4-2002

# A class of nonlinear stochastic volatility models

Jun YU
*Singapore Management University*, yujun@smu.edu.sg

Zhenlin YANG
*Singapore Management University*, zlyang@smu.edu.sg

# A Class of Nonlinear Stochastic Volatility Models*

Jun Yu† and Zhenlin Yang‡

April 18, 2002

**Abstract**

This paper proposes a class of nonlinear stochastic volatility models based on the Box-Cox transformation which offers an alternative to the one introduced in Andersen (1994). The proposed class encompasses many parametric stochastic volatility models that have appeared in the literature, including the well known lognormal stochastic volatility model, and has an advantage in the ease with which different specifications on stochastic volatility can be tested. In addition, the functional form of transformation which induces marginal normality of volatility is obtained as a byproduct of this general way of modeling stochastic volatility. The efficient method of moments approach is used to estimate model parameters. Empirical results reveal that the lognormal stochastic volatility model is rejected for daily index return data but not for daily individual stock return data. As a consequence, the stock volatility can be well described by the lognormal distribution as its marginal distribution, consistent with the result found in a recent literature (cf Andersen et al (2001a)). However, the index volatility does not follow the lognormal distribution as its marginal distribution.

*JEL classification:* G12, C22, C52
*Key words*: Box-Cox Transformation; GARCH; EMM; Stochastic Volatility.

*We have benefited greatly from conversations with Ronald Gallant, Genshiro Kitagawa and Xibin Zhang, and from comments and suggestions of Charles Corrado, Steve Heston, Neil Shephard and Yiu Kuen Tse. The first author gratefully acknowledges financial support from the Royal Society of New Zealand Marsden Fund under number 01-UOA-015. This research was started while the first author was visiting National University of Singapore in 2001 and he would like to thank Department of Economics at National University of Singapore for hospitality. We also thank Calvin Chan and Shirley Huang for computing support. Corresponding author is Jun Yu.

†Department of Economics, University of Auckland, Private Bag 92019, Auckland, New Zealand; email: j.yu@auckland.ac.nz

‡Department of Economics, National University of Singapore, Singapore; email: ecsyzl@nus.edu.sg

# 1 Introduction

Modeling the volatility of financial time series via stochastic volatility (SV) models has received a great deal of attention in the theoretic finance literature as well as in the empirical finance literature. Prices of options based on SV models are shown to be more accurate than those based on the Black-Scholes model (see, for example, Melino and Turnbull (1990)). Moreover, the SV model offers a powerful alternative to GARCH-type models to explain the well documented time varying volatility. Empirical successes of the lognormal SV model relative to GARCH-type models are documented in Danielsson (1994), Geweke (1994), and Kim, Shephard and Chib (1998) in terms of in-sample fitting, and in Yu (2002) in terms of out-of-sample forecasting.

In the theoretical finance literature on option pricing, the SV model is often formulated in terms of stochastic differential equations. For instance, Wiggins (1987), Chesney and Scott (1989), and Scott (1991) specify the following model for the asset price $P(t)$ and the corresponding volatility $\sigma^2(t)$,

$$dP(t)/P(t) = \alpha dt + \sigma(t) dB_1(t), \tag{1.1}$$

$$d \ln \sigma^2(t) = \lambda(\xi - \ln \sigma^2(t))dt + \gamma dB_2(t), \tag{1.2}$$

where $B_1(t)$ and $B_2(t)$ are two standard Brownian motions.

In the empirical literature, the above continuous time model is often discretized. The discrete time SV model may be obtained, for example, via the Euler-Maruyama approximation. The approximation, after a location shift and reparameterization, leads to the so-called *lognormal* SV model given by

$$X_t = \sigma_t e_t, \tag{1.3}$$

$$\ln \sigma_t^2 = \mu + \phi(\ln \sigma_{t-1}^2 - \mu) + \sigma v_t, \tag{1.4}$$

where $X_t$ is a continuously compounded return and $e_t, v_t$ are two uncorrelated sequences of independent and identically distributed (iid) $N(0,1)$ random variables. The above model is equivalently represented, in the majority of empirical literature, by

$$X_t = \exp(\frac{1}{2} h_t) e_t, \tag{1.5}$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma v_t. \tag{1.6}$$

This most widely used SV model is built upon the models of Clark (1973) and Tauchen and Pitt (1983) and first introduced by Taylor (1982, 1986 and 1994). One implication of its specification is that the marginal distribution of logarithmic volatility is normal. This assumption has very important implications for financial economics and risk management.

Alternative SV models have appeared in the theoretical literature as well as in the empirical literature. For example, Stein and Stein (1991) and Johnson and Shanno (1987) assume $\sigma(t)$ follows, respectively, an Ornstein-Uhlenbeck (OU) process and a geometric Brownian motion, while Hull and White (1987) and Heston (1993) assume a geometric Brownian motion and a square-root process for $\sigma^2(t)$. In the discrete time case, various SV models can be regarded as generalizations to corresponding GARCH models. For example, a polynomial SV model is a generalization of GARCH(1,1) (Bollerslev (1986)) while a square root polynomial SV model is a generalization of standard deviation (SD)-GARCH(1,1). Andersen (1994) introduces a general class of SV models, of which a class of polynomial SV models has been emphasized. This class encompasses most of the discrete time SV models in the literature. Other more recent classes of SV models include those proposed by Barndorff-Nielsen and Shephard (2001b) and by Meddahi (2001).

Despite all these alternative specifications, there is a lack of procedure for selecting appropriate functional form of stochastic volatility.[1] The specification of the correct stochastic volatility function, on the other hand, is very important in several respects. First, different functional forms lead to different formulae for option pricing. Misspecification of the stochastic volatility function can result in incorrect option prices. Second, the marginal distribution of volatility depends upon the function form of stochastic volatility.

In this paper, we propose a new class of SV models, namely, nonlinear SV models. Like the class of Andersen (1994), it includes as special cases many discrete time SV

[1]It is well known that a GARCH process converges to a relevant stochastic volatility process (Nelson (1990)). A specification test based on a GARCH family can be suggestive of an appropriate stochastic volatility specification; see for example, Hentschel (1995). Such a test, however, is by no mean a direct test of stochastic volatility specifications.

models that have appeared in the literature. It overlaps with but does not encompass the class of Andersen. An advantage of our proposed class is the ease with which different specifications on stochastic volatility can be tested. In fact, the specification test is based on a single parameter. Another advantage of our proposed class is that, as a byproduct of this general way of modeling stochastic volatility, one obtains the functional form of transformation which induces marginal normality of volatility. Section 2 presents this class of nonlinear SV models. In Section 3, we use an efficient method of moments (EMM) approach to estimate the proposed class of models. In Section 4, the class is fitted to daily observations on an individual stock return and a stock index return and in Section 5 we present conclusions and possible extensions.

# 2 A Class of Nonlinear SV Models

The lognormal SV model specifies that the logarithmic volatility follows an AR(1) process. However, this relationship may not always be warranted by the data. A natural generalization to this relationship is to allow a general (nonlinear) smooth function of volatility to follow an AR(1) process. That is,

$$X_t = \sigma_t e_t, \tag{2.7}$$

$$h(\sigma_t^2, \delta) = \mu + \phi[h(\sigma_{t-1}^2, \delta) - \mu] + \sigma v_t, \tag{2.8}$$

where $e_t$ and $v_t$ are two uncorrelated $N(0,1)$ sequences, and $h$ is a smooth function indexed by a parameter $\delta$. A nice choice of this function is the Box-Cox power function (Box and Cox (1964)):

$$h(t,\delta) = \begin{cases} (t^\delta - 1)/\delta, & \text{if } \delta \neq 0, \\ \log t, & \text{if } \delta = 0. \end{cases} \tag{2.9}$$

As the function $h$ is specified as a general nonlinear function, the model is thus termed in this paper the *nonlinear* SV model. Several attractive features of this new class of SV models include: i) as we will show below it includes the lognormal SV model and the other "classical" SV models as special cases, ii) it adds great flexibility on the functional form, and iii) it allows a simple test for the lognormal SV specification, i.e., a test of

$H_0 : \delta = 0$, and some other "classical" SV specifications. If we write $h_t = h(\sigma_t^2, \delta)$, then we can re-write the nonlinear SV models as

$$
\begin{aligned}
X_t &= [g(h_t, \delta)]^{1/2} e_t, && (2.10)\\
h_t &= \mu + \phi(h_{t-1} - \mu) + \sigma v_t, && (2.11)
\end{aligned}
$$

where $g(h_t, \delta)$ is the inverse Box-Cox transformation of the form

$$
g(h_t, \delta) = \begin{cases} (1 + \delta h_t)^{1/\delta}, & \text{if } \delta \neq 0, \\ \exp(h_t), & \text{if } \delta = 0. \end{cases} \tag{2.12}
$$

Equivalently we can re-write them in a form of

$$
X_t = \sigma_t e_t, \tag{2.13}
$$

$$
\frac{(\sigma_t^2)^\delta - 1}{\delta} = \mu + \phi[\frac{(\sigma_{t-1}^2)^\delta - 1}{\delta} - \mu] + \sigma v_t. \tag{2.14}
$$

Denote the parameters of interest by $\theta = (\mu, \delta, \phi, \sigma)$.

