

An Analysis of Nikkei 225 Call and Put Option Price Differences between Market Price and Theoretical Price

※Satoru KANO[†] and Asuka TAKEUCHI-Nogimori^{††}

Abstract

In this article, we investigate the Nikkei 225 options market and analyze the difference between the market price and the theoretical price. In the analyses, the following points are identified as characteristics of the option market. First, a multiple number of options with different maturities are traded on the same day. Second, in the Nikkei 225 option data set, option prices together with trading volumes frequently take a value of zero. In order to take these characteristics of the option data into consideration, the data set must be handled as panel data with a certain rotation structure. We propose a new approach to estimate the truncated model using simulation.

The estimation results show that the differences in the call options depend on the moneyness and the survival period, and the differences in the put options depend on the moneyness, the survival period, and the trading periods. It is also seen that the variance of these difference depends on the strike price, the transaction date and survival period. Five models, BS, GARCH, EGARCH, GJR, and APGARCH, were used to calculate the theoretical price. There were no notable differences in the signs of estimated coefficients and their relative sizes in the regression models of the difference between the market price and the theoretical price in the result of the put option. As for the call option result, a common estimate was obtained in the EGARCH, GJR, and APGARCH models that take the asymmetry into consideration.

Key words: options, panel data, truncation model, GARCH, EGARCH, GJR, APGARCH

日経225 コール・プットオプションの市場価格と 理論価格の乖離の実証分析

加納 悟・竹内(野木森)明香

要 旨

本稿は、オプション市場価格と理論価格の乖離（価格差）を、日経225 コール・オプションとプット・オプションのデータを用いて分析したものである。分析の中で、オプション市場の特徴として、次の2点を考慮している。1点目は、オプション市場では、権利行使価格と満期の条件が異なる複数のオプションが取引されていることである。2点目は、日経225オプションでは、オプション取引が成立せず取引量がゼロのときオプション価格が0円として記載される点である。これらの特徴を考慮すると、オプションデータは切断データとなり、ローテーション構造を含むパネルデータとなる。このような不均一分散をもつ切断パネル・データ・モデルの推定方法として、本稿ではシミュレーションを用いた推定方法を提案する。

推定結果から、コール・オプションの価格差はマネネスと残存期間に、プット・オプションの価格差はマネネス、残存期間、取引日に依存していた。価格差の分散は、権利行使価格、取引日、残存期間に依存している。理論価格の算出にBS, GARCH, EGARCH, GJR, APGARCHモデルの5種類を使用した。異なるモデルから計算された理論価格を用いても、プット・オプションの価格差の分析結果は変わらなかった。コール・オプションの結果では、ボラティリティの非対称性を考慮したEGARCH, GJR, APGARCHモデルから算出される理論価格で、共通した推定結果が得られている。

キーワード：オプション、パネルデータ、切断モデル、GARCH, EGARCH, GJR, APGARCH

※本論文は、竹内が一橋大学大学院経済研究科博士課程在学時に、故加納悟教授と行った共同研究の結果である。ご指導いただいた加納悟先生に深く感謝の意を示したい。ここに謹んで哀悼の意を表し、心からご冥福をお祈り申し上げます。

投稿受付日 2010年9月30日
採択決定日 2011年2月23日

†一橋大学経済研究所教授
††一橋大学経済研究所非常勤研究員

1. Introduction

Stock index options made their debut in Japan in June 1989, when they were first introduced at the Osaka Securities Exchange. Since then, trading rules have been amended frequently, giving the market the structure it has today. Among the most actively traded options in Japan are the Nikkei 225 options. In this article, we investigate the Nikkei 225 option market and analyze the difference between the market price and the theoretical price.

The earliest attempt to model option pricing was undertaken by Black and Scholes (1973). Since then, various models have been proposed to improve the original BS (Black and Scholes) model by extending it. The principal objective of the proposed models has been to describe the movement of the underlying asset price more accurately. As option prices are primarily determined by the underlying asset prices and as the volatility of the underlying asset price is an important factor, much effort has been expended on describing the volatility movement of the underlying asset precisely. For example, the variance changing models such as ARCH (autoregressive conditional heterogeneity) type and SV (stochastic volatility) models have been applied and a comparison made of how well they can approximate the underlying assets volatility. If there is no arbitrage, the value of an option is calculated from above the underlying stock price itself.

However, some empirical studies show that the option price is not simply determined by the movements of the underlying asset. That is, the prices observed in the market may differ from the prices calculated from the above volatility models under the no-arbitrage condition. Analyzing the rate of return of options on the Nikkei 225 index, Nishina and Nabil (1997), for example, show that the put-call parity does not hold. Bookstaber (1981) suggests that a bias exists due to non-simultaneity of the option market and the underlying asset market. Easley, O'Hara and Srinivas (1998) point out that the trading volume of a stock option correlates with the future price of the underlying assets. Thus, it can be concluded that option prices are affected by factors different from the underlying asset. The option theoretical prices calculated from the above volatility models by the assumption of no arbitrage do not take these facts into account.

If option prices are not determined only by the underlying asset values, there would be a difference between the theoretical price under assumption of no arbitrage and the market price. Renault (1997) indicates the difference between the theoretical price and the market price and raises the following four facts as the reasons. The first is the difference that occurs by assuming that the estimated value of volatility is a true value. Second, an error in the for-

malization of underlying asset process is mentioned. The third factor is the hypothesis that the theoretical price is the expected value of an option payoff at the expiration day. The fourth factor is an error not covered by any of the reasons listed above. In this paper, the four reasons for the option price difference indicated in Renault (1997) are put into two categories. We will consider the first three as model error and the last one as error without consideration in the no-arbitrage condition. As mentioned before, the model error has frequently been considered in the formula for calculating the theoretical price, but the error without consideration in the no-arbitrage condition has not been analyzed in depth in option empirical studies.

