Department of Economics

Stochastic Volatility Driven by Large Shocks

George Kapetanios and Elias Tzavalis

Working Paper No. 568 September 2006 ISSN 1473-0278

Stochastic Volatility Driven by Large Shocks

George Kapetanios*and Elias Tzavalis†
Department of Economics,
Queen Mary, University of London
London E1 4NS, UK

This version July, 2006

Abstract

This paper presents a new model of stochastic volatility which allows for infrequent shifts in the mean of volatility, known as structural breaks. These are endogenously driven from large innovations in stock returns arriving in the market. The model has a number of interesting properties. Among them, it can allow for shifts in volatility which are of stochastic timing and magnitude. This model can be used to distinguish permanent shifts in volatility coming from large pieces of news arriving in the market, from ordinary volatility shocks.

Keywords: Stochastic volatility, structural breaks, *JEL Classification*: C22, C15

1 Introduction

There is recently considerable evidence indicating the existence of structural breaks in the conditional variance process (volatility) of many economic and financial series. These breaks appear to be associated with extraordinary economic events, such as financial crises, monetary regime changes and exchange rate realignments.[1] Such events are viewed as large shocks in the literature. If these shocks are not accounted for, they overstate the evidence of persistence in the volatility process. Most of the above evidence is supported by testing procedures designed to identify the presence of structural breaks in volatility based on the intervention (dummy) analysis of Box and Tiao. These procedures cannot however diagnose if these large shocks can cause shifts in volatility of the series. To this end, in this paper

*g.kapetanios@qmul.ac.uk
†e.tzavalis@qmul.ac.uk

[1]See Diebold and Pauly (1987), Lamourex and Lastrapes (1990), Tzavalis and Wickens (1995), Hamilton and Lin (1996), Diebold and Inoue (2001), Ang and Bekaert (2002), Sensier and Van Dijk (2004) and Morana and Beltratti (2004), *inter alia*.

we suggest a parametric model of structural breaks in volatility which allows them to be endogenously driven by stock return innovations which are larger in size than a threshold parameter.

The suggested model captures two essential features of breaks: their rarity and varying magnitude, over time. In other worlds, it allows for breaks in stochastic volatility which are stochastic in both time and magnitude. The second feature of the model distinguishes it from other models considering shifts in the volatility of fixed magnitude (see Hamilton (1989) and Glosten, Jagannathan, and Runkle (1993), *inter alia*). To model such types of breaks, we adopt the framework of discrete time stochastic volatility models (SV) (see Taylor (1986), and Harvey, Ruiz, and Shephard (1994), *inter alia*). Our choice to model the process of breaks in volatility within this framework stems from the fact that, unlike ARCH-type models, SV ones specify the volatility process as a separate random process driven by its own shocks. Due to this extra randomness, the SV models can offer extra flexibility to disentangle the economic sources of volatility breaks from the ordinary volatility shocks.

The paper is organised as follows. Section 2 presents our model and discuss some of its properties. In Section 3, we present alternative estimation procedures of the model, while in Section 4 we report the results of a Monte Carlo exercise to assess its performance to adequately trace structural breaks in volatility. Section 5 presents the results of an empirical application of the model to examine if breaks in the volatility of the S&P 500 index implied return driven by large shocks in the stock market. Finally, Section 6 concludes the paper.

2 Model specification

We start our analysis with a simple version of the model. As we proceed, we discuss some possible extensions which may be of use in practice. Consider the following extension of the stochastic volatility model which allows for structural breaks in conditional variance

$$y_t = \mu + \sigma e^{h_t^{1/2}} \epsilon_t, \tag{1}$$

with

$$h_t = \beta_{t-1} + \gamma h_{t-1} + \eta_{1,t} \tag{2}$$

and

$$\beta_t = \beta_{t-1} + I(|\epsilon_{t-1}| > r)\eta_{2,t}, \tag{3}$$

where h_t is the logarithm of the conditional variance (volatility) of an observed economic series (e.g. a stock return) y_t, at time t, $\epsilon_t \sim NID(0,1)$ and $\eta_{1,t}, \eta_{2,t} \sim NID(0, \sigma^2_{\eta_i})$,

$i = 1, 2$, are innovations (shocks) which can be allowed to be correlated, and $\mathcal{I}(\mathcal{A}_t)$ is an indicator function taking the value 1 if the event $\mathcal{A} = \{|\epsilon_{t-1}| > r\}$ occurs, where r is a threshold parameter, and zero otherwise. The events captured by set $\mathcal{A}_t$ can be thought of as reflecting outliers in the level of series y_t. These can be attributed to large pieces of news arriving in the stock market at any point in time due, for instance, to monetary regime changes and financial market crises. In the case that ϵ_t and $\eta_{1,t}$ are correlated, the model can be extended to capture the well known leverage effects in financial markets (see Section 5). Another interesting extension of the model would be towards a multivariate direction, where the innovations driving the changes in β_t can be allowed to come from different sources than y_t, such as macroeconomic news.

The model given by equations (1)-(3) allows for the intercept β_t of the volatility process h_t (and, hence, h_t itself) to be subject to abrupt, discontinuous random changes over time, given by $\beta_t - \beta_{t-1} = I(|\epsilon_{t-1}| > r)\eta_{1,t}$. These changes accord with the common perception of structural breaks referred to in the literature. They are endogenously driven by innovations ϵ_{t-1} (or the standardised level of series y_t itself, if $\mu = 0$) which are larger in size than a threshold parameter r.[2] The specification of the stochastic process governing the shifts in β_t, given by (3), implies that both their timing and magnitude are stochastic in nature. The timing of a structural break in β_t is controlled by innovations ϵ_{t-1} and, more specifically, depends on the occurrence of the event $\mathcal{A}_t = \{|\epsilon_{t-1}| > r\}$, while the magnitude depends on the innovation $\eta_{2,t}$. This last feature of the model clearly distinguishes it from existing models in the literature that consider shifts in volatility of fixed magnitude driven by exogenous variables, or innovations in stock returns.[3] The presence of the innovation $\eta_{2,t}$ in process (3) constitutes a more flexible approach of modeling random shifts in volatility, as it leaves the data at hand to decide above (or below) which values of the threshold parameter r (including $r = 0$) innovations ϵ_{t-1} can have an impact on h_t. As such, it also allows for the possibility that not all the large innovations ϵ_{t-1} have an impact on volatility. Note that, when $\sigma^2_{\eta_2} = 0$, large values of innovations ϵ_{t-1} do not cause any structural change in β_t. In this case, our

[2]Note that by specifying β_t as

$$\beta_t = \beta_{t-1} + I(\epsilon_{t-1} > r_1)\eta^+_{2,t} + I(\epsilon_{t-1} < r_2)\eta^-_{2,t}, \tag{4}$$

where $\eta^+_{2,t}$ and $\eta^-_{2,t}$ are $NIID$ innovations and r_1 and r_2 are two different threshold parameters, our model can also allow for large positive and negative innovations ϵ_{t-1} to have asymmetric effects on h_t.

[3]See the Markov regime switching model of Hamilton (1989) and its various extensions or the extensions of GARCH, EGARCH and SV models by Glosten, Jagannathan, and Runkle (1993), Asai and McAleer (2004) and Yu (2005) , respectively. For instance, the model of Asai and McAleer (2004) assumes that β_t is given as $\beta_t = \gamma\{I(\epsilon_{t-1} < 0) - E[I(\epsilon_{t-1} < 0)]\}$.

model reduces to the standard SV model, with no breaks.

As it stands, model (1)-(3) generates a non-stationary pattern for the volatility process h_t, as the variance of the process governing the breaks in β_t grows with the time-interval of the data. If stationarity of h_t is a desirable property of the data, then stationarity of β_t would be required for this. There are a number of restrictions which can be imposed on β_t to make it stationary (see Cogley and Sargent (2002)). A straightforward one is the following

$$\beta_t = \delta_t \beta_{t-1} + I(|\epsilon_{t-1}| > r)\eta_{2,t}, \tag{5}$$

where

$$\delta_t = \begin{cases} 1 & \text{if } I(|\beta_{t-1}| < \beta) \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

This condition implies that β_t is bounded by β and, hence, it renders h_t stationary, too.[4] In the next theorem, we prove that restriction (5) implies strict stationarity of h_t provided that $|\gamma| < 1$.

Theorem 1 *If $|\gamma| < 1$ and condition (5) hold, then h_t is strictly stationary.*

The proof of the theorem is given in the Appendix.

3 Model Estimation

Estimation of model (1)-(3) requires an algorithm of sequentially updating estimates of the two state variables h_t and β_t. One natural choice for this is the Kalman filter. However, the model as it stands is clearly nonlinear, and thus application of the Kalman filter is not feasible. We will therefore approach estimation of the model from a number of angles which have different levels of ease of use and accuracy. Our first approach follows the work of Harvey, Ruiz, and Shephard (1994) in writing the measurement equation of the model in logarithmic form as

$$\ln \tilde{y}_t^2 = \left[\ln \sigma^2 + E(\ln \epsilon_t^2)\right] + h_t + u_t \left[\ln \epsilon_t^2 - E(\ln \epsilon_t^2)\right] \tag{7}$$

where $\tilde{y}_t = y_t - 1/T \sum_{t=1}^{T} y_t$ and $u_t = \ln \epsilon_t^2 - E(\ln \epsilon_t^2)$ Then, the model defined by (7) and the transition equations (2) and (3) is linear and amenable to the analysis by the standard

[4]Further restrictions could be placed on the process β_t so that, if the bound β is exceeded, the process returns to some prespecified level. We do not advocate a particular mechanism for making the process β_t stationary. We simply wish to indicate that there exist specifications which give a stationary β_t process. The exact specification of the process may be left to the empirical researcher depending on their priors on the particular issue at hand.

