Long Memory and Structural Breaks in Realized Volatility: An Irreversible Markov Switching Approach

Nima Nonejad

CREATES Research Paper 2013-26

Long Memory and Structural Breaks in Realized Volatility: an Irreversible Markov Switching Approach

Nima Nonejad[†*]

Aarhus University and CREATES

Abstract

We propose an ARFIMA (autoregressive fractionally integrated moving average) model that is able to capture long memory and incorporate structural breaks in the model parameters. We model structural breaks through irreversible Markov switching or so-called change-point dynamics. Monte Carlo simulations demonstrate that our approach is effective in estimating the model parameters, identifying and dating structural breaks. Applied to daily S&P 500 data we find evidence of four structural breaks. The evidence of structural breaks is robust to different specifications including a GARCH specification for the conditional volatility of realized volatility.

Keywords: Change-points, Gibbs sampling, Long memory, Structural breaks
(*JEL:* C11, C22, C52, G10)

1 Introduction

Measuring and modeling volatility is a very important issue in many pricing and risk management problems. Recently, a new observable measure of volatility, called realized volatility (RV) has been proposed. Realized volatility uses high-frequency data information and has been shown to be an accurate estimate of expost volatility. RV is constructed from the sum of intraday squared returns and converges to quadratic variation for a broad class of continuous time models. A well-known property of RV is the strong serial dependence as evidenced for instance in Andersen et al. (2001) and Andersen et al. (2007). For this reason long memory models such as the ARFIMA (autoregressive fractionally integrated moving average) model have been applied to RV data.

In this paper we provide a Bayesian analysis of structural breaks in daily S&P 500 realized volatility. In particular, we propose an ARFIMA model in which the level, persistence and volatility of realized volatility parameters are subject to structural breaks. The basis of the analysis is on an ARFIMA model which builds on the Hidden Markov Chain (HMC) formulation of the multiple change-point model proposed by Chib (1998). Breaks are captured through an integer-valued state variable, s_t that tracks the regime from which a particular observation, y_t is drawn. s_t is modeled as a discrete first order Markov process with a constrained transition probability matrix. At each point in time, s_t can either remain in the current state or jump to the next state.

We investigate specifications which allow for all parameters, as well as only for a subset of parameters to change due to structural breaks. This allows one to isolate the impact of structural breaks on individual parameters and use all data in estimation of parameters that are not affected by structural breaks. Each change-point ARFIMA model is estimated conditional on $0, 1, ..., m$ breaks occurring and for each of these specifications the marginal likelihood (ML) and the deviance information criterion (DIC) are calculated and used to determine the number of change points. Specifically, we can compare marginal likelihoods using Bayes factors and use differences in DIC between different specifications to compare models and determine the number of structural breaks. It is important to note that DIC can be considered as a compelling alternative to ML. Furthermore, calculation of DIC in our MCMC scheme is trivial as the likelihood with s_t, $t = 1, ..., T$ integrated out is easily obtainable.

Our contributions in this paper are two-fold. First, we provide an efficient Markov chain Monte Carlo sampling scheme to draw s_t, $t = 1, ..., T$ and the parameters within each regime, θ_k, $k = 1, ..., m$ from their respective conditional posteriors. Furthermore, instead of using traditional approaches to evaluate the likelihood function such as Chan and Palma (1998), we build upon previous works on precision-based algorithms as in Chan and Jeliazkov (2009), Chan (2013) and use a direct approach to evaluate the likelihood function. In fact, we believe that incorporating the precision-based algorithm of Chan and Jeliazkov (2009) and Chan (2013) along with the change-

point specification of Chib (1998) within the ARFIMA setting is the most important contribution that we make. Furthermore, we also conduct an extensive Monte Carlo experiment to investigate if our methods work well in identifying data generating parameters, true structural break dates and the correct number of structural breaks. With regards to the last point we compare the ability of ML and DIC to detect the correct number of structural breaks. Our simulations, based on empirically reasonable scenarios, show that ML and DIC perform very well in identifying the true number of structural breaks. The higher the number of parameters that are affected by a break, the more likely it is that structural breaks are correctly identified. Finally, it becomes more difficult to identify one or more breaks when only the persistence parameter changes.

Empirical results for S&P 500 volatility provide strong evidence in favor of four structural breaks based on data from January 2nd, 2000 to December 31st, 2009, for a total of 2515 trading days. The effect of structural breaks is mainly confined to the conditional mean and variance with weaker evidence that the persistence parameter is also subject to structural breaks. Finally, in order to investigate if the existence of breaks is spurious due to neglected conditional variance dynamics, we also consider breaks in an ARFIMA-GARCH model. Again, evidence is strong in favor of structural breaks in the model parameters and the estimated change-point dates are close to that of the change-point ARFIMA model.

The structure of this paper is as follows. In Section 2, we present the change-point ARFIMA model. Bayesian estimation techniques and model comparison methods are presented in section 3. Section 4 presents the Monte Carlo results. In Section 5, we briefly review the theory behind realized volatility. Section 6 is the application on S&P 500 volatility while section 7 concludes. An appendix explains how to evaluate the likelihood using the precision-based algorithm of Chan and Jeliazkov (2009) and Chan (2013).

2   Change-point ARFIMA Model

Consider an ARFIMA (AutoRegressive Fractionally Integrated Moving Average) model

$$y_t \;=\; \mu + (1-L)^{-d}\varepsilon_t, \quad \varepsilon_t \sim N\left(0,\sigma^2\right) \tag{2.1}$$

for $t = 1,...,T$. y_t is the actual observation, L is the lag operator such that $L\varepsilon_t = \varepsilon_{t-1}$ and d determines the long memory of the process. The fractional difference operator, $(1-L)^{-d}$ in (2.1) is defined as

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

3

where $\Gamma(.)$ is the Gamma function. Equation (2.1) is a generalization of the moving average (MA) model to non-integer values of d. Specifically, if $d > 0$ the process is said to have long memory since the autocorrelations die out at a hyperbolic rate. For $0 < d < 0.5$, (2.1) is a stationary long-memory process with non-summable autocorrelation functions. For $d = 0$, we have that $y_t = \mu + \varepsilon_t$.

There are many ways to estimate (2.1), see for example Beran (1994) and Robinson (2003). In this paper we focus on MCMC methods for inference, in particular Gibbs sampling. We rely on the idea of Chan and Palma (1998). Specifically, Chan and Palma (1998) consider an approximation of (2.1) based on a truncation lag of order M. Thereafter, the likelihood is computed using the Kalman filter. However, instead of using the Kalman filter we take a different approach to evaluate the likelihood function. Our approach extends previous works on precision-based algorithms as in Chan and Jeliazkov (2009) and Chan (2013) using a direct approach to evaluate the likelihood function.

The aforementioned method exploits the special structure of (2.1), particularly that the covariance matrix for the joint distribution of $Y_T = (y_1, ..., y_T)'$ is sparse, i.e. it contains only few non-zero elements. By exploiting the sparse structure of the covariance matrix of Y_T, we are able to develop an easy and fast method for evaluating the likelihood function.

Conditional on the model parameters, $\theta = (\mu, d, \sigma^2)'$ and M we can write (2.1) as $Y_T = u + H\varepsilon$ where $u = \mu\iota$, ι is a $T \times 1$ vector of 1's, $\varepsilon = (\varepsilon_1, ..., \varepsilon_T)' \sim N(0, S_Y)$ and $S_{Y_T} = \sigma^2 I_T$. H is a $T \times T$ lower triangular matrix with ones on the main diagonal and

$$
H = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\
\pi_1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\
\pi_2 & \pi_1 & 1 & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \pi_1 & 1 & 0 & 0 & \cdots & 0 \\
\pi_M & \pi_{M-1} & \vdots & \pi_1 & 1 & 0 & \cdots & 0 \\
0 & \pi_M & \pi_{M-1} & \vdots & \pi_1 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & \pi_M & \pi_{M-1} & \cdots & \pi_1 & 1
\end{bmatrix}
$$

where $\pi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}$. It is important to note that in general H is a banded $T \times T$ matrix that contains only $(T-M/2)(M+1) < T(M+1))$ non-zero elements. Using the algorithm of Chan (2013) it is shown in the appendix that $p(Y_T \mid \theta, M)$ has a closed form solution given as

$$
\log p(Y_T \mid \theta, M) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2}(Y_T - u)'\Omega_{Y_T}^{-1}(Y_T - u) \tag{2.2}
$$

where $\Omega_{Y_T} = HS_{Y_T}H'$ and $p(.)$ denotes the density of all random quantities.

Allowing for structural breaks in the parameters of (2.1) is straightforward. Assume that there are $m-1$, $m \in \{1,2,...\}$ change points at unknown times $\{\tau_1, \tau_2, ..., \tau_{m-1}\}$. Separated by those change points, there are m different regimes. The density of y_t depends on $\theta_k = \{\mu_k, d_k, \sigma_k^2\}$, $k = 1,2,...m$ whose value changes at the change points $\{\tau_1, \tau_2, ..., \tau_{m-1}\}$ and remains constant otherwise

$$
\theta = \begin{cases}
\theta_1 & \text{if} & t < \tau_1 \\
\theta_2 & \text{if} & \tau_1 \le t < \tau_2 \\
\vdots & \vdots & \vdots \\
\theta_{m-1} & \text{if} & \tau_{m-2} \le t < \tau_{m-1} \\
\theta_m & \text{if} & \tau_{m-1} \le t
\end{cases}
\tag{2.3}
$$

Let $S = (s_1, ..., s_T)'$ where $s_t = k$ indicates that y_t is from regime k. The one-step-ahead transition probability matrix for s_t is given as

$$
P = \begin{bmatrix}
p_{11} & p_{12} & 0 & \cdots & 0 \\
0 & p_{22} & p_{23} & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & 0 & p_{m-1,m-1} & p_{m-1,m} \\
0 & 0 & \cdots & 0 & 1
\end{bmatrix}
\tag{2.4}
$$

where $p_{lk} = p(s_t = k \mid s_{t-1} = l)$ with $k = l$ or $k = l+1$ is the probability of moving from regime l at time $t-1$ to regime k at time t. P ensures that given $s_t = k$ at time t, the next period, $t+1$, $s_t + 1$ remains in the same state or jumps to the next state. For instance, given $s_t = k$, one has $s_{t+1} = k$ or $s_{t+1} = k+1$ with $p_{k,k} + p_{k,k+1} = 1$. Once the last regime is reached, one stays there forever, that is $p_{m,m} = 1$. This structure enforces the ordering (2.3) on the change points.

