



Munich Personal RePEc Archive

**Particle Gibbs with Ancestor Sampling  
Methods for Unobserved Component  
Time Series Models with Heavy Tails,  
Serial Dependence and Structural Breaks**

Nonejad, Nima

1 May 2014

Online at <https://mpra.ub.uni-muenchen.de/55664/>

MPRA Paper No. 55664, posted 02 May 2014 07:09 UTC

# Particle Gibbs with Ancestor Sampling Methods for Unobserved Component Time Series Models with Heavy Tails, Serial Dependence and Structural Breaks

Nima Nonejad<sup>†\*</sup>

Aarhus University and CREATES

## Abstract

Particle Gibbs with ancestor sampling (PG-AS) is a new tool in the family of sequential Monte Carlo methods. We apply PG-AS to the challenging class of unobserved component time series models and demonstrate its flexibility under different circumstances. We also combine discrete structural breaks within the unobserved component model framework. We do this by modeling and forecasting time series characteristics of postwar US inflation using a long memory autoregressive fractionally integrated moving average model with stochastic volatility where we allow for structural breaks in the level, long and short memory parameters contemporaneously with breaks in the level, persistence and the conditional volatility of the volatility of inflation.

**Keywords:** Ancestor sampling, Bayes, Particle filtering, Structural breaks

(*JEL*: C11, C22, C52, C63)

---

\*The author acknowledges support from CREATES-Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. † Department of Economics and Business, Aarhus University, Fuglesangs Alle 4, 8210 Aarhus V, Denmark. Phone: +45 87165325. E-mail: [nnonejad@creates.au.dk](mailto:nnonejad@creates.au.dk), [nimanonejad@gmail.com](mailto:nimanonejad@gmail.com). I would like to thank Joshua Chan, Siem Jan Koopman, Fredrik Lindsten and Asger Lunde for useful comments. This version 05/01/2014.

# 1 Introduction

Unobserved component (UC) models are widely used to model time series and dynamical systems. In this paper we define UC models as a general class of linear, nonlinear, Gaussian and non-Gaussian state space models that include at least a component that is unobservable. The UC itself can either be a single continuous component as in section 3 or we can both have a continuous along with a discrete unobserved component, see section 4. Therefore, our models should not be confused solely with the UC model of Stock and Watson (2007).

We apply a relatively new tool in the family of sequential Monte Carlo methods which is particularly useful for inference in UC models, namely, particle Gibbs with ancestor sampling (PG-AS), suggested in Lindsten et al. (2012). PG-AS builds on the particle Gibbs (PG) sampler proposed by Andrieu et al. (2010). In PG, we start by running a sequential Monte Carlo (SMC) sampler in which one particle trajectory is set deterministically to a reference trajectory that is specified a priori. After a complete run of the SMC algorithm, a new trajectory is obtained by selecting one of the particle trajectories with probabilities given by their importance weights. The effect of the reference trajectory is that the target distribution of the resulting Markov kernel remains invariant, regardless of the number of particles used in the underlying SMC algorithm. However, PG suffers from a serious drawback, which is that the underlying mixing can be very poor when there is path degeneracy in the SMC sampler. In some cases this problem can be addressed by adding a backward simulation step to the PG sampler, yielding a method denoted as PG with *backward simulation*, see for instance Lindsten and Schön (2013). PG-AS alleviates the path degeneracy problem in a very computationally elegant fashion. Specifically, the original PG kernel is modified using a so-called *ancestor sampling* step. This way the same effect as backward sampling is achieved, but without an explicit backward pass.

In this paper we aim to show that PG-AS provides a very compelling and computationally fast framework for estimating rather advanced econometric models. We start by reproducing the results of Chan and Hsiao (2013) which is a book chapter in Bayesian Inference in the Social Sciences published by John Wiley & Sons. The aforementioned article uses “pure” Gibbs sampling methods to estimate different stochastic volatility model specifications. We briefly summarize their applications, estimate the same models using PG-AS and compare our results with the results of Chan and Hsiao (2013). The main difference between Gibbs sampling and PG-AS is that where in traditional Gibbs sampling problems we resort to converting nonlinear or non-Gaussian models to linear and Gaussian state space models in order to draw the latent states, in the PG-AS framework we draw these latent states directly using the nonlinear or non-Gaussian framework.

We also provide extensions where we combine discrete structural breaks within the UC framework using both simulated and macroeconomic time-series data. Structural breaks are modeled

through irreversible Markov switching or so-called change-point dynamics, see Chib (1998). For instance, we model time series characteristics of postwar US inflation using a long memory autoregressive fractionally integrated moving average model with stochastic volatility where we allow for structural breaks in the level, autoregressive (AR), moving average (MA) parameters, long memory parameter,  $d$  contemporaneously with breaks in the level, persistence and the conditional volatility of the volatility of inflation.

Overall, we believe that applying PG-AS to unobserved component models, especially structural break specifications is the most important contribution that we make as to our knowledge there has not yet been any attempts made to use PG-AS in the econometric analysis of these types of models. As we shall see for these type of models, PG-AS requires limited design effort on the user's part especially if one desires to change some features in a particular model. On the other hand, estimating the same type of models using "pure" Gibbs sampling would require relatively more programming effort.

The remaining of this paper is as follows: in section 2 we briefly describe the steps of PG-AS. In sections 3 and 4 we present our empirical applications. Finally, the last section concludes.

## 2 Particle Gibbs with Ancestor Sampling

Consider the following stochastic volatility (SV) model

$$y_t = \mu + \exp(h_t/2) \varepsilon_t, \quad \varepsilon_t \sim N(0, 1) \quad (2.1)$$

$$h_t = \mu_h + \phi_h (h_{t-1} - \mu_h) + \zeta_t, \quad \zeta_t \sim N(0, \sigma_h^2) \quad (2.2)$$

where  $y_t$  is the observed data,  $h_{1:T} = (h_1, \dots, h_T)'$  are the unobserved log-volatilities,  $\mu_h$  is the drift term in the state equation,  $\sigma_h$  is the volatility of log-volatility and  $\phi_h$  is the persistence parameter. Typically, we would impose that  $|\phi_h| < 1$  so that we have a stationary process with the initial condition,  $h_1 \sim N(\mu_h, \sigma_h^2 / (1 - \phi_h^2))$ . We collect the model parameters in  $\theta = (\mu, \mu_h, \phi_h, \sigma_h^2)'$  and let  $Y_T = (y_1, \dots, y_T)'$ . The above stochastic volatility model is an example of a nonlinear state space model where the measurement equation, (2.1) is nonlinear in the states,  $h_{1:T}$ . The major challenge of estimating this model is that while sampling  $p(\theta | h_{1:T}, Y_T)$  is relatively easy, sampling  $h_{1:T} \sim p(h_{1:T} | \theta, Y_T)$  is often difficult.

Within the Gibbs sampling framework, the most popular approach for estimating (2.1)-(2.2) is the so-called auxiliary mixture sampler, see Kim et al. (1998). The idea is to approximate the nonlinear stochastic volatility model using a mixture of linear Gaussian models. Specifically, we can square both sides of (2.1) and take the logarithm such that  $y_t^* = h_t + \varepsilon_t^*$  where  $y_t^* = \log(y_t - \mu)^2$  and  $\varepsilon_t^* = \log \varepsilon_t^2$ . Kim et al. (1998) show that  $\varepsilon_t^*$  can be approximated by a seven-component Gaussian mixture density. We can then write this mixture density in terms of an auxiliary random

variable,  $z_t \in \{1, \dots, 7\}$  that serves as a mixture component indicator. Hence,  $\varepsilon_t^* | z_t \sim N(m_i, s_i^2)$  with  $p(z_t = i) = \omega_i$ . The values of  $m_i$ ,  $s_i^2$  and  $\omega_i$ ,  $i = 1, \dots, 7$  are all fixed and given in Kim et al. (1998). Using this Gaussian mixture approximation the SV model can be expressed as a linear Gaussian state space model. Bayesian estimation can then be performed using standard Gibbs sampling techniques for linear Gaussian state space models, see for instance Kim and Nelson (1999). Finally, notice that using this specification we sample from the posterior,  $p(\theta, h_{1:T}, z_1, \dots, z_T | Y_T)$  augmented to include  $z_1, \dots, z_T$  and not from  $p(\theta, h_{1:T} | Y_T)$ .

Within the PG-AS framework we need not use the above approximation. On the contrary, we approach estimating (2.1)-(2.2) directly by first drawing  $h_{1:T} \sim p(h_{1:T} | \theta, Y_T)$  using the conditional particle filter with ancestor sampling. Thereafter, we draw  $p(\theta | h_{1:T}, Y_T)$ . Notice that once we obtain  $h_{1:T}$ , sampling each element of  $\theta$  is straightforward. In the following we describe the steps of the conditional particle filter with ancestor sampling (CPF-AS) which is used to draw  $h_{1:T}$  from  $p(h_{1:T} | \theta, Y_T)$ . For more details the reader is referred to Lindsten et al. (2012). Let  $i = 1, \dots, N$  denote the number of Gibbs sampling iterations,  $j = 1, \dots, M$  denote the number of particles and let  $p(y_t | \theta, h_t, Y_{t-1})$  denote the density of  $y_t$  given  $\theta$ ,  $h_t$  and  $Y_{t-1}$ . Finally, let  $h_{1:T}^{(i-1)}$  be a fixed reference trajectory of  $h_{1:T}$  sampled at iteration  $i - 1$  of the Gibbs sampler. The steps of CPF-AS (particle filter conditional on  $h_{1:T}^{(i-1)}$ ) for the SV model are as follows

1. **if**  $t = 1$

(a) Draw  $h_1^{(j)} | h_0^{(j)}, \theta$  for  $j = 1, \dots, M - 1$  and set  $h_1^M = h_1^{(i-1)}$ .

(b) Set  $w_1^{(j)} = \tau_1^{(j)} / \sum_{k=1}^M \tau_1^{(k)}$  where  $\tau_1^{(j)} = p(y_1 | \theta, h_1^{(j)}, Y_0)$  for  $j = 1, \dots, M$ .

2. **else for**  $t = 2$  **to**  $T$  **do**

(a) Resample  $\{h_{t-1}^{(j)}\}_{j=1}^{M-1}$  using indices  $a_t^{(j)}$  where  $p(a_t^{(j)} = j) \propto w_{t-1}^{(j)}$  for  $j = 1, \dots, M - 1$ .

(b) Draw  $h_t^{(j)} | h_{t-1}^{(a_t^{(j)})}, \theta$  for  $j = 1, \dots, M - 1$ .

(c) Set  $h_t^M = h_t^{(i-1)}$ .

(d) Draw  $a_t^{(M)}$  from  $p(a_t^{(M)} = j) \propto w_{t-1}^{(j)} p(h_t^{(i-1)} | h_{t-1}^{(j-1)}, \theta)$ .

(e) Set  $h_{1:t}^{(j)} = \left( h_{1:t-1}^{(a_t^{(j)})}, h_t^{(j)} \right)$ ,  $w_t^{(j)} = \tau_t^{(j)} / \sum_{k=1}^M \tau_t^{(k)}$  where  $\tau_t^{(j)} = p(y_t | \theta, h_t^{(j)}, Y_{t-1})$ .

3. **end for**

4. Sample  $h_{1:T}^{(i)} | \theta, Y_T$  with  $p(h_{1:T}^{(i)} = h_{1:T}^{(j)} | \theta, Y_T) \propto w_T^{(j)}$ .

Notice that CPF-AS is akin to a standard particle filter, but with the difference that  $h_{1:T}^M$  is specified a priori and serves as a reference trajectory. Hence, at each step we use only  $M - 1$  particles. Furthermore, whereas in the particle Gibbs algorithm of Andrieu et al. (2010) we set  $a_t^{(M)} = M$ , in PG-AS we sample a new value for the index variable,  $a_t^{(M)}$  in an *ancestor sampling* step, (d).

Even though this is a small modification of the algorithm, improvements in mixing can be quite considerable, see Lindsten et al. (2012). Once we sample  $h_{1:T}$ , we use standard methods and sample each element of  $\theta$ , see for instance Kim et al. (1998) and Chan and Hsiao (2013).

