Stochastic Volatility Driven by Large Shocks

George Kapetanios∗
Queen Mary, University of London

Elias Tzavalis†
Department of Economics,
Athens University of Economics and Business

January, 2007

Abstract

This paper presents a new model of stochastic volatility which allows for infrequent shifts in the mean of volatility, referred to as structural breaks in the literature. 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. It can be also used to distinguish long-term shifts in volatility due to large pieces of bad or good news arriving in the market from ordinary volatility shocks.

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

∗Department of Economics, Queen Mary University of London, Mile End Road, London E1 4NS, UK. email: G.Kapetanios@qmul.ac.uk. Tel: +44(0)20 78825097.
†Department of Economics, Athens University of Economics and Business, Athens 104 34, Greece. email: etzavalis@aeub.gr.
1 Introduction

There is recently considerable evidence indicating the existence of structural breaks in the conditional variance process (volatility) of many economic or financial series. These breaks constitute discontinuous shifts in volatility which appear to be associated with extraordinary economic events, such as financial crises, monetary regime changes and exchange rate realignments (see Diebold and Pauly (1987), Lamoureux and Lastrapes (1990), Tzavalis and Wickens (1995), Hamilton and Lin (1996), Diebold and Inoue (2001), Ang and Bekaert (2002), Sensier and Van Dijk (2004), Mikosch and Starica (2004) and Morana and Beltratti (2004), \textit{inter alia}). The literature recognize such events as large shocks, or pieces of news, arriving in the market at a given point in time. If these shocks are not accounted for, they tend to overstate 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 (1975).

To capture the long-term effects of these type of shocks in volatility, in this paper we suggest a parametric model of structural breaks in volatility which are driven by stock return innovations which are larger in size than a threshold parameter. Our model can disentangle the long from the short-term shifts in volatility caused by stock market news. The first can be attributed to the large shocks of the market, while the second to the ordinary volatility shocks. Separating these two different type of shifts in the volatility function will have important implications for long-run portfolio management and hedging, as it should bring more focus on controlling large risks based on the long-term shifts in volatility leaving aside its short-term ones. Our model can also be used to study if the well known financial leverage effects (see Black (1976), Christie (1992) and more recently Yu (2005)), i.e. the negative relationship between volatility and stock market news, can be associated with larger pieces of stock market news rather than smaller or medium sized ones, as is often assumed in the literature.

From the point of view of the literature on structural breaks, the suggested model captures two essential features of breaks: their rarity and varying magnitude, over time. In other words, 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 in separating the long from the short-term shifts in volatility.

The paper is organised as follows. Section 2 presents our model and discusses some of its properties. Section 3 suggests alternative estimation procedures of the model. In Section 4 we report the results of a Monte Carlo exercise assessing the performance of the suggested estimation procedures to adequately trace structural breaks in volatility generated by our model. 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 returns are 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

\[ y_t = \mu + \sigma (e^{h_t})^{1/2} \epsilon_t, \]  

with

\[ h_t = \beta_{t-1} + \gamma h_{t-1} + \eta_{1,t} \]  

and

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

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 \( \eta_{1,t}, \eta_{2,t} \sim NID(0, \sigma_{\eta_i}^2), i = 1, 2 \), are innovations (shocks) which can be allowed to be correlated. \( I(A_t) \) is an indicator function taking the value 1 if the event \( A = \{|\epsilon_{t-1}| > r\} \) occurs, where \( r \) is a threshold parameter, and zero otherwise. The events captured by \( A_t \) can be thought of as reflecting a large piece of positive or negative news (shock) arriving at time \( t - 1 \) in the market. These news are
sometimes recognized as outliers in the level of series $y_t$.

The model of stochastic volatility (SV) given by equations (1)-(3) allows for long-term shifts in volatility $h_t$ endogenously driven by shocks $\epsilon_{t-1}$, at time $t-1$ that are larger than a threshold parameter $r$. These shifts, parametrically captured through the process of the changes in $\beta_t$ given by (3) in our model, accord with the common perception of structural breaks in volatility, referred to in the introduction.$^1$ There are many economic reasons for which larger stock market innovations $\epsilon_{t-1}$ can cause structural breaks in the level of market volatility $h_t$. These can reflect changes in agents’ beliefs about the optimal asset allocation due to monetary regime changes or realignments, financial market crises, institutional changes and fads or other market events. The long-term shifts in volatility considered by our model through the stochastic process of $\beta_t$ can reflect persistent financial leverage effects that large pieces of innovations, $\epsilon_{t-1}$, may have on future volatility. As was stated in the introduction, in the stochastic volatility literature these effects are not separated into long and short-term ones. Both of them are captured by the correlation between the mean and volatility innovations $\epsilon_t$ and $\eta_{1,t}$, of all sizes, respectively.

