A Semiparametric Stochastic Volatility Model

Jun Yu†

July 31, 2008

Abstract

This paper examines how volatility responds to return news in the context of stochastic volatility (SV) using a nonparametric method. The correlation structure in the classical leverage SV model is generalized based on a linear spline. In the new model the correlation between the return innovation and volatility innovation is dependent on the type of news arrived to the market. Theoretical properties of the proposed model are examined. A simulation-based maximum likelihood method is developed to estimate the new model. Simulations show that the estimation method provides reliable parameter estimates. The new model is fitted to daily and weekly data in the US and compared with the classical SV models in terms of their in-sample and out-of-sample performances. Empirical results suggest strong evidence in favor of the proposed model. In particular, the new model finds strong evidence of leverage effect when the classical model fails to identify it. Also, the new model provides better out-of-the-sample forecasts of volatility than the classical model.

JEL classification: G12, C11, C15

Keywords: Leverage effect; Simulated maximum likelihood; Laplace approximation; Spline; Realized volatility

1 Introduction

How volatility responds to return news has long been an active research topic; see Black (1976), Christie (1982), French, Schwert, and Stambaugh (1987), Campbell and Hentschel (1992), Engle and Ng (1993), Bekaert and Wu (2000), Linton and Mammen (2005) for a rather incomplete list of studies in the literature. Answer to this question has important implications for financial decision making and financial 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, Nelson and Sunier, 1995). Furthermore, an option

*I gratefully acknowledge financial support from the Singapore Ministry of Education AcRF Tier 2 fund under Grant No. T206B4301-RS.
†School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; email: yujun@smu.edu.sg.
contract would be substantially mis-priced when the relationship is misspecified (Hull and White, 1987, Duan, 1995).

It is now well accepted in the volatility literature that equity volatility responds to return news in an asymmetric fashion, namely, a piece of bad news has different impact on future volatility from the good news of the same magnitude. Arguably the most popular empirical method for examining the asymmetric volatility response is via some form of ARCH-type models.\(^1\) 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. In the context of dynamic general equilibrium economy, Aydemir, Gallmeyer and Hollifield (2006) provided a theoretical framework to examine the various scenarios for the financial leverage effect.

Volatility response can also be studied using stochastic volatility (SV) models (see Ghysels, Harvey, and Renault, 1996 for a review of SV models). Unlike ARCH-type models, SV models specify volatility as a separate random process. Due to this extra randomness, there are certain advantages in the SV models over the ARCH-type models for modeling the dynamics of asset returns (Kim, Shephard, and Chib, 1998). The third method for studying volatility response is to use realized volatility. There has been a surge of interest in this approach recently and this approach requires the availability of ultra high frequency data; see, for example, Andersen, Bollerslev, Diebold and Ebens (2001), Bollerslev, Litvinova and Tauchen (2006), Bandi and Reno (2008).

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 remain constant, regardless of price movements. On the other hand, Figlewski and Wang (2000) and 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 correlation coefficient.

---

\(^1\) Contributions in this area include Nelson (1991), Glosten, Jagannathan, and Runkle (1993), Engle and Ng (1993), Bekaert and Wu (2000), Linton and Mammen (2005), and Mishra, Su and Ullah (2008).
The central focus of the present paper is to provide a more general framework to investigate the asymmetric relationship between volatility and return news in the context of SV models. In particular, using the linear spline, we allow the correlation coefficient between the two innovations to be dependent on the size and the direction of the previous price movement nonparametrically. Since our model nests the leverage SV model, we can easily check the validity of this classical specification. Empirical applications using data on Microsoft reveal strong evidence against the classical specification both in-the-sample and out-of-the-sample and echo the empirical results found in Mishra, Su and Ullah (2008) in the GARCH framework using a nonparametric technique.

Our model extends the model studied in Harvey and Shephard (1996), Yu (2005), Omori et al (2007). Unlike Harvey and Shephard (1996) that used a quasi maximum likelihood (ML) estimation method and Yu (2005) and Omori et al (2007) that used Bayesian MCMC method, we estimate the SV models using the simulation-based maximum likelihood (SML) method. Our model is closely related to the model of Wu and Xiao (2002) where a nonparametric model was used to examine the relationship between the log implied volatility and the lagged return innovation. Our work is different from Wu and Xiao in two aspects. First, different nonparametric methods are employed. While we use the spline-based smoother, Wu and Xiao (2002) used the Nadaraya-Watson kernel method in a partial linear framework. One of the main advantages for the kernel method is its simpler theoretical analysis. However, relative to the spline-based smoother, the kernel method is slow to compute. This computational advantage is very important to us because volatility is not assumed to be an observable in our approach and hence has to be integrated out. Such an integration is of high dimension and has to be done numerically. Second, the relationship between return and log-volatility is in the physical measure in our study but is in the risk-neutral measure in theirs. 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 in the SV framework while Engle and Ng used it as a regression tool to describe relationship between the volatility and the lagged return innovation in the GARCH framework. Robinson (1991), Linton and Mammen (2005) and Yang (2006) provided more general ARCH models. All the models are of an additive structure and hence do not nest ours.

The remainder of the article is organized as follows. In Section 2 we introduce the semi-
parametric SV model, develop some statistical properties of the model, and discuss two special cases. Section 3 develops a simulation-based maximum likelihood (SML) method for estimating the proposed model and documents the performance of SML in simulations. Empirical results based on US data are presented and discussed in Section 4. Section 5 concludes. Appendices collect the proof of theorem and the derivation of likelihood.

2 The Proposed SV Model

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

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

(2.1)

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

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

(2.2)

where $1(y_t < 0) = 1$ if $y_t < 0$ and 0 otherwise. In this model, the asymmetry is induced by the indicator function. However, based on a nonparametric technique, Mishra, Su and Ullah (2008) have found evidence of further asymmetry in the residuals of fitted threshold GARCH(1,1), suggesting that with single threshold the GARCH(1,1) cannot capture all the asymmetry in the data.

