University of Toronto
Department of Economics

Working Paper 304

Are there Structural Breaks in Realized Volatility?

By Chun Liu and John M Maheu

December 18, 2007

Are there Structural Breaks in Realized Volatility?

Chun Liu

School of Economics and Management

Tsinghua University

John M. Maheu[*]

Dept. of Economics

University of Toronto

November 2007

Abstract

Constructed from high-frequency data, realized volatility (RV) provides an efficient estimate of the unobserved volatility of financial markets. This paper uses a Bayesian approach to investigate the evidence for structural breaks in reduced form time-series models of RV. We focus on the popular heterogeneous autoregressive (HAR) models of the logarithm of realized volatility. Using Monte Carlo simulations we demonstrate that our estimation approach is effective in identifying and dating structural breaks. Applied to daily S&P 500 data from 1993-2004, we find strong evidence of a structural break in early 1997. The main effect of the break is a reduction in the variance of log-volatility. The evidence of a break is robust to different models including a GARCH specification for the conditional variance of $\log(RV)$.

key words: realized volatility, change point, marginal likelihood, Gibbs sampling, GARCH

JEL: C22, C11, G10

1 Introduction

The econometric modeling of volatility is an important issue in empirical finance. Instability of the volatility process has important implications for decisions in risk management,

1

portfolio choice and derivative pricing. Many papers have studied the effect of neglected parameter changes on the level and persistence of volatility[1]. Another strand of the literature has focused on testing for structural breaks[2], or directly modeling parameter change.[3] A common feature of this research is the use of stock returns to infer structural changes in the latent volatility process.

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 ex post volatility. RV is constructed from the sum of intraday squared returns, and converges to quadratic variation which is the sum of integrated volatility plus a jump component for a broad class of continuous time models. This is an essentially nonparametric estimate of ex post volatility. Barndorff-Nielsen and Shephard (2004) show how the continuous component can be separated from the jump component of volatility. Empirical and theoretical features of RV are discussed by Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold and Labys (2001), Andersen, Bollerslev, Diebold and Ebens (2001), Barndorff-Nielsen and Shephard (2002a,2002b), and Meddahi (2002). The purpose of this paper is to investigate the evidence for structural breaks within the context of popular time-series models of RV. The use of realized volatility, in contrast to daily returns, will provide us with more power to detect structural changes.

An important feature of the time-series of RV is the strong serial dependence.[4] Corsi (2004) shows that the heterogeneous autoregressive (HAR) model can capture the strong persistence in the data with a parsimonious linear structure. This is not a true long-memory model but it does provide a good approximation to the dynamics of long-memory. These features make this specification popular and easy to estimate. HAR type parameterizations appear in recent research including Andersen, Bollerslev and Diebold (2006), Andersen, Bollerslev and Huang (2007), Bollerslev et al. (2007), Corsi et al. (2005), Forsberg and Ghysels (2006), and Maheu and McCurdy (2006).

This paper provides a Bayesian analysis of structural breaks in daily S&P 500 realized

---

[1]Examples include Chu (1995), Hillebrand (2005), Hwang and Pereira (2004), Lamoureux and Lastrapes (1990), Mikosch and Starica (2004), and Starica and Granger (2005) among others.

[2]Andreou and Ghysels (2002), Pastor and Stambaugh (2001), Rapach and Strauss (2006), Ray and Tsay (2002), and Smith (2006).

[3]For instance, Engle and Rangel (2005).

[4]Different models that account for this include: autoregressive fractional integrated moving average (ARFIMA) models by Andersen, Bollerslev, Diebold and Labys (2003), Giot and Laurent (2004), and Martins, van Dijk and de Pooter (2004); Markov switching in Maheu and McCurdy (2002); unobserved component models, Barndorff-Nielsen and Shephard (2002a) and Koopman, Jungbacker and Hol (2005); the mixed data sampling approach of Ghysels et al. (2003); and the heterogeneous autoregressive model of Corsi (2004).

volatility. We focus on a logarithmic HAR specification similar to Andersen et al. (2006), and an extension to include GARCH effects. We test for multiple structural breaks using the change-point (or structural break) model of Chib (1998) which is estimated with efficient Markov chain Monte Carlo sampling methods. We investigate specifications which allow for all parameters, as well as only a subset of parameters, to change due to a structural break. This allows us to isolate the impact of a break on individual model parameters and to use all data in the estimation of parameters not affected by breaks. Each model is estimated conditional on $0, 1, 2, ..., k_{\max}$ breaks occurring. For each of these we calculate the marginal likelihood, and the evidence for the number of breaks can be compared using Bayes factors.[5]

A first step is to investigate the power of the Chib change-point model to detect the correct number of structural breaks. The simulations, based on empirically realistic models, show the method to perform well in estimating the true number of breaks when there are 1 or more breaks, as well as when there are no breaks. The larger the number of parameters that are affected by a break, the more likely the structural breaks are correctly identified. In addition, it is relatively more difficult to identify several structural breaks, particularly when only one model parameter changes.

We find the marginal likelihood has an advantage over traditional model comparison criteria, such as sum of squared errors, in identifying the true number of structural breaks. For example, in-sample loss functions of one period ahead forecasts of RV are less sensitive to structural breaks. In fact, the R^2 from commonly used Mincer and Zarnowitz (1969) forecast regressions strongly favor models with more change-points than there are. Forecasting loss functions are less powerful in detecting breaks and can be a misleading indicator of structural breaks.

The empirical results for S&P 500 provide strong evidence of a structural break in the logarithm of RV (log-RV) during February 1997 based on data from 1993-2004. The effect of the structural break is mainly confined to the variance with weaker evidence that the regression parameters are effected. The structural break results in a smaller variance after 1997. The conclusions are robust to the inclusion/exclusion of jumps, and asymmetric terms in the model.

Concurrent work by Andersen et al. (2006), Andersen et al. (2007), Corsi et al. (2005), and Bollerslev et al. (2007) using a similar model for S&P 500 volatility ignore the possibility of structural breaks. We demonstrate that accounting for the structural break improves out-of-sample density forecasts of log-volatility dramatically after the

---

[5]Related applications of Bayesian change-point analysis include Kim et al. (2005), Martin (2000), Pastor and Stambaugh (2001), Pettenuzzo and Timmermann (2006), and Ray and Tsay (2002).

break point. Corsi et al. (2005), Andersen et al. (2007) and Bollerslev et al. (2007) use a HAR model for volatility and allow for GARCH effects in the conditional variance. To investigate if our finding of a break is spurious due to neglected conditional variance dynamics, we consider breaks in a HAR-GARCH model for log-RV. Again the evidence is strong for a break in the long-run variance from the GARCH model and the estimated break point is identical to that of the homoskedastic model. Moreover, ignoring this break results in persistence estimates in the GARCH model that are too high. In summary, we find a permanent reduction in the variance of log-RV in February 1997 for all of our specifications. Failure to model this break results in biased parameter estimates and poor density forecasts.

The structure of this paper is as follows. In Section 2 we review the construction of RV and the jump component, and the empirical model for daily log-RV. The change-point model and Bayesian estimation are detailed in Section 3. Section 4 presents simulation results on the ability of the change-point model to detect breaks, date change points and other features. Section 5 is the application to S&P 500 volatility while Section 6 concludes. An appendix provides details on the estimation of the marginal likelihood and the HAR-GARCH model.

2   Realized Volatility

2.1   Estimation

We assume that the price process belongs to the class of special semimartingales, which is a very broad class of processes including Ito and jump processes. In this environment Andersen, Bollerslev, Diebold and Labys (2001) show that the quadratic variation of the process, which is defined as integrated volatility plus the jump component, provides a natural measure of ex post volatility. RV is constructed from the sum of intraday squared returns and is a consistent estimate of quadratic variation as the intraday sampling frequency goes to infinity. In contrast to traditional measures of volatility, such as daily squared returns, realized volatility is more efficient. Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002b) and Meddahi (2002) discuss the precision of RV.

Market microstructure dynamics contaminate the price process with noise. The noise can be time dependent and may be correlated with the efficient price (Hansen and Lunde (2006)). RV can be a biased and inconsistent estimator. For instance see Bandi and Russell (2006), Hansen and Lunde (2006), Oomen (2005) and Zhang et al. (2005) for more details on the effects of market microstructure noise on volatility estimation. To

reduce the effect of market microstructure noise, we employ a kernel-based estimator that utilizes autocovariances of intraday returns. Specifically, we follow Hansen and Lunde (2006) to provide a bias correction to realized volatility as follows,[6]

$$RV_t = \sum_{i=1}^{M} r_{t,i}^2 + 2 \sum_{h=1}^{q} \left(1 - \frac{h}{q+1}\right) \sum_{i=1}^{M-h} r_{t,i} r_{t,i+h} \tag{1}$$

where $r_{t,i}$ is the ith logarithmic return during day t, M is the number of returns in each day, and $q = 1$ in our calculations. This estimator is guaranteed to be non-negative and is almost identical to the subsample-based estimator of Zhang et al. (2005).

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

$$RBP_t = \mu_1^{-2} \sum_{i=1}^{M-1} |r_{t,i}||r_{t,i+1}|, \tag{2}$$

where $\mu_1 = \sqrt{2/\pi}$. Asymptotically as $M \to \infty$, realized bipower variation converges to integrated volatility. The difference between RV_t and RBP_t is an estimate of the daily jump component. In our paper, we follow Andersen et al. (2006) and estimate jumps by,

$$J_t = \begin{cases} \log\left(RV_t - RBP_t + 1\right) & \text{when } RV_t - RBP_t > 0 \\ 0 & \text{Otherwise} \end{cases} \tag{3}$$

where we add 1 to ensure that $J_t \geq 0$.

2.2   Models

Consider the heterogeneous autoregressive model (HAR) of realized volatility proposed by Corsi (2004). Corsi shows that this model can capture many of the features of volatility including long memory. Empirically the distribution of $\log(RV_t)$ is approximately bell shaped[7], so we consider a logarithmic version of the HAR similar to that implemented by Andersen et al. (2006) and defined as

$$v_t = \beta_0 + \beta_1 v_{t-1} + \beta_2 v_{t-5,t-1} + \beta_3 v_{t-22,t-1} + \epsilon_t, \quad \epsilon_t \sim NID(0,\sigma^2). \tag{4}$$

---

[6]The adjusted RV_t from Hansen and Lunde produced several days with a negative estimate for RV_t. We follow their suggestion and use Bartlett weights to ensure that RV_t is positive. The estimator is no longer unbiased, but may have a smaller MSE.

[7]See Andersen, Bollerslev, Diebold and Labys (2001), Andersen, Bollerslev, Diebold and Ebens (2001), and Giot and Laurent (2006).

where $v_t = \log(RV_t)$ and

$$v_{t-5,t-1} = \frac{\log(RV_{t-1}) + \log(RV_{t-2}) + \cdots + \log(RV_{t-5})}{5} \tag{5}$$

$$v_{t-22,t-1} = \frac{\log(RV_{t-1}) + \log(RV_{t-2}) + \cdots + \log(RV_{t-22})}{22} \tag{6}$$

This model postulates three factors that affect volatility: daily log-volatility v_{t-1}, weekly log-volatility $v_{t-5,t-1}$ and monthly log-volatility $v_{t-22,t-1}$.

