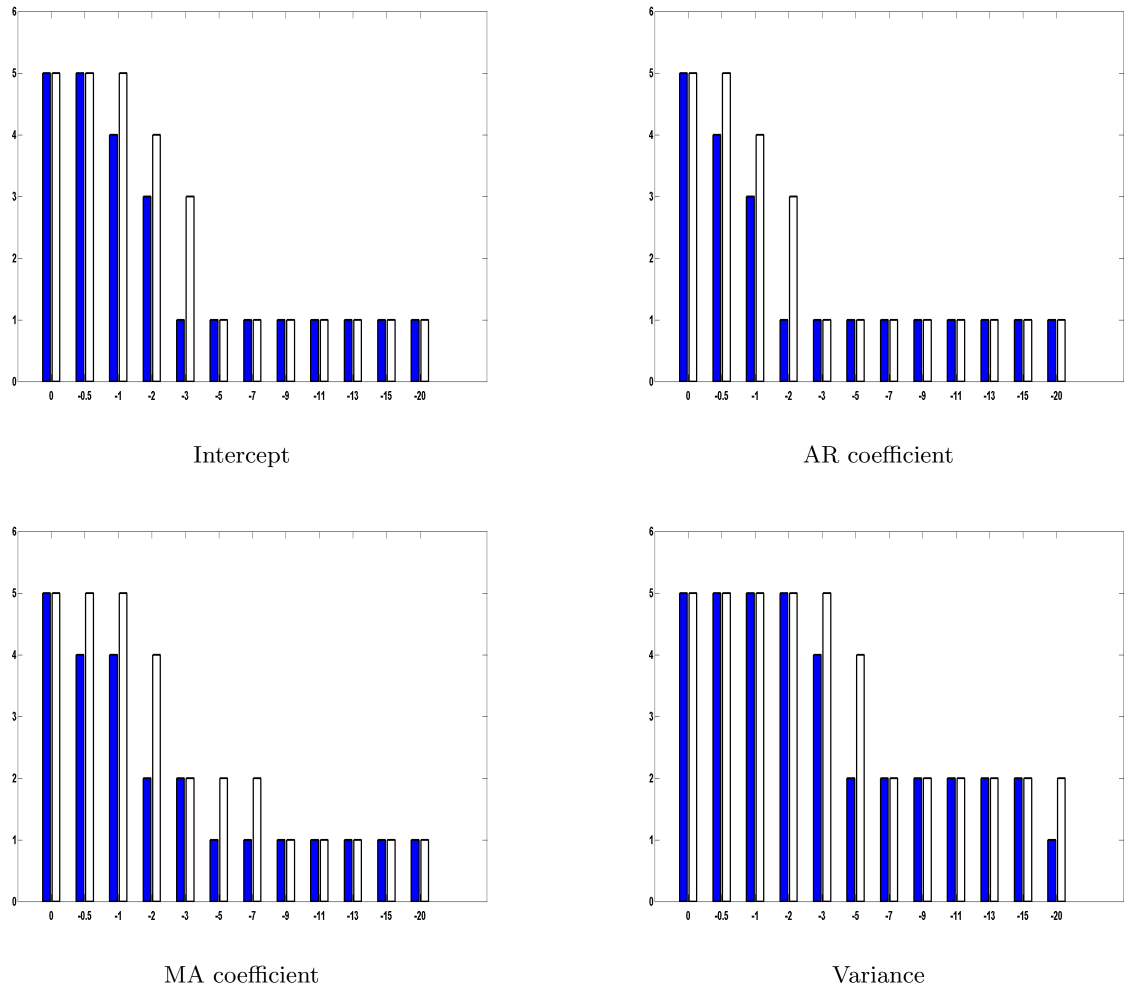
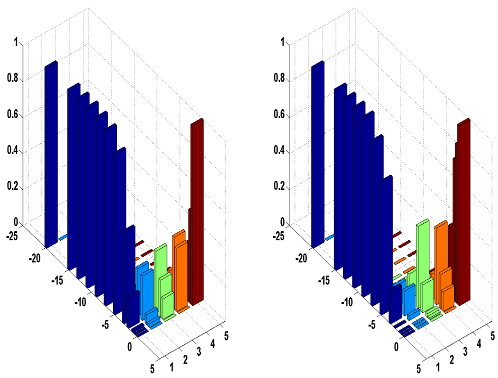
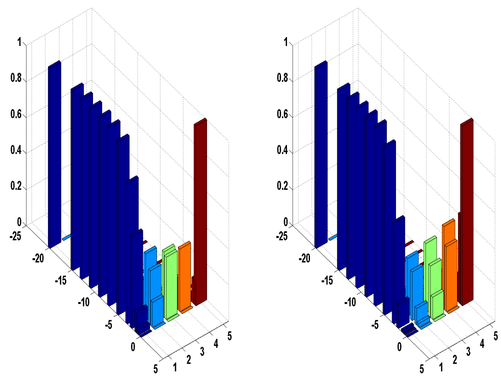


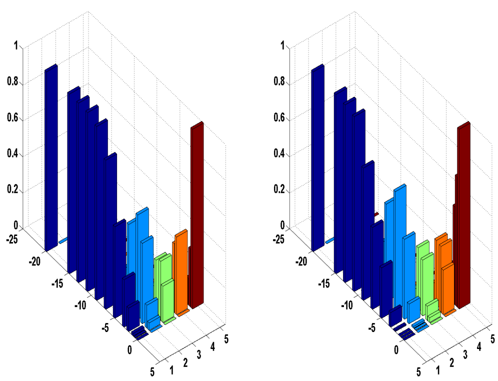
Figure 9: US GDP growth rate - Posterior mode of the number of regimes per parameters for several values of the penalty. The result of the 2MU prior is detailed in blue while the 3MN prior is given in white.



