國立交通大學

統計學研究所

博 士 論 文

關於選擇權定價的三則短論

A Petit Trio on Option Pricing

研 究 生：牛維方

指導教授：盧鴻興 教授

中 華 民 國 九 十 七 年 七 月

關於選擇權定價的三則短論

學生：牛維方　　　　指導教授：盧鴻興

國立交通大學統計學研究所 博士班

摘要

選擇權定價是當代財務科學最重要也最引人入勝的的課題之一，各種複雜的價格模型與演算方法均可能派上用場，讓不同背景但同樣精於計量方法的學者在此一展身手。

在實務上，定價也是多種套利交易策略的核心技術。為求精確的定價，通常需要使用較複雜且具有結構的模型，其中之一就是連續時間隨機波動率模型。然而，這類模型中重要的狀態變數-波動率實際上是不可觀察的，同時整個系統的概率函數也無法以解析形式表達。

這一系列論文探討利用隨機波動率模型進行選擇權定價相關的統計問題，特別將波動率不可觀察這一問題納入考量。基於 GARCH模型會收斂到隨機波動率模型這一性質，只要建構一個僅可部分觀察的GARCH模型，其概率函數可運用蒙地卡羅馬可夫鍊方法計算，且為原有模型之良好近似，因此可用於相關的統計推論，包括估計與定價。在此脈絡下，既有做法的一些不足與缺失均可獲得改善。

最後針對選擇權定價中的損失函數也作一討論。簡言之，雖然定價理論並未對選擇損失函數設下絕對標準，但損失函數必須基於足以表現選擇權價格資訊內涵的統計量(如隱含波動率)，而由於定價均基於均衡的論述，由模型獲得的該統計量僅需作為實繼價格獲得的統計量進行均數回歸的目標，而非價格的絕對參照。

A Petit Trio on Option Pricing

Student: Wei-Fang Niu   Advisor：Henry Horng-Shing Lu

Institute of Statistics

National Chiao Tung University

ABSTRACT

Option pricing may be one of the most important and fascinating topics in modern finance. Complex models and algorithms can be applied here so that researchers and practitioners may bring their quantitative skills into full play.

In practice, pricing is also the core of different types of arbitrage strategies. For more precise pricing, structural models are generally necessary where continuous time stochastic volatility model can be one of the candidates. However, volatility, the most important state variable, is in fact unobservable and the likelihood cannot be available in close form for the stochastic volatility models.

The sets of articles explore related statistical issues about option pricing with stochastic volatility models. Especially, the unobservability of volatilities is taken into considerations. By the fact that a GARCH model would converge weakly to the corresponding stochastic volatility model, statistical inference including estimation and pricing can be made based on a specially designed partially observed GARCH model whose likelihood will be obtained through MCMC methods. In this context, some drawbacks from the current practices can be improved.

Finally, an investigation on the loss functions for option pricing is also made. Although the pricing theory does not restrict to any specific loss functions, the statistics that correspond to the information contents, for example implied volatilities, should be used as a basis for the construction of loss functions. Furthermore, due to the fact that pricing is generally based on some equilibrium conditions, the model implied statistics would just play as the target of mean reversion of the real price implied process, instead of an absolute reference of prices.

誌謝

能夠完成這份論文，首先我要感謝李昭勝老師，是他帶領我進入財金這個充滿機會的領域，並極度寬容我的諸多無禮；而這份論文沒能在李老師在世時完整呈獻，一直是我心中至大的缺憾。衷心期盼這份論文的付梓，能稍稍告慰老師的在天之靈。

特別感謝盧鴻興老師在論文最後階段這一年半期間的大力協助，包括提供我各個角度的思維，及拓展我的學術視野，它們豐富了論文的廣度及深度。交大統計所的洪志真老師、陳志榮老師、洪慧念老師、陳鄰安老師及所辦公室郭碧芬小姐，在這七年間無私地提供知識、生活及各種繁瑣事務上的協助，讓我萬分感激；而財金所鍾惠民老師及王克陸老師不時的鼓勵，亦讓我銘感在心。

另外，我也必須感謝鉅融資本管理的執行長鄭振和博士，引領我更貼切地觀察金融與經濟，對於研究的主題有更深入觀點，當然，也要感謝王南傑先生、林詩珊小姐及眾多朋友的協助。

這份論文當中許多想法來自多年前在交大李昭勝老師的辦公室中與段錦泉教授的討論，在此特別對他說聲謝謝，同時恭喜他當選中研院院士。

這七年間所遇到學長姐、同學與學弟妹已難以歷數，很高興在最後一段學生生涯可以結識他們 – 包括許英麟、林淑惠、黃景業、吳漢銘，阿賓、芒果、顏家玲、闕棟鴻、小強及其他許許多多一同打混、打球的統計人，感謝他們，因為他們讓這七年不至於太難熬，因為他們讓我有堅持下去的勇氣。

當然，還是要謝謝幾位口試委員。事實上，這些文章寫好後並沒有經過太多深入的討論與檢視，感謝他們在口試過程中提供的意見，特別是姚怡慶老師及徐南蓉老師。

最後，我特別要感謝我的內人羅慧綺，在這七年來無怨無悔地給予我種種支持。感謝我的家人 – 我的母親陳玉鸞女士、岳父母及我和羅慧綺的兄弟姐妹們。

感謝以上提及與當提及卻未提及的每一個人，有了他們，然後才有我這一點小小的成就。

Contents

Table Contents

Figure Contents

1. Introduction

Since the epochal work of Black and Scholes (1973) and Merton (1973), financial derivative markets grew up rapidly and quantitative methods were widely used in finance. To lax the assumptions of the Black-Scholes model, large amounts of complex structural models were introduced for option pricing, including the so called continuous time stochastic volatility models.

These types of models have been widely utilized in industries and academia. However, some features and difficulties make the currently popular practices about these models desultory. First, the likelihood functions for such systems are seldom available and thus some indirect methods such as EMM and indirect inference shall be applied. But to the extent searches, the effectiveness and efficiency for these methods applied to the stochastic volatility models are not formally concluded.

Next, as the volatility processes are in fact never observable, an “estimate” for volatility at each point in time should be filtered with the prices. However, if this approach really makes sense, why not just use GARCH-type models? Third, an exogenous loss function is generally required for the determinations of some subset of parameters, especially risk premium of volatility. What matters is that almost no commonly used loss functions are provided with significant economic interpretation and statistical properties.

This dissertation is engaged in the problems mentioned above. By introducing a partially observed GARCH model and utilizing MCMC method, the likelihood function for the stochastic volatility model can be properly approximated numerically. The byproduct of simulated likelihood would be a large amount of paths that connecting the observations, which can be used as a basis for further inferences such as predicting or pricing.

As for loss functions, since the option prices may range widely, commonly used error sum of squares or similar measures cannot be an adequate candidate. In fact, the pricing theory tells no specific loss functions. But there are at least two questions to be answered for the choice of loss functions. First, what are the information contents of option prices? Second, what characteristics will be represented by the information contents? The answers clearly lead to a reasonable choice of loss functions.

In Chapter 2 the estimation method for the stochastic volatility models is discussed. Following the context above, calculations for option prices become trivial. Some interesting results are thus documented in Chapter 3. Beyond the apparent prejudice about “losses” for pricing options, outlines for choosing proper loss functions are drawn in Chapter 4. These investigations and discussions complete a schema for statistical inference about option pricing.

2. Estimating the continuous time stochastic volatility models with partially observed GARCH models

2.1 Introduction

In the last decade, Continuous time stochastic volatility models have been proposed in finance area, especially option pricing. These models generally consist of two stochastic differential equations,

$$dX_t = \mu_X(X_t, V_t, t)dt + \sigma_X(X_t, t) \cdot \sqrt{V_t} \cdot dW_{1,t} \quad , \tag{2.1}$$

$$d\phi(V_t) = \mu_V(V_t, t)dt + \sigma_V(V_t, t)dW_{2,t}, \tag{2.2}$$

The first equation involves the dynamics of X_t, the logarithm of the prices S_t, and the second the volatilities V_t. Here $(W_{1,t}, W_{2,t})$ is assumed to be a correlated two-dimensional Brownian motion and $\theta \in \boldsymbol{H} \subseteq \boldsymbol{R}^d$ is the unknown parameter. This class of models covers a large range of applications in literatures concerning equity prices and interest rates, for example, Heston (1993) and Anderson and Lund (1997). Like most practical applications of diffusion models, observations for continuous time stochastic volatility models are taken at discrete time and generally neither explicit forms of the transition probability functions nor the likelihood functions are available. Furthermore, the volatilities are in fact unobservable here and thus there are always missing values at least as many as the observations. Therefore, likelihood inference for these models turns out to be much more complicated.

In this chapter, a procedure is proposed for approximating the likelihood functions of continuous time stochastic volatility models in which the two driving processes are correlated. The major idea here is to fit the discretely observed data drawn from a bivariate diffusion process with a partially observed GARCH process whose frequency of construction is much higher. The approximated density functions will converge to the joint density function of the model (2.1) and (2.2) under certain conditions and thus can be used as a basis for statistical inference. However, such practice will induce large amounts of missing values between observations. With Markov Chain Monte Carlo (MCMC) techniques, paths conditional on the observations can be sampled to compute the approximate likelihood function based on the GARCH model. Then the EM algorithm can be applied to get the MLE's.

Among literatures estimators based on the generalized method of moments (GMM, Hansen, 1982) are often used in estimating parameters of the stochastic volatility models. Especially, there are a large number of literatures with the efficient methods of the moments (EMM, Gallant and Tauchen, 1996) as a major tool for inferences. For example, Anderson and Lund (1997) discussed a stochastic volatility model for interest rates and EMM with a semi-nonparametric (SNP) density, based on GARCH score generators, is used to estimate the continuous time model. Indirect inference proposed by Gourieroux et al. (1994) provides another approach, and a practical application can be referred to Fiorentini et al. (2002). Besides, under the assumptions that the two Brownian motions driving the price and the volatility processes are independent, in Genon-Catalot et al. (1999) the return is treated as a subordinated process with which the moment conditions can be derived.

Estimating functions provide another approach to this estimation problem, for example, Kessler (2000) and Sørensen (1999). Also there are authors who drew inferences via a Bayesian method based on Euler approximations to diffusions, for example, Eraker (1998).

On the side of likelihood inference, simulated likelihood method for diffusion models with discrete observed data is first proposed in Pedersen (1995). Likelihood functions can be numerically approximated through simulating paths between observed points under the Euler scheme. Importance sampling techniques play important role in implementing the idea, and various strategies for generating paths have been suggested, for example, Kessler (1997), Elerian (1998) and Elerian et al. (2001).

When the two processes are assumed to be uncorrelated the returns are still normally distributed. With this property, the likelihood functions can be approximated by simulated paths of volatilities, for example Sørensen (2003). But unfortunately this method is also not applicable when the two driving processes ($W_{1,t}$,$W_{2,t}$) are assumed correlated and the resulting distributions for the returns are not available.

On the other hand, since Bollerslov (1986) GARCH models provide as alternative models that describe heteroskedasticity in financial time series. It has been shown in Nelson (1990) that a GARCH model will converge weakly to some bivariate diffusion model as time interval becomes infinitesimal. A family of augmented GARCH(1,1) processes and their limiting diffusion processes are proposed by Duan (1997). Another family of CEV-ARCH models and their limiting processes can be found in Fornari and Mele (2004).

From the viewpoint of statistical inference, Wang (2002) showed asymptotic non-equivalence of GARCH and continuous time stochastic volatility models at the basic frequency of construction. On the other hand, similar to Pedersen(1995) on approximating diffusions with Euler expansion, Brown, Wang and Zhao (2003) suggested that these two models are asymptotically equivalent at frequencies lower than the square root of the basic frequency of construction.

These results point out alternative access to approximate of the likelihood function for the continuous time stochastic volatility model beyond the traditional Euler and Milstein schemes. In short, to inference about a stochastic volatility model is equivalent to inference about a GARCH model constructed at a higher frequency.

The rest of the chapter is organized as follows. In section 2.2, background knowledge about the model is reviewed. Major results will be stated in section 2.3. Numerical illustrations are given in section 1.4 and then in section 2.5 are conclusions.

2.2 The model and likelihood inference about diffusions

The stochastic volatility models have been presented in the introduction. Generally, equation (2.1) just corresponds to the commonly used geometric Brownian motion or the CKLS model with coefficient of volatility replaced by the square root of a positive random variable V_t. Since in documented literatures, for example Engle and Patton (2001), it is suggested that volatilities possess the property of mean reversion, equation (2.2) is usually specified with a term like $\mu_V(V_t,t) = -\kappa(V_t - \xi)$, in which κ represents the speed of mean reversion and ξ is the long-term equilibrium level of volatility.

For further discussions, the following conditions are assumed.

Assumption 1. $\mu_X(X_t,V_t,t)$, $\sigma_X(X_t,t)$, $\mu_V(V_t,t)$ and $\sigma_V(V_t,t)$ are functions which satisfy regularity conditions such that there exists a unique strong solution (X_t, V_t) of the systems (2.1) and (2.2).

Assumption 2. The process V_t is stationary and ergodic with a distribution π.

With these conditions, the stationary distribution for V_0 and the transition probability density function for X_t, $V_t| X_{t-1}$, V_{t-1} exist. Then the likelihood function can be expressed in a proper form.

2.2.1 Likelihood function

Data for this problem consist of discrete-time observations $X_0, X_1, \cdots, X_T$ only, while the process V_t remains completely unobservable. With the Markovian property of the diffusion process, the likelihood function can be expressed as

$$L(\eta;\widetilde{X}^{(1)}) = \int \pi(V_0)\prod_{t=1}^{T} p(X_t, V_t \mid X_{t-1}, V_{t-1};\eta)d\widetilde{V}^{(1)},$$

$$= \int \pi(V_0)\, p(\widetilde{X}^{(1)};\eta)dV_0 \qquad (2.3)$$

where η denotes the vector of parameters, $\widetilde{X}^{(1)} = (X_0,\cdots,X_T)$, $\widetilde{V}^{(1)} = (V_0,\cdots,V_T)$,

$p(X_t, V_t| X_{t-1}, V_{t-1};\eta)$ the transition probability density function, $p(\widetilde{X}^{(1)};\eta)$ the joint density function of $\widetilde{X}^{(1)}$ and $\pi(V_0)$ the stationary distribution for V_0.. Taking expectation in (2.3) is due to that $\widetilde{V}^{(1)}$ is not observable. Note that in general the transition density functions are quite complicated so the integral cannot be decomposed into factors concerning pivotal quantities such as $X_t - X_{t-1}$. This means that, when $\widetilde{V}^{(1)}$ is not observable, X_t will be dependent on $(X_0,\cdots,X_{t-1})$ rather than only on X_{t-1}.

Explicit forms of transition probabilities have been identified only for some specific univariate processes, for example the Ornstein-Uhlenbeck process and the Cox-Ingersoll-Ross process. However, it is almost infeasible to find explicit form of transition densities for the whole system (2.1) and (2.2) even when $W_{1,t}$ and $W_{2,t}$ are not correlated.

A practical means to compute the likelihood functions and find MLEs is through numerical methods. When equation (2.1) is set as a geometric Brownian motion for prices and $W_{1,t}$ and $W_{2,t}$ are assumed to be independent, each increment $X_t - X_{t-1}$ is normally distributed and the likelihood function can be obtained through simulating large number of paths of volatilities (Sørensen, 2003).

2.2.2 Simulated maximum likelihood and MCMC methods for inferences about diffusions

When observations for all processes are available, Pedersen (1995) shows the approximate likelihood under the Euler expansion converges to the true likelihood function in probability as the subdivision length between observations approaches 0. That is, to obtain a good approximation of the likelihood function data augmentation is necessary and different paths connecting two consecutive observations should be simulated.

Consider the model consisting of equation (2.1) with V_t as a constant. The discretized version with subdivision length $\Delta=1/n$ would be

$$X_{t+\Delta} = X_t + \mu_X(X_t, V_t, t)\Delta + \sigma_X(X_t, t)\Delta W_t, \quad (2.4)$$

where ΔW_t is a normally distributed random variable with mean 0 and variance Δ. Let the observed data be $\widetilde{X}^{(1)} = (X_0, \cdots, X_T)$ and denote the augmented data as

$$^*\widetilde{X}^{(n)} = (X_0, {}^*\widetilde{X}_1^{(n)}, X_1, \cdots, X_T)$$

$$= (X_0, {}^*X_\Delta, \cdots, {}^*X_{1-\Delta}, X_1, \cdots, X_T),$$

where ${}^*\widetilde{X}_t^{(n)} = ({}^*X_{t-1+\Delta}, \cdots, {}^*X_{t-\Delta})$ is the n-1 augmented data points lying between X_{t-1} and X_t.

The joint density function for ${}^*\widetilde{X}^{(n)}$ under the Euler Scheme is

$$q({}^*\widetilde{X}^{(n)}; \eta) = \prod_{t=1}^{T} q\left(X_t, {}^*X_t^{(n)} \mid X_t; \eta\right)$$

$$= \prod_{s=1}^{nT} \left[\frac{1}{\sqrt{{}^*\sigma_{s\Delta}}} \phi({}^*z_{s\Delta}) \right],$$

where

$${}^*z_{s\Delta} = \left({}^*X_{s\Delta} - {}^*X_{(s-1)\Delta} - \mu_S({}^*X_{(s-1)\Delta}, {}^*V_{(s-1)\Delta}, (s-1)\Delta)\right) \Big/ {}^*\sigma_{(s-1)\Delta},$$

$${}^*\sigma_s = \sigma_X({}^*X_s, t),$$

and $\phi(\cdot)$ is the density function of the standard normal distribution.

Since $({}^*\widetilde{X}_1^{(n)}, \cdots, {}^*\widetilde{X}_T^{(n)})$ are auxiliary variables that are not observed, the likelihood function for $\widetilde{X}^{(1)}$ would be

$$L(\eta \mid \widetilde{X}^{(1)}) = \lim_{n\to\infty} \prod_{t=1}^{T} q^{(n)}\left(X_t \mid X_{t-1}; \eta\right)$$

$$= \lim_{n\to\infty} \int \prod_{t=1}^{T} q\left(X_t, {}^*\widetilde{X}_t^{(n)} \mid X_{t-1}; \eta\right) d\,{}^*\widetilde{X}_1^{(n)} \cdots d\,{}^*\widetilde{X}_T^{(n)}$$

in which the expectation is taken over $({}^*\widetilde{X}_1^{(n)}, \cdots, {}^*\widetilde{X}_T^{(n)})$.

Generally numerical procedures such importance sampling shall be used for the calculation of the likelihood function. For an importance sampler φ and L repetitions of paths, the likelihood function maybe approximated as

$$\frac{1}{\mathrm{L}}\sum_{\mathrm{i=1}}^{\mathrm{L}}\prod_{t=1}^{T}\frac{q\left(X_t, {}_i^*\widetilde{X}_t^{(n)} \mid X_{t-1};\eta\right)}{\pi\left({}_i^*\widetilde{X}_t^{(n)} \mid X^{(1)};\eta\right)}.$$

The original suggestion of Pedersen is quite simple. The required augmented data can be simply generated with the Euler expansion (2.4), that is, an importance sampler like

$$\varphi\left({}^*\widetilde{X}_t^{(n)} \mid X_{t-1};\eta\right)=\prod_{s=1}^{n-1} q\left({}^*X_{t-1+s\Delta} \mid {}^*X_{t-1+(s-1)\Delta};\eta\right).$$

Clearly, a major drawback about this method is that it tends to lead to large jumps between the last augmented points and the consecutive observed data point.

Based on the Brownian bridge, Pedersen's approach can be modified with the following scheme

$$X_{t-1+s\Delta} = X_{t+(s-1)\Delta} + (x_t - x_{t-1})\Delta + \sigma_X (X_{t+(s-1)\Delta}, t)\Delta W_{s\Delta},$$

where (x_t - x_{t-1}) represents an estimate for the drift speed from time *t*-1 to *t*.

Elerian et al. (2001) proposed alternative importance sampler for the problem. The advantage of the approach is drawing paths at one shot and eliminating huge jumps. The augmented data between X_{t-1} and X_t can be sampled from a multivariate normal distribution $N(\mu^*, \Sigma^*)$ where

$$\mu^* = \arg\max_{{}^*X_t^{(n)}} \log q\left({}^*X_t^{(n)} \mid X_{t-1}, X_t\right),$$

$$\Sigma^* = -\left[\frac{\partial^2}{\partial^* X_t^{(n)} \partial^* X_t^{(n)'}} \log q\left({}^*X_t^{(n)} \mid X_{t-1}, X_t\right)\right]^{-1},$$

and $q\left({}^*X_t^{(n)} \mid X_{t-1}, X_t\right)$ denote the conditional density of ${}^*X_t^{(n)}$ under the Euler expansion (2.4).

More bias correction and variance reduction methods and a summary discussion may be found in Durham and Gallant (2001). Most of the methods mentioned may be applied to the stochastic volatility models, especially when the two driving Brownian motions are uncorrelated.

2.2.3 Asymptotic equivalence of stochastic volatility models and GARCH models

Since Engle(1982) and Bollerslov (1986), GARCH models have been widely used for modeling financial time series with stochastic volatilities. Nelson (1990) first investigated the convergence of GARCH processes to bivariate diffusions as the length of time intervals between observations goes to zero. Up to now diffusion limits for a variety of GARCH type processes have been found, for example, Duan (1997) and Fornari and Mele (2004). The relation between the two categories of models becomes very elaborate, especially when they both are essentially one-dimensional processes.

But even though the GARCH processes converge to their diffusion limits in distribution, it is not trivial that inferences through the two processes are equivalent. A major distinction between the two types of models is observability of volatility processes. Thus, once by subtle arrangement a GARCH model may maintain its availability of likelihood but its volatility process unobservable, it may work well to approximate the continuous counterpart. In fact, some recent researches have shown that the equivalence of the two types of models depends on the sampling frequency and the basic frequency of construction of the processes.

