Journal of Econometrics 167 (2012) 473–482

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A semiparametric stochastic volatility model☆

Jun Yu

Sim Kee Boon Institute for Financial Economics, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore

ARTICLE INFO

Article history:
Available online 1 October 2011

JEL classification:
G12
C11
C15

Keywords:
Leverage effect
Simulated maximum likelihood
Laplace approximation
Spline

ABSTRACT

In this paper the correlation structure in the classical leverage stochastic volatility (SV) model is generalized based on a linear spline. In the new model the correlation between the return and volatility innovations is time varying and depends nonparametrically on the type of news arrived to the market. Theoretical properties of the proposed model are examined. The model estimation and comparison are conducted by Bayesian methods. The performance of the estimates are examined in simulations. The new model is fitted to daily and weekly US data and compared with the classical SV and GARCH models in terms of their in-sample and out-of-sample performances. Empirical results suggest evidence in favor of the proposed model. In particular, the new model finds strong evidence of time varying leverage effect in individual stocks when the classical model fails to identify the leverage effect.

1. Introduction

How volatility responds to return news has long been an active research topic; see Black (1976), Christie (1982), Engle and Ng (1993) and Wu and Xiao (2002) for a rather incomplete list of studies in the literature. Answer to this question has important implications for financial decision making and asset pricing. For example, predictability of volatility critically depends on the relationship between the return shock and volatility. Moreover, there are important implications of the relationship for portfolio selection and risk management (Bekaert and Wu, 2000) and for "betas" (Braun et al., 1995). Furthermore, an option contract would be substantially mis-priced when the relationship is misspecified (Duan, 1995).

It is now well accepted in the volatility literature that equity volatility responds asymmetrically to return news, namely, a piece of bad news has different impact on future volatility from the good news of the same magnitude. The most popular and convenient empirical method for examining the asymmetric volatility response is via some form of ARCH-type models. The motivation mainly comes from the so-called leverage hypothesis originally put forward by Black (1976). According to the leverage hypothesis, when bad news arrives, it decreases the value of a firm's equity and hence increases its leverage. Consequently, the equity becomes more risky and its volatility increases. Likewise the volatility decreases after good news arrives.

Volatility response can also be studied using stochastic volatility (SV) models. Unlike ARCH-type models, SV models specify volatility as a separate random process, which provides certain advantages over the ARCH-type models for modeling the dynamics of asset returns (Kim et al., 1998). The third method for studying volatility response is to use realized volatility; see, for example, Andersen et al. (2001, ABDE hereafter), Bandi and Reno (forthcoming) and Hansen et al. (2010). In this literature some important asymmetries are well documented in market-wide equity index returns but not in individual stocks. This observation leads some researchers to conclude that the significant asymmetries in equity index returns are due to volatility feedback effect but not leverage effect; see ABDE.

In the SV literature, the asymmetric volatility response is often studied by specifying a negative correlation between the return innovation and the volatility innovation. This classical leverage SV model was first estimated by Harvey and Shephard (1996). The model specification requires the correlation coefficient between the two innovations remains constant, regardless of price movements. On the other hand, Daouk and Ng (2007) reported evidence of stronger leverage effect in down markets than in up markets. Obviously, this empirical result cannot be explained by the classical leverage SV model with a constant leverage effect.

The central focus of the present paper is to provide a more general framework to investigate the asymmetric relationship

☆ I wish to thank the co-editors, referees, Yong Li, Peter Phillips, Neil Shephard, Liangjun Su, Anthony Tay, Yiu Kuen Tse, seminar participants at Singapore Management University, FEMES, NZESG, SESG, London School of Economics, FEMES, SETA2008, and CREST for comments on an earlier version of the paper. Excellent research assistantship from Chao Luo and Xin Chen are acknowledged. I gratefully acknowledge financial support from the Singapore Ministry of Education AcRF Tier 2 fund under Grant No. T206B4301-RS.

E-mail address: yujun@smu.edu.sg.

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.09.029

between volatility and return news in the context of SV models. Using the linear spline, we allow the correlation coefficient between the two innovations to be time varying and depend nonparametrically on the size and the direction of the previous price movement. Since our model nests the SV model with the constant leverage, we can easily check the validity of this classical specification. Empirical applications reveal strong evidence against the classical specification both in-the-sample and out-of-the-sample.

Our model extends the specification studied in Harvey and Shephard (1996), Yu (2005) and Omori et al. (2007). Following Meyer and Yu (2000), the Bayesian Markov chain Monte Carlo (MCMC) methods are used to estimate and compare alternative models. Our model is closely related to the model of Wu and Xiao (2002) where a flexible nonparametric model was used to relate the log implied volatility and the lagged return innovation. However, our work is different from Wu and Xiao in four aspects. First, Wu and Xiao is an ARCH-type model while ours is an SV. The two models do not nest each other. Although the model of Wu and Xiao allows for a very general news impact function, it assumes an additive functional form and cannot even nest the simplest SV model. Second, different nonparametric methods are employed. While we use the spline-based smoother, Wu and Xiao used the Nadaraya–Watson kernel method in a partial linear framework. One of the main advantages for the kernel method lies in its simpler theoretical analysis. However, the kernel method cannot be used in the context of SV due to the curse-of-dimensionality problem. Third, the relationship between return and log-volatility is in the physical measure in our study but is in the risk-neutral measure in theirs. The risk-neutral measure is more useful for pricing whereas the physical measure allows one to forecast volatility. Finally, volatility is latent in our method whereas Wu and Xiao assumed that the volatility of the US market index is well approximated by the volatility index, VIX. For individual stocks, VIX is no longer a valid approximation to the volatility.

Our model is somewhat related to that of Engle and Ng (1993) in the sense that the linear spline is used. However, we use the linear spline to model the correlation between the two innovations while Engle and Ng used it as a regression tool to relate volatility to the lagged return innovation. Robinson (1991) and others provided more general ARCH models. All the models are of an additive structure and hence do not nest ours. Finally, our model is related to Bandi and Reno (forthcoming) where the time varying leverage effect is estimated using a nonparametric method with intra-day data. Unlike Bandi and Reno who tie the strength of the leverage effect to the current level of volatility, we assume the driving factor for the time varying leverage is the lagged return.

The article is organized as follows. In Section 2 we introduce the semiparametric SV model and develop some statistical properties of the model. Section 3 discusses the MCMC methods for parameter estimation and for model comparison and documents the performance of MCMC in simulations. Empirical results based on US data are presented and discussed in Section 4. Section 5 concludes. Appendix proves the theorem.

2. The proposed SV model

Let y_t be the rate of return of a stock or a market portfolio in time period t, σ_t^2 be the conditional variance of y_t, $h_t = \ln \sigma_t^2$, ϵ_t be the return innovation. GARCH models specify a deterministic relationship between σ_{t+1}^2 and y_t (or ϵ_t). Different models coexist to capture the asymmetric volatility response. For example, EGARCH(1, 1) of Nelson (1991) assumes

$$h_{t+1} = \alpha + \varphi h_t + \beta_0 \epsilon_t + \beta_1 |\epsilon_t|, \tag{1}$$

where the asymmetry is induced by the term $\beta_0 \epsilon_t$. Threshold GARCH(1, 1) of Glosten et al. (1993) assumes

$$\sigma_{t+1}^2 = \alpha + \varphi \sigma_t^2 + \beta y_t^2 + \beta^* y_t^2 \mathbf{1}(y_t < 0), \tag{2}$$

where $\mathbf{1}(y_t < 0) = 1$ if $y_t < 0$ and 0 otherwise. In this model, the asymmetry is induced by $\mathbf{1}(\cdot)$. However, based on a nonparametric technique, Mishra et al. (2010) have found the evidence of further asymmetry in the residuals of fitted threshold GARCH(1, 1).

Engle and Ng (1993) introduced a partially nonparametric model of the form

$$\sigma_{t+1}^2 = \alpha + \varphi \sigma_t^2 + m(\epsilon_t) \tag{3}$$

where $m(\cdot)$ is an unknown function. Engle and Ng estimated $m(\cdot)$ using the linear spline

$$m(\epsilon_t) = \sum_{i=0}^{m^+} \theta_i \mathbf{1}(\epsilon_t > \tau_i)(\epsilon_t - \tau_i) + \sum_{i=0}^{m^-} \delta_i \mathbf{1}(\epsilon_t < \tau_{-i})(\epsilon_t - \tau_{-i}),$$

where τ_i are the predetermined knots associated with the linear spline.