3   Bayesian Estimation

To conduct model estimation we jointly estimate the long memory dynamics and S. However, model estimation is not straightforward. First, S is not observable. Second, there is no standard method to draw d_k or μ_k from their conditional posterior densities. Although the joint posterior density of the model, $p(P, \theta, M, S \mid Y_T)$ is not a well-known density, samples from it can be obtained using Gibbs sampling and Metropolis-Hastings (M-H). The parameters are divided into four blocks: $\theta = \{\theta_k\}_{k=1}^m$, S, M and P. The Gibbs sampler requires the following steps: first, choose starting values for P, θ and M, i.e. $P^{(0)}$, $\theta^{(0)}$, $M^{(0)}$ and set $i = 1$. Then iterate from

1. $S^{(i)} \mid P^{(i-1)}, \theta^{(i-1)}, M^{(i-1)}, Y_T$

2. $\left\{ \theta_k^{(i)} \right\}_{k=1}^{m} \mid P^{(i-1)}, M^{(i-1)}, S^{(i)}, Y_T$

3. $M^{(i)} \mid P^{(i-1)}, \theta^{(i)}, S^{(i)}, Y_T$

4. $P^{(i)} \mid \theta^{(i)}, M^{(i)}, S^{(i)}, Y_T$

5. Set $i = i+1$ and go to the first step.

Notice that in step 2 of iteration i of the Gibbs sampler each element of θ is updated one-at-a-time. After dropping a set of burn-in samples, the remaining draws are collected for inference. For N large enough, any function of interest can be consistently estimated. For instance,

$$\widehat{f(\theta)} \;=\; \frac{1}{N} \sum_{i=1}^{N} f\left(\theta^{(i)} \right)$$

is a consistent estimate of $E\left[f(\theta) \mid Y_T \right]$, the posterior mean of $f(\theta)$. Below more details are provided on each step of the Gibbs sampling procedure.

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

$$p(S \mid P, \theta, M, Y_T) \;=\; p(s_T \mid P, \theta, M, Y_T) \prod_{t=1}^{T-1} p(s_t \mid s_{t+1}, P, \theta, M, Y_t) \tag{3.1}$$

in which one samples sequentially from each density on the right-hand-side of (3.1) beginning with $p(s_T \mid P, \theta, M, Y_T)$, and then $p(s_t \mid s_{t+1}, P, \theta, M, Y_t)$, $t = T-1, ..., 1$. 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 (3.1) are obtained based on the following steps:

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

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

$$p(s_t = k \mid P, \theta, M, Y_{t-1}) \;=\; \sum_{l=k-1}^{k} p(s_{t-1} = l \mid P, \theta, M, Y_{t-1}) \, p_{lk}, \quad k = 1, ..., m \tag{3.2}$$

$$p(s_t = k \mid P, \theta, M, Y_t) \;=\; \frac{p(s_t = k \mid P, \theta, M, Y_{t-1}) \, p(y_t \mid \theta, M, Y_{t-1}, s_t = k)}{\sum_{l=1}^{m} p(s_t = l \mid P, \theta, M, Y_{t-1}) \, p(y_t \mid \theta, M, Y_{t-1}, s_t = l)},$$
$$k = 1, ..., m$$

The last equation is obtained from Bayes' rule. Note that in (3.2) the summation is only from $k-1$ to k, due to the restricted nature of the transition matrix and $p(y_t \mid \theta, M, Y_{t-1}, s_t = k)$ has a closed

form solution, see (2.2).

(c) Finally, Chib (1998) shows that the individual densities in (3.1) are

$$p\left(s_t \mid s_{t+1}, P, \theta, M, Y_t\right) \;\; \propto \;\; p\left(s_t \mid P, \theta, M, Y_t\right) p\left(s_{t+1} \mid s_t, P\right)$$

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

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

where

$$c_t \;\; = \;\; \frac{p\left(s_t = k \mid P, \theta, M, Y_t\right) p\left(s_{t+1} = k \mid s_t = k, P\right)}{\sum_{l=k-1}^{k} p\left(s_t = l \mid P, \theta, M, Y_t\right) p\left(s_{t+1} = k \mid s_t = l, P\right)}$$

Finally, note that $p\left(s_1 = 1 \mid s_2, P, \theta, M, Y_1\right) = 1$.

Step 2: Simulation of $\theta_k \mid \theta_1, ..., \theta_{k-1}, \theta_{k+1}, ..., \theta_m, M, S, Y_T$. For each regime, the conditional posterior of θ_k depends only on information in regime k. Furthermore, compared to σ_k^2 sampling μ_k and d_k is more complicated since their conditional posteriors do not have closed form. Therefore, the Metropolis-Hastings algorithm is used. Let $\hat{Y}_k = \{y_t : s_t = k\}$ denote the observations in regime k. We sample μ_k and d_k, $k = 1, ..., m$ one-at-a-time. For example, μ_k is sampled in the following way:

1. Sample a candidate, μ_k^* from a random walk proposal $q\left(\mu_k^* \mid \mu_k^{(i-1)}\right) \sim N\left(\mu_k^{(i-1)}, \Sigma_k\right)$ where Σ_k is chosen by the researcher in a manner to ensure a sufficient acceptance rate. We follow Koop (2003, page 98) and adjust Σ_k to get an acceptance rate roughly around 30 to 40%. We do this by experimenting with different values of Σ_k until we find one which yields a reasonable acceptance rate probability.

2. Define the acceptance probability of μ_k^* as

$$a_{\text{MH}}\left(\mu_k^*, \mu_k^{(i-1)}\right) = \min\left\{1, \frac{p\left(\mu_k^* \mid d_k^{(i-1)}, \sigma_k^{2(i-1)}, M^{(i-1)}, \hat{Y}_k\right) q\left(\mu_k^{(i-1)} \mid \mu_k^*\right)}{p\left(\mu_k^{(i-1)} \mid d_k^{(i-1)}, \sigma_k^{2(i-1)}, M^{(i-1)}, \hat{Y}_k\right) q\left(\mu_k^* \mid \mu_k^{(i-1)}\right)}\right\} \quad (3.3)$$

3. Draw u from the standard Uniform distribution, U. If $u \leq a_{\text{MH}}\left(\mu_k^*, \mu_k^{(i-1)}\right)$ then set $\mu_k^{(i)} = \mu_k^*$ else set $\mu_k^{(i)} = \mu_k^{(i-1)}$.

Finally, $\sigma_k^2 \mid \mu_k, d_k, M, \hat{Y}_k \sim IG\left(v_k/2, l_k/2\right)$ where $IG\left(\frac{\cdot}{2}, \frac{\cdot}{2}\right)$ stands for the Inverse-gamma density, see Kim and Nelson (1999), $v_k = T_k + v_0$, $l_k = \hat{\varepsilon}_k' \hat{\varepsilon}_k + l_0$, T_k is the number of observations in regime

k and $\hat{\varepsilon}_k = \{\hat{\varepsilon}_t : s_t = k\}$. v_0 and l_0 are the prior hyperparameter values.

Step 3: Simulation of $M \mid P, \theta, S, Y_T$. In order to sample M from its conditional posterior we use the same method as in Raggi and Bordignon (2012). A truncation parameter, M^* is proposed from a discretized Laplace proposal, $q\left(M^* \mid M^{(i-1)}\right) = \frac{1}{2\lambda} \exp\left(-\lambda \left| M^* - M^{(i-1)} \right|\right)$ where $\lambda = 0.1$ in order to obtain small moves. The Metropolis-Hastings acceptance probability is given as

$$a_{\text{MH}}\left(M^*, M^{(i-1)}\right) = \min\left\{1, \frac{p\left(M^* \mid \theta^{(i)}, S^{(i)}, Y_T\right) q\left(M^{(i-1)} \mid M^*\right)}{p\left(M^{(i-1)} \mid \theta^{(i)}, S^{(i)}, Y_T\right) q\left(M^* \mid M^{(i-1)}\right)}\right\} \qquad (3.4)$$

Step 4: Simulation of $P \mid S$. Assume that $p_{kk} \sim Beta\left(a_0, b_0\right)$. The conditional posterior for each diagonal component of P is then

$$p_{kk} \mid S \quad \sim \quad Beta\left(a_0 + n_{kk}, b_0 + 1\right), \quad k = 1, ..., m-1$$

where n_{kk} is the number of one-step transitions from state k to state k in a sequence of S.

3.1 Breaks in μ and d

Suppose that only μ and d are subject to structural breaks. Thus, we have μ_k and d_k for $k = 1, ..., m$ while the conditional variance, σ^2 is constant trough time. Modeling this specification is straightforward because as before we can use the conditioning properties of the Gibbs sampler.

Specifically, in order to sample $\mu_k^{(i)}$ and $d_k^{(i)}$ for $k = 1, ..., m$ we use $S^{(i)}, M^{(i-1)}, \hat{Y}_k = \{y_t : s_t = k\}$ and perform M-H to obtain $\mu_k^{(i)}$ and $d_k^{(i)}$ using (3.3). Conditional on $S^{(i)}, \mu_1^{(i)}, ..., \mu_m^{(i)}, d_1^{(i)}, ..., d_m^{(i)}$, $M^{(i-1)}$ and Y_T we then draw $\sigma^{2(i)}$ from the Inverse-gamma density. The remaining parameters, $M^{(i)}$ and $P^{(i)}$ are also sampled conditional on $S^{(i)}, \mu_k^{(i)}, d_k^{(i)}, k = 1, ..., m, \sigma^{2(i)}$ and Y_T using **Step 3** and **Step 4** from the previous section.

3.2 Only breaks in σ^2

Now suppose only σ^2 changes between regimes, while μ and d are constant. For this case, we draw $\sigma_k^{2(i)}$ from the Inverse-gamma density, $IG\left(\frac{v_k}{2}, \frac{l_k}{2}\right)$ for $k = 1, ...m$ using $S^{(i)}$ from **Step 1**, $\mu^{(i-1)}, d^{(i-1)}, M^{(i-1)}$ and Y_T. Thereafter, we stack $\sigma_k^{2(i)}$'s using $s_t = k$ and construct a vector of time-varying conditional variances, $\sigma_t^{2(i)}$. To complete the cycle we sample $\mu^{(i)}$ and $d^{(i)}$ conditional on $\sigma_1^{2(i)}, ..., \sigma_T^{2(i)}, M^{(i-1)}, Y_T$.