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

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

(2.3)

where $m(\cdot)$ is a smooth but unknown function. Engle and Ng estimated $m(\cdot)$ using the following linear spline

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

where $\tau_i$ are the predetermined knots associated with the linear spline.
This approach is related to that of Wu and Xiao (2002) and that of Linton and Mammen (2005). In Wu and Xiao (2002), $m(\cdot)$ was estimated by the kernel method, for which asymptotic results are available; see, for example, Robinson (1988), Linton (1995) and Fan and Li (1999). Linton and Mammen (2005) replaced Equation (2.3) by the following additive specification:

$$
\sigma^2_{t+1} = \alpha_t + \sum_{j=0}^{+\infty} \phi_j(\theta)m(\epsilon_{t-j}).
$$

In contrast to GARCH-type models, the SV models specify a stochastic relationship between $\sigma^2_{t+1}$ (or $h_{t+1}$) and $y_t$ by using an additional innovation. To account for volatility asymmetry, the classical leverage SV model takes the form of, in discrete time,

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

and

$$
h_{t+1} = \varphi h_t + \gamma v_t,
$$

where $v_t$ is i.i.d. $N(0, 1)$, $\text{corr}(\epsilon_t, v_t) = \rho$. Equation (2.5) can be equivalently represented by

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

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

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

implying that on average $\ln \sigma^2_{t+1}$ 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 $\ln \sigma^2_{t+1}$ and $y_t$ is independent of the sign and the size of $\epsilon_t$, the standardized return innovation.

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 $\ln \sigma^2_{t+1}$ 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 with a more flexible relation between $\ln \sigma^2_{t+1}$ and $y_t$, we first choose $m$ knots, denoted by $\tau_1, \ldots, \tau_m$ with $\tau_1 > \cdots > \tau_m$, from the support
of $\epsilon_t$. Let $\tau_0/\tau_{m+1}$ be the right/left bound of the support of $\epsilon_t$.\(^2\) That is, the support of $\epsilon_t$ is divided into $m + 1$ intervals. Note that the sizes of the intervals need not be the same. The volatility equation is defined by

$$h_{t+1} = \varphi h_t + \sum_{i=1}^{m+1} \left( \rho_i \epsilon_t + \sqrt{1 - \rho_i^2} w_t \right) 1(\tau_{i-1} > \epsilon_t \geq \tau_i), \quad (2.8)$$

where $w_t$ is i.i.d. $\text{N}(0,1)$ and $\text{corr}(\epsilon_t, w_t) = 0$. Together with the mean equation (2.4), it defines our semiparametric SV model.

Let

$$v_t = \sum_{i=1}^{m+1} \left( \rho_i \epsilon_t + \sqrt{1 - \rho_i^2} w_t \right) 1(\tau_{i-1} > \epsilon_t \geq \tau_i) \quad (2.9)$$

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

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

Obviously the construction of $v_t$ is based on the linear spline with the basis functions,

$$(x - \tau_1)_+, \cdots, (x - \tau_m)_+,$$

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 $\rho_i = \rho_i \forall i$, $v_t = \rho \epsilon_t + \sqrt{1 - \rho^2} w_t$ and the specification becomes the classical leverage SV model. In general $\rho_i$ can have different sizes and even different signs. Following the same approach to deriving (2.7), we have

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

Clearly, on average $\ln \sigma_t^2$ is a piecewise linear function in $y_t$ with kinks at the $\tau_i$s. Between $\tau_1$ and $+\infty$ the slope of the linear function is $\rho_1$ while between $\tau_2$ and $\tau_1$ it is $\rho_2$. Below $\tau_m$, the slope is $\rho_{m+1}$.

There are two choices that one has to make in the model specification, $m$ and $\tau_1$s. Regarding $m$, ideally one should allow it to increase with the sample size. When this is the case, the

\(^2\)If the support of $\epsilon_t$ is the entire real line, then $\tau_0 = +\infty$ and $\tau_{m+1} = -\infty$. 

6
rate of convergence and the asymptotic variance may be different from standard ML results. Essentially increasing \( m \) trades off smaller bias with larger variance. The reason why the variance increases with \( m \) is because less effective observations are used to estimate \( \rho_i \) with larger \( m \). Moreover, the computational cost increases with \( m \). This is because \( h \) is not observable and need to be integrated out numerically. The larger the \( m \), the more parameters in the model and hence the more numerical iterations in the ML estimation. If \( m \) is fixed, one can resort to the classical asymptotic theory for ML to make statistical inferences. The inferences are valid as long as the model is correctly specified. A more detailed discussion of ML is provided in Section 3. The choice of \( \tau \)s could be based on trial and error or more formally a model selection criterion. However, the exercise will be computationally expensive if many sets of \( \tau \)s are considered.

Now let’s discuss some simple but important special cases of the model and establish certain statistical properties. First, if \( m = 1 \), there are only two regimes. If we further choose \( \tau_1 = 0 \), then corr(\( \epsilon_t, v_t \)) = \( \rho_1 \) if \( \epsilon_t \geq 0 \) and corr(\( \epsilon_t, v_t \)) = \( \rho_2 \) if \( \epsilon_t < 0 \). This model nests the classical leverage effect model and is called Spline1 SV model in this paper. The probability density
function (pdf) of \( v_t \) is provided in Lemma 1 of Appendix A. In Figure 1 we plot the pdf of \( v_t \) for three sets of \((\rho_1, \rho_2)\), namely, \((-0.5, -0.5), (0, -0.5), (0.5, -0.5)\). When \( \rho_1 = \rho_2 \), \( v_t \) is the standard normal and hence symmetric. When \( \rho_2 \neq \rho_1 \), \( v_t \) becomes skewed. In particular, \( v_t \) is negatively skewed when \( \rho_2 > \rho_1 \) and positively skewed when \( \rho_2 < \rho_1 \). The larger the difference between \( \rho_1 \) and \( \rho_2 \), the more skewness in \( v_t \).

We now establish statistical properties of this special case, such as stationarity, ergodicity, and moments.

**Theorem 2.1.** Define the Spline1 SV model by

\[
y_t = \sigma \exp(h_t/2)\epsilon_t, \quad \epsilon_t \sim \text{i.i.d. N}(0,1),
\]

where \( h_{t+1} = \varphi h_t + \gamma v_t \).

\[
v_t = \left( \rho_1 \epsilon_t + \sqrt{1-\rho_1^2}w_t \right)1(\epsilon_t \geq 0) + \left( \rho_2 \epsilon_t + \sqrt{1-\rho_2^2}w_t \right)1(\epsilon_t < 0)
\]

\( w_t \) is i.i.d. \( \text{N}(0,1) \) and \( \text{corr}(w_t, \epsilon_t) = 0 \). 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 given by

\[
E(y_t^{2i-1}) = 0, i = 1, 2, \cdots,
\]

\[
E(y_t^{2i}) = \frac{(2i)!}{2^i i!} \sigma^{2i} G(i, \rho_1, \rho_2, \gamma, \varphi), i = 1, 2, \cdots,
\]

where \( G(i, \rho_1, \rho_2, \gamma, \varphi) \) is defined by

\[
G(s, \rho_1, \rho_2, \gamma, \varphi) = \prod_{j=0}^{\infty} \left\{ \exp \left( \frac{1}{2}s^2 \gamma^2 \varphi^{2j} \right) \left[ \Phi (s \gamma \varphi^{j} \rho_1) + \Phi (-s \gamma \varphi^{j} \rho_2) \right] \right\}, \quad (2.10)
\]

and \( \Phi(\cdot) \) is the cumulative distribution function of the standard normal variable.