As set in the previous section, let $\widetilde{X}^{(n)} = (X_0, X_\Delta, \cdots, X_{n\Delta T})$ be observations from the stochastic model and $\widetilde{Y}^{(n)} = (Y_0, Y_\Delta, \cdots, Y_{n\Delta T})$ from the corresponding GARCH model at the basic frequency of construction. With the notation $D(X, Y)$ for L_1 distance of the joint density functions of the two processes X and Y, Wang (2002) showed $D(\widetilde{X}^{(n)}, \widetilde{Y}^{(n)})$ does not converge to 0 as n→∞. In other words, the likelihood processes have different asymptotic distributions and consequently the two types of models are not asymptotically equivalent.

However, as the frequency of observations become much lower than that of construction, the result goes quite different. Specifically speaking, let observations be $^{+}\widetilde{X}^{(n)} = (X_\Delta, X_{2l\Delta}, \cdots, X_{N^*\cdot l\Delta})$ and $^{+}\widetilde{Y}^{(n)} = (Y_{l\Delta}, Y_{2l\Delta}, \cdots, Y_{N^*\cdot l\Delta})$, where l represents the period between observations and N^* is the largest integer not larger than nT/l. Brown, Wang and Zhao (2003) illustrated the asymptotic equivalence of the MGARCH model and its diffusion limit with the dataset as $n\to\infty$ and $l/n^{1/2} \to \infty$.

These seemingly contradicting results in fact sketch the relation between the stochastic volatility model and its GARCH counterpart elaborately. Even though the GARCH process converges to the stochastic volatility model, the GARCH process is still composed of normally distributed innovations and determined volatilities. Augmentation of data deprives the GARCH process of these properties so that it may look like generated by a stochastic volatility model.

In other words, the implications are very similar to those among Lo (1988) and Pedersen (1995) on the univariate processes or multivariate process that are completely observable. In short, even though the GARCH models provide as good approximations to stochastic volatility models, likelihood functions for the stochastic volatility models cannot be obtained through the corresponding GARCH model at the frequency of observation, $1/T$. However, by the GARCH processes constructed at higher frequencies, the approximate likelihood function can be calculated with simulating all missing values.

2.3 Simulated likelihood for stochastic volatility models

Although Brown et al. (2002) had only investigated and proved the asymptotic equivalence of the MGARCH model and its diffusion limit, it may be well expected to extend the result to general cases with finite samples. In practice, it is reasonable to expect that the data considered from a stochastic volatility model would look as generated from some specified GARCH process whose frequency of construction is much higher, say n, but observations are taken every n period. Obviously there will be a large number of missing values between the original observations, but the MCMC techniques used in Elerian et al. (2001) can easily help solve this problem.

2.3.1 Approximations with partially observed GARCH

Let $\left(\tilde{X}^{(1)}, \tilde{V}^{(1)}\right)$ be drawn from equations (2.1) and (2.2) and $\left(\tilde{Y}^{(1)}, \tilde{h}^{(1)}\right)$ from the corresponding GARCH with length of construction interval Δ. Note that the observed data is indeed $\tilde{X}^{(1)}$, or, equivalently, $\tilde{Y}^{(1)}$. The following assumption is essentially necessay.

Assumption 3. The GARCH processes converge to its diffusion counterpart in distribution. That is, the conditional distribution $F^{(n)}(y_t, h_t \mid y_{t-1}, h_{t-1})$ converges to $F(x_t, v_t \mid x_{t-1}, v_{t-1})$ uniformly as $n \to \infty$.

Assumption 4. Each term of the sequence $q^{(n)}(Y_1, Y_2, \cdots, Y_T \mid Y_0, h_0)$ exists and the sequence converges to a proper density function.

It should be noted that Assumption 3 does not restrict the models under consideration to a narrow extent. In fact, both the approximating augmented GARCH(1,1) process in Duan(1997) and the CEV-ARCH models in Fornari and Mele (2006) satisfy this condition.

The likelihood function for $\tilde{X}^{(1)}$ has been shown in (3). With Assumption 4 that may be generally feasible for well defined GARCH models, the likelihood function for $\tilde{Y}^{(1)}$ can be written down in a similar form:

$$L^{(n)}(\eta;\tilde{Y}^{(1)}) = \int \pi(h_0) q^{(n)}(Y_1, Y_2, \cdots, Y_T \mid Y_0, h_0) dh_0 \ . \qquad (2.5)$$

Similar to Elerian et al. (2001) and as a direct result of Brown et al. (2003), equation (5) can be used as an approximation to the likelihood of the stochastic volatility model. For practical implementation, the convergence of the sequence of density functions $q^{(n)}(y_1, y_2, \cdots, y_T \mid y_0, h_0)$ would be the major consideration. However, in spite of the weak convergence of the processes, the convergence of the joint density function is not necessary under general conditions. Assumption 4 is thus necessary here. The convergence result of $q^{(n)}(y_1, y_2, \cdots, y_T \mid y_0, h_0)$ to $p(y_1, y_2, \cdots, y_T \mid y_0, h_0)$ is then stated as follows.

Theorem 2.3.1. With Assumptions 3 and 4, $q^{(n)}(y_1, y_2, \cdots, y_T \mid y_0, h_0)$ converges to $p(y_1, y_2, \cdots, y_T \mid y_0, h_0)$ uniformly as $n \to \infty$.

The proof of the theorem is a direct application of Lemme A1. Convergence results can be also obtained for specific models, for example Brown et al. (2003). However, with Theorem 3.1, a proper approximation of the likelihood (2.3) can be obtained from (2.5). Generally the close form of the transition density $q^{(n)}$ in (2.5) is not available, but by the Markovian property of the GARCH process, there can be found alternative expression of (2.5) as

$$L(\eta;\tilde{Y}^{(1)}) = \int \pi(V_0) p^{(n)}(\tilde{Y}^{(1)} \mid V_0;\eta) dV_0$$

$$= \int \pi(V_0) \prod_{s=1}^{nT} q^{(n)}({}^*Y_{s\Delta} \mid {}^*Y_0, \cdots {}^*Y_{(s-1)\Delta}, V_0;\eta) d{}^*\tilde{Y}_1 \cdots d{}^*\tilde{Y}_T dV_0$$

(2.6)

where ${}^*\tilde{Y}_t$ are the augmented data between Y_t and Y_{t-1}.

The problem becomes very similar to the simulated likelihood of Pedersen (1995).

However, due to the complex structure of the GARCH models, it would not necessarily so easy to find a sampler that mimics the conditional distribution of ${}^*\widetilde{Y}_t$ well. Thus, alternative approach is to generate paths from $p^{(n)}({}^*Y^{(n)} \mid \widetilde{Y}^{(1)}, V_0; \eta)$ and apply EM algorithm to compute the MLE, though this may be computationally costly. With each simulated paths, the object function in the Maximization step is then

$$\sum_{i=1}^{L} \log\left(\pi({}_i h_0) \cdot \prod_{s=1}^{nT}\left[\frac{1}{\sqrt{{}_i^* h_{s\Delta}^{(n)}}}\phi({}_i^* \varepsilon_{s\Delta}^{(n)})\right]\right), \tag{2.7}$$

where ${}_i^*\varepsilon_{s\Delta}^{(n)}$ and ${}_i^* h_{s\Delta}^{(n)}$ are innovations and variances at time $s\Delta$ on the i-th path.

The algorithm to compute MLE for η is then,

1. Initialize V_0 and ${}^*\widetilde{Y}^{(n)}$;
2. Update ${}_i h_0$ from $h_0 | {}^*\widetilde{Y}$;
3. Update sequentially ${}^*\widetilde{Y}_t^{(\Delta)}$ from ${}^*\widetilde{Y}_t^{(\Delta)} \mid \widetilde{Y}^{(1)}, {}^*\widetilde{Y}_1^{(n)}, \cdots, {}^*\widetilde{Y}_{t-1}^{(n)}, {}^*\widetilde{Y}_{t+1}^{(n)}, \cdots, {}^*\widetilde{Y}_T^{(n)}; \eta$;
4. Repeat step 2 and 3 and take L independent paths;
5. Maximize (2.7) with respect to η;
6. Repeat Steps 2 to 5 until convergence.

MCMC methods play important roles in this approach. Since the conditional density in step 3 becomes complex and generally no well-established algorithm is available, Metrapolis- Hasting algorithm can be used for sampling ${}^*\widetilde{Y}_t^{(n)}$. And the iterative sampling of ${}^*\widetilde{X}_t^{(n)} \mid X_{t-1}, X_t; \eta$ in Steps 2 and 3 is in fact a realization of Gibbs sampling.

It should be pointed out here that the discretization error for the transition probability function is indeed a function of the number of subdivision n rather than solely a function of the time between observations.

A major feature of this method is that the bivariate diffusion process with the second process unobservable is approximated by a univariate GARCH process. This method preserves the missing value problem from simulating the whole unobservable process to simulating paths connecting observations. In other words, any operations about the volatilities can be avoided. Furthermore, the GARCH process provides analytical form of the likelihood functions. What is left concerns only implementation of the MCMC algorithm.

There are three points to be emphasized. First, how equation (2.5) approaches the true likelihood function is in fact a matter of n, the number of subdivisions between observations, rather than the time between observations or values of $\widetilde{X}^{(1)}$. This means that however small the observation time interval is, subdivision is always necessary. This can be seen from the fact that the GARCH model always provides normally distributed innovations while the leptokurtosis of the conditional distribution for the return under the stochastic volatility models denied the possibilities of the normal distribution.

Second, this approach provides as another aspect to the continuous time stochastic volatility models. This class of models can be viewed as a partially observed GARCH process that has higher observation frequency and many missing values. On this point statistical inference for these continuous time models will be made feasible significantly.

Third, unlike Sørensen (2003) in which only observations previous to time t are used in constructing the conditional density at t, the method here employs all observations in determining each segment of paths. Perhaps, at a first glance, this does not seem so natural. However, on seeing that all observations $\widetilde{X}^{(1)}$ are correlated and thus each segment of the path provides information about those previous to or after it, this approach is indeed much more reasonable.

2.3.2 Why not Euler expansion?

It can be seen that the GARCH models cannot be the only class of models that satisfy theorem 2.3.1. In addition, as is acquainted with many researcher and practitioner in related fields, discretization through the Euler expansion provides another approach to approximate the stochastic volatility models.

The approximate likelihood function herein then involves a functional form of the bivariate normal distributions. For the stochastic volatility models (2.1) and (2.2), denote the augmented data as $({}^{*}\widetilde{X}^{(n)}, {}^{*}\widetilde{V}^{(n)})$. Since the process V_t is completely unobservable and there does not exist a similar form like (2.6) under the Euler scheme, the whole path of V_t must be treated as missing values and be simulated conditional on the observed price processes. Similar to (2.4) and (2.7), the likelihood function can be approximated by L independent paths as

$$L(\eta;\widetilde{X}^{(1)}) \approx \sum_{i=1}^{L}\left(\pi({}_iV_0)\prod_{s=\Delta}^{n\Delta T} MN({}_i^*Z_s^{(n)};\ {}_i^*\mu_s^{(n)},\ {}_i^*\Sigma_s^{(n)})\right), \tag{2.8}$$

where

$${}_i^*Z_s = \left({}_i^*X_s - {}_i^*X_{s-\Delta}, \phi({}_i^*V_{s-\Delta}) - \phi({}_i^*V_{s-\Delta})\right)$$

$${}_i^*\mu_s = \left(\mu_X({}_i^*X_s, {}_i^*V_s, s)\cdot\Delta, \mu_V({}_i^*V_s, s)\cdot\Delta\right)'$$

$${}_i^*\Sigma_s = \begin{bmatrix} {}_i^*\Sigma_s(1,1) & {}_i^*\Sigma_s(1,2) \\ {}_i^*\Sigma_s(2,1) & {}_i^*\Sigma_s(2,2) \end{bmatrix}$$

$${}_i^*\Sigma_s(1,1) = {}_i^*V_s \cdot \sigma_X^2({}_i^*X_s, s),$$

$${}_i^*\Sigma_s(1,2) = {}_i^*\Sigma_s(2,1) = \rho\cdot\sqrt{{}_i^*V_s}\cdot\sigma_X({}_i^*X_s, {}_i^*V_s, s)\cdot\sigma_V({}_i^*X_s, s),$$

$${}_i^*\Sigma_s(2,2) = \sigma_V^2({}_i^*V_s, s),$$

and $MN(\cdot;\cdot,\cdot)$ represents the density function of the multivariate normal distribution.

This approach may still work to obtain some values of the approximate likelihood function. However, the question apparent is: will this approximate likelihood function converge to the true one as the length of subdivision Δ tends to 0? The answer seems not so trivial.

The above numerical procedure in fact approximates the following integral

$$\int \pi(V_0)\prod_{t=1}^{T} p(S_t \mid S_{t-1}, V_{t-1};\eta)\, p(V_t \mid S_t, S_{t-1}, V_{t-1};\eta)\, d\widetilde{V}.$$

More precisely, these calculations involve the conditional expectations of volatilities given prices. Theoretically, what would go wrong may be that the conditional distribution under the Euler expansion will not necessarily converge to that of the original model. Some discussions and necessary conditions on the convergence of conditional expectations can be found in Goggin (1994) and Crimaldi and Pratelli (2005). And extra efforts are generally required to build artfully transformed processes to meet the required conditions.

To illustrate the property of (2.8), consider the following model where the observable process X_t has no drift and the volatility process is the limit of the GARCH(1,1) model (Nelson, 1990):

$$dX_t = \sqrt{V_t}\, dW_{1,t}, \tag{2.9}$$

$$dV_t = -\kappa(V_t - \xi)dt + \sigma \cdot V_t dW_{2,t}\,. \tag{2.10}$$

To avoid possible singularities by direct discretization for V_t, the following discretization scheme for Xt and log(V_t) by a transforma of Ito's lemma is used:

$$X_t = X_{t-\Delta t} + \sqrt{V_{t-\Delta t}} \cdot \Delta Z_{1,t}\,, \tag{2.11}$$

$$\log(V_t) = \log(V_{t-\Delta t}) + \left(\frac{\kappa\xi}{V_{t-\Delta t}} - \kappa - \frac{\sigma^2}{2}\right)dt + \sigma \cdot \Delta Z_{2,t}\,, \tag{2.12}$$

where $\Delta Z_{1,t}$ and $\Delta Z_{2,t}$ are normally distributed random variables with mean 0, variance Δ and coefficient of correlation ρ.

For $t<s<t+1$, the posterior distribution of ${}^*V_s$ given $\tilde{X}^*$ and other ${}^*V_s$'s can be expressed as

$$-2\log p({}^*V_s | {}^*X_{s-1}, {}^*V_{s-1}, {}^*X_s, {}^*X_{s+1}, {}^*V_{s+1})$$

$$\sim \frac{\left(\log({}^*V_s) - \log({}^*V_{s-1}) + \left(\frac{\kappa\xi}{{}^*V_{s-1}} - \kappa - \frac{\sigma^2}{2}\right)\Delta\right)^2}{\sigma^2\Delta}$$

$$-\frac{2\rho\left(X_s - X_{s-1}\right)\left(\log({}^*V_s) - \log({}^*V_{s-1}) + \left(\frac{\kappa\xi}{{}^*V_{s-1}} - \kappa - \frac{\sigma^2}{2}\right)\Delta\right)}{\sigma\sqrt{{}^*V_{s-1}}\Delta}$$

$$+\log{}^*V_s + \frac{\left(X_{s+1} - X_s\right)^2}{{}^*V_s\Delta} + \frac{\left(\log({}^*V_{s+1}) - \log({}^*V_s) + \left(\frac{\kappa\xi}{{}^*V_s} - \kappa - \frac{\sigma^2}{2}\right)\Delta\right)^2}{\sigma^2\Delta}$$

$$-\frac{2\rho\left(X_{s+1} - X_s\right)\left(\log({}^*V_{s+1}) - \log({}^*V_s) + \left(\frac{\kappa\xi}{{}^*V_s} - \kappa - \frac{\sigma^2}{2}\right)\Delta\right)}{\sigma\sqrt{{}^*V_s}\Delta}\,.$$

As most of these terms involve the dynamics *Vs only, they can be assume to be less cariational. So note the terms that are related to data first, that is

$$\log{}^*V_s + \frac{\left(X_{s+1} - X_s\right)^2}{{}^*V_s\Delta}\,.$$

While X_{t+1}-X_t is viewed as data and thus fixed, this means that $X_{s+1} - X_s$ as one of the n segments is approximately as large as $1/n$, which in term implies that ${}^{*}V_s$ is approximately of the scale n. That is, the simulated variance process under the Euler expansion may diverge as the number of subdivisions n approaches infinity!

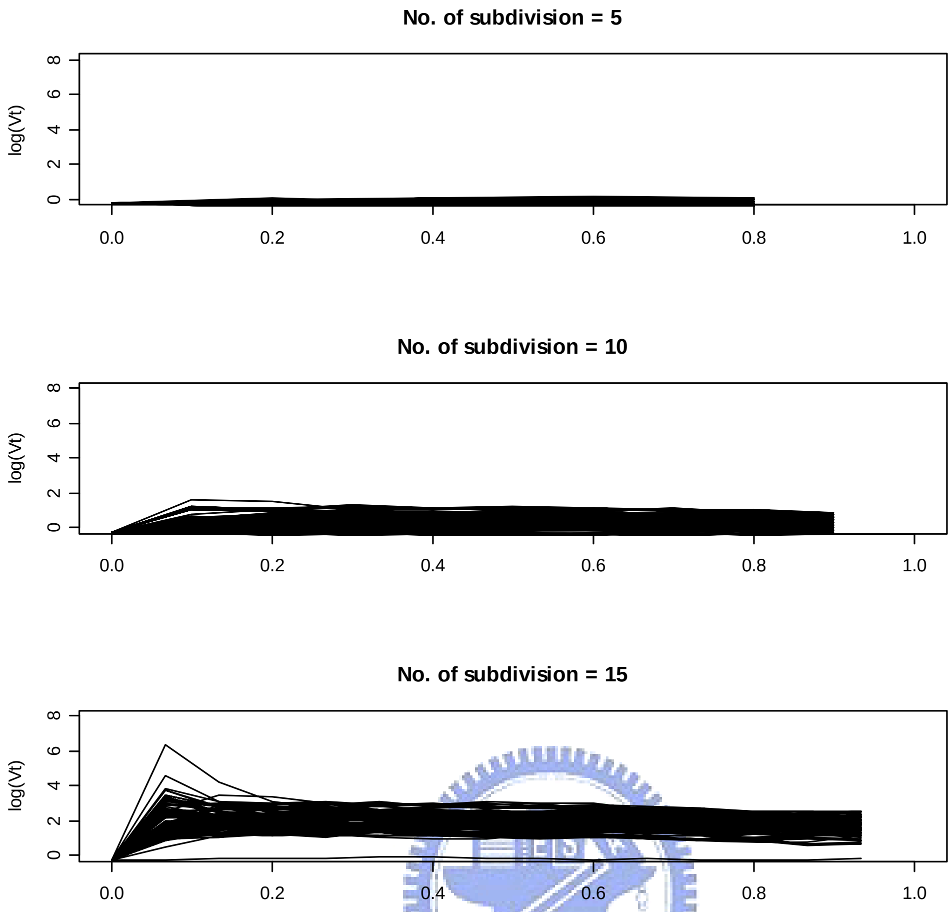
Table 2.1: Parameters used in the simulation study.

parameter	Value
κ	0.06988
ξ	0.7928
σ	0.1772
ρ	-0.65

Figure 2.1: Simulated paths of V_t conditional on X_0=0, X_1=1 and V_0=0.8.

Figure 2.1 illustrates this effect. The setting of parameters is the same as in the next section and shown in Table 2.1. Simulated paths of V_t between the period of 0 to 1 given X_0=0, V_t=0.8 and X_1= 1 are generated. The three panels correspond respectively to the cases with subdivisions 5, 10, 15. In each simulation, linear interpolations for X_t at each time point and constant V_t equal to 0.8 are used as initial values. After 200 burn-in iterations, paths of V_t are taken every two iteration from the next 200 iterations. It can be seen that the level of the Vt paths go up significantly. This illustrates the inadequacy in estimating stochastic volatility models based on the Euler scheme.

Figure 2.1. Simulated paths of V_t conditional on X_0=0, X_1=1 and V_0=0.8.


No. of subdivision = 5
log(Vt)
No. of subdivision = 10
log(Vt)
No. of subdivision = 15
log(Vt)
0.0
0.2
0.4
0.6
0.8
1.0

2.4 Numerical illustrations

With the model (2.9) and (2.10) for a numerical illustration of parameter estimation, ten datasets of length 500 and time between observations as 1 are generated according to the Euler approximation (2.11) and (2.12) with Δ=0.001. The initial value of X_t is X_0=0 and V_0 is generated from the stationary distribution of V_t:

$$\pi(v_t) = \frac{\lambda^a}{\Gamma(a)} v_t^{-a-1} e^{-\frac{\lambda}{v_t}},$$

where $a = 1 + 2\kappa/\sigma^2$ and $\lambda = 2\kappa\xi/\sigma^2$.

From the results shown in section 2.4, the approximating GARCH process corresponding to the model specified by (2.9) and (2.10) is

$$^*Y_s^{(n)} = \sqrt{^*h_s^{(n)}}\,\varepsilon_s \sqrt{\Delta},$$

$$^*\phi_s^{(n)} = \kappa\xi \cdot \Delta + {}^*\phi_s^{(n)}\left[1 - \kappa \cdot \Delta\right]$$

$$+ {}^*\phi_s^{(n)} \cdot \sigma \sqrt{\frac{1-\rho^2}{2}} \cdot \left[\left(\varepsilon_s - \sqrt{\frac{1}{2}\frac{\rho^2}{1-\rho^2}} \right)^2 - \frac{1}{2}\frac{2-\rho^2}{1-\rho^2} \right] \sqrt{\Delta},$$

$$h_{ks}^{(n)} = \left| ^*\phi_s^{(n)} \right| .$$

An approximation for the transition density of this type of GARCH model can be found in Duan et al. (1999). Their result also implies that Assumption 5 is satisfied for this model.