Finally, note that we can consider breaks in the conditional variance with only partial breaks in μ or d by combining the methods in the last two subsections.

3.3 Bayes factors and marginal likelihood computation

Let $\mathscr{M}$ denote a model parametrization in which some or all parameters are subject to breaks. The marginal likelihood (ML) of model $\mathscr{M}$ is defined as

$$p(Y_T \mid \mathscr{M}) = \int p(Y_T \mid P, \theta, M, \mathscr{M}) p(P, \theta, M \mid \mathscr{M}) dP d\theta dM \qquad (3.5)$$

The marginal likelihood is a measure of the success the model has in accounting for the data after parameter uncertainty has been integrated out over the prior, $p(P, \theta, M \mid \mathscr{M})$. $p(Y_T \mid P, \theta, M, \mathscr{M})$ is the likelihood function with S integrated out. It is calculated as

$$\log p(Y_T \mid P, \theta, M, \mathscr{M}) = \sum_{t=1}^{T} \log p(y_t \mid P, \theta, M, \mathscr{M}, Y_{t-1}) \qquad (3.6)$$

where

$$p(y_t \mid P, \theta, M, \mathscr{M}, Y_{t-1}) = \sum_{k=1}^{m} p(y_t \mid \theta, M, \mathscr{M}, Y_{t-1}, s_t = k) \qquad (3.7)$$
$$p(s_t = k \mid P, \theta, M, \mathscr{M}, Y_{t-1})$$

The last term on the right-hand-side of (3.7) is computed from (3.2). In the following steps the model index, $\mathscr{M}$ is suppressed for conciseness. Gelfand and Dey (1994), henceforth G-D propose a method to calculate ML based on

$$\frac{1}{N} \sum_{i=1}^{N} g\left(\theta^{(i)}\right) / \left[p\left(Y_T \mid P^{(i)}, \theta^{(i)}, M^{(i)}\right) p\left(P^{(i)}, \theta^{(i)}, M^{(i)}\right)\right] \rightarrow p(Y_T)^{-1} \quad \text{as} \quad N \rightarrow \infty \quad (3.8)$$

It applies to any posterior simulator, no matter what algorithm is used. The prior, $p\left(P^{(i)}, \theta^{(i)}, M^{(i)}\right)$ can be evaluated directly and $p\left(Y_T \mid P^{(i)}, \theta^{(i)}, M^{(i)}\right)$ is calculated by substituting $\theta^{(i)}$ into the likelihood function with S integrated out, (3.6). Gelfand and Dey (1994) show that if $g\left(\theta^{(i)}\right)$ is thin-tailed relative to $p\left(Y_T \mid P^{(i)}, \theta^{(i)}, M^{(i)}\right) p\left(P^{(i)}, \theta^{(i)}, M^{(i)}\right)$ then (3.8) is bounded and the estimator is consistent. Following Geweke (2005) a truncated Normal distribution, $N(\theta^*, \Sigma^*)$ is used for $g(\theta)$ where θ^* and Σ^* are the posterior sample moments calculated as $\theta^* = N^{-1} \Sigma_{i=1}^{N} \theta^{(i)}$ and $\Sigma^* = N^{-1} \Sigma_{i=1}^{N} \left(\theta^{(i)} - \theta^*\right) \left(\theta^{(i)} - \theta^*\right)'$ whenever $\theta^{(i)}$ is in the domain of the truncated Normal. The domain, Θ is defined as

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

where z is the dimension of the parameter vector and $\chi^2_\alpha(z)$ is the αth percentile of the Chi-squared distribution with z degrees of freedom. In practice, 0.75, 0.95 and 0.99 are popular selections for α. High values of α work best since then more draws will be included when estimating (3.8).

We estimate (3.8) using the aforementioned values of α. In general, we find that different values of α lead overwhelmingly to very similar results, see section 4.2. It was also suggested to compute the marginal likelihood using the method of Sims et al. (2008) because as pointed out in Sims et al. (2008) the G-D method may not work for models with time-varying parameters as the posterior density tends to be non-Gaussian. We also calculate ML for the empirical part using the method of Sims et al. (2008). However, we do not find any significant qualitative changes compared to G-D. Therefore, we choose to retain these values. Furthermore, Monte Carlo results clearly indicate that G-D correctly identifies the true model in the presence of structural breaks.

Once we calculate the marginal likelihood for different specifications, we can compare models across the number of regimes as well as across the subset of parameters using Bayes factors. The Bayes factor (BF) for model $\mathcal{M}_A$ versus model $\mathcal{M}_B$ is

$$BF_{\mathcal{M}_{AB}} = \frac{p(y_1,...,y_T \mid \mathcal{M}_A)}{p(y_1,...,y_T \mid \mathcal{M}_B)}$$

This odds ratio is the factor by which the data considers $\mathcal{M}_A$ more probable than $\mathcal{M}_B$. Kass and Raftery (1995) suggest interpreting the evidence for $\mathcal{M}_A$ as: not worth more than a bare mention for $1 \le BF_{\mathcal{M}_{AB}} < 3$; positive for $3 \le BF_{\mathcal{M}_{AB}} < 20$; strong for $20 \le BF_{\mathcal{M}_{AB}} < 150$; and very strong for $BF_{\mathcal{M}_{AB}} \ge 150$.

4 Monte Carlo

In this section, a set of Monte Carlo simulations is conducted to investigate the ability of the change-point ARFIMA model to detect the correct number of change points. The effect of different sample sizes and the ability of the deviance information criterion (DIC) to detect the correct number of change points is also considered. Specifically, we compare the performance of DIC with ML to find out if DIC is just as capable of identifying the true model from which the data is generated as ML. We do this because computing DIC for the change-point ARFIMA model is almost trivial, see section 4.4. However, as pointed out by Spiegelhalter et al. (2002) we must be cautions against using ML as a basis against which to assess DIC. ML addresses how well the prior has predicted the observed data whereas DIC addresses how well the posterior might predict future data generated by the same parameters that give rise to the observed data.

4.1 Setting

The ARFIMA model based on equation (2.1), in which change points affect one or more model parameters is considered. Table 1 lists all the specifications used in the simulations and the empirical application. Specifically, $\mathcal{M}_0$ is the simple ARFIMA model without any structural changes, $\mathcal{M}_1$ to $\mathcal{M}_5$ are models in which different parameters change and $\mathcal{M}_6$ is the model in which all parameters change from a structural break. For each model, a sample of $T = 1000$ observations is generated. The true models considered include cases of no, one and two change points. In settings with change points the positions of the change points follow a Uniform distribution, U. For instance, when there is one change point, the position of this change point follows $U(0.25 \times T, 0.75 \times T)$. When there are two change points, the first one follows $U(0.25 \times T, 0.40 \times T)$ and the second follows $U(0.60 \times T, 0.80 \times T)$.

The parameter values of the data generating process (DGP) under different scenarios are listed in Table 2. For example, for $\mathcal{M}_1$ only μ changes while the other parameters remain constant. In specifications $\mathcal{M}_4$ to $\mathcal{M}_6$, the time series properties change greatly and the change points should be identified rather easily. Specifications $\mathcal{M}_1$ to $\mathcal{M}_3$ are rather challenging because the model structure is more stable.

We specify the priors as: $d_k \sim N(0, 100)$ truncated such that $0 < d_k < 0.5$, $\mu_k \sim N(0, 100)$ and $\sigma_k^2 \sim IG\left(\frac{4}{2}, \frac{0.02}{2}\right)$, $k = 1, ..., m$. A suitable prior for M is the Poisson truncated distribution with $M \in \{M_{min}, ..., M_{max}\}$ where in this paper $M_{min} = 10$ and $M_{max} = 50$, see also Raggi and Bordignon (2012). Finally, we assume that $p_{kk} \sim Beta(8, 0.1)$, $k = 1, ..., m - 1$. In this setting, d_k, μ_k, σ_k^2 and M are very uninformative, while the prior for p_{kk} favors infrequent structural breaks. We conduct a prior sensitivity analysis with regards to the S&P 500 data and report the results in section 6.4.

4.2 Change point identification

It is assumed that model specification $\mathcal{M}_i$ is known but the number or dates of the structural breaks are not known. For each draw from the data generating process (DGP), the change-point ARFIMA model assuming 0, 1, 2 and 3 structural breaks is estimated. Evidence for the number of break points is then ranked according to the highest marginal likelihood (ML). We calculate ML using G-D with $\alpha = 0.99$. The best model specification has the highest marginal likelihood, the second best has next highest, etc. Thereafter, a new data from the DGP is generated and the procedure is repeated until 100 repetitions are completed. The frequency over repetitions in which each specification is best according to the marginal likelihood criterion is then reported.

Table 3 lists the results for each specification. For convenience, bold entries in these cells should be 100 in case of perfect classification. For example, the second row for $\mathcal{M}_1$ says that for a DGP with one change point, 99 times is correctly identified as one change point in terms of ML while

1 is identified incorrectly as two change points. The next entry in the table repeats this for a DGP with two change points. Here, 100 times two change points is correctly identified.

Overall, the change-point ARFIMA model works very well. When there is no change point, this is correctly selected most of the time. Looking at these cases (first row in each panel) they are: $100/100$ for $\mathcal{M}_1$, $87/100$ for $\mathcal{M}_2$ and $100/100$ for $\mathcal{M}_3 - \mathcal{M}_6$. When the process contains change points, the marginal likelihood method correctly identifies existence of the change points in most cases. For example, the probability of correctly identifying instability of the process is $0.99 \, (86 + 5 + 8)/100$ for $\mathcal{M}_2$ and one change point, and is 1 for two change points. The correct number of change points is found most of the times. Looking at the numbers in bold, many of them are close to 100. However, relatively smaller numbers associated with $\mathcal{M}_2$ show that ML is less powerful when there are changes only in d. Therefore, for DGPs where d is subject to structural breaks it becomes easier to identify the breaks when more parameters undergo a change. For example, compare $\mathcal{M}_2$ with the better performance of $\mathcal{M}_4$ in which both μ and d change from a structural break.