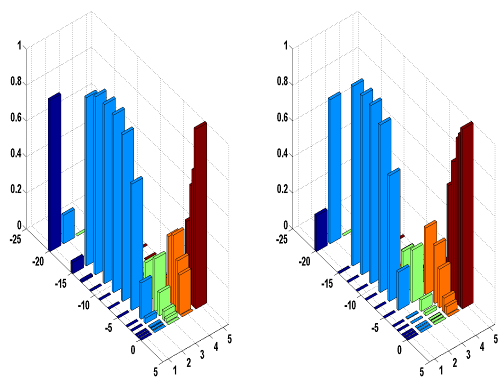
Intercept



AR coefficient



MA coefficient



Variance

Figure 10: US GDP growth rate - Probabilities of the number of regimes per parameters for several values of the penalty. The 2MU prior is displayed on the left Graphic while the 3MN prior is provided on the right one.

Table 5: PPT's results (US 3M Tbill)

PPT's result: CP-AR model with 7 regimes							
Dates	$\rightarrow$ 1957	$\rightarrow$ 1960	$\rightarrow$ 1966	$\rightarrow$ 1979	$\rightarrow$ 1982	$\rightarrow$ 1989	$\rightarrow$ 2002
μ_k	0.021	0.252	0.017	0.220	0.412	0.246	-0.004
	(0.034)	(0.208)	(0.067)	(0.161)	(0.521)	(0.211)	(0.054)
β_k	1.002	0.895	1.006	0.969	0.958	0.968	0.992
	(0.020)	(0.071)	(0.020)	(0.026)	(0.045)	(0.27)	(0.011)
σ_k^2	0.023	0.256	0.015	0.260	2.558	0.161	0.048
	(0.003)	(0.068)	(0.003)	(0.031)	(0.671)	(0.027)	(0.005)

Posterior means and standard deviations for the CP-AR model of PPT.

that the mean parameters hardly vary over time and that the regimes are mainly generated by switches in the variance. Figure 11 displays the marginal probabilities of having a specific number of breaks for each model parameters given the two shrinkage priors. Besides a likely break in the AR dynamics in the 3MN prior case that is not detected by the 2MU prior case, the two models lead to very similar results.

Table 6: Parameter configurations of CP-ARMA models (US 3M Tbill)

	CP-ARMA-2MU	CP-ARMA-3MN	CP-AR
$\{\mu, \beta, \phi, \sigma^2\}$ configuration	{1,1,1,9}-85%	{1,1,1,9}-54%	{7,7,7,7}-100%
	{1,1,1,10}-12%	{1,2,1,9}-36%	—
	{1,1,2,9}-1%	{1,1,1,10}-3%	—

Posterior probabilities for the most likely break date configurations. The probabilities do not add up to one since there are several configurations associated with very small probabilities.

Tables 7 report results of the Sparse CP-ARMA model with the two different priors given the most likely setting (8 breaks in the variance). The two prior distributions lead almost identical results. Although richer in model dynamics by allowing for MA terms, our two specifications are more parsimonious than PPT. In fact, the number of model parameters amounts only to 12. In contrast to the CP-AR model, the Sparse-ARMA model shows explicitly that the US 3M Tbill series behaves as an heteroskedastic random walk. Also, the fact that the MA parameter is significant indicates the relevance of using a richer dynamics that typically involve path dependence. Interestingly, all the breaks detected by PPT are also identified by the Sparse CP-ARMA model. Since the conditional mean parameters are