One interesting feature of our model is that the specification of the stochastic break process $\beta_t$ allows for both the timing and magnitude of breaks to be stochastic. The timing of a structural break in $\beta_t$ is controlled by innovations $\epsilon_{t-1}$ and, more specifically, depends on the occurrence of the event $A_t = \{ |\epsilon_{t-1}| > r \}$, 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 which consider long-term, or instant, shifts in volatility of fixed magnitude.$^2$ 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 $\epsilon_{t-1}$ can have an impact on $h_t$. As such, it also allows for the possibility that not all the large innovations $\epsilon_{t-1}$

\[ \beta_t = \beta_{t-1} + I(\epsilon_{t-1} > r_1)\eta_{2,t}^+ + I(\epsilon_{t-1} < r_2)\eta_{2,t}^-, \]

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 $\epsilon_{t-1}$ to have asymmetric effects on $h_t$.

$^1$Note that by specifying $\beta_t$ as

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

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 $\epsilon_{t-1}$ to have asymmetric effects on $h_t$.

$^2$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 $\beta_t$ is given as $\beta_t = \gamma \{ I(\epsilon_{t-1} < 0) - E[I(\epsilon_{t-1} < 0)] \}$.
have an impact on volatility. Note that, when $\sigma_n^2 = 0$, large values of innovations $\epsilon_{t-1}$ do not cause any structural change in $\beta_t$. In this case, our model reduces to the standard SV model, with no breaks. Finally, note that another interesting extension of our model would be towards a multivariate direction, where the innovations driving the changes in $\beta_t$ can be allowed to come from different sources than $y_t$.

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 $\beta_t$ grows with the time-interval of the data. If stationarity of $h_t$ is a desirable property of the data, then stationarity of $\beta_t$ would be required for this. There are a number of restrictions which can be imposed on $\beta_t$ to make it stationary (see Cogley and Sargent (2002)). A straightforward one is the following

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

where

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

This condition implies that $\beta_t$ is bounded by $\beta$ and, hence, it renders $h_t$ stationary, too. 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 $\beta_t$. 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 straightforward. 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

---

3Further restrictions could be placed on the process $\beta_t$ so that, if the bound $\beta$ is exceeded, the process returns to some prespecified level. We do not advocate a particular mechanism for making the process $\beta_t$ stationary. We simply wish to indicate that there exist specifications which give a stationary $\beta_t$ process. The exact specification of the process may be left to the empirical researcher depending on their priors on the particular issue at hand.
of Harvey, Ruiz, and Shephard (1994) in writing the measurement equation of the model in logarithmic form as
\[
\ln \hat{y}_t^2 = \left[ \ln \sigma^2 + E(\ln \epsilon_t^2) \right] + h_t + u_t
\]
where \( \hat{y}_t = y_t - 1/T \sum_{t=1}^{T} y_t \) 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) has a linear transition equation and can be made amenable to analysis by the standard Kalman filter. One adjustment which is needed to this end involves substituting \( I(\hat{\epsilon}_{t-1|t-1} > r) \) for \( I(|\epsilon_{t-1}| > r) \), where \( \hat{\epsilon}_{t-1|t-1} \) denotes the conditional expectation of \( \epsilon_{t-1} \) given the past values of \( \hat{y}_1, \ldots, \hat{y}_{t-1} \). Estimates of \( |\hat{\epsilon}_{t-1|t-1}| \) can be provided through equation (1), based on estimates of the state variables \( h_{t-1} \) and \( \beta_{t-1} \) at time \( t-1 \). In particular, since, on assuming Gaussianity for \( \epsilon_t \), \( E(\ln \epsilon_t^2) = -1.27 \), we easily get that \( |\hat{\epsilon}_{t-1|t-1}| = \sqrt{\exp(\{\hat{\epsilon}_{t-1|t-1} \} + 1.27)} \). 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 \( z_t = \ln \hat{y}_t^2 - 1/T \sum_{t=1}^{T} \ln \hat{y}_t^2 \). Under the above assumptions, we can write model (1)-(3) in a state space form as
\[
\begin{align*}
    z_t &= X_t b_t + u_t, \quad t = 1, \ldots, T, \\
    b_t &= A_t b_{t-1} + R_t \eta_t \quad \eta_t \sim i.i.d. N(0, \Sigma_{\eta})
\end{align*}
\]
where \( X_t = (1, 0)' \), \( b_t = (h_t, \beta_t)' \), \( \eta_t = (\eta_{1,t}, \eta_{2,t})' \),
\[
A_t = \begin{pmatrix} \gamma & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad R_t = \begin{pmatrix} 1 & 0 \\ 0 & I(|\hat{\epsilon}_{1,t-1|t-1}| > r) \end{pmatrix}
\]

