

# **A Mixture Innovation Heterogeneous Autoregressive Model for Structural Breaks and Long Memory**

**Nima Nonejad**

**CREATES Research Paper 2013-24**

# A Mixture Innovation Heterogeneous Autoregressive Model for Structural Breaks and Long Memory

Nima Nonejad<sup>†\*</sup>

Aarhus University and CREATES

## Abstract

We propose a flexible model that is able to simultaneously approximate long memory behavior as well as incorporate structural breaks in the model parameters. Our model is an extension of the heterogeneous autoregressive (HAR) model, which is designed to model and forecast volatility of financial time series. In an extensive empirical evaluation involving several volatility series, we demonstrate presence of structural breaks and their importance for forecasting. Furthermore, we find that the choice of how to model break processes is important in achieving good forecast performance.

**Keywords:** Bayesian inference, Forecasting, Mixture innovation models, Realized volatility  
(*JEL*: C11, C22, C51, C53)

---

\*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. I would like to thank Asger Lunde for his useful considerations, comments and providing me with the high frequency data. First draft: 08/16/2013, this version 04/09/2014.

# 1 Introduction

Measuring and modeling the conditional variance or volatility of financial time series is an important issue in econometrics. General approaches of estimating volatility are based on parametric models such as generalized autoregressive conditional heteroskedasticity (GARCH) models proposed by Engle (1982) and Bollerslev (1986) or stochastic volatility (SV) models as in Kim et al. (1998). However, during the past two decades, the approach of using improved measures of ex post volatility constructed from high-frequency data has become very popular. This measure is called realized volatility (RV) and it is discussed formally by Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002 a,b).

This paper proposes a simple model that merges long memory dynamics and nonlinearities. The specification that is put forward is a generalization of the Heterogeneous Autoregressive (HAR) model of Corsi (2009). The HAR model has been applied with success in modeling and forecasting realized volatility, see for example Andersen et al. (2007). Our model is called the Mixture Innovation Heterogeneous Autoregressive (MHAR) model and combines ingredients from HAR and mixture innovation (MIA) models, see Gerlach et al. (2000), Giordani and Kohn (2008) and Groen et al. (2012). This approach builds on the state space representation, modeling the breaks through mixture distributions in state innovations of linear Gaussian state space models. Indeed, this approach is very intuitive and has several desirable features such as: the possibility of including a random number of breaks that can occur in the sample, the possibility of jointly modeling small possibly frequent and large possibly less frequent breaks, while allowing different parameters to change at different points in time.

It is an open question if the MHAR model will perform well when dealing with the sort of structural changes present in realized volatility series. The main purpose of this paper is to shed light on this question. Therefore, we compare the performance of the MIA specification with an existing method of modeling structural breaks, namely, the change-point specification of Chib (1998), henceforth CPHAR, given in for example Liu and Maheu (2008). In addition, we also include alternative forecasting procedures such as: recursively estimating the HAR model and a random walk time-varying parameter HAR model.

We believe that applying the MIA specification using realized volatility data and comparing its performance with the change-point specification which in the literature is considered as a “state-of-the-art” structural break model is in fact the most important contribution that we make since basically no work has been done on comparing these methods. MHAR and CPHAR differ in important aspects, like for instance their treatment of the break process. The change-point specification imposes the restriction that a precise number of breaks occurs in the sample whereas for the MIA specification the number of breaks is treated as unknown.

For eleven realized volatility series between 2004 and 2009 we consider the aforementioned models and produce daily, weekly and biweekly forecasts. We evaluate forecast performance using two criteria: predictive likelihood (PL), see Geweke (2005) and root mean squared error (RMSE) both which are easily obtained within the Bayesian estimation procedure. It turns out that these two loss functions lead to similar qualitatively conclusions. Overall, structural breaks play an important role for forecasting in all of the volatility series that we consider. Specifically, we find that each structural break specification outperforms the HAR model regardless of criterion or forecast horizon. Furthermore, we find that for longer forecast horizons the MHAR model with time-varying volatility tends to perform better. However, we find that there is no one single method which can be recommended universally, i.e. for all series and all forecast horizons. For some series CPHAR performs better than MHAR specifications and vice versa.

The structure of this paper is as follows. Section 2 discusses the econometric issues for Bayesian estimation. Section 3 reviews the theory behind the volatility measures used in this paper. Section 4 briefly presents the HAR model. Details on the data is presented in section 5. Section 6 discusses the empirical results and finally the last section concludes.

## 2 Modeling Structural Breaks

### 2.1 Change-point model

The models considered in this paper use the framework of a Gaussian linear regression model

$$y_t = X_{t-1}\beta + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (2.1)$$

for  $t = 1, \dots, T$ . Let  $Y_T = (y_1, \dots, y_T)'$  be a vector of size  $T$  and  $X_T$  be a  $T \times k$ , matrix of regressors with row  $X_{t-1}$  which can also include lags of  $y_t$ . Obviously, different structural break models vary in the way they model breaks in (2.1) by allowing  $\beta$  and possibly  $\sigma^2$  to vary with time. To begin with, we focus mainly on structural breaks in the regression coefficients,  $\beta$  and assume that  $\sigma^2$  is fixed through time. However, breaks in  $\sigma^2$  can also be modeled rather easily. We consider two main approaches for modeling structural breaks in  $\beta$  and later in both  $\beta$  and  $\sigma^2$ .

We start by considering the change-point (or structural break) specification proposed by Chib (1998). This specification uses a hidden Markov model with a restricted transition matrix to model the change points. A test for the number of structural breaks is then a test of the dimension of the hidden Markov chain.

Assume that there are  $m - 1$ ,  $m \in \{1, 2, \dots\}$  change points at unknown times,  $\Omega_m = \{\tau_1, \tau_2, \dots, \tau_{m-1}\}$ . Separated by those change points, there are  $m$  different regimes. The density of observation  $y_t$ ,  $t = 1, \dots, T$  depends on  $\beta_j$ ,  $j = 1, 2, \dots, m$  whose value changes at the change points,  $\Omega_m$  and  $\sigma^2$ .

Thus,

$$\beta_t = \begin{cases} \beta_1 & \text{if } t < \tau_1 \\ \beta_2 & \text{if } \tau_1 \leq t < \tau_2 \\ \vdots & \vdots \\ \beta_{m-1} & \text{if } \tau_{m-2} \leq t < \tau_{m-1} \\ \beta_m & \text{if } \tau_{m-1} \leq t \end{cases} \quad (2.2)$$

Let  $S = (s_1, \dots, s_T)'$  denote the unobserved state system where  $s_t = j$  indicates that  $y_t$  is from regime  $j$  and follows the conditional distribution  $p(y_t | \beta_j, \sigma^2, Y_{t-1})$ . The one-step ahead transition probability matrix for  $s_t$ ,  $P$  is assumed to be

$$\begin{aligned} p(s_t = j | s_{t-1} = j) &= p_j \\ p(s_t = j+1 | s_{t-1} = j) &= 1 - p_j \end{aligned} \quad (2.3)$$

for  $j = 1, \dots, m-1$  and  $p(s_t = m | s_{t-1} = m) = 1$ . The other elements of  $P$  are set to zero. Hence, if regime  $j$  holds at time  $t-1$ , then at time  $t$  the process can either remain in regime  $j$  (with probability  $p_j$ ) or a break occurs and the process moves to regime  $j+1$  (with probability  $1-p_j$ ). Once the last regime is reached, one stays there forever. This structure enforces the ordering (2.2) on the change points<sup>1</sup>. Technical details about estimation of (2.1) using (2.2) is provided in the appendix<sup>2</sup>.

As stated in the introduction, (2.2) requires a fixed number of structural breaks to occur<sup>3</sup>. In this paper we follow Pesaran et al. (2006), Liu and Maheu (2008), Bauwens et al. (2011) and estimate models for different number of structural breaks (0 to 7 in our empirical application). Then, we compare results across these models using the marginal likelihood criterion. Specifically, let models with  $i$  and  $j$  structural breaks be denoted by  $M_i$  and  $M_j$  respectively. For each specification we can calculate the marginal likelihood,  $p(y_1, \dots, y_T | M_i)$  following Chib (1995) and rank models by means of their Bayes factor,  $BF_{ij} = p(y_1, \dots, y_T | M_i) / p(y_1, \dots, y_T | M_j)$ . Large values of  $BF_{ij}$  indicate that data supports  $M_i$  over  $M_j$ . In the appendix, we show how to compute the marginal likelihood of the change-point regression model using the method of Chib (1995).

<sup>1</sup>Theoretically, we could allow for breaks in each element of  $\beta$  to occur independently. In this case,  $s_t$  will be a  $k \times 1$  discrete random variable with the first element controlling breaks in  $\beta_1$ , the second element controlling breaks in  $\beta_2$  and so on. Furthermore, in order to ease the notation conditioning on  $X_T$  is suppressed.

<sup>2</sup>Notice that our specification is identical to Liu and Maheu (2008) but differs slightly from the hierarchical prior specification (on the conditional mean, variance and the prior on the regime durations) of Pesaran et al. (2006). However, since we perform direct forecasting it does not make any difference which specification is used.

<sup>3</sup>Koop and Potter (2007) argue that (2.3) may be restrictive in some situations. They suggest the use of the more flexible Poisson distribution for the durations. However, in order to avoid the very heavy additional computational cost inherent to the use of the Poisson prior we choose only to implement the change-point model using (2.3).

## 2.2 Mixture innovation model

An alternative specification that allows for structural breaks in the regression parameters of (2.1) can be defined in the following way

$$y_t = X_{t-1}\beta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (2.4)$$

where  $\beta_t = (\beta_{1t}, \dots, \beta_{kt})'$  is a vector of time-varying regression parameters and each element of  $\beta_t$  evolves according to the following form

$$\beta_{it} = \beta_{it-1} + \kappa_{it}\eta_{it}, \quad \eta_{it} \sim N(0, q_i^2), \quad i = 1, \dots, k \quad (2.5)$$

where  $\kappa_{it} = \{0, 1\}$  is an unobserved process with  $p(\kappa_{it} = 1) = \pi_i$ ,  $K_t = (\kappa_{1t}, \dots, \kappa_{kt})$ ,  $K = \{K_t\}_{t=1}^T$  and  $B = \{\beta_t\}_{t=1}^T$ . In this framework the  $i$ th parameter,  $\beta_{it}$  remains the same as its previous value,  $\beta_{it-1}$  unless  $\kappa_{it} = 1$  in which it changes with  $\eta_{it}$ .

This specification implies that  $\beta_{it}$  in (2.5) is allowed to change every time period, but it does not necessarily need to change at all. Also, changes in the separate parameters are not restricted to coincide as in the change-point model. Rather, changes in each  $\beta_{it}$  are allowed to occur at different points in time. Furthermore, as stated in section 2.1, for the change-point model we specify the number of structural breaks by comparing Bayes factors for different specifications. Here, we estimate only one specification and allow the data to determine the nature of the structural breaks in each element of  $\beta_t$ . In the appendix, we provide details on how to estimate the MIA model using Gibbs sampling. Finally, we note that if  $\kappa_{1t} = \kappa_{2t} \dots \kappa_{kt} = 1$  for  $t = 1, \dots, T$  then we get the very well-known time-varying parameter (TVP) model where  $\beta_{it}$  varies according to a random walk.

## 2.3 Breaks in the conditional variance

In this section we describe how to model structural breaks in the conditional variance,  $\sigma^2$  for CP and MIA specifications.

First, for the change-point model in order to estimate the parameters for each regime we use only observations in regime  $j$ . Therefore, let  $\hat{y}_j = \{y_t : s_t = j\}$  and  $\hat{X}_j = \{X_{t-1} : s_t = j\}$ . Once we sample  $\beta_j$  for  $j = 1, \dots, m$  then we can use  $\hat{\varepsilon}_j = \hat{y}_j - \hat{X}_j\beta_j$  and thus sample  $\sigma_j^2$  from the Inverse gamma density, see Liu and Maheu (2008). Ideally, it would be desirable to allow  $\sigma^2$  to vary independently from  $\beta$ . However, for the change-point model this would be relatively more computationally demanding and will probably not provide any significant improvements.