We also investigate the ability of G-D to identify the true model for different values of α. Therefore, besides $\alpha = 0.99$ we repeat our Monte Carlo experiment with $\alpha = 0.75$ and $\alpha = 0.95$. Results for $\mathcal{M}_1$, $\mathcal{M}_5$ and $\mathcal{M}_6$ are summarized in Table 4. We get identical results for these specifications and for $\mathcal{M}_0, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4$ (not reported) as well. However, the Bayes factor in favor of the true model in each case varies slightly across the values of α.

4.3 Parameter estimates

Given a full MCMC run we calculate the mean, median and mode of the posterior draws, $\theta_k^{(i)}$, $i = 1, ... N$. We then take the mean of these quantities over the number of Monte Carlo replications. Finally, we also consider the root mean squared error (RMSE) for each parameter in each regime

$$RMSE = \sqrt{\frac{1}{R} \frac{1}{N} \sum_{h=1}^{R} \sum_{i=1}^{N} \left(\theta_{k,h}^{(i)} - \theta_k \right)^2}$$

where $\theta_{k,h}^{(i)}$ is the ith posterior draw of θ_k at the hth Monte Carlo iteration and θ_k is the vector of the true DGP parameters in regime k. We summarize results for $\mathcal{M}_4$ with 2 change points for $T = 500$, $T = 1000$ and $T = 2000$ in Table 5.

Overall, we see that the change-point ARFIMA model works very well as on average the estimates are very close to their true values. Compared to $T = 500$, as we increase the number of observations in the DGP to $T = 1000$ the RMSE for each parameter drops on average by 10 to 40%. The RMSE for each parameter drops even more when we increase the sample size to $T = 2000$.

4.4 Deviance information criterion

Another approach for comparing the evidence for the number of change points is by using the deviance information criterion (DIC) of Spiegelhalter et al. (2002). It is a compelling alternative to AIC or BIC and it can be applied to nested or non-nested models. Calculation of DIC in a MCMC scheme is trivial. Contrary to AIC or BIC it does not require maximization over the parameter space. DIC is a combination of $p(Y_T \mid P, \theta, M)$ and a penalty term, p_D which describes the complexity of the model and serves as a penalization term that corrects deviance's propensity towards models with more parameters. More precisely, $p_D = \overline{D(P, \theta, M)} - D(\bar{P}, \bar{\theta}, \bar{M})$ where $\overline{D(P, \theta, M)}$ is approximated by $N^{-1} \Sigma_{i=1}^{N} - 2 \log p\left(Y_T \mid P^{(i)}, \theta^{(i)}, M^{(i)}\right)$ and $D(\bar{P}, \bar{\theta}, \bar{M}) = -2 \log p\left(Y_T \mid \bar{P}, \bar{\theta}, \bar{M}\right)$ where $\bar{P}$, $\bar{\theta}$ and $\bar{M}$ are estimated from the Gibbs output using mean or mode of the posterior draws. The DIC is defined as

$$DIC \;=\; D\left(\bar{P}, \bar{\theta}, \bar{M}\right) + 2 p_D$$

It is worth mentioning that the best model is the one with the smaller DIC. However, it is difficult to say what would constitute a significance difference in DIC. Very roughly, differences of more than 10 might definitely rule out the model with the higher DIC.

The DIC is calculated for specifications $\mathscr{M}_1$, $\mathscr{M}_3$, $\mathscr{M}_6$ and results are summarized in Table 6. Similar results are obtained for other cases. The number of times out of 100 repetitions that a specific change-point model is selected as best according to DIC and ML is reported. For instance, in the top row, for DGP $\mathscr{M}_6$ with no change points, 84 out of 100 repetitions the no change-point model has the smallest DIC, while $12/100$ times the one change-point model is best and $2/100$ times two and three change-point models are best. When there are two structural breaks in the DGP for $\mathscr{M}_6$, DIC correctly identifies the true model $97/100$ times whereas $3/100$ the three change-point model is best. On the other hand, for this DGP, ML identifies the correct model $100/100$ times.

4.5 Higher number of change points

In section 4.2 we evaluated the performance of the change-point ARFIMA model under 0, 1 and 2 structural breaks. In this section we evaluate the performance of the change-point ARFIMA model when the DGP contains more change points. Specifically, we set $T = 2000$ and simulate data containing 4 change points. As before, the position of the change points follow a Uniform distribution. We consider specifications $\mathscr{M}_1$, $\mathscr{M}_3$, $\mathscr{M}_4$, $\mathscr{M}_6$ and for each specification we estimate the change-point ARFIMA model assuming $0, ..., 5$ structural breaks.

We find that DIC and ML correctly identify the true model in most cases. However, we are also interested in whether or not our method is able to correctly identify the position of the change

points. Therefore, for each Monte Carlo iteration we calculate the position of the change points using the mode of $\left\{ S^{(i)} \right\}_{i=1}^{N}$ and compare it with the true position of the change points. Specifically, for each change point we calculate

$$DIFF_k = \frac{1}{R} \sum_{h=1}^{R} \left| \overline{cp}_k^{(h)} - cp_k^{(h)} \right|$$

for $k = 1, ..., m-1$ where $\overline{cp}_k^{(h)}$ is the kth estimated change point date and $cp_k^{(h)}$ is the true position of the kth change point for the hth Monte Carlo iteration. For results to give meaning, we focus only on specifications that have the correct number of change points, i.e. four change points.

Table 7 reports $DIFF_k$ for each specification when the underlying process contains 4 change points. For example, for $\mathcal{M}_6$ we find that on average we miss the correct date of the second change point by 0.6 time periods (or $0.6 \approx 1$ day if we work with daily data). On the other hand, we correctly identify the correct date of the last change point at every Monte Carlo iteration as $\left| \overline{cp}_4^{(1)} - cp_4^{(1)} \right| = \left| \overline{cp}_4^{(2)} - cp_4^{(2)} \right| = ... = \left| \overline{cp}_4^{(R)} - cp_4^{(R)} \right| = 0$. The same happens for the third and forth change point for $\mathcal{M}_1$ and $\mathcal{M}_4$.

Overall, we find that the change-point ARFIMA model works very well in identifying the true dates of the change points. Specifications $\mathcal{M}_1$, $\mathcal{M}_4$ and $\mathcal{M}_6$ are more accurate than $\mathcal{M}_3$.

4.6 Sample size

In order to assess the robustness of the ML criterion with respect to different sample sizes the more challenging specifications are considered with sample sizes of 500, 1000 and 2000. Therefore, results for $\mathcal{M}_2$, with one and two change points along with $\mathcal{M}_4$ with two change points are reported in Table 8. Increasing the number of observations improves identification of the true number of change points. The distribution is also more concentrated on the true model. For instance, for DGP, $\mathcal{M}_2$, 2 CP with $T = 500$ we identify the correct specification only $54/100$ times whereas the for $T = 1000$ the correct specification is selected $95/100$ times and for $T = 2000$ the correct specification is selected $98/100$ times.

5 Realized Volatility

Suppose that, along day t, the logarithmic prices of a given asset follow a continuous-time diffusion process

$$dp(t+s) = \mu(t+s)\,dt + \sigma(t+s)\,dW(t+s), \quad 0 \le s \le 1, \quad t = 1, 2, ...$$

where $p(t+s)$ is the logarithmic price at time $t+s$, $\mu(t+s)$ is the drift component, $\sigma(t+s)$ is the instantaneous volatility and $W(t+s)$ is a standard Brownian motion. In addition, suppose that $\sigma(t+s)$ is orthogonal to $W(t+s)$, such that there is no leverage effect. This assumption is standard in the realized volatility literature. Andersen et al. (2003), Barndorff-Nielsen and Shephard (2002a) show that daily returns, defined as $r_t = p(t) - p(t-1)$ are conditionally Gaussian. That is

$$r_t \mid \mathscr{F}_t \sim N\left(\int_0^1 \mu(t+s-1)\,ds, \int_0^1 \sigma^2(t+s-1)\,ds\right)$$

The true volatility for day t is defined as $IV_t = \int_0^1 \sigma^2(t+s-1)\,ds$ and is known as the integrated volatility. In the absence of microstructure noise, *realized volatility* is a consistent estimator of IV_t as the intraday sampling frequency goes to infinity. Realized volatility (RV) is constructed from the sum of intraday squared returns, $\Sigma_{j=1}^n r_{j,t}^2$ where $r_{j,t} = p_{j,t} - p_{j-1,t}$. $p_{j,t}$ is the jth intraday price and n is the number of intra-daily observations. As pointed out by for example Andersen et al. (2003), RV_t is more efficient than traditional measures of volatility, such as daily squared returns.

Market microstructure dynamics contaminate the price process with noise. Hence, RV_t can be a biased and inconsistent estimator of IV_t, see Hansen and Lunde (2006) for more details on the effects of market microstructure noise on volatility estimation. In order to reduce the effects of market microstructure noise, we employ a kernel-based estimator that utilizes autocovariances of intraday returns. Specifically, we follow Hansen and Lunde (2006) and provide a bias correction to realized volatility in the following way

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

where q is a small positive integer and we set $q = 1$. Henceforth, RV_t^q is referred to as RV_t.

6 Application to S&P 500 Volatility

6.1 Data

The empirical application is based on S&P 500 index data using the Spyder (SPY) fund. The data consists of 5 minutes intra-daily observations from January 2nd, 2000 to December 31st, 2009, for a total of $T = 2515$ 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 returns are constructed. These returns are used to construct realized volatility. Following Raggi and Bordignon (2012), the annualized realized standard deviation, $y_t = \frac{\sqrt{252RV_t}}{100}$ is considered.

15

However, there are outliers in y_t, and therefore we risk that a single outlier may wrongly be identified as a separate regime. To rule this out, we follow Kim et al. (2005) and Liu and Maheu (2008) imposing the assumption that each regime lasts at least 66 days. Specifically, we perform the following: when we simulate a draw of S which has a regime shorter than 66 days, we discard it and resample until each regime is 66 days or more in length. We find that our results are robust to different duration restrictions, see section 6.3. The first 1000 draws are discarded and the next 5000 are used for posterior inference.

6.2 Results

To conduct estimation, we use the same priors as in section 4.1. We investigate models under the structural change configurations in Table 1. Table 9 displays log(ML) and DIC (indicated inside the parentheses) for specifications with no change point up to six change points.