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 \( I_{t-1} \) 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:

$$\hat{b}_{t|t-1} = A_t \hat{b}_{t-1}$$  \hspace{1cm} (10)
$$\hat{P}_{t|t-1} = A_t \hat{P}_{t-1} A_t' + R_t \Sigma \eta R_t,$$

known as the prediction equations, and

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

known as the updating equations, where

$$f_t = X_t' P_{t|t-1} X_t + \zeta_t$$  \hspace{1cm} (12)

(see, Hamilton (1994), \textit{inter alia}) and $\zeta_t = E(u_t^2)$. By the normality assumption for $\epsilon_t$ 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 $L(r)$, can be written in terms of the prediction errors $v_t = y_t - X_t' \hat{b}_{t|t-1}$ as

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

$L(r)$ 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 $I_t$ known as filter estimates, and (ii) estimates conditional on the information of the whole sample, denoted as $I_T$, 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

$$\hat{b}_{t|T} = \hat{b}_t + P^*_t (\hat{b}_{t+1|T} - A_{t+1} \hat{b}_t)$$  \hspace{1cm} (14)
and

$$P_{t|T} = P_t + P^*_t (P_{t+1|T} - P_{t+1|t}) P^*_t$$  \hspace{1cm} (15)

where $P^*_t = P_t A_{t+1} P_{t+1|t}^{-1}$. The filtered estimates of $b_t$, denoted as $\hat{b}_{t|t}$, can reveal agents’ perceptions about the current state of $\beta_t$ and $h_t$ in the economy, at time $t$. The set of smoothed
estimates of $b_t$, denoted as $\hat{b}_{0:T}$, can be used to statistically appraise the impact of large innovations $\epsilon_t$ on $\beta_t$ 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 $\eta_{1,t}, \eta_{2,t} \sim NID(0, \sigma^2_{\eta_1})$, 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 $\epsilon_{t-1}$ 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 $L(r)$ 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 next theorem 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 and stationary 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, e.g., 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), \( \epsilon_1 \), such as its 99-th centile.\(^4\)

As estimation of model (1)-(3) through the Kalman filter described above is suboptimal given that the error term \( u_t \) is not expected to be Gaussian, in what follows we suggest an alternative estimation procedure based on importance sampling along the lines of Durbin and Koopman (2001). The relative performance of the two estimation procedures in small samples will be assessed in the next section through a Monte Carlo study.

Let \( b = (b_1, \ldots, b_T) \), \( \tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_T) \) 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

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

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 \( b^{(1)}, \ldots, b^{(B)} \). Then, an estimator for \( E(b|\tilde{y}) \) based on importance sampling is given as

\[
\hat{b} = \frac{\sum_{i=1}^{B} b^{(i)} w(b^{(i)}, \tilde{y})}{\sum_{i=1}^{B} w(b^{(i)}, \tilde{y})}
\]

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 \( \prod_{t=1}^{T} p(\tilde{y}_t)p(\tilde{y}_{t-1}|b_t) \), where \( \tilde{y}_t = R_t \tilde{y}_t \). Due to the discontinuity of \( p(\tilde{y}_t) \) 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