The idea of our proposed SV models is similar to that made in Higgins and Bera (1992) from the linear ARCH model (Engle (1982)) to the nonlinear ARCH (NARCH) model. Obviously, our model provides a stochastic volatility generalization of a nonlinear GARCH(1,1) model.

It can be seen as $\delta \to 0$, $(1 + \delta h_t)^{1/(2\delta)} \to \exp(0.5h_t)$ and $((\sigma_t^2)^\delta - 1)/\delta \to \ln \sigma_t^2$. Hence the proposed nonlinear SV model includes the lognormal SV model as a special case. If $\delta = 1$, the variance equation (2.14) becomes

$$
\sigma_t^2 = \mu' + \phi(\sigma_{t-1}^2 - \mu') + \sigma v_t, \tag{2.15}
$$

where $\mu' = \mu + 1$. This is a polynomial SV model in Andersen (1994). According to this specification, volatility follows a normal distribution as its marginal distribution. If $\delta = 0.5$, the variance equation (2.14) becomes

$$
\sigma_t = \mu'' + \phi(\sigma_{t-1} - \mu'') + 0.5\sigma v_t, \tag{2.16}
$$

where $\mu'' = 0.5\mu + 1$. This is a square root polynomial SV model in Andersen (1994) and can be regarded as a discrete time version of the continuous time SV model in Scott

(1987) and Stein and Stein (1991). As a result, the marginal distribution of the square root of volatility is Gaussian.

In Table 1 we summarize some well-known stochastic volatility models and show their parameter relations with our model. For the continuous time stochastic volatility models, their Euler discrete time versions are considered. It can be seen that all these models can be obtained from our model by placing the appropriate restrictions on the three parameters $\delta, \mu$ and $\phi$. In fact, all the models except our model require $\delta$ to be 0, 0.5, or 1.[2] For a general $\delta$, our model is different from any of them and $\delta$ provides some idea about the degree of departure from a "classical" parametric SV model. See Figure 1 for the comparison of the square root of inverse Box-Cox transformation, $(1+\delta h_t)^{1/(2\delta)}$, for various values of $\delta$ over the interval $[-2, 2]$, a possible range that actual $h_t$ may lie within in the framework of lognormal SV model.

The Box-Cox transformation has been applied in various areas in finance. One of the most relevant applications to our work may be that proposed by Higgins and Bera (1992) who introduce the nonlinear ARCH model. Another relevant application is Hentschel (1995) who introduces a family of GARCH models by applying the Box-Cox transformation to the conditional standard deviation. A nice feature of our proposed class is that it provides a simple way to test the null hypothesis of polynomial SV specifications, including the lognormal SV specification, against a variety of non-polynomial alternatives. In fact, this specification test is based entirely on a single parameter, $\delta$. Moreover, as a consequence of specification testing, our proposed class provides an effective channel to check the marginal distribution of unobserved volatility. Therefore, our method serves as an alternative approach to studying marginal distribution of daily volatility from that which appeared in a recent literature based on ultra-high frequency data (cf Andersen et al (2001a, b)).

To conclude this section, we establish some basic properties of the proposed class of

[2]Some specifications in Table 1 may be different from the actual specifications used in the original references. However, they are equivalent to each other via Ito's lemma. For example, Heston (1993) adopts a square root specification for $\sigma_t^2$ which is identical to assuming $\sigma_t$ follows a particular OU process.

SV models. It is easy to see that $h_t$ is stationary and ergodic if $\phi < 1$ and that if so

$$\mu_h \equiv E(h_t) = \mu,\ \sigma_h^2 \equiv Var(h_t) = \frac{\sigma^2}{1-\phi}, \text{ and } \rho(\ell) \equiv Corr(h_t, h_{t-\ell}) = \phi^\ell.$$

It follows that $X_t$ is stationary and ergodic as it is the product of two stationary and ergodic processes. For the moments of $X_t$, a distributional constraint has to be imposed on $v_t$ or $h_t$. As $\sigma_t^2$ is nonnegative, the exact normality of $v_t$ is incompatible unless $\delta = 0$ or $1/\delta$ is an even integer.[3] Our experience suggests that, as far as parameter estimation is concerned, the assumption of the exact normality of $v_t$ works well for all the empirically possible values of parameters that we have encountered.[4] However, to derive some theoretical results, we assume in general that $u_t = \sigma_t^2 = (1+\delta h_t)^{1/\delta}$ follows a generalized lognormal distribution as defined in Chen (1995) with pdf

$$f(u_t;\delta,\mu,\sigma_h) = \begin{cases} \sigma_h^{-1}\psi[(u_t(\delta)-\mu)/\sigma_h]u_t^{\delta-1}/\Psi(\theta), & \text{if } \delta > 0, \\ \sigma_h^{-1}\psi[(u_t(\delta)-\mu)/\sigma_h]u_t^{\delta-1}, & \text{if } \delta = 0, \\ \sigma_h^{-1}\psi[(u_t(\delta)-\mu)/\sigma_h]u_t^{\delta-1}/\Psi(-\theta), & \text{if } \delta < 0, \end{cases} \tag{2.17}$$

where $u_t(\delta)$ is the Box-Cox power transformation of $u_t$, and $\psi$ and $\Psi$ are, respectively, the standard normal pdf and cdf with $\theta = (1+\delta\mu)/\delta\sigma_h$. Chen (1995) shows that

$$\begin{aligned} \text{If } \delta = 0, E(u_t^k) &= \exp(k\mu + \frac{1}{2}k^2\sigma_h^2), k = 1, 2, \cdots, \\ \text{If } \delta > 0, E(u_t^k) &= \frac{1}{\sqrt{2\pi}}\int_\delta^\infty \exp(-\frac{1}{2}v^2)[1+\delta(\mu+\sigma_h v)]^{k/\delta}dv < \infty, \text{for } k = 1, 2, \cdots, \\ \text{If } \delta < 0, E(u_t^k) &= \frac{1}{\sqrt{2\pi}}\int_\infty^{-\delta} \exp(-\frac{1}{2}v^2)[1+\delta(\mu+\sigma_h v)]^{k/\delta}dv < \infty, \text{iff } \delta < -k. \end{aligned}$$

Combining Chen's results with Gaussianity of $e_t$ and independence between $e_t$ and $v_t$, moments of $X_t$ can easily be found. In particular, all the odd moments are zero and the even moments are

$$E(X_t^k) = E(u_t^{k/2})E(e_t^k) = \frac{k!}{2^{k/2}(k/2)!}E(u_t^{k/2}), k = 2, 4, \cdots.$$

[3] This is the well known truncation problem with the Box-Cox power transformation. The truncation effect is negligible if $\delta\sigma_h/(1+\delta\mu)$ is small, which is achieved when i) $\delta$ is small, or ii) $\mu$ is large, or iii) $\sigma_h$ is small. See Yang (1999) for a discussion on this. Furthermore, a small value of $\delta\sigma_h/(1+\delta\mu)$ can be achieved by re-scaling $X_t$. For example, one can denote a return by $(\ln P_t - \ln P_{t-1}) \times 10$ rather than by $(\ln P_t - \ln P_{t-1}) \times 100$.

[4] The same problem occurs in the model proposed by Stein and Stein (1991). They claim that "for a wide range of empirically reasonable parameter values, the probability of passing the barrier at $\sigma = 0$ is so small as to be of no significant consequence."