Andersen et al. (2006), Huang and Tauchen (2005), and Tauchen and Zhou (2005) conclude that jumps are an important component of realized volatility. Therefore, we consider a model with a jump component as well as asymmetric terms motivated by the models in Bollerslev et al. (2007),

$$v_t = \beta_0 + \beta_1 v_{t-1} + \beta_2 v_{t-5,t-1} + \beta_3 v_{t-22,t-1} + \beta_J J_{t-1} + \beta_{A1} \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} + \beta_{A2} \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} I(r_{t-1} < 0) + \epsilon_t \tag{7}$$

where r_t is the daily return, $I(r_{t-1} < 0)$ is an indicator function, which equals 1 when $r_{t-1} < 0$ and equals 0 otherwise. J_{t-1} is the jump component defined above. The last two terms allow for asymmetric effects from positive and negative returns similar to the EGARCH model of Nelson (1991).

To simplify the notation, we note that the model can be cast into a standard regression form

$$y_t = X_t \beta + \epsilon_t$$

where $y_t = v_t$, $\beta = [\beta_0 \ \beta_1 \ \beta_2 \ \beta_3 \ \beta_J \ \beta_{A1} \ \beta_{A2}]'$ and

$$X_t = \left[1 \ v_{t-1} \ v_{t-5,t-1} \ v_{t-22,t-1} \ J_{t-1} \ \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} \ \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} I(r_{t-1} < 0) \right].$$

In the following denote $Y_t = \{v_1, v_2, \ldots, v_t\}$, and $I_t = \{y_1, X_1, \ldots, y_t, X_t\}$.

3   Change-point Model

We consider the change-point (or structural break) model proposed by Chib (1998). This uses a hidden Markov model with a restricted transition matrix to model change-points. A test for the number of breaks is then a test of the dimension of the hidden Markov chain. As Chib shows there are efficient posterior simulation methods available for this type of model. Model parameters and change points are jointly estimated conditional on a fixed number of change points. Bayes factors can be used to compare the evidence for

the number of breaks.

In the following we review a Gibbs sampling approach for the unrestricted model in which all parameters are subject to change after a break. This is followed by a discussion of the restricted model in which some parameters are restricted to be constant across breaks. This latter specification allows us to isolate which features of the model are most likely affected by a structural break.

Assume there are $m - 1$, $m \in \{1, 2, ...\}$ change points at unknown times $\Omega_m = \{\tau_1, \tau_2, \ldots, \tau_{m-1}\}$. Separated by those change points, we have m different regimes. The density of observation y_t, $t = 1, 2, \cdots, T$, depends on the parameter $\theta_k = \{\beta_k, \sigma_k^2\}$, $k = 1, ..., m$, whose value changes at the change points Ω_m and remains constant otherwise

$$
\theta_t = \begin{cases}
\theta_1 & \text{if} & t < \tau_1 \\
\theta_2 & \text{if} & \tau_1 \leq t < \tau_2 \\
\vdots & \vdots & \vdots \\
\theta_{m-1} & \text{if} & \tau_{m-2} \leq t < \tau_{m-1} \\
\theta_m & \text{if} & \tau_{m-1} \leq t.
\end{cases}
\tag{8}
$$

Denote the state of the system at each time by $S = \{s_1, s_2, \ldots, s_T\}$ where $s_t = k$ indicates that the observation y_t is from regime k and follows the conditional distribution $f(y_t | I_{t-1}, \theta_k)$. The one-step ahead transition probability matrix for s_t is assumed to be

$$
P = \begin{pmatrix}
p_{11} & p_{12} & 0 & \cdots & & 0 \\
0 & p_{22} & p_{23} & \cdots & & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots \\
\vdots & \vdots & 0 & p_{m-1,m-1} & p_{m-1,m} \\
0 & 0 & \cdots & & 0 & 1
\end{pmatrix}
\tag{9}
$$

where $p_{ij} = \Pr(s_t = j | s_{t-1} = i)$ with $j = i$ or $i + 1$, and this is the probability of moving from regime i at time $t - 1$ to regime j at time t. The transition matrix 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 $s_t = k$, we have $s_{t+1} = k$ or $s_{t+1} = k + 1$ and $p_{k,k} + p_{k,k+1} = 1$. Once we move to the last regime m, we stay there for ever, that is $p_{m,m} = 1$. This structure enforces the ordering (8) on the change-points.

3.1 Bayesian Estimation

To conduct model estimation, we use Bayesian methods. We specify independent conditionally conjugate priors for the parameters. They are

$$\beta_k \sim N(d_0, B_0), \; \sigma_k^2 \sim IG\left(\frac{v_0}{2}, \frac{l_0}{2}\right), \; p_{ii} \sim Beta\left(a_0, b_0\right) \qquad (10)$$

for $k = 1, ..., m$, and $i = 1, ..., m - 1$, and where $IG(\cdot, \cdot)$ denotes the inverse Gamma distribution. For a detailed discussion of prior selection in the context of change-point models see Koop and Potter (2004).

3.1.1 Gibbs Sampling

Although the posterior for the model is not a well known distribution, we can obtain samples from the posterior based on a Gibbs sampling scheme. Good introductions to Gibbs sampling and Markov chain Monte Carlo methods can be found in Chib (2001), and Geweke (2005).

We divide the parameters into 3 blocks: parameters of the HAR model $\Theta = \{\theta_1, ..., \theta_m\}$, the state of the system S and the transition matrix P. We augment the parameter space to include the states so that we sample from the full posterior $p(\Theta, S, P | I_T)$. After choosing a starting value, $\Theta^{(0)}$, and $P^{(0)}$, Gibbs sampling requires we iterate sampling from the following conditional densities

- $S | P, \Theta, I_T$

- $P | S, \Theta, I_T$

- $\Theta | S, P, I_T$

After dropping a set of burn-in samples to ensure convergence, we collect the remaining draws $\{P^{(h)}, \Theta^{(h)}, S^{(h)}\}_{h=1}^R$ to approximate the posterior distribution. To ensure these draws have converged to the stationary distribution, we investigate running the chain from different starting values and compute convergence diagnostics such as Geweke (1992).[8]

For a sufficiently large posterior sample, any function of interest can be consistently estimated. For example,

$$\widehat{g(\theta_k)} = \frac{1}{R} \sum_{h=1}^R g(\theta_k^{(h)}) \qquad (11)$$

---

[8]All results in this paper are robust to different starting values. For instance, we consider many different starting values, including values drawn randomly over a wide range of the parameter space.

is a simulation consistent estimate of $E\left[g\left(\theta_k\right)|I_T\right]$, the posterior mean of $g(\theta_k)$.

Below we provide more details on each step of the Gibbs sampling.

Step 1: Simulation of $S|P,\Theta,I_T$. Chib (1996) shows that a joint draw of all states can be achieved using

$$p(S|\Theta,P,I_T) = p(s_T|\Theta,P,I_T)\prod_{t=1}^{T-1}p(s_t|s_{t+1},\Theta,P,I_t) \tag{12}$$

in which we sample sequentially from each density on the right hand side of (12) beginning with $p(s_T|\Theta,P,I_T)$, and then $p(s_t|s_{t+1},\Theta,P,I_t)$ $t=T-1,...,1$. At each step we condition on the previously drawn state s_{t+1}, until we obtain a full draw S. The individual densities in (12) are obtained based on the following three steps.

(a) Initialization: At $t=1$, set $p(s_1 = 1|\Theta,P,I_1) = 1$.

(b) Compute the Hamilton (1989) filter, $p\left(s_t = k|\Theta,P,I_t\right)$. This involves a prediction and an update step in which we iterate on the following from $t=2,...,T$,

$$p\left(s_t = k|\Theta,P,I_{t-1}\right) = \sum_{l=k-1}^{k}p(s_{t-1}=l|\Theta,P,I_{t-1})p_{lk}, \quad k=1,...,m, \tag{13}$$

$$p\left(s_t = k|\Theta,P,I_t\right) = \frac{p\left(s_t = k|\Theta,P,I_{t-1}\right)f\left(y_t|\Theta,I_{t-1},s_t = k\right)}{\sum_{l=1}^{m}p\left(s_t = l|\Theta,P,I_{t-1}\right)f\left(y_t|\Theta,I_{t-1},s_t = l\right)}, \quad k=1,...,m.$$

The last equation is obtained from Bayes' rule. Note that in (13) the summation is only from $k-1$ to k, due to the restricted nature of the transition matrix, and $f\left(y_t|\Theta,I_{t-1},s_t = k\right) \sim N\left(X_t\beta_k,\sigma_k^2\right)$.

(c) Finally, Chib (1996) has shown that the individual densities in (12) are,

$$p\left(s_t|s_{t+1},\Theta,P,I_t\right) \propto p\left(s_t|\Theta,P,I_t\right)p\left(s_{t+1}|s_t,P\right). \tag{14}$$

Thus, given $s_T = m$, we can draw s_t backwards over $t=T-1,T-2,\ldots,2$ as

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

where

$$c_t = \frac{p\left(s_t = k|\Theta,P,I_t\right)p\left(s_{t+1}=k|s_t=k,P\right)}{\sum_{l=k-1}^{k}p\left(s_t = l|\Theta,P,I_t\right)p\left(s_{t+1}=k|s_t=l,P\right)}.$$

c_t is similar to Bayes' rule and the filter $p\left(s_t = k|\Theta,P,I_t\right)$ is computed in step (b).

9

Finally, note that $P(s_1 = 1|I_1, s_2, \Theta, P) = 1$.

Step 2: Simulation of $P|S, \Theta, I_T$. The conditional posterior for each component of P is

$$p_{ii}|S \sim Beta\,(a_0 + n_{ii}, b_0 + 1) \tag{15}$$

where n_{ii} is the number of one-step transitions from state i to state i in the sequence S.

Step 3: Simulation of $\Theta|S, P, I_T$. The conditional posterior density of θ_k only depends on the data in regime k. Therefore, let $\widehat{Y}_k = \{y_t : s_t = k\}$ and $\widehat{X}_k = \{X_t : s_t = k\}$ and use standard Gibbs sampling methods for the linear model,

$$\beta_k|\widehat{Y}_k, \widehat{X}_k, \sigma_k^2 \sim N(M_k, V_k) \tag{16}$$

where

$$M_k = V_k(\sigma_k^{-2}\widehat{X}_k'\widehat{Y}_k + B_0^{-1}d_0),\ V_k = (\sigma_k^2\widehat{X}_k'\widehat{X}_k + B_0^{-1})^{-1},$$

and

$$\sigma_k^2|\widehat{Y}_k, \widehat{X}_k, \beta_k \sim IG\left(\frac{v_k}{2}, \frac{l_k}{2}\right) \tag{17}$$

where $v_k = T_k + v_0$, $l_k = (\widehat{Y}_k - \widehat{X}_k\beta_k)'(\widehat{Y}_k - \widehat{X}_k\beta_k) + l_0$, and T_k is the number of observations in regime k.