\(^4\)That is, rather than using \( I(\mid \tilde{\epsilon}_{1,t-1} \mid > r) \) as the trigger of a break in \( \beta_t \), one can use \( I(|| \tilde{\epsilon}_{1,t-1} \mid > r/\sigma_{\epsilon_1} \) = \( I(|| \tilde{\epsilon}_{1,t-1} \mid > r^* \), where \( r^* = r/\sigma_{\epsilon_1} \) and \( \sigma_{\epsilon_1} \) is the standard deviation of \( \epsilon_{1,t} \). This is simply a standardisation. However, if we assume, e.g. normality of \( \epsilon_{1,t} \), the grid search can then focus on a few quantiles of the normal distribution as values for \( r^* \). So for example the grid search can examine the 0.5%, 1.0% and 2.0% quantiles of the normal distribution which can considerably reduce the search time. This also immediately implies a frequency for the occurrence of breaks. We do not examine this strategy further in the rest of the paper. But, we feel that, in cases where the standard grid search is empirically problematic, it might be useful.
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 the initial value \( \bar{\theta}_t = 0 \) as

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

Then, define

\[
\tilde{y}_t = \bar{\theta}_t - \ddot{s}_t^{-1} \dot{s}_t, \quad \text{with} \quad \bar{\zeta}_t = \ddot{s}_t^{-1}
\]

Apply the Kalman filter and smoother as defined by (10), (11), (12), (14) and (15), setting \( z_t = \tilde{y}_t \) and \( \zeta_t = \bar{\zeta}_t \).\(^5\) This returns a value for \( \hat{b}_{i|T} \) 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 final value of \( \hat{b}_{i|T} \) 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 also investigate if this estimator 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 the estimation procedures suggested in the previous section to adequately capture structural breaks in volatility process \( h_t \) generated by our model. This is done for samples where the number of breaks is relatively small. Our exercise can also show if our model can generate patterns of stochastic volatility breaks as those observed in practice.

\(^5\)It is straightforward to show that for the simple stochastic volatility model

\[
\tilde{y}_t = \bar{\theta}_t + 1 - \frac{\exp(\bar{\theta}_t)}{(\tilde{y}/\sigma)^2} \quad \text{and} \quad \bar{\zeta}_t = \frac{2\exp(\bar{\theta}_t)}{(\tilde{y}/\sigma)^2}.\]
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 \( \beta_t \) 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 \( \beta_t \), in our experiments we generate data according to model (1)-(3) where \( \gamma = 0.3, \sigma_{\eta_1}^2 = 1, \sigma_{\eta_2}^2 = 0.25 \) and the initial value of \( \beta_t \) is set to \( \beta_0 = 0 \). 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 \( \beta_t \) and the smoothed estimates of \( \beta_t \) obtained by the Kalman filter, denoted as \( \text{Corr}(\beta_t, \hat{\beta}_{KF}^T) \), 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 based on \( \hat{b}_{uT} \) at the end of the iterations, using relationship (18). The former is denoted as \( \text{Corr}(\beta_t, \hat{\beta}_{IS}^T) \), while the latter as \( \text{Corr}(\beta_t, \hat{\beta}_{AIS}^T) \).

To better see how closely the suggested estimation procedures can capture structural breaks in \( \beta_t \) over the sample, we also report pictorial results for particular replications; see Figures 1 to 4. 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 \( \beta_t \).

<table>
<thead>
<tr>
<th>(( r, T ))</th>
<th>(1.96, 500)</th>
<th>(1.96, 2000)</th>
<th>(2.24, 500)</th>
<th>(2.24, 2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( nb )</td>
<td>30</td>
<td>120</td>
<td>12</td>
<td>50</td>
</tr>
<tr>
<td>\text{Corr}(\beta_t, \hat{\beta}_{KF}^T)</td>
<td>0.740</td>
<td>0.906</td>
<td>0.566</td>
<td>0.826</td>
</tr>
<tr>
<td>\text{Corr}(\beta_t, \hat{\beta}_{IS}^T)</td>
<td>0.777</td>
<td>0.928</td>
<td>0.479</td>
<td>0.770</td>
</tr>
<tr>
<td>\text{Corr}(\beta_t, \hat{\beta}_{AIS}^T)</td>
<td>0.692</td>
<td>0.905</td>
<td>0.403</td>
<td>0.749</td>
</tr>
</tbody>
</table>

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 \( \beta_t \) 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 \(\beta_t\) and their true values reaches its highest value which is close to 90\%, using either the Kalman filter or 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 a number of \(nb = 30\) 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. \((r, T) = \{(1.96, 500), (1.96, 2000), (2.24, 2000)\}\)), the two 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. As expected, this improves greatly with the sample size.

