Estimation of Stochastic Volatility with High and Low Prices

Suk Joon Byun, Jung-Soon Hyun, Woon Jun Sung*

October, 2014

ABSTRACT

This paper suggests stochastic volatility models incorporating both the leverage effect and information on the daily high/low prices of stocks. The leverage effect is measured using open-to-close returns and two distinct intraday data, ranges, defined by the differences between daily high and low log-prices, and extreme prices in order to detect asymmetric volatility behavior. The likelihood-based inferences of Markov Chain Monte Carlo (MCMC) are conducted to estimate parameters and volatility. The simulation study reveals that the proposed model is superior to a traditional stochastic volatility model using returns only but there is little difference between estimators using ranges or high/low prices. Performing an empirical analysis using the E-mini S&P 500 and the Nasdaq 100 Futures, we find strong evidence of the leverage effect even when information of high/low prices is incorporated.

* Byun: KAIST Business School, 85 Hoegiro, Dongdaemun-gu, Seoul, 130-722, Korea (email: sjbyun@business.kaist.ac.kr); Hyun: KAIST Business School, 85 Hoegiro, Dongdaemun-gu, Seoul, 130-722, Korea (email: jshyun@business.kaist.ac.kr); Sung: KAIST Business School, 85 Hoegiro, Dongdaemun-gu, Seoul, 130-722, Korea (email: assmir@business.kaist.ac.kr).

1. Introduction

Volatility represents the variability of markets, and measuring variability is an important issue in finance. Modeling volatility should reflect the empirical characteristics of the variability such as the stochastic change of volatilities over time or the correlated movement between stock prices and volatilities. Heston (1993) suggests a continuous-time square-root stochastic volatility model, in which innovations to volatility are partially correlated with innovations to the price of stock. This model has been extended to several other models by two approaches. One approach adds jumps into the stock returns or the volatility process (See Bates (2000), Pan (2002), Eraker et al. (2003)). Eraker et al. (2003) find strong evidence for jumps in volatility as well as jumps in returns. Another approach supplements the Heston model by extending the number of factors in the model. Using a two-factor stochastic volatility model and time-varying weights of the factors, Christoffersen et al. (2009) enable modeling more flexible volatility structure. Unlike the affine extensions, Jacquier et al. (2004) and Yu (2005) propose non-affine lognormal autoregressive conditional volatility models which are not easy for being applied to option pricing but are adept at empirically fitting underlying stocks. Chernov et al. (2003) test affine stochastic volatility and logarithmic models. They show the necessity of extended models such as affine stochastic volatility models with jumps and a two-factor logarithmic model.

Including the aforementioned models, most models assume prices evolve in continuous time. Yet an empirical test is conducted using daily or weekly close-to-close prices. For instance, suppose a stock shows large movement during trading hours but the closing price is at a similar level of the previous day's closing price. Using only the close-to-close returns may not accurately measure the variability of the asset prices in continuous-time framework. To consider such sharp intraday rises or falls of prices into a model, Gallant et al. (1999) use close-to-close returns along with the range (calculated as the difference between the daily high and low log-prices). Alizadeh et al. (2002) also show range-based volatility proxies are more efficient than estimated volatility based on absolute daily returns and robust to market microstructure noise. Though high-frequency data are also available these days to explain prices' intraday movement, using high/low data has a couple of advantages. As Brandt and Diebold (2006) have pointed out, daily open, close, high and low prices (H/L prices hereafter) are not only widely available with lengthy time-series, but yield empirical results that are fairly robust independent of market microstructure noise such as bid-ask bounce and asynchronous trading. Horst et al. (2012) also recently suggest a stochastic volatility model including full opening, closing and H/L prices and demonstrate that the use of full information can improve estimation.

Until now, all models using H/L prices assume, to the best of our knowledge, independent innovations between stock returns and changes of its volatility, i.e., no correlation, perhaps due to the joint likelihood of H/L prices and close prices revealed hardly when prices and volatilities are correlated. However, a negative correlation is not only empirically observed in stock markets but theoretically meaningful since the negative correlation is well known as the leverage effect in finance. According to Black (1976), when stock prices decrease, the company's debt to equity ratio becomes high and then investors will consider the company to be at higher risk of default leading eventually to increases in the stock's volatility. This leverage effect is a stylized fact observed in stock markets.

Without considering H/L prices, some studies suggest stochastic log-volatility models with the leverage effect. Allowing an inter-temporal dependence between disturbances of the processes to their basic stochastic volatility model with independent innovations (Jacquier et al. (1994)), Jacquier et al. (2004) estimate the model using the likelihood-based estimation method called Markov Chain Monte Carlo (MCMC). Yu (2005) compares the Jacquier et al. (2004) model with another discretized stochastic volatility model with contemporaneous dependence between disturbances of the processes. Yu (2005) asserts that his model can reflect the leverage effect better than Jacquier et al.'s (2004). There are also multi-factor stochastic volatility models with leverage effect such as Durham (2006).

In this context, this paper suggests a stochastic volatility model allowing correlation between returns and volatility processes but we use the additional information on H/L prices of assets. Our model is closely related to models of Alizadeh et al. (2002), Brandt and Jones (2005) and Horst et al. (2012) but is extended to include the leverage effect. It is the first try to deal with stochastic volatility models with leverage effect in the framework using extreme values. To reflect the information on extreme values, we apply the likelihood-based MCMC estimation method. The idea is an approximate decomposition of the entire likelihood into two parts applying the Bayes rule. The first part is the joint likelihood of closing, high, and low returns. The second is the joint likelihood of daily returns and volatilities.

We suggest two estimators to incorporate H/L prices to our model. The first one uses returns and ranges, and the other does opening, high, low and closing prices separately. To verify the performance, we compare our suggested methods with a return-based basic stochastic volatility model with correlation and H/L prices-based models without correlation. We conduct both simulation study and empirical analysis using the E-mini S&P 500 and the Nasdaq 100 Futures.

Through simulation study, first, the parameters estimated by our model are closer to the true parameters than those from the return-based basic model with correlation. Second, the difference between two models is shown by the root mean squared errors (RMSEs) of the estimated volatilities. RMSEs of the basic stochastic volatility model increase $1.54 - 1.76$ times of those of our model on average. Thirds, in the comparison with H/L prices-based model without correlation, it is interesting to see that the parameters estimated show minor difference even though there exists a significant correlation. However, the existence of correlation makes a major difference in volatility estimation. When the correlation is absolutely larger than -0.4, the RMSEs of our model are 4-10% smaller than those of the previous models such as Brandt and Jones (2005) and Horst et al. (2012) and the difference is statistically significant.

In empirical analysis with market data, we find negative correlation of around -0.3 - -0.7 between returns and changes of volatility for both the E-mini S&P 500 and the Nasdaq 100 futures. The existence of strong leverage effect suggests that the correlation must be incorporated in the model when high and low prices are considered. We also confirm that there is asymmetric information of H/L prices but the effect is meager.

The rest of this paper is organized as follows. In Section 2, we introduce the stochastic volatility models and explain how the information on H/L prices is incorporated into the model. In Section 3 we present the estimation method. In Section 4 and 5, the estimation results are given. First, in Section 4, we demonstrate our models through simulation as evidence. Next, in Section 5, we estimate the parameters of the stochastic volatility models from actual market data. Finally, Section 6 summarizes the contents of this paper and suggests directions for further research.

2. Stochastic Volatility Models

2.1. The basic model

Among the various stochastic volatility models in continuous-time economy, affine or log-volatility models are widely used. While people prefer affine models to log-volatility models from the perspective of pricing options, log-volatility models are often used for examining statistical behavior of an underlying asset. We also choose a log-volatility process in a continuous time economy. Let Y_t be the logarithm of an asset's price S_t and σ_t be volatility of the log-return at time t. We assume that the log-returns of stock prices evolve below:

$$\begin{aligned} dY_t &= \mu dt + \sigma_t dW_t^1 \\ d\ln\sigma_t &= \kappa(\theta - \ln\sigma_t)dt + \upsilon dW_t^2 \end{aligned} \tag{1}$$

where W_t^1 and W_t^2 are Wiener processes correlated with $< dW_t^1, dW_t^2 >= \rho dt$ for a constant ρ and μ is the instantaneous expected rate of the log return. If $\rho = 0$, it coincides with the models proposed by Alizadeh et al. (2002) and Horst et al. (2012). The log-volatility process of equation (1) is assumed that it follows a mean-reverting Ornstein–Uhlenbeck process. The parameter κ is a mean-reverting factor, θ is a long-term mean of the log-volatility, and υ is the volatility of stock returns' volatility.

For empirical practice, following Yu (2005), we assume a basic stochastic volatility model discretized as follows: For $t = 1,2,\cdots,T$, trading days

$$\begin{aligned} y_t &= \sigma_{t-1}\varepsilon_t^y \\ \ln\sigma_t &= \alpha + \delta\ln\sigma_{t-1} + \upsilon\varepsilon_t^\sigma \end{aligned} \tag{2}$$

where ε_t^y and ε_t^σ have a bivariate standard normal distribution with the correlation ρ. Here y_t becomes a daily log return and α equals $\kappa\theta$, and δ corresponds to $1-\kappa$ which represents the autocorrelation of log volatilities. If δ is positively large, autocorrelation is strong so that the persistency of volatility is also high. We name this discretized model as SV. It is worthy to note that the return at t is conditioned on the volatility estimated at the previous day $t-1$. We neglect the drift term in equation (2) because short-term data like daily prices are used, so that we focus on parameters of volatility and volatility itself. Horst et al. (2012) also support a driftless model given by equation (2) by estimating the drift of weekly log-prices near 0.