Extending the standard SV model within the PG-AS framework is straightforward. For example, assume that  $\varepsilon_t \sim St(v)$ , where  $St$  stands for the Student-t distribution with  $v > 2$  degrees of freedom. For this specification, at the  $i$ th iteration of the Gibbs sampler for  $p(y_t | \theta, h_t, Y_{t-1})$  we can use

$$p(y_t | \theta, h_t, Y_{t-1}) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2}) \sqrt{(v-2)\pi} \sigma_t} \frac{1}{\sigma_t} \left( 1 + \frac{(y_t - \mu)^2}{(v-2)\sigma_t^2} \right)^{-(v+1)/2} \quad (2.3)$$

inside the CPF-AS algorithm and obtain  $h_{1:T}^{(i)}$ . We then sample  $\mu^{(i)}$ ,  $\mu_h^{(i)}$ ,  $\phi_h^{(i)}$  and  $\sigma_h^{2(i)}$  as in Chan and Hsiao (2013). Finally, in order to draw  $v^{(i)} \sim p(v | \mu^{(i)}, h_{1:T}^{(i)}, Y_T)$  we use (2.3) and perform Metropolis-Hastings (M-H). Hence, we generate a candidate  $v^*$  from  $q(v) \sim TN_{]2, \infty[}(\hat{v}_{ML}, V)$  where  $TN_{]2, \infty[}$  stands for the truncated Normal density on the domain  $]2, \infty[$  and  $\hat{v}_{ML}$  is obtained by maximizing (2.3) with respect to  $v$  using the already obtained values of  $h_{1:T}^{(i)}$  and  $\mu^{(i)}$ . We set  $V = c \cdot \text{var}(\hat{v}_{ML})$ , where  $c \in \mathbb{R}^+$  and fine-tune  $V$  by adjusting  $c$  such that we can get a decent M-H acceptance ratio around 50 to 60%. We then draw  $u$  from a standard Uniform distribution,  $U$  and accept  $v^*$ , i.e.  $v^{(i)} = v^*$  if  $a_{\text{MH}}(v^*, v^{(i-1)}) > u$  where

$$a_{\text{MH}}(v^*, v^{(i-1)}) = \min \left\{ 1, \frac{p(v^* | \mu^{(i)}, h_{1:T}^{(i)}, Y_T) q(v^{(i-1)})}{p(v^{(i-1)} | \mu^{(i)}, h_{1:T}^{(i)}, Y_T) q(v^*)} \right\} \quad (2.4)$$

else we set  $v^{(i)} = v^{(i-1)}$ . On the other hand, if we were to use pure Gibbs sampling to estimate the SV model with Student-t distributed errors then we would be forced to follow Chan and Hsiao (2013) and convert the model into a conditionally Gaussian state space model by defining the measurement error in (2.1) as  $\varepsilon_t = \lambda_t^{-1/2} e_t$  where  $e_t \sim N(0, 1)$ ,  $\lambda_t \sim IG(v/2, v/2)$  and has a closed form conditional posterior. We would then follow the steps in Chan and Hsiao (2013) and sample from the augmented posterior,  $p(\theta, v, h_{1:T}, \lambda_1, \dots, \lambda_T, z_1, \dots, z_T | Y_T)$ .

## 2.1 Model comparison using the output from PG-AS

One of the main outputs from CPF-AS is the loglikelihood of  $Y_T$  with  $h_{1:T}$  integrated out,  $p(Y_T | \theta)$ . This quantity is the product of the individual *integrated* likelihood contributions

$$p(Y_T | \theta) = \prod_{t=1}^T p(y_t | \theta, Y_{t-1}) \quad (2.5)$$

(2.5) can be used to for instance compute the marginal likelihood (ML) for a particular model. The marginal likelihood is defined as

$$p(Y_T) = \int_{\Theta} p(Y_T | \theta) p(\theta) d\theta \quad (2.6)$$

and is a measure of the success the model has in accounting for the data after the parameter uncertainty has been integrated out over the prior,  $p(\theta)$ .

Gelfand and Dey (1994) propose a very compelling and general method to calculate ML. It is efficient and utilizes the same routines when calculating ML for different models. The Gelfand-Dey (G-D) estimate of the marginal likelihood is based on

$$\frac{1}{N} \sum_{i=1}^N g(\theta^{(i)}) / [p(Y_T | \theta^{(i)}) p(\theta^{(i)})] \rightarrow p(Y_T)^{-1} \quad \text{as } N \rightarrow \infty \quad (2.7)$$

where  $N$  is the number of PG-AS iterations. Gelfand and Dey (1994) show that if  $g(\theta^{(i)})$  is thin-tailed relative to  $p(Y_T | \theta^{(i)}) p(\theta^{(i)})$  then (2.7) is bounded and the estimator is consistent. Following Geweke (2005) a truncated Normal distribution,  $N(\theta^*, \Sigma^*)$  is used for  $g(\theta)$ . The quantities  $\theta^*$  and  $\Sigma^*$  are the posterior sample moments calculated as  $\theta^* = N^{-1} \sum_{i=1}^N \theta^{(i)}$  and  $\Sigma^* = N^{-1} \sum_{i=1}^N (\theta^{(i)} - \theta^*) (\theta^{(i)} - \theta^*)'$  whenever  $\theta^{(i)}$  is in the domain of the truncated Normal. This domain,  $\Theta$  is defined as

$$\Theta = \left\{ \theta : (\theta^{(i)} - \theta^*)' (\Sigma^*)^{-1} (\theta^{(i)} - \theta^*) \leq \chi_{\alpha}^2(z) \right\}$$

where  $z$  is the dimension of the parameter vector and  $\chi_{\alpha}^2(z)$  is the  $\alpha$ th percentile of the Chi-squared distribution with  $z$  degrees of freedom. In practice, 0.5, 0.75, 0.95 and 0.99 are popular selections for  $\alpha$ . Once the marginal likelihood for different specifications has been calculated, we can compare them using Bayes factors, BF. The relative evidence for model  $\mathcal{M}_A$  versus  $\mathcal{M}_B$  is  $BF_{\mathcal{M}_A \mathcal{M}_B} = p(Y_T | \mathcal{M}_A) / p(Y_T | \mathcal{M}_B)$ . This odds ratio is the factor by which the data considers  $\mathcal{M}_A$  more probable than  $\mathcal{M}_B$ . Kass and Raftery (1995) recommend considering twice the logarithm of the Bayes factor for model comparison and suggest a rule-of-thumb of support for  $\mathcal{M}_A$  based on  $2 \log BF_{\mathcal{M}_A \mathcal{M}_B}$ : 0 to 2 not worth more than a bare mention, 2 to 6 positive, 6 to 10 strong, and greater than 10 as very strong.

We can also use  $p(Y_T | \theta)$  and compute the deviance information criterion (DIC) of Spiegelhalter et al. (2002). DIC is a compelling alternative to AIC or BIC and it can be applied to nested or non-nested models. Calculation of DIC in a PG-AS scheme is trivial. Contrary to AIC or BIC it does not require maximization over the parameter space. DIC is a combination

of  $p(Y_T | \theta)$  and a penalty term,  $p_D$  which describes the complexity of the model and serves as a penalization term that corrects deviance's propensity toward models with more parameters. More precisely,  $p_D = \overline{D(\theta)} - D(\bar{\theta})$  where  $\overline{D(\theta)}$  is approximated by  $N^{-1} \sum_{i=1}^N -2 \log p(Y_T | \theta^{(i)})$  and  $D(\bar{\theta}) = -2 \log p(Y_T | \bar{\theta})$ .  $\bar{\theta}$  is estimated from the PG-AS output using the mean or mode of the posterior draws. The DIC is defined as  $D(\bar{\theta}) + 2p_D$ .

It is worth noting that the best model is the one with the smaller DIC. Very roughly, for differences of more than 10 we might definitely rule out the model with the higher DIC.

### 3 Stochastic Volatility Models

In this section we start by estimating the basic stochastic volatility model, (2.1)-(2.2) using PG-AS. Thereafter, we extend the SV model to allow for moving average errors and heavy tails. Furthermore, for each specification we compare our results to the results of Chan and Hsiao (2013)<sup>1</sup>.

We use the exact data and prior settings as Chan and Hsiao (2013). Thus, we ought to be able to compare the performance of PG-AS with traditional Gibbs sampling in a rather easy and intuitive way. As stated earlier, the main difference between the Gibbs sampling approach of Chan and Hsiao (2013) and PG-AS is that where Chan and Hsiao (2013) resort to converting nonlinear or non-Gaussian models to linear and Gaussian state space models in order to draw  $h_{1:T}$ , in the PG-AS framework we draw  $h_{1:T}$  directly from  $p(h_{1:T} | \theta, Y_T)$  using the nonlinear or non-Gaussian framework. In the following we briefly summarize their applications labeled as (1), (2), (3) and reproduce their results using PG-AS:

(1) Modeling daily AUD/USD exchange rate returns from January 2005 to December 2012 using (2.1)-(2.2).

(2) Modeling daily PHP/USD exchange rate returns from July 2007 to December 2012 using a moving average stochastic volatility (MASV) model. This model is defined as

$$\begin{aligned} y_t &= \mu + e^{h_t/2} \varepsilon_t + \psi_1 e^{h_{t-1}/2} \varepsilon_{t-1}, & \varepsilon_t &\sim N(0, 1) \\ h_t &= \mu_h + \phi_h (h_{t-1} - \mu_h) + \zeta_t, & \eta_t &\sim N(0, \sigma_h^2) \end{aligned}$$

where  $\varepsilon_t$  is independent from  $\zeta_t$ . Furthermore, for identification, we impose that the root of the characteristic polynomial associated with the MA coefficient,  $\psi_1$  is outside the unit circle. Compared to (2.1)-(2.2) we have an additional parameter,  $\psi_1$ . The conditional posterior of  $\psi_1$  does not have a closed form solution. Therefore, we sample  $\psi_1$  using Metropolis-Hastings. We follow Chan and Hsiao (2013), choose the prior for  $\psi_1$  as a standard Normal truncated in the interval  $] -1, 1[$  and employ rejection sampling to ensure that  $\psi_1^{(i)} \in ] -1, 1[$ .

<sup>1</sup>The article along with some additional material is available at <http://people.anu.edu.au/joshua.chan/research.html>.



(3) Modeling daily returns on the silver spot price from January 2005 to December 2012 using a moving average stochastic volatility model with Student-t distributed errors, MASVt

$$\begin{aligned} y_t &= \mu + e^{h_t/2} \varepsilon_t + \psi_1 e^{h_{t-1}/2} \varepsilon_{t-1}, \quad \varepsilon_t \sim St(v) \\ h_t &= \mu_h + \phi_h (h_{t-1} - \mu_h) + \zeta_t, \quad \eta_t \sim N(0, \sigma_h^2) \end{aligned}$$

For each of these models we use the exact same prior hyperparameter values as Chan and Hsiao (2013) to sample  $\theta$  and  $\psi_1$  from their respective conditional posteriors. With regards to  $v$ , we choose  $p(v) \sim U(2, 128)$ , where  $U$  stands for the Uniform distribution with lower (upper) endpoint of 2(128) and sample  $v$  using (2.4), see also Chib et al. (2002). Finally, since  $p(Y_T | \theta)$  is easily available from the particle filter, we also report the logarithm of the marginal likelihood,  $\log(\text{ML})$  and DIC. We set  $M = 100$  and follow Chan and Hsiao (2013) obtaining  $N = 20000$  draws from the posterior distribution, after a burn-in period of 1000. The PG-AS estimates along with the original results of Chan and Hsiao (2013) are reported in Tables 1 to 3. Overwhelmingly, these estimates are very similar to the Gibbs sampling estimates of Chan and Hsiao (2013), referred to and labeled as ‘‘MCMC’’ in the text below.

In the top row of Figure 1 we compare posterior estimates of  $\exp(h_t/2)$ ,  $t = 1, \dots, T$  for the SV model using AUD/USD returns. The left-hand-side displays the posterior mean (solid line) and 90% credibility intervals (dashed lines) for PG-AS while the right-hand-side displays the same quantities for the Gibbs sampling algorithm of Chan and Hsiao (2013). We see that these estimates are basically identical. To further compare the performance of the samplers, we also report the inefficiency factors,  $R_B$  of the posterior draws of  $\theta$ ,  $\psi_1$ ,  $v$ ,  $h_{1:T}$  for PG-AS and Chan and Hsiao (2013).  $R_B$  is defined as

$$R_B = 1 + \frac{2B}{B-1} \sum_{l=1}^B K\left(\frac{l}{B}\right) \hat{\rho}(l)$$

where  $\hat{\rho}(l)$  is an estimate of the autocorrelation at lag  $l$  of the sampler,  $B$  is the bandwidth and  $K$  is the Parzen Kernel, see also Kim et al. (1998) for a further background on this measure.  $R_B$  displays the relative variance of the posterior sample draws when adapting for correlation between iterations, as compared to the variance without accounting for correlation and serves as a useful diagnostic for measuring how well the chain mixes. In these calculations, we choose a bandwidth,  $B$  of 100. Overall, we see that both methods perform similarly for  $\theta$ ,  $\psi_1$  and  $v$  as  $R_B$  does not vary that much across estimation methods. However, note that for  $h_{1:T}$  each vector is of length  $T$ , so we have a total of  $2T$  inefficiency factors for each application. Therefore, we use box plots to report this information. In panels (c) and (d) of Figure 1 we report results for AUD/USD with similar results for the other applications. The middle line of the box denotes the median, while the lower

Table 1: Posterior mean (mean), standard deviation (std. dev), 5%-tile and 95%-tile of the model parameters, AUD/USD daily returns data

Parameter	mean	std. dev	5%-tile	95%-tile	$R_B$
Original results, Chan and Hsiao (2013)					
$\mu$	-0.029	0.013	-0.051	-0.006	1.496
$\mu_h$	-0.748	0.351	-1.275	-0.229	1.094
$\phi_h$	0.989	0.004	0.982	0.995	12.248
$\sigma_h^2$	0.017	0.003	0.012	0.023	43.519
Reproducing original results using PG-AS					
$\mu$	-0.029	0.013	-0.052	-0.007	1.567
$\mu_h$	-0.713	0.354	-1.239	-0.197	1.154
$\phi_h$	0.989	0.004	0.982	0.996	13.147
$\sigma_h^2$	0.018	0.003	0.012	0.024	47.231
log(ML)	-2336.395				
$\alpha=0.99$					
DIC	4640.389				

$R_B$ : inefficiency factors (using a bandwidth  $B$  of 100). log(ML): logarithm of the marginal likelihood for the corresponding value of  $\alpha$ . DIC: deviance information criterion.

