2016-07

Sparse Change-point HAR Models for Realized Variance

Arnaud Dufays
Jeroen V.K. Rombouts

décembre / december 2016

Centre de recherche sur les risques
les enjeux économiques et les politiques publiques

www.crrep.ca

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 switches in all parameters. This implies that all parameters have to change when a break happens. To gauge whether and where there are structural breaks in realized variance, we introduce the sparse change-point HAR model. The approach controls for model parsimony by limiting the number of parameters which evolve from one regime to another. Sparsity is achieved thanks to employing a nonstandard shrinkage prior distribution. We derive a Gibbs sampler for inferring the parameters of this process. Simulation studies illustrate the excellent performance of the sampler. Relying on this new framework, we study the stability of the HAR model using realized variance series of several major international indices between January 2000 and August 2015.

Keywords: Realized variance, Bayesian inference, Time series, Shrinkage prior, Change-point model, Online forecasting

JEL Classification: C11, C15, C22, C51.

Arnaud Dufays: Université Laval https://sites.google.com/site/websiteofarnauddufays/

Jeroen V.K. Rombouts : ESSEC Business School, Av. B. Hirsch, Cergy Pontoise, rombouts@essec.edu

1 Introduction

A central theme in financial econometrics over the last two decades has been estimating volatility using high frequency data. For general price diffusion dynamics it is possible to precisely estimate daily integrated variance using intra-day data in a model free and "realized" way. In particular, the standard daily realized variance estimator proposed by Andersen, Bollerslev, Diebold, and Labys (2001) is simply the sum of squared intra-day observations over the day. The fact that there is microstructure noise and jumps has given rise to various realized variance estimators, see Ait-Sahalia and Jacod (2014) for textbook material and more references on the topic. See also McAleer and Medeiros (2008b) for an earlier review on realized volatility.

Apart from the estimation of historical daily realized variance, there have been efforts to develop time series models that easily allow forecasting future variance. Taking into account that the dynamics of realized variance has long memory features, Corsi (2009) proposes a model easy to estimate with forecasting performance difficult to beat. In fact, his heterogenous autoregressive (HAR) model explains tomorrows expected volatility as a linear function of the volatilities of yesterday, last week and last month. Extensions of the HAR model can be found in Corsi, Audrino, and Reno (2012). Apart from many variance forecasting papers using the HAR model (including multivariate extensions), the model is also omnipresent in the literature on variance risk premia, see for example Bollerslev, Marrone, Xu, and Zhou (2014) and Bekaert and Hoerova (2014), and in option pricing using realized volatility as in Corsi, Fusari, and Vecchia (2013).

This paper studies the stability of the HAR model using realized variance series of eleven international indices between January 2000 and August 2015. We enrich the HAR model by allowing for multiple structural breaks in the parameters. The structural breaks in our model can be such that they appear only in some of the mean and/or variance parameters. Our method allows estimating parameters with no breaks with the full sample and others parameters with the relevant subsamples. Which parameters break, is identified automatically with our approach. Information on the breaks and corresponding parameters is obtained in one single estimation algorithm using Bayesian inference. By providing such a general methodology to detect relevant breaks, we can gauge the performance of the standard HAR model in-sample by investigating which parameters of the model break, and out-of-sample by computation of the predictive likelihoods and root mean squared errors.

Other papers have studied the structure of the HAR model. Audrino and Knaus (2016) question if the daily, weekly and monthly frequencies used in the HAR model are inherent in individual stock data. They apply the LASSO algorithm as model selection tool on an unrestricted linear autoregressive process and cannot recover the HAR structure. However when comparing out-of-sample the model implied by the LASSO algorithm with the HAR model they find similar performance. An extension is Audrino, Camponovo, and Roth (2015) who construct a test to investigate the optimal lag structure, allowing them to show for several individual stocks that this lag structure is time varying.

This paper uses a new approach to model time series subject to multiple structural breaks since 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. Furthermore, our shrinkage methodology alleviates the search of an optimal number of regimes and does only rely on one estimation output. We use a new shrinkage prior based on a finite mixture of Uniform distributions as proposed in Dufays and Rombouts (2015). In fact, their new shrinkage prior allows for changes in subsets of model parameters that are virtually zero, implying that only relevant parameters break. Regarding the selection of the prior hyper parameters, it is possible to state explicitly how much additional log-likelihood is required for triggering a parameter break.

The paper closest to ours on the topic of breaks in realized variance is Liu and Maheu (2008). They investigate the evidence of structural breaks in the HAR model by using the change-point approach introduced by Chib (1998). The estimation of the number of breaks and their corresponding dates requires computing marginal likelihood for respectively models with no break, one break, two breaks, etc. This procedure requires to estimate the CP model multiple times. Using daily S&P500 data between 1993-2004, Liu and Maheu (2008) find evidence of one break in the variance of the HAR model in early 2007. Other papers modelling the dynamics of realized volatility using switching HAR type models have been investigated for example by McAleer and Medeiros (2008a) and Gallo and Otranto (2015).

We validate our new approach in a simulation study in two settings based on a data generating process (DGP) of Chan, Yip, and Zhang (2014) and our estimation results found in the application. The estimation algorithm for the Sparse CP model detects on average the correct break dates and infers which parameters are changing over time. The standard CP model of Chib (1998) detects the correct number of breaks for the first DGP but the parameters estimates are less precise. For the second DGP, the standard CP model systematically fails to detect the correct number of breaks. In contrast, the Sparse CP model performs well and can even detect relatively small and short lived regimes.

With respect to our empirical results for realized variance on the full sample of eleven international stock markets, we first compare the standard CP-HAR model with our new sparse CP-HAR model. It turns out that both approaches find almost identical breaks in realized variance, detecting up to six short and long lived regimes. However, since a break in the standard CP-HAR model implies a full new set of HAR parameters, we cannot identify which parameters are changing and therefore we loose parsimony. In fact, the sparse CP-HAR reveals that there is typically no change at all in the HAR model parameters, i.e. the breaks are almost uniquely in the variance of the realized variance series. The second part of our empirical results focuses on an out-of-sample forecast evaluation comparing the standard HAR model of Corsi (2009) with the sparse CP-HAR model. As expected, since structural breaks are almost uniquely in the variance, we find similar performance between the two models when the forecast accuracy is measured in root mean squared error or in average predictive likelihoods. When zooming in on a subsample in the forecast exercise, we find that for some stock markets the average predictive likelihood is slightly better for the sparse CP-HAR model for the period from the global financial crisis and the subsequent two years.

The rest of the paper is organized as follows. Section 2 describes the data and standard models for realized volatility. Section 3 defines the sparse CP-HAR model. We detail the tailored to breaks shrinkage prior and describe the sampler to perform inference. We show the excellent performance of the approach in a simulation study in Section 4. Section 5 estimates first the model on the full sample of international stock market realized variance. This allows us to verify where exactly are breaks in realized variance. Second we do an out-of-sample forecast evaluation comparing the standard HAR model with the sparse CP-HAR model. Section 6 concludes.

2 Data and standard models for realized volatility

We use daily realized variance measures in natural logarithm for the following eleven international stock marked indices: AEX, CAC, DAX, DJI, EUROSTOXX (EURO), FTSE, NASDAQ (NASD), NIKKEI (NIK), RUSSELL (RUS), SMI and S&P500. The nontransformed realized variances (RV) are computed using five-minutes returns. The series start on January 3, 2000 ending on August 5, 2015 and are available from Oxford-Man Institute of Quantitative Finance. Descriptive statistics for the realized variance measures in logs are given in Table 1.

Table 1: Descriptive statistics for log realized variance

	AEX	CAC	DAX	DJI	EURO	FTSE	NASD	NIK	RUS	SMI	S&P500
Mean	-0.35	-0.35	-0.13	0.03	-0.49	-0.02	-0.71	-0.33	-0.32	-0.43	-0.47
Std.Dev.	1.01	1.01	0.97	1.03	1.05	1.01	1.03	1.07	0.86	0.98	1.07
Skewness	0.49	0.49	0.32	0.37	0.53	0.10	0.47	0.45	0.30	0.51	0.46
Kurtosis	3.17	3.17	3.16	3.10	3.54	5.07	3.11	2.95	3.54	3.70	3.38
Min	-3.80	-3.80	-3.20	-2.99	-3.46	-7.67	-3.27	-3.12	-2.95	-3.53	-4.12
Max	3.59	3.59	3.94	4.07	4.46	4.68	3.83	3.76	3.47	4.07	4.35
AR(1)	0.84	0.84	0.84	0.84	0.76	0.77	0.84	0.86	0.78	0.74	0.79
Nobs	3967	3968	3945	3894	3945	3914	3776	3899	3896	3893	3892

Notes: Descriptive statistics for the log realized variance measures. Daily data from January 3, 2000 and end on August 5, 2016. AR(1) is the first order sample autocorrelation; Nobs signifies the number of observations.