The main objective of this paper is to examine whether the error without consideration in the no-arbitrage condition exists in the option market and investigate empirically what characteristics are seen in this error by analyzing the difference between the market price and the theoretical price. In the analyses, the following points are identified as characteristics of the option market.

First, a multiple number of options with different maturities are traded on the same day in the market. It is necessary to classify the option price data not only by the traded days but also by the maturity days. As a result, the data set must be handled as panel data with a certain rotation structure. Accordingly, it becomes necessary to take the covariance of the price data into consideration when the price movements are explained by various factors such as the survival period and the moneyness.

Second, in the Nikkei 225 option data set, option prices together with trading volumes frequently take a value of zero. This indicates that for some reasons the options are not traded on that day and such observations have been excluded from the empirical analysis in the literature. However, as such observations where the price is zero reflect the investors' decision, exclusion of them will lead to loss of information. More importantly, the estimation of any relationship may be biased due to this exclusion.

The remainder of this paper is organized as follows. Section 2 provides a brief overview of the characteristics of the Nikkei 225 options market and briefly explains how to calculate the theoretical option prices. In Section 3, the difference between the market price and the theoretical price in the Nikkei 225 option market is investigated empirically taking the abovementioned characteristics of the data into account. Section 4 provides the estimation results. Section 5 concludes the paper.

2. Illustration of the Nikkei 225 option data

2.1 A brief outline of the Nikkei 225 options market

First of all, let us briefly look at the features of the Nikkei 225 options market. Eight types of options with different expiration dates are tradeded on the same day. There are three kinds of transaction time horizons: 15 months, 5 months, and 4 months. When the exercise day of an option arrives, a new option is created on that day. Usually, for options with the same maturity, five strike prices are initially set symmetrically around the underlying asset price. Therefore, there are at least 40 options running on every trading day. However, the number of strike prices may increase when the underlying asset price exceeds the highest or lowest strike price. In the following sections, in order to simplify the analysis, monthly data are created from the daily data by picking up the prices on the expiration day in each month. Further, the data of each option are extracted for four months before the expiration date. For convenience, the four months during the survival period are labeled as $\tau = 1, 2, 3, 4$. As a result, time series data with a rotation structure are created. In this paper, we analyze the market prices of the Nikkei 225 call option C_m and put options P_m with maturity from January 2000 to April 2002.

2.2 Estimation of the theoretical price

The theoretical prices corresponding to the market prices C_m and P_m do not exist as data. The theoretical option prices (C_m and P_m) below are calculated by using the BS, GARCH (generalized ARCH), EGARCH (exponential GARCH), GJR, and APGARCH (asymmetric power GARCH) models. In the following, the strike price of an option is written as K , the daily closing price of the underlying asset as S , the survival period as τ , and moneyness ($= K/S$) as M . Further, the number of strike prices for options in the i -th group maturity is denoted as K_i .

A number of models have been proposed to describe the volatility fluctuation of the underlying assets accurately. The GARCH model by Bollerslev (1986) considers the persistence of volatility shock. The EGARCH model by Nelson (1991), GJR model by Glosten, Jagannathan and Runkle (1993), APGARCH model by Ding et al. (1993) consider the asymmetry as well as the persistency of the shock. Comparisons of option pricing have been made among the volatility fluctuation models. For example, Crouhy (1994), Duan and Zhang (2001) made comparison between the ARCH type model and the BS model. They conclude that the predictive performance of the option price by the ARCH type model is better than that of the BS model.

More concretely, the GARCH model

$$\begin{aligned}
 R_{s,u} &= \mu + \epsilon_u, \\
 \epsilon_u &= \sigma_u z_u, \quad z_u \sim i.i.d.N(0,1), \\
 \sigma_u^2 &= \omega + \alpha \epsilon_{u-1}^2 + \beta \sigma_{u-1}^2
 \end{aligned} \tag{1}$$

is estimated by maximum likelihood using the daily return rate of the Nikkei 225 ($R_{s,u}$), for 1,000 business days before the trading day. For the EGARCH model, the variance fluctuation is expressed as;

$$\ln(\sigma_u^2) = \omega + \alpha[\theta z_{u-1} + \gamma(|z_{u-1}| - E(|z_{u-1}|))] + \beta \ln(\sigma_{u-1}^2). \tag{2}$$

For the GJR model, the variance fluctuation is expressed as;

$$\sigma_u^2 = \omega + \alpha \epsilon_{u-1}^2 + \gamma D_{t-1} \epsilon_{u-1}^2 + \beta \sigma_{u-1}^2. \tag{3}$$

Now, let D_{t-1} of GJR model a dummy variable that becomes 1 when $\epsilon_{u-1} < 0$ and 0 otherwise. And for the APGARCH model, the variance fluctuation is expressed as;

$$\sigma_u^\delta = \omega + \alpha(|\epsilon_{u-1}| - \gamma \epsilon_{u-1})^\delta + \beta \sigma_{u-1}^\delta. \tag{4}$$

Then, let $-1 < \gamma < 1$. Further, for the constant volatility model, the volatility is assumed to be: $\sigma_u^2 = \omega$ in (1), and this model is regarded as the BS model in this paper.

As a result of estimation, the parameters that expresses the asymmetry of EGARCH, GJR, and APGARCH models were estimated to be significant at 5%. The persistence of the volatility shock and the asymmetry are existent in the data used in this paper.

After the estimated ARCH type models are transformed to the models where local risk neutrality is assumed (Duan, 1995; Bauwens & Lubrano, 2002), the underlying assets prices at the maturity of the option are generated by the Monte Carlo simulation. Using the simulated Nikkei 225 prices at maturity, we can simulate the payoff of the call and put options, too. When the average of these payoffs is discounted at the transaction day, it becomes the theoretical option prices, C_m and P_m , from these volatility models.