Table 2: Posterior mean (mean), standard deviation (std. dev), 5%-tile and 95%-tile of the model parameters, PHP/USD daily returns data

Parameter	mean	std. dev	5%-tile	95%-tile	$R_B$
Original results, Chan and Hsiao (2013)					
$\mu$	-0.009	0.010	-0.024	0.007	1.684
$\mu_h$	-2.187	0.290	-2.589	-1.793	1.340
$\phi_h$	0.983	0.008	0.970	0.994	29.688
$\sigma_h^2$	0.018	0.004	0.012	0.025	48.748
$\psi_1$	0.141	0.026	0.098	0.184	2.115
Reproducing original results using PG-AS					
$\mu$	-0.007	0.009	-0.023	0.008	1.352
$\mu_h$	-2.054	0.521	-2.667	-1.319	3.869
$\phi_h$	0.986	0.007	0.972	0.997	31.392
$\sigma_h^2$	0.019	0.005	0.012	0.031	50.455
$\psi_1$	0.141	0.025	0.098	0.183	1.991
log(ML)	-554.076				
$\alpha=0.99$					
DIC	1075.639				

$R_B$ : inefficiency factors (using a bandwidth  $B$  of 100). log(ML): logarithm of the marginal likelihood for the corresponding value of  $\alpha$ . DIC: deviance information criterion.

Table 3: Posterior mean (mean), standard deviation (std. dev), 5%-tile and 95%-tile of the model parameters, silver spot price daily returns data

Parameter	mean	std. dev	5%-tile	95%-tile	$R_B$
Original results, Chan and Hsiao (2013)					
$\mu$	0.125	0.037	0.064	0.186	2.262
$\mu_h$	1.102	0.232	0.744	1.436	3.879
$\phi_h$	0.984	0.006	0.974	0.993	20.469
$\sigma_h^2$	0.017	0.005	0.011	0.024	51.696
$\psi_1$	-0.088	0.022	-0.120	-0.050	4.498
$\nu$	6.675	1.235	5.071	8.894	6.153
Reproducing original results using PG-AS					
$\mu$	0.123	0.035	0.065	0.182	3.335
$\mu_h$	1.422	0.252	1.036	1.781	1.933
$\phi_h$	0.985	0.005	0.975	0.993	21.914
$\sigma_h^2$	0.016	0.003	0.011	0.023	50.697
$\psi_1$	-0.091	0.022	-0.108	-0.034	2.269
$\nu$	6.469	0.853	5.201	7.993	6.753
log(ML)	-4574.536				
$\alpha=0.99$					
DIC	9108.228				

$R_B$ : inefficiency factors (using a bandwidth  $B$  of 100). log(ML): logarithm of the marginal likelihood for the corresponding value of  $\alpha$ . DIC: deviance information criterion.

and upper lines represent respectively the 25% and 75%-tiles. The whiskers extend to the maximum and minimum. For example, the box plot associated with  $h_{1:T}$  for PG-AS (MCMC) indicates that about 75% of the log-volatilities have inefficiency factors less than 1.6 (3.75), and the maximum is close to 2.1 (4.8). Overall, results suggest that PG-AS is quite efficient in terms of producing posterior draws of  $h_{1:T}$  that are not highly autocorrelated. Furthermore, we see that PG-AS provides better mixing compared to MCMC for  $h_{1:T}$  as  $R_B$  is on average lower.

[Figure1 about here]

Overall, with regards to reproducing the results of Chan and Hsiao (2013) we find that the only minor difference between PG-AS and Gibbs sampling estimates is in  $\mu_h$  for the MASVt model as PG-AS estimates  $\mu_h$  at a higher rate. However, this does not affect any qualitative conclusions drawn from the results. In order to explore this difference we perform a Monte Carlo analysis. We consider two data generating processes (DGP), one for a SVt model and one for a MASVt model. We generate 100 data sets of  $T = 1000$  observations for each case, re-estimate, compare the estimated parameters of PG-AS and MCMC with the parameters under the DGP. We choose the DGP parameters close to the estimates that we obtain using the silver data. With regards to MCMC

estimation we use the codes of the authors. For each DGP we first generate  $h_{1:T}$  through (2.2) with  $h_1 \sim N(\mu_h, \sigma_h^2 / (1 - \phi_h^2))$ . Thereafter, we use  $\varepsilon_t \sim St(\nu)$  and generate  $y_t$ . Given a full PG-AS and MCMC run we calculate the mean, median and mode of the posterior draws,  $\theta^{(i)}$ ,  $i = 1, \dots, N$ . We then take the mean of these quantities over the number of Monte Carlo repetitions. We also consider the root mean squared error (RMSE), defined as

$$RMSE = \sqrt{\frac{1}{100} \frac{1}{N} \sum_{r=1}^{100} \sum_{i=1}^N (\theta_r^{(i)} - \theta)^2}$$

where  $\theta_r^{(i)}$  is the  $i$ th posterior draw of the  $r$ th Monte Carlo repetition and  $\theta$  is the vector of the true DGP parameters. Results are summarized in Table 4. Overall, both methods work very well. However, in both cases we notice reductions in the RMSE of  $\mu_h$  by about 75% when we use PG-AS compared to MCMC. More importantly, as for the silver data, in the Monte Carlo repetitions we (correctly) estimate  $\mu_h$  at a higher rate using PG-AS compared to MCMC. Furthermore, we also see a slight reduction in the RMSE of  $\nu$  for PG-AS compared to MCMC. Thus, Monte Carlo results show that we can be very confident with regards to using PG-AS. We also repeat this Monte Carlo for a plain SV model and report the results in the bottom part of Table 4. Here, we do not find any differences worth mentioning.

Table 4: Monte Carlo results

DGP	true	MCMC	PG-AS	MCMC	PG-AS	MCMC	PG-AS	MCMC	PG-AS
		mean		median		mode		RMSE	
SVt									
$\mu_h$	1.1	0.787	0.998	0.793	1.009	-1.639	-1.703	0.393	0.224
$\phi_h$	0.98	0.967	0.967	0.969	0.969	0.967	0.977	0.021	0.022
$\sigma_h^2$	0.018	0.022	0.024	0.021	0.023	0.009	0.010	0.005	0.008
$\nu$	8	12.017	10.464	10.638	9.980	16.981	10.720	8.930	8.393
MASVt									
$\mu$	0.1	0.098	0.097	0.098	0.097	0.068	0.093	0.038	0.037
$\mu_h$	1.1	0.787	0.996	0.792	1.007	-1.869	-1.775	0.391	0.222
$\phi_h$	0.98	0.965	0.966	0.969	0.968	0.963	0.976	0.032	0.031
$\sigma_h^2$	0.018	0.022	0.024	0.021	0.022	0.009	0.009	0.005	0.007
$\psi_1$	-0.09	-0.089	-0.088	-0.089	-0.088	-0.095	-0.093	0.033	0.033
$\nu$	8	12.141	10.766	10.689	10.203	19.065	9.962	8.426	8.215
SV									
$\mu_h$	-0.8	-0.791	-0.755	-0.796	-0.761	-3.090	-5.622	0.190	0.191
$\phi_h$	0.98	0.974	0.973	0.975	0.974	0.970	0.980	0.012	0.013
$\sigma_h^2$	0.017	0.020	0.021	0.019	0.020	0.009	0.009	0.004	0.005

This table reports the true values of the DGP together with the mean, median, mode and root-mean-squared error (RMSE) of the estimated parameters for the Monte Carlo simulations.

Finally, in Figure 2 we plot the marginal posterior densities using the output of PG-AS for the MASVt model using the silver data. The marginal posteriors are bell shaped and centered around the mean. We also report Markov chain output of the model parameters. The chain mixes well with relatively fast decaying autocorrelation functions.

[Figure2 about here]

### 3.1 Sensitivity of PG-AS with respect to $M$

We often find that the choice of  $M$  is important because it ensures that the estimate of  $h_{1:T}$  is not too jittery or imprecise. Therefore, we experiment with different values of  $M$  to find out its effect on estimation results. We do this by considering a more challenging specification than (2.1)-(2.2). From a computational point of view, this extension also allows us to demonstrate the flexibility of PG-AS to adapt to more complicated model structures without any major computational costs.

Specifically, we consider the stochastic volatility in mean (SVM) model of Koopman and Hol Uspensky (2002). Hence, in this specification,  $e^{h_t}$  appears in both the conditional mean and the conditional variance. We follow the same notation as before and define the SVM model as

$$y_t = \mu + \lambda \exp(h_t) + \exp(h_t/2) \varepsilon_t, \quad \varepsilon_t \sim N(0, 1) \quad (3.1)$$

where  $h_t$  follows (2.2). Estimation of this specification is nontrivial using pure Gibbs sampling. This is because drawing  $h_{1:T} \sim p(h_{1:T} | \theta, \lambda, Y_T)$  is computationally more demanding as (3.1) and (2.2) cannot be written in linear state space form, see for instance Chan (2014). However, within the PG-AS context, estimating the SVM model is straightforward. We note that  $p(y_t | \theta, \lambda, h_t, Y_{t-1}) \sim N(\mu + \lambda \exp(h_t), \exp(h_t))$ . Incorporating this specification is very easy in CPF-AS as we only need to modify step (e) of the algorithm and use  $\tau_t^{(j)} = N\left(\mu + \lambda \exp\left(h_t^{(j)}\right), \exp\left(h_t^{(j)}\right)\right)$  instead of  $\tau_t^{(j)} = N\left(\mu, \exp\left(h_t^{(j)}\right)\right)$  in the case of the SV model. The PG-AS sampler then provides us with  $h_{1:T}$  conditional on  $\theta, \lambda$  and  $Y_T$ . Furthermore, we add another layer to our sampler and also sample  $\lambda | \mu, h_{1:T}, Y_T$  from its conditional posterior.

We simulate  $y_t, t = 1, \dots, 1000$  using (3.1)-(2.2). Thereafter, we re-estimate the SVM model using our simulated data for  $M = 2, 10, 100$  and 1000. We report parameter estimates of the SVM model using the above mentioned number of particles in Table 5. Besides these estimates we also report the inefficiency factors ( $R_B$ ) of the parameters and  $h_{1:T}$  for each case, see Figure 3. Furthermore, we compute Geweke's convergence statistics and present estimation time (in seconds) of the SVM model for each  $M$ . In each case, we sample  $N = 20000$  draws from  $p(\theta, \lambda, h_{1:T} | Y_T)$  after a burn-in period of 1000.

Table 5: Sensitivity of the PG-AS sampler with respect to  $M$ 

Parameter	true	mean	std. dev	5%-tile	95%-tile	$R_B$	Geweke
$M = 2$							
$\mu$	0.1	0.102	0.044	0.030	0.222	35.157	-0.843
$\lambda$	0.3	0.381	0.115	0.204	0.751	34.732	0.606
$\mu_h$	-0.8	-0.909	0.270	-1.250	0.621	3.682	-0.791
$\phi_h$	0.97	0.977	0.010	0.960	0.999	30.353	-2.516
$\sigma_h^2$	0.02	0.017	0.004	0.011	0.035	58.685	1.928
log(ML), $\alpha = 0.99$	-972.594						
DIC	1912.139						
Time (seconds)	5005.062						
$M = 10$							
$\mu$	0.1	0.103	0.046	0.026	0.226	24.655	0.428
$\lambda$	0.3	0.377	0.120	0.191	0.743	26.412	-0.227
$\mu_h$	-0.8	-0.908	0.255	-1.240	0.395	1.816	0.685
$\phi_h$	0.97	0.977	0.010	0.959	0.998	17.403	2.523
$\sigma_h^2$	0.02	0.018	0.004	0.012	0.036	42.590	-1.385
log(ML), $\alpha = 0.99$	-971.142						
DIC	1911.540						
Time (seconds)	6849.859						
$M = 100$							
$\mu$	0.1	0.102	0.046	0.027	0.231	21.098	-1.279
$\lambda$	0.3	0.381	0.119	0.194	0.757	22.693	1.344
$\mu_h$	-0.8	-0.914	0.251	-1.245	0.351	1.474	-0.345
$\phi_h$	0.97	0.977	0.010	0.960	0.999	14.932	-0.030
$\sigma_h^2$	0.02	0.017	0.004	0.012	0.035	40.968	0.155
log(ML), $\alpha = 0.99$	-969.906						
DIC	1909.094						
Time (seconds)	7049.614						
$M = 1000$							
$\mu$	0.1	0.103	0.046	0.027	0.222	20.420	0.319
$\lambda$	0.3	0.378	0.121	0.187	0.763	21.768	-0.492
$\mu_h$	-0.8	-0.910	0.262	-1.246	0.622	1.835	1.412
$\phi_h$	0.97	0.976	0.011	0.957	0.999	14.176	0.692
$\sigma_h^2$	0.02	0.018	0.005	0.011	0.039	40.983	-0.478
log(ML), $\alpha = 0.99$	-970.003						
DIC	1909.364						
Time (seconds)	29001.682						

$R_B$ : inefficiency factors (using a bandwidth  $B$  of 100). Geweke: Geweke's convergence statistics. log(ML): logarithm of the marginal likelihood for the corresponding value of  $\alpha$ . DIC: deviance information criterion.

Overall, we see that PG-AS performs very well as parameter estimates are close to their respective DGP values. Furthermore, the choice of  $M = 100$  is very sensible. In fact, we get almost identical results for  $M = 10$  and  $M = 100$  but  $R_B$  falls as we set  $M = 100$ . For instance, in Figure 3, for  $M = 10$ , 75% of  $h_{1:T}$ s have inefficiency factors less than 8 while for  $M = 100$  this number is close to 2. Compared to  $M = 100$  we do not obtain any significant gains in  $R_B$  when we set  $M = 1000$ . However, as  $M$  increases the computation time also increases and from this point of view  $M = 1000$  seems computationally very demanding.