The benchmark model for forecasting the realized variance is the HAR model proposed by Corsi (2009). The standard HAR model is given by

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-1}^{(5)} + \beta_3 y_{t-1}^{(22)} + \varepsilon_t \tag{1}$$

with $y_{t-1}^{(n)} = (y_{t-1} + y_{t-2} + \cdots + y_{t-n})/n$, and with $\varepsilon_t \sim N(0, \sigma^2)$. If we replace the definitions of $y_{t-1}^{(5)}$ and $y_{t-1}^{(22)}$ in (1) we see that the HAR model is an autoregressive model of order 22, denoted by AR(22), with the following restrictions

$$y_t = \beta_0 + [\beta_1 + \beta_2/5 + \beta_3/22] y_{t-1} + [\beta_2/5 + \beta_3/22] \sum_{i=2}^{5} y_{t-i} + [\beta_3/22] \sum_{i=6}^{22} y_{t-i} + \epsilon_t. \tag{2}$$

The reduction from 22 parameters to 3 parameters which drive the dynamics of realized variance has empirically shown to strike a good balance between model flexibility and parameter uncertainty.

Table 2 reports parameter estimates for the standard HAR model, using OLS, for the eleven indices. All parameters are strongly significant, on average the implied persistence equals 0.95. A standard chi-square Box-Pierce test for residual autocorrelation is strongly rejected for all series indicating that the restrictions on the AR(22) model do not allow to fully capture the serial dependence in the realized variance series. In contrast, estimating an unrestricted AR(22) model on the eleven indices does not reject the residual autocorrelation tests anymore.

Table 2: standard HAR model estimates

	β_0	β_1	β_2	β_3	σ^2	B-P
AEX	-0.02	0.36	0.46	0.13	0.24	78.12
	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	
CAC	-0.01	0.35	0.46	0.15	0.23	94.12
	(0.37)	(0.00)	(0.00)	(0.00)	(0.00)	
DAX	-0.00	0.35	0.43	0.18	0.26	53.24
	(0.89)	(0.00)	(0.00)	(0.00)	(0.00)	
DJI	-0.03	0.25	0.49	0.20	0.38	53.80
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	
EURO	-0.00	0.29	0.45	0.20	0.35	41.44
	(0.77)	(0.00)	(0.00)	(0.00)	(0.00)	
FTSE	-0.03	0.32	0.47	0.18	0.24	42.03
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
NAS	-0.01	0.42	0.35	0.19	0.25	41.71
	(0.09)	(0.00)	(0.00)	(0.00)	(0.00)	
NIKK	-0.02	0.38	0.34	0.22	0.25	40.37
	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	
RUS	-0.03	0.28	0.44	0.21	0.36	63.74
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	
SMI	-0.03	0.36	0.48	0.12	0.19	56.25
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
SP500	-0.03	0.30	0.45	0.19	0.35	57.49
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	

Notes: OLS parameter estimates of the HAR model and the HARSQ process for log realized variance with p-values in brackets. B-P is the Box-Pierce autocorrelation test.

In the next section, we present a more flexible HAR model that allows for breaks in the parameters. In fact, this new change-point model can detect which parameters change over time. This allows to check the stability of the HAR model and forecast realized variance with the relevant historical data.

3 Sparse change-point HAR model

3.1 Model formulation

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 $[1, K+1]$. The CP-HAR model with $K+1$ regimes is defined by

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

for $s_t = 1, \ldots, K+1$. 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

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

This specification stands for the state-of-the-art CP setup in the literature. Its estimation for a fixed number of breaks K is typically carried out by the Gibbs sampler if we assume the following prior distributions:

$$\begin{aligned} \{\beta_{0,i}, ..., \beta_{3,i}\}' &\sim N(\underline{0}, \Sigma_\beta) \ \ \forall i \in [1, K+1], \\ \sigma_i^{-2} &\sim IG(k_\sigma, \theta_\sigma) \ \ \forall i \in [1, K+1], \\ p_i &\sim Beta(\alpha_p, \beta_p) \ \ \forall i \in [1, K], \end{aligned}$$

where IG denotes the Inverse Gamma distribution. The number of breaks is then determined by selecting the model exhibiting the highest marginal likelihood in a range of CP models which differ on the number of regimes. We refer to Chib (1998) for more details on the Gibbs sampler and to Chib (1995) for more information on the estimation of the marginal likelihood. Note that the standard CP model assumes that all the parameters change when a break is detected. This makes the number of parameters a linear function of the number of regimes. In our specific case of the HAR model, this number amounts to $5(K+1)+K$. Note that Koop and Potter (2007), Maheu and Gordon (2008), Geweke and Jiang (2011) and Maheu and Song (2013) have extended this change-point process. However, all these papers stick to the assumption that all the parameters change when a break is detected.

To apply our shrinkage prior that encourages sparsity in the CP model, we re-write (3) relative to the first regime as follows

$$
\begin{aligned}
y_t &= (\beta_{0,1} + \sum_{i=2}^{s_t} \Delta\beta_{0,i}) + (\beta_{1,1} + \sum_{i=2}^{s_t} \Delta\beta_{1,i}) y_{t-1} \\
&\quad + (\beta_{2,1} + \sum_{i=2}^{s_t} \Delta\beta_{2,i}) y_{t-1}^{(w)} + \ (\beta_{3,1} + \sum_{i=2}^{s_t} \Delta\beta_{3,i}) y_{t-1}^{(m)} + \varepsilon_t \\
&\quad \text{with } \varepsilon_t \sim N(0, \sigma_1^2 \prod_{i=2}^{s_t} \sigma_i^2) \ \text{ for } \ s_t = 2, \dots, K+1, \qquad (4)
\end{aligned}
$$

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 $s_t = 2, \dots, K+1$ and likewise for the other parameters. To set the notation below, we define the mean parameter set $\Theta = (\beta_{0,1}, \beta_{1,1}, \beta_{2,1}, \beta_{3,1}, \Delta\beta_{0,2}, \Delta\beta_{1,2}, \Delta\beta_{2,2}, \Delta\beta_{3,2}, ...)$, the variance parameter set $\Sigma = \{\sigma_1^2, \sigma_2^2, ..., \sigma_{K+1}^2\}$ and $p_{1:K} = \{p_1, ..., p_K\}$, the set of transition probabilities.

3.2 The shrinkage prior for CP-HAR models

In the applications below, we estimate the Sparse CP-HAR model which exhibits several parameters in each regime. However, to simplify the exposition here, we focus only on 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 $\Delta\mu_2 = \mu_2 - \mu_1$. 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.

We use the new shrinkage prior consisting of a mixture of two Uniform distribution (denoted by 2MU(a,b,P) or briefly 2MU) that is tailored to time series CP models as proposed in Dufays and Rombouts (2015). The $2MU$ belongs to the spike-and-slab priors (e.g. Mitchell and Beauchamp (1988) and Ishwaran and Rao (2005)) and its general form is as follows

$$
\Delta\mu \ \sim \ \omega \ \ U[\frac{-a}{2}, \frac{a}{2}] + (1-\omega) U[\frac{-b}{2}, \frac{b}{2}], \qquad (5)
$$

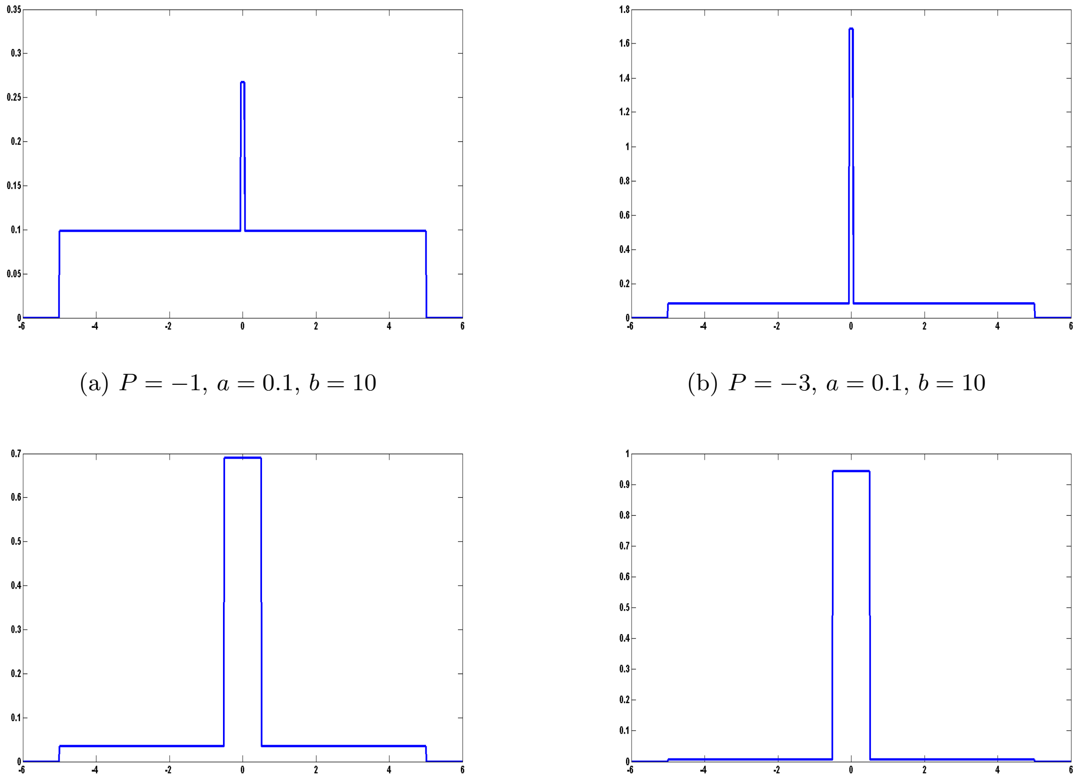
where $U[\frac{-a}{2}, \frac{a}{2}]$ denotes the Uniform distribution with bounds depending on a. In practice, we will take a small and b large. We define P the penalty on the log-likelihood function for detecting a new regime and by x any point in the wider uniform component (i.e. $|x| > \frac{a}{2}$). Assuming that the 2MU verifies $\log f(x) - \log f(0) = P$, then it yields