Results suggest existence of four change points according to ML and DIC. The log marginal likelihood for no change point is -1946.84 and most specifications with structural breaks improve on this. The difference between the best structural break specification ($\mathcal{M}_5$, 4 CP) and $\mathcal{M}_0$ is large with a Bayes factor of $\exp(689.74)$ in favor of four structural breaks. This is very strong evidence. For all model settings, except $\mathcal{M}_2$, the largest ML (lowest DIC) occurs at four change points. There is some posterior support for four change points for $\mathcal{M}_2$, but it is outperformed by its one change point counterpart with a Bayes factor of $\exp(3.03)$.

We also compare models across parameter specifications. The highest log(ML) and lowest DIC model across all cases is -1257.10 and 2360.45 respectively for $\mathcal{M}_5$ with four change points. This means that there are four structural breaks in μ and σ^2. Comparing $\mathcal{M}_5$, 4 CP with the second highest ML and second lowest DIC, which are -1290.64 and 2378.46 for $\mathcal{M}_6$, also with four change points, we find that the Bayes factor for $\mathcal{M}_5$, 4 CP versus $\mathcal{M}_6$, 4 CP is $\exp(33.54)$. At the same time we find that the difference in DIC between these two models is 18.

Figure 1 displays the data and the estimated change-point dates shown by vertical lines using the posterior mode of $\left\{S^{(i)}\right\}_{i=1}^{N}$ for $\mathcal{M}_5$, 4 CP. The change-point dates are given as: 04-10-2003, 07-17-2007, 09-16-2008 and 03-19-2009. As seen in Figure 1 the last two break points occur during the financial crises (fall of 2008 and spring of 2009). The first and second break points correspond roughly to the beginning of the Iraq war in 2003 and the beginning of the subprime crisis in the US in 2007.

Table 10 reports some summary statistics concerning posterior distribution of the key parameters for $\mathcal{M}_5$, 4 CP and some diagnostics. Specifically, after discarding the first 1000 iterations we collect the final sample and compute the posterior mean, of θ, 95% credibility intervals (indicated inside the brackets), inefficiency measures, R_B, Geweke's convergence statistics, Metropolis-

Hastings acceptance ratios, DIC and log(ML)[1].

None of the density intervals for the parameters include 0. The order of integration, d, is estimated at 0.45 (compared to 0.48 for $\mathcal{M}_0$). This implies that S&P 500 data exhibits long memory behavior. When we look at parameters that change from a break we find that there are sensible differences across regimes. For instance, we see that μ_k increases from 0.15 to 0.39 during the regime that covers the financial crisis and then subsequently falls to 0.17 from the last change point to the end of the sample. The same happens for the conditional volatility of realized volatility, σ_k. Specifically, we estimate σ_4 at 0.82 during the 2008 financial crises which is twice as large as σ_3 which we estimate at 0.42. Thereafter, σ_k falls to 0.28 from the last change point to the end of the sample.

6.3   Robustness to minimum duration restrictions

In this section we test the robustness of the minimum regime durations. We follow Liu and Maheu (2008) and estimate the best model according to ML and DIC, $\mathcal{M}_5$ under different minimum duration lengths. Hence, besides the minimum 66 days (3 months) we consider the following duration lengths: 44 days (2 months), 88 days (4 months), 110 days (5 months) and 132 days (6 months).

Table 11 reports ML values for $\mathcal{M}_5$ from 1 to 6 change points under the mentioned duration lengths. Overall, we see that the marginal likelihoods are almost identical across different cases, except for duration lengths of 44 and 132 days. First, when we set the minimum duration length to 44 we obtain the exact same ML values as in Table 9 for 1 to 4 change points. For 5 and 6 change points we get lower ML values. However, this does not change the main conclusion.

For 132 days we find evidence in favor of 3 change points at: 04-10-2003, 07-25-2007 and 03-19-2009. Furthermore, the ML values for $\mathcal{M}_5$, 4 to 6 CP differ significantly than those in Table 9. However, this is understandable. The explanations is as follows: in section 6.2 we found evidence of four structural breaks. Furthermore, the last two break dates occur at 09-16-2008 and 03-19-2009 such that there is basically 127 days between these to dates. Hence, when we set the minimum duration length to 132 we are automatically forcing the model to find different break dates than those dates. In fact, for the specification with four change-points we find that the last two breaks occur at 09-05-2008 and at 03-19-2009 whereas for the other duration lengths they occur at 09-16-2008 and 03-19-2009. Therefore, as a consequence we obtain different ML values. Furthermore, this indicates that getting a different estimate of the third change point (09-05-2008 vs 09-16-2008) worsens ML considerably.

---

[1] R_B displays the relative variance of the posterior sample draws when adapting for correlation between iterations, as compared to the variance without accounting for correlation. In these calculations, a bandwidth B of 100 is used. See Kim et al. (1998) for a further background on this measure.

6.4 Prior sensitivity analysis

In this section sensitivity of the results to prior specification is evaluated by investigating alternative priors on the transition probabilities, p_{kk}, keeping prior values of the other parameters the same as in section 4. p_{kk}, $k = 1, ..., m-1$ is one of the key parameters of the model because it controls the duration of each regime in a sequence of S.

In Table 12, we experiment with different hyperparameter values for p_{kk} and report log(ML) for each of these hyperparameters by estimating $\mathcal{M}_5$ from 1 to 6 change points. For instance, the first alternative prior that is considered is $p_{kk} \sim Beta(0.5, 0.5)$ which is relatively flat. With this prior we find evidence of three change points in the data. The change-point dates are given as: 04-10-2003, 07-25-2007 and 03-19-2009. For this prior specification, compared to $\mathcal{M}_5$, 4 CP the Bayes factor in favor of $\mathcal{M}_5$, 3 CP is $\exp(1.35)$ which is very weak evidence. Hence, contrary to the results in Table 9, with $p_{kk} \sim Beta(0.5, 0.5)$ we get uncertain results with regards to the correct number of change points, since we do not have substantial posterior evidence in favor of 3 or 4 change points. For $p_{kk} \sim Beta(10, 2)$ we find strong evidence in favor of four change points and the change-point dates correspond exactly to those in section 6.2.

Finally, for $p_{kk} \sim Beta(20, 0.1)$ which is a relatively tighter prior (a priori the expected duration of each regime is about 201 days) we find that $\mathcal{M}_5$, 5 CP performs best. The first break occurs at 07-26-2002, the second at 04-14-2003 while the remaining break dates correspond exactly to those in section 6.2.

Evidently, there remains some uncertainty regarding the correct number of change points given the hyperparameter values of p_{kk} for very uninformative ($p_{kk} \sim Beta(0.5, 0.5)$) and very tight ($p_{kk} \sim Beta(20, 0.1)$) priors. However, results overwhelmingly suggest existence of structural breaks during the financial crisis of 2008/2009.

6.5 Forecasts

In this section, we compare the out-of-sample performance of $\mathcal{M}_5$ (break) with $\mathcal{M}_0$ (no-break). Specifically, we compare the out-of-sample predictive likelihood (PL) and predictive mean between these two models. Given data up to time $t-1$, $Y_{t-1} = (y_1, ..., y_{t-1})'$, the predictive likelihood, $p(y_t, .., y_T \mid Y_{t-1})$ is the predictive density evaluated at the realized outcome, $y_t, ..., y_T$, $t \leq T$, see Geweke (2005). It contains the out-of-sample prediction record of a particular model, making it the essential quantity of interest for model evaluation. For instance, the predictive likelihood for $\mathcal{M}_5$ is given as

$$p(y_t, .., y_T \mid Y_{t-1}, \mathcal{M}_5) = \int p(y_t, .., y_T \mid P, \theta M, Y_{t-1}, \mathcal{M}_5) \, p(P, \theta M \mid Y_{t-1}, \mathcal{M}_5) \, dP d\theta dM \quad (6.1)$$

Notice that the terms on the right-hand-side of (6.1) have parameter uncertainty integrated out. If $t = 1$ this would be the marginal likelihood and (6.1) changes to (3.5). Hence, the sum of log predictive likelihoods can be interpreted as a measure similar to the log of the marginal likelihood, but ignoring the initial $t - 1$ observations. The predictive likelihood can be used to order models according to their predictive abilities. In a similar fashion to Bayes factors, one can also compare the performance of models based on a specific out-of-sample period by predictive Bayes factors, PBF. The PBF for model A versus B is given as

$$PBF_{AB} \quad = \quad p\left(y_t,..,y_T \mid Y_{t-1}, \mathcal{M}_A\right) / p\left(y_t,..,y_T \mid Y_{t-1}, \mathcal{M}_B\right)$$

and summarizes the relative evidence of the two models over the out-of-sample data, $y_t, ..., y_T$. Calculating the predictive likelihood within a Gibbs sampling scheme is easy as we can use the output from the Gibbs sampler. These 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 $p\left(y_{t+1} \mid Y_t, \mathcal{M}_A\right)$ can then be calculated.

In the context of forecasting with $\mathcal{M}_5$ we perform the following: for the first out-of-sample observation at time t we calculate the marginal likelihood for various number of change points, $(1,...,n)$ using Y_{t-1}. Thereafter, we choose the optimal change point number, n_1 using Bayes factors. We calculate the predictive likelihood, $p\left(y_t \mid Y_{t-1}, \mathcal{M}_5\right)$ and the predictive mean, $E\left[y_t \mid Y_{t-1}, \mathcal{M}_5\right]$ using the parameters associated with specification n_1. Thereafter, we increase the out-of-sample with one observation, calculate marginal likelihoods for $(1,...,n_1+1)$ change points, choose the optimal change point number, n_2, repeat the above forecasting procedure to obtain $p\left(y_{t+1} \mid Y_t, M_2\right)$ and $E\left[y_{t+1} \mid Y_t, \mathcal{M}_5\right]$. We choose the out-of-sample period from January 23rd, 2006 to the end of the sample.

Although we focus on the predictive likelihood to measure predictive content, it is also interesting to consider out-of-sample point forecasts of RV_t based on the predictive mean. Therefore, we also report mean absolute error (MAE) and root mean squared error (RMSE) for the predictive mean. The out-of-sample period corresponds exactly to the period used to calculate the predictive likelihood. Furthermore, in addition to MAE and RMSE, forecasts are also compared using the linear exponential (LINEX) loss function of Zellner (1986). This loss function is defined as $L\left(y_t, \hat{y}_t\right) = b_{LINEX} \left[\exp\left(a_{LINEX}\left(\hat{y}_t - y_t\right)\right) - a_{LINEX}\left(\hat{y}_t - y_t\right) - 1\right]$, where $\hat{y}_t$ is the point forecast. $L\left(y_t, \hat{y}_t\right)$ ranks overprediction (underprediction) more heavily for $a_{LINEX} > 0$ $(a_{LINEX} < 0)$. We use $b_{LINEX} = 1$, with $a_{LINEX} = 1$ and $a_{LINEX} = -1$ in our calculations.