Modeling structural breaks in  $\sigma^2$  for the MIA model is a bit more complicated. In this paper we take the same approach as Giordani and Kohn (2008) and Groen et al. (2012). The approach is as follows:

- Initialize the sampler with a time series of conditional variances,  $\sigma_1^2, \dots, \sigma_T^2$ .
- Conditional on  $\sigma_1^2, \dots, \sigma_T^2$  draw  $K, B, q_1^2, \dots, q_k^2$  and  $\pi_1, \dots, \pi_k$  from their respective conditional posteriors. Compute the residual for time  $t$  as  $\varepsilon_t = y_t - X_{t-1}\beta_t = \sigma_t u_t$  where  $u_t \sim N(0, 1)$ . We can then square both sides and take the logarithm such that  $\log \varepsilon_t^2 = \log \sigma_t^2 + \log u_t^2$  where  $\log u_t^2$  is  $\log \chi_1^2$  distributed and can be very accurately approximated by a mixture of Normals with seven components, see Kim et al. (1998). We follow the stochastic volatility literature and model structural breaks in  $\sigma^2$  as

$$\log \sigma_t^2 = \log \sigma_{t-1}^2 + \kappa_t^{SV} \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

Here,  $\kappa_t^{SV} = \{0, 1\}$  and evolves independently from  $K_t$ . Estimating this specification is straightforward since we can write the above equations in state space form and again use the conditioning features of the Gibbs sampler to sample  $\log \sigma_t^2, \kappa_t^{SV}, t = 1, \dots, T, \pi^{SV}$  and  $\sigma_\eta^2$  from their respective conditional posteriors. Specifically, we draw  $K^{SV} = (\kappa_1^{SV}, \dots, \kappa_T^{SV})'$  using the algorithm of Gerlach et al. (2000). Thereafter, we sample  $\log \sigma_t^2, t = 1, \dots, T$  using Carter and Kohn (1994) conditional on  $K^{SV}$  and  $\sigma_\eta^2$ .

## 2.4 Model comparison

In this paper we compare the performance of models using a specific out-of-sample period. Consider the universe  $\mathcal{M} = (M_1, \dots, M_n)$  of models. Let  $p(y_t | \theta_n, Y_{t-1}, M_n)$  denote the conditional data density of model  $M_n$  given  $Y_{t-1}$  and the model parameters,  $\theta_n$ . Conditional on  $Y_{t-1} = (y_1, \dots, y_{t-1})'$ , the predictive likelihood (PL) of model  $M_n$  for  $y_t, \dots, y_T, t < T$  is defined as

$$p(y_t, \dots, y_T | Y_{t-1}, M_n) = \int_{\Theta_n} p(y_t, \dots, y_T | \theta_n, Y_{t-1}, M_n) p(\theta_n | Y_{t-1}, M_n) d\theta_n \quad (2.6)$$

Note that, if  $t = 1$  this would be the marginal likelihood and (2.6) changes to

$$p(y_1, \dots, y_T | M_n) = \int_{\Theta_n} p(y_1, \dots, y_T | \theta_n, M_n) p(\theta_n | M_n) d\theta_n$$

where  $p(y_1, \dots, y_T | \theta_n, M_n)$  is the likelihood and  $p(\theta_n | M_n)$  is the prior density of model  $M_n$ . 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 indicates how well model  $M_n$  accounts for the realizations  $y_t, \dots, y_T$ , such that the best model is the one which achieves its maximum value. Hence, it can be used to order models according to their

predictive abilities. For instance, (2.6) is simply the product of the individual predictive likelihoods

$$p(y_t, \dots, y_T | Y_{t-1}, M_n) = \prod_{s=t}^T p(y_s | Y_{s-1}, M_n) \quad (2.7)$$

where each of the terms  $p(y_s | Y_{s-1}, M_n)$  has parameter uncertainty integrated out. The relative value of density forecasts can be compared using the realized data  $y_t, \dots, y_T$  with the predictive likelihoods for two or more models. The Bayesian framework allows also for comparison and ranking of models using the predictive Bayes factors. The predictive Bayes factor (*PBF*) for  $M_1$  versus  $M_2$  is

$$PBF_{12} = p(y_t, \dots, y_T | Y_{t-1}, M_1) / p(y_t, \dots, y_T | Y_{t-1}, M_2)$$

$PBF_{12}$  provides an estimate of the relative evidence for model  $M_1$  versus  $M_2$  over  $y_t, \dots, y_T$ . The predictive Bayes factors include Occam's razor effect in that they penalize highly parametrized models that do not deliver improved predictive content. Kass and Raftery (1995) recommend considering twice the logarithm of *PBF* for model comparison. Evidence in favor of model  $M_1$  can be interpreted as: not worth more than a bare mention for  $0 \leq 2 \log(PBF_{12}) < 2$ ; positive for  $2 \leq 2 \log(PBF_{12}) < 6$ ; strong for  $6 \leq 2 \log(PBF_{12}) < 10$  and very strong for  $2 \log(PBF_{12}) > 10$ .

## 2.5 Calculating the predictive likelihood and the predictive mean

Calculating the predictive likelihood within a Gibbs sampling scheme is easy as we can use the predictive decomposition, (2.7) along with the output from the Gibbs sampler draws,  $\theta^{(1)}, \dots, \theta^{(N)}$ . Specifically, each term on the right-hand-side of (2.7) can be consistently estimated from the Gibbs sampler output as

$$p(y_t | Y_{t-1}, M_n) \approx \frac{1}{N} \sum_{i=1}^N p(y_t | \theta_n^{(i)}, Y_{t-1}, M_n) \quad (2.8)$$

For example, in the context of (2.1),  $\theta_n^{(i)} = (\beta^{(i)}, \sigma^{2(i)})'$  and  $p(y_t | \theta_n^{(i)}, Y_{t-1}, M_n)$  denotes the Normal density with mean  $X_{t-1}\beta^{(i)}$ , variance  $\sigma^{2(i)}$  evaluated at  $y_t$ . The Gibbs sampler draws are obtained based on the information set  $Y_{t-1}$ . As a new observation enters the information set, the posterior is updated through a new round of Gibbs sampling and the predictive density of  $p(y_{t+1} | Y_t, M_n)$  can then be calculated.

Calculating the predictive likelihood for the change-point or MIA specifications is a bit more complicated because one must also consider uncertainty regarding the timing of the structural breaks. This uncertainty is accounted for by using draws of  $S^{(i)}$ ,  $K^{(i)}$  or  $K^{SV(i)}$ ,  $i = 1, \dots, N$  which



are all available from the Gibbs output.

We can also compare forecasts of models based on the predictive mean. Similar to the predictive likelihood, the predictive mean can be computed using Gibbs draws. For instance, in the context of (2.1) we calculate the predictive mean of  $y_t$  conditional on  $Y_{t-1}$  as

$$E[y_t | Y_{t-1}, M_n] \approx \frac{1}{N} \sum_{i=1}^N X_{t-1} \beta_n^{(i)} \quad (2.9)$$

### 3 Realized Volatility

Assume that the price process belongs to the class of special semi-martingales which is a very broad class of processes including Ito and jump processes. Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002a, b) show that the *quadratic variation* of the process which is defined as *integrated volatility* plus a jump component provides a natural measure of ex post volatility. Hence, consider the following logarithmic price process

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + J(t)dq(t), \quad 0 \leq t \leq T \quad (3.1)$$

where  $\mu(t)$  is the drift term,  $\sigma(t)$  is the stochastic volatility process,  $W(t)$  is a standard Wiener process,  $dq(t)$  is a Poisson process with  $dq(t) = 1$  which corresponds to a jump at time  $t$  and  $dq(t) = 0$  which corresponds to no jump, a jump intensity  $\lambda(t)$  and  $J(t)$  refers to the size of a realized jump. The increment in *quadratic variation* from time 0 to  $t$  is

$$QV_t = \int_0^t \sigma^2(s) ds + \sum_{0 \leq s \leq t, dq(s)=1} J^2(s)$$

where the first term, integrated volatility, is from the continuous component of equation (3.1), and the second term is the contribution from discrete jumps.

To consider estimation of  $QV_t$ , the daily time interval is normalized to unity and divided into  $n$  periods. Each period has length  $\Delta = 1/n$ . The  $\Delta$ period return is defined as  $r_{t,j} = p(t + j\Delta) - p(t + (j-1)\Delta)$ ,  $j = 1, \dots, n$ . The daily return is simply given as  $r_t = \sum_{j=1}^n r_{t,j}$ . Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002a, b) apply the following estimator called *realized volatility*. It is defined as

$$RV_t = \sum_{j=1}^n r_{t,j}^2 \xrightarrow{P} QV_t$$

and converges to quadratic variation as  $n \rightarrow \infty$ . Therefore,  $RV_t$  is the relevant quantity to focus on with regards to modeling and forecasting of volatility.

Barndorff-Nielsen and Shephard (2004) also show how the continuous component can be separated from the jump component of volatility. They define realized bipower variation as

$$RBP_t = \mu_1^{-2} \sum_{j=2}^n |r_{t,j-1}| |r_{t,j}|$$

where  $\mu_1 = \sqrt{2/\pi}$ . As  $n \rightarrow \infty$

$$RBP_t \xrightarrow{p} \int_0^t \sigma^2(s) ds$$

The difference between  $RV_t$  and  $RBP_t$  is an estimate of the daily jump component.

Market microstructure dynamics contaminate the price process with noise. In some instances the noise can be time dependent and may be correlated with the efficient price. Hence,  $RV_t$  can be a biased and inconsistent estimator of  $QV_t$ . Hansen and Lunde (2006) provide a bias correction to realized volatility in the following way

$$\widehat{RV}_t^q = \sum_{j=1}^n r_{t,j}^2 + 2 \sum_{w=1}^q \left(1 - \frac{w}{1+q}\right) \sum_{j=1}^{n-w} r_{t,j} r_{t,j+w}$$

We set  $q = 1$  in this paper. Market microstructure also contaminates bipower variation, see Andersen et al. (2007). Therefore, we adjust bipower variation as

$$\widehat{RBP}_t = \frac{\pi}{2} \frac{n}{n-2} \sum_{j=3}^n |r_{t,j-2}| |r_{t,j}|$$

In the following  $\widehat{RV}_t^q$  is referred to as  $RV_t$  and  $\widehat{RBP}_t$  is referred to as  $RBP_t$ .

## 4 Model

The Heterogeneous Autoregressive (HAR) model is proposed by Corsi (2009) and provides a flexible way to model and forecast realized volatility. According to the framework of Corsi (2009) partial volatility is defined as the volatility generated by a certain market component and the model is an additive cascade of different partial volatilities. By straightforward recursive substitution of the partial volatilities Corsi (2009) shows that the additive volatility cascade leads to a simple restricted linear autoregressive model. The HAR model can approximate many of the features of volatility, including long-memory. Our benchmark model is given as

$$y_{t,h} = \beta_0 + \beta_J J_{t-1} + \beta_d y_{t-1,1} + \beta_w y_{t-5,5} + \beta_m y_{t-22,22} + \varepsilon_{t,h}, \quad \varepsilon_{t,h} \sim N(0, \sigma^2) \quad (4.1)$$

where  $y_{t,h} = h^{-1} \sum_{i=1}^h RV_{t+i-1}$  is the average realized volatility,  $h \geq 1$  periods ahead. Evidently,  $y_{t,1} = y_t$ . This model postulates that three factors affect  $y_{t,h}$ : daily volatility,  $y_{t-1}$ , weekly volatility,  $y_{t-5,5}$  and monthly volatility,  $y_{t-22,22}$ . For  $h = 1$  we have a HAR model for the daily volatility, for  $h = 5$  a HAR model for the weekly volatility and so forth. Following Andersen et al. (2007), we also include a jump term in (4.1) defined as  $J_t = \max\{0, RV_t - RBP_t\}$ .

The HAR model with a jump component can be cast into (2.1). Furthermore, besides (4.1) we can also estimate structural breaks versions of the HAR model using the techniques from section 2.