The parameters used in this simulation study are shown in Table 2.1. And this setting is compliant with the estimates for the NGARCH model with the NYSE composite index returns in Duan (1997).

Summary statistics for each dataset are listed in Table 2.2. Figure 2.2 shows the trend chart for the simulated values of X_t-X_{t-1} and V_t from the first 4 datasets. QQ-plots for the same 4 datasets are shown in Figure 2.3. The phenomena of volatility clustering can be easily observed. With the results of B-J tests, normality for X_t-X_{t-1} is seen to be very different through datasets. Datasets 1, 6 and 10 are very close to being normally distributed, while others differ from normality significantly. On the other hand, the Ljung-Box test indicates that in principle the increments X_t-X_{t-1} are statistically independent.

Table 2.2: Summary statistics for the increments X_t-X_{t-1}.

dataset	mean	variance	skewness	kurtosis	BJ stat	p value	Box-Ljung statistic	p-value
1	0.0075	0.7930	-0.0149	3.0649	0.11	0.948	2.3751	0.1233
2	-0.0260	0.7256	-0.2144	3.5662	10.51	0.005	0.0115	0.9146
3	0.0270	1.0351	-0.5145	6.5380	282.84	0	3.8780	0.0489
4	-0.0158	0.8316	-0.3243	4.6880	68.13	0	0.0000	0.9993
5	0.0236	0.7856	-0.0366	3.7860	12.98	0.002	0.1225	0.7263
6	0.0152	0.7894	-0.0426	3.1772	0.81	0.669	0.1910	0.6621
7	-0.0023	0.8312	-0.0689	3.6905	10.33	0.006	1.4088	0.2352
8	-0.0705	1.0778	-0.6887	6.8981	356.10	0	0.7640	0.3821
9	0.0421	0.8233	0.1272	5.3566	117.05	0	0.0659	0.7974
10	0.0333	0.6941	0.0116	3.4174	3.64	0.162	0.8251	0.3637

Figure 2.2: Trends of X_t-$X_{t-1}$of the first 4 datasets.

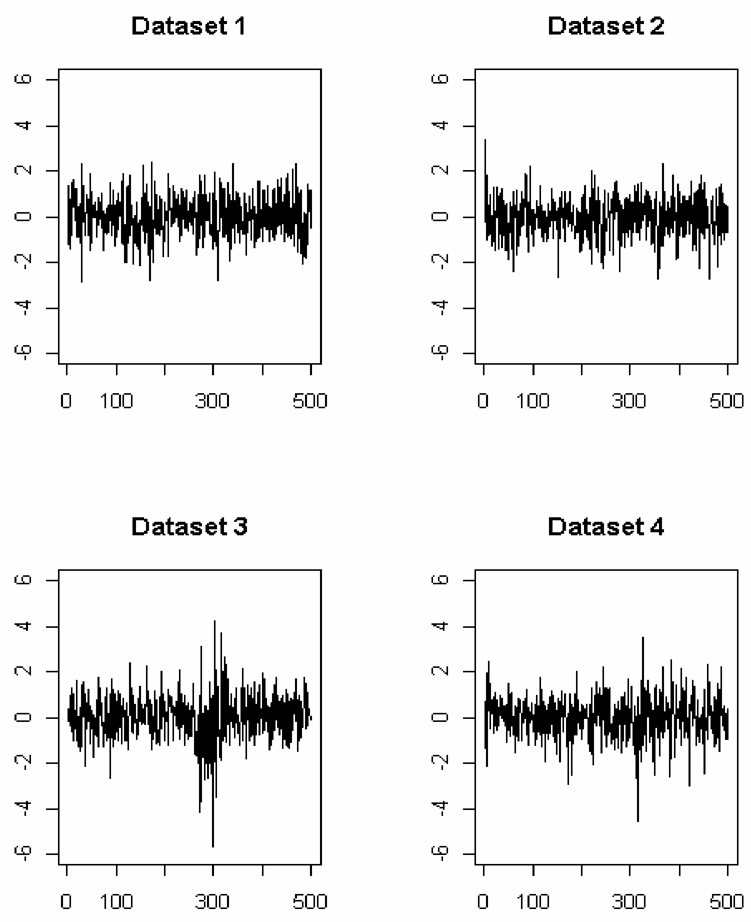
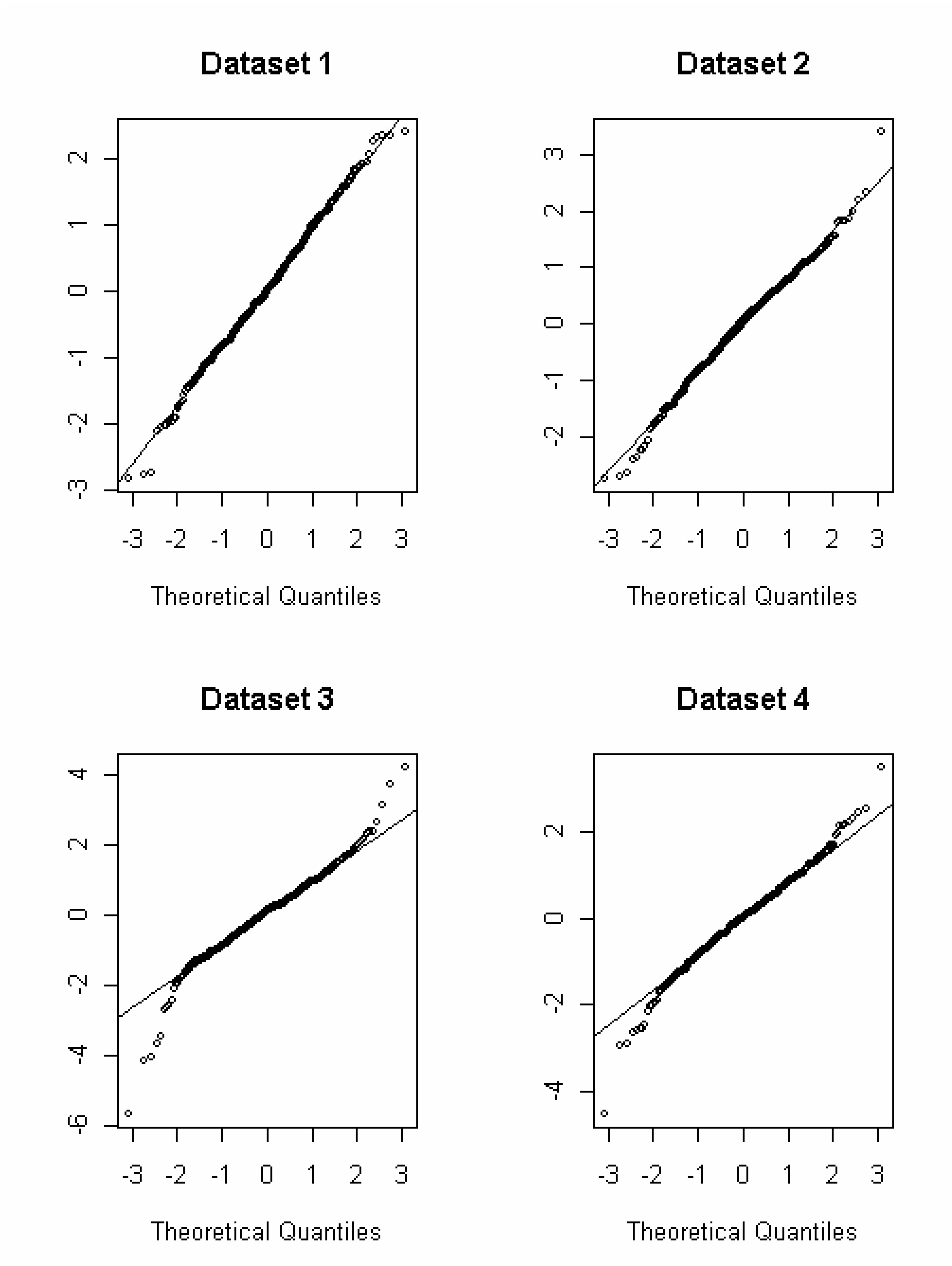


Figure 2.3: QQ-plots of X_t-$X_{t-1}$of the first 4 datasets.



Parameter estimation based on (2.7) with these datasets proceeds with Δ=0.2. By the method in Raftery and Lewis (1992), a small size experiment is conducted for determining the iteration times required to achieve convergence in each step.

For the Metropolis algorithm used in step 2, 150 burn-in iterations are exercised before taking one sample for each ${}^*\widetilde{X}_t^{(n)}$. Similarly, sample paths for ${}^*\widetilde{X}_t^{(n)}$ are taken every two iterations from the 200 iterations following 20 burn-in iterations.

Table 2.3: Parameter estimates. A comparison between the simulated likelihood, GARCH approximation and EMM.

Method	Parameter	True Value	Mean	Median	Max	Min	STD
Simulated Likelihood	κ	0.0699	0.0637	0.0587	0.1287	0.0423	0.0241
	ξ	0.7928	0.8098	0.8063	0.9379	0.6994	0.0678
	σ	0.1772	0.1928	0.1863	0.2255	0.1685	0.0218
	ρ	-0.6500	-0.6955	-0.7079	-0.5164	-0.8524	0.0974
GARCH	κ	0.0699	0.1174	0.0732	0.6060	0.0160	0.1734
	ξ	0.7928	0.8321	0.8187	0.9587	0.7212	0.0801
	σ	0.1772	0.1500	0.1523	0.2505	0.0781	0.0502
	ρ	-0.6500	-0.9213	-0.9661	-0.5747	-1.0000	0.1314
EMM	κ	0.0699	0.0670	0.0557	0.1176	0.0161	0.0369
	ξ	0.7928	0.8364	0.8135	1.1231	0.6964	0.1274
	σ	0.1772	0.1899	0.1738	0.3357	0.0972	0.0853
	ρ	-0.6500	-0.7727	-0.7575	-0.6358	-0.9855	0.1237

Table 2.4: Parameter estimates. Estimates with the likelihood function approximated through the Euler approximation and V_t assumed to be observed.

	Parameter	True Value	Mean	Median	Max	Min	STD
Δ=1	κ	0.0699	0.0632	0.0601	0.0927	0.0379	0.0164
	ξ	0.7928	0.8065	0.7924	0.9047	0.7236	0.0678
	σ	0.1772	0.1696	0.1689	0.1735	0.1656	0.0026
	ρ	-0.6500	-0.6507	-0.6507	-0.6165	-0.6812	0.0182
Δ =0.2	κ	0.0699	0.0659	0.0622	0.1031	0.0421	0.0181
	ξ	0.7928	0.8121	0.8011	0.8996	0.7318	0.0676
	σ	0.1772	0.1767	0.1769	0.1786	0.1726	0.0017
	ρ	-0.6500	-0.6504	-0.6508	-0.6365	-0.6600	0.0070

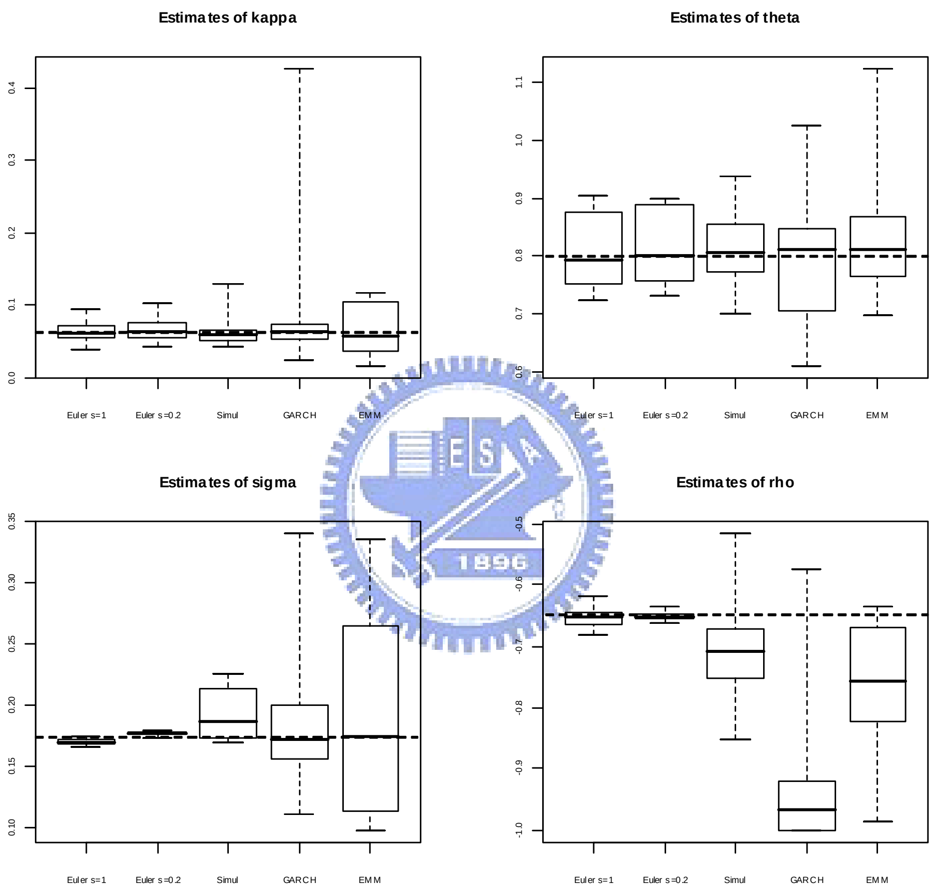
Estimates via NGARCH approximations and EMM methods for each dataset are also calculated for comparisons. The AR-NGARCH and AR-EGARCH processes are considered as the score generator for the EMM method. As in Anderson and Lund (1997), each dataset is first fitted with the two classes of processes. By the AIC criterion the model that fits better is selected. It is found that AR terms in the mean are generally not necessary, and the EGARCH process fits 6 of 10 datasets better than the NGARCH process although the latter has this diffusion models (2.9) and (2.10) as its limit.

The SNP densities are then set with K_X=0 and K_Z=1 and thus have the form

$$f_K(S_t \mid \eta) = \frac{(1+\alpha_1 z_t)^2 \phi(z_t)}{\int (1+\alpha_1 z_t)^2 \phi(z_t) dz_t}.$$

On computing the expectation of the score, a sequence of length 100,000 is simulated under the Euler scheme in which Δ=0.04.

Figure 2.4: Box plots for estimates from different methods. The dash lines indicate the true values.



The summary of estimation results is listed in Table 2.3 and graphically represented by box plots in Figure 2.4. The estimates for ξ, σ, and ρ from the approximate likelihood are concentrated around the true value. However, κ seems to be systematically underestimated. This may be explained as follows. First it should be noted that the GARCH processes only provide as an approximation but not an exact distribution. Since κ is related to mean reversion, estimating κ precisely requires more information about the "events" of mean reversion. However, taking discrete observations implies that those very quick mean reverse events will be dropped. In other words, the sampled "events" for estimating κ is biased under the discrete time scheme. An indirect evidence is shown in Table 2.4. The estimates are obtained through maximizing the likelihood of all X_t and V_t under the Euler scheme for Δ =1 and 0.2 respectively. It is seen that κ also tends to be underestimated, and estimates for κ with Δ=1 is even lower than those with Δ=0.2.

It can be seen that estimates from NGARCH approximations are much poorer than the estimates from the simulated likelihood. In fact, there is almost at least one parameter wildly estimated for each dataset. This result conforms to Wang (2002) and should not be too surprising since NGARCH approximations are indeed based on normally distributed innovations, while the bivariate diffusion models generate much more complicated distributions.

The EMM method provides more reasonable estimates than those from the NGARCH approximations. However, the EMM estimator is seen to be less efficient than that from the simulated likelihood. This result is conformable to the nature of EMM estimators discussed in Gallant and Tauchen (1996), since the diffusion models (2.9) and (2.10) is not embedded in either the NGARCH model or the EGARCH model.

2.5 Conclusions and extensions

In this chapter, a method to approximate the likelihood functions of the continuous time stochastic volatility models is proposed. Although this method requires some knowledge about GARCH approximation process, it is in fact easy in coding and practical implementations. Furthermore, it serves as a basis for statistical inference for this class of models. Commonly used methods and criteria including likelihood ratio, AIC and BIC can be conducted through this approach.

In addition to the continuous time stochastic volatility models, the method proposed here may also be applied to the estimation of jump-diffusion models. Duan et al. (2005) propose a family of GARCH-Jump models whose limits are just jump-diffusion models. It is reasonable to expect that the same procedures can also be applied to the estimation of this richer class of processes.

Furthermore, this approach on estimation problems gives alternative access to the understanding of the stochastic volatility models. Since, as emphasized in the introduction of the model, the volatilities are never observed, any operations such as calculating option prices should not be conditional on the filtration $\sigma\left(\widetilde{S}^{(1)},\widetilde{V}^{(1)}\right)$, and in fact $\sigma\left(\widetilde{S}^{(1)}\right)$ is obviously a more reasonable candidate. More specifically, calculating derivative prices should be based on the conditional distribution $P(S_{T+\tau} \mid \widetilde{S}^{(1)})$ instead of $P(S_{T+\tau} \mid S_T, \hat{V}_T)$, where $\hat{V}_T$ is filtered by $\widetilde{S}^{(1)}$. Clearly, the partially observed GARCH process approach provides as a solution, although there are technical details to be solved.

For example, the partially observed GARCH process may be also used as a filter that produces the conditional distributions of the unobserved variances. The properties of the filter obviously need more investigations.

Finally, the introduction of the partially observed GARCH process may largely widen the uses of continuous time stochastic volatility models in financial time series modeling. In fact, the type of GARCH models elaborately bridges the two clusters of models, and possesses the advantages of the two ends. An alternative may be the discrete time stochastic volatility models, which are asymptotically equivalent to the continuous time stochastic volatility models. Clearly, inferences about the two types of discrete time models involve computational costly simulations to integrate out latent variables. Some kind of unification and the extents these models are feasible may be interesting for future studies.

2.6 Appendix

2.6.1 Convergence of a sequence of derivatives

Let $\{f_n\}$ be a sequence of integrable functions in an open set $E \subset R^d$ and $F_n = \int f_n d\mu$, where μ is the Lebesque measure in R^d.

Lemma 2.6.1. Suppose that there exists functions g and $F = \int f d\mu$ such that f_n converges uniformly to g and F_n converges uniformly to F. Then $f = g$ a.e.

Proof. Assume that $c \in E$. Define a sequence $\{g_n\}$as follows:

$$g_n(x) = \begin{cases} \frac{1}{\mu(B)} \int_B f_n d\mu & \text{if } x \neq c, \\ f_n(c) & \text{if } x = c. \end{cases}$$

Convergence of $\{g_n(c)\}$ comes from the fact that $\{f_n(c)\}$ converges. For $x \neq c$, $\{g_n(x)\}$ also converges since $\{F_n\}$ converges. Thus $\{g_n\}$ converges uniformly and denote $G(x) = \lim_{n \to \infty} g_n(x)$.

As f_n exists and is integrable, $\lim_{x \to c} g_n(x) = g_n(c)$. So each g_n is continuous at c. Since g_n converges uniformly to G, G is also continuous at c. Thus

$$G(c) = \lim_{x \to c} G(x).$$

But for $x \neq c$,

$$G(x) = \lim_{n \to \infty} g_n(x) = \lim_{n \to \infty} \frac{1}{\mu(B)} \int_B f_n d\mu = \frac{1}{\mu(B)} \int_B f d\mu.$$

Thus the derivative $f(c)$ is equal to $G(c)$. But

$$G(c) = \lim_{n \to \infty} g_n(c) = \lim_{n \to \infty} g_n(c) = g(c),$$

hence $f(c)=g(c)$. Since c is arbitrary, this finishes the proof.

2.6.2 Augmented GARCH process

Duan (1997) proposed a family of parametric GARCH models and identified their limits. For s=1, 2, ⋯, nT, the approximating augmented GARCH(1,1) process is defined as

$$Y_{s\Delta}^{(n)} = \left[\omega_0(h_{s\Delta}^{(n)}) + \omega_1(h_{s\Delta}^{(n)})Y_{(s-1)\Delta}^{(n)}\right]\cdot\Delta + Y_{(s-1)\Delta}^{(n)} + \sqrt{h_{s\Delta}^{(n)}}\,\varepsilon_{s\Delta}\sqrt{\Delta}\,,$$

$$\phi_{s+\Delta}^{(n)} = (\alpha_0 + q_4)\Delta + \phi_{s\Delta}^{(n)}\left[1 + (\alpha_1 + \alpha_2 q_2 + \alpha_3 q_3 - 1)\cdot\Delta\right] +$$

$$\phi_{s\Delta}^{(n)}\left[\alpha_2(Z_{s\Delta}^{(2)} - q_2) + \alpha_3(Z_{s\Delta}^{(3)} - q_3)\right]\sqrt{\Delta} + (Z_{s\Delta}^{(4)} - q_4)\sqrt{\Delta}\,,$$

$$h_{s\Delta}^{(n)} = \left|\lambda\phi_{s\Delta}^{(n)} - \lambda + 1\right|^{1/\lambda} \quad \text{if } \lambda > 0$$

$$\exp(\phi_{s\Delta}^{(n)} - 1) \qquad \text{if } \lambda = 0\,,$$

where ω_0 and ω_1 are Borel measurable functions, $\varepsilon_{s\Delta}$, s= s=1, 2, ⋯, nT, are i.i.d. standard normal random variables,

$$Z_{s\Delta}^{(2)} = \left|\varepsilon_{s\Delta} - c\right|^{\delta},$$

$$Z_{s\Delta}^{(3)} = \max(0, c - \varepsilon_{s\Delta})^{\delta},$$

$$Z_{s\Delta}^{(4)} = \alpha_3 f(|\varepsilon_{s\Delta} - c|;\delta) + \alpha_4 f(\max(0, c - \varepsilon_{s\Delta});\delta)\,,$$

$$q_i = E[Z_{s\Delta}^{(i)}]\,,\ i=1, 2, 3,$$

and

$$f(z;\delta) = (z^{\delta} - 1)/\delta \quad \text{for} \quad z \ge 0\,.$$

For Δ=1 and various values of parameters λ, c, α_0, α_1, α_2, α_3 and α_4, the model corresponds to different named GARCH models. The major result can be summarized in the following theorem.