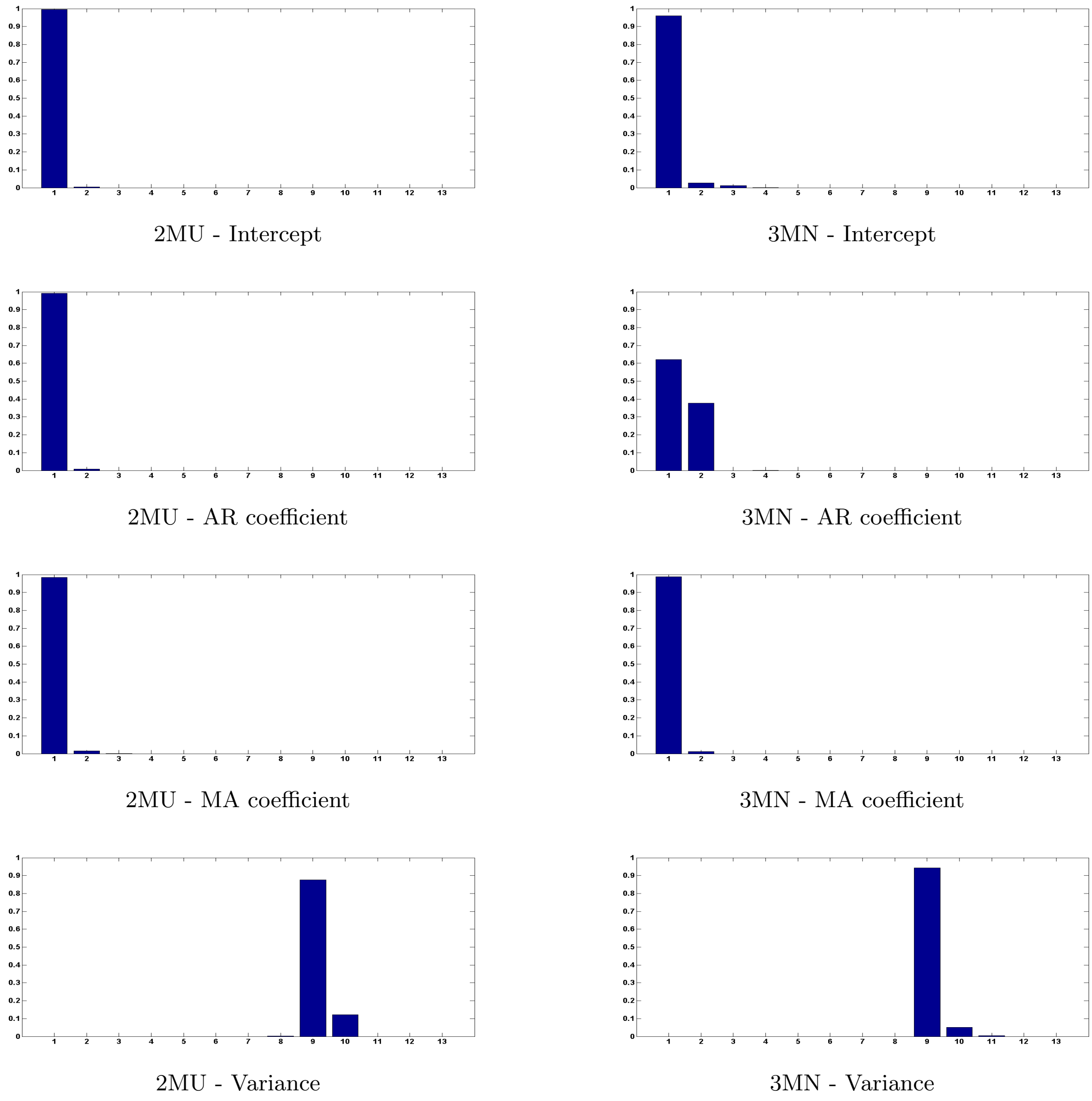


Figure 11: US 3M Tbill - Posterior probabilities of the number of breaks per parameters.

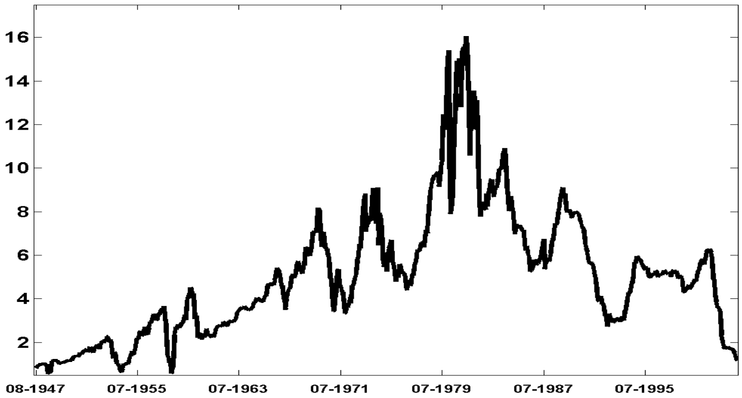
Table 7: Sparse CP-ARMA model (US 3M Tbill)

			2MU shrinkage prior		
Intercept			0.09		
			(0.04)		
AR(1) Coeff.			0.99		
			(0.01)		
MA(1) Coeff.			0.15		
			(0.05)		
Period	$\rightarrow$ 08-1957	$\rightarrow$ 10-1960	$\rightarrow$ 06-1966	$\rightarrow$ 03-1973	$\rightarrow$ 01-1975
Variance	0.02	0.24	0.01	0.17	0.86
	(0.01)	(0.05)	(0.00)	(0.06)	(0.23)
Period	$\rightarrow$ 07-1979	$\rightarrow$ 08-1982	$\rightarrow$ 06-1989	$\rightarrow$ 12-2002	
Variance	0.12	1.87	0.17	0.04	
	(0.03)	(0.34)	(0.03)	(0.01)	
			3MN shrinkage prior		
Intercept			0.05		
			(0.02)		
AR(1) Coeff.			0.99		
			(0.00)		
MA(1) Coeff.			0.16		
			(0.04)		
Period	$\rightarrow$ 07-1957	$\rightarrow$ 02-1961	$\rightarrow$ 06-1966	$\rightarrow$ 03-1973	$\rightarrow$ 01-1975
Variance	0.02	0.22	0.01	0.17	0.87
	(0.00)	(0.03)	(0.00)	(0.03)	(0.18)
Period	$\rightarrow$ 07-1979	$\rightarrow$ 08-1982	$\rightarrow$ 06-1989	$\rightarrow$ 12-2002	
Variance	0.13	2.12	0.16	0.04	
	(0.03)	(0.27)	(0.02)	(0.01)	

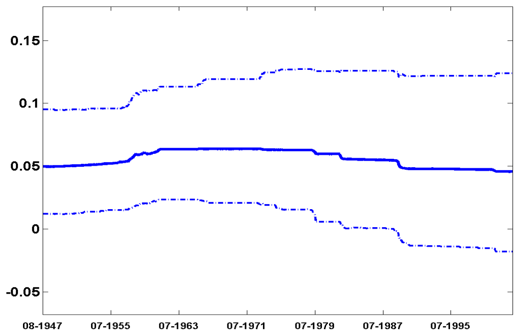
Posterior means and standard deviations for the regimes of the most likely Sparse CP-ARMA model.

estimated on longer window sizes, the Sparse CP-ARMA model is expected to provide also better forecasts. This feature is documented in Section 5.

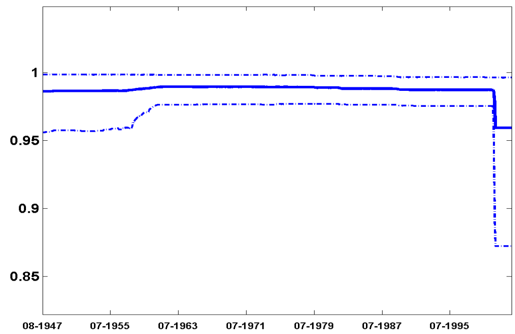
For respectively the 3MN and 2MU priors, Figures 12 and 13 display the posterior means of the model parameters over time and their corresponding 95% confidence intervals. Again, the time-varying variance behaves analogously for the two shrinkage priors. The differences between the 3MN and 2MU priors arise in the short AR regime at the end of the sample that comes up in the second best configuration. We also observe that the intercept slowly evolves over time indicating some smooth transition behavior.