## 5 Data

The data consists of high-frequency observations of trades on the S&P 500 index using the Spyder (SPY) fund, Boeing (BA), Bank of America (BAC), Caterpillar (CAT), General Electric (GE), IBM, Johnson & Johnson (JNJ), JP Morgan (JPM), Pepsi (PEP), Walmart (WMT) and Exxon (XOM) from January 2, 2004 to December 31, 2009, for a total of 1511 trading days. The cleaning of the data is carried out using the steps in Barndorff-Nielsen et al. (2009). After cleaning a 5-minute grid from 9:30 to 16:00 is constructed using previous-tick method, see Hansen and Lunde (2006). From this grid, 5-minute intraday log-returns are constructed. The log-returns are then used to construct realized volatility and realized bipower variation. Conditioning on the first 22 observations, the final data consists of  $T = 1488$  observations. We define  $y_{t,h}$  as  $\sqrt{252RV_{t,h}}/100$ . Table 1 presents summary statistics for  $y_t$ .

## 6 Results

### 6.1 Priors

For HAR, CPHAR and MHAR models estimation and forecasting is performed conditional on the following priors

- HAR

$$\beta \sim N(0, 100I) \quad , \quad IG(4/2, 0.2/2)$$

- CPHAR

$$\beta_j \sim N(0, 100I) \quad , \quad p(\sigma_j^2) \sim IG(4/2, 0.2/2)$$

$$p_j \sim Beta(20, 0.1) \quad , \quad j = 1, \dots, m, \quad p_m = 1$$

- MHAR

$$p(q_i^2) \sim IG(4/2, 0.2/2) \quad , \quad p(\sigma^2) \sim IG(4/2, 0.2/2)$$

$$\pi_i \sim Beta(0.5, 37) \quad , \quad i = d, w, m$$

where  $IG(\frac{1}{2}, \frac{1}{2})$  stands for the Inverse-gamma density, see Kim and Nelson (1999). In section 7 we conduct a prior sensitivity analysis for CPHAR as well as for MHAR. Overall, for the MHAR model we see that results are slightly sensitive to prior settings on  $\pi$  whereas results are very robust with respect to different hyperparameter values on  $P$  for CPHAR.

Finally, besides these models we also estimate a MHAR model with time-varying volatility, see section 2.3. Henceforth, we label this model as MHAR-SV. For this model, we choose the same prior hyperparameter values as the MHAR model for  $q_i^2$  and  $\pi_i$ . With regards to the additional parameters we set  $\pi_i^{SV} \sim Beta(0.5, 42)$  and  $p(\sigma_\eta^2) \sim IG(0.2/2, 10/2)$ . We also experiment with different hyperparameter values for  $\pi_i^{SV}$  and do not find any significant changes worth mentioning.

## 6.2 Full sample estimation

We estimate daily CPHAR, MHAR and MHAR-SV models using the volatility series listed in Table 1. This way we obtain a better understanding with regards to nature and dates of possible structural breaks.

For each series we estimate the CPHAR model from 0 to 7 change points. We relax the assumption of homoskedastic errors of section 2.1 and thus incorporate structural breaks in  $\sigma^2$  as well as  $\beta$ , see section 2.3. We then choose the change-point specification with the highest marginal likelihood value. With regards to MHAR specifications we find that  $\beta_{dt}$ ,  $\beta_{wt}$  and  $\beta_{mt}$  are able to capture structural changes in  $y_t$  as  $\beta_{0t}$  and  $\beta_{Jt}$  are basically constant through time. Therefore, we choose to estimate the intercept and the jump component as being constant coefficients. Hence, the MHAR (MHAR-SV) model is respecified as

$$y_{t,h} = \beta_0 + \beta_J J_{t-1} + \beta_{dt} y_{t-1,1} + \beta_{wt} y_{t-5,5} + \beta_{mt} y_{t-22,22} + \varepsilon_{t,h} \quad (6.1)$$

Estimating  $\beta_0$  and  $\beta_J$  within the Gibbs sampling scheme is straightforward as we can sample these parameters from their Gaussian conditional posterior using sampled draws of the regression parameters and the conditional variance.

In Table 2 we report structural break dates found in daily CPHAR models using the complete sample. For most of the series we find evidence in favor 2 change points. On the other hand, for IBM, XOM and SPY we find evidence in favor of 4 change points.

The next question is how large are the parameter changes when breaks occur and which param-

eters are mostly affected. Table 3 contains posterior mean and standard deviation of parameters for the CPHAR equations of each regime for BAC, IBM and SPY, over the full sample period. Focusing on these series we observe that the more sensitive parameters are  $\beta_d$ ,  $\beta_w$  and  $\sigma^2$ . As expected,  $\sigma^2$  increases during the financial crises of 2008. This increase is noticeable while on the other hand changes in the regression coefficients are less spectacular than changes in  $\sigma^2$ . In Figure 1 we picture the break probabilities for BAC, IBM and SPY.

For MHAR specifications we do not obtain exact change-point dates or estimates of  $\beta_{dt}$ ,  $\beta_{wt}$  and  $\beta_{mt}$ . Instead, we plot structural break probabilities,  $\kappa_{dt}$ ,  $\kappa_{wt}$  and  $\kappa_{mt}$  along with estimates of  $\beta_{dt}$ ,  $\beta_{wt}$  and  $\beta_{mt}$ , for  $t = 1, \dots, T$  using the mean of the Gibbs sampler draws. They are given in Figures 2 to 7.

Overall, looking at the break probabilities and the estimates of the regression coefficients we can clearly see structural breaks occurring during the fall of 2008 and the beginning of 2009. For instance, for BAC we see small breaks in the beginning of 2008 for  $\beta_{dt}$  and  $\beta_{wt}$  while  $\beta_{mt}$  is pretty much constant. However, all coefficients change abruptly from structural breaks during the financial crises. When we allow for time-variation in the volatility of realized volatility we see that the regression coefficients tend to behave more smoothly with more frequent and smaller changes in the beginning of the sample for BAC. Furthermore, we clearly confirm an almost linear increase in the level of  $\sigma^2$  in the beginning of 2007 and a subsequent gradual fall from the beginning of 2009 to the end of the sample.

### 6.3 Forecasts

We present results for forecasting daily ( $h = 1$ ), weekly ( $h = 5$ ) and biweekly ( $h = 10$ ) realized volatility using the direct method of forecasting, see Marcellino et al. (2005). The following is a list of the forecasting models used in this paper along with their acronyms.

1.  $M_1$ : HAR: constant parameter HAR model.
2.  $M_2$ : CPHAR: structural break HAR model using the specification of Chib (1998).
3.  $M_3$ : MHAR: mixture innovation HAR model.
4.  $M_4$ : MHAR-SV: MHAR model with structural breaks in the volatility of realized volatility.
5.  $M_5$ : TVPHAR: random walk time-varying parameter HAR model. This specification is a standard time-varying parameter model where  $\beta_{it} = \beta_{it-1} + \eta_{it}$  for  $i = d, w, m$ . As before, we estimate the intercept and the jump component as being constant coefficients. This specification assumes typically small and gradual breaks in  $\beta_t$ . Finally, notice that TVPHAR is a restricted version of MHAR when  $\kappa_{dt} = \kappa_{wt} = \kappa_{mt} = 1$ , for  $t = 1, \dots, T$ .

Tables 4 to 6 present results of our forecasting exercise for  $h = 1$ ,  $h = 5$  and  $h = 10$ , respectively. For each series listed in Table 1, we choose the out-of-sample period from February 2, 2008 to December 31, 2009 for a total of 483 observations. We first estimate the models using the initial sample and forecast. Then, we add one data point, update and forecast again, until the end of the out-of-sample data. This strategy works for HAR, MHAR, MHAR-SV and TVPHAR specifications as we do not need to specify the number of structural breaks at each point in time over the out-of-sample data.

In the context of forecasting with CPHAR we follow Bauwens et al. (2011) and perform the following: for the first out-of-sample observation at time  $t$  we calculate the marginal likelihood for various number of change points,  $(0, \dots, n)$  using  $Y_{t-1}$ . Thereafter, we choose the optimal change point number,  $n_1$  using Bayes factors. We calculate the predictive likelihood,  $p(y_t | Y_{t-1}, M_2)$  and the predictive mean,  $E[y_t | Y_{t-1}, M_2]$  using the parameters associated with specification  $n_1$ . Thereafter, we increase the out-of-sample with one observation, calculate marginal likelihoods for  $(0, \dots, n_1 + 1)$  change points, choose the optimal change point number,  $n_2$ , repeat the above forecasting procedure to obtain  $p(y_{t+1} | Y_t, M_2)$  and  $E[y_{t+1} | Y_t, M_2]$ .

We report the logarithm of  $PBF$  for  $M_2, \dots, M_5$  versus  $M_1$  and the ratio of RMSE for  $M_2, \dots, M_5$  over  $M_1$  for the out-of-sample data. Overall, we see that structural break specifications outperform the HAR model both in terms of predictive likelihood and point forecasts, especially as forecast horizon lengthens. For instance, for SPY, the  $\log(PBF)$  in favor of MHAR over HAR is 50.54 for  $h = 1$ , 60.64 for  $h = 5$  and 72.54 for  $h = 10$  (64.58 for  $h = 1$ , 62.84 for  $h = 5$  and 74.50 for  $h = 10$  for CPHAR). With regards to point forecasts we find that MHAR and CPHAR on average outperform the HAR model by 5% to 10% for  $h = 1$ , 10% to 25% for  $h = 5$  and  $h = 10$ . The TVPHAR model outperforms the HAR model in terms of the predictive likelihood regardless of forecast horizon. On the other hand, TVPHAR performs slightly worse than HAR in terms of point forecasts for  $h = 1$ .

We also compare results between models that allow for structural breaks. Here, we get very interesting results. For example, when we compare CPHAR with MHAR we find that for some series and forecast horizons the CPHAR model performs better while for other series the MHAR model performs better. For instance, for BAC we find that MHAR outperforms CPHAR regardless of forecast horizon or criterion. Compared to CPHAR, the  $\log(PBF)$  in favor of MHAR is 5.29 for  $h = 1$ , 15.08 for  $h = 5$  and 29.61 and  $h = 10$ . In terms of point forecasts, compared to CPHAR we see a reduction of 20% for  $h = 1$ , 24% for  $h = 5$  and 26% for  $h = 10$  in RMSE when we use the MHAR model. On the other hand, for XOM, we see that CPHAR outperforms MHAR by 8.14 for  $h = 1$ , 1.87 for  $h = 5$  and 4.98 for  $h = 10$  in terms of  $\log(PBF)$ . Furthermore, these models tend to perform better than TVPHAR in terms of the predictive likelihood regardless of forecast horizon. However, TVPHAR performs relatively better especially in terms of point forecasts, for

$h = 5$  and  $h = 10$ . The difference between the predictive likelihood and the predictive mean is that the predictive likelihood criterion takes into account the whole shape of the predictive density whereas the predictive mean does not.

Finally, the MHAR-SV model tends to dominate its homoskedastic counterpart as well as the CPHAR model. Density forecasts show the most improvements, while point forecasts often show only modest gains over MHAR or CPHAR. For example, MHAR-SV improves upon MHAR with increases in the  $\log(\text{PL})$  of 7.17 for  $h = 1$ , 1.07 for  $h = 5$  and 4.01 for  $h = 10$  for SPY.

## 7 Prior Sensitivity Analysis

### 7.1 CPHAR

In this section sensitivity of the results to prior specification is evaluated by investigating alternative priors on the transition probability,  $p_j \sim \text{Beta}(a_0, b_0)$ ,  $j = 1, \dots, m - 1$  keeping the prior values for the other parameters the same as in section 6.1.  $p_j$  is one of the key parameters of the model because it controls the duration of each regime. Models with different hyperparameter values for  $\beta_j$  and  $\sigma_j^2$  are also estimated and results are very similar to those given in Tables 4 to 6.

In Table 7 we repeat the forecasting exercise of section 6.3 while experimenting with different hyperparameter values for  $p_j$ . For instance, the first alternative prior that is considered is  $p_j \sim \text{Beta}(0.5, 0.5)$  which is relatively flat while the last alternative prior is  $p_j \sim \text{Beta}(100, 0.1)$  which is relatively very tight. In fact, for  $p_j \sim \text{Beta}(100, 0.1)$  we assume a priori that the expected duration of each regime is about 1000 days before we see the data.

Overall, we see that we get almost identical results regardless the values of  $a_0$  and  $b_0$ . Results overwhelmingly suggest existence of structural breaks over the out-of-sample data. Furthermore, the choice of prior hyperparameters is of limited importance both in terms of PL and RMSE.