$$P = \log \frac{(1-\omega)/b}{\omega/a + (1-\omega)/b}. \tag{6}$$

Re-arranging (6) to obtain an expression for ω in relation to a chosen penalty P gives

$$\omega = \frac{a(1-e^P)}{be^P + a(1-e^P)}, \tag{7}$$

which belongs to [0,1] for any positive values of a,b and any negative value of P. When P is set to 0, the 2MU prior reduces to a Uniform distribution with support $[-\frac{b}{2}, \frac{b}{2}]$ which results in a standard CP model with diffuse priors. Figure 1 displays examples of the density function for different values of P, a and b.



(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 1: Mixture of two uniform distributions with different values of P, a and b.

Regarding the variance parameter, we also use a 2MU prior with positive support which shrinks the parameter around one. This setting can be written as follows

$$\sigma_i^2 \sim \omega \ U[1-\frac{a}{2},1+\frac{a}{2}]+(1-\omega)U[0,b] \ \text{ for } i=2,...,K+1. \quad (8)$$

3.3 Setting the 2MU hyper-parameters

The 2MU shrinkage prior depends on three hyper-parameters, i.e. the penalty parameter P and the Uniform bounds a, b. The latter defines when a break is detected since all parameter deviations inside the narrow Uniform bounds would not be considered as a break. Obviously, the hyper-parameter a should be set as small as possible while the wider Uniform component should cover all the admissible support of the parameters. In practice, we set b to 10 for the mean parameters and to 200 for the variance. The parameter a is fixed as follows. We first estimate the HAR model without break in the parameters and for each parameter, the parameter a is set to 10% of the difference between the posterior mean and the 5% quantile of the posterior distribution.

The penalty parameter should be carefully chosen such that it performs as the break detection procedure based on the marginal likelihood. That is, its value should imply detecting a similar amount of breaks as if we were using a state-of-the-art marginal likelihood method. In particular, we want to detect a break if and only if the Bayes factor of this model would have been preferred at 95% compared to a model without the break. This requires to compute ex-ante the marginal likelihood of the two models. Since this is not feasible, we rely on the easier to compute Bayesian information criterion. Let us denote by M_2 a model with one break that we compare through the Bayes factor to the process M_1, the same model without a break. According to our 95% rule and the BIC approximation, it leads to the following inequality

$$\frac{f(y_{1:T}|M_2)}{f(y_{1:T}|M_1)} \geq \frac{0.95}{1-0.95}, \quad (9)$$

$$\approx \log\frac{f_{M_2}(y_{1:T}|\hat{\mu},\Delta\hat{\mu},\hat{p}_1)}{f_{M_1}(y_{1:T}|\bar{\mu})} \geq \log\frac{0.95}{1-0.95}+\log T, \quad (10)$$

where $\{\bar{\mu}\}$ and $\{\hat{\mu},\Delta\hat{\mu},\hat{p}_1\}$ stand for the maximum likelihood estimates of Model 1 and Model 2 respectively. The model without a break is similar to a process where the deviation of the parameter $\Delta\mu$ is equal to 0 whatever the transition probability p_1. Using Bayes' rule, the ratio of the posterior densities evaluated at two distinctive points of the support, we have

$$\log \frac{f(\mu, \Delta\mu, p_1|y_{1:T})}{f(\mu^*, \Delta\mu^*, p_1^*|y_{1:T})} = \log \frac{f(y_{1:T}|\mu, \Delta\mu, p_1)f(\mu)f(p_1)}{f(y_{1:T}|\mu^*, \Delta\mu^*, p_1^*)f(\mu^*)f(p_1^*)} + \log \frac{f(\Delta\mu)}{f(\Delta\mu^*)}. \quad (11)$$

Note that for any $\Delta\mu \notin [\frac{-a}{2}, \frac{a}{2}]$ (and within the wider Uniform bound) and $\Delta\mu^* \in [\frac{-a}{2}, \frac{a}{2}]$, the ratio of the prior densities $\log \frac{f(\Delta\mu)}{f(\Delta\mu^*)}$ gives the penalty parameter P. Evaluating the parameters at their maximum likelihood estimates sheds light on how to set the penalty value since

$$\log \frac{f(\hat{\mu}, \Delta\hat{\mu}, \hat{p}_1|y_{1:T})}{f(\bar{\mu}, \Delta\mu = 0, \hat{p}_1|y_{1:T})} = \log \frac{f(y_{1:T}|\hat{\mu}, \Delta\hat{\mu}, \hat{p}_1)f(\hat{\mu})f(\hat{p}_1)}{f(y_{1:T}|\bar{\mu}, \Delta\mu = 0, \hat{p}_1)f(\bar{\mu})f(\hat{p}_1)} + P, \quad (12)$$

$$= \log \frac{f(y_{1:T}|\hat{\mu}, \Delta\hat{\mu}, \hat{p}_1)f(\hat{\mu})}{f(y_{1:T}|\bar{\mu}, \Delta\mu = 0, \hat{p}_1)f(\bar{\mu})} + P. \quad (13)$$

Using Equation (10) in Equation (13), the fact that $\log \frac{f(\hat{\mu}, \Delta\hat{\mu}, \hat{p}_1|y_{1:T})}{f(\bar{\mu}, \Delta\mu=0, \hat{p}_1|y_{1:T})}$ should at least be greater or equal to 0 when Equation (10) holds and that $\frac{f(\hat{\mu})}{f(\bar{\mu})} \approx 1$, we get

$$P \leq -\log \frac{0.95}{1 - 0.95} - \log T. \quad (14)$$

The penalty parameter should be lower or equal to $-\log \frac{0.95}{1-0.95} - \log T$ if we desire to have a higher posterior mode in the break support (i.e. $\Delta\mu \notin [\frac{-a}{2}, \frac{a}{2}]$) than the posterior mode in the no break domain in the situation where the Bayes factor in favor of the process M_2 would have been preferred at 95%. As this penalty rule is based on the BIC approximation, we denoted it as $P_{\text{BIC}} = -\log \frac{0.95}{1-0.95} - \log T$.

3.4 Inference

3.4.1 Prior elicitation

The Sparse CP-HAR model detects which parameters experience structural breaks. Only one estimation is needed to capture the breaks in the model parameters. The Sparse CP-HAR model only requires to set the number of breaks K to a sufficiently large value. This method avoids the standard procedure which consists of estimating a CP process given an increasing number of structural breaks and selecting the best model according to the marginal likelihood.

Table 3 displays the prior distributions of the Sparse CP-HAR parameters. Besides the HAR parameters that are driven by the 2MU distribution, the other parameters follow standard distributions. In particular, the transition probabilities and the state vector are specified

Table 3: Prior distributions and hyper-parameters.