3.1.2 Only Breaks in Conditional Mean Parameters

Suppose we have a subset of regression parameters η with n elements that are subject to breaks denoted by η_k, $k = 1, ..., m$, while the remainder δ and the variance σ^2 are not. If we partition the regression parameter as $\beta_k = [\eta_k'\ \delta']'$ and reorder $X_t = [U_t\ V_t]$ to conform to this we carry out sampling as follows.

- To sample from $\eta_k|\delta, \sigma^2, S, I_T$, for each regime $k = 1, \cdots, m$, we use the observations only in regime k and let $\widehat{y}_k = \{y_t - V_t\delta : s_t = k\}$ and $\widehat{X}_k = \{U_t : s_t = k\}$ and use standard Gibbs sampling methods as in step 3.

- To sample from $\delta|\eta_1, ..., \eta_k, \sigma^2, S, I_T$, we use the model $\widehat{y}_t = V_t\delta + \epsilon_t$, where $\widehat{y}_t = y_t - \sum_{k=1}^{m} U_t\eta_k I(s_t = k)$ where $I(\cdot)$ is the indicator function. Once again standard methods can be used to sample δ.

- For $\sigma^2|\eta_1, ..., \eta_k, \delta, S, I_T$, we sample the variance from the model $y_t = \sum_{k=1}^{m} U_t\eta_k I(s_t = k) + V_t\delta + \epsilon_t$ with $\epsilon_t \sim NID(0, \sigma^2)$ which is a draw from the inverse gamma density.

3.1.3 Only Breaks in the Variance

Suppose only the variance changes between regimes, and β is constant.

- Draw of $\sigma_k^2|\beta, S, I_T$, $k = 1, ..., m$, only uses data in each regime, this is standard as in step 3 above.

- $\beta|\sigma_1^2, ..., \sigma_k^2, S, I_T$. Due to the heteroskedasticity, we construct new data as $\widehat{y}_t = y_t/\sigma_k$, and $\widehat{X} = X/\sigma_k$ in each regime. Then we have the standard linear model with an error variance of 1.

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

3.2 Bayes Factors

Let A denote a model parameterization in which some or all parameters are subject to breaks and conditional on k breaks occurring. Then the marginal likelihood (ML) is defined as

$$p(Y_T|A) = \int p(Y_T|A, \Theta, P)p(\Theta, P|A)d\Theta dP$$

which is a measure of the success the model has in accounting for the data after the parameter uncertainty has been integrated out over the prior. The term $p(\Theta, P|A)$ is the prior and $p(Y_T|A, \Theta, P)$ is the likelihood function and calculated as

$$\log p\left(Y_T|A, \Theta, P\right) = \sum_{t=1}^{T} \log f\left(y_t|I_{t-1}, \Theta, P\right) \tag{18}$$

where

$$f\left(y_t|I_{t-1}, \Theta, P\right) = \sum_{k=1}^{m} f\left(y_t|I_{t-1}, \Theta, s_t = k\right) p\left(s_t = k|\Theta, P, I_{t-1}\right). \tag{19}$$

The last term in the right hand side of Equation (19) is computed from Equation (13). For this model we can estimate the marginal likelihood following Chib (1995), the details of which are reported in the Appendix.

Once we calculate the marginal likelihood for different specifications we can compare them using Bayes factors. Note that we can compare models across the number of breaks, as well as the type of breaks (restricted versus unrestricted). Bayes factors can be used to compare any model configuration. The relative evidence for model A versus B is $BF_{AB} = p(Y_T|A)/p(Y_T|B)$. Kass and Raftery (1995) suggest interpreting the evidence for A as: not worth more than a bare mention for $1 \leq BF_{AB} < 3$; positive for $3 \leq BF_{AB} < 20$;

strong for $20 \leq BF_{AB} < 150$; and very strong for $BF_{AB} \geq 150$. Equivalently, based on a log scale, $\log(BF_{AB}) > 0$ is evidence in favor of A versus B.

4   Monte Carlo Simulation

In this section we use Monte Carlo simulations to investigate the ability of the change-point model to detect change-points, and date the time of a change-point. We also consider the effect of different sample sizes, and the in-sample forecasting performance. As mentioned above, a key ingredient of the analysis is the marginal likelihood which we estimate following Chib (1995). We have also calculated the marginal likelihood following Gelfand and Dey (1994) but do not report it to save space. We found a very close correspondence between the two estimates and identical conclusions from Bayes factors.

4.1   Setting

We consider the HAR model based on equation (4), in which change points affect one or more model parameters. Table 1 lists all the specifications used in the simulations and the empirical application. Specifically, M_0 is the simple HAR model without any structural change, $M_1 - M_7$ are models in which different parameters change and M_8 is the model in which all parameters change from a structural break.

For each model, we generate $T = 1000$ observations, which is much less than the actual data in our empirical application. Thus, we test the robustness of the method under more challenging conditions. The true models we consider include cases of no change point, 1 and 2 change points. When there is 1 change point, the position of this change point follows a uniform distribution $U(0.25 \times T, 0.75 \times T)$. When there are 2 change points, the first one follows $U(0.2 \times T, 0.4 \times T)$ and the second follows $U(0.6 \times T, 0.8 \times T)$. This setting allows us to account for the randomness of the change points as well as ensuring sufficient observations in each regime to conduct estimation.

To make our simulation empirically realistic, we select parameter values of the HAR model which are close to those reported in Andersen et al. (2006). They are listed in Table 2. For example, in M_0 there are no structural breaks, all parameters are constant across regimes, while for M_1 only β_0 changes. If there is 1 break, β_0 changes from -0.1 to -0.4. If there are 2 breaks, it changes from -0.1 to -0.4 and finally to -0.1.

The priors are $\beta_k \sim N(0, 100I)$, $\sigma_k^2 \sim IG(0.001, 0.001)$, $k = 1, ..., m$, and $p_{ii} \sim Beta(20, 0.1)$, $i = 1, ..., m-1$. In this setting, β_k and σ_k^2 are very uninformative, while the prior for p_{ii} favors infrequent structural breaks. Specifically, it means we assume the expected duration of each regime is about 201 before we see the data. However, if

there are outliers in the time series, a single point may be identified as a separate regime. To rule this out and allow for only infrequent breaks, we impose the assumption that each regime lasts at least 66 days, which is roughly 3 months in actual trading time. In practice, when we simulate a draw of S which has a regime shorter than 66, we discard it and resample until each regime is 66 or more in length. Our results are robust to different duration restrictions which we discuss in more detail for the empirical application in Section 5. A similar strategy is followed by Kim et al. (2005) using financial data. The first 5000 samples from the posterior simulator are discarded and the next 15000 are used for posterior inference.

4.2   Identifying the Number of Change Points

In the following, the model specification M_i is assumed known but the number of structural breaks and model parameters are not known. For each draw from the data generating process (DGP), we estimate the change-point model assuming $0, 1, 2,$ and 3 structural breaks. We rank the evidence for the number of break points according to the largest marginal likelihood. The best change point specification has the largest marginal likelihood, while the second best has the next largest, etc. We draw a new sample of data from the DGP and repeat this until 100 repetitions are completed. Then we report the frequency over repetitions in which each specification is *best* according to the marginal likelihood.

Table 3 lists the results for each of the different specifications. For convenience, bold entries in these cells should be 100 if classification is perfect. For example, the first entry (row) for M_1 says that for a DGP with no change points, out of 100 repetitions 90 are correctly identified as no change point, while 10 (8+1+1) are identified incorrectly as change points. The next entry in the table repeats this for a DGP with 1 change point. Here, 92 times 1 change point is correctly identified.

Overall the Chib model works very well. When there is no change point, this is correctly selected most of the time. Looking at the cases when data are generated from the model with no change point (first row in each panel), they are $90/100$ for M_1, $96/100$ for M_2, $99/100$ for M_3 and $100/100$ for $M_4 - M_8$. Except for a few cases, this method will correctly rule out structural breaks when the underlying process is actually stable.

On the other hand, when the process contains change points, the Chib model correctly identifies the existence of the change points in most cases. For example, the probability of correctly identifying instability of the process is 0.98 ((92+3+3)/100) for M_1 and 1 change point, and is 0.79 for 2 change points. Furthermore, the correct number of change points are found most of the time. Looking at the numbers in bold, many of them are close

13

to 100. However, the relatively low numbers associated with M_1 and M_2 show that the method is not as powerful when there are changes only in the intercept β_0 or only in β_1. Therefore, for this DGP it is easier to identify structural breaks when more parameters undergo a change. For example, compare M_1, and M_2, with the better performance of M_4 in which all β change from a break. Similarly, we see better results when the variance is subject to structural breaks. The best results of the table correspond to models M_5–M_8. In summary, the change-point model provides a very reliable method for the identification of structural breaks using a marginal likelihood criteria.

4.3 Dating the Change Points

After finding the number of structural breaks, we are interested in whether this method can identify the position of the change points correctly. Given S, we define the jth change-point as, $cp_j = \inf\{t|s_t \neq s_{t-1}, t > cp_{j-1}\}$ with $cp_0 = 0$. Given a full MCMC run, we use the posterior mean $\overline{cp}_j = \frac{1}{R} \sum_{h=1}^{R} cp_j^{(h)}$, where $cp_j^{(h)}$ is the jth change point based on the draw $S^{(h)}$.

For each of the 100 repetitions we consider two measures of accuracy, mean absolute error (MAE), and root mean squared error (RMSE) defined as,

$$MAE = \frac{1}{100} \frac{1}{W} \sum_{i=1}^{100} \sum_{j=1}^{W} \left| \overline{cp}_j(i) - cp_j(i) \right|, RMSE = \sqrt{\frac{1}{100} \frac{1}{W} \sum_{i=1}^{100} \sum_{j=1}^{W} \left(\overline{cp}_j(i) - cp_j(i) \right)^2}$$

(20)

where W is the number of change points. $cp_j(i)$ is the true position of change point j, and $\overline{cp}_j(i)$ is the estimate, from the ith draw of the DGP. Recall that change points are randomly drawn from the DGP each repetition.

Table 4 reports the MAE and RMSE when the underlying process contains 1 or 2 change points. For the results to be meaningful we focus only on models which have the correct number of change points. For example, if the DGP has 2 change points, we focus on specifications that assume 2 change points only. Consistent with previous results, M_1 and M_2 perform the worst, and we see improvements when more parameters are subject to structural breaks. Specification M_8 in which all parameters change is the most accurate.

4.4 Sample Size

We are also concerned about the robustness of this method with respect to different sample sizes. Here we consider one of the more challenging specifications with sample sizes of $500, 1000, 2000$ and 5000. We report results for only breaks in β_1, change-point

specification M_2, with 2 change points in Table 5. Increasing the number of observations improves the identification of the true number of change points. The distribution is also more concentrated on the true model. We obtain similar, but generally better results for other models which are not reported here.