Kalman filter. One further adjustment which is needed to this end involves substituting $I(|\hat{u}_{t-1|t-1}| > r)$ for $I(|u_{t-1}| > r)$, where $\hat{u}_{t-1|t-1}$ denotes the conditional expectation of u_{t-1} given $\ln \tilde{y}_1^2, \ldots, \ln \tilde{y}_{t-1}^2$. This ensures that now the model can be estimated through the standard Kalman filter procedure assuming Gaussianity for the error term u_t. Although Gaussianity of u_t does not hold under the assumptions of our model, the estimates retrieved by the Kalman filter have important properties, as it is a minimum mean square estimator of the state variables among all other linear estimators (see Harvey, Ruiz, and Shephard (1994)).

To carry out the estimation of the model through the Kalman filter, first we assume that the threshold parameter r is known. We discuss estimation of r later in this section. Let $z_t = \ln \tilde{y}_t^2 - 1/T \sum_{t=1}^{T} \ln \tilde{y}_t^2$. Under the above assumptions, we can write model (1)-(3) in a state space form as

$$z_t = X_t b_t + u_t, \quad t = 1, \ldots, T \tag{8}$$

$$b_t = A_t b_{t-1} + R_t \eta_t \quad \eta_t \sim i.i.d.N(0, \Sigma_\eta) \tag{9}$$

where $X_t = (1, 0)'$, $b_t = (h_t, \beta_t)'$, $\eta_t = (\eta_{1,t}, \eta_{2,t})'$,

$$A_t = \begin{pmatrix} \gamma & 1 \\ 0 & 0 \end{pmatrix} .\text{and } R_t = \begin{pmatrix} 1 & 0 \\ 0 & I(|\hat{u}_{1,t-1|t-1}| > r) \end{pmatrix}$$

Below, we abstract from issues arising from the estimation of the parameters of the model and concentrate on the estimation of the state vector $b_t = (h_t, \beta_t)'$ conditional on the parameters being known. Let us denote the estimator of b_t conditional on the information set $\mathcal{I}_{t-1}$ as $\hat{b}_{t|t-1}$ and that conditional on the information set up to and including time t as by $\hat{b}_t$. Denote the covariance matrices of the estimators $\hat{b}_{t|t-1}$ and $\hat{b}_t$ as $P_{t|t-1}$ and P_t, respectively. Then, estimation of $\hat{b}_t$ by the Kalman filter comprises sequential application of the following two sets of equations:

$$\begin{aligned} \hat{b}_{t|t-1} &= A_t \hat{b}_{t-1} \\ \hat{P}_{t|t-1} &= A_t \hat{P}_{t-1} A_t' + R_t \Sigma_\eta R_t, \end{aligned} \tag{10}$$

known as the prediction equations, and

$$\begin{aligned} \hat{b}_t &= \hat{b}_{t|t-1} + \hat{P}_{t|t-1} X_t \left(\frac{y_t - X_t' \hat{b}_{t|t-1}}{f_t} \right) \\ P_t &= P_{t|t-1} - P_{t|t-1} X_t \left(\frac{1}{f_t} \right) X_t' P_{t|t-1}, \end{aligned} \tag{11}$$

known as the updating equations, where

$$f_t = X_t' P_{t|t-1} X_t + \zeta_t \tag{12}$$

(see, Hamilton (1994), *inter alia*) and $\zeta_t = E(u_t^2)$. By the normality assumption for ϵ_t and the assumption that $E(\epsilon_t^2) = 1$, it can be shown that $E(u_t^2) = 4.93$. For a given value of r, the log-likelihood function for the observation equation (8), denoted as $\mathcal{L}(r)$, can be written in terms of the prediction errors $v_t = y_t - X_t' \hat{b}_{t|t-1}$ as

$$\mathcal{L}(r) = -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log f_t - \frac{1}{2} \sum_{t=1}^{T} v_t^2 / f_t. \tag{13}$$

$\mathcal{L}(r)$ can be used to estimate recursively the unknown parameters of the model (apart from r which is assumed to be known at the moment). In summary, the Kalman filter can be used to obtain the following sets of estimates of the state variables b_t: (i) estimates conditional on $\mathcal{I}_t$ known as filter estimates, and (ii) estimates conditional on the information of the whole sample, denoted as $\mathcal{I}_T$, known as smoothed estimates. The second set of estimates and their respective covariance matrices are denoted by $\hat{b}_{t|T}$ and $P_{t|T}$ and are given by

$$\hat{b}_{t|T} = \hat{b}_t + P_t^* (\hat{b}_{t+1|T} - A_{t+1} \hat{b}_t) \tag{14}$$

and

$$P_{t|T} = P_t + P_t^* (P_{t+1|T} - P_{t+1|t}) P_t^{*\prime} \tag{15}$$

where $P_t^* = P_t A_{t+1}' P_{t+1|t}^{-1}$. The filtered estimates of b_t can reveal agents' perceptions about the current state of β_t in the economy, at time t. The set of smoothed estimates of b_t can be used to statistically appraise the impact of large innovations ϵ_t on β_t using information over the whole sample. Finally, using the general state space model (8)-(9), forecasts of the state at time $t + h$ can be produced conditional on information available at time t. For our model specification, where $\eta_{1,t}, \eta_{2,t} \sim NID(0, \sigma_{\eta_i}^2)$, multi-step forecasts need to be produced using stochastic simulations due to the nonlinear nature of the model.

The above estimation procedure assumes that r is known which may not be true in practice. In addition, from an economic point of view it will be useful to estimate the threshold parameter r endogenously from the data employing our model. This will enable us to evaluate the magnitude of a structural innovation ϵ_{t-1} which can cause permanent shifts in the volatility function h_t. As in other threshold models (see, e.g. Kapetanios (2000)), to estimate r we will adopt a grid search procedure over a range of possible values of r.

According to this, the loglikelihood function $\mathcal{L}(r)$ will be maximized for every point of the grid and the point which gives the maximum likelihood, over the grid, will be considered as an optimum estimate r. The estimates of the unknown parameters of the model and the state vector b_t corresponding to this estimate of r will constitute the maximum likelihood estimates of the state space model (8)-(9). These estimates will be consistent provided that the threshold parameter will be consistently estimated. The last result is stated in Theorem 1 and proven in the Appendix.

Theorem 2 *Assume that the structural break model may be written as in (8)-(9) where* $\eta_{1,t}$ *and* $\eta_{2,t}$ *are* $NID(0, \sigma^2_{\eta_1})$ *and* $NID(0, \sigma^2_{\eta_2})$ *respectively, and* A_t *is specified so that* b_t *is a geometrically ergodic process. Then, the estimator of* r*, denoted* $\hat{r}$*, obtained via grid search, is consistent.*

The proof of the theorem is given in the Appendix. Below, we make some remarks concerning the estimation of the threshold parameter in practice.

Remark 1 *The normality assumption is not necessary for the consistency proof. It can be replaced with the assumption that the fourth moment of innovations* $\eta_{1,t}$ *and* $\eta_{2,t}$ *exist and an assumption about continuity of the density functions of these innovations.*

Remark 2 *Since estimation of the threshold parameter is problematic in small samples in general (see Kapetanios (2000)) and since this problem is exacerbated by the rarity of breaks in the present context, the grid search can be considerably simplified if we consider values of* r *which correspond to extreme quantiles of the normalised error of (1),* ϵ_t*, such as its 95-th or 99-th centile.*

As the estimation procedure through the Kalman filter described above is suboptimal given the fact that u_t is not Gaussian, in what follows we suggest an alternative procedure based on importance sampling along the lines of Durbin and Koopman (2001). As we will see in Section 5, apart from non-Gaussianity this estimation procedure can account for possible leverage effects when estimating the volatility of many financial series. These can not be handled by the application of the Kalman filter to the linearised form of the model given by (8)-(9).