2.2. The model incorporating ranges and H/L prices

Daily opening, high, low and closing prices are publically known data. They are easily obtained via internet data providers without any cost. On October 17, 2008 S&P 500 index started with 942.29 and ended up with 940.55, so the daily percentage change was no more than 0.2%. However the H/L prices were 984.64 and 918.74 respectively, so that the range was almost 7%. It may be interesting to see how extreme prices such as high and low prices influence the parameters in equation (1). In order to incorporate intraday data into the model, the continuous model in equation (1) cannot be discretized such as equation (2). Instead we keep a continuous framework as follows: Under the same notation as in equation (2) and for $t = 1,2,\cdots,T$, trading days,

$$\begin{aligned} y_s &= \sigma_t\varepsilon_s^y, \qquad \text{for } t-1 < s \le t, \\ \ln\sigma_t &= \alpha + \delta\ln\sigma_{t-1} + \upsilon\varepsilon_t^\sigma. \end{aligned} \tag{3}$$

Note that unlike the basic model, returns continuously move with constant conditional volatility

σ_t over intraday $(t-1,t]$ but volatility is approximately discretized. Since volatility is constant during a day, the correlation measuring leverage effect is not taken into account over intraday but it is considered over inter-day via the correlation ρ at day t where the two innovations ε_t^y and ε_t^σ follow a bivariate standard normal. Two discretized models in equation (2) and equation (3) are slightly different in the conditional volatility affecting the return. Yu (2005) argues that the contemporaneous dependence between the two disturbances of equation (2) makes y_t a martingale difference whereas y_t in equation (3) is not.[1]. Thus the model of equation (3) is not consistent in the efficient market hypothesis. The reason that we mix discretization of volatility with continuous movement of returns is to incorporate both information on extreme values and the leverage effect simultaneously. This will be explained in more detail in the next section.

3. Estimation Methodology

In this section we embody the model given by equation (3) with intraday data through two different rules. In the first rule, we exploit ranges to supplement daily returns, precisely open-to-close returns. Since range is defined by the value of subtraction of the lowest log-price from the highest log-price, it does not distinguish each level of the extreme values. The model of equation (3) with the estimator incorporating returns and ranges is denoted by RR (open-to-close returns and ranges). This RR estimator is the correlated proxy corresponding to the model of Alizadeh et al. (2002) which uses ranges only. Another rule is to reflect levels of extreme prices separately. The RHL (open-to-close returns, open-to-high and open-to-low returns) is named by the estimator which is the correlated counterpart corresponding to the CHLO model of Horst et al. (2012) which demonstrate information on levels of extreme values improves the estimation of volatility, although the difference is small. We test whether the same phenomena is observed in a correlated case. Within the frame of information, using separate returns rather than ranges seems more useful but has its pros and cons. Ranges only use high and low prices which occur mainly during consecutive trades, so these prices may be considered as the values from a theoretical continuous-time series. On the contrary, the RHL exploits both the H/L and opening prices but opening prices may be easily influenced by market microstructure due to

[1] Instead of the log-volatility process, Yu (2005) assumes the log-variance process. In the process, the conditional volatility of a return is just a square root of updated variance, so the difference of the two processes is little.

trading mechanisms of stock markets.[2]

3.1 The RHL estimator

We first explain the RHL estimator. To exploit information on the high and low prices during trading hours, we define a daily return y_t as $\log S_t - \log S_{t-1}$. Similarly, we define the maximum and the minimum returns as $H_t = \max_{t-1<s\leq t} y_s$ and $L_t = \min_{t-1<s\leq t} y_s$, respectively. The joint density of the maximum, the minimum conditioned on the return and the volatility can all be derived from the results of Feller (1951), Freedman (1971), and Klebaner (2005). Since the return y_t is a driftless Brownian Motion with a constant volatility σ_t over the interval $(t-1, t]$, the joint density of H_t and L_t is

$$p(H_t \in db,\ L_t \in da |\ y_t = y, \sigma_t)$$

$$= \frac{1}{\sigma_t^2} \frac{1}{\phi\left(\frac{y}{\sigma_t}\right)} \sum_{n=-\infty}^{\infty} \left[4n^2 \left(\frac{(2n(b-a)-y)^2}{\sigma_t^2} - 1 \right) \phi\left(\frac{2n(b-a)-y}{\sigma_t} \right) \right. \\ \left. -4n(n-1) \left(\frac{(2n(b-a)+y-2b)^2}{\sigma_t^2} - 1 \right) \phi\left(\frac{2n(b-a)+y-2b}{\sigma_t} \right) \right] \quad (4)$$

where $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$. The joint density of equation (4) involves a calculation of an infinite sum but the sum quickly converges to a small number (see Choi and Roh (2013)).

In order to incorporate extreme prices with returns that are correlated with volatilities, we decompose the likelihood of return data into two terms using Bayes rule:

$$\begin{aligned} P(y, H, L|\Theta, \sigma) &= P(H, L|y, \Theta, \sigma) P(y|\Theta, \sigma) \\ &\approx P\big(H, L|\mathrm{y}, \Theta, \sigma, (-\rho)\big) P(y|\Theta, \sigma) \end{aligned} \quad (5)$$

where y, H and L are vectors of time series data of open-to-close, open-to-high and open-to-low returns. Θ represents a vector of model parameters $\{\alpha, \delta, \upsilon\}$ and σ is a vector of a time series of

[2] Amihud and Mendelson (1987) show that open-to-open returns exhibit greater dispersion and non-normality by the trading mechanism. There are studies that opening or closing prices do not represent appropriate stock values but are affected by market structures (e.g., Harris (1989), Stoll and Whaley (1990)), and those prices may be noisy.

volatility. The first term, $P(H, L|y, \Theta, \sigma)$, is the joint density of high and low returns conditioned on returns, volatility and parameters of the stochastic volatility model. Due to this term, information on the high and low prices can be incorporated when the model parameters are estimated. The second term, $P(y|\Theta, \sigma)$, is the likelihood of open-to-close returns given volatility and parameters. In equation (3), the model assumes the daily correlation between the return process and the volatility process, so the leverage effect can be measured from the second term as in the case of general stochastic volatility models. The distribution, $P(y|\Theta, \sigma)$, can be calculated exactly from return data, but in this model we approximate $P(H, L|y, \Theta, \sigma)$ to $P\big(H, L|y, \Theta, \sigma, (-\rho)\big)$. As in the distribution of equation (4), the joint density is conditioned on constant volatility, and correlation is not considered in the distribution of extreme values. Thus, the symbol $(-\rho)$ means that there is no correlation. If innovations are uncorrelated, equation (5) holds exactly.

Since we measure the leverage effect only from the $P(y|\Theta, \sigma)$ term, correlation may be underestimated. Nonetheless we may argue that the inter-temporal dependence between the return process and the volatility process in equation (3) gauges the leverage effect appropriately. Suppose volatility increases from σ_{t-1} to σ_t, i.e., $\sigma_{t-1} < \sigma_t$. Then the return y_t is more likely to decrease but the increase in σ_t can cause a large gap between the maximum and the minimum returns, which means that returns and ranges are negatively correlated. For the case of the contmeporaneous dependence like equation (2), the gap $H_t - L_t$ tends to have a small value because the likelihood of returns and extreme returns are affected by the lower value of σ_{t-1} instead of σ_t. The gap, which is defined as a range, can be considered as a type of variability measure, so that the negative correlation between returns and changes of ranges represents a type of leverage effect and can be measured from equation (3).[3]

3.2 The RR estimator

Instead of using the level of extreme values we use ranges, differences of maximum and minimum returns such as $H_t - L_t$ to exploit the symmetric information of high and low price levels. Our RR estimator is in a line with Brandt and Jones (2005) in reflecting both returns and ranges in the frame of stochastic volatility models, but differ by taking into consideration the correlation that can measure the leverage effect. By a change of variables, the density of range conditioned on constant

[3] We tested the contemporaneous dependence model, but leverage effect was considerably underestimated when high and low prices are used.

volatility and an absolute value of return can be derived from equation (4):

$$p((H_t - L_t) \in dR\ |\sigma_t, |y_t| = |y|)$$

$$= \frac{1}{\sigma_t^2 \phi\left(\frac{|y|}{\sigma_t}\right)} \sum_{n=-\infty}^{\infty} \left[4n^2 \left(\frac{(2nR - |y|)^2}{\sigma_t^2} - 1 \right) \phi\left(\frac{2nR - |y|}{\sigma_t} \right) (R - |y|) \right.$$
$$-2n(n-1)(2(n-1)R + |y|)\phi\left(\frac{2(n-1)R + |y|}{\sigma_t} \right)$$
$$\left. +2n(n-1)(2nR - |y|)\phi\left(\frac{2nR - |y|}{\sigma_t} \right) \right]. \tag{6}$$

Again we use Bayes rule to incorporate both range information and leverage effect

$$\begin{aligned} P(y, R|\Theta, \sigma) &= P(R|y, \Theta, \sigma)P(y|\Theta, \sigma) \\ &\approx P\big(R|y, \Theta, \sigma, (-\rho)\big)P(y|\Theta, \sigma) \end{aligned} \tag{7}$$

where R is a vector of a time series of ranges and other notations are the same with those of equation (5). From the first term, $P(R|y, \Theta, \sigma)$, we can exploit the information on ranges by using the density in equation (6). Correlation can be taken into account from the second term, $P(y|\Theta, \sigma)$. Like the case of using price levels, the likelihood $P(R|y, \Theta, \sigma)$ is approximated to $P\big(R|y, \Theta, \sigma, (-\rho)\big)$ because piecewise constant volatility is assumed.

3.3 MCMC Method

The Markov Chain Monte Carlo (MCMC) method is an exact likelihood based inference and highly efficient making it widely used to estimate model parameters and unobservable variables such as volatility (see Eraker et al. (2003), Eraker (2004)). In addition to estimating latent variables, MCMC provides estimation risk and is a useful tool for estimation under complex distribution. Jacquier et al. (1994) insist MCMC outperforms GMM and QMLE in the estimation of stochastic volatility models. Johannes and Polson (2002) also document an overview of MCMC methodology in a continuous-time framework.

The MCMC draws samples from each conditional posterior distribution which is factored into likelihood and prior by the Bayes rule. In the case of the stochastic volatility model without considering extreme values, the likelihood of $P(y|\Theta,\sigma)$ is well known, so that sampling is routine. This is not true for the RR and RHL estimators, specifically when the return and volatility are correlated. Bayes rule makes the posteriors factored below:

$$RHL: P\left(\Theta_i|\Theta_{(-i)}, y, H, L, \sigma\right) \propto P(y, H, L|\Theta, \sigma)P(\sigma|\Theta)P(\Theta_i)$$
$$RR: P\left(\Theta_i|\Theta_{(-i)}, y, R, \sigma\right) \propto P(y, R|\Theta, \sigma)P(\sigma|\Theta)P(\Theta_i) \tag{8}$$

where Θ_i represents each parameter of the volatility process and $\Theta_{(-i)}$ indicates the parameter set except the parameter Θ_i. The likelihoods, $P(y, H, L|\Theta, \sigma)$ and $P(y, R|\Theta, \sigma)$, are calculated from the approximated likelihoods of equation (5) and equation (7).