4.5 Forecasts

Another approach to compare the evidence for the number of change points is by forecast quality through the associated loss functions RMSE, MAE and the R^2. The R^2 is obtained from a regression of realized observations on a constant and a model forecast (Mincer and Zarnowitz (1969)). In our simulation exercise, we calculate these quantities as well. That is, using the change-point model estimated conditional on $m-1$ breaks, we calculate the in-sample model forecast for RV_t using the posterior mean. Since we assume RV_t is conditionally log-normal, we use the following partially analytical results for the posterior mean:

$$E\left[RV_t|I_{t-1}\right] \approx \frac{1}{R}\sum_{h=1}^{R}\exp\left[X_t\beta^{(h)} + \frac{1}{2}\sigma^{2(h)}\right] \tag{21}$$

where $\beta^{(h)}, \sigma^{2(h)}$ for $h = 1, \ldots, R$ are draws from the posterior simulator conditional on all the data.

We calculate the in-sample RMSE, MAE and R^2 for specifications M_3 and M_8 and summarize the results in Table 6. Similar results are obtained for other cases. For each criteria we report the number of times out of 100 repetitions that the specific change-point model is selected as best based on a forecast criterion and the marginal likelihood. For instance, in the top row, given a DGP with no change points, 90 out of 100 (90/100) repetitions the no change-point model has the smallest MAE, while 6/100 times the 1 change-point model is best, and 4/100 times the 2 change-point model is best. If there were perfect classification the bold numbers should be 100.

The RMSE criterion generally identifies the correct number of structural breaks. The MAE deteriorates with more change points for M_3. Note that the posterior mean is only an optimal estimate for RMSE, and we include the MAE for comparison only. The R^2 criterion is particularly bad in both cases. This shows that forecasts alone provide weaker, and in the case of R^2, misleading evidence about structural breaks. The marginal likelihood, that uses the whole distribution and hence more information, performs better. In the case of M_8, we would expect the marginal likelihood to be the best since it is sensitive to breaks in the variance while the RMSE, MAE and R^2 should not be. However,

it also performs well for M_3 in which the variance is not subject to breaks.

In conclusion, we have presented simulation evidence that the Chib change-point model performs well in identifying the number of structural breaks, dating change points, and the superiority of the marginal likelihood criterion in identifying the number of breaks.

5   Application to S&P 500 Volatility

5.1   Data

We consider the S&P 500 index by using the Spyder (Standard & Poor's Depository Receipts), which is an Exchange Traded Fund that represents ownership in the S&P 500 Index. The ticker symbol is SPY. Since this asset is actively traded, it avoids the stale price effect of the S&P 500 index.[9] The Spyder trades at about 1/10th the value of the index. There is a very close correspondence between the series and generally the tracking error is well below 1%.

The Spyder price transaction data are obtained from the Trade and Quotes (TAQ) database. We removed any price transaction change that was larger than 3% which were obvious errors and keep those records with a TAQ correction indicator of 0 (regular trade) and when possible a 1 (trade later corrected). We also excluded any transaction with a sale condition of Z, which is a transaction reported on the tape out of time sequence. A 5 minute grid from 9:30 to 16:00 was constructed using previous-tick method[10] which means if there is no transaction at that time grid, the nearest previous transaction record is used. The first observation of the day occurring just after 9:30 was used for the 9:30 grid time. From this grid, 5 minute intraday log returns are constructed. Then the return data was used to construct daily returns, realized volatility, and realized bipower variation. The data ranges from January 29, 1993 to March 30, 2004. Conditioning on the first 22 observations, we have $T = 2791$ observations. Table 7 displays summary statistics for daily RV_t and $\log(RV_t)$.

5.2   Results

To conduct Bayesian estimation we use the same prior[11] as in our simulation exercises but focus on the HAR model with jumps and asymmetric terms, see equation (7). We investigate this model under the structural change configurations in Table 1. Table 8

---

[9]See Atchison et al. (1987) for a discussion of the spurious autocorrelation in index returns due to nonsynchronous trading.

[10]For this and other sampling methods see Hansen and Lunde (2006).

[11]The results in Section 5 are robust to different hyperparameters in our prior distributions.

displays the log marginal likelihood for specifications with no change point up to 7 change points.

The results suggest the existence of a change point for S&P 500 data. The log marginal likelihood for no change point is -2333.28, but all specifications with structural breaks improve on this except for M_1 and M_2. The difference in the best structural break specification (M_5, and one break) and the no-break model is large with a Bayes factor of $\exp(57.99)$ in favor of a single break. This is very strong evidence. Taking a closer look at the table, the results suggest one change point. For most model settings, the largest marginal likelihood occurs at 1 change point. There is some posterior support for 2 change points for M_5, but it is dominated by its 1 change point counterpart, and the Bayes factor is 29.67 favoring 1 change point versus 2.

More interesting findings come from the comparison between models. The largest log marginal likelihood across all models is -2275.29 from M_5 with 1 change point. This is a structural break in σ^2. Considering the second largest number in this table, which is -2277.51 from M_7 with 1 change point, the Bayes factor for M_5 vs M_7 is 9.22. Therefore, we conclude the effect of the break is mainly variance, but there is the possibility that the break affects the first 2 regression parameters as well as the variance of log-realized volatility.[12] Our results also suggest that there is no structural break in β_2 or β_3, which can be seen from the relative small value for M_4 (compared to M_3) and M_8 (compared to M_7).

Figure 1 displays the data, the posterior density of the change point date and the cumulative probability of the change point for M_5 conditional on one change point. There is some uncertainty as to the change point date. The posterior mode of the change point density is associated with February 6, 1997.

There are several events that may have contributed to the change in volatility dynamics. The New York Stock Exchange (NYSE) launched real-time stock tickers on CNBC and CNN-FN in 1997. Previously, market data had been delayed 20 minutes. The NYSE changed the smallest price increment to 1/16 from 1/8 in June 1997.[13] Finally, the East

---

[12]In M_7 with 1 change point, β_0 goes from -0.4341, ($s_t = 1$) to -0.0971, ($s_t = 2$), and β_1 goes from 0.1202 ($s_t = 1$) to 0.3089 ($s_t = 2$), and σ^2 goes from 0.4581, ($s_t = 1$) to 0.2395 ($s_t = 2$). This is an increase in the long-run mean of log-RV along with a reduction in the variance of the innovations to log-RV.

[13]The empirical evidence shows that the smaller ticker size resulted in a smaller bid-ask spread, but there is no conclusive evidence about the relationship between the tick size and trading costs. Goldsteina and Kavajecz (2000) argue both spreads and depths declined after the change of the tick size, and the combined effect of smaller spreads and reduced cumulative limit order book depth has made those investors trading small orders better off. Similarly, Portniaguina et al. (2006) demonstrate trading cost are minimized at larger tick sizes for larger market orders, creating an incentive to submit smaller orders when tick size is reduced. As a result, investors will tend to submit smaller orders and execute their order patiently after the reduction in tick size.

Asian financial crisis began in the summer of 1997.

The first section of Table 9 list the posterior mean and 95% density intervals for the stable model and the best model M_5 with 1 structural break. None of the density intervals for the parameters include 0 except for the coefficient on $\frac{|r_{t-1}|}{\sqrt{RV_{t-1}}}$. The jump coefficient is negative, and means that a jump this period will cause log-volatility to revert to a lower level next period. In other words, volatility is less persistent after a jump. The coefficient on $\frac{|r_{t-1}|}{\sqrt{RV_{t-1}}}I(r_t < 0)$ is positive and very large with a value of 0.2583. A negative return leads to higher log-volatility the next day, while a positive return has little to no effect. Figure 2 plots the marginal posterior densities for the variance. There is considerable difference in the variances from the two regimes. Marginal posterior densities for other parameters are bell shaped.

5.3   Forecasts

Structural breaks can have an important effect on forecasting. In this section, we compare the predictive likelihood of the model M_5 with 1 structural break and the model without structural breaks. The predictive likelihood contains the out-of-sample prediction record of a model, making it the central quantity of interest for model evaluation (Geweke and Whiteman (2005)).

Given the data up to time $s-1$, $Y_{s-1} = \{y_1, ..., y_{s-1}\}$, the predictive likelihood $p(y_s, ..., y_t|Y_{s-1})$ (Geweke (1995,2005)) is the predictive density evaluated at the realized outcome $y_s, ..., y_t$, $s \leq t$. The predictive likelihood for model A can be calculated by

$$p(y_s, ..., y_t|Y_{s-1}, A) = \frac{p(y_1, ..., y_t|A)}{p(y_1, ..., y_{s-1}|A)} \tag{22}$$

where the nominator and the denominator at the right hand side are the marginal likelihood for all the data Y_t, and for the in-sample data Y_{s-1}, respectively. In a similar fashion to the Bayes factor which is based on all the data, we can compare the performance of models on a specific out-of-sample period by predictive Bayes factors. The predictive Bayes factor for model A versus B is

$$PBF = p(y_s, ..., y_t|Y_{s-1}, A)/p(y_s, ..., y_t|Y_{s-1}, B). \tag{23}$$

and summarizes the relative evidence of the two models over the out-of-sample data $y_s, ..., y_t$.

To compare the out-of-sample density forecasts of the model with and without a structural break, we calculate the predictive Bayes factor for data after the break point

February 6, 1997. In other words, $s - 1 =$ February 5, 1997, and $t \geq s$. Specifically, for one observation $(t = s)$ out-of-sample, $PBF = 1.12$; 1 month $(t = s + 21)$ $PBF = 2.17$; 3 months $(t = s + 65)$ $PBF = 20.02$; and 5 months $(t = s + 109)$ $PBF = 176.49$; each in favor of the break specification. The improvements continue till the end of sample.

5.4   Identifying the Structural Break in Real Time

To investigate how quickly we can identify the structure break, we re-estimate the model using data up to the date of the identified structural break, Feb. 6, 1997, then using data up to Feb. 7, 1997, etc. For each run, the Bayes factor is computed for the 1 change-point model M_5 versus the stable no-break model.[14] The recursive $\log(BF)$ is shown in Figure 3. For the S&P 500 data, it takes 300 observations (around 14 months) to provide compelling evidence for a structural break. After this the evidence for a break continues to grow till the final BF at the end of the sample (not shown in the figure) is $\exp(57.99)$.

5.5   Jumps and Asymmetric Terms

Our finding of a structural break is robust to the inclusion or exclusion of jumps and asymmetric terms. Table 10 reports results for our favored specification. The candidate models include a model with neither jump nor asymmetric terms ("No jump, No Asym" column), a model with only jumps ("with Jumps" column), a model with only asymmetric terms ("with Asym" column) and a model with both jump and asymmetric terms (the last column). We also consider a model with an alternative asymmetric term $\log(RV_{t-1})I(r_{t-1} < 0)$, whose results are listed in the "with Alt-Asym" column. According to log-marginal likelihood, the best candidate model M_5 contains both jumps and asymmetric terms. Note that the evidence for a single break is still strong in the other model formulations.