Let $b = (b_1, ..., b_T)$, $\tilde{y} = (\tilde{y}_1, ..., \tilde{y}_T)$ and the conditional density of b given $\tilde{y}$ be denoted by $p(b|\tilde{y})$. Then, importance sampling constitutes an estimation method of the conditional

mean of b given $\tilde{y}$, defined as

$$E(b|\tilde{y}) = \int bp(b|\tilde{y})db = \int b\left[\frac{p(b|\tilde{y})}{g(b|\tilde{y})}\right] g(b|\tilde{y})db, \tag{16}$$

where $g(b|\tilde{y})$ is a density that approximates $p(b|\tilde{y})$, based on simulation. Following Durbin and Koopman (2001), $g(b|\tilde{y})$ can be set to the Gaussian density that has the same conditional mode as $p(b|\tilde{y})$. If $g(b|\tilde{y})$ is known, then $E(b|\tilde{y})$ can be estimated by simulation. In particular, let a set of B random draws from $g(b|\tilde{y})$ be denoted as $b^{(1)}, ..., b^{(B)}$. Then, an estimator for $E(b|\tilde{y})$ based on importance sampling is given as

$$\hat{b} = \frac{\sum_{i=1}^{B} b^{(i)} w(b^{(i)}, \tilde{y})}{\sum_{i=1}^{B} w(b^{(i)}, \tilde{y})} \tag{17}$$

where $w(b^{(i)}, \tilde{y}) = \frac{p(b^{(i)}, \tilde{y})}{g(b^{(i)}, \tilde{y})}$, $p(b^{(i)}, \tilde{y})$ is the true joint density of $b^{(i)}$ and $\tilde{y}$, $g(b^{(i)}, \tilde{y})$ is its Gaussian approximation consistent with $g(b|\tilde{y})$, and $p(b^{(i)}, \tilde{y})$ is given by $\prod_{t=1}^{T} p(\tilde{\eta}_t)p(\tilde{y}_t|b_t)$, where $\tilde{\eta}_t = R_t\eta_t$. Due to the discontinuity of $p(\tilde{\eta}_t)$ at zero, $\frac{p(b^{(i)}, \tilde{y})}{g(b^{(i)}, \tilde{y})}$ can be approximated by $\frac{p(b^{(i)}|\tilde{y})}{g(b^{(i)}|\tilde{y})}$. This essentially means that the marginal density of b, $p(b)$, is approximated by a Gaussian density, $g(b)$. This approximation is such that the first and second moments of $p(b)$ and $g(b)$ coincide. To complete the importance sampling estimation procedure, it remains to discuss how to obtain $g(b, \tilde{y})$. To this end, we follow the iterative method suggested by Durbin and Koopman (2001). Let $\theta_t = X_t b_t$ and $s_t(\tilde{y}_t|\theta_t) = -\log p(\tilde{y}_t|\theta_t)$. Define the first and second derivatives of $s_t(\tilde{y}_t|\theta_t)$ at $\bar{\theta}_t = 0$ as

$$\dot{s}_t = \left.\frac{\partial s_t}{\partial \theta}\right|_{\theta_t=\bar{\theta}_t} \text{ and } \ \ddot{s}_t = \left.\frac{\partial^2 s_t}{\partial \theta \partial \theta'}\right|_{\theta_t=\bar{\theta}_t}.$$

Then, define

$$\bar{y}_t = \bar{\theta}_t - \ddot{s}_t^{-1}\dot{s}_t, \text{ with } \bar{\zeta}_t = \ddot{s}_t^{-1} \tag{18}$$

Apply the Kalman filter and smoother as defined by (10),(11), (12), (14) and (15), setting $z_t = \bar{y}_t$ and $\zeta_t = \bar{\zeta}_t$.[5] This returns a value for $\hat{b}_{t|T}$ which is used as a new value for $\bar{\theta}_t$ in (18) and the Kalman filter and smoother until convergence. The output of the Kalman filter when the iterations converge defines a normal distribution which is used as an estimate of $g(b, \tilde{y})$. The

[5]It is straightforward to show that for the simple stochastic volatility model

$$\tilde{y}_t = \bar{\theta}_t + 1 - \frac{\exp(\bar{\theta}_t)}{(\tilde{y}/\sigma)^2} \text{ and } \quad \tilde{\zeta}_t = \frac{2\exp(\bar{\theta}_t)}{(\tilde{y}/\sigma)^2}.$$

final value of $\hat{b}_{t|T}$ from this set of iterations can also be considered as a possible estimator of the state. We will refer to this estimator as the approximate importance sampling estimator. Our Monte Carlo study will show that this estimation procedure has desirable properties in small samples. Parameter estimation through importance sampling can be carried out straightforwardly by maximising the likelihood given by

$$\ell = \frac{1}{B}\sum_{i=1}^{B} w(b^{(i)}, \tilde{y}).$$

4 Monte Carlo Study

In this section, we carry out a small scale Monte Carlo study to investigate the performance of our model to track structural breaks in volatility process h_t adequately. This is done for samples where the number of breaks is relatively small. The presence of the threshold and the fact that breaks occur infrequently raises the question of how well these breaks can be captured by the Kalman filter or the importance sampling estimation procedures. As the main aim of our Monte Carlo exercise is to assess the performance of these procedures, we concentrate on the estimation of the state variable β_t assuming that the parameters of the model are known. It is reasonable to expect that the state variable driving the breaks is hard to carry inference on given that there are only a few observations which will contain information about the breaks.

To evaluate the accuracy of the estimates of β_t, in our experiments we generate data according to model (1)-(3) where $\gamma = 0.3$, $\sigma^2_{\eta_1} = 1$, $\sigma^2_{\eta_2} = 0.25$ and $b_0 = 0$. For the threshold parameter, we consider two cases: $r = 1.96$ and $r = 2.24$ implying on average a break on every 20 and 40 periods, respectively. We set the sample size to either $T = 500$ or $T = 2000$. In each experiment, we run 500 replications and we report the average correlation coefficient between the true β_t and the smoothed estimates of β_t obtained by the Kalman filter and the importance sampling estimation procedure. For the latter, we report two sets of correlation coefficients. One for the final smoothed estimate, given by (17), where $B = 500$ and the second for the smoothed estimate of the approximate Gaussian model as given by $\hat{b}_{t|T}$ at the end of the iterations using (18).

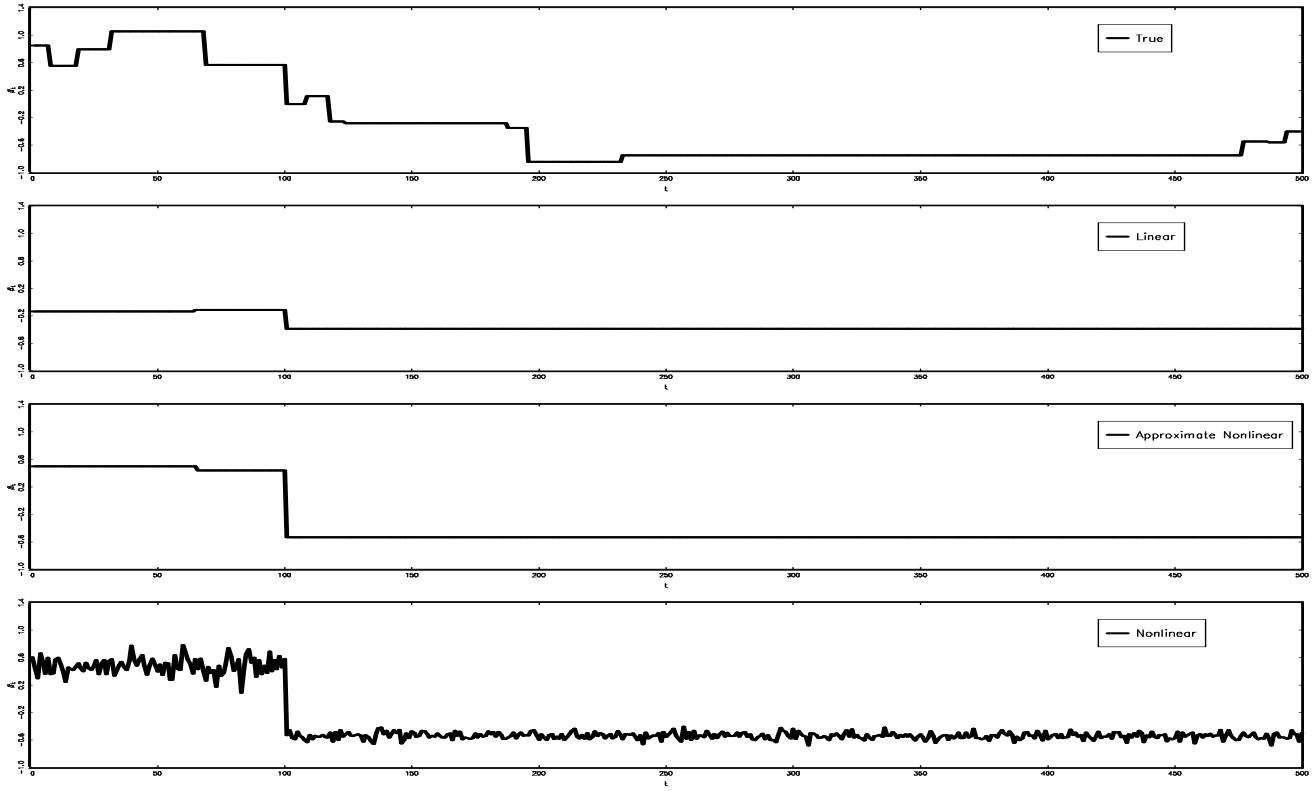
To better see how closely our model can capture structural breaks in β_t over the sample, we also report pictorial results for particular replications. These replications correspond to the 50% quantile of the empirical distribution of the correlations between the smoothed estimates of the Kalman filter and true values of β_t. These results are presented in Figures 1 to 4.