[Figure3 about here]

## 4 Structural Breaks and PG-AS

In this section we demonstrate the flexibility of PG-AS by combing PG-AS with the change-point specification of Chib (1998). Specifically, we believe that this combination is in fact the most important contribution that we make since relatively only few papers have dealt with combining models with SV effects with what in the literature is considered as a “state-of-the-art” structural break model.

We start by using a linear regression Markov switching model with a specific structure on the transition probabilities combined with SV effects. Testing for the number of regimes,  $m \in \{1, 2, \dots\}$  is then a test for the number of structural breaks, see Chib (1998). We follow Liu and Maheu (2008) and estimate each model conditional on  $0, 1, \dots, m$  breaks occurring in the sample. For each of these we calculate ML (DIC) and use them to determine the optimal number of change points. Specifically, we can compare marginal likelihoods using Bayes factors and use differences in the DIC between different specifications to determine the number of structural breaks.

The  $m$ -state MS linear regression model with SV effects which allows for  $m - 1$  structural breaks in the regression coefficients and level of volatility is defined as

$$y_t = X_{t-1}\beta_{s_t} + \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \gamma_{s_t} e^{h_t}\right) \quad (4.1)$$

$$h_t = \phi_h h_{t-1} + \zeta_t, \quad \zeta_t \sim N\left(0, \sigma_h^2\right) \quad (4.2)$$

where  $s_t = 1, \dots, m$ ,  $s_{1:T} = (s_1, \dots, s_T)'$  and  $s_t = k$  indicates that  $y_t$  is from regime  $k$ . In section 4.3 we also consider structural breaks in  $\phi_h$  and  $\sigma_h^2$ . Incorporating breaks in  $\phi_h$  and  $\sigma_h^2$  is straightforward as we can use the conditioning features of PG-AS and thus sample these parameters using the log-volatilities within each regime. As before,  $Y_T = (y_1, \dots, y_T)'$  is  $T \times 1$ ,  $X_T$  is a  $T \times n$  matrix of regressors with row  $X_{t-1}$  which can also include lags of  $y_t$  and  $\beta_{s_t}$  is  $n \times 1$ . The parameter  $\gamma_{s_t} = \exp(\mu_{h,s_t})$  is a measure of the average variance in regime  $k$ . The one-step-ahead transition

probability matrix for  $s_t$  is assumed to be

$$P = \begin{bmatrix} p_{11} & p_{12} & 0 & \cdots & 0 \\ 0 & p_{22} & p_{23} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 0 & p_{m-1,m-1} & p_{m-1,m} \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \quad (4.3)$$

where  $p_{lk} = p(s_t = k \mid s_{t-1} = l)$  with  $k = l$  or  $k = l + 1$  is the probability of moving from regime  $l$  at time  $t - 1$  to regime  $k$  at time  $t$ .  $P$  ensures that given  $s_t = k$  at time  $t$ , the next period  $t + 1$ ,  $s_{t+1}$  remains in the same state or jumps to the next state. For instance, given  $s_t = k$ , one has that  $s_{t+1} = k$  or  $s_{t+1} = k + 1$  with  $p_{k,k} + p_{k,k+1} = 1$ . Once the last regime is reached, we stay there forever, that is  $p_{m,m} = 1$ . This structure enforces the following ordering

$$\theta = \begin{cases} \theta_1 & \text{if } t < \tau_1 \\ \theta_2 & \text{if } \tau_1 \leq t < \tau_2 \\ \vdots & \vdots \\ \theta_{m-1} & \text{if } \tau_{m-2} \leq t < \tau_{m-1} \\ \theta_m & \text{if } \tau_{m-1} \leq t \end{cases}$$

where  $\theta_k = (\beta_k, \gamma_k)'$ ,  $k = 1, \dots, m$ ,  $\theta = (\{\beta_k\}_{k=1}^m, \{\gamma_k\}_{k=1}^m, \phi_h, \sigma_h^2)'$  and  $\tau_1, \tau_2, \dots, \tau_{m-1}$  are the unknown structural break dates. Modeling (4.1)-(4.2) using (4.3) is straightforward as we can use the conditioning features of PG-AS. Specifically, we proceed by cycling through the following steps

1.  $s_{1:T} \mid \beta = \beta_1, \dots, \beta_m, \gamma = \gamma_1, \dots, \gamma_m, \phi_h, \sigma_h^2, P, Y_T$
2.  $h_{1:T} \mid \beta_1, \dots, \beta_m, \gamma_1, \dots, \gamma_m, \phi_h, \sigma_h^2, s_{1:T}, Y_T$
3.  $\beta_1, \dots, \beta_m \mid \gamma_1, \dots, \gamma_m, h_{1:T}, s_{1:T}, Y_T$
4.  $\gamma_1, \dots, \gamma_m \mid \beta_1, \dots, \beta_m, h_{1:T}, s_{1:T}, Y_T$
5.  $\phi_h \mid \sigma_h^2, h_{1:T}$
6.  $\sigma_h^2 \mid \phi_h, h_{1:T}$
7.  $P \mid s_{1:T}$

In step 1, we use the algorithm of Chib (1998) to draw  $s_{1:T}$ , see for instance Liu and Maheu (2008).  $h_{1:T}$  is sampled using CPF-AS from section 2 conditional on the newly draws of  $s_{1:T}$  and  $\theta$  from the



previous Gibbs iteration. The parameters:  $\beta$ ,  $\gamma$ ,  $\phi_h$  and  $\sigma_h^2$  are all sampled using standard techniques conditional on the newly draws of  $s_{1:T}$  and  $h_{1:T}$ . The conditional posterior of  $p_{kk}$ ,  $k = 1, \dots, m-1$  is  $Beta(a_0 + n_{kk}, b_0 + 1)$  where  $n_{kk}$  is the number of one-step transitions from state  $k$  to state  $k$  in a given sequence of  $s_{1:T}$ .

We use  $N(0_n, I_n)$  prior for  $\beta_k$ ,  $\gamma_k \sim IG(\frac{4}{2}, \frac{0.2}{2})$ ,  $\phi_h \sim Beta(20, 1.5)$ ,  $\sigma_h^2 \sim IG(\frac{4}{2}, \frac{0.2}{2})$  where  $IG(\frac{\cdot}{2}, \frac{\cdot}{2})$  stands for the Inverse-gamma density, see Kim and Nelson (1999). Finally, with regards to  $p_{kk}$ , we choose  $p_{kk} \sim Beta(a_0 = 20, b_0 = 0.1)$  for  $k = 1, \dots, m-1$ . We start with a simulation example in section 4.1 and then provide an empirical application using real US GDP growth rate.

## 4.1 Simulation example

Consider data generated according to the following model

$$\begin{aligned} y_t &= 1 + \varepsilon_t, & \varepsilon_t &\sim N\left(0, 1.4e^{h_t}\right) & \text{for } 1 \leq t < 100 \\ y_t &= 0.1 + \varepsilon_t, & \varepsilon_t &\sim N\left(0, 0.2e^{h_t}\right) & \text{for } t \geq 100 \end{aligned}$$

where  $h_t = 0.9h_{t-1} + \zeta_t$  and  $\zeta_t \sim N(0, 0.02)$  for  $t = 1, \dots, 200$ . We estimate the above model conditional on  $m = 0, 1, 2$  change points and determine the optimal number of change points using ML and DIC. Furthermore, notice that in order to obtain  $p(Y_T | \theta, P)$  we use that

$$\log p(Y_T | \theta, P) = \sum_{t=1}^T \log p(y_t | \theta, P, Y_{t-1})$$

where

$$p(y_t | \theta, P, Y_{t-1}) = \sum_{k=1}^m p(y_t | \theta, Y_{t-1}, s_t = k) p(s_t = k | \theta, P, Y_{t-1})$$

The first term,  $p(y_t | \theta, Y_{t-1}, s_t = k)$  is obtained from CPF-AS using  $\theta_k = (\beta_k, \gamma_k)'$ ,  $\phi_h$ ,  $\sigma_h^2$  and the last term is computed from

$$\begin{aligned} p(s_t = k | \theta, P, Y_{t-1}) &= \sum_{l=k-1}^k p(s_{t-1} = l | \theta, P, Y_{t-1}) p_{lk}, & k &= 1, \dots, m \\ p(s_t = k | \theta, P, Y_t) &= \frac{p(s_t = k | \theta, P, Y_{t-1}) p(y_t | \theta, Y_{t-1}, s_t = k)}{\sum_{l=1}^m p(s_t = l | \theta, P, Y_{t-1}) p(y_t | \theta, Y_{t-1}, s_t = l)}, & k &= 1, \dots, m \end{aligned} \quad (4.4)$$

The last equation is obtained from Bayes' rule. Note that in (4.4) the summation is only from  $k-1$  to  $k$ , due to the restricted nature of the transition matrix.

We run our sampler for 25000 iterations and discard the first 5000 draws. In Table 6 we report

log(ML) and DIC for specifications with no change point up to two change points. As expected, the specification with one change point obtains the highest (lowest) ML (DIC). We also compute the marginal likelihood for the change-point models using the method of Sims et al. (2008) because as pointed out in Sims et al. (2008) the G-D method may not work for models with time-varying parameters as the posterior density tends to be non Gaussian. However, we do not find any significant changes compared to G-D. Therefore, we choose to retain these values. We also conduct a Monte Carlo analysis (not reported) generating data from (4.1)-(4.2) for  $0, \dots, m$  change points and comparing ML between different specifications. Results indicate that the G-D method correctly identifies the true number of structural breaks at almost every Monte Carlo iteration.

Table 6: Change-point stochastic volatility model

# CP	log(ML) $\alpha=0.50$	log(ML) $\alpha=0.75$	log(ML) $\alpha=0.95$	log(ML) $\alpha=0.99$	DIC
0	-266.721	-266.430	-266.291	-266.278	509.786
1	-242.227	-242.303	-242.112	-242.074	445.991
2	-252.231	-251.926	-251.716	-251.677	467.307

This table displays the full sample marginal likelihood and DIC estimates for the change-point stochastic volatility model. # CP is the number of change points that are conditioned on.

We compare the performance of our model (break) to a recursive OLS specification (no-break). The top panel of Figure 4 shows the predictive mean of these models along with  $y_t$ . Both predictive means are similar before the break at  $t = 100$ . However, after the break, we see a quick reduction in the predictive mean from the break model while the predictive mean from the no-break model remains high for a long time. In the top right panel, we plot the density of the change point along with the true position of the change point. Clearly, our model is able to detect the date of the structural break. We also compare our estimates of  $\sigma_t^2 = \gamma_{s_t} \exp(h_t)$ ,  $t = 1, \dots, T$  with the true conditional variance process. In panel (f), we plot the marginal posterior densities of  $\gamma_1$  and  $\gamma_2$ . The marginal posteriors are bell shaped and centered around their means (1.34 for  $\gamma_1$  and 0.15 for  $\gamma_2$ ). We also report the Markov chain output and autocorrelation function for  $\gamma_1 | Y_T$  and  $\gamma_2 | Y_T$ . The chain mixes well with relatively fast decaying autocorrelation functions.

[Figure4 about here]

Finally, Table 7 displays out-of-sample results for one-period-ahead direct forecasts (see Marcellino et al. (2005)) for the no-break and the break model. In the context of forecasting with the break model we perform the following: for the first out-of-sample observation at time  $t$  we calculate ML for 1 and 2 change points using  $Y_{t-1}$ . Thereafter, we choose the optimal change-point number,  $n_1$  using Bayes factors and calculate the predictive mean,  $E[y_t | Y_{t-1}]$  using the parameters associated with specification  $n_1$ . Thereafter, we increase the out-of-sample with one observation, calculate

ML for 1 and 2 change points, choose the optimal  $n_2$  and repeat the above forecasting procedure to obtain  $E[y_{t+1} | Y_t]$ . Furthermore, in addition to MAE and RMSE, forecasts are also compared using the linear exponential (LINEX) loss function of Zellner (1986). This loss function is defined as  $L(y_t, \hat{y}_t) = b[\exp(a(\hat{y}_t - y_t)) - a(\hat{y}_t - y_t) - 1]$ , where  $\hat{y}_t$  is the forecast.  $L(y_t, \hat{y}_t)$  ranks overprediction (underprediction) more heavily for  $a > 0$  ( $a < 0$ ). The table includes  $b = 1$ , with  $a = 1$  and  $a = -1$ .

Table 7: Out-of-sample forecasts

Model	MAE	RMSE	LINEX, $a = 1, b = 1$	LINEX, $a = -1, b = 1$
No-break	0.779	0.916	0.452	0.553
Break	0.528	0.722	0.305	0.340

This table reports mean absolute error (MAE) and root mean squared error (RMSE) for the forecasts based on the predictive mean for one-step-ahead. Furthermore, average LINEX loss function is reported for  $a = 1$ ,  $a = -1$  and  $b = 1$ . The out-of-sample period is from  $t = 51, \dots, 200$ .

Overall, the break model offers improvements in terms of MAE and RMSE compared to the no-break model. When the LINEX loss function is used, the break model's ability to capture variations in higher moments provides gains in terms of point forecasts.

## 4.2 Real output

Recent literature documents a structural break in the volatility of US GDP growth rate, see for instance Kim and Nelson (1999) and Gordon and Maheu (2008). In this section we follow Gordon and Maheu (2008) and consider structural break estimates of AR(2) models in real US GDP growth rate using the framework of (4.1)-(4.2).