In the aspect of estimation, using the approximated likelihoods is advantageous. Note that the joint densities given in equation (4) and equation (6) are not dependent on any parameters except for volatility. Therefore the Gibbs sampler can be used for sampling the parameters α, κ, and υ. That is, these parameters are sampled directly from known distributions whose conjugate priors are given by

$$(\alpha, \kappa) \sim BVN\left(0, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \text{ and } \upsilon^2 \sim IG(2.5, 0.1)$$

where BVN and IG refer to a bivariate normal and an inverse gamma distribution, respectively. In order to sample the parameter ρ, indicating the leverage effect, we use an independent Metropolis-Hastings algorithm. The proposal density we use in the Metropolis-Hastings algorithm is $U(-1,1)$ where U represents a uniform distribution.

As for volatility estimation, it is well-known that the conditional posterior of volatility is not recognizable, so the Gibbs sampler is no longer applicable. In its place, a random walk Metropolis-Hastings algorithm is commonly used. We sample volatilities using the same algorithm for the RR and RHL estimators. The conditional posteriors are calculated as follows

$$
\begin{aligned}
RHL&: P(\sigma_t|\Theta, \sigma_{t-1}, \sigma_{t+1}, y_t, H_t, L_t, y_{t+1}, H_{t+1}, L_{t+1}) \\
&\propto P(y_{t+1}, H_{t+1}, L_{t+1}|\Theta, \sigma_t, \sigma_{t+1}) P(y_t, H_t, L_t|\Theta, \sigma_t, \sigma_{t-1}) \\
&\times P(\sigma_{t+1}|\Theta, \sigma_t) P(\sigma_t|\Theta, \sigma_{t-1}) \\
RR&: P(\sigma_t|\Theta, \sigma_{t-1}, \sigma_{t+1}, y_t, R_t, y_{t+1}, R_{t+1}) \\
&\propto P(y_{t+1}, R_{t+1}|\Theta, \sigma_t, \sigma_{t+1}) P(y_t, R_t|\Theta, \sigma_t, \sigma_{t-1}) \\
&\times P(\sigma_{t+1}|\Theta, \sigma_t) P(\sigma_t|\Theta, \sigma_{t-1}).
\end{aligned}
\tag{9}
$$

In the MCMC algorithm, the sampling of parameters and volatility is iterated from the equation (8) and equation (9). Before the Markov chain converges, initial sampling values are discarded. We use the first 10,000 samplings as a burn-in period and 90,000 samplings for estimation after the burn-in period. Through trace plots, we check the convergence of the algorithm in the next simulation and empirical analysis sections.

4. Simulation Analysis

The RR and the RHL are the first estimators considering both the leverage effect and information on extreme prices. Through the simulation, we examine performance on two aspects. In one aspect, we see how close the estimated parameters are to the realized parameters. Another aspect is to contrast the RMSEs between volatilities estimated and volatilities calculated from simulation. We compare the RR and the RHL with the basic SV model and existing estimators, the RR$(-\rho)$ and the RHL$(-\rho)$ which are the counter parts with no correlation to the RR and the RHL, respectively. In other words, the RR$(-\rho)$ and the $\left(\text{RHL}(-\rho)\right)$ are the same estimators as the RR (RHL) except for the imposition of zero correlation.

First we test the case that each simulation path has 500 lengths which mean 500 trading days. Next we extend the total number of trading days to 1,000 days. For both cases, we generate 500 paths. Each trading day consists of 1,000 sub-periods to produce intraday returns whose maximum and minimum returns correspond to the high and the low returns, respectively. In order to compare against the results of Brandt and Jones (2005), we use the same parameter values and test three levels of persistency of volatility: low, medium, and high. The persistency is determined by the parameter δ which is the autocorrelation coefficient of the volatility process. Because the leverage effect is taken into account in this paper, we also test three cases of correlation, -0.2, -0.4 and -0.6. Since this paper focuses on the leverage effect, we do not test cases of positive correlation.

Table 1 provides interesting features of estimation results for the case with the simulation length

T=500. The mean of estimates are reported and the values in parentheses are the root mean square errors (RMSEs). Regardless of correlations and persistency, the parameters of the volatility process, α, δ, and υ, are better estimated in estimators relating intraday data than in the SV model both in terms of the means and the RMSEs. Specifically, the parameters measuring persistency and variation of volatility are much more improved. This result is because additional information on H/L prices can lead to more accurate inference of the volatility dynamics. One special feature of Table 1 is that the SV model overestimates the volatility of volatility parameter υ considerably and the extent of overestimation also severely grows as persistency increases. Brandt and Jones (2005) argue a highly persistent volatility process generates less volatile daily volatility, which can impede the inference of the volatility of volatility parameter υ. As for the estimate of α, High/low prices alter the parameter significantly, but α just adjusts average of long-term volatility level. The SV, RR, and RHL have almost analogous long-term means of volatility when these values are calculated from the estimated parameters. The average level of volatility is well measured without H/L prices. This result seems natural since the intraday data movement has a small effect on the average value of volatilities. In the comparison between the RR and the RHL, it seems there is no difference as we find no evidence that there is an asymmetric impact of H/L price levels in the simulation. Horst et al. (2012) also document similar evidence.

******************** INSERT Table 1 HERE ********************

Panel A to C show the results according to the correlation levels -0.2, -0.4, and -0.6. We confirm that the absolute value of estimates of ρ increases in the SV, RR, and RHL as the absolute value of the true ρ increases. While all three estimators are able to estimate various magnitudes of the leverage effect suitably, the estimates of ρ are slightly underestimated when compared to the true values across all the estimators. Although no distinction between the RR and the RHL appear, the mean values of estimates of ρ in the SV model are closer to the true values in the low and medium persistence cases. This may occur because the simulated paths are not continuous but discrete or due to the approximation of joint densities given by equation (5) and equation (7). For high persistence cases, regardless of how large the correlation is, the mean values of estimated ρ from the RR and the RHL are closer to the true values than those of the SV. In the high persistence case of the SV model, since the parameters defining the volatility dynamics are poorly estimated, the estimation of correlation is also affected. As for RMSEs of ρ, the RR and the RHL outperform the SV model in all cases.

Panel A to C also show the parameter estimates by the RR($-\rho$) and the RHL($-\rho$) i.e., no correlation assumed, whereas samples are generated with correlations -0.2, -0.4, and -0.6. In Panel A, since correlation is low, the parameter estimates by the RR($-\rho$) and the RHL($-\rho$) are very close to those estimated by the RR and the RHL in low and medium persistence. For high persistence, estimates of RR($-\rho$) are not significant while estimates of RHL($-\rho$) are significant. From Panel B and C, we find that the parameter estimates are closer to the true value in the RR and the RHL than in the RR($-\rho$) and RHL($-\rho$). However, the difference is not large.

******************** INSERT Table 2 HERE ********************

In Table 2 the results of the estimations are given, where sample paths are extended to 1,000 days. The results show similar patterns of the parameter estimation seen in Table 1 but provide better estimates in terms of both the means and the RMSEs. However, the degree of improvement is more noticeable in the SV model than in other cases. The same phenomenon is observed in Alizadel et al. (2002). This can be interpreted as the effect of information. While the RR and the RHL estimators exploit the additional information on H/L prices more effectively in a short period, it appears the improvements of estimates are not large enough given a longer period.[4] This is anticipatable in light of Parkinson (1980) which shows theoretically that in order to obtain same amount of variance of a continuous random walk using daily returns or ranges, the number of observations for returns is needed about 2.5 times more than that for similar ranges. In the examination of correlation parameter, mean values of SV are a little bit closer to the true values except for the high persistence cases, but the RMSEs of the RR and the RHL are mostly smaller than those of SV.

******************** INSERT Table 3 HERE ********************

In addition to inferring parameters, estimating volatility is also important in stochastic volatility models. Table 3 reports ratios of the RMSEs between models. Each simulated volatility series with 500 lengths is produced from the simulation in Table 1. Using the simulated volatility series and the estimated volatility series, RMSEs are calculated from per path. The table shows the means and the

[4] We also tested 2,000 lengths of the sample period. There is not much difference with the results of T=1,000. When the sample period increases, the benefit of using high and low prices still appears.

values at the 5% and 95% percentiles (values in parenthesis). "SV/RR" denotes the value of the RMSE from SV over that from RR. Other symbols have similar meaning. The closer to 1 the ratio, the more similar the models are in sense of volatility estimates. We find the ratios SV/RR and SV/RHL moderately increase as persistency increases but the RR/RHL is almost identical. It is worthy to note RMSEs of SV are at least 1.5 times larger than those of RR or RHL, which implies that volatilities are estimated more accurately when intraday data are added rather than when only returns are used. Since the SV model shows poor results in the case of high persistence, this also affects volatility estimates. Within the same persistence level, the ratios remain stable across correlations. These finding suggests improvements of using extreme prices in volatility estimates are not heavily dependent upon correlation levels but rather persistency. The $\mathrm{RR}(-\rho)/\mathrm{RR}$ reports the comparison of our RR estimator with the $\mathrm{RR}(-\rho)$, a proxy used by Alizadeh et al. (2002). When correlation is low such as $\rho = -0.2$, the two estimators are statistically identical with the 95% confidence level. When the correlation is larger than -0.4, the RMSEs of RR are 4-10% smaller on average than those of $\mathrm{RR}(-\rho)$ and the difference is statistically significant. As we can see, the similar results hold for RHL and $\mathrm{RHL}(-\rho)$.

Table 4 shows the same result as in Table 3 when the length of paths increases to 1,000 days. Unlike enhancement of parameter estimates for the SV model shown in Table2, SV/RR and SV/RHL ratios do not improve much. Hence the effect of incorporating H/L prices appears more precious in estimation of volatilities than of parameters.

******************* INSERT Table 4 HERE *******************

5. Empirical Analysis

Many empirical documents report the leverage effect on index data such as the S&P 500 and the Nasdaq 100 even though the sample periods of the indexes are different by paper. In this section, we consider those indexes to observe whether there is leverage effect or not when H/L prices are added.

5.1. Data

In the RR and RHL, the daily high and low prices occur from the opening time to the closing time. As Tsiakas (2008) states, the New York Stock Exchange requires price continuity obligation, which mandates specialists to maintain the opening price close to the closing price of the previous trading day so that the open-to-close returns are calculated as if returns occur in a 24 hour cycle. Due to this market microstructure, the expected value of extreme values such as the high and low prices from the S&P 500 are underestimated because the overnight information is not reflected. To avoid this, we use the E-mini S&P 500 Futures instead of the S&P 500 index. E-mini futures contracts are traded from 5:00 p.m. the previous day to 4:15 p.m. the following day, so that trading occurs for nearly 24 hours. Although the Nasdaq Stock Market does not have price continuity obligation, the market also operates only during the trading hours. Similarly, we use the E-mini Nasdaq 100 Futures instead of the Nasdaq 100 index.