Prior Distributions of the mean parameters

$$\{\beta_0, \beta_1, \beta_2, \beta_3\}' \sim N(\begin{pmatrix}0 & 0 & 0 & 0\end{pmatrix}', I_4)$$

For each regime $j \in [2, K+1]$ and for each mean parameter $i \in [0, 3]$, $\Delta\beta_{i,j} \sim 2MU(a_i, 10, P^{i,j}_{\text{Mean}})$

Prior Distributions of the variances

$$\sigma_1^{-2} \sim \text{G(1/5,5)}$$

For each regime $j \in [2, K+1]$, $\sigma_j^2 \sim 2MU(0, 100, P^j_\sigma)$

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 $i \in [0, 3]$ in each regime $j \in [2, K+1]$ $P^{i,j}_{\text{Mean}} \sim N(P_{BIC}, 0.5)$

For each regime $j \in [2, K+1]$ $P^j_\sigma \sim N(P_{BIC}, 0.5)$

The d-dimensional identity matrix is denoted by I_d. The shrinkage prior is denoted by $2MU(\,)$. The Gamma distribution is written $G(\,)$ and Beta( ) stands for the Beta distribution.

as in Chib (1998) and Pesaran, Pettenuzzo, and Timmermann (2006). The shrinkage priors are characterized by the penalty parameter P and the Uniform bounds a, b. The penalty parameter is taken as random and driven by a Normal distribution centered at the penalty value implied by the BIC as outlined in Section 3.3. Finally, the maximum number of breaks K, i.e. K_{Max}, is set to 8. In fact, since irrelevant regimes are shrunk to zero, this number is not relevant as long as it is above the observed number of breaks in the time series.

3.4.2 Algorithm

Given a number of breaks K, one can build an MCMC algorithm by sampling from the following posterior distributions $\pi(s_{1:T}|y_{1:T}, p_{1:K}, \Theta, \Sigma)$,$\pi(p_{1:K}|y_{1:T}, s_{1:T}, \Theta, \Sigma)$,$\pi(\Theta|y_{1:T}, s_{1:T}, p_{1:K}, \Sigma)$ and $\pi(\Sigma|y_{1:T}, s_{1:T}, p_{1:K}, \Theta)$ in a sequential way. Drawing from the first two conditional distributions are standard as explained in Chib (1998). However, the sparse prior distributions used for the mean and the variance paramters complicate the sampling from the conditional distributions $\pi(\Theta|y_{1:T}, s_{1:T}, p_{1:K}, \Sigma)$ and $\pi(\Sigma|y_{1:T}, s_{1:T}, p_{1:K}, \Theta)$. We explain in more detail how to sample from $\pi(\Theta|y_{1:T}, s_{1:T}, p_{1:K}, \Sigma)$, the sampling of $\pi(\Sigma|y_{1:T}, s_{1:T}, p_{1:K}, \Theta)$ is done similarly.

Sampling from $\pi(\Theta|y_{1:T}, s_{1:T}, p_{1:K}, \Sigma)$.

Due to the mixture of two components of our prior distributions, we cannot sample the mean parameter of a particular regime in one step. To circumvent the issue, we draw one at a time each mean parameter from its full conditional distribution. Consider a general sparse CP regression that embeds the Sparse CP-HAR model as a particular case:

$$y_t = \sum_{i=1}^{N}(\beta_i + \sum_{j=2}^{s_t} \Delta\beta_{i,j})x_{t,i} + \sqrt{\sigma_1^2 \prod_{j=2}^{s_t} \Delta\sigma_j^2 \epsilon_t} \quad \text{for} \quad s_t = 1, \ldots, K+1.$$

Sampling from $\pi(\beta_{1:N}|y_{1:T}, \Theta_{/\beta_{1:N}}, s_{1:T}, p_{1:K}, \Sigma)$ is standard if we assume a multivariate normal prior. The vector $\beta_{1:N}$ stands for $(\beta_1, \ldots, \beta_N)'$ and $\Theta_{/\beta_{1:N}}$ denotes the set of the mean parameters excluding $\beta_{1:N}$. The sampling complicates when it comes to drawing the parameters of the other regimes (i.e. $\Delta\beta_{i,j}$ with $j > 1$). To do so, we sample one parameter at a time. Let us focus on the density $\pi(\Delta\beta_{i,j}|y_{1:T}, \Theta_{/\Delta\beta_{i,j}}, s_{1:T}, p_{1:K}, \Sigma)$. To ease the sampling, we introduce the discrete variable $Pr(z_{\Delta\beta_{i,j}} = 0) = \omega$ that indicates which component of the

2MU mixture is active such that

$$f(\Delta\beta_{i,j}|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma) = \sum_{k=0}^{1} f(\Delta\beta_{i,j},z_{\Delta\beta_{i,j}}=k|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma)$$

$$= \sum_{k=0}^{1} f(\Delta\beta_{i,j}|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}}=k) f(z_{\Delta\beta_{i,j}}=k|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma).$$

To sample from $f(\Delta\beta_{i,j}|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma)$, one can draw $z_{\Delta\beta_{i,j}} \sim f(z_{\Delta\beta_{i,j}}|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma)$ and then $\Delta\beta_{i,j} \sim f(\Delta\beta_{i,j}|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}})$. Note that the latter density is a truncated normal distribution with known mean and variance. Sampling the discrete random variable can be done by first noting that

$$f(z_{\Delta\beta_{i,j}}=1|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma) = \frac{\frac{f(y_{1:T}|\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}}=1)}{f(y_{1:T}|\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}}=0)} f(z_{\Delta\beta_{i,j}}=1)}{\frac{f(y_{1:T}|\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}}=1)}{f(y_{1:T}|\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}}=0)} f(z_{\Delta\beta_{i,j}}=1) + f(z_{\Delta\beta_{i,j}}=0)}.$$

Secondly, for any $\Delta\beta_{i,j} \in [\frac{-a}{2},\frac{a}{2}]$, we have that

$$\frac{f(y_{1:T}|\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}}=1)}{f(y_{1:T}|\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}}=0)} = \frac{f(\Delta\beta_{i,j}|z_{\Delta\beta_{i,j}}=1)}{f(\Delta\beta_{i,j}|z_{\Delta\beta_{i,j}}=0)} \cdots \frac{f(\Delta\beta_{i,j}|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}}=0)}{f(\Delta\beta_{i,j}|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}}=1)}. \quad (15)$$

Finally, the ratio (15) is given by the ratio of the prior of the mean parameter times the ratio of two truncated normal densities evaluated at any point of the set $\Delta\beta_{i,j} \in [\frac{-a}{2},\frac{a}{2}]$. Algorithm 1 summarizes how to sample $\Delta\beta_{i,j}$ from its full conditional distribution.

Algorithm 1 Sampling one mean parameter $\Delta\beta_{i,j}$ from its full conditional distribution.

[1] Compute $\frac{f(y_{1:T}|\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}}=1)}{f(y_{1:T}|\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}}=0)}$ using Equation (15).

[2] Compute the probability $f(z_{\Delta\beta_{i,j}}=1|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma)$

[3] Sample $q \sim U[0,1]$. If $q \leq f(z_{\Delta\beta_{i,j}}=1|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma)$, set $z_{\Delta\beta_{i,j}}=1$. Otherwise set $z_{\Delta\beta_{i,j}}=0$.

[4] Sample $\Delta\beta_{i,j} \sim f(\Delta\beta_{i,j}|y_{1:T},\Theta_{/\Delta\beta_{i,j}},s_{1:T},p_{1:K},\Sigma,z_{\Delta\beta_{i,j}})$.

Sampling the variance parameters are done in a similar way. Algorithm 2 describes the full Gibbs sampler.

Algorithm 2 Gibbs sampler

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 $\Theta \sim f(\Theta | y_{1:T}, s_{1:T}, p_{1:K}, \Sigma)$:

- For the first regime, sample $\beta_{1:N} \sim f(\beta_{1:N} | y_{1:T}, s_{1:T}, p_{1:K}, \Sigma, \Theta_{/\beta_{1:N}})$.
- For each parameter $\Delta\beta_{i,j}$ of the other regimes (i.e. $j \geq 2$), sample $\Delta\beta_{i,j}$ using algorithm 1.

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

- For the first regime, sample $\sigma_1^2 \sim f(\sigma_1^2 | y_{1:T}, s_{1:T}, p_{1:K}, \Sigma_{/\sigma_1^2}, \Theta)$.
- For each parameter σ_j^2 of the other regimes (i.e. $j \geq 2$), sample σ_j^2 by adapting algorithm 1.

[3] - Sample $p_{1:K} \sim \prod_{i=1}^{K} Beta(\alpha_p + \sum_{t=1}^{T} \mathbb{1}_{s_t=i}, \beta_p + 1)$ 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 forward-backward sampler (see Chib (1998))

end for

The estimation scheme exposed in Algorithm 2 is sufficient to draw from the posterior distribution. In fact, our in-sample results documented in Section 5 are obtained using this algorithm. Nevertheless, MCMCs require lots of iterations to converge to the stationary distribution. While this drawback is not problematic per se, if we need several estimations of the model, it becomes an obstacle when hundreds estimations are required as it is usually the case for forecasting exercises. To solve this issue, our forecast evaluation is carried out by a Sequential Monte Carlo (SMC) sampler, see Del Moral, Doucet, and Jasra (2006). The technique combines Markov-Chain Monte Carlo (MCMC) and SMC methods and uses the importance sampling method to update the posterior approximation in light of additional observations. We refer to Dufays (2016) for more information.

