Asymmetry in Stochastic Volatility Models: Threshold or Correlation?

Daniel R. Smith*

*Simon Fraser University, drsmith@sfu.ca

Copyright ©2009 The Berkeley Electronic Press. All rights reserved.
Asymmetry in Stochastic Volatility Models: Threshold or Correlation?*

Daniel R. Smith

Abstract

We compare the ability of correlation and threshold effects in a stochastic volatility model to capture the asymmetric relationship between stock returns and volatility. The parameters are estimated using maximum likelihood based on the extended Kalman filter and uses numerical integration over the latent volatility process. The stochastic volatility model with only correlation does a better job of capturing asymmetry than a threshold stochastic volatility model even though it has fewer parameters. We develop a stochastic volatility model that includes both threshold effects and correlated innovations. We find that the general model with both threshold effects and correlated innovations dominates purely threshold and correlated models. In this augmented model volatility and returns are negatively correlated, and volatility is more persistent, less volatile and higher following negative returns even after counting for the negative correlation.

*I would like to thank the editor Bruce Mizrach and two anonymous referees for their comments. All remaining errors are my own responsibility. Part of this paper was written while I was a Visiting Fellow in the Faculty of Business at the Queensland University of Technology in Brisbane, Australia. I gratefully acknowledge financial support from the Social Sciences and Humanities Research Council of Canada.
1 Introduction

There is a remarkable amount of empirical evidence that the relationship between stock returns and volatility is asymmetric. In GARCH models volatility increases more following large negative returns than following large positive returns (Glosten, Jagannathan, and Runkle, 1993) and in stochastic volatility models returns are found to be correlated with either future (Harvey and Shephard, 1996) or contemporaneous (Jacquier, Polson, and Rossi, 2004) volatility. There are two explanations for this relationship: the “leverage” effect in which equity volatility increases following the increase in financial leverage induced by negative returns (Black, 1976, and Christie, 1982), and the volatility feedback effect (French, Schwert, and Stambaugh, 1987, Campbell and Hentschel, 1992, Smith, 2006).

The stochastic volatility model introduced by Taylor (1986) is an important tool for modeling conditional volatility in a wide range of financial time series. The standard way to incorporate asymmetric volatility is to allow the shocks to returns and future volatility to be correlated. Correlated stochastic volatility models were introduced by Wiggins (1989) and Chesney and Scott (1989), and there are many algorithms that have been proposed to estimate the model’s parameters.\(^1\)

However, allowing for correlation between returns and volatility is not the only way to build an asymmetric stochastic volatility model. An alternative approach to incorporating asymmetry has recently been introduced by So, Li, and Lam (2003) who develop a threshold stochastic volatility model. In their model the intercept and persistence in the autoregressive model for log volatility (and the conditional mean parameters) change with the sign of the lagged stock returns. They present a Bayesian Markov-Chain Monte Carlo algorithm to estimate the model’s parameters. They find that volatility is higher following negative returns on the S&P500 index and that volatility persistence doesn’t change with the sign of lagged returns. Their model does not allow for correlation between returns and volatility.

In this paper we compare the ability of threshold effects and correlation to capture the asymmetric volatility-return relationship. In particular we develop a stochastic volatility model for stock returns that incorporates both threshold and correlation to allow for asymmetry. We develop an algorithm to estimate the model parameters by maximum likelihood. We show how the brute-force numerical integration filter of Fridman and Harris (1998), which

---

\(^1\)Interestingly, stochastic volatility models that account for this correlation are often labeled as capturing the “leverage effect” (see, e.g., Omori, Chib, Shephard, and Nakajima, 2007, Wu, 2001, and Yu, 2005) even though financial leverage is almost never included.
applies the exact nonlinear filter of Kitagawa (1987) to stochastic volatility models, can be extended to estimating the correlated-threshold stochastic volatility model. An advantage of this filter is the ease with which it can be extended to include quite complex models relatively simply. We find that the correlated stochastic volatility model does a better job of modeling return asymmetry than a purely threshold model even though it has fewer parameters. However, we find that a model that allows for both correlation and threshold effects dominates both restricted models. In this augmented model volatility and returns are negatively correlated, and volatility is more persistent, less volatile and higher following negative returns even after accounting for the negative correlation.

The remainder of the paper proceeds as follows. In Section 2 we discuss two alternative approaches to allowing for asymmetric relations between returns and volatility. We present our approach to estimating the model parameters in Section 3. Our empirical results are presented in Section 4. In Section 5 we present a small Monte Carlo simulation experiment to study the finite sample behavior of our tests and assess the robustness of our results to conditional non-normality. Finally, we conclude in Section 6.

2 Asymmetric Stochastic Volatility Models

The standard discrete-time univariate lognormal stochastic volatility model (SV) for the return series \( r_t \) models log-volatility, denoted by \( x_t = \log \sigma_t^2 \), as a latent AR(1) process:

\[
\begin{align*}
\rho_t &= \mu + \phi \rho_{t-1} + \exp(x_t/2)z_t \\
x_{t+1} &= \omega + \beta x_t + \sigma_v v_t
\end{align*}
\]

(2.2)

where the innovations \( z_t \) and \( v_t \) are standard normal random variables. The SV model fits stock returns and many other financial time series very well.\(^2\) The so-called leverage effect (Black, 1976, Christie, 1982) is modeled by allowing \( z_t \) and \( v_t \) to be correlated with correlation coefficient \( \rho = E(z_t v_t) \). The leverage effect is captured when \( \rho < 0 \) since conditional volatility increases on average following negative returns. This asymmetric relationship is an important feature of stock returns and is found in both GARCH (Nelson, 1991, Glosten et al., 1993) and SV models (Harvey and Shephard, 1996, Jacquier

et al., 2004, Omori et al., 2007, Yu, 2005) for equity returns. We choose an AR(1) specification of the conditional mean for simplicity but the model is easily generalized to more complex dynamics (see, e.g., Smith, 2006, for volatility-in-mean models with and without volatility feedback effects). It is well-known that this type of model, just like the GARCH model, generates fat tails in returns (see Carnero, Peña, and Ruiz (2004) for a comparison of SV and GARCH models in generating leptokurtosis) and by making the innovations nonnormal further deviations from normality can be captured (Sandmann and Koopman, 1998, and Jacquier et al., 2004).

So et al. (2003) propose a threshold model to capture the asymmetric relationship in which the parameters driving volatility dynamics vary with the sign of lagged stock returns:

\[ r_t = \mu_{s_{t-1}} + \phi_{s_{t-1}} r_{t-1} + \exp(x_t/2) z_t \]  
\[ x_{t+1} = \omega_{s_t} + \beta_{s_t} x_t + \sigma_{\nu,s_t} v_t, \]  

where \( z_t \) and \( v_t \) are independent standard normal random variables, \( s_t \) is the indicator function

\[ s_t = \begin{cases} 1 & \text{if } r_t < 0 \\ 0 & \text{if } r_t \geq 0 \end{cases}, \]

and the time-varying coefficients satisfy

\[ \mu_{s_{t-1}} = \mu_0 + s_{t-1} \cdot \mu_1 \]
\[ \phi_{s_{t-1}} = \phi_0 + s_{t-1} \cdot \phi_1 \]
\[ \omega_{s_t} = \omega_0 + s_t \cdot \omega_1 \]
\[ \beta_{s_t} = \beta_0 + s_t \cdot \beta_1 \]
\[ \sigma^2_{\nu,s_t} = \sigma^2_{\nu,0} + s_t \cdot \sigma^2_{\nu,1}. \]

This model has two components: a SETAR (Self-Exciting Threshold AutoRegressive model, pioneered by Tong, 1983) specification for the conditional mean with a threshold of zero, and a threshold specification for the conditional volatility dynamics. The model extends the basic SETAR model by allowing for stochastic volatility in which the volatility dynamics depend on the sign of lagged returns. Models with threshold effects can account for asymmetry through the intercept \( \omega \), which switches with the sign of lagged returns: when \( \omega_1 > 0 \) volatility will tend to be higher following negative returns than following positive returns.

Sun (2005) also finds an asymmetric effect in a stochastic volatility model of interest rates, and Asai and McAleer (2004) find asymmetry in exchange rates.
We extend this model to allow for both correlation and threshold effects by allowing $v_t$ and $z_t$ to be correlated:\textsuperscript{4,5}

$$E(z_tv_t) = \rho.$$ 

It is important to include correlation along with the threshold effect since they capture potentially different features. In a purely threshold model volatility increases after negative returns, but this increase is independent of the size of the negative return. On the other hand, in a model with only correlation the response of log-volatility is symmetric for positive and negative unexpected returns.\textsuperscript{6} Including both correlation and threshold permits a richer specification of the asymmetric stochastic volatility model.

3 Parameter Estimation

Estimating the parameters of stochastic volatility models is complicated because volatility is latent and the likelihood function is defined as the following $T$-dimensional integral

$$L(\theta; r) = \int f(r|x, \theta)f(x|\theta)dx$$

where $r$ and $x$ denote the $T$-dimensional sample paths of returns and volatility. Many estimation strategies have been proposed to estimate the parameters, including moment matching (Taylor, 1986, and Wiggins, 1989), Generalized Method of Moments (Melino and Turnbull, 1990) quasi-maximum likelihood (Harvey et al., 1994 and Harvey and Shephard, 1996), Bayesian-based Markov-Chain Monte Carlo methods (Jacquier, Polson and Rossi, 1994, 2004), Simulated Maximum Likelihood (Danielsson, 1994), Monte-Carlo Maximum Likelihood (Sandmann and Koopman, 1998), Efficient Method of Moments (Andersen et al., 1999), and numerical maximum likelihood using the exact nonlinear filter (Fridman and Harris, 1998, and Watanabe, 1999). The results of Monte

\textsuperscript{4}We don’t allow $\rho$ to change with the sign of returns though this could be done trivially. One would need to choose whether the correlation depends on the returns at time $t$ (as in the variance) or $t-1$ (as in the mean).