The E-mini S&P 500 and the Nasdaq 100 Futures started trading in September 1997 and June 1999, respectively. We use the overlapping time series of daily data from June 1999 to June 2007, which about 9 years. Data is obtained from Bloomberg and Table 5 summarizes the statistics for open-to-close, -high, -low returns, and ranges. To compare the E-mini futures with stock indexes, we also report basic statistics for the S&P 500 and the Nasdaq 100 indexes for the same period. Most statistics seem to be similar between the E-mini futures and the indexes. All moments of ranges for indexes are smaller than those of ranges for the E-mini futures. Annualized standard deviations are about 17% for both S&P 500 index and E-mini S&P 500 Futures but wild as about 32-33% for E-mini Nasdaq 100 Futures and Nasdaq 100 indexes. The ranges have a distributional property as in equation (10) below if the returns follow a driftless Brownian Motion process with standard deviation σ over a time period τ as shown by Feller (1951).

$$E[R(\tau)] = 2\left(\frac{2\tau\sigma^2}{\pi}\right)^{1/2}. \tag{10}$$

Equation (10) guarantees that the ratio between the expectation of ranges and standard deviation of returns over a unit period, $E[R]/\sigma$, has a constant value of 1.5958. From Table 5, when this ratio is calculated by using the average value of ranges and standard deviation of returns, the S&P 500 and the Nasdaq 100 have 1.24 and 1.22, respectively. Both values are smaller than the constant, which means the average value of the ranges is relatively small. When the ratio is calculated using the E-mini S&P 500 and Nasdaq 100 Futures, the values increase to 1.41 and 1.33, respectively. Hence the E-mini index futures fis the model better than index itself. Therefore, volatility might be underestimated when H/L prices of the indexes are incorporated.[5]

[5] Using S&P 500 and Nasdaq 100, we estimated volatility from the SV, RR, and RHL. Averages of volatility from the RR and RHL were lower than that from the SV model. The difference of annual volatility was about 4 percent in both of the two indexes, which was not a small difference.

******************* INSERT Table 5 HERE *******************

5.2. Estimates from the stochastic volatility models

Using the E-mini S&P 500 and Nasdaq 100 Futures, we estimate both parameters and volatility from the SV, RR, and RHL. The results of parameter estimates are given in Table 6. First, we can identify the parameters of correlation ρ are negative in all models, and it confirms that there is leverage effect in the indexes. Since parameters change depending on the sample period, it may be difficult to compare our models with others in some papers, but a large portion of our sample period is overlapped with the period of Horst et al. (2012). Horst et al. (2012) also tested models that incorporate high/low prices. One of them uses ranges and open-to-close returns like our RR, and the other that correspondence of our RHL uses high/low returns and open-to-close returns. However, they do not consider leverage effect of S&P 500, that is to say they assume the independence between returns and changes of volatility. If zero correlation is assumed when there is leverage effect, it can change volatility estimation. The RR and RHL have negative correlations significantly in both of E-mini S&P 500 and Nasdaq 100 futures, so leverage effect should be measured in the indexes even when H/L prices are incorporated.

In the estimated parameters of volatility dynamics, the persistence parameters δ of the RR and RHL are lower than those of the SV model. When H/L prices or ranges are used, it lowers the persistency of volatility. In the case of υ which is the variability parameter of the volatility process, information on H/L prices increases it. These results are also found in Alizadeh et al. (2002). When they use log ranges as volatility proxy in their one-factor stochastic volatility model, the persistence parameters decrease sharply and the volatility of volatility parameters increase. Even though they test exchange rates, similar patterns occur as in our indexes once range data is used. Since we incorporate ranges and returns together, changes of the parameters are not that big.

Though negative correlations appear from all the three models, the difference in ρ values between the SV model and the other two seems to be quite large. We also tested whether a two-factor stochastic volatility model may be proper for the empirical data or not. In the two-factor model, just volatility process is different like

$$\begin{gathered} \ln\sigma_t = \ln p_t + \ln q_t \\ where\ \ln p_t = \alpha + \delta_1 \ln p_{t-1} + \upsilon_1 \varepsilon_t^p \ and\ \ln q_t = \delta_2 \ln q_{t-1} + \upsilon_2 \varepsilon_t^q . \end{gathered} \tag{11}$$

To measure leverage effect, ε_t^p which follows a standard normal distribution is correlated with the return innovation ε_t^y. The innovation ε_t^q also follows a standard normal distribution, but it is independent with ε_t^y and ε_t^p. The estimation result, not reported here, is almost same as the one-factor models. The correlation difference between the SV model and the other two models still appeared. The two-factor model just distributed the parameter values in the one-factor models. δ_1, the persistence parameter of the correlated process, catches high persistency and δ_2 has small values. To conclude, when returns and H/L prices or ranges are jointly incorporated, the magnitude of leverage effect diminishes in E-mini S&P 500 and Nasdaq 100 Futures.

******************** INSERT Table 6 HERE ********************

The RR and RHL have almost same parameter estimates like in the simulation analysis, which imply that the effect of information on H/L price levels and ranges is not much different in parameter estimation. Besides parameters of the models, it is needed to investigate the effect of H/L prices on volatility estimates. Comparisons of estimated volatility between the models are shown in Table 7. Correlations are high and values of some statistics are also similar, so we know that the SV, RR, and RHL models analogously measure the variability of markets to some degree. In terms of the indexes, Nasdaq 100 is more volatile than S&P 500, which is already widely known. To future analyze, the estimates of volatility are plotted in Figure 1 and Figure 2. Volatility estimates of the RR and RHL fluctuate quite differently with those of the SV model by the top and middle panels of the figures. There are several times where the difference of estimated volatility is more than 5 percent in both of E-mini S&P 500 and Nasdaq 100 futures. It is never small differences, so we see that H/L prices affect volatility. The graphs also show that the volatility estimates of the RR and RHL are higher than those of the SV model at some peaks. For example, on April 4, 2000, volatility estimates from the RR and RHL are higher than those from the SV model over 20 percent. On that day, S&P 500 started at 1505.98 and ended at 1494.73, but the low price was 1417.22. Since our RR and RHL can reflect that crush, the volatility estimates are high on that day.

******************** INSERT Table 7 HERE ********************

******************** INSERT Figure 1 HERE ********************

******************** INSERT Figure 2 HERE ********************

The RR and RHL are the first trial that considers both information on H/L prices and leverage effect, so we diagnose applying the models to empirical data. Statistics of the return residuals are checked in Table 7. Moments indicate that the residuals have a similar distribution to a standard normal in all the three. When autocorrelations are examined, serial correlation rarely remains. Therefore, a definite evidence of model misspecification does not appear. To further investigate model misspecification, distributions of the residuals are given in Figure 3 and Figure 4. The extent of nonnormality is less severe in the RR and RHL. In the case of the SV model, several residuals show large shocks, and it is also presented in Eraker et al. (2003), where in the case of the stochastic volatility model without jumps . As a result, our RR and RHL are not misspecified even reflecting extreme prices and leverage effect.

******************** INSERT Table 8 HERE ********************

******************** INSERT Figure 3 HERE ********************

******************** INSERT Figure 4 HERE ********************

Using additional high/low prices would be beneficial in the aspect of augmented information, but it could also be advantageous in other facets. There are studies that opening or closing prices do not represent stock values but are affected by market structures (e.g., Harris (1989), Stoll and Whaley (1990)), and those prices may be noisy. Whereas range is less noisy than returns as volatility measures (see Alizadeh et al. (2002)). It is also well known that using range or H/L prices in volatility estimation gives efficiencies (e.g., Parkinson (1980), Garman and Klass (1980)). Considering these things, utilizing extreme prices might enable more accurate volatility estimation.

6. Conclusions and Future research

Volatility changes over time, so stochastic volatility models have become widely known in academics and practice. It is also well known that volatility increases when market returns fall. Many researches model this financial leverage effect with negative correlation between returns and changes of volatility. In this context, we have suggested stochastic volatility models with leverage effect and incorporated the information on high /low prices.

To compare the performance of models, we executed the MCMC simulation study. With changes in the correlation, the model using the additional information on ranges or high/low prices gives better estimates of the true parameters than that of the basic stochastic volatility model. The same was also true when estimating volatility. As the correlation level increased, our proposed model better estimated volatility than the proxy models of Alizadeh et al. (2002) and Horst et al. (2012) and it was statistically significant. Using the E-mini S&P 500 and Nasdaq 100 Futures, we find strong evidence of the leverage effect for both the E-mini S&P 500 and the Nasdaq 100 Futures and confirm that there is asymmetric information of H/L prices but the effect is meager.

Volatility estimation is important especially for option pricing. In this paper, it is the first attempt that the stochastic volatility model with leverage effect incorporates ranges or high/low prices. So examining the effect on option prices from the model might be interesting works. In respect of complementing our models, adding jumps to the return process or volatility process can be tried as Eraker et al. (2003) did under the simple stochastic models. Thus, it is needed to attempt linking jumps to high/low prices for future research.

REFERENCES

Anderson, Torben, Luca Benzoni, and Jesper Lund, 2002, Towards an empirical foundation for continuous-time equity return models, *The Journal of Finance* 57, 1239-1284.

Alizadeh, Sassan, Michael W Brandt, and Francis X Diebold, 2002, Range-based estimation of stochastic volatility models, *The Journal of Finance* 57, 1047-1091.

Amihud, Yakov, and Haim Mendelson, 1987, Trading mechanisms and stock returns: An empirical investigation, *The Journal of Finance* 42, 533-553.

Bates, Davis, 2000, Post-'87 Crash fears in S&P 500 futures options, *Journal of Econometrics* 94, 181-238.

Black, F, 1976, Studies of stock price volatility changes, *In: Proceedings of the 1976 Meetings of the American Statistical Association* 171-181.

Brandt, Michael W, and Christopher S Jones, 2005, Bayesian range-based estimation of stochastic volatility models, *Finance Research Letters* 2, 201-209.

Chernov, Mikhail, A Ronald Gallant, Eric Ghysels, and George Tauchen, 2003, Alternative models for stock price dynamics, *Journal of Econometrics* 116, 225-257.

Choi, ByoungSeon, and JeongHo Roh, 2013, On the trivariate joint distribution of brownian motion and its maximum and minimum, *Statistics & Probability Letters 83, 1046-1053*.