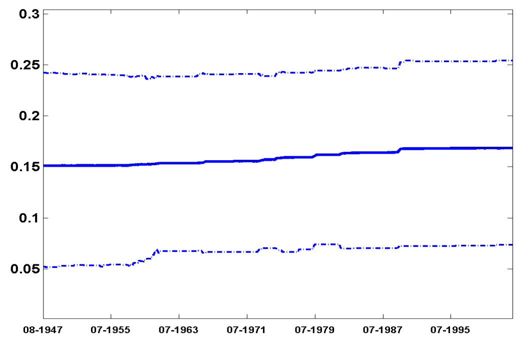
US 3M Tbill



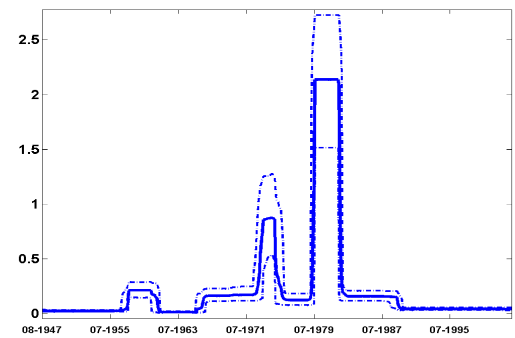
Intercept time variation



AR time variation



MA time variation



Variance time variation

Figure 12: US 3M Tbill with the 3MN shrinkage prior - Posterior means of the parameters over time and their corresponding 95% confidence intervals (in dashed line).

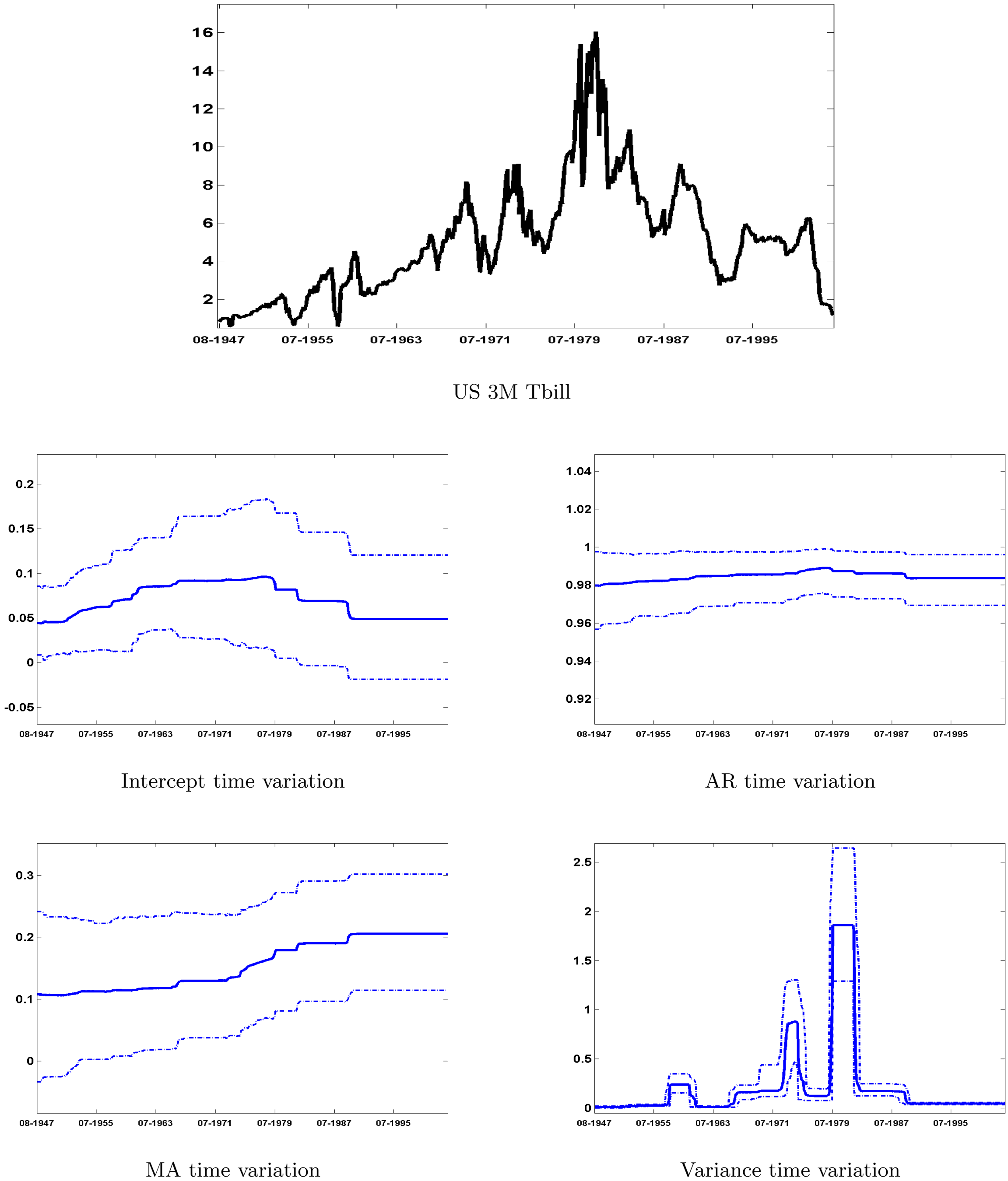


Figure 13: US 3M Tbill with the 2MU shrinkage prior - Posterior means of the parameters over time and their corresponding 95% confidence intervals (in dashed line).

5 Forecasting performance

The Sparse CP-ARMA model exhibits the appealing feature of detecting which parameter of the model evolves when a structural break occurs. Since the approach encompasses ARMA and CP-ARMA models as special cases and therefore strikes a nice balance between fit and parsimony, the Sparse CP-ARMA model potentially produces also more precise predictive densities. To investigate this, we provide a small forecasting exercise for the US GDP and (updated) US 3M Tbill series. A comparison with other models using time varying parameters combined with regressor variables, see for example Belmonte, Koop, and Korobilis (2014), is beyond the scope of this paper.