Theorem 2.6.2 As $n \to \infty$, the distribution function converges uniformly to that of the limiting stochastic volatility models

$$dX_t = \left[\omega_0(h_t) + \omega_1(h_t)\cdot X_t\right]dt + \sqrt{h_t}\,dB_{1,t}\,,$$

$$d\phi_t = \left[\alpha_0 + q_4 + (\alpha_1 + \alpha_2 q_2 + \alpha_3 q_3 - 1)\phi_t\right]dt + v_t\rho_t dB_t^1 + v_t\sqrt{1-\rho^2}\,dB_{2,t}$$

$$h_t = \left|\lambda\phi_t - \lambda + 1\right|^{1/\lambda} \quad \text{if } \lambda > 0$$

$$\exp(\phi_t - 1) \qquad \text{if } \lambda = 0\,,$$

where

$$v_t = \left[\sigma_4^2 + 2(\alpha_2\sigma_{24} + \alpha_3\sigma_{34})\phi_t + (\alpha_2^2\sigma_2^2 + \alpha_3^2\sigma_3^2 + 2\alpha_2\alpha_3\sigma_{23})\phi_t^2\right]^{1/2},$$

$$\rho_t = v_t^{-1}\left[\sigma_{14} + \phi_t(\alpha_2\sigma_{12} + \alpha_3\sigma_{13})\right],$$

$$Var(\varepsilon_s, Z_s^{(2)}, Z_s^{(3)}, Z_s^{(4)}) = \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^3 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^4 \end{bmatrix},$$

and $B_{1,t}$ and $B_{2,t}$ are two independent Wiener processes.

Note that the convergence holds for $0 \leqq t < \infty$, not merely for some fixed time T. And weak convergence of the processes implies the joint distributions of

$\left((Y_0^{(n)}, h_0^{(n)}), (Y_1^{(n)}, h_1^{(n)}), \cdots, (Y_1^{(n)}, h_1^{(n)})\right)$ converge to that of

$\left((X_0, V_0), (X_1, V_1), \cdots, (X_T, V_T)\right)$.

Corollary.2.6.3 The joint distributions $P(Y_0^{(n)}, Y_1^{(n)}, \cdots, Y_T^{(n)})$ converge weakly to $P(X_0, X_1, \cdots, X_T)$.

Similar results for modeling interest rate related processes can be found in Fornari and Mele (2006). A model of GARCH class that approximates stochastic volatility CEV diffusions has been developed. And it is seen that approximating GARCH processes for major stochastic volatility models discussed in literatures have been well established.

2.6.2 Efficient Method of Moment

Based on the GMM principle (Hansen, 1982), Gallant and Tauchen (1997) propose the EMM method for estimating complicated models. They show that if the score generator encompasses the maintained model, then EMM is as efficient as maximum likelihood. Results of Tauchen (1997) suggest that the EMM estimator will be nearly as efficient as maximum likelihood when the score generator is a good statistical approximation to the observed process. Gallant and Long (1997) support this conjecture by showing that if the score generator is the SNP density, then efficiency of the EMM estimator can be made arbitrarily close to that of maximum likelihood. An empirical illustration for EMM can be found in Anderson and Lund (1997).

As implementing the method, an auxiliary model which closely approximates the conditional returns distribution must be selected first. For stochastic volatility models, GARCH type models are naturally the candidates. The SNP densities are then used to make high-order approximations.

The SNP densities, proposed by Gallant and Nychka (1987), is an approximation based upon a Hermite series expansion:

$$h_K(y_t|F_t;\psi)=\frac{[P_K(z_t,x_t)]^2}{\int[P_K(z_t,x_t)]^2\phi(z_t)}\frac{\phi(z_t)}{\sqrt{h_t}}$$

where

$$z_t=\frac{y_t-\mu_t}{\sqrt{h_t}},$$

$$P_K(z_t,x_t)=\sum_{i=0}^{K_Z}\left(\sum_{j=0}^{K_X}a_{ij}x_t^j\right)z_t^i,\qquad a_{00}=1,$$

And $\phi(\cdot)$ is the density function of the standard normal distribution, μ_t and h_t follow the specification of the auxiliary GARCH model.

Let ξ denote the parameters in the auxiliary model. An estimate $\hat{\xi}$ can be easily obtained through maximizing the likelihood function of the auxiliary model. So is the estimate of the inverse asymptotic variance matrix of the score function, $\hat{W}_N$,.

The EMM estimator is obtained by minimizing the moment condition:

$$m_{T(N)}^{'}V_{T(N)}^{-1}m_{T(N)},$$

where $m_{T(N)}$ is the expectation of the score function, evaluated at $\hat{\xi}$ with a sample of length $T(N)$ under the stochastic volatility model by Monte Carlo simulation.

3. Computing option prices under stochastic volatility models

3.1 Introduction

This chapter proposes a new approach to calculate option prices under stochastic volatility models. A partially observed GARCH model is employed to approximate the joint distribution of the series of prices, so the calculation proceeds without estimating the volatility process. Through simulation experiments there are also more fascinating properties found for the nature of the type of models.

Stochastic volatility models arose for the relief of the constant volatility assumption of the Black-Scholes formula. Varieties of specifications can be found in documented literatures, such as Hull and White (1987) and Heston (1993). Empirical studies also piled up since the last years of 1990's, for example, Bakshi, Cao and Chen(1997) and Bates(2000).

Generally, two stochastic processes are used for describing the dynamics of the asset prices. However, one important issue which is seldom mentioned or emphasized in literatures is that volatilities are in fact unobserved. This means that inference with a stochastic volatility model is in nature a missing value problem.

Typical statistical approach to deal with these missing values is taking expectations conditional on the observed values – the prices. However, in all the literatures reviewed, the basic idea for pricing options under the stochastic volatility models is originated from the Markovian property of the model, so the following formula is generally used,

$$E^{Q}[e^{-r(T-t)}g(S_T)|S_t, V_t]. \tag{3.1}$$

where $g(S_T)$ is the payoff at maturity. It should be noted that the formula above in fact leads to a function of the random variable V_t, so it is obviously a random variable and possesses specific distribution. Then, besides the structural parameters of the asset prices, the risk premium of volatility and especially the unobserved volatilities need to be estimated with the observed prices.

Thus, the formula becomes

$$E^{Q}\left[e^{-r(T-t)}g(S_T) \mid S_t, \hat{V}_t\right],$$

where $\hat{V}_t$ is the estimate of V_t. Generally a two-stage strategy is usually taken. First the structural parameters are estimated through GMM/EMM methods or other approaches. Then a loss function or some moment conditions are used for the determination of the risk premium of volatility and the instantaneous volatilities at each time. It follows that the estimation of the unobserved volatilities will depend on the choice of loss functions. Unfortunately, almost all commonly used loss functions are in lack of economic interpretations or even do not satisfy certain fundamental statistical requirements, so it can be asserted that the properties of the estimated volatilities are suspicious, not to mention the resulting option prices.

Some authors use the Kalman filter or other nonlinear filter as proxies for volatilities as the structural parameters are obtained. These estimates may have certain asymptotic properties with very high sampling frequencies, but arbitrary biases exist for observations from finite sampling frequencies.

To find the key to these problems the identification of the filtration up to time t, Ω_t, must be made first. As mentioned earlier, the volatilities are indeed unobservable, so there's no doubt that Ω_t should be $\sigma\{S_0,\ldots,S_t\}$ instead of $\sigma\{(S_0, V_0)',\ldots,(S_t, V_t)'\}$. More specifically, pricing options should rely on the conditional distribution

$$p^Q(S_T \mid S_0,\ldots,S_t).$$

Based on the above arguments, a method is proposed to compute option prices under the stochastic volatility models without plugging in estimates of volatilities. Brown et al. (2003) suggest the statistical equivalence of GARCH and stochastic volatility models (under certain conditions). With the simulated likelihood approach in chapter 2, paths connecting observations can be generated with a partially observed GARCH model and MCMC algorithm. Then future asset prices can be simulated by extending these paths using general Monte Carlo methods. Since the method is based on the sequence of prices only, it is not necessary to estimate volatilities and there exists one-to-one relation between the price and the premium.

Furthermore, numerical results also suggest some interesting and inspiring properties of the method. For example, for a lower and intermediate level of the true variance, tendency toward a positive bias exists on the conditional distribution of variance given past path of prices and thus the implied option prices. And for ascent and descent paths with very close level of the true variances, the method also suggests distinct option values. These properties meet the facts in most of the markets but cannot be achieved with the traditional approach.

This chapter is organized as follows. A brief discussion on the current practice is in Section 3.2. Careful investigation on the proper formulation of the option prices is included in Section 3.3. In Section 3.4 the proposed algorithm is presented and illustrated with examples. In Section 3.5 discussion and further extension are drawn.

3.2 A quick review on the current practice for computing option prices

Among all related literatures, Heston's pricing model with a closed form solution might be the most frequently referred. Heston (1993) extends the Black-Scholes model with a stochastic factor:

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_{1,t},$$

$$dV_t = -\kappa (V_t - \theta) dt + \sigma \sqrt{V_t} dW_{2,t},$$

where S_t and V_t are the asset price and instantaneous volatility at time t respectively, and $W_{1,t}$ and $W_{2,t}$ are standard Brownian motions with coefficient of correlation ρ. The solution for a call option is then of the form

$$S_t \cdot P_1(S_t, V_t; T-t, K, \eta, \lambda) - K \cdot P(t,T) \cdot P_2(S_t, V_t; T-t, K, \eta, \lambda)$$

where K is the strike price, T is maturity, η is the parameter vector $(\kappa, \theta, \sigma, \rho)'$, λ is the risk premium of volatilities and P, P_1, P_2 are appropriate functions.

It is noted that there are parameters, η and λ, and random variables V_t's to be estimated for the implementation the formula. The structural parameters η can be estimated with the price series of the underlying asset. As the likelihood function for this type of system is not available, GMM/EMM or similar methods are the common choices. The risk premium λ is in principle disassociated to the price dynamics directly, so the information for estimating it generally should include option prices. Intuitively the instantaneous variance V_t's could be estimated with a filter on the series of the asset prices, however, by minimizing a loss function for the option prices also seems to be common.

3.2.1 Loss functions in option pricing

Loss functions play important roles in option pricing. As the two-stage estimation procedure is generally required, loss functions are usually the objective function for the estimation of the risk premium and even the unobserved volatilities at each time period, for example Bakshi et al (1997). This means that the role of the loss functions is far beyond the evaluation of the pricing models.

However, the choice of the loss functions is an annoying problem in practical exercise pricing options, just as in Bakshi et al (1997):

“The objective function in equation (17) is defined as the sum of squared dollar pricing errors, ⋯.An alternative could be to minimize the sum of squared percentage pricing errors ⋯. Based on this and other considerations, we choose to adopt the object function in equation (17)”

Clearly there does not exist a loss function that satisfies fundamental requirements and applies to all categories of option contracts. It is unfortunate that the standard option valuation theory implies a unique option price, but mentions nothing about how to specify the error term (Renault, 1997). However, there are two points about the loss functions to be emphasized. First, consistency of the choice of loss functions in the two stage of exercises is essentially necessary (Christoffersen and Jacobs, 2004). Next, the choice of loss function implicitly defines the model under consideration, see Engle (1993).

Christoffersen and Jacobs (2004) shows that the estimates of parameters and volatilities will depend on the choice of the loss function. They also suggest that the choice of the loss function may depend on the purpose of the empirical exercise. This does not matter from the view of practical purposes as long as the users may actually benefit from it. However, from an academic view, is it acceptable that the volatility process for one specific asset should be different for a speculator and a hedger?

Moreover, the commonly used loss functions are also commonly in lack of any economic interpretations or any supports of statistical properties. For example, the mean squared dollar errors that is most frequently used in most literatures,

$$\frac{1}{N}\sum_{i=1}^{N}\left(C_i - \hat{C}_i\right)^2 ,$$

implicitly assume homogeneity of variance in the pricing errors. This seemingly means that the errors may only come from those factors providing constant impact, such as transaction costs, etc. Thus this would also raise questions like: does the pricing error of one dollar have the same implications for options of one dollar and 1000 dollar? Other popular loss functions obviously have similar difficulties or violate their statistical assumptions, see chapter 4.

In other words, forecasting the unobserved volatilities based on the loss functions can lead to logical and theoretical contradictions. These random variables should be forecasted or filtered with other properly established methods.

3.2.2 Filtering the volatilities with asset prices

There can be found some methods to filter variance estimates under the stochastic volatility model using the information of the prices only, for example Nelson and Foster (1994).

Let X_t denote the logarithm of S_t. Then for the Heston model the filter of Nelson and Foster can be derived as below:

$$\hat{u}_{t+1} = \hat{u}_t + \left[-\kappa + e^{-\hat{u}_t} \left(\kappa\theta - \frac{\sigma^2}{2} \right) \right] \Delta$$

$$+ e^{-\frac{\hat{u}_t^2}{2}} \left[\rho \hat{\xi}_t e^{-\hat{u}_t} + \left(\frac{1-\rho^2}{2} \right)^{1/2} \left(\hat{\xi}_t^2 e^{-\hat{u}_t} - 1 \right) \right] \sigma \sqrt{\Delta},$$

where Δ is the sampling period and ξ_t is the normalized innovation, that is,

$$\xi_{t+1} = \frac{1}{\sqrt{\Delta}} \left(X_{t+1} - X_t - E\left(X_{t+1} - X_t \right) \right).$$

This filter is indeed an UMVUE. That is, its bias for estimating V_t is 0. And note that filters derived from the Euler expansion of the differential equation need not to provide as an unbiased or even consistent estimator, see Goggin (1994) and Crimaldi and Pratelli (2005).

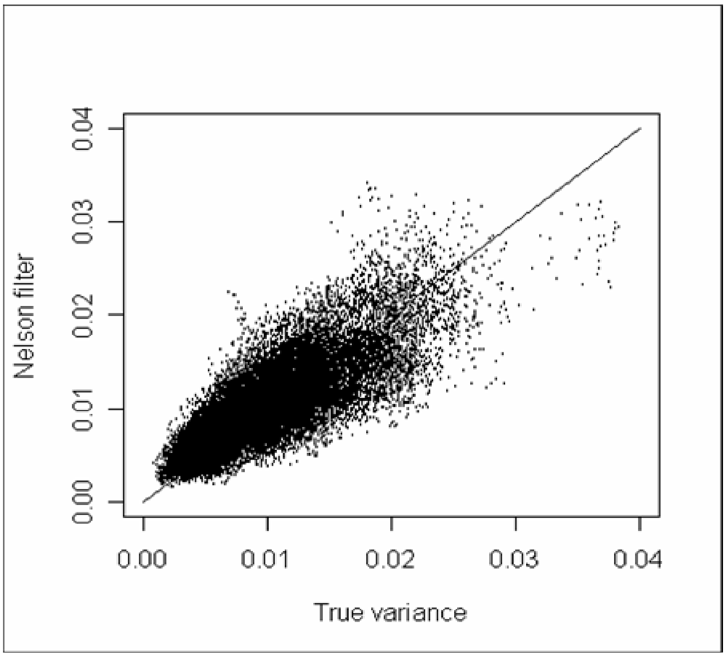
To illustrate how the estimation errors would impact the pricing of options, a simulation experiment is conducted with the Heston model and the following setting of parameters.

Table 3.1: Parameters used in the simulation study.

parameter	Value
r	0.02
κ	2
θ	0.01
σ	0.1
ρ	-0.65

A sequence of (X_t, V_t) of length 25000 and time between observations as 0.004 are generated according to the Euler approximation with Δt=0.0004. Estimates of variances at every time point are then calculated with respect to the above filters. Figures 3.1 displays the scatter plot of the estimated and true variances.

Figure 3.1:. Scatter plot of the estimated and true variances.



Option prices with V_t estimated by the Nelson and Foster filter and the corresponding standard deviations for different levels of volatilities are listed in Table 3.2. The estimation error at the equilibrium of the variance, 0.01, is roughly distributed with a standard deviation 2.75×10^{-3}, which in term implies a standard deviation of 0.249 for the price 2.793. In fact, the ratios of the standard deviation to the price for the series of option contracts are around 9%. Obviously this is not acceptable in practice.

Table 3.2: Call option prices for different levels of volatilities.

V_t	Standard deviation of estimates of V_t	Call price	Standard deviation of estimated call prices
0.005	0.00202	2.301	0.217
0.0075	0.00288	2.558	0.281
0.01	0.00275	2.793	0.249
0.0125	0.00333	3.009	0.278
0.015	0.00392	3.212	0.310
0.0175	0.00144	3.403	0.331
0.02	0.00489	3.585	0.345

These results just show that pricing options with estimated variances indeed ignore the uncertainties of the estimates and can lead to unacceptable errors. Intuitively, a stochastic volatility model is necessary only when the volatilities change over time considerably. Under such circumstances, it is generally difficult to estimate the level of volatilities with the prices. On the other hand, when the volatilities can be successfully filtered with the prices, the volatilities would be locally constant or at least changing slowly. Thus plugging in the formula with estimated volatilities indeed involves logical inconsistencies.

All the above conflicts come from improper formulation of option prices. In the next section, the point will be discussed rigorously. Clearly, the filtration up to the current time t, Ω_t, plays an important role, and all the inference including option pricing should be based on the joint distribution of $(S_0, \cdots, S_t)$.

3.3 Which prices for options?

In the previous section, it is pointed out that filtering volatilities makes it arduous for pricing options under the stochastic volatilities. In this section, a proper formulation of the option prices will be carefully investigated to overcome the difficulties mentioned above. Based on a new scheme for estimation of the stochastic models in chapter 2, pricing options may proceed without plugging in estimates of variances. The algorithm will be presented in the next section.

From the very start, the school of mathematical finance accustomed themselves to diffusion-type models. The Morkovian properties of the models naturally lead to the formulation for the prices of options as (3.1). Since V_t is in fact an unobservable random variable, it is necessary to take expectations over it. That is, the formula,

$$E_{V_t}^{P}\left[E_{S_t}^{Q}\left(e^{-r(T-t)}g(S_T) \mid S_t, V_t\right)\right], \tag{3.3}$$

would be a more reasonable choice and so the commonly used one, (3.2) should be viewed as an approximation for (3.3). But unfortunately, the estimation error can be unacceptable for practical purposes as shown in the previous section.

To overcome this problem, it should be made clear that the information up to time t is in fact no more than the observed asset prices. So the correct formulation of the option prices ought to be

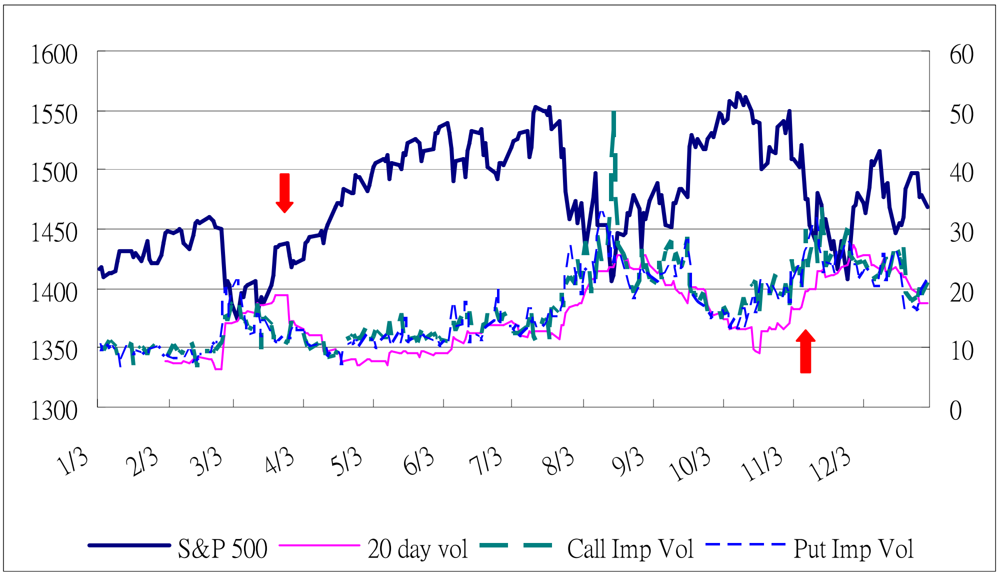
$$E_{S_t}^{Q}\left(e^{-r(T-t)}g(S_T) \mid S_0, \cdots, S_t\right).$$

This idea may be not so institutional. However, taking into consideration that volatilities are unobservable and thus ($S_0, \cdots, S_t$) are correlated and contain information about V_t, the formula is indeed trivial and more reasonable.

A further implication is that the option prices will depend on the past paths. In other words, two price processes may lead to different option prices even though they indeed have the same level of volatility at time t. Such property also cannot be deduced from the approach based on the Markovian property, especially when an unbiased estimator for V_t is used. However, the asymmetry for the upturn and down turn of the markets does exist in the real world.