The second special case corresponds to \( m = 2 \) where there are three regimes. This model is called Spline2 SV model in this paper. It is known in the GARCH literature that when \( \epsilon_t \) is very close to zero, volatility does not respond to \( \epsilon_t \) in a significant manner (Engle and Ng (1993) and Linton and Mammen (2005)). As a result, it seems reasonable to choose \( \tau_1 \) to be a small, positive number, \( \tau_2 \) to be a small, negative number. However, if \( \tau_3 \) is too close to zero, there are too few observations to estimate \( \rho_2 \); if \( \tau_1 \) (or \( \tau_2 \)) is too far away from zero, there are too few observations to estimate \( \rho_1 \) (or \( \rho_3 \)). 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 \( \rho \)s. A drawback with such a choice is that the Spline2 SV model does not nest the Spline1 SV model.
3 Estimation and Forecasting Methods and Sampling Performance

3.1 Estimation Method

To calculate the likelihood function of SV models, one has to deal with a high-dimensional integral since the latent process $h$ needs to be integrated out from the joint probability density function, $pdf(y, h)$, where $y = (y_1, \cdots, y_T)$ and $h = (h_1, \cdots, h_{T+1})$. Unfortunately, such an integral cannot be solved analytically in general and methods are called for to approximate the integral numerically. Built upon work by Shephard and Pitt (1997), Durbin and Koopman (1997), Durham (2006) and many others, we demonstrate the effectiveness of a simulation-based maximum likelihood (SML) method which first approximates the integral via importance sampling techniques and then maximizes the log-likelihood function.

To fix the idea, denote by $\theta$ the vector of model parameters. Let the likelihood function of the parameter vector $\theta$ be

$$
pdf(y|\theta) = \int pdf(y, h|\theta)dh = \int \frac{pdf(y, h; \theta)}{q(h)}dQ(h)
$$

(3.11)

where $Q(h)$ is an importance distribution function and $q(h) = Q'(h)$ is an importance density function. The idea of the SML method is to draw samples $h^{(1)}, \ldots, h^{(S)}$ from $q$ so that we can approximate $pdf(y|\theta)$ by

$$
pdf(y|\theta) \approx \frac{1}{S} \sum_{s=1}^{S} \frac{pdf(y, h^{(s)}; \theta)}{q(h^{(s)})}
$$

(3.12)

The almost sure convergence of the right hand side to the left hand side as $S \to \infty$ follows directly from the strong law of large numbers.

The Monte Carlo sampling variance is given by

$$
\frac{1}{S} \int \left[ \frac{pdf(y, h|\theta)}{q(h)} - pdf(y|\theta) \right] q(h)dh.
$$

(3.13)

Obviously one needs to choose $q$ to minimize this variance. If there exists a $q$ so that

$$
\frac{pdf(y, h|\theta)}{q(h)} = pdf(y|\theta),
$$

the variance of Monte Carlo sampling is zero regardless of $S$. In this case $S = 1$ is large enough to accurately calculate the integral. For SV models, this is generally not possible. Since $pdf(y|\theta)$ is a constant with respect to $h$, to reduce the variance, one can match $pdf(y, h; \theta)$
and \( q(h) \) as closely as possible while ensuring that it is easy to simulate from \( q \). Following suggestions made by Shephard and Pitt (1997) and Durbin and Koopman (1997), we choose \( q \) to be the Laplace approximation to \( \text{pdf}(y, h; \theta) \) (see for example, Laplace, 1986). That is \( h^{(s)} \sim N(h^{*}, -\Omega^{-1}) \) where

\[
h^* = \arg \max_h \log \text{pdf}(y, h; \theta)
\]

and

\[
\Omega = \frac{\partial^2 \log \text{pdf}(y, h^*; \theta)}{\partial h \partial h'}
\]

Hence, the importance density is given by the multivariate ((\( T + 1 \))-dimensional) normal with mean \( h^* \) and variance \(-\Omega^{-1}\). The term Importance Sampling emphasizes the fact that the importance density \( q(h) \) is designed to sample mostly from the important part of \( \text{pdf}(y, h; \theta) \).

To approximate \( \text{pdf}(y|\theta) \) via the importance sampling method, one needs the expression for \( \log \text{pdf}(y, h; \theta) \), which we develop now.

\[
\text{pdf}(y, h) = \text{pdf}(h_{T+1}|y, h_1, \ldots, h_T)\text{pdf}(y, h_1, \ldots, h_T)
\]

\[
= \text{pdf}(h_{T+1}|y_T, h_T)\text{pdf}(y_T|y_1, \ldots, y_{T-1}, h_1, \ldots, h_T)\text{pdf}(y_1, \ldots, y_{T-1}, h_1, \ldots, h_T)
\]

\[
= \cdots
\]

\[
= \text{pdf}(h_1) \prod_{t=1}^{T} \{\text{pdf}(h_{t+1}|y_t, h_t)\text{pdf}(y_t|h_t)\}.
\]

Taking the log, we get

\[
\log \text{pdf}(y, h) = \log \text{pdf}(h_1) + \sum_{t=1}^{T} \log \text{pdf}(h_{t+1}|y_t, h_t) + \sum_{t=1}^{T} \log \text{pdf}(y_t|h_t).
\]

In Appendix B, we give the detailed expression for each component on the right hand side of Equation (3.16) for the Spline1 SV model.

Based on the expression of \( \log \text{pdf}(y, h) \), we obtain the first and second derivatives of \( \log \text{pdf}(y, h) \) with respect to \( h \). The mode, \( h^* \), is obtained by iterating Newton’s method until the full convergence. We use the standard asymptotic results for ML, such as the asymptotic normality, to make statistical inferences.

Two numerical issues are in order. First, it is important to note that one needs apply the Common Random Numbers (CRNs) technique during the numerical optimization so that the objective function has a smooth surface. That is, the fixed set of canonical random numbers
are always obtained from standardized normals during the numerical optimization. Second, as in other simulation-based methods, it is important to examine the simulation induced errors. To do so we follow the suggestion made by Durham (2006) by fitting the same model to the same dataset repeatedly with different random seeds. In this paper, we measure the simulation error of each estimator (including the log-likelihood function) by the variability (ie the standard deviation) of the estimator across different random seeds. We term the simulation error the Monte Carlo standard error (MCSE). Moreover, instead on reporting the estimates and the log-likelihood value based on single random seed, we report the mean of the each estimator and the mean of the log-likelihood value, across different random seeds.