### 7.2 MHAR

In order to illustrate sensitivity of the results to the prior hyperparameter values on  $\pi_i$ ,  $p(\pi_i) \sim \text{Beta}(a_{i0}, b_{i0})$  for  $i = d, w, m$ , we consider the MHAR model using the SPY data. The forecasting exercise of section 6.3 is repeated using different prior hyperparameter values for  $p(\pi_i)$ . For instance, for the first prior specification we set  $p(\pi_i) \sim \text{Beta}(0.5, 27)$  keeping the other hyperparameter values the same as in section 6.1.

The  $\log(\text{PBF})$  and ratio of RMSE for MHAR over HAR using these different prior settings are displayed in Table 8. It can be seen that these different prior hyperparameter values are overall yielding fairly similar results. For  $p(\pi_i) \sim \text{Beta}(2, 8)$  we get better results in terms of  $\log(\text{PBF})$  and RMSE for  $h = 10$ .

## 8 Conclusion

In this paper we compare different forecasting procedures which allow for structural breaks in the model parameters using different realized volatility time series. Our set of forecasting models is divided into three groups: constant parameter model, HAR, one which formally specifies the number of structural breaks, CPHAR and those which determine the nature of structural changes in the parameters using MIA and random walk specifications.

The empirical application provides some interesting results. First, we add to the literature establishing existence of structural breaks in realized volatility time series. Second, our results also show the importance of using a forecasting method which allows for some sort of structural changes in the parameters. Furthermore, perhaps as expected, we cannot establish that there is one single forecasting method that always is to be preferred. On the contrary, we find that for some series the MHAR model outperforms the CPHAR model whereas for other series the CPHAR model works better.

Finally, when we account for structural breaks in the volatility of realized volatility in the MHAR model we find that this specification tends to dominate its homoskedastic counterpart as well as the CPHAR model. Density forecasts show the most improvement, while point forecasts show only modest gains.



## References

- [1] Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens. 2001. "The Distribution of Realized Stock Return Volatility." *Journal of Financial Economics* 61: 43-76.
- [2] Andersen, T. G., T. Bollerslev, and F. X. Diebold. 2007. "Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility." *Review of Economics and Statistics* 89: 701-720.
- [3] Barndorff-Nielsen, O. E., and N. Shephard. 2002a. "Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models." *Journal of the Royal Statistical Society Series B* 64: 253-280.
- [4] Barndorff-Nielsen, O. E., and N. Shephard. 2002b. "Estimating Quadratic Variation using Realised Variance." *Journal of Applied Econometrics* 17: 457-477.
- [5] Barndorff-Nielsen, O. E., and N. Shephard. 2004. "Power and Bipower Variation with Stochastic Volatility and Jumps." *Journal of Financial Econometrics* 2(1): 1-37.
- [6] Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard. 2009. "Realized kernels in practice: trades and quotes." *The Econometrics Journal* 12(3): 1-32.
- [7] Bauwens, L., G. Koop, D. Korobilis, and V. K. Rombouts. 2011. "A Comparison of Forecasting Models for Macroeconomics Series: The Contribution of Structural Break Models ." Working paper, University of Strathclyde.
- [8] Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. E. Tauchen. 2007. "A Discrete-Time Model for Daily S&P 500 Returns and Realized Variations: Jumps and Leverage Effects." *Journal of Econometrics* 150: 151-166.
- [9] Carter, C., and R. Kohn. 1994. "On Gibbs sampling for state space Models." *Biometrika* 81: 541-553.
- [10] Chib, S. 1998. "Estimation and Comparison of Multiple Change-Point Models." *Journal of Econometrics* 86: 221-241.
- [11] Corsi, F. 2009. "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics* 7: 174-196.
- [12] Durbin, J., and S. J. Koopman. 2002. "A simple and efficient simulation smoother for state space time series analysis." *Biometrika* 89: 603-616.

- [13] Gerlach, R., C. Carter, and R. Kohn. 2000. "Efficient Bayesian Inference for Dynamic Mixture Models." *Journal of the American Statistical Association* 95: 819-828.
- [14] Geweke, J. 2005. *Contemporary Bayesian Econometrics and Statistics*. Wiley.
- [15] Giordani, P., and R. Kohn. 2008. "Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models." *Journal of Business and Economic Statistics* 26: 66-77.
- [16] Groen, J., R. Paap, and F. Ravazzolo. 2012. "Real-time Inflation Forecasting in a Changing World." *Journal of Business and Economic Statistics* 31(1): 29-44.
- [17] Hansen, P. R., and A. Lunde. 2006. "Realized Variance and Market Microstructure Noise." *Journal of Business and Economic Statistics* 24: 127-218.
- [18] Huang, X., and G. Tauchen. 2005. "The Relative Contribution of Jumps to Total Price Variance." *Journal of Financial Econometrics* 3: 456-499.
- [19] Kass, R. E., and A. E. Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90: 773-795.
- [20] Kim, J. C., and C. R. Nelson. 1999. *State-Space Models with Regime Switching: Classical and Gibbs-sampling Approaches with Applications*. MIT press.
- [21] Koop, G. 2003. *Bayesian Econometrics*. John Wiley & Sons Ltd.
- [22] Koopman, S. J., B. Jungbacker, and E. Hol. 2005. "Forecasting Daily Variability of the S&P 100 Stock Index Using Historical, Realised and Implied Volatility Measurements." *Journal of Empirical Finance* 12(3): 445-475.
- [23] Liu, C., and J. Maheu. 2008. "Are there structural breaks in Realized Volatility?" *Journal of Financial Econometrics* 6: 326-360.
- [24] Liu, C., and J. Maheu. 2009. "Forecasting Realized Volatility: A Bayesian Model Averaging Approach." *Journal of Applied Econometrics* 24: 709-733.
- [25] 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: 499-526.
- [26] Pesaran, H., D. Pettenuzzo, and A. Timmermann. 2006. "Forecasting Time Series Subject to Multiple Structural Breaks." *Review of Economic Studies* 73: 1057-1084.

# A Appendix

## A.1 Estimation of the change-point model

To conduct estimation of the change-point model, we start by specifying independent conditionally conjugate priors for the parameters in each regime. They are

$$\beta_j \sim N(n_0, N_0), \quad \sigma_j^2 \sim IG\left(\frac{\nu_0}{2}, \frac{s_0}{2}\right), \quad p_j \sim \text{Beta}(a_0, b_0)$$

for  $j = 1, \dots, m$  and  $p_m = 1$ . Notice that here we consider the case where both  $\beta$  and  $\sigma^2$  change due to structural breaks. For the sake of notation, let  $\theta_j = (\beta_j, \sigma_j^2)'$ . Furthermore, in order to ease the notation conditioning on  $X_T$  is suppressed.

In order to perform Gibbs sampling we divide the parameter space into three blocks:  $\theta = \{\theta_1, \dots, \theta_m\}$ , the state of the system  $S = (s_1, \dots, s_T)'$ , and the transition matrix,  $P$ . Below, we provide more details on each step of the Gibbs sampler.

**Step 1:** Simulation of  $S \mid \theta, P, Y_T$ . Chib (1998) shows that a joint draw of  $S$  can be achieved in one step using

$$p(S \mid \theta, P, Y_T) = p(s_T \mid \theta, P, Y_T) \prod_{t=1}^{T-1} p(s_t \mid s_{t+1}, \theta, P, Y_t) \quad (\text{A.1})$$

in which one samples sequentially from each density on the right-hand-side of (A.1) beginning with  $p(s_T \mid \theta, P, Y_T)$ , and then  $p(s_t \mid s_{t+1}, \theta, P, Y_t)$   $t = T-1, \dots, 1$ . At each step one conditions on the previously drawn state  $s_{t+1}$ , until a full draw of  $S$  is obtained. The individual densities in (A.1) are obtained based on the following steps:

(a) Initialization: At  $t = 1$ , set  $p(s_1 = 1 \mid \theta, P, Y_1) = 1$ .

(b) Compute the Hamilton (1989) filter,  $p(s_t = j \mid \theta, P, Y_t)$ . This involves a prediction and an update step in which one iterates on the following from  $t = 2, \dots, T$ ,

$$p(s_t = j \mid \theta, P, Y_{t-1}) = \sum_{l=j-1}^j p(s_{t-1} = l \mid \theta, P, Y_{t-1}) p_{lj}, \quad j = 1, \dots, m \quad (\text{A.2})$$

$$p(s_t = j \mid \theta, P, Y_t) = \frac{p(s_t = j \mid \theta, P, Y_{t-1}) p(y_t \mid \theta, Y_{t-1}, s_t = j)}{\sum_{l=1}^m p(s_t = l \mid \theta, P, Y_{t-1}) p(y_t \mid \theta, Y_{t-1}, s_t = l)}, \quad (\text{A.3})$$

$j = 1, \dots, m$

The last equation is obtained from Bayes' rule. Note that in (A.2) the summation is only from  $j-1$  to  $j$ , due to the restricted nature of the transition matrix and  $p(y_t \mid \theta, Y_{t-1}, s_t = j) \sim N(X_{t-1}\beta_j, \sigma_j^2)$  has a closed form solution.

(c) Finally, Chib (1998) shows that the individual densities in (A.1) are

$$p(s_t | s_{t+1}, \theta, P, Y_t) \propto p(s_t | \theta, P, Y_t) p(s_{t+1} | s_t, P)$$

Thus, given  $s_T = m$ ,  $s_t$  is drawn backwards over  $t = T - 1, T - 2, \dots, 2$  as

$$s_t | s_{t+1}, \theta, P, Y_t = \begin{cases} s_{t+1} & \text{with probability } c_t \\ s_{t+1} - 1 & \text{with probability } 1 - c_t \end{cases}$$

where

$$c_t = \frac{p(s_t = j | \theta, P, Y_t) p(s_{t+1} = j | s_t = j, P)}{\sum_{l=j-1}^j p(s_t = l | \theta, P, Y_t) p(s_{t+1} = j | s_t = l, P)}$$

Finally, note that  $p(s_1 = 1 | s_2, \theta, P, Y_t) = 1$ .

**Step 2:** Simulation of  $\theta | S, Y_T$ . The conditional posterior density of  $\theta_j$  depends only on the data in regime  $j$ . Therefore, let  $\hat{Y}_j = \{y_t : s_t = j\}$ ,  $\hat{X}_j = \{X_{t-1} : s_t = j\}$  and use Gibbs sampling methods for a linear regression model. Hence,  $\beta_j | \hat{Y}_j, \hat{X}_j, \sigma_j^2 \sim N(n_j, N_j)$  where

$$N_j = \left( \sigma_j^{-2} \hat{X}_j' \hat{X}_j + N_j^{-1} \right)^{-1}, \quad n_j = N_j \left( \sigma_j^{-2} \hat{X}_j' \hat{Y}_j + N_j^{-1} n_j \right)$$

and  $\sigma_j^2 | \hat{X}_j, \hat{Y}_j, \beta_j \sim IG\left(\frac{v_j}{2}, \frac{s_j}{2}\right)$  where  $v_j = T_j + v_0$ ,  $s_j = (\hat{Y}_j - \hat{X}_j \beta_j)' (\hat{Y}_j - \hat{X}_j \beta_j) + s_0$  and  $T_j$  is the number of observations in regime  $j$ .

**Step 3:** Simulation of  $P | S$ . The conditional posterior for each diagonal component of  $P$  is very simple and given by  $p_j | S \sim \text{Beta}(a_0 + n_j, b_0 + 1)$  where  $n_j$  is the number of one-step transitions from state  $j$  to state  $j$  in a sequence of  $S$ .