In contrast to ARCH-type models, the SV models specify a stochastic relationship between σ_{t+1}^2 (or h_{t+1}) and y_t by using an additional innovation. It is very important to point out that the meaning of σ_{t+1}^2 in SV models is NOT the same as that in ARCH-type models. By assuming σ_{t+1}^2 is a conditional variance, ARCH-type models adopt the one-step-ahead prediction approach to volatility modeling. Whereas, due to the presence of an additional innovation in the state equation of SV, σ_{t+1}^2 is not measurable with respect to the natural filtration and hence is not a conditional variance. This difference has an important implication for the analysis of the news impact, which will be discussed in detail later.

To account for volatility asymmetry, the classical leverage SV model takes the form of

$$y_t = \mu_y + \sigma \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim \text{i.i.d. } N(0, 1), \tag{4}$$

$$h_{t+1} = \varphi h_t + \gamma v_t, \quad v_t \sim \text{i.i.d. } N(0, 1), \tag{5}$$

where $\text{corr}(\epsilon_t, v_t) = \rho$. Eq. (5) can be equivalently represented by

$$h_{t+1} = \varphi h_t + \gamma(\rho \epsilon_t + \sqrt{1-\rho^2} w_t), \tag{6}$$

where w_t is i.i.d. $N(0, 1)$ and $\text{corr}(\epsilon_t, w_t) = 0$. Consequently, we have

$$\begin{aligned} h_{t+1} &= \varphi h_t + \gamma \rho \epsilon_t + \gamma \sqrt{1-\rho^2} w_t \\ &= \varphi h_t + \rho \frac{\gamma}{\sigma} \exp(-h_t/2)(y_t - \mu_y) + \gamma \sqrt{1-\rho^2} w_t, \end{aligned} \tag{7}$$

implying that on average $\ln \sigma_{t+1}^2$ is a linear function in y_t. When $\rho < 0$, the linear function is downward sloping and this feature is often referred to as the leverage effect. Clearly the relationship between $\ln \sigma_{t+1}^2$ and y_t is independent of the sign and the size of ϵ_t and hence the leverage effect, captured by ρ, is a constant in this model.

There is ample evidence that the effect of bad news on volatility is different from the good news of the same magnitude. Using the firm level accounting data, Figlewski and Wang (2000) reported a more remarkable leverage effect in down markets than in up markets. A similar pattern of asymmetry found in Daouk and Ng (2007) using unleveled firm volatility. The evident suggests that a global linear relationship between $\ln \sigma_{t+1}^2$ and y_t may be too restrictive and there is a clear need for a more general SV model for the volatility asymmetry.

To introduce our semiparametric SV model, we first choose m knots, denoted by $\tau_1, \ldots, \tau_m$ with $\tau_1 > \cdots > \tau_m$, from the support of ϵ_t. Let τ_0/τ_{m+1} be the right/left bound of the support of ϵ_t.[1] That is, the support of ϵ_t is divided into $m+1$ intervals. Note that the sizes of the intervals need not be the same. The volatility equation is defined by

$$h_{t+1} = \varphi h_t + \gamma \sum_{i=1}^{m+1} (\rho_i \epsilon_t + \sqrt{1-\rho_i^2} w_t) \mathbf{1}(\tau_{i-1} \geq \epsilon_t > \tau_i), \quad w_t \sim \text{i.i.d. } N(0,1) \tag{8}$$

where $\text{corr}(\epsilon_t, w_t) = 0$. Together with Eq. (4), it defines our semiparametric SV model.

Let

$$v_t = \sum_{i=1}^{m+1} (\rho_i \epsilon_t + \sqrt{1-\rho_i^2} w_t) \mathbf{1}(\tau_{i-1} \geq \epsilon_t > \tau_i) \tag{9}$$

be the innovation in the variance equation. It can be shown that

$$v_t = \begin{cases} \rho_1 \epsilon_t + \sqrt{1-\rho_1^2} w_t & \text{if } \tau_0 \geq \epsilon_t > \tau_1 \\ \vdots & \\ \rho_{m+1} \epsilon_t + \sqrt{1-\rho_{m+1}^2} w_t & \text{if } \tau_m \geq \epsilon_t > \tau_{m+1}. \end{cases}$$

Obviously, the construction of v_t is based on the linear spline with the basis functions, $(x-\tau_1)_+, \ldots, (x-\tau_m)_+$, where x_+ is equal x if x is positive and 0 otherwise. See Ruppert et al. (2003) for a detailed account of spline smoothing.

When $\rho_i = \rho, \forall i$, $v_t = \rho\epsilon_t + \sqrt{1-\rho^2} w_t$ and the specification becomes the classical leverage SV model. In general, ρ_i can have different sizes and even different signs. Following the same approach to deriving (7), we have

$$h_{t+1} = \begin{cases} \varphi h_t + \rho_1 \dfrac{\gamma}{\sigma} \exp(-h_t/2)(y_t - \mu_y) & \\ \quad + \gamma \sqrt{1-\rho_1^2} w_t & \text{if } \tau_0 \geq \epsilon_t > \tau_1 \\ \vdots & \\ \varphi h_t + \rho_{m+1} \dfrac{\gamma}{\sigma} \exp(-h_t/2)(y_t - \mu_y) & \\ \quad + \gamma \sqrt{1-\rho_{m+1}^2} w_t & \text{if } \tau_m \geq \epsilon_t > \tau_{m+1}. \end{cases}$$

Clearly, on average $\ln \sigma_{t+1}^2$ is a piecewise linear function in y_t with kinks at the τ_is. Between τ_1 and $+\infty$ the slope of the linear function is ρ_1 while between τ_2 and τ_1 it is ρ_2. Below τ_m, the slope is ρ_{m+1}. In this model, the leverage effect is time varying and the magnitude of the leverage effect is determined by ϵ_t. We now establish statistical properties for the model.

Theorem 2.1. *Define the SV model by*

$$\begin{cases} y_t = \sigma \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim \text{i.i.d. } N(0,1) \\ h_{t+1} = \varphi h_t + \gamma \displaystyle\sum_{i=1}^{m+1} (\rho_i \epsilon_t + \sqrt{1-\rho_i^2} w_t) \\ \qquad \times \mathbf{1}(\tau_{i-1} \geq \epsilon_t > \tau_i), \quad w_t \sim \text{i.i.d. } N(0,1) \end{cases}$$

where $\text{corr}(w_t, \epsilon_t) = 0$ *and* $\tau_0 = +\infty$, $\tau_{m+1} = -\infty$. *Then* $\{y_t\}$ *and* $\{h_t\}$ *are covariance stationarity, strictly stationary and ergodic if and only if* $|\varphi| < 1$. *Also,* $\{y_t\}$ *possesses finite moments of arbitrary order and the expression for the moments of* y_t *is*

$$E(y_t^{2i-1}) = 0, \qquad E(y_t^{2i}) = \frac{(2i)!}{2^i i!} \sigma^{2i} G(i, \boldsymbol{\rho}, \gamma, \varphi), \quad i = 1, 2, \ldots,$$

where $G(s, \boldsymbol{\rho}, \gamma, \varphi)$ *is defined by*

$$G(s, \boldsymbol{\rho}, \gamma, \varphi) = \prod_{j=0}^{\infty} \Bigg\{ \exp\left(\frac{1}{2} s^2 \gamma^2 \varphi^{2j}\right) \Bigg[\Phi(\tau_m - s\gamma\varphi^j \rho_{m+1}) + 1 - \Phi(\tau_1 - s\gamma\varphi^j\rho_1) + \sum_{i=2}^{m} (\Phi(\tau_{i-1} - s\gamma\varphi^j\rho_i) - \Phi(\tau_i - s\gamma\varphi^j\rho_i)) \Bigg] \Bigg\},$$

with $\Phi(\cdot)$ *being the cumulative distribution function of* $N(0,1)$, $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{m+1})'$.

Remark 2.1. The closed form expression for moments facilitates calculations of all moments and model comparison. This result holds true for any value of m. When $m \to \infty$, the moments involve an infinite product and an infinite sum, and hence truncations are inevitable.