Figure 3.2 shows the level of S&P 500 index, 20-day volatilities, implied volatilities of near-the-money calls and puts in 2007. It is clear that the levels of 20-day volatilities are very close around the bottom of March and the middle of September, but the levels of implied volatilities (for call and put options) actually differ a lot.

Figure 3.2: S&P 500 index in 2007: level of index, 20-day volatilities, implied volatilities for call and put options.



To sum up, calculating option prices, from the view of statistical inference, is then based on the joint distribution of (S_0, ···, S_t). As it is not available, an approximation scheme using partially observed GARCH models will be presented next.

3.4 Pricing options with the filtration consisting of prices

In this section, a new scheme for inference and pricing under the stochastic volatility models is proposed. The above two subtle difficulties may be overcome with this approach.

3.4.1 An algorithm for ML estimation of parameters and option pricing

In contrast to the stochastic volatility models, GARCH models provide as a tool that are feasible in practical operations. There are some facts about the two types of models to be pointed out.

1. For specific sequence of observations, the joint distributions of GARCH models are analytical, while generally the stochastic volatility models are not.
2. As the period of observations approaches 0, the GARCH models converge to stochastic volatility models in distribution, see Nelson (1990), Duan (1997) and Fornari and Mele (2004).

The first point implies that likelihood inference with the stochastic volatility models is generally infeasible. However, in light of the second argument, approximating stochastic volatility models with some specially designed GARCH is possible.

For this point, Brown et al. (2003) show the asymptotic equivalence of GARCH and stochastic volatility models. And the partially observed GARCH with MCMC methods for statistical inference is proposed in chapter 2.

Let $\left(\widetilde{X}^{(1)},\widetilde{V}^{(1)}\right)$ be drawn from the stochastic volatility model and $\left(\widetilde{Y}^{(1)},\widetilde{h}^{(1)}\right)$ from the corresponding GARCH with length of construction interval Δ. Denote ${}^{*}\widetilde{Y}_t=\left(Y_{t-1+\Delta},\cdots,Y_{t-\Delta}\right)$ as the augmented data between Y_t and Y_{t-1}. Note that the observed data is indeed $\widetilde{X}^{(1)}$, or equivalently $\widetilde{Y}^{(1)}$.

Past paths ${}^{*}\widetilde{Y}$ conditional on the observed data $\widetilde{Y}^{(1)}$ can be generated with the following algorithm.

1. Initialize V_0 and ${}^{*}\widetilde{Y}^{(\Delta)}$.
2. Update ${}_{i}h_0$ from $h_0|{}^{*}\widetilde{Y}$.

3. Update sequentially ${}^{*}\widetilde{Y}_t^{(\Delta)}$ from ${}^{*}\widetilde{Y}_t^{(\Delta)} \mid \widetilde{Y}^{(1)}, {}^{*}\widetilde{Y}_1^{(\Delta)}, \cdots, {}^{*}\widetilde{Y}_{t-1}^{(\Delta)}, {}^{*}\widetilde{Y}_{t+1}^{(\Delta)}, \cdots, {}^{*}\widetilde{Y}_T^{(\Delta)}; \eta$.

4. Repeat step 2 and 3 until convergence and take L independent paths.

The likelihood function can be approximated as

$$\sum_{i=1}^{L}\left(\pi({}_i h_0)\cdot\prod_{s=\Delta}^{n\Delta T}\left[\frac{1}{\sqrt{{}_i^{*}h_s^{(\Delta)}}}\phi({}_i^{*}\varepsilon_s^{(\Delta)})\right]\right).$$

By maximizing the above function, maximum likelihood estimator for the parameters can be obtained. Option prices then can be calculated based on the estimated parameters and the L independent paths.

5. Extending the L paths using general Monte Carlo methods with the risk-neutral measure under the GARCH model to get L samples of Y_T and thus the option prices.

This method converts the bivariate diffusion process with the second process unobservable into a partially observed univariate GARCH process. So it does provide an approximation for the formulation of the option prices (3.3), without requiring the estimation of the current variance, V_t or h_t. It is also applicable to various types of derivatives as long as the payoff function at maturity involves the prices only.

Furthermore an important implication is that the option prices at time t may depend on $(S_0, \cdots, S_t)$ in a much more complicated manner, instead of the estimated volatilities only. This important characteristic provides a way to distinct valuations of options conditional on different patterns of past paths. Next, a simple numerical experiment will be conducted for further studies.

3.4.2 Empirical performance of the algorithm

First, 12 disjoint paths whose variances at terminal ranges from 0.004 to 0.026 are selected. Both the Nelson and Foster filter and the partially observed GARCH methods are applied to estimate the variances. For the partially observed GARCH method, each observation period of 0.004 is divided into 5 subdivisions. Metropolis-Hasting algorithm is applied to each segment with 200 iterations of burn-in. After 100 iterations for the whole path, 1200 samples of paths are taken every two iterations.

Since there exists a monotone relation between the options prices, it suffices to compare the estimated variances at time t. From Figure 3.3, it can be seen that the

Nelson and Foster filter actually provides as an asymptotic unbiased estimator, although it tends to have larger variance. In contrast, the partially observed GARCH estimates seem to be less volatile but biased upward at lower and intermediate level of the true variance at terminal point.

Figure 3.3: Comparison between the filters.

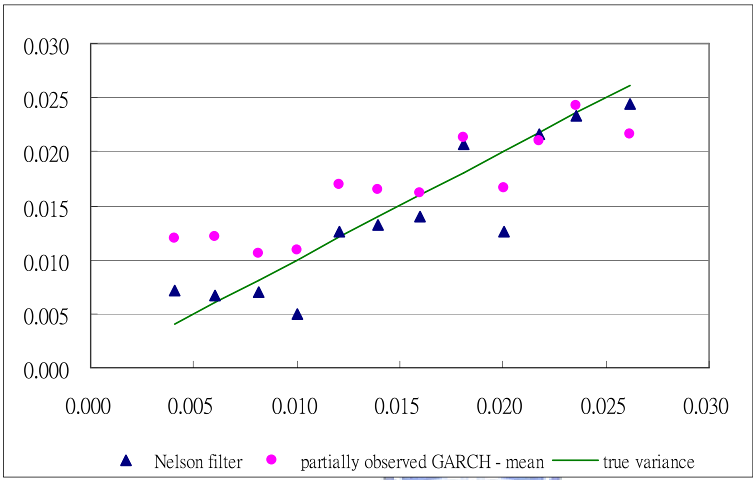
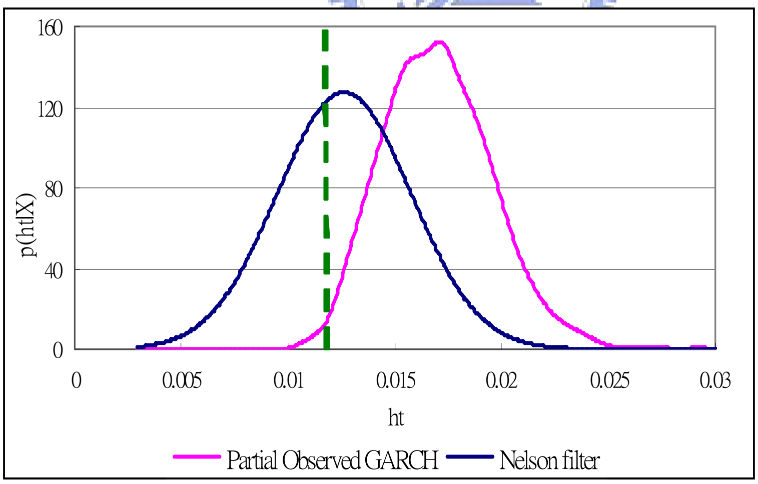


Figure 3.4: Conditional distribution of V_t given $(S_0,\cdots,S_t)'$ for the path with terminal point having level of variance about 0.012.

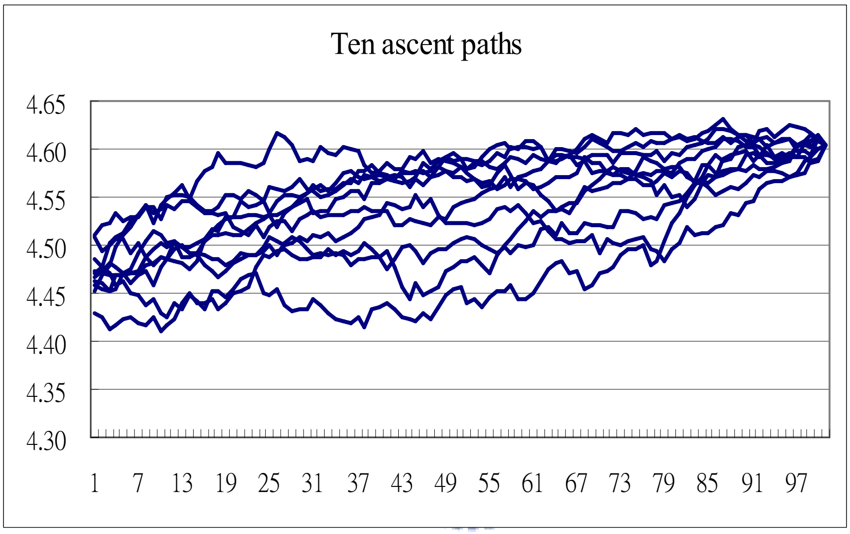


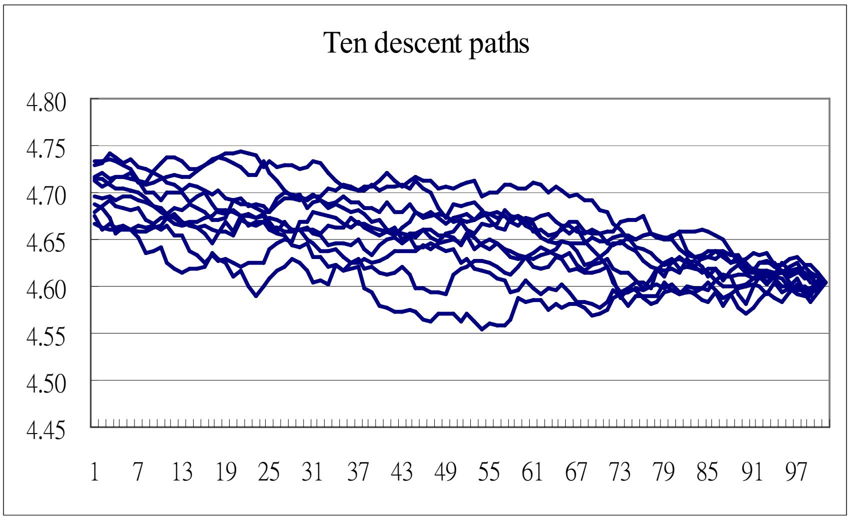
The observed unbiasedness means that the option prices obtained with the Nelson and Foster filter essentially match the prices by assuming that the true variances are known. On the other hand, the option prices from the partially observed GARCH tend to be higher compared to the "true" prices as the level of the true variance at terminal is under or near the long term equilibrium level of the variance process. The source of the bias may be partly due to the discretization error. But more probable, it may indicate that the conditional distribution of V_t given $(S_0,\cdots,S_t)'$ is positively biased as in Figure 3.3. The property coincides with the phenomenon that the implied volatilities are generally higher than the historical volatilities or most of model-implied volatilities. Of course, more details are worth further investigations.

Next, from the very long simulated path, 10 ascent and 10 descent disjoint paths of length 100 are selected according to the conditions below:

1. The variance at the last point is between 0.0097 and 0.013.
2. The absolute change rate through the 100 periods is the largest.

Figure 3.5: Ascent and descent paths.

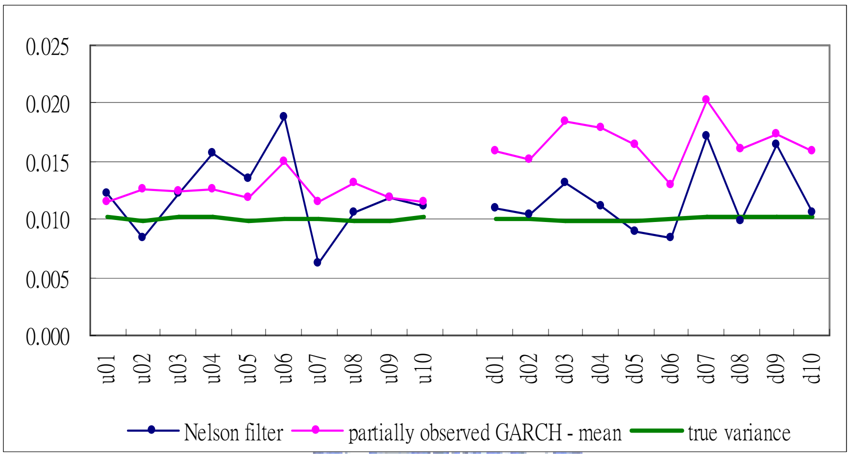




These paths are shown in Figure 3.5. And in Figure 3.6, the true volatilities, the Nelson and Foster filter and partially GARCH, implied volatilities are plotted together.

Again, the unbiasedness of the Nelson and Foster filter is revealed, indifferent to ascent or descent paths. However, in addition to the positive bias, the partially observed GARCH implied volatilities are definitely dependent on the past paths. There clearly exist significant differences between the volatilities from ascent paths and descent paths. This property certainly may help to explain the asymmetry of the markets shown in Figure 3.2.

Figure 3.6: Variance estimates for ascent and descent paths.



In addition to the consistency in theory and integration in practical operation, the results summarized above show that the partially observed GARCH approach provides huge potential to rationalize the behavior of the participants in the option markets, while the traditional method does not. A major key is of course the fact that volatilities are unobservable and should not simply be estimated. Furthermore, the method leads to a one-to-one relation between the risk premium of volatilities and option prices. As recently many researches are focused on the dynamics of implied volatility surfaces, alternative choice such as premium surface can be explored to further understanding of the behavior of the financial markets.

3.5. Conclusions

This study provides as an application of the partially observed GARCH for the inference about the stochastic volatility models. As emphasized in documented literatures, volatility could be the most important latent variable in modeling the financial time series. But as it is not observed, the prices are in fact correlated and inferences generally become much more complex.

Currently, the common approach to obtain these unobserved volatilities is by minimizing some loss functions or by filtering with prices. Such practices by intuition evidently lead into logical difficulties or contradictions with the reality. Moreover, this even casts away the advantages of the elegant modeling as shown in 4.2. Fortunately, it is illustrated here that inferences about the (option) prices without estimating the volatility processes may proceed well.

Off with absurd loss functions and piece-together calibration procedures, pricing derivatives can be more scientific and systematic. The results of simulations also suggest that there can be found more fantastic properties that really meet the real-world markets. What come next would be further studies about the premium surface, which is reasonably expected to reflect the psychological response of the market participants.

To sum up, the study proposes and demonstrates a totally new scheme for pricing options under the stochastic volatility models. Further studies should be conducted for the properties of the partially observed GARCH model. For example, the possible biases for filtering volatilities should be well explored since it may have significant implications as stated in 4.2. Combining the advantages of GARCH and diffusion types of models, alternative approach for modeling financial time series may be initiated.

3.6 Appendix

2.5.1 CEV GARCH family and their limiting processes

Fornari and Mele (2004) considered the following processes for short-term rate with its volatility having a constant elasticity of variance process,

$$dr_t = (\mu - \xi\, r_t)dt + \sigma_t \sqrt{r_t}\, dW_t^{(1)},$$

$$d\sigma_t^{\delta} = (\omega - \varphi\sigma_t^{\delta})dt + \psi\sigma_t^{\delta\eta} dW_t^{(2)}.$$

The model can be modified to the Heston' model by setting the price process as the geometric Brownian motion and (δ, η)=(2,1/2) in the volatility process.

The discrete time counterpart with length of construction Δ is then

$$r_{(k+1)\Delta} = r_{k\Delta} + \mu_{\Delta} - \xi_{\Delta} r_{k\Delta} + \sigma_{(k+1)\Delta}\sqrt{r_{k\Delta}}\, u_{(k+1)\Delta}$$

$$\sigma_{(k+1)\Delta}^{\delta} = \sigma_{k\Delta}^{\delta} + \omega_{\Delta} - \left(1 - \Delta^{-\frac{\delta\eta}{2}} E\left[\left|u_{k\Delta}\right|^{\delta\eta}(1-\gamma s_k)^{\delta\eta}\right]\alpha_{\Delta} - \beta_{\Delta}\right)\sigma_{k\Delta}^{\delta}$$

$$+ \alpha_{\Delta}\left(\left|u_{k\Delta}\right|^{\delta\eta}(1-\gamma s_k)^{\delta\eta} - E\left[\left|u_{k\Delta}\right|^{\delta\eta}(1-\gamma s_k)^{\delta\eta}\right]\right)\Delta^{-\frac{\delta\eta}{2}}\sigma_{k\Delta}^{\delta\eta},$$

where $u_{k\Delta}/\sqrt{\Delta}$ is general error distributed with shape parameter v and s_k is the sign of $u_{k\Delta}/\sqrt{\Delta}$.

Denote the following quantities,

$$\nabla_v^2 = \frac{\Gamma(v^{-1})}{2^{2/v}\Gamma(2v^{-1})},$$

$$m_{\delta,v} = \frac{2^{2\delta/v}\nabla_v^{2\delta}\Gamma(\frac{2\delta+1}{v})}{\Gamma(v^{-1})},$$

$$n_{\delta,v}\ \frac{2^{2\delta/v-1}\nabla_v^{\delta}\Gamma(\frac{\delta+1}{v})}{\Gamma(v^{-1})},$$

$$\varphi_{\Delta} = 1 - n_{\delta\eta,v}\left((1-\gamma)^{\delta\eta} + (1+\gamma)^{\delta\eta}\right)\alpha_{\Delta} - \beta_{\Delta}$$

$$\psi_{\Delta} = \sqrt{(m_{\delta\eta,v} - n_{\delta\eta,v}^{2})\left((1-\gamma)^{2\delta\eta} + (1+\gamma)^{2\delta\eta}\right) - 2n_{\delta\eta,v}^{2}(1-\gamma)^{\delta\eta}(1-\gamma)^{\delta\eta}}\,\alpha_{\Delta}$$

and $$\rho = \frac{2^{\frac{\delta\eta - v+1}{v}} \nabla_{v}^{\delta\eta+1}\Gamma\left(\frac{\delta\eta+2}{v}\right)\left((1-\gamma)^{\delta\eta} - (1+\gamma)^{\delta\eta}\right)}{\Gamma(v^{-1})\sqrt{\left(m_{\delta\eta,v} - n_{\delta\eta,v}^{2}\right)\left((1-\gamma)^{2\delta\eta} + (1+\gamma)^{2\delta\eta}\right) - 2n_{\delta\eta,v}^{2}(1-\gamma)^{\delta\eta}(1-\gamma)^{\delta\eta}}}.$$

Suppose that $\lim_{\Delta\to 0}\Delta^{-1}\mu_{\Delta} = \mu$, $\lim_{\Delta\to 0}\Delta^{-1}\xi_{\Delta} = \xi$ and

$$\lim_{\Delta\to 0}\Delta^{-1}\omega_{\Delta} = \omega \in R^{+},$$

$$\lim_{\Delta\to 0}\Delta^{-1}\varphi_{\Delta} = \varphi < \infty$$

$$\lim_{\Delta\to 0}\Delta^{-1/2}\psi_{\Delta} = \psi < \infty.$$

Then, the process $(r_{k\Delta}, \sigma_{k\Delta}^{\delta})$ converges to (r_t, σ_t^{δ}) in distribution as $\Delta \to 0$.

4. Which Loss Functions for Option Pricing?

4.1 Introduction

Since Black-Scholes (1973, henceforth BS) and Merton (1973) originally constructed their option pricing formulas, a number of different new models have been developed for option valuation. Each of these models relaxes some of the restrictive BS assumptions for the unrealistic assumptions of the BS world. For instance, an important class of models specifies the heteroskedasticity of the price returns (see, e.g., the stochastic volatility models of Hull and White (1987), Scott (1987), Stein and Stein (1991) and Heston (1993), the stochastic volatility jump diffusion models of Bates (1996) and Scott (1997) and discrete-time GARCH model of Duan (1995) and Heston and Nandi (2000)). In most of these articles it is emphasized that their model is more practical and that the corresponding model-implied option prices can fit the market-observed option prices better. However, a solid foundation for model valuations does not seem to exist. In this chapter, some commonly used loss functions are investigated and a procedure to construct reasonable loss function is proposed.

It is well accepted in the literatures that the choice of loss function is important for parameter estimation and model evaluation. First of all, consistency in the two stage is essentially necessary (Christoffersen and Jacobs, 2004). Next, Engle (1993) already argued that the choice of loss function implicitly defines the model under consideration. However, it is unfortunate that the standard option valuation theory implies a unique option price, but mentions nothing about how to specify the error term (Renault, 1997). The choice of loss function is a vital step in model estimation and evaluation as it implicitly assumes some specific error structures. For example, using loss function such as sums of squares of pricing errors implies the assumption of white noise errors.

There are many loss functions commonly used for model comparison such as root mean squared dollar errors ($RMSE), root mean squared relative errors (%RMSE), mean absolute dollar errors ($MAE) and mean absolute relative errors (%MAE). These criteria consist mainly of differences between model-implied prices and market prices. That is, ($\hat{C}_{model} - C_{market}$).

A quickly and easily seen conflict for this class of loss functions is that the option price or the pricing error does not contain any information about the market. More specifically, two different option contracts can have the same price and even pricing error under some models, but they in fact reflect very different market scenarios. That is, only when accompanied with its specification and prices of underlying asset, the price of an option contract can be informative. In fact, when considering the information contents of option prices, implied volatilities play an important role in all major aspects.