Table 1: Monte Carlo results				
(r, T)	$(1.96, 500)$	$(1.96, 2000)$	$(2.24, 500)$	$(2.24, 2000)$
nb	30	120	12	50
Corr. Coeff. (Kalman Filter)	0.740	0.906	0.566	0.826
Corr. Coeff. (Apprx. Imp. Sampling)	0.777	0.928	0.479	0.770
Corr. Coeff. (Imp. Sampling)	0.692	0.905	0.403	0.749
Notes: nb denotes the number of breaks per sample				

The results of the table clearly indicate that both estimation procedures discussed in the paper can satisfactorily capture structural breaks in β_t generated by our model. As expected, this is more evident for the cases of (r, T) where there is an adequate number of breaks per sample and the size of sample is large enough. For instance, for the case of $(r, T) = (1.96, 2000)$ the correlation coefficients between the smoothed estimates of the changes in β_t and their true values reaches its highest value which is close to 90%, using either the Kalman filter or and the importance sampling estimates. Apart from this case, the performance of our estimation procedures is also satisfactory when the size of sample is smaller but the number of breaks is substantial, e.g. $(r, T) = (1.96, 500)$ implying $nb = 30$ structural break on average per sample. The second interesting conclusion that can be drawn from the results of the table is about the performance of the standard Kalman filter estimation procedure relative to that based on importance sampling. Our results suggest that, for the case of frequent breaks and/or large enough size of T (i.e. $(r, T) = \{(1.96, 500), (1.96, 2000), (2.24, 2000)\}$), all methods have similar performance. For the case of a smaller number of breaks (i.e. for $(r, T) = (2.24, 500)$, where $nb = 12$), it seems that the Kalman filter performs better. Interestingly for this case, the final estimate from the importance sampling algorithm although comparable to the approximate importance sampling estimator is slightly worse. However, it improves greatly with the sample size.

The above conclusions can be confirmed by inspecting the pictorial output reported in Figures 1 to 4. Inspection of these figures show that the smoothed estimates of β_t track quite well the true break process, when T is large and/or the number of breaks per T is adequate enough [see Figures 1, 2, and 4]. Note that the different estimators of β_t considered can capture quite well the trend of the true break process even for the case that T is smaller and the number of breaks per T is less. This happens despite the fact that correlation coefficients

Figure 1: $T = 500$, $r = 1.96$

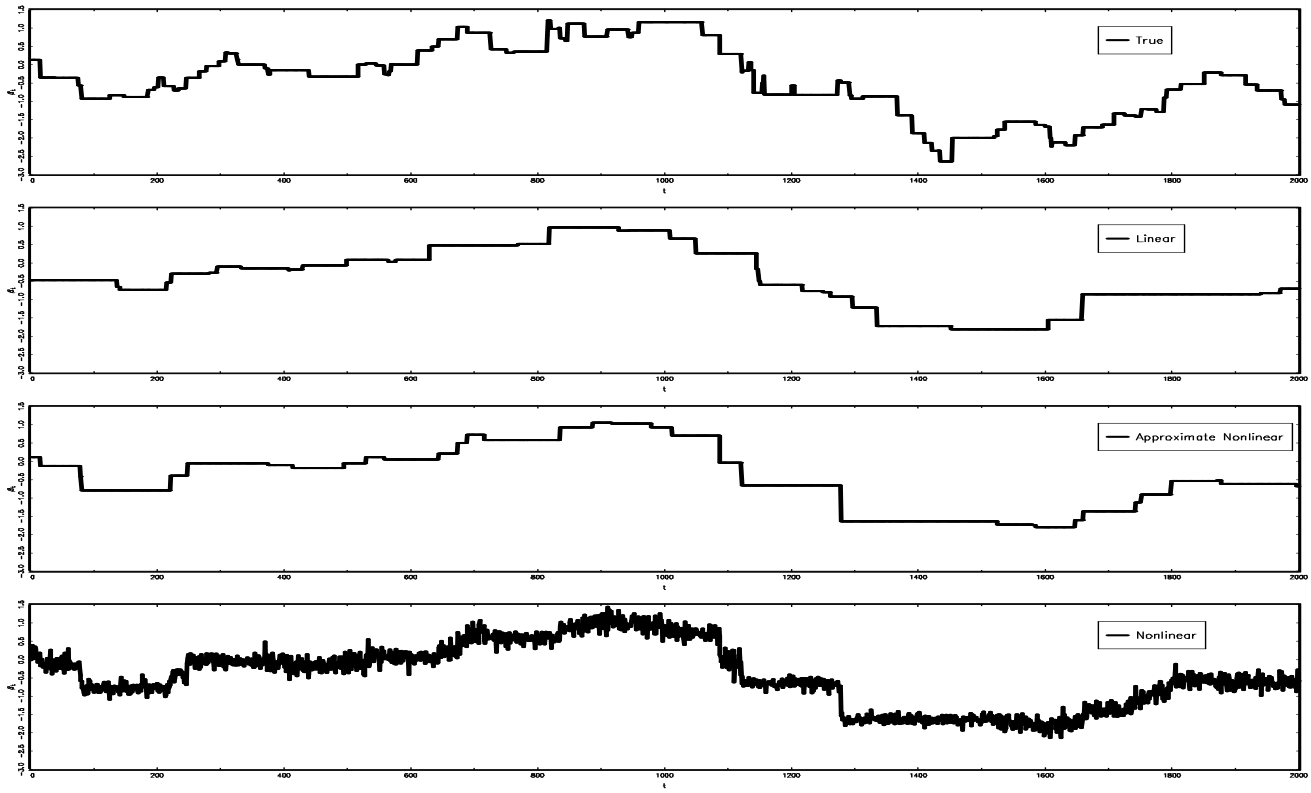


between the estimates and the true values of β_t reported in Table 1 are much smaller for this case.

5 Empirical Application

As an empirical application of our model, we employ it to trace out possible structural breaks occurred in stochastic volatility of the implied return by the S&P 500 index driven by large return innovations (news) in the US stock market. The data we use are daily and cover the period between the 2nd of January 1992 and the 14th of June 2005. During this period, extreme events occurred in the US stock market. Examples include the burst of the market bubble which began in the spring of year 2000 and the fall of the share prices due to the

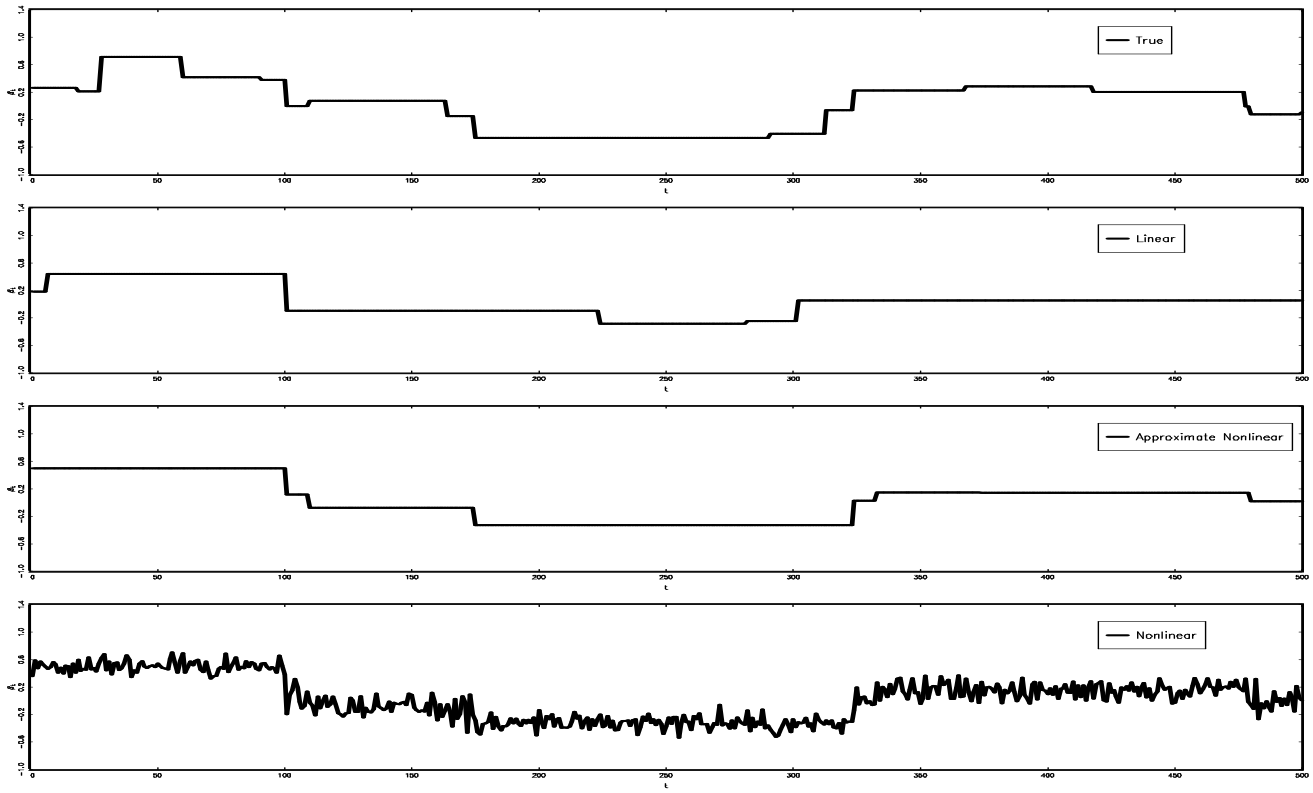
Figure 2: $T = 2000$, $r = 1.96$



collapse of Enron and WorldCom corporations.