Let  $y_t = 100[\log(q_t/q_{t-1}) - \log(p_t/p_{t-1})]$ , where  $q_t$  is quarterly US GDP seasonally adjusted and  $p_t$  is the US GDP price index. Our data ranges from 1947q4 to 2013q3, for a total of 264 observations. In the following we compare the performance of (4.1)-(4.2) with 2 lags of  $y_t$ , CP( $m$ )-AR(2)-SV to the change-point AR(2) model, henceforth CP( $m$ )-AR(2) where  $\sigma^2$  changes with  $s_t$

$$y_t = \beta_{1,s_t} + \beta_{2,s_t}y_{t-1} + \beta_{3,s_t}y_{t-2} + \sigma_{s_t}\varepsilon_t, \quad \varepsilon_t \sim N(0, 1) \quad (4.5)$$

(4.5) is estimated using pure Gibbs sampling, see for instance Liu and Maheu (2008). We estimate (4.1)-(4.2), (4.5) conditional on  $m = 0, 1, 2$  change points and determine the optimal number of change points using ML and DIC. We find that both change-point models produce similar results suggesting that one structural break has occurred. The posterior density of the change point for each model is plotted in Figure 5. They show a very close correspondence between these two models. Using the posterior mode of  $\left\{s_{1:T}^{(i)}\right\}_{i=1}^N$  both models point that the break date is 1983q3, identical to the break date of Gordon and Maheu (2008) and close to the break date of Kim and Nelson (1999).

Specifically, Kim and Nelson (1999) find evidence of a break in 1984q1. However, they consider a sample from 1953q2 to 1997q1.

Overall, we find a significant one-time drop in the volatility of  $y_t$ . For CP(1)-AR(2) the first regime implies an unconditional variance for US GDP growth rate of 1.38 while for the second regime it is 0.35. Furthermore, for CP(1)-AR(2)-SV we find that  $\gamma_2 = 0.27$  is estimated at a lower rate than  $\gamma_1 = 1.19$  indicating a fall in the average variance of real US GDP growth rate since the beginning of the 1980s.

In Figure 5 we also report the estimates of  $\sigma_t^2 = \gamma_{s_t} \exp(h_t)$ ,  $t = 1, \dots, T$  for CP(1)-AR(2)-SV. These estimates are very interesting indeed. Specifically, besides a significant reduction in the volatility of real US GDP growth rate since the 1980s our results also point to a gradual reduction in the volatility of  $y_t$  during the 1960s followed by a subsequent increase from the beginning of the 1970s until the break point at 1983q3.

[Figure5 about here]

### 4.3 Structural break ARFIMA-SV model

In this section we propose a structural break ARFIMA model with SV effects. As before, we use the framework of Chib (1998) to model structural breaks. Specifically, our model allows for structural breaks in  $\mu$ ,  $d$ , autoregressive (AR), moving average (MA) coefficients,  $\gamma$ ,  $\phi_h$  and  $\sigma_h^2$ . Our change-point ARFIMA-SV model is as follows

$$y_t - \mu_{s_t} = \frac{\Psi(L)}{\Phi(L)} (1-L)^{-d_{s_t}} \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \gamma_{s_t} e^{h_t}\right) \quad (4.6)$$

$$h_t = \phi_{h,s_t} h_{t-1} + \zeta_t, \quad \zeta_t \sim N\left(0, \sigma_{h,s_t}^2\right) \quad (4.7)$$

where  $s_t = 1, \dots, m$ ,  $\Phi(L) = (1 - \phi_{1,s_t}L - \dots - \phi_{p,s_t}L^p)$  and  $\Psi(L) = (1 + \psi_{1,s_t}L + \dots + \psi_{q,s_t}L^q)$  are AR and MA polynomials in the lag operator,  $L$  where  $L^p y_t = y_{t-p}$  for  $p = 0, 1, \dots$  with integer orders  $p \geq 0$  and  $q \geq 0$ . The fractional difference operator,  $(1-L)^{-d_{s_t}}$  with  $d_{s_t} \in \mathbb{R}$  is given by

$$(1-L)^{-d_{s_t}} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d_{s_t})}{\Gamma(j+1)\Gamma(d_{s_t})} L^j$$

where  $\Gamma(\cdot)$  is the Gamma function. Equation (4.6) is a generalization of the ARMA model to non-integer values of  $d_{s_t}$ . Specifically, if  $d_{s_t} > 0$  the process is said to have long memory since the autocorrelations die out at an hyperbolic rate. For  $0 < d_{s_t} < 0.5$ , (4.6) is a stationary long-memory process with non-summable autocorrelation functions. For  $d_{s_t} = 0$ , we have a structural break ARMA( $p, q$ ) model with SV effects. We assume that  $0 < d_{s_t} < 0.5$  and that  $\Phi(z) = 0$ ,  $\Psi(z) = 0$  for unknown  $z$  do not have common roots.

In order to estimate (4.6)-(4.7) we rely on the idea of Chan and Palma (1998). Specifically, Chan and Palma (1998) consider an approximation of (4.6) based on a truncation lag of order  $L$ . We follow their framework and proceed to draw  $\theta = \{\theta_k\}_{k=1}^m$  where  $\theta_k = (\mu_k, d_k, \phi_{1,k}, \dots, \phi_{p,k}, \psi_{1,k}, \dots, \psi_{q,k})'$ ,  $\gamma_1, \dots, \gamma_m, \phi_{h,1}, \dots, \phi_{h,m}, \sigma_{h,1}^2, \dots, \sigma_{h,m}^2, P$  and  $L$  from their respective conditional posteriors. However, the conditional posteriors of most parameters do not have closed form solution and are therefore sampled using Metropolis-Hastings.

We note that the conditional posteriors of  $\{\theta_k, \gamma_k, \phi_{h,k}, \sigma_{h,k}^2\}_{k=1}^m$  depend only on information in regime  $k$ . Let  $\hat{Y}_k = \{y_t : s_t = k\}$ ,  $\hat{h}_k = \{h_t : s_t = k\}$  denote observations in regime  $k$  and let  $\theta_k^{-d_k} = (\mu_k, \phi_{1,k}, \dots, \phi_{p,k}, \psi_{1,k}, \dots, \psi_{q,k})'$ . At the  $i$ th iteration of PG-AS we can sample each element of  $\theta_k$  one-at-a-time. For example,  $d_k$  is sampled in the following way:

1. Sample a candidate,  $d_k^*$  from a Gaussian random walk proposal  $q(d_k^* | d_k^{(i-1)}) \sim N(d_k^{(i-1)}, \Sigma_k)$  truncated such that  $0 < d_k^* < 0.5$ .  $\Sigma_k$  is chosen by the researcher in a manner to ensure a sufficient acceptance rate. We follow Koop (2003, page 98) and adjust  $\Sigma_k$  to get an acceptance rate roughly around 30 to 40% by experimenting with different values of  $\Sigma_k$  until we find one which yields a reasonable acceptance rate probability.

2. Define the acceptance probability for  $d_k^*$  as

$$a_{\text{MH}}(d_k^*, d_k^{(i-1)}) = \min \left\{ 1, \frac{p(d_k^* | \theta_k^{-d_k^{(i-1)}}, \gamma_k^{(i-1)}, L^{(i-1)}, \hat{h}_k^{(i)}, \hat{Y}_k) q(d_k^{(i-1)} | d_k^*)}{p(d_k^{(i-1)} | \theta_k^{-d_k^{(i-1)}}, \gamma_k^{(i-1)}, L^{(i-1)}, \hat{h}_k^{(i)}, \hat{Y}_k) q(d_k^* | d_k^{(i-1)})} \right\}$$

3. Draw  $u \sim U(0, 1)$ . If  $u \leq a_{\text{MH}}(d_k^*, d_k^{(i-1)})$  then set  $d_k^{(i)} = d_k^*$  else set  $d_k^{(i)} = d_k^{(i-1)}$ .

We sample  $\phi_{h,k} | \sigma_{h,k}^2, \hat{h}_k$  using M-H, see Kim et al. (1998).  $\sigma_{h,k}^2 | \phi_k, \hat{h}_k \sim IG(v_k/2, l_k/2)$ ,  $v_k = T_k + v_0$ ,  $l_k = \hat{\zeta}_k' \hat{\zeta}_k + l_0$ ,  $T_k$  is the number of observations in regime  $k$  and  $\hat{\zeta}_k = \{\zeta_t : s_t = k\}$ .  $v_0$  and  $l_0$  are prior hyperparameter values.  $\gamma_k | \theta_k, L, \hat{h}_k, \hat{Y}_k \sim IG(r_k/2, g_k/2)$  where  $r_k = T_k + r_0$  and  $g_k = \hat{\epsilon}_k' \hat{\epsilon}_k / \exp(\hat{h}_k) + g_0$ . As before  $p_{kk} | S \sim \text{Beta}(a_0 + n_{kk}, b_0 + 1)$ ,  $k = 1, \dots, m-1$ . Finally, in order to sample  $L$  from its conditional posterior we use the method of Raggi and Bordignon (2012).

We apply our model to a monthly time series of inflation, using the US City Average core consumer price index (CUUR0000SA0L1E) of the Bureau of Labor Statistics (BLS). Our series excludes the direct effects of price changes for food and energy. We denote the series by  $P_t$  and use data from 1960:1 until 2013:12, for a total of 648 months. We follow Bos et al. (2012) and construct the monthly US core inflation as  $\pi_t = 100 \log(P_t/P_{t-1})$ . To adapt for part of the seasonality in the series, we regress the inflation on a series of seasonal dummies,  $D$ , as in  $\pi = D\beta + u$ . Instead of using the original inflation,  $\pi_t$ , we use  $y_t = \hat{u}_t + \bar{\pi}$ , where  $\hat{u}_t$  is the residual of adapting the inflation for the major seasonal effects at time  $t$  and  $\bar{\pi}$  is the average inflation level.

We estimate (4.6)-(4.7) from 0 to 4 change points and choose the optimal number of change points using ML and DIC. Both in terms of ML and DIC results point that the specification with 2 change points fits the data best<sup>2</sup>. However, as pointed out by Spiegelhalter et al. (2002) we must be cautious against using ML as a basis against which to assess DIC. ML addresses how well the prior has predicted the observed data, whereas DIC addresses how well the posterior might predict future data generated by the same parameters that give rise to the observed data. Tables 8 and 9 present estimation results for: a homoscedastic ARFIMA model, ARFIMA-SV and ARFIMA-SV conditional on 2 change points, henceforth labeled as CP(2)-ARFIMA-SV. With regards to model estimation we experiment with different AR and MA lags. We find  $d_{s_t}$ ,  $\phi_{11,s_t}$  and  $\phi_{12,s_t}$  to be significant. Thus, we set  $p(d_{s_t}) \sim N(0, 1)$  truncated such that  $0 < d_{s_t} < 0.5$ ,  $p(\phi_{1,s_t}) \sim N(0, 1), \dots, p(\phi_{12,s_t}) \sim N(0, 1)$  and employ rejection-sampling to ensure that the roots of  $(1 - \phi_{1,s_t}L - \dots - \phi_{12,s_t}L^{12})$  lie outside the unit circle. However,  $\phi_{1,s_t}, \dots, \phi_{10,s_t}$  are not significant in our applications and are therefore fixed at zero. A suitable prior for  $L$  is the Poisson truncated distribution with  $L \in \{L_{min}, \dots, L_{max}\}$  where in this paper  $L_{min} = 10$  and  $L_{max} = 50$ , see Raggi and Bordignon (2012).

Table 8: Full sample estimation results for US core inflation

Parameter	ARFIMA(12,d,0)			ARFIMA(12,d,0)-SV			IMA(1,1)-SV		
	mean	std. dev	$R_B$	mean	std. dev	$R_B$	mean	std. dev	$R_B$
$d_1$	0.384	0.030	7.850	0.299	0.035	8.950			
$\psi_1$							-0.840	0.027	12.547
$\mu_1$	0.171	0.070	7.454	0.115	0.020	15.248			
$\phi_{11,1}$	0.116	0.037	21.052	0.149	0.036	19.757			
$\phi_{12,1}$	0.241	0.038	22.777	0.368	0.038	20.998			
$\sigma_1^2$	0.032	0.001	1.063						
$\gamma_1$				0.025	0.006	32.979	0.028	0.006	32.036
$\phi_{h,1}$				0.975	0.010	16.637	0.965	0.014	17.289
$\sigma_{h,1}^2$				0.035	0.011	39.745	0.040	0.012	38.035
$L$				29.545	4.499	30.984			
log(ML) $\alpha=0.50$	168.895			271.701			224.363		
log(ML) $\alpha=0.75$	168.867			272.105			224.768		
log(ML) $\alpha=0.95$	168.837			272.342			225.004		
log(ML) $\alpha=0.99$	168.828			272.383			225.046		
DIC	-362.286			-521.928			-422.079		

This table reports the posterior mean (mean), standard deviation (std. dev) and inefficiency factors,  $R_B$  with  $B = 100$  for different ARFIMA( $p,d,q$ ) type models. log(ML): logarithm of the marginal likelihood using the corresponding value of  $\alpha$ . DIC: deviance information criterion. The plain ARFIMA( $p,d,q$ ) model is estimated using Gibbs sampling.

<sup>2</sup>As before, we also compute the marginal likelihood for the change-point ARFIMA( $p,d,q$ )-SV models using the method of Sims et al. (2008) and do not find any significant changes compared to G-D.