\textsuperscript{5}Asai and McAleer (2004) develop a model that is a special case of our model in which $\mu_1 = \phi_1 = \beta_1 = \sigma_{v,1} = 0$ and allows only $\omega$ to change with the sign of lagged returns. They label a model without correlation as the TE model and the model including correlation the DLTE model.

\textsuperscript{6}Though we note that Harvey and Shephard (1996, p. 431) demonstrate that when inference in the basic correlated SV model is conducted using the Kalman filter, the sign of returns does affect future volatility.

http://www.bepress.com/snde/vol13/iss3/art1
Carlo studies have found that quasi-maximum likelihood and GMM tend to be inefficient relative to the more computationally intensive approaches (i.e., EMM, MCMC, numerical integration, see appropriate references above).

Algorithms to estimate restricted versions or our correlated threshold stochastic volatility model have been proposed by So et al. (2003), who suggest a Bayesian Markov-Chain Monte Carlo algorithm, and Asai and McAleer (2004), who use the Monte Carlo Maximum Likelihood framework of Sandmann and Koopman (1998). In this section we show how to estimate the model parameters using maximum likelihood, which we implement with a nonlinear numerical integration-based filter. One of the key advantages of the exact nonlinear filter we employ below is the ease with which it can be extended.

### 3.1 The Basic Algorithm

Because we cannot use a standard filtering scheme we will estimate the parameters of this model with an algorithm that implements brute force numerical integration over the unobserved latent volatility using the conditional probability formulae following Kitagawa (1987) and Fridman and Harris (1998). Where the standard Kalman filter summarizes the entire distribution of the latent volatility process using its conditional mean and mean-squared-error, our algorithm keeps track of the conditional density of log-volatility at a finite set of $N$ points. The main advantage of this algorithm is that it reduces the $T$-dimensional integral to $T$ different 1-dimensional integrals, which makes maximum likelihood estimation feasible. Monte Carlo studies by Fridman and Harris (1998) and Watanabe (1999) show that such numerical integration-based algorithms have impressive finite sample properties.

To introduce the algorithm suppose that at time $t$ the conditional density of log-volatility is known to be $f(x_t|I_t)$, where $I_t$ denotes the information available to the econometrician at time $t$. The conditional normality of $x_{t+1}$ then implies that we can construct the forecast density of $x_{t+1}$:

$$f(x_{t+1}|I_t) = \int f(x_{t+1}|x_t)f(x_t|I_t)dx_t, \quad (3.1)$$

where $f(x_{t+1}|x_t) = \phi(x_{t+1}; \omega + \beta x_t, \sigma^2_t)$ and $\phi(.; a, b)$ is the pdf of a normal random variable with mean $a$ and variance $b$. The joint density of $x_{t+1}$ and the data $r_{t+1}$ is given by

$$f(r_{t+1}, x_{t+1}|I_t) = f(r_{t+1}|x_{t+1}, r_t)f(x_{t+1}|I_t), \quad (3.2)$$

where $f(r_{t+1}|x_{t+1}) = \phi(r_{t+1}; \mu + r_t, \exp(x_{t+1}))$. The log-likelihood function

Published by The Berkeley Electronic Press, 2009
requires the conditional density be known

\[ f(r_{t+1}|I_t) = \int f(r_{t+1}, x_{t+1}|I_t) dx_{t+1} \tag{3.3} \]

and is then calculated as

\[ \mathcal{L}(r; \theta) = \sum_{t=0}^{T-1} \log f(r_{t+1}|I_t). \tag{3.4} \]

We can update our inference about the distribution of \( x_{t+1} \) by conditioning on the new information \( r_{t+1} \) using

\[ f(x_{t+1}|I_{t+1}) = \frac{f(r_{t+1}, x_{t+1}|I_t)}{f(r_{t+1}|I_t)}. \tag{3.5} \]

The difficulty we now face is that we cannot summarize the whole conditional density from one iteration to the next since returns and volatility are not jointly normally distributed. In the standard Kalman filter the entire distribution is described by the filtered estimate and its mean squared error. However, in our case although returns are conditionally normal, volatility is lognormal and we therefore cannot use the standard Kalman filter. We overcome this limitation following Kitagawa (1987) and Fridman and Harris (1998) and use a numerical integration scheme which only requires that we keep track of the conditional densities at a finite number \( N \) points. In particular we employ a numerical scheme to calculate the general integral

\[ \int f(x)dx \approx \sum_{i=1}^{N} w_i f(x_i) \tag{3.6} \]

where \( w_i \) are a set of weights and \( f(x_i) \) denotes the value of the function \( f \) evaluated at only \( N \) different points \( x_i \) for \( i = 1, \ldots, N \), which are referred to as nodes. There are many schemes that approximate integrals with finite summation including the mid-point, trapezoidal, Simpson’s rule as well as the various quadrature rules. We follow Fridman and Harris (1998) and use Gauss-Legendre integration, which approximates \( f \) over a finite interval (set to plus and minus five standard deviations of the unconditional distribution of \( x_t \)) using \( N \) orthonormal polynomials. In this scheme the weights and nodes at which \( f \) is approximated solve a set of linear equations and are thus prespecified and common to all integrals (see Judd, 1998). This scheme is quite efficient and to ensure the accuracy of our results we use 50 nodes,
though the results are similar with either 25 or 75 nodes.\footnote{When \( N = 25 \) there is a small upward bias in the estimate of \( \sigma_\nu \), which is also noted by Watanabe (1999).} We index time \( t \) volatility by \( i = 1, \ldots, N \) and time \( t + 1 \) volatility by \( j = 1, \ldots, N \).

Below we present the algorithm to compute the log-likelihood function and generates the filtered density of latent volatility as a by-product. The algorithm takes the density of volatility evaluated at \( N \) points \( f(x_t = x_j|I_t) \) as input, and generates the marginal density of returns \( f(r_{t+1}|I_t) \), which we use to compute the log-likelihood, and inference about volatility \( f(x_{t+1} = x_i|I_{t+1}) \) that is then used as the input for the subsequent iteration.

**Step 1.** Given the density of volatility \( f(x_t = x_j|I_t) \) evaluated at \( N \) nodes as input from the previous iteration, construct the forecast density of volatility:

\[
f(x_{t+1} = x_i|I_t) \approx \sum_{j=1}^{N} w_j \phi(x_i; \omega + \beta x_j, \sigma^2_\nu) f(x_t = x_j|I_t). \tag{3.7}
\]

Note that the transition density \( \phi(x_i; \omega + \phi x_j, \sigma^2_\nu) \) only needs to be computed once.

**Step 2.** Compute the joint density of returns and volatility:

\[
f(r_{t+1}, x_{t+1} = x_i|I_t) = \phi(r_{t+1}; \mu + \phi r_t, \exp(x_i)) f(x_{t+1} = x_i|I_t). \tag{3.8}
\]

**Step 3.** Compute and store the marginal density of returns

\[
f(r_{t+1}|I_t) \approx \sum_{i=1}^{N} w_i f(r_{t+1}, x_{t+1} = x_i|I_t), \tag{3.9}
\]

which is used to construct the log-likelihood as in (3.4).

**Step 4.** Update the conditional density of log-volatility as

\[
f(x_{t+1} = x_i|I_{t+1}) = \frac{f(r_{t+1}, x_{t+1} = x_i|I_t)}{f(r_{t+1}|I_t)}, \tag{3.10}
\]

which is then used as the input for the next iteration in the algorithm.

**Step 5.** Goto Step 1.

We initialize the algorithm using the unconditional density of \( x_{t+1} \):

\[
f(x_1 = x_i|I_0) = \phi(x_i \left| \frac{\omega}{1 - \beta}, \frac{\sigma^2_\nu}{1 - \beta^2} \right). \tag{3.11}
\]
3.2 The Correlated Model

When returns are correlated with future volatility we need to integrate over both current and future stock return volatility since the return depends on the change in volatility:

\[ f(r_{t+1}|I_t) = \int \int f(r_{t+1}, x_{t+1}, x_{t+2}|I_t)dx_{t+1}dx_{t+2} \quad (3.12) \]

and the joint density of \((r_{t+1}, x_{t+1}, x_{t+2})\) depends on two correlated normally distributed innovations \(z_{t+1}\) and \(\nu_{t+1}\). We can simplify the exposition by using the Cholesky decomposition of the correlation matrix, which allows us to express the innovations as functions of two independent standard normal random variables \(u_{1,t}\) and \(u_{2,t}\):

\[
\begin{aligned}
z_{t+1} &= u_{1,t+1} \\
\nu_{t+1} &= \rho u_{1,t+1} + \sqrt{1-\rho^2}u_{2,t+1}.
\end{aligned}
\]