Overall, the break model offers improvements compared to the no-break model. For one observation out-of-sample, $\log\left(PBF\right) = 2.26$, 1 month $\log\left(PBF\right) = 5.14$, 3 months $\log\left(PBF\right) = 59.06$ each in favor of the break specification. The improvements continue till the end of sample, see

Table 12. Finally, Table 12 also displays out-of-sample results for one-day-ahead point forecasts for the no-break and break model. The break model offers improvements in terms of MAE and RMSE compared to the no-break model. When the LINEX loss function is used, the break model also provides gains in terms of point forecasts. However, compared to density forecasts point forecasts show only modest improvements. The difference between the predictive likelihood and the predictive mean is that the predictive likelihood criterion takes into account the whole shape of the density whereas the predictive mean does not.

6.6 Structural breaks and GARCH effects

In this section we model changes in the volatility of realized volatility, see for instance Bollerslev et al. (2007), Corsi et al. (2008) and Liu and Maheu (2008). To investigate if the presence of structural breaks is due to neglected conditional variance dynamics, the following ARFIMA-GARCH model is considered

$$(1-L)^d (y_t - \mu) = \gamma_k \sigma_t \varepsilon_t, \ \varepsilon_t \sim N(0,1) \quad , \quad \sigma_t^2 = \omega + a\varepsilon_{t-1}^2 + b\sigma_{t-1}^2 \tag{6.2}$$

where $\omega > 0$, $a > 0$, $b > 0$ and $a+b < 1$ are also imposed. The parameter γ_k is a scaling constant, which has a direct effect on y_t. In the following $\gamma_1 = 1$ and it is assumed that $\gamma_k > 0$ for $k = 2,...,m$. Thus, in regime 1, this is a standard ARFIMA model with GARCH effects, see Baillie et al. (1996) while in later regimes, the conditional variance of y_t can be larger or smaller than σ_t^2 depending of course on $\gamma_k > 1$ or $\gamma_k < 1$. As noted by Liu and Maheu (2008) the advantage of this specification is that one can model permanent changes in the volatility of realized volatility but avoid the path dependence in σ_t^2 induced by parameter changes in ω, a, and b. Equation (6.2) is estimated using the AR representation of the ARFIMA model. The likelihood is evaluated using the method of Beran (1994). Let $\theta = (\mu, d, \omega, a, b, \gamma_2,...\gamma_m)'$, the Gibbs sampler requires iteration of

- $S^{(i)} \mid P^{(i-1)}, \theta^{(i-1)}, Y_T$

- $\theta^{(i)} \mid P^{(i-1)}, S^{(i)}, Y_T$

- $P^{(i)} \mid \theta^{(i)}, S^{(i)}, Y_T$

We use Metropolis-Hastings to sample each element of θ. For the GARCH parameters, we sample $\psi = (\omega, a, b)'$ all-at-once using the independent chain Metropolis-Hastings algorithm. Specifically, conditional on S, $\gamma_2,...\gamma_m$, d and μ at each iteration of the Gibbs sampler we maximize (6.2) with respect to ψ and set the candidate generating density to $q^* \sim N(\hat{\psi}_{ML}, c \cdot var(\hat{\psi}_{ML}))$ where $c \in \mathbb{R}^+$. The priors of θ are independent Normals with mean 0, variance 100, truncated (except for μ) to satisfy the restrictions on each parameter. Furthermore, we ensure that $a+b < 1$ by resampling

$a^{(i)} > 0$ and $b^{(i)} > 0$ until $a^{(i)} + b^{(i)} < 1$. We follow section 6.2 and estimate a stable ARFIMA-GARCH model as well as its structural break version with 1 to 5 change points. We find strong evidence in favor of structural breaks for (6.2). Specifically, we find that the specification with 3 change points performs best both in terms of ML and DIC. The change-point dates are given as: 11-20-2007, 09-04-2008 and 03-19-2009.

Parameter estimates of the ARFIMA-GARCH specification conditional on 3 change points are listed in Table 10. Compared to $\mathscr{M}_5$, 4 CP we find that d increases (from 0.44 to 0.48). The scaling parameter, γ_k also changes between regimes. Specifically, γ_k rises during the second and third phase which start from late 2007 and last until 03-18-2009. γ_k falls from 1.68 for 09-04-2008/03-18-2009 to 0.93 for 03-19-2009/12-31-2009. Finally, the unconditional volatility of volatility, $\sqrt{\gamma_{S_t}^2 \omega / (1 - a - b)}$ increases during the financial crises (1.91 for the period 09-04-2008 to 03-18-2009 compared to 0.82 for the period 11-20-2007 to 09-03-2008) and falls to 0.58 from the last change point to the end of the sample.

7 Conclusion

We present a Bayesian method for joint analysis of long-memory and structural breaks using change-point ARFIMA models. We estimate different specifications and determine the number of change points using ML and DIC. Monte Carlo simulations demonstrate that our MCMC sampler works very well as the estimated parameters are close to their true values. Furthermore, we find that ML and DIC are powerful in detecting the correct number of structural breaks.

Applying the model to daily S&P 500 data from 2000 to 2009 shows that there is robust evidence of four structural breaks. We demonstrate that accounting for structural breaks improves density and point forecasts. Finally, an ARFIMA model with GARCH effects is also considered. This model provides evidence in favor of three structural breaks.

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] Baillie, R. T., C. F. Chung, and M. A. Tieslau. 1996. "Analysing inflation by the fractionally integrated ARFIMA-GARCH model." *Journal of Applied Econometrics 11*: 23-40.

[4] 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 B 64*: 253-280.

[5] Barndorff-Nielsen, O. E., and N. Shephard. 2002b. "Estimating Quadratic Variation using Realised Variance." *Journal of Applied Econometrics 17*: 457-477.

[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] Beran, J. 1994. *Statistics for Long-Memory Processes*. Chapman and Hall.

[8] Chan, N. H., and W. Palma. 1998. "State space modeling of long-memory processes." *Annals of Statistics 26*: 719-740.

[9] Chan, J., and I. Jeliazkov. 2009. "Efficient Simulation and Integrated Likelihood Estimation in State Space Models." *International Journal of Mathematical Modelling and Numerical Optimisation 1-101*: 101-120.

[10] Chan, J. 2013. "Moving Average Stochastic Volatility Models with Application to Inflation Forecast." *Journal of Econometrics 176*(2): 162-172.

[11] Chib, S., and E. Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *The American Statistician 49(4)*: 327-335.

[12] Chib, S. 1996. "Calculating posterior distributions and model estimates in Markov mixture models." *Journal of Econometrics 75*: 79-97.

[13] Chib, S. 1998. "Estimation and Comparison of Multiple Change-Point Models." *Journal of Econometrics 86*: 221-241.

[14] Corsi, F., U. Kretschmer, S. Mittnik, and C. Pigorsch. 2008. "The Volatility of Realized Volatility." *Econometric Reviews 27*: 46-78.

[15] Gelfand, A., and D. Dey. 1994. "Bayesian Model Choice: Asymptotics and Exact Calculations." *Journal of the Royal Statistical Society 56*: 501-514.

[16] Geweke, J. 2005. *Contemporary Bayesian Econometrics and Statistics.* Wiley.

[17] Hamilton, J. D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica 57(2)*: 357-384.

[18] Hansen, P. R., and A. Lunde. 2006. "Realized Variance and Market Microstructure Noise." *Journal of Business and Economic Statistics 24(2)*: 127-161.

[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] Kim, J. C., C. Morley, and C. R. Nelson. 2005. "The Structural Break in the Equity Premium." *Journal of Business and Economic Statistics 23(2)*: 181-191.

[22] Koop, G. 2003. *Bayesian Econometrics*. John Wiley & Sons Ltd.

[23] Liu, C., and J. Maheu. 2008. "Are There Structural Breaks in Realized Volatility?" *Journal of Financial Econometrics 6*: 326-360.

[24] Raggi, D., and S. Bordignon. 2012. "Long Memory and Nonlinearities in Realized Volatility: A Markov Switching Approach." *Computational Statistics and Data Analysis 56*: 3730-3742.

[25] Sims, C. A., D. F. Waggoner, and T. Zha. 2008. "Methods for inference in large multiple-equation Markov-switching models." *Journal of Econometrics 146(2)*: 255-274.

[26] Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde. 2002. "Bayesian Measures of Model Complexity and Fit (with comments)". *Journal of the Royal Statistical Society B 64*: 583-639.

[27] Zellner, A. 1986. "Bayesian estimation and prediction using asymmetric loss functions." *Journal of the American Statistical Association 81*: 446-451.

A Appendix

This appendix details a direct approach for evaluating the likelihood function of the ARFIMA model using the precision-based algorithm of Chan and Jeliazkov (2009) and Chan (2013). Consider the following ARFIMA model

$$y_t = \mu + (1-L)^{-d}\varepsilon_t, \quad \varepsilon_t \sim N\left(0,\sigma^2\right) \tag{A.1}$$

Conditional on M we write (A.1) as $Y_T = u + H\varepsilon$ where u, H and ε follow directly from the main text. The special structure of H can be exploited to speed up computation. For instance, obtaining the Cholesky decomposition of a banded $T \times T$ matrix with fixed bandwidth involves only $O(T)$ operations as opposed to $O\left(T^3\right)$ for a full matrix of the same size. Similar computational savings can be generated in operations such as multiplication, forward and backward substitution by using block-banded or sparse matrix algorithms. These banded and sparse matrix algorithms are implemented in Matlab.