Table 9: Full sample estimation results for US core inflation, change-point specifications

Parameter	CP(2)-ARFIMA(12,d,0)-SV			CP(2)-IMA(1,1)-SV		
	mean	std. dev	$R_B$	mean	std. dev	$R_B$
$d_1$	0.202	0.069	8.326			
$d_2$	0.417	0.053	5.759			
$d_3$	0.154	0.049	9.130			
$\psi_1$				-0.847	0.042	5.894
$\psi_2$				-0.569	0.109	7.635
$\psi_3$				-0.857	0.063	7.862
$\mu_1$	0.135	0.042	19.157			
$\mu_2$	0.536	0.113	33.068			
$\mu_3$	0.072	0.017	24.069			
$\phi_{11,1}$	0.203	0.081	26.967			
$\phi_{12,1}$	0.302	0.080	23.580			
$\phi_{11,2}$	0.027	0.089	34.852			
$\phi_{12,2}$	0.097	0.093	27.149			
$\phi_{11,3}$	0.150	0.046	25.748			
$\phi_{12,3}$	0.505	0.048	35.564			
$\gamma_1$	0.035	0.006	4.724	0.034	0.007	6.728
$\gamma_2$	0.060	0.018	15.854	0.063	0.020	13.747
$\gamma_3$	0.014	0.003	38.034	0.021	0.010	55.629
$\phi_{h,1}$	0.792	0.113	17.233	0.821	0.108	15.533
$\phi_{h,2}$	0.913	0.055	12.243	0.925	0.046	10.165
$\phi_{h,3}$	0.838	0.086	24.056	0.854	0.083	29.536
$\sigma_{h,1}^2$	0.049	0.024	25.156	0.047	0.022	20.573
$\sigma_{h,2}^2$	0.059	0.029	26.603	0.063	0.029	24.137
$\sigma_{h,3}^2$	0.054	0.024	38.123	0.054	0.024	39.649
$p_{11}$	0.993	0.006	1.028	0.992	0.006	1.070
$p_{22}$	0.992	0.006	1.021	0.992	0.007	1.076
$L$	20.570	5.379	23.945			
log(ML)	346.612			259.751		
$\alpha=0.50$						
log(ML)	347.018			260.156		
$\alpha=0.75$						
log(ML)	347.254			260.392		
$\alpha=0.95$						
log(ML)	347.295			260.434		
$\alpha=0.99$						
DIC	-549.385			-434.130		

This table reports the posterior mean (mean), standard deviation (std. dev) and inefficiency factors,  $R_B$  with  $B = 100$  for different structural break ARFIMA( $p,d,q$ )-SV type models. The parameters associated with each regime are labeled with subscript 1, ...,  $m$ . log(ML): logarithm of the marginal likelihood using the corresponding value of  $\alpha$ . DIC: deviance information criterion.

Finally, we follow section 4.2 and assume that  $\gamma_k, \sigma_{h,k}^2 \sim IG\left(\frac{4}{2}, \frac{0.2}{2}\right)$ ,  $\phi_{h,k} \sim Beta(20, 1.5)$ ,  $k = 1, \dots, m$ , and  $p_{kk} \sim Beta(20, 0.1)$ ,  $k = 1, \dots, m - 1$ . In this setting, the prior for each element of  $\theta_k$  is very standard, while the prior for  $p_{kk}$  favors infrequent structural breaks. In the appendix we evaluate sensitivity of the results to different prior specifications by investigating alternative prior hyperparameter values on  $p_{kk}$ .

We also report results for a change-point integrated moving average (IMA) model of order 1 with SV effects, CP( $m$ )-IMA( $1, 1$ )-SV. This model corresponds to a CP( $m$ )-ARFIMA( $0, 1, 1$ )-SV model for  $y_t$  or a CP( $m$ )-ARIMA( $0, 0, 1$ )-SV model for  $\Delta y_t$ . In this model changes in the long run persistence are captured by changes in the MA(1) parameter,  $\psi_{s_t}$ <sup>3</sup>. We estimate CP( $m$ )-IMA( $1, 1$ )-SV from 0 to 4 change points and choose the optimal number of change points using ML and DIC. Again, results point that the specification with 2 change points fits the data best. We report estimation results for IMA( $1, 1$ )-SV and CP(2)-IMA( $1, 1$ )-SV in Tables 8 and 9.

We find that specifications with change points dominate the ARFIMA( $12, d, 0$ ) model, both in terms of ML and DIC. Furthermore, compared to ARFIMA( $12, d, 0$ )-SV, the log  $BF$  in favor of CP(2)-ARFIMA( $12, d, 0$ )-SV is 75. For the ARFIMA( $12, d, 0$ ) model the order of integration,  $d_1$ , is estimated at 0.38. This implies that US core inflation exhibits long memory behavior.  $\phi_{12,1}$  captures the main seasonal effects. The average inflation rate,  $\mu_1$  is estimated at 0.17%. Furthermore, the residual standard deviation of the ARFIMA( $12, d, 0$ ) model,  $\sigma_1$  is at 0.18% per month. When we compare ARFIMA( $12, d, 0$ ) with ARFIMA( $12, d, 0$ )-SV we find that  $d_1$  drops from 0.38 to 0.29. The AR coefficients,  $\phi_{11,1}$  and  $\phi_{12,1}$  increase from 0.11 and 0.24 to 0.14 and 0.36 respectively. The estimate of  $\mu_1$  is also affected, being more precisely estimated at a lower value. The SV component itself is nearly nonstationary as the autoregressive coefficient of volatility,  $\phi_{h,1}$ , is close to one and  $\sigma_{h,1}^2$ , is well identified at 0.03 with a standard error of 0.01. The average variance of inflation,  $\gamma_1$  is estimated at 0.025% per month.

Figure 6 displays the data, estimates of  $\sigma_t^2 = \gamma_{s_t} \exp(h_t)$ ,  $t = 1, \dots, T$  and the posterior density of the change-point dates for CP(2)-ARFIMA( $12, d, 0$ )-SV. The posterior mode of the change-point density is associated with 1973:7 and 1984:2. Furthermore, from the top right panel of Figure 6 we see that the volatility decrease in the early 1980s is noticeable and persistent. This phenomenon is often labeled as “The Great Moderation”, see for example Stock and Watson (2007).

Overall, results from (4.6)-(4.7) conditional on 2 change points show that we can divide the evolution of  $y_t$  into three subsequent phases: the period from 1960:1-1973:6, 1973:7-1984:1 and 1984:2 to the end of the sample. Both  $d_{s_t}$  and  $\mu_{s_t}$  are much smaller after the last change point,  $d_2 = 0.41$  versus  $d_3 = 0.15$  and  $\mu_2 = 0.53$  versus  $\mu_3 = 0.07$ . On the other hand,  $\phi_{11,s_t}$  and  $\phi_{12,s_t}$  increase from 0.02 and 0.09 to 0.15 and 0.50 respectively. At the same time, the estimate of  $\gamma_{s_t}$  almost doubles in the second regime while on the other hand it falls from 0.06 to 0.01 after the last

<sup>3</sup>IMA( $1, 1$ )-SV without structural breaks is equivalent to the UC model of Stock and Watson (2007).



structural break. Furthermore, the unconditional volatility of volatility,  $\sigma_{h,s_t}/\sqrt{1-\phi_{s_t}^2}$  rises from 0.36 in the first regime to 0.60 in the second regime and falls to 0.42 from the last change point to the end of the sample.

The last two columns of Table 9 report results for CP(2)-IMA(1,1)-SV. The estimate of  $\psi_{1,s_t}$  rises from  $-0.84$  in the first phase to  $-0.56$  in the second phase and drops to  $-0.85$  in the subsequent phase. Furthermore, similar to CP(2)-ARFIMA(12,d,0)-SV the unconditional volatility of volatility drops from the last change point to the end of the sample. However, CP(2)-IMA(1,1)-SV performs worse both in terms of ML and DIC compared CP(2)-ARFIMA(12,d,0)-SV.

To summarize: we find evidence of structural breaks in the dynamics of  $y_t$ . As expected, most significant changes in the model parameters occur during the Great Moderation. More importantly, it is also cautiously evident that the long memory characteristics of US inflation might have not remained significant after the Great Moderation, as indicated by the lower persistence estimate for the last regime,  $d_3$ . We seem to agree with Stock and Watson (2007) that US inflation may have become harder to model and forecast as the month-to-month memory has dropped.

[Figure6 about here]

We follow section 4.1 and compare the out-of-sample performance of CP(2)-ARFIMA(12,d,0)-SV (break) with ARFIMA(12,d,0)-SV (no-break). Specifically, we compare the out-of-sample predictive likelihood (PL) and predictive mean between these two models.

Given the data up to time  $t-1$ ,  $Y_{t-1}$ , the predictive likelihood (PL),  $p(y_t, \dots, y_T | Y_{t-1})$  is the predictive density evaluated at the realized outcome,  $y_t, \dots, y_T$ ,  $t \leq T$ , see Geweke (2005). The PL for model  $\mathcal{M}_A$  is given as

$$\begin{aligned} p(y_t, \dots, y_T | Y_{t-1}, \mathcal{M}_A) &= \int_{\Theta_A} p(y_t, \dots, y_T | \theta_A, Y_{t-1}, \mathcal{M}_A) p(\theta_A | Y_{t-1}, \mathcal{M}_A) d\theta_A \\ &= \prod_{s=t}^T N^{-1} \sum_{i=1}^N p(y_s | \theta_A^{(i)}, Y_{s-1}, \mathcal{M}_A) \end{aligned} \quad (4.8)$$

Notice that the terms on the right-hand-side of (4.8) have parameter uncertainty integrated out. If  $t=1$  this would be the marginal likelihood and (4.8) changes to (2.6). Hence, the sum of log predictive likelihoods can be interpreted as a measure similar to the log of the marginal likelihood, but ignoring the initial  $t-1$  observations.

The predictive likelihood can be used to order models according to their predictive abilities. In a similar fashion to Bayes factors, one can also compare the performance of models based on a specific out-of-sample period by predictive Bayes factors, PBF. The PBF for model  $A$  versus  $B$  is given as  $PBF_{AB} = p(y_t, \dots, y_T | Y_{t-1}, \mathcal{M}_A) / p(y_t, \dots, y_T | Y_{t-1}, \mathcal{M}_B)$  and summarizes the relative evidence of the two models over the out-of-sample data,  $y_t, \dots, y_T$ .

In order to compare the out-of-sample density forecasts of the models, we calculate the predictive Bayes factor for data at and after the first break point, 1973:7. Hence,  $t - 1 = 1973:6$ . For one observation out-of-sample,  $\log(PBF) = 2.36$ , 6 months  $\log(PBF) = 2.81$ , 1 year  $\log(PBF) = 2.94$ , 5 years  $\log(PBF) = 5.11$ , 10 years  $\log(PBF) = 7.33$  and 15 years  $\log(PBF) = 10.34$  each in favor of the break specification. The improvements continue till the end of sample, see Table 10. Finally, Table 10 also displays out-of-sample results for one-month-ahead point forecasts for the no-break and the break model. Overall, the break model offers improvements in terms of MAE and RMSE compared to the no-break model.

Table 10: Out-of-sample forecasts for US core inflation

Model	MAE	RMSE	LINEX, $a = 1, b = 1$	LINEX, $a = -1, b = 1$	log(PL)
No-break	0.129	0.180	0.017	0.016	116.048
Break	0.121	0.168	0.014	0.015	130.556

This table reports mean absolute error (MAE), root mean squared error (RMSE) and average LINEX for the forecasts based on the predictive mean for one-month-ahead. Furthermore, the one-month-ahead log predictive likelihood, log(PL) is also reported. The out-of-sample period is from 1973:7 to the end of the sample.

## 5 Conclusion

In this paper we apply PG-AS to the challenging class of unobserved component time series models and demonstrate its flexibility under different circumstances. We start by reproducing the results of Chan and Hsiao (2013). Although we use PG-AS instead of Gibbs sampling, our results are very similar to those of Chan and Hsiao (2013) and any qualitative conclusion and inference drawn from the results is the same.

We also show the flexibility of PG-AS by combining PG-AS with the change-point specification of Chib (1998). Our simulation and empirical example using US GDP data show that this combination provides reliable results, in terms of estimation, change point identification and forecasting.

Finally, we analyze the behavior of US monthly core inflation rate using a structural break ARFIMA-SV model. We find evidence of two structural breaks. Furthermore, we find considerable differences in parameter estimates within in each regime. We also demonstrate that accounting for structural breaks improves density and point forecasts.

# A Appendix

## A.1 Prior sensitivity analysis

In this section sensitivity of the results to prior specification is evaluated by investigating alternative prior hyperparameter values,  $a_0$  and  $b_0$  on the transition probabilities,  $p_{kk} \sim \text{Beta}(a_0, b_0)$ , keeping prior hyperparameter values of the other parameters the same as in the main text.  $p_{kk}$ ,  $k = 1, \dots, m - 1$  is one of the key parameters of the model because it controls the duration of each regime in  $S$ .

In Table 11, we experiment with different hyperparameter values for  $p_{kk}$  and report the break dates for each of them by estimating CP(2)-ARFIMA-SV using the corresponding values of  $a_0$  and  $b_0$ . For instance, the first alternative prior that is considered is  $p_{kk} \sim \text{Beta}(0.1, 0.1)$  which is relatively flat. With this prior we still find that the change-point dates correspond to 1973:7 and 1984:2. In fact, regardless the values of  $a_0$  and  $b_0$  we still find that the change-point dates for each of these specifications correspond to 1973:7 and 1984:2. In Table 11, we also report  $\log BF$  of CP(2)-ARFIMA-SV versus ARFIMA-SV using the corresponding values of  $\alpha$  and the difference in DIC between CP(2)-ARFIMA-SV and ARFIMA-SV. These results overwhelmingly suggest existence of structural breaks. More importantly, we find that the choice of prior hyperparameter values on  $P$  is of relatively limited importance.

