Volatility estimation by combining stock price data and option data

Yi Liu and Yazhen Wang[*][†]

Volatility modeling and analysis are traditionally based on either historical price data or option data. Finance theory shows that option prices heavily depend on the underlying stocks' prices, thus the two kinds of data are related. This paper explores the approach that combines both stock price data and option data to perform the statistical analysis of volatility. We investigate the Black-Scholes model and an exponential GARCH model and derive the relationship among the Fisher information for volatility estimation based on stock price data alone or option data alone as well as joint volatility estimation for combining stock price data and option data. Under the Black-Scholes model an asymptotic theory for the joint estimation is established, and a simulation study is conducted to check finite sample performances of the proposed joint estimation.

AMS 2000 subject classifications: Primary 62M10; secondary 62F10, 62F12, 91B25, 91B84.
Keywords and phrases: The Black-Scholes model, Diffusion, GARCH model, Option data, Stock data, Volatility estimation.

1. INTRODUCTION

Volatility is central to modern finance, and there is much literature on the study of volatility in finance, econometrics and statistics. Two data sources, historical stocks' price data and option data, are frequently employed to estimate and forecast volatility in financial studies and practices, and it is often the case that either price data or option data are used. Since the option pricing theory shows that the prices of options depend on the price models of the underlying stocks, option data are related to the stocks' price data. Instead of choosing one source against the other, this paper explores a combined statistical approach for volatility estimation based on both stock price data and option data. We consider the Black-Scholes model and an exponential GARCH model and derive an exact or approximate joint likelihood for the combined two types of data under the models. The volatility parameter estimation is developed by maximizing the obtained joint likelihood. We derive

the Fisher information and establish the asymptotic theory for the proposed estimators to show the advantage of the joint estimation approach. A simulation study is conducted to investigate the finite sample performances of the estimators. Assuming a geometric Brownian motion model for the stock price, Black and Scholes (1973) established the so-called Black-Scholes formula for the European option prices. The model and formula have closed form expressions that allow us to develop explicit asymptotic results for the combined estimation. For financial models with varying volatilities, there is a great amount of volatility studies and option pricing on discrete-time GARCH models [Bollerslev (1986), Engle (1982), Gouriéroux (1997), Lyuu and Wu (2005), Nelson (1991)] and continuous-time diffusion models [Fouque et al. (2000), Heston (1993), Hull and White (1987), Melino and Turnbull (1990), Rossi (1996), Scott (1987), Stein and Stein (1991), Wiggins (1987), Duan (1997), Nelson (1990), Yu, Li and Wells (2010)]. GARCH models use lagged observation errors to model volatility and obtain the closed form likelihood functions of historical price data, but the GARCH option pricing is very complicated. While it is very hard to derive the likelihood for the price data generated from the diffusion models, their option prices often have some sort of explicit expressions. Based on the diffusion limit of an exponential GARCH model, we take advantage of the explicit expansion of the diffusion option price to approximate the GARCH option price, and hence we build an implied volatility model and develop a joint likelihood for both stock price data and option price data. We obtain the estimators of model parameters by maximizing the joint likelihood. We derive the Fisher information to show the advantage of the parameter estimation based on the combined data.

The rest of the paper is arranged as follows. Sections 2 and 3 feature the combined estimation for the Black-Scholes model and an exponential GARCH model, respectively. Section 4 presents a simulation study.

2. ESTIMATION UNDER THE BLACK-SCHOLES PARADIGM

2.1 Constant volatility estimation by either price data or option data

Let S_t be the underlying price process of a stock, and assume

$$(1) \qquad S_t = S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma B_t},$$

where B_t is a standard Brownian motion, μ is drift, and σ is volatility. Black and Scholes (1973) derived the formula for the European call option price as a function of spot price S at time t, strike price K, maturity duration $T - t$, interest rate r and volatility σ, as follows,

$$(2)\ \ C_{BS}(t, S; T, K; \sigma) = \Phi(d_1)S - \Phi(d_2)Ke^{-r(T-t)},$$

where Φ is the cdf of standard normal distribution and

$$d_1 = \frac{\log(S/K) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}},$$
$$d_2 = d_1 - \sigma\sqrt{T - t}.$$

Suppose that the stock price governed by (1) is observed at discrete time $t_j = j/n$, $j = 1, \ldots, n$. Let $X_j = \sqrt{n}(\log(S_{t_j}) - \log(S_{t_{j-1}}))$ be the scaled stock returns, $j = 1, \ldots, n$. Then $X_1, \ldots, X_n$ are i.i.d normal and the maximum likelihood estimator of σ is derived as follows,

$$(3)\ \ \hat{\sigma}_S = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})^2}, \qquad \bar{X} = \frac{1}{n}\sum_{j=1}^{n}X_j.$$