Forecasting using the Sparse CP-ARMA model is straightforward as for the ARMA and CP-ARMA models used here. For each of the M posterior draws of the model parameters, one can simulate the distribution of the future observations by assuming that the model remains in the last regime. More specifically, to sample the marginal posterior distribution of the h-ahead forecast, we have that

$$f(y_{T+h}|y_{1:T}) = \sum_{s_{1:T}} \int f(y_{T+1:T+h}|y_{1:T}, s_{1:T}, p_{1:K}, \Theta, \Sigma) f(s_{1:T}, p_{1:K}, \Theta, \Sigma|y_{1:T}) dy_{T+1:T+h-1} dP d\Theta d\Sigma,$$

where the sum is over all the possible state vectors and $f(y_{T+1:T+h}|y_{1:T}, s_{1:T}, p_{1:K}, \Theta, \Sigma)$ can be recursively decomposed into h products as follows $\prod_{q=1}^{h} f(y_{T+q}|y_{1:T+q-1}, s_{1:T}, p_{1:K}, \Theta, \Sigma)$. Consequently, for each of the M posterior draws of the model parameters, one can simulate a h-ahead prediction by simulating sequentially for $q = 1, ..., h$ as follows

$$y_{T+q}|y_{1:T+q-1}, s_{1:T}, p_{1:K}, \Theta, \Sigma \sim N(\mu_{K+1} + \beta_{K+1} y_{T+q-1} + \phi_{K+1} \epsilon_{T+q-1}, \sigma^2_{K+1}).$$

Note that as in Pesaran, Pettenuzzo, and Timmermann (2006) we can also implement forecasting subject to future breaks. Since our sparse CP model performs very well, the empirical forecasting performance of incorporating out-of-sample breaks is left for future research.

Following the same forecasting setup as in Bauwens, Koop, Koribilis, and Rombouts (2015), our study consists in starting with the first forty percent of the observations, estimating each model on this time-span, producing forecasts and then iterates this operation by adding one by one the remainder of observations. For example, considering the inflation time series (646 observations), this procedure requires 389 model estimations for the ARMA and the Sparse CP-ARMA processes while it amounts to 389*($K_{\max}$+1) estimations of the

Table 8: Forecast exercise : RMSEs and Average of predictive densities for different horizons given three models (ARMA(1,1), CP-ARMA(1,1) and the Sparse CP-ARMA(1,1)).

Forecast horizon	1	2	4	8	12	16
US GDP growth rate - forecasting from 1973Q3 to 2014Q1						
			Average predictive likelihood			
ARMA	0.3068	0.2821	0.2658	0.2602	0.2592	0.2593
CP-ARMA ($K_{\max} = 3$)	0.4041	0.3837	0.3618	0.3570	0.3553	0.3619
Sparse CP-ARMA - 3MN	**0.4141**	**0.3965**	**0.3732**	0.3641	0.3629	**0.3673**
Sparse CP-ARMA - 2MU	0.4111	0.3926	0.3732	**0.3692**	**0.3693**	0.3665
			RMSE			
ARMA	0.6505	0.7577	0.8826	0.9467	0.9651	0.9585
CP-ARMA ($K_{\max} = 3$)	0.6425	0.7119	0.8241	0.8347	0.8561	0.8324
Sparse CP-ARMA - 3MN	**0.6090**	0.6776	0.7440	0.7899	**0.8035**	0.8108
Sparse CP-ARMA - 2MU	0.6110	**0.6630**	**0.7314**	**0.7688**	0.8051	**0.7979**
US 3M Tbill - forecasting from 1978M11 to 2008M12						
			Average predictive likelihood			
ARMA	0.6911	0.3990	0.2562	0.1729	0.1448	0.1285
CP-ARMA ($K_{\max} = 8$)	1.0373	0.5643	0.3458	0.2115	0.1619	0.1346
Sparse CP-ARMA - 3MN	**1.1506**	**0.6558**	**0.4058**	**0.2449**	**0.1872**	**0.1560**
Sparse CP-ARMA - 2MU	1.1495	0.6506	0.4003	0.2410	0.1856	0.1552
			RMSE			
ARMA	0.2625	0.8159	1.7240	2.8194	3.9539	4.5659
CP-ARMA ($K_{\max} = 8$)	0.2815	0.8747	1.8326	2.9876	4.4173	5.4948
Sparse CP-ARMA - 3MN	0.2437	0.7493	1.5383	**2.5882**	3.7698	4.6421
Sparse CP-ARMA - 2MU	**0.2381**	**0.7289**	**1.5284**	2.6056	**3.6621**	**4.4835**

CP-ARMA model (where $K_{\max}$ denotes the maximum number of breaks allowed in the series). Note that the Sparse CP-ARMA model limits the number of estimations compared to the CP-ARMA one although its specification is more general. The feasibility of this forecast exercise is due to the SMC sampler which allows to update the posterior realizations in light of adding new observations without having to run an entire estimation procedure each time. More details are given in Appendix B.

We consider several forecast horizons (1,2,4,8,12,16), e.g. in the case of the quarterly US GDP growth rate, the horizons correspond to one quarter, two quarters and one, two, three and four years. We updated the US 3M Treasury bill time series up to until 12-2008 so that it includes the global financial crisis. Table 8 documents the average of the predictive densities obtained at the different horizons (i.e. $\forall h \in \{1, 2, 4, 8, 12, 16\}\,,\quad \sum_{t=\tau}^{T-h} f(y_{t+h}|y_{1:t})/(T-h-\tau+1)$ with τ indicating the first out-of-sample date) as well as the root mean squared error (RMSE) derived from the posterior predictive median.

For both applications, the Sparse CP-ARMA model dominates in terms of average predictive density and RMSE and this holds for all horizons. For example, the US 3M Treasury bill results show that the average predictive likelihood for the single regime ARMA model amounts to 0.6911 for the one step ahead forecasts (horizon 1). Taking into account breaks by forecasting using a CP-ARMA model, the average predictive likelihood improves to 1.0373. The sparse CP-ARMA model allows the average predictive likelihood further increase to 1.1506.

6 Conclusion

This paper introduces the Sparse change point time series model, a new model class in the change point framework, that detects which parameters of the model change when a structural break occurs. The proposed approach has several advantages compared to the current methodology.