Remark 2.2. Two choices have to be made in the proposed model, m and τs. Ideally, one should allow m to increase with the sample size (such as $o(n)$). However, the larger the m, the more parameters in the model and hence the higher the computational cost. Essentially increasing m trades off smaller bias with larger variance. The reason that the variance increases with m is because less effective observations are used to estimate ρ_i with a larger m. To control the computational cost, we fix m in this paper. The choice of τs could be based on trial and error or more formally a model selection criterion. However, the exercise will be computationally expensive if a large set of τs is considered.

Remark 2.3. If $m = 1$, there are only two regimes. When we set $\tau_1 = 0$, $\text{corr}(\epsilon_t, v_t) = \rho_1$ if $\epsilon_t > 0$ and $\text{corr}(\epsilon_t, v_t) = \rho_2$ if $\epsilon_t \leq 0$. This model nicely nests the classical leverage effect model and is called *Spline1 SV* in this paper. The moments of y_t is given by

$$E(y_t^{2i-1}) = 0,$$

$$E(y_t^{2i}) = \frac{(2i)!}{2^i i!} \sigma^{2i} G_1(i, \rho_1, \rho_2, \gamma, \varphi), \quad i = 1, 2, \ldots,$$

where $G_1(s, \rho_1, \rho_2, \gamma, \varphi) = \prod_{j=0}^{\infty} \{\exp\left(\frac{1}{2}s^2\gamma^2\varphi^{2j}\right) [\Phi(s\gamma\varphi^j\rho_1) + \Phi(-s\gamma\varphi^j\rho_2)]\}$.

Remark 2.4. If $m = 2$ there are three regimes. This model is called *Spline2 SV* in this paper. It is known in the GARCH literature that when ϵ_t is very close to zero, volatility does not respond to ϵ_t in a significant manner (Engle and Ng, 1993). As a result, it is reasonable to choose τ_1 to be a small, positive number, τ_2 to be a small, negative number. However, if τs are too close to zero, there are too few observations to estimate ρ_2; if τ_1 (or τ_2) is too far away from zero, there are too few observations to estimate ρ_1 (or ρ_3). In the empirical applications, we choose $\tau_{1,2} = \pm 0.4$. Since $\Pr(\epsilon_t > 0.4) = \Pr(\epsilon_t < -0.4) = 34.5\%$, $\Pr(|\epsilon_t| > 0.4) = 31\%$, we have a nearly equal split of observations to estimate the ρs. A drawback with such a choice is that the *Spline2 SV* does not nest the *Spline1 SV*. Since our model comparison method is Bayesian-based, such a drawback does not impose any problem to us.

Nearly all the existing ARCH models assume an additive functional form to relate the conditional variance (or log-variance) to the return news and the lagged conditional variance. Such an additive structure greatly facilitates the news impact analysis. The news impact function (NIF), first introduced in Pagan and Schwert (1990) and extended by Engle and Ng (1993), treats the conditional variance as a function of the return news lagged one-period, holding constant the other lagged variables. Consequently,

[1] If the support of ϵ_t is the entire real line, then $\tau_0 = +\infty$ and $\tau_{m+1} = -\infty$.

in the ARCH-type models the functional form of the NIF is explicitly specified and determined solely by the term that contains the lagged return news and hence time invariant. For example, in the GARCH(1, 1), σ_{t+1}^2 is a quadratic function of ε_t which centered at the original, i.e.,

$$\sigma_{t+1}^2 = \alpha + \varphi\sigma_t^2 + \beta\sigma_t^2\varepsilon_t^2|_{\sigma_t^2=\overline{\sigma}^2} := \alpha^* + \beta^*\varepsilon_t^2,$$

where $\overline{\sigma}^2$ is the unconditional mean of σ_t^2. The NIF for EGARCH (1, 1) is

$$\sigma_{t+1}^2 = \begin{cases} \exp(\alpha^* + (\beta_0 + \beta_1)\varepsilon_t), & \text{if } \varepsilon_t \geq 0 \\ \exp(\alpha^* + (\beta_0 - \beta_1)\varepsilon_t), & \text{if } \varepsilon_t < 0. \end{cases} \quad (10)$$

The NIF for threshold GARCH(1, 1) is

$$\sigma_{t+1}^2 = \begin{cases} \alpha^* + \beta\bar{\sigma}_t^2\varepsilon_t^2, & \text{if } \varepsilon_t \geq 0 \\ \alpha^* + (\beta + \beta^*)\bar{\sigma}^2\varepsilon_t^2, & \text{if } \varepsilon_t < 0. \end{cases} \quad (11)$$

Hence, the NIF combines two quadratic functions of ε_t in the threshold GARCH model and two exponential functions in the EGARCH model. When $\beta_1 \neq 0$ (or $\beta^* \neq 0$), the response of volatility is asymmetric in the EGARCH model (or the threshold GARCH model). In Wu and Xiao (2002) and Linton and Mammen (2005) the parametric dependence of σ_{t+1}^2 on ε_t is replaced by a kernel function which enters the conditional variance function additively.

In the SV models, the conditional variance is not explicitly specified but implied from the structure of the model specification. Unfortunately, even for the simplest SV model, the NIF does not have a closed form expression, nor has an additive structure in the measurable variables. Although h_{t+1} is not measurable in the SV models, one may be attempted to define the functional relation between h_{t+1} and y_t as the NIF, holding h_t and other variables constant, as in (7). While this definition of NIF is analytically attractive for the SV models, it has a number of drawbacks. First, in the basic SV model, $\rho = 0$ and hence (7) implies that h_{t+1} is not a function of y_t. This result seems to be at odd with the intuition that the return news must have some impact on future volatility. Second, $\exp(h_{t+1})$ is not a conditional variance in the SV model and hence the NIF defined on $\exp(h_{t+1})$ is not comparable with what has been used in the ARCH literature. Due to these two drawbacks, we decide to follow the ARCH literature and define NIF to be the function that relates the conditional variance to the return innovation lagged one period, holding constant of other variables.

When $\varphi < 1$, the model is strictly stationary and the conditional distribution is

$$\text{pdf}(y_t|y_1, \ldots, y_{t-1}) = \frac{\text{pdf}(y_1, \ldots, y_{t-1}, y_t)}{\text{pdf}(y_1, \ldots, y_{t-1})}$$
$$= \frac{\int \cdots \int \text{pdf}(y_1, \ldots, y_t, h_1, \ldots, h_t)dh_1 \cdots dh_t}{\int \cdots \int \text{pdf}(y_1, \ldots, y_{t-1}, h_1, \ldots, h_{t-1})dh_1 \cdots dh_{t-1}}.$$

From the conditional distribution, one can obtain the conditional mean and the conditional variance. Since the conditional mean is zero, the conditional variance is

$$\text{Var}(y_t|y_1, \ldots, y_{t-1})$$
$$= \frac{\int \cdots \int y_t^2 \text{pdf}(y_1, \ldots, y_t, h_1, \ldots, h_t)dh_1 \cdots dh_t dy_t}{\int \cdots \int \text{pdf}(y_1, \ldots, y_{t-1}, h_1, \ldots, h_{t-1})dh_1 \cdots dh_{t-1}}. \quad (12)$$

As an example, for the basic SV model (i.e. $\rho = 0$ in Eq. (6)), this becomes Eq. (13) (see Box I) where $\phi(x; \mu, \sigma^2)$ denotes the density function of $N(\mu, \sigma^2)$.

As in the ARCH literature, the NIF is defined as $\text{Var}(y_t|y_1 = 0, \ldots, y_{t-2} = 0, y_{t-1})$. It can be shown that NIF is continuous and symmetric in y_{t-1} for the basic SV model. Unfortunately, the integrals in (12) and (13) cannot be solved analytically. As a result, $\text{Var}(y_t|y_1, \ldots, y_{t-1})$ does not have an analytic form. In general, $\text{Var}(y_t|y_1, \ldots, y_{t-1})$ is not an additive function in $y_1, \ldots, y_{t-1}$. Also, $\text{Var}(y_t|y_1, \ldots, y_{p-1}) \neq \text{Var}(y_t|y_1, \ldots, y_{q-1})$ when $p \neq q$. Moreover, when $t \neq s$, $\text{Var}(y_t|y_1, \ldots, y_{t-1}) \neq \text{Var}(y_s|y_1, \ldots, y_{s-1})$. All these observations suggest that the news impact analysis in the SV model is much more complicated than that in the ARCH-type model. The SV models, including the basic SV model, are not nested by any existing ARCH-type models that assume an additive functional form.