The above conclusions can be confirmed by inspecting the pictorial output reported in Figures 1 to 4. These figures show that the smoothed estimates of \(\beta_t\) obtained through the two different estimation procedures track quite well the true shifts in the volatility process and its nonlinear long-term trend, when \(T\) is large and/or the number of breaks per \(T\) is adequate enough [see Figures 1, 2, and 4]. This is true for all estimation procedures considered and holds even for the case of \(T\) being small and the number of breaks per \(T\) being also small.

5 Empirical Application

As an empirical application of our SV model, given by equations (1)-(3), in this section we trace out possible structural breaks that occurred in the volatility of the S&P 500 index return 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 collapse of Enron and WorldCom corporations.

In our empirical exercise, we provide results based on both estimation procedures suggested in Section 3. Firstly, we estimate the model using the linear Kalman filter algorithm. A grid search for the value of the threshold parameter $r$ implying a probability of 5%, 2.5%, 2% and 1.5% for the event $\mathcal{A} = \{|\epsilon_{t-1}| > r\}$ occurring, suggests that the threshold value corresponding to 2.5% best describes the data since it corresponds to the largest value of the loglikelihood. Second, we estimate our model based on the importance sampling procedure, using the above value of the threshold parameter. We provide two different sets of results for this procedure. The first is based on a version of our model that does not allow for structural breaks in $h_t$, but allows for a possible correlation between innovations $\epsilon_t$ and $\eta_{1,t}$.
denoted by $\rho$. This specification constitutes the standard SV model allowing for leverage effects. Within our framework, it can be estimated by setting $\sigma_{\eta_2} = 0$. The second set of the importance sampling results are based on the full specification of our model, which also allows for correlation between $\epsilon_t$ and $\eta_{1,t}$. This correlation coefficient can capture any short-term leverage effects left unaccounted by process (3), filtering the long-term effects large innovations $\epsilon_t$ on volatility function $h_t$. Note that these short-term leverage effects can not be handled by estimating our model through the Kalman filter based on the linearized system of equations (8)-(9). To simplify the above estimation procedures, we have estimated $\sigma^2$ as $\exp\left(\frac{1}{T}\sum_{t=1}^{T} \ln \tilde{y}_t^2 + 1.27\right)$, where $E(\ln \epsilon_t^2) = -1.27$. This is an important simplification, as numerical maximization 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 maximization should be kept at a minimum.

To allow for possible correlation between innovations $\epsilon_t$ and $\eta_{1,t}$, we have slightly modified the initial specification of our model following Koopman (2005). In particular, we use the following version of our model

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

with

$$h_t \equiv \beta_{t-1} + \gamma h_{t-1} + \sigma m\{\eta_{1,t} + \eta_{3,t}\}$$  \hspace{1cm} (20)

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

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

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

\[
\hat{s}_t = \left( \frac{\partial^2 s_t}{\partial h_t \partial h_{t+1}} \right) = \left( \frac{1}{4\sigma^2 \bar{y}_t} \left( 2 \bar{y}_t^2 \beta - \bar{y}_t \beta \sigma \kappa \eta_{3,t} \varepsilon_{2}^2 \varepsilon_t \right) \right)
\]

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

**Table 2: Parameter Estimates**

<table>
<thead>
<tr>
<th>Model</th>
<th>( \sigma_{m2} )</th>
<th>( \sigma_{n2} )</th>
<th>( \gamma )</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV( ^{KF} ), with breaks</td>
<td>0.949</td>
<td>0.188</td>
<td>0.374</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.158)</td>
<td>(0.059)</td>
<td>(0.162)</td>
<td>-</td>
</tr>
<tr>
<td>SV( ^{IS} ), with breaks</td>
<td>0.630</td>
<td>0.126</td>
<td>0.116</td>
<td>-0.028</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.0028)</td>
<td>(0.052)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>SV( ^{IS} ), with no breaks</td>
<td>0.723</td>
<td>-</td>
<td>0.633</td>
<td>-0.206</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.012)</td>
<td>(0.015)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Standard errors are reported in parentheses.