## A.2 Marginal likelihood

To compute the marginal log-likelihood for the change-point model we use the method of Chib (1995) which is based on

$$p(Y_T) = \frac{p(Y_T | \theta, P) p(\theta, P)}{p(\theta, P | Y_T)} \quad (\text{A.4})$$

where  $p(Y_T | \theta, P)$  is the likelihood function with  $S$  integrated out,  $p(\theta, P)$  is the prior density, and  $p(\theta, P | Y_T)$  is the posterior density. As before, we set  $Y_T = (y_1, \dots, y_T)'$  and follow the notation from the previous sections. In principle any value of  $(\theta, P)$  can be used to compute (A.4). Here,

we follow Liu and Maheu (2008) and use the posterior mean denoted as  $(\bar{\theta}, \bar{P})$ . Then

$$\log p(Y_T) = \log p(Y_T | \bar{\theta}, \bar{P}) + \log p(\bar{\theta}, \bar{P}) - \log p(\bar{\theta}, \bar{P} | Y_T) \quad (\text{A.5})$$

$p(\bar{\theta}, \bar{P})$  is evaluated directly and the likelihood function  $p(Y_T | \bar{\theta}, \bar{P})$  is calculated as

$$\log p(Y_T | \theta, P) = \sum_{t=1}^T \log p(y_t | \theta, P, Y_{t-1}) \quad (\text{A.6})$$

where

$$p(y_t | P, \theta, Y_{t-1}) = \sum_{j=1}^m p(y_t | \theta, P, Y_{t-1}, s_t = j) p(s_t = j | \theta, P, Y_{t-1}) \quad (\text{A.7})$$

The most difficult and demanding part of (A.5) is the computation of the  $p(\bar{\theta}, \bar{P} | Y_T)$  since it must be computed numerically. We use the decomposition

$$p(\bar{\theta}, \bar{P} | Y_T) = p(\bar{\beta} | Y_T) p(\bar{\sigma}^2 | \bar{\beta}, Y_T) p(\bar{P} | \bar{\beta}, \bar{\sigma}^2, Y_T) \quad (\text{A.8})$$

where each term on the right-hand-side can be estimated from MCMC simulations. The first term can be estimated as

$$p(\bar{\beta} | Y_T) \approx \frac{1}{N} \sum_{i=1}^N p(\bar{\beta} | \sigma^{2(i)}, S^{(i)}, Y_T)$$

where  $p(\bar{\beta} | \sigma^{2(i)}, S^{(i)}, Y_T) = \prod_{j=1}^m p(\bar{\beta}_j | \sigma^{2(i)}, S^{(i)}, Y_T)$  and the draws  $\{\sigma^{2(i)}, S^{(i)}\}_{i=1}^N$  are directly available from the Gibbs output. The second term in (A.8) is equal to

$$p(\bar{\sigma}^2 | \bar{\beta}, Y_T) = \int p(\bar{\sigma}^2 | \bar{\beta}, S, Y_T) p(S | \bar{\beta}, Y_T) dS$$

where  $p(\bar{\sigma}^2 | \bar{\beta}, S, Y_T) = \prod_{j=1}^m p(\bar{\sigma}_j^2 | \bar{\beta}, S, Y_T)$ . To obtain the draws from  $p(S | \bar{\beta}, Y_T)$ , we run an additional reduced Gibbs sampling conditional on  $\bar{\beta}$ , that is, we run a Gibbs sampling scheme where we do not draw from  $\beta$  but fix them to be  $\bar{\beta}$ . Thereafter, we use  $\{S^{(i)}\}_{i=1}^N$  and calculate  $p(\bar{\sigma}^2 | \bar{\beta}, Y_T)$  as

$$p(\bar{\sigma}^2 | \bar{\beta}, Y_T) \approx \frac{1}{N} \sum_{i=1}^N p(\bar{\sigma}^2 | \bar{\beta}, S^{(i)}, Y_T)$$

Finally, for  $p(\bar{P}|\bar{\beta}, \bar{\sigma}^2, Y_T) = \prod_{j=1}^{m-1} p(\bar{p}_j|\bar{\beta}, \bar{\sigma}^2, Y_T)$ ,  $j = 1, \dots, m-1$  we sample  $\{S^{(i)}\}_{i=1}^N$  from  $p(S|\bar{\beta}, \bar{\sigma}^2, Y_T)$  and set

$$p(\bar{p}_j|\bar{\beta}, \bar{\sigma}^2, Y_T) \approx \frac{1}{N} \sum_{i=1}^N p(\bar{p}_j|\bar{\beta}, \bar{\sigma}^2, S^{(i)}, Y_T)$$

### A.3 Estimation of the MIA model

Consider the following model

$$y_t = X_{t-1}\beta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (\text{A.9})$$

where  $\beta_t = (\beta_{1t}, \dots, \beta_{kt})'$  and

$$\beta_{it} = \beta_{it-1} + \kappa_{it}\eta_{it}, \quad \eta_{it} \sim N(0, q_i^2), \quad i = 1, \dots, k \quad (\text{A.10})$$

where  $\kappa_{it} = \{0, 1\}$  with  $p(\kappa_{it} = 1) = \pi_i$ . The parameters of (A.9)-(A.10) are: the structural break probabilities,  $\pi = (\pi_1, \dots, \pi_k)'$ , the magnitude of the breaks in the state parameters,  $q = (q_1^2, \dots, q_k^2)'$  and  $\sigma^2$ . These quantities are all collected in  $\theta$ . As before, let  $K_t = (\kappa_{1t}, \dots, \kappa_{kt})$ ,  $K = \{K_t\}_{t=1}^T$  and  $B = \{\beta_t\}_{t=1}^T$ . The Gibbs sampling scheme for (A.9)-(A.10) is as follows

- Sample  $K | \theta, Y_T$

The structural breaks are sampled using the algorithm of Gerlach et al. (2000). In particular, Gerlach et al. (2000) has two important features. First,  $K$  is generated without conditioning on the states,  $B$ . Second, the number of operations required to obtain a draw of  $K$  is reduced from  $O(T^2)$  to  $O(T)$ . Define  $K_{-t} = \{K_s\}_{s=1, s \neq t}^T$ .

The conditional posterior of  $K_t$  is defined as

$$\begin{aligned} p(K_t | K_{-t}, Y_T, \theta) &\propto p(Y_T | K, \theta) p(K_t | K_{-t}, \theta) \\ &\propto p(y_{t+1}, \dots, y_T | K, y_1, \dots, y_t, \theta) \\ &\times p(y_t | K_1, \dots, K_t, y_1, \dots, y_{t-1}, \theta) p(K_t | K_{-t}, \theta) \end{aligned} \quad (\text{A.11})$$

The term  $p(K_t | K_{-t}, \theta)$  is obtained from the prior and  $p(y_t | K_1, \dots, K_t, y_1, \dots, y_{t-1}, \theta)$  is computed using the Kalman filter. The important contribution of Gerlach et al. (2000) is that the second term in (A.11),  $p(y_{t+1}, \dots, y_T | K, Y_t, \theta)$  can be obtained in one step after an initial set of backward recursions. Finally, since  $K_t$  can only take a finite number of values, it can be drawn by computing the right-hand-side of (A.11) for all possible values of  $K_t$  and then normalizing. For more details on implementation of the algorithm we refer the reader to Gerlach et al. (2000).

- Sample  $B \mid K, \theta, Y_T$

$B$  is sampled from its conditional posterior using the simulation smoother of Carter and Kohn (1994). The algorithm of Durbin and Koopman (2002) is also an interesting alternative.

- Sample  $\theta \mid K, B, Y_T$

To sample  $q_i^2$  and  $\sigma^2$  Inverse gamma priors are used, see Kim and Nelson (1999). Finally, assume that  $p(\pi_i) \sim \text{Beta}(a_{i0}, b_{i0})$ . Then, the conditional posterior of  $\pi_i$  is  $\text{Beta}(a_i, b_i)$  where  $a_i = a_{i0} + \sum_{t=1}^T \kappa_{it}$  and  $b_i = b_{i0} + T - \sum_{t=1}^T \kappa_{it}$ .

Table 1: Summary statistics for  $RV_t$ 

Series	mean	median	std. dev.	kurtosis	skewness
BA	0.022	0.018	0.013	16.354	2.970
BAC	0.030	0.015	0.036	16.130	3.135
CAT	0.026	0.021	0.016	12.914	2.691
GE	0.022	0.015	0.020	16.313	3.222
IBM	0.018	0.014	0.011	20.611	3.464
JNJ	0.013	0.011	0.008	36.515	4.335
JPM	0.028	0.017	0.026	13.886	2.773
PEP	0.015	0.013	0.009	43.877	4.656
SPY	0.013	0.010	0.010	23.864	3.640
WMT	0.018	0.015	0.009	29.579	3.772
XOM	0.020	0.017	0.012	43.574	4.714

Summary statistics for  $RV_t$  from January 2, 2004 to December 31, 2009. In total 1510 observations.

Table 2: Change-point dates based on full sample

Series	# CP	dates			
BA	2	07-25-07	09-11-08		
BAC	3	02-26-07	10-30-07	09-11-08	
CAT	2	07-17-07	09-02-08		
GE	3	05-14-04	06-18-07	09-08-08	
IBM	4	07-18-07	03-24-08	09-05-08	12-09-08
JNJ	2	12-31-07	09-05-08		
JPM	2	07-18-07	09-02-08		
PEP	2	07-18-07	09-15-08		
SPY	4	02-27-07	03-19-08	09-08-08	12-29-08
WMT	2	09-29-06	09-09-08		
XOM	4	07-19-07	02-12-08	09-08-08	12-08-08

This table reports the change-point dates for each series. The first column lists the volatility series. The second column lists the number of the change points (CP) that are conditioned on and the last column shows the change-point dates. The change-point dates are defined as the first observation of the new regime, using the mode of  $\left\{S^{(i)}\right\}_{i=1}^N$ . Sample period starts from February 2, 2004 to December 31, 2009 (1488 observations).



Table 3: Parameter estimates: CPHAR model

Regime	Posterior mean (standard deviation)					
	$\beta_0$	$\beta_J$	$\beta_d$	$\beta_w$	$\beta_m$	$\sigma^2$
<i>BAC</i>						
1	0.234 (0.011)	-0.190 (0.137)	0.395 (0.046)	0.198 (0.070)	0.084 (0.081)	0.009 (0.001)
2	0.273 (0.012)	-0.467 (0.210)	0.492 (0.084)	0.297 (0.130)	0.057 (0.120)	0.038 (0.004)
3	0.312 (0.011)	-0.382 (0.155)	0.652 (0.075)	0.265 (0.101)	-0.030 (0.078)	0.174 (0.017)
4	0.315 (0.026)	-0.200 (0.187)	0.433 (0.066)	0.321 (0.101)	0.189 (0.087)	1.032 (0.079)
<i>IBM</i>						
1	0.190 (0.006)	-0.313 (0.112)	0.208 (0.045)	0.498 (0.069)	-0.063 (0.079)	0.011 (0.001)
2	0.230 (0.006)	-0.606 (0.147)	0.438 (0.089)	0.195 (0.157)	0.131 (0.156)	0.055 (0.006)
3	0.210 (0.006)	-0.029 (0.244)	0.319 (0.115)	0.386 (0.198)	-0.020 (0.173)	0.018 (0.002)
4	0.3469 (0.056)	-0.263 (0.328)	0.224 (0.139)	0.418 (0.242)	0.008 (0.189)	0.339 (0.062)
5	0.217 (0.003)	-0.320 (0.181)	0.349 (0.079)	0.489 (0.113)	0.081 (0.078)	0.021 (0.001)
<i>SPY</i>						
1	0.137 (0.003)	-0.229 (0.135)	0.231 (0.069)	0.480 (0.107)	0.049 (0.074)	0.004 (0.001)
2	0.155 (0.004)	-0.298 (0.215)	0.553 (0.083)	0.219 (0.123)	0.062 (0.132)	0.025 (0.004)
3	0.149 (0.003)	-0.001 (0.223)	0.497 (0.107)	0.309 (0.144)	-0.103 (0.104)	0.007 (0.001)
4	0.260 (0.053)	-0.172 (0.334)	0.279 (0.132)	0.415 (0.215)	-0.045 (0.180)	0.309 (0.055)
5	0.151 (0.003)	-0.197 (0.151)	0.340 (0.081)	0.484 (0.098)	0.152 (0.094)	0.015 (0.001)

This table reports posterior mean and standard deviations (indicated inside the parentheses) for model parameters from the preferred CPHAR model. Sample period starts from February 2, 2004 to December 31, 2009 (1488 observations).