We carry out a number of different estimations. Firstly, we estimate the simple stochastic volatility model with the break mechanism using the linear Kalman filter algorithm, described by equations (10)-(15). A grid search for the value of r implying a probability of 5%, 2.5, 2% and 1.5% for the event $\{|u_{t-1}| > r\}$ occuring, suggests that the threshold value corresponding to 5% best describes the data since it corresponds to the largest log-likelihood. We simplify our estimation procedures by estimating σ^2 as $\frac{1}{T}\sum_{t=1}^{T} \ln \tilde{y}_t^2 + 1.27$, where $E(\ln \epsilon_t^2) = -1.27$. This is important simplification of our estimation procedure, as numerical maximisation of the likelihood is not a trivial numerical exercise, especially for the importance sampling procedure for which the number of parameters that needs to be estimated by likelihood maximisation should be kept at a minimum.

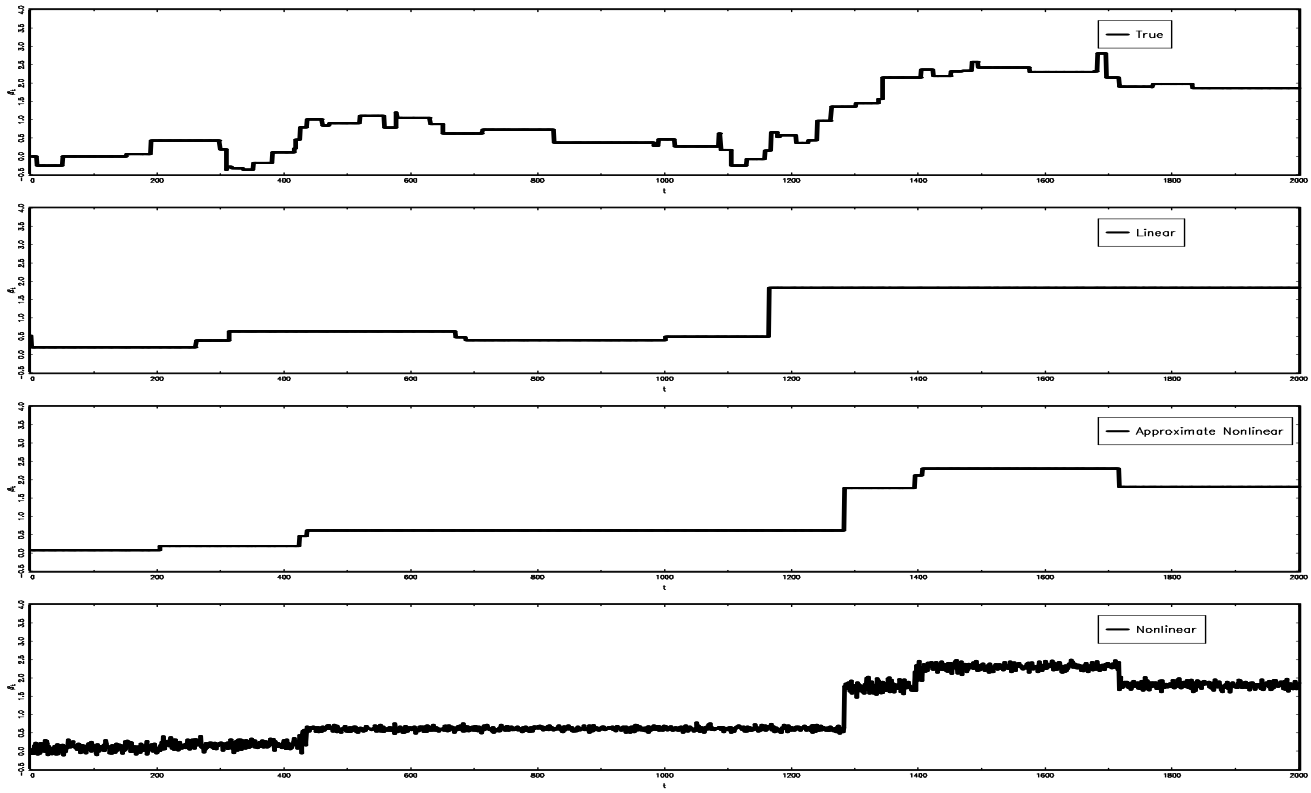
Figure 3: $T = 500$, $r = 2.24$



In Table 2, we report parameter estimates for the remaining parameters of the model, namely γ, σ_{η_1} and σ_{η_2}, with their standard errors in parentheses, based on the importance sampling procedure. The table presents two sets of results. The first does not allow for structural breaks in h_t, but allows for leverage effects, i.e. possible correlation between innovations ϵ_t and $\eta_{1,t}$, denoted by ρ. This specification constitutes the standard SV model and can be estimated within our framework by setting $\sigma_{\eta_2} = 0$. The second set presents the results for the full specification of our model, given by (1)-(3), which also allows for leverage effects. To this end, we slightly modified the initial specification of our model following Koopman (2005). In particular, we use the following version of our model

$$y_t = \mu + \sigma e^{h_t^{1/2}} \{\epsilon_t + sign(\rho)\eta_{3,t}\} \tag{19}$$

Figure 4: $T = 2000$, $r = 2.24$



with

$$h_t \equiv \beta_{t-1} + \gamma h_{t-1} + \sigma_{\eta_1}\{\eta_{1,t} + \eta_{3,t}\} \tag{20}$$

$$\beta_t = \beta_{t-1} + I(|\epsilon_{t-1}| > r)\eta_{2,t} \tag{21}$$

where $\epsilon_t \sim NID(0, 1-|\rho|)$, $\eta_{1,t} \sim NID(0, 1-|\rho|)$, $\eta_{2,t} \sim NID(0, \sigma_{\eta_2})$ and $\eta_{3,t} \sim NID(0, |\rho|)$, where all errors are all mutually and serially independent. The above specification of our model when is written in state space form consists of three state variables, where $\eta_{3,t}$ constitutes the new one. This model can be estimated through the importance sampling procedure where now $\theta_t = (h_t, \eta_{3,t})$**.** Then, $s_t(\tilde{y}_t|\theta_t)$ becomes

$$s_t(\tilde{y}_t|\theta_t) = \frac{1}{2}h_t + \frac{1}{2}\sigma^{-2}\exp(-h)\,\beta\left(\tilde{y}_t - \sigma\exp(\frac{1}{2}h)\kappa\eta_{3,t}\right)^2,$$

and it has first and second order derivatives given by

$$\dot{s}_t = \begin{pmatrix} \frac{\partial s_t}{\partial h_t} \\ \frac{\partial s_t}{\partial \eta_{3,t}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2\sigma^2 e^{h_t}} \left(-\beta \tilde{y}_t^2 + \beta\kappa\eta_{3,t}\tilde{y}_t\sigma e^{\frac{1}{2}h_t} + \sigma^2 e^{h_t} \right) \\ -\frac{1}{\sigma e^{\frac{1}{2}h_t}} \left(\tilde{y}_t\beta\kappa - \sigma\beta\kappa^2\eta_{3,t}e^{\frac{1}{2}h_t} \right) \end{pmatrix}$$

and

$$\ddot{s}_t = \begin{pmatrix} \frac{\partial^2 s_t}{\partial h_t^2} & \frac{\partial^2 s_t}{\partial h_t \partial \eta_{3,t}} \\ \frac{\partial^2 s_t}{\partial h_t \partial \eta_{3,t}} & \frac{\partial^2 s_t}{\partial \eta_{3,t}^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{4\sigma^2 e^{h_t}} \left(2\tilde{y}_t^2\beta - \tilde{y}_t\sigma\beta\kappa\eta_{3,t}e^{\frac{1}{2}h_t} \right) & 0.5\frac{\tilde{y}_t}{\sigma}\beta\kappa e^{-\frac{1}{2}h_t} \\ \frac{1}{2}\frac{\tilde{y}_t}{\sigma}\beta\kappa e^{-\frac{1}{2}h_t} & \beta\kappa^2 \end{pmatrix},$$

respectively, where $\beta = (1 - |\rho|)^{-1}$ and $\kappa = sign(\rho)$.

The results of the table lead to a number of interesting conclusions. First, as was expected, accounting for structural breaks in stochastic volatility h_t substantially reduces the degree of linear persistence as measured by the estimate of the autoregressive coefficient γ. This drops significantly from 0.63 to 0.11. These results clearly indicate that the high degree of persistence observed in many empirical studies mentioned in the introduction can be attributed to the lack of accounting for structural changes in the volatility process, h_t. Note that together with the drop in the estimate of γ there is also a significant decrease of the value of the correlation coefficient ρ, capturing the leverage effects. This happens because the impact of news on volatility, implied by the leverage effects, is captured through the break process (3) in our model.