Christoffersen, Peter, Steven Heston, and Kris Jacobs, 2009, The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well, *Management Science* 55, 1914-1932.

Durham, Garland B, 2006, Monte carlo methods for estimating, smoothing, and filtering one-and two-factor stochastic volatility models, *Journal of Econometrics* 133, 273-305.

Eraker, Bjørn, 2004, Do stock prices and volatility jump? Reconciling evidence from spot and option prices, *The Journal of Finance* 59, 1367-1404.

Eraker, Bjørn, Michael Johannes, and Nicholas Polson, 2003, The impact of jumps in volatility and returns, *The Journal of Finance* 58, 1269-1300.

Feller, William, 1951, The asymptotic distribution of the range of sums of independent random variables, *The Annals of Mathematical Statistics* 427-432.

Gallant, A Ronald, Chien-Te Hsu, and George Tauchen, 1999, Using daily range data to calibrate volatility diffusions and extract the forward integrated variance, *Review of Economics and Statistics* 81, 617-631.

Harris, Lawrence, 1989, A day-end transaction price anomaly, *Journal of Financial and Quantitative Analysis* 24, 29-45.

Harvey, Andrew C, and Neil Shephard, 1996, Estimation of an asymmetric stochastic volatility model for asset returns, *Journal of Business & Economic Statistics* 14, 429-434.

Horst, Enrique Ter, Abel Rodriguez, Henryk Gzyl, and German Molina, 2012, Stochastic volatility models including open, close, high and low prices, *Quantitative Finance* 12, 199-212.

Jacquier, Eric, Nicholas G Polson, and Peter E Rossi, 1994, Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics* 12, 413-417.

Jacquier, Eric, Nicholas G Polson, and Peter E Rossi, 2004, Bayesian analysis of stochastic volatility models with fat-tails and correlated errors, *Journal of Econometrics* 122, 185-212.

Johannes, Michael, and Nick Polson, 2003, Mcmc methods for continuous-time financial econometrics, *Available at SSRN 480461*.

Pan, Jun, 2002, The jump-risk premia implicit in options: Evidence from an integrate time-series study, *Journal of Financial Economics* 63, 3-50.

Parkinson, Michael, 1980, The extreme value method for estimating the variance of the rate of return, *Journal of Business* 53, 61.

Roberts, Gareth O, and Jeffrey S Rosenthal, 2001, Optimal scaling for various metropolis-hastings algorithms, *Statistical science* 16, 351-367.

Rogers, Leonard CG, Stephen E Satchell, and Y Yoon, 1994, Estimating the volatility of stock prices:

A comparison of methods that use high and low prices, *Applied Financial Economics* 4, 241-247.

Stoll, Hans R, and Robert E Whaley, 1990, Stock market structure and volatility, *Review of Financial Studies* 3, 37-71.

Tsiakas, Ilias, 2008, Overnight information and stochastic volatility: A study of European and us stock exchanges, *Journal of Banking & Finance* 32, 251-268.

Yang, Dennis, and Qiang Zhang, 2000, Drift-independent volatility estimation based on high, low, open, and close prices*, *The Journal of Business* 73, 477-492.

Yu, Jun, 2005, On leverage in a stochastic volatility model, *Journal of Econometrics* 127, 165-178.

Table 1

Parameter Estimation Results of Simulation with T=500

This table shows the result of parameter estimation of simulated 500 paths. Each path is produced from a discretized stochastic volatility model:

$$y_t = \sigma_{t-1}\varepsilon_t^y$$
$$\ln\sigma_t = \alpha + \delta\ln\sigma_{t-1} + \upsilon\varepsilon_t^\sigma$$

where ε_t^y and ε_t^σ have a bivariate standard normal distribution with correlation ρ. The return and volatility series of each path has 500 lengths, so a trading day t has a value from 1 to 500. During each time step from $t-1$ to t, there are 1,000 sub-periods. We set three different levels of persistence: "Low", "Medium", and "High", by changing the values of volatility parameters. We vary the level of correlation: the case of -0.2 (Panel A), -0.4 (Panel B), and -0.6 (Panel C). Parameter estimation is conducted through different models. "SV" denotes the basic stochastic volatility model. "RR" denotes the stochastic volatility model incorporating ranges. "RHL" denotes the stochastic volatility model incorporating high and low prices. "RR$(-\rho)$" $\left(\text{RHL}(-\rho)\right)$ corresponds to the RR (RHL) with imposition of zero correlation. These values are means of estimates and the values in parenthesis are root mean square errors.

Panel A: Parameter Estimation with ρ=-0.2

	Low persistence				Medium persistence				High persistence			
	α	δ	ν	ρ	α	δ	ν	ρ	α	δ	ν	ρ
True	-0.368	0.900	0.182	-0.200	-0.184	0.950	0.130	-0.200	-0.074	0.980	0.083	-0.200
SV	-0.549	0.851	0.202	-0.165	-0.364	0.902	0.171	-0.145	-0.276	0.928	0.146	-0.125
	(0.274)	(0.074)	(0.033)	(0.138)	(0.242)	(0.064)	(0.045)	(0.158)	(0.273)	(0.071)	(0.065)	(0.179)
RR (-ρ)	-0.393	0.894	0.175	Null	-0.243	0.935	0.138	Null	-0.238	0.936	0.138	Null
	(0.092)	(0.024)	(0.015)	Null	(0.097)	(0.025)	(0.013)	Null	(0.092)	(0.024)	(0.013)	Null
RHL (-ρ)	-0.394	0.894	0.175	Null	-0.244	0.935	0.138	Null	-0.154	0.960	0.106	Null
	(0.093)	(0.024)	(0.015)	Null	(0.098)	(0.025)	(0.013)	Null	(0.105)	(0.027)	(0.024)	Null
RR	-0.390	0.895	0.175	-0.141	-0.242	0.935	0.139	-0.138	-0.152	0.960	0.106	-0.134
	(0.090)	(0.024)	(0.015)	(0.094)	(0.096)	(0.025)	(0.013)	(0.108)	(0.104)	(0.027)	(0.024)	(0.122)
RHL	-0.391	0.895	0.175	-0.140	-0.242	0.935	0.139	-0.137	-0.153	0.960	0.106	-0.133
	(0.091)	(0.024)	(0.015)	(0.094)	(0.097)	(0.025)	(0.013)	(0.108)	(0.105)	(0.027)	(0.024)	(0.123)

Panel B: Parameter Estimation with ρ=-0.4												
	Low persistence				Medium persistence				High persistence			
	α	δ	ν	ρ	α	δ	ν	ρ	α	δ	ν	ρ
True	-0.368	0.900	0.182	-0.400	-0.184	0.950	0.130	-0.400	-0.074	0.980	0.083	-0.400
SV	-0.526	0.857	0.201	-0.309	-0.363	0.902	0.172	-0.306	-0.283	0.925	0.147	-0.265
	(0.235)	(0.064)	(0.031)	(0.155)	(0.241)	(0.065)	(0.046)	(0.169)	(0.289)	(0.075)	(0.066)	(0.205)
RR (-ρ)	-0.386	0.896	0.175	Null	-0.238	0.936	0.138	Null	-0.144	0.962	0.105	Null
	(0.089)	(0.024)	(0.014)	Null	(0.092)	(0.024)	(0.013)	Null	(0.094)	(0.024)	(0.023)	Null
RHL (-ρ)	-0.386	0.896	0.176	Null	-0.238	0.936	0.138	Null	-0.153	0.960	0.106	Null
	(0.089)	(0.024)	(0.014)	Null	(0.092)	(0.024)	(0.013)	Null	(0.104)	(0.027)	(0.024)	Null
RR	-0.367	0.900	0.173	-0.286	-0.225	0.939	0.136	-0.294	-0.144	0.962	0.105	-0.286
	(0.079)	(0.021)	(0.015)	(0.135)	(0.078)	(0.021)	(0.012)	(0.136)	(0.094)	(0.024)	(0.023)	(0.152)
RHL	-0.368	0.900	0.173	-0.284	-0.226	0.939	0.136	-0.292	-0.144	0.962	0.106	-0.284
	(0.079)	(0.021)	(0.015)	(0.136)	(0.078)	(0.021)	(0.012)	(0.137)	(0.094)	(0.024)	(0.024)	(0.154)

Panel C: Parameter Estimation with ρ=-0.6												
	Low persistence				Medium persistence				High persistence			
	α	δ	ν	ρ	α	δ	ν	ρ	α	δ	ν	ρ
True	-0.368	0.900	0.182	-0.600	-0.184	0.950	0.130	-0.600	-0.074	0.980	0.083	-0.600
SV	-0.509	0.862	0.197	-0.477	-0.358	0.904	0.169	-0.458	-0.269	0.929	0.145	-0.394
	(0.209)	(0.057)	(0.026)	(0.168)	(0.225)	(0.060)	(0.043)	(0.192)	(0.262)	(0.068)	(0.064)	(0.254)
RR (-ρ)	-0.382	0.897	0.174	Null	-0.245	0.934	0.138	Null	-0.135	0.964	0.104	Null
	(0.089)	(0.024)	(0.015)	Null	(0.097)	(0.025)	(0.013)	Null	(0.084)	(0.022)	(0.022)	Null
RHL (-ρ)	-0.383	0.897	0.175	Null	-0.245	0.934	0.138	Null	-0.156	0.959	0.106	Null
	(0.090)	(0.024)	(0.015)	Null	(0.097)	(0.025)	(0.013)	Null	(0.109)	(0.028)	(0.024)	Null
RR	-0.340	0.907	0.167	-0.454	-0.215	0.941	0.133	-0.458	-0.135	0.964	0.104	-0.446
	(0.074)	(0.020)	(0.018)	(0.162)	(0.066)	(0.018)	(0.009)	(0.162)	(0.084)	(0.022)	(0.022)	(0.180)
RHL	-0.340	0.907	0.167	-0.453	-0.215	0.941	0.133	-0.457	-0.135	0.964	0.104	-0.444
	(0.074)	(0.020)	(0.018)	(0.164)	(0.066)	(0.018)	(0.009)	(0.164)	(0.084)	(0.022)	(0.022)	(0.181)

Table 2

Parameter Estimation Results of Simulation with T=1,000

This table shows the result of parameter estimation of simulated 500 paths. Each path is produced from a discretized stochastic volatility model:

$$y_t = \sigma_{t-1}\varepsilon_t^y$$
$$\ln\sigma_t = \alpha + \delta\ln\sigma_{t-1} + \upsilon\varepsilon_t^\sigma$$

where ε_t^y and ε_t^σ have a bivariate standard normal distribution with correlation ρ. The return and volatility series of each path has 1,000 lengths, so a trading day t has a value from 1 to 1,000. During each time step from $t-1$ to t, there are 1,000 sub-periods. We set three different levels of persistence: "Low", "Medium", and "High", by changing the values of volatility parameters. We vary the level of correlation: the case of -0.2 (Panel A), -0.4 (Panel B), and -0.6 (Panel C). Parameter estimation is conducted through different models. "SV" denotes the basic stochastic volatility model. "RR" denotes the stochastic volatility model incorporating ranges. "RHL" denotes the stochastic volatility model incorporating high and low prices. These values are means of estimates and the values in parenthesis are root mean square errors.

Panel A: Parameter Estimation with ρ=-0.2

	Low persistence				Medium persistence				High persistence			
	α	δ	ν	ρ	α	δ	ν	ρ	α	δ	ν	ρ
True	-0.368	0.900	0.182	-0.200	-0.184	0.950	0.130	-0.200	-0.074	0.980	0.083	-0.200
SV	-0.442	0.880	0.188	-0.159	-0.274	0.926	0.153	-0.151	-0.171	0.954	0.122	-0.140
	(0.139)	(0.037)	(0.022)	(0.107)	(0.129)	(0.034)	(0.028)	(0.116)	(0.126)	(0.033)	(0.040)	(0.134)
RR	-0.364	0.902	0.172	-0.144	-0.211	0.943	0.133	-0.142	-0.114	0.970	0.097	-0.143
	(0.061)	(0.016)	(0.014)	(0.077)	(0.056)	(0.015)	(0.008)	(0.086)	(0.055)	(0.014)	(0.015)	(0.094)
RHL	-0.365	0.902	0.173	-0.143	-0.212	0.943	0.133	-0.141	-0.114	0.970	0.097	-0.142
	(0.061)	(0.016)	(0.014)	(0.077)	(0.056)	(0.015)	(0.008)	(0.086)	(0.056)	(0.014)	(0.015)	(0.094)

Panel B: Parameter Estimation with ρ=-0.4												
	Low persistence				Medium persistence				High persistence			
	α	δ	ν	ρ	α	δ	ν	ρ	α	δ	ν	ρ
True	-0.368	0.900	0.182	-0.400	-0.184	0.950	0.130	-0.400	-0.074	0.980	0.083	-0.400
SV	-0.429	0.883	0.186	-0.330	-0.269	0.927	0.153	-0.327	-0.168	0.955	0.123	-0.307
	(0.120)	(0.033)	(0.019)	(0.116)	(0.115)	(0.031)	(0.028)	(0.123)	(0.114)	(0.030)	(0.041)	(0.149)
RR	-0.341	0.907	0.169	-0.297	-0.198	0.946	0.131	-0.305	-0.108	0.971	0.096	-0.309
	(0.062)	(0.017)	(0.016)	(0.116)	(0.043)	(0.011)	(0.008)	(0.111)	(0.047)	(0.012)	(0.014)	(0.114)
RHL	-0.342	0.907	0.169	-0.296	-0.199	0.946	0.131	-0.303	-0.108	0.971	0.096	-0.307
	(0.061)	(0.017)	(0.016)	(0.117)	(0.043)	(0.012)	(0.008)	(0.112)	(0.047)	(0.012)	(0.014)	(0.116)

Panel C: Parameter Estimation with ρ=-0.6												
	Low persistence				Medium persistence				High persistence			
	α	δ	ν	ρ	α	δ	ν	ρ	α	δ	ν	ρ
True	-0.368	0.900	0.182	-0.600	-0.184	0.950	0.130	-0.600	-0.074	0.980	0.083	-0.600
SV	-0.433	0.882	0.184	-0.502	-0.272	0.926	0.151	-0.493	-0.167	0.955	0.122	-0.454
	(0.111)	(0.030)	(0.018)	(0.131)	(0.111)	(0.030)	(0.025)	(0.141)	(0.111)	(0.029)	(0.040)	(0.179)
RR	-0.322	0.912	0.164	-0.465	-0.189	0.948	0.127	-0.473	-0.101	0.972	0.094	-0.478
	(0.065)	(0.017)	(0.020)	(0.144)	(0.036)	(0.010)	(0.007)	(0.139)	(0.040)	(0.011)	(0.012)	(0.139)
RHL	-0.322	0.911	0.164	-0.463	-0.189	0.948	0.127	-0.471	-0.102	0.972	0.094	-0.476
	(0.065)	(0.017)	(0.020)	(0.146)	(0.036)	(0.010)	(0.007)	(0.140)	(0.041)	(0.011)	(0.012)	(0.141)

Table 3

Ratio of Root Mean Square Error for Volatility with T=500

This table reports means of the ratio of root mean square errors for volatility. Each simulated volatility series with 500 lengths is produced from the simulation in Table 1. Throughout the parameter estimation in Table 1, latent volatilities are estimated from the three SV, RR, and RHL. Using the simulated volatility series and the estimated volatility series, root mean square error is calculated per path. "SV/RR" denotes the value of the root mean square error from SV over that from RR. "SV/ RHL" denotes the value of the root mean square error from SV over that from RHL. "RR/ RHL" denotes the value of the root mean square error from RR over that from RHL. "RR (-ρ) / RR" denotes the value of the root mean square error from RR over that from RR (-ρ). "RHL (-ρ) / RHL" denotes the value of the root mean square error from RR over that from RR (-ρ). From the 500 simulated paths, this table shows means and 5% and 95% percentiles (values in parenthesis) of the ratios. Varying the levels of persistence and correlation, the ratio of root mean square error is calculated.

Panel A: Low Persistence			
Model	ρ = -0.2	ρ = -0.4	ρ = -0.6
SV / RR	1.55 (1.37, 1.78)	1.55 (1.36, 1.78)	1.54 (1.33, 1.77)
SV / RHL	1.55 (1.36, 1.78)	1.55 (1.36, 1.77)	1.53 (1.32, 1.76)
RR / RHL	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)
RR (-ρ) / RR	1.01 (1.00, 1.02)	1.04 (1.02, 1.06)	1.10 (1.07, 1.14)
RHL (-ρ) / RHL	1.01 (1.00, 1.02)	1.04 (1.02, 1.06)	1.10 (1.07, 1.14)
Panel B: Medium Persistence			
Model	ρ = -0.2	ρ = -0.4	ρ = -0.6
SV / RR	1.63 (1.38, 1.92)	1.62 (1.37, 1.91)	1.64 (1.38, 1.94)
SV / RHL	1.63 (1.38, 1.92)	1.62 (1.37, 1.90)	1.64 (1.37, 1.95)
RR / RHL	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)
RR (-ρ) / RR	1.01 (1.00, 1.02)	1.04 (1.02, 1.06)	1.11 (1.07, 1.15)
RHL (-ρ) / RHL	1.01 (1.00, 1.02)	1.04 (1.02, 1.06)	1.11 (1.07, 1.15)
Panel C: High Persistence			
Model	ρ = -0.2	ρ = -0.4	ρ = -0.6
SV / RR	1.72 (1.39, 2.12)	1.72 (1.38, 2.09)	1.76 (1.4, 2.21)
SV / RHL	1.71 (1.39, 2.12)	1.72 (1.38, 2.09)	1.76 (1.40, 2.20)
RR / RHL	1.00 (0.99, 1.01)	1.00 (0.98, 1.01)	1.00 (0.98, 1.01)
RR (-ρ) / RR	1.01 (0.99, 1.02)	1.04 (1.01, 1.07)	1.10 (1.06, 1.15)
RHL (-ρ) / RHL	1.01 (0.99, 1.02)	1.04 (1.01, 1.07)	1.10 (1.06, 1.15)

Table 4

Ratio of Root Mean Square Error for Volatility with T=1,000

This table reports means of the ratio of root mean square errors for volatility. Each simulated volatility series with 1,000 lengths is produced from the simulation in Table 2. Throughout the parameter estimation in Table 2, latent volatilities are estimated from the three SV, RR, and RHL models. Using the simulated volatility series and the estimated volatility series, root mean square error is calculated per path. "SV/RR" denotes the value of the root mean square error from SV over that from RR. "SV/ RHL" denotes the value of the root mean square error from SV over that from RHL. "RR/ RHL" denotes the value of the root mean square error from RR over that from RHL. From the simulated 500 paths, this table shows means and 5% and 95% percentiles (values in parenthesis) of the ratios. Varying the levels of persistence and correlation, the ratio of root mean square error is calculated.

Panel A: Low Persistence			
Model	ρ = -0.2	ρ = -0.4	ρ = -0.6
SV / RR	1.54 (1.39, 1.71)	1.54 (1.39, 1.70)	1.51 (1.36, 1.68)
SV / RHL	1.54 (1.39, 1.70)	1.54 (1.39, 1.69)	1.51 (1.36, 1.68)
RR / RHL	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)
Panel B: Medium Persistence			
Model	ρ = -0.2	ρ = -0.4	ρ = -0.6
SV / RR	1.60 (1.44, 1.80)	1.59 (1.41, 1.79)	1.58 (1.40, 1.79)
SV / RHL	1.60 (1.44, 1.80)	1.59 (1.42, 1.79)	1.58 (1.39, 1.79)
RR / RHL	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)
Panel C: High Persistence			
Model	ρ = -0.2	ρ = -0.4	ρ = -0.6
SV / RR	1.65 (1.42, 1.92)	1.66 (1.44, 1.93)	1.67 (1.41, 1.96)
SV / RHL	1.65 (1.42, 1.92)	1.66 (1.43, 1.93)	1.67 (1.40, 1.95)
RR / RHL	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)

Table 5

Statistics for Indexes and E-mini Futures of S&P 500 and Nasdaq 100

From June 21, 1999 to June 30, 2007, statistics for daily returns and ranges of S&P 500, Nasdaq 100, and their E-mini futures are reported. "Return" refers to open-to-close log returns, which are calculated from opening and closing prices. "High" refers to open-to-high log returns, which are calculated from the maximum and opening prices of each day. "Low" refers to open-to-low log returns, which are calculated from the minimum and opening prices of each day. "Range" refers to the difference between "High" and "Low".