We can easily compute the Fisher information $2n/\sigma^2$ for estimating σ and derive the asymptotical normality of $\hat{\sigma}_S$ as follows,

$$(4)\ \ \ \sqrt{n}(\hat{\sigma}_S - \sigma) \to N(0, \sigma^2/2), \qquad \text{as } n \to \infty.$$

Since the option price $C_{BS}(t, S, T, K; \sigma)$ depends on known (t, S, T, K) and unknown σ, we may estimate volatility σ from the observed option price data. Assume that the observed option price data are equal to the true option prices calculated from (2) plus random errors, that is, the observed option price data obey a nonlinear regression model,

$$(5)\ \ C_i = C_{BS}(t_i, S_{t_i}, T_i, K_i; \sigma) + \epsilon_i, \quad i = 1, \ldots, m,$$

where t_i is the point of purchase and T_i is the time of expiration, S_{t_i} and K_i represent the stock and strike prices, respectively, and ϵ_i's are random errors with mean zero and variance η^2. For model (5), we estimate volatility σ by the non-linear least squares estimator $\hat{\sigma}_{BS}$ which minimizes the sum of squared residuals

$$\sum_{i=1}^{m}[C_i - C_{BS}(t_i, S_{t_i}, T_i, K_i; \sigma)]^2.$$

Following Lai (1994) we may derive the asymptotic distribution of $\hat{\sigma}_{BS}$ as follows,

$$(6)\ \ \ D^{1/2}(\hat{\sigma}_{BS} - \sigma) \to N(0, \eta^2), \qquad \text{as } m \to \infty,$$

where $D = \sum\limits_{i=1}^{m} E[(\frac{\partial C_{BS}(x_i, \hat{\sigma}_{BS})}{\partial \sigma})^2]$ and $x_i = (t_i, S_{t_i}; T_i, K_i)$.

As the Black-Scholes formula indicates a one-to-one correspondence between the option price and the volatility when we know the starting time t, expiration time T, spot price S, striking price K, and interest rate r, we define the implied volatility to be the volatility corresponding to a given option price for (t, T, S, K, r), and practitioners often use the implied volatility instead of the corresponding option prices in financial applications. Alternative to (5) we may adopt a model based on implied volatility by assuming the observed implied volatilities are equal to the true volatilities plus some random errors,

$$(7)\ \ \ \ \ \ \ \ \ I_i = \sigma + z_i, \qquad i = 1, \ldots, m,$$

where σ denotes the true volatility, I_i's are observed implied volatilities, and z_i's are i.i.d. random errors with mean 0 and variance τ^2. For (7), the Fisher information for σ is m/τ^2, we estimate σ by sample mean $\bar{I} = \frac{1}{m}\sum_{i=1}^{m}I_i$, and its asymptotic distribution is given as follows,

$$(8)\ \ \ \ \sqrt{m}(\bar{I} - \sigma) \to N(0, \tau^2), \qquad \text{as } m \to \infty.$$

For model (5) we need to solve the optimization problem and find the non-linear least squares estimator $\hat{\sigma}_{BS}$. Numerical algorithms such as the Newton-Raphson method can be used to compute the estimator $\hat{\sigma}_{BS}$. For model (7) the estimator $\bar{I}$ is easy to compute once the observed implied volatilities are given. However, we may need numerical methods such as the Newton-Raphson method to invert the Black-Scholes formula and calculate the observed implied volatilities I_i.

2.2 Estimation by combining stock price data and option price data

In finance, volatilities are often obtained from two estimation approaches based on either stock price data or option data. A natural question is which approach is better: Should we collect individual stock data and analyze the data to scrutinize the degree of variability, or would it be more beneficial to explore option data traded within the market at a supposed fair value for volatility estimation? From the statistical point of view, rather than trying to select one of the approaches, it is statistically more efficient to combine the two types of data and develop a volatility estimation scheme based on the combined data.

We consider the stock price data generated from the geometric Brownian motion model (1) and the option data observed from the implied volatility model (7). For the stock price data, we have the scaled return data $X_j = \sqrt{n}(\log(S_{t_j}) - \log(S_{t_{j-1}})) \sim N(\nu, \sigma^2)$, where $\nu = (\mu - \sigma^2/2)/\sqrt{n}$. We assume that random errors $z_1, \ldots, z_m$ in model (7) are i.i.d. normal and independent of $X_1, \ldots, X_n$. The joint likelihood function based on the observed implied volatility data $I_1, \ldots, I_m$ and the scaled stock returns $X_1, \ldots, X_n$ is given by

$$(9) \quad (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left\{ -\frac{\sum_{j=1}^{n}(x_j - \nu)^2}{2\sigma^2} \right\}$$
$$\times (2\pi)^{-\frac{m}{2}} \tau^{-m} \exp\left\{ -\frac{\sum_{i=1}^{m}(I_i - \sigma)^2}{2\tau^2} \right\}.$$

We maximize the likelihood to find the maximum likelihood estimator of σ and denote the resulting maximum likelihood estimator by $\hat{\sigma}_I$. The following theorem derives the asymptotic distribution of $\hat{\sigma}_I$.