It follows from (A.1) that $p(Y_T \mid \theta, M) \sim N(u, \Omega_{Y_T})$ where $\Omega_{Y_T} = HS_{Y_T}H'$ and $S_{Y_T} = \sigma^2 I_T$. Since S_{Y_T} is a diagonal matrix and H is a lower triangular sparse matrix, the product, Ω_{Y_T} is sparse. Moreover, since $|H| = 1$ for any $\pi_1, ..., \pi_M$ one has that $|\Omega_{Y_T}| = |S_{Y_T}|$. The joint density of Y_T is therefore given by

$$\log p(Y_T \mid \theta, M) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\left(\sigma^2\right) - \frac{1}{2}(Y_T - u)'\Omega_{Y_T}^{-1}(Y_T - u) \tag{A.2}$$

As stated in Chan (2013) it is important to note that we need not obtain the $T \times T$ inverse matrix $\Omega_{Y_T}^{-1}$ in order to evaluate (A.2) which would involve $O\left(T^3\right)$ operations. Instead, it can be computed in three steps, each of which requires only $O(T)$ operations. Therefore, the following notation is introduced, see also Chan (2013): given a lower (upper) triangular $T \times T$ non-singular matrix A and a $T \times 1$ vector c, let $A \backslash c$ denote the unique solution to the triangular system $Ax = c$ obtained by forward (backward) substitution, i.e. $A \backslash c = A^{-1}c$.

Now, obtain the Cholesky decomposition, C_{Y_T} of Ω_{Y_T} such that $C_{Y_T}C'_{Y_T} = \Omega_{Y_T}$. This involves only $O(T)$ operations. Compute $x_1 = C'_{Y_T} \backslash (C_{Y_T} \backslash (Y_T - u))$ by forward followed by backward substitution, each of which require $O(T)$ operations since C_{Y_T} is also banded. Then, by definition

$$x_1 = C_{Y_T}^{-1'}\left(C_{Y_T}^{-1}(Y_T - u)\right) = \left(C_{Y_T}C'_{Y_T}\right)^{-1}(Y_T - u) = \Omega_{Y_T}^{-1}(Y_T - u)$$

Finally, compute $x_2 = -\frac{1}{2}(Y_T - u)'x_1 = -\frac{1}{2}(Y_T - u)'\Omega_{Y_T}^{-1}(Y_T - u)$ which gives the quadratic term in (A.2).

24

Table 1: Change-point model specifications

Model index	parameters that change from a break
$\mathcal{M}_0$	None
$\mathcal{M}_1$	μ
$\mathcal{M}_2$	d
$\mathcal{M}_3$	σ^2
$\mathcal{M}_4$	μ, d
$\mathcal{M}_5$	μ, σ^2
$\mathcal{M}_6$	All parameters

This table labels the various change-point specifications. The first column is the model index and the second column lists model parameters that change due to structural breaks.

Table 2: Parameter values for Monte Carlo simulations

Regime		$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$
1		1	1	1	1	1	1	1
2	μ	1	2	1	1	2	2	2
3		1	0.8	1	1	0.8	0.8	0.8
1		0.3	0.3	0.20	0.3	0.3	0.3	0.2
2	d	0.3	0.3	0.45	0.3	0.45	0.3	0.45
3		0.3	0.3	0.05	0.3	0.05	0.3	0.05
1		0.1	0.1	0.1	0.1	0.1	0.1	0.1
2	σ^2	0.1	0.1	0.1	0.18	0.1	0.18	0.18
3		0.1	0.1	0.1	0.36	0.1	0.36	0.36

This table lists the parameter values for the Monte Carlo simulations. The first column is the index of the regimes. The second column lists the parameters. The first row is the model index. If there is one break then the DGP parameters are first from regime 1 and then regime 2.

Table 3: Change-point identification by marginal likelihood

DGP		# of CP	0 CP	1 CP	2 CP	3 CP
			frequency by ML			
$\mathcal{M}_1$ μ		0	**100**	0	0	0
		1	0	**99**	1	0
		2	0	0	**100**	0
$\mathcal{M}_2$ d		0	**87**	0	0	13
		1	1	**86**	5	8
		2	0	0	**95**	5
$\mathcal{M}_3$ σ^2		0	**100**	0	0	0
		1	0	**100**	0	0
		2	0	1	**99**	0
$\mathcal{M}_4$ μ d		0	**100**	0	0	0
		1	0	**97**	3	0
		2	0	0	**98**	2
$\mathcal{M}_5$ μ σ^2		0	**100**	0	0	0
		1	1	**99**	0	0
		2	0	2	**98**	0
$\mathcal{M}_6$ All parameters		0	**100**	0	0	0
		1	0	**100**	0	0
		2	0	0	**100**	0

The first column lists the true model along with the parameters that change due to a structural break. CP, change points; ML, marginal likelihood. The "0 CP" displays the number of times in the repetitions when the specification with no change points has the highest ML ect. Each row sums to 100. In this table $\alpha = 0.99$.

Table 4: Model comparison using ML

# of CP	measure	0 CP	1 CP	2 CP	3 CP
		Frequency by ML for different values of α			
		DGP, $\mathcal{M}_1$			
	$\alpha = 0.75$	**100**	0	0	0
0	$\alpha = 0.95$	**100**	0	0	0
	$\alpha = 0.99$	**100**	0	0	0
	$\alpha = 0.75$	0	**99**	1	0
1	$\alpha = 0.95$	0	**99**	1	0
	$\alpha = 0.99$	0	**99**	1	0
	$\alpha = 0.75$	0	0	**100**	0
2	$\alpha = 0.95$	0	0	**100**	0
	$\alpha = 0.99$	0	0	**100**	0
		DGP, $\mathcal{M}_5$			
	$\alpha = 0.75$	**100**	0	0	0
0	$\alpha = 0.95$	**100**	0	0	0
	$\alpha = 0.99$	**100**	0	0	0
	$\alpha = 0.75$	1	**99**	0	0
1	$\alpha = 0.95$	1	**99**	0	0
	$\alpha = 0.99$	1	**99**	0	0
	$\alpha = 0.75$	0	2	**98**	0
2	$\alpha = 0.95$	0	2	**98**	0
	$\alpha = 0.99$	0	2	**98**	0
		DGP, $\mathcal{M}_6$			
	$\alpha = 0.75$	**100**	0	0	0
0	$\alpha = 0.95$	**100**	0	0	0
	$\alpha = 0.99$	**100**	0	0	0
	$\alpha = 0.75$	0	**100**	0	0
1	$\alpha = 0.95$	0	**100**	0	0
	$\alpha = 0.99$	0	**100**	0	0
	$\alpha = 0.75$	0	0	**100**	0
2	$\alpha = 0.95$	0	0	**100**	0
	$\alpha = 0.99$	0	0	**100**	0

The evidence for the number of change points is determined according ML. "0 CP" column reports the number of times in the 100 repetitions when the specification with no change point has the best performance. Similarly for the other columns.

Table 5: Monte Carlo parameter estimates

Regime	parameter	true	mean	median	mode	RMSE
		DGP, $\mathscr{M}_4$, $T = 500$				
1		1	0.9880	0.9877	0.9812	0.0771
2	μ	2	1.9883	1.9883	1.9455	0.1393
3		0.8	0.7985	0.7985	0.7941	0.0269
1		0.3	0.2798	0.2782	0.2511	0.0783
2	d	0.45	0.4223	0.4258	0.4117	0.0519
3		0.05	0.0854	0.0780	0.0752	0.1194
1-3	σ^2	0.1	0.0953	0.0951	0.0762	0.0074
		DGP, $\mathscr{M}_4$, $T = 1000$				
1		1	0.9967	0.9968	0.9878	0.0575
2	μ	2	2.0012	2.0014	1.9961	0.1164
3		0.8	0.8018	0.8019	0.8014	0.0245
1		0.3	0.2983	0.2974	0.2846	0.0458
2	d	0.45	0.4430	0.4453	0.4437	0.0267
3		0.05	0.0686	0.0643	0.0705	0.0336
1-3	σ^2	0.1	0.0988	0.0986	0.0847	0.0047
		DGP, $\mathscr{M}_4$, $T = 2000$				
1		1	1.0033	1.0040	0.9956	0.0407
2	μ	2	1.9896	1.9900	1.9929	0.0735
3		0.8	0.7979	0.7978	0.7957	0.0101
1		0.3	0.3089	0.3082	0.2983	0.0264
2	d	0.45	0.4479	0.4490	0.4446	0.0239
3		0.05	0.0638	0.0621	0.0551	0.0241
1-3	σ^2	0.1	0.0991	0.0993	0.0898	0.0030

This table reports the true value of the DGP parameters along with the mean, median, mode and root-mean-squared error (RMSE) of the estimated parameters for the Monte Carlo simulations, generated from $\mathscr{M}_4$, 2 CP with parameters as indicated using samples of $T = 500$, $T = 1000$ and $T = 2000$.

Table 6: Model comparison using different criteria

| # of CP | measures | Frequency by DIC and ML | | | |
		0 CP	1 CP	2 CP	3 CP
		DGP, $\mathcal{M}_1$			
0	DIC	**100**	0	0	0
	ML	**100**	0	0	0
1	DIC	0	**98**	0	2
	ML	0	**99**	1	0
2	DIC	0	0	**98**	2
	ML	0	0	**100**	0
		DGP, $\mathcal{M}_3$			
0	DIC	**100**	0	0	0
	ML	**100**	0	0	0
1	DIC	0	**93**	0	7
	ML	0	**100**	0	0
2	DIC	0	0	**96**	4
	ML	0	1	**99**	0
		DGP, $\mathcal{M}_6$			
0	DIC	**84**	12	2	2
	ML	**100**	0	0	0
1	DIC	0	**96**	2	2
	ML	0	**100**	0	0
2	DIC	0	0	**97**	3
	ML	0	0	**100**	0

The evidence for the number of change points is determined according to DIC and ML. "0 CP" column reports the number of times in the 100 repetitions when the specification with no change point has the best performance (lowest DIC and highest ML). Similarly for the other columns.

Table 7: Dating the change points

DGP	# CP	DIFF 1st CP	2nd CP	3rd CP	4th CP
$\mathcal{M}_1$ μ	4	0.0053	0.0105	0	0
$\mathcal{M}_3$ σ^2	4	2.2000	4.9375	4.3500	1.3250
$\mathcal{M}_4$ μ d	4	0.3913	0.2391	0	0
$\mathcal{M}_6$ All parameters	4	0.0822	0.6027	0.1644	0

This table reports simulation results for $\mathcal{M}_1$, $\mathcal{M}_3$, $\mathcal{M}_4$ and $\mathcal{M}_6$ under 4 change points. Specifically, we report the average difference (for each change point) between the estimated change point and true change point date. The definition of DIFF is given in the text.