	Mean	Std Dev	Skewness	Kurtosis	Min	Max
E-mini S&P 500						
Return (%)	0.0030	1.0836	-0.0016	5.7194	-6.3224	5.9496
High (%)	0.7417	0.7174	2.9476	20.5970	0.0000	9.4103
Low (%)	-0.7909	0.7813	-2.1029	10.5684	-7.8328	0.0000
Range *100	1.5327	0.9371	2.1188	11.8624	0.1897	9.4602
S&P 500						
Return (%)	0.0064	1.1034	0.0791	5.6533	-6.0045	5.5720
High (%)	0.6614	0.6985	2.2347	10.8393	0.0000	5.6786
Low (%)	-0.7051	0.7801	-2.1019	10.1275	-7.2775	0.0000
Range *100	1.3665	0.8295	2.0272	10.6274	0.2474	8.4792
E-mini Nasdaq 100						
Return (%)	-0.0124	2.1409	0.0040	6.9938	-11.0383	14.8467
High (%)	1.4038	1.4408	2.8688	17.8567	0.0000	17.2104
Low (%)	-1.4443	1.5379	-2.1510	9.5464	-13.2220	0.0000
Range *100	2.8481	1.9894	1.8402	9.0674	0.2594	20.5918
Nasdaq 100						
Return (%)	-0.0284	2.0278	0.3586	9.4212	-9.8182	19.1698
High (%)	1.2046	1.3193	3.2077	26.0899	0.0000	19.2548
Low (%)	-1.2650	1.3868	-2.4140	13.3041	-14.8486	0.0000
Range *100	2.4696	1.7127	2.0979	11.9902	0.3842	19.2548

Table 6

Parameter Estimates of E-mini S&P 500 and Nasdaq 100 Futures

This table reports parameter estimates for E-mini S&P 500 and Nasdaq 100 Futures from June 21, 1999 to June 30, 2007. The estimates are derived from the MCMC method and daily-scaled values. "SV" represents the model using open-to-close returns. "RR" represents the model using open-to-close returns and ranges. "RHL" represents the model using open-to-close, -low, and - high returns. Parentheses are the standard deviation of samplings.

	α	δ	υ	ρ
E-mini S&P 500				
SV	-0.1430 (0.0302)	0.9697 (0.0064)	0.1095 (0.0120)	-0.7645 (0.0501)
RR	-0.4996 (0.0634)	0.8936 (0.0133)	0.2141 (0.0112)	-0.3835 (0.0314)
RHL	-0.4933 (0.0625)	0.8949 (0.0132)	0.2110 (0.0112)	-0.3809 (0.0321)
E-mini Nasdaq 100				
SV	-0.0437 (0.0133)	0.9895 (0.0032)	0.0799 (0.0071)	-0.5273 (0.0772)
RR	-0.2485 (0.0376)	0.9399 (0.0090)	0.1949 (0.0101)	-0.3065 (0.0328)
RHL	-0.2396 (0.0375)	0.9420 (0.0089)	0.1905 (0.0103)	-0.3083 (0.0344)

Table 7

Statistics for Estimated Volatility of E-mini S&P 500 and Nasdaq 100 Futures

This table shows some statistics and correlation of estimated volatility from SV, RR, and RHL. Using the MCMC estimation method, latent volatilities of E-mini S&P 500 and Nasdaq 100 futures from June 21, 1999 to June 30, 2007 are estimated. Statistics are annualized percent values.

Correlation				Mean	Std Dev	Min	Max
E-mini S&P 500	SV	RR	RHL				
SV	1	0.9091	0.9099	15.71	6.87	5.81	53.02
RR		1	0.9974	15.31	7.40	4.52	63.19
RHL			1	15.32	7.32	5.00	62.93
E-mini Nasdaq 100	SV	RR	RHL				
SV	1	0.9459	0.9457	29.32	16.86	8.67	87.76
RR		1	0.9981	28.48	16.57	6.14	105.10
RHL			1	28.54	16.56	7.34	103.94

Table 8

Residual Diagnostics

Statistics for residuals from SV, RR, and RHL are reported. Residuals are produced from each measurement equation of the discretized return processes by applying the estimation to daily E-mini S&P 500 and Nasdaq 100 Futures data from June 21, 1999 to June 30, 2007. Moments and autocorrelations for residuals are shown in this table.

	Moments				Autocorrelations		
Model	Mean	Std Dev	Skewness	Kurtosis	Lag 1	Lag 5	Lag 10
E-mini S&P 500							
SV	0.001	0.961	-0.241	2.912	-0.007	-0.015	0.007
RR	0.071	0.980	0.084	2.718	-0.015	-0.025	0.015
RHL	0.071	0.978	0.080	2.713	-0.014	-0.025	0.014
E-mini Nasdaq 100							
SV	0.005	0.967	-0.145	2.819	-0.001	-0.007	-0.012
RR	0.055	0.984	0.066	2.760	0.003	-0.008	-0.010
RHL	0.055	0.982	0.068	2.758	0.002	-0.008	-0.009

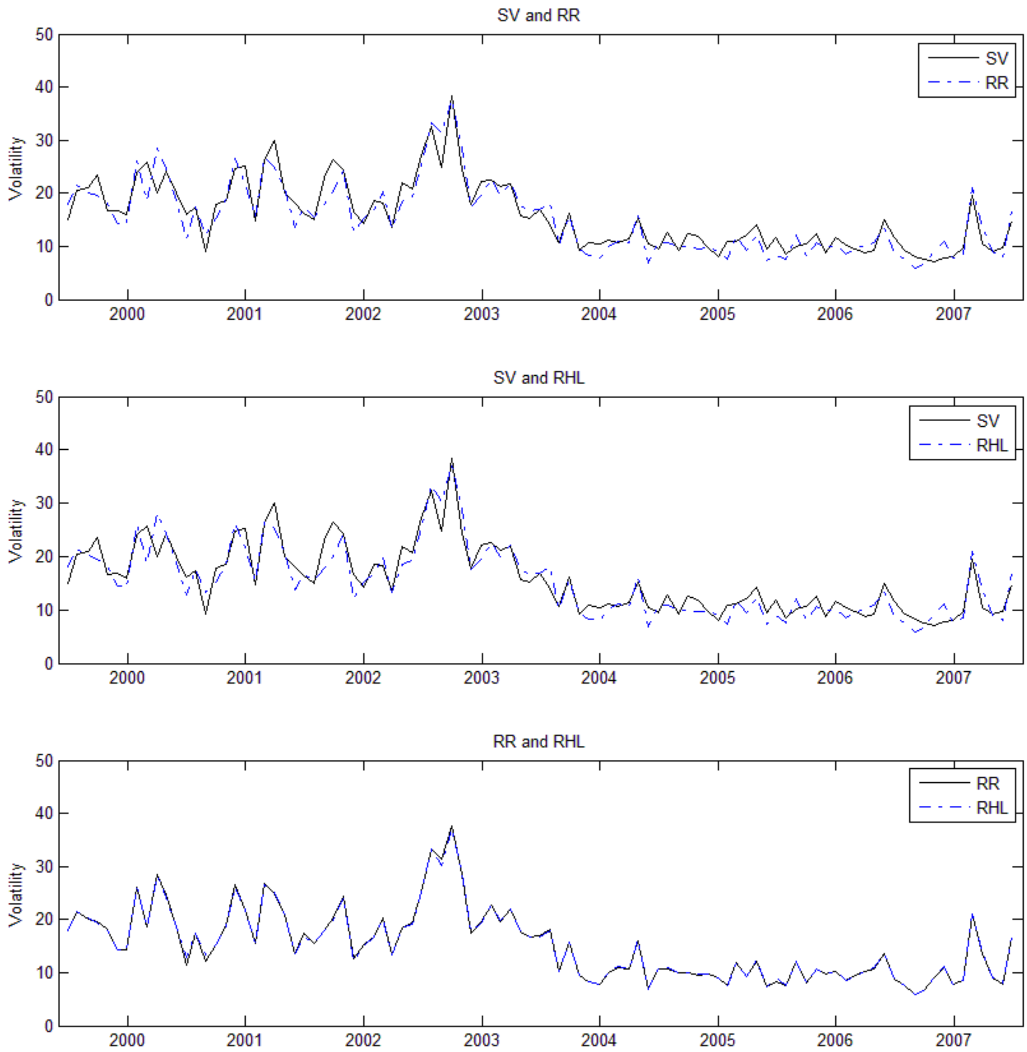

Figure 1. Monthly estimated volatility of E-mini S&P 500 Futures from SV, RR, and RHL. This figure plots monthly volatility estimates of E-mini S&P 500 Futures from June 21, 1999 to June 30, 2007. The volatility is estimated from daily data, and then plotted monthly. The scale of estimates is yearly percent. The top panel is a comparison between volatility of SV model and that of RR. The middle panel is a comparison between volatility of SV model and that of RHL. The lower panel is a comparison between volatility of RR and that of RHL.