When $m = 1$, our model is related to the asymmetric SV model recently proposed by Asai and McAleer (2006), where the specification of the volatility equation is given by

$$h_{t+1} = \varphi h_t + \gamma I(y_t < 0) + \rho\sigma_v\epsilon_t + \sigma_v\sqrt{1-\rho^2}w_t. \quad (14)$$

It can be shown that when $\gamma \neq 0$, the conditional variance is discontinuous in y_t at $y_t = 0$. This restriction is yet to be empirically justified.

The distinctive difference in the NIF between the SV models and the ARCH-type models is surprising. From Nelson (1990) it is known that the two classes of models have the same diffusion limit. Hence, the conditional distributions should have the same limit. However, as pointed out by Wang (2002), when discrete observations are available, the two classes of models are NOT asymptotically equivalent in terms of Le Cam's deficiency distance. This is because there is an important difference between the structure with respect to noise propagation in their conditional variances and hence conditional distributions.

3. Econometric analysis

3.1. Model estimation and comparison method

The SV models belong to the family of nonlinear non-Gaussian state space models. To do the maximum likelihood (ML) estimation and calculate the likelihood function of SV models, one has to deal with a high-dimensional integral since the latent process $\mathbf{h}$ needs to be integrated out from the joint density function, $\text{pdf}(\mathbf{y}, \mathbf{h})$, where $\mathbf{y} = (y_1, \ldots, y_T)$ and $\mathbf{h} = (h_1, \ldots, h_{T+1})$. Unfortunately, such an integral cannot be solved analytically in general. Several numerical methods have been proposed to approximate the integral via importance sampling techniques; see, for example, Shephard and Pitt (1997), Durbin and Koopman (1997) and Richard and Zhang (2007). However, the asymptotic properties of the ML estimate remain largely unknown, with the exception of consistency which was recently developed in Douc et al. (2011). For this reason, we decide to adopt Bayesian MCMC as the inferential framework because MCMC does not rely on asymptotic approximations to conduct inference.

Various MCMC methods have been developed to sample the parameters in the context of the SV models, including the single-move Metropolis–Hastings algorithm of Jacquier et al. (1994) and the multi-move algorithms of Kim et al. (1998) and Omori et al. (2007). Following Meyer and Yu (2000), we make use of a freely available Bayesian software, WinBUGS, to do the single-move Gibbs sampling. The Gibbs sampler, first proposed in Geman and Geman (1984), generates iterative samples from all the full conditional distributions. It can be justified by the Clifford–Hammersley theorem (Hammersley and Clifford, 1970). WinBUGS provides an easy and efficient implementation of the Gibbs sampler and has been widely used to estimate latent variables model. Both the simulation studies and the out-of-the-sample forecasting exercise carried out in present paper are implemented using R2WinBUGS (Sturtz et al., 2005).

To fix the idea of the MCMC, let $p(\boldsymbol{\theta})$ be the prior distribution of the unknown parameter $\boldsymbol{\theta}$. Bayesian methods overcome the

$$\frac{\int\cdots\int y_t^2 \prod_{i=1}^{t}\phi(y_i;0,e^{h_i})\prod_{i=2}^{t}\phi(h_i;\varphi h_{i-1},\gamma^2)\phi\left(h_1;0,\frac{\gamma^2}{1-\phi^2}\right)dh_1\cdots dh_t dy_t}{\int\cdots\int \prod_{i=1}^{t-1}\phi(y_i;0,e^{h_i})\prod_{i=2}^{t-1}\phi(h_i;\varphi h_{i-1},\gamma^2)\phi\left(h_1;0,\frac{\gamma^2}{1-\phi^2}\right)dh_1\cdots dh_{t-1}} \quad (13)$$

Box I.

difficulty in ML by the data-augmentation strategy (Tanner and Wong, 1987), namely, the parameter space is augmented from $\boldsymbol{\theta}$ to $(\boldsymbol{\theta}, \mathbf{h})$. By successive conditioning, the joint prior density is

$$p(\boldsymbol{\theta}, \mathbf{h}) = p(\boldsymbol{\theta})p(h_0)\prod_{t=1}^{T+1} p(h_t|h_{t-1}, \boldsymbol{\theta}). \quad (15)$$

The likelihood function is

$$p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{h}) = \prod_{t=1}^{T} p(y_t|h_t, \boldsymbol{\theta}). \quad (16)$$

Obviously, both the joint prior density and the likelihood function are available analytically, provided that the analytical expressions for the prior distributions of $\boldsymbol{\theta}$ are supplied. By Bayes' theorem, the joint posterior distribution of the unobservables given the data is given by:

$$p(\boldsymbol{\theta}, \mathbf{h}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(h_0)\prod_{t=1}^{T+1} p(h_t|h_{t-1}, y_t, \boldsymbol{\theta})\prod_{t=1}^{T} p(y_t|h_t, \boldsymbol{\theta}). \quad (17)$$

The Gibbs sampler, used to generate a Markov chain (i.e. correlated samples) whose stationary distribution is the joint posterior distribution (17), works as follows in the first step. Given the initialization $(\boldsymbol{\theta}^{(0)}, \mathbf{h}^{(0)})$, we draw from each of the following distributions:

$$\begin{aligned}
&\theta_1^{(1)} \sim p(\theta_1|\theta_2^{(0)}, \ldots, \theta_K^{(0)}, \mathbf{h}^{(0)}, \mathbf{y});\\
&\vdots\\
&\theta_K^{(1)} \sim p(\theta_K|\theta_1^{(1)}, \ldots, \theta_{K-1}^{(1)}, \mathbf{h}^{(0)}, \mathbf{y});\\
&h_1^{(1)} \sim p(h_1|\boldsymbol{\theta}^{(1)}, h_2^{(0)}, \ldots, h_{T+1}^{(0)}, \mathbf{y});\\
&\vdots\\
&h_{T+1}^{(1)} \sim p(h_{T+1}|\boldsymbol{\theta}^{(1)}, h_1^{(1)}, \ldots, h_T^{(1)}, \mathbf{y}).
\end{aligned}$$

For a Markov chain to have a unique stationary distribution, it has to be irreducible and aperiodic. If, in addition, the chain is positive Harris-recurrent, it is ergodic and a central limit theorem is applicable for sample-path averages. Roberts and Smith (1994) established general conditions under which the Markov chain generated from the Gibbs sampler converges to the target posterior distribution when more and more iterations are obtained.

By the ergodic law of large numbers, posterior moments of $\boldsymbol{\theta}$ and $\mathbf{h}$ and posterior marginal densities may be estimated by averaging the corresponding functions over the sample after convergence is achieved. For example, let $\{\boldsymbol{\theta}^{(j)}, \mathbf{h}^{(j)}, j = 1, \ldots, J\}$ be a simulated sample form the joint posterior distribution $p(\boldsymbol{\theta}, \mathbf{h}|\mathbf{y})$. The posterior mean of $\boldsymbol{\theta}$ can be estimated by

$$\hat{\boldsymbol{\theta}} = \frac{1}{J}\sum_{j=1}^{J}\boldsymbol{\theta}^{(j)}. \quad (18)$$

Since any measurable function of a stationary and ergodic sequence is stationary and ergodic, the ergodic law of large numbers is applicable to $f(\boldsymbol{\theta}, \mathbf{h})$ where f is any measurable function. For example, the posterior variance–covariance of $\boldsymbol{\theta}$ may be estimated by

$$\widehat{\mathrm{Var}}(\boldsymbol{\theta}|\mathbf{Y}) = \frac{1}{J}\sum_{j=1}^{J}(\boldsymbol{\theta}^{(j)} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta}^{(j)} - \hat{\boldsymbol{\theta}})'. \quad (19)$$

Similarly, one can obtain the estimate of any posterior credible interval.

For any simulation-based estimation method, it is important to assess the numerical quality of a point estimate and in this paper we calculate and report the Monte Carlo standard error (MCSE). Let the parameter of interest be θ_1. Applying the ergodic central limit theory to a converged Markov chain of θ_1, we have

$$\sqrt{J}(\widehat{\theta}_1 - E(\theta_1|\mathbf{y})) \xrightarrow{d} N\left(0, \mathrm{Var}(\theta_1^{(1)}) + 2\sum_{j=2}^{\infty}\mathrm{Cov}(\theta_1^{(1)}, \theta_1^{(j)})\right). \quad (20)$$

The MCSE, defined by $\sqrt{\mathrm{Var}(\theta_1^{(1)}) + 2\sum_{j=2}^{\infty}\mathrm{Cov}(\theta_1^{(1)}, \theta_1^{(j)})/J}$, tells how reliable the estimate of the posterior mean of θ_1 is. It can be estimated by the method of batch means or by the spectral method (Geweke, 1992). It is obvious that conceptually MCSE is very different from the posterior standard error.