3. Analysis of the difference between the market price and the theoretical price

3.1 Model

In this study, we consider the difference between the market prices C_m and P_m and the theoretical prices C_s and P_s . As mentioned above, the theoretical prices are calculated from the volatility models under the assumption of no arbitrage, and the market prices are affected by factors both with and without consideration in the no-arbitrage condition.

Multiple options exist in a single trading day in the option market. The investors must select from among these options. Each of the options can be categorized according to strike prices, survival period, and the expiration day. In trading, investors have information on the theoretical price, which they usually trust, the strike price, the survival periods, and the underlying asset's price. They can make their decisions based on this information. If an investor decides which option to trade by considering the relationship of it with the other options, that investor selects the most undervalued or overvalued options by comparing the market price with the theoretical price. Therefore, the market price C_m and P_m will be affected not only by the theoretical price C_s and P_s calculated from the underlying asset movements but also by the other option prices. When such as investor's decision behavior is considered, the market price would include the part which depends on the survival period and strike prices.

In addition, since the number of strike prices in one expiration day increase when fluctuation of the underlying asset is large, options traded during a long period and options traded during a short period show a difference in the number of their strike prices. Since a largest number of strike prices will be set during a 15-month trading period, it is possible to select an option from among a greater number of options with the same survival period. It can be seen from the above that the lengths of trading periods (15 months, 5 months, and 4 months) are also likely to affect the market price.

In this research, we consider a simple model in which the market price depends on the moneyness, the survival period, and the trading period in addition to the theoretical price: for the call options,

$$C_{m, ikt\tau} = C_{s, ikt\tau} + \beta_M M_{ikt\tau} + \sum_{s=1}^4 \beta_{\tau, s} D_{\tau, s} + \sum_{j=4}^5 \beta_{op, j} D_{op, j} + u_{ikt\tau} \quad (5)$$

and for the put options,

$$P_{m, ikt\tau} = P_{s, ikt\tau} + \beta_M M_{ikt\tau} + \sum_{s=1}^4 \beta_{\tau, s} D_{\tau, s} + \sum_{j=4}^5 \beta_{op, j} D_{op, j} + u_{ikt\tau}. \quad (6)$$

Here, $D_{\tau, s}$ ($s = 1, 2, 3, 4$) is the dummy variable that takes one when the survival period is s months and $D_{op, j}$ ($j = 4, 5$) is the dummy variable that takes one when the option trading period is j months. The suffix i denotes a maturity ($i = 1, \dots, 28$), t is a trading day, k ($k = 1, \dots, k_i$) is a strike price, and τ is a survival period, which $\tau = t - 4(i - 1)$. The error term $u_{ikt\tau}$ is assumed to be decomposed into the error ε_t that depends on the transaction date, the error ε_k that depends on the strike price, the error ε_τ that depends on the survival period within the same maturity group, and the error term $\varepsilon_{ikt\tau}$ that is purely random: $u_{ikt\tau} = \varepsilon_t + \varepsilon_k + \varepsilon_\tau + \varepsilon_{ikt\tau}$. It is assumed that ε_t , ε_k , ε_τ and $\varepsilon_{ikt\tau}$ are mutually independent. The variances of these four error terms are

Σ_s^i ($s = 2, 3, 4$) are the variance matrices of the options on the same trading day with different maturities. The structure is expressed as follows.

$$\Sigma_2^i = \begin{pmatrix} 0 & \Lambda_t & 0 & 0 \\ 0 & 0 & \Lambda_t & 0 \\ 0 & 0 & 0 & \Lambda_t \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \Sigma_3^i = \begin{pmatrix} 0 & 0 & \Lambda_t & 0 \\ 0 & 0 & 0 & \Lambda_t \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \Sigma_4^i = \begin{pmatrix} 0 & 0 & 0 & \Lambda_t \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (11)$$

Another characteristic that can be observed is that the variance becomes smaller as moneyness becomes out-of-the money. Here, for call option, the adjustment for this heteroskedasticity is made simply by multiplying the moneyness M by both sides of (5). Also, an adjustment is done by dividing both sides of (6) in the put option.

3.2 Treatment of 0-value data

In the Nikkei 225 option market, 0-value is recorded as a price when no trading is established. In conventional research, such data are excluded from the analysis. It is noted, however, that the investors' decisions are somehow reflected in these data and, therefore, exclusion of them will lead to loss of information. More importantly, exclusion may cause bias in the estimation. The difficulty is that the reasons why market prices become 0 are unknown. Also, the trading volume, which reflects the investors' demand, becomes 0 simultaneously with the price. A method of considering 0-value data in the model is explained below. Among the 968 observations of call option prices that are used in this paper, 74 observations are with $C_m = 0$. 0-value data had existed in the data before 2000 in the put option. However, trades were established in all of the data of 968 observations used in this paper. Because of this fact, a manipulation of 0-value data explained below has not been done for the put option.

In this paper, the following two possibilities of C_m being 0 are considered. First, an investor's value of option that the investor thinks reasonable after considering the three factors, such as the theoretical price which the investor trust, the strike price, and the survival period, can be negative when the investor requires a risk premium and C_s is small. Since the market price cannot be negative, the option price is truncated at 0 yen under such a circumstance. That is, a zero market price of a call ($C_m = 0$) is observed when the value of the option, C_m^* , that all investors consider appropriate satisfies:

$$i) C_{m,itr}^* < 0. \quad (12)$$

Moreover, since zero data $C_m = 0$ are observed in all of the moneyness, it is seen that C_m becomes zero even when C_s is large. These data can not be explained by the truncation (12).