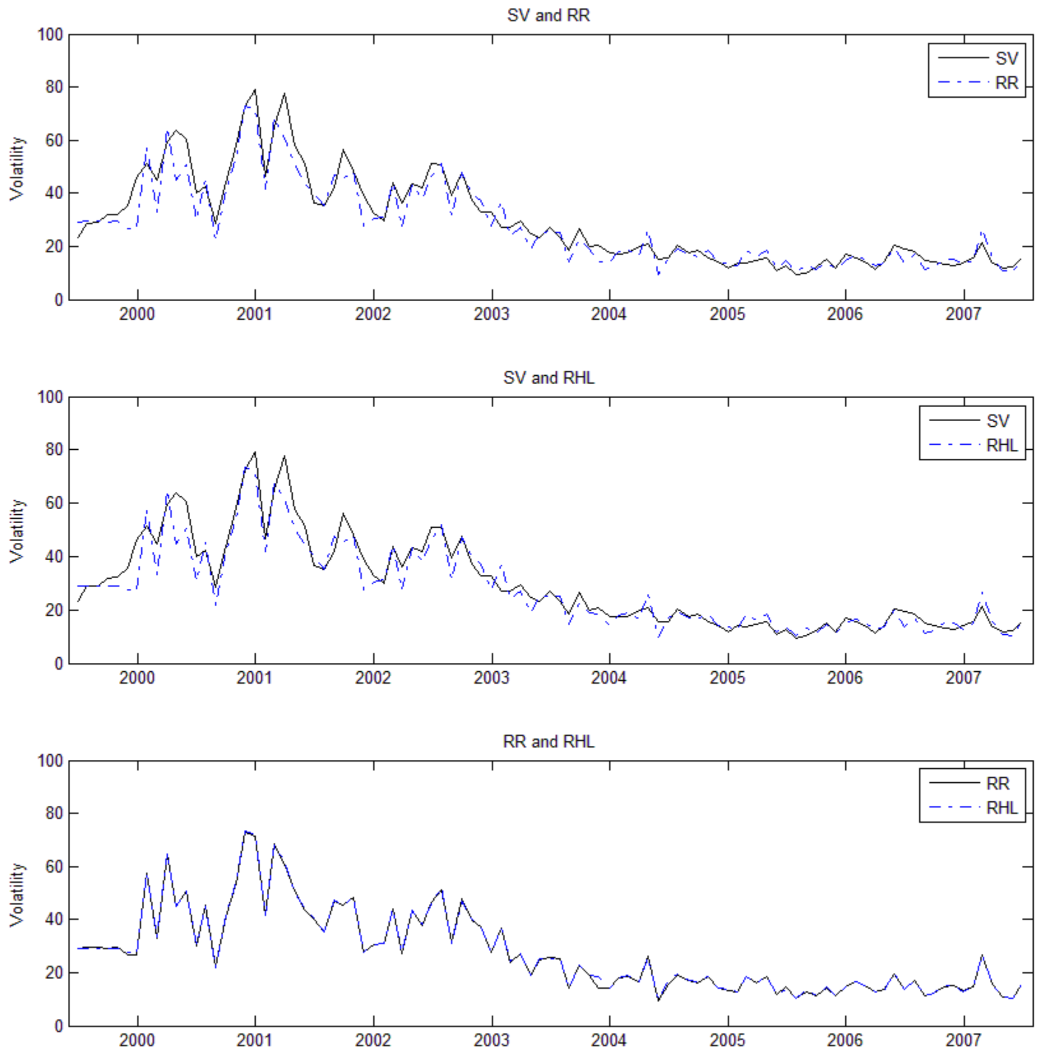

Figure 2. Monthly estimated volatility of E-mini Nasdaq 100 Futures from SV, RR, and RHL. This figure plots monthly volatility estimates of E-mini Nasdaq 100 Futures from June 21, 1999 to June 30, 2007. The volatility is estimated from daily data, and then plotted monthly. The scale of estimates is yearly percent. The top panel is a comparison between volatility of SV model and that of RR. The middle panel is a comparison between volatility of SV model and that of RHL. The lower panel is a comparison between volatility of RR and that of RHL.

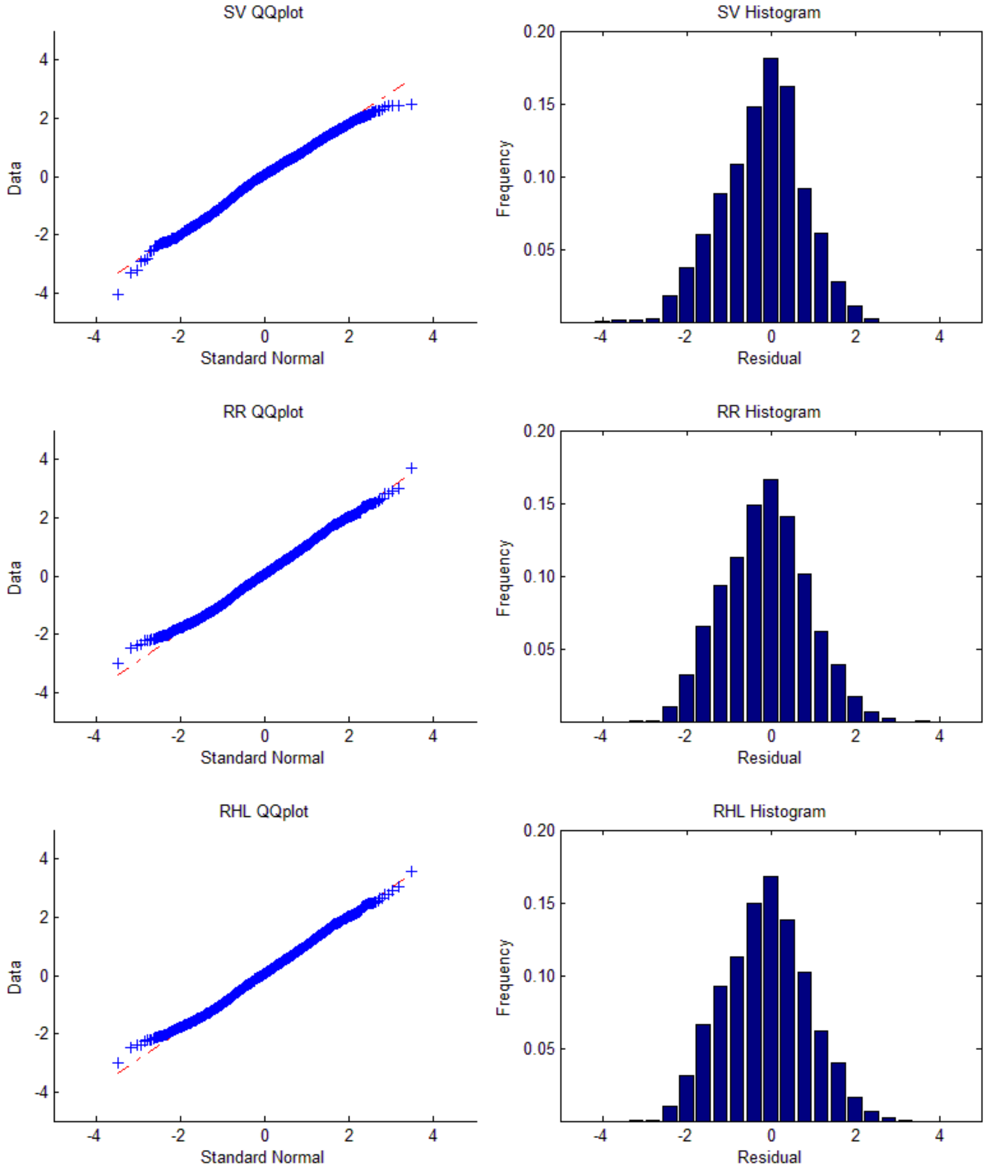

Figure 3. Distributions of return residuals of E-mini S&P 500 Futures. This figure shows the distribution of return residuals of E-mini S&P 500 Futures from June 21, 1999 to June 30, 2007. The residuals are standardized innovations of measurement equations. QQ plots are on the left and histograms are on the right. The top panel is the distribution of residuals from SV model. The middle panel is the distribution of residuals from RR. The lower panel is the distribution of residuals from RHL.

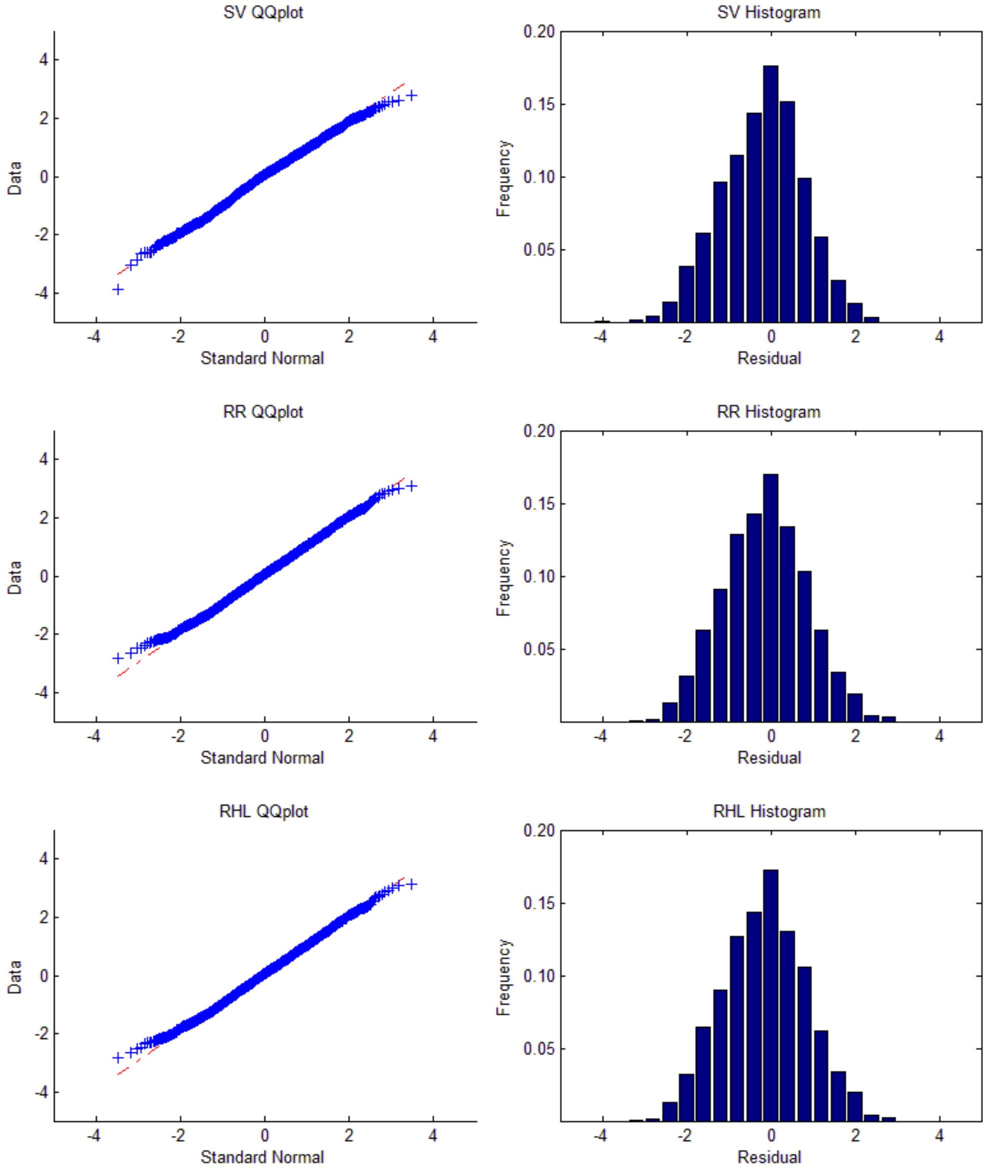

Figure 4. Distributions of return residuals of E-mini Nasdaq 100 Futures. This figure shows the distribution of return residuals of E-mini Nasdaq 100 Futures from June 21, 1999 to June 30, 2007. The residuals are standardized innovations of measurement equations. QQ plots are on the left and histograms are on the right. The top panel is the distribution of residuals from SV model. The middle panel is the distribution of residuals from RR. The lower panel is the distribution of residuals from RHL.