The results of both estimation procedures are reported in Table 2. From these results and Figures 5-7, presenting estimates of the state variables of breaks \( \beta_t \) and volatility \( h_t \) based on the alternative estimation procedures, one can draw a number of interesting conclusions. First, both the Kalman filter and importance sampling estimates of \( \sigma_{n2} \) provide evidence that there are persistent structural shifts in \( h_t \), driven by large innovations in the S&P500 stock return. This can be supported by comparing the values of the loglikelihood function of the models with and without the break process, estimated through the importance sampling procedure. These values are found to be 28.97 and -639.42, respectively, suggesting that the model with the breaks provides a much better fit of the data than the standard SV model, with no breaks. Further support for our model can be obtained by inspecting the smoothed estimates of \( \beta_t \) and \( h_t \), given by all different estimation procedures and graphically presented by Figures 5-7. These figures clearly indicate that the stochastic volatility function is subject to a number of distinct breaks in its intercept \( \beta_t \), over the whole sample. These are of different size and follow a nonlinear long-term trend, as our model predicts. These results are consistent across the different estimation procedures adopted. As was expected, the Kalman filter and the approximating importance sampling estimates of \( \beta_t \) and \( h_t \) are smoother than those based on the final estimates of the importance sampling procedure.

As the above figures show, during our sample period, there are persistent cyclical shifts in \( \beta_t \) (and, hence, \( h_t \)). Until the middle of nineties, \( \beta_t \) remains almost stable around its lowest level over the whole sample. After that period, it steadily increases and reaches its highest
level in year 2002 and, then, it starts declining towards a lower lever. Both the increasing and decreasing parts of the process $\beta_t$ are triggered by a sequence of negative or positive large shocks in the stock index return after the middle of the 90’s and in the beginning of year 2003, respectively. Note that these shifts in $\beta_t$ are not abrupt as is assumed by other modeling procedures of structural breaks volatility, but smoother. Our estimates show that it takes some years for volatility to shift from its lowest to its highest level.

To investigate if the above long-term shifts in volatility correspond to those estimated through dummy intervention procedures, we have carried out the Bai and Perron (1998) unit root tests for the log-squared return $\ln \tilde{y}_t^2$. These tests can trace out multiple shifts in stochastic volatility, which, if ignored, tend to produce spurious evidence of unit roots in its level. The results of these tests have shown that there are three cyclical shifts of volatility through our sample. These correspond to the pattern of volatility shifts found by our model and are dated at the following dates: 19/11/96, 19/03/02 and 22/07/03. As can be seen from Figures 5-7, the first date is in the middle of the shift from the lowest level of volatility to the highest one, while the other two dates indicate the two consequent falls in volatility towards a lower level which occurred after 2002.

A second conclusion which can be drawn from the results of Table 2 is that the presence of the break process (3) in the volatility function $h_t$ driven by large stock return innovations substantially reduces the absolute value of the estimate of the correlation coefficient $\rho$ between $\epsilon_t$ and $\eta_{1,t}$, capturing leverage effects. This coefficient becomes almost equal to zero, when the full specification of our model is considered. This drop in the absolute value of $\rho$ reveals that documented leverage effects in volatility due to stock return news may be attributed to large pieces of stock market news rather than small ones. This is consistent with evidence in Campbell and Hentschel (1992). Finally, note that together with the drop in the absolute value of $\rho$, the presence of structural breaks in $h_t$ reduces substantially the autoregressive coefficient of the volatility process $\gamma$. This coefficient falls significantly from 0.63 to 0.11, when our model allows for breaks. The last result adds to the substantial evidence in the literature, as reviewed in our introduction, supporting that a high degree of persistency in volatility function $h_t$ can be attributed to the lack of accounting for structural changes in the volatility process, $h_t$. 

18
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 $\beta_t$. The final panel reports the actual return data.

6 Conclusions

This paper has introduced a new parametric model of stochastic volatility allowing for structural breaks in its intercept (mean). Following substantial evidence in the literature that shifts in volatility are associated with extraordinary events in economic series (e.g. stock returns), the suggested model considers a break process in the level of stochastic volatility which is endogenously driven by innovations (news) arriving in the stock market. These innovations are larger than a threshold parameter in absolute value. The timing and size of the breaks that our model considers are stochastic in nature, and thus can allow for random shifts in volatility. Apart from being a stylized fact, there may exist a number of theoreti-
ical reasons that large pieces of news arriving in the market can cause long-term effects in stochastic volatility. These can reflect the longer term financial leverage effects that are documented in the literature.