Therefore, the truncation (12) is not the sole reason for the data with $C_m = 0$. When the transaction vanishes, either the investors' demand or supply becomes nil. The results in Parlour (1998) and Anshuman and Kalay (1998) are considered one of the possible reasons for this situation. They analyzed the market behavior with discrete pricing restrictions like the Nikkei 225 option market, and found that the spread between the bid/ask price and the preset limited price causes no trading. That is, we define the deviation as:

$$\begin{aligned} DV &= C_{m,itkr}^* - \mathbb{E} \left(C_{s,itkr} + \beta_M M_{itkr} + \sum_{s=1}^4 \beta_{\tau,s} D_{\tau,s} + \sum_{j=4}^5 \beta_{op,j} D_{op,s} + u_{itkr} \right) \\ &= C_{m,itkr}^* - \left(C_{s,itkr} + \beta_M M_{itkr} + \sum_{s=1}^4 \beta_{\tau,s} D_{\tau,s} + \sum_{j=4}^5 \beta_{op,j} D_{op,s} \right). \end{aligned} \quad (13)$$

It is plausible that C_m could be 0 when:

$$ii) \quad DV > a \quad \text{or} \quad DV < b (< 0). \quad (14)$$

where a is a positive constant and b is a negative constant.

In summary, the model and the truncation mechanism are expressed as follows.

$$\begin{aligned} M_{itkr} C_{m,itkr}^* &= M_{itkr} \left(C_{s,itkr} + \beta_M M_{itkr} + \sum_{s=1}^4 \beta_{\tau,s} D_{\tau,s} + \sum_{j=4}^5 \beta_{op,j} D_{op,s} + u_{itkr} \right) \\ u_{itkr} &= \epsilon_l + \epsilon_k + \epsilon_\tau + \epsilon_{itkr} \\ C_{m,itkr} &= \begin{cases} C_{m,itkr}^* \\ 0 & \text{if } i) \text{ or } ii). \end{cases} \end{aligned} \quad (15)$$

For these truncated data, OLS using the entire data or OLS using the subsample for which $C_m < 0$ are both inconsistent estimators of the coefficients in the model, so that we should consider these truncated conditions to estimate the parameters (Wooldridge, 2002). If the data was cross-section data and the variance structure was homoscedastic and the truncation mechanisms were expressed only by (12), the familiar Tobit model could be applied and we can apply the maximum likelihood estimator. In the case of the panel data including individual effects, we have to maximize the log-likelihood with respect to each individual effect's parameter. However, in short panels, this estimator is inconsistent. On the other hand, for the random effects models including individual effects, the likelihood function included one-dimensional integral (Cameron and Trivedi, 2005). Note that as this model has a complicated variance covariance structure as well as truncation conditions, therefore, there are multi-dimensional integrations in the likelihood function and the maximum likelihood estimations of the parameters of this model cannot be analytically obtained (see Appendix)². Because of the complicated structure of the

variance matrix and truncation mechanism, however, we are obliged to resort to a simulation method. In this estimated procedure, we assume the conditional distribution of the unobserved data and simulated this data, after we estimated the model.

Let C_{m1} be a vector of the option prices in the market for which $C_m \neq 0$ and C_{m0}^* be a vector of the latent investor's value corresponding to C_{m0} for which $C_m = 0$. Note here that C_{m0}^* is unobservable. Assume also that (C_{m1}, C_{m0}^*) distributes according to a multivariate normal distribution with the mean vector μ and the variance matrix Σ , respectively:

$$\mu = E \begin{pmatrix} C_{m1} \\ C_{m0}^* \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_0 \end{pmatrix}, \quad \Sigma = \text{Var} \begin{pmatrix} C_{m1} \\ C_{m0}^* \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \text{where } \Sigma_{12} = \Sigma_{21}' . \quad (16)$$

Then, the distribution of C_{m0}^* conditional on C_{m1} is also a multivariate normal distribution whose density is:

$$f(C_{m0}^* | C_{m1}) = (2\pi)^{-\frac{k_0}{2}} \left(\frac{|\Sigma_{11}|}{|\Sigma|} \right)^{\frac{1}{2}} \exp \left(-\frac{Q_0}{2} \right), \quad (17)$$

where

$$Q_0 = (C_{m0}^* - v_0)' (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12})^{-1} (C_{m0}^* - v_0)$$

and

$$v_0 = \mu_0 + \Sigma_{21} \Sigma_{11}^{-1} (C_{m1} - \mu_1).$$

We generate C_{m0}^* using this property.

Concretely, the following steps are iterated.

1. Estimate the model by the OLS using the observations for which $C_m \neq 0$.
2. Estimate $\sigma_i^2, \sigma_k^2, \sigma_\tau^2, \sigma_{itkr}^2$ from the OLS residuals \hat{u}^1 using by OLS regressions.
3. The parameters in (17) are replaced by their estimates as:

$$\hat{\mu}_i^1 = C_s + \hat{\beta}_M^1 + M_{itkr} + \sum_{s=1}^4 \hat{\beta}_{\tau,s}^1 D_{\tau,s} + \sum_{j=4}^5 \hat{\beta}_{op,j}^1 D_{op,s} \quad (i = 0, 1), \quad \hat{v}_0^1 = \hat{\mu}_0 + \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} (C_{m1} - \hat{\mu}_1).$$

Then, generate random variables η from the standard normal distribution, and set: $\hat{C}_{m0} = A\eta_{itkr} + \hat{v}_0^1$, where A is a matrix that satisfies: $\hat{\Sigma}^{-1} = A'A$.

4. The truncation points, a and b , in the second type truncation ii) are set to the minimum value of OLS residuals \hat{u}^1 for the lower bound a and the maximum value of OLS residuals \hat{u}^1 for the upper bound b (see Appendix). As a result, when \hat{C}_{m0} satisfies one of the truncation conditions, we set $\hat{C}_{m0}^{*1} = \hat{C}_{m0}^*$.