As a result, the expression for kurtosis is easily derived. Of particular interest are cases where $1/\delta$ is a positive, even integer, which give rise to some polynomial SV models of Anderson (1994). In these cases, no truncation is necessary and exact normality assumption can be given to the distribution of $v_t$. Under Gaussianity of $v_t$, we have $h_t \sim N(\mu, \sigma_h^2)$, and hence $1 + \delta h_t \sim N(1 + \delta\mu, \delta^2\sigma_h^2)$. The recursive formulas of Katz (1999) can be used for finding the moments of $u_t$ as well as the product moments of $u_t$ and $u_{t-\ell}$, which can then be converted to give moments and product moments for the $X_t$ process. Define

$$\gamma(i, j) = E(Z_1^i Z_2^j), i, j = 0, 1, \cdots ,$$

where $Z_1$ and $Z_2$ are two normal random variables with means $\mu_1$ and $\mu_2$, standard deviations $\sigma_1$ and $\sigma_2$, and the correlation coefficient between them $\rho$. The recursive formulae take the form

$$\begin{aligned}
\gamma(i, 0) &= \mu_1\gamma(i-1, 0) + (i-1)\sigma_1^2\gamma(i-2, 0) \\
\gamma(0, j) &= \mu_2\gamma(0, j-1) + (j-1)\sigma_2^2\gamma(0, j-2) \\
\gamma(1, j) &= \mu_2\gamma(1, j-1) + \sigma_1\sigma_2\rho\gamma(0, j-1) + (j-1)\sigma_2^2\gamma(1, j-2) \\
\gamma(i, j) &= \mu_1\gamma(i-1, j) + j\sigma_1\sigma_2\rho\gamma(i-1, j-1) + (i-1)\sigma_1^2\gamma(i-2, j),
\end{aligned}$$

where $i, j = 2, 3, \cdots$. We now consider several special SV models to give explicit expressions for the moments and autocorrelation functions.

**The square root polynomial SV model.** When $\delta = 0.5$, we have the square root polynomial SV model and $u_t = (1 + 0.5h_t)^2$. Let $Z_1 = 1 + 0.5h_t$ and $Z_2 = 1 + 0.5h_{t-\ell}$. Then, we have $\mu_1 = \mu_2 = 1 + \mu/2$, $\sigma_1^2 = \sigma_2^2 = \sigma_h^2/4$ and $\rho = \phi^\ell$. The odd moments of $X_t$ are zero. The even moments can be found from the expression of $\gamma(i, 0)$. In particular,

$$E(X_t^2) = (1 + \mu/2)^2 + \sigma_h^2/4, \text{ and } E(X_t^4) = 3[(1 + \mu/2)^4 + 3(1 + \mu/2)^2\sigma_h^2/2 + 3\sigma_h^4/16].$$

Further, the autocorrelation function of $X_t^2$ is found using expressions of $\gamma(4, 0)$ and $\gamma(2, 2)$, which is

$$\frac{(1 + \mu/2)^4 + (1 + \mu/2)^2\sigma_h^2\phi^\ell + \sigma_h^4\phi^{2\ell}/8 + \sigma_h^2((1 + \mu/2)^2 + \sigma_h^2/4)}{3[(1 + \mu/2)^4 + 3(1 + \mu/2)^2\sigma_h^2/2 + 3\sigma_h^4/16]}.$$

**The lognormal SV model.** The moments and dynamic properties of other polynomial SV models with $1/\delta$ an even integer can be found in a similar way (though more tedious) to the above. It should be pointed out that the popular lognormal SV model can also be considered as a special case of the above model with $1/\delta$ being a very large, positive, even integer. Its moment and dynamic properties can be found in Shephard (1996) and Knight, Satchell and Yu (2001).

# 3 Estimation by Efficient Method of Moments

The literature on estimating SV models is vast. This is in part due to the fact that the likelihood function has no closed form expression for the SV model and hence the maximum likelihood approach is extremely difficult to implement. As a consequence, the SV model becomes a central example to compare the relative merits of alternative estimation procedures.

To estimate the discrete time SV model, Melino and Turnbull (1990) propose generalized method of moments (GMM) which is further improved by Andersen and Sorensen (1996). For the continuous time SV model, a GMM approach is developed by Hansen and Scheinkman (1995). The idea behind GMM is to match a number of sample moments with model moments. Harvey, Ruiz and Shephard (1994) and Ruiz (1994) suggest the quasi maximum likelihood (QML) approach. The main idea is to approximate non-normal disturbances by normal disturbances and then maximize the Gaussian likelihood function. Observing that the joint and conditional characteristic functions of the SV model have closed form expression, Yu (1998), Knight et al (2001) propose to estimate the discrete time SV model via the empirical characteristic function, while Singleton (2001) and Jiang and Knight (2002) use the empirical characteristic function method to estimate the continuous time SV model. More efficient estimation methods involve the whole family of simulation based methods. These include the simulated maximum likelihood method proposed by Danielson and Richard (1993) and Danielsson (1994); the Markov Chain Monte Carlo (MCMC) method proposed by Jacquier, Polson and Rossi (1994) and improved by Kim et al (1998); the maximum likelihood Monte Carlo method (Sandmann and Koopman (1998)); the simulation method using important sampling and antithetic variables proposed by Durbin and Koopman (2000); and the efficient method of moments (EMM) procedure by Gallant and Tauchen (1996).

The relative merits of the alternative methods depend not only on the finite sample efficiency but also on the flexibility to adapt to modifications of model specification. Moreover, in the framework of SV models, a good method should also allow one to extract the unobserved volatility model with a low cost and to do simple but useful model diagnostics. Judged by these criteria, EMM is our choice for inferences since it provides a flexible and reasonably efficient approach to analyzing the SV model. The asymptotic efficiency of EMM is provided in Gallant and Tauchen (1996) for Markov processes, and Gallant and Long (1997) and Tauchen (1997) for non-Markov processes. Andersen, Chung and Sorensen (1999) document a finite sample comparison of various methods for estimating the lognormal SV model in Monte Carlo studies and find that the EMM procedure performs quite well in comparison with other estimation procedures. Gallant, Hsieh, and Tauchen (1997) and Gallant and Tauchen (2001c) discuss flexibility of modeling modifications of the SV model. Using a nonlinear Kalman filtering technique, Gallant and Tauchen (1998) propose a reprojection method to infer the unobserved state vector. One advantage of the EMM approach lies in its diagnostics. For example, it allows for a model diagnostic suggestive of the dimension along which the model may be inadequate and provides simple overall model specification checking (cf Gallant and Tauchen (1996) and Tauchen (1997)).

EMM is first introduced by Gallant and Tauchen (1996) and has now found many successful applications in economics and finance; see Gallant and Tauchen (2001a) for a brief review of the literature. It is closely related to GMM of Hansen (1982). An important difference between them is that while GMM relies on an *ad hoc* chosen set of moment conditions, EMM is based on a judiciously chosen set of moment conditions. The moment conditions EMM employs is the expectation of the score of an auxiliary model which is often referred to as the score generator.

Let the SV model of interest be the structural model. The conditional density of the structural model is defined by

$$p_t(x_t|y_t, \theta),$$

where the true value of $\theta$ is $\theta_0$ and $\theta_0 \in \Theta \subset \Re^{\ell_\theta}$ with $\ell_\theta$ being the length of $\theta_0$. Denote the conditional density of an auxiliary model by

$$f_t(x_t|y_t, \beta), \beta \in R \subset \Re^{\ell_\beta}$$

where $y_t$ is a vector of lagged $x_t$. Further define the expected score of the auxiliary model under the structural model as

$$m(\theta, \beta) = \int \cdots \int \frac{\partial}{\partial \beta} \ln f(x|y, \beta) p(x|y, \theta) p(y|\theta) dx dy.$$

Obviously, in the context of the SV model, the integration cannot be solved analytically since neither $p(x|y, \theta)$ nor $p(y|\theta)$ has closed form. However, it is easy to simulate from an SV model so that one can approximate the integral by Monte Carlo simulations. That is

$$m(\theta, \beta) \approx m_N(\theta, \beta) \equiv \frac{1}{N} \sum_{\tau=1}^{N} \frac{\partial}{\partial \beta} \ln f(\hat{x}_\tau(\theta)|\hat{y}_\tau(\theta), \beta),$$

where $\{\hat{x}_\tau, \hat{y}_\tau\}$ are simulated from the structural model. The EMM estimator is a minimum chi-squared estimator which minimizes the following quadratic form,