To estimate the model, the paper suggests two different procedures. The first is based on the importance sampling method. This can handle the nonlinear nature of the model and possible non-Gaussianity of the error terms. The second method is based on the Kalman filter which linearises and assumes Gaussianity of the error terms. In a Monte Carlo exercise, the paper assess the ability of the above estimation procedures to sufficiently track a true break process, generated by our model. Our results show that both the importance sampling and Kalman filter estimation procedures perform satisfactorily provided that there is a reasonable number of observations for a given number of breaks. Finally, in an empirical application of our model to real data it is shown that it can capture smoother cyclical shifts in volatility. These shifts can be given the economic interpretation of a generalised form of leverage effects. One of the findings of our empirical exercise is that these effects seem to last for a substantial periods and can be attributed to large piece of bad, or good, news arriving in the market at a particular point in time.

References


Appendix

Proof of Theorem 1

We now prove strict stationarity for $h_t$, given by

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

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

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

where

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

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

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

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

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

which is finite as long as $\beta_t$ is strictly stationary and $E(\beta_{t-j-1})^\alpha < \infty$ and $E(\eta_{1,t-j})^\alpha < \infty$. Thus it suffices to prove that $\beta_t$ is strictly stationary and $E(\beta_{t-j-1})^\alpha < \infty$. $E(\beta_{t-j-1})^\alpha < \infty$ follows easily from strict stationarity and $E(\eta_{2,t})^\alpha < \infty$. Thus we only need to prove strict stationarity for $\beta_t$. To do that we prove geometric ergodicity of $\beta_t$, 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, $\beta_t$ is geometrically ergodic if there exists constants $0 < \vartheta < 1$, $B, L < \infty$, and a small set $C$ such that

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

$$E[\|\beta_t\| \mid \beta_{t-1} = d] \leq B, \quad \forall d \in C,$$  \quad (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 the following condition holds

\[
E(\delta_t) < 1.
\]  

(25)

To prove (25) it suffices to show that

\[
\Pr(|\beta_{t-1}| > \beta) > 0.
\]

This follows easily by the independence of \( \epsilon_{t-1} \) and \( \eta_{2,t} \), the fact that \( \Pr(|\epsilon_{t-1}| > r) > 0 \) and the fact that \( \Pr(|\eta_{2,t-1}| > 2\beta) > 0 \) for all finite \( \beta \).

**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 minimization of the sum of squares function \( S(\psi) = \sum_{t=1}^{T} v_t^2 \) where \( v_t = y_t - X_t' \hat{b}_{t-1} \) are the prediction errors of the model. For univariate models such a minimization is equivalent to maximum likelihood estimation (see Harvey (1989) (pp. 129)). 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 establish that the data \( y_t \) are geometrically ergodic and hence covariance stationary (Condition C1). Secondly, we need to prove (Condition C2)

\[
E_{\theta_0}(v_t|t-1)^2 < E_{\theta}(v_t|t-1)^2 \quad \forall \theta \neq \theta_0,
\]

(26)

where \( \theta_0 \) denotes the true parameter vector. Thirdly, we need to prove (Condition C3)

\[
\lim_{\delta \to 0} E \left( \sup_{\theta \in B(\theta_0, \delta)} |v_t(\theta_0) - v_t(\theta)| \right) = 0,
\]

(27)

where \( B(a, b) \) is an open ball of radius \( b \) centered around \( a \). These three conditions together imply the uniform convergence of the objective function \( S(\psi) \) to the limit objective function which is the key to establishing consistency. Condition C1 is required for obtaining a law of large numbers needed for Claim 1 of Chan (1993), and hence for convergence of the objective function. Condition C3 is needed for uniformity of the convergence and finally condition C2
is needed to show that the limiting objective function is minimized at the true parameter values. More specifically, 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.

Condition 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 $\beta_t$. This can be easily obtained using the drift condition of Tweedie (1975) as in Theorem 1. A model for $\beta_t$ that is easily seen to satisfy the drift condition is given by (22). An alternative could be the following

$$\beta_t = I(\|\beta_{t-1} - 1\| > \beta) \beta_1 + I(\|\beta_{t-1} - 1\| < \beta) I(\|\epsilon_{1,t-1}\| > r) \eta_{2,t-1},$$

where $\beta_{t-1} = \beta_{t-1} + I(\|\epsilon_{1,t-1}\| > r) \eta_{2,t-1}$ and $\beta > \beta_1$ are finite constants. This model simply restricts the process $\beta_t$ to return to a prespecified level $\beta_1$ if its expected value at time $t - 1$ exceeds $\beta$.