The second conclusion that can be drawn from the results of the table is that the estimate of $\sigma_{\eta_2} = 0.63$ is different than zero, meaning that our model can identify significant structural breaks in h_t driven by larger than $r = 2.5\%$ innovations in the stock return y_t. Note that the standard error reported cannot be used for testing the hypothesis that $\sigma_{\eta_2} = 0$ since, under the null hypothesis, the parameter takes a value on the boundary of the parameter space. However, the magnitude of σ_{η_2} and its estimated standard error strongly suggest that this is the case. Further support for our model compared with the standard SV, which does not allow for structural breaks, can be also gained by the values of the log-likelihood function for these two models. These are found to be 28.97 and -639.42, respectively, suggesting that our model provides a better fit of the data than the standard SV model, with no breaks.

Table 2: Parameter Estimates				
Model	σ_{η_1}	σ_{η_2}	γ	ρ
Nonl. Kalman Filter w/out Leverage	0.949	0.188	0.374	-
	(0.158)	(0.059)	(0.162)	-
Nonl. Imp. Sampling with Leverage	0.630	0.126	0.116	-0.028
	(0.008)	(0.0028)	(0.052)	(0.009)
Imp. Sampling with Leverage	0.723	-	0.633	-0.206
	(0.006)	-	(0.012)	(0.015)
Notes: Standard errors are reported in parentheses.				

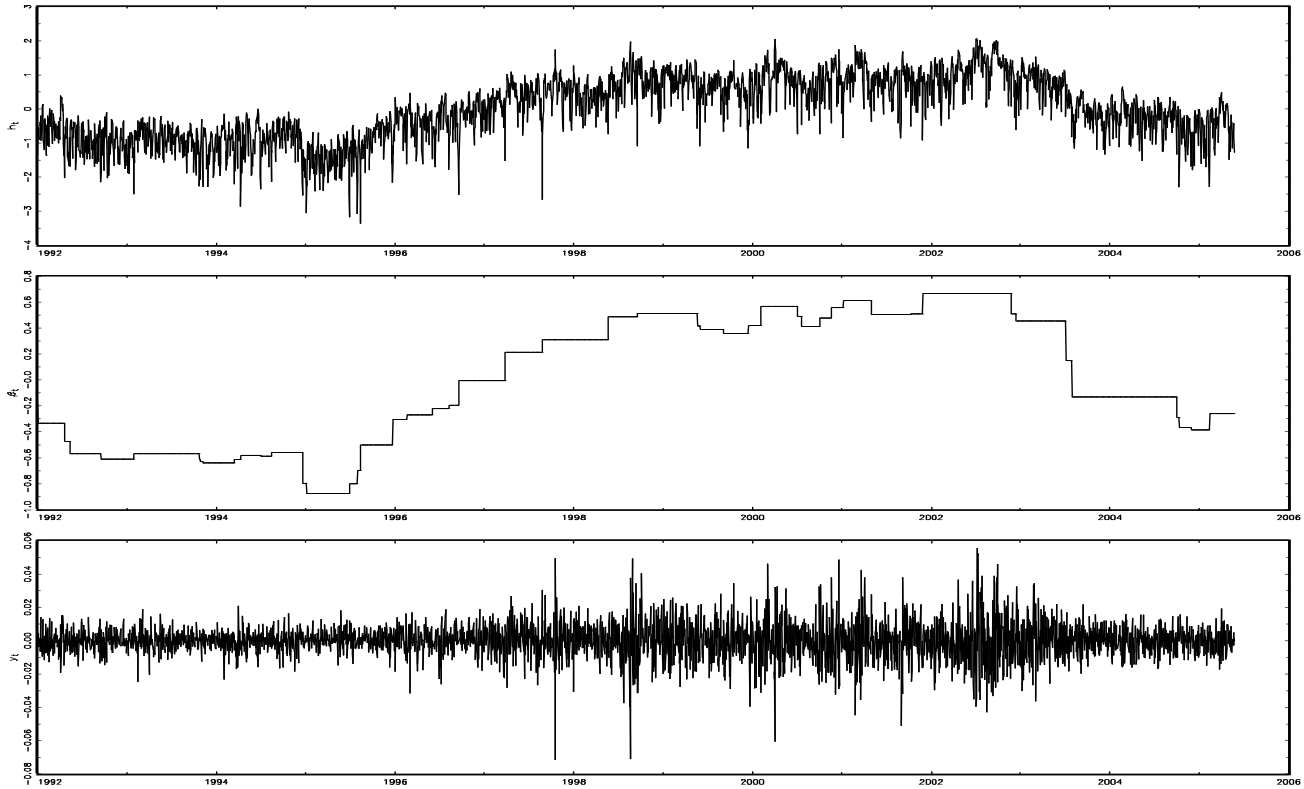
To investigate the patterns of the break process β_t and the stochastic volatility h_t, as well as to identify large stock return news which generated the breaks we also report some pictorial results. In Figures 5-7 we present the estimates of the Kalman filter and importance sampling procedures for β_t and h_t together with the actual series of the SP&500 stock return. In the importance sampling figures, we also report smoothed estimates of β_t taken from the Gaussian approximation of our model, as described in Section 3. Inspection of the figures indicates that stochastic volatility follows a nonlinear pattern due to the presence of a substantial number of breaks in its intercept β_t of different size. Our results show that, during our sample, β_t has reached its lowest level between years 1993 and 1996, while its highest level was reached from the middle of 1998 to the beginning of 2002.

6 Conclusions

This paper has introduced a new model of stochastic volatility allowing for structural breaks in its intercept (mean). On the basis of evidence that shifts in volatility are associated with extraordinary events in economic series (e.g. stock returns), the model considers a break process that is endogenously driven by innovations (news) which are larger than a threshold parameter. These breaks are stochastic in time and size, and thus can allow for random shifts in volatility.

To estimate the model, the paper suggests a procedure based on importance sampling. This can handle the nonlinear nature of the model and the non Gaussianity of the error terms. In a Monte Carlo exercise, the paper assess the performance of this estimation procedure to sufficiently track a true break process. It also compare this with an estimation procedure based on the Kalman filter which linearises and assumes Gaussianity. The results of this exercise indicate that both the importance sampling and Kalman filter estimation procedures can adequately capture the true break process, when the number of breaks is substantial for

Figure 5: Results from Kalman Filter Estimation. The first panel reports the smoothed estimate of h_t. The second panel reports the smoothed estimate of β_t. The final panel reports the actual return data.



a given sample size. Finally, in an empirical illustration of our model it is shown that the evidence of structural breaks in the mean of volatility found by many recent studies can be attributed to large news arriving in the stock market. This can cause permanent shifts in the level of volatility.

References

Ang, A., and G. Bekaert (2002): "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics*, 20, 163–182.

Asai, M., and M. McAleer (2004): "Dynamic Asymmetric Leverage in Stochastic Volatility Models," *Faculty of Economics, Tokyo Metropolitan University*.

Chan, K. S. (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21(1), 520–533.

Cogley, T., and T. S. Sargent (2002): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US," *New York University, Mimeo*.

Diebold, F. X., and A. Inoue (2001): "Long Memory and Structural Change," *Journal of Econometrics*, 105, 131–159.

Diebold, F. X., and P. Pauly (1987): "Structural Change and the Combination of Forecasts," *Journal of Forecasting*, 6, 21–40.

Durbin, J., and S. J. Koopman (2001): *Time Series Analysis by State Space Methods*. Oxford University Press.

Glosten, L. R., R. Jagannathan, and D. Runkle (1993): "Relationship between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48, 1779–1802.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Time Series," *Econometrica*, 57(2), 357–384.

——— (1994): *Time Series Analysis*. Princeton University Press.

Hamilton, J. D., and G. Lin (1996): "Stock Market Volatility and the Business Cycle," *Journal of Applied Econometrics*, 16, 39–56.

Harvey, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A. C., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

Kapetanios, G. (2000): "Small sample properties of the conditional least squares estimator in SETAR models," *Economics Letters*, 69(3), 267–276.

Koopman, S. J. (2005): "On Importance Sampling for State Space Model," *Timbergen Institute Discussion Paper No. 2005-117.*

Lamourex, C. C., and W. Lastrapes (1990): "Persistence in Variance, Structural Change and the GARCH Model," *Journal of Business and Economic Statistics*, 8, 225–234.

Ling, S., and M. McAleer (1996): "Stationarity and the Existence of Moments of a Family of GARCH Processes," *Journal of Econometrics*, 106, 109–117.

Morana, C., and A. Beltratti (2004): "Structural Change and Long-Range Dependence in Volatility of Exchange Rates: Either, Neither or Both?," *Journal of Empirical Finance*, 11, 629–658.

Sensier, M., and D. Van Dijk (2004): "Testing for Changes in Volatility of US Macroeconomic Time Series," *Review of Economics and Statistics*, 86, 833–839.

Taylor, S. J. (1986): *Modeling Financial Time Series.* Wiley.

Tweedie, R. L. (1975): "Sufficient Conditions for Ergodicity and Recurrence of Markov Chains on a General State Space," *Stochastic Processes Appl.*, 3, 385–403.

Tzavalis, E., and M. Wickens (1995): "The Persistence in Volatility of the US Term Premium," *Economics Letters*, 49, 381–398.

Yu, J. (2005): "On Leverage in a Stochastic Volatility Model," *Journal of Econometrics*, 127, 165–178.