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} m'_N(\theta, \hat{\beta}_n)(I_n)^{-1} m_N(\theta, \hat{\beta}_n),$$

where $\hat{\beta}_n$ is a quasi maximum likelihood estimator of the auxiliary model and $I_n$ is an estimate of

$$I_0 = \lim_{n \to \infty} Var\left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ \frac{\partial}{\partial \beta} \ln f_t(x_t|y_t, \beta^*) \} \right)$$

with $\beta^*$ being the pseudo true value of $\beta$. Under regularity conditions, Gallant and Tauchen (1996) show that the EMM estimator is consistent and has the following asymptotic normal distribution,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \frac{\partial}{\partial \theta} m(\theta_0, \beta^*)(I_0)^{-1} \frac{\partial}{\partial \theta'} m(\theta_0, \beta^*)).$$

For specification testing, we have

$$J_n = n m'_N(\hat{\theta}_n, \hat{\beta}_n)(I_n)^{-1} m_N(\hat{\theta}_n, \hat{\beta}_n) \xrightarrow{d} \chi^2_{\ell_\beta - \ell_\theta}$$

under the null hypothesis that the structural model is correct. When a model fails the above specification test one may wish to examine the quasi-t-ratios and/or t-ratios to look for some suggestion as to what is wrong with the structural model. The quasi-t-ratios are defined as

$$\hat{T}_n = S_n^{-1}\sqrt{n}m_N(\hat{\theta}_n, \hat{\beta}_n)$$

where $S_n = [diag(I_n)]^{1/2}$. It is well known that the elements of $\hat{T}_n$ are downward biased in absolute value. To correct the bias one can use the t-ratios defined by

$$\tilde{T}_n = Q_n^{-1}\sqrt{n}m_N(\hat{\theta}_n, \hat{\beta}_n)$$

where

$$Q_n = \left( diag\{I_n - \frac{\partial}{\partial\theta'}m_N(\hat{\theta}_n, \hat{\beta}_n)[m_N'(\hat{\theta}_n, \hat{\beta}_n)(I_n)^{-1}m_N(\hat{\theta}_n, \hat{\beta}_n)]^{-1}\frac{\partial}{\partial\theta}m_N(\hat{\theta}_n, \hat{\beta}_n)\} \right)^{1/2}.$$

Large quasi-t-ratios and t-ratios reveal the features of the data that the structural model cannot approximate.

Furthermore, Gallant and Tauchen (1996) show that if the auxiliary model nests the data generating process, under regularity conditions the EMM estimator has the same asymptotic variance as the maximum likelihood estimator and hence is fully efficient. If the auxiliary model can closely approximate the data generating process, the EMM estimator is nearly fully efficient (Gallant and Long (1997) and Tauchen (1997)).

To choose an auxiliary model, the seminonparametric (SNP) density proposed by Gallant and Tauchen (1989) can be used since its success has been documented in many applications. As to SNP modeling, six out of eight tuning parameters are to be selected, namely, $L_u$, $L_g$, $L_r$, $L_p$, $K_z$, and $K_y$. The other two parameters, $I_z$ and $I_x$, are irrelevant for univariate time series and hence set to be 0. $L_u$ determines the location transformation whereas $L_g$ and $L_r$ determine the scale transformation. Altogether they determine the nature of the leading term of the Hermite expansion. The other two parameters $K_z$ and $K_y$ determine the nature of the innovation. To search for a good auxiliary model, one can use the Schwarz BIC criterion to move along an upward expansion path until an adequate model is found, as outlined in Bansal et al (1995). To preserve space we refer readers to Gallant and Tauchen (2001b) for further discussion about the role of the tuning parameters and how to design an expansion path to choose them.

# 4 Empirical Applications

In this section we consider two applications using an individual stock price series and a stock index series. The stock price data consist of 3,778 observations on the daily price of a share of Microsoft, adjusted for stock split, for the period from March 13, 1986 to February 23, 2001. The same data have been used in Gallant and Tauchen (2001a) to fit a continuous time SV model. The stock index data consist of 4044 observations on 100 times the log-first difference of the daily S&P 500 index for the period from January 4, 1977 to December 31, 1992. The same data have been used in Gallant and Tauchen (2001a) to fit the SV model of Clark (1973). Let $P_t$ represent the stock price of Microsoft at period $t$. Define the daily return $X_t$ as $(\ln P_t - \ln P_{t-1}) \times 100$. There are 3,777 observations for Microsoft return data.

Figure 2 displays the two return series and Table 2 reports some descriptive statistics for them. From Table 2, it can be seen that, from maximums, minimums and variances, the Microsoft returns are more volatile than the S&P500 returns. However, the distribution of Microsoft returns is less lepkurtotic than that of S&P500 returns.

Neither return series is mean-adjusted. To allow for a possible no-zero mean and also some dynamics in mean, we introduce an AR(1) structure in the mean equation. As a consequence, we fit the following two models to each of the return series:

$$X_t = \mu_0 + c(X_{t-1} - \mu_0) + \exp(\frac{1}{2}h_t)e_t, \tag{4.18}$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma v_t; \tag{4.19}$$

and

$$X_t = \mu_0 + c(X_{t-1} - \mu_0) + (1 + \delta h_t)^{1/(2\delta)} e_t, \tag{4.20}$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma v_t. \tag{4.21}$$

We call them the lognormal SV model and the proposed SV model respectively.

The same sets of tuning parameters in the SNP model are employed as in Gallant and Tauchen (2001a), since identical datasets are used. We report these tuning parameters in the following order

$$(L_u, L_g, L_r, L_p, K_z, I_z, K_y, I_y).$$

To ensure that the chosen SNP model is reasonable, we have compared the BIC value with those from many alternative sets of tuning parameters and find that the BIC value from the chosen SNP model is one of the smallest. The set of the tuning parameters, the corresponding BIC value, the leading term in the Hermite expansion, the characterization of $X_t$ and the number of parameters in the auxiliary model are presented in Table 3 for both series. A GARCH leading term is used for the Microsoft returns whereas an ARCH leading term is used for the S&P500 returns.

Since the sample sizes for both series are large, we believe that the choice of leading term is not crucial as long as a form of conditional heteroskedasticity has been accommodated. In a Monte Carlo study, Andersen et al (1999) find that the EMM efficiency approaches that of maximum likelihood for larger sample size when various forms of conditional heteroskedasticity are used as the leading term. Moreover, they find that the EMM-based inferences, such as the $t$-statistic and $J_n$ statistic, are robust to the choice of auxiliary model when the sample size is large.

Table 4 and Table 5 report the empirical results for Microsoft returns. To ensure a global minimum is obtained, we perturb starting values when minimizing quadratic expression and estimating SNP density. Furthermore, we simulate 100,000 observations from the SV models, of which first 10,000 observations are discarded in order to let the effect from initialization die off. Table 4 reports the estimates, the numerical Wald standard errors, the 95% approximate criterion-difference confidence intervals, the value of statistic $J_n$, and the degrees of freedom and the $p$-value of $J_n$ for the lognormal SV model and for the proposed SV model. Table 5 reports the quasi-t-ratios and t-ratios from the score generator for both models.

A few results emerge from these two tables. First, the point estimate of $\delta$ in the proposed SV model is $-0.0526$ which is insignificantly different from 0 but significantly less than 0.5 and 1. Consequently, one cannot reject the lognormal SV model but can comfortably reject the other polynomial SV specifications, including the Stein-Stein and Heston specifications. The marginal distributions of volatility implied from the estimated lognormal and proposed models are plotted in Figure 3. It appears that they are quite close to each other. Second, the point estimate of $\phi$ (0.9476) is close to 1 and in the stationary region when the lognormal model is fitted. In the proposed SV model, it decreases to 0.7260 which is significantly less than 1 and greater than 0. As a result, one has to reject the Hull-White and Clark specifications. Third, although our specification test cannot reject the lognormal SV model, the minimum $\chi^2$ criterion provides some evidence against the lognormal specification. One can reject it at the one percent level. This evidence is further reinforced by the diagnostic quasi-t-ratios and t-ratios. There are large t-ratios on the scores corresponding to the polynomial part of the SNP score when the lognormal model is fitted. These t-statistics indicate that $\exp(0.5h_t)$ may not be the correct transformation. When the proposed SV model is fitted, although this specification is not statistically significantly different from the lognormal specification, the minimum $\chi^2$ criterion is quite encouraging. One can accept the proposed model at the 5 percent level. We are not, of course, suggesting the proposed model is completely satisfactory. In fact, one should note that the t-ratios are not entirely clean. However, if we compare the t-ratios with those from the lognormal model, our model is overall superior. For example, although there are large t-ratios on the scores corresponding to the ARCH part of the SNP score in the proposed model, these compare with large t-ratios on the scores corresponding to both the ARCH part and the polynomial part of the SNP score in the lognormal model. Finally, $\delta$ seems to be more difficult to estimate than other parameters with the Wald standard error being the largest.