Table 4: Out-of-Sample forecast results,  $RV_{t,h}$ ,  $h = 1$ 

Series	log(PBF)				RMSE			
	CPHAR	MHAR	MHAR-SV	TVP HAR	CPHAR	MHAR	MHAR-SV	TVP HAR
BA	20.09	17.52	19.23	17.51	0.97	0.98	0.97	1.03
BAC	90.14	95.44	101.27	92.62	0.97	0.78	0.77	0.83
CAT	26.36	22.76	26.23	17.90	0.96	1.00	0.98	1.01
GE	81.62	75.51	83.02	74.00	0.96	0.89	0.88	0.91
IBM	36.77	26.88	32.38	24.63	0.94	0.97	0.96	1.01
JNJ	28.68	36.35	41.68	35.83	0.95	0.86	0.89	1.01
JPM	56.06	62.00	65.36	56.68	0.97	0.86	0.85	0.90
PEP	42.01	47.43	52.68	49.19	0.96	0.88	0.87	0.94
SPY	64.58	50.54	57.72	47.93	0.93	0.91	0.90	1.00
WMT	19.12	27.06	35.62	27.31	0.95	0.86	0.86	0.95
XOM	47.61	39.47	45.67	37.73	0.93	0.89	0.88	0.99

This table reports the log predictive Bayes factor,  $\log(PBF)$  of the model of interest versus the HAR model and the out-of-sample root mean squared error, RMSE for the predictive mean of the model of interest over the HAR model. The out-of-sample period is from February 2, 2008 to December 31, 2009.

Table 5: Out-of-Sample forecast results,  $RV_{t,h}$ ,  $h = 5$ 

Series	log(PBF)				RMSE			
	CPHAR	MHAR	MHAR-SV	TVP HAR	CPHAR	MHAR	MHAR-SV	TVP HAR
BA	42.80	43.25	41.97	38.92	0.86	0.88	0.85	0.89
BAC	110.81	125.89	126.28	110.12	0.91	0.69	0.67	0.81
CAT	47.09	50.31	48.67	46.19	0.85	0.87	0.86	0.87
GE	114.89	117.22	114.02	105.46	0.81	0.80	0.77	0.96
IBM	43.69	44.08	49.02	38.29	0.89	0.92	0.91	0.85
JNJ	58.30	53.73	64.30	50.34	0.87	0.89	0.84	1.02
JPM	91.58	96.45	99.46	86.80	0.79	0.75	0.71	0.91
PEP	65.13	63.61	70.49	58.66	0.90	0.85	0.84	1.04
SPY	62.84	60.64	61.71	59.08	0.88	0.93	0.87	0.93
WMT	49.99	47.73	50.56	43.12	0.87	0.84	0.82	1.07
XOM	55.79	53.92	59.26	49.13	0.90	0.85	0.80	0.89

This table reports the log predictive Bayes factor,  $\log(PBF)$  of the model of interest versus the HAR model and the out-of-sample root mean squared error, RMSE for the predictive mean of the model of interest over the HAR model. The out-of-sample period is from February 2, 2008 to December 31, 2009.

Table 6: Out-of-Sample forecast results,  $RV_{t,h}$ ,  $h = 10$

Series	log(PBF)				RMSE			
	CPHAR	MHAR	MHAR-SV	TVP HAR	CPHAR	MHAR	MHAR-SV	TVP HAR
BA	43.70	55.75	55.53	46.81	0.80	0.80	0.81	0.75
BAC	120.79	150.40	144.17	127.49	0.85	0.63	0.64	0.72
CAT	64.62	72.21	64.59	67.34	0.77	0.79	0.77	0.78
GE	128.75	135.01	124.37	115.61	0.71	0.70	0.72	0.85
IBM	55.58	57.53	57.30	54.22	0.83	0.85	0.82	0.73
JNJ	66.98	68.59	79.90	62.66	0.82	0.82	0.79	1.10
JPM	96.42	107.35	107.19	84.41	0.72	0.68	0.68	0.80
PEP	72.52	72.06	78.91	65.56	0.86	0.83	0.80	1.09
SPY	74.50	72.54	76.54	68.31	0.81	0.86	0.84	0.79
WMT	56.41	58.51	64.77	51.20	0.88	0.82	0.81	0.97
XOM	68.19	63.20	69.12	60.65	0.86	0.77	0.72	0.80

This table reports the log predictive Bayes factor,  $\log(PBF)$  of the model of interest versus the HAR model and the out-of-sample root mean squared error, RMSE for the predictive mean of the model of interest over the HAR model. The out-of-sample period is from February 2, 2008 to December 31, 2009.

Table 7: Prior sensitivity analysis, CPHAR model, SPY

	log(PBF)			RMSE		
	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$
$p_j \sim Beta(0.1, 0.1)$	64.28	62.71	72.60	0.93	0.88	0.81
$p_j \sim Beta(8, 0.1)$	64.13	62.72	73.27	0.94	0.88	0.82
$p_j \sim Beta(10, 0.1)$	64.06	65.00	72.60	0.94	0.88	0.81
$p_j \sim Beta(8, 2)$	63.70	63.89	72.61	0.94	0.88	0.81
$p_j \sim Beta(20, 2)$	64.57	65.10	72.60	0.93	0.88	0.81
$p_j \sim Beta(100, 0.1)$	64.56	65.16	74.70	0.93	0.88	0.81

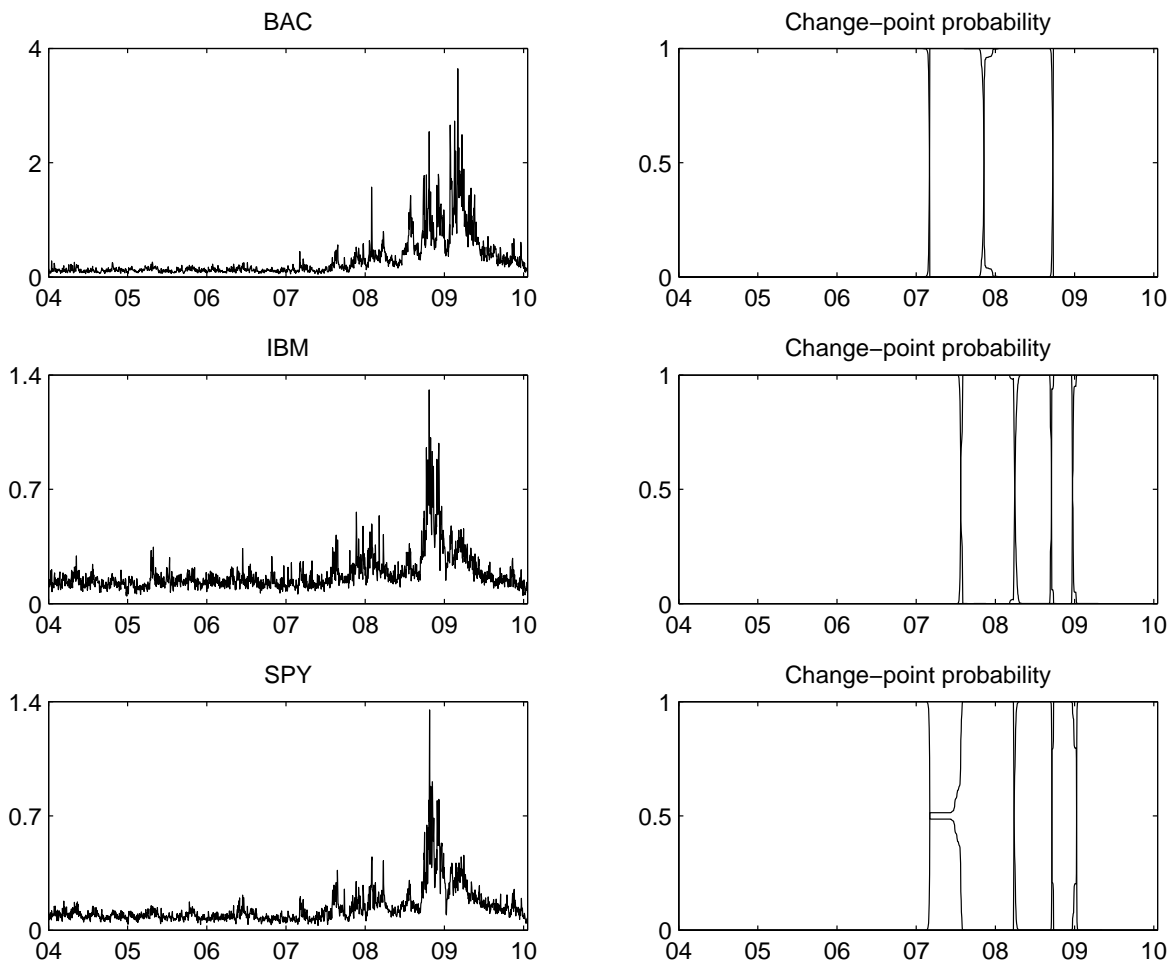
This table reports the log predictive Bayes factor,  $\log(PBF)$  of the model of interest versus the HAR model and the out-of-sample root mean squared error, RMSE for the predictive mean of the model of interest over the HAR model. The results are for the CHAR model considering six different prior hyperparameter values for  $p_j$ . The out-of-sample period is from February 2, 2008 to December 31, 2009.

Table 8: Prior sensitivity analysis, MHAR model, SPY

	log(PBF)			RMSE		
	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$
$\pi \sim \text{Beta}(0.5, 27)$	50.58	60.87	73.80	0.88	0.93	0.86
$\pi \sim \text{Beta}(0.5, 47)$	50.54	61.24	73.22	0.89	0.93	0.86
$\pi \sim \text{Beta}(0.5, 8)$	50.04	61.34	73.96	0.90	0.93	0.86
$\pi \sim \text{Beta}(2, 8)$	49.12	61.79	71.69	0.90	0.92	0.82
$\pi \sim \text{Beta}(0.5, 100)$	50.45	60.23	73.35	0.91	0.93	0.87
$\pi \sim \text{Beta}(0.5, 1000)$	50.54	60.96	72.12	0.92	0.93	0.87

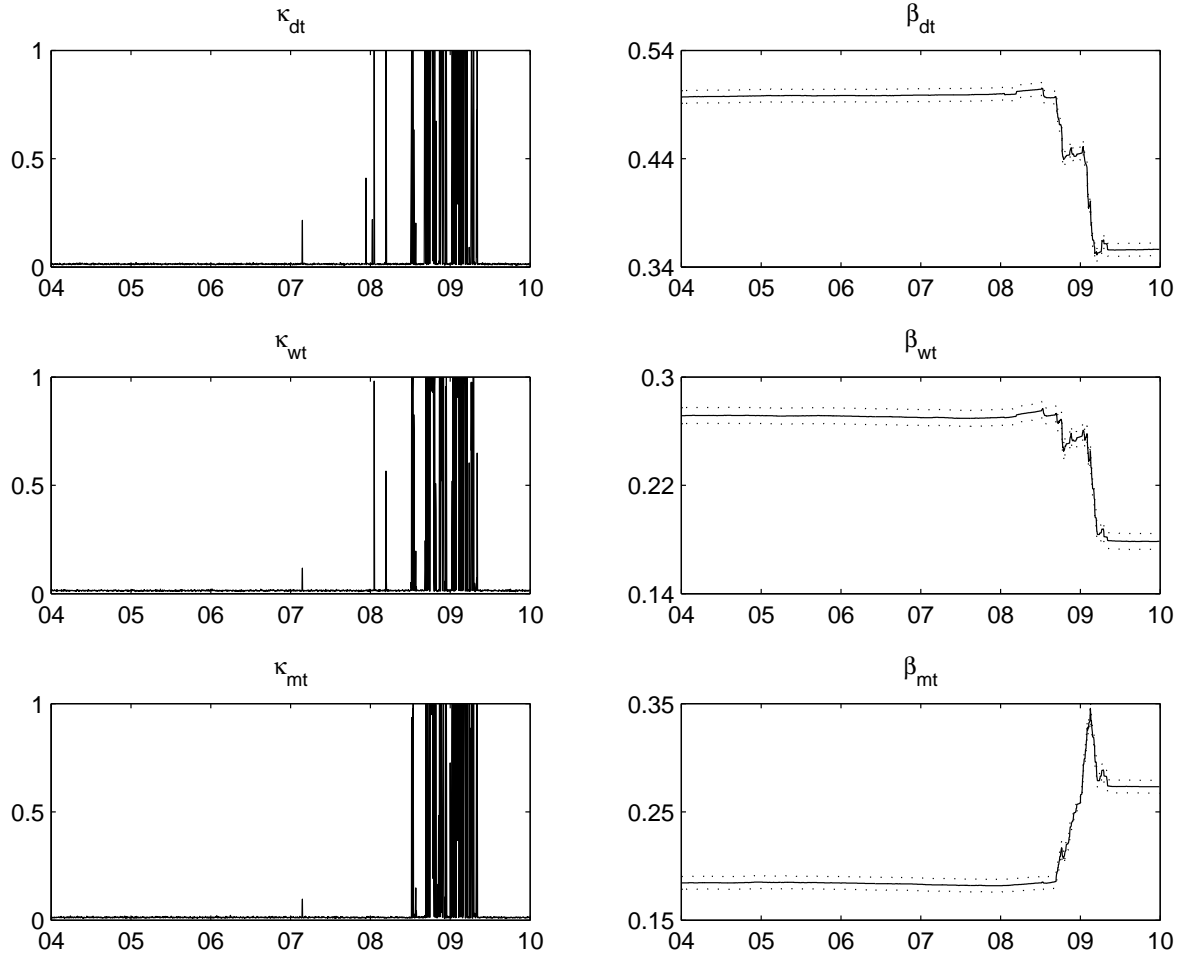
This table reports the log predictive Bayes factor,  $\log(PBF)$  of the model of interest versus the HAR model and the out-of-sample root mean squared error, RMSE for the predictive mean of the model of interest over the HAR model. The results are for the MHAR model considering different prior hyperparameter values for  $\pi$ . The out-of-sample period is from February 2, 2008 to December 31, 2009.