Although jumps are less significant, the asymmetric terms are very important. Including them increases the marginal likelihood greatly. Nevertheless, the best model for $\log(RV_t)$ is the structural break model with both jumps and asymmetric terms.

5.6   Robustness of Minimum Regime Duration

To test the robustness of the minimum regime duration, we estimate the preferred model M_5 with different minimum regime lengths. The results are listed in Table 11. Besides

---

[14]The Bayes factor computed here provides a measure of the model fit to the data *before and after* the break point. The predictive Bayes factor in Section 5.3 provides a measure of the model fit only to data *after* the break point.

the minimum 66 observations (3 months) setting in our previous sections, we also consider the minimum length as short as 10 observations (2 weeks) or 22 observations (1 month), and the longer setting of 132 observations (6 months). In this table, the first row ("Minimum Length" row) reports the lowest bound for the number of the observations in each regime. The marginal likelihoods are almost identical across the different cases. Each of the different regime duration restrictions favor the 1 change point which has a mode at February 6, 1997. The choice of the minimum regime length has little effect on our results.

5.7 The Relationship between Structural Breaks and GARCH Effects

The change in the variance of log-volatility is related to the recent finding of heteroskedasticity in models of S&P 500 realized volatility (Corsi et al. (2005), Andersen et al. (2007), Bollerslev et al. (2007)).[15] To investigate if our finding of a break is spurious due to neglected conditional variance dynamics, we consider breaks in a HAR-GARCH model. Based on the previous results we focus on breaks in the conditional variance only. The HAR model extended to include a GARCH structure is

$$y_t = X_t\beta + \gamma_{s_t}\sigma_t u_t, \ \ u_t \sim N(0,1) \tag{24}$$

$$\sigma_t^2 = \omega + a(y_{t-1} - X_{t-1}\beta)^2 + b\sigma_{t-1}^2, \tag{25}$$

where γ_{s_t} is a scaling constant which has a direct effect on the unconditional variance of log-RV. We normalize $\gamma_1 = 1$ and assume all $\gamma_k > 0, k = 2, ..., m$. Thus, in regime 1, this is a standard HAR model with a GARCH conditional variance. While in later regimes, the conditional variance of log-RV can be larger or smaller than σ_t^2 depending on $\gamma_{s_t} > 1$ or $\gamma_{s_t} < 1$. The advantage of this parameterization is that we can model permanent changes in the variance but avoid the path dependence in the conditional variance[16] induced by parameter changes in $\omega, a,$ and b. As a result, the efficient full block Gibbs sampling for $S|P, \Theta, I_T$ is still available.

The estimation process is more involved than the HAR model. Denoting the model parameters $\Theta = \{\beta, \omega, a, b, \gamma_2, \cdots, \gamma_m\}$, the Gibbs sampler requires us to iterate on 1)

---

[15]Both Corsi et al. (2005) and Bollerslev et al. (2007) use S&P 500 index futures over 1985-2004, while Andersen et al. (2007) use the shorter sample 1990-2005. Corsi et al. (2005) model GARCH effects for the variance of $\sqrt{RV_t}$ while Bollerslev et al. (2007) model GARCH effects in the variance of the $\log(RBP_t)$, and Andersen et al. (2007) similarly model $\log(RBP_t)$ adjusted by a testing procedure for jumps.

[16]See Bauwens et al. (2006) for a discussion of path dependence in regime switching GARCH models.

$S|P, \Theta, I_T$, 2) $P|S, \Theta, I_T$, and 3) $\Theta|S, P, I_T$. The first and second sampling steps are identical to the basic HAR model. However, Gibbs sampling is not available for all parameters of $\Theta|S, P, I_T$. Therefore, we adopt a random walk Metropolis-Hastings (M-H) algorithm to sample Θ. The details of the estimation and calculation of the marginal likelihood are presented in the appendix. In estimation the restrictions $\omega > 0, a \geq 0, b \geq 0$, are imposed for positivity of variance and $a + b < 1$ for stationarity.

Given our previous results, we concentrate on a stable HAR-GARCH model as well as its structural break versions with 1–3 change points. The log marginal likelihood for both the HAR model and the HAR-GARCH model are listed in Table 12. Incorporating GARCH effects, improves the model with the $\log(ML)$ increasing from -2333.28 for the stable HAR to -2276.10 for the stable HAR-GARCH parameterization. The best HAR model with constant variance within regimes and 1 change point is still marginally better than the stable HAR-GARCH model. In other words, a single structural break in the homoskedastic model provides a similar description of the data as the stable GARCH model.

There is strong evidence of parameter change for the HAR-GARCH model. Conditional on 1 break, the $\log(ML)$ increases from -2276.10 to -2253.38, providing a Bayes factor of $\exp(22.72)$ in favor of 1 structural break. This is consistent with our previous results. For instance, the unconditional variance of log-RV drops from 0.4482 to 0.2627 in the second regime.[17] The posterior mode of the change point is exactly the same date, February 6, 1997. Finally, there is some evidence of multiple breaks in this model. Bayesian model averaging could be used to provide forecasts that incorporate uncertainty about the number of change points.

The parameter estimates of the HAR-GARCH specification are listed in Table 9. The most notable change is the dramatic reduction in the persistence measure $a + b$ of the conditional variance. This drops from 0.9856 for no change point, to 0.6874 for 1 change point. Not only does the no break HAR-GARCH model have the wrong long-run variance, but it also overstates the persistence of the variance of log-RV.

6   Summary and Conclusion

This paper explores structural breaks in realized volatility. We focus on the popular heterogeneous autoregressive (HAR) models of the logarithm of realized volatility (log-RV) and test for breaks using Bayesian methods. Our Monte Carlo simulations demonstrate that this method is powerful in detecting and dating structural breaks. We show that

---

[17]The unconditional variance within a regime is $\gamma_{s_t}^2 \omega / (1 - a - b)$.

detecting breaks using traditional forecast criteria may be misleading. Applying our model to daily S&P 500 data, we find strong, robust evidence of a structural break in log-RV. The main effect of the break is on the variance of log-volatility, but there is weaker evidence for a change in both the regression parameters and variance. We demonstrate that accounting for the structural break improves density forecasts. Finally, we consider a HAR model with GARCH effects and also find a structural break in the unconditional variance of log-RV. There is a permanent reduction in the variance of log-RV in February 1997 for all of our specifications using data from 1993-2004. Failure to model this break results in biased parameter estimates and poor density forecasts.

7 Appendix

7.1 Marginal Likelihood

In this appendix we review the estimation of the marginal likelihood following Chib (1995). Given observations $Y_T = \{y_1, y_2, \ldots, y_T\}$, we follow the notation in the previous sections except that conditioning on X_t is suppressed. A rearrangement of Bayes rule gives the marginal likelihood, $p(Y_T)$, as

$$p(Y_T) = \frac{f(Y_T|\Theta, P) \, p(\Theta, P)}{p(\Theta, P|Y_T)} \tag{26}$$

where $f(Y_T|\Theta, P)$ is the likelihood function, $p(\Theta, P)$ is the prior density and $p(\Theta, P|Y_T)$ is the posterior ordinate. Although this equation is valid for any value (Θ, P) in the parameter space, a point with high posterior mass will tend to provide a more accurate estimate. Therefore, we select the posterior mean denoted as (Θ^*, P^*). Then

$$\ln p(Y_T) = \log f(Y_T|\Theta^*, P^*) + \log p(\Theta^*, P^*) - \log p(\Theta^*, P^*|Y_T) \tag{27}$$

The prior density $p(\Theta^*, P^*)$ can be evaluated directly and the likelihood function $f(Y_T|\Theta^*, P^*)$ can be calculated as in (18).

To calculate the posterior ordinate $p(\Theta^*, P^*|Y_T)$, note the decomposition

$$p(\Theta^*, P^*|Y_T) = p(\beta^*|Y_T) \, p(\sigma^{2*}|Y_T, \beta^*) \, p(P^*|Y_T, \beta^*, \sigma^{2*}) \tag{28}$$

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

$$p(\beta^*|Y_T) = \int p(\beta^*|Y_T, S, \sigma^2) \, p(S, \sigma^2|Y_T) \, dS d\sigma^2 \tag{29}$$

and can be estimated as

$$p\widehat{(\beta^*|Y_T)} = \frac{1}{R}\sum_{h=1}^{R} p\left(\beta^*|Y_T, \sigma^{2(h)}, S^{(h)}\right) \tag{30}$$

where the draws $\{\sigma^{2(h)}, S^{(h)}\}_{h=1}^{R}$ are available directly from our estimation step, and $p\left(\beta^*|Y_T, \sigma^{2(h)}, S^{(h)}\right) = \prod_{k=1}^{m} p\left(\beta_k^*|Y_T, \sigma^{2(h)}, S^{(h)}\right)$ where each conditional density is determined by (16). The second term in (28) is equal to

$$p\left(\sigma^{2*}|Y_T, \beta^*\right) = \int p\left(\sigma^{2*}|Y_T, \beta^*, S\right) p\left(S|, Y_T, \beta^*\right) dS, \tag{31}$$

where $p\left(\sigma^{2*}|Y_T, \beta^*, S\right) = \prod_{k=1}^{m} p\left(\sigma_k^{2*}|Y_T, \beta^*, S\right)$ and each conditional posterior has density according to (17). To obtain the draws from $p\left(S|Y_T, \beta^*\right)$, we run an additional reduced Gibbs sampling conditional on β^*; i.e. we run all the sampling in Section 3.1 except that we do not draw the values for β but fix them to be β^*. From this reduced run simulation, $\{S^{(h)}\}_{h=1}^{R}$ is used in

$$p\widehat{(\sigma^{2*}|Y_T}, \beta^*) = \frac{1}{R}\sum_{h=1}^{R} p\left(\sigma^{2*}|Y_T, \beta^*, S^{(h)}\right). \tag{32}$$

The last term of (29) is

$$p\left(P^*|Y_T, \beta^*, \sigma^{2*}\right) = \prod_{i=1}^{m-1} p\left(p_{ii}^*|Y_T, \beta^*, \sigma^{2*}\right) \tag{33}$$

and each term equal to

$$p\left(p_{ii}^*|Y_T, \beta^*, \sigma^{2*}\right) = \int p\left(p_{ii}^*|Y_T, S, \beta^*, \sigma^{2*}\right) p\left(S|\beta^*, \sigma^{2*}, Y_T\right) dS$$

Sampling $\{S^{(h)}\}_{h=1}^{R}$ from $p\left(S|\beta^*, \sigma^{2*}, Y_T\right)$ the estimate of each term is

$$p\widehat{(p_{ii}^*|Y_T}, \beta^*, \sigma^{2*}) = \frac{1}{R}\sum_{h=1}^{R} p\left(p_{ii}^*|Y_T, S^{(h)}, \beta^*, \sigma^{2*}\right). \tag{34}$$

7.2 Estimation of HAR-GARCH model

We set all priors in the regression equation as before, they are independent normal $N(0, 100)$. The GARCH parameters have independent normal $N(0, 100)$ truncated to $\omega > 0, a \geq 0, b \geq 0$, and $a + b < 1$. The priors for the scaling parameters γ_k $(k = 2, \cdots, m)$

23

are also truncated $N(0, 100)$ with positive supports. These priors are uninformative.