Appendix

Proof of Theorem 1

We now prove strict stationarity for h_t, given by

$$y_t = e^{1/2h_t}\epsilon_{1t}$$

$$h_t = \beta_{t-1} + \gamma h_{t-1} + \epsilon_{2t}$$

$$\beta_t = \delta_t \beta_{t-1} + I(|\epsilon_{1t-1}| > r)\epsilon_{3t}$$

where

$$\delta_t = \begin{cases} 1 \text{ if } I(|\beta_{t-1}| < \beta) \\ 0 \quad \text{otherwise} \end{cases} \tag{22}$$

The first step is to derive a recursive representation for h_t. This is given by

$$h_t = \sum_{j=0}^{\infty} \gamma^j (\beta_{t-j-1} + \epsilon_{2,t-j})$$

Following Theorem 2.1 of Ling and McAleer (1996) the result will follow if we show that for some $\alpha \in (0,1)$

$$E(h_t^\alpha) < \infty$$

By The Marcinkiewicz-Zygmund inequality we have that

$$E(h_t^\alpha) = E\left(\left(\sum_{j=0}^{\infty} \gamma^j (\beta_{t-j-1} + \epsilon_{2,t-j})\right)^\alpha\right) \leq c\left(\sum_{j=0}^{\infty} \gamma^{2j}\right)^{\alpha/2} E\left(\beta_{t-j-1} + \epsilon_{2,t-j}\right)^\alpha$$

which is finite as long as β_t is strictly stationary and $E\left(\beta_{t-j-1}\right)^\alpha < \infty$ and $E\left(\epsilon_{2,t-j}\right)^\alpha < \infty$. Thus it suffices to prove that β_t is strictly stationary and $E\left(\beta_{t-j-1}\right)^\alpha < \infty$. $E\left(\beta_{t-j-1}\right)^\alpha < \infty$ follows easily from strict stationarity and $E\left(\epsilon_{3,t}\right)^\alpha < \infty$. Thus we only need to prove strict stationarity for β_t. To do that we prove geometric ergodicity of β_t which implies strict stationarity. asymptotically. To prove geometric ergodicity we use the drift criterion of Tweedie (1975). This condition states that a process is ergodic under the regularity condition that disturbances have positive densities everywhere if the process tends towards the center of its state space at each point in time. More specifically, β_t is geometrically ergodic if there exists constants $\delta < 1$, $B, L < \infty$, and a small set C such that

$$E\left[\|\beta_t\| \mid \beta_{t-1} = d\right] \leq \vartheta \|d\| + L, \quad \forall d \notin C, \tag{23}$$

$$E\left[\|u_t\| \mid u_{t-1} = u\right] \leq B, \quad \forall d \in C, \tag{24}$$

where $\|\cdot\|$ is the Euclidean norm. The concept of the small set is the equivalent of a discrete Markov chain state in a continuous context. It is clear that (24) follows easily. We need to show (23). (23) follows if

$$E(\delta_t) < 1 \tag{25}$$

To show (25) it suffices to show that

$$\Pr(|\beta_{t-1}| > \beta) > 0$$

But this follows easily by the independence of ϵ_{1t-1} and ϵ_{3t}, the fact that $\Pr(|\epsilon_{1t-1}| > r) > 0$ and the fact that $\Pr(|\epsilon_{3t-1}| > 2\beta) > 0$ for all finite β.

Proof of Theorem 2

In this appendix, we give a proof of the consistency of the threshold parameter r, which can be estimated via a grid search procedure. To simplify matters we suggest estimation of the threshold parameter via minimisation of the sum of squares function $S(\psi) = \sum_{i=1}^{T} z_t^2$ where z_t are given by

$$z_t = y_t - x_t\hat{\beta}_{t|t-1} - \hat{\epsilon}_{t|t-1}$$

and are the prediction errors of the model. Harvey (1989) (pp. 129) states that for univariate models such a minimisation is equivalent to maximum likelihood estimation. For simplicity we also assume $k = 1$ without loss of generality.

Following the proof of consistency of the threshold parameter estimates by Chan (1993) we see that three conditions need to be satisfied for consistency. Firstly, we need to show that the data y_t are geometrically ergodic and hence covariance stationary (Condition C1). Secondly we need to show that (Condition C2)

$$E_{\theta^0}(z_t|t-1)^2 < E_{\theta}(z_t|t-1)^2 \quad \forall \theta \neq \theta^0 \tag{26}$$

where θ^0 denotes the true parameter vector, and thirdly we need to show that (Condition C3)

$$\lim_{\delta \to 0} E\left(sup_{\theta \in B(\theta^0,\delta)}|z_t(\theta^0) - z_t(\theta)|\right) = 0 \tag{27}$$

where $B(a, b)$ is an open ball of radius b centered around a. C1 is needed for obtaining a law of large numbers needed for Claim 1 of Chan (1993). C1 can be obtained in a number of ways for a strictly exogenous geometrically ergodic processes x_t. For that we simply need geometric ergodicity of β_t. This can be easily obtained using the drift condition of Tweedie (1975)as in Theorem 1. A model for β_t that is easily seen to satisfy the drift condition is

$$\beta_t = I(|\beta^*_{t-1}| > \beta)\beta_1 + I(|\beta^*_{t-1}| < \beta)\beta_{t-1} + I(|\beta^*_{t-1}| < \beta)I(|\epsilon_{1,t-1}| > r)\epsilon_{2,t-1} \tag{28}$$

where $\beta^*_{t-1} = \beta_{t-1} + I(|\epsilon_{1,t-1}| > r)\epsilon_{2,t-1}$ and $\beta > \beta_1$ are finite constants. This model simply restricts the process β_t to return to a prespecified level β_1 if its expected value at time $t-1$ exceeds β. A wide variety of other models, such as the closely related but simpler model given in (22), are possible.

We need condition C2 to get a similar expression to (3.7) of Chan (1993) and condition C3 to prove Lemma 1 of Chan (1993). Condition C3 is a stochastic equicontinuity type

condition and is particularly important in view of the discontinuity involved with respect to the threshold parameter.

To prove condition C2 we focus on the general state space in the main body of the paper repeated here for convenience.

$$y_t = X_t b_t \tag{29}$$

$$b_t = A_t(\theta) b_{t-1} + \eta_t \tag{30}$$

We assume that the parameters of interest appear only in the matrix A_t. This is only for notational convenience. The proof can easily go though if the parameters also appear in the variance of η_t. We have that

$$z_t(\theta) = y_t - \hat{y}_{t|t-1}(\theta) = X_t b_t - X_t b_{t|t-1}(\theta) = \tag{31}$$

$$X_t A_t(\theta^0) b_{t-1} + X_t \eta_t - X_t \hat{b}_{t|t-1}(\theta) = X_t A_t(\theta^0) b_{t-1} - X_t A_t(\theta) \hat{b}_{t-1|t-1}(\theta) + X_t \eta_t \tag{32}$$

It is clear that the value of θ enters recursively through $\hat{b}_{t-i|t-i}(\theta)$. But for showing C2 it suffices to show that $E_{\theta^0}(z_t|t-1)^2 < E_\theta(z_t|t-1)^2$ for the case where θ^0 enters in $\hat{b}_{t-1|t-1}(\theta)$ both for $z_t(\theta)$ and $z_t(\theta^0)$.

So let us define

$$\tilde{z}_t(\theta) = X_t A_t(\theta^0) b_{t-1} - X_t A_t(\theta) \hat{b}_{t-1|t-1}(\theta^0) + X_t \eta_t \tag{33}$$

and

$$\tilde{z}_t(\theta^0) = X_t A_t(\theta^0) b_{t-1} - X_t A_t(\theta^0) \hat{b}_{t-1|t-1}(\theta^0) + X_t \eta_t = X_t A_t(\theta^0)(b_{t-1} - \hat{b}_{t-1|t-1}(\theta^0)) + X_t \eta_t \tag{34}$$

If we show that $E(\tilde{z}_t(\theta)|t-1)^2 > E(\tilde{z}_t(\theta^0)|t-1)^2$ then C2 is proven. But, noting that A_t depends only on data available up to $t-1$ and that

$$E((b_{t-1} - \hat{b}_{t-1|t-1}(\theta^0))(b_{t-1} - \hat{b}_{t-1|t-1}(\theta^0))') = P_{t-1|t-1}$$

we get

$$E(\tilde{z}_t(\theta^0)|t-1)^2 = A_t(\theta^0)' P_{t-1|t-1} A_t(\theta^0) + \Sigma_\eta$$

Also

$$\tilde{z}_t(\theta) = X_t A_t(\theta^0) b_{t-1} - X_t A_t(\theta^0) \hat{b}_{t-1|t-1}(\theta^0) + X_t A_t(\theta^0) \hat{b}_{t-1|t-1}(\theta^0) - X_t A_t(\theta) \hat{b}_{t-1|t-1}(\theta^0) + X_t \eta_t =$$

$$X_t A_t(\theta^0)(b_{t-1} - \hat{b}_{t-1|t-1}) + \hat{b}_{t-1|t-1}(X_t A_t(\theta^0) - X_t A_t(\theta)) + X_t \eta_t$$

Noting that $\hat{b}_{t-1|t-1}$ is fixed given data at $t-1$ gives

$$E(\tilde{z}_t(\theta)|t-1)^2 = A_t(\theta^0)' P_{t-1|t-1} A_t(\theta^0) + \Sigma_\eta +$$

$$(X_t A_t(\theta^0) - X_t A_t(\theta)) b_{t-1|t-1} b'_{t-1|t-1} (X_t A_t(\theta^0) - X_t A_t(\theta))'$$

Hence, C2 is proven.