Table 6 and Table 7 report the empirical results for S&P500 returns. As for Microsoft returns, we perturb starting values when doing the optimizations. Furthermore, we simulate 101,000 observations from the SV models, of which first 1,000 observations are discarded in order to let transients die out. Table 6 reports the estimates, the numerical Wald standard errors, the 95% approximate criterion-difference confidence intervals, the value of statistic $J_n$, and the degrees of freedom and the p-value of $J_n$ for the lognormal SV model and for the proposed SV model. Table 7 reports the quasi-t-ratios and t-ratios from the score generator for both models.

A few results emerge from these two tables. First and most interestingly, the point estimate of $\delta$ is $-0.4597$. It is significantly less than 0 and hence significantly less than 0.5 and 1, although its Wald standard error remains the largest. As a consequence, one has to reject the lognormal SV model and all the other SV models in Table 1. Observing that $\delta$ is not significantly different from $-0.5$, to gain some idea about our estimated results, we approximate $\delta \approx -0.5$, plug the estimates into Equation (2.14) and get the following estimated variance equation:

$$\frac{1}{\sigma_t} = 1.2237 + 0.9840(\frac{1}{\sigma_{t-1}} - 1.2237) + 0.057v_t.$$

This compares to the estimated variance equation in the lognormal model,

$$\ln \sigma_t^2 = -0.3425 + 0.9846(\ln \sigma_{t-1}^2 + 0.3425) + 0.1022v_t.$$

The marginal distributions of volatility implied from these two fitted models are plotted in Figure 3. It is evident that these two distributions are not close to each other and hence the lognormal distribution is not a good approximation to the marginal distribution of volatility. Furthermore, as argued in Section 2, in theory, a distributional constraint has to be imposed for general $\delta$ in the proposed SV models to ensure non-negativeness of $\sigma_t$. In the empirical applications, however, we still adopt the assumption of exact normality. To understand how restricted this assumption is, we calculate Prob($\sigma_t < 0$)=Prob($1/\sigma_t < 0$)=0.0000065 which is a very small value.

Second, the point estimate of $\phi$ (0.984) is close to 1 and just in the stationary region. In the proposed SV model, it remains at a similar level. In fact all the estimated parameters have similar magnitude and similar standard errors across both models. The only exceptions are $\mu$ which decreases from 1.4229 to 1.1781 and $\sigma$ which increases from 0.1022 to 0.1140. This is because $\mu$ and $\sigma$ are closely related to $\delta$ in the proposed model. Since the estimated $\delta$ is far away from 0 in the proposed model, this translates to large discrepancies between the estimated $\mu$'s and the estimated $\sigma$'s. Third, the minimum $\chi^2$ criterion provides mild evidence against the lognormal specification. It is rejected at the 5 percent level but accepted at the 1 percent level. The evidence is consistent with the diagnostic quasi-t-ratios and t-ratios. There are large quasi-t-ratios and t-ratios on the scores corresponding to the polynomial part of the SNP score. These t-statistics indicate that $\exp(0.5h_t)$ may not be the correct transformation. When the proposed SV model is fitted, the $p$-value of $J_n$ statistic increases by about 80%. One can accept the proposed model at the 5% percent level. Furthermore, all the quasi-t-ratios become insignificant in the proposed model. Although some of the t-ratios on the scores corresponding to the polynomial part of the SNP score are still too large, they are clearly smaller than those in the lognormal model.

We can briefly summarize the empirical results. Although the EMM diagnostics suggest that the lognormal SV model cannot adequately capture some features in the Microsoft returns, the specification test based on the proposed nonlinear SV model indicates that the nonlinear specification is not significantly different from the lognormal specification. Therefore, the logarithmic volatility can be well approximated by the normal distribution as its marginal distribution. For S&P500 returns, although the EMM diagnostics only provide mild evidence against the lognormal SV model, the specification test based on the proposed nonlinear SV model rejects it and all the other SV specifications. As a consequence, the logarithmic volatility does not follow the normal distribution as its marginal distribution. These results compare interestingly with those reached in Andersen et al (2001a, b), where ultra-high frequency data are used to calibrate daily volatility via realized volatility and the distribution of daily volatility is then nonparametrically estimated. Andersen et al (2001b) have justified the approach by showing that the realized volatility constructed from ultra-high frequency data converges to daily volatility as the sampling frequency goes to infinity. Despite this appealing theoretical property, more recent studies suggest that the approach via ultra-high frequency data has to be used with caution. For example, Barndorff-Nielsen and Shephard (2001a) find that realized volatility can be a noise estimator of daily volatility even when the sampling frequency is reasonably high. The properties of realized volatility as an estimator of daily volatility are further complicated by microstructure problems in transaction data; see the interesting work of Bai, Russell and Tiao (2000) and Andreou and Ghysel (2001) in this context. Based entirely on the daily observations, not surprisingly, our approach is not subject to these criticisms.

The empirical inadequacy of lognormal SV specification for financial time series has been reported in many other works. Examples include Andersen and Lund (1997a) and Gallant and Tauchen (2001c). To improve the overall specification, many extensions have been suggested. Most of these extensions are based on the introduction of a third factor into the structural model. For example, Andersen and Lund (1997b) suggest the third factor should be associated with the mean level while Gallant and Tauchen (2001c) make use of another volatility factor. As an alternative way to extend the lognormal SV specification, our proposed SV model stays within the two-factor family and hence is conceptually simpler than three-factor models. Furthermore, the EMM-based diagnostics indicate that our extension is quite encouraging for the data sets that we have considered.

# 5 Conclusions and Extensions

In this paper a class of nonlinear stochastic volatility models has been proposed. The new class is based on the Box-Cox power transformation and encompasses many parametric stochastic volatility models which have appeared in the literature, including the well known lognormal stochastic volatility model. There are two advantages of our proposed class. First, it facilitates a specification test of "classical" SV models. Second, the functional form of transformation, which induces marginal normality of volatility, is obtained. The EMM approach of Gallant and Tauchen (1996) is used to estimate model parameters. Empirical applications are performed using an individual stock return series and an index return series. Empirical results show that the lognormal SV model is not rejected for the stock returns but it has to be rejected for the index returns. In fact, our result suggests that all the polynomial stochastic volatility models previously used in the literature are rejected for the index returns. As a result, the daily logarithmic stock volatility is well described by a normal distribution as its marginal distribution, consistent with the results found in a recent literature (Andersen et al (2001a)). However, the daily logarithmic index volatility does not follow the normal distribution as its marginal distribution.

One of the important questions in finance is how a superior specification can lead to more accurate option prices. To address this question in the context of nonlinear SV models, one has to first estimate the continuous time SV models. The approach suggested in this paper can be applied. With estimated results, comparison of option prices based on the nonlinear SV models and the "classical" SV models would be of considerable interest. There are some other possible extensions to our work. One possibility is to allow the so-called leverage effect in the nonlinear SV model. This can be done by introducing a negative correlation between two disturbances (cf Harvey and Shephard (1996) and Meyer and Yu (2000)). Another interesting extension would be to incorporate jumps into the model; see for example Barndorff-Nielsen and Shephard (2001b). Finally, it would be interesting to evaluate the out-of-sample forecasting performances of the nonlinear SV models relative to other models.

# References


[1] Andersen, T. (1994). Stochastic autoregressive volatility: A framework for volatility modeling. *Mathematical Finance* 4, 75–102.

[2] Andersen, T., T. Bollerslev, F.X. Diebold, and H. Ebens (2001a). The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76.

[3] Andersen, T., T. Bollerslev, F.X. Diebold, and P. Labys (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Asscociation* 96, 42–55.

[4] Andersen, T., H. Chung, and B. Sorensen (1999). Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Econometrics* 91, 61–87.

[5] Andersen, T. and J. Lund (1997a). Estimating continuous time stochastic volatility models of the short-term interest rate. *Journal of Econometrics* 77, 343–377.

[6] Andersen, T. and J. Lund (1997b). Stochastic volatility and mean drift in the short term interest rate diffusion: Implications for the yield curve. Manuscript, Northwestern University.

[7] Andersen, T. and B. Sorensen (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics* 14, 329–352.

[8] Andreous, E. and E. Ghysel (2001). Rolling-sampling volatility estimators: Some new theoretical, simulation and empirical results. Manuscript, Department of Economics, University of North Carolina

[9] Bai, X., J.F. Russell and G.C. Tiao (2000). Beyond Merton's utopia: Effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data. Manuscript, University of Chicago

[10] Bansal, R., A.R. Gallant, R. Hussey and G. Tauchen (1995) Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics* 66, 251–287.

[11] Barndorff-Nielsen, N. and N. Shephard (2001a). Realized power variation and stochastic volatility models. Manuscript, Nuffield College

[12] Barndorff-Nielsen, N. and N. Shephard (2001b). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B* 63, 167–241.

[13] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

[14] Box, G.E.P. and D.R. Cox (1964). An analysis of transformations. *Journal of Royal Statistical Society B* 26, 211–243.

[15] Chen, G. (1995). Generalized lognormal distribution with reliability application. *Computational Statistics and Data Analysis* 19, 309–319.

[16] Chesney, M. and L.O. Scott (1989). Pring European currency options: A comparison of the modified Black-Scholes model and a random variance model. *Journal of Financial and Quantitative Analysis* 24, 267–284.

[17] Clark, P.K. (1973). A subordinated stochastic process model with finite variance for speculative price. *Econometrica* 41, 135–155.

[18] Danielsson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *Journal of Econometrics* 64, 375–400.

[19] Danielsson, J. and J.F. Richard (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics* 8, s153–s173.

[20] Durbin, J. and S.J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *Journal of the Royal Statistical Society Series B* 62, 3–56.

[21] Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

[22] Gallant, A.R., D. Hsieh and G. Tauchen (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics* 81, 159-192.

[23] Gallant, A.R. and L.R. Long (1997). Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika* 84, 125-141.

[24] Gallant, A.R. and G. Tauchen (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica* 57, 1091-1120.

[25] Gallant, A.R. and G. Tauchen (1996). Which moments to match? *Econometric Theory* 12, 657-681.

[26] Gallant, A.R. and G. Tauchen (1998). Reprojecting partially observed systems with application to interest rate diffusions. *Journal of the American Statistical Association* 93, 10-24.

[27] Gallant, A.R. and G. Tauchen (2001a). EMM: A program for efficient method of moments estimation. User's Guide, Department of Economics, University of North Carolina.

[28] Gallant, A.R. and G. Tauchen (2001b). SNP: A program for nonparametric time series analysis. User's Guide, Department of Economics, University of North Carolina.

[29] Gallant, A.R. and G. Tauchen (2001c). Efficient method of moments. Working paper, Department of Economics, University of North Carolina.

[30] Geweke, J. (1994). Bayesian comparison of econometric models. Working Paper, Federal Reserve Bank of Minneapolis, Minnesota.

[31] Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

[32] Hansen, L.P. and J.A. Scheinkman (1995). Back to the future: Generating moment implications for continuous-time Markov processes. *Econometrica* 63, 767–804.

[33] Harvey, A.C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies* 61, 247–264.

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

[35] Hentschel, L. (1995). All in the family: Nesting symmetric and asymmetric GARCH models. *Journal Financial Economics* 39, 71-104.

[36] Heston, S.L. (1993). A closed-form solution for options with stochastic volatility, with application to bond and currency options. *Review of Financial Studies* 6, 327–343.

[37] Higgins, M.L. and A. Bera (1992). A class of nonlinear ARCH models. *International Economic Review* 33, 137–158.

[38] Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42, 281–300.

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

[40] Jiang, G. and J. L. Knight. (2002). Estimation of continuous time processes via the empirical characteristic function. *Journal of Business and Economic Statistics* 12, 198–212.

[41] Johnson, H. and D. Shanno (1987). Option pricing when the variance is changing. *Journal of Financial and Quantitative Analysis* 22, 143–152.

[42] Katz, R.W. (1999). Moments of power transformed time series. *Environmetrics* 10, 301–307.

[43] Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.

[44] Knight, J.L., S.S. Satchell and J. Yu (2001). Estimation of stochastic volatility model by the empirical characteristic function method. *Australian and New Zealand Journal of Statistics*, forthcoming.

[45] Meddahi, N. (2001). An eigenfunction approach for volatility modeling. Manuscript, Department of Economics, University of Montreal.

[46] Melino, A. and S.M. Turnbull (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 239–265.

[47] Meyer, R. and J. Yu (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal*, 3, 198–215.

[48] Nelson, D. (1990). ARCH models as diffusion approximations. *Journal of Econometrics* 45, 7–38.

[49] Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics.* 63, 289–306.

[50] Sandmann, G. and S.J. Koopman (1998). Maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* 63, 289–306.

[51] Scott, L.O. (1987). Option pricing when the variance changes randomly: Theory, estimation and an application. *Journal of Financial and Quantitative Analysis* 22, 419–439.

[52] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In Cox, D. R., Hinkley, D. V. and Barndorff-Nielson, O. E. (eds), *Time Series Models in Econometrics, Finance and Other Fields,* pp. 1-67. London: Chapman & Hall.

[53] Singleton, K.J. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* 102, 111–141.

[54] Stein, E.M. and J.C. Stein (1991). Stock price distributions with stochastic volatility: An analytical approach. *Review of Financial Studies* 4, 727–752.

[55] Tauchen, G. (1997). New minimum chi-square methods in empirical finance. In Kreps, D. and K.F. Wallis, editors, *Advances in Econometrics, Seventh World Congress*. Cambridge, UK: Cambridge University Press.

[56] Tauchen, G. and M. Pitts (1983). The price variability-volume relationship on speculative markets. *Econometrica* 51, 485–505.

[57] Taylor, S.J (1982). Financial returns modelled by the product of two stochastic processes — a study of the daily sugar prices 1961-75. In Anderson, O.D., Editor, *Time Series Analysis: Theory and Practice, 1*, 203–226. North-Holland, Amsterdam.

[58] Taylor, S.J (1986). *Modelling Financial Time Series*. Chichester, Wiley

[59] Taylor, S.J (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance* 4, 183–204.

[60] Wiggins, J.B. (1987) Option values under stochastic volatility: Theory and empirical estimate. *Journal of Financial Economics* 19, 351–372.

[61] Yang, Z. (1999) Estimating a transformation and its effect on Box-Cox T-ratio. *Test* 8, 167-190.

[62] Yu, J. (1998) *The Empirical Characteristic Function in Time-Series Estimation*. PhD thesis, The University of Western Ontario.

[63] Yu, J. (2002) Forecasting volatility in the New Zealand stock market. *Applied Financial Economics* 12, 193–202.

Table 1: Alternative Stochastic Volatility Models and Parameter Relationship

| | Models | $\delta$ | $\mu$ | $\phi$ |
|---|---|---|---|---|
| Wiggins (1987)<br>Scott (1987)<br>Chesney and Scott (1989)<br>Taylor (1994)<br>Jacquier et al (1994)<br>Kim et al (1998) | $\ln \sigma_t^2 = \mu + \phi(\ln \sigma_{t-1}^2 - \mu) + \sigma v_t$ | 0 | | |
| Scott (1987)<br>Stein and Stein (1991)<br>Andersen (1994) | $\sigma_t = \mu + \phi(\sigma_{t-1} - \mu) + \sigma v_t$ | 0.5 | | |
| Heston(1993) | $\sigma_t = \phi \sigma_{t-1} + \sigma v_t$ | 0.5 | 0 | |
| Hull and White (1987)<br>Johnson and Shanno (1987) | $\ln \sigma_t^2 = \mu + \ln \sigma_{t-1}^2 + \sigma v_t$ | 0 | | 1 |
| Andersen(1994) | $\sigma_t^2 = \mu + \phi(\sigma_{t-1}^2 - \mu) + \sigma v_t$ | 1 | | |
| Clark (1973) | $\ln \sigma_t^2 = \mu + \sigma v_t$ | 0 | | 0 |
| Nonlinear SV | $\frac{(\sigma_t^2)^\delta - 1}{\delta} = \mu + \phi[\frac{(\sigma_{t-1}^2)^\delta - 1}{\delta} - \mu] + \sigma v_t$ | | | |

Table 2: Summary statistics for S&P500 Returns and Microsoft Returns

| | SP500 | Microsoft |
|---|---|---|
| Sample Size | 4044 | 3777 |
| Mean | 0.03472 | 0.1503 |
| Variance | 0.9799 | 6.4579 |
| Excess Kurtosis | 40.154 | 14.999 |
| Maximum | 8.2470 | 17.869 |
| Minimum | $-19.3488$ | $-35.828$ |

Note: The S&P 500 index return series is for the period from January 4, 1977 to December 31, 1992. The Microsoft return series is for the period from March 13, 1986 to February 23, 2001.

Table 3: Tuning parameters for SNP modeling, BIC, Leading Term, Characterization of $X_t$ and the Number of Parameters in the Auxiliary Model

| Data | Tuning parameters | BIC | Leading term | Characterization of $X_t$ | $\ell_\beta$ |
|---|---|---|---|---|---|
| Microsoft | $(1,1,1,1,6,0,0,0)$ | 1.32087 | GARCH | Semiparametric GARCH | 11 |
| SP500 | $(2,0,11,1,4,0,0,0)$ | 1.32715 | ARCH | Semiparametric ARCH | 19 |

Table 4: Parameter estimates, standard errors, confidence intervals, $\chi^2$ criterion for Microsoft

| | Lognormal SV model | Proposed SV model |
|---|---|---|
| $\mu_0$ | 0.1683<br>(0.0313)<br>$[0.1080, 0.2315]$ | 0.1821<br>(0.0320)<br>$[0.1811, 0.1954]$ |
| $c$ | 0.0278<br>(0.0174)<br>$[-0.0066, 0.0622]$ | 0.0249<br>(0.0174)<br>$[0.0092, 0.0366]$ |
| $\mu$ | 1.4229<br>(0.0824)<br>$[1.2581, 1.5730]$ | 1.1781<br>(0.1418)<br>$[1.0204, 1.2752]$ |
| $\phi$ | 0.9476<br>(0.0121)<br>$[0.9053, 0.9625]$ | 0.7260<br>(0.0177)<br>$[0.7084, 0.7372]$ |
| $\sigma$ | 0.1989<br>(0.0121)<br>$[0.1848, 0.2298]$ | 0.4265<br>(0.1081)<br>$[0.3044, 0.5099]$ |
| $\delta$ | NA | $-0.0526$<br>(0.1629)<br>$[-0.2351, 0.0887]$ |
| $\chi^2$ | 19.91 | 11.09 |
| df | 6 | 5 |
| $p-$ value | 0.0029 | 0.050 |

Note: The number in parentheses is the standard error. The number in brackets is the confidence interval. The results are based on 100,000 runs with the first 10,000 runs discarded. The lognormal SV model is defined by Equations (4.18) and (4.19). The proposed SV model is defined by Equations (4.20) and (4.21).

Table 5: Quasi-t-ratios and t-ratios for Microsoft

| | | Lognormal SV model | | Proposed SV model | |
|---|---|---|---|---|---|
| | | Quasi-t-ratio | T-ratio | Quasi-t-ratio | T-ratio |
| VAR | $b_1$ | 0.108 | 0.734 | 0.135 | 0.611 |
| | $b_2$ | −0.063 | −0.292 | −0.113 | −0.912 |
| ARCH | $r_1$ | 1.022 | 2.753 | 1.619 | 3.075 |
| | $r_2$ | 1.091 | 2.122 | 2.220 | 2.963 |
| | $r_3$ | 1.175 | 2.577 | 2.028 | 3.092 |
| SNP | $s_1$ | 0.451 | 0.707 | 0.122 | 0.204 |
| | $s_2$ | 1.924 | 3.272 | 1.100 | 1.314 |
| | $s_3$ | 0.538 | 0.586 | 0.454 | 0.511 |
| | $s_4$ | 3.155 | 3.813 | 0.996 | 1.065 |
| | $s_5$ | 0.187 | 0.202 | 0.303 | 0.396 |
| | $s_6$ | 2.981 | 3.543 | 0.552 | 0.685 |

Note: The VAR and ARCH quasi-t-ratios and t-ratios correspond to the conditional mean equation and conditional variance equation of the SNP specification, respectively. The SNP quasi-t-ratios and t-ratios correspond to the coefficients of the polynomial of the SNP specification.

Table 6: Parameter estimates, standard errors, confidence intervals, $\chi^2$ criterion for S&P500

| | Lognormal SV model | Proposed SV model |
|---|---|---|
| $\mu_0$ | 0.0389<br>(0.0136)<br>$[0.0123, 0.0651]$ | 0.0387<br>(0.0137)<br>$[0.0122, 0.0650]$ |
| $c$ | 0.0880<br>(0.0159)<br>$[0.0583, 0.1179]$ | 0.0875<br>(0.0158)<br>$[0.0579, 0.1177]$ |
| $\mu$ | $-0.3425$<br>(0.0613)<br>$[-0.4593, -0.2291]$ | $-0.4474$<br>(0.0812)<br>$[-0.5066, -0.3441]$ |
| $\phi$ | 0.9846<br>(0.0120)<br>$[0.9657, 0.9966]$ | 0.9840<br>(0.0106)<br>$[0.9736, 0.9946]$ |
| $\sigma$ | 0.1022<br>(0.0456)<br>$[0.0566, 0.1719]$ | 0.1140<br>(0.0405)<br>$[0.0734, 0.1532]$ |
| $\delta$ | NA | $-0.4597$<br>(0.1807)<br>$[-0.6316, -0.2885]$ |
| $\chi^2$ | 25.01 | 21.65 |
| df | 14 | 13 |
| $p-$ value | 0.034 | 0.061 |

Note: The number in parentheses is the standard error. The number in brackets is the confidence interval. The results are based on 101,000 runs with the first 1,000 runs discarded. The lognormal SV model is defined by Equations (4.18) and (4.19). The proposed SV model is defined by Equations (4.20) and (4.21).

Table 7: Quasi-t-ratios and t-ratios for S&P500

| | | Lognormal SV model | | Proposed SV model | |
|---|---|---|---|---|---|
| | | Quasi-t-ratio | T-ratio | Quasi-t-ratio | T-ratio |
| VAR | $b_1$ | $-0.096$ | $-0.334$ | $-0.089$ | $-0.302$ |
| | $b_2$ | $-1.243$ | $-1.249$ | $-1.243$ | $-1.259$ |
| | $b_3$ | $-0.446$ | $-1.540$ | $-0.436$ | $-1.571$ |
| ARCH | $r_1$ | 0.781 | 1.391 | 0.676 | 1.471 |
| | $r_2$ | 0.627 | 0.909 | 0.513 | 0.896 |
| | $r_3$ | 0.130 | 0.224 | 0.028 | 0.051 |
| | $r_4$ | 1.122 | 1.228 | 1.047 | 1.160 |
| | $r_5$ | 1.053 | 1.498 | 0.933 | 1.399 |
| | $r_6$ | 0.253 | 0.312 | 0.132 | 0.179 |
| | $r_7$ | 0.066 | 0.088 | $-0.052$ | $-0.071$ |
| | $r_8$ | 1.032 | 1.521 | 0.861 | 1.489 |
| | $r_9$ | 0.881 | 1.040 | 0.754 | 0.945 |
| | $r_{10}$ | 0.549 | 0.588 | 0.448 | 0.494 |
| | $r_{11}$ | $-0.243$ | $-0.280$ | $-0.388$ | $-0.500$ |
| | $r_{12}$ | 0.257 | 0.270 | 0.145 | 0.159 |
| SNP | $s_1$ | 0.033 | 0.063 | 0.093 | 0.179 |
| | $s_2$ | 0.977 | 3.091 | 0.706 | 2.477 |
| | $s_3$ | 0.205 | 0.232 | 0.258 | 0.291 |
| | $s_4$ | 2.303 | 3.767 | 1.798 | 3.332 |

Note: The VAR and ARCH quasi-t-ratios and t-ratios correspond to the conditional mean equation and conditional variance equation of the SNP specification, respectively. The SNP quasi-t-ratios and t-ratios correspond to the coefficients of the polynomial of the SNP specification.

