



BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2009-31

Real Time Detection of Structural Breaks in GARCH Models

by Zhongfang He and John M. Maheu

Bank of Canada Working Paper 2009-31

November 2009

Real Time Detection of Structural Breaks in GARCH Models

by

Zhongfang He¹ and John M. Maheu²

¹Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
zhe@bankofcanada.ca

²Department of Economics
University of Toronto and RCEA
jmaheu@chass.utoronto.ca

Bank of Canada working papers are theoretical or empirical works-in-progress on subjects in economics and finance. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada.

Abstract

A sequential Monte Carlo method for estimating GARCH models subject to an unknown number of structural breaks is proposed. Particle filtering techniques allow for fast and efficient updates of posterior quantities and forecasts in real time. The method conveniently deals with the path dependence problem that arises in these type of models. The performance of the method is shown to work well using simulated data. Applied to daily NASDAQ returns, the evidence favors a partial structural break specification in which only the intercept of the conditional variance equation has breaks compared to the full structural break specification in which all parameters are subject to change. The empirical application underscores the importance of model assumptions when investigating breaks. A model with normal return innovations results in strong evidence of breaks; while more flexible return distributions such as t-innovations or a GARCH-jump mixture model still favors breaks but indicates much more uncertainty regarding the time and impact of them.

JEL classification: C11, C15, C22, C53

Bank classification: Econometric and statistical methods; Financial markets

Résumé

Pour estimer les modèles GARCH susceptibles de compter un nombre indéterminé de ruptures structurelles, les auteurs proposent une méthode séquentielle de Monte-Carlo. Celle-ci fait appel à des techniques de filtrage particulière qui permettent l'actualisation rapide et efficace de valeurs postérieures et de prévisions en temps réel. Cette méthode apporte une solution commode au problème de la dépendance du sentier, présent dans les modèles GARCH. Elle fonctionne bien lorsque les données utilisées sont issues de simulations. Appliquée aux rendements quotidiens des titres du NASDAQ, elle fournit des résultats qui avantagent plus une spécification où les ruptures ne concernent que l'ordonnée de l'équation de la variance conditionnelle qu'une spécification dans laquelle l'ensemble des paramètres sont variables. Cette application empirique fait ressortir l'importance revêtue par les hypothèses de modélisation pour l'étude des ruptures. Dans un modèle où les rendements sont soumis à des chocs distribués selon une loi normale, les données obtenues confirment l'existence de ruptures. Ces dernières se retrouvent également dans les distributions plus souples, c'est-à-dire en présence de chocs distribués selon une loi de Student, ou dans un modèle GARCH mixte avec processus de saut; seulement, la date et l'incidence des ruptures font alors l'objet d'une incertitude bien plus grande.

Classification JEL : C11, C15, C22, C53

Classification de la Banque : Méthodes économétriques et statistiques; Marchés financiers

1 Introduction

This paper addresses the econometric challenges of estimating GARCH models subject to structural breaks by using a Bayesian sequential Monte Carlo approach. We present a particle filtering method to sequential estimation that builds on the change-point model of Chib (1998). The approach allows the number of breaks as well as model parameters to be estimated jointly in one run and can conveniently handle models with path dependence.

A particle filter is a class of sequential Monte Carlo filtering methods that approximate the posterior distribution of the latent state variables by a set of *particles* and associated probability weights. One approach to learning about fixed parameters is the mixture kernel smoothing method of Liu and West (2001). This has been successfully applied in Markov switching stochastic volatility (Carvalho and Lopes (2007), Casarin and Trecroci (2006)) and stochastic volatility with jumps in a continuous time setting (Raggi and Bordignon (2008)). Other work with sequential Monte Carlo filtering methods include Johannes and Polson (2006) and Polson, Stroud, and Muller (2008).

A challenging problem for MCMC analysis of structural breaks is how to handle models with path dependence, e.g. GARCH models. For example, a change in a parameter of the conditional variance at time t will affect all future conditional variances (Gray 1996). Due to this path dependence the dimension of the state space grows over time and becomes computationally challenging for typical datasets used in finance. The problem is more severe in the context of Markov switching. Bauwens, Preminger, and Rombouts (2007) develop a single-move Gibbs sampler for a Markov switching GARCH model with a fixed number of regimes. Our algorithm can jointly estimate the model parameters and the number of structural breaks at each point in time based on a single run of the particle filter algorithm making it computationally efficient.

Chopin (2007) proposes a particle filtering algorithm for estimating structural break models in which the fixed model parameters are formulated as part of the state variables. Local MCMC sampling based on Gilks and Berzuini (2001) is used to reduce degeneracy. Our approach differs in that we use Chib's formulation of structural breaks and the fixed

parameters are treated separately from state variables. This has the advantage that it can easily incorporate the case where only a subset of parameters have breaks while in the Chopin approach, it is not straightforward, if not impossible, due to the path dependence problem that occurs in local MCMC moves. In contrast, we are able to compare GARCH specifications that allow breaks in all parameters with versions that allow breaks only to the intercept of the conditional variance. It is this latter specification that isolates breaks in the long-run variance and has empirical support (Starica and Granger 2005).

The proposed algorithm is applied to structural break GARCH (SB-GARCH) models. We investigate 4 specifications, namely, a partial SB-GARCH model in which only the intercept of the volatility equation has breaks with normal and t return innovations, and a full SB-GARCH models in which all parameters are subject to breaks with normal and t return innovations. Based on simulated data, we find that the algorithm performs well in estimating the number and location of breaks as well as the fixed parameters. Furthermore, the marginal likelihoods, computed as a byproduct of the particle filter, can be used to identify the true model specification correctly. As an empirical application, we analyze the daily NASDAQ returns from January 3, 1995 to December 29, 2006. Based on marginal likelihoods, a partial SB-GARCH model with t -innovations has the highest cumulative predictive power. The structural break GARCH model with student- t return innovations outperforms the normal specification and indicates much more uncertainty regarding the time and impact of breaks on the model.

We find it important to use a flexible model for daily returns; failure to do so may result in the false identification of structural change. Our benchmark GARCH model with normal innovations identifies two breaks in late 1997 and early 2004, which are associated with changes in the long-run variance of returns. According to the filter estimates of the states the evidence of structural change with normal return innovations is largely removed once we consider t -innovations or a GARCH-jump mixture model. Nevertheless, Bayes factors favor the structural break model even with the large amount of uncertainty about breaks.

The paper is organized as follows. Section 2 presents the general structural break model we consider. Section 3 provides a brief review of particle filtering methods and the algorithm for the general structural break model. In Section 4, we detail the SB-GARCH models we estimate. Section 5 gives the simulation results of the algorithm and Section 6 is the empirical application on daily NASDAQ returns. The conclusions are presented in Section 7. Additional results can be found in the working paper version, He and Maheu (2008).

2 A General Structural Break Model

We consider a general structure break model in which the observation y_t is drawn from an analytically tractable density $p(y_t|D_{t-1}, \theta_t)$, where D_{t-1} is the set of information known at time $t - 1$ and θ_t is the set of parameters. The changing parameters θ_t are assumed to be driven by an unobservable state variable s_t such that

$$\theta_t = \theta_k \quad \text{when } s_t = k, \quad k \in \{1, 2, \dots, \bar{K}\}, \quad (1)$$

where θ_k is the parameter value in state k , and \bar{K} is a preset value for the maximum number of states. Some of the elements of θ_t may be constant across states. The state variable s_t is modeled as a Markov chain with the transition probability matrix

$$P = \begin{pmatrix} p_{11} & 1 - p_{11} & 0 & \dots & 0 \\ 0 & p_{22} & 1 - p_{22} & \dots & 0 \\ & & & \dots & \\ 0 & 0 & \dots & p_{\bar{K}-1, \bar{K}-1} & 1 - p_{\bar{K}-1, \bar{K}-1} \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}. \quad (2)$$

This specification imposes many 0 restrictions on the more general first-order Markov chain with \bar{K} states. A structural break at time t occurs when $s_t \neq s_{t-1}$. This formulation of structural breaks was originally proposed by Chib (1998) and is used extensively in

subsequent papers in the literature, e.g. Pastor and Stambaugh (2001), Kim, Morley, and Nelson (2005) and Liu and Maheu (2008). It has two major benefits. First, it automatically enforces an ordering of break points in the series y_t . Moreover, when viewed as a hidden Markov model (HMM), it facilitates the marriage with the existing large literature on HMM's and hence the development of efficient estimation methods (Scott (2002)). The regime-switching model of Hamilton (1988) can be viewed as a special case of this setup if identical states are assumed to recur (Pesaran, Pettenuzzo, and Timmermann (2006)).

Chib (1998) has developed a general MCMC algorithm for structural break models with a fixed number of states. But there are important limitations to this approach. As a smoothing algorithm, estimates of state variables rely on the information from the entire sample. It is not computationally feasible to update estimates as each new observation arrives in high frequency since the whole algorithm has to be re-run to incorporate this new information. It is inconvenient to determine the number of states via the existing MCMC methods. The usual practice is to run the algorithm repeatedly conditional on a fixed number \bar{K} of states specified *a priori* and calculate their marginal likelihoods for several \bar{K} . The number of states is then determined by comparing these marginal likelihoods. Computing the marginal likelihood is unfortunately often a complicated issue in the MCMC context. The whole estimation process can be time-consuming and impractical in real applications where inference needs to be updated frequently. In the case of models with path dependence such as the structural break GARCH model, which we discuss next, the Chib algorithm is not directly applicable.