For the sampling step $\Theta|S, P, I_T$, the conditional distributions for some of the model parameters are unknown, and therefore Gibbs sampling is not available. Instead we use a random walk Metropolis-Hastings algorithm. If we denote all the parameters except for θ_l as $\Theta_{-l} = \{\theta_1, \cdots, \theta_{l-1}, \theta_{l+1}, \cdots, \theta_L\}$, we sample a new θ_l given Θ_{-l} fixed. With Θ as the previous value of the chain we iterate on the following steps:

Step 1: Propose a new Θ' according to $\Theta'_{-l} = \Theta_{-l}$, with element l determined as

$$\theta'_l = \theta_l + e_l, \ e_l \sim N(0, \xi_l^2). \tag{35}$$

Step 2: Accept Θ' with probability

$$\min\left\{\frac{p(Y_T|\Theta', P, S)p(\Theta', P)}{p(Y_T|\Theta, P, S)p(\Theta, P)}, 1\right\} \tag{36}$$

and otherwise reject. $p(\Theta, P)$ is the prior, and

$$\log p\left(Y_T|\Theta, P, S\right) = \sum_{t=1}^{T}\left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\gamma_{s_t}^2\sigma_t^2) - \frac{(y_t - X_t\beta)^2}{2\gamma_{s_t}^2\sigma_t^2}\right] \tag{37}$$

where $\sigma_t^2 = \omega + a(y_{t-1} - X_{t-1}\beta)^2 + b\sigma_{t-1}^2$.[18] Each ξ_l^2 is selected to give an acceptance frequency between 0.3–0.5. Running Step 1-2 above for all the parameters $l = 1, \cdots, L$, we obtain a new draw Θ which is one iteration. We perform 200,000 iterations and use the last 100,000 for posterior inference.

For the marginal likelihood we use the method of Gelfand and Dey (1994) adapted by Geweke (2005) (Section 8.2.4). This estimate is based on $\frac{1}{R}\sum_{h=1}^{R} g(\Theta^{(h)})/[p(Y_T|\Theta^{(h)}, P^{(h)})p(\Theta^{(h)}, P^{(h)})] \to p(Y_T)^{-1}$ as $R \to \infty$, where $p(Y_T|\Theta, P)$ is the likelihood with S integrated out as in (18)-(19), and $g(\Theta^{(h)})$ is a truncated multivariate Normal. Note that the prior, likelihood and $g(\Theta)$ must contain all integrating constants. Finally, to avoid underflow/overflow we use logarithms in this calculation and above in (36).

References

[1] Andersen, T. G. and T. Bollerslev. (1998). "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts." *International Economic Review*, 39, 885-905.

---

[18]To start up the conditional variance we set $\sigma_0^2 = \omega/(1 - a - b)$.

[2] Andersen, T. G., T. Bollerslev and F. X. Diebold. (2006). "Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility." *Review of Economics and Statistics*, forthcoming.

[3] 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.

[4] Andersen T. G., T. Bollerslev, F. X. Diebold and P. Labys. (2001). "The Distribution of Exchange Rate Volatility." *Journal of the American Statistical Association*, 96, 42-55.

[5] Andersen T. G., T. Bollerslev, F. X. Diebold and P. Labys. (2003). "Modeling and Forecasting Realized Volatility." *Econometrica* 71.2, 579-625.

[6] Andersen, T. G., T. Bollerslev and X. Huang. (2007). "A Reduced Form Framework for Modeling Volatility of Speculative Prices based on Realized Variation Measures." Working paper, Duke University.

[7] Andreou, E. and E. Ghysels. (2002). "Detecting Multiple Breaks in Financial Market Volatility Dynamics." *Journal of Applied Econometrics*, 17(5), 579-600.

[8] Atchison, M. D., K. Butler and R. R. Simonds. (1987). "Nonsynchronous Security Trading and Market Index Autocorrelation." *The Journal of Finance*, 42(1), 111-118.

[9] Bandi F. M. and J. R. Russell. (2006). "Separating Microstructure Noise from Volatility." *Journal of Financial Economics*, 79, 655-692.

[10] 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.

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

[12] 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.

[13] Bauwens, L., A. Preminger and J. V.K. Rombouts. (2006). "Regime Switching GARCH Models." CORE Discussion paper 2006/11.

[14] 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." working paper, Duke University.

[15] Chib, S. (1995). "Marginal likelihood from the Gibbs output." *Journal of the American Statistical Association*, 90(432), 1313-1321.

[16] Chib, S. (1996). "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models." *Journal of Econometrics*, 75, 79-97.

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

[18] Chib, S. (2001). "Markov Chain Monte Carlo Methods: Computation and Inference." In J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, volume 5, North Holland, Amsterdam, 3569-3649.

[19] Chu, CS J. (1995). "Detecting Parameter Shift in GARCH Models." *Econometric Reviews*, 14, 241-266.

[20] Corsi, F. (2004). "A simple long memory model of realized volatility." Working paper, University of Southern Switzerland.

[21] Corsi, F., U. Kretschmer, S. Mittnik and C. Pigorsch. (2005). "The Volatility of Realized Volatility." Center for Financial Studies, working paper 2005/33.

[22] Engle, R. F. and J. G. Rangel. (2005). "The Spline-GARCH Model for Unconditional Volatility and its Global Macroeconomic Causes." Working Paper, New York University and UCSD.

[23] Forsberg, L. and E. Ghysels. (2006). "Why do Absolute Returns Predict Volatility so well." working paper, University of North Carolina, Chapel Hill.

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

[25] Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to calculating posterior moments (with discussion)." In J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.), *Bayesian Statistics*, Vol. 4, 169-193, Oxford: Oxford University Press.

[26] Geweke J. (1995). "Bayesian comparison of Econometric Models." Working paper, Research Department, Federal Reserve Bank of Minneapolis.

[27] Geweke J. (2005). *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons Ltd.

[28] Geweke J. and C. Whiteman. (2005). "Bayesian Forecasting," Forthcoming in Handbook of Economic Forecasting, Edited by Elliott, Granger and Timmermann, Elsevier.

[29] Ghysels, E., P. Santa-Clara and R. Valkanov. (2003). "Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies." *Journal of Econometrics* (forthcoming).

[30] Giot P. and S. Laurent. (2004). "Modelling daily Value-at-Risk using realized volatility and ARCH type models." *Journal of Empirical Finance*, 11, 379-398.

[31] Giot P. and S. Laurent. (2006). "The information content of implied volatility in light of the jump/continuous decomposition of realized volatility." *Journal of Futures Markets*, forthcoming.

[32] Goldsteina, M. A. and K. A. Kavajecz. (2000). "Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE." *Journal of Financial Economics*, 56, Issue 1, April 125-149.

[33] Hamilton, J. D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica*, Vol. 57, No. 2, 357-384.

[34] Hansen P. R. and A. Lunde. (2006). "Realized Variance and Market Microstructure Noise." *Journal of Business and Economic Statistics*, 24, 127–218.

[35] Hillebrand, E. (2005). "Neglecting parameter changes in GARCH models." *Journal of Econometrics*, 127 (1-2), 121-138.

[36] Huang, X. and G. Tauchen. (2005). "The Relative Contribution of Jumps to Total Price Variance." *Journal of Financial Econometrics*, 3, 456-499.

[37] Hwang, S. and P. L. V. Pereira. (2004). "The Effects of Structural Breaks in ARCH and GARCH Parameters on Persistence of GARCH Models." working paper, Cass Business School.

[38] Kass, R. E. and A. E. Raftery. (1995). "Bayes factors and model uncertainty." *Journal of the American Statistical Association*, 90, 773-795.

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

[40] Koop G. and S. M. Potter. (2004). "Prior Elicitation in Multiple Change-point Models." Working Paper, University of Strathclyde.

[41] 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.

[42] Lamoureux, C. G. and W. D. Lastrapes. (1990). "Persistence in Variance, Structural Change, and the GARCH Model." *Journal of Business and Economic Statistics*, 8(2), 225-34.

[43] Maheu, J. M. and T. H. McCurdy. (2002). "Nonlinear Features of Realized FX Volatility." *Review of Economics and Statistics*, 84 (4), 668-681.

[44] Maheu, J. M. and T. H. McCurdy. (2006). "Do High-frequency Measures of Volatility Improve Forecasts of Return Distributions?" working paper, University of Toronto.

[45] Martens, M., D. van Dijk and M. de Pooter. (2004). "Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity." Tinbergen Institute Discussion Paper.

[46] Martin, G.M. (2000). "US Deficit Sustainability: a New Approach Based on Multiple Endogenous Breaks." *Journal of Applied Econometrics*, 15, 83-105.

[47] Meddahi N. (2002). "A Theoretical Comparison Between Integrated and Realized Volatility." *Journal of Applied Econometrics*, 17, 479-508.

[48] Mikosch, T. and C. Starica. (2004). "Changes of structure in financial time series and the GARCH model." working paper, University of Copenhagen and Chalmers University of Technology.

[49] Mincer J.A. and V. Zarnowitz. (1969). "The evaluation of economic forecasts." In J.A. Mincer (eds.), *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, NBER, New York, 3-46.

[50] Nelson, D. B. (1991). "Conditional heteroskedasticity in asset returns: A new approach." *Econometrica*, 59, 347-370.

[51] Oomen, R. (2005). "Properties of bias corrected realized variance under alternative sampling schemes." *Journal of Financial Econometrics,* 3, 555-577.

[52] Pastor, L. and R. F. Stambaugh. (2001). "The Equity Premium And Structural Breaks." *Journal of Finance*, 56(4), 1207-1239.

[53] Pettenuzzo, D. and A. Timmermann. (2006). "Predictability of Stock Returns and Asset Allocation under Structural Breaks." Working paper, University of California, San Diego.

[54] Portniaguina, E., D. Bernhardtb and E. Hughson. (2006). "Hybrid markets, tick size and investor trading costs." *Journal of Financial Markets*, 9(4), November, 433-447.

[55] Ray B. K. and R. S. Tsay. (2002). "Bayesian methods for change-point detection in long-range dependent processes." *Journal of Time Series Analysis*, 23(6), 687-705.

[56] Rapach, D. and J. K. Strauss. (2006). "Structural Breaks and GARCH Models of Exchange Rate Volatility." Working paper, Saint Louis University and Saint Louis University.

[57] Smith, D. R. (2006) "Testing for Structural Breaks in GARCH Models." working paper, Simon Fraser University.

[58] Starica, C. and C. Granger. (2005). "Nonstationarities in Stock Returns." *The Review of Economics and Statistics*, 87(3), 503-522.

[59] Tauchen G. and H. Zhou. (2005). "Identifying Realized Jumps on Financial Markets." Working Paper, Duke University.

[60] Zhang L., P. Mykland and Y. Ait-Sahalia. (2005). "A tale of two time scales: determining integrated volatility with noisy high-frequency data." *Journal of American Statistical Association*, 100, 1394-1411.

Table 1: Change-point Model Specifications

Model Index	Parameters that change from a break
M_0	None
M_1	β_0
M_2	β_1
M_3	β_0, β_1
M_4	All β
M_5	σ^2
M_6	β_0, σ^2
M_7	$\beta_0, \beta_1, \sigma^2$
M_8	All Parameters

This table labels the various structural break specifications. The first column is the index of the model, the second lists the parameters that can change due to a structural break. All other parameters are kept constant across structural breaks.

Table 2: Parameter Values for Simulation

Regime		M_0	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
1		-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
2	β_0	-0.1	-0.4	-0.1	-0.4	-0.4	-0.1	-0.4	-0.4	-0.4
3		-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
1		0.4	0.4	0.1	0.1	0.1	0.4	0.4	0.1	0.1
2	β_1	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
3		0.4	0.4	0.1	0.1	0.1	0.4	0.4	0.1	0.1
1		0.25	0.25	0.25	0.25	0.4	0.25	0.25	0.25	0.4
2	β_2	0.25	0.25	0.25	0.25	0.15	0.25	0.25	0.25	0.15
3		0.25	0.25	0.25	0.25	0.4	0.25	0.25	0.25	0.4
1		0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.2	0.1
2	β_3	0.2	0.2	0.2	0.2	0.4	0.2	0.2	0.2	0.4
3		0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.2	0.1
1		0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
2	σ^2	0.2	0.2	0.2	0.2	0.2	0.5	0.5	0.5	0.5
3		0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2

This table lists the parameter values for our Monte Carlo simulation. 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 no break the DGP parameters are M_0. If there is 1 break then the DGP parameters are first from regime 1 and then regime 2. Similarly for 2 breaks.

Table 3: Ranking the Number of Change Points by Marginal Likelihood

DGP		Frequency by ML			
	# of CP	0 CP	1 CP	2 CP	3 CP
M_1	0	**90**	8	1	1
β_0	1	2	**92**	3	3
	2	21	1	**40**	38
M_2	0	**96**	4	0	0
β_1	1	1	**55**	37	7
	2	2	3	**56**	39
M_3	0	**99**	1	0	0
β_0, β_1	1	0	**90**	8	2
	2	0	1	**95**	4
M_4	0	**100**	0	0	0
All β	1	0	**91**	7	2
	2	0	0	**87**	13
M_5	0	**100**	0	0	0
σ^2	1	0	**100**	0	0
	2	1	1	**98**	0
M_6	0	**100**	0	0	0
β_0, σ^2	1	0	**100**	0	0
	2	0	0	**100**	0
M_7	0	**100**	0	0	0
$\beta_0, \beta_1, \sigma^2$	1	0	**100**	0	0
	2	0	0	**100**	0
M_8	0	**100**	0	0	0
All Pars	1	0	**100**	0	0
	2	0	0	**100**	0

The first column lists the true model along with the parameters that change from a structural break. CP=change points, ML=marginal likelihood. The "0 CP" column displays the number of times in the 100 repetitions when that specification had the largest marginal likelihood. The "1 CP" column tells the number of times in the 100 repetitions when that specification had the largest marginal likelihood, etc. Each row sums to 100. With perfect classification bold entries would be 100.

Table 4: Dating the Change Points

DGP	# of CP	MAE	RMSE
M_1	1	27.7699	63.2031
β_0	2	26.1847	56.5047
M_2	1	36.7686	63.5820
β_1	2	38.6881	73.1602
M_3	1	6.6670	13.1058
β_0, β_1	2	4.2328	9.6509
M_4	1	7.2147	12.7888
All β	2	6.4533	21.9577
M_5	1	10.5404	16.7350
σ^2	2	10.4418	16.8038
M_6	1	7.5043	11.9126
β_0, σ^2	2	7.4723	12.1060
M_7	1	6.0532	9.7271
$\beta_0, \beta_1, \sigma^2$	2	4.1134	7.6212
M_8	1	5.6092	4.4879
All Pars	2	3.2063	2.3007

The table reports the MAE and RMSE of the dating of the change points over the 100 repetitions. The definition of MAE and RMSE are in the text. The first column lists the true model along with the parameters that change from a structural break, # of CP lists the number of change points in the DGP.

Table 5: Effect of Sample Size on Identification of Change Points

Sample Size	Frequency by ML			
	0 CP	1 CP	2 CP	3 CP
500	16	34	**47**	3
1000	2	3	**56**	39
2000	0	0	**57**	43
5000	0	0	**68**	32

This table reports the simulation results for M_2 (only β_1 changes) with 2 change points. The "0 CP" column records the number of times in the 100 simulations when the specification with no change point has the largest marginal likelihood. The "1 CP" column records the number of times in the 100 simulations when the specification with 1 change point has largest marginal likelihood, etc. With perfect classification bold entries would be 100.

Table 6: Model Comparison using Different Criteria

DGP, M_3		Frequency by Loss Function and ML			
# of CP	Measures	0 CP	1 CP	2 CP	3 CP
0	MAE	**90**	6	4	0
	RMSE	**85**	9	6	0
	R^2	**3**	7	20	70
	ML	**99**	1	0	0
1	MAE	80	**19**	1	0
	RMSE	36	**62**	1	0
	R^2	0	**21**	23	56
	ML	0	**90**	8	2
2	MAE	4	73	**23**	0
	RMSE	0	20	**76**	4
	R^2	0	0	**46**	54
	ML	0	1	**95**	4
DGP, M_8		Frequency by Loss Function and ML			
# of CP	Measures	0 CP	1 CP	2 CP	3 CP
0	MAE	**79**	14	7	0
	RMSE	**68**	13	16	3
	R^2	**10**	8	22	60
	ML	**100**	0	0	0
1	MAE	0	**89**	9	2
	RMSE	0	**88**	10	10
	R^2	0	**19**	23	58
	ML	0	**100**	0	0
2	MAE	0	0	**89**	11
	RMSE	0	0	**84**	16
	R^2	0	0	**35**	65
	ML	0	0	**100**	0

In all cases parameter estimates and change points are obtained from the Chib model estimated conditional on 0,1,2, and 3 change points. The evidence for the number of change points is determined according to MAE, RMSE, R^2 for the in-sample one period ahead forecasts of RV_t using specification M_3, and M_8. Also reported is the marginal likelihood (ML). The R^2 is calculated from the regression $RV_t = a + bE_{t-1}[RV_t] + \text{error}_t$. $E_{t-1}[RV_t]$ is the model forecast computed conditional on the number of change points in row two of the table. "0 CP" column reports the number of times in the 100 repetitions when the specification with no change point has the best performance (lowest MAE and RMSE and highest R^2 and ML). Similarly for the other columns.

Table 7: Summary statistics for RV_t and $\log(RV_t)$

	RV_t	$\log(RV_t)$
Mean	1.1143	-0.3540
Median	0.7423	-0.2980
Std Dev.	1.5777	0.9478
Skewness	7.4730	-0.0133
Excess Kurtosis	91.4800	0.2789
Min	0.0357	-3.3338
Max	30.7842	3.4270

Spyder data Mar 03, 1993 to March 24, 2004 (2791 observations).

Table 8: Comparison by Marginal Likelihood for S&P 500 Volatility

# CP	M_1 β_0	M_2 β_1	M_3 β_0,β_1	M_4 All β	M_5 σ^2	M_6 β_0,σ^2	M_7 β_0,β_1,σ^2	M_8 All Par.
0	**-2333.28**	**-2333.28**	-2333.28	-2333.28	-2333.28	-2333.28	-2333.28	-2333.28
1	-2336.97	-2338.26	**-2326.19**	-2329.20	**-2275.29**	**-2280.00**	**-2277.51**	**-2286.78**
2	-2341.45	-2342.57	-2328.92	-2339.98	-2278.68	-2287.70	-2286.85	-2300.15
3	-2345.52	-2346.76	-2337.53	-2352.40	-2282.83	-2296.95	-2297.92	-2313.09
4	-2349.16	-2351.30	-2341.32	-2362.65	-2289.00	-2306.24	-2308.19	-2336.99
5	-2353.32	-2355.27	-2349.18	-2377.90	-2294.86	-2315.63	-2319.21	-2356.49
6	-2357.99	-2359.84	-2355.99	-2390.73	-2301.71	-2327.58	-2333.86	-2371.96
7	-2364.19	-2364.40	-2363.98	-2404.00	-2308.00	-2335.83	-2349.12	-2394.57

This table reports the log marginal likelihood from HAR models with both jump and asymmetric terms. The first row lists the index of models. The second row lists the parameters subject to structural breaks in each specification. The first column is the number of the change points (CP) that are conditioned on to estimate the marginal likelihood. The bold numbers are the largest marginal likelihood within each specification.

36

Table 9: Parameter Estimates for S&P 500 Volatility

| | HAR | | | HAR-GARCH | |
	0 CP	M_5, 1 CP		0 CP	1 CP
		$s_t = 1$	$s_t = 2$	0 CP	1 CP
β_0	-0.0988	-0.0934		-0.0912	-0.0933
	$(-0.1356, -0.0592)$	$(-0.1288, -0.0570)$		$(-0.1265, -0.0561)$	$(-0.1319, -0.0548)$
β_1	0.2605	0.2737		0.2618	0.2745
	$(0.2134, 0.3076)$	$(0.2304, 0.3248)$		$(0.2127, 0.3117)$	$(0.2223, 0.3259)$
β_2	0.4011	0.4025		0.4151	0.3971
	$(0.3264, 0.4735)$	$(0.3229, 0.4694)$		$(0.3371, 0.4932)$	$(0.3188, 0.4746)$
β_3	0.2674	0.2546		0.2532	0.2559
	$(0.2007, 0.3326)$	$(0.1925, 0.3222)$		$(0.1859, 0.3204)$	$(0.1924, 0.3230)$
β_J	-0.2840	-0.2424		-0.2342	-0.2731
	$(-0.4261, -0.1340)$	$(-0.3824, -0.1038)$		$(-0.3788, -0.0857)$	$(-0.4355, -0.1192)$
β_{A1}	0.0116	0.0041		0.0097	0.0085
	$(-0.0329, 0.0517)$	$(-0.0353, 0.0465)$		$(-0.0319, 0.0478)$	$(-0.0331, 0.0498)$
β_{A2}	0.2596	0.2583		0.2361	0.2394
	$(0.2114, 0.3035)$	$(0.2134, 0.3004)$		$(0.1926, 0.2797)$	$(0.1963, 0.2825)$
σ^2	0.3009	0.4450	0.2290		
	$(0.2864, 0.3176)$	$(0.4010, 0.4928)$	$(0.2124, 0.2450)$		
γ_2					0.7656
					$(0.7192, 0.8137)$
ω				0.0045	0.1401
				$(0.0004, 0.0101)$	$(0.0625, 0.2313)$
a				0.0319	0.1133
				$(0.0121, 0.0537)$	$(0.0661, 0.1698)$
b				0.9537	0.5741
				$(0.9158, 0.9861)$	$(0.3468, 0.7791)$