First, it solves the over-parametrization issue exhibited by all the CP models assuming that all the parameters vary when a break is detected. Moreover, our method requires only one estimation of the model to select the number of regimes and does not rely on the marginal likelihood. This leads to a substantial gain in computational time and coding efforts. Finally, the detection of breaks depends on a controlled parameter that penalizes the log-likelihood function, the advantage being its simple interpretation compared to a selection based on the marginal likelihood. All these improvements are obtained by using shrinkage priors on the standard CP parameters. As the existing priors are not appropriate to our methodology, we introduce new shrinkage ones, detail their properties and discuss selection of their hyper parameters.

The empirical exercise based on an ARMA model with path dependence emphasize the ability of our framework to detect which parameters change when a break occur. It also highlights that standard CP models lead to an over-parametrization issue that is ruled out with the sparse approach. As a result, we also obtain sizeable improvements in the predictive densities compared to the ARMA and the CP-ARMA models.

This paper uses the new shrinkage priors to model univariate time series. Given the recent developments of time varying parameter vector autoregressive models, it would be interesting to extend the approach to multivariate time series.

A MCMC for CP-ARMA estimation

As shown in Algorithm 1, each MCMC iteration consists in four steps: sampling from $\pi(\Theta|y_{1:T}, s_{1:T}, p_{1:K}, \Sigma)$, from $\pi(\Sigma|y_{1:T}, s_{1:T}, p_{1:K}, \Theta)$, from $\pi(p_{1:K}|y_{1:T}, s_{1:T}, \Theta, \Sigma)$ and finally from $\pi(s_{1:T}|y_{1:T}, p_{1:K}, \Theta, \Sigma)$. Since the first two full conditional distributions are nonstandard, we employ the DREAM method of Vrugt, ter Braak, Diks, Robinson, Hyman, and Higdon (2009) which basically is a Random-Walk Metropolis Hastings algorithm with adaptive covariance matrices. More precisely, given M parallel MCMC chains, a candidate for the parameter of the chain j is given by (omitting subscripts referring to MCMC iterations):

- Sample $\Theta^j \sim \Theta|y_{1:T}, s^j_{1:T}, p^j_{1:K}, \Sigma^j$:
  - Draw a proposal $\tilde{\Theta}$ from

$$\tilde{\Theta} = \Theta^j + 2.38/\sqrt{2\delta d}\left(\sum_{g=1}^{\delta} \Theta_{r_1(g)} - \sum_{h=1}^{\delta} \Theta_{r_2(h)}\right) + \zeta$$

    with $\forall g, h = 1, 2, ..., \delta$, $i \neq r_1(g)$, $r_2(h)$; $r_1(.)$ and $r_2(.)$ stand for random integers uniformly distributed on the support $[1, M]_{-j}$ with the requirement that $r_1(g) \neq r_2(h)$ when $g = h$; $\zeta \sim N(0, \eta^2_\Theta I_d)$; and d is the number of elements in Θ. The standard deviation η_Θ is set to 0.00001 and δ is randomly taken from $[1, 4]$. As the dimension of Θ can be high, we apply the DREAM on sub-blocks of random size as in Chib and Ramamurthy (2010) and in Dufays (2012).

– Accept or reject the draw according to the Metropolis probability:

$$\min\left\{\frac{\pi(\tilde{\Theta}|y_{1:T}, s^j_{1:T}, p^j_{1:K}, \Sigma^j)}{\pi(\Theta^j|y_{1:T}, s^j_{1:T}, p^j_{1:K}, \Sigma^j)}, 1\right\}.$$

The conditional distribution $\pi(\Sigma|y_{1:T}, s_{1:T}, p_{1:K}, \Theta)$ is also sampled by the DREAM algorithm as $\Sigma = \{\sigma^2_1, \Delta\sigma^2_2, ..., \Delta\sigma^2_K\}$ is a vector of variance parameters. The third posterior distribution $\pi(p_{1:K}|y_{1:T}, s_{1:T}, \Theta, \Sigma)$ is naturally conjugated and is given by the Beta distribution. Sampling from $(s_{1:T}|y_{1:T}, p_{1:K}, \Theta, \Sigma)$ is challenging due to the path dependence issue exhibited by the CP-ARMA model. We use an approximate model without path dependence which enables the use of the forward-backward algorithm. This approximation is given by

$$y_t \;=\; \mu_{s_t} + \beta_{s_t} y_{t-1} + \phi_{s_t}\tilde{\epsilon}_{t-1,s_t} + \epsilon_t \;\text{ with }\; \epsilon_t \;\sim\; N(0, \sigma^2_{s_t})$$

where $\tilde{\epsilon}_{t-1,s_t} = \sum_{i=1}^{K+1}(y_{t-1} - \mu_i + \beta_i y_{t-2} + \phi_i\tilde{\epsilon}_{t-2,i})\tilde{f}(s_{t-1} = i|y_{1:t}, s_t, p_{1:K+1}, \Theta, \Sigma)$ with $\tilde{f}(\ )$ denoting the filtering density of the approximate model. The draw obtained from this method is accepted or rejected using the Metropolis-Hastings ratio, see Dufays (2012) for details.

B Time and Tempered algorithm

The TNT sampler, see Dufays (2014), is a combination of the SMC sampler (Del Moral, Doucet, and Jasra (2006)) and the Re-sample Move (RM) algorithm (see Gilks and Berzuini (2001)). It decomposes the posterior distribution into the product :

$$\pi(s_{1:T}, p_{1:K+1}, \Theta, \Sigma|y_{1:T}) = \pi(s_{1:\tau}, p_{1:K+1}, \Theta, \Sigma|y_{1:\tau})\frac{f(y_{\tau+1:T}, s_{\tau+1:T}|s_{1:\tau}, p_{1:K+1}, \Theta, \Sigma, y_{1:\tau})}{f(y_{\tau+1:T}|y_{1:\tau})} \quad (5)$$

where $y_{i:j} = \{y_i, ..., y_j\}$ for $i \leq j$ and τ stands for a user-defined parameter belonging to [1,T]. The TNT sampler uses the SMC sampler for numerically approximating the first term of (5) and then switches to the RM for estimating the sequence of posterior distributions from $\pi(s_{1:\tau}, p_{1:K+1}, \Theta, \Sigma|y_{1:\tau})$ to $\pi(s_{1:T}, p_{1:K+1}, \Theta, \Sigma|y_{1:T})$. The TNT sampler is therefore well-suited for estimating many posterior distributions by keeping updating previous estimations. Moreover, since the first distribution of interest $\pi(s_{1:\tau}, p_{1:K+1}, \Theta, \Sigma|y_{1:\tau})$ is approximated by a SMC sampler, we avoid particle degeneracy issues in the early stages that can arise in standard SMC samplers that only iterate on time.