Theorem 2.1. *The Fisher information for σ based on the combined data is equal to $2n/\sigma^2 + m/\tau^2$, and as $n, m \to \infty$, we have*

$$(10) \quad \left(\frac{2n}{\sigma^2} + \frac{m}{\tau^2} \right)^{1/2} (\hat{\sigma}_I - \sigma) \to N(0, 1).$$

Proof. Since both samples, stock returns, $X_1, \ldots, X_n$, and option data, $I_1, \ldots, I_m$, are i.i.d. normal, and $(X_1, \ldots, X_n)$ and $(I_1, \ldots, I_m)$ are independent, the joint distribution of the two samples follows an exponential family. Define $S_x = \frac{1}{n}(X_j - \bar{X})^2$. From (9) we have the log likelihood

$$l(\sigma, \nu, \tau) \propto$$
$$-n \log \sigma - \frac{\sum_{j=1}^{n}(x_j - \nu)^2}{2\sigma^2} - \frac{\sum_{i=1}^{m}(I_i - \sigma)^2}{2\tau^2} - m \log \tau.$$

It is easy to derive that the MLE of ν is $\bar{X}$, and given σ, the MLE of τ^2 is $\frac{1}{m} \sum_{i=1}^{m}(I_i - \sigma)^2$. We find the derivatives of the log likelihood with respect to σ,

$$\frac{dl}{d\sigma} = -\frac{n}{\sigma} + \frac{\sum_{j=1}^{n}(x_j - \nu)^2}{\sigma^3} + \frac{\sum_{i=1}^{m}(I_i - \sigma)}{\tau^2},$$

$$\frac{d^2l}{d\sigma^2} = \frac{n}{\sigma^2} - \frac{3\sum_{j=1}^{n}(x_j - \nu)^2}{\sigma^4} - \frac{m}{\tau^2}.$$

The Fisher information for σ is equal to

$$-E\left[\frac{d^2l}{d\sigma^2} \right] = -\frac{n}{\sigma^2} + \frac{3\sum_{j=1}^{n}E(X_j - \nu)^2}{\sigma^4} + \frac{m}{\tau^2} = \frac{2n}{\sigma^2} + \frac{m}{\tau^2}.$$

For the exponential family, the regularity conditions are met. Applying the large sample theory of MLE we obtain that as $n, m \to \infty$,

$$\left[\frac{2n}{\sigma^2} + \frac{m}{\tau^2} \right]^{1/2} (\hat{\sigma}_I - \sigma) \to N(0, 1). \qquad \square$$

Theorem 2.1 shows that the Fisher information for σ based on the combined data is equal to the sum of the Fisher information based on the stock data and the Fisher information based on the option data, and thus the MLE for the combined data has asymptotic variance equal to half of the harmonic average of those of the volatility estimators based on either stock return data or option data. It suggests that the MLE for the combined data has smaller asymptotic

variance than the MLEs for either stock data alone or option data alone, and thus the joint estimation based on the combined data is statistically more efficient.

We may also combine the stock price data generated from the geometric Brownian motion model (1) and the option data observed from the non-linear option regression model (5) instead of implied volatility model. The joint likelihood for this case is more complicated, with the form

$$(11)$$
$$(2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{n}(x_j - \nu)^2 \right\}$$
$$\times (2\pi)^{-\frac{m}{2}} \eta^{-m} \exp\left\{ -\frac{1}{2\eta^2} \sum_{i=1}^{m}(C_i - C_{BS}(x_i, \sigma))^2 \right\}.$$

We maximize the likelihood to find the maximum likelihood estimator of σ and denote the resulting maximum likelihood estimator by $\hat{\sigma}$. Similar to the limiting distribution in Theorem 2.1 we may obtain that as $n, m \to \infty$,

$$(12) \quad J_{n,m}^{1/2}(\hat{\sigma} - \sigma) \to N(0, 1),$$

where

$$J_{n,m} = \frac{2n}{\sigma^2} + \frac{\sum_{i=1}^{m} E[(\frac{d}{d\sigma}C_{BS}(x_i; \hat{\sigma}))^2]}{\eta^2}, \quad x_i = (t_i, S_{t_i}, T_i, K_i).$$

Again we see that the asymptotic variance of the volatility estimator for stock return data is $\sigma^2/(2n)$, and that the asymptotic variance of the volatility estimator for option data from model (5) is $\eta^2 / \sum_{j=1}^{m} E[\frac{\partial C_{BS}(x_i, \hat{\sigma})}{\partial \hat{\sigma}}]^2$. The MLE, $\hat{\sigma}$, for the combined data has asymptotic variance equal to the half of the harmonic average of those of the volatility estimators based on either stock return data or option data. It indicates that the MLE for the combined data has smaller asymptotic variance than the MLEs for either stock data alone or option data alone.