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

$$y_t = X_t b_t$$

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

(29)  

(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 through if the parameters also appear in the variance of $\eta_t$. We have that

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

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

(31)

(32)

It is clear that the value of $\theta$ enters recursively through $\hat{b}_{\theta|t-1}(\theta)$. To prove C2, it suffices to show that $E_{\theta^0} (v_t|t-1)^2 < E_{\theta} (v_t|t-1)^2$ for the case where $\theta^0$ enters in $\hat{b}_{\theta|t-1}(\theta)$ both for $v_t(\theta)$ and $v_t(\theta^0)$. To this end, let us define

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

(33)
and
\[ \hat{v}_t(\theta^0) = X_t A_t(\theta^0) b_{t-1} - X_t A_t(\theta^0) \hat{b}_{t-1|t-1}(\theta^0) + X_t \eta_t \] 
\[ = X_t A_t(\theta^0) (b_{t-1} - \hat{b}_{t-1|t-1}(\theta^0)) + X_t \eta_t. \] 

If we show that \( E(\hat{v}_t(\theta)|t-1)^2 > E(\hat{v}_t(\theta^0)|t-1)^2 \), then C2 is proven. This can be done as follows. First, 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, after taking conditional expectations of \( \hat{v}_t(\theta)^2 \), the following
\[ E(\hat{v}_t(\theta)^2|t-1) = A_t(\theta^0)' P_{t-1|t-1} A_t(\theta^0) + \Sigma_\eta. \]

Also note that
\[ \hat{v}_t(\theta) = X_t A_t(\theta^0) b_{t-1} - X_t A_t(\theta^0) \hat{b}_{t-1|t-1}(\theta^0) + X_t A_t(\theta^0) \hat{b}_{t-1|t-1}(\theta^0) - X_t A_t(\theta^0) + X_t \eta_t. \]

As \( \hat{b}_{t-1|t-1} \) is fixed given data at \( t-1 \), we get
\[ E(\hat{v}_t(\theta)^2|t-1) = A_t(\theta^0)' P_{t-1|t-1} A_t(\theta^0) + \Sigma_\eta + 
\]
\[ (X_t A_t(\theta^0) - X_t A_t(\theta)) b_{t-1|t-1} b_{t-1|t-1}' (X_t A_t(\theta^0) - X_t A_t(\theta))'. \]

Given the above results, this can be rewritten as
\[ E(\hat{v}_t(\theta)^2|t-1) = E(\hat{v}_t(\theta^0)|t-1) \]
\[ + (X_t A_t(\theta^0) - X_t A_t(\theta)) b_{t-1|t-1} b_{t-1|t-1}' (X_t A_t(\theta^0) - X_t A_t(\theta))'. \]

which proves that \( E(\hat{v}_t(\theta)|t-1)^2 > E(\hat{v}_t(\theta^0)|t-1)^2 \). Hence, C2 is proven.

We next 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. To prove condition C3, we need to show the following
\[ \lim_{\delta \to 0} E \left( \sup_{\theta \in B(\theta^0, \delta)} |v_2(\theta^0) - v_2(\theta)| \right) = 0, \]
or equivalently that
\[ \lim_{\delta \to 0} E \left( \sup_{\theta \in B(\theta^0, \delta)} |\hat{v}_2(\theta^0) - \hat{v}_2(\theta)| \right) = 0. \] 

(35)
Given (33) and (34), (35) follows if the following holds

\[ \lim_{\delta \to 0} E \left( \sup_{\theta \in B(\theta_0, \delta)} |b_{11}(A_1(\theta^0) - A_1(\theta))| \right) = 0, \]

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. Let \( A_t(\theta) = A_t(r) = I(|\epsilon_t| > r) \). 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 without loss of generality that \( r > r_0 \). Condition C3 follows immediately by noting that

\[ \lim_{r \to r_0} Pr(|\epsilon_t| \in (r, r_0)) = 0. \]
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 $\beta_t$ from the Gaussian approximating model. The third panel reports the final smoothed estimate of $\beta_t$. 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 $\beta_t$ from the Gaussian approximating model. The third panel reports the final smoothed estimate of $\beta_t$. The final panel reports the actual return data.