We can now write the joint density as

\[ f(r_{t+1}, x_{t+1}, x_{t+2}|I_t) = \sigma \nu^{-1} \exp(-x_{t+1}/2)\phi(u_{1,t+1})\phi(u_{2,t+1}) f(x_{t+1}|I_t) \quad (3.13) \]

where

\[
\begin{aligned}
u_{1,t+1} &= (r_{t+1} - \mu - \phi x_t) \cdot \exp(-x_{t+1}/2) \\
u_{2,t+1} &= (x_{t+2} - \omega - \beta x_t - \rho \nu_t u_{1,t+1}) \cdot \sigma \nu^{-1}.
\end{aligned}
\]

The updated inference of volatility is then given by

\[ f(x_{t+1}|I_{t+1}) = \frac{\int f(r_{t+1}, x_{t+1}, x_{t+2}|I_t)dx_{t+2}}{f(r_{t+1}|I_t)} \quad (3.14) \]

and the input for the next conditional density is

\[ f(x_{t+2}|I_{t+1}) = \frac{\int f(r_{t+1}, x_{t+1}, x_{t+2}|I_t)dx_{t+1}}{f(r_{t+1}|I_t)}. \quad (3.15) \]

Just as in the uncorrelated model, we implement estimation using numerical integration over the common grid of \(N\) different volatility points \(\{x_i\}_{i=1}^N\) for both \(x_{t+1}\) and \(x_{t+2}\). For expositional clarity we index time \(t+1\) volatility by \(i = 1, \ldots, N\) and time \(t+2\) volatility by \(j = 1, \ldots, N\). The input for the \((t+1)\)th iteration is the vector of conditional densities \(f(x_{t+1} = x_i|I_t)\). The algorithm takes as input the forecast density of volatility \(f(x_{t+1} = x_i|I_t)\), and produces the density of returns \(f(r_{t+1}|I_t)\), updated inference about period \(t+1\)’s volatility \(f(x_{t+1} = x_i|I_{t+1})\), and a forecast of future volatility \(f(x_{t+2} = x_i|I_{t+1})\), which is then the input for the next iteration.

http://www.bepress.com/snde/vol13/iss3/art1
Step 1. Compute the joint density for the \(N^2\) possible values of \(x_{t+1}\) and \(x_{t+2}\):

\[
f(r_{t+1}, x_{t+1} = x_i, x_{t+2} = x_j | I_t) = \sigma_{\nu}^{-1} \exp(-x_i/2) \phi(u_{1,t+1}^{(i)}) \phi(u_{2,t+1}^{(i,j)}) f(x_{t+1} = x_i | I_t), \text{ for } i,j = 1, \ldots, N
\]

(3.16)

with

\[
u(i)_{1,t+1} = (r_{t+1} - \mu - \phi r_t) \cdot \exp(-x_i/2)
\]

\[
u(i,j)_{2,t+1} = (x_j - \omega - \beta x_i - \rho \sigma_{\nu} u_{1,t+1}^{(i)}) \cdot \sigma_{\nu}^{-1}.
\]

Step 2. Calculate and store the marginal density of returns:

\[
f(r_{t+1} | I_t) \approx \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j f(r_{t+1}, x_{t+1} = x_i, x_{t+2} = x_j | I_t),
\]

(3.17)

which is used to construct the log-likelihood as in (3.4).

Step 3. Update inference about latent volatility using the density from Step 2

\[
f(x_{t+1} = x_i | I_{t+1}) \approx \frac{\sum_{j=1}^{N} w_j f(r_{t+1}, x_{t+1} = x_i, x_{t+2} = x_j | I_t)}{f(r_{t+1} | I_t)}.
\]

(3.18)

Step 4. Compute inference about future volatility and use in the next step, again using the density of \(r_{t+1}\) from Step 2

\[
f(x_{t+2} = x_i | I_{t+1}) \approx \frac{\sum_{i=1}^{N} w_i f(r_{t+1}, x_{t+1} = x_i, x_{t+2} = x_j | I_t)}{f(r_{t+1} | I_t)}.
\]

(3.19)

Step 5. Goto Step 1.

Again we initialize the algorithm using the unconditional density of \(x_{t+1}\) given in equation (3.11). The algorithm requires a little more computational effort since the joint density must be computed at the \(N^2\) combinations of current and future volatility at each time point, i.e., \(T \cdot N^2\) times.
3.3 Threshold Models

An important advantage of the exact nonlinear filter is the ease with which it can be extended. In particular, we only need to slightly redefine the transition densities in the algorithms presented in Sections 3.1 and 3.2 to allow the threshold model parameters to be estimated by maximum likelihood with numerical integration. To estimate the threshold SV model without correlation, denoted the TSV model, only requires that we modify Steps 1 and 2 in Section 3.1 above, by replacing equation (3.7) in Step 1 with

\[
f(x_{t+1} = x_i | I_t) \approx \sum_{j=1}^{N} w_j \phi(x_i; \omega_{s_t} + \beta_{s_t} x_j, \sigma_{\nu,s_t}^2) f(x_i = x_j | I_t),
\]

and equation (3.8) in Step 2 by

\[
f(r_{t+1}, x_{t+1} = x_i | I_t) = \phi(r_{t+1}; \mu_{s_t} + \phi_{s_t} x_i, \exp(x_i)) f(x_{t+1} = x_i | I_t),
\]

noting that \( r_t \in I_t \). In the correlated threshold model, denoted TSVA, we only need to modify equation (3.12) in Step 1 of Section 3.2 by using the alternative definition of the uncorrelated normal innovations as

\[
u_{1,t+1}^{(i)} = \left( r_{t+1} - \mu_{s_t} - \phi_{s_t} r_t \right) \cdot \exp(-x_i/2)
\]

\[
u_{2,t+1}^{(i,j)} = \left( x_j - \omega_{s_t} - \beta_{s_t} x_i - \rho \sigma_{\nu,s_{t+1}} u_{1,t+1}^{(i)} \right) \cdot \sigma_{\nu,s_{t+1}}^{-1}.
\]

All other steps proceed as before.

The exact nonlinear filter is a powerful tool for estimating stochastic volatility models because it is so easy to generalize to quite complex models. The extra coding time required to accommodate threshold effects is negligible. It also allows us to estimate a model which nests both the TSV and SVA models, which is not a trivial feat. So et al. (2003) develop a Bayesian

\[\text{As pointed out by a referee, Markov chain Monte Carlo methods can also be modified quite easily.}\]

\[\text{The unconditional density is constructed as follows. Define the } N \times N \text{ matrices } \rho_0 \text{ and } \rho_1 \text{ with element } (j,i)\\ \rho_k^{(j,i)} = u_j \phi(x_i; \omega_k + \beta_k x_j, \sigma_{\nu,k}^2) \text{ for } k = 0, 1\\ \text{using these two conditional transition matrices, we define the unconditional density as}\\ \pi = \lim_{n \to \infty} (0.5 \cdot \rho_0 + 0.5 \cdot \rho_1)^n \times \pi,\\ \text{which is computed setting } n \text{ to a suitably large value and a suitable starting value.}\]
Markov-Chain Monte Carlo algorithm to estimate a restricted version of this model that omits leverage effects and does not allow the variance of log-volatility to vary. Similarly, Asai and McAleer (2004) use Monte Carlo maximum likelihood to estimate a model with correlated innovations but only allow for threshold effects in the intercept. Fortunately, our numerical integration-based maximum likelihood scheme allows for correlation in a general threshold stochastic volatility model and we find that both effects are important.

Sandmann and Koopman (1998) note that this algorithm of Kitagawa (1987) and Fridman and Harris (1998) has some disadvantages. It is known to be inefficient relative to the particle filter, and requires one to pre-specify the grid over which volatility is to be integrated. To have a wide enough interval we must either choose a sparse grid with poor numerical properties, or miss extreme volatility episodes such as the crash of October 1987. However, as noted above, we find that the computational inefficiencies are minor. On the other hand, the Monte Carlo Maximum Likelihood (MCML) algorithm suggested by Sandmann and Koopman (1998) takes the Kalman filter as its base and corrects the distortion of the conditional density. Because the Kalman filter uses the log-squared return as its input, this algorithm is affected by the so-called “inlier problem”, which occurs because small returns result in very large negative log-squared-returns. These outliers can complicate the finite sample properties. A more important issue is that the mean parameters are not estimated jointly with the variance dynamics. Although Sandmann and Koopman (1998) show that the MCML algorithm can be extended relatively simply, the exact nonlinear filter is even easier to extend in many interesting directions with virtually no effort. For example, Smith (2006) shows how the algorithm can be used to incorporate a volatility feedback effect into a stochastic volatility-in-mean model using the log-linear decomposition of Campbell and Shiller (1988). Although other estimation methods such as EMM and

\[10\] In Section 5.1 we undertake a Monte Carlo study in which we estimate a range of stochastic volatility models with \(T = 1,000\) observations. The calculations were performed using Matlab release 13 on dual processor Athlon MP 2800+ processors (2.133 GHz) with 1GB RAM per processor running Linux. The log-likelihoods were maximized numerically using the algorithm fminunc.m. We estimate two models: the computationally efficient quasi-maximum likelihood model based on the Kalman filter that took on average 17.5 seconds to estimate the parameters and compute the standard errors by numerical differentiation and the numerical integration-based filter that took almost twice as long at 32.0 seconds to estimate the model parameters and standard errors, yet this is still quite feasible. The asymmetric model, which allows for correlation between returns and volatility, took 40.1 seconds on average to estimate using the Kalman filter-based algorithm of Harvey and Shephard (1996), and the exact numerical integration-based filter took about six times as long at 264.2 seconds, which is still less than five minutes—again more than feasible.
4 Empirical Results

In our empirical analysis we use continuously compounded returns on the value-weighted CRSP portfolio from July 1963 to December 2004, giving a total of 10,698 daily returns. We plot the returns in Figure 1 and observe the well-known volatility clustering phenomena. We observe the extreme returns of October 1987 and the higher levels of volatility during the recessions of 1970 and 1974 and the early 1980s, as well as the protracted period of higher volatility associated with the Tech bubble and its subsequent collapse.

We estimate a range of stochastic volatility models with various specifications for asymmetric volatility. Because daily stock index returns are known to be serially correlated due to nonsynchronous trading we model the condi-
tional mean as an AR(1) process. Our main focus is on threshold effects as a source of asymmetry so we estimate a simplified threshold model in which only volatility dynamics change with the sign of lagged returns.

We report the parameter estimates in Table 1 for the full sample, and Table 2 focuses on the shorter post-1989 data, which avoids the crash of 1987. As expected, we find that returns are moderately autocorrelated with $\phi$ being around 0.15 in the full sample. Return autocorrelation is a little lower in the post-1989 sample with $\phi$ being around 0.08, but it is still highly statistically significant. We also find that volatility is extremely persistent. In the symmetric SV models we find $\beta$ to be a little over 0.98 in both sub-samples. The point estimates of 0.9881 and 0.9862 imply that a shock to log-volatility has a half-life ($\log(\beta)/\log(0.5)$) of 58 and 50 trading days respectively. Accounting for correlation between returns and volatility does little to the persistence of volatility, which is still more than 0.98 in both samples. However, the correlation between shocks to returns and future volatility are both large (between around -0.55 and -0.65) and highly significant, improving the log likelihoods by more than 45 in both cases and is more than 10 times its standard error. We plot the conditional volatility from a range of models in Figures 2 and 3 and note that the models generate reasonably similar volatility dynamics.
Table 1: Parameter Estimates for TSV Model: 1962-2004

<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th>SVA</th>
<th>RTSV</th>
<th>RTSVA</th>
<th>TSV</th>
<th>TSVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0005</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.1705</td>
<td>0.1771</td>
<td>0.1718</td>
<td>0.1770</td>
<td>0.1724</td>
<td>0.1735</td>
</tr>
<tr>
<td></td>
<td>(0.0099)</td>
<td>(0.0099)</td>
<td>(0.0099)</td>
<td>(0.0105)</td>
<td>(0.0100)</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\omega_0$</td>
<td>-0.1500</td>
<td>-0.1814</td>
<td>-0.2617</td>
<td>-0.1820</td>
<td>-0.4387</td>
<td>-0.3089</td>
</tr>
<tr>
<td></td>
<td>(0.0222)</td>
<td>(0.0193)</td>
<td>(0.0234)</td>
<td>(0.0257)</td>
<td>(0.0330)</td>
<td>(0.0488)</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.9850</td>
<td>0.9819</td>
<td>0.9810</td>
<td>0.9819</td>
<td>0.9627</td>
<td>0.9642</td>
</tr>
<tr>
<td></td>
<td>(0.0022)</td>
<td>(0.0020)</td>
<td>(0.0023)</td>
<td>(0.0023)</td>
<td>(0.0034)</td>
<td>(0.0052)</td>
</tr>
<tr>
<td>$\sigma^2_{0,\nu}$</td>
<td>0.0253</td>
<td>0.0307</td>
<td>0.0203</td>
<td>0.0306</td>
<td>0.0239</td>
<td>0.0625</td>
</tr>
<tr>
<td></td>
<td>(0.0022)</td>
<td>(0.0026)</td>
<td>(0.0025)</td>
<td>(0.0039)</td>
<td>(0.0044)</td>
<td>(0.0091)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.5690</td>
<td>-0.5678</td>
<td>-0.6630</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0284)</td>
<td>(0.0490)</td>
<td>(0.0317)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.1575</td>
<td>0.0008</td>
<td>0.5842</td>
<td>0.3871</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0132)</td>
<td>(0.0246)</td>
<td>(0.0717)</td>
<td>(0.0615)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.0413</td>
<td>0.0376</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0077)</td>
<td>(0.0057)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{1,\nu}$</td>
<td>-0.0110</td>
<td>-0.0546</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0062)</td>
<td>(0.0089)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td>37463.8</td>
<td>37571.0</td>
<td>37534.9</td>
<td>37571.0</td>
<td>37543.6</td>
<td>37594.9</td>
</tr>
<tr>
<td>AIC</td>
<td>-74917.5</td>
<td>-75130.0</td>
<td>-75057.7</td>
<td>-75128.0</td>
<td>-75071.3</td>
<td>-75171.7</td>
</tr>
<tr>
<td>BIC</td>
<td>-74881.2</td>
<td>-75086.4</td>
<td>-75014.1</td>
<td>-75077.1</td>
<td>-75013.0</td>
<td>-75106.2</td>
</tr>
<tr>
<td>Joint</td>
<td>26.26</td>
<td>7.02</td>
<td>7.3061</td>
<td>7.0628</td>
<td>6.3196</td>
<td>10.3980</td>
</tr>
<tr>
<td>Bias</td>
<td>10.9908</td>
<td>2.2973</td>
<td>5.0994</td>
<td>2.3154</td>
<td>3.8289</td>
<td>0.4429</td>
</tr>
<tr>
<td>N. Size</td>
<td>10.9908</td>
<td>2.2973</td>
<td>5.0994</td>
<td>2.3154</td>
<td>3.8289</td>
<td>0.4429</td>
</tr>
</tbody>
</table>

NOTES: We present the parameter estimates and standard errors (computed using a numerically evaluated Hessian matrix) in parenthesis for a range of stochastic volatility models, which are all special cases of

$$r_t = \mu + \phi r_{t-1} + \exp(x_t/2)u_{1,t},$$

$$x_{t+1} = \omega_0 + s_t \omega_1 + (\beta_0 + s_t \beta_1)x_t + \sqrt{\sigma^2_{0,\nu} + s_t \sigma^2_{1,\nu}} \left( \rho u_{1,t} + \sqrt{1-\rho^2} u_{2,t} \right).$$

We also report the AIC and BIC for each model, along with conditional moment tests for

$$E \left( \frac{z_{t+1}^2}{z_{t+1}^2 \cdot s_t \cdot r_t} \right) = 0$$

where $z_{t+1} = \Phi^{-1}(F_t(r_{t+1}))$, which is labeled the joint test, the sign bias test for $E(z_{t+1}^2 \cdot s_t) = 0$, and the size-bias test for $E(z_{t+1}^2 \cdot s_t \cdot r_t) = 0$. 

http://www.bepress.com/snde/vol13/iss3/art1
Table 2: Parameter Estimates for TSV Model: 1990-2004

<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th>SVA</th>
<th>RTSV</th>
<th>RTSVA</th>
<th>TSV</th>
<th>TSVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.0007</td>
<td>0.0004</td>
<td>0.0005</td>
<td>0.0004</td>
<td>0.0006</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>( 0.0001)</td>
<td>( 0.0001)</td>
<td>( 0.0001)</td>
<td>( 0.0001)</td>
<td>( 0.0002)</td>
<td>( 0.0002)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.0758</td>
<td>0.0879</td>
<td>0.0778</td>
<td>0.0884</td>
<td>0.0784</td>
<td>0.0857</td>
</tr>
<tr>
<td></td>
<td>( 0.0167)</td>
<td>( 0.0168)</td>
<td>( 0.0168)</td>
<td>( 0.0173)</td>
<td>( 0.0169)</td>
<td>( 0.0170)</td>
</tr>
<tr>
<td>$\omega_0$</td>
<td>-0.1149</td>
<td>-0.1715</td>
<td>-0.2266</td>
<td>-0.1683</td>
<td>-0.7169</td>
<td>-0.5447</td>
</tr>
<tr>
<td></td>
<td>( 0.0101)</td>
<td>( 0.0378)</td>
<td>( 0.0433)</td>
<td>( 0.0388)</td>
<td>( 0.0743)</td>
<td>( 0.0528)</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.9881</td>
<td>0.9823</td>
<td>0.9844</td>
<td>0.9820</td>
<td>0.9337</td>
<td>0.9400</td>
</tr>
<tr>
<td></td>
<td>( 0.0011)</td>
<td>( 0.0039)</td>
<td>( 0.0042)</td>
<td>( 0.0040)</td>
<td>( 0.0077)</td>
<td>( 0.0082)</td>
</tr>
<tr>
<td>$\sigma_{0,\nu}^2$</td>
<td>0.0197</td>
<td>0.0287</td>
<td>0.0150</td>
<td>0.0305</td>
<td>0.0217</td>
<td>0.0655</td>
</tr>
<tr>
<td></td>
<td>( 0.0031)</td>
<td>( 0.0058)</td>
<td>( 0.0044)</td>
<td>( 0.0078)</td>
<td>( 0.0068)</td>
<td>( 0.0259)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.6457</td>
<td>-0.6595</td>
<td>-0.6455</td>
<td>1.0778</td>
<td>0.6460</td>
<td>0.0934</td>
</tr>
<tr>
<td></td>
<td>( 0.0464)</td>
<td>( 0.0577)</td>
<td>( 0.0820)</td>
<td>( 0.0630)</td>
<td>( 0.0091)</td>
<td>( 0.0065)</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.1644</td>
<td>-0.0127</td>
<td>-0.0114</td>
<td>0.0014</td>
<td>3.5110</td>
<td>1.2331</td>
</tr>
<tr>
<td></td>
<td>( 0.0220)</td>
<td>( 0.0378)</td>
<td>( 0.0104)</td>
<td>( 0.9703)</td>
<td>( 0.0610)</td>
<td>( 0.0367)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.0934</td>
<td>0.0637</td>
<td>0.0587</td>
<td>0.0065</td>
<td>0.0029</td>
<td>0.6118</td>
</tr>
<tr>
<td></td>
<td>( 0.0065)</td>
<td>( 0.0029)</td>
<td>( 0.0083)</td>
<td>( 0.0028)</td>
<td>( 0.0028)</td>
<td>( 0.0028)</td>
</tr>
<tr>
<td>$\sigma_{1,\nu}^2$</td>
<td>-0.0114</td>
<td>-0.0587</td>
<td>6.2965</td>
<td>6.2965</td>
<td>6.2965</td>
<td>6.2965</td>
</tr>
<tr>
<td></td>
<td>( 0.0042)</td>
<td>( 0.1204)</td>
<td>( 0.0098)</td>
<td>( 0.0610)</td>
<td>( 0.0060)</td>
<td>( 0.0060)</td>
</tr>
</tbody>
</table>

NOTES: We present the parameter estimates and standard errors (computed using a numerically evaluated Hessian matrix) in parenthesis for a range of stochastic volatility models, which are all special cases of

\[ r_t = \mu + \phi r_{t-1} + \exp(x_t/2)u_{1,t} \]
\[ x_{t+1} = \omega_0 + s_t \omega_1 + (\beta_0 + s_t \beta_1)x_t + \sqrt{\sigma_{0,\nu}^2 + s_t \sigma_{1,\nu}^2} \left( \rho u_{1,t} + \sqrt{1 - \rho^2} u_{2,t} \right). \]

We also report the AIC and BIC for each model, along with conditional moment tests for

\[ E\left( z_{t+1}^2 \cdot s_t \right) = 0 \quad \text{where} \quad z_{t+1} = \Phi^{-1}(F_t(r_{t+1})) \], which is labeled the joint test, the sign bias test for \( E(z_{t+1}^2 \cdot s_t) = 0 \), and the size-bias test for \( E(z_{t+1}^2 \cdot s_t \cdot r_t = 0 \).
To assess the extent to which our results depend on the use of numerical integration-based maximum likelihood estimation we consider an alternative estimation technique: Bayesian Markov chain Monte Carlo estimation. We estimate the Bayesian model parameters using the program WinBUGS following Meyer and Yu (2000) and Yu (2005).\textsuperscript{11} The algorithm generates draws from the posterior distribution of the parameters given the data sequentially as a Markov Chain. We simulate and discard the first 20,000 iterations as a burn-in period that is followed by a subsequent series of 10,000 draws. Details of the algorithm along with the assumed prior distributions are presented in the appendix. To save space we only report results for the post-1989 sample. We report the average and standard deviation (in parenthesis) across the draws from the posterior distribution for the symmetric stochastic volatility

\textsuperscript{11}The WinBUGS-based Bayesian model works well for relatively simple models, which is the reason we restrict ourselves to the non-threshold models.
model for the post-1989 sample:

\[
\begin{align*}
    r_t &= 0.0007203 \pm 0.0001133 + 0.07394 \pm 0.01672 \times r_{t-1} + \exp(x_t/2) \times z_t \\
    x_{t+1} &= -0.110469 \pm 0.030719 + 0.9885 \pm 0.003158 \times x_t + 0.1388 \pm 0.01286 \times v_t
\end{align*}
\]

and similarly for the correlated asymmetric stochastic volatility model:

\[
\begin{align*}
    r_t &= 0.0004321 \pm 0.0001193 + 0.0874 \pm 0.01696 \times r_{t-1} + \exp(x_t/2) \times z_t \\
    x_{t+1} &= -0.163735 \pm 0.032784 + 0.9830 \pm 0.003425 \times x_t + 0.1643 \pm 0.01253 \times v_t
\end{align*}
\]

The average correlation coefficient was -0.6532 and the standard deviation is 0.04433. The average parameter estimates across the simulations are remarkably close to the maximum likelihood estimates as reported in Table
2, which suggests that our conclusions are not driven by our choice of the numerical integration-based estimation method. To evaluate the importance of correlation we use the Deviance Information Criteria\textsuperscript{12} (DIC hereafter) of Spiegelhalter, Carlin, and van der Linde (2002) that is automatically computed by WinBUGS, which is related to the ML-based information criteria AIC and BIC. A better model is indicated by a smaller DIC and Berg, Meyer, and Yu (2004) find that the DIC is able to identify correctly specified stochastic volatility models quite well. The DIC for the basic symmetric SV model is -25573.1, which drops by over 650 to -26256.5 after allowing for correlation, providing strong evidence of asymmetry.

We next ask the question, does allowing for correlation or threshold effects do a better job in capturing the asymmetric relationship between returns and volatility? To address this issue we estimate a restricted threshold model that sets the correlation parameter $\rho$ to zero, which is denoted as RTSV. This model has the same number of parameters as the correlated stochastic volatility model. In both data sets the SVA model produces a higher log-likelihood (the difference is about 40 in the full sample and 20 in the post-1989 sample) than the RTSV model, suggesting that correlation does a better job than threshold effects.

The parameter $\omega_1$ captures the difference in log-volatility following negative rather than following positive returns. It can be interpreted as approximately how much higher the conditional variance of stock returns is, measured as a percentage, following negative returns. We find in both samples that $\omega_1$ is around 0.15, which implies that the conditional variance of

\textsuperscript{12}The DIC is based on the deviance

$$D(\theta) = -2 \log f(y|\theta) + 2 \log g(y)$$

where $f(y|\theta)$ is the likelihood function and $g(y)$ is a model-independent normalizing term. The DIC is defined as

$$DIC = \bar{D} + p_D$$

and consists of two parts: the posterior average of the deviance

$$\bar{D} = E_{\theta|y} (D(\theta)) = E_{\theta|y} (-2 \log f(y|\theta)),$$

which is a measure of model fit (better models are implied by smaller values of $\bar{D}$) and the effective number of parameters

$$p_D = D - D(\bar{\theta}),$$

which is the difference between the posterior average of the deviance and the deviance evaluated at the posterior average of the parameters. Smaller values of DIC result from better fitting models (i.e., smaller $\bar{D}$), and more parsimonious parameterizations (i.e., smaller $p_D$).
stock returns is about 15 percent higher following negative returns than it is following positive returns:

$$E(\log \sigma_{t+1}^2 | r_t < 0) - E(\log \sigma_{t+1}^2 | r_t > 0) = \omega_1. \quad (4.1)$$

In the correlated stochastic volatility model (i.e., SVA), volatility is higher following negative returns than positive returns, since the standardized unexpected return $z_t$ is negatively correlated with the shock to $u_t$. The importance of the sign of $z_t$ can be gauged by computing

$$E(\log \sigma_{t+1}^2 | z_t < 0) - E(\log \sigma_{t+1}^2 | z_t > 0) = \sigma_\nu (E(u_t | z_t < 0) - E(u_t | z_t > 0)),$$

which can be simplified using the Cholesky orthogonalization for $u_t = \rho z_{1,t} + \sqrt{1 - \rho^2} z_{2,t}$ with $z_{1,t} = z_t$ independent of $z_{2,t}$, giving

$$E(\log \sigma_{t+1}^2 | z_t < 0) - E(\log \sigma_{t+1}^2 | z_t > 0) = \sigma_\eta \rho \{ E(z_t | z_t < 0) - E(z_t | z_t > 0) \} = -\sigma_\eta \rho \sqrt{2/\pi}. \quad (4.2)$$

This equation obtains by noting that $E(u_t | z_t < 0) = E(\rho z_{1,t} + \sqrt{1 - \rho^2} z_{2,t} | z_t < 0) = \rho E(z_t | z_t < 0)$, and since $z_t$ is symmetric, $E(z_t | z_t < 0) - E(z_t | z_t > 0) = -E(|z_t| | z_t < 0) - E(|z_t| | z_t > 0) = -E(|z_t|)$, and $E(|z_t|) = \sqrt{2/\pi}$ since $z_t \sim N(0, 1)$. Using this result we find that volatility is 5.7% higher following negative returns in the full-sample case, and 8.73% in the post-1989 data.

The threshold model captures a larger increase in volatility following negative returns, but ignores the size of the negative return, which matters greatly in the correlated model. The threshold model will create a larger difference in volatility following small to modest returns, while the correlated model will generate larger differences following large losses.

To formally compare the SVA and RTSV models we estimate an augmented model, denoted RTSVA that nests both as special cases. The log-likelihood improvement over the SVA model is trivial in both samples. Furthermore, the point estimate of $\omega_1$ is now trivially different from zero both statistically and economically—being less than 1 percent in the full sample, and slightly negative in the post-1989 sample.\(^{14}\)

\(^{13}\)To see this, note that $E(\log \sigma_{t+1}^2 | r_t < 0) = \omega_0 + \omega_1 + \phi E(\log \sigma_t^2)$ and $E(\log \sigma_{t+1}^2 | r_t > 0) = \omega_0 + \phi E(\log \sigma_t^2)$, taking the difference gives the result.

\(^{14}\)Asai and McAleer (2004) report a very similar effect in their analysis of daily S&P500 returns between June 1986 and April 2000. They find that the log-likelihood in the RTSV (they label as DLTE) and SVA (they label as DL) are identical to one decimal. The point estimate of the threshold parameter the volatility equation $\omega_1$ ($\gamma$ in their notation) in their paper is quite small at 0.0136, which is only roughly one-third its standard error of 0.0452.
An alternative approach to comparing the non-nested standard SVA model with the two threshold models RTSV and TSV we employ the non-nested likelihood ratio test of Vuong (1989), who derives a suitably normalized likelihood ratio test statistic that has a standard normal asymptotic distribution. In particular, Vuong (1989) shows that under certain regularity conditions the appropriately normalized likelihood ratio statistic converges to a standard normal random variable:

\[ T^{-1/2} LR_T / \hat{\omega}_T \overset{D}{\longrightarrow} N(0, 1) \]  

(4.3)

where \( LR_T = L^A_T - L^B_T \), and \( L^A_T \) and \( L^B_T \) are the log-likelihoods of two non-nested models, \( T \) is the number of observations, and

\[ \hat{\omega}^2_T = \frac{1}{T} \sum_{t=1}^{T} \left( \log \frac{f_A(r_t)}{f_B(r_t)} \right)^2 - \left( \frac{1}{T} \sum_{t=1}^{T} \log \frac{f_A(r_t)}{f_B(r_t)} \right)^2 \]  

(4.4)

is a consistent estimator for the variance of the likelihood ratio statistic. The SVA model only accounts for correlation between the shocks and universally produce higher log-likelihoods than the exclusively threshold models. In the full sample period we can reject both RTSV and TSV in favor of the simple SVA model: the Vuong (1989) test statistics are respectively 4.0287 (\( p \)-value < 0.0001) and 2.7629 (\( p \)-value = 0.0057). In the recent post-1989 sample we can reject the restricted threshold model RTSV in favor of SVA as the Vuong (1989) test statistic is 3.3001 (\( p \)-value < 0.0001), though there is insufficient evidence to statistically distinguish the full TSV and SVA models as the value of the test statistic is only 1.3781 (\( p \)-value = 0.1682).

To this point we have only allowed for very restrictive threshold effects and we relax these restrictions to allow both volatility persistence (i.e., \( \beta \)) and the variance of conditional volatility (i.e., \( \sigma^2_\eta \)) to vary with the sign of lagged returns. We initially estimate a model (denoted TSV) that allows all three volatility parameters to change and find that this augmented model is a dramatic improvement over the restricted threshold model RTSV, producing log-likelihoods that are about 10 higher in both samples. However, despite the fact that the TSV model has two parameters more than SVA, it still produces a lower log-likelihood.

Our most comprehensive model (TSVA) nests both the general threshold effects and correlated models. It fits the data better than all special cases, improving the log-likelihood of the heretofore best model (SVA) by more than 23 in the full sample and 11 in the post-1989 sample. We find that volatility is more persistent (i.e., \( \beta_1 > 1 \)) and less volatile (i.e., \( \sigma^4_{\nu} < 0 \)) following negative
returns. Interestingly, we find that the negative correlation between returns and volatility is higher in the asymmetric threshold model and the intercept is higher following negative returns (i.e., $\omega_1 > 0$). These results hold in both sub-samples and suggests that the RTSVA model is misspecified. They also suggest that threshold effects and correlation are complementary.

To compare the models we use the Akaike and Bayesian Information Criteria: $AIC = 2k - 2 \log L(\theta; y)$ and $BIC = k \log(T) - 2 \log L(\theta; y)$ where $k$ is the number of parameters in the model. A model is preferred when either criteria is minimized. Both criteria penalize models for having many parameters and the BIC penalizes models more heavily than does the AIC. In both sub-periods the AIC suggests that the full TSVA model is the best, followed by the simple SVA model. The BIC also suggests the full model is best in the full 1962-2004 sample period, though the BIC favors the more parsimonious SVA model for the recent post-1989 sample (the TSVA model is the next lowest BIC). Overall, the evidence suggests that the TSVA model fits the data the best.

4.1 Misspecification Tests

An important part of building any econometric model is specification testing. In particular we are interested in testing whether a particular stochastic volatility model is able to fully capture the effect of past returns, and especially negative returns, on conditional volatility. In their sign and negative size and bias tests Engle and Ng (1993) test whether future squared standardized residuals from a GARCH model are uncorrelated with the sign of today’s return, i.e. $s_t = 1_{(r_t<0)}$, and the size of the lagged negative return, i.e. $s_t \cdot r_t$. Unfortunately these tests cannot be directly applied in our context because unlike the GARCH model volatility is unobserved in the stochastic volatility model. Our solution to this problem is to use the probability integral transformation or conditional cdf (see, e.g., Rosenblatt, 1952) to construct a pseudo-standardized residual following Berkowitz (2001) (see also Smith (2007) for an application to stochastic volatility models) that will be iid $N(0, 1)$ if the model is correctly specified. Berkowitz (2001, Theorem 2) demonstrates that inaccuracies in the conditional density of the true returns are preserved in the transformed data, so diagnostic tests based on the pseudo-standardized residual are useful to detect deficiencies in the underlying model.

If a particular stochastic volatility model is correctly specified then the model’s probability integral transformation or cdf, i.e. $F_t(r_{t+1})$, will be an independent uniform random variable. Using this we can deduce that the associated quantile of a normal random variable $Z_{t+1} = \Phi^{-1}(F_t(r_{t+1}))$ will be independent standard normal random variables and we can thus think of
Z_{t+1} as a pseudo-standardized residual. In all cases the probability integral transform of \( r_{t+1} \) can be written as

\[
F_t(r_{t+1}) = \int_{-\infty}^{r_{t+1}} \int f(r|x_t = x) f(x_t = x|I_t) dx \, dr \\
\approx \sum_{i=1}^{N} w_i F(r_{t+1}|x_i) f(x_t = x_i|I_t).
\] (4.5)

Our goal is now to test whether \( Z_{t+1} \) is orthogonal to past negative returns, which will indicate that the volatility model adequately captures the information in the sign and magnitude of lagged returns. In particular, motivated by the sign and negative size bias test of Engle and Ng (1993), we test if \( Z_{t+1}^2 \) is uncorrelated with \( s_t \) and \( r_t \cdot s_t \). To compute the test we follow Breunig, Najarian, and Pagan (2003), who consider conducting general specification tests based on the moments \( m(Z_{t+1}; \theta) \) where the null hypothesis of correct specification implies that \( E(m_t) = \tau_0 \). The moment conditions we consider are

\[
m(Z_{t+1}; \theta) = \begin{pmatrix} Z_{t+1}^2 \cdot s_t \\ Z_{t+1}^2 \cdot s_t \cdot r_t \end{pmatrix}.
\] (4.6)

The specification test is based on the sample moments (using a consistent estimator of the parameter vector denoted \( \hat{\theta} \))

\[
\hat{\tau} = \frac{1}{T} \sum_{t=0}^{T-1} m(Z_{t+1}; \hat{\theta}),
\] (4.7)

which under suitable regulatory conditions follows an asymptotic normal distribution

\[
T^{1/2}(\hat{\tau} - \tau_0) \xrightarrow{D} N(0, V_{\tau})
\] (4.8)

with covariance matrix

\[
V_{\tau} = V_{mm} - M_{\theta} I_{\theta\theta}^{-1} M_{\theta}
\] (4.9)

where \( V_{mm} = \lim_{T \to \infty} E\left( \frac{1}{T} \sum_{t=1}^{T} m(r_t; \theta)m(r_t; \theta)' \right) \) is the asymptotic covariance matrix of the moment conditions, \( I_{\theta\theta} = \lim_{T \to \infty} E\left( -\frac{1}{T} \frac{\partial^2}{\partial \theta \partial \theta'} \right) \), and \( M_{\theta} = \lim_{T \to \infty} E\left( \frac{\partial m(r_t; \theta)}{\partial \theta} \right) \) is the asymptotic Jacobian of the moment conditions. The asymptotic variance-covariance matrix of the moment conditions accounts for estimation error in the parameter vector \( \theta \) (see also Chen, 2007).

The point estimates of Wald-based tests for the joint hypothesis, and the sign and negative size tests are reported in the lower part of Tables 1 and

http://www.bepress.com/snde/vol13/iss3/art1
2. Consider first the full-sample results in Table 1. Perhaps the most striking result in these tables is the fact that the p-value for the joint test decreases quite dramatically after the introduction of correlation between returns and volatility, even after accounting for threshold effects. The model not rejected as misspecified using the joint test in the recent sample is the RTSVA model, which incidentally is also the model favored by both the AIC and BIC. This model also passes the univariate sign and size bias tests. We note that the p-value also increases after the introduction of correlation in the full-sample results, though none of the models pass the joint diagnostic tests.

5 Extentions

5.1 Monte Carlo Experiment

The finite-sample performance of the numerical integration-based estimates of the plain vanilla stochastic volatility model has been studied using Monte Carlo experiments by both Fridman and Harris (1998) and Watanabe (1999). The method performs roughly as well as any of the other relatively efficient methods, i.e. Bayesian MCMC (Jacquier, Polson, and Rossi, 1994), Efficient Method of Moments (Andersen et al., 1999), and Monte Carlo maximum likelihood (Sandmann and Koopman, 1998). We extend this simulation-based evidence to models that allow for asymmetric relationship between returns and volatility in correlated and threshold SV models. In particular we generate returns from a lognormal AR(1) stochastic volatility model with zero mean

\[ r_t = \exp(x_t/2)z_t \]

where \( z_t \sim N(0, 1) \), and conditional log-volatility is modeled in the base simulation experiment by

\[ x_{t+1} = -0.1370 + 0.9860 \cdot x_t + 0.1513 \cdot v_t, \]

where \( v_t \sim N(0, 1) \). In the correlated SVA model log-volatility is generated as

\[ x_{t+1} = -0.1370 + 0.9860 \cdot x_t + 0.1513 \cdot (-0.5670 \cdot z_t + 0.8237 \cdot v_t). \]

Finally, in the RTSV model we simulate log-volatility as

\[ x_{t+1} = -0.2617 + 0.1575 \cdot s_t + 0.9810 \cdot x_t + 0.1425 \cdot v_t. \]

We use \( N = 50 \) nodes in our empirical applications and the Monte Carlo study. Watanabe (1999) documents an upward bias in \( \sigma_\nu \) when \( N = 25 \)
Table 3: Monte Carlo Study of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>$\omega_0$</th>
<th>$\phi$</th>
<th>$\sigma^2_\nu$</th>
<th>$\rho$</th>
<th>$\omega_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: SV Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>-0.1370</td>
<td>0.9860</td>
<td>0.1513</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ave. NML</td>
<td>-0.2056</td>
<td>0.9790</td>
<td>0.1624</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE NML</td>
<td>0.1252</td>
<td>0.0126</td>
<td>0.0281</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ave. SE</td>
<td>0.0871</td>
<td>0.0089</td>
<td>0.0283</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel B: SVA Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>-0.1370</td>
<td>0.9860</td>
<td>0.1513</td>
<td>-0.5670</td>
<td>0</td>
</tr>
<tr>
<td>Ave. NML</td>
<td>-0.1957</td>
<td>0.9803</td>
<td>0.1549</td>
<td>-0.4648</td>
<td></td>
</tr>
<tr>
<td>RMSE NML</td>
<td>0.1075</td>
<td>0.0108</td>
<td>0.0255</td>
<td>0.1676</td>
<td></td>
</tr>
<tr>
<td>Ave. SE</td>
<td>0.0673</td>
<td>0.0069</td>
<td>0.0249</td>
<td>0.1287</td>
<td></td>
</tr>
<tr>
<td><strong>Panel C: RTSV Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>-0.2617</td>
<td>0.9810</td>
<td>0.0203</td>
<td>0</td>
<td>0.1575</td>
</tr>
<tr>
<td>Ave.</td>
<td>-0.3816</td>
<td>0.9731</td>
<td>0.0262</td>
<td>0.2359</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1541</td>
<td>0.0113</td>
<td>0.0353</td>
<td>0.1407</td>
<td></td>
</tr>
<tr>
<td>Ave. SE</td>
<td>0.0616</td>
<td>0.0042</td>
<td>0.0117</td>
<td>0.0715</td>
<td></td>
</tr>
</tbody>
</table>

NOTES: We generate 500 artificial sample paths with $T = 1000$ time-series observations from three conditionally normal SV models: a symmetric model, a model where shocks to returns are correlated with future volatility, and a purely threshold model. Each model is a special case of

$$
    r_t = \exp \left( x_t / 2 \right) z_t \\
    x_{t+1} = \omega_0 + \omega_1 \cdot s_t + \beta \cdot x_t + \sigma_\nu \cdot \left( \rho \cdot z_t + \sqrt{1 - \rho^2} \cdot \nu_t \right),
$$

where $z_t \sim N(0, 1)$ and $\nu_t \sim N(0, 1)$ are independent shocks. For each model we report the true parameter values followed by: 1) the mean parameter estimates from the exact numerical integration-based filter (averaged across simulations), 2) the root-mean squared error (RMSE) $\sqrt{\sum_{i=1}^{500} (\theta_{NML}^{(i)} - \theta_{true})'(\theta_{NML}^{(i)} - \theta_{true})}$, 3) the average standard error.

that is largely eliminated by increasing the number of nodes to $N = 50$. We find similar results in experiments that we omit to save space. We report the results of these experiments in Table 3. In each panel we report the true parameter estimates, followed by the average point estimate across 500 simulations, the root mean squared error of the sample estimates, and the average standard error. We find that all three models perform quite well, though there is a slight upward bias in the estimated size of the threshold effect, and a small downward bias in the size of the correlation coefficient.
We also study the ability of the Vuong (1989) non-nested likelihood ratio test to differentiate between the SVA and RTSV models. We use the data simulated from both the SVA and RTSV models and compute the LR test comparing the two. We plot a nonparametric Kernel estimate of the density across the 500 simulations in Figure 4. The model clearly has some power to differentiate between the models. Using a one-tailed 5 percent critical value of 1.65 we find evidence in favor of the RTSV model 16.6 percent of the time when it is true, but only 0.4 percent of the time when the SVA model is true. When the SVA model is true we find evidence in favor of it 26.6 percent of the time, but in only 1.2 percent of simulations when the RTSV model is true. This is quite impressive given that we are only simulating $T = 1000$ observations and the power will likely improve as the sample size increases.

5.2 Fat Tails

It is well-known that returns from a conditionally normal SV model will exhibit fat tails relative to a normal distribution, but this effect is insufficient to explain the excess kurtosis that is observed in financial data. A number of authors have found that heavy tailed standardized distributions are required to adequately model stock returns, including Harvey et al. (1994), Fridman and Harris (1998), Watanabe (1999), Chib, Nardari, and Shephard (2002), and Jacquier et al. (2004). As both Fridman and Harris (1998) and Watanabe (1999) show, it is very easy to extend the basic stochastic volatility model to account for fat tails in the numerical integration-based stochastic volatility model. All one needs to do is modify the marginal density of returns conditional on volatility $f(r_{t+1}|x_{t+1} = x_t)$, say using a Student’s $t$ distribution in place of the Gaussian density in equation (3.8). To account for both correlation and fat tails we must work a little harder. Following Smith (2007) we solve this problem by constructing the transition density $f(r_t, x_{t+1}|x_t)$ used in equation (3.16) in Step 1 of the correlated model outlined in Section 3.2 in a flexible way. In particular we use Copula theory to specify the marginal distribution of stock returns using the heavy tailed Student $t$ distribution and maintain the log-normal distribution for volatility and then choose a Copula that captures the dependencies or correlation between them. A copula is a function that stitches together two or more marginal distributions (Sklar, 1959): $F_{XY}(x, y) = C(F_X(x), F_Y(y))$, and $f_{XY}(x, y) = f_X(x)f_Y(y)c(F_X(x), F_Y(y))$ ($f(\cdot)$ denotes the pdf and $F(\cdot)$ denotes the cdf) and when the marginal dis-

---

15 Jacquier et al. (2004) solve the problem using a Bayesian approach, but allow returns and volatility to be contemporaneously correlated. Yu (2005) notes that allowing for correlation between returns and future volatility is preferable both theoretically and empirically.
Figure 4: Plot of density for Vuong’s (1989) non-nested likelihood ratio test.

NOTES: The nonparametric Kernel density estimate of the Vuong nonnested likelihood ratio test comparing the SVA and RTSV models:

\[ T^{-1/2} \frac{L_{SVA} - L_{RTSV}}{\hat{\omega}_T}, \]

where

\[ \hat{\omega}_T^2 = \frac{1}{T} \sum_{t=1}^{T} \left[ \log \frac{f_{SVA}(r_t)}{f_{RTSV}(r_t)} \right]^2 - \left[ \frac{1}{T} \sum_{t=1}^{T} \log \frac{f_{SVA}(r_t)}{f_{RTSV}(r_t)} \right]^2 \]

when the SVA model is true (the blue line), and when the RTSV model is true (the red line). Data is generated using the true parameters listed in Panels B and C of Table 3.

distributions are continuous the copula is simply a \textit{cdf} with uniform marginals. Copulas are particularly useful in this application as we have complete flexibility to choose any distribution for the marginal distribution of returns and
then handle the correlation independently. Closed-form densities typically place strong restrictions on the relationship between the marginal distributions. For example in the multivariate t distribution all marginal distributions share the same degrees of freedom, which may not always be appropriate. By modeling the joint density using copulas we have complete flexibility when choosing the shapes of the marginal distributions. We model the marginal distribution of stock returns using a Student t random variable and stitch this together with the conditionally normal distribution of log-volatility using the Gaussian copula:

$$f(r_t, x_{t+1} | x_t) = t_{\nu} \left(r_t - \mu_{t-1} - \phi_{s_{t-1}} r_{t-1}; \exp(x_t)\right) \cdot \phi(x_{t+1}; \omega, \beta, \sigma^2) \cdot c_N(\eta_t, \zeta_t; \rho)$$  \hspace{1cm} (5.1)

where $t_{\nu}(\cdot; b)$ denotes the pdf of a scaled Student’s t distribution with $\lambda$ degrees of freedom and variance $b$, $\eta_t = T_{\lambda} \left((r_{t+1} - \mu_t - \phi_{s_t} r_t) / \exp(x_t/2)\right)$ denotes the marginal cdf of stock returns conditional on $x_t$, $T_{\lambda}(\cdot)$ denotes the cdf of a standardized Student t distribution with $\lambda$ degrees of freedom, $\zeta_t = \Phi \left((x_{t+1} - \omega_t - \beta x_t) / \sigma_{t, \nu}\right)$ denotes the marginal cdf of future volatility conditional on $x_t$, $\Phi(\cdot)$ the cdf of a standard normal random variable, and $c_N(\eta_t, \zeta_t; \rho)$ is the Gaussian copula with correlation coefficient $\rho$, which can be expressed as:

$$c_N(\eta_{t+1}, \zeta_{t+1}; \rho) = \frac{1}{\sqrt{1-\rho^2}} \cdot \exp \left\{ -\frac{\Phi^{-1}(\eta_{t+1})^2 + \Phi^{-1}(\zeta_{t+1})^2 - 2\rho \Phi^{-1}(\eta_{t+1}) \Phi^{-1}(\zeta_{t+1})}{2(1-\rho^2)} \right\}. \hspace{1cm} (5.2)$$

We choose this specification because it nests the jointly normal density used earlier in this paper in the limit as $\lambda \to \infty$. The algorithm is then implemented by replacing equation (3.16) in Step 1 of the correlated model outlined in Section 3.2 with

$$f(r_{t+1}, x_{t+1} = x_t, x_{t+2} = x_{j+1} | I_t) = t_{\nu} \left(r_{t+1} - \mu_t - \phi_{s_t} r_t; \exp(x_t/2)\right) \cdot \phi(x_j; \omega, \beta, \sigma^2) \cdot c_N(\eta^{(i,j)}_{t+1}, \zeta^{(i,j)}_{t+1}; \rho), \text{ for } i, j = 1, \ldots, N \hspace{1cm} (5.3)$$

with $\eta^{(i,j)}_{t+1} = T_{\lambda} \left((r_{t+1} - \mu_t - \phi_{s_t} r_t) / \exp(x_t/2)\right)$, and $\zeta^{(i,j)}_{t+1} = \Phi \left(x_j - \omega_t - \beta x_t / \sigma_{t, \nu}\right)$.

In Table 4 we present the parameter estimates for the fat-tailed models based on the Student t distribution. The behavior of volatility and the point estimate of the correlation between returns and future volatility is practically unchanged after allowing for fat tails. Interestingly the importance of threshold effects after accounting for correlation is greatly reduced. The log-
likelihood improves by only around 0.3 after allowing for all three volatility parameters to change with the sign of prior returns.\textsuperscript{16}

\section{Conclusions}

In this paper we have compared the ability of correlation between returns and volatility and threshold effects to capture the asymmetric relationship between returns and volatility. We show how to estimate the model’s parameters by maximum likelihood using a nonlinear filter for latent stochastic volatility models that allows for both correlation and threshold effects and uses numerical integration. We find that correlation alone does a better job than threshold alone, generating a higher log-likelihood even though it has fewer parameters. However, a model which includes both correlation and threshold dominates both restricted models. In this augmented model we find negative correlation between returns and future volatility. We also find that volatility is higher following negative returns even after accounting for the negative correlation and that the variance of volatility is lower and volatility is more persistent following negative returns than following positive returns. We compare correlation with threshold effects using information criterial and non-nested hypothesis tests and conclude that the uncorrelated models are misspecified and are dominated by models that allow returns to be correlated with future volatility, though threshold effects are still important. Finally, we undertake a range of robustness checks of the model and find that our results are remarkably similar when we estimate the model by Bayesian-based MCMC algorithm or allow for fat tails in the marginal distribution of returns. We study the finite sample behavior of our estimator using a Monte Carlo study and find that it performs quite well.

\section*{Appendix}

\section{Markov Chain Monte Carlo-Based Bayesian Estimation}

Jacquier et al. (1994) demonstrate how stochastic volatility models may be estimated by Markov Chain Monte Carlo (MCMC) methods. They augment the parameter space with the conditional volatilities whose distribution is

\textsuperscript{16}The fat-tailed asymmetric restricted threshold model RTSVA-T converged to the same parameter estimate as the model without threshold effects SVA-T, so we only report five models.
Table 4: Parameter Estimates for TSV Model With Student $t$ Marginal Density: 1990-2004

<table>
<thead>
<tr>
<th></th>
<th>SV-T</th>
<th>SVA-T</th>
<th>RTSV-T</th>
<th>TSV-T</th>
<th>TSVA-T</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.0007</td>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.0649</td>
<td>0.0747</td>
<td>0.0668</td>
<td>0.0698</td>
<td>0.0576</td>
</tr>
<tr>
<td></td>
<td>(0.0160)</td>
<td>(0.0160)</td>
<td>(0.0160)</td>
<td>(0.0161)</td>
<td>(0.0153)</td>
</tr>
<tr>
<td>$\omega_0$</td>
<td>-0.0731</td>
<td>-0.1665</td>
<td>-0.3746</td>
<td>-0.4710</td>
<td>-0.1230</td>
</tr>
<tr>
<td></td>
<td>(0.0167)</td>
<td>(0.0175)</td>
<td>(0.0129)</td>
<td>(0.0093)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.9925</td>
<td>0.9828</td>
<td>0.9756</td>
<td>0.9721</td>
<td>0.9876</td>
</tr>
<tr>
<td></td>
<td>(0.0017)</td>
<td>(0.0018)</td>
<td>(0.0003)</td>
<td>(0.0002)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>$\sigma_{0,\nu}^2$</td>
<td>0.0181</td>
<td>0.0247</td>
<td>0.0300</td>
<td>0.0097</td>
<td>0.0171</td>
</tr>
<tr>
<td></td>
<td>(0.0063)</td>
<td>(0.0047)</td>
<td>(0.0040)</td>
<td>(0.0010)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.2874</td>
<td>0.3970</td>
<td>-0.0149</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0268)</td>
<td>(0.0232)</td>
<td>(0.0007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.0006</td>
<td>-0.0021</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0016)</td>
<td>(0.0001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{1,\nu}^2$</td>
<td>0.0338</td>
<td>0.0068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0049)</td>
<td>(0.0009)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.0928</td>
<td>0.0575</td>
<td>0.0770</td>
<td>0.0732</td>
<td>0.0532</td>
</tr>
<tr>
<td></td>
<td>(0.0207)</td>
<td>(0.0206)</td>
<td>(0.0211)</td>
<td>(0.0177)</td>
<td>(0.0225)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.6840</td>
<td>-0.6262</td>
<td></td>
<td>-0.6262</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0439)</td>
<td>(0.0349)</td>
<td></td>
<td>(0.0349)</td>
<td></td>
</tr>
</tbody>
</table>

NOTES: We present the parameter estimates and standard errors (computed using a numerically evaluated Hessian matrix) in parenthesis for a range of stochastic volatility models, which are all special cases of

\begin{equation*}
\begin{aligned}
    r_t &= \mu + \phi r_{t-1} + \exp(x_t/2)u_{1,t} \\
    x_{t+1} &= \omega_0 + s_t \cdot \omega_1 + (\beta_0 + s_t \cdot \beta_1) \cdot x_t + \sqrt{\sigma_{0,\nu}^2 + s_t \cdot \sigma_{1,\nu}^2} \left( \rho u_{1,t} + \sqrt{1-\rho^2} u_{2,t} \right).
\end{aligned}
\end{equation*}

We also report the log-likelihood, AIC and BIC for each model. The data spans the period January 1, 1990 to December 31, 2004.
simulated along with the structural parameters by the MCMC algorithm. They exploit the hierarchical structure of the system by writing

\[ x_t | x_{t-1}, \omega, \phi, \sigma^2_v \sim N(\omega + \phi x_{t-1}, \sigma^2_v) \] (A.1)

\[ r_t | x_t, \mu \sim N(\mu, e^{x_t}). \] (A.2)

Suitable specification of the prior distributions for the parameters allows the computationally intensive simulations from the posterior distribution such as the Gibbs sampler and Metropolis-Hastings algorithms. The model can be extended to allow returns and future volatility to be correlated \( \rho = \text{Correl}(e_t, \eta_t) \), which can factored as

\[ r_t = \mu + e^{x_t/2} \left( \rho z_{1,t} + \sqrt{1 - \rho^2} z_{2,t} \right) \] (A.3)

\[ x_{t+1} = \omega + \phi x_t + \sigma_v z_{1,t} \] (A.4)

where \( z_{1t} \) and \( z_{2t} \) are iid standard normal random variables. The hierarchical distribution of latent volatility and returns is given by

\[ x_{t+1} | x_t, \omega, \phi, \sigma^2_v \sim N(\omega + \phi x_t, \sigma^2_v) \] (A.5)

\[ r_t | x_{t+1}, x_t, \omega, \phi, \sigma^2_v, \rho, \mu \sim N \left( \mu + \rho e^{x_t/2} \left( \frac{x_{t+1} - \omega - \phi x_t}{\sigma_v} \right), e^{x_t} (1 - \rho^2) \right). \] (A.6)

Following Meyer and Yu (2000) and Yu (2005) we specify the prior distributions for the parameters as:

- \( \phi = 2\phi^* - 1 \) where \( \phi^* \sim \text{Beta}(3, 3) \) which has a mean of 0.5 and standard deviation of 0.19.
- \( \mu(1 - \phi) \sim N(0, 25) \), which is close to uninformative.
- \( \sigma^2_v \sim \text{Inverse-Gamma}(2.5, .025) \) which has mean 0.167 and standard deviation 0.024.
- \( \beta = 2\beta^* - 1 \) where \( \beta^* \sim \text{Beta}(20, 1.5) \) which has a mean of 0.93 and standard deviation of 0.054.
- \( \omega \sim (1 - \beta)N(0, 25) \), which is close to uninformative.
- \( \rho \sim U(-1, 1) \).

We specify the distribution of unconditional volatility

\[ x_1 \sim N \left( \frac{\omega}{1 - \beta}, \frac{\sigma^2_v}{1 - \beta^2} \right), \]

though the results are not sensitive to using uninformative priors.
References