5. After combining the generated data \hat{C}_{m0}^{*1} with C_{m1} , conduct GLS.
6. Using the residual from GLS estimation, the variance matrix as well as other parameters (μ_i , ν_θ) are reestimated and set to $\hat{\Sigma}^2$, $\hat{\mu}_i^2$, $\hat{\nu}_\theta^2$. These estimated parameters are used to generate \hat{C}_{m0}^{*2} . The obtained parameter estimates after convergences are regarded as the final estimates.

4. Estimation results

In this section, we will first report the estimation results of call options and put options, and then discuss whether the error without consideration in the no-arbitrage condition exists.

Models were estimated by using five theoretical prices that are calculated from BS, GARCH, EGARCH, GJR, and APGARCH models.

The results of FGLS for call options after convergence are shown in Table 1. The estimated model is:

$$M_{itkr}(C_{m,itkr} - C_{s,itkr}) = M_{itkr} \left(\beta_M M_{itkr} + \sum_{s=1}^4 \beta_{\tau,s} D_{\tau,s} + \sum_{j=4}^5 \beta_{op,j} D_{op,s} + u_{itkr} \right) \quad (18)$$

The significance level is 5%.

From the table, the significance and the sign of the coefficients are different between volatility models. However, in the EGARCH, GJR, and APGARCH models that take the asymmetry into consideration, the sign and size relationship of the coefficient are similar. First, the estimate value of β_M is estimated as negative in BS and GARCH models, but is positive in EGARCH, GJR, and APGARCH models. That is, when using the theoretical prices considering the asymmetry of the volatility, $C_m - C_s$ is large as K/S becomes out-of-the-money. Next, the estimated values of $\beta_{\tau,s}$ ($s = 1, 2, 3, 4$) increase as the survival period shortens. This is a common result in all of the volatility models. And $\beta_{op,4}$ is significantly positive in BS and GARCH model, however, it is not estimated to be significant in the other volatility models.

Estimates of σ_i^2 , σ_k^2 , σ_τ^2 are significant in all volatility models. The estimates of σ_i^2 are relatively bigger than σ_k^2 and σ_τ^2 . That is, the size of the variance component depending on the transaction date is large. From these estimates, it is confirmed that heteroskedasticity exists and depends on the transaction date as well as the strike price.

The result of the put option is summarized in Table 2. The estimated model is,

$$\frac{1}{M_{itkr}}(P_{m,itkr} - P_{s,itkr}) = \frac{1}{M_{itkr}} \left(\beta_M M_{itkr} + \sum_{s=1}^4 \beta_{\tau,s} D_{\tau,s} + \sum_{j=4}^5 \beta_{op,j} D_{op,s} + u_{itkr} \right).$$

Table 1: CALL: Estimation results in FGLS

	BS	GARCH	EGARCH	GJR	APGARCH
β_M	-110.96* (29.48)	-65.92* (32.82)	32.32 (32.23)	119.03* (27.56)	70.86* (31.61)
$\beta_{\tau,1}$	38.19 (45.72)	-63.64 (47.17)	-166.74* (47.26)	-289.83* (37.26)	-254.19* (41.93)
$\beta_{\tau,2}$	74.50 (46.23)	-14.84 (47.78)	-116.70* (47.85)	-231.60* (37.78)	-193.22* (42.55)
$\beta_{\tau,3}$	77.75* (46.75)	1.10 (48.38)	-95.64* (48.44)	-181.82* (38.33)	-141.77* (43.17)
$\beta_{\tau,4}$	95.06* (47.18)	31.25 (48.92)	-64.53 (48.95)	-157.91* (38.78)	-109.19* (43.72)
$\beta_{op,4}$	16.43 (18.74)	17.38 (20.03)	14.35 (20.64)	10.23 (19.30)	10.65 (20.93)
$\beta_{op,5}$	34.98* (19.45)	34.85* (20.85)	34.94 (21.44)	28.14 (20.02)	26.49 (21.77)
σ_t^2	24705.60* (483.25)	20975.90* (387.61)	21585.50* (416.96)	6620.37* (179.93)	8764.08* (194.38)
σ_k^2	1753.29 (1430.83)	3894.85* (1147.65)	3272.65* (1234.55)	3199.99* (532.74)	5620.95* (575.54)
σ_τ^2	3730.64* (961.82)	4079.20* (771.47)	4608.56* (829.88)	4491.19* (358.12)	4888.26* (386.89)
σ_{itk}^2	12688.89* (2936.69)	9166.77* (2355.49)	10585.38* (2533.85)	4637.06* (1093.42)	1663.94 (1181.26)
a	967.50	902.01	915.47	883.60	867.61
b	-607.56	-552.29	-585.28	-723.13	-741.11
R^2	0.013	0.020	0.034	0.092	0.077

The estimation was conducted using 968 observations from January 2000 to April 2000. The numerical value in parentheses shows the standard deviation. In all models, the number of iteration until convergence was 4. “*” indicates significant at 5%.