2.1 Structural Breaks and GARCH

In this section we review some of the computational issues in estimating structural breaks in GARCH specifications and the resulting path dependence. The GARCH model of Engle (1982) and Bollerslev (1986) has been widely used in practice for estimating the

volatility of asset returns. One typical form is

$$y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim i.i.d.(0, 1), \quad \sigma_t^2 = c + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (3)$$

where y_t is the return of some financial asset. The sum $\alpha + \beta$ measures the persistence of the volatility process. A common finding in the literature is that estimates of this sum tend to be close to one, indicating that volatility is highly persistent. However, it has been argued that this high persistence may be due to structural breaks in the volatility process, e.g. see Diebold (1986), Lamoureux and Lastrapes (1990) and Mikosch and Starica (2004), and the omission of possible shifts in parameters would bias upward the estimate of persistence parameters and impair volatility forecasting (see Hamilton and Susmel (1994), Gray (1996) and Klaassen (2002), among others).

Mikosch and Starica (2004) showed theoretically that structural breaks in the *unconditional variance* of the GARCH volatility process could cause spuriously high persistence estimates. This motivates specifying a partial structural break GARCH (SB-GARCH) model

$$y_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = c_{s_t} + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (4)$$

where s_t is the unobserved state variable governed by (2) and is independent of ϵ_t . The unconditional variance in a regime $c_{s_t}/(1 - \alpha - \beta)$, can change over time.

Estimating SB-GARCH models is a challenging problem since the likelihood of y_t depends on the entire sequence of past states up to time t due to the recursive structure of its volatility. For example, consider the evolution of the conditional variance given a start-up value of σ_1^2 and $s_1 = 1$. At $t = 2$ there can be a break with $s_2 = 2$ and $\sigma_2^2 = c_2 + \alpha y_1^2 + \beta \sigma_1^2$ or no break with $s_2 = 1$, $\sigma_2^2 = c_1 + \alpha y_1^2 + \beta \sigma_1^2$. It is clear that the variance is a function of s_2 so we denote it as $\sigma_2^2(s_2)$. Now at time $t = 3$ there is again the possibility of a break or no break. In this case the conditional variance is a function of s_2 and s_3 , i.e. $\sigma_3^2(s_2, s_3)$. In general, the likelihood at time t is a function of the entire history

of states $\{s_2, s_3, \dots, s_t\}$. Efficient MCMC methods assume a first-order Markov chain for the discrete state (Chib 1998) to implement the forward-backward smoother to draw the states. Therefore, $\{s_2, s_3, \dots, s_t\}$ must be recast to a first-order Markov chain whose dimension is increasing in t making the Chib (1998) approach infeasible. In addition, the particle filter approach of Chopin and Pelgrin (2004) which is based on forward-backward smoother is also not applicable due to the path dependence problem. The path dependence in the conditional variance also occurs when α and/or β change from a break.

To circumvent the path dependence problem in the context of Markov switching GARCH, a variety of alternative tractable approximations have been proposed in the literature, e.g. Gray (1996), Klaassen (2002) and Haas, Mittnik, and Paoletta (2004). Bauwens, Preminger, and Rombouts (2007) develops a single-move MCMC algorithm that could be adapted to estimate SB-GARCH models with a known fixed number of states. However, this would not provide real time estimates, nor is it feasible to estimate SB-GARCH models with an unknown number of states via existing MCMC methods. As far as we know, no methods for computing marginal likelihoods of this class of models are available. So it is effectively infeasible to estimate this class of models via existing MCMC methods unless one is willing to assume that the number of break points is known a priori.

By focusing on the sequential filtering problem rather than the smoothing problem (MCMC), the path dependence that structural breaks induce in GARCH models is eliminated since only the one-step-ahead predictive distribution is needed in computation, which is an integral over two possible states conditional on parameters in the proposed structural break model.

3 Particle Filter

The foundational particle filtering algorithm is proposed by Gordon, Salmond, and Smith (1993). For detailed discussions of the particle filter, see the books edited by Doucet, de Freitas, and Gordon (2001) and Ristic, Arulampalam, and Gordon (2004). Roughly

speaking, the particle filter is a class of sequential Monte Carlo filtering methods which approximate the posterior distribution of the state variables, $p(s_t|D_t)$, by a set of *particles* $\{s_t^{(i)}\}_{i=1}^N$, with probability weights $\{w_t^{(i)}\}_{i=1}^N$, where N is the number of particles and $\sum_{i=1}^N w_t^{(i)} = 1$. This relation is conventionally denoted as $\{s_t^{(i)}, w_t^{(i)}\}_{i=1}^N \sim p(s_t|D_t)$. Any fixed model parameters θ , are assumed to be known and suppressed in the following.

Given a set of particles and weights, the posterior mean of any function of the state variable $f(s_t)$ can be directly estimated as

$$E[f(s_t)|D_t] \approx \sum_{i=1}^N f(s_t^{(i)})w_t^{(i)}.$$

The predictive density is approximated as

$$p(s_{t+1}|D_t) = \int p(s_{t+1}|s_t)p(s_t|D_t)ds_t \approx \sum_{i=1}^N p(s_{t+1}|s_t^{(i)})w_t^{(i)}, \quad (5)$$

and the filtering density is approximated as

$$p(s_{t+1}|D_{t+1}) \propto p(y_{t+1}|s_{t+1})p(s_{t+1}|D_t) \approx p(y_{t+1}|s_{t+1}) \sum_{i=1}^N p(s_{t+1}|s_t^{(i)})w_t^{(i)}. \quad (6)$$

The centerpiece of a particle filter algorithm is how to propagate particles forward from $\{s_t^{(i)}, w_t^{(i)}\}_{i=1}^N$ to $\{s_{t+1}^{(i)}, w_{t+1}^{(i)}\}_{i=1}^N$. This step however is often vulnerable to the weight degeneracy problem, that is, only a small subset of the particles is assigned appreciable weights in the propagation stage and hence the effective size of particles is reduced, which leads to greater approximation errors. A variety of particle filter algorithms have been proposed to reduce the weight degeneracy problem, e.g. see Kitagawa (1996), Carpenter, Clifford, and Fearnhead (1999) and Johannes and Polson (2006). In this paper, we focus on the auxiliary particle filter (APF) developed in Pitt and Shephard (1999) which is widely used and reliable. By giving more importance to particles with large predictive values, the APF improves sampling efficiency while significantly reducing weight degeneracy problems.

3.1 Particle Filter with Unknown Parameters

We follow Liu and West (2001) to incorporate parameter learning which has been successfully used in several papers including Carvalho and Lopes (2007), Casarin and Trecroci (2006) and Raggi and Bordignon (2008). The posterior density $p(\theta|D_t)$ is approximated by a mixture of multivariate normals

$$p(\theta|D_t) \approx \sum_{i=1}^N w_t^{(i)} N(\theta|a\theta_t^{(i)} + (1-a)\bar{\theta}_t, b^2V_t),$$

where $\bar{\theta}_t = \sum_{i=1}^N w_t^{(i)} \theta_t^{(i)}$ and $V_t = \sum_{i=1}^N w_t^{(i)} (\theta_t^{(i)} - \bar{\theta}_t)(\theta_t^{(i)} - \bar{\theta}_t)'$, and $N(\theta|.,.)$ is the multivariate normal pdf. The constants a and b measure the extent of the shrinkage and are determined via a discount factor $\delta \in (0, 1)$ as $a = \sqrt{1 - b^2}$ and $b^2 = 1 - [(3\delta - 1)/2\delta]^2$.

Conditional on samples of θ drawn from the mixture, the usual particle filters can be applied to estimate the state variables. As shown in Liu and West (2001), the kernel smoothing approach combined with an efficient particle filter such as the APF produces efficient estimates. Casarin and Marin (2007) find this method to be the best among several other alternatives for the estimation of fixed parameters and states for a stochastic volatility model.

3.2 A Particle Filter for Structural Break Models

We combine the APF with the kernel smoothing approach to design a sequential Monte Carlo algorithm for the general structural break model. The state at time t , s_t , will equal the number of states that have actually appeared in the studied time series up to time t . In contrast to the MCMC method of Chib (1998), this approach does not enforce that all states be visited. The MCMC method in Chib (1998) samples the state variables backward from the end of sample and assumes that the final state is \bar{K} and the first state in the sample is 1 so that all states are visited. For the particle filter there can be redundant states and we follow Chopin and Pelgrin (2004) to specify an upper bound \bar{K} on the number of states to facilitate computation. As long as the upper bound

on the number of states is large enough we can jointly estimate the model parameters and the number of structural breaks at each point in time. Therefore, the problem of determining the number of states is automatically solved by sequentially learning about the state variables over time. We consider this a major benefit of sequential Monte Carlo methods for structural break models.

The parameters θ should be reparameterized to take values on the real line since they will be sampled through a mixture of normal kernels. For example, the probability p_{ij} can be reparameterized as $\log(\frac{p_{ij}}{1-p_{ij}})$. To simplify notation, we will use θ to denote the transformed parameters as well in the following discussion.