Figure 1: Change-point probabilities: CPHAR model



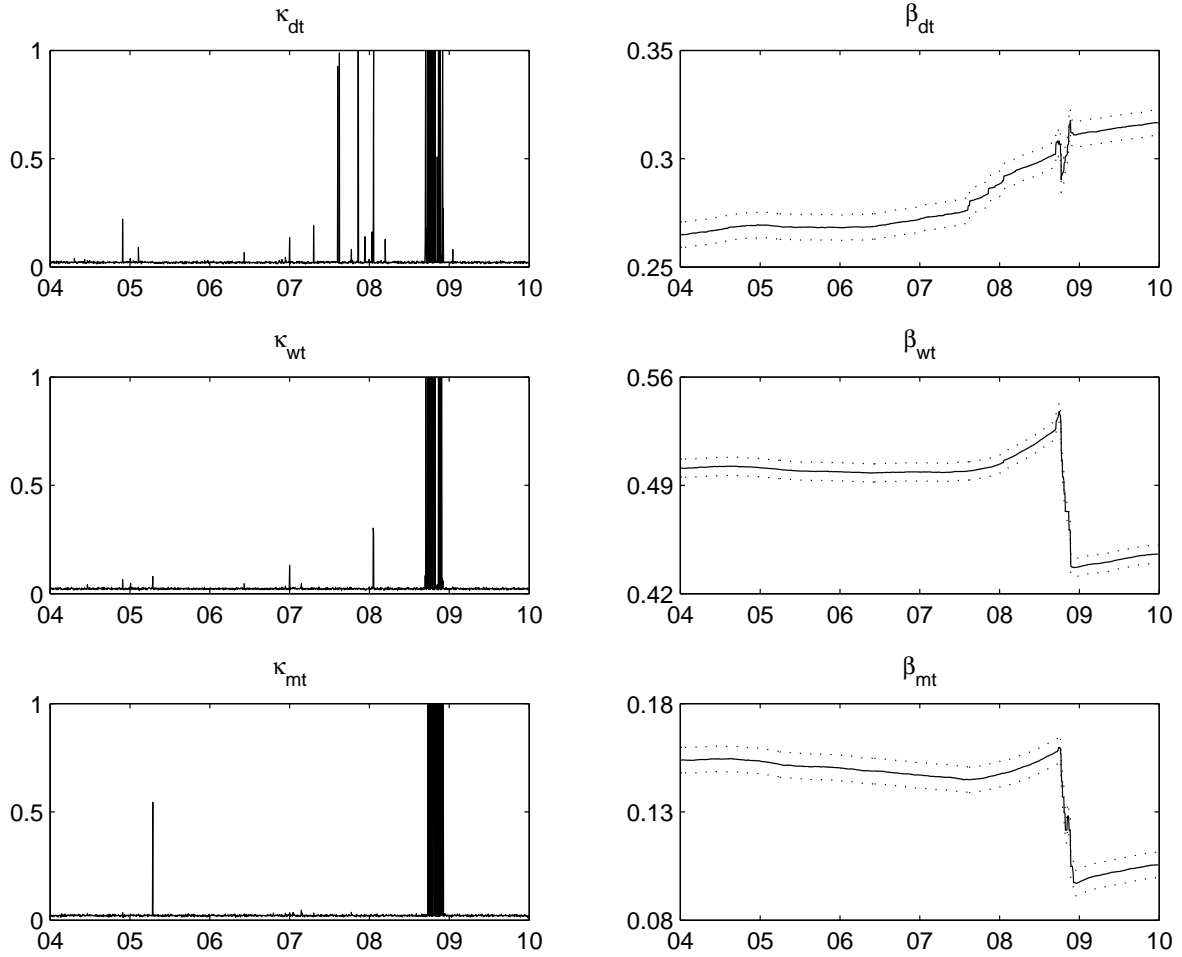
Left: annual realized volatility. Right: change-point probability

Figure 2: Posterior estimates: MHAR model, BAC



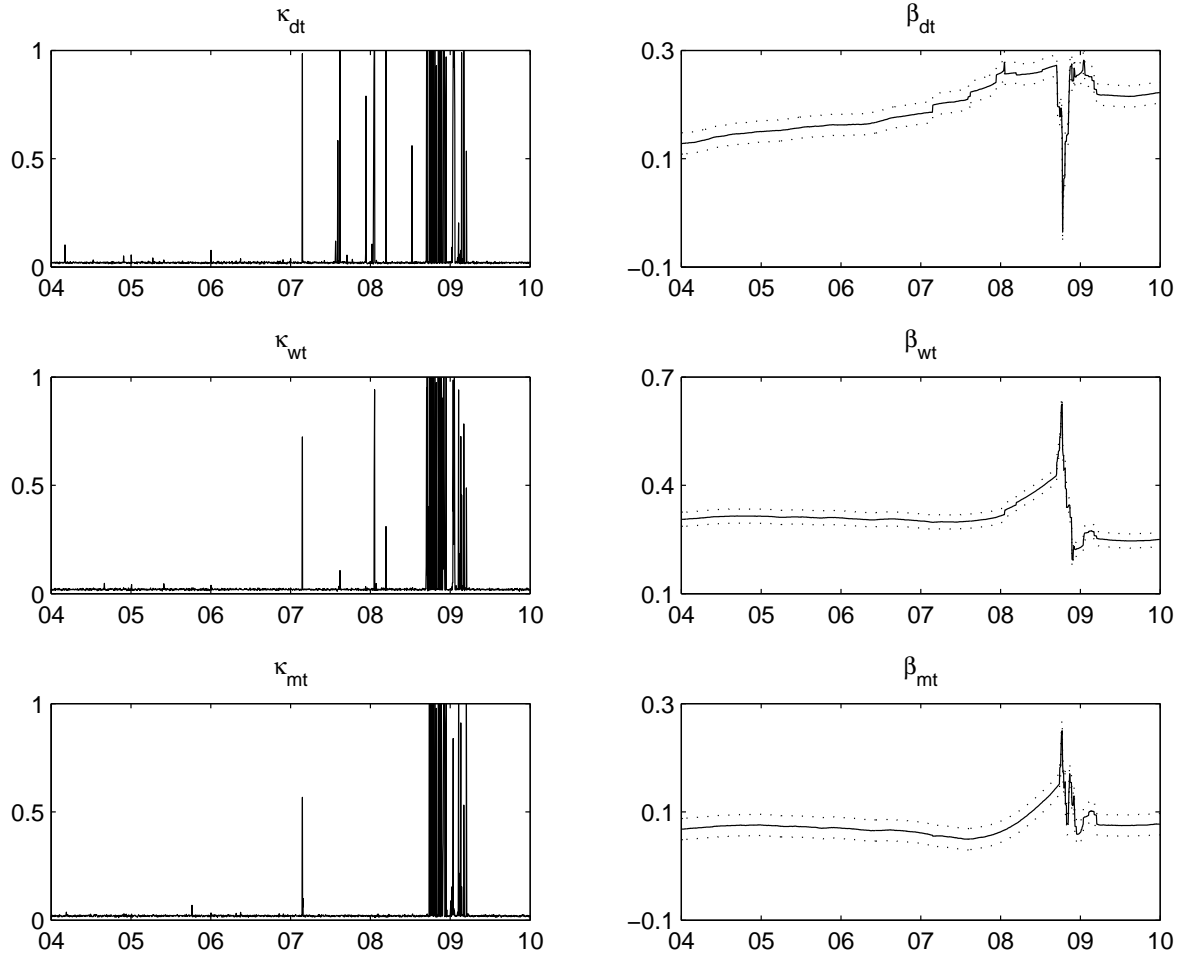
Left: posterior structural break probabilities. Right: posterior estimates of the regression coefficients,  $\beta_{dt}$ ,  $\beta_{wt}$  and  $\beta_{mt}$ .

Figure 3: Posterior estimates: MHAR model, IBM



Left: posterior structural break probabilities. Right: posterior estimates of the regression coefficients,  $\beta_{dt}$ ,  $\beta_{wt}$  and  $\beta_{mt}$ .

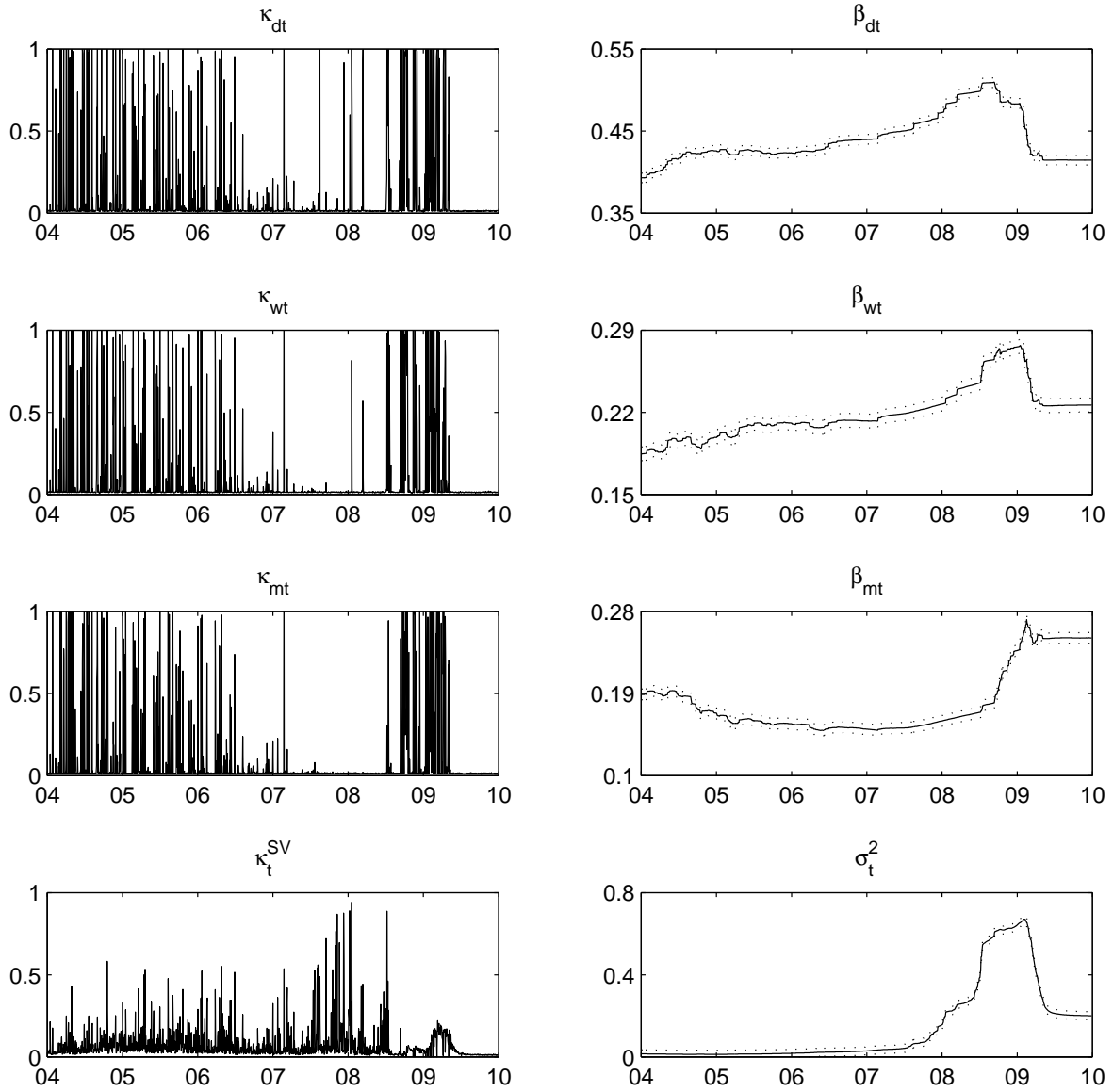
Figure 4: Posterior estimates: MHAR model, SPY



Left: posterior structural break probabilities. Right: posterior estimates of the regression coefficients,  $\beta_{dt}$ ,  $\beta_{wt}$  and  $\beta_{mt}$ .