So there appear in recent years new performance criteria such as implied volatility root mean squared error (IVRMSE). Clearly, such criteria can be viewed as designed specifically for option model. In fact, many authors in the literature felt inadequate using only dollar based performance criteria for model selection so that they supplemented with the implied volatility graphs to help them evaluate models. Rubinstein (1985) and Bakshi, Cao and Chen (1997) diagnosed the relative model misspecification by comparing the implied-volatility patterns of each model across both moneyness and maturity. Besides, Canina and Figlewski (1993) pointed out that if a model is to provide a plausible explanation of market price, then it needs to be consistent with the observed "smile" across strike prices in the BS model. These arguments all support the use of implied volatility loss function for model evaluation instead of dollar based loss functions. Pan (2002) used IVMSE to measure model performance instead of dollar based loss function. He thinks that this avoids placing undue weight on expensive options, generally those are options deep-in the-money with longer time to maturity. Duan (1996) used IVMSE not only to evaluate models but also to estimate option models.

In this chapter all the commonly used performance criteria mentioned above are carefully investigated. Statistical properties are examined using the ad-hoc Black-Scholes model and GARCH pricing models applied to TAIEX options. All evidences show that there exist significant violations to the fundamental assumptions for the loss functions.

On the other hand, once it is reached the consensus that the loss function should be based on the implied volatilities, some well known properties may be used to construct the required loss function. Among all of them, mean reversion would be the first choice as it can be quantitatively characterized. Furthermore, since all structural pricing models are based on some equilibrium conditions, the role of the implied volatilities induced by the model prices would be just the target of mean reversion for the real implied volatilities naturally. Then the speed of mean reversion is meaningful and can be used as the loss function for model evaluation.

The remainder of this chapter is organized as follows. Section 4.2 introduces these common performance criteria. Section 4.3 briefly provides a description of the TAIEX option data, then demonstrates that the performance criteria of loss functions based on pricing errors have heteroskedastic pattern. Section 4.4 is a discussion on how to build a reasonable loss function. Finally, concluding remarks are offered in Section 4.5.

4.2 Commonly used loss functions and their characteristics

In this section some commonly used loss functions are introduced. They can be divided into two major classes based on pricing errors and implied volatility errors.

The first class of loss functions for model evaluation can be divided into two subclasses further, which are the dollar loss functions relate only to the pricing error , $\hat{C}_{model} - C_{market}$, and the relative error loss functions to the relative error, $(\hat{C}_{model} - C_{market}) / C_{market}$.

Within the first subclass, one frequently used loss functions is the root mean squared dollar errors (\$RMSE) given by

$$\$RMSE \equiv \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{C}_i - C_i)^2} \qquad (4.1)$$

where $\hat{C}_i$, C_i and n are the model-implied option price, market-observed option price and the number of option contracts used. An immediate alternative is the mean absolute dollar errors (\$MAE) defined by

$$\$MAE \equiv \frac{1}{n}\sum_{i=1}^{n}\left|\hat{C}_i - C_i\right|. \qquad (4.2)$$

The error structure for the dollar loss functions can be written as

$$\hat{C} = C + \varepsilon_{\$}. \qquad (4.3)$$

where $\varepsilon_{\$}$ is a white noise. It is seen that the dollar loss functions are advantageous in being easily interpreted. However, it is expected that the relatively wide range of option prices across moneyness and maturities would raise the problem of heteroskedasticity. In-the-money and long-term contracts usually have higher option prices and thus tend to correspond to higher pricing errors, so dollar loss functions implicitly assign more weight to this group of contracts and thus would tend to choose the model which is outstanding in high value option contracts.

To overcome this problem some researchers prefer using loss functions based on relative errors. Examples include the root mean squared relative errors (%MRAE), defined as

$$\%RMSE \equiv \sqrt{\frac{1}{n}\sum_{i=1}^{n}((\hat{C}_i - C_i)/C_i)^2}\ , \tag{4.4}$$

whereas the mean absolute relative errors (%MAE), given by

$$\%MAE \equiv \frac{1}{n}\sum_{i=1}^{n}\left|(\hat{C}_i - C_i)/C_i\right| \tag{4.5}$$

And the error structure of relative error loss functions is then

$$\hat{C} = C + C\varepsilon_{\%}\,. \tag{4.6}$$

The relative error loss functions are also easy to understand and comparatively conform to common sense if all market participants have their portfolios completely constructed by options. Then $1 error on a $10 option is more serious than $1 error on a $100 option, isn't it? However, since options behave very differently from the spots, the error structure (4.6) could overcorrect in practice, and the out-of-the money and short term options with value very low will implicitly be assigned too much weight.

Dollar loss functions and relative error loss functions are both widely used in the literatures and sometimes even both criteria are applied together. Chernov and Ghysels (2000) used $RMSE and %RMSE, Dumas, Fleming, and Whaley (DFW) (1998) used $RMSE and $MAE, Heston and Nandi (2000) used $RMSE, %RMSE and $MAE (only when the valuation error does not fall within the bid-ask spread), and Bakshi, Cao and Chen (1997) applied $MAE and %MAE to evaluate their option pricing model. Although these criteria appear in most of articles concerning empirical investigations on option pricing theories, the error structures (4.3) and (4.6) are seldom examined.

On the other hand, for the nonlinearity of the payoffs of options, there are loss functions designed specifically for option valuation problems. With the convention of quoting option price in terms of volatility on the market, some researchers favor estimating option pricing models by minimizing the mean squared errors of the implied volatility of the BS formula. It is therefore the implied volatility root mean squared errors (IVRMSE) defined as

$$IVRMSE \equiv \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\sigma}_i - \sigma_i)^2} \qquad (4.7)$$

where the implied volatilities are

$$\sigma_i = BS^{-1}(C_i, T_i, X_i, S, r)$$

and

$$\hat{\sigma}_i = BS^{-1}(\hat{C}_i, T_i, X_i, S, r)$$

and BS^{-1} is the inverse of BS formula, T_i, X_i, S and r are time-to-maturity, strike price, the price of underlying asset and risk free-rate respectively. Similarly, the implied volatility mean absolute errors (IVMAE) is given by

$$IVMAE \equiv \frac{1}{n}\sum_{i=1}^{n}\left|\hat{\sigma}_i - \sigma_i\right|. \qquad (4.8)$$

And the error structure of implied volatility loss functions is then

$$\hat{\sigma} = \sigma + \varepsilon_{IV} \qquad (4.9)$$

or equivalently

$$\hat{C} = C^{BS}(\sigma + \varepsilon_{IV}). \qquad (4.10)$$

It is noted that the option price becomes a nonlinear function of the implied volatility error term. This reflects the fact that option prices are expectation values of nonlinear functions of asset prices, and utilizing this class of loss functions also implies that prices do not contribute to information discovery directly.

There can be also found some articles in which the implied volatility based loss functions are used, for example Duan (1996) and Pan (2002).

4.3 Empirical investigation of the error structures

To investigate empirically the error structures implied by the loss functions, a dataset from TAIFEX is analyzed with three pricing models, ad hoc BS, NGARCH and

EGARCH models. All the corresponding properties mentioned above will be investigated.

4.3.1 Data description

The sample contains reported prices of TAIEX options traded on the Taiwan Futures Exchange (TAIFEX) over the period July 2002 through June 2004. The underlying asset is the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX). TAIEX options are European-style and expire at the open of the market on the trading day following the third Wednesday of the delivery month. More details about the specification of TAIEX options can be found at the website http://www.taifex.com.tw.

The raw data is collected directly from the website of the exchange. Similar to Bakshi, Cao and Chen (1997) and Dumas, Fleming and Whaley (1998), several criteria are used to construct the dataset. First, quotes for call options from 1:15 p.m. to 1:25 p.m. in every trading day are collected. Although the same option may be quoted again in the time window (with the same or different index levels) on a given day, only the last record of that option is included in our sample. Second, option data with less than six days or more than one hundred days to expiration are eliminated to avoid the expiration-related price effects. Third, option data whose absolute "moneyness" ($\left|\frac{S}{K}-1\right|$) is greater than 10 percent are also eliminated. These options may induce liquidity-related biases because they are usually not actively traded in the market. Forth, the option contracts quoted less than 1 point are excluded because of the transaction cost effect. Finally, option quotes lower than their intrinsic values are excluded.

These criteria yield a sample of 5957 observations. Table 4.1 describes the sample characteristics of the call option prices employed in this work. Average prices and the number of available calls are reported for each category. Moneyness is defined as the ratio of the spot price to the exercise price (S/K). A call option is said to be deep out-of-the money if moneyness belongs to the interval (0.90, 0.97); out-of-the-money (OTM) if $0.97 \le S/K < 0.99$; at-the-money (ATM) when $0.99 \le S/K < 1.01$; in-the-money (ITM) when $1.01 \le S/K < 1.03$; and deep-in-the-money if $1.03 \le S/K < 1.10$. In terms to maturity, option contacts can be classified as short-term ($\le$30 days), medium-term (30-60 days) and long-term (>60 days).

The annualized 1-Month Deposit Rates obtained from the Central Bank of China are used as the proxy of the risk-free interest rates in pricing these options.

Table 4.1: Sample characteristics of TAIEX index options

S/K		T-t<=30	31<T-t<60	T-t>60
0.90-0.97	N	1225	928	266
	Average Price	32.66	76.58	112.11
	Price STD	27.50	40.60	58.49
	Average IV	0.2796	0.2577	0.2232
	IV STD	0.0563	0.0622	0.0521
0.97-0.99	N	467	347	73
	Average Price	76.40	134.22	178.42
	Price STD	39.21	48.48	65.08
	Average IV	0.2580	0.2415	0.2178
	IV STD	0.0597	0.0590	0.0463
0.99-1.01	N	455	342	65
	Average Price	118.81	178.80	232.40
	Price STD	45.49	52.90	74.52
	Average IV	0.2529	0.2360	0.2192
	IV STD	0.0613	0.0602	0.0517
1.01-1.03	N	426	245	35
	Average Price	182.54	242.91	307.17
	Price STD	49.10	61.59	87.56
	Average IV	0.2544	0.2388	0.2293
	IV STD	0.0627	0.0612	0.0516
1.03-1.10	N	703	328	52
	Average Price	338.86	369.14	472.06
	Price STD	108.07	104.16	137.83
	Average IV	0.2906	0.2405	0.2578
	IV STD	0.0912	0.0610	0.0696

4.3.2 Statistical properties of the errors

As many authors, for example Bakshi, Cao and Chen (1997), point out, dollar-error based loss functions generally put more weights on certain groups of options. In fact, there actually exist obvious patterns of pricing errors and relative errors to option prices, which in nature induce different weights on each contract. The following investigations will demonstrate the phenomenon and thus lead to violations to the assumptions (3) and (6) for these criteria are not satisfied.

In fact, the investigations proceed just as stated in the elementary statistics textbooks, and all we have to do is simply see if there exist certain patterns in the scatter plots of pricing errors or relative pricing errors to prices.

Following Christoffersen and Jacobs (2004) on using consistent loss function at both the estimation and evaluation stages, the parameters of the ad hoc Black-Scholes model are estimated daily by minimizing $RMSE, %RMSE and IVRMSE respectively. On each day parameters are estimated with data from the past five days. The residual plots with respect to the option prices under the ad hoc Black-Scholes model are presented in Figure 4.1. In Figure 4.1a, the heteroskedastic pattern of the pricing errors versus the observed option prices is obvious. When the call price is close to zero, the errors tend to be smaller. The amplitude of the errors increases as the call price increases until the call price reaches around $100, and then decreases slightly. Figure 4.1b also shows that the relative errors still have a significant heteroskedastic pattern. Especially, when the option prices are small, the relative pricing errors become terribly large. This means that using relative error measure tend to overcorrect. In Figure 1c, it can be seen that the residuals are distributed with seemingly larger variances while the option prices are large. These results reveal that all the above loss functions are not consistent with the homogeneity assumptions.

Beside the simple model, it is expected that the same results can be obtained under the GARCH option pricing models. In applying GARCH models, the likelihood functions based on observations from asset prices can be expressed explicitly, so the MLE's of the parameters (λ, α_0, α_1, β_1, γ) are obtained by numerical methods and then applied to the pricing of options. No information from option prices will be used to determine any parameters so the inconsistency problem mentioned in Christoffersen and Jacobs (2004) shall not be confronted.

The parameters are estimated month by month using the maximum likelihood method on the TAIEX daily closing prices in the past year and then volatilities are updated day by day. Residual plots of the GARCH option pricing models are presented in Figures 4.2 and 4.3. Similar patterns to those in Figure 4.1 can be easily found. And furthermore, the IV errors for higher option prices tend to be negative instead of concentrating around zero.

Figure 4.4 shows the scatter plots of the IV errors versus the real implied volatilities for all the three models. Obvious patterns can be seen in all the three panels. In fact, due to the temporal structure of volatilities, there certainly exist autocorrelations among all these volatilities and the errors cannot be independent so as to exhibit patterns of correlations.

All the above results state that the commonly used loss functions may be in lack of solid statistical grounds, no mention economic interpretations. An important issue is the role of the pricing models to be expected – a good predictor or an equilibrium level under the assumption of reasonable behaviors.

Figure 4.1: Residual plots of the ad hoc BS model versus call price.

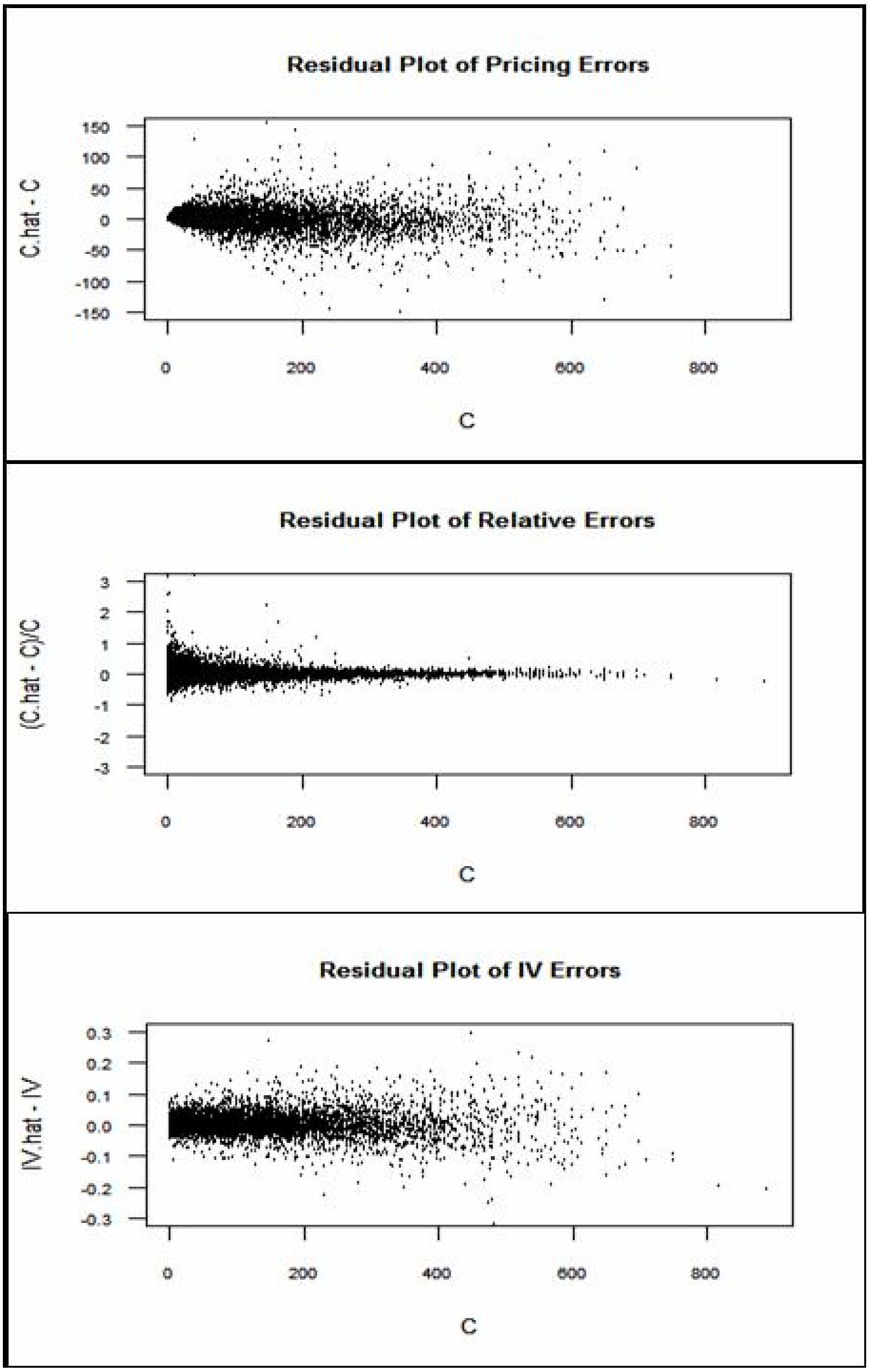


Figure 4.2: Rresidual plots of the NGARCH model versus call price.

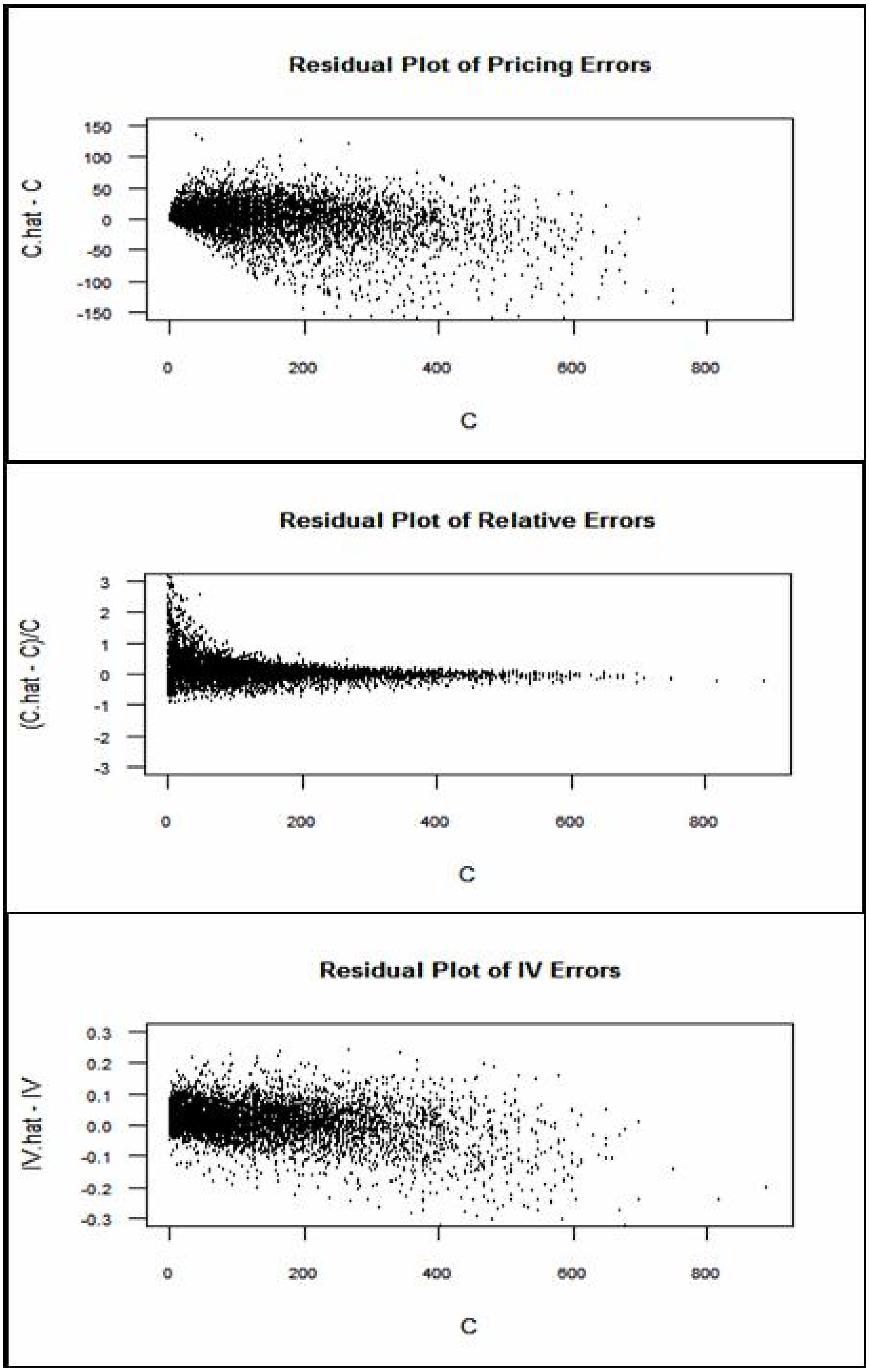


Figure 4.3: Residual plots of the EGARCH model versus call price.