Several alternative methods are available to estimate the posterior marginal density. One way is to use the kernel method. The second way is the so-called Rao–Blackwellization. When the simulation size is large, the marginal density can be regarded as exact, enabling exact finite sample inferences.

It is crucial to check if convergence has been achieved because all the estimates are supposed to be obtained from a converged sample coming from the stationary distribution. In this paper we make use of Heidelberger and Welch test (Heidelberger and Welch, 1983) to check convergence and all the results we report in the present paper are based on samples which have passed the Heidelberger and Welch convergence test.

To compare two models, say M_0 and M_1, let $\pi(M_k)$ be the prior model probability density, $p(\mathbf{y}|M_k)$ the marginal likelihood of model k, and $p(M_k|\mathbf{y})$ the posterior probability density, where $k = 0, 1$. Comparison of two models in the Bayesian framework is amount to calculate the posterior odds ratio. By Bayes' theorem, we have

$$\frac{p(M_0|\mathbf{y})}{p(M_1|\mathbf{y})} = \frac{p(\mathbf{y}|M_0)}{p(\mathbf{y}|M_1)} \times \frac{\pi(M_0)}{\pi(M_1)}, \quad (21)$$

that is,

$$\ln(\text{Posterior Odds Ratio}) = \ln(\text{BF}) + \ln(\text{Prior Odds Ratio}), \quad (22)$$

where the BF is referred to as the Bayes factor which is defined as the ratio of the marginal likelihood values of the two competing models. If the prior odds ratio is set to 1, as is done in much of the Bayesian literature and also in the present paper, the posterior odds ratio takes the same value as the BF. BFs are the leading method of Bayesian model comparison. They are the Bayesian analogues of the classical likelihood ratio tests. Jeffreys (1961) gave a scale for interpretation of BFs. If ln(BF) is less (bigger) than 0, there is evidence for (against) M_0. Moreover, if $\ln(\text{BF}) \in (0, 1)$, the evidence against M_0 is barely worth mentioning; if $\ln(\text{BF}) \in (1, 3)$, the evidence against M_0 is positive; if ln(BF) $\in$ (1, 3) (or (3, ∞)), the evidence against M_0 is strong (or very strong). However, model comparison via BFs is not the same as the classical likelihood ratio tests because BFs automatically includes a penalty for including too much model structure and hence guards against overfitting.

Calculating BFs requires the marginal likelihood $p(\mathbf{y}|M_k)$ to be evaluated. By definition, $p(\mathbf{y}|M_k)$ is the integral of the likelihood function with respect to the prior density

$$p(\mathbf{y}|M_k) = \int p(\mathbf{y}|\boldsymbol{\theta}, M_k)p(\boldsymbol{\theta}, M_k)d\boldsymbol{\theta}. \quad (23)$$

To calculate the marginal likelihood, we follow Chib (1995) by rearranging Bayes' theorem

$$p(\mathbf{y}|M_k) = \frac{p(\mathbf{y}|\boldsymbol{\theta}, M_k)p(\boldsymbol{\theta}|M_k)}{p(\boldsymbol{\theta}|\mathbf{y}, M_k)}.$$

Thus, the log-marginal likelihood can be calculated by

$$\ln(p(\mathbf{y}|M_k)) = \ln(p(\mathbf{y}|\boldsymbol{\theta}, M_k)) + \ln(p(\boldsymbol{\theta}|M_k)) - \ln(p(\boldsymbol{\theta}|\mathbf{y}, M_k)), \quad (24)$$

where $\boldsymbol{\theta}$ is an appropriately selected high density point which is chosen to be the posterior mean ($\bar{\boldsymbol{\theta}}$) in the present paper. The first term on the right hand sight of (24) is the log-likelihood and the second term is the log prior density, both evaluated at $\bar{\boldsymbol{\theta}}$. The third quantity involves the posterior density which is only known up to a normality constant. Following Kim et al. (1998), we use a multivariate kernel density estimate to approximate it.

Regarding the prior distributions of $\boldsymbol{\theta}$, we follow the literature, namely, all components of $\boldsymbol{\theta}$ are assumed to be independent; $\gamma^2 \sim$ Inverse-Gamma(2.5, 0.025); $\varphi^* \sim$ Beta(20, 1.5), where $\varphi^* = (\varphi + 1)/2$; $\mu, \mu_y \sim N(0, 25)$, where $\mu = \exp(\sigma/2)$; $\rho, \rho_i \sim$ Uniform(−1, 1), for all i.

The log-likelihood function $p(\mathbf{y}|\boldsymbol{\theta}, M_k)$ has no analytical form for the SV models as it is marginalized over the latent states $\mathbf{h}$. However, it is possible to approximate it with simulation techniques. In the paper, we follow Shephard and Pitt (1997) and Skaug and Yu (2008) by using the importance sampling method based on the Laplace-approximation.

Once the model is estimated by MCMC, the one-period-ahead volatility can easily be calculated as a by-product because the smoothed estimates of h_{T+1} and $\exp(h_{T+1})$ are obtainable as the posterior means of the Markov chains of h_{T+1} and $\exp(h_{T+1})$. To obtain the K-period-ahead volatility, we only need to redefine $\mathbf{h} = (h_1, \ldots, h_{T+K})$.

3.2. Sampling performance

To check the reliability of the proposed estimation method, we simulate data from three SV models, the basic SV, the leverage SV and the *Spline1 SV*. In the simulations, the values of φ, σ and γ are always set to 0.9, 1, 0.135, respectively. We fix μ_y to zero and assume it is known. In the leverage SV model, we set $\rho = -0.3$. In the *Spline1 SV*, two sets of parameter values are selected for (ρ_1, ρ_2) : $(-0.5, -0.5), (-0.5, 0)$. As it will be clear from the empirical studies reported below, these parameter values are practically realistic.

For each parameter setting, 1000 observations are simulated from the true model. We then replicate the experiment for 500 times to obtain the mean and the standard error for each parameter estimate across 500 replications. For the basic SV and the leverage SV, the number of the total iterations in MCMC is 30,000 with the first 10,000 iterations used as the burn-in. For the *Spline1 SV* and *Spline2 SV*, the number of the total iterations is 150,000 with the first 50,000 iterations used as the burn-in. Table 1 summarizes the results. The most important finding from this table is that MCMC is reliable for all the parameters in all cases, reinforcing what has been found in the literature; see Jacquier et al. (1994) and Yu (2005). Moreover, it seems more difficult to estimate σ, ρ_1 and ρ_2 when ρ_1 and ρ_2 are further away from each other.

4. Empirical results

4.1. Estimation results from daily data

In this subsection we fit the basic SV, the classical leverage SV, the *Spline1 SV* and the *Spline2 SV* to two continuously compounded daily return series in the US, namely the S&P500 from January 2, 1985 to December 31, 1989 and the Microsoft (MSFT) from January 2, 1987 to December 31, 1991. The number of observations is 1263 and 1264 for the two series. In both series, the data is subtracted by the sample mean and hence μ_y is not estimated.

Estimation results, including the parameter estimates, the posterior standard errors (in parenthesis), the MCSE (in bracket) and the minus log-marginal-likelihood values are reported in Table 2. Several conclusions can be drawn. First, φ is highly significant in all cases. Second, in the leverage SV, ρ is estimated to be negative for both series, featured by −0.3691 and −0.1343. However, ρ is significant only for the S&P500. That is why the leverage SV provides a significant improvement over the basic SV in terms of the marginal likelihood only for the S&P500. For the MSFT, ρ is insignificant and ln(BF) of the basic SV against the leverage SV is less than 0. Although the fact that the estimated ρ is much larger in indices than in stocks is in odd with the leverage hypothesis, this finding is consistent with those documented in the literature; see for example, Tauchen et al. (1996) and ABDE (2001). The result indicates that one would conclude the absence of the leverage effect if only the leverage SV is fitted to the MSFT. As it will be clear below, this conclusion is misleading.