4 Simulation study

To validate the MCMC algorithm, we explore the model estimation on two data generating processes (DGPs). We start with a DGP presented in Chan, Yip, and Zhang (2014) and reproduced by Equations (16)-(18) :

$$y_t = 0.9y_{t-1} + \epsilon_t \text{ when } t \leq 512, \quad (16)$$
$$= 1.69y_{t-1} - 0.81y_{t-2} + \epsilon_t \text{ if } 512 < t \leq 768, \quad (17)$$
$$= 1.32y_{t-1} - 0.81y_{t-2} + \epsilon_t \text{ when } 768 < t \leq 1024, \quad (18)$$

where $\epsilon_t \sim N(0,1)$. The process consists of an AR(2) process which experiences two change-points. Interestingly, the process exhibits a zero unconditional mean in all regimes. The first break appears in two parameters, namely the AR(1) coefficient and the AR(2) parameter while the second break only causes a switch in the AR(1) parameter. We simulate 100 series from the DGP and for each series, we estimate the Sparse CP model with an AR(2) specification and a maximum number of regimes $K_{Max} = 8$. The number of MCMC iterations is set to 55,000 and we keep the last 5,000 draws to summarize the posterior distribution. For the empirical applications below, we increase these numbers respectively to 120,000 and 20,000. Simulations with an AR(3) specification, not reported here to save space, give very similar results. In particular, 98% of the detected number of regimes for the (zero) AR(3) coefficient is equal to one.

Table 4: AR(2) DGP defined in (16)-(18)

Percentage of detected number of regimes									Break dates		
#Regime	1	2	3	4	5	6	7	8	Break	1	2
Shrinkage CP-AR(2) with $K_{\max} = 8$											
Intercept	**98**	1	1	0	0	0	0	0	Intercept	No break	
AR(1) coef.	0	0	**98**	1	1	0	0	0	AR(1) coef.	510.94 (5.76)	767.02 (4.19)
AR(2) coef.	0	**100**	0	0	0	0	0	0	AR(2) coef.	510.66 (5.72)	—
Variance	**100**	0	0	0	0	0	0	0	Variance	No break	
Standard CP-AR(2)											
	0	0	**99**	1	0	0	0	0		511.07 (6.00)	767.04 (4.39)

Notes: Left panel: Percentage of detected number of regimes per parameter obtained from 100 simulations and model estimations. Values in bold correspond to the correct number of regimes in the parameter. Right panel: Averages over each model estimation of the posterior mean of the break date as well as the standard deviations (in brackets) when the right number of regimes is detected.

Table 4 (left panel) summarizes the percentage of detected number of regimes, computed as the posterior mode of the marginal probabilities of having a specific number of regimes, per model parameter. In this specific example, the highest marginal probabilities coincide at least at 98% to the correct number of regimes. Regarding the break date estimation, Table

4 (right panel) documents the averages over each model estimation of the posterior mean of the break as well as its corresponding standard deviations when the right number of regimes is detected. These averages highlight the accurate estimation of the process.

It is worth mentioning that only 100 model estimations (carried out with a high number of regimes) have been required to obtain these results. This feature appears as a sharp improvement compared to the standard Bayesian CP methodology (e.g. Eo (2016)) which would have been computationally infeasible. In fact, it would have needed more than 6.5 billions model estimations conjugated with the estimation of the marginal likelihood to solve the same problem.[1] Nevertheless, we can still compare our results with the standard CP model of Chib (1998) presented in Section 3.1 which assumes that all the parameters vary when a break is detected. By fixing $K = 8$, this assumption limits the number of models to estimate to 800.

Table 4 (bottom) also summarises the estimation results obtained from the standard CP model on the same 100 simulated series. The results, here and for the next sections when referred to standard CP, are obtained with hyperparameters equal to $\Sigma_\beta = I_4$, $\{k_\sigma, \theta_\sigma\} = \{1/5, 5\}$,$\{\alpha_p, \beta_p\} = \{100, 1\}$. According to the marginal likelihood, 99% of the estimated model with the true number of regimes fit best the data. The average of the posterior break dates over these models (shown on the right panel) indicates an accurate detection of the break dates.

Table 5 sheds further light on the advantage of detecting which parameters switch from one regime to another by comparing the parameter estimates of the Sparse CP-AR(2) model with those of the standard CP-AR(2) process. As expected, the averages of the posterior means and their corresponding standard deviations are very precise for the Sparse setup. It turns out that although the standard CP-AR(2) model performs reasonably well in the parameter estimation on average, the standard deviations are substantially larger than those of the Sparse one.

The second DGP exposed in (19)-(20) is analogous to an estimated HAR model from the

[1] Assuming an identical $K_{Max} = 8$ gives $100(4^8) = 6553600$.

Table 5: AR(2) DGP defined in (16)-(18)

Sparse CP-AR(2) with $K_{\max} = 8$				Standard CP-AR(2)			
#Regime	1	2	3	#Regime	1	2	3
Intercept	0.00	No break		Intercept	-0.00	-0.00	0.00
	(0.03)	No break			(0.05)	(0.08)	(0.06)
AR(1) coef.	0.89	1.68	1.31	AR(1) coef.	0.90	1.66	1.31
	(0.05)	(0.03)	(0.03)		(0.05)	(0.13)	(0.05)
AR(2) coef.	-0.00	-0.81	No break	AR(2) coef.	-0.00	-0.79	-0.80
	(0.05)	(0.03)	No break		(0.05)	(0.10)	(0.04)
Variance	1.00	No break		Variance	1.01	1.37	1.01
	(0.04)	No break			(0.06)	(3.61)	(0.08)

Notes: Averages of the posterior means per parameter obtained from the simulations which detect the correct DGP for the two different models. True parameters are given in (16)-(18).

application below. This HAR model with time-varying variances is given by

$$y_t = 0 + 0.25 y_{t-1} + 0.45 y_{t-1}^{(w)} + 0.2 y_{t-1}^{(m)} + \epsilon_t \text{ with } \epsilon_t \sim N(0, \sigma_t^2), \tag{19}$$

$$\sigma_t^2 = \begin{cases} 0.27, & \text{if } t \leq 1734 \\ 0.39, & \text{if } 1734 < t \leq 2243 \\ 0.17, & \text{if } 2243 < t \leq 2433 \\ 0.45, & \text{if } 2433 < t \leq 3950. \end{cases} \tag{20}$$

The variance experiences three breaks corresponding to four regimes. The first break is the most difficult to detect because of the break magnitude and the short segment of the regime. The third regime only lasts 190 observations but the size of the break makes its detection more likely.

Table 6 provides the number of times over 100 simulations that a specific number of regimes per parameter is detected. As before, the detected number of regimes corresponds to the regime related to the mode of the marginal posterior probabilities over the regimes. We find that 99% percent of the estimations spells out that no break exists in the mean parameters which is in agreement with the DGP. Additionally, four regimes in the variance parameter have been detected 84% of the time and 100% of the estimations claims that there is a break in the variance. Interestingly, the number of breaks in the variance is never overestimated. Table 6 also documents the results for the standard CP model. Relying on the marginal likelihood, the standard CP approach mostly detect two regimes in the series and only 19% of the estimations recovers the true DGP. Note that the results would have been even more distorted if the prior distributions of the CP parameters had been more diffuse. In fact, the marginal likelihood acts as an Ockham's razor through the discrepancy between the posterior and the prior distributions of the parameters. In fact, applying the Bayes' theorem, the marginal log-likelihood is given by $\log f(y_{1:T}) = \log f(y_{1:T}|\theta) + [\log f(\theta) - \log f(\theta|y_{1:T})]$ when θ denotes the model parameters.

Table 6: HAR DGP defined in (19)-(20): Percentage of detected regimes

#Regime	1	2	3	4	5	6	7	8
Sparse CP-HAR with $K_{\max} = 8$								
Intercept	**99**	1	0	0	0	0	0	0
AR(1) coef.	**100**	0	0	0	0	0	0	0
Weekly coef.	**100**	0	0	0	0	0	0	0
Monthly coef.	**100**	0	0	0	0	0	0	0
Variance	0	8	8	**84**	0	0	0	0
Standard CP-HAR model								
	0	**63**	18	19	0	0	0	0

Notes: Percentage of detected number of regimes per parameter obtained from 100 simulations and model estimations. Values in bold correspond to the correct number of regimes in the parameter.

Table 7: HAR DGP defined in (19)-(20): Break dates

Break	1	2	3
Sparse CP-HAR with $K_{\max} = 8$			
Intercept		No break	
AR(1) coef.		No break	
Weekly coef.		No break	
Monthly coef.		No break	
Variance	1667.86	2188.73	2383.60
	(83.58)	(17.66)	(13.60)
Standard CP-HAR model			
	1798.11	2198.95	2385.84
	(76.03)	(16.14)	(9.38)

Notes: Averages over each model estimation of the posterior mean of the break date as well as the standard deviations (in brackets) when the right number of regimes is detected.