Table 8: Effect of sample size on identification of change points

Sample size	Frequency by ML 0 CP	1 CP	2 CP	3 CP
DGP, $\mathcal{M}_2$, 1 CP				
500	9	**76**	6	9
1000	1	**86**	5	8
2000	0	**95**	3	2
DGP, $\mathcal{M}_2$, 2 CP				
500	0	0	**54**	46
1000	0	0	**95**	5
2000	0	0	**98**	2
DGP, $\mathcal{M}_4$, 2 CP				
500	0	0	**83**	17
1000	0	0	**98**	2
2000	0	0	**98**	2

This table reports simulation results for $\mathcal{M}_2$ and $\mathcal{M}_4$ under different specifications. The "0 CP" column records the number of times in the 100 repetitions when the model with no change point has the largest marginal likelihood. The "1 CP" column records the number of times in the 100 repetitions when the specification with one change point has the largest marginal likelihood, etc.

Table 9: Model comparison by marginal likelihood and DIC for S&P 500 volatility

# CP	$\mathcal{M}_1$ μ	$\mathcal{M}_2$ d	$\mathcal{M}_3$ σ^2	$\mathcal{M}_4$ μ, d	$\mathcal{M}_5$ μ, σ^2	$\mathcal{M}_6$ All parameters
0	-1946.84 (3860.65)	-1946.84 (3860.65)	-1946.84 (3860.65)	-1946.84 (3860.65)	-1946.84 (3860.65)	-1946.84 (3860.65)
1	-1942.64 (3852.58)	**-1942.42 (3851.81)**	-1918.87 (3778.32)	-1951.32 (3854.58)	-1927.45 (3803.89)	-1929.60 (3801.83)
2	-1938.71 (3845.17)	-1946.55 (3859.81)	-1549.75 (2777.14)	-1941.20 (3836.48)	-1378.12 (2705.06)	-1531.20 (2701.23)
3	-1933.58 (3814.83)	-1945.08 (3854.55)	-1504.93 (2728.29)	-1949.16 (3816.87)	-1283.65 (2514.36)	-1379.35 (2511.32)
4	**-1915.71 (3781.36)**	-1945.45 (3853.45)	**-1404.15 (2562.78)**	**-1921.52 (3764.23)**	**-1257.10 (2360.45)**	**-1290.64 (2378.46)**
5	-1933.43 (3819.57)	-1945.58 (3855.29)	-1670.04 (2859.83)	-1937.13 (3774.48)	-1259.75 (2368.46)	-1999.48 (3898.11)
6	-1940.84 (3833.59)	-1947.59 (3858.50)	-1682.08 (2983.70)	-1941.23 (3801.87)	-1737.02 (2653.20)	-2053.65 (4109.60)

This table reports the log marginal likelihood and DIC (indicated in the parentheses) of different ARFIMA models. The first row lists the index of the models. The second row lists the parameters that are subject to structural breaks within each specification. The first column is the number of the change points (CP) that are conditioned on. Bold numbers indicate the highest marginal likelihood and lowest DIC within each specification.

Table 10: Parameter estimates for S&P 500 volatility

Parameter	$\mathcal{M}_5$, 4 CP					ARFIMA-GARCH, 3 CP				
	mean	95% credibility	R_B	Geweke	M-H ratio	mean	95% credibility	R_B	Geweke	M-H ratio
d	0.4445	[0.4181,0.4722]	6.86	-0.50	0.35	0.4806	[0.4612,0.4897]	7.95	-0.19	0.32
μ_1	0.1890	[0.1685,0.2089]	4.20	-1.22	0.37	0.1128	[0.0669,0.1590]	7.76	-0.61	0.39
μ_2	0.0945	[0.0876,0.1015]	10.11	-0.15	0.34					
μ_3	0.1554	[0.1266,0.1836]	6.09	-0.61	0.41					
μ_4	0.3922	[0.3270,0.4589]	9.82	1.11	0.42					
μ_5	0.1702	[0.1517,0.1890]	4.45	-2.19	0.34					
σ_1^2	0.3238	[0.2944,0.3572]	1.03	0.76						
σ_2^2	0.0501	[0.0450,0.0549]	1.76	0.58						
σ_3^2	0.1806	[0.1515,0.2152]	0.88	1.41						
σ_4^2	0.6809	[0.5345,0.8755]	1.02	0.01						
σ_5^2	0.0822	[0.0676,0.0999]	1.38	-0.52						
M	31	[24,38]	15.41	-1.85	0.28					
γ_2						1.1014	[0.8552,1.3937]	7.04	0.01	0.39
γ_3						1.6845	[1.2693,2.2280]	7.04	-0.01	0.37
γ_4						0.9303	[0.7590,1.1251]	4.53	-0.21	0.33
ω						0.0029	[0.0023,0.0036]	3.71	-1.04	0.42
a						0.1644	[0.1449,0.1833]	4.95	-1.95	0.42
b						0.8313	[0.8124,0.8505]	4.87	1.47	0.42
DIC	2360.45					2375.11				
log(ML)	-1257.10					-1261.83				

This table reports the posterior mean (mean), 95% credibility intervals, inefficiency factors, R_B, Geweke's convergence statistics (Geweke), M-H acceptance ratios, DIC and log(ML) using $\alpha = 0.99$ for $\mathcal{M}_5$, 4 CP and ARFIMA-GARCH, 3 CP. The parameters associated with each regime are labeled with subscript $1, \dots, m$.

Table 11: Robustness to minimum regime duration lengths

# CP	44 days (2 months)	66 days (3 months)	88 days (4 months)	110 days (5 months)	132 days (6 months)
1	-1927.45	-1927.45	-1927.45	-1927.45	-1927.45
2	-1378.12	-1378.12	-1380.89	-1378.12	-1378.12
3	-1283.65	-1283.65	-1283.65	-1283.65	**-1283.65**
4	**-1257.10**	**-1257.10**	**-1257.10**	**-1257.10**	-1662.56
5	-2033.78	-1259.75	-1259.75	-1259.75	-1542.71
6	-2036.12	-1737.02	-1737.02	-1737.02	-1702.11

This table compares log(ML) values of different minimum regime durations for $\mathcal{M}_5$. Each column reports the lower bound for the number of observations in each regime.

Table 12: Prior sensitivity analysis, $\mathcal{M}_5$

| # CP	$p_{kk} \sim Beta(0.5,0.5)$	$p_{kk} \sim Beta(8,0.1)$	$p_{kk} \sim Beta(10,2)$	$p_{kk} \sim Beta(20,0.1)$
1	-1928.12	-1927.45	-1935.21	-1927.35
2	-1384.52	-1378.12	-1393.53	-1379.41
3	**-1290.85**	-1283.65	-1669.50	-1616.89
4	-1292.20	**-1257.10**	**-1289.23**	-1444.74
5	-1869.91	-1259.75	-1869.29	**-1237.21**
6	-1905.09	-1737.02	-1888.65	-1604.79

This table compares log(ML) values of different prior hyperparameter values for p_{kk}. The priors of the other parameters are fixed according to section 4.

Table 13: Out-of-sample forecasts for S&P 500 volatility

Model	MAE	RMSE	LINEX $a_{LINEX}=1, b_{LINEX}=1$	LINEX $a_{LINEX}=-1, b_{LINEX}=1$	log(PL)
No-break	0.3553	0.6241	0.2582	3.2456	-923.73
Break	0.3403	0.5993	0.2475	2.5680	-544.48

This table reports mean absolute error (MAE), root mean squared error (RMSE) and average LINEX for the forecasts based on the predictive mean for one-day-ahead. Furthermore, the one-day-ahead log predictive likelihood, log(PL) is also reported.

Figure 1: Change-point dates for $\mathcal{M}_5$, 4 CP
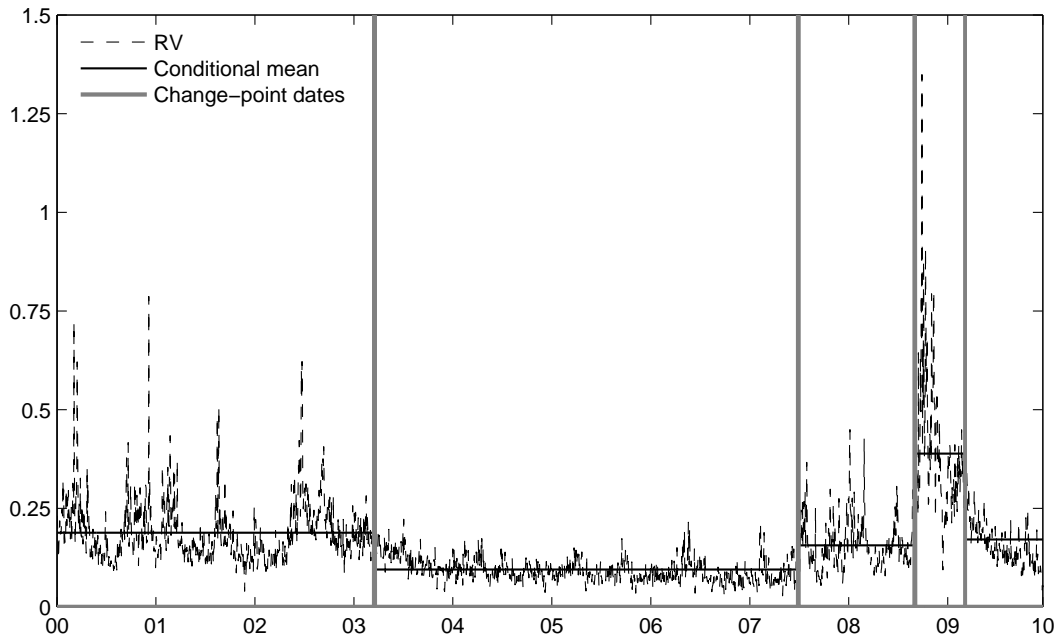
S&P 500 annualized volatility, conditional mean and estimated change-point dates indicated as vertical lines for $\mathcal{M}_5$, 4 CP.

Research Papers
2013

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
2013-25:	Nima Nonejad: Time-Consistency Problem and the Behavior of US Inflation from 1970 to 2008
2013-26:	Nima Nonejad: Long Memory and Structural Breaks in Realized Volatility: An Irreversible Markov Switching Approach