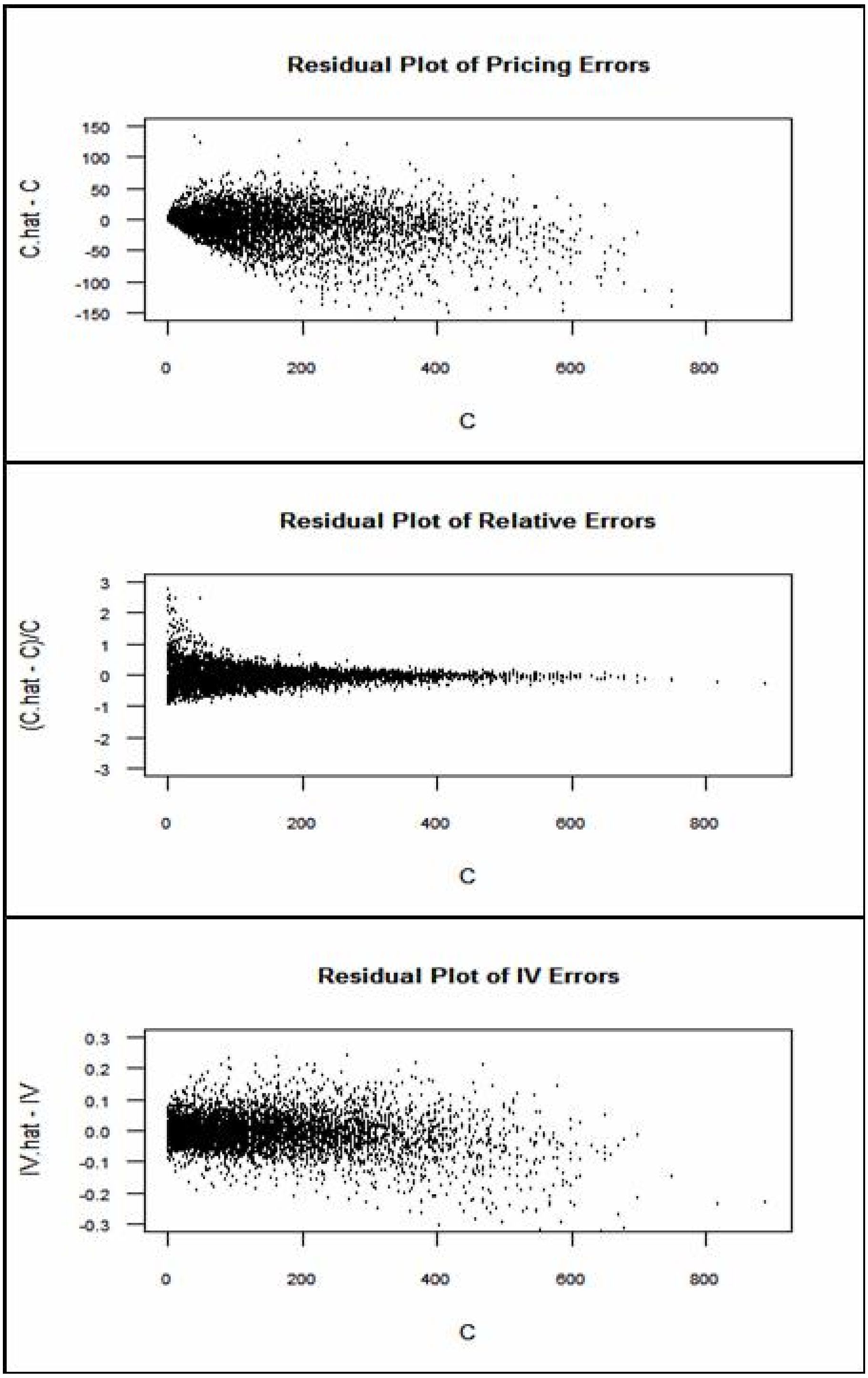
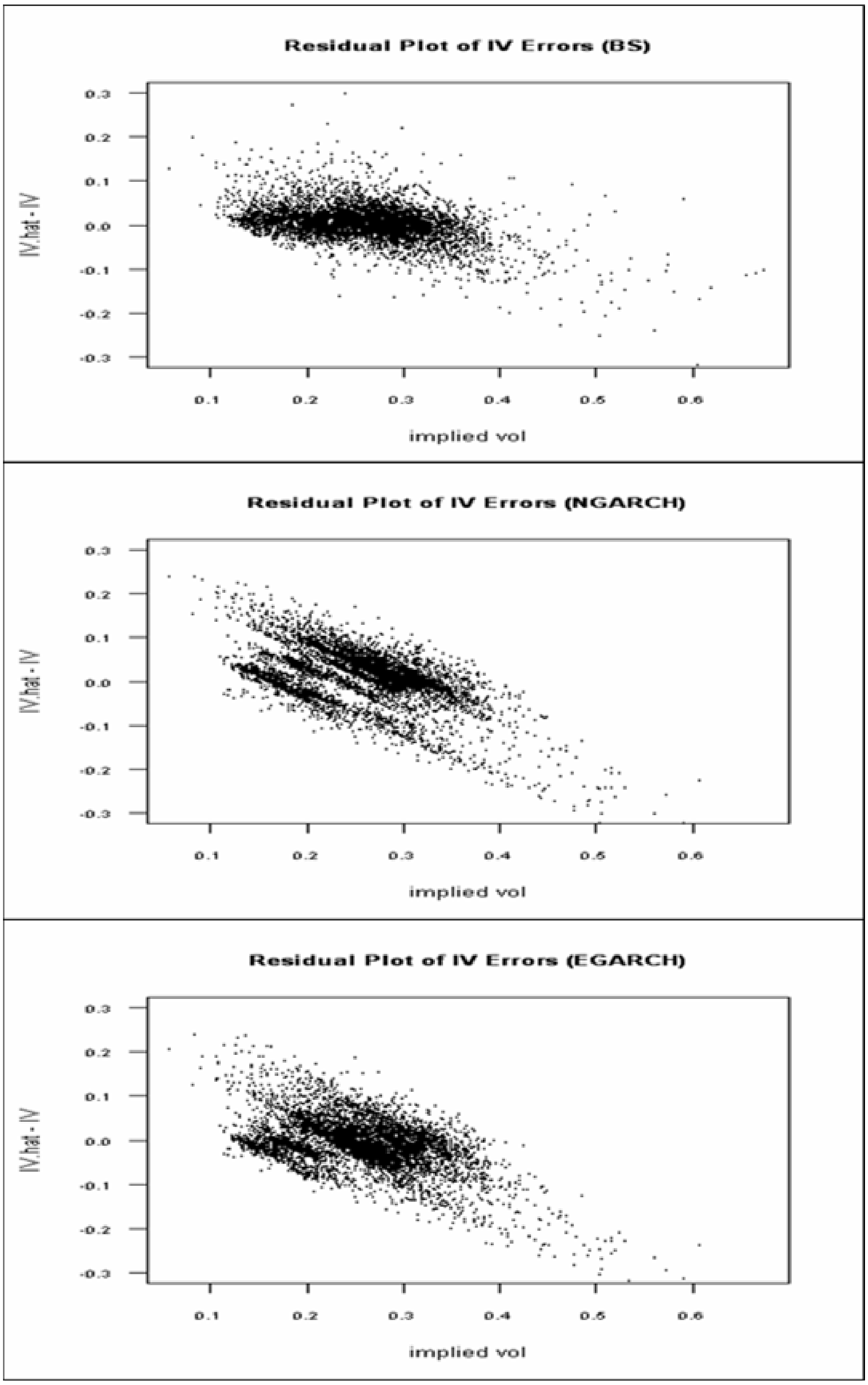


Figure 4.4: Implied volatility residual plots.



4.4 Aspects toward a rational loss function

4.4.1 Inconsistency of dollar based loss functions in information revealing

On investigating the performances of option pricing models, it should be noted that pricing errors or even the prices themselves are meaningless without incorporating current prices of underlying assets and the specifications of the options such as the strike price and time to maturity.

Consider the quotes for the two contracts listed in Table 4.2. The two contracts have the same price but very different implied volatilities. Suppose that a trader obtains his volatility as 20% and then his fair prices for the two contracts will be 5.64. Clearly all values of the dollar based loss functions for both contracts will be almost identical. However, pricing contract 2 with volatility 20% can lead to much larger loss (or gain) since the two values appear to be very different than those of contract 1.This example illustrates that loss functions such as $MSE, $MAE, %RMSE or %RMAE can be very insensitive to the parameters of the model.

Table 4.2: Comparison of two call option contracts with different specification.

	Contract 1	Contract 2
Stock price	100	100
Strike	100	95
time to maturity	0.5	0.09
Interest rate	0	0
Actual price	5	5
Implied volatility	17.75%	7.50%
Price estimated with σ=20%	5.64	5.64

The difficulty confronted here for $MSA etc is just that the information about the specification of each option contract is not well incorporated. More specifically, pricing an option contract with an error of 5% or 1 dollar, say, does not help judge how a model or method has satisfactory performance.

In the real world, the price is not the only thing to be considered for the market. A basket of options can be traded as if people are trading volatilities. That is, volatility in some circumstances may be viewed as some kind of tradables. Therefore, options for them are not presented as the instruments to bet on or hedge against the direction of an underlying risk. Instead, options are motivated as the instrument of volatility. So volatility provides more information than option price. In practice, there are even option contracts quoted by BS implied volatilities instead of the prices.

4.4.2 Information contents of option prices

By the arbitrage pricing theory, in a frictionless and dynamically complete market, options would be redundant securities. This means that option prices contain no more information than stock prices. However, interestingly, there are still piled amount of literatures for exploring the information contents of option prices.

A major concern of these articles is the insider trading in the option market. It is noted that in this category of literatures the underlying asset and the corresponding option are usually considered as simultaneously traded securities. So the causality relation of the observed and option implied stock price changes would be put into comparisons. Typical examples via this approach include Stephan and Whaley (1990) and Chan, Chung and Johnson (1993). Chakravarty, Gulen and Mayhew (2004), with the information share of Hasbrouck (1995), measured relative contribution of each market to price discovery.

It should be emphasized again that in all of the above studies the option price itself is never used as the proxy of information. What is taken into comparisons with the price of the underlying asset is the implied price through the implied volatilities! This implicitly points to the invalidity of the dollar loss functions.

Another main theme considering the information content of option prices is if the implied volatility could provide a good forecast of the future volatility, for example Christensen and Prabhala (1998). From this context, a causality test is usually applied to detect the relation between the implied volatilities and future realized volatilities. The conclusions of these studies are very mixed. But what is important is that the implied volatilities play a major role in all of the above studies instead of the option prices. That is, even in documented academic literatures, implied volatilities have been viewed as the major characteristics of the option prices.

4.4.3 Black-Scholes formula as a self-fulfilling prophecy

As discussed above, the BS implied volatilities are indeed the sole of the information content of the option prices in academic researches. The dominance of the BS model is reflected in the fact that the implied volatility becomes the standard method of quoting option prices in industry.

Option traders routinely use the BS formula, although it is well known that the BS assumptions are not realistic. However, since there exists a one-to-one correspondence relation between the option price and the only parameter - volatility, the BS formula becomes easy to understand and remember. Thus practitioners would rather use the BS model instead of other complicated model.

The BS formula has become a standard among professionals and also in computer platforms. It provides a way to transform a volatility quote to a dollar value attached to this quote. This helps develop common platforms for hedging, risk managing, and trading volatility. Thus, once we accept that the use of the BS formula amounts to a convention, and that traders differ in the selection of the value of the parameter σ, the critical quantity is no longer the option price, but the volatility. This is also why for derivatives such as caps, floors, and swaptions the quote is just in term of volatility.

What is left is then the question: if the prices of the underlying assets do not possess a lognormal distribution as the BS model implies, will the market prices of options satisfy the formula just because the participants believe and use it? In a complete market, the answer for the above question cannot be positive as there will certainly produce opportunities for arbitrage.

However, in an incomplete market, the answer for the above question could be very different compared with in a complete market. Cherian and Jarrow (1998) demonstrated that the BS model can be a self-fulfilling prophecy. That is, in an incomplete market, even though the underlying asset's objective distribution is not lognormal, as long as all participants believe it is, the BS formula is still the equilibrium eventually. Therefore, even from the academic viewpoint, it is reasonable to use the implied volatility of the BS model as the characteristics of option prices.

4.4.4 Setting up a reasonable loss function

As it is assured that a reasonable loss function must be built through the implied volatilities or other similar quantities, what is left would be which property can be used to construct a loss function.

There are at least twenty years since the researchers in the finance and economics areas note the temporal structure of volatilities, see Engle and Patton (2001). Some major characteristics of volatilities have been well identified, for example persistence and mean reverting. The famous ARCH model by Engle and its variations are all addressing these properties. And furthermore the continuous time volatility models generally contain their drifting parts corresponding to the properties, especially mean reversion. These properties also explain the patterns of IV errors in Figure 4.4 since these observed and model projected volatilities would certainly be highly correlated.

On the other hand, it is noted that all pricing models are based on some kind of equilibrium arguments but there are in fact lots of noises in reality that cannot be covered by these theories. Thus, the pricing models should not provide as an absolute standard so that the real prices must be always close to the model projected. A more feasible approach is that the pricing models only reveal the equilibrium level at the specific moment. That is, the implied volatilities deduced from the structural model are viewed as the dynamic target of mean reversion for the observed implied volatilities. So a reasonable loss function from this context is just the speed of mean reversion. The parameters such as the risk premium can be obtained by maximization of the speed of mean reversion.

The simplest formulation for this problem can be obtained by assuming that the logarithm of the implied volatilities $U_t^{BS} = \log(\sigma_t^{BS})$ follows the Ornstein-Uhlenbeck process with a time varying target of mean reversion that is deduced from the structural model. More precisely, it may be assumed

$$dU_t^{BS} = -\kappa\left(U_t^{BS} - U_t^{M}\right)dt + \delta dW_t, \tag{4.11}$$

where κ is the speed of mean reversion, $U_t^{M} = \log(\sigma_t^{M})$ is the logarithm of implied volatility induced by the model price and W_t is a standard Brownian motion. An approximate discrete time version may be expressed as

$$U_t^{BS} = (1 - e^{-\kappa})U_{t-1}^{BS} + e^{-\kappa}U_t^{M} + \varepsilon_t, \tag{4.12}$$

where ε_t is an error term with constant variance. The ML estimator for κ is then

$$\hat{\kappa} = \frac{\sum\left(U_t^{BS} - U_{t-1}^{BS}\right)\left(U_t^{M} - U_{t-1}^{BS}\right)}{\left(U_t^{BS} - U_{t-1}^{BS}\right)^2}.$$

The above equations (4.11) and (4.12) indicate that the observed implied volatilities will always tend to move toward the level of model implied volatility from its previous level. Thus, a model is said to have better performance when it corresponds to higher speed of mean reversion, and versa. So a natural loss function would be just the estimated speed of mean reversion $\hat{\kappa}$.

Beyond constructing loss functions, an approach to examine if the model is effective can be also provided. By replacing U_t^{M} with a constant parameter, equation (4.12) just assumes simply an AR(1) structure for the implied volatility process U_t^{BS}. So what is left is just standard procedures for model selection. And obviously, the above simple modeling for the volatility can be replaced by another process that better describes that of volatilities, for example the square root process by Cox, Ingersol and Rubinstein(19??). However, more computation burden is of course necessary.

4.4.5 Investigations on the loss function

In practical implementation, a volatility index should be constructed first since the sample generally contains multiple contracts at each time period. The (old or new) CBOE volatility index would be the candidate to construct the necessary values of

$\log(\sigma_t^{BS})$ and $\log(\sigma_t^{M})$.

For a numerical illustration, the index is constructed using the same dataset in the previous section. A weighted average for the volatilities of the near term contracts are calculated day by day with the reciprocal of absolute moneyness as the weights.

Figure 4.5: Implied volatility index, real and by models.

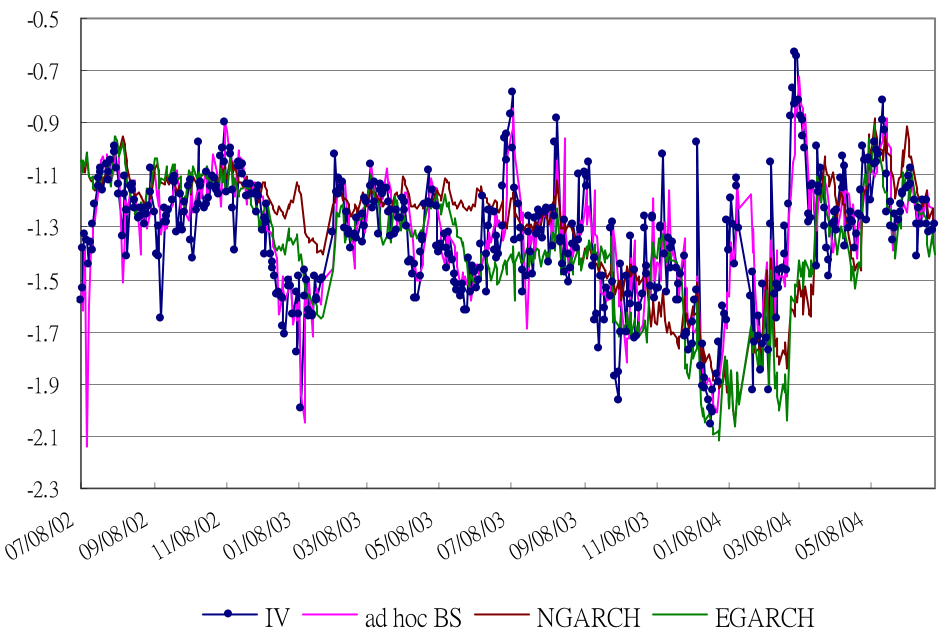


Figure 4.5 shows the dynamics of the real volatility index and the model implied ones by the three pricing models. It is easily seen that the ad hoc BS implied index proceed with the real implied volatilities adaptively since much more information are utilized with this method. On the other hand, the index form the two GARCH models just duteously play as the target of mean reversion for the long term. Theses facts are also reflected in Figures 4.6.a, b and c. The speeds of mean reversion are estimated day by

day with a rolling window of 65 days. Clearly, the ad hoc BS has much larger speeds of mean reversion over the whole period. And the two GARCH models are averagely lower but comparable to each other.

Figure 4.6a: Real and model implied volatility index and estimated speed of mean reversion with moving window by the ad hoc BS model.

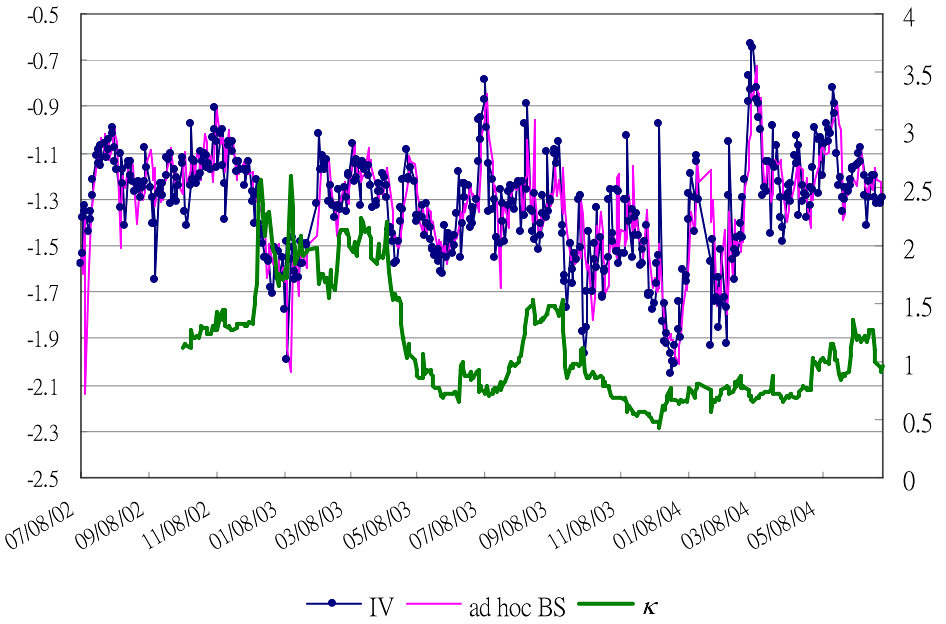


Figure 4.6b: Real and model implied volatility index and estimated speed of mean reversion with moving window by the NGARCH model.

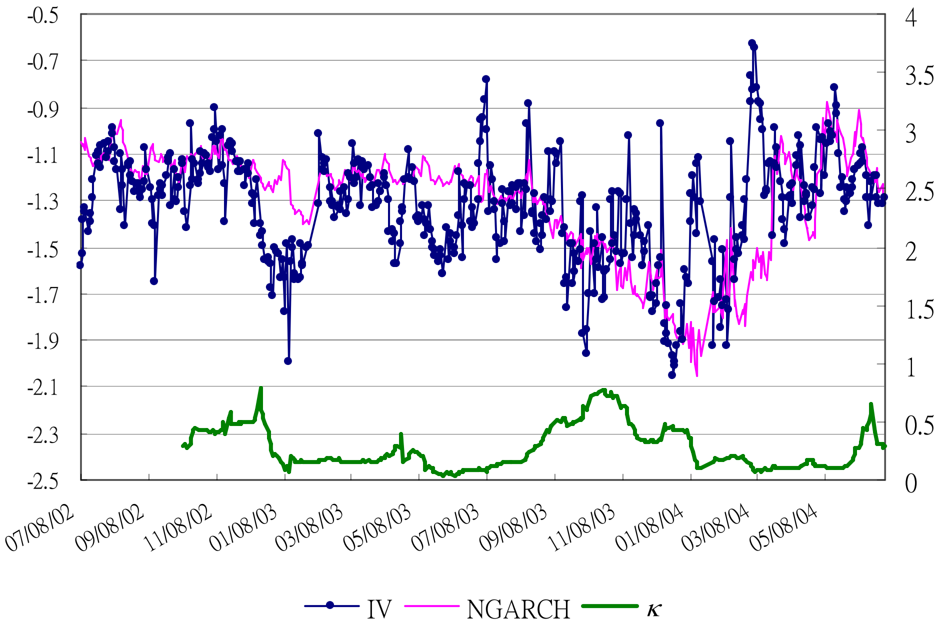


Figure 6c: Real and model implied volatility index and estimated speed of mean reversion with moving window by the EGARCH model.