The significance level was 5%. In the put option, similarities in the significance of β and in the relationship of the sign are seen in BS, GARCH, EGARCH, GJR, and APGARCH models. First, the coefficient of moneyness, β_M , was estimated significantly positive. Therefore, the positive bias of the price difference is large as it becomes in-the-money. Next, $\beta_{\tau,s}$ ($s = 1, 2, 3, 4$) is made significantly negative. The value of $\beta_{\tau,s}$ ($s = 1, 2, 3, 4$) decreases as the survival period becomes shorter, and the negative bias becomes larger as it approaches the expiration day. Next, the

Table 2: PUT: Estimation results in GLS

	BS	GARCH	EGARCH	GJR	APGARCH
β_M	143.72* (26.56)	184.57* (24.05)	267.99* (26.97)	253.55* (25.75)	218.18* (27.14)
$\beta_{\tau,1}$	16.68 (35.26)	-8.93 (32.58)	-90.62* (35.69)	-65.34* (34.49)	-44.68 (36.03)
$\beta_{\tau,2}$	-33.92 (35.77)	-62.37* (33.03)	-140.44* (36.21)	-116.81* (34.98)	-95.99* (36.55)
$\beta_{\tau,3}$	-64.27* (36.28)	-89.98* (33.49)	-179.41* (36.74)	-156.46* (35.48)	-126.82* (37.07)
$\beta_{\tau,4}$	-124.41* (36.77)	-157.12* (33.90)	-246.50* (37.21)	-226.72* (35.92)	-195.52* (37.56)
$\beta_{op,4}$	43.88* (13.82)	39.86* (13.89)	47.13* (15.40)	42.41* (14.78)	44.73* (14.87)
$\beta_{op,5}$	20.71 (14.52)	18.20 (14.50)	22.44 (16.09)	22.42 (15.43)	24.12 (15.57)
σ_t^2	8978.74* (212.60)	7368.01* (188.33)	7672.71* (199.29)	7748.37* (196.97)	8694.69* (209.79)
σ_k^2	2002.29* (646.04)	1034.54* (572.30)	1688.74* (605.61)	1364.36* (598.54)	1961.69* (637.51)
σ_τ^2	1448.75* (416.38)	1807.58* (368.86)	2188.01* (390.33)	2027.44* (385.77)	1889.31* (410.88)
σ_{itk}^2	8287.55* (1340.77)	8985.45* (1187.73)	9835.70* (1256.86)	9627.29* (1242.19)	9033.34* (1323.06)
R^2	0.167	0.197	0.215	0.216	0.193

The estimation was conducted using 968 observations from January 2000 to April 2000. The numerical value in parentheses shows the standard deviation. In all models, the number of iteration until convergence was 4. "*" indicates significant at 5%.

coefficient of the trading period, $\beta_{op,j}$ ($j = 4, 5$), was estimated significantly positive. Positive bias is larger for a shorter trading period. Finally, σ_t^2 , σ_k^2 , σ_τ^2 were assumed significant in positive values. The heteroskedasticity of the variance exists.

Let us compare the results for both type of options. First, the relationship between money-ness and the price differences is considered. In the put option and in the call option that takes the asymmetry into consideration, the coefficients, β_M , are estimated significantly positive. The result indicates that the option price differences increase as it becomes in-the-money for the put option and out-of-the-money for the call option. In other words, the market price is set to

be higher than the theoretical price as the underlying asset becomes lower than the strike prices. From these results, it can be supposed that investors demand risk premium due to the fall of the underlying asset. Next, the relationship between survival period and price differences is examined. $\beta_{t,s}$ ($s = 1, 2, 3, 4$) was estimated significantly negative both in the call option and put option, but the relationship of the coefficient became opposite. The option price differences decrease as the survival period became shorter in the call option, but conversely increase in the put option. Finally, the significances of variance are compared. σ_r^2 , σ_h^2 , σ_r^2 are estimated significant. Heteroskedasticity exists in both of the options. Thus, it can be found that the option prices differences are correlated to the past value of the option itself and to the other option prices that are traded at the same time.

From these estimation results of call and put options, we will discuss the model error and the error without consideration in the no-arbitrage condition. We will explain the effects of the model error for the option price difference, after that we discuss about the estimation results of call and put options.

If a model error exists in the market, the price difference should vary depending on the volatility model for the following reasons. The five models, BS, GARCH, EGARCH, GJR, and APGARCH models are used to calculate the theoretical price. The above-mentioned five volatility models can be compared as follows. BS model interprets volatility as a constant. In contrast, GARCH, EGARCH, GJR, and APGARCH models are enhanced models in which volatility varies with time. Additionally, among the three models, EGARCH, GJR, and APGARCH take the asymmetry of volatility into consideration. In this data, since the asymmetry parameters in EGARCH, GJR, and APGARCH models are estimated significantly, the most accurate models of volatility are EGARCH, GJR, and APGARCH. If the market price can mostly be predicted from a volatility process of the underlying asset, the option price prediction made by EGARCH, GJR, and APGARCH models should be more reliable than that made by BS and GARCH models. Moreover, EGARCH, GJR, and APGARCH models that take the asymmetry into consideration are formulated differently. That is, if a model error exists, the estimated results change with BS, GARCH, EGARCH, GJR, and APGARCH models.

On the other hand, an error without consideration in the no-arbitrage condition should exist independently of the volatility model. In the case of the put option price difference models, there were no notable differences in the signs of estimated coefficients and their relative sizes. As for the call options, however, a common estimate is obtained only in the EGARCH, GJR, and APGARCH models that take the asymmetry into consideration. As similar consequences were obtained from the these typical models, it is expected that the estimated

relationship of the price difference with the moneyness, the survival period, the trading period, and the striking price would not change drastically depending on these typical volatility models. Thus, the common characteristics of $C_m - C_s$ and $P_m - P_s$ obtained in multiple volatility models. This result can be considered that the error without consideration in the no arbitrage condition exists.

5. Conclusions

In this article, we investigated the Nikkei 225 option market and analyzed the difference between the market option price and the theoretical option price. The estimation results showed that the option price differences depend on the moneyness and the survival period, and also on the trading period. Moreover, the variance of these differences depends on the strike price, the transaction date and the survival period. As the survival period became longer, a negative bias of the difference increased in call option and decreased in put option. As a result, the call option is consistent with Long and Officer (1997). The five models, BS, GARCH, EGARCH, GJR, and APGARCH models are used to calculate the theoretical price, and we estimated the model of $C_m - C_s$ and $P_m - P_s$ by using five theoretical prices. From the estimation results, the common characteristics obtained among the call option price differences $C_m - C_s$ using the C_s calculated from the EGARCH, GJR, and APGARCH models and the put option price differences $P_m - P_s$ using P_s calculated from BS, GARCH, EGARCH, GJR, and APGARCH models. Thus, since there are common characteristics of the option price difference obtained in multiple volatility models, we can be considered that error without consideration in the no-arbitrage condition exists.

Some future research topics are as follows. First, we suggested a new estimation method, but did not consider the asymptotic theory of the estimator. It is important to test the covariance structure and model specification. Second, it is worthwhile interpreting this estimated results in context of the informational inefficiency (Fama, 1970), and the truncation conditions, for example the term of a and b in this condition, in context of the market microstructure theory.

A Estimation of truncation point

To estimate the truncation point in the Tobit model, let us consider the likelihood function. We define:

$$y_i = \begin{cases} y_i^* \\ 0 & \text{if } y_i^* - \mu_i < b \end{cases} \quad (19)$$

where $y_i^* \sim N(\mu_i, \sigma^2)$, y_i^* is the latent variable and b is the truncation point. The log-likelihood function for a constant dispersion is:

$$\ln L = \sum_{0 < y_i} -\frac{1}{2} \left(\ln(2\pi) + \ln \sigma^2 + \frac{(y_i - \mu)^2}{\sigma^2} \right) + \sum_{y_i=0} \ln \left(\int_{-\infty}^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y_i - \mu)^2}{2\sigma^2} \right\} dy_i \right). \quad (20)$$

By differentiating the log-likelihood function with respect to b , it becomes:

$$\frac{\partial \ln L}{\partial b} = \sum_{y_i=0} \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(b - \mu_i)^2}{2\sigma^2} \right\}}{\int_{-\infty}^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y_i - \mu_i)^2}{2\sigma^2} \right\} dy_i} > 0. \quad (21)$$

It can be seen that the log-likelihood function is monotonically increasing with respect to b . In addition, since the data whose values are smaller than b do not exist, the minimum value of the untruncated data becomes the estimated value of b . That is: $\hat{b} = \min(y_i - \mu_i \mid y_i \text{ is observed})$.

Next, consider the heteroscedasticity of y_i and;

$$y_i = \begin{cases} y_i^* \\ 0 & \text{if } \frac{y_i^* - \mu_i}{\sigma_i} < b, \text{ where } y_i^* \sim N(\mu_i, \sigma_i). \end{cases} \quad (22)$$

Truncation point b is defined for a normalized y_i^* . For simplification the data are converted as;

$$\tilde{y}_i = \begin{cases} \tilde{y}_i^* \\ 0 & \text{if } \tilde{y}_i^* < b, \end{cases} \quad \text{where } \tilde{y}_i^* = \frac{y_i^* - \mu}{\sigma_i} \sim N(0, 1). \quad (23)$$

The Jacobean of this transformation is; $d\tilde{y}_i/dy_i = 1/\sigma_i$, and the log-likelihood function for the transformed data can be shown as

$$\ln L = \sum_{0 < y_i} -\frac{1}{2} \left[\ln(2\pi) + (\tilde{y}_i)^2 \right] + \sum_{y_i=0} \ln \left(\int_{-\infty}^b \sqrt{\frac{1}{2\pi}} \exp \left[-(\tilde{y}_i)^2 \right] dy_i^* \right). \quad (24)$$

This log-likelihood function is monotonically increasing with respect to b , and the value of b is estimated as: $\hat{b} = \min(\tilde{y}_i \mid y_i \text{ is observed})$.

In a general case where the correlation among y_i exists, and y_i is treated as in this paper, we must consider:

$$Y \sim N(\mu, \Sigma), \quad y^i = \begin{cases} y_i^* \\ 0 \end{cases} \quad \text{if } \frac{y_i^* - \mu_i}{\sigma_i} < b \quad (25)$$

where $Y = \{y_1, y_2, \dots, y_n\}$. In this case, the log-likelihood function cannot be expressed as simply as in equation (24). When the correlation exists, the log-likelihood function is expressed by: $\ln L = \Lambda_1 + \Lambda_2$,

$$\begin{aligned} \Lambda_1 &= \sum_{y_i > 0} \left(-\frac{n}{2} \ln(2\pi) - \frac{1}{2} |\Sigma| - \frac{1}{2} (Y - \mu)' \Sigma^{-1} (Y - \mu) \right) \\ \Lambda_2 &= \sum_{y_i = 0} \ln \int_{-\infty}^{b_i} \dots \int_{-\infty}^{b_i} (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (Y - \mu)' \Sigma^{-1} (Y - \mu) \right\} dy_1 \dots dy_n \end{aligned} \quad (26)$$

where $b_i = b\sigma_i + \mu_i$. It is difficult to maximize such a log-likelihood function directly with respect to the parameters because the integration of the second term is complicated. We have transformed y_i in the same way as above: $\tilde{Y}^* = \tilde{\Sigma}^{-1/2} (Y^* - \mu) \sim N(0, R)$, where $R = \tilde{\Sigma}^{-1/2} \Sigma \tilde{\Sigma}^{-1/2}$,

$$\tilde{\Sigma} = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix} \quad \tilde{y}_i = \begin{cases} y_i^* \\ 0 \end{cases} \quad \text{if } \tilde{y}_i^* < b, \quad (27)$$