The SMC sampler relies on an increasing tempered function $\phi(z) : [1, ..., p] \to [0, 1]$ such that $\phi(p) = 1$ to build a sequence of artificial distributions $\pi_n(s_{1:\tau}, p_{1:K+1}, \Theta, \Sigma | y_{1:\tau}) \propto [f(y_{1:\tau}|s_{1:\tau}, p_{1:K+1}, \Theta, \Sigma) f(\Theta, \Sigma)]^{\phi(n)} \bar{f}(\Theta, \Sigma)^{1-\phi(n)} f(s_{1:\tau}, p_{1:K+1}, \Theta, \Sigma)$ which coincides with the targeted posterior distribution when $\phi(n) = 1$. The auxiliary density $\bar{f}(\Theta, \Sigma)$ allows to introduce step by step the shrinkage prior into the sequence of importance sampling. The auxiliary distribution of the mean parameters is given by $\{\mu_1, \beta_1, \phi_1\}' \sim N(\begin{pmatrix} 0 & 0 & 0 \end{pmatrix}', 0.5I_3)$ for the first regime and by $\{\Delta\mu_k, \Delta\beta_k, \Delta\phi_k\}' \sim N(\begin{pmatrix} 0 & 0 & 0 \end{pmatrix}', 0.5I_3)$ for each other regime $k \in [2, K+1]$. The auxiliary variance distribution is set to $\sigma_1^{-2} \sim$ G(1,1) and to $\Delta\sigma_k^2 \sim N(0, 0.5)1_{\sigma_k^2 > 0}$ for $k \in [2, K+1]$.

The TNT algorithm is initiated by drawing M particles from the prior distributions with associated uniform weights $\{x_0^i, W_0^i\}_{i=1}^M$ where $x_0^i = \{s_{1:\tau}^i, p_{1:K+1}^i, \Theta^i, \Sigma^i\} \sim \bar{f}(\Theta, \Sigma) f(p_{1:K}) f(s_{1:\tau}|p_{1:K})$ and then iterates from $n = 1, \ldots, p, p+1, ..., p + (T - \tau) + 1$ as follows

- **Correction step:** $\forall i \in [1, M]$, Re-weight each particle with respect to the nth posterior distribution

  - If in tempered domain $(n \leq p)$:

$$\tilde{w}_n^i = [f(y_{1:\tau}|x_{n-1}^i) f(\Theta^i, \Sigma^i)]^{\phi(n)-\phi(n-1)} \bar{f}(\Theta^i, \Sigma^i)^{\phi(n-1)-\phi(n)} \quad (6)$$

  - If in time domain $(n > p)$: Set $x_n^i = \{\Theta^i, \Sigma^i, p_{1:K}^i, s_{1:\tau+n-p-1}^i, s_{\tau+n-p}\}$ with $s_{\tau+n-p} = K + 1$, i.e. the last regime and compute the weights

$$\tilde{w}_n^i = \frac{f(y_{1:\tau+n-p}|x_n^i) f(x_n^i)}{f(y_{1:\tau+n-p-1}|x_{n-1}^i) f(x_{n-1}^i)} \quad (7)$$

  Compute the unnormalized weights : $\tilde{W}_n^i = \tilde{w}_n^i W_{n-1}^i$.
  Normalize the weights : $W_n^i = \frac{\tilde{W}_n^i}{\sum_{j=1}^M \tilde{W}_n^j}$

- **Re-sampling step:** Compute the Effective Sample Size (ESS) as

$$ESS = \frac{M}{\sum_{i=1}^M (W_n^i)^2}$$

- If $ESS \in [\lambda_1, \lambda_2]$ where λ_1 and λ_2 are user-defined thresholds then apply a stratified re-sampling (see Carpenter, Clifford, and Fearnhead (1999)) on the particles and reset the weight uniformly.
- if $ESS < \lambda_1$[1], re-run a TNT sampler with $\tau = \tau + \max(n - p, 0)$.

- **Mutation step:** $\forall i \in [1, M]$, run N steps of the MCMC kernel presented in Appendix A with invariant distribution $\pi_n(x_n|y_{1:\tau})$ for $n \leq p$ and $\pi(x_n|y_{1:\tau+n-p})$ for $n > p$.

The user-defined parameters are set to $\lambda_1 = 0.25M$, $\lambda_2 = 0.75M$, $M = 2.000$ and $N = 50$. The function ϕ is adapted at each iteration to insure that the sequential artificial distributions are similar (see Algorithm 2, more details in Jasra A., Doucet, and Tsagaris (2011)).

References

Andrews, D. (1993): "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856.

Bai, J., and P. Perron (1998): "Estimating and Testing Linear Models With Multiple Structure Changes," *Econometrica*, 66, 47–78.

Bauwens, L., J.-F. Carpantier, and A. Dufays (2015): "Autoregressive Moving Average Infinite Hidden Markov-Switching Models," *Journal of Business and Economic Statistics*, Forthcoming.

Bauwens, L., A. Dufays, and J. Rombouts (2013): "Marginal Likelihood for Markov Switching and Change-point GARCH Models," *Journal of Econometrics*, 178 (3), 508–522.

Bauwens, L., G. Koop, D. Koribilis, and J. Rombouts (2015): "The Contribution of Structural Break Models to Forecasting Macroecomic Series," *Journal of Applied Econometrics*, 30(4), 596–620.