3.2 Forecasting Volatility

Following the suggestion of Skaug and Yu (2007), we now show how to forecast volatility out-of-the-sample. The one-period-ahead volatility (denoted by $\hat{\sigma}_{T+1}$) can be obtained as a by-product of the Laplace approximation, ie,

$$\hat{\sigma}_{T+1} = \hat{\sigma} \exp(h^*_{T+1}/2),$$

where $h^*_{T+1}$ is the last element in $h^*$ evaluated at the ML estimates.

To forecast volatility $K$-period-ahead (denoted by $\hat{\sigma}_{T+K}$), we just need to re-define $h$ by $(h_1, \ldots, h_T, h_{T+1}, \ldots, h_{T+K})$ and then calculate

$$\hat{\sigma}_{T+K} = \hat{\sigma} \exp(h^*_{T+K}/2),$$

where $h^*_{T+K}$ is the last element in $h^*$ evaluated at the ML estimates.

3.3 Sampling Performance

Since the Spline1 SV model performs the best in our empirical applications, we choose it as a benchmark model to check the reliability of the SML method in simulations. In the simulations, the values of $\varphi$, $\sigma$ and $\gamma$ are set to 0.95, 1, 0.20, respectively. We fix $\mu_y$ to zero and assume it is known. For $(\rho_1, \rho_2)$, we choose the following two combinations: $(-0.5, -0.5)$, $(0.5, -0.5)$. As it will be clear from empirical studies reported below, these parameter values are empirically realistic.

For each parameter setting, 1,000 observations, simulated from the true model, are used to estimate the five parameters, $\varphi$, $\sigma$, $\gamma$, $\rho_1$, and $\rho_2$. We then replicate the experiment for 500
times to obtain the mean and the root mean square error (RMSE) for each parameter estimate across 500 replications. In all cases, we choose \( S = 64 \) for the importance sampler.

Table 1 summarize the results. There are two panels in the table, corresponding to the two cases for \((\rho_1, \rho_2)\). For each parameter setting, we report the true value, the mean and the RMSE. The general picture that emerges from this table is that SML provides good performance for all the parameters in both cases. Moreover, it seems more difficult to estimate \(\sigma, \rho_1, \rho_2\) when \(\rho_1\) is further away from \(\rho_2\).

<table>
<thead>
<tr>
<th>(\sigma)</th>
<th>(\varphi)</th>
<th>(\gamma)</th>
<th>(\rho_1)</th>
<th>(\rho_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>1</td>
<td>0.95</td>
<td>0.2</td>
<td>-0.5</td>
</tr>
<tr>
<td>Mean</td>
<td>1.0164</td>
<td>0.9435</td>
<td>0.2045</td>
<td>-0.5071</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1544</td>
<td>0.0213</td>
<td>0.0402</td>
<td>0.1484</td>
</tr>
<tr>
<td>True</td>
<td>1</td>
<td>0.95</td>
<td>0.2</td>
<td>-0.5</td>
</tr>
<tr>
<td>Mean</td>
<td>1.1053</td>
<td>0.9434</td>
<td>0.2055</td>
<td>-0.5023</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.4027</td>
<td>0.0213</td>
<td>0.0429</td>
<td>0.2506</td>
</tr>
</tbody>
</table>

4 Empirical Results

4.1 Estimation Results from Daily Data

In this subsection we fit the basic SV model, the classical leverage SV model, the Spline1 SV model and the Spline2 SV model to two continuously compounded daily return series in the US, namely S&P500 from January 1985 to December 1989 and Microsoft from January 1987 to December 1991.\(^3\) The number of observations for these three series is 1,263 and 1,264, respectively. The data are plotted in Figure 3.

Estimation results, including the parameter estimates, the asymptotic standard deviations (in parenthesis), the Monte Carlo standard errors (in bracket) and the minus log-likelihood values are reported in Table 2.\(^4\) Several conclusions can be drawn. First, the log-volatility persistence \(\varphi\) is highly significant in all cases, with estimated values ranging from 0.91 to 0.95.

\(^3\) All the daily closing prices can be downloaded from yohoo.finance. For Microsoft, since no data are available before March 13, 1986, we begin our sample from January 1987. Both return series have been multiplied by 100 and demeaned so that \(\mu_y\) is not estimated.

\(^4\) The results are obtained from 50 random seeds.
Second, in the leverage SV model, $\rho$ is estimated to be negative for both series, featured by -0.3107 and -0.1717. However, $\rho$ is significant only for S&P500. That is why for S&P500 the leverage SV model provides a significant improvement over the basic SV model in terms of likelihood. For Microsoft $\rho$ is insignificant and the log-likelihood value between the two SV models differs by only 0.003. While the fact that estimated $\rho$ 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 Andersen, Bollerslev, Diebold and Ebens (2001). The result indicates that one would conclude the absence of leverage effect if only the leverage SV model is fitted to the Microsoft data. As it will be clear below, this conclusion is a mistake.

Third, the Spline1 SV model provides a significant improvement over the leverage SV model. In particular, the likelihood ratio (LR) statistic for leverage SV against Spline1 SV is 17.02 and 27.64, respectively. The p-value of LR statistic (measured against the $\chi^2_1$ distribution) is 0.000 for both series. This result echoes the empirical result reached in Mishra,
Su and Ullah (2008) where the single threshold GARCH model is found to be inadequate to explain all the asymmetry in volatility response. More interest results emerge if one examines the estimates of $\rho_1$ and $\rho_2$. In the estimated Spline1 SV model, $\rho_1$ is more negative than $\rho$ in the estimated leverage SV and statistically significant for both series. On the other hand, $\rho_2$ is estimated to be positive and statistically significant for both series. It is rather surprising to find positive estimates for $\rho_2$. We will examine the out-of-the-sample performance of the estimated model in Section 4.3. While we fail to find a significant leverage effect in the leverage SV model for Microsoft, we do find the strong evidence of leverage effect in the Spline1 SV model. Fourth, in the estimated Spline2 SV model, as expected, for S&P500 $\rho_2$ is close to 0. It has a very large standard error. The estimates for the other two $\rho$s are similar to those in the estimated Spline1 SV model. Not surprisingly, the likelihood value improve little by adding one more knot to the Spline1 SV model. The results in the estimated Spline2 SV model are slightly different for Microsoft. The estimate of $\rho_2$ is -0.9556. While this is significantly different from zero but not significantly different from the estimate of $\rho_1$ in the same model (-0.7446) nor from $\rho_1$ in the estimated Spline1 SV model (-0.7813). The estimate of $\rho_3$ is 0.6908 which is similar to the estimate of $\rho_2$ is the Spline1 SV model (0.7161). While we cannot use the LR statistic to compare Spline1 SV and Spline2 SV, the difference in the two log-likelihood values is so small, indicating weak evidence to support Spline2 SV for both series.