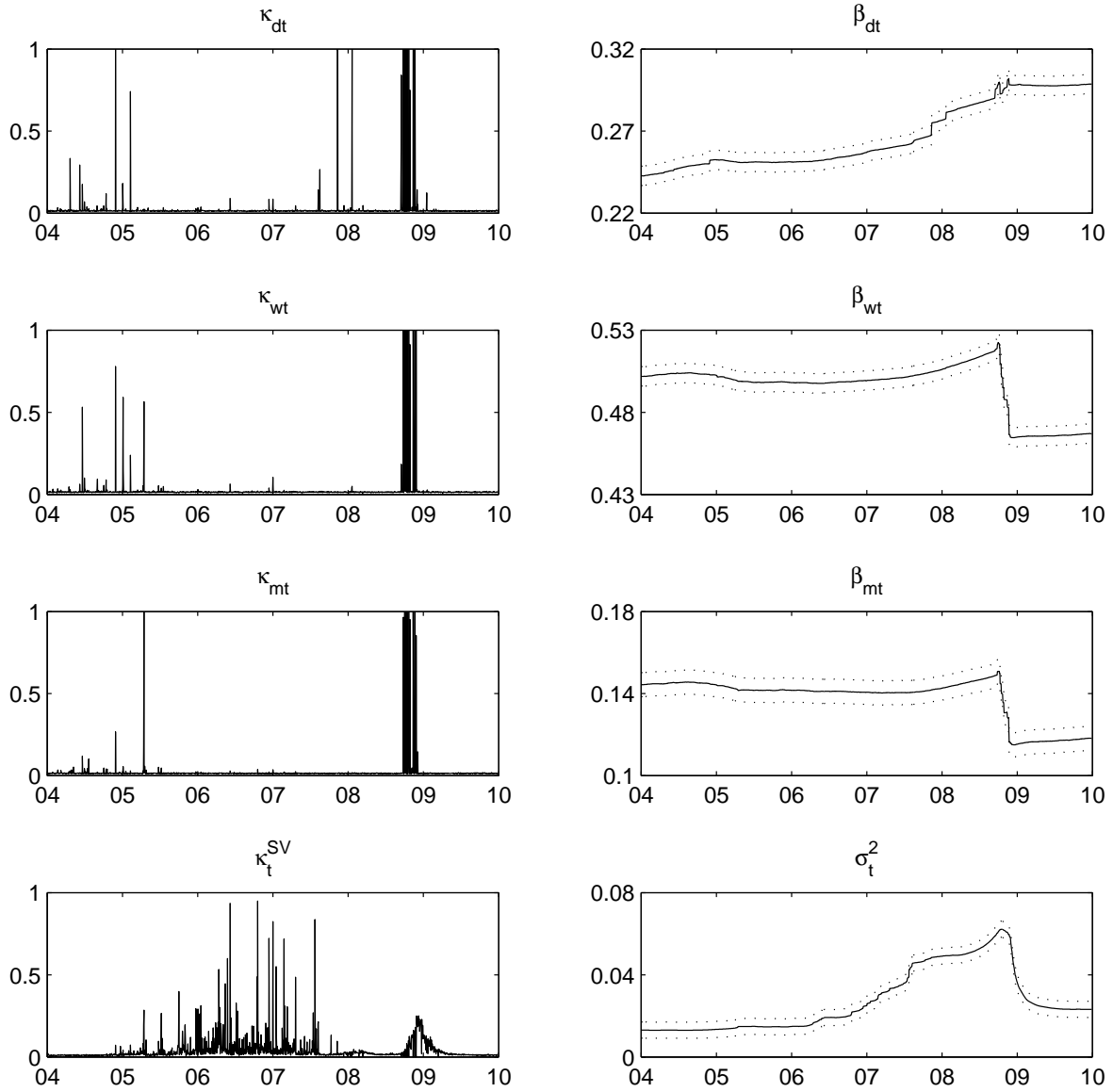


Figure 5: Posterior estimates: MHAR-SV model, BAC



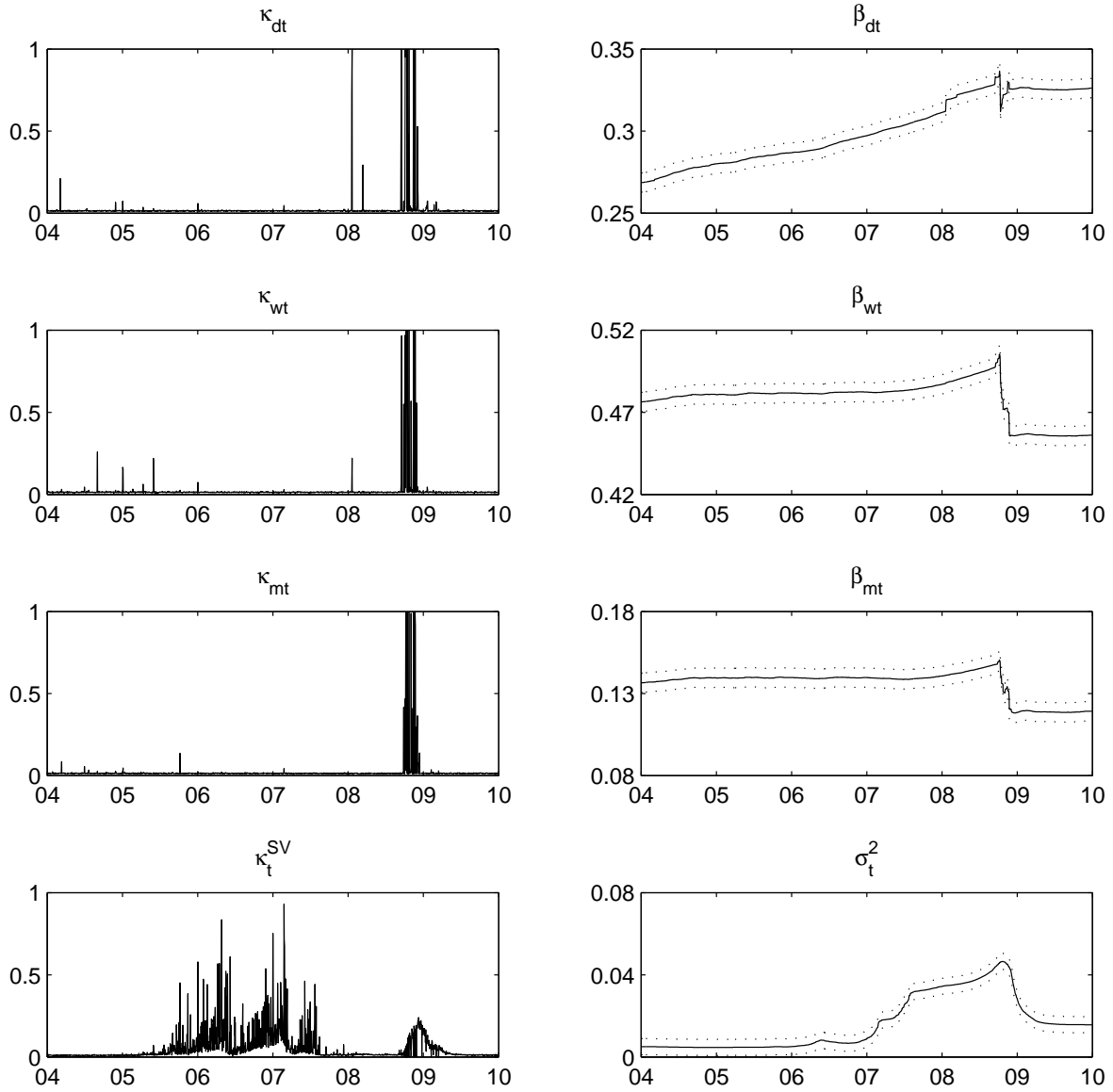
Left: posterior structural break probabilities. Right: posterior estimates of the regression coefficients,  $\beta_{dt}$ ,  $\beta_{wt}$  and  $\beta_{mt}$  and the time-varying conditional variance,  $\sigma_t^2$ .

Figure 6: Posterior estimates: MHAR-SV model, IBM



Left: posterior structural break probabilities. Right: posterior estimates of the regression coefficients,  $\beta_{dt}$ ,  $\beta_{wt}$  and  $\beta_{mt}$  and the time-varying conditional variance,  $\sigma_t^2$ .

Figure 7: Posterior estimates: MHAR-SV model, SPY



Left: posterior structural break probabilities. Right: posterior estimates of the regression coefficients,  $\beta_{dt}$ ,  $\beta_{wt}$  and  $\beta_{mt}$  and the time-varying conditional variance,  $\sigma_t^2$ .

# Research Papers 2013



- 2013-07: Sílvia Gonçalves, Ulrich Hounyo and Nour Meddahi: Bootstrap inference for pre-averaged realized volatility based on non-overlapping returns
- 2013-08: Katarzyna Lasak and Carlos Velasco: Fractional cointegration rank estimation
- 2013-09: Roberto Casarin, Stefano Grassi, Francesco Ravazzolo and Herman K. van Dijk: Parallel Sequential Monte Carlo for Efficient Density Combination: The Deco Matlab Toolbox
- 2013-10: Hendrik Kaufmann and Robinson Kruse: Bias-corrected estimation in potentially mildly explosive autoregressive models
- 2013-11: Robinson Kruse, Daniel Ventosa-Santaulària and Antonio E. Noriega: Changes in persistence, spurious regressions and the Fisher hypothesis
- 2013-12: Martin M. Andreasen, Jesús Fernández-Villaverde and Juan F. Rubio-Ramírez: The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications
- 2013-13: Tom Engsted, Stig V. Møller and Magnus Sander: Bond return predictability in expansions and recessions
- 2013-14: Charlotte Christiansen, Jonas Nygaard Eriksen and Stig V. Møller: Forecasting US Recessions: The Role of Sentiments
- 2013-15: Ole E. Barndorff-Nielsen, Mikko S. Pakkanen and Jürgen Schmiegel: Assessing Relative Volatility/Intermittency/Energy Dissipation
- 2013-16: Peter Exterkate, Patrick J.F. Groenen, Christiaan Heij and Dick van Dijk: Nonlinear Forecasting With Many Predictors Using Kernel Ridge Regression
- 2013-17: Daniela Osterrieder: Interest Rates with Long Memory: A Generalized Affine Term-Structure Model
- 2013-18: Kirstin Hubrich and Timo Teräsvirta: Thresholds and Smooth Transitions in Vector Autoregressive Models
- 2013-19: Asger Lunde and Kasper V. Olesen: Modeling and Forecasting the Volatility of Energy Forward Returns - Evidence from the Nordic Power Exchange
- 2013-20: Anders Bredahl Kock: Oracle inequalities for high-dimensional panel data models
- 2013-21: Malene Kallestrup-Lamb, Anders Bredahl Kock and Johannes Tang Kristensen: Lassoing the Determinants of Retirement
- 2013-22: Johannes Tang Kristensen: Diffusion Indexes with Sparse Loadings
- 2013-23: Asger Lunde and Anne Floor Brix: Estimating Stochastic Volatility Models using Prediction-based Estimating Functions
- 2013-24: Nima Nonejad: A Mixture Innovation Heterogeneous Autoregressive Model for Structural Breaks and Long Memory