Table 7 documents the averages over the posterior means of the break dates as well as the corresponding standard deviations when the right number of regimes is detected. The detected change-points in the variance parameters closely track the true ones. The averages over the 19 posterior means of the break dates obtained from the standard CP model give similar results as the sparse setup although we expect that the parameter estimates are less accurately estimated. Table 8 confirms this. The parameter estimates are very accurate for the sparse model (left panel) compared to the values given by the standard CP approach (right panel).

Table 8: HAR DGP defined in (19)-(20)

Sparse CP-HAR with $K_{\max} = 8$					Standard CP-HAR				
#Regime	1	2	3	4	#Regime	1	2	3	4
Intercept	0.00		No break		Intercept	0.00	-0.01	0.01	-0.00
	(0.01)		No break			(0.01)	(0.12)	(0.07)	(0.01)
AR(1) coef.	0.25		No break		AR(1) coef.	0.25	0.24	0.08	0.05
	(0.02)		No break			(0.03)	(0.07)	(0.12)	(0.10)
Weekly coef.	0.44		No break		Weekly coef.	0.44	0.41	0.15	0.08
	(0.03)		No break			(0.04)	(0.15)	(0.22)	(0.18)
Monthly coef.	0.20		No break		Monthly coef.	0.20	0.15	0.04	0.04
	(0.03)		No break			(0.04)	(0.16)	(0.14)	(0.08)
Variance	0.27	0.40	0.17	0.45	Variance	0.28	0.47	0.11	0.09
	(0.01)	(0.03)	(0.02)	(0.02)		(0.01)	(0.47)	(0.17)	(0.18)

Notes: Averages of the posterior means per parameter obtained from the simulations having detected the correct DGP for the two different models.

5 Application to international stock market realized variances

In the empirical application, we first estimate the CP-HAR models on the full sample to evaluate the presence of breaks in realized variance. Next, we evaluate the forecast performance of the sparse CP-HAR model by comparing it to the standard HAR model.

5.1 Are there relevant breaks in realized volatility?

We start by estimating the standard CP-HAR model on the full sample all the series. Table 9 documents the number of detected regimes and the break dates. Besides DAX and AEX, there is clear evidence in favour of CP models making relevant to allow abrupt changes in the parameter values. However, some of the regimes are very short indicating that they capture extreme values. In particular, the EuroStoxx realized log-variances exhibit two very short regimes while the FTSE, the SMI and the NIKKEI realized log-variances present both one outlier. If we do not account for these short regimes, we still have 9 over 11 series that exhibit at least two long regimes. This encourages the use of CP models for fitting and forecasting the series.

Table 9: Number of detected regimes according to the standard CP-HAR approach.

Series	# Regimes	Break dates				
AEX	1					
CAC	4	2001-10-31	2005-08-03	2005-08-26		
DAX	1					
DJI	5	2007-02-26	2007-03-06	2009-03-10	2009-09-30	
EuroStoxx	6	2013-04-30	2013-05-02	2014-03-25	2014-05-01	2014-06-05
FTSE	4	2004-02-18	2005-07-06	2005-07-08		
NASDAQ	5	2000-04-26	2007-05-11	2009-01-26	2009-09-03	
NIKKEI	4	2011-03-11	2011-03-24	2011-12-30		
Russell	6	2001-09-07	2001-09-24	2007-03-14	2009-01-28	2009-09-21
SMI	6	2003-04-02	2005-07-06	2005-07-08	2015-01-14	2015-01-15
SP500	4	2007-02-05	2009-02-10	2009-09-30		

Notes: Number of regimes according to the marginal likelihood.

Inference on the Sparse CP models reveals a different picture on the structural breaks in realized variance series. Table 10 exhibits the posterior mode of the number of regimes per parameter as well as the posterior probability of detecting this respective number of regimes. Interestingly, besides in the Russell, no break is detected in the mean parameters. Moreover, as opposed to the standard CP process, results on the DAX and on the AEX emphasize that breaks do occur over the period but only in the variance.

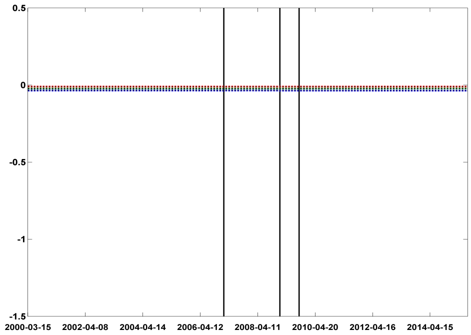
To appreciate visually these results, we first focus on the case of the S&P 500 realized variance. Figure 2 displays the posterior mean of the break dates detected by the standard CP-HAR model with respect to the median (as well as the 10% and the 90% quantiles) of the HAR parameters over time obtained from the Sparse CP-HAR model. The two processes detect the same regimes. While the standard CP-HAR model exhibits 20 parameters, the shrinkage process abates this number by only changing the variance parameter when a break is detected. In fact, with only 8 parameters differing from zero, the Sparse CP-HAR model automatically balances between parsimony and flexibility.

To pursue our analysis further, Figure 3 presents the same results as for the S&P 500 but for the Russell realized variance. Interestingly, the persistence of the series changes in 2006 as emphasized by the switch of the AR(1) coefficient over that period. We observe that the

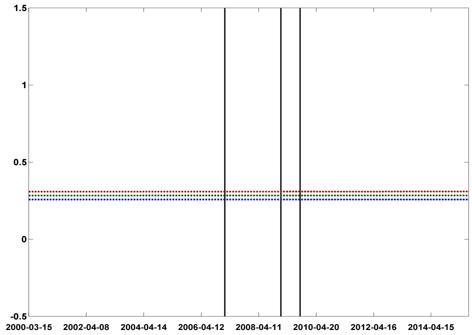
Table 10: Posterior mode of the number of regimes as well as the related posterior probability for each model parameter.

Series	Shrinkage CP-HAR ($K_{Max} = 8$)					# Regimes
	β_0	β_1	β_2	β_3	σ^2	Standard CP
AEX	1	1	1	1	4	1
	(0.99)	(1.00)	(1.00)	(1.00)	(0.72)	
CAC	1	1	1	1	5	4
	(1.00)	(1.00)	(0.99)	(0.99)	(0.54)	
DAX	1	1	1	1	5	1
	(0.98)	(0.99)	(1.00)	(1.00)	(0.71)	
DJI	1	1	1	1	4	5
	(0.99)	(1.00)	(1.00)	(1.00)	(0.99)	
EuroStoxx	1	1	1	1	8	6
	(0.99)	(0.99)	(0.99)	(0.99)	(1.00)	
FTSE	1	1	1	1	6	4
	(0.99)	(0.93)	(1.00)	(1.00)	(0.70)	
NASDAQ	1	1	1	1	5	5
	(0.98)	(1.00)	(1.00)	(1.00)	(0.98)	
NIKKEI	1	1	1	1	5	4
	(0.94)	(0.99)	(0.99)	(1.00)	(0.43)	
Russell	2	2	1	1	4	6
	(0.53)	(1.00)	(0.59)	(0.99)	(0.82)	
SMI	1	1	1	1	6	6
	(0.99)	(1.00)	(1.00)	(0.99)	(0.87)	
SP500	1	1	1	1	4	4
	(0.99)	(1.00)	(1.00)	(1.00)	(0.59)	

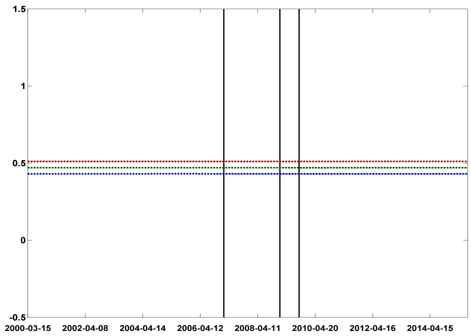
Notes: Number of regimes per parameters according to the Sparse CP process. The probabilities are in brackets.