4.2 Estimation Results from Weekly Data

In this subsection we fit the basic SV model, the classical SV model with leverage effect, the Spline1 SV model and the Spline2 SV model to two continuously compounded weekly return series in the US, namely S&P500 and Microsoft, both from April 11, 1986 to December 28, 2007. The number of observations for is 1,133. The data are plotted in Figure 3.

Estimation results, including the parameter estimates, the asymptotic standard deviations (in parenthesis), the Monte Carlo standard errors (in bracket) and the log-likelihood values are reported in Table 3. Several conclusions can be drawn. First, the log-volatility persistence

\footnote{We use the closing price of each week. Data can be downloaded from yohoo.finance. Neither return series is demeaned and $\mu_y$ is estimated.}

\footnote{The results are obtained from 50 random seeds.}
Table 2: Estimation Results from Daily Data. The number is in parenthesis is the asymptotic standard deviation. The number in bracket is the Monte Carlo standard error.

<table>
<thead>
<tr>
<th>Data</th>
<th>Model</th>
<th>-Log L</th>
<th>$\sigma$</th>
<th>$\varphi$</th>
<th>$\gamma$</th>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>$\rho_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP500</td>
<td>Basic</td>
<td>1659.68</td>
<td>.8386</td>
<td>.9453</td>
<td>.2610</td>
<td>(.059)</td>
<td>(.019)</td>
<td>(.0057)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.044)</td>
<td>(.0166)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leverage</td>
<td>1654.11</td>
<td>.8431</td>
<td>.9161</td>
<td>.3232</td>
<td>(.048)</td>
<td>(.027)</td>
<td>(.0036)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.056)</td>
<td>(.0121)</td>
<td>(.0052)</td>
</tr>
<tr>
<td></td>
<td>Spline1</td>
<td>1645.60</td>
<td>2.129</td>
<td>.9118</td>
<td>.3916</td>
<td>(.748)</td>
<td>(.021)</td>
<td>(.0057)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.039)</td>
<td>(.0214)</td>
<td>(.0133)</td>
</tr>
<tr>
<td></td>
<td>Spline2</td>
<td>1645.51</td>
<td>2.101</td>
<td>.9097</td>
<td>.4028</td>
<td>(.160)</td>
<td>(.021)</td>
<td>(.0038)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.062)</td>
<td>(.0212)</td>
<td>(.0133)</td>
</tr>
<tr>
<td>MSFT</td>
<td>Basic</td>
<td>3045.79</td>
<td>2.5380</td>
<td>.9757</td>
<td>.2368</td>
<td>(.170)</td>
<td>(.020)</td>
<td>(.0049)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.061)</td>
<td>(.0154)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leverage</td>
<td>3043.92</td>
<td>2.5350</td>
<td>.9437</td>
<td>.2473</td>
<td>(.166)</td>
<td>(.023)</td>
<td>(.0039)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.049)</td>
<td>(.0134)</td>
<td>(.0046)</td>
</tr>
<tr>
<td></td>
<td>Spline1</td>
<td>3030.10</td>
<td>12.004</td>
<td>.9314</td>
<td>.4031</td>
<td>(5.362)</td>
<td>(.016)</td>
<td>(.0054)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.3994)</td>
<td>(.029)</td>
<td>(.0121)</td>
</tr>
<tr>
<td></td>
<td>Spline2</td>
<td>3028.51</td>
<td>11.987</td>
<td>.9330</td>
<td>.3889</td>
<td>(.5611)</td>
<td>(.017)</td>
<td>(.0050)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.4142)</td>
<td>(.017)</td>
<td>(.0153)</td>
</tr>
</tbody>
</table>

$\varphi$ is highly significant in all cases and higher than those reported in Table 2, with estimated values ranging from 0.95 to 0.99. Second, in the leverage SV model, $\rho$ is estimated to be $-0.4729$, $-0.059$, respectively, for the two series. As in the daily data, $\rho$ is statistically significant for the index but not for the stock. Once again the results are reinforced by the the LR statistic, 13.2 for S&P500 but only 0.12 for Microsoft. The p-value of the LR statistic is 0.000, and 0.729, respectively. Hence one would conclude the absence of the leverage effect by fitting the leverage SV model to Microsoft.

Third, the Spline1 SV model provides no significant improvement over the leverage SV model for the S&P500 with the log-likelihood value remaining essentially unchanged. The story is completely different for Microsoft. The LR statistic is 8.02 which is significant at the 1% level. The estimated $\rho_1$ is negative (-0.3472) and the estimated $\rho_2$ is positive (0.2986), both being significant at the 10% level. This estimated signs for $\rho$s corroborates well with those in the daily data. Fourth, in the estimated Spline2 SV model, the estimate of $\rho_1$ is very close to the estimate of $\rho_2$ for both series. They are both close to the estimate of $\rho_1$ in the Spline1 SV model. Also, the estimate of $\rho_3$ in the Spline2 SV model is close to the estimate of $\rho_2$ in the Spline1 SV model. Not surprisingly, the likelihood value improves little by adding one more knot to the Spline1 SV model for both series. Hence, there is no evidence to support
the Spline2 SV model in the weekly data.

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 two weekly return series from the last subsection, namely the weekly return series for S&P500 and Microsoft from April 11, 1986 to December 28, 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. While we focus on one-period ahead weekly volatility in this paper, multi-period-ahead forecasts and forecasts in other frequencies can be obtained in the same manner.

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
Table 3: Estimation Results from Weekly Data. The number in parenthesis is the asymptotic standard deviation. The number in bracket is the Monte Carlo standard error.