The corresponding Jacobean is: $d\tilde{Y}/dY = \tilde{\Sigma}^{-1/2}$. The log-likelihood function of the transformed variable \tilde{Y} is: $\ln L = \Lambda_3 + \Lambda_4$,

$$\begin{aligned} \Lambda_3 &= \sum_{\tilde{y}_i > 0} |\tilde{\Sigma}|^{-\frac{1}{2}} \left(-\frac{n}{2} \ln(2\pi) - \frac{1}{2} |R| - \frac{1}{2} \tilde{Y}' R^{-1} \tilde{Y} \right) \\ \Lambda_4 &= \sum_{\tilde{y}_i = 0} \ln \int_{-\infty}^b \dots \int_{-\infty}^b |\tilde{\Sigma}|^{-\frac{1}{2}} (2\pi)^{-\frac{n}{2}} |R|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \tilde{Y}' R^{-1} \tilde{Y} \right\} d\tilde{y}_1 \dots d\tilde{y}_n. \end{aligned}$$

At this time, where the log-likelihood function of equation (26) is differentiated with respect to b , it can be easily seen that:

$$\frac{\partial \ln L}{\partial b} = \sum_{y_i = 0} \frac{|\tilde{\Sigma}|^{-\frac{1}{2}} (2\pi)^{-\frac{n}{2}} |R|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \tilde{Y}' R^{-1} \tilde{Y} \right\}}{\int_{-\infty}^b \dots \int_{-\infty}^b |\tilde{\Sigma}|^{-\frac{1}{2}} (2\pi)^{-\frac{n}{2}} |R|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \tilde{Y}' R^{-1} \tilde{Y} \right\} d\tilde{y}_1 \dots d\tilde{y}_n} > 0. \quad (28)$$

That is, the log-likelihood function is monotonically increasing with b . Clearly from the definition of b , we can choose: $\hat{b} = \min(\tilde{y}_i | y_i \text{ is observed})$, in order to maximize the log-likelihood function.

Acknowledgement

We thank Colin McKenzie, Toshiaki Watanabe, and Taku Yamamoto for their valuable comments. Financial support from the Hitotsubashi University Research Unit for Statistical Analysis in Social Sciences, a 21st Century COE Program is greatly appreciated.

Footnotes

- 1 When the new strike price sets within three months before expiration day, these option data do not exist consecutively for four months. These data are excluded from our analysis.
- 2 Some literature, Honore (1992), Kyriazidou (1997), and Charlier, Melenberg and van Soest (2001) suggested the semiparametric estimation for fixed panel model and limited dependent variables.

References

- Anshuman, V. and A. Kalay (1998), "Market Making Rents under Discrete Prices," *Review of Financial Studies* 11, pp.81-109.
- Bauwens, L. and M. Lubrano (2002), "Bayesian Option Pricing Using Asymmetric GARCH Models," *Journal of Empirical Finance* 9, pp.321-342.
- Black, F. and M. Scholes (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, pp.673-659.
- Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 31, pp.673-659.
- Bookstaber, R. M. (1981), "Observed Option Mispricing and the Nonsimultaneity of Stock and Option Quotations," *The Journal of Business* 54, pp.141-155.
- Cameron, A. C. and P. K. Trivedi (2005), *Microeconometrics Methods and Applications*, Cambridge University Press.
- Charlier, E., B. Melenberg and A. van Soest (2001), "An Analysis of Housing Expenditure using Semiparametric Models and Panel Data," *The Journal of Econometrics* 101, pp.71-107.
- Crouhy, M. (1994), "Extensions of the Black-Scholes Model to the case of Stochastic Volatility," unpublished manuscript.
- Ding, Z., C. W. J. Granger, and R. F. Engle (1993), "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance* 1(1), pp.83-106.
- Duan, J. C. (1995), "The GARCH Option Pricing Model," *Mathematical Finance* 5, pp.13-32.
- Duan, J. C. and H. Zhang (2001), "Pricing Hang Seng Index Options around the Asian Financial Crisis . A GARCH Approach," *Journal of Banking and Finance* 25, pp.1989-2014.
- Easley, D., M. O'Hara, and P. S. Srinivas (1998), "Option Volume and Stock Prices: Evidence on Where Informed Traders Trade," *The Journal of Finance* 53, pp.431-465.
- Fama, E. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work," *The Journal of Finance* 25, pp.383-417.
- Glosten, L. R., R. Jagannathan, and D. Runkle (1993), "On the Relation between the Expected Value and the Volatility of Nominal Excess Returns on Stocks," *The Journal of Finance* 48, pp.1179-1801.
- Honore, B. E. (1992), "Trimmed LAD and Least Squared Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica* 60, pp.533-565. Kyriazidou, E. (1997), "Estimation of a Panel Data Sample Selection Model," *Econometrica* 65, pp.1335-1364.
- Long, D. M. and D. T. Officer (1997), "The Relation between Option Mispricing and Volume in the Black Scho-

- les Option Model,” *Journal of Financial Research* 20(1), pp.1-12.
- Nishina, K. and M. M. Nabil (1997), “Return Dynamics of Japanese Stock Index Options,” *The Japanese Economic Review* 48, pp.43-64.
- Nelson, D. B. (1991), “Conditional Heteroskedasticity in Asset Returns: A New Approach,” *Econometrica* 59, pp.347-370.
- Parlour, C. (1998), “Price Dynamics in Limit Order Markets,” *Review of Financial Studies* 11, pp.789-816.
- Renault, E. (1997), “Econometric Models of Option Pricing Errors,” in D. M. Kreps and K. F. Wallis eds., *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress, Vol.3*, Cambridge University Press, Chapter 9, pp.223-278.
- Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

※When I was a doctoral student at Hitotsubashi University, I started this research with my supervisor late professor Satoru Kanoh. While I am deeply grateful to professor Satoru Kanoh for his constant guidance during his life, I am thinking of you and praying for you in this time of loss.