Third, the *Spline1 SV* provides a significant improvement over the leverage SV in both cases. In particular, the log-marginal-likelihood improves by 19.73 and 65.47 from the leverage SV to the *Spline1 SV*, indicating that the single threshold model is inadequate to explain all the asymmetry in volatility response. More interesting results emerge if one examines the estimates of ρ_1 and ρ_2. In the *Spline1 SV* model, ρ_1 is more negative than ρ in the leverage SV and statistically significant for both series. On the other hand, ρ_2 is estimated to be positive. While not reported, the inference based on $\rho_1 - \rho_2$ suggests that $\rho_1 \neq \rho_2$ in the estimated *Spline1 SV* model for both series. It is rather surprising to find positive estimates for ρ_2. We will examine the out-of-the-sample performance of *Spline1 SV* in Section 4.3. While we fail to find a significant leverage effect in the leverage SV for the MSFT, we do find the strong evidence of leverage effect in *Spline1 SV*. Fourth, in *Spline2 SV*, as expected, ρ_2 is close to 0 with a large standard error for the S&P500. The estimates for the other two ρs are similar to those in *Spline1 SV*. Not surprisingly, the marginal likelihood value decreases by adding one more knot to *Spline1 SV*. The results in the estimated *Spline2 SV* are slightly different for the MSFT. The estimate of ρ_2 is −0.963. While this is significantly different from zero but not significantly different from the estimate of ρ_1 in the same model (−0.7541). The marginal likelihood values of *Spline1 SV* and *Spline2 SV* indicate there is little evidence to support *Spline2 SV* over *Spline1 SV* for both series.

4.2. Estimation results from weekly data

In this subsection we first fit the SV models to the weekly return series of MSFT from April 4, 1986 to December 24, 2007. Only the returns for the individual stock are selected due to the empirical results reported earlier. The number of observations is 1133 in the

Table 1
Finite sample properties of MCMC for three SV models based on 500 simulated sample paths of 1000 observations in each path. For the basic SV and the leverage SV, the number of total iterations in MCMC is 30,000 with the first 10,000 iterations used as the burn-in. For the *Spline1 SV*, the number of total iterations in MCMC is 150,000 with the first 50,000 iterations used as the burn-in.

	σ	φ	γ	ρ	σ	φ	γ	ρ_1	ρ_2
	Basic SV				Spline1 SV				
True	1	0.9	0.135		1	0.9	0.135	−0.5	−0.5
Mean	1.0031	0.8806	0.1208		1.0115	0.8733	0.1251	−0.4086	−0.5276
Std	0.0318	0.0466	0.0225		0.1104	0.0436	0.0285	0.2716	0.2192
	Leverage SV				Spline1 SV				
True	1	0.9	0.135	−0.3	1	0.9	0.135	−0.5	0
Mean	1.0081	0.8649	0.1222	−0.2895	0.9486	0.8584	0.1130	−0.4116	−0.0737
Std	0.0353	0.0639	0.0224	0.2107	0.1326	0.0483	0.0139	0.2060	0.2352

Table 2
Estimation results from daily data. The number in parenthesis is the posterior standard error. The number in bracket is the Monte Carlo standard error.

Data	Model	-Log MargLik	σ	φ	γ	ρ_1	ρ_2	ρ_3
S&P500	Basic	1730.26	0.9302 (0.067) [0.001]	0.9426 (0.019) [0.0008]	0.2621 (0.042) [0.0021]			
	Leverage	1720.13	0.9336 (0.055) [0.0014]	0.9212 (0.024) [0.0012]	0.3066 (0.049) [0.0030]	−0.3691 (0.087) [0.0041]		
	Spline1	1700.40	2.077 (0.3961) [0.015]	0.9135 (0.019) [0.0008]	0.3689 (0.058) [0.0026]	−0.8386 (0.090) [0.0133]	0.1435 (0.137) [0.005]	
	Spline2	1704.38	1.874 (0.3537) [0.014]	0.9157 (0.019) [0.0007]	0.3458 (0.051) [0.0022]	−0.8446 (0.1079) [0.0046]	0.2059 (0.3484) [0.011]	0.1429 (0.1672) [0.0066]
MSFT	Basic	3017.76	2.685 (0.274) [0.0088]	0.9444 (0.029) [0.0016]	0.2422 (0.063) [0.0037]			
	Leverage	3020.49	2.669 (0.2892) [0.0131]	0.9397 (0.030) [0.0018]	0.2603 (0.063) [0.004]	−0.1343 (0.097) [0.0045]		
	Spline1	2955.02	24.42 (6.57) [0.2795]	0.7778 (0.0317) [0.0014]	1.509 (0.1921) [0.009]	−0.9724 (0.0243) [0.001]	0.9003 (0.046) [0.0021]	
	Spline2	2967.30	8.304 (2.38) [0.099]	0.7136 (0.051) [0.0022]	1.627 (0.1407) [0.006]	−0.7541 (0.069) [0.0026]	−0.963 (0.024) [0.0009]	0.6941 (0.064) [0.0021]

return series. In this empirical exercise, μ_y is estimated but not reported to save space.

Estimation results are reported in Table 3. Several conclusions can be drawn. First, in the leverage SV, ρ is estimated to be −0.075. As in the daily MSFT, ρ is statistically insignificant. Once again the results are reinforced by a small difference in the marginal likelihood values of the basic SV and the leverage SV models. Hence, one would conclude the absence of the leverage effect if the leverage SV is fitted, consistent with the usual claim for individual stocks.

Second, the *Spline1 SV* provides a significant improvement over the leverage SV model with $\ln(\mathrm{BF}) \gg 0$. The estimated ρ_1 is negative (−0.2968) and the estimated ρ_2 is positive (0.2678). The 10% credible interval of ρ_1 excludes 0. This signs for estimated ρs corroborate well with those in the daily data. Third, in the estimated *Spline2 SV*, ρ_1 is close to ρ_2. They are both close to the estimate of ρ_1 in the *Spline1 SV*. Also, ρ_3 in the estimated *Spline2 SV* is close to ρ_2 in the estimated *Spline1 SV*. Not surprisingly, adding one more knot to the *Spline1 SV* decreases the marginal likelihood value. Hence, there is no evidence to support the *Spline2 SV* in the weekly data.

To check the robustness of the empirical results, we fit the classical leverage SV and the *Spline1 SV* to three weekly return series of Johnson and Johnson (JnJ), 3M, and Kellogg, all from April 4, 1986 to December 24, 2007. The number of observations is 1133 in all cases. To save space, we choose not to report results on the basic SV and the *Spline2 SV* because both models are found to be outperformed by the *Spline1 SV*.

Estimation results are reported in Table 4. Similar conclusions can be drawn from Table 4 as from Table 3. For example, in all cases, ρ is statistically insignificant in the leverage SV. The *Spline1 SV* provides a significant improvement over the leverage SV in all cases. The ln(BF)s are 13.14, 8.06 and 5.24, suggesting very strong evidence in favor of *Spline1 SV*. The estimated ρ_1 is always very negative (−0.5161, −0.4408, and −0.5409) and significantly less than 0. On the other hand, the estimated ρ_2 is always insignificantly different from 0. Therefore, the leverage effect is found to be significant in one regime albeit not globally.

Table 3
Estimation results from weekly data of Microsoft. The number in parenthesis is the posterior standard error. The number in bracket is the Monte Carlo standard error.

Data	Model	-LogMargLik	σ	φ	γ	ρ_1	ρ_2	ρ_3
MSFT	Basic	3297.75	7.353 (1.885) [0.097]	0.9936 (0.0062) [0.0003]	0.1133 (0.0265) [0.0016]			
	Leverage	3299.11	7.337 (1.694) [0.0959]	0.9938 (0.0049) [0.0002]	0.1202 (0.0207) [0.0013]	−0.075 (0.1425) [0.0080]		
	Spline1	3289.99	10.86 (2.468) [0.1034]	0.9855 (0.0055) [0.0002]	0.1245 (0.0236) [0.0011]	−0.2968 (0.1565) [0.0058]	0.2678 (0.1766) [0.0071]	
	Spline2	3291.02	10.63 (2.402) [0.0099]	0.9864 (0.0055) [0.0002]	0.124 (0.027) [0.0013]	−0.2529 (0.1782) [0.0071]	−0.3358 (0.4106) [0.0139]	0.320 (0.1801) [0.007]

Table 4
Estimation results from weekly data of JnJ, 3M and Kellogg. The number in parenthesis is the posterior standard error. The number in bracket is the Monte Carlo standard error.