<table>
<thead>
<tr>
<th>Data</th>
<th>Model</th>
<th>Log L</th>
<th>( \mu_y )</th>
<th>( \sigma )</th>
<th>( \nu )</th>
<th>( \rho_1 )</th>
<th>( \rho_2 )</th>
<th>( \rho_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP500</td>
<td>Basic</td>
<td>2860.11 (0.0015)</td>
<td>0.00242 (0.0005)</td>
<td>0.0186 (0.0001)</td>
<td>9734 (0.0018)</td>
<td>1675 (0.0132)</td>
<td>0.411 (0.0001)</td>
<td>0.001 (0.0005)</td>
</tr>
<tr>
<td></td>
<td>Leverage</td>
<td>2867.71 (0.0016)</td>
<td>0.00188 (0.0005)</td>
<td>0.0184 (0.0001)</td>
<td>9512 (0.0005)</td>
<td>221 (0.0002)</td>
<td>-0.4729 (0.0012)</td>
<td>-0.424 (0.0003)</td>
</tr>
<tr>
<td></td>
<td>Spline1</td>
<td>2867.71 (0.0017)</td>
<td>0.00188 (0.0005)</td>
<td>0.0184 (0.0001)</td>
<td>9512 (0.0005)</td>
<td>221 (0.0002)</td>
<td>-0.4729 (0.0012)</td>
<td>-0.424 (0.0003)</td>
</tr>
<tr>
<td></td>
<td>Spline2</td>
<td>2867.71 (0.0017)</td>
<td>0.0019 (0.0005)</td>
<td>0.0186 (0.0001)</td>
<td>9644 (0.0005)</td>
<td>191 (0.0002)</td>
<td>-0.4172 (0.0012)</td>
<td>-0.4275 (0.0012)</td>
</tr>
<tr>
<td>MSFT</td>
<td>Basic</td>
<td>1903.40 (0.0001)</td>
<td>0.0043 (0.0002)</td>
<td>0.0450 (0.0001)</td>
<td>9886 (0.0009)</td>
<td>1202 (0.0255)</td>
<td>0.001 (0.0006)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leverage</td>
<td>1903.46 (0.0001)</td>
<td>0.0042 (0.0002)</td>
<td>0.0452 (0.0001)</td>
<td>9888 (0.0008)</td>
<td>1207 (0.0296)</td>
<td>0.001 (0.0006)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spline1</td>
<td>1907.47 (0.0002)</td>
<td>0.0042 (0.0002)</td>
<td>0.0480 (0.0001)</td>
<td>9858 (0.0005)</td>
<td>1213 (0.0320)</td>
<td>0.001 (0.0006)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spline2</td>
<td>1907.47 (0.0002)</td>
<td>0.0042 (0.0002)</td>
<td>0.0480 (0.0001)</td>
<td>9860 (0.0005)</td>
<td>1301 (0.0242)</td>
<td>0.001 (0.0006)</td>
<td></td>
</tr>
</tbody>
</table>

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 comes directly from standard stochastic process theory, according to which the empirical quadratic variation converges to integrated volatility as the infill sampling frequency goes to zero (Andersen, Bollerslev, Diebold and Labys (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 Friday, April 11, 1986 to Friday, 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 28, 2005. Therefore, the final forecast is for the week that begins December 31, 2008. In total, we need to make
114 forecasts from each model. We match each forecasted volatility with the corresponding realized volatility.

In Table 4, we report the Mean Absolute Error (MAE) and the Mean Square Error (MSE) in order to evaluate forecast accuracy. In all cases, the Spline1 SV model performs the best, followed by the leverage SV model and then by the basic SV model. For S&P500, the improvement from the leverage SV model over the basic SV model is more remarkable than that from the Spline1 SV model over the leverage SV model. However, for Microsoft, the improvement from the Spline1 SV model over the leverage SV model is more remarkable, suggesting that the Spline1 SV model indeed leads to superior forecasts.
Figure 5: Recursive estimates of $\rho_1$ and $\rho_2$ from the Spline1 SV model from S&P500.

Figure 6: Recursive estimates of $\rho_1$ and $\rho_2$ from the Spline1 SV model from Microsoft.
5 Conclusion

Using the linear spline we propose a semiparametric stochastic volatility model so that volatility can respond to lagged return news in a more flexible way. On average the log-volatility may respond differently to the lagged return news depending on the sign and the size of the previous price movements. The model nests the basic SV model and the leverage SV model. Statistical properties of the proposed model are discussed. The model is fitted to daily and weekly US index and stock returns. For the daily index and stock returns and the weekly stock return, we found the superior in-the-sample performance of the proposed model. While we could not find significant leverage effect in the leverage SV model for the daily and weekly stock returns, we found strong evidence of leverage effect in the new model. Not only does the model perform better in-the-sample, but also it yields more accurate forecasts than the classical models.

APPENDIX A

To prove the theorem, we need the following lemma. Denote by \( pdf_X(x) \) be the pdf of random variable \( X \). Similarly one defines \( pdf_Y(y) \), \( pdf_Z(z) \), \( pdf_{X,Y}(x,y) \), \( pdf_{X|Y=y}(x) \) and the cdf counterparts. Denote by \( \phi(\cdot) \) and \( \Phi(\cdot) \) the pdf and the cdf of standard normal, respectively.

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

\[
Z = \begin{cases} 
\rho_1 X + \sqrt{1 - \rho_1^2} Y, & \text{if } X \geq 0 \\
\rho_2 X + \sqrt{1 - \rho_2^2} Y, & \text{if } X < 0,
\end{cases}
\]

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

\[
m_Z(s) = \exp \left( s^2 / 2 \right) \left( \Phi(s \rho_1) + \Phi(-s \rho_2) \right), \quad (A.1)
\]

and the pdf of \( Z \) is

\[
pdf_Z(z) = \phi(z) \left\{ \Phi \left( \frac{\rho_1 z}{\sqrt{1 - \rho_1^2}} \right) + \Phi \left( -\frac{\rho_2 z}{\sqrt{1 - \rho_2^2}} \right) \right\}. \quad (A.2)
\]

Moreover, the joint pdf of \( (Z, X) \) is given by

\[
pdf_{Z,X}(z, x) = \begin{cases} 
\phi(z) \phi \left( \frac{x - \rho_1 z}{\sqrt{1 - \rho_1^2}} \right) \frac{1}{\sqrt{1 - \rho_1^2}}, & \text{if } x \geq 0 \\
\phi(z) \phi \left( \frac{x - \rho_2 z}{\sqrt{1 - \rho_2^2}} \right) \frac{1}{\sqrt{1 - \rho_2^2}}, & \text{if } x < 0.
\end{cases} \quad (A.3)
\]
and the conditional pdf of $Z|X = x$ is given by

$$
pdf_{Z|X=x}(z) = \begin{cases} 
\frac{\phi(z)}{\phi(x)} \phi \left( \frac{x-\rho_1 z}{\sqrt{1-\rho_1^2}} \right) \frac{1}{\sqrt{1-\rho_1^2}}, & \text{if } x \geq 0 \\
\frac{\phi(z)}{\phi(x)} \phi \left( \frac{x-\rho_2 z}{\sqrt{1-\rho_2^2}} \right) \frac{1}{\sqrt{1-\rho_2^2}}, & \text{if } x < 0 
\end{cases}
$$

(A.4)