3. ESTIMATION UNDER MODELS WITH STOCHASTIC VOLATILITY

Typical models with stochastic volatility are GARCH models and continuous time stochastic volatility models such as bivariate diffusions. GARCH models are convenient for statistical inference, and continuous stochastic volatility models can better handle financial theory such as option pricing. It turns out that there is a link between a discrete GARCH model and a continuous time bivariate diffusion model [Nelson (1990), Wang (2002), and Duan, Wang and Zou (2009)]. This section will illustrate the joint maximum likelihood estimation approach by combining both stock price data and option data. Since Fisher information usually determines the asymptotic variances of the efficient estimators such as MLE, we derive and compare the Fisher information for the three approaches based on stock price data, option data or combined data.

3.1 GARCH and diffusion models

To define a discrete GARCH model over time interval [0, 1], we divide [0, 1] into n subinterval of length $1/n$ and set $t_k = k/n$, $k = 0, 1, \ldots, n$. The GARCH model assumes price process S_{G,n,t_k}, $k = 1, \ldots, n$, to follow

$$(13) \quad \begin{aligned} \log S_{G,n,t_k} - \log S_{G,n,t_{k-1}} = & (\mu - \sigma_{G,n,t_k}^2/2)n^{-1} \\ & + \sigma_{G,n,t_k} n^{-1/2}\epsilon_k, \end{aligned}$$

where μ is drift, ϵ_k is a sequence of i.i.d. standard normal random variables, and the conditional variance (or volatility), σ_{G,n,t_k}^2, depends on lagged errors. Existing GARCH models differ in volatility specification. The representative exponential GARCH(1, 1) model has the following ARMA structured volatility,

$$(14) \\ \begin{aligned} \log \sigma_{G,n,t_k}^2 = & \beta_0 + \beta_1 \log \sigma_{G,n,t_k} + \beta_2\{|\epsilon_{k-1}| - (2/\pi)^{1/2}\} \\ & + \beta_3 \epsilon_{k-1}, \end{aligned}$$

where

$$\begin{aligned} \beta_0 &= \vartheta_0 n^{-1}, \quad \beta_1 = 1 + \vartheta_1 n^{-1}, \\ \beta_2 &= (1 - 2/\pi)^{-1/2}\vartheta_2 n^{-1/2}, \quad \beta_3 = \vartheta_3 n^{-1/2}. \end{aligned}$$

To associate the GARCH model with a diffusion, extend $(S_{G,n,t_k}, \sigma_{G,n,t_k}^2)$ to [0, 1] by letting

$$\begin{aligned} (S_{G,n,s}, \sigma_{G,n,s}^2) &= (S_{G,n,t_k}, \sigma_{G,n,t_k}^2), \\ & \text{for } s \in [t_k, t_{k+1}), k = 0, \ldots, n-1. \end{aligned}$$

Then as $n \longrightarrow \infty$, the GARCH process $(S_{G,n,s}, \sigma_{G,n,s}^2)$ weakly converges to the bivariate diffusion process $(S_{D,s}, \sigma_{D,s}^2)$ governed by the following stochastic differential equation system,

$$(15) \quad d\log S_{D,s} = (\mu - \sigma_{D,s}^2/2)ds + \sigma_{D,s}dW_{1,s},$$

$$(16) \quad \begin{aligned} d\log \sigma_{D,s}^2 = & (\vartheta_0 + \vartheta_1 \log \sigma_{D,s}^2)ds + \vartheta_2 dW_{2,s} \\ & + \vartheta_3 dW_{1,s}, \end{aligned}$$

where $(W_{1,s}, W_{2,s})$ are two independent standard Brownian motions. The diffusion process $(S_{D,s}, \sigma_{D,s}^2)$ is referred to as the diffusion limit of the exponential GARCH model (13)–(14). See Nelson (1990) and Wang (2002).

3.2 Estimation for return data

We consider an exponential GARCH model with the following specification,

$$(17) \quad X_{t_k} = \log S_{G,n,t_k} - \log S_{G,n,t_{k-1}} = \sigma_{G,n,t_k} n^{-1/2}\epsilon_k,$$

$$(18) \\ \begin{aligned} \log \sigma_{G,n,t_k}^2 = & \beta_0 + \beta_1 \log \sigma_{G,n,t_{k-1}}^2 + \beta_2\{|\epsilon_{k-1}| - (2/\pi)^{1/2}\} \\ & + \beta_3 \epsilon_{k-1}. \end{aligned}$$

where

$$\begin{aligned} \beta_0 &= \vartheta_0 n^{-1}, \quad \beta_1 = 1 + \vartheta_1 n^{-1}, \\ \beta_2 &= (1 - 2/\pi)^{-1/2}\vartheta_2 n^{-1/2}, \quad \beta_3 = \vartheta_3 n^{-1/2}. \end{aligned}$$