Data	Model	-LogMargLik	σ	φ	γ	ρ_1	ρ_2
JnJ	Leverage	2833.57	3.564 (0.6939) [0.0365]	0.9728 (0.0205) [0.0012]	0.163 (0.0424) [0.0027]	−0.1545 (0.1214) [0.006]	
	Spline1	2820.43	0.6.375 (1.38) [0.049]	0.964 (0.0119) [0.0005]	0.1912 (0.0381) [0.0018]	−0.5161 (0.1536) [0.0056]	0.2319 (0.1605) [0.0057]
3M	Leverage	2778.22	2.767 (0.1803) [0.0049]	0.9423 (0.027) [0.0016]	0.2122 (0.0548) [0.0035]	−0.1681 (0.1294) [0.0069]	
	Spline1	2770.16	4.212 (1.044) [0.042]	0.9279 (0.0267) [0.0011]	0.2657 (0.067) [0.0032]	−0.4408 (0.1778) [0.0072]	0.1206 (0.1983) [0.0083]
Kellogg	Leverage	2795.49	3.585 (0.7277) [0.035]	0.9881 (0.0097) [0.0004]	0.1186 (0.0336) [0.0018]	−0.1915 (0.131) [0.0073]	
	Spline1	2790.25	5.758 (1.216) [0.051]	0.9839 (0.0077) [0.0003]	0.1192 (0.03) [0.0014]	−0.5409 (0.1746) [0.0075]	−0.0115 (0.1975) [0.0083]

As in the weekly MSFT, the leverage effect cannot be identified in the leverage SV model but becomes prominent in *Spline1 SV* for all three series.

4.3. Forecasting results from weekly data

Superior in-the-sample performance does not necessarily lead to any gain out-of-the-sample. In this section, we compare the out-of-the-sample performance of the proposed model against the classical models for forecasting volatility using the four weekly return series from the last subsection, Microsoft, JnJ, 3M and Kellogg from April 7, 1986 to December 24, 2007. Three competing models, namely, the basic SV, the leverage SV and the *Spline1 SV*, are fitted to the return data and used to obtain one-period-ahead out-of-sample forecasts of weekly volatility.

We measure weekly volatility using the so-called realized volatility (RV) obtained from daily data. Let RV_t denote the weekly RV and $p(t,k)$ denote the daily log-price. Then RV_t is defined by $\sqrt{\sum_{k=1}^{N_t}(p(t,k)-p(t,k-1))^2}$ where N_t is the number of trading days in week t and $p(t,0) = p(t-1,N_{t-1})$. The theoretical justification of RV as a measure of volatility can be found in Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002).

We split the weekly sample into an 'in-sample' estimation period and an 'out-of-sample' forecast evaluation period. For estimation we use the rolling window scheme, where the size of the sample, which is used to estimate the competing models, is fixed at 990. Therefore, we first estimate all the competing models with weekly returns over the period from April 7, 1986 to April 1, 2005. The first forecast is made for the week beginning April 4, 2005. When a new observation is added to the sample, we delete the first observation and re-estimate all the models. The re-estimated models are then used to forecast volatility. This process is repeated until we reach the end of the sample, December 24, 2007. Therefore, the final forecast is for the week that begins December 31, 2007. In total, we need to make 144 forecasts from each model. We match each forecasted volatility with the corresponding realized volatility. Following the suggestion of a referee, we also forecast the weekly volatility with three GARCH models, namely, GARCH(1, 1), GJR-GARCH(1, 1) and EGARCH(1, 1). All three GARCH models are estimated by ML.

In Table 5, we report the Mean Absolute Error (MAE) to evaluate forecast accuracy. Most importantly, in all four cases, *Spline1 SV* performs the best, not only better than the two classical SV models but also better than the three GARCH models. In particular, the percentage improvement of *Spline1 SV* over the leverage SV ranges between 1% and 4%. Interestingly, GARCH(1, 1) always performs better than the two GARCH rivals except in one case. Within the SV family, in 3 out of 4 cases (MSFT, JnJ and 3M), the leverage SV performs worse than the basic SV. This is not surprising to us since

Table 5
Forecasting results from weekly data.

	MAE × 1000					
	Basic SV	Leverage SV	Spline1 SV	GARCH	GJR-GARCH	EGARCH
MSFT	9.866	9.906	9.671	11.392	10.714	10.860
JnJ	10.681	10.723	10.377	10.406	11.472	10.609
3M	10.984	11.112	10.969	10.988	11.422	11.682
Kellogg	6.901	6.889	6.725	6.739	6.816	7.381

ρ is insignificant in the leverage SV in these three cases. In all cases, the improvement from the *Spline1 SV* over the leverage SV is more remarkable than the difference between the leverage SV and the basic SV. The empirical result suggests that the new model not only provides a better in-the-sample fit to the data, but also gains on predicting volatility out-of-the-sample.

5. Conclusion

Using the linear spline we introduce a semiparametric SV model with time varying leverage effects. The driving factor for time varying leverage is the size and the sign of the lagged return. The model nests the basic SV and the leverage SV models. Statistical properties of the proposed model are discussed. The model is fitted to daily and weekly US index and stock returns and found to have the superior in-the-sample performance. Although one could not find a significant leverage effect in the classical leverage SV for the daily and weekly stock returns, strong evidence of leverage effect was found in the new model when the leverage effect is allowed to be time varying. Not only does the new model perform better in-the-sample, but also it yields more accurate forecasts of volatility than the classical models.

This paper focuses on models in univariate. While it is perhaps desirable to consider multivariate extensions of the proposed model, this task is beyond the scope of the current paper. It is worth pointing out there are several recent studies where alternative asymmetric multivariate SV models have been introduced (Asai and McAleer, 2009).

Appendix

To prove the theorem, we first give a lemma.

Lemma 1. *Suppose* $X, Y \sim$ i.i.d. $N(0,1)$ *and* X *and* Y *are independent. Define*

$$Z = \begin{cases} \rho_1 X + \sqrt{1-\rho_1^2}\,Y, & \text{if } \infty \geq X > \tau_1 \\ \vdots & \\ \rho_{m+1} X + \sqrt{1-\rho_{m+1}^2}\,Y, & \text{if } \tau_m \geq X > -\infty. \end{cases}$$

Then the moment generate function (mgf) of Z *is*

$$m_Z(s) = \exp(s^2/2)\Bigg[\Phi(\tau_m - s\rho_{m+1}) + 1 - \Phi(\tau_1 - s\rho_1) + \sum_{i=2}^{m}(\Phi(\tau_{i-1} - s\rho_i) - \Phi(\tau_i - s\rho_i))\Bigg].$$

Proof of Lemma 1. Let $\phi(\cdot)$ and $\Phi(\cdot)$ be the pdf and the cdf of $N(0,1)$. The mgf of Z is

$$\begin{aligned} m_Z(s) &= \mathrm{E}\{\exp(sz)\} = \iint \exp(sz) pdf_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty} pdf_Y(y)\Bigg\{\exp(sy\sqrt{1-\rho_{m+1}^2}) \int_{-\infty}^{\tau_m} \exp(s\rho_{m+1}x)\,pdf_X(x)\mathrm{d}x + \cdots + \exp(sy\sqrt{1-\rho_1^2}) \\ &\quad \times \int_{\tau_m}^{+\infty} \exp(s\rho_1 x) pdf_X(x)\mathrm{d}x\Bigg\} dy. \end{aligned}$$

Since

$$\int_{-\infty}^{\tau_m} \exp(s\rho_{m+1}x)pdf_X(x)\mathrm{d}x = \exp(s^2\rho_{m+1}/2)\Phi(\tau_m - s\rho_{m+1}),$$

$$\int_{\tau_i}^{\tau_{i-1}} \exp(s\rho_i x)pdf_X(x)\mathrm{d}x = \exp(s^2\rho_i/2)[\Phi(\tau_{i-1} - s\rho_i) - \Phi(\tau_i - s\rho_i)],$$

$$\int_{\tau_1}^{\infty} \exp(s\rho_1 x)pdf_X(x)\mathrm{d}x = \exp(s^2\rho_1/2)(1 - \Phi(\tau_1 - s\rho_1)),$$

we get the mgf of Z. If $\rho_i = \rho, \forall i$, this mgf becomes $e^{s^2/2}$ which is the mgf of $N(0,1)$. □

Proof of Theorem 2.1. Since $|\varphi| < 1$, we may rewrite Eq. (5) as $h_{t+1} = \gamma \sum_{j=0}^{\infty} \varphi^j v_{t-j}$. Thus, the mgf of h_{t+1} is

$$\begin{aligned} \mathrm{E}(\exp(sh_{t+1})) &= \prod_{j=0}^{\infty}\Bigg\{\exp\left(\frac{1}{2}s^2\gamma^2\varphi^{2j}\right)[\Phi(\tau_m - s\gamma\varphi^j\rho_{m+1}) + 1 \\ &\quad - \Phi(\tau_1 - s\gamma\varphi^j\rho_1) + \sum_{i=2}^{m}(\Phi(\tau_{i-1} - s\gamma\varphi^j\rho_i) \\ &\quad - \Phi(\tau_i - s\gamma\varphi^j\rho_i))]\Bigg\} \\ &:= G(i, \boldsymbol{\rho}, \gamma, \varphi). \end{aligned} \tag{A.1}$$

The existence of mgf for v_t implies that the variance of v_t is finite. Since h_{t+1} is a linear process with finite innovation variance, the stationarity and ergodicity are ensured if and only $|\varphi| < 1$.