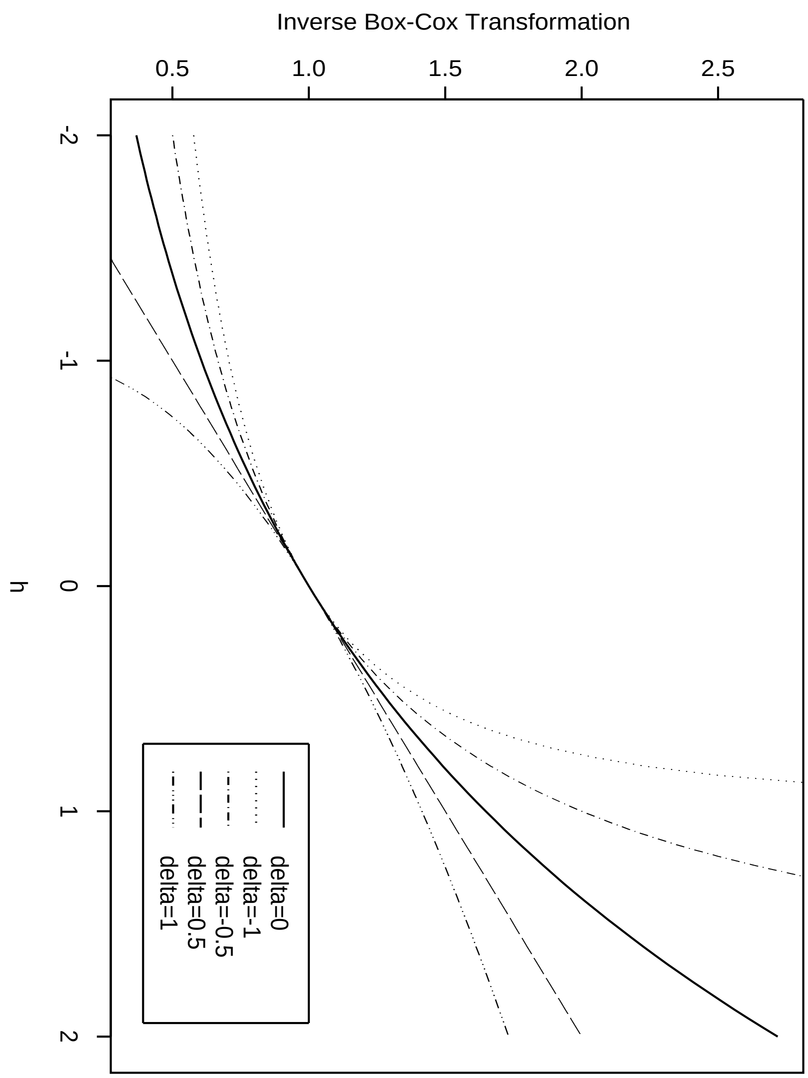


Figure 1: Inversion Box-Cox Transformation for Various Values of $\delta$

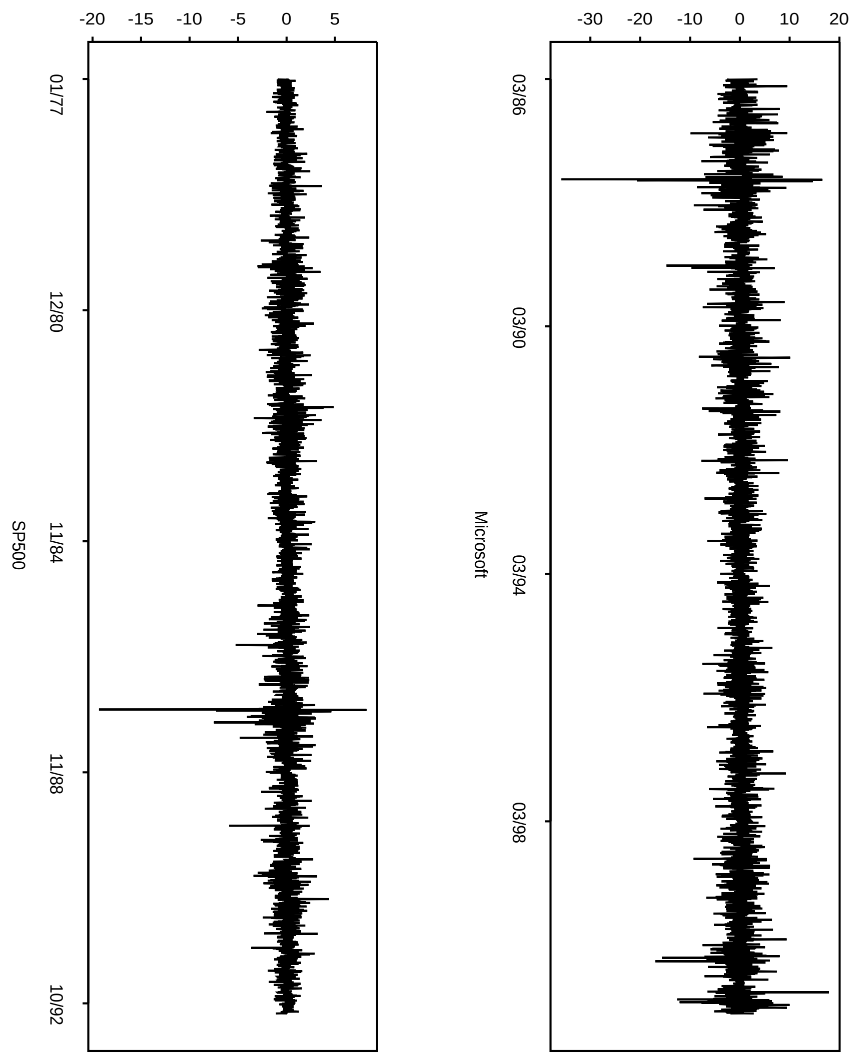


Figure 2: Time Series Plots for Microsoft Returns and S&P500 Returns

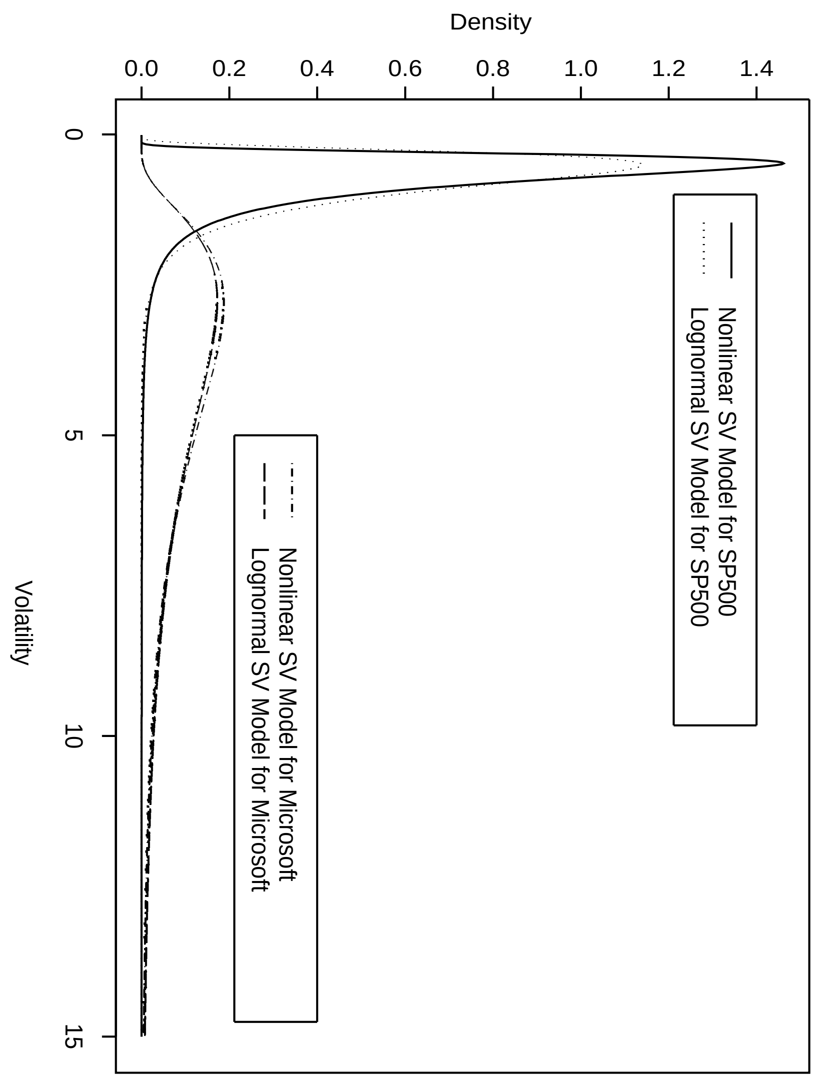


Figure 3: Marginal Densities of Volatility