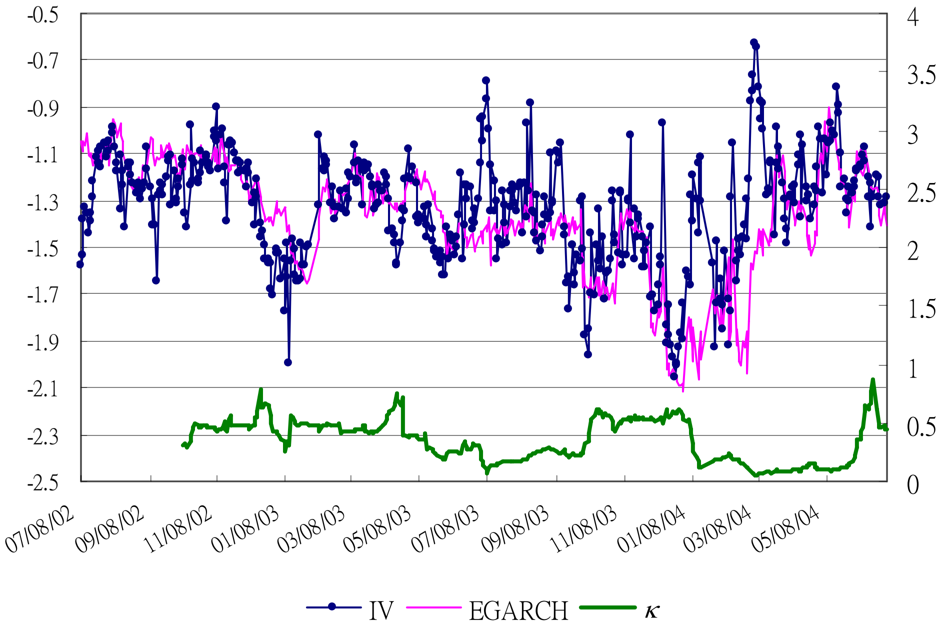


Figure 4.7 represents residual series from a simple AR(1) model and the mean reversion model in (4.12) with the target provided by the above three methods. It is clearly seen that different levels of serial correlation and volatility cluster exist among all four charts, especially obvious for the two from NGARCH and EGARCH. These evidence imply that the loss function based on (4.13) could be still too simple and not satisfactory, but acceptable for practical implementation since the ad hoc BS, commonly believed well behaved, has insignificant serial correlation and roughly local homogeneity of variance.

Figure 4.7a: Residual series of volatility index with AR(1) model.

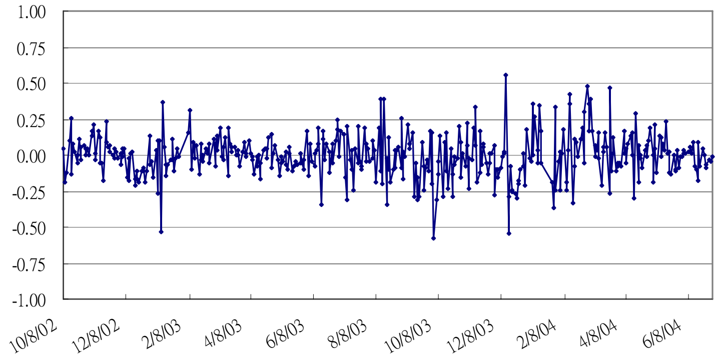


Figure 4.7b: Residual series of volatility index regressed with the ad hoc BS as the target of mean reversion.

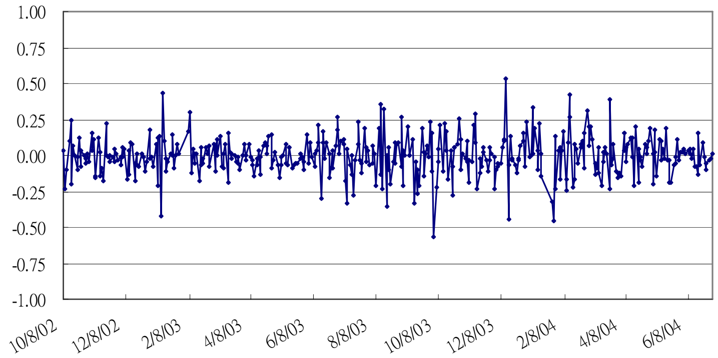


Figure 4.7c: Residual series of volatility index regressed with the NGARCH as the target of mean reversion.

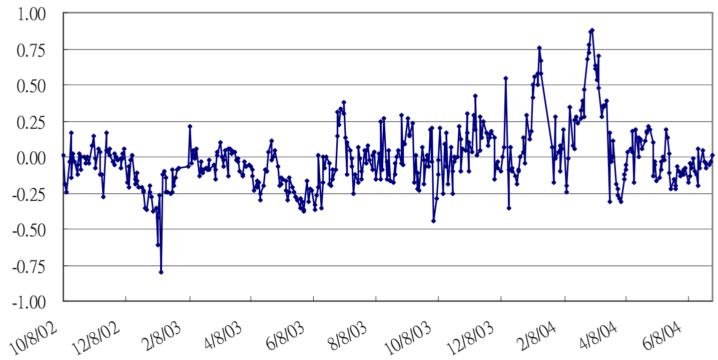
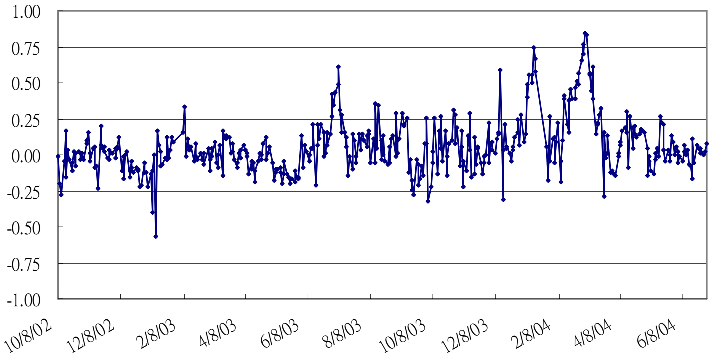


Figure 7d:Residual series of volatility index regressed with the EGARCH as the target of mean reversion.



4.5 Concluding Remarks

In this chapter a procedure for choosing appropriate loss functions for evaluating option pricing models is sketched. Inappropriate loss function may cause bias for model selection. And it has been shown that the traditional loss functions that are currently widely used are generally inappropriate. They are in lack of economic interpretations at first. Furthermore, significant heteroskedastic patterns of residuals exist so as to violate fundamental statistical assumptions of the underlying model.

There are two points to be emphasized. First, the loss functions composed of the option prices or pricing errors only are meaningless. As indicated in 4.1, the same prices and pricing errors can be easily made by options with different specifications. Implied volatilities, which in fact incorporate the information of prices and specifications, are the possible candidate for the construction of loss function. In fact, whether in industry or in academics, the Black-Scholes implied volatilities are commonly used as the advocates of option prices. In other words, the Black-Scholes may not be so “right”, but it actually provides as a platform to the investigation of options.

Second, a pricing model need not provide the correct price, but the level of the characteristics of the information content should tend to be. As mentioned earlier, all pricing models are based on some kind of equilibrium arguments. However, there are indeed large amounts of noises in the financial markets. And more important, market participants who are really risk neutral are rarely seen. More or less, people tend to be risk averse or seeking so that patterns of mean reversion would always exists in the financial markets. Thus, it is natural that the loss functions are constructed by the assumption that the models provide the level of mean reversion for the characteristics of information contents.

Obviously, further investigations on the detailed properties of the real and model implied characteristics should be made. And more elaborate modeling for their relations would be necessary in the future. For example, multivariate approaches such as implied volatility surface, see Benko et al. (2007), certainly provide much richer contents and possibilities than volatility index. And it should not be expected that a perfect loss function could be obtained before the information contents of option prices are well clarified.

4.6 Appendix

4.6.1 The ad hoc Black-Scholes model

The BS model assumes that the volatility of the underlying asset return is constant, which is not realistic. However, in practical applications, analysts and traders always use the market price to extract the implied volatility based on the BS model. One way to adjust the classic BS model is to allow each option to have its own BS implied volatility depending on the strike price and time-to-maturity. Following Dumas, Fleming and Whaley (1998) and Derman (1999) with some modifications, the following functional form for the options implied volatility is considered:

$$\sigma = a_0 + a_1 \frac{S}{K} + a_2 \left(\frac{S}{K} \right)^2 + a_3 T + a_4 T^2 + a_5 \left(\frac{S}{K} \right) T + \varepsilon_{IV}, \qquad (4.6.1)$$

where σ denotes the implied volatility, S is underlying asset price, K is strike price and T is time-to-maturity. For each contract with different exercise price and maturity the fitted value for volatility can be plugged back into the BS formula to obtain the model price.

In the empirical analysis of Section 4.3.2, heteroskedasticity of the traditional loss function with respect to the option prices is illustrated using the ad hoc Black-Scholes model estimated daily, because this approach is simple and widely used as a benchmark in the existing literature. Furthermore, to make sure that the ad hoc Black-Scholes model is not a special case for heteroskedasticity of traditional loss function, the properties of loss functions under the GARCH models are also investigated.

4.6.2 The GARCH pricing models

Since Engle (1982) and Bollerslev (1986), GARCH models have become one of the standard method for modeling financial time series. For obtaining pricing measures under the GARCH models, Duan (1995) first proposed the criteria Local Risk-Neutral Valuation Relationship (LRNVR). In this chapter, two types of GARCH models, NGARCH and EGARCH, are taken into consideration as illustrations. Following Duan (1996b), the two GARCH models and the corresponding pricing measures can be well defined.

The NGARCH model is defined as

$$\log S_t = \log S_{t-1} + r + \lambda\sqrt{h_t} - \frac{h_t}{2} + \varepsilon_t \,, \qquad (4.6.2)$$

$$h_t = \alpha_0 + \alpha_1 \cdot h_{t-1} \cdot \varepsilon_{t-1}^2 + \beta_1 \cdot h_{t-1} \,, \qquad (4.6.3)$$

where S_t and h_t are the spot price and volatility at time t, r is the risk-free interest rate, λ, α_0, α_1, and β_1 are nonnegative parameters, and ε_t's are i.i.d. normally distributed innovations with variances h_t. The corresponding pricing measure is then

$$\log S_t = \log S_{t-1} + r - \frac{h_t}{2} + \xi_t \,, \qquad (4.6.4)$$

$$h_t = \alpha_0 + \alpha_1 \cdot h_{t-1} \cdot \left(\xi_t - \lambda\sqrt{h_t}\right)^2 + \beta_1 \cdot h_{t-1} \,, \qquad (4.6.5)$$

where ξ_t is normally distributed with mean 0 and variance h_t.

Similarly, for the EGARCH model (4.6.3) is replaced by

$$\log(h_t) = \alpha_0 + \alpha_1\left(\left|\varepsilon_{t-1}\right| + \gamma \cdot \varepsilon_{t-1}\right) + \beta_1 \cdot \log(h_{t-1}) \,, \qquad (4.6.6)$$

and (4.6.5) by

$$\log(h_t) = \alpha_0 + \alpha_1 \cdot \left[\left|\xi_{t-1} - \lambda\sqrt{h_{t-1}}\right| + \gamma \cdot \left(\xi_{t-1} - \lambda\sqrt{h_{t-1}}\right)\right]$$

$$+ \beta_1 \cdot \log(h_{t-1}) \qquad (4.6.7)$$

5. Conclusions and discussions

This series of studies are focused on two important issues about option pricing. The first one concerns the implementation of continuous time stochastic volatility models, including procedures for estimation and pricing. The second is about the loss functions which are important both in parameter estimation and model evaluation for pricing options.

For the estimation of stochastic volatility models, a partially observed GARCH model is proposed to approximate the likelihood function. Beyond the purpose of estimation, this type of models indeed bridges the gap between the continuous time stochastic volatility models and the discrete time GARCH models. Furthermore, inferences including estimation and pricing can be done under the same schema.

The most fascinating is that the method provides an approach for pricing options based on an approximated conditional distribution of the future price given the prices up to now, instead of the current price and a “filtered” volatility. And this approach obviously leads to some interesting results, for example higher option prices accompanied by descent paths of prices and lower prices by ascent paths.

Clearly the properties of the partially observed GARCH models shall be worth further investigations. It is not clear if the filtered variance via this method will be an (asymptotic) unbiased estimator of the true variance. If not, it may help explain partly why option prices are generally higher than that by the Black-Scholes formula with volatility estimated by past prices, which in turn is about the behaviors of market participants when they cannot observe some state variables such as volatilities.

Loss functions are important and critical for option pricing, even though no theory can be found to formally address it. However, as pointed out in the third chapter, it is essential that the loss function should not be composed of option prices only since there are certainly much more other elements to constitute the information contents together.

From a practical point of view, the Black-Scholes implied volatility or other nonparametric volatility index is certainly to provide as a basis for constructing loss functions. Then mean reversion as a major characteristic of volatilities should be quantified for the introduction of loss functions. That is, a precise definition of mean reversion may be the next step to a more compact and meaningful loss functions.

Some elementary statistical concepts are incorporated into all of the studies. For example, error sum of squares used as loss functions should be based on homogeneity of variance, and taking expectations over unobserved random variables instead of just “estimating” it. For financial engineering as “engineering”, the introduction of these concepts may help increase the proportionality of science.

5. References

Anderson, T. G. and Lund, J. "Estimating continuous-time stochastic volatility models of the short-term interest rate". Journal of Econometrics, 77, 343-377, 1997.

Armstrong, J. S., and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons", International Journal of Forecasting, 8, 69-80, 1992.

Bakshi, G., Cao, C., and Chen, Z. "Empirical performance of alternative option pricing models", Journal of Finance, 52, 2003-2049, 1997.

Bates, D. "Jumps and stochastic volatility: exchange rate processes implicit in Deutsche mark options", Review of Financial Studies,9, 69-108, 1996.

Benko, M., Fengler, M. R., Hardle, W., and Kopa, M. "On extracting information implied in options", Computational Statistics, 22, 4, 543-553, 2007.

Black, F. and Scholes, M. "The pricing of options and corporate liabilities". Journal of Political Economy, 81, 637-651, 1973.

Bollerslov, T., (1986) Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, 307-327.

Brown, L. D., Wang, Y. and Zhao, L. H. "On the statistical equivalence at suitable frequencies of GARCH and stochastic volatility models with the corresponding diffusion models". Statistica Sinica, 13, 993-1013, 2003.

Canina, L., and Figlewski, S. "The informational content of implied volatility". The Review of Financial Studies 6, 659-681, 1993.

Carbone, R. and Armstrong, J. "Evaluation of extrapolative forecasting methods: Results of a survey of academicians and practitioners". Journal of Forecasting 1, 215-217, 1982.

Chan, K., Chung, Y. and Johnson, H. "Why option prices lag stock prices: a trading based explanation." Journal of Finance, 48, 1957-1968, 1993.

Chakravarty, S., Gulen, H., and Mathew, S. "Informed trading in stock and option markets." Journal of Finance, 59, 1235-1257, 2004.

Cherian, J., and Jarrow, R., "Options markets, self-fulfilling prophecies, and implied volatilities", Review of Derivatives Research, 2, 5-37, 1998.

Chernov, M., and Ghysels, E. "A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of option valuation". Journal of Financial Economics, 56, 407-458, 2000.

Christie, A."The stochastic behavior of common stock variances: Value, leverage and interest effects", Journal of Financial Economics, 10, 407-432, 1982.

Christoffersen, P., and Jacobs, K. "The importance of the loss function in option valuation", Journal of Financial Economics, 72, 291-318, 2004.

Cox, J. C., Ingersol, J. E., Jr., and Ross, R. A., "A theory of the term structure of interest rate". Econometrica, 53, 2, 385-407, 1985.

Crimaldi, I. and Pratelli, L. "Convergence results for conditional expectations". Bernoulli, 11, 737–745, 2005.

Derman, E. “Regimes of volatility”. Risk, 4, 55-59, 1999.

Detlefsen, K. and Härdle, W. “Calibration risk for exotic options”. SFB Discussion Paper, 2006.

Doob, J. L. Stochastic process. John Willy & Sons, New York, 1953.

Duan, J. “The GARCH option pricing model”. Mathematical Finance, 5, 13-32, 1995.

—— “Cracking the smile”. Risk, 9, 55-59, 1996a.

—— “A unified theory of option pricing under stochastic volatility- from GARCH to diffusion”. research paper, 1996b.

—— “Augmented GARCH(p,q) process and its diffusion limit”. Journal of Econometrics, 79, 97-127, 1997.

Duan, J.C., Gauthier, G. and Simonato, J.G.. “An Analytical Approximation for the GARCH Option Pricing Model”. Journal of Computational Finance, 2, 75-116, 1999.

Duan, J.C., Ritchken, P. and Sun, Z. “Approximating GARCH-jump models, jump-diffusion processes, and option pricing”. Mathematical Finance, 16, 21-52, 2006.

Dumas, B., Fleming, J. and Whaley, R. “Implied volatility functions: empirical tests”, Journal of Finance, 53, 2059-2106, 1998.

Durham, G. B. and Gallant, A. R. “Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes”, Journal of Business & Economic Statistics, 20, 3, 297-316, 2002.

Elerian, O. “A note on the existence of a closed form conditional transition density for the Milstein scheme”. Working paper, 1998.

Elerian, O., Chib, S. and Shephard, N. “Likelihood inference for discretely observed non-linear diffusions”. Econometrica, 69, 959-993, 2001.

Engle, R. “Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation”. Econometrica, 50, 987-1008, 1982.

—— “A comment on Hendry and Clements on the limitations of comparing mean square forecast errors”. Journal of Forecasting, 12, 642-644, 1993.

Engle, R., and Ng, V. “Measuring and testing the impact of news on volatility”. Journal of Finance, 43, 1749-1778, 1993.

Engle, R. and Patton, A. “What good is a volatility model?” Quantitative Finance, 1, 237-245, 2001.

Eraker, B. “MCMC analysis of diffusion models with application to finance”. Journal Business & Economic Statistics, 19, 177-191, 2001.

Fiorentini, G., León, A. and Rubio, G. “Estimation and empirical performance of Heston's stochastic volatility model: the case of a thinly traded market”. Journal of Empirical Finance, 9, 2, , 225-255, 2002.

Fornari, F. and Mele, A. "Approximating volatility diffusions with CEV-ARCH models". Journal of Economic Dynamics and Control, 30, 6, 931-966, 2004.

Gallant, R. and Tauchen, G. E. "Which moment to match?" Econometric Theory, 12, 4, 657-681, 1996.

—— "Reprojecting partially observed systems with application to interest rate diffusions". Journal of the American Statistical Association, 93, 441, 10-24, 1998.

Genon-Catalot, V., Jeantheau, T. and Laredo, C. "Parameter estimation for discretely observed stochastic volatility models". Bernoulli, 5, 855-872, 1999.

Goggin, E. "Convergence in distribution of conditional expectations". Annals of Probability, 22,1097-1114, 1994.

Gourieroux, C. , Monfort, A. and Renault, E. "Indirect Inference". Journal of Applied Econometrics, 8S, S85-118, 1993.

Hasbrouck, J. "One security, many markets: determining the location of price discovery." Journal of Finance, 50, 1175-1199, 1995.

Hansen, L. P. "Large sample properties of generalized method of moments estimators". Econometrica, 50, 1029-1054, 1982.

Harvey, A. C. and Shepard, N. "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns". Journal of Business & Economic Statistics, 20, 2, 198-212, 1996.

Heston, S. "A closed-form solution for options with stochastic volatility with application to bond and currency options". Review of Financial Studies, 6, 327-343, 1993.

Heston, S., and S. Nandi, "A closed-form GARCH option pricing model". Review of Financial Studies, 13, 585-625, 2000.

Hull, J., and White, A. "The pricing of options with stochastic volatilities". Journal of Finance, 42, 281-300, 1987.

Kessler, M. "Estimation of an ergodic diffusion from discrete observations", Scandinavian Journal of Statistics, 24, 211-229, 1997.

—— "Simple and explicit estimating functions for a discretely observed diffusion process". Scandinavian Journal of Statistics, 27, 65-82, 2000.

Lo, A. "Maximum Likelihood Estimation of Generalized Ito Processes with Discretely-Sampled Data". Econometric Theory, 4, 231-247, 1988.

Merton, R. "Theory of rational option pricing". Bell Journal of Economics, 4, 141-183, 1973.

Nelson, D. B. "ARCH models diffusion approximation". J. Econometrics, 45, 7-38, 1990.

Nelson, D. B. and Foster, P. "Asymptotic Filtering Theory for ARCH Models". Econometrica, 62, 1-41, 1994.

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

Pedersen, A. R. "A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations”. Scandinavian Journal of Statistics, 22, 55-71, 1995.

Raftery, A. and Lewis, S. “How many iterations in the Gibbs sampler?” Bayesian Statistics, 4, 763-773, 1992.

Renault, E., “Econometric models of option pricing errors”. Advances in Economics and Econometrics, Seventh World Congress, Cambridge University Press, 223-278, 1997.

Rubinstein, M. “Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE options classes from August 23,1976 through August 31,1978”. Journal of Finance, 40,455-480, 1985.

Scott, L. “Option pricing when the variance changes randomly: Theory, estimators, and applications”. Journal of Financial and Quantitative Analysis, 22, 419-438, 1987.

Scott, L. “Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: Application of Fourier inversion methods”. Mathematical Finance, 7, 413-426, 1997.

Sørensen, H. “Simulated likelihood approximations for stochastic volatility models”. Scandinavian Journal of Statistic,30, 257-276, 2003.

Sørensen, M. “Prediction-based estimating functions”. *Econometric J.*, 3, 123-147, 1999.

Stephan, J. S. and Whaley, R. E. “Intraday price change and trading volume relations in the stock and stock option markets.” Journal of Finance, 45, 1, 191-220, 1990.

Stein, E., and J. Stein, “Stock price distributions with stochastic volatility”. Review of Financial Studies, 4, 727-752, 1991.

Wang, Y. “Asymptotic nonequivalence of GARCH models and diffusions”. Annals of Statistics, 30, 754-783, 2002.