The moments of h_t can be obtained by differentiating the log mgf. To obtain the moments of y_t, note that for $i = 1, 2, \ldots,$

$$\begin{aligned} \mathrm{E}(y_t^{2i-1}|h_t) &= \mathrm{E}\left(\sigma^{2i-1}\exp\left(\frac{2i-1}{2}h_t\right)\epsilon_t^{2i-1}|h_t\right) \\ &= \sigma^{2i-1}\exp\left(\frac{2i-1}{2}h_t\right)\mathrm{E}(\epsilon_t^{2i-1}|h_t) = 0, \end{aligned}$$

$$\mathrm{E}(y_t^{2i}|h_t) = \mathrm{E}(\sigma^{2i}\exp(ih_t)\epsilon_t^{2i}|h_t) = \sigma^{2i}\exp(ih_t)\frac{(2i)!}{2^i i!}.$$

Hence,

$$\begin{aligned} \mathrm{E}(y_t^{2i-1}) &= \mathrm{E}(\mathrm{E}(y_t^{2i-1}|h_t)) = 0, \\ \mathrm{E}(y_t^{2i}) &= \mathrm{E}(\mathrm{E}(y_t^{2i}|h_t)) \\ &= \sigma^{2i}\mathrm{E}(\exp(ih_t))\frac{(2i)!}{2^i i!} = \sigma^{2i}\frac{(2i)!}{2^i i!}G(i, \boldsymbol{\rho}, \gamma, \varphi). \end{aligned}$$

When $m = 1$, $\boldsymbol{\rho} = (\rho_1, \rho_2)$ and it is easy to verify that

$$G(i, \boldsymbol{\rho}, \gamma, \varphi) = \prod_{j=0}^{\infty} \left\{ \exp\left(\frac{1}{2} s^2 \gamma^2 \varphi^{2j}\right) \times [\Phi(s\gamma\varphi^j \rho_1) + \Phi(-s\gamma\varphi^j \rho_2)] \right\}. \quad \square$$

References

Andersen, T., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. The distribution of stock return volatility measurement. Journal of Financial Economics 61, 43–76.

Andersen, T., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42–55.

Asai, M., McAleer, M., 2006. Asymmetric multivariate stochastic volatility. Econometric Reviews 25, 453–473.

Asai, M., McAleer, M., 2009. The structure of dynamic correlations in multivariate stochastic volatility models. Journal of Econometrics 25, 182–192.

Bandi, F., Reno, R., 2010. Time-varying leverage effects, Journal of Econometrics (forthcoming).

Barndorff-Nielsen, O., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, Series B 64, 253–280.

Bekaert, G., Wu, G., 2000. Asymmetric volatility and risk in equity markets. Review of Financial Studies 13, 1–42.

Black, F., 1976. Studies of stock market volatility changes. Proceedings of the American Statistical Association, Business and Economic Statistics Section 177–181.

Braun, P.A., Nelson, D., Sunier, A., 1995. Good news, bad news, volatility and betas. Journal of Finance 50, 1575–1603.

Chib, S., 1995. Marginal likelihood from the Gibbs output. The Journal of the American Statistical Association 90, 1313–1321.

Christie, A.A., 1982. The stochastic behavior of common stock variances. Journal of Financial Economics 10, 407–432.

Daouk, H., Ng, D., 2007. Is unlevered firm volatility asymmetric? Working Paper. Cornell University.

Douc, R., Moulines, E., Olsson, J., van Handel, E., 2011. Consistency of the maximum likelihood estimator for general hidden Markov models. Annals of Statistics 39, 474–513.

Duan, J., 1995. The GARCH option pricing model. Mathematical Finance 5, 13–32.

Durbin, J., Koopman, S.J., 1997. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika 84, 669–684.

Engle, R., Ng, V., 1993. Measuring and testing the impact of news in volatility. Journal of Finance 43, 1749–1778.

Figlewski, S., Wang, X., 2000. Is the 'leverage effect' a leverage effect? Working Paper. New York University.

Geman, D., Geman, S., 1984. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. IEEE Transaction on Pattern Analysis and Machine Intelligence 12, 609–628.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 4. Oxford University Press, Oxford, pp. 169–193.

Glosten, L.R., Jagannanthan, R., Runkle, D., 1993. Relationship between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1802.

Hammersley, J., Clifford, P., 1970. Markov fields on finite graphs and lattices. Unpublished Manuscript.

Hansen, P., Huang, Z., Shek, H., 2010. Realized GARCH: a complete model of returns and realized measures of volatility. Working Paper. Stanford University.

Harvey, A.C., Shephard, N., 1996. The estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business and Economic Statistics 14, 429–434.

Heidelberger, P., Welch, P., 1983. Simulation run length control in the presence of an initial transient. Operations Research 31, 1109–1144.

Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–389.

Jeffreys, H., 1961. The Theory of Probability, Oxford.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.

Linton, O., Mammen, E., 2005. Estimating semiparametric ARCH(∞) models by kernel smoothing methods. Econometrica 73, 771–836.

Meyer, R., Yu, J., 2000. BUGS for a Bayesian analysis of stochastic volatility models. Econometrics Journal 3, 198–215.

Mishra, S., Su, L., Ullah, A., 2010. Semiparametric estimator of time series conditional variance. Journal of Business and Economic Statistics 28, 256–274.

Nelson, D., 1990. ARCH models as diffusion approximations. Journal of Econometrics 45, 7–38.

Nelson, D., 1991. Conditional heteroskedasticity in asset pricing: a new approach. Econometrica 59, 347–370.

Omori, Y., Chib, S., Shephard, N., Nakajima, J., 2007. Stochastic volatility with leverage: fast likelihood inference. Journal of Econometrics 140, 425–449.

Pagan, A.R., Schwert, G.W., 1990. Alternative models for conditional stock volatility. Journal of Econometrics 45, 267–290.

Richard, J.F., Zhang, W., 2007. Efficient high-dimensional importance sampling. Journal of Econometrics 141, 1385–1411.

Roberts, G.O., Smith, A.F.M., 1994. Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms. Stochastic Processes and their Applications 49, 207–216.

Robinson, P.M., 1991. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. Journal of Econometrics 47, 67–84.

Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric Regression. Cambridge University Press.

Shephard, N., Pitt, M., 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika 84, 653–667.

Skaug, H., Yu, J., 2008. Automated likelihood based inference for stochastic volatility models. Working Paper. Singapore Management University.

Sturtz, S., Ligges, U., Gelman, A., 2005. R2WinBUGS: a package for running WinBUGS from R. Journal of Statistical Software 12, 1–16.

Tanner, T.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–549.

Tauchen, G., Zhang, H., Liu, M., 1996. Volume, volatility and leverage: a dynamic analysis. Journal of Econometrics 74, 177–208.

Wang, Y., 2002. Asymptotic nonequivalence of GARCH models and diffusions. Annals of Statistics 30, 754–783.

Wu, G., Xiao, Z., 2002. A generalized partially linear model of asymmetric volatility. Journal of Empirical Finance 9, 287–319.

Yu, J., 2005. On leverage in a stochastic volatility model. Journal of Econometrics 127, 165–178.