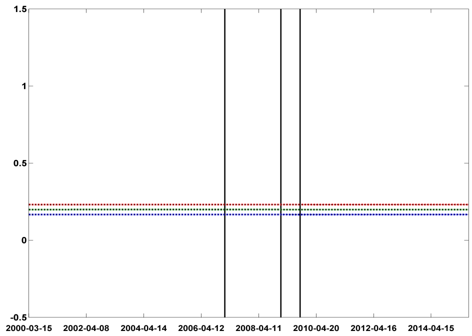
(a) β_0



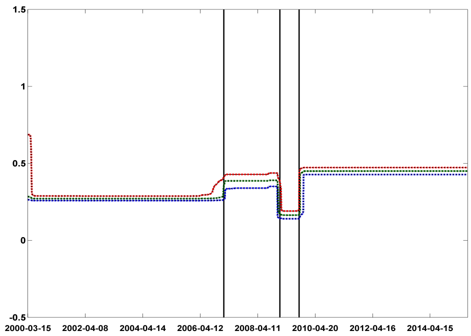
(b) β_1



(c) β_2

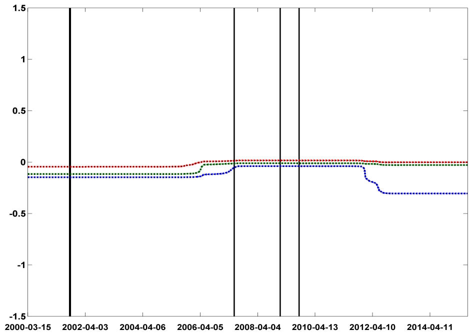


(d) β_3

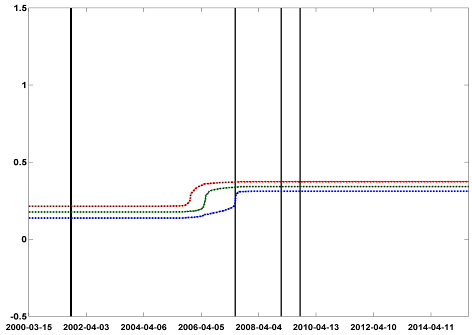


(e) Variance

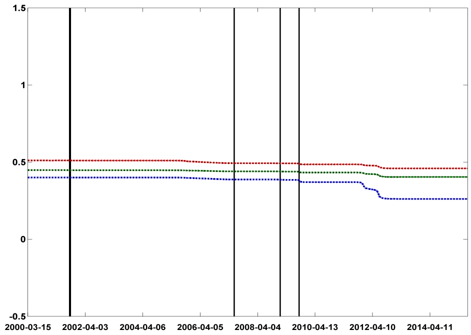
Figure 2: S&P 500 log realized variances: posterior mean of the break dates estimated with the standard CP-HAR model (vertical lines) with respect to the posterior median (as well as the 10% and the 90% quantiles) of the HAR parameters over time given the Sparse CP-HAR process.



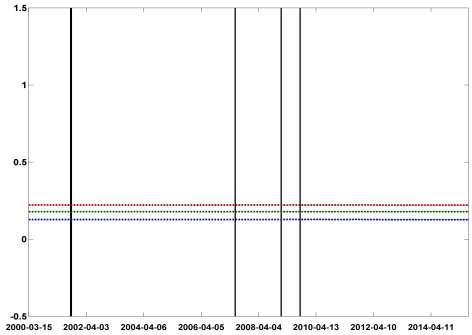
(a) β_0



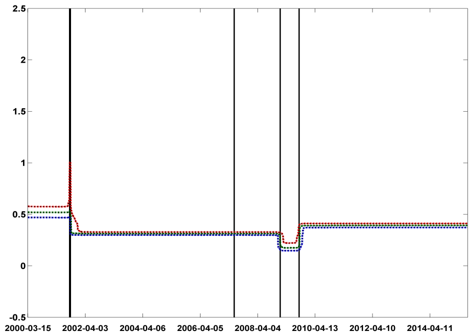
(b) β_1



(c) β_2



(d) β_3



(e) Variance

Figure 3: Russell log realized variances: posterior mean of the break dates estimated with the standard CP-HAR model (vertical lines) with respect to the posterior median (as well as the 10% and the 90% quantiles) of the HAR parameters over time given the Shrinkage CP-HAR process.

standard CP-HAR model captures the same breaks as the Sparse CP model. In particular, the sparse approach highlights that all but the second break appear in the variance parameter.

Table 11 gives a clear picture of the trade-off between the parsimony and the flexibility reached by the Shrinkage process in case of the Russell index. We document the posterior means of the parameters, and the corresponding standard deviations, for the draws having the number of breaks equivalent to the mode of the marginal probabilities presented in Table 10. As expected, the Sparse CP process (involving 10 parameters) is parsimonious compared to the standard CP process (involving 30 parameters). However, inspecting the parameter estimates, the regimes of the two processes are similar. This holds particularly well for the variance parameters which experience most of the breaks as revealed by the Sparse CP model.

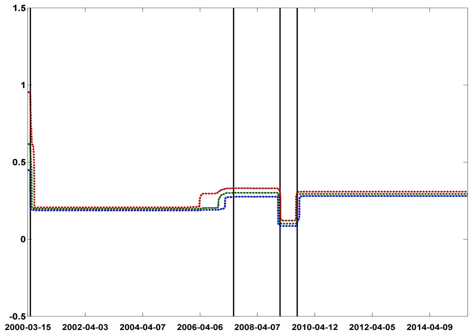
Table 11: Russell realized variances: Shrinkage and CP-HAR parameters

Sparse CP-HAR with $K_{\max} = 8$					Standard CP-HAR					
#Regime	1	2	3	4	1	2	3	4	5	6
Intercept	-0.12	-0.01	No break		-0.08	0.63	-0.12	0.06	-0.10	-0.06
	(0.02)	(0.02)	No break		(0.05)	(0.75)	(0.03)	(0.04)	(0.06)	(0.02)
AR(1) coef.	0.18	0.34	No break		0.20	-0.32	0.13	0.46	-0.05	0.38
	(0.03)	(0.02)	No break		(0.06)	(0.46)	(0.04)	(0.06)	(0.10)	(0.03)
Weekly coef.	0.43	No break			0.60	0.52	0.49	0.40	0.45	0.30
	(0.03)	No break			(0.10)	(0.89)	(0.06)	(0.08)	(0.18)	(0.05)
Monthly coef.	0.18	No break			0.07	-0.36	0.17	0.07	0.62	0.24
	(0.04)	No break			(0.09)	(0.79)	(0.06)	(0.07)	(0.17)	(0.04)
Variance	0.52	0.40	0.18	0.39	0.49	3.28	0.31	0.32	0.15	0.39
	(0.04)	(0.39)	(0.04)	(0.02)	(0.04)	(305.10)	(0.01)	(0.02)	(0.02)	(0.01)

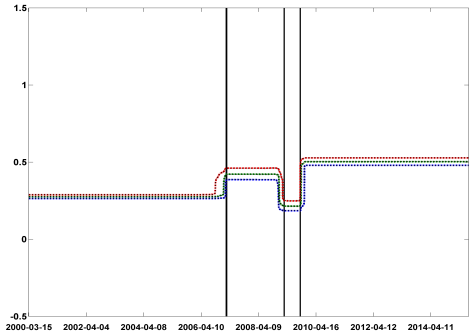
We finally emphasise that the similarity in the break patterns between the standard CP model and the Sparse CP model is not a particular case of the S&P 500 and the Russell realized variances only. Figure 4 shows the variance estimates over time of the Shrinkage model for four other stock market indices (Nasdaq, Dow Jones, SMI and EuroStoxx). The break dates obtained by the standard CP model (and depicted by the vertical lines) are also captured by the Sparse CP process. In addition to that, the Sparse CP model also enables to see that only switches in the variance would have been sufficient for these series.

5.2 Forecast evaluation

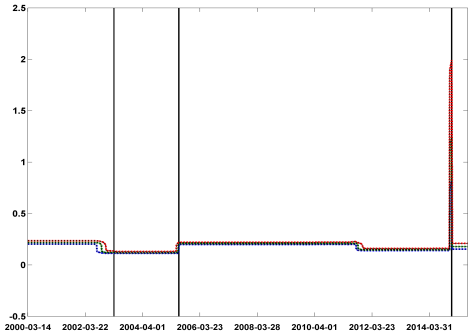
Forecasting using the Sparse CP-HAR model is a straightforward byproduct of the sampling scheme. For each of the posterior draws of the model parameters, we simulate the distribution of the future observations by assuming that the model remains in the last regime. More specifically, to sample the marginal posterior distribution of the h-ahead forecast, we have



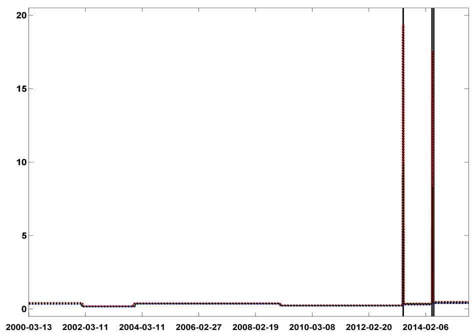
(a) NASDAQ: Variance over time



(b) Dow Jones: Variance over time



(c) SMI: Variance over time



(d) EuroStoxx: Variance over time

Figure 4: Posterior mean of the break dates estimated with the standard CP-HAR model (vertical lines) with respect to the posterior median (as well as the 10% and the 90% quantiles) of the variance parameter over time given the Sparse CP-HAR process. The mean parameters of these series do not experience breaks.