Belmonte, M. A., G. Koop, and D. Korobilis (2014): "Hierarchical Shrinkage in Time-Varying Parameter Models," *Journal of Forecasting*, 33(1), 80–94.

[1]it can only happen when $n > p$ due to the adaption of the tempered function during the SMC sampler. See Algorithm 2 for more details.

Billio, M., R. Casarin, and A. Osuntuyi (2014): "Efficient Gibbs sampling for Markov switching {GARCH} models," *Computational Statistics & Data Analysis*, (0), –.

Billio, M., A. Monfort, and C. Robert (1999): "Bayesian estimation of switching ARMA models," *Journal of Econometrics*, 93(2), 229 – 255.

Carpenter, J., P. Clifford, and P. Fearnhead (1999): "Improved particle filter for nonlinear problems," *IEE Proceedings-Radar, Sonar and Navigation*, 146(1), 2–7.

Chan, N., C. Yip, and R. Zhang (2014): "Group LASSO for Structural Break Time Series," *Journal of the American Statistical Association*, 109(506), 590–599.

Chib, S. (1998): "Estimation and comparison of multiple change-point models," *Journal of Econometrics*, 86, 221–241.

Chib, S., and S. Ramamurthy (2010): "Tailored randomized block MCMC methods with application to DSGE models," *Journal of Econometrics*, 155(1), 19–38.

Del Moral, P., A. Doucet, and A. Jasra (2006): "Sequential Monte Carlo samplers," *The Royal Statistical Society: Series B(Statistical Methodology)*, 68, 411–436.

Dufays, A. (2012): "Infinite-State Markov-switching for Dynamic Volatility and Correlation Models," *CORE discussion paper, 2012/43*.

——— (2014): "On the Conjugacy of Off-Line and On-Line Sequential Monte Carlo Samplers," *National Bank of Belgium Working Paper No. 263*, p. Available at SSRN: http://ssrn.com/abstract=2495351 or http://dx.doi.org/10.2139/ssrn.2495351.

Durham, G., and J. Geweke (2012): "Adaptive Sequential Posterior Simulators for Massively Parallel Computing Environments," *http://www.censoc.uts.edu.au/pdfs/gewekepapers/gpu2full.pdf*.

Eo, Y. (2012): "Bayesian Inference about the Types of Structural Breaks When There Are Different Breaks in Many Parameters," *Available at SSRN: http://ssrn.com/abstract=2011825 or http://dx.doi.org/10.2139/ssrn.2011825*, Available at SSRN: http://ssrn.com/abstract=2011825 or http://dx.doi.org/10.2139/ssrn.2011825.

George, E. I., and R. E. McCulloch (1993): "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88(423), 881–889.

Geweke, J., and Y. Jiang (2011): "Inference and prediction in a multiple-structural-break model," *Journal of Econometrics*, 163(2), 172 – 185.

Gilks, W. R., and C. Berzuini (2001): "Following a moving target - Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 127–146.

Giordani, P., and R. Kohn (2008): "Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models," *Journal of Business & Economic Statistics*, 26(1), 66–77.

Griffin, J., and P. Brown (2010): "Inference with normal-gamma prior distributions in regression problems," *Bayesian Analysis*, 5, 171–188.

Hamilton, J. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Herbst, E., and F. Schorfheide (2012): "Sequential Monte carlo sampling for DSGE models," *Working Paper No12-27, Federal reserve Bank of Philadelphia.*

Inoue, A., and L. Kilian (2008): "How useful is bagging in forecasting economic time series? A case study of U.S. Consumer Price Inflation," *Journal of the American Statistical Association*, 103(482), 511–522.

Ishwaran, H., and J. S. Rao (2005): "Spike and Slab Variable Selection: Frequentist and Bayesian Strategies," *The Annals of Statistics*, 33(2), 730–773.

Jasra, A., D. A. Stephens, and C. C. Holmes (2007): "On population-based simulation for static inference," *Statistics and Computing*, 17(3), 263–279.

Jasra A., Stephens, D. A., A. Doucet, and T. Tsagaris (2011): "Inference for Lévy-Driven Stochastic Volatility Models via Adaptive Sequential Monte Carlo," *Scandinavian Journal of Statistics*, 38(1), 1–22.

Kalli, M., and J. E. Griffin (2014): "Time-varying sparsity in dynamic regression models," *Journal of Econometrics*, 178(2), 779 – 793.

Kass, R., and A. Raftery (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

Koop, G., and S. Potter (2007): "Estimation and Forecasting with Multiple Breaks," *Review of Economic Studies*, 74, 763–789.

Maheu, J. M., and S. Gordon (2008): "Learning, forecasting and structural breaks," *Journal of Applied Econometrics*, 23(5), 553–583.

Mitchell, T. J., and J. J. Beauchamp (1988): "Bayesian Variable Selection in Linear Regression," *Journal of the American Statistical Association*, 83(404), 1023–1032.

Park, T., and G. Casella (2008): "The Bayesian Lasso," *Journal of the American Statistical Association*, 103(482), 681–686.

Pastor, L., and R. F. Stambaugh (2001): "The Equity Premium and Structural Breaks," *Journal of Finance*, 56, 1207–1239.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006): "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, 73, 1057–1084.

Rabiner, L. R. (1989): "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, pp. 257–286.

Scheipl, F., L. Fahrmeir, and T. Kneib (2012): "Spike-and-Slab Priors for Function Selection in Structured Additive Regression Models," *Journal of the American Statistical Association*, 107(500), 1518–1532.

Stock, J. H., and M. W. Watson (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 14, 11–30.

——— (2012): "Generalized Shrinkage Methods for Forecasting Using Many Predictors," *Journal of Business & Economic Statistics*, 30(4), 481–493.

Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009): “Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptative Randomized Subspace Sampling,” *International Journal of Nonlinear Sciences and Numerical Simulations*, 10, 271–288.

West, M. (1987): “On scale mixtures of normal distributions,” *Biometrika*, 74, 646–648.