Proof of Lemma 1: The mgf of $Z$ is

$$m_Z(s) = E \{ \exp(sz) \} = E \left[ \exp \left( s \left( \rho_1 x + \sqrt{1-\rho_1^2} y \right) 1(x \geq 0) + s \left( \rho_2 x + \sqrt{1-\rho_2^2} y \right) 1(x < 0) \right) \right]$$

$$= E \left[ \exp \left( s \left( \rho_1 x + \sqrt{1-\rho_1^2} y \right) 1(x \geq 0) \right) \exp \left( s \left( \rho_2 x + \sqrt{1-\rho_2^2} y \right) 1(x < 0) \right) \right]$$

$$= \int_{-\infty}^{\infty} \int_{0}^{\infty} \exp \left( s \rho_1 x + s \sqrt{1-\rho_1^2} y \right) pdf_X(x) pdf_Y(y) dxdy + \int_{-\infty}^{\infty} \int_{0}^{\infty} \exp \left( s \rho_2 x + s \sqrt{1-\rho_2^2} y \right) pdf_X(x) pdf_Y(y) dxdy$$

$$= \int_{-\infty}^{\infty} \exp \left( s \sqrt{1-\rho_1^2} y \right) pdf_Y(y) dy \int_{0}^{\infty} \exp (sp_1 x) pdf_X(x) ddx + \int_{-\infty}^{\infty} \exp \left( s \sqrt{1-\rho_2^2} y \right) pdf_Y(y) dy \int_{0}^{\infty} \exp (sp_2 x) pdf_X(x) ddx$$

$$= \exp (s^2(1-\rho_1^2)) \Phi(sp_1) \exp (s^2 \rho_1^2) + \exp (s^2(1-\rho_2^2)) \Phi(-sp_2) \exp (s^2 \rho_2^2)$$

$$= \exp (s^2/2) \left( \Phi(sp_1) + \Phi(-sp_2) \right).$$

When $\rho_1 = \rho_2 = \rho$, this mgf becomes $\exp (s^2/2)$ which is the mgf of the standard normal.

To find $pdf_Z(z)$, we first calculate $cdf_Z(z)$ as

$$cdf_Z(z) = \Pr(Z < z) = \Pr \left( \left( \rho_1 X + \sqrt{1-\rho_1^2} Y \right) 1(X \geq 0) + \left( \rho_2 X + \sqrt{1-\rho_2^2} Y \right) 1(X < 0) < z \right)$$

$$= \int_{Z < z} pdf_{X,Y}(x,y) dxdy$$

$$= \int_{0}^{\infty} \int_{-\infty}^{\infty} \sqrt{1-\rho_1^2} pdf_X(x) pdf_Y(y) dxdy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sqrt{1-\rho_2^2} pdf_X(x) pdf_Y(y) dxdy$$

$$= \int_{0}^{\infty} pdf_X(x) \Phi \left( \frac{z - \rho_1 x}{\sqrt{1-\rho_1^2}} \right) dx + \int_{-\infty}^{0} pdf_X(x) \Phi \left( \frac{z - \rho_2 x}{\sqrt{1-\rho_2^2}} \right) dx.$$
Hence, the pdf of $Z$ is

$$
pdf_Z(z) = \frac{\partial pdf_Z(z)}{\partial z}
= \int_0^\infty pdf_X(x) \phi\left(\frac{z - \rho_1 x}{\sqrt{1 - \rho_1^2}}\right) \frac{1}{\sqrt{1 - \rho_1^2}} dx + \int_{-\infty}^0 pdf_X(x) \phi\left(\frac{z - \rho_2 x}{\sqrt{1 - \rho_2^2}}\right) \frac{1}{\sqrt{1 - \rho_2^2}} dx
= \frac{1}{\sqrt{1 - \rho_1^2}} \int_0^\infty \phi(x) \phi\left(\frac{z - \rho_1 x}{\sqrt{1 - \rho_1^2}}\right) dx + \frac{1}{\sqrt{1 - \rho_2^2}} \int_{-\infty}^0 \phi(x) \phi\left(\frac{z - \rho_2 x}{\sqrt{1 - \rho_2^2}}\right) dx.
$$

(A.5)

Note that

$$
\phi(x) \phi\left(\frac{z - \rho_1 x}{\sqrt{1 - \rho_1^2}}\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \exp\left(-\frac{x^2}{2}\right) \exp\left(-\frac{(z - \rho_1 x)^2}{2(1 - \rho_1^2)}\right) = \frac{1}{\sqrt{2\pi(1 - \rho_1^2)}} \exp\left(-\frac{(x - \rho_1 z)^2}{2(1 - \rho_1^2)}\right) \frac{1 - \rho_1^2}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).
$$

So we have

$$
\int_0^\infty \phi(x) \phi\left(\frac{z - \rho_1 x}{\sqrt{1 - \rho_1^2}}\right) dx = \Phi\left(\frac{\rho_1 z}{\sqrt{1 - \rho_1^2}}\right) \frac{1 - \rho_1^2}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).
$$

Similarly

$$
\int_{-\infty}^0 \phi(x) \phi\left(\frac{z - \rho_2 x}{\sqrt{1 - \rho_2^2}}\right) dx = \Phi\left(-\frac{\rho_2 z}{\sqrt{1 - \rho_2^2}}\right) \frac{1 - \rho_2^2}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).
$$

Substituting them into (A.5), we obtain the expression for $pdf_Z(z)$.  

22
To find \( pdf_{Z,X}(z,x) \), we first calculate \( cdf_{Z,X}(z,x) \) as

\[
cdf_{Z,X}(z,x') = \Pr(Z < z, X < x')
\]

\[
= \Pr \left( \left( \rho_1 X + \sqrt{1 - \rho_1^2} Y \right) 1(X \geq 0) + \left( \rho_2 X + \sqrt{1 - \rho_2^2} Y \right) 1(X < 0) < z, X < x' \right)
\]

\[
= \int_{\{Z < z\} \cap \{X < x'\}} \text{pdf}_{X,Y}(x,y) \, dxdy
\]

\[
= \begin{cases} 
\int_{-\infty}^{x'} \int_{-\infty}^{\frac{z - \rho_1 x}{\sqrt{1 - \rho_1^2}}} \phi(x) \phi(y) dxdy + \int_{-\infty}^{0} \int_{-\infty}^{\frac{z - \rho_2 x}{\sqrt{1 - \rho_2^2}}} \phi(x) \phi(y) dxdy, & \text{if } x \geq 0 \\
\int_{-\infty}^{x'} \int_{-\infty}^{\frac{z - \rho_2 x}{\sqrt{1 - \rho_2^2}}} \phi(x) \phi(y) dxdy, & \text{if } x < 0,
\end{cases}
\]

\[
= \begin{cases} 
\int_{0}^{x'} \phi(x) \Phi \left( \frac{z - \rho_1 x}{\sqrt{1 - \rho_1^2}} \right) dx + \int_{-\infty}^{0} \phi(x) \Phi \left( \frac{z - \rho_2 x}{\sqrt{1 - \rho_2^2}} \right) dx, & \text{if } x \geq 0 \\
\int_{-\infty}^{x'} \phi(x) \Phi \left( \frac{z - \rho_2 x}{\sqrt{1 - \rho_2^2}} \right) dx, & \text{if } x < 0
\end{cases}
\]