To initialize the algorithm, we need to specify an upper bound \overline{K} for the possible number of states in the studied dataset. The only requirement for \overline{K} is that it should be larger than the estimated number of states so that the number of states is freely determined. Let $\{\theta_{k,t}^{(i)}\}_{i=1}^N$, $k = 1, 2, \dots, \overline{K}$, denote the particles of the parameters in state k and let $\theta_t^{(i)} \equiv \{\theta_{1,t}^{(i)}, \dots, \theta_{\overline{K},t}^{(i)}\}$. Let $p_t^{(i)}$ denote the transition probability components of $\theta_t^{(i)}$. Note the subscript t indicates that the parameter values are learned at time t , and not necessarily time-varying. Initial particles and weights are set as $s_1^{(i)} = 1$, $w_1^{(i)} = \frac{1}{N}$, for $i = 1, 2, \dots, N$ and parameters are drawn from the prior. The prior $\{\theta_1^{(i)}\}_{i=1}^N$ will depend on the specific model.

Our initial algorithm which employed the standard multinomial resampling performed poorly in practice, delivering rather unstable estimates of the parameters and states over repeated runs. An efficient sampling scheme for drawing the auxiliary indices is necessary to reduce the Monte Carlo variation and stabilize the estimates.

To stabilize estimates over multiple runs we use stratified sampling (Carpenter, Clifford, and Fearnhead 1999). To produce a new sample of size m from a population $\{x_t\}_{t=1}^N$ with weights $\{w_t\}_{t=1}^N$, stratified sampling first produces stratified uniform random variables $\{u_t\}_{t=1}^m$ by drawing $u_t \sim U(\frac{t-1}{m}, \frac{t}{m})$ independently. From each of these draws x_t is selected based on multinomial sampling. This method is fast and effective in stabilizing estimates across different runs.

Given a set of particles and weights $\{\theta_t^{(i)}, s_t^{(i)}, w_t^{(i)}\}_{i=1}^N \sim p(s_t, \theta | D_t)$, the following is a general algorithm for structural break models.

1. For $i = 1, 2, \dots, N$, let r be the mode of $p(s_{t+1} | s_t^{(i)}, \theta_t^{(i)})$, i.e. the state with the highest probability in $p(s_{t+1} | s_t^{(i)}, \theta_t^{(i)})$. Compute the weights for the auxiliary index $\tilde{w}_{t+1}^{(i)} \propto p(y_{t+1} | D_t, a\theta_{r,t}^{(i)} + (1-a)\bar{\theta}_{r,t})w_t^{(i)}$, where $\bar{\theta}_{r,t} = \sum_{i=1}^N w_t^{(i)}\theta_{r,t}^{(i)}$.
2. Draw stratified uniform random variables $\{u_j\}_{j=1}^N$ with $u_j \sim U(\frac{j-1}{N}, \frac{j}{N})$.
3. Produce the auxiliary indices $\{r^i\}_{i=1}^N$, $r^i \in \{1, 2, \dots, N\}$, by retaining N^i copies of i , $i = 1, 2, \dots, N$, where N^i is the number of $\{u_j\}_{j=1}^N$ falling in the interval $(\sum_{j=0}^{i-1} \tilde{w}_j, \sum_{j=0}^i \tilde{w}_j]$ with $\tilde{w}_0 = 0$.
4. For $i = 1, 2, \dots, N$, sample $\theta_{t+1}^{(i)}$ from $N(m_t^{(r^i)}, b^2V_t)$, where $m_t^{(r^i)} = a\theta_t^{(r^i)} + (1-a)\bar{\theta}_t$, $V_t = \sum_{i=1}^N w_t^{(i)}(\theta_t^{(i)} - \bar{\theta}_t)(\theta_t^{(i)} - \bar{\theta}_t)'$.
5. For $i = 1, 2, \dots, N$, sample $s_{t+1}^{(i)} \in \{1, 2, \dots, \bar{K}\}$ from $p_{t+1}^{(i)}(s_{t+1} | s_t^{(r^i)})$.
6. Let $s = s_{t+1}^{(i)}$. Compute $w_{t+1}^{(i)} \propto p(y_{t+1} | D_t, \theta_{s,t+1}^{(i)}) / p(y_{t+1} | D_t, a\theta_{r,t}^{(r^i)} + (1-a)\bar{\theta}_{r,t})$.

This gives the sample $\{s_{t+1}^{(i)}, \theta_{t+1}^{(i)}, w_{t+1}^{(i)}\}_{i=1}^N \sim p(s_{t+1}, \theta | D_{t+1})$.

3.3 Predictive Likelihoods and Model Selection

The predictive likelihood $p(y_{t+1} | D_t)$ can be computed from

$$p(y_{t+1} | D_t) = \int \int \int p(y_{t+1} | s_{t+1}, \theta) p(s_{t+1} | s_t, \theta) p(s_t, \theta | D_t) ds_{t+1} ds_t d\theta. \quad (7)$$

Once a set of particles and weights $\{s_t^{(i)}, \theta_t^{(i)}, w_t^{(i)}\}_{i=1}^N \sim p(s_t, \theta | D_t)$ is available, one can compute the approximation

$$p(y_{t+1} | D_t) \approx \sum_{i=1}^N w_t^{(i)} \int p(y_{t+1} | s_{t+1}, \theta_t^{(i)}) p(s_{t+1} | s_t^{(i)}, \theta_t^{(i)}) ds_{t+1},$$

where

$$\int p(y_{t+1}|s_{t+1}, \theta_t^{(i)})p(s_{t+1}|s_t^{(i)}, \theta_t^{(i)})ds_{t+1} = p(y_{t+1}|s_{t+1} = s_t^{(i)}, \theta_t^{(i)})p(s_{t+1} = s_t^{(i)}|s_t^{(i)}, \theta_t^{(i)}) + p(y_{t+1}|s_{t+1} = s_t^{(i)} + 1, \theta_t^{(i)})p(s_{t+1} = s_t^{(i)} + 1|s_t^{(i)}, \theta_t^{(i)}). \quad (8)$$

If the information set is $D_t = \{y_1, y_2, \dots, y_t\}$, the marginal likelihoods can be computed sequentially via predictive likelihoods

$$p(y_1, \dots, y_t) = p(y_1) \prod_{\tau=2}^t p(y_\tau|y_1, \dots, y_{\tau-1}), \quad t = 2, \dots, T.$$

By construction, the marginal likelihood can be interpreted as a measure of the cumulative out-of-sample predictive power of the model under investigation. Sequential Bayes factors (the ratio of marginal likelihoods of two specifications) can be then used to conduct model selection. Given a model A with marginal likelihood $p(y_1, \dots, y_T|A)$, and model B with marginal likelihood $p(y_1, \dots, y_T|B)$, the Bayes factor in favor of model A versus model B is $BF_{AB} = \frac{p(y_1, \dots, y_T|A)}{p(y_1, \dots, y_T|B)}$. Kass and Raftery (1995) suggest interpreting the evidence for A as: not worth more than a bare mention if $0 \leq BF_{AB} < 3$; positive if $3 \leq BF_{AB} < 20$; strong if $20 \leq BF_{AB} < 150$; and very strong if $BF_{AB} \geq 150$. To produce real time estimates of the Bayes factor, two particle filters could be run in parallel for two different models.

4 Structural Break GARCH Models

As discussed above, the final state $s_T = k$, where $k \leq \bar{K}$, is determined by the data rather than being specified *a priori*. The state variable s_t evolves according to the transition probability matrix in Equation 2. The chain can either stay in the current state or go to the next one. The final state, k , is equal to the number of in-sample states and the number of break points is $k - 1$.

Let $\theta_k = [c_k, \alpha, \beta, P]$ be the model parameters in state k and $\theta = \{\theta_1, \dots, \theta_{\bar{K}}\}$. The

likelihood of y_t for the partial break model in Equation 4, when $\epsilon_t \sim NID(0, 1)$ is

$$p(y_t|D_{t-1}, s_t, \theta) = p(y_t|D_{t-1}, \theta_{s_t}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{y_t^2}{2\sigma_t^2}\right), \quad (9)$$

where $\sigma_t^2 = c_{s_t} + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$. We refer to this structural break specification as a partial SB-GARCH-N model since only the intercept of the conditional variance is subject to breaks.

Empirical studies often suggest fat tails in the distribution of asset returns, therefore, an alternative specification for the return innovation ϵ_t would be a student- t distribution with v degrees of freedom. Let SB-GARCH-t denote this model and if $\theta_k = [c_k, \alpha, \beta, v, P]$, then the data density of y_t becomes

$$p(y_t|D_{t-1}, s_t, \theta) = p(y_t|D_{t-1}, \theta_{s_t}) = \frac{\Gamma(\frac{v+1}{2})}{\sigma_t \Gamma(\frac{v}{2}) \sqrt{\pi v}} \left(1 + \frac{y_t^2}{v\sigma_t^2}\right)^{-\frac{v+1}{2}}, \quad (10)$$

where $\sigma_t^2 = c_{s_t} + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$.