For model (17)–(18), the log likelihood function is of the following form

$$(19) \quad l_n(\vartheta) = -\frac{1}{2}\sum_{k=1}^{n}\left(\log \sigma_{G,n,t_k}^2 + \frac{nX_{t_k}^2}{\sigma_{G,n,t_k}^2}\right) - \frac{n}{2}\log(2\pi),$$

where $\vartheta = (\vartheta_0, \vartheta_1, \vartheta_2, \vartheta_3)$. We may maximize the likelihood function to find the MLE of ϑ. The Fisher information for ϑ is given by

$$(20) \quad - E\left[\frac{\partial^2 l_n(\vartheta)}{\partial \vartheta^2}\right],$$

where $\frac{\partial^2 l_n(\vartheta)}{\partial \vartheta^2}$ denotes the second partial derivative matrix of $l_n(\vartheta)$ with respect to ϑ.

3.3 Estimation for option data

For the bivariate diffusion model (15)–(16), we have the following formula for the price of the European call option (with stock price S at time t, strike price K, maturity duration $T - t$, interest rate r, and initial value $\sigma_0 = y$),

$$C_D(t, S; T, K; y) = E\left[C_{BS}\left(t, SM_D; T, K; \sqrt{\overline{\sigma_D^2}}\right)\right],$$

where C_{BS} is the Black-Scholes formula in (2) and

$$\begin{aligned} \overline{\sigma_D^2} &= \frac{1 - \rho^2}{T - t}\int_t^T \sigma_{D,s,t,y}^2 ds, \\ M_D &= \exp\left(\rho\int_t^T \sigma_{D,s,t,y}dB_{V,s} - \frac{\rho^2}{2}\int_t^T \sigma_{D,s,t,y}^2 ds\right), \\ B_{V,s} &= \rho W_{1,s} + sign(\vartheta_2)\sqrt{1 - \rho^2}W_{2,s}, \quad \rho = \frac{\vartheta_3}{\sqrt{\vartheta_2^2 + \vartheta_3^2}}, \end{aligned}$$

$\sigma_{D,s,t,y}$ denotes $\sigma_{D,s}$ beginning at $\sigma_{D,t} = y$, and the expectation is taken with respect to the random source $B_{V,s}$. See Fouque et al. (2000), Heston (1993), Hull and White (1987), Scott (1987), Stein and Stein (1991), and Wiggins (1987). Furthermore, we have the following approximation for the option price,

$$(21) \\ \begin{aligned} C_D(t, S; T, K; y) \approx & C_{BS}(t, SM_D; T, K; \bar{\sigma}) \\ & - (T - t)\left(V_2 S^2\frac{\partial^2 C_{BS}}{\partial^2 S^2} + V_3 S^3\frac{\partial^3 C_{BS}}{\partial S^3}\right), \end{aligned}$$

and using (21) to solve $C_D(t, S; T, K; y) = C_{BS}(t, SM_D; T, K; I_D)$ we obtain an approximate implied volatility I_D as follows,

$$(22) \quad I_D \approx a + b\frac{\log(K/S)}{T - t},$$

where a and b are given by

(23) $\quad a = \bar{\sigma} + \dfrac{V_3}{\bar{\sigma}^3}\left(r + \dfrac{3}{2}\bar{\sigma}^2\right) - \dfrac{V_2}{\bar{\sigma}}, \qquad b = \dfrac{V_3}{\bar{\sigma}^3},$

and

$$
\begin{aligned}
\bar{\sigma}^2 &= \exp\left(-\frac{\theta_0 + \theta_2^2/4}{\theta_1}\right), \\
\frac{V_3}{\bar{\sigma}^3} &= \frac{\rho}{\theta_2}\left[\exp\left(\frac{\theta_2^2}{16\theta_1}\right) - \exp\left(-\frac{3\theta^2}{16\theta_1}\right)\right], \\
\frac{V_2}{\bar{\sigma}} &= \frac{2\rho}{\theta_2}\exp\left(-\frac{\theta_0 + \theta_2^2/4}{\theta_1}\right)\left[\exp\left(\frac{\theta_2^2}{16\theta_1}\right)\right. \\
&\quad \left. - \exp\left(-\frac{3\theta_2^2}{16\theta_1}\right)\right], \\
\theta_0 &= \vartheta_0, \theta_1 = \vartheta_1, \theta_2 = \sqrt{\vartheta_2^2 + \vartheta_3^2}, \rho = \frac{\vartheta_3}{\sqrt{\vartheta_2^2 + \vartheta_3^2}}.
\end{aligned}
$$

See Fouque et al. (2000, chapter 5).

Formula (22) provides an approximate relationship of the implied volatility I_D with (t, S, T, K). It suggests that the observed implied volatility may obey the following model,

(24) $\quad I_i = a + b\dfrac{\log(K_i/S_i)}{T_i - t_i} + z_i, \quad i = 1, \ldots, m,$

where a and b are defined as above and z_i's are independent normal random variables with mean zero and variance τ^2. Let $Y_i = \frac{\log(K_i/S_i)}{T_i - t_i}$. Then model (24) is a simple linear regression model with parameters a and b, and the log-likelihood function is given by

(25) $\quad l_m(a,b) = \begin{aligned}[t] &-\frac{1}{2\tau^2}\sum_{i=1}^m [I_i - (a + bY_i)]^2 \\ &-m\log\tau - \frac{m}{2}\log(2\pi).\end{aligned}$

Denote the MLE of (a, b) by $(\hat{a}, \hat{b})$. Then

$$\hat{a} = \bar{I} - \hat{b}\bar{Y}, \quad \hat{b} = \frac{\sum_{i=1}^m Y_i(I_i - \bar{I})}{\sum_{i=1}^m Y_i(Y_i - \bar{Y})}.$$

Note that a and b are functions of $\vartheta_0, \vartheta_1, \vartheta_2,$ and ϑ_3 (or ρ), which are denoted by $a(\vartheta_0, \vartheta_1, \vartheta_2, \vartheta_3)$ and $b(\vartheta_0, \vartheta_1, \vartheta_2, \vartheta_3)$, respectively, to stress their dependence on $\vartheta = (\vartheta_0, \vartheta_1, \vartheta_2, \vartheta_3)$, and we can estimate only a and b from model (24). Suppose that $a(\vartheta_0, \vartheta_1, \vartheta_2, \vartheta_3)$ and $b(\vartheta_0, \vartheta_1, \vartheta_2, \vartheta_3)$ are differentiable, and assume that we know ϑ_0 and ϑ_3 (or ρ). With the given ϑ_0 and ϑ_3, and the MLE, $(\hat{a}, \hat{b})$, of (a, b) from model (24) we may use the invariant principle of MLE to find the MLE of $(\vartheta_1, \vartheta_2)$. Denote by $(\hat{\vartheta}_1, \hat{\vartheta}_2)$ the MLE of $(\vartheta_1, \vartheta_2)$ based on $I_1, \ldots, I_m$ from model (24).

The Fisher information for (a, b) is equal to

$$\frac{1}{\tau^2}\begin{pmatrix} m & \sum_{i=1}^m EY_i \\ \sum_{i=1}^m EY_i & \sum_{i=1}^m EY_i^2 \end{pmatrix},$$

and the Fisher information for $(\vartheta_1, \vartheta_2)$ is given by

(26) $\dfrac{1}{\tau^2}\,\dfrac{\partial(a,b)}{\partial(\vartheta_1,\vartheta_2)}\begin{pmatrix} m & \sum_{i=1}^m EY_i \\ \sum_{i=1}^m EY_i & \sum_{i=1}^m EY_i^2 \end{pmatrix}\left(\dfrac{\partial(a,b)}{\partial(\vartheta_1,\vartheta_2)}\right)',$

where $\frac{\partial(a,b)}{\partial(\theta_1,\theta_2)}$ denotes the partial derivative matrix of (a, b) with respect to $(\vartheta_1, \vartheta_2)$.

3.4 Estimation for both return data and option data

Option pricing for GARCH models is very hard. For example, GARCH option pricing has no explicit formula such as the one described in Section 3.3. As illustrated in Section 3.1, the limit of a GARCH model is a bivariate diffusion, and the European option prices of the GARCH model converge to the European option prices of its diffusion limit given in Section 3.3 [see Duan, Wang and Zou (2009)]. Thus, we may consider joint parameter estimation for both return data from model (17)–(18) and option data from model (24), where the price process S_i in model (22) are equal to S_{G,n,t_i} in (17).

By conditional arguments, we combine the stock data from the GARCH model (17)–(18) and the option data from the implied volatility model (24) and derive the joint log likelihood for $\vartheta = (\vartheta_0, \vartheta_1, \vartheta_2, \vartheta_3)$ as follows,

(27)
$$
\begin{aligned}
l_{m,n}(\vartheta) = &-\sum_{i=1}^m \frac{[I_i - (a + bY_i)]^2}{2\tau^2} - m\log\tau \\
&-\frac{1}{2}\sum_{k=1}^n\left(\log\sigma_{G,n,t_k}^2 + \frac{nX_{t_k}^2}{\sigma_{G,n,t_k}^2}\right) - \frac{m+n}{2}\log(2\pi),
\end{aligned}
$$

where (a, b) are functions of ϑ described by (23). We maximize the log likelihood to find the MLE of ϑ. From (19), (25) and (27) we see that $l_{m,n}(\vartheta) = l_n(\vartheta) + l_m(a(\vartheta), b(\vartheta))$, which indicates the Fisher information for ϑ based on the combined data is equal to the sum of the Fisher information matrix (20) for the stock data and

$$\frac{1}{\tau^2}\,\frac{\partial(a,b)}{\partial(\vartheta_0,\vartheta_1,\vartheta_2,\vartheta_3)}\begin{pmatrix} m & \sum_{i=1}^m EY_i \\ \sum_{i=1}^m EY_i & \sum_{i=1}^m EY_i^2 \end{pmatrix}\left(\frac{\partial(a,b)}{\partial(\vartheta_0,\vartheta_1,\vartheta_2,\vartheta_3)}\right)',$$

which is an extension of the information matrix (26) for the option data, where $\frac{\partial(a,b)}{\partial(\vartheta_0,\vartheta_1,\vartheta_2,\vartheta_3)}$ denotes the partial derivative matrix of (a, b) with respect to $(\vartheta_0, \vartheta_1, \vartheta_2, \vartheta_3)$. We conclude that the Fisher information relationship among those different data sources for the varying volatility models is similar to the case of constant volatility, and the joint estimation based on the combined data is statistically more efficient.

4. A SIMULATION STUDY

We conducted a simulation study to check the finite sample performances of the proposed estimators under the Black-Scholes set-up. We simulated stock prices over a year

Table 1. Simulation results for volatility estimation based on 1000 replications

Model	$\mathrm{Var}(\hat{\sigma})$	Error	$\mathrm{Mean}(\hat{\sigma})$	$\mathrm{Var}(\hat{\sigma})$
Stock ($n = 252$)	3.0386×10^{-5}		0.12348	3.0267×10^{-5}
Implied Volatility	5×10^{-5}	Normal	0.12400	5.14×10^{-5}
($m = 200$)		Uniform	0.12400	4.83×10^{-5}
		Exponential	0.12400	4.94×10^{-5}
		t (df = 10)	0.12402	6.21×10^{-5}
B-S	2.843×10^{-8}	Normal	0.12375	2.7129×10^{-8}
($m = 200$)		Uniform	0.12375	2.3935×10^{-8}
		Exponential	0.12375	2.9317×10^{-8}
		t (df = 10)	0.12375	3.4969×10^{-8}
Stock & Implied volatility	4.919×10^{-7}	Normal	0.12401	4.823×10^{-7}
Stock & B-S volatility	2.466×10^{-8}	Normal	0.12376	2.3169×10^{-8}

(252 trading days) from model (1) with a mean return of 1.2841 and a standard deviation of 0.12375 annually. We calculated $\mathrm{Var}(\hat{\sigma})$ to be $\frac{0.12375^2}{2 \times 252} = 3.0386 \times 10^{-5}$, and repeated the simulation procedure 1,000 times. The simulated results are summarized in the second row of Table 1. From the table we see that the mean and variance of $\hat{\sigma}_S$ based on the 1,000 repetitions are almost equal to their corresponding true values.

The values of options placed upon this stock were recorded. These options have a life remaining between three and six months until expiration, and the strike price is some value close to the stock price, a certain percentage either above or below the current amount. The risk free rate of interest is set to be 3.31%. We simulated 200 option data from each of the option model (5) and the implied volatility model (7). Four random errors considered in the models are $N(0, \tau^2)$, $U(-\frac{\sqrt{0.0012}}{2}, \frac{\sqrt{0.0012}}{2})$, $E(-0.01, 100)$, and t_{10}, where $\tau = 0.01$, and $\eta = 0.04$. The simulation procedure was repeated 1,000 times, and the simulation results are displayed in the third and fourth rows of Table 1. The simulation indicates that the averages of $\hat{\sigma}_{BS}$ and $\bar{I}$ based on the 1,000 repetitions are very close to the true value 0.12375, with very small variances for all four error distributions.

Finally we generated together stock price data over a year from model (1) and 200 option data from each of the option model (5) and the implied volatility model (7) as described in Section 2.2, where the model parameters σ, η, τ are the same as above. As the results are very similar, we considered only normal random errors in models (5) and (7). According to the description in Section 2.2 we combined the stock data with 200 option data from either (5) or (7) to compute estimators $\hat{\sigma}_I$ and $\hat{\sigma}$, respectively. The last two rows of Table 1 show the simulation results for the volatility estimator with combined stock price data and option data. Again the simulation shows excellent performances of $\hat{\sigma}_I$ and $\hat{\sigma}$, with their averages over the 1000 repetitions very close to the true value and extremely small variances. In comparison with estimators based on either stock data or option data, the combined estimators $\hat{\sigma}_I$ and $\hat{\sigma}$ have much smaller variances than estimator $\hat{\sigma}_S$ based on stock data alone and the

respective volatility estimators $\bar{I}$ and $\hat{\sigma}_{BS}$ based on option data alone. The simulation findings confirm the theoretical results established in Section 2.

5. CONCLUSION

Volatility analysis in empirical financial studies is often developed for either historical price data or option data. As option prices heavily depend on the underlying stocks' prices, and financial models for the two kinds of data are highly related, in this paper we study volatility analysis by combining historical stock price data and option data. We consider the Black-Scholes model and an exponential GARCH model and establish the relationship among the Fisher information for volatility estimation based on stock price data alone or option data alone and for joint volatility estimation based on combining stock price data and option data. An asymptotic theory for the joint estimation under the Black-Scholes model is provided. A simulation study is conducted to check finite sample performances of the proposed joint estimation.

REFERENCES

AÏT-SAHALIA, Y. (1996). Nonparametric pricing of interest rate derevative securities. *Econometrica* **64**, 527–560.

BLACK, F. and SCHOLES, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* **81**, 637–659.

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327. MR0853051

BOLLERSLEV, T., CHOU, R. Y. and KRONER, K. F. (1992). ARCH modeling in finance. A review of the theory and empirical evidence. *Journal of Econometrics* **52**, 5–59.

DOTHAN, M. U. (1990). *Prices in Financial Markets.* Oxford University Press. MR1070714

DROST, F. C. and KLAASSEN, C. A. J. (1997). Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* **81**, 193–221. MR1484585

DUAN, J. C. (1995). The GARCH option pricing model. *Mathematical Finance* **5**, 13–32. MR1322698

DUAN, J. C. (1997). Augmented GARCH(p, q) process and its diffusion limit. *Journal of Econometrics* **79**, 97–127. MR1457699

DUAN, J. C., WANG, Y. and ZOU, J. (2009). Convergence speed of GARCH option price to diffusion option price. *International Journal of Theoretical and Applied Finance* **12**, 359–391. MR2536492

DUFFIE, D. (1992). *Dynamic Asset Pricing Theory.* Princeton University Press, Princeton, N.J. Second Edition.

DUMAS, B., FLEMING, F. and WHALEY, R. (1998). Implied volatility functions: Empirical tests. *Journal of Finance* **53**, 2059–2106.

ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007. MR0666121

ENGLE, R. F. and BOLLERSLEV. T. (1986). Modeling the persistence of conditional variances (with discussion). *Econometric Review* **5**, 1–50. MR0876792

ENGLE, R. F. and NG, V. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance* **48**, 1749–1778.

FOUQUE, J. P., PAPANICOLAOU, G. and SIRCAR, K. R. (2000). *Derivatives in Financial Markets with Stochastic Volatility.* Cambridge University Press. MR1768877

GOURIÉROUX, C. (1997). *ARCH Models and Financial Applications.* Springer. MR1439744

GEWEKE, J. (1986). Modeling the persistence of conditional variances: A comment. *Econometric Review* **5**, 57–61. MR0876792

HARRISON, M. and PLISKA, S. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Its Application* **11**, 215–260. MR0622165

HESTON, S. (1993). A closed form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**, 327–343.

HULL, J. (1997). *Options, Futures, and Other Derivatives.* Princeton, N.J. Third edition.

HULL, J. and WHITE, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance* **42** 281–300.

LAI, T. L. (1994). Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *Ann. Statist.* **22**, 1917–1930. MR1329175

LI, H., WELLS, M. and YU, C. (2008). A Bayesian analysis of time-changed Lévy processes of return dynamics. *Review of Financial Studies* **21**, 2345–2378.

LYUU, Y. and WU, C. (2005). On accurate and provably efficient GARCH option pricing algorithms. *Quantitative Finance* **5**, 181–198. MR2240245

MELINO, A. and TURNBULL, S. (1990). Pricing foreign currency options with stochastic volatility. *J. Econometrics* **45**, 239–265.

METRON, R. (1973). The theory of rational option pricing. *Bell Journal of Economics and Management Science* **4**, 141–183. MR0496534

NELSON, D. B. (1992). Filtering and forecasting with misspecified ARCH models. I. Getting the right variance with the wrong model. *Journal of Econometrics* **52**, 61–90. MR1165644

RITCHKEN, P. and TREVOR, R. (1999). Pricing options under generalized GARCH and stochastic volatility processes. *Journal of Finance* **54**, 377–402.

ROSSI, P. E. (1996). *Modeling Stock Market Volatility.* Academic Press. MR1434088

SCOTT, L. (1987). Option pricing when the variance changes randomly: Theory, estimation and an application. *J. Financial and Quantitative Analysis* **22**, 419–438.

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

WANG, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *The Annals of Statistics* **30**, 754–783. MR1922541

WIGGINS, J. (1987). Option values under stochastic volatility: Theory and empirical estimates. *J. Financial Economics* **19**, 351–372.

YU, C., LI, H. and WELLS, M. (2011). MCMC estimation of Lévy jump models using stock and option prices. *Mathematical Finance* **21**, 383–422. MR2830427

Yi Liu
Department of Statistics
University of Wisconsin
Madison, WI 53706
USA
E-mail address: liuyi@stat.wisc.edu

Yazhen Wang
Department of Statistics
University of Wisconsin
Madison, WI 53706
USA
E-mail address: yzwang@stat.wisc.edu