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 posterior draws of the parameters, one can simulate a h-ahead prediction by simulating sequentially for $q = 1, ..., h$ as follows

$$y_{T+q}|y_{1:T+q-1}, s_{1:T}, p_{1:K}, \Theta, \Sigma \sim N(\beta_0 + \beta_{1,K+1} y_{T+q-1} + \beta_{2,K+1} y^{(5)}_{T+q-1} + \beta_{3,K+1} y^{(22)}_{T+q-1}, \sigma^2_{K+1}),$$

in which $\beta_{i,K+1} = \beta_{i,1} + \sum_{j=2}^{K+1} \Delta\beta_{i,j}$ for $i \in [0, 3]$. Note that as in Pesaran, Pettenuzzo, and Timmermann (2006) we could also implement forecasting subject to future breaks.

Following a similar forecasting setup as in Bauwens, Koop, Koribilis, and Rombouts (2015), our study consists in starting with the first eighty 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. We consider daily forecast horizons (h) of 1, 2, 5, 10, and 25. As loss functions, we compute the root mean squared error (RMSE) derived from the posterior predictive mean as well as the average of the predictive densities obtained at the different horizons ,i.e. $\sum_{t=\tau}^{T-h} f(y_{t+h}|y_{1:t})/(T-h-\tau+1)$ with τ indicating the first out-of-sample date.

Table 12 presents the forecast evaluation results. In terms of RMSE, the HAR and Sparse CP-HAR models show similar performance for all the horizons. This holds true for each of the eleven stock market indices. In fact, since for most series the breaks in realized variances are in the conditional variance rather than the conditional mean, we can expect that both approaches deliver very similar point forecasts. Regarding the average predictive likelihood, we have similar findings as for the RMSE loss function. For some indices and short horizons, the average predictive likelihood is slightly higher for the standard HAR model, the differences being smaller than five percent though.

Finally, we also did the forecast evaluation on a subsample covering the start of the global financial crisis (the day after the Lehman Brother's collapse) and the two subsequent years. In terms of RMSE, the results are close to the full forecast evaluation results above. In terms of average predictive likelihood, it turns out that for horizons up to 10 days the Sparse CP-HAR model mildly outperforms the HAR model.

6 Conclusion

We enrich the HAR model by allowing for multiple structural breaks in the parameters. The structural breaks in our model can be such that they appear only in some of the mean and/or variance parameters. Our method allows estimating parameters with no breaks with the full sample and others parameters with the relevant subsamples. Which parameters break, is identified automatically with our approach. Sparsity is achieved thanks to employing a shrinkage prior tailored to change-point models. We propose a Bayesian sampler to perform inference on this new sparse CP-HAR model. A detailed simulation study illustrates the excellent performance of the sampler compared to standard CP approaches where all parameters change

Table 12: RMSE and Average predictive likelihood

	RMSE					Average predictive likelihood				
Horizon	1	2	5	10	25	1	2	5	10	25
SP500										
HAR	0.67	0.74	0.84	0.89	0.96	0.45	0.41	0.37	0.34	0.30
Sparse CP-HAR	0.67	0.74	0.84	0.89	0.96	0.42	0.38	0.34	0.32	0.28
Russell										
HAR	0.64	0.70	0.77	0.80	0.87	0.45	0.41	0.38	0.35	0.32
Sparse CP-HAR	0.64	0.70	0.77	0.81	0.88	0.45	0.40	0.37	0.34	0.30
NASDAQ										
HAR	0.56	0.63	0.71	0.77	0.84	0.52	0.46	0.42	0.39	0.34
Sparse CP-HAR	0.56	0.63	0.71	0.77	0.83	0.51	0.45	0.41	0.38	0.34
SMI										
HAR	0.44	0.48	0.56	0.64	0.72	0.66	0.61	0.53	0.46	0.40
Sparse CP-HAR	0.44	0.48	0.56	0.64	0.72	0.69	0.63	0.55	0.48	0.42
NIKKEI										
HAR	0.58	0.64	0.69	0.73	0.82	0.54	0.49	0.45	0.42	0.36
Sparse CP-HAR	0.58	0.64	0.70	0.74	0.83	0.51	0.47	0.43	0.40	0.34
FTSE										
HAR	0.47	0.51	0.57	0.61	0.69	0.58	0.54	0.48	0.44	0.38
Sparse CP-HAR	0.48	0.51	0.57	0.61	0.68	0.58	0.54	0.49	0.44	0.38
EuroStoxx										
HAR	0.74	0.76	0.81	0.86	0.91	0.50	0.46	0.42	0.38	0.34
Sparse CP-HAR	0.74	0.76	0.81	0.86	0.91	0.48	0.44	0.40	0.36	0.32
DJI										
HAR	0.70	0.75	0.83	0.88	0.94	0.43	0.40	0.37	0.34	0.30
Sparse CP-HAR	0.69	0.75	0.83	0.88	0.94	0.40	0.38	0.34	0.32	0.28
DAX										
HAR	0.57	0.61	0.65	0.72	0.78	0.53	0.49	0.44	0.39	0.36
Sparse CP-HAR	0.57	0.61	0.65	0.72	0.78	0.52	0.48	0.44	0.39	0.35
CAC										
HAR	0.52	0.57	0.63	0.69	0.75	0.57	0.52	0.47	0.42	0.37
Sparse CP-HAR	0.52	0.56	0.63	0.69	0.75	0.56	0.51	0.46	0.41	0.36
AEX										
HAR	0.53	0.57	0.64	0.70	0.77	0.57	0.51	0.46	0.40	0.35
Sparse CP-HAR	0.53	0.57	0.64	0.70	0.77	0.57	0.51	0.46	0.41	0.35

when there is a break.

We study the stability of the HAR model using realized volatilities series of eleven international indices between January 2000 and August 2015. In-sample, the sparse CP-HAR model reveals that there is typically no change at all in the HAR model parameters, i.e. the breaks are almost uniquely in the variance of the realised variance series. This feature would be left undetected when estimating CP models with the standard approach. In an out-of-sample forecasting exercise, we find similar performance between the standard HAR model and the sparse CP-HAR model.

References

Ait-Sahalia, Y., and J. Jacod (2014): *High-Frequency Financial Econometrics*. Princeton University Press, Princeton.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96(453), 42–55.

Audrino, F., L. Camponovo, and C. Roth (2015): "Testing the lag structure of assets' realized volatility dynamics," *SSRN Working paper*.

Audrino, F., and S. D. Knaus (2016): "Lassoing the HAR Model: A Model Selection Perspective on Realized Volatility Dynamics," *Econometric Reviews*, 35(8-10), 1485–1521.

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.

Bekaert, G., and M. Hoerova (2014): "The VIX, the variance premium and stock market volatility," *Journal of Econometrics*, 183(2), 181–192.

Bollerslev, T., J. Marrone, L. Xu, and H. Zhou (2014): "Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence," *Journal of Financial and Quantitative Analysis*, 49, 633–661.

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. (1995): "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.

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

Corsi, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7(2), 174–196.

Corsi, F., F. Audrino, and R. Reno (2012): *HAR Modelling for Realized Volatility Forecasting*. In Volatility Models and Their Applications. Wiley, Princeton.

Corsi, F., N. Fusari, and D. L. Vecchia (2013): "Realizing smiles: Options pricing with realized volatility," *Journal of Financial Economics*, 107(2), 284 – 304.

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. (2016): "Evolutionary Sequential Monte Carlo for Change-point models," *Econometrics*, 4(1).

Dufays, A., and J. Rombouts (2015): "Sparse change-point time series models," CORE Discussion paper 2015/32.

Eo, Y. (2016): "Structural changes in inflation dynamics: multiple breaks at different dates for different parameters," *Studies in Nonlinear Dynamics and Econometrics*, 20 (3).

Gallo, G. M., and E. Otranto (2015): "Forecasting realized volatility with changing average levels," *International Journal of Forecasting*, 31(3), 620 – 634.

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

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

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

LIU, C., AND J. M. MAHEU (2008): "Are there structural breaks in realized volatility?," *Journal of Financial Econometrics*, 6(3), 326–360.

MAHEU, J., AND Y. SONG (2013): "A new structural break model, with an application to canadian inflation forecasting," *International journal of forecasting*, 30, 144–160.

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

MCALEER, M., AND M. C. MEDEIROS (2008a): "A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries," *Journal of Econometrics*, 147(1), 104 – 119, Econometric modelling in finance and risk management: An overview.

——— (2008b): "Realized Volatility: A Review," *Econometric Reviews*, 27(1-3), 10–45.

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

PESARAN, M. H., D. PETTENUZZO, AND A. TIMMERMANN (2006): "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, 73, 1057–1084.