Table 11: Prior sensitivity analysis, CP(2)-ARFIMA-SV

Prior	break dates	$\log BF$	$\log BF$	$\log BF$	$\log BF$	diff(DIC)
		$\alpha = 0.50$	$\alpha = 0.75$	$\alpha = 0.95$	$\alpha = 0.99$	
$\text{Beta}(0.1, 0.1)$	1973:7, 1984:2	59.345	59.347	59.346	59.346	-33.251
$\text{Beta}(8, 0.1)$	1973:7, 1984:2	67.807	67.808	67.808	67.808	-33.385
$\text{Beta}(20, 0.1)$	1973:7, 1984:2	74.911	74.913	74.912	74.912	-27.454
$\text{Beta}(100, 0.1)$	1973:7, 1984:2	60.123	60.125	60.124	60.124	-29.880

This table compares the performance of CP(2)-ARFIMA-SV for different values of  $a_0$  and  $b_0$  where  $p_{kk} \sim \text{Beta}(a_0, b_0)$ . The priors of the other parameters are set according to the main text.  $\log BF$ : logarithm of the Bayes factor for CP(2)-ARFIMA-SV over ARFIMA-SV using the corresponding value of  $\alpha$ .  $\text{diff(DIC)}$ : difference in DIC between CP(2)-ARFIMA-SV and ARFIMA-SV.

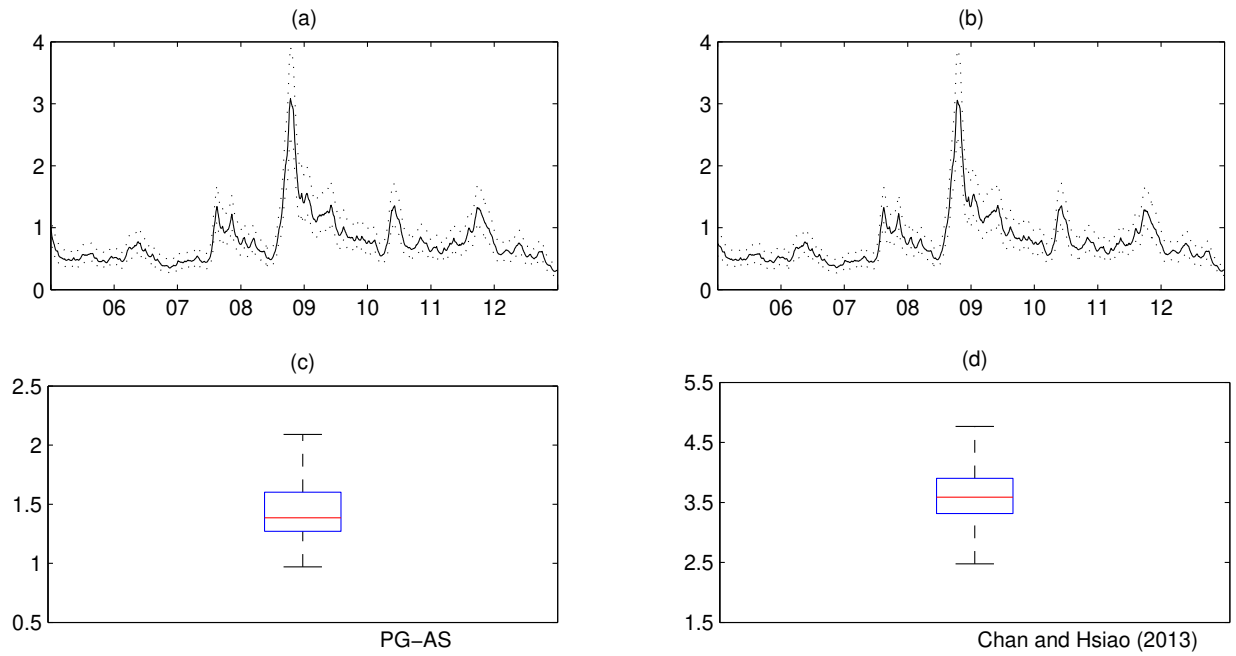
## References

- [1] Andrieu, C., and A. Doucet. 2002. "Particle filtering for partially observed Gaussian state space models." *Journal of the Royal Statistical Society B* 64(4): 827-836.
- [2] Andrieu, C., A. Doucet, and R. Holenstein. 2010. "Particle Markov chain Monte Carlo methods (with discussion)." *Journal of the Royal Statistical Society B* 72(3): 1-33.
- [3] Bos, C. S., S. J. Koopman, and M. Ooms. 2012. "Long memory with stochastic variance model: A recursive analysis for U.S. inflation." *Computational Statistics and Data Analysis*, forthcoming.
- [4] Chan, J. 2013. "Moving Average Stochastic Volatility Models with Application to Inflation Forecast." *Journal of Econometrics* 176(2): 162-172.
- [5] Chan, J. 2014. "The Stochastic Volatility in Mean Model with Time-Varying Parameters: An Application to Inflation Modeling." Working paper, Research School of Economics, Australian National University.
- [6] Chan, J., and C. Hsiao. 2013. "Estimation of Stochastic Volatility Models with Heavy Tails and Serial Dependence." in *Bayesian Inference in the Social Sciences*. John Wiley & Sons, New York.
- [7] Chan, N. H., and W. Palma. 1998. "State space modeling of long-memory processes." *Annals of Statistics* 26(2): 719-740.
- [8] Chib, S., E. Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *The American Statistician* 49(4): 327-335.
- [9] Chib, S. 1998. "Estimation and Comparison of Multiple Change-Point Models." *Journal of Econometrics* 86(2): 221-241.
- [10] Chib, S., F. Nadari, and N. Shephard. 2002. "Markov chain Monte Carlo methods for stochastic volatility models." *Journal of Econometrics* 108(2): 281-316.
- [11] Creal, D. 2012. "A survey of sequential Monte Carlo methods for economics and finance." *Econometric Reviews* 31(3): 245-296.
- [12] Doucet, A., S. J. Godsill, and C. Andrieu. 2000. "On sequential Monte Carlo sampling methods for Bayesian filtering." *Statistics and Computing* 10(3): 197-208.

- [13] Doucet, A., and A. Johansen. 2011. “A tutorial on particle filtering and smoothing: Fifteen years later.” in *The Oxford Handbook of Nonlinear Filtering*. D. Crisan and B. Rozovsky, Eds. Oxford University Press.
- [14] Gelfand, A., and D. Dey. 1994. “Bayesian Model Choice: Asymptotics and Exact Calculations.” *Journal of the Royal Statistical Society B* 56(3): 501-514.
- [15] Geweke, J. 2005. *Contemporary Bayesian Econometrics and Statistics*. Wiley.
- [16] Gordon, S., and J. Maheu. 2008. “Learning, Forecasting and Structural Breaks.” *Journal of Applied Econometrics* 23(5): 553-583.
- [17] Kass, R. E., and A. E. Raftery. 1995. “Bayes Factors.” *Journal of the American Statistical Association* 90: 773-795.
- [18] Kim, C. J., and C. R. Nelson. 1999. *State Space Models with Regime Switching Classical and Gibbs Sampling Approaches with Applications*. MIT Press.
- [19] Kim, C. J., and C. R. Nelson. 1999. “Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of Business Cycle.” *Review of Economics and Statistics* 81(4): 608-616.
- [20] Kim, C. J., C. R. Nelson, and J. Piger. 2004. “The Less Volatile U.S. Economy: A Bayesian Investigation of Timing, Breadth, and Potential Explanations.” *Journal of Business and Economic Statistics* 22(1): 80-93.
- [21] Kim, S., N. Shephard, and S. Chib. 1998. “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models.” *Review of Economic Studies* 65(3): 361-393.
- [22] Koop, G. 2003. *Bayesian Econometrics*. John Wiley & Sons Ltd.
- [23] Koopman, S. J., and E. H. Uspensky. 2002. “The stochastic volatility in mean model: empirical evidence from international stock markets.” *Journal of Applied Econometrics* 17(6): 667-689.
- [24] Lindsten, F., M. I. Jordan, and T. B. Schön. 2012. “Ancestor Sampling for Particle Gibbs.” *Advances in Neural Information Processing Systems (NIPS)* 25: 2600-2608.
- [25] Lindsten, F., and T. B. Schön. 2013. “Backward Simulation Methods for Monte Carlo Statistical Inference.” *Foundations and Trends in Machine Learning* 6(1): 1-14.
- [26] Liu, C., and J. Maheu. 2008. “Are There Structural Breaks in Realized Volatility?” *Journal of Financial Econometrics* 6(3): 326-360.

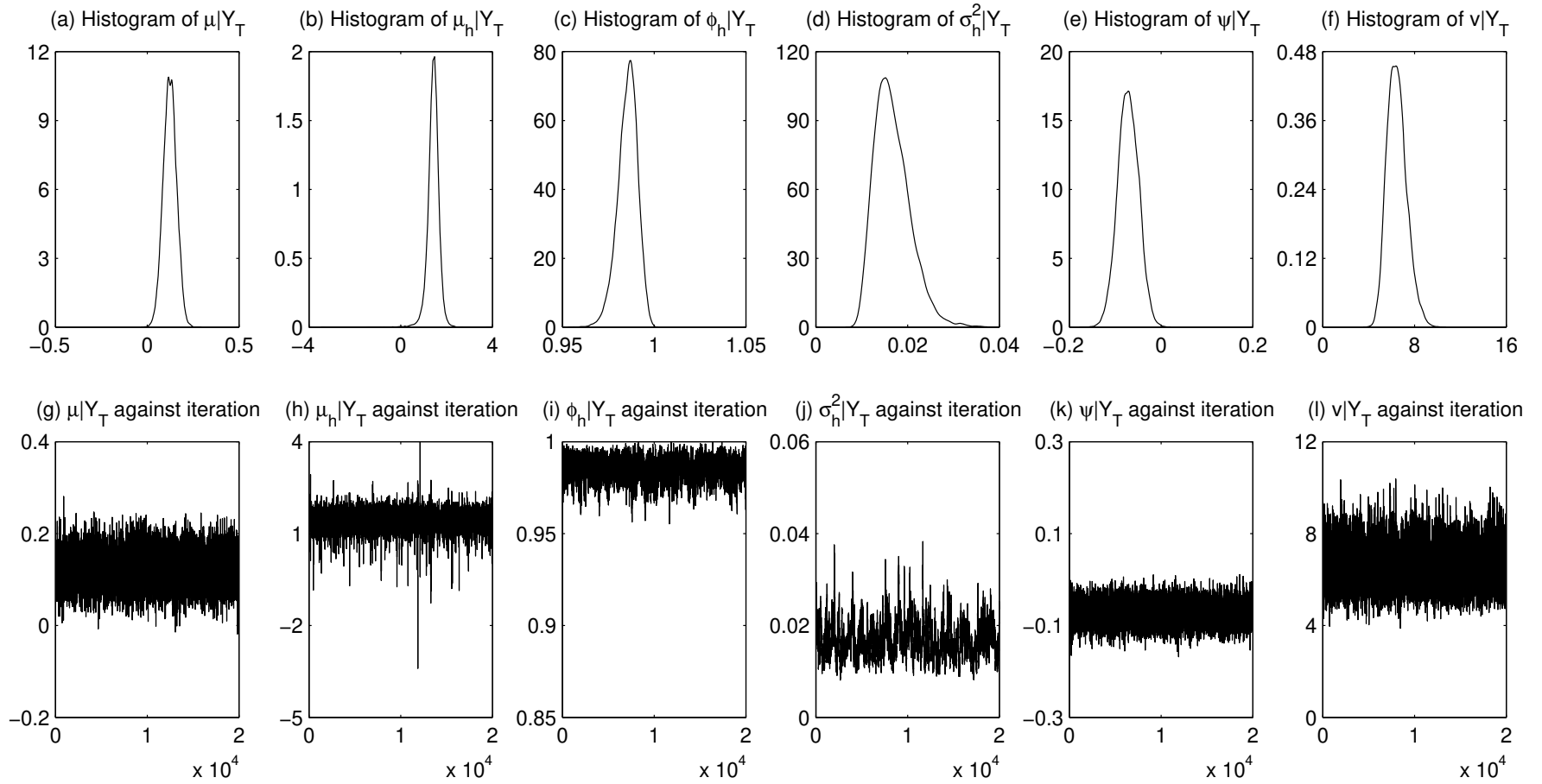
- [27] Marcellino, M., J. H. Stock, and M. W. Watson. 2005. "A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Series h-Steps Ahead." *Journal of Econometrics* 135(1-2): 499-526.
- [28] Pesaran, H., D. Pettenuzzo, and A. Timmermann. 2006. "Forecasting Time Series Subject to Multiple Structural Breaks." *Review of Economic Studies* 73(4): 1057-1084.
- [29] Raggi, D., and S. Bordignon. 2012. "Long Memory and Nonlinearities in Realized Volatility: A Markov Switching Approach." *Computational Statistics and Data Analysis* 56(11): 3730-3742.
- [30] Sims, C. A., D. F. Waggoner, and T. Zha. 2008. "Methods for Inference in Large Multiple-Equation Markov-Switching Models." *Journal of Econometrics* 146(2): 255-274.
- [31] Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde. 2002. "Bayesian Measures of Model Complexity and Fit (with comments)". *Journal of the Royal Statistical Society B* 64(4): 583-639.
- [32] Stock, J. H., and M. W. Watson. 2007. "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit, and Banking* 39(1): 3-34.
- [33] Zellner, A. 1986. "Bayesian Estimation and Prediction Using Asymmetric Loss Functions." *Journal of the American Statistical Association* 81: 446-451.