It is possible that all parameters of the volatility process, not just the unconditional variance, may be subject to structural breaks. So we also consider a full SB-GARCH model (SB-GARCH-N and SB-GARCH-t)

$$y_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = c_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2, \quad (11)$$

where $\theta_k = [c_k, \alpha_k, \beta_k, P]$, or $\theta_k = [c_k, \alpha_k, \beta_k, v, P]$, depending on the specification of the innovation ϵ_t . The algorithm presented in Section 3.2 can be applied to estimate these models and compute their marginal likelihoods for model comparison purposes.

A possibility is that outliers in returns may be identified as permanent structural breaks. To investigate this we consider the following partial break GARCH-jump mixture

$$y_t = \sigma_t \epsilon_t + J_t \eta_t, \quad \sigma_t^2 = c_{s_t} + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (12)$$

$$J_t \sim B(q), \quad \eta_t \sim N(0, h). \quad (13)$$

$B(q)$ denotes a Bernoulli distribution where q is the probability of $J_t = 1$ and otherwise $J_t = 0$. $J_t, \epsilon_t \sim N(0, 1)$ and η_t , are independent. Estimation of the model (SB-GARCH-J) can be done by augmenting the latent states with the jump indicator J_t and applying the algorithm in Section 3 with straightforward extensions.

5 Simulation Evidence

To analyze the performance of the proposed algorithm, we simulated 3000 observations from the partial SB-GARCH-N model with 2 break points. The parameter values are as follows $c_1 = 0.2, c_2 = 0.6, c_3 = 0.1, \alpha = 0.1, \beta = 0.8$ with the break points at 1000 and 2000. A plot of the simulated data is presented in Figure 1.

The size of particles is set to be 300,000 and the upper bound of possible states $\bar{K} = 5$. Following Liu and West (2001), the discount factor is set as $\delta = 0.99$. In the empirical work we restrict $p_{ii} = p$ for parsimony. The priors are as follows $c_i \sim Ga(1, 0.2)$, $i = 1, \dots, \bar{K}$, $\alpha \sim Be(1, 8), \beta \sim Be(4, 1)$, $\log\left(\frac{p}{1-p}\right) \sim N(10, 1)$ and when using the student- t distribution for returns the prior on the degrees of freedom is set as $v \sim Ga(1.2, 30)$.

In our experiments, we find that a prior on the transition probability p close to one is critical for obtaining sensible results since estimates of states are sensitive to outliers. A false estimate of states usually leads to poor estimates of parameters. A prior for $\log\left(\frac{p}{1-p}\right)$ centered on lower values (i.e. 5) quickly results in all states being visited and being stuck in the final state \bar{K} for the rest of the sample. Therefore, to avoid state saturation it is necessary to have an informative prior. As such, our prior favors no breaks. On the other hand the priors for remaining parameters cover a wide range of values consistent with existing empirical studies on GARCH models.

All estimations in this paper are performed using the GNU scientific library in C. Each filtering iteration takes about 2-3 seconds. The estimates of states are presented in the middle panel of Figure 1. The lower panel of Figure 1 provides the distribution of the state estimates (the filter $P(s_t|D_t)$) over time, which shows clearly the evolution of the state particles. The algorithm is able to successfully identify the number and locations of

break points. There are a few spikes in the filter estimates but they are quickly corrected. Intuitively, these spikes are caused by outlier observations. The state particles in these "false" states are subsequently dominated by particles which stay in the "true" states as the algorithm updates information by using new observations and gives higher weights to these correct particles.

The full-sample parameter estimates are presented in Table 1. For all parameters, the posteriors collapse to their true values as time increases and the estimates of the parameters c_2 and c_3 , which are specific to state 2 and 3 respectively, change sharply upon the break points with their credible intervals quickly shrinking afterwards.

Table 1: Parameter Estimates for Simulated Data

| | True Value | Partial SB-GARCH-N | Partial SB-GARCH-t | Full SB-GARCH-N | GARCH-N |
|------------|------------|------------------------|---------------------------|------------------------|------------------------|
| c_1 | 0.2 | 0.167 (0.102,0.259) | 0.169 (0.103,0.262) | 0.229 (0.076,0.541) | 0.088 (0.064,0.117) |
| c_2 | 0.6 | 0.580 (0.333,0.946) | 0.573 (0.331,0.922) | 0.480 (0.242,0.852) | |
| c_3 | 0.1 | 0.098 (0.055,0.162) | 0.098 (0.055,0.159) | 0.160 (0.071,0.306) | |
| α_1 | 0.1 | 0.069 (0.039,0.112) | 0.066 (0.039,0.103) | 0.034 (0.006,0.108) | 0.098 (0.069,0.135) |
| α_2 | | | | 0.071 (0.036,0.124) | |
| α_3 | | | | 0.072 (0.029,0.145) | |
| β_1 | 0.8 | 0.834 (0.751,0.898) | 0.829 (0.748,0.893) | 0.821 (0.617,0.941) | 0.875 (0.834,0.908) |
| β_2 | | | | 0.846 (0.757,0.912) | |
| β_3 | | | | 0.755 (0.582,0.884) | |
| v | | | 42.734 (21.239,77.134) | | |
| <i>LML</i> | | -5463.023 | -5464.364 | -5466.228 | -5507.791 |

This table reports the full-sample posterior means of parameters for simulated data. Numbers in parenthesis are 95% credible sets. LML stands for log marginal likelihood.

We also investigated the effective sample size, computed as $1/\sum_{i=1}^N (w_t^{(i)})^2$, to check for weight degeneracy. There were sporadic drops in the effective sample size, usually

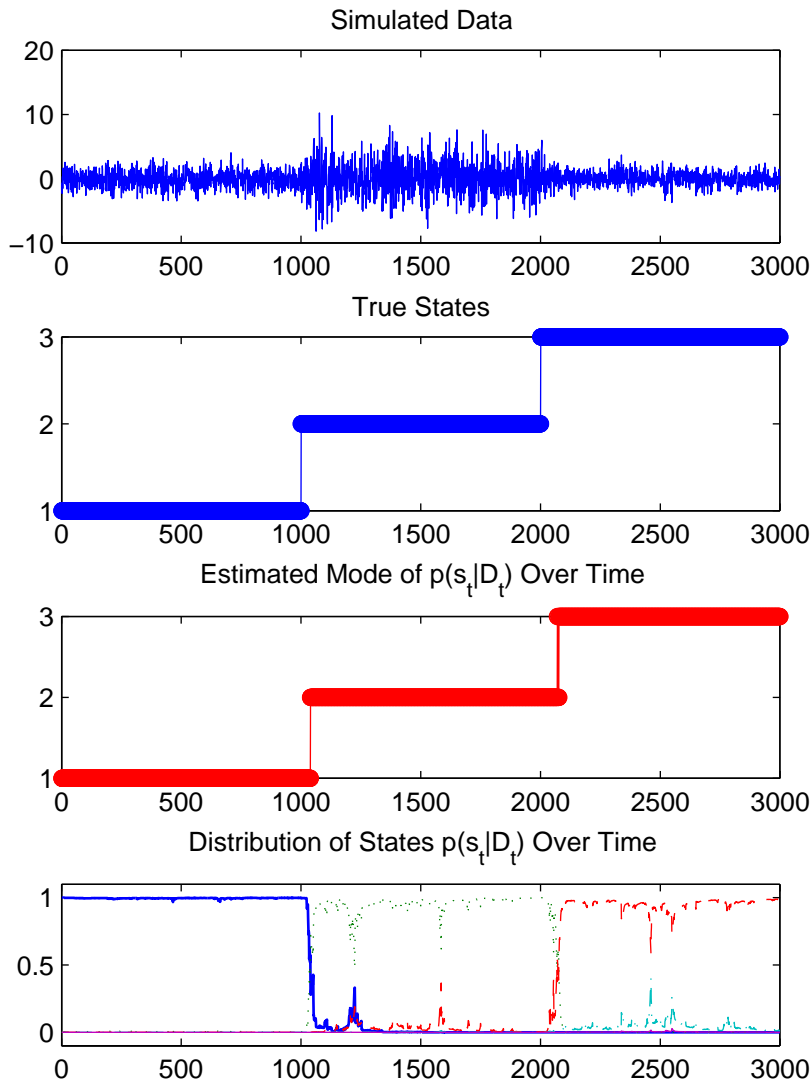


Figure 1: State Estimates of the Partial SB-GARCH-N Model, Simulated Data. For the distribution of states, bold solid: $p(s_t = 1|D_t)$, dot: $p(s_t = 2|D_t)$, dash: $p(s_t = 3|D_t)$, dash-dot: $p(s_t = 4|D_t)$, solid: $p(s_t = 5|D_t)$

around large outliers and break points, but they return to normal quickly afterwards, suggesting that weight degeneracy is minor in this study.

To see if the algorithm has the power to detect the true model when there are competing ones, we also estimate a partial SB-GARCH- t model, a full SB-GARCH-N model and a standard GARCH model with normal innovations for the simulated data. The results are presented in Table 1. The partial SB-GARCH-N model has the largest marginal likelihood, with the partial SB-GARCH- t model closely behind. The no-break GARCH model has a much smaller marginal likelihood than the structural break alternatives. Based on the time series of cumulative log-predictive likelihoods (not shown), the partial SB-GARCH-N model consistently outperforms the full SB-GARCH-N and no-break GARCH models whereas it does slightly better than the partial SB-GARCH- t model. In additional simulation experiments we found the correct model was identified by the largest marginal likelihood when the underlying DGP was a full SB-GARCH model.

6 Empirical Application

In this section, we apply the proposed algorithm to the daily NASDAQ composite returns from January 3, 1995 to December 29, 2006 (3022 observations). The data source is the Center for Research in Security Prices (CRSP). In estimation, returns are demeaned and scaled up by 100. The priors, number of particles, \bar{K} , a , b , and δ are the same as in the simulation experiments.

6.1 Estimation Results

To compare the performance of competing SB-GARCH models for the NASDAQ returns, we consider estimating 4 models: the partial SB-GARCH models with normal and t innovations and the full SB-GARCH models with normal and t innovations. To explore the importance of modeling structural breaks, we also fix $\bar{K} = 1$ and estimate two no-break GARCH models with normal and t innovations respectively. The estimated full-sample marginal likelihoods of the six models are presented in Table 2. Based on these

Table 2: Marginal Likelihoods

| | Partial | Partial | Full | Full |
|-----|------------|------------|------------|------------|
| | SB-GARCH-N | SB-GARCH-t | SB-GARCH-N | SB-GARCH-t |
| LML | -5231.25 | -5225.46 | -5228.40 | -5228.41 |
| | GARCH-N | GARCH-t | SB-GARCH-J | |
| LML | -5254.24 | -5234.92 | -5227.26 | |

This table reports the full-sample log marginal likelihoods for NASDAQ return series from Jan 03,1995 to Dec 29,2006. LML stands for log marginal likelihood.

estimates, the partial SB-GARCH-t model is the most favored while the versions in which all parameters change after a break have lower marginal likelihoods.

We first provide a comparison of the SB-GARCH-t models with the SB-GARCH-N models, which are the ones most commonly studied in the existing literature. The full-sample marginal likelihoods of Table 2 provide overwhelming evidence in favor of the specification with t-innovations: the t models have larger marginal likelihoods than the normal models in both the partial SB-GARCH and the no-break categories. Second each of the break models provides a large improvement over the no-break alternative. For instance, the log-Bayes factors are 22.99 (normal innovations) and 9.46 (t-innovations) in favor of the partial break model.

The sequentially filtered estimates of states of the partial SB-GARCH-N model are presented in Figures 2. This model identifies two break points associated with a sustained increase in NASDAQ volatility lasting from October 27, 1997 to January 23, 2004. The full-sample posterior means and 95% credible sets of parameter estimates are reported in Table 3. The resulting persistence estimate $\alpha + \beta = 0.966$ is lower than the estimate of 0.983 by the no-break GARCH model with normal innovations. The full SB-GARCH-N model has much smaller values of $\alpha + \beta$ in each regime. This finding is consistent with the existing studies of GARCH models that find lower persistence once structural breaks in the volatility process are taken into account.

The results are different when we fit the more flexible partial SB-GARCH-t model to the data. Figure 3 and 4 present the sequentially filtered estimates of states and parameters of this model. It identifies a similar pattern of breaks, but there is more

Table 3: Parameter Estimates for NASDAQ Returns

| | Partial SB-GARCH-N | Partial SB-GARCH-t | Full SB-GARCH-N | GARCH-N | GARCH-t | Partial SB-GARCH-J |
|------------|------------------------|---------------------------|------------------------|------------------------|---------------------------|------------------------|
| c_1 | 0.070 (0.038,0.113) | 0.043 (0.027,0.067) | 0.274 (0.062,0.796) | 0.052 (0.032,0.071) | 0.033 (0.023,0.0448) | 0.048 (0.022,0.082) |
| c_2 | 0.171 (0.090,0.303) | 0.211 (0.097,0.391) | 0.170 (0.100,0.273) | | | 0.231 (0.115,0.417) |
| c_3 | 0.055 (0.006,0.324) | 0.171 (0.006,0.947) | 0.073 (0.029,0.153) | | | |
| α_1 | 0.074 (0.054,0.098) | 0.062 (0.043,0.085) | 0.137 (0.034,0.341) | 0.095 (0.072,0.124) | 0.070 (0.044,0.088) | 0.064 (0.046,0.090) |
| α_2 | | | 0.087 (0.054,0.129) | | | |
| α_3 | | | 0.023 (0.007,0.054) | | | |
| β_1 | 0.892 (0.851,0.924) | 0.890 (0.848,0.924) | 0.632 (0.230,0.922) | 0.888 (0.859,0.912) | 0.909 (0.835,0.908) | 0.875 (0.846,0.919) |
| β_2 | | | 0.882 (0.832,0.921) | | | |
| β_3 | | | 0.872 (0.749,0.946) | | | |
| v | | 27.190 (16.021,43.337) | | | 19.268 (12.797,27.874) | |
| q | | | | | | 0.041 (0.012,0.101) |
| h | | | | | | 2.149 (0.695,5.135) |

This table reports the full-sample posterior means of parameters for NASDAQ return series from Jan 03,1995 to Dec 29,2006. Numbers in parenthesis are 95% credible sets.

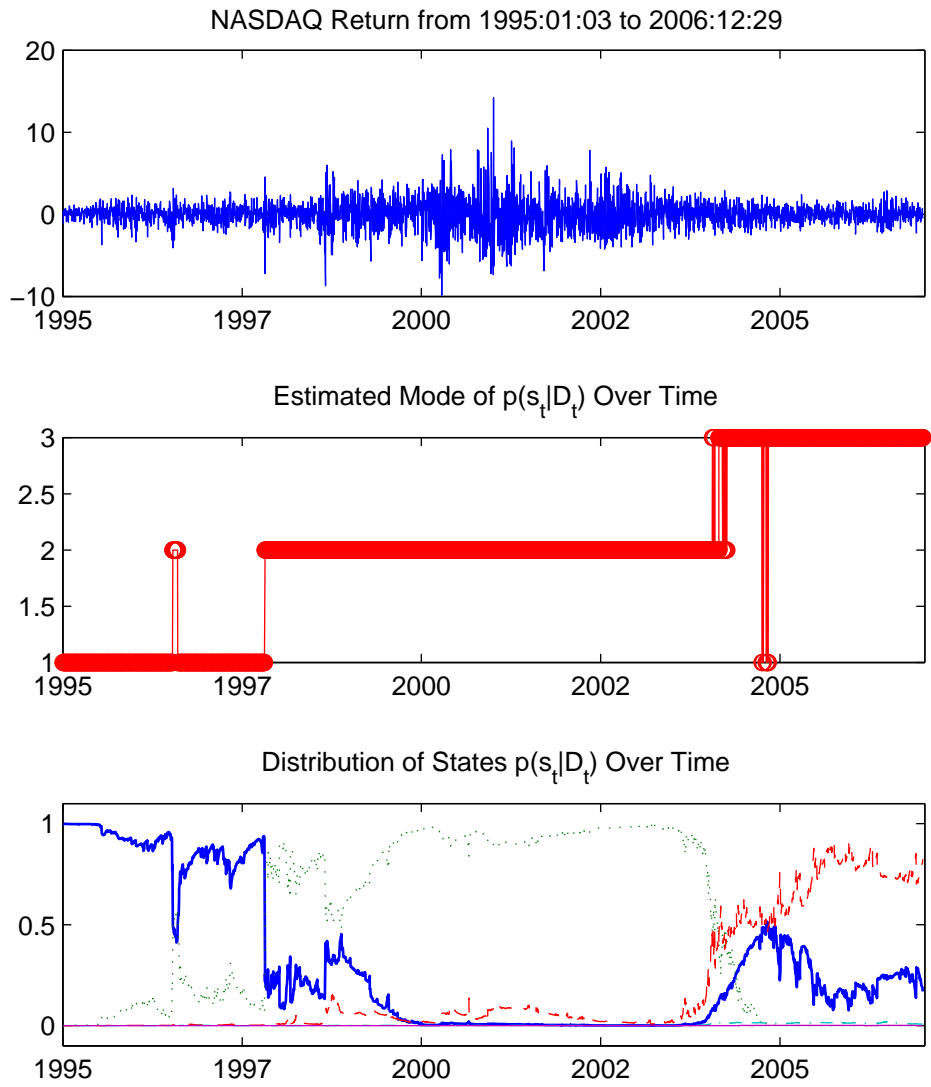


Figure 2: State Variable Estimates of the Partial SB-GARCH-N Model, NASDAQ Data. For the distribution of states, bold solid: $p(s_t = 1|D_t)$, dot: $p(s_t = 2|D_t)$, dash: $p(s_t = 3|D_t)$, dash-dot: $p(s_t = 4|D_t)$, solid: $p(s_t = 5|D_t)$

uncertainty. Recall that our analysis addresses the real time filtering problem, and as such smoothed historical estimates of breaks may be more precise. By the end of the sample there is substantial probability that no break has occurred. For example, at the end of the sample $p(s_T = 1|D_T) = 0.671$, $p(s_T = 2|D_T) = 0.0002$, and $p(s_T = 3|D_T) = 0.321$. Compared to the specification with normal innovations this model has a built in robustness from the fat-tailed t-innovations. A structural break can temporarily appear as tail observations and it may take many observations to disentangle these effects. This is consistent with Maheu and McCurdy (2009) who model the unconditional distribution of returns and find less breaks when fat-tailed innovations are used instead of normal innovations. Nevertheless, Bayes factors favor the structural break model even with the large amount of uncertainty about breaks.

Sequential parameter estimates appear in Figure 4. Early in the sample the degree of freedom parameter takes a large drop. The first break found in the SB-GARCH-N model (Oct. 27,1997) is classified as an increase in tail thickness by this model (v drops in Figure 4). It is not till closer to mid-sample that the evidence for a structural break increases.

In common with the other break models Table 3 shows that $\alpha + \beta$ is lower for the t-innovation specifications when breaks are allowed. We also find that by the end of the sample the degree of freedom parameter is larger, (27 versus 19 for the no break model) for the SB-GARCH-t model.

The filtered conditional standard deviations through time are displayed in Figure 5 for 2 alternative models. Although the volatility estimates are broadly similar, panel B which displays their relative difference, show that the no break GARCH-N tend to produce larger estimates and is persistently larger near the end of the sample. As was mentioned above, model estimates may attribute higher persistence to volatility when breaks are not modeled.

Figures 6 and 7 provide information on the predictive value of modeling breaks. The first figure displays the cumulative log-Bayes factor in favor of the partial SB-GARCH-

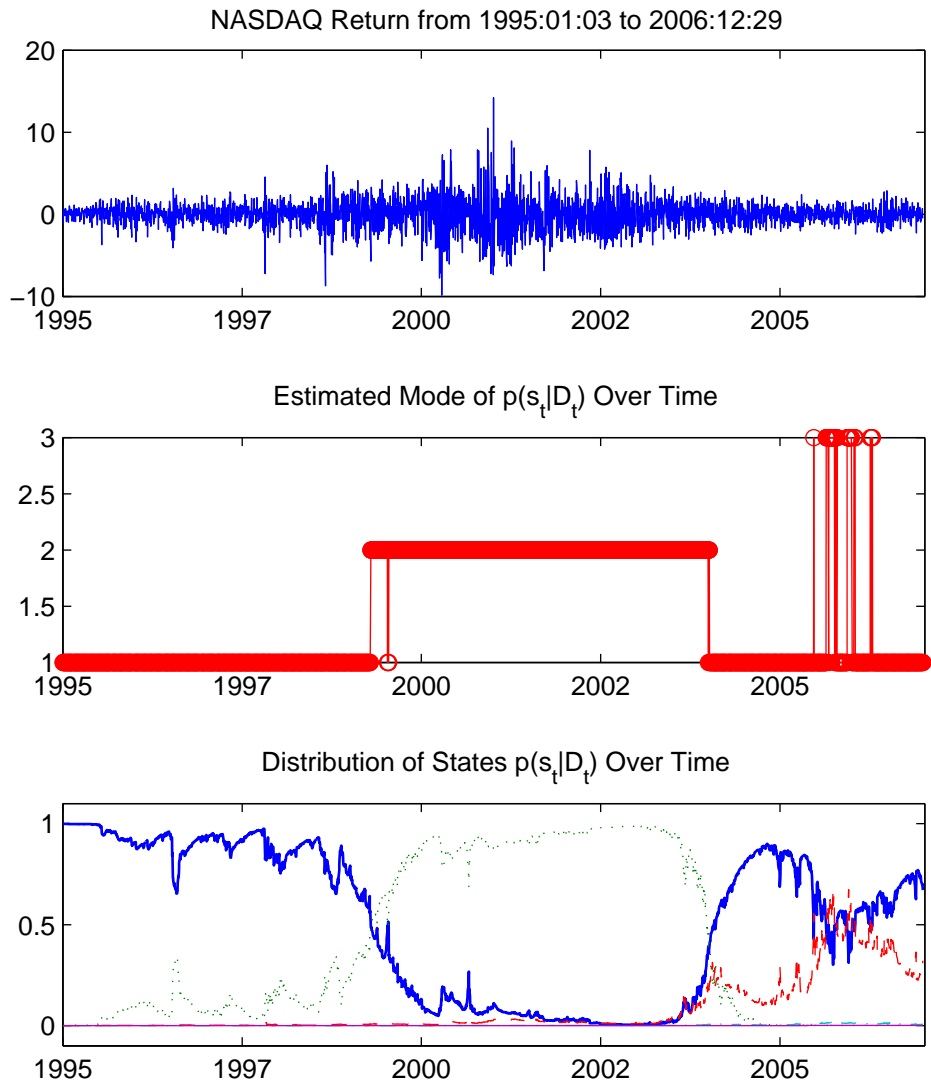


Figure 3: State Variable Estimates of the Partial SB-GARCH-t Model, NASDAQ Data. For the distribution of states, bold solid: $p(s_t = 1|D_t)$, dot: $p(s_t = 2|D_t)$, dash: $p(s_t = 3|D_t)$, dash-dot: $p(s_t = 4|D_t)$, solid: $p(s_t = 5|D_t)$

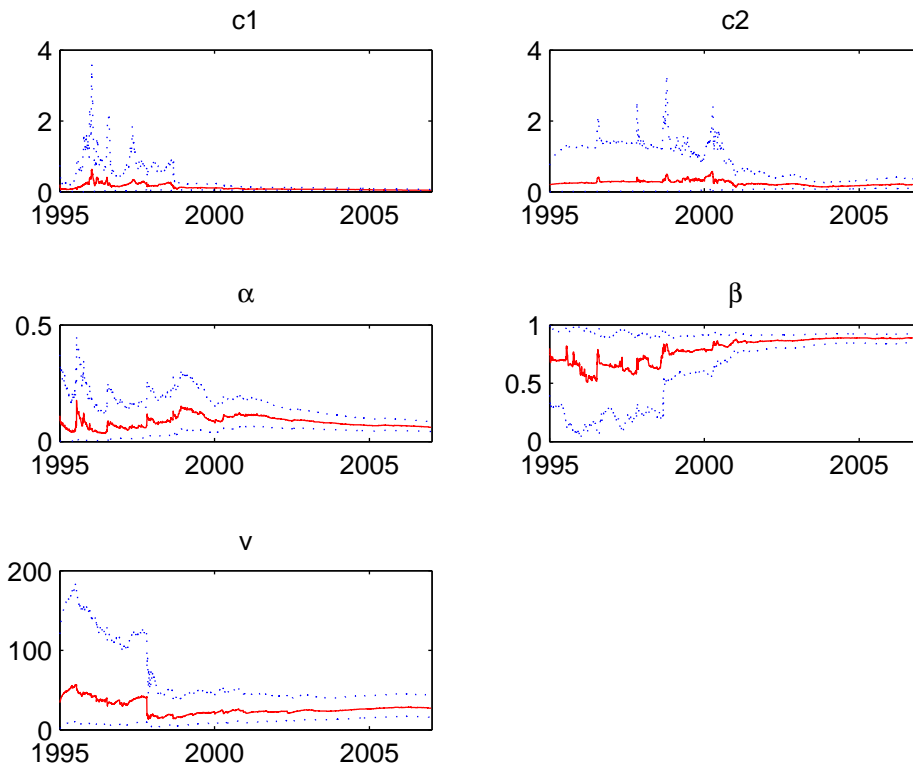


Figure 4: Parameters Estimates of the Partial SB-GARCH-t Model, NASDAQ Data

N against the no break alternative. There are some minor gains mid-sample but the real improvements occur at the latter part of the sample. Here the improvements in the predictive densities are consistent and ongoing. Similarly for the partial SB-GARCH-t model, the gains come at the end of the sample. In the middle of the sample there are some penalties incurred from the increased model complexity of the break specification. There is always a trade-off between modeling the structural change and the resulting increase in parameter uncertainty around break points.

6.2 Robustness

We found no evidence of weight degeneracy in our application (see He and Maheu (2008)). We experimented running the programs under different seeds of the random number generator to check the stability of estimates. The difference between parameter estimates across runs is generally less than 0.01 while the difference between estimated marginal

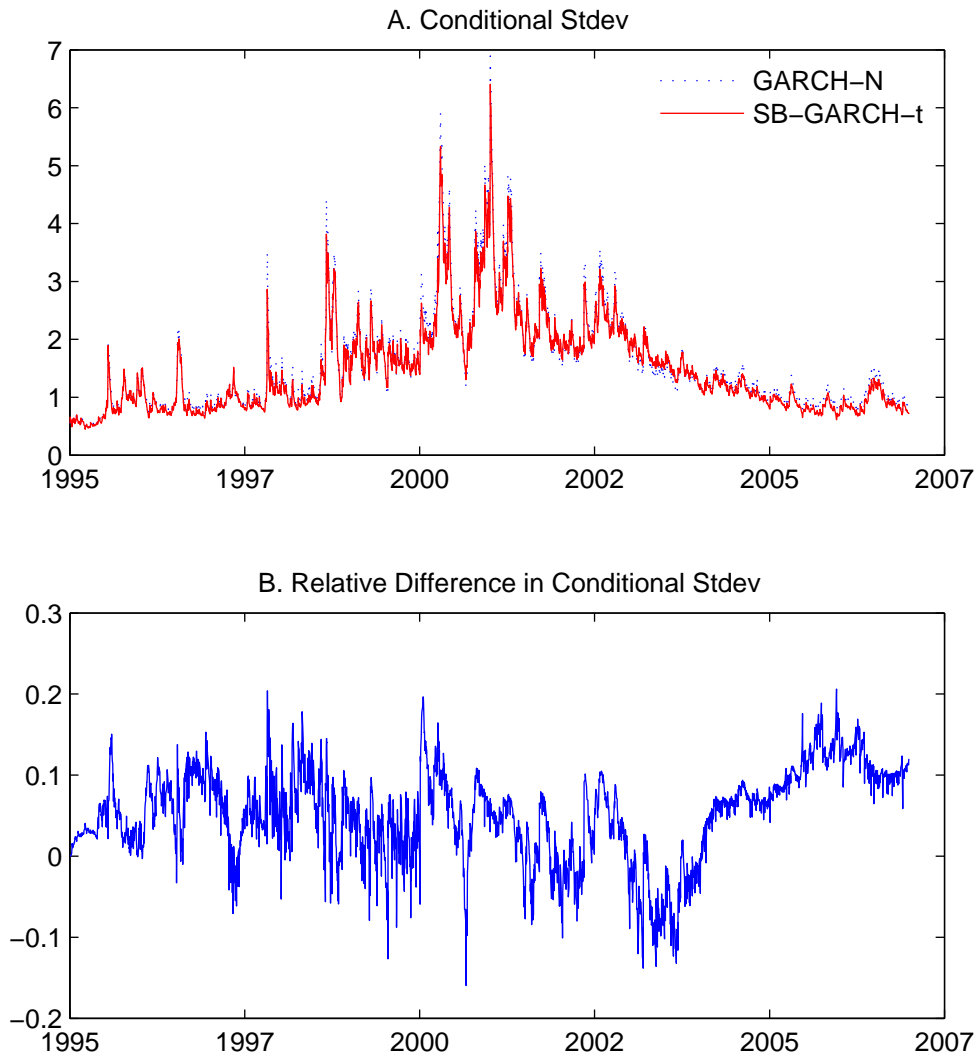


Figure 5: Conditional Standard Deviations: A. No-break GARCH-N v.s. partial SB-GARCH-t, and B. their relative difference, $(\sigma_{t,\text{GARCH-N}} - \sigma_{t,\text{SB-GARCH-t}}) / \sigma_{t,\text{GARCH-N}}$.

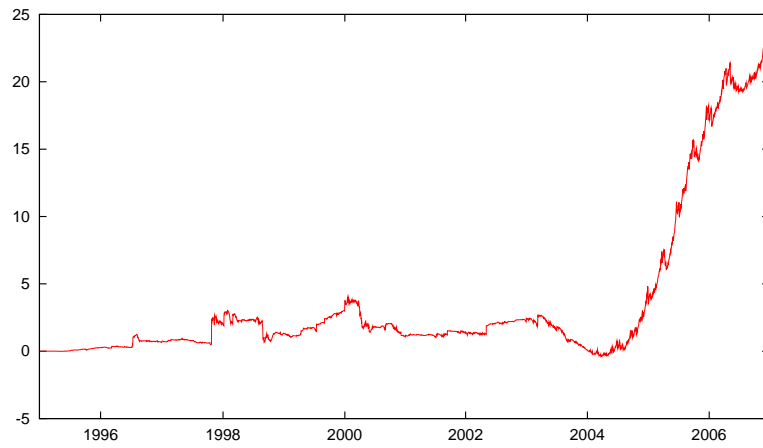


Figure 6: Cumulative Log-Bayes Factor: Partial SB-GARCH-N vs No Break GARCH-N

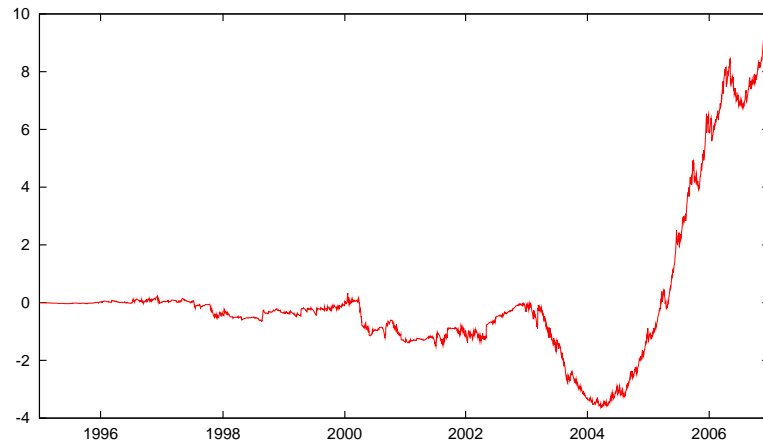


Figure 7: Cumulative Log-Bayes Factor: Partial SB-GARCH-t vs No Break GARCH-t

likelihoods is less than 1. We found that 100,000 or more particles produced reliable results while less than this could be unstable across repeated runs.

For the SB-GARCH-J model, assuming independent priors $q \sim Be(40, 2)$, $h \sim Ga(2, 1)$, while keeping priors of other parameters unchanged, we obtain similar estimates to the SB-GARCH-t model. Model estimates appear in Table 3 while state and jump estimates are plotted in Figure 8. In particular, the break points identified by the SB-GARCH-J model are close to those of the SB-GARCH-t model. The marginal likelihood of the SB-GARCH-J model is larger than that of the SB-GARCH-N model but slightly lower than the SB-GARCH-t model. The parameter estimates of the SB-GARCH-J model are similar to those of the SB-GARCH-N model. We conclude that it is important to use a flexible model for daily returns, failure to do so may result in false identification of structural change. Finally, we note that modeling jumps generally tends to reduce the effect of temporary outliers and hence enables lowering the strong prior on the transition probabilities of states. For example, using a prior of the transition probability $\log\left(\frac{p}{1-p}\right) \sim N(8.5, 1)$ produces similar results.

Table 4 reports the sensitivity of the marginal likelihood for the preferred partial SB-GARCH-t model for different priors. In each case, a change is made to the benchmark prior and this is listed in the first column of the table. Although there are some changes in parameter estimates and state inference, the improved predictions this model provides are robust to different priors.

Table 4: Marginal Likelihood Estimates for Different Priors: Partial SB-GARCH-t

| Prior | LML |
|---|----------|
| benchmark | -5225.46 |
| $c_i \sim Ga(1, 0.4)$ | -5227.18 |
| $\alpha \sim U(0, 1), \beta \sim U(0, 1)$ | -5226.46 |
| $\log(p_{ii}/(1 - p_{ii})) \sim N(8, 1)$ | -5220.04 |
| $\bar{K} = 7$ | -5225.91 |
| $\nu \sim Ga(1.2, 10)$ | -5224.41 |

This table reports the log-marginal likelihood (LML) for the partial SB-GARCH-t model based on the benchmark prior in Section 5 with any changes listed in the first column.

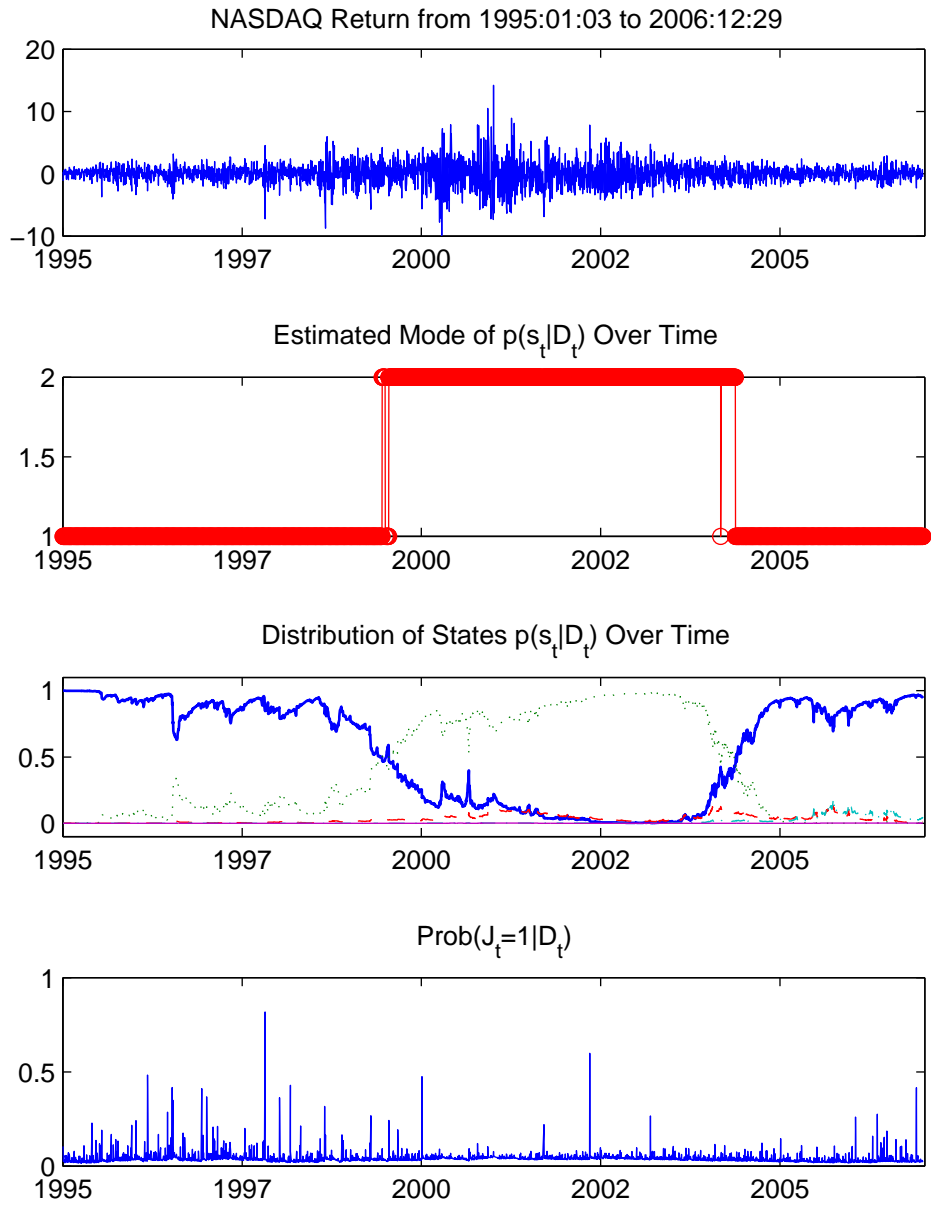


Figure 8: State Variable Estimates of the Partial SB-GARCH-Jump Mixture Model with NASDAQ Data. For the distribution of states, bold solid: $p(s_t = 1|D_t)$, dot: $p(s_t = 2|D_t)$, dash: $p(s_t = 3|D_t)$, dash-dot: $p(s_t = 4|D_t)$, solid: $p(s_t = 5|D_t)$

7 Conclusion

This paper proposes a sequential Monte Carlo filtering algorithm to estimate GARCH models subject to structural breaks. There are several notable features of the proposed algorithm: the number of breaks is estimated simultaneously with other model parameters and states in a single run; the estimates of parameters and states are fast and efficiently updated once new observations become available; and by focusing on the sequential filtering problem the path dependence that structural breaks induce in GARCH models does not cause any problems for estimation.

Simulation examples show that the algorithm is able to perform accurate sequential inference. Our empirical application underscores the importance of model assumptions when investigating breaks. A model with normal return innovations results in strong evidence of breaks; while more flexible return distributions such as t-innovations or adding jumps to the model still favors breaks but indicates much more uncertainty regarding the time and impact of them. We also find that the partial structural break specification delivers better performance than the full structural break specification in which all parameters change from a break.

Acknowledgements

We thank the co-editor Erricos John Kontogiorgos, an associate editor and two anonymous referees for many constructive comments. We are grateful for helpful comments from Victor Aguirregabiria, Chuan Goh, Tom McCurdy, Angelo Melino and seminar participants at the RCEA conferences on Bayesian Econometrics and Small Open Economies in a Globalized World. Maheu thanks SSHRC for financial support.

References

BAUWENS, L., A. PREMINGER, AND J. ROMBOUTS (2007): “Regime Switching GARCH Models,” CORE Discussion Paper 2007-55.

- BOLLERSLEV, T. (1986): “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31, 309–328.
- CARPENTER, J., P. CLIFFORD, AND P. FEARNHEAD (1999): “An improved particle filter for nonlinear problems,” *IEE Proceedings-Radar, Sonar and Navigation*, 146, 2–7.
- CARVALHO, C., AND H. LOPES (2007): “Simulation-based sequential analysis of Markov switching stochastic volatility models,” *Computational Statistics and Data Analysis*, 51, 4526–4542.
- CASARIN, R., AND J.-M. MARIN (2007): “Online data processing: comparison of Bayesian regularized particle filters,” University of Brescia, Department of Economics, working paper n. 0704.
- CASARIN, R., AND C. TRECROCI (2006): “Business Cycle and Stock Market Volatility: A Particle Filter Approach,” Cahier du CEREMADE N. 0610, University Paris Dauphine.
- CHIB, S. (1998): “Estimation and Comparison of Multiple Change Point Models,” *Journal of Econometrics*, 86, 221–241.
- CHOPIN, N. (2007): “Dynamic detection of change points in long time series,” *Annals of the Institute of Statistical Mathematics*, 59, 349–366.
- CHOPIN, N., AND F. PELGRIN (2004): “Bayesian inference and state number determination for hidden Markov models: an application of the information content of the yield curve about inflation,” *Journal of Econometrics*, 123(2), 327–344.
- DIEBOLD, F. (1986): “Modeling the persistence of conditional variances: a comment,” *Econometric Reviews*, 5, 51–56.
- DOUCET, A., N. DE FREITAS, AND N. GORDON (eds.) (2001): *Sequential Monte Carlo Methods in Practice*. Springer-Verlag.
- ENGLE, R. F. (1982): “Autoregressive Conditional Heteroskedasticity with Estimates of the UK inflation,” *Econometrica*, 50, 987–1008.

- GILKS, W. R., AND C. BERZUINI (2001): “Following a Moving Target-Monte Carlo Inference for Dynamic Bayesian Models,” *Journal of the Royal Statistical Society. Series B*, 63(1), 127–146.
- GORDON, N., D. SALMOND, AND A. SMITH (1993): “Novel approach to nonlinear/non-Gaussian Bayesian state estimation,” *IEE Proceedings*, F-140, 107–113.
- GRAY, S. F. (1996): “Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process,” *Journal of Financial Economics*, 42, 27–62.
- HAAS, M., S. MITTNIK, AND M. PAOLELLA (2004): “A New Approach to Markov Switching GARCH Models,” *Journal of Financial Econometrics*, 2, 493–530.
- HAMILTON, J. D. (1988): “Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates,” *Journal of Economic Dynamics and Control*, 12, 385–423.
- HAMILTON, J. D., AND R. SUSMEL (1994): “Autoregressive Conditional Heteroskedasticity and Changes in Regime,” *Journal of Econometrics*, 64, 307–333.
- HE, Z., AND J. M. MAHEU (2008): “Real Time Detection of Structural Breaks in GARCH Models,” University of Toronto, Department of Economics, Working Paper 336.
- JOHANNES, M., AND N. POLSON (2006): “Exact Bayesian particle filtering and parameter learning,” Working Paper, University of Chicago.
- KASS, R. E., AND A. E. RAFTERY (1995): “Bayes Factors,” *Journal of the American Statistical Association*, 90(420), 773–795.
- KIM, C., J. MORLEY, AND C. NELSON (2005): “The structural breaks in the equity premium,” *Journal of Business and Economic Statistics*, 23, 181–191.
- KITAGAWA, G. (1996): “Monte Carlo filter and smoother for non-Gaussian nonlinear state space models,” *Journal of Computational and Graphical Statistics*, 5, 1–25.

- KLAASSEN, F. (2002): “Improving GARCH volatility forecasts with regime switching GARCH,” *Empirical Economics*, 27, 363–394.
- LAMOUREUX, G. C., AND W. D. LASTRAPES (1990): “Persistence in Variance, Structural Change, and the GARCH Model,” *Journal of Business & Economic Statistics*, 8, 225–234.
- LIU, C., AND J. M. MAHEU (2008): “Are there structural breaks in realized volatility?,” *Journal of Financial Econometrics*, 6(3), 326–360.
- LIU, J., AND M. WEST (2001): “Combined parameter and state estimation in simulation-based filtering,” in *Sequential Monte Carlo Methods in Practice*, ed. by A. Doucet, N. de Freitas, and N. Gordon. Springer-Verlag.
- MAHEU, J. M., AND T. H. MCCURDY (2009): “How useful are historical data for forecasting the long-run equity return distribution?,” *Journal of Business and Economic Statistics*, 27(1), 95–112.
- MIKOSCH, T., AND C. STARICA (2004): “Nonstationarities in financial time series, the long-range dependence and the IGARCH effects,” *Review of Economics and Statistics*, 86, 378–390.
- PASTOR, L., AND R. F. STAMBAUGH (2001): “The Equity Premium and Structural Breaks,” *Journal of Finance*, 4, 1207–1231.
- PESARAN, M. H., D. PETTENUZZO, AND A. TIMMERMANN (2006): “Forecasting Time Series Subject to Multiple Structural Breaks,” *Review of Economic Studies*, 73(4), 1057–1084.
- PITT, M. K., AND N. SHEPHARD (1999): “Filtering via simulation: auxiliary particle filters,” *Journal of the American Statistical Association*, 94, 590–599.

- POLSON, N. G., J. R. STROUD, AND P. MULLER (2008): “Practical Filtering with Sequential Parameter Learning,” *Journal of the Royal Statistical Society: Series B*, 70(2), 413–428.
- RAGGI, D., AND S. BORDIGNON (2008): “Sequential Monte Carlo Methods for Stochastic Volatility Models with Jumps,” working paper, Department of Economics, University of Bologna.
- RISTIC, B., S. ARULAMPALAM, AND N. GORDON (eds.) (2004): *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, Boston.
- SCOTT, S. L. (2002): “Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century,” *Journal of the American Statistical Association*, 97(457), 337–351.
- STARICA, C., AND C. GRANGER (2005): “Nonstationarities in Stock Returns,” *Review of Economics and Statistics*, 87(3), 503–522.