This table reports posterior means and 0.95 probability intervals for the stable model, the best HAR model M_5 given 1 change point, and the HAR-GARCH model with 0 and 1 change point. The first and second regime are denoted by $s_t = 1, 2$ respectively for the HAR and $\gamma_1 = 1$, and γ_2 for the HAR-GARCH.
The HAR model is

$$v_t = \beta_0 + \beta_1 v_{t-1} + \beta_2 v_{t-5,t-1} + \beta_3 v_{t-22,t-1} + \beta_J J_{t-1} + \beta_{A1} \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} + \beta_{A2} \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} I\left(r_{t-1} < 0\right) + \epsilon_t$$

where $\epsilon_t \sim NID(0, \sigma^2)$. While the HAR-GARCH model is identical except

$$\varepsilon_t = \gamma_{s_t} \sigma_t u_t$$

where $u_t \sim NID(0, 1)$, $\gamma_1 = 1$, $\gamma_k > 0$, for $k \geq 2$, and

$$\sigma_t^2 = \omega + a(v_{t-1} - E_{t-2} v_{t-1})^2 + b\sigma_{t-1}^2.$$

Table 10: Jumps and Asymmetric terms

# CP	No Jump, No Asym	with Jumps	with Asym	with Alt-Asym	with Jump and Aysm
0	-2405.23	-2401.42	-2335.59	-2342.47	-2333.28
1	**-2353.80**	**-2351.50**	**-2276.00**	**-2283.09**	**-2275.29**
2	-2355.84	-2354.33	-2279.34	-2286.50	-2278.68
3	-2359.47	-2357.41	-2284.06	-2290.90	-2282.83
4	-2363.14	-2363.80	-2289.34	-2297.26	-2289.00
5	-2369.90	-2369.60	-2295.86	-2303.01	-2294.86
6	-2375.27	-2375.54	-2300.57	-2309.91	-2301.71
7	-2381.27	-2379.44	-2306.60	-2314.50	-2308.00

This table compares the log marginal likelihood for M_5 for the HAR model without jumps and asymmetric (Asym) terms (column 2, "No Jump, No Asym" Column), with jumps (column 3), with asymmetric terms (column 4), with an alternative asymmetric term (column 5), and with both jumps and asymmetric terms (column 6).

Table 11: Robustness for Minimum Regime Duration

Minimum Duration	10 obs (2 weeks)	22 obs (1 month)	66 obs (3 months)	132 obs (6 months)
# CP				
0	-2333.28	-2333.28	-2333.28	-2333.28
1	**-2275.29**	**-2275.29**	**-2275.29**	**-2275.30**
2	-2278.67	-2278.67	-2278.68	-2278.67
3	-2283.38	-2283.35	-2282.83	-2282.87

This table compares the log marginal likelihood for different minimum regime durations for the model M_5. The first row reports the lower bound for the number of observations in each regime.

Table 12: Comparison of the HAR M_5 and the HAR-GARCH Model

# of CP	HAR	HAR-GARCH
0	-2333.28	-2276.10
1	$\mathbf{-2275.29}$	$\mathbf{-2253.38}$
2	-2278.68	-2253.56
3	-2282.83	-2255.69

This table compares the log marginal likelihoods for the stable HAR M_5 model and the stable HAR-GARCH model as well as their structural break versions with 1, 2 and 3 change points.

A. Time Series of log Realized Volatility

B. Posterior Density of Change Point

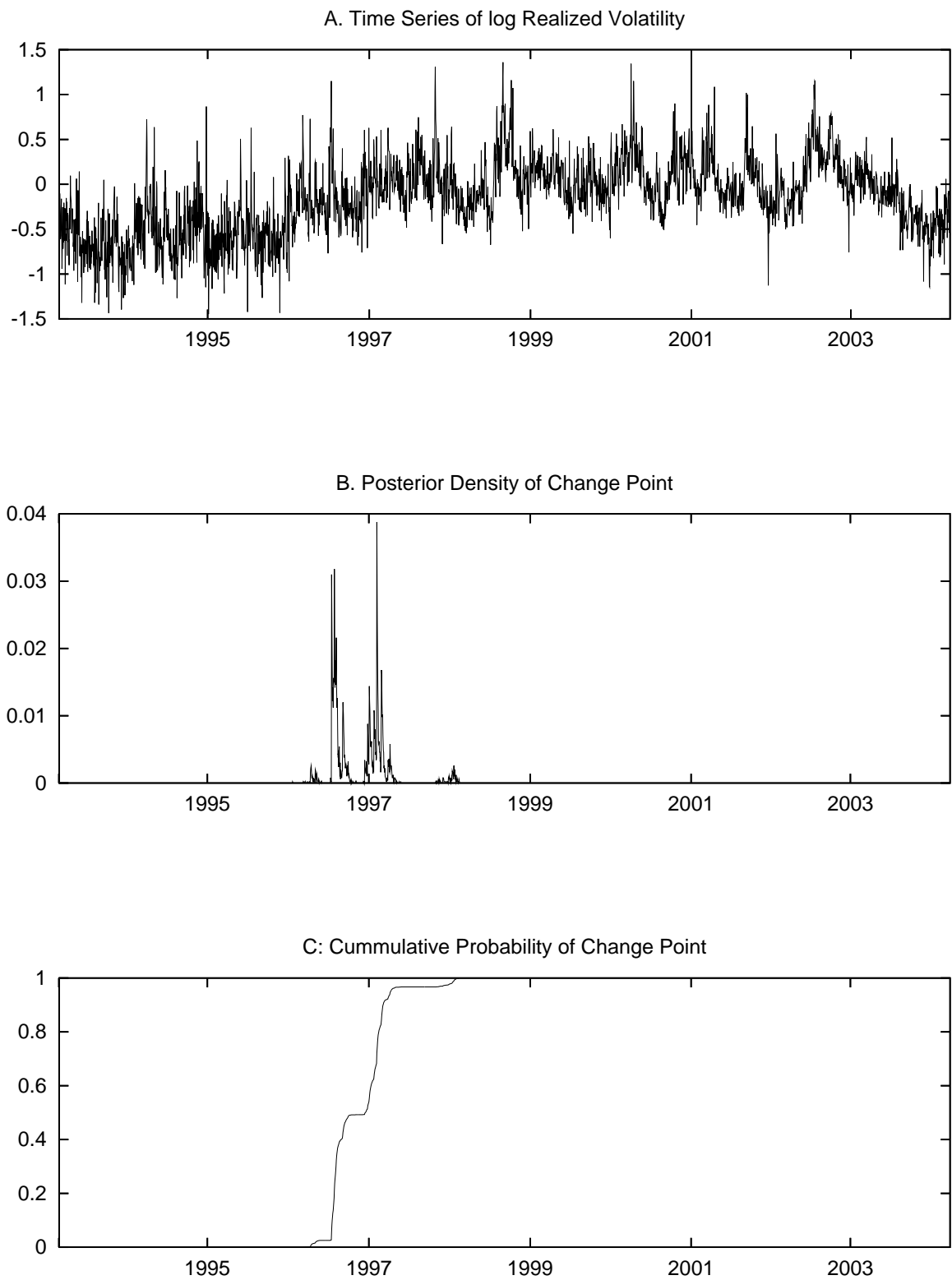C: Cummulative Probability of Change Point
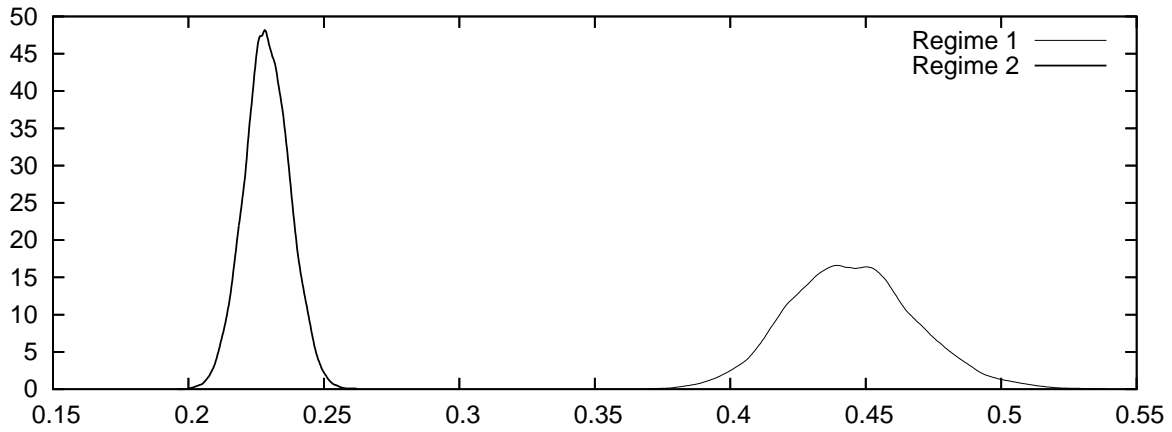
Figure 1: Change Point Date for M_5

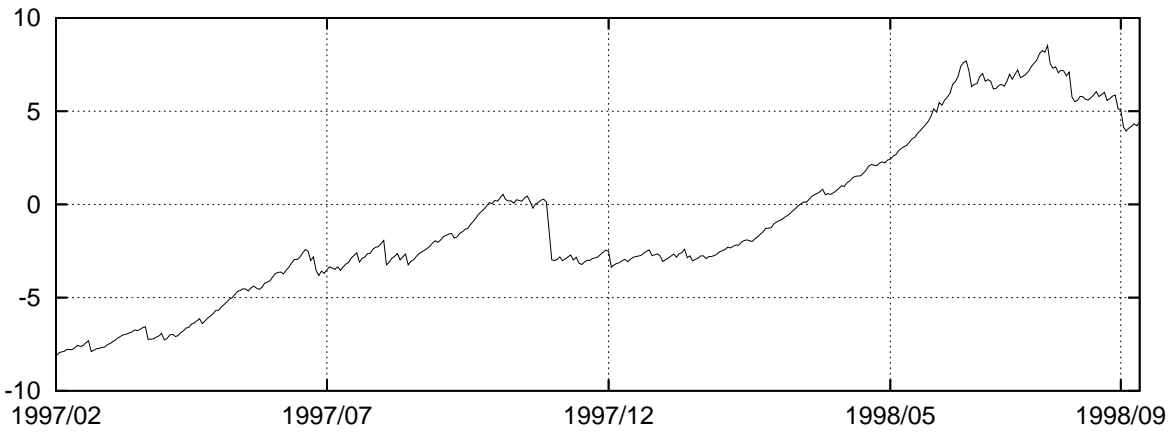Figure 2: Marginal Posterior Densities for σ^2

Figure 3: Recursive Log-Bayes Factor for Structural Break Model M_5 vs Stable Model

40