Figure 1: PG-AS and Gibbs sampling results for AUD/USD daily returns data



Graph (a): posterior mean (solid line) and 90% credibility intervals (dashed lines) of  $\exp(h_t/2)$  using PG-AS, (b): posterior mean (solid line) and 90% credibility intervals (dashed lines) of  $\exp(h_t/2)$  using the Gibbs sampling approach of Chan and Hsiao (2013). Graphs (c) and (d): box plots of the inefficiency factors of  $h_{1:T}$  for PG-AS and the Gibbs sampling approach of Chan and Hsiao (2013).

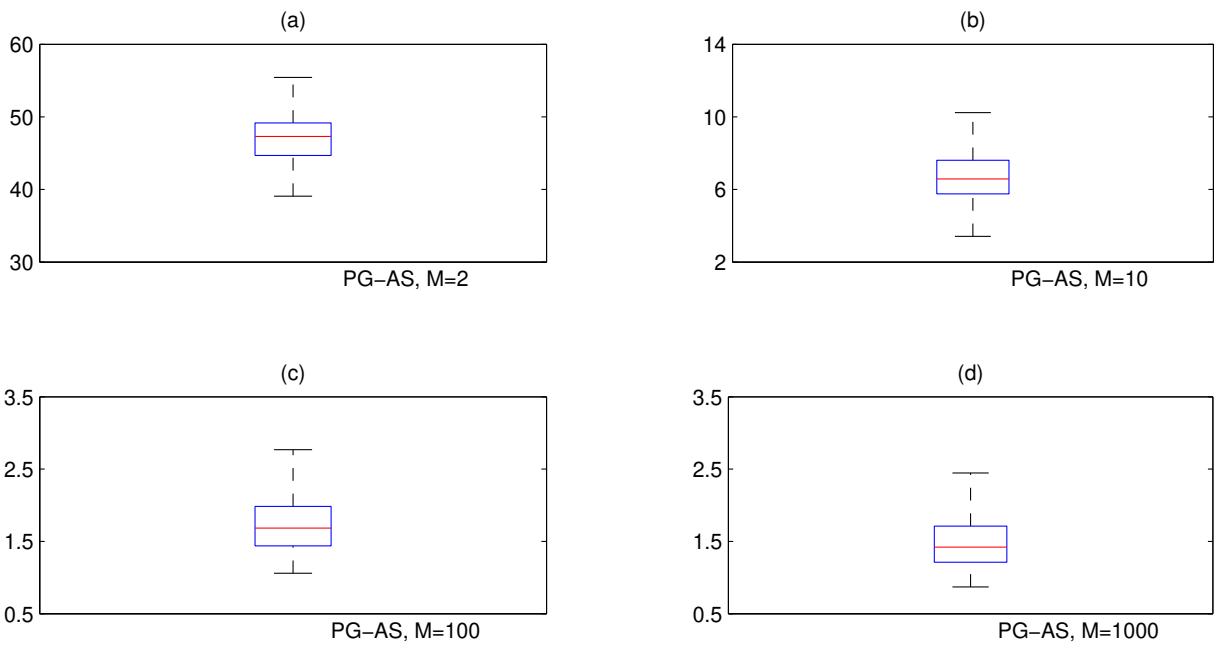
Figure 2: PG-AS sampler for daily returns on the silver spot price



Graphs (a)-(f): histograms of marginal distributions. Graphs (g)-(l): simulation against iterations

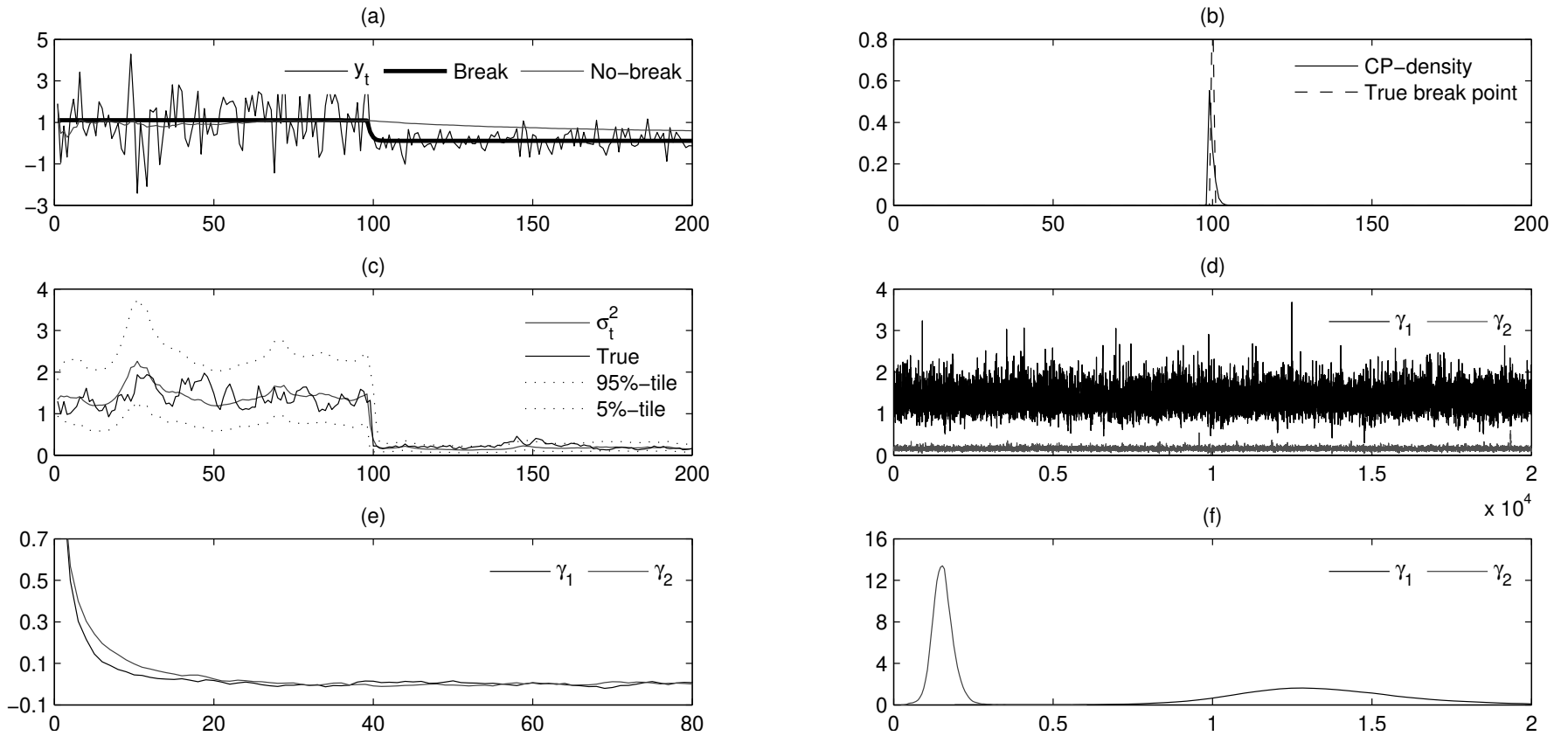


Figure 3: Sensitivity of the PG-AS sampler with respect to  $M$



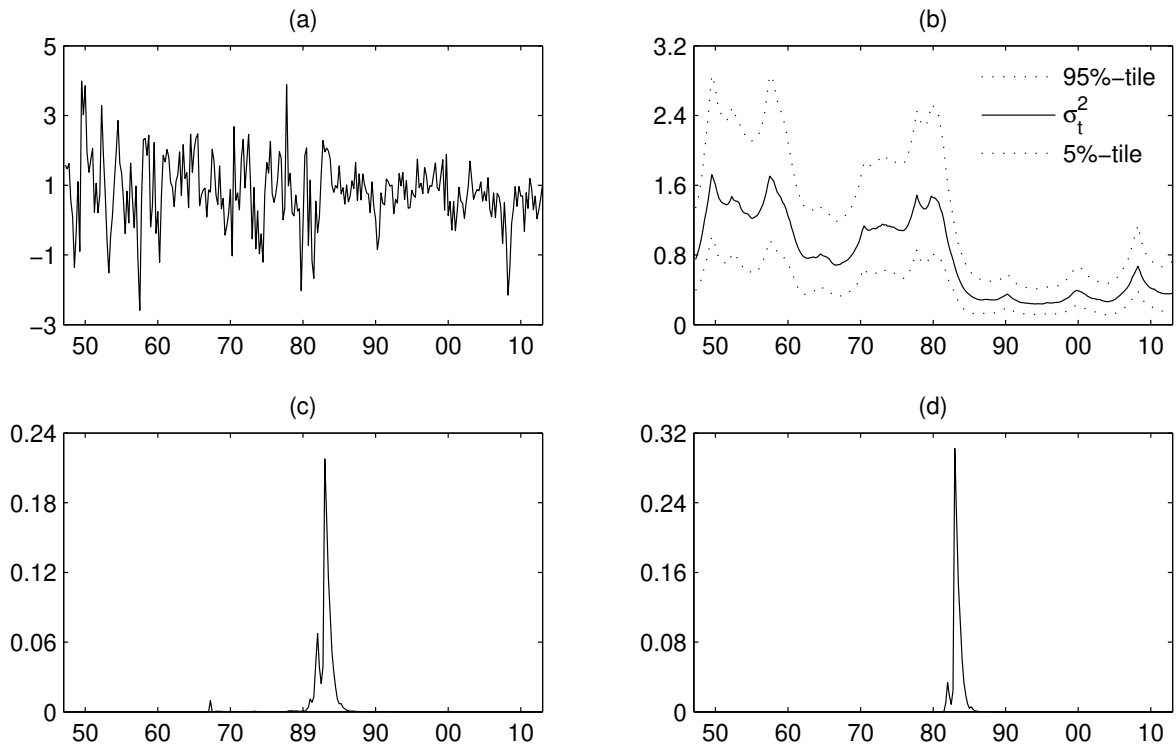
Graphs (a)-(d): box plots of the inefficiency factors of  $h_{1:T}$  using the corresponding number of particles.

Figure 4: Simulation results



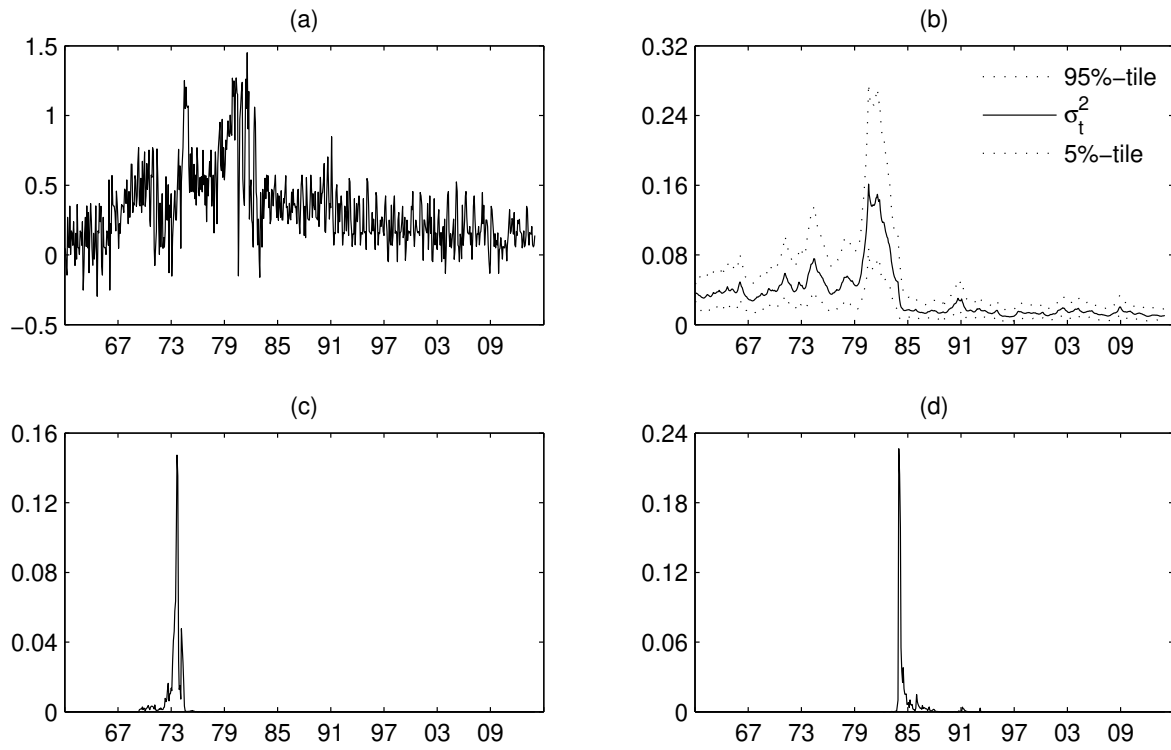
Graph (a): data and predictive means, (b): change point density, (c): posterior mean (solid line) and 90% credibility intervals (dashed lines) of  $\sigma_t^2$ ,  $t = 1, \dots, T$ , (d): simulation against iterations, (e): autocorrelation functions, (f): histograms of marginal distributions.

Figure 5: PG-AS sampler for US quarterly real GDP growth rate from 1947q4 to 2013q3



Graph (a): US quarterly real GDP growth rate, (b): posterior estimates of  $\sigma_t^2$ ,  $t = 1, \dots, T$  for CP-AR(2)-SV, (c): change point density for CP-AR(2)-SV, (d): change point density for CP-AR(2).

Figure 6: PG-AS sampler for US monthly core inflation rate from 1960:1 to 2013:12



Graph (a): US monthly core inflation adjusted for fixed seasonals,  $y_t$ , (b): posterior estimates of the conditional variance of inflation, (c): posterior density of the first change point, (d): posterior density of the second change point.