Hence, the joint pdf of \( Z \) and \( X \) is

\[
\text{pdf}_{Z,X}(z,x) = \frac{\partial \text{cdf}_{Z,X}(z,x)}{\partial z \partial x}
\]

\[
= \begin{cases} 
\phi(z) \phi \left( \frac{(x - \rho_1 z)}{\sqrt{1 - \rho_1^2}} \right) \frac{1}{\sqrt{1 - \rho_1^2}}, & \text{if } x \geq 0 \\
\phi(z) \phi \left( \frac{(x - \rho_2 z)}{\sqrt{1 - \rho_2^2}} \right) \frac{1}{\sqrt{1 - \rho_2^2}}, & \text{if } x < 0,
\end{cases}
\]

and the conditional pdf of \( Z \) on \( X = x \) is

\[
\text{pdf}_{Z|X=x}(z) = \frac{\text{pdf}_{Z,X}(z,x)}{\text{pdf}_{X}(x)}
\]

\[
= \begin{cases} 
\frac{\phi(z)}{\phi(x)} \phi \left( \frac{(x - \rho_1 z)}{\sqrt{1 - \rho_1^2}} \right) \frac{1}{\sqrt{1 - \rho_1^2}}, & \text{if } x \geq 0 \\
\frac{\phi(z)}{\phi(x)} \phi \left( \frac{(x - \rho_2 z)}{\sqrt{1 - \rho_2^2}} \right) \frac{1}{\sqrt{1 - \rho_2^2}}, & \text{if } x < 0
\end{cases}
\]

**Proof of Theorem 2.1:** First let us find the moment generating function of \( h_{t+1} \). Since \( |\varphi| < 1 \), substituting recursively for the \( h_t \) terms lets us rewrite Equation (2.5) as

\[
h_{t+1} = \gamma \sum_{j=0}^{\infty} \varphi^j v_{t-j}.
\]
Therefore, the mgf of $h_{t+1}$ is

$$E(\exp(sh_{t+1})) = E \left[ \exp \left( s\gamma \sum_{j=0}^{\infty} \varphi^j v_{t-j} \right) \right]$$

$$= \prod_{j=0}^{\infty} E \left[ \exp \left( s\gamma \varphi^j v_{t-j} \right) \right]$$

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

$$\equiv G(i, \rho_1, \rho_2, \gamma, \varphi), \quad (A.6)$$

where the last step is from (A.1) of Lemma 1.

Now let us show that the variance of $v_t$ is finite.

$$\text{Var}(v_t) = \frac{\partial^2 \log m_{v_t}(s)}{\partial s^2} \bigg|_{s=0} = 1 - \frac{(\rho_1 - \rho_2)^2}{2\pi} < \infty$$

Since $h_{t+1}$ is a linear process with finite innovation variance, the stationarity and ergodicity are ensured if and only in $|\varphi| < 1$.

As argued above, the moments of $h_t$ can be obtained by differentiating the log moment generating function. To obtain the moments of $y_t$, note that for $i = 1, 2, \cdots$,

$$E(y_t^{2i-1} | h_t) = E \left( \sigma^{2i-1} \exp \left( \frac{2i - 1}{2} h_t \right) \epsilon_t^{2i-1} | h_t \right)$$

$$= \sigma^{2i-1} \exp \left( \frac{2i - 1}{2} h_t \right) E \left( \epsilon_t^{2i-1} | h_t \right) = 0$$

and

$$E(y_t^{2i} | h_t) = E \left( \sigma^{2i} \exp (ih_t) \epsilon_t^{2i} | h_t \right)$$

$$= \sigma^{2i} \exp (ih_t) E \left( \epsilon_t^{2i} | h_t \right)$$

$$= \sigma^{2i} \exp (ih_t) \frac{(2i)!}{2^i i!}$$

Hence,

$$E(y_t^{2i-1}) = E \left( E(y_t^{2i-1} | h_t) \right) = 0$$

and

$$E(y_t^{2i}) = E \left( E(y_t^{2i} | h_t) \right)$$

$$= \sigma^{2i} E \left( \exp (ih_t) \right) \frac{(2i)!}{2^i i!}$$

$$= \sigma^{2i} \frac{(2i)!}{2^i i!} G(i, \rho_1, \rho_2, \gamma, \varphi)$$

24
APPENDIX B

Expression of $\log pdf(y,h)$: First note that $y_t|h_t \sim N(0, \sigma \exp(0.5h_t))$, we have

$$\log pdf(y_t|h_t) = -\log \sigma - \frac{1}{2} h_t - \frac{1}{2} \log 2\pi - \frac{\epsilon_t^2}{2}$$

Moreover, by (A.4) of Lemma 1 and the change-of-variable technique, we get

$$pdf(h_{t+1}|y_t,h_t) = pdf(\varphi h_t + \gamma v_t|y_t,h_t)$$

$$= \begin{cases} \phi(v_t) \phi(\frac{\epsilon_t - \rho h_t v_t}{\sqrt{1-\rho^2}}) \frac{1}{\sqrt{1-\rho^2}} & \text{if } \epsilon_t \geq 0 \\ \phi(v_t) \phi(\frac{\epsilon_t - \rho h_t v_t}{\sqrt{1-\rho^2}}) \frac{1}{\sqrt{1-\rho^2}} & \text{if } \epsilon_t < 0 \end{cases} ,$$

which implies

$$\log pdf(h_{t+1}|y_t,h_t) = \begin{cases} -\frac{1}{2} \log(2\pi \gamma^2 (1-\rho^2)) - \frac{1}{2} v_t^2 - \frac{\rho_1 \epsilon_t^2}{2(1-\rho_1^2)} \left[ \rho_1 \epsilon_t^2 + \rho_2 v_t^2 - 2\epsilon_t v_t \right] , & \text{if } \epsilon_t \geq 0 \\ -\frac{1}{2} \log(2\pi \gamma^2 (1-\rho^2)) - \frac{1}{2} v_t^2 - \frac{\rho_2 \epsilon_t^2}{2(1-\rho_2^2)} \left[ \rho_2 \epsilon_t^2 + \rho_2 v_t^2 - 2\epsilon_t v_t \right] , & \text{if } \epsilon_t < 0. \end{cases}$$

While tedious it is straightforward to obtain the first and second derivatives of $\log pdf(y,h)$, both with respect to $h$, which are omitted for brevity.

References