We move on to condition C3. We show this result for z_2 assuming without loss of generality that the initial conditions are given by $b_0 = 0$ and $P_0 = 0$. Then it is easy to show the same result for any t working recursively. So we have to show that

$$\lim_{\delta \to 0} E\left(sup_{\theta \in B(\theta^0, \delta)} |z_2(\theta^0) - z_2(\theta)|\right) = 0$$

or equivalently that

$$\lim_{\delta \to 0} E\left(sup_{\theta \in B(\theta^0, \delta)} |\tilde{z}_2(\theta^0) - \tilde{z}_2(\theta)|\right) = 0$$

Given (33) and (34) we need to show that

$$\lim_{\delta \to 0} E\left(sup_{\theta \in B(\theta^0, \delta)} |b_{1|1}(A_1(\theta^0) - A_1(\theta))|\right) = 0$$

or

$$\lim_{\delta \to 0} E\left(sup_{\theta \in B(\theta^0, \delta)} |(A_1(\theta^0) - A_1(\theta)|\right) = 0$$

We use a simple model for A_t to illustrate the proof although more complicated models can be similarly treated. We use $A_t(\theta) = A_t(r) = I(|\epsilon_t| > r)$. Then we need to show that

$$\lim_{\delta \to 0} E\left(sup_{r \in B(r^0, \delta)} (I(|\epsilon_1| > r) - I(|\epsilon_1| > r^0))\right) = 0$$

This is simply equal to $Pr(|\epsilon_t| \in (r, r^0))$ where we have assumed with loss of generality that $r > r^0$. But

$$\lim_{r \to r^0} Pr(|\epsilon_t| \in (r, r^0)) = 0$$

proving C3.

Figure 6: Results from the importance sampling algorithm without leverage effects. The first panel reports the smoothed estimate of h_t. The second panel reports the smoothed estimate of β_t from the Gaussian approximating model. The third panel reports the final smoothed estimate of β_t. The final panel reports the actual return data.

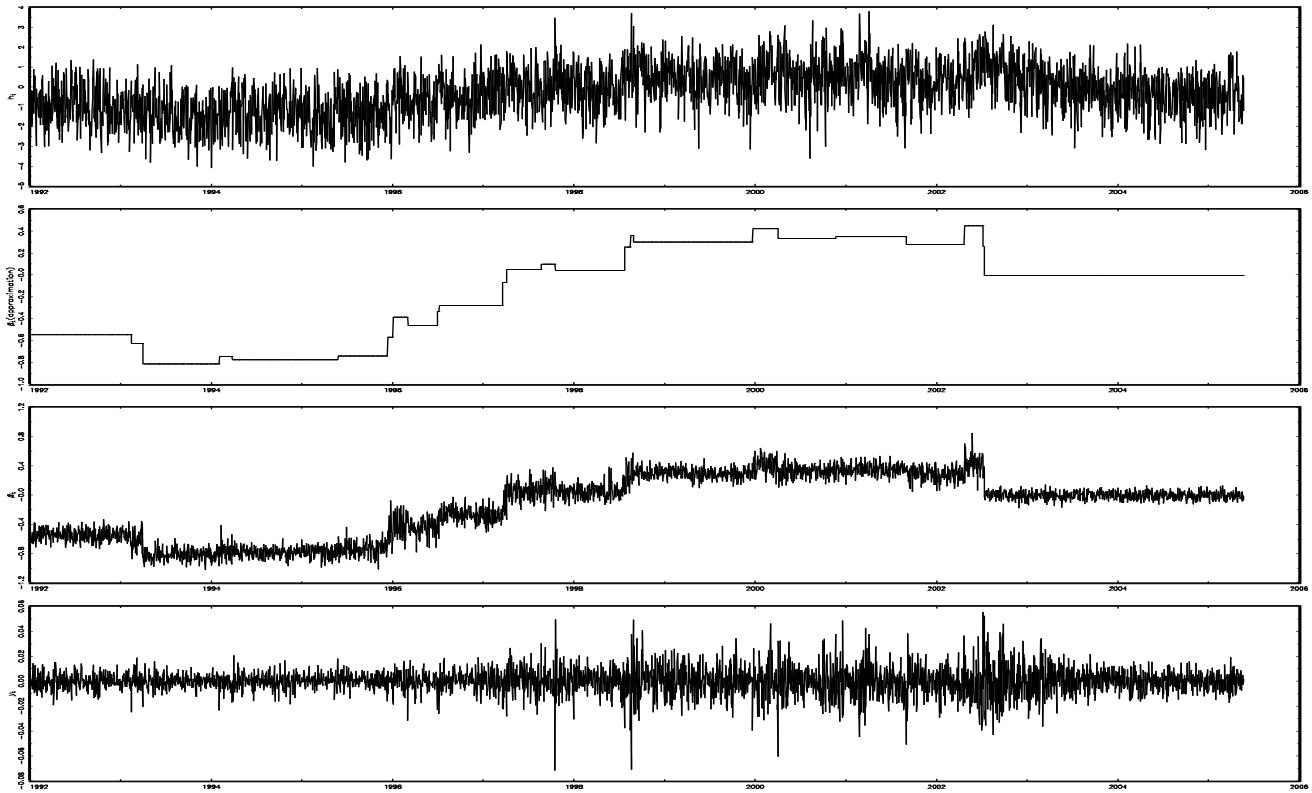
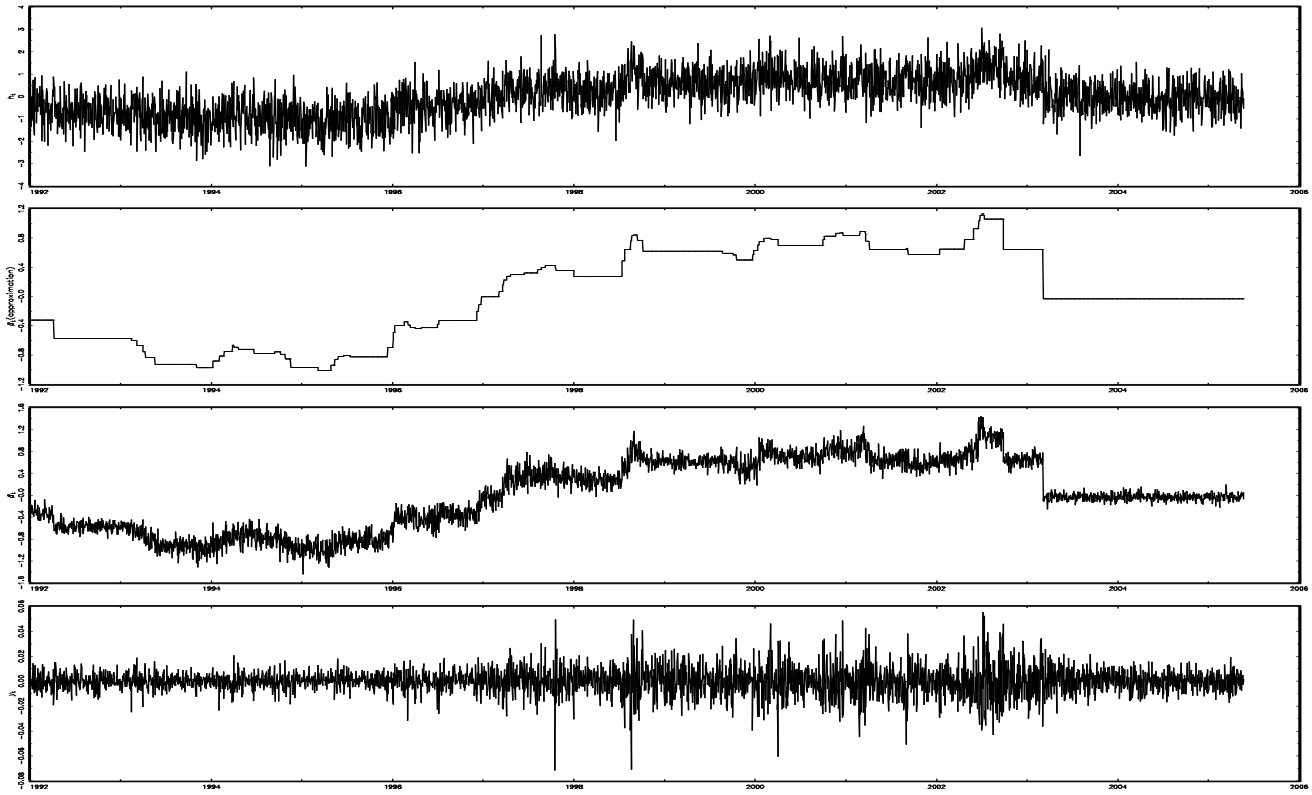


Figure 7: Results from the importance sampling algorithm with leverage effects. The first panel reports the smoothed estimate of h_t. The second panel reports the smoothed estimate of β_t from the Gaussian approximating model. The third panel reports the final smoothed estimate of β_t. The final panel reports the actual return data.



This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm