Nonparametric Option Pricing by Transformation

Jin-Chuan Duan*

March 2002

Abstract

This paper develops a nonparametric option pricing theory and numerical method for European, American and path-dependent derivatives. In contrast to the nonparametric curve fitting techniques commonly seen in the literature, this nonparametric pricing theory is more in line with the canonical valuation method developed Stutzer (1996) for pricing options with only a sample of asset returns. Unlike the canonical valuation method, however, our nonparametric pricing theory characterizes the asset price behavior period-by-period and hence is able to price European, American and path-dependent derivatives. This nonparametric theory relies on transformation to normality and can deal with asset returns that are either i.i.d. or dynamic. Applications to simulated and real data are provided and implications discussed.

Key words: Risk-neutralization, relative entropy, Markov chain, GARCH, empirical distribution, volatility, smile/smirk, extreme value distribution.

*Rotman School of Management, University of Toronto; Tel: 416-946-5653, Fax: 416-971-3048, E-mail: jcduan@rotman.utoronto.ca; Web: http://www.rotman.utoronto.ca/~jcduan. Duan acknowledges the support received as Manulife Chair in Financial Services at University of Toronto. This research is funded by grants from Social Sciences and Humanities Research Council of Canada and Natural Sciences and Engineering Council of Canada. The author thanks Michael Carey for computing assistance and the seminar participants at University of Chicago and Western Ontario for comments.

1 Introduction

Nonparametric techniques in typical option pricing applications use a sample of option prices to calibrate the relationship between the option price and the underlying asset price. Hutchinson, *et al* (1994), Derman and Kani (1994), Rubinstein (1994), Aït-Sahalia and Lo (1995), Buchen and Kelly (1996), Jackwerth and Rubinstein (1996), Jacquier and Jarrow (2000), Broadie, *et al* (2000) and Garcia and Gencay (2001) are some examples. In a nutshell, these papers apply sophisticated curve fitting techniques to hopefully extract the true option pricing function from the observed option prices and the underlying asset value. Some of these approaches are limited in the sense that a large amount of data is needed, for example, Ait-Sahalia and Lo (1995), and they are subject to the so-called curse of dimensionality. Others, such as Buchen and Kelly (1996) are limited in their applicability to one maturity at a time because the nonparametric risk-neutral distributions can only be identified separately for different maturities. Consequently, they are not suitable for interpolation across maturity. Most of them cannot be used to price the derivative contracts that are not already covered in the data set in terms of contract features, because these methods are essentially interpolation devices. In other words, one cannot calibrate the model to European option data over a particular strike price range and hope to price an option with a strike price outside the range or to price path-dependent derivatives such as barrier options. For those can be used to extrapolate such as Derman and Kani (1994), Rubinstein (1994) and Jackwerth and Rubinstein (1996), the pricing system are too constrained to capture some important option price features. A common feature of these valuation approaches is the need to have some option data in order to implement the models.

The only nonparametric option pricing model that can value options solely based upon a sample of the underlying asset prices is, to our knowledge, the canonical valuation method of Stutzer (1996). It derives a risk-neutral distribution using the relative entropy principle. The risk-neutral distribution is a distribution function closest to the empirical distribution for the gross return over the maturity of interest subject to the condition that the expected return equals the risk-free rate. The canonical valuation method cannot deal with early exercise, however. This means that American style options cannot be priced using this method. The reason is that the canonical valuation method is silent on the period-by-period risk-neutral price dynamic. For the same reason, the canonical valuation method cannot be used to price path-dependent derivatives.

In a spirit similar to the canonical valuation method of Stutzer (1996), we develop a nonparametric option pricing theory which establishes the pricing relationship solely based upon the price data of the underlying asset without resorting to option prices. This nonparametric option pricing theory is not another curve fitting technique because it formalizes the risk-neutralization process so that one can infer directly from the price dynamic of the underlying asset to establish the risk-neutral pricing dynamic. The typical nonparametric option pricing technique requires of calibration of the model to option data. But the nonparametric pricing theory proposed in this paper does not need to be calibrated to option data. This feature has a key advantage because one can price derivatives for which no comparable contingent claims are traded and the theory can also be subject to a direct test in terms of its ability to price exchange traded options. Since the entire risk-neutral pricing dynamic is fully characterized, one can use the theory to price European, American and a variety of path-dependent derivatives. It is in this regard that our nonparametric option pricing theory differs from the canonical valuation method of Stutzer (1996).

Transforming one-period asset return to normality is a key step in constructing our nonparametric option pricing theory. This transformation can be easily accomplished by first applying the empirical distribution and then inverting it by the standard normal distribution function. Applying the relative entropy principle with the condition that the expected asset return equals the risk-free rate, one can derive the risk-neutral distribution for the normalized asset return. This risk-neutral distribution turns out to be again the normal distribution but with a mean shift to absorb the asset risk premium. We show that the nonparametric theory is in a complete agreement with the Black-Scholes (1973) model if one assumes normality for continuously compounded asset returns. For a dynamic asset return model such as GARCH, we show that the nonparametric option pricing theory yields the same pricing conclusion as the GARCH option pricing model of Duan (1995) when conditional normality is imposed.

Our nonparametric option pricing theory also differs from the canonical valuation method of Stutzer (1996) in its ability to reflect the market condition. Directly applying past returns to construct future returns as in the canonical valuation method inevitably fails to reflect the important elements of market conditions. Consider, for example, using the canonical valuation method to price one-month options on a particular day of high market volatility. Suppose that a sample of 250 one-month past returns are constructed by shifting one day backward at a time. The empirical distribution from this sample of overlapping one-month past returns is unlikely to properly reflect what is likely to happen in the coming month upon which the one-month options critically depend. A sample containing returns which ran up to a high volatility state on the option valuation day is most likely to differ from the return characteristics during the subsequent one-month period.

To operationalize an option pricing theory, one needs to develop numerical schemes for pricing various kinds of derivatives. The nonparametric pricing theory developed in this paper can be easily implemented using Monte Carlo method for pricing European and many path-dependent derivatives. Recent advancements in the Monte Carlo method by Carriere (1996), Tsitsiklis and van Roy (1999), Longstaff and Schwartz (2001), Rogers (2001) and Andersen and Broadie (2001) make it possible to compute American options as well. For lower-dimensional valuation problems, the Monte Carlo is less attractive, however. We thus adapt the Markov chain method of Duan and Simonato (2001) to the nonparametric setting in computing European and American style options. We favor the Markov chain method because it is a more efficient numerical method for valuation problems which can be expressed as a one- or two-dimensional Markov system. There are two cases examined in this paper.

The i.i.d. case constitutes a one-dimensional Markov system whereas the dynamic case can be expressed as a two-dimensional Markov system.

Our results in the i.i.d. case show that the nonparametric pricing theory performs reasonably well in the simulation environment when the data generating system is based on the Black-Scholes model. For the S&P 500 index data, the nonparametric method is found to produce volatility smile/smirk for short-maturity options. In other words, it reflects the skewness and kurtosis properties of the real data. This is encouraging because the S&P 500 index options are known to exhibit this pattern. The decaying pattern of the produced volatility smile/smirk over the maturity dimension is, however, inconsistent with our knowledge about the S&P 500 index options. The implied volatility surface flattens out too fast and lacks the complexity of observed volatility surface. Moreover, smile/smirk at short-maturity does not appear to be as steep as one typically observes on the S&P 500 index options.

Such a result is actually expected. Financial returns are known to exhibit clustering stochastic volatilities. Empiricists often use GARCH processes to model them. In short, the i.i.d. assumption, implicitly in many theoretical models, is incompatible with data. Since our nonparametric pricing theory is not constrained by the i.i.d. assumption, we apply it to the S&P 500 index data by adopting a GARCH process as the description of the dynamic volatility structure. The results suggest that the implied volatility surface is indeed very different in comparison to that under the i.i.d. assumption. The surface flattens out much slower when maturity is increased. Even up to a maturity of six months, smile/smirk is still clearly present. The smile/smirk pattern for shorter maturities is much steeper as compared to that under the i.i.d. assumption. Qualitatively, these features are consistent with the stylized facts associated with index options. Our nonparametric pricing model using the dynamic volatility structure also distinguishes itself in terms of its ability to reflect the prevailing market condition. We show the implied volatility surface responds in a significant way to the level of market volatility at the time of option valuation. Although the general surface decaying pattern is same, the smile/smirk pattern at shorter maturities is greatly affected by the market condition.

2 The non-parametric option pricing theory

Consider a sequence of continuously compounded asset returns, denoted by $\{R_t; t = 1, 2, \cdots\}$. To develop an operational option pricing theory, there are two critical issues need to be addressed. First, one must derive a corresponding risk-neutral distribution for R_t, which is a distribution function that can be used to price options as if economic agents were risk-neutral. Second, one must come up with a scheme for computing values for European, American and exotic derivatives. The first issue is the core of an option pricing theory whereas the second has an important operational significance and usually requires numerical methods. We now deal with the first issue and leave the second one to the next section.

Financial returns are well known to have some dynamic features. The most notable one is the volatility clustering phenomenon. In other words, $\{R_t; t = 1, 2, \cdots\}$ may be a stationary ergodic sequence but need not be an independent one. In order to develop the nonparametric option pricing theory we need to filter out the dynamic feature. Here we assume the dynamic feature occurs only in the one-period conditional mean and variance of R_t, denoted by μ_t and σ_t^2. We further assume that they are functions of past asset returns. This further assumption is needed because we want asset returns to form a self-determining stochastic system. Due to the assumptions, $\left\{\frac{R_t-\mu_t}{\sigma_t}; t = 1, 2, \cdots\right\}$ forms an i.i.d. sequence.

Let $G(\cdot)$ be the distribution function of $\frac{R_t-\mu_t}{\sigma_t}$ and assume it is a continuous distribution. Define $Z_t \equiv \Phi^{-1}\left(G\left(\frac{R_t-\mu_t}{\sigma_t}\right)\right)$ where $\Phi(\cdot)$ stands for the standard normal distribution function. It can be verified straightforwardly that Z_t is a standard normal random variable and the transformed variables $\{Z_t; t = 1, 2, \cdots\}$ form an i.i.d. sequence. We will refer to Z_t as the normalized asset return. The probability law governing R_t (or Z_t) is typically referred to as the physical probability measure. In parametric models, $G(\cdot)$ is a function under the physical measure and can usually be deduced from assuming a stochastic process for the asset price dynamic. Since we deal with the valuation problem nonparametrically, $G(\cdot)$ will be obtained via some nonparametric (or semiparametric) means. A concrete method will be provided later in the paper.

Knowing that the risk-neutral distribution is not the same as the physical distribution, it is natural to identify their differences. A key feature of the risk-neutral distribution is its expected return equal to the risk-free rate. In principle, one can estimate the expected return of the physical distribution which is typically different from the risk-free rate. It allows for at most one degree of freedom in risk-neutralization if one is to completely characterize the risk-neutral distribution. Other than the expected return condition, there is no *a priori* reason for the risk-neutral distribution to deviate from the physical distribution. In other words, it is natural to have a risk-neutral distribution that is as close to the physical distribution as possible while satisfying the condition on its expected value. To operationalize this concept, one can call upon the information theory to find the risk-neutral density function that minimizes the so-called relative entropy subject to its expected value condition. It is well known in information theory that the relative entropy principle can be justified axiomatically and is consistent with Bayesian method of statistical inference. For option pricing, the notion of relative entropy was previously utilized in Buchen and Kelly (1996) and Stutzer (1996), among others. Specifically, we deal with the physical and risk-neutral density functions for the normalized return and introduce one degree of freedom so that it can be used to match the required expected return correctly. Note that the normalized return has a standard normal density function, $\phi(\cdot)$. Using the relative entropy principle, the risk-neutral density for the normalized return, Z_t, is the solution to the following problem: for some value c_t,

$$\min_{f(x)} \int_{-\infty}^{\infty} f(x) \ln \frac{f(x)}{\phi(x)}\, dx \tag{1}$$

$$\begin{aligned} \text{subject to} \qquad \int_{-\infty}^{\infty} f(x)\, dx &= 1 \\ \int_{-\infty}^{\infty} x f(x)\, dx &= c_t. \end{aligned}$$

Note that the subscript of c_t is meant to reflect the fact that the risk-neutral density for the one-period normalized return may be a function of time because of the potential dynamic structure in the conditional mean and variance. Indeed, we will show later that the risk-neutral density for the one-period normalized return is the same for all period if the mean and variance are constant.

It is well known in the information theory that the above programming problem has the solution in the form of

$$\begin{aligned} f(x;\lambda_t) &= \frac{\phi(x)\exp(\lambda_t x)}{\int_{-\infty}^{\infty} \phi(x)\exp(\lambda_t x)\, dx} \\ &= \phi(x-\lambda_t). \end{aligned} \tag{2}$$

Note that the first condition of being a density function is satisfied by the above solution and the value of λ_t simply corresponds to a given value of c_t. In other words, we might just as well ignore c_t and view the density function as parameterized by λ_t. The value of λ_t is of course determined by the fact that the risk-neutral density must give rise to an expected asset return equal to the risk-free rate r (continuously compounded) minus the dividend yield d (continuously compounded). That is, λ_t^* solves

$$\int_{-\infty}^{\infty} \exp\left[\sigma_t G^{-1}(\Phi(x)) + \mu_t\right] \phi(x-\lambda_t^*)\, dx = \exp(r-d). \tag{3}$$

Note that the above equation is due to $R_t = \sigma_t G^{-1}(\Phi(Z_t)) + \mu_t$ so that term $\exp\left[\sigma_t G^{-1}(\Phi(x)) + \mu_t\right]$ is the one-period gross return of the asset. It can be shown that the left hand side of the above equation is a monotonically increasing function of λ_t^*, which implies a unique solution if the solution does exist. In general, the solution exists because the value of μ_t must be tied to r in any sensible market equilibrium.

Using the transformation, the sequence of continuously compounded asset return can be expressed as $\{R_t; t = 1, 2, \cdots\} = \{\sigma_t G^{-1}(\Phi(Z_t)) + \mu_t; t = 1, 2, \cdots\}$ where Z_t is the normalized return and has the physical density function of $\phi(x)$ and the risk-neutral density function of $\phi(x-\lambda_t^*)$. Consequently, the risk-neutral asset value dynamic becomes

$$S_t = S_{t-1}\exp\left[\sigma_t G^{-1}(\Phi(Z_t)) + \mu_t\right] \tag{4}$$

where Z_t is a normal random variable with mean λ_t^* and variance 1, which can be used to value contracts contingent upon the path of S_t. In short, we have succeeded in characterizing the risk-neutralized valuation system in terms of the normalized return.

Example 1: the Black-Scholes option pricing model

This nonparametric option pricing theory is compatible with the celebrated Black-Scholes (1973) model. We now substantiate this claim. Under the geometric Brownian motion assumption, the one-period continuously compounded return has a normal distribution with constant mean μ and variance σ^2, which implies $G(x) = \Phi(x)$ and $G^{-1}(\Phi(Z_t)) = Z_t$. Assume $d = 0$. According to our nonparametric pricing theory, Z_t has the risk-neutral density of $\phi(x - \lambda^*)$ where by the condition in (3), λ^* satisfies

$$\int_{-\infty}^{\infty} \exp(\sigma x + \mu)\phi(x - \lambda^*)\ dx = \exp(r). \tag{5}$$

It is obvious that λ^* will be a constant. Using the moment generating function of normal random variable, we obtain

$$\lambda^* = -\frac{\mu - r + \frac{\sigma^2}{2}}{\sigma}. \tag{6}$$

By equation (4), the risk-neutral asset price dynamic becomes

$$\begin{aligned} S_t &= S_{t-1}\exp\left(\sigma Z_t + \mu\right) \\ &= S_{t-1}\exp\left(r - \frac{\sigma^2}{2} + \sigma\varepsilon_t\right) \end{aligned} \tag{7}$$

where $\varepsilon_t = Z_t - \lambda^*$ is a standard normal random variable under the risk-neutral distribution. This result is in a complete agreement with the Black-Scholes model.

In reality, one does not know whether the asset price dynamic is governed by a geometric Brownian motion even if it were the case. What will happen if one applies our nonparametric option pricing theory in this situation? In other words, how well will the nonparametric option pricing theory perform if one only observes a sequence of realized returns generated by a geometric Brownian motion yet without knowing so? This issue will be addressed after we discuss a method of obtaining nonparametrically the function $G(\cdot)$.

Example 2: the GARCH option pricing model

Suppose we assume the asset return volatility follows a linear GARCH dynamic: $\sigma_t^2 = \beta_0 + \beta_1\sigma_{t-1}^2 + \beta_2(R_{t-1} - \mu_{t-1})^2$. Moreover, the conditional mean has the form as $\mu_t = r + \eta\sigma_t - \frac{1}{2}\sigma_t^2$ and the conditional distribution is normal. Note that we have implicitly assumed $d = 0$. Under these assumptions, $G(x) = \Phi(x)$ and the condition in (3) implies that λ_t^* solves

$$\int_{-\infty}^{\infty} \exp(\sigma_t x + r + \eta\sigma_t - \frac{1}{2}\sigma_t^2)\phi(x - \lambda_t^*)\ dx = \exp(r). \tag{8}$$

By the moment generating function of the normal random variable, it becomes

$$\exp(r + \eta\sigma_t + \lambda_t^*\sigma_t) = \exp(r), \tag{9}$$

which in turn implies $\lambda_t^* = -\eta$. The risk-neutral asset price dynamic becomes

$$\begin{aligned} S_t &= S_{t-1} \exp\left(\sigma_t Z_t + \mu_t\right) \\ &= S_{t-1} \exp\left(r - \frac{\sigma_t^2}{2} + \sigma_t \varepsilon_t\right) \end{aligned} \tag{10}$$

where $\varepsilon_t = Z_t + \eta$ is a standard normal random variable with respect to the risk-neutral distribution and the volatility dynamic becomes

$$\begin{aligned} \sigma_t^2 &= \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 (R_{t-1} - \mu_{t-1})^2 \\ &= \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-1}^2 (\varepsilon_{t-1} - \eta)^2. \end{aligned} \tag{11}$$

Thus, the nonparametric option pricing theory yields a result agreeing with the GARCH option pricing model of Duan (1995).

Duan (1999) assumes a parametric form for the conditional distribution to allow for conditional leptokurtosis and derives an option pricing theory using an equilibrium argument similar to Duan (1995). In contrast, the nonparametric option pricing theory does not require any prior knowledge on the conditional distribution and uses the relative entropy principle and transformation to normality as the basis for deriving the theory. In the next section, we will implement a version of the GARCH model to real data without assuming any prior knowledge on the conditional distribution.

3 Implementing algorithm for the nonparametric pricing theory

In order to implement the nonparametric pricing theory, one must first obtain a nonparametric (or semiparametric) distribution function for the continuously compounded return. There are many ways of constructing a nonparametric distribution function from a sequence of data. In this section, we consider a simple procedure of using the empirical distribution, which we find it convenient for our purpose of option valuation. One important feature required of the construction method is to be able to invert the distribution function quickly because its inversion is required in identifying λ_t^*. We will proceed with the i.i.d. case first and later move on to the dynamic model.

3.1 I.I.D. case

This is a simpler case in terms of identifying the nonparametric distribution function as well as for valuing derivatives numerically. We adopt the following procedure:

Step 1: Identify the empirical distribution from a sample of one-period continuously compounded asset returns $\{R_i; i = 1, \cdots N\}$. We first compute the sample mean and standard deviation, denoted by $\overline{\mu}$ and $\overline{\sigma}$, respectively. The empirical distribution function for a sample, $\mathbf{R} = \{\frac{R_i - \overline{\mu}}{\overline{\sigma}}; i = 1, \cdots N\}$, is formally defined as $\widehat{G}(x; \mathbf{R}) \equiv \frac{1}{N} \sum_{i=1}^{N} 1\{\frac{R_i - \overline{\mu}}{\overline{\sigma}} \leq x\}$ where $1\{\cdot\}$ is an indicator function giving a value of 1 if the condition is true and 0 otherwise. Note that $\widehat{G}(x; \mathbf{R})$ is actually a step function and is not invertible. The empirical distribution is also subject to sampling variation and has a bounded support. It is therefore preferable to use a smoothed version, denoted by $G(x; \widehat{\Theta})$, to make it invertible, to dampen out the sampling fluctuation and to allow for the possibility of having an infinite support. The particular smoothing technique used in this paper is described in Appendix. Figures 1a and 1b show the empirical distribution and its smoothed version for the simulated samples with 252 and 1260 standard normal random variates. The functions for the S&P 500 index return data sample are given in Figure 1c.

Step 2: Solve for λ^* numerically in

$$\int_{-\infty}^{\infty} \exp\left[\sigma G^{-1}\left(\Phi(x); \widehat{\Theta}\right) + \mu\right] \phi(x - \lambda^*)\, dx = \exp(r - d). \tag{12}$$

We use a bisection search to find λ^* where the integral is evaluated numerically and $\Phi(x)$ is evaluated using the standard polynomial approximation formula. Note that μ and σ do not need to share the same values as $\overline{\mu}$ and $\overline{\sigma}$. One may want to use, for example, two years worth of past return data to come up with $G(x; \widehat{\Theta})$ but decide to use other value of μ and σ for the future to better reflect the new market conditions. This can be justified if, for example, the interest rate has gone up substantially relative to the interest rates during the period the sample is taken. The future expected continuously compounded return μ need to reflect the new level of interest rates even if one keeps the same volatility level. One can, for example, use a value of μ equal to the sum of the new interest rate and the historical risk premium minus the anticipated dividend yield for the period to come.

Step 3: Choose a numerical scheme to generate the asset price until the maturity of the contingent contract according to the following system:

$$S_t = S_{t-1} \exp\left[\sigma G^{-1}\left(\Phi(Z_t); \widehat{\Theta}\right) + \mu\right] \tag{13}$$

where Z_t is a sequence of independent normal random variables with mean λ^* and variance 1. Then compute the expected value of the contingent payoff. For example, one may use Monte Carlo simulation to perform this task. In this paper, we will use the Markov chain method for the valuation task because it is a more efficient algorithm and can be used for European and American style derivatives. The technical details are given in Appendix.

We consider two examples. First, we implement the nonparametric pricing technique using the artificial data sets generated according to a geometric Browning motion. We then implement the method on the real data set of the S&P 500 index returns. In the first case, we assess the values for a set of European call options using the nonparametric pricing technique for each of the simulated data sets. As a comparison, we apply the Black-Scholes model to the same data sets to obtain the corresponding pricing results. The comparison in this case is to assess how well the Black-Scholes theoretical values can be uncovered by our nonparametric method. In the case of real data, we compute option values using both our nonparametric pricing technique and the Black-Scholes model. This comparison sheds light on the consequence of assuming normality for the real data that are known to be negatively skewed and fat-tailed. In addition to European calls, we also compute American puts to demonstrate that our nonparametric method is applicable to derivatives with early exercise possibilities.

To simulate the data according to the geometric Brownian motion, we assume $\mu = 0.1$ and $\sigma = 0.15$ (annualized). We also assume $d = 0$ and $r = 0.05$ in the simulation study. The nonparametric method is subject to sampling errors in a way similar to implementing the Black-Scholes model with the estimated sample standard deviation. We consider two sample sizes: 1 year (252 days) and 5 years (1260 days). Note that the parameters need to be converted to the ones suitable for daily frequency because one trading day is regarded as the length of one period in this analysis. The statistics on mean, median, standard deviation, maximum and minimum are calculated using 200 Monte Carlo repetitions. The results are summarized in Tables 1 and 2. In these tables, we have in the first row, corresponding to each maturity, the Black-Scholes theoretical values using the true parameter value, i.e., $\sigma = 0.15$. Two groups of results are reported with the first computed using our nonparametric pricing method and the second using the sample standard deviation. In each group, we report five numbers: mean, median, standard deviation, minimum and maximum of the estimated option prices obtained in 200 simulation runs. For the sample size of 252, the statistics indicate that the nonparametric technique performs reasonably well in comparison to the results using the Black-Scholes model with the estimated volatility, keeping in mind that using the Black-Scholes model is expected to perform better because the additional knowledge about the true nature of the data generating process is utilized in its implementation. Even taking advantage of the additional information about the simulated data, the improvement over the nonparametric pricing technique is not great as measured by the standard deviation. Both methods seem to yield upward biased price estimates with a larger bias for the nonparametric method. The general properties remain the same when the sample size is increased to 1260. Not surprisingly, the standard deviation decreases as the sample size increases.

We now turn to the implementation using real data. The data set consists of the S&P500 index and the three-month Treasury bill rates on a daily basis from the last trading day of 1995 to the last trading day of 2000. This yields 1263 daily excess returns in the sample. We conduct the option valuation on December 29, 2000 which is the last day of the data sample. The prevailing interest rate was 5.89%. After converting it to the continuously compounded rate, we have $r = 0.057231$. For option valuation under the i.i.d. assumption, we thus use $\mu = 0.101397769 + 0.057231$ and $\sigma = 0.184390836$. We have assumed $d = 0$ in both estimation and option valuation. The results for European calls are summaried in Figure 2 where the option values produced by the nonparametric method are converted to the implied volatilities using the Black-Scholes formula. Such a plot is usually referred to as the implied volatility surface. Whenever the implied volatility is higher (lower) than 0.184390836, the nonparametric method yields a higher (lower) option value relative to the Black-Scholes model. Figure 2 reports the implied volatilities from 7 trading days to 6 months over the moneyness range from 0.85 to 1.15. It is clear that option prices inferred from the real data differ from those suggested by the Black-Scholes model. For short-term options, there is a clear pattern of volatility smile/smirk, suggesting that for in-the-money call options, the nonparametric method yields values higher than those by the Black-Scholes model. For out-of-the-money calls, the increase is much smaller in magnitude. This pattern is not at all surprising given that the sample skewness and kurtosis are -0.340393637 and 6.568075038, respectively. Fat-tails are expected to give rise to higher values for in- and out-of-the money call options relative to the Black-Scholes model values. The negative skewness, however, makes in-the-money calls even more valuable but lessens the effect of fat-tails on out-of-the-money calls. Since this is the typical pattern exhibited by market prices of the S&P 500 index options, the nonparametric method which truthfully reflects the actual empirical distribution appears to be a superior way of approaching option valuation.

The smile/smirk pattern quickly disappears when maturity is increased. This feature is, however, at odd with the empirical regularities of the S&P 500 index options for which the smile/smirk pattern persists for longer maturities. Our nonparametric pricing results are, of course, computed under the i.i.d. assumption. The empirical evidence has long suggested that a dynamic structure such as the GARCH effect is clearly present in financial data. In other words, it is not surprising to produce an option pricing result that is at odd with the empirical regularity. Flattening of smile/smirk is driven by the Central Limit Theorem. With the typical dynamic structure observed in data, convergence to normality is expected to take place slower. The implication of this result is not trivial because any model with i.i.d. returns is expected to behave the same way in the maturity dimension. Different models only produce different degrees of smile/smirk for a given maturity. When maturity is increased, the same rate of reversion to normality applies to all models with i.i.d. returns, suggesting that these models will not be able to fit the implied volatility surface well.

For American options, we use puts instead of calls because American calls are effectively European calls when there is no dividends. American puts are priced using the nonparametric method and the Black-Scholes model and the results are reported in Table 3. As opposed to the method of Stutzer (1996), we are able to price American options nonparametrically because the nonparametric pricing system is developed on a period-by-period basis instead of fixing the risk-neutral distribution for a given maturity. For any maturity, one simply goes through our nonparametric pricing system period-by-period and recursively assesses the early exercise possibility. The numerical technique described in Appendix is suitable for European and American options. It can also be used to price barrier options using the idea of Duan, *et al* (1999).

For ease of comparison, all American option values reported are based on assuming the index level of 100 and these values can be easily translated to reflect the actual index level. All values are computed by allowing early exercise on a daily basis because one trading day is considered to be one basic period. The results are grouped according to the pricing method. For in-the-money puts at the highest strike price of 110, the nonparametric method yields higher values across all maturities. In the case of out-of-the-money puts, the Black-Scholes approach gives rise to higher option values. The intuitive reason for these results is not entirely clear. Although it is obvious that out-of-the-money European puts should have higher values under the nonparametric method due to negative skewness of the S&P 500 index return, early exercise has significantly complicated the intuition. It may be helpful to consider a two-period out-of-the-money put option. The stock price at time 1 is likely to be high because the original high price makes it out-of-the-money. This makes early exercise unwise. The value of holding on to the option is lower, however, if the return distribution is negatively skewed. That is to say the nonparametric method will assign a lower value than does the Black-Scholes model. This perhaps explain why American out-of-the-money puts have higher values using the Black-Scholes model. In terms of dollar values, two valuation methods do not generate significant differences. Percentage wise, however, the difference can be substantial for out-of-the-money options.

3.2 Dynamic case

We now describe the implementation steps of the nonparametric option pricing theory for asset returns exhibiting a dynamic structure. Recall our assumption that the dynamic structure is only present in the conditional mean and variance and they must be some functions of past returns.

Step 1: Identify econometrically suitable dynamic structures for the conditional mean μ_t and variance σ_t^2 using a sample of one-period continuously compounded asset returns $\{R_i; i = 1, \cdots N\}$; for example, use a GARCH-in-mean model without specifying the conditional distribution. To make the numerical valuation task more manageable, it is advisable to assume that the conditional mean is some function of conditional standard deviation because of the need to solve for λ_t^*.

Step 2: Construct a smoothed version of the empirical distribution just as in Step 1 of the preceding section using the normalized sample, $\mathbf{R} = \{\frac{R_i - \mu_i}{\sigma_i}; i = 1, \cdots N\}$. Specifically, we have $\widehat{G}(x; \mathbf{R}) \equiv \frac{1}{N}\sum_{i=1}^{N} 1\{\frac{R_i-\mu_i}{\sigma_i} \leq x\}$ where $1\{\cdot\}$ is an indicator function giving a value of 1 if the condition is true and 0 otherwise. Again, we use the smoothed version of the empirical distribution function $G(x; \widehat{\Theta})$.

Step 3: Solve for λ_t^* numerically in

$$\int_{-\infty}^{\infty} \exp\left[\sigma_t G^{-1}\left(\Phi(x); \widehat{\Theta}\right) + \mu_t\right] \phi(x - \lambda_t^*)\ dx = \exp(r - d). \tag{14}$$

We use a bisection search to find λ_t^* where the integral is evaluated numerically and $\Phi(x)$ is evaluated using the standard polynomial approximation formula. Note that λ_t^* must be solved for every pair of (μ_t, σ_t) and that is why parametrizing μ_t as a function of σ_t will simplify the numerical task. Note that the numerical method used later partitions σ_t into n states. As a result, λ_t^* only needs to be evaluated n times under the assumption that μ_t is a function of σ_t.

Step 4: Choose a numerical scheme to generate the asset price until the maturity of the contingent contract according to the following system:

$$S_t = S_{t-1} \exp\left[\sigma_t G^{-1}\left(\Phi(Z_t); \widehat{\Theta}\right) + \mu_t\right] \tag{15}$$

where Z_t is a sequence of independent normal random variables with mean λ_t^* and variance 1. Note that μ_t, σ_t and λ_t^* are known at time $(t-1)$ because of the assumption that μ_t and σ_t are functions of past returns. (For example, this is indeed the case for the GARCH model.) Then compute the expected value of the contingent payoff.

To demonstrate the use of the nonparametric option pricing theory in a dynamic setting, we assume the following nonlinear asymmetric GARCH-in-mean model:

$$R_t = \mu_t + \sigma_t \epsilon_t \tag{16}$$

$$\mu_t = r - d + \eta\sigma_t - \frac{1}{2}\sigma_t^2 \tag{17}$$

$$\sigma_t^2 = \beta_0 + \beta_1\sigma_{t-1}^2 + \beta_2\sigma_{t-1}^2(\epsilon_{t-1} - \theta)^2 \tag{18}$$

where ϵ_t's are i.i.d. random variables with mean 0 and variance 1 without specifying its distribution. Thus, $Z_t = \Phi^{-1}\left(G(\epsilon_t; \widehat{\Theta})\right)$ or $\epsilon_t = G^{-1}\left(\Phi(Z_t); \widehat{\Theta}\right)$. We use the Markov chain method for the valuation task and the technical details for this particular dynamic model are given in Appendix. The parameters in the above system are estimated using the quasi-maximum likelihood method because we do not assume any conditional distribution function. The data set is the S&P500 index excess return described in the preceding section. The parameter values obtained by the quasi-maximum likelihood estimation are $\eta = 0.0381459263$, $\beta_0 = 0.0000072571$, $\beta_1 = 0.7026496515$, $\beta_2 = 0.0748155199$ and $\theta = 1.5299656423$. As a by-product of estimation, the sample $\mathbf{R} = \{\frac{R_i - \mu_i}{\sigma_i}; i = 1, \cdots 1263\}$ is also obtained. Note that we have assumed $d = 0$ in this estimation.

Assume that we conduct the option valuation on December 29, 2000 which is the last day of the data sample. The prevailing interest rate was 5.89%, which is translated into $r = 0.057231$. The conditional standard deviation for the next day was estimated to be 0.22237484 (annualized), which will be used in the option valuation as the initial value of the volatility. Again, for ease of comparison all option values are computed based on assuming the index value of 100, and the parameters need to be converted for daily frequency because one trading day is regarded as the length of one basic period. In the case of European options, we use the implied volatility surface over the same range of moneyness and maturity as in Figure 2 to examine the impact of using a dynamic return model. The results are presented in Figure 3.

In comparison with Figure 2, there are two striking features. First, the smile/smirk is clearly steeper for shorter maturities with in-the-money calls being substantially higher than out-of-the-money calls. Second, the rate of flattening for smile/smirk is far slower than the results under the i.i.d. assumption. The fact that the volatility surface becomes flatter when maturity is increased is consistent with the Central Limit Theorem, suggesting that eventually the standardized cumulative continuously compounded return becomes normally distributed. What the dynamic model such as the GARCH process does is to generate a dependence structure which makes the cumulative return more skewed and fat-tailed initially. Beyond some point in the maturity dimension, convergence to normality begins to take hold but the rate of convergence is slow due to the dependence structure. It is clear that using the dynamic model makes a significant difference in option valuation. Given that the return data exhibit GARCH-like dynamic features, it is imperative to account for them in option pricing. Interestingly, the pattern in Figure 3 is qualitatively consistent with the results of using market prices of options commonly reported in literature.

The dynamic model allows one to consider the effect of the market condition at the time of option valuation. On December 20, 2000, the next day conditional volatility actually reached 0.35018012 due to a 3.1% drop in the market value. This volatility is substantially higher than the one a few days later previously used as our day of option valuation. We now conduct option valuation on December 20, 2000 to examine the impact of a higher initial market volatility. The volatility surface over the same range of moneyness and maturity is given in Figure 4. Comparing it to Figure 3, the surface is higher overall and is clearly steeper for shorter maturities. Again, the surface does not flatten quickly as in the i.i.d. case.

Figures 3 and 4 together suggest that the general property as to how smile/smirk evolves in the maturity dimension does not differ much for different levels of initial volatility. A higher (lower) initial volatility simply places the surface higher (lower).

The impact of dynamic dependence on American option values is examined using Table 4. There are two panels in this table with the top panel corresponding to the last data point of the sample and the bottom panel to a higher volatility state a few days earlier. Since the initial volatility for the bottom panel is higher, it is not surprising to see the values in the bottom panel higher than their corresponding ones in the top panel. The comparison simply indicates the importance of accounting for the market condition at the time of option valuation. A more interesting comparison is to consider the results reported in Table 3 under the i.i.d. assumption. The nonparametric method under the i.i.d. assumption tends to generate lower values for out-of-the-money puts but higher value for in-the-money puts (using the top panel of Table 4). This result indicates the importance of dynamic dependence, for it alters the cumulative return distribution as well as changes the early exercise decision.

4 Extension to derivatives on multiple assets

In order to price the derivatives on more than one asset, we generalize the theory to the case of multiple assets. Consider a sequence of k-dimensional vector of continuously compounded asset returns, denoted by $\{\mathbf{R}_t; t = 1, 2, \cdots\}$. We again assume that the dynamic feature occurs only in the one-period conditional mean and variance for each element of $\mathbf{R}_t$, denoted by $\mu_{i,t}$ and $\sigma^2_{i,t}$. We further assume that they are functions of past asset returns so that asset returns form a self-determining k-dimensional stochastic system. Due to the assumptions, $\left\{\left(\frac{R_{1,t}-\mu_{1,t}}{\sigma_{1,t}}, \frac{R_{2,t}-\mu_{2,t}}{\sigma_{2,t}} \cdots \frac{R_{k,t}-\mu_{k,t}}{\sigma_{k,t}}\right); t = 1, 2, \cdots\right\}$ forms an i.i.d. sequence of random vectors. Let $G_i(\cdot)$ be the marginal distribution function of $\frac{R_{i,t}-\mu_{i,t}}{\sigma_{i,t}}$. Define $Z_{i,t} \equiv \Phi^{-1}\left(G_i\left(\frac{R_{i,t}-\mu_{i,t}}{\sigma_{i,t}}\right)\right)$ where $\Phi(\cdot)$ again stands for the standard normal distribution function. Similar to the earlier result, each $Z_{i,t}$ is a standard normal random variable, but together the k-dimensional vector of transformed returns need not follow a multivariate normal distribution. Here we assume they form a multivariate normal distribution. This assumption amounts to assuming a normal copula in forming a joint distribution from marginal ones. Let Ω be the $k \times k$ correlation matrix. Denote by $\phi(\mathbf{x}; \Omega)$ and $\Phi(\mathbf{x}; \Omega)$ be the k-dimensional multivariate normal density and distribution functions with mean vector $\mathbf{0}$ and covariance matrix Ω.

Similar to the development in the earlier section, the risk-neutral density for the normalized returns is the solution to the following problem: for some set of values $\{c_{1,t}, c_{2,t}, \cdots c_{k,t}\}$ reflecting potentially time-varying nature of these values,

$$\min_{f(\mathbf{x})} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x}) \ln \frac{f(\mathbf{x})}{\phi(\mathbf{x}; \Omega)} \, d\mathbf{x} \tag{19}$$

$$\text{subject to} \quad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x})\, d\mathbf{x} = 1$$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f(\mathbf{x})\, d\mathbf{x} = c_{i,t}, \ \text{for } i = 1, 2, \cdots, k.$$

The solution to the above programming problem is in the form of

$$\begin{aligned} f(\mathbf{x}; \boldsymbol{\lambda}_t) &= \frac{\phi(\mathbf{x}; \Omega) \exp\left(\sum_{i=1}^{k} q_{i,t} x_i\right)}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(\mathbf{x}; \Omega) \exp\left(\sum_{i=1}^{k} q_{i,t} x_i\right) d\mathbf{x}} \\ &= \phi(\mathbf{x} - \boldsymbol{\lambda}_t; \Omega) \end{aligned} \tag{20}$$

where $\boldsymbol{\lambda}_t$ is a k-dimensional vector corresponding to $\{q_{1,t}, q_{2,t}, \cdots q_{k,t}\}$, which in turn corresponds to $\{c_{1,t}, c_{2,t}, \cdots c_{k,t}\}$. In other words, we might just as well ignore $\{c_{1,t}, c_{2,t}, \cdots c_{k,t}\}$ and view the density function as parameterized by $\boldsymbol{\lambda}_t$. The value of $\boldsymbol{\lambda}_t$ is of course determined by the fact that the risk-neutral density must give rise to an expected asset return equal to the risk-free rate r (continuously compounded) minus the dividend yield d_i (continuously compounded) on an asset-by-asset basis. Due to multivariate normality, this can be solved individually using the risk-neutral marginal distribution; that is, $\lambda^*_{i,t}$ solves

$$\int_{-\infty}^{\infty} \exp\left[\sigma_{i,t} G_i^{-1}\left(\Phi(x)\right) + \mu_{i,t}\right] \phi(x - \lambda^*_{i,t})\ dx = \exp(r - d_i). \tag{21}$$

Note that the above equation is again due to $R_{i,t} = \sigma_{i,t} G_i^{-1}\left(\Phi(Z_{i,t})\right) + \mu_{i,t}$ so that term $\exp\left[\sigma_{i,t} G_i^{-1}\left(\Phi(x)\right) + \mu_{i,t}\right]$ is the one-period gross return of the asset.

Using the transformation, the sequence of continuously compounded return for the i-th asset can be expressed as $\{R_{i,t}; t = 1, 2, \cdots\} = \left\{\sigma_{i,t} G_i^{-1}\left(\Phi(Z_{i,t})\right) + \mu_{i,t}; t = 1, 2, \cdots\right\}$ where $Z_{i,t}$ is the normalized return and has the physical density function of $\phi(x)$ and the risk-neutral density function of $\phi(x - \lambda^*_{i,t})$. Note that the correlation between $Z_{i,t}$ and $Z_{j,t}$ equals the (i,j)-th element of Ω under both the physical and risk-neutral distributions. As a result, the risk-neutral asset value dynamic becomes

$$S_{i,t} = S_{i,t-1} \exp\left[\sigma_{i,t} G_i^{-1}\left(\Phi(Z_{i,t})\right) + \mu_{i,t}\right], \text{ for } i = 1, 2, \cdots, k \tag{22}$$

where $Z_{i,t}$ is a normal random variable with mean $\lambda^*_{i,t}$ and variance 1 and the correlation between $Z_{i,t}$ and $Z_{j,t}$ equal to the (i,j)-th element of Ω. The multivariate option pricing system is thus completely characterized.

5 Appendix

5.1 Smoothing the empirical distribution function

For this purpose, we piece together three functions to form the smoothed version of the empirical distribution function. First we define a function that is the standard normal distribution function modified by a cubic polynomial to allow for departure from normality:

$$Q(x) = a\Phi(x)^3 + b\Phi(x)^2 + c\Phi(x) + d.$$

Next we define the general Pareto distribution, for $\omega > 0$,

$$H_{\delta,\omega}(y) = \begin{cases} 1 - \left(1 + \frac{\delta y}{\omega}\right)^{-\frac{1}{\delta}} & \text{if } \delta \neq 0 \\ 1 - \exp\left(-\frac{y}{\omega}\right) & \text{if } \delta = 0 \end{cases},$$

where the support is $y \geq 0$ if $\delta \geq 0$ and $0 \leq y \leq -\frac{\omega}{\delta}$ if $\delta < 0$. The distributions corresponding to $\delta > 0$ are fat-tailed whereas those corresponding to $\delta < 0$ have bounded tails. By the extreme value theory, most distribution functions have their exceedance distribution functions (distribution function conditional on a tail) well approximated by the general Pareto distribution. Technically, the following result, due to Balkema and de Haan (1974) and Pickands (1975), serves as the theoretical basis for this assertion: there is a positive measurable function $\omega(u)$ such that

$$\lim_{u\to\infty} \sup_{0\leq y<\infty} \left| \Pr\{Y - u \leq y | Y > u\} - H_{\delta,\omega(u)}(y) \right| = 0.$$

This provides an flexible way of fixing the functional form for both the right and left tails. For the middle portion, we use the cubic polynomial modified normal distribution function. Piecing three functions together in a continuous and differential manner gives rise to

$$G(x;\Theta) = \begin{cases} Q(u_1) - Q(u_1)H_{\delta_1,\omega_1}(e^{u_1-x} - 1) & \text{if } x < u_1 \\ Q(x) & \text{if } u_1 \leq x \leq u_2 \\ Q(u_2) + [1 - Q(u_2)]\, H_{\delta_2,\omega_2}(e^{x-u_2} - 1) & \text{if } x > u_2 \end{cases}$$

for $u_2 > u_1$.

The tail probabilities are modeled by the exceedance distribution on the exponential of x to ensure that the expected gross return is finite. If the expected gross return does not exist, the expected future stock price must be infinity and all call option values also become unbounded. Since x is a normalized continuously compounded return, the expected gross return amounts to the moment generating function evaluated at σ. For $\delta > 0$, the moment generating function does not exist because the tail probability rate is governed by a power function. Using the exponential of x in the exceedance distribution ensures existence of the moment generating function.

Note that there are eight parameters in the system, but six of them are free due to the smooth pasting requirement. These free parameters are $\Theta = (a, b, c, d, \delta_1, \delta_2)$. Continuity is a natural result of the construction but differentiability at u_1 and u_2 constrains two parameters - ω_1 and ω_2. In the implementation, we use $u_1 = -1.5$ and $u_2 = 1.5$. We find $\widehat{\Theta}$ by solving the following nonlinear regression problem:

$$\min_{\Theta} \sum_{i=1}^{N} \left[\widehat{G}\left(\frac{R_i - \overline{\mu}}{\overline{\sigma}}; \mathbf{R}\right) - G\left(\frac{R_i - \overline{\mu}}{\overline{\sigma}}; \Theta\right) \right]^2.$$

5.2 Option valuation: the I.I.D. case

We use the method of Duan and Simonato (2001) to come up an m-state time-homogeneous Markov chain to approximate the stochastic process of the transformed variable $X_t = \ln S_t - (r-d)t$ under the risk-neutral distribution. Let m be an odd integer to simplify the construction. Let x_U and x_L denote the largest and smallest state values used and make the center of the interval equal to the initial value of the target chain X_0. Equally partition the interval into $m-1$ cells and denote the set of m states by $\{x_1, x_2, \cdots, x_m\}$ with $x_1 = x_L$, $x_{(m+1)/2} = X_0$ and $x_m = x_U$. To facilitate the following derivation, we conveniently let $x_0 = -\infty$ and $x_{m+1} = \infty$. The m-state Markov chain has the transition probability from the ith state to jth state defined as

$$\begin{aligned}
\pi_{ij} &= \Pr\left\{ \frac{x_{j-1}+x_j}{2} < X_{t+1} \leq \frac{x_j + x_{j+1}}{2} \,\middle|\, X_t = x_i \right\} \\
&= \Pr\left\{ \frac{x_{j-1}+x_j}{2} < x_i - r + d + \sigma G^{-1}\left(\Phi(Z_{t+1}); \widehat{\Theta}\right) + \mu \leq \frac{x_j + x_{j+1}}{2} \right\} \\
&= \Pr\left\{ \begin{array}{c} \frac{1}{\sigma}\left(\frac{x_{j-1}+x_j}{2} - x_i + r - d - \mu\right) < G^{-1}\left(\Phi(Z_{t+1}); \widehat{\Theta}\right) \\ \leq \frac{1}{\sigma}\left(\frac{x_j+x_{j+1}}{2} - x_i + r - d - \mu\right) \end{array} \right\} \\
&= \Pr\left\{ \begin{array}{c} \Phi^{-1}\left[G\left(\frac{1}{\sigma}\left(\frac{x_{j-1}+x_j}{2} - x_i + r - d - \mu\right); \widehat{\Theta}\right)\right] < Z_{t+1} \\ \leq \Phi^{-1}\left[G\left(\frac{1}{\sigma}\left(\frac{x_j+x_{j+1}}{2} - x_i + r - d - \mu\right); \widehat{\Theta}\right)\right] \end{array} \right\} \\
&= \Phi\left\{ \Phi^{-1}\left[G\left(\frac{1}{\sigma}\left(\frac{x_{j-1}+x_j}{2} - x_i + r - d - \mu\right); \widehat{\Theta}\right)\right] - \lambda^* \right\} \\
&\quad -\Phi\left\{ \Phi^{-1}\left[G\left(\frac{1}{\sigma}\left(\frac{x_j+x_{j+1}}{2} - x_i + r - d - \mu\right); \widehat{\Theta}\right)\right] - \lambda^* \right\}.
\end{aligned}$$

Since π_{ij} does not depend on t, the Markov chain is time-homogeneous. Denote the transition probability matrix by Π with its (i,j)-th entry being π_{ij}. Note that Π is a highly sparse matrix by the nature of the problem. (Sparsity is obtained in our implementation by treating any value less than 10^{-6} as 0. We then normalize each row to have the row sum equal to 1.) Let the m-dimensional value vector at time t be V_t. Denote the contingent payoff function by $f(S_t)$ and the payoff vector corresponding to m values of x_i by F_t. In other words, the ith element of F_t is $f(\exp(x_i + (r-d)t))$. At maturity, $V_T = F_T$. The following recursive system can be used to value American style options:

$$V_t = \max\left\{F_t, e^{-r}\Pi V_{t+1}\right\}.$$

For European style options, it can be simplified to $V_t = e^{-r(T-t)}\Pi^{T-t}F_T$. The option value corresponding to the current stock price is the center element of V_t.

In our implementation, we let $x_g = 3\sigma\sqrt{T-t}$. The value for x_U and x_L are determined by

$$
\begin{aligned}
x_U &= X_0 + x_g \sqrt{\frac{m-1}{100}} \\
x_L &= X_0 - x_g \sqrt{\frac{m-1}{100}}.
\end{aligned}
$$

Note that we treat $m = 101$ as the base case for which the Markov chain is constructed to cover the target variable at the maturity of the option over the range of three standard deviations in each direction. Such a construction can ensure that the partition condition given in Duan and Simonato (2001) is satisfied so that the approximation algorithm will converge to the right theoretical value as m tends to infinity. The results reported in the paper are computed using $m = 501$.

5.3 Option valuation: the dynamic case

In the following description, we use the return dynamic given in (16)-(18). These restrictions make the target system into a two-dimensional Markov process. We again follow the idea of Duan and Simonato (2001) to come up an $m \times n$-state time-homogeneous Markov chain to approximate the stochastic process for the pair of transformed variable $X_{1t} = \ln S_t - (r-d)t$ and $X_{2t} = \ln \sigma_{t+1}^2$ under the risk-neutral distribution (with m values for X_{1t} and n values for X_{2t}). We make minor modifications to suit our specific problem whenever it is appropriate. Note that the price dynamic is a univariate non-Markovian system but has been changed to a bivariate Markovian system so that the Markov chain method can apply. Again m and n are odd integers to simplify the construction of the Markov chain.

Let x_{1U} and x_{1L} denote the largest and smallest state values used for X_{1t}. Similarly, we have x_{2U} and x_{2L} for X_{2t}. We center the transformed asset price interval at X_{10}, the initial value of the transformed asset price. For the logarithmic volatility, we center the interval at $\overline{X}_2$, which is the unconditional mean of X_{2t}. Note that we do not center the interval at X_{20}, the initial value of the logarithmic volatility, because the process X_{2t} is a stationary system which mean reverts to $\overline{X}_2$. Equally partition these intervals into $m-1$ and $n-1$ cells and denote the Cartesian product of $m \times n$ states by $\{(x_{1i}, x_{2j}) : i = 1, \cdots, m; j = 1, \cdots n\}$ where $x_{11} = x_{1L}$, $x_{1(m+1)/2} = X_{10}$, $x_{1m} = x_{1U}$, $x_{21} = x_{2L}$, $x_{2(n+1)/2} = \overline{X}_2$ and $x_{2m} = x_{2U}$. To facilitate the following derivation, we conveniently let $x_{10} = -\infty$ and $x_{1(m+1)} = \infty$. The $m \times n$-state Markov chain has the transition probability from the (i,j)-th state to (k,l)-th state defined as

$$
\begin{aligned}
\pi(i,j;k,l) &= \Pr\left\{ \frac{x_{1(k-1)} + x_{1k}}{2} < X_{1(t+1)} \leq \frac{x_{1k} + x_{1(k+1)}}{2} \,\middle|\, X_{1t} = x_{1i},\ X_{2t} = x_{2j} \right\} \\
&\quad \times \mathbf{1}\left\{ \frac{x_{2(l-1)} + x_{2l}}{2} < \Gamma\left(x_{1i}, x_{1k}, x_{2j}\right) \leq \frac{x_{2l} + x_{2(l+1)}}{2} \right\}
\end{aligned}
$$

where

$$\Gamma(x,y,z) = \ln\left[\beta_0 + \beta_1 e^z + \beta_2\left(y - x - (\eta+\theta)e^{z/2} + \frac{1}{2}e^z\right)^2\right].$$

Note that the indicator function comes into play because $X_{2(t+1)} = \Gamma\left(X_{1t}, X_{1(t+1)}, X_{2t}\right)$ due to equations (16)-(18).

Numerically solve for λ_j^* corresponding to each x_{2j} in

$$\int_{-\infty}^{\infty} \exp\left[e^{x_{2j}/2} G^{-1}\left(\Phi(x); \widehat{\Theta}\right) + \eta e^{x_{2j}/2} - \frac{1}{2}e^{x_{2j}}\right] \phi(x - \lambda_j^*)\, dx = 1,$$

which is derived from equation (14). Thus,

$$\begin{aligned}
& \Pr\left\{ \frac{x_{1(k-1)} + x_{1k}}{2} < X_{1(t+1)} \leq \frac{x_{1k} + x_{1(k+1)}}{2} \middle| X_{1t} = x_{1i},\ X_{2t} = x_{2j} \right\} \\
= & \Pr\left\{ \frac{x_{1(k-1)} + x_{1k}}{2} < x_{1i} + e^{x_{2j}/2} G^{-1}\left(\Phi(Z_{t+1}); \widehat{\Theta}\right) + \eta e^{x_{2j}/2} - \frac{1}{2}e^{x_{2j}} \leq \frac{x_{1k} + x_{1(k+1)}}{2} \right\} \\
= & \Pr\left\{ \begin{array}{c} e^{-x_{2j}/2}\left(\frac{x_{1(k-1)}+x_{1k}}{2} - x_{1i} + \frac{1}{2}e^{x_{2j}}\right) - \eta < G^{-1}\left(\Phi(Z_{t+1}); \widehat{\Theta}\right) \\ \leq e^{-x_{2j}/2}\left(\frac{x_{1k}+x_{1(k+1)}}{2} - x_{1i} + \frac{1}{2}e^{x_{2j}}\right) - \eta \end{array} \right\} \\
= & \Pr\left\{ \begin{array}{c} \Phi^{-1}\left[G\left(e^{-x_{2j}/2}\left(\frac{x_{1(k-1)}+x_{1k}}{2} - x_{1i} + \frac{1}{2}e^{x_{2j}}\right) - \eta; \widehat{\Theta}\right)\right] < Z_{t+1} \\ \leq \Phi^{-1}\left[G\left(e^{-x_{2j}/2}\left(\frac{x_{1k}+x_{1(k+1)}}{2} - x_{1i} + \frac{1}{2}e^{x_{2j}}\right) - \eta; \widehat{\Theta}\right)\right] \end{array} \right\} \\
= & \Phi\left\{\Phi^{-1}\left[G\left(e^{-x_{2j}/2}\left(\frac{x_{1k} + x_{1(k+1)}}{2} - x_{1i} + \frac{1}{2}e^{x_{2j}}\right) - \eta; \widehat{\Theta}\right)\right] - \lambda_j^*\right\} \\
& -\Phi\left\{\Phi^{-1}\left[G\left(e^{-x_{2j}/2}\left(\frac{x_{1(k-1)} + x_{1k}}{2} - x_{1i} + \frac{1}{2}e^{x_{2j}}\right) - \eta; \widehat{\Theta}\right)\right] - \lambda_j^*\right\}.
\end{aligned}$$

It is clear that $\pi(i,j;k,l)$ does not depend on t and thus the Markov chain is time-homogeneous. Denote the $mn \times mn$ transition probability matrix by Π with $\pi(i,j;k,l)$ as its entry. It should be pointed out that Π is a highly sparse matrix by the nature of the problem. (Sparsity is in part obtained by the fact that theoretically there are at most m nonzero elements in any row due to the GARCH structure and in part due to our implementation of treating any value less than 10^{-6} as 0. We always normalize each row to have the row sum equal to 1.) Let the mn-dimensional value vector at time t be V_t. Denote the contingent payoff function by $f(S_t)$ and the mn-dimensional payoff vector corresponding to m values of x_{1i} by F_t. (Note that the payoff function only depends on price not volatility. Thus, the mn-dimensional payoff factor consists of the m-dimensional vector repeated for n times.) In other words, the element of F_t corresponding to x_{1i} is always $f(\exp(x_{1i}+(r-d)t))$. At maturity, $V_T = F_T$. The following recursive system can be used to value American style options:

$$V_t = \max\left\{F_t, e^{-r}\Pi V_{t+1}\right\}.$$

For European style options, it can be simplified to $V_t = e^{-r(T-t)}\Pi^{T-t}F_T$. We take the option value corresponding to the current stock price and volatility as the linearly interpolated value using two elements of V_t where both of them correspond to the current stock price but one has a volatility immediately above the current volatility level and the other immediately below it. The reason for performing interpolation is due to the fact that the logarithmic current volatility is not necessarily a value in the discretized set because of centering the interval at $\overline{X}_2$.

In our implementation, we let $x_{1g} = 3\psi_1\sqrt{T-t}$ and $x_{2g} = 3\psi_2$ where ψ_1 is the standard deviation of the sample $\{R_i; i = 1, \cdots N\}$ and ψ_2 is the standard deviation of estimated values of $\ln(\sigma_i^2)$ using the GARCH model to filter the sample $\{R_i; i = 1, \cdots N\}$. Note that we take the sample mean of $\ln(\sigma_i^2)$ as $\overline{X}_2$. We do not multiply ψ_2 by $\sqrt{T-t}$ because the volatility system is stationary and its variability does not grow with time. We also need to take into account the possibility that the logarithmic initial volatility may be outside the range defined around its stationary value. The value for x_{iU} and x_{iL} $(i = 1, 2)$ are thus determined by

$$\begin{aligned} x_{1U} &= X_{10} + x_{1g}\sqrt{\frac{m-1}{100}}, \quad x_{1L} = X_{10} - x_{1g}\sqrt{\frac{m-1}{100}}, \\ x_{2U} &= \max\left\{X_{20}, \overline{X}_2 + x_{2g}\sqrt{\frac{n-1}{50}}\right\}, \quad x_{2L} = \min\left\{X_{20}, \overline{X}_2 - x_{2g}\sqrt{\frac{n-1}{50}}\right\}. \end{aligned}$$

Note that we use $m = 101$ and $n = 51$ as the base case. Such formulas can ensure that the partition condition given in Duan and Simonato (2001) is satisfied so that the approximation algorithm will converge to the right theoretical value as both m and n tend to infinity. The results reported in the paper are computed using $m = 501$ and $n = 101$.

References

[1] Aït-Sahalia, Y. and A. Lo, 1998, "Nonparametric estimation of state-price densities implicit in financial asset prices", *Journal of Finance 53, 499-547.*

[2] Andersen, L. and M. Broadie, 2001, "A primal-dual simulation algorithm for pricing multi-dimensional American options", Columbia University working paper.

[3] Balkema, A. and L. de Haan, 1974, "Residual life time at great age", *Annals of Probability* 2, 792-804.

[4] Black, F. and M. Scholes, 1973, "The pricing of options and corporate liabilities", *Journal of Political Economy 81, 637-659.*

[5] Broadie, M., J. Detemple, E. Ghysels and O. Torres, 2000, “American options with stochastic volatility and stochastic dividends: a nonparametric investigation”, *Journal of Econometrics* 94, 53-92.

[6] Buchen, P. and M. Kelly, 1996, “The maximum entropy distribution of an asset inferred from option prices”, *Journal of Financial and Quantitative Analysis* 31, 143-159.

[7] Carriere, J., 1996, “Valuation of the early-exercise price for derivative security using simulations and splines”, *Insurance: Mathematics and Economics* 19, 19-30.

[8] Duan, J.-C., 1995, “The GARCH option pricing model,” *Mathematical Finance* 5, 13-32.

[9] Duan, J.-C., 1999, “Conditionally fat-tailed distributions and the volatility smile in options”, Hong Kong University of Science and Technology working paper.

[10] Duan, J.-C., E. Dudley, G. Gauthier and J.G. Simonato, 1999, “Pricing discretely monitored barrier options by a Markov chain” Hong Kong University of Science and Technology working paper.

[11] Duan, J.-C. and J.-G. Simonato, 2001, “American option pricing under GARCH by a Markov chain approximation”, *Journal of Economic Dynamics and Control* 25, 1689-1718.

[12] Garcia, R. and R. Gencay, 2001, “Pricing and hedging derivative securities with neural networks and a homogeneity hint”, *Journal of Econometrics* 94, 93-115.

[13] Hutchinson, J., A. Lo and T. Poggio, 1994, “A nonparametric approach to the pricing and hedging of derivative securities via learning networks”, *Journal of Finance* 49, 851-889.

[14] Jackwerth, J. and M. Rubinstein, 1996, “Recovering probability distributions from option prices”, *Journal of Finance* 51, 1611-1631.

[15] Jacquier E. and R. Jarrow, 2000, “Bayesian analysis of contingent claim model error”, *Journal of Econometrics 94, 145-180.*

[16] Longstaff, F. and E. Schwartz, 2001, “Valuing American options by simulation: a simple least squares approach”, *Review of Financial Studies* 14, 113-148.

[17] Pickands, J., 1975, “Statistical inference using extreme value order statistics”, *Annals of Statistics* 3, 119-131.

[18] Rogers, C., 2001, “Monte Carlo valuation of American options”, University of Bath working paper.

[19] Rubinstein, M., 1994, “Implied binomial trees”, *Journal of Finance* 49, 771-818.

[20] Stutzer, M., 1996, “A simple nonparametric approach to derivative security valuation”, *Journal of Finance* 51, 1633-1652.

[21] Tsitsiklis, J. and B. van Roy, 1999, “Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives”, *IEEE Transactions on Automatic Control* 44, 1840-1851.

Table 1. A comparison of the nonparametric method with the Black-Scholes model using 252 simulated returns. The following parameter values are used: $\mu = 0.1$ and $\sigma = 0.15$ and $r = 0.05$. Maturity in terms of trading days is determined by using 252 trading days per year and rounding it to the nearest integer. Statistics are computed using 200 simulation runs.

T = 1 month	**Strike price/stock price**				
	0.90	0.95	1.00	1.05	1.10
Black-Scholes (True parameter)	10.3817	5.5947	1.9396	0.3479	0.0289
NP (i.i.d.; N = 252)					
Mean	10.3843	5.6152	1.9786	0.3680	0.0338
Std	0.1178	0.1138	0.1139	0.0615	0.0125
Median	10.3920	5.6280	1.9876	0.3740	0.0340
Min	9.8420	5.0962	1.6090	0.2270	0.0104
Max	10.7723	5.9291	2.2735	0.5435	0.0794
Black-Scholes (Est. parameter; N = 252)					
Mean	10.3825	5.5995	1.9475	0.3539	0.0308
Std	0.0035	0.0360	0.0831	0.0504	0.0099
Median	10.3823	5.6013	1.9548	0.3571	0.0307
Min	10.3764	5.5152	1.7340	0.2319	0.0111
Max	10.3979	5.7223	2.2045	0.5200	0.0704
T = 6 months	**Strike price/stock price**				
	0.90	0.95	1.00	1.05	1.10
Black-Scholes (True parameter)	12.7467	8.7561	5.5271	3.1837	1.6694
NP (i.i.d.; N = 252)					
Mean	12.7987	8.8356	5.6243	3.2810	1.7528
Std	0.6476	0.5803	0.4891	0.3824	0.2733
Median	12.8541	8.8937	5.6696	3.3207	1.7805
Min	9.8909	6.3083	3.6301	1.8720	0.8646
Max	14.8741	10.5822	7.0186	4.3906	2.5767
Black-Scholes (Est. parameter; N = 252)					
Mean	12.7585	8.7723	5.5459	3.2030	1.6873
Std	0.0904	0.1522	0.1967	0.2021	0.1703
Median	12.7632	8.7839	5.5629	3.2205	1.7005
Min	12.5450	8.3924	5.0421	2.6852	1.2603
Max	13.0646	9.2576	6.1555	3.8291	2.2276

Table 2. A comparison of the nonparametric method with the Black-Scholes model using 1260 simulated returns. The following parameter values are used: $\mu = 0.1$ and $\sigma = 0.15$ and $r = 0.05$. Maturity in terms of trading days is determined by using 252 trading days per year and rounding it to the nearest integer. Statistics are computed using 200 simulation runs.

T = 1 month	**Strike price/stock price**				
	0.90	0.95	1.00	1.05	1.10
Black-Scholes (True parameter)	10.3817	5.5947	1.9396	0.3479	0.0289
NP (i.i.d.; N = 1260)					
Mean	10.3902	5.6231	1.9914	0.3747	0.0346
Std	0.0241	0.0271	0.0406	0.0246	0.0053
Median	10.3906	5.6235	1.9939	0.3754	0.0347
Min	10.2267	5.4413	1.8061	0.2829	0.0179
Max	10.5411	5.7496	2.0979	0.4410	0.0487
Black-Scholes (Est. parameter; N = 1260)					
Mean	10.3819	5.5959	1.9417	0.3494	0.0294
Std	0.0015	0.0164	0.0379	0.0229	0.0044
Median	10.3819	5.5962	1.9432	0.3501	0.0293
Min	10.3782	5.5490	1.8275	0.2825	0.0178
Max	10.3884	5.6562	2.0739	0.4324	0.0471
T = 6 months	**Strike price/stock price**				
	0.90	0.95	1.00	1.05	1.10
Black-Scholes (True parameter)	12.7467	8.7561	5.5271	3.1837	1.6694
NP (i.i.d.; N = 1260)					
Mean	12.8338	8.8728	5.6612	3.3127	1.7753
Std	0.1371	0.1334	0.1277	0.1136	0.0899
Median	12.8350	8.8793	5.6673	3.3152	1.7778
Min	11.8658	7.9520	4.8506	2.6699	1.3226
Max	13.6547	9.5823	6.2235	3.7171	2.0385
Black-Scholes (Est. parameter; N = 1260)					
Mean	12.7497	8.7603	5.5322	3.1889	1.6741
Std	0.0411	0.0693	0.0896	0.0921	0.0776
Median	12.7506	8.7627	5.5356	3.1924	1.6768
Min	12.6313	8.5545	5.2624	2.9117	1.4433
Max	12.9006	9.0065	5.8454	3.5106	1.9488

Table 3. American put prices based on the nonparametric method (under the i.i.d. assumption) and the Black-Scholes model. The data set contains daily S&P 500 index excess returns from the first trading day of 1996 to the last trading day of 2000 totaling 1263 data points. The current stock price is 100 and $r = 0.057231$. The sample standard deviation equals 0.184390836. Maturity in terms of trading days is determined by using 252 trading days per year and rounding it to the nearest integer.

Maturity	Strike price				
	90	95	100	105	110
NP (i.i.d.; N = 1263)					
1 month	0.0331	0.3675	1.8665	5.2233	10.0000
2 months	0.1815	0.8363	2.5380	5.6325	10.0000
3 months	0.3749	1.2227	3.0181	5.9821	10.0691
4 months	0.5722	1.5530	3.3986	6.2802	10.1885
5 months	0.7623	1.8372	3.7151	6.5377	10.3159
6 months	0.9429	2.0884	3.9904	6.7673	10.4430
Black-Scholes (Est. parameter; N = 1263)					
1 month	0.0380	0.3959	1.9055	5.2263	10.0000
2 months	0.1996	0.8725	2.5672	5.6271	10.0000
3 months	0.3994	1.2552	3.0330	5.9627	10.0454
4 months	0.5974	1.5767	3.3978	6.2444	10.1496
5 months	0.7844	1.8499	3.6981	6.4854	10.2605
6 months	0.9593	2.0887	3.9568	6.6984	10.3706

Table 4. American put prices based on the nonparametric method under the dynamic model assumption. The data set contains daily S&P 500 index excess returns from the first trading day of 1996 to the last trading day of 2000 totaling 1263 data points. The current stock price is 100 and $r = 0.057231$. Maturity in terms of trading days is determined by using 252 trading days per year and rounding it to the nearest integer.

Maturity	Strike price				
	90	95	100	105	110
NP (initial volatility = 0.22237484)					
1 month	0.2763	0.7668	2.1014	5.1576	10.0000
2 months	0.6346	1.3367	2.7929	5.5510	10.0000
3 months	0.9358	1.7744	3.3113	5.9379	10.0000
4 months	1.2060	2.1484	3.7451	6.2953	10.0662
5 months	1.4544	2.4797	4.1215	6.6214	10.2028
6 months	1.6852	2.7777	4.4557	6.9200	10.3619
NP (initial volatility = 0.35018012)					
1 month	0.7007	1.4882	3.0297	5.7944	10.0000
2 months	1.1678	2.1060	3.7065	6.3043	10.1302
3 months	1.4716	2.5044	4.1608	6.6997	10.3143
4 months	1.7240	2.8338	4.5379	7.0437	10.5137
5 months	1.9527	3.1267	4.8687	7.3515	10.7148
6 months	2.1652	3.3926	5.1656	7.6310	10.9112

Figure 1a. The empirical distribution and its smoothed version for a simulated sample of 252 standard normal random variates.

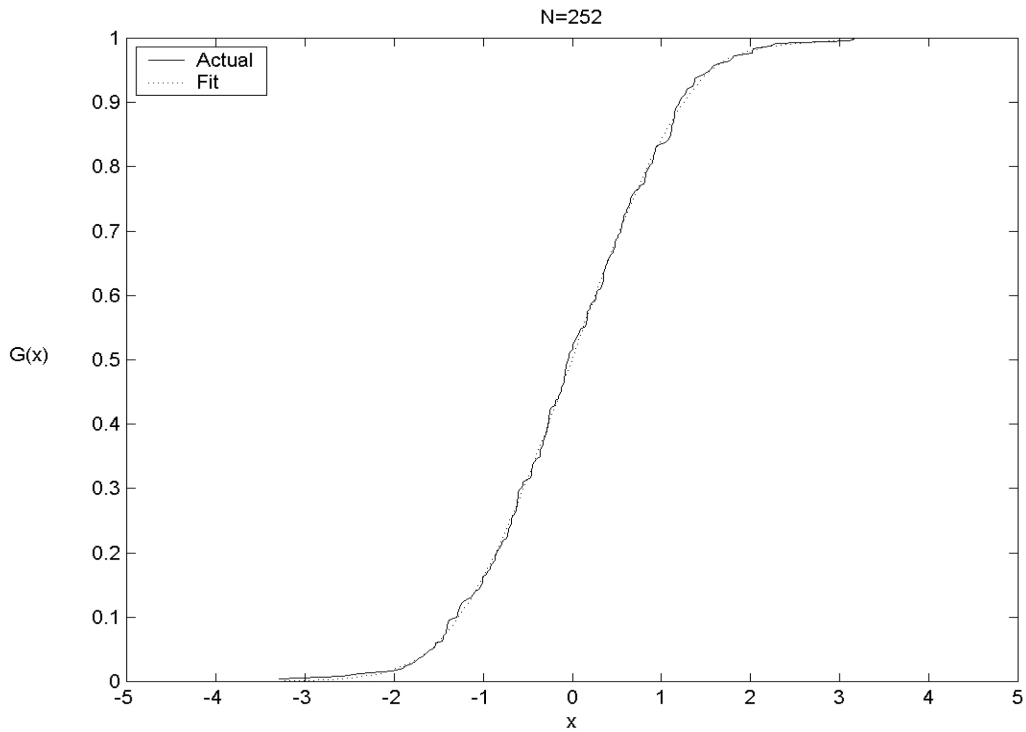

Figure 1b. The empirical distribution and its smoothed version for a simulated sample of 1260 standard normal random variates.

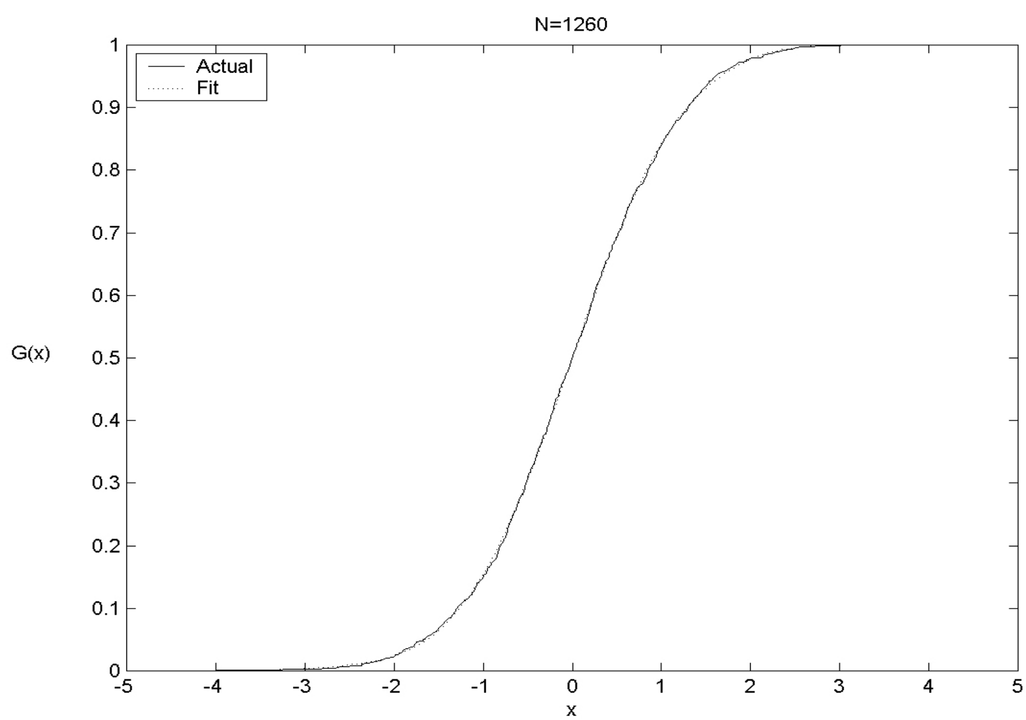

Figure 1c. The empirical distribution and its smoothed version for the S&P 500 index return data (standardized).

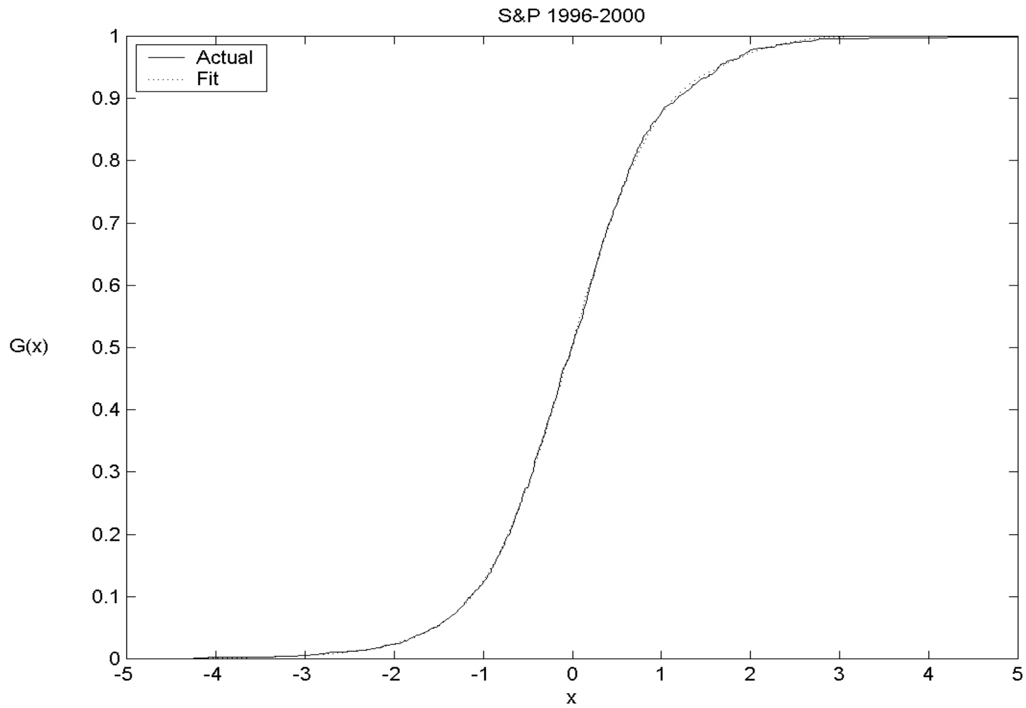

Figure 2. The implied volatility surface of the European call option values computed with the nonparametric pricing method (under the i.i.d. assumption). The data set contains daily S&P 500 index excess returns from the first trading day of 1996 to the last trading day of 2000 totaling 1263 data points. Maturity is stated in fractions of one year and stock-to-strike price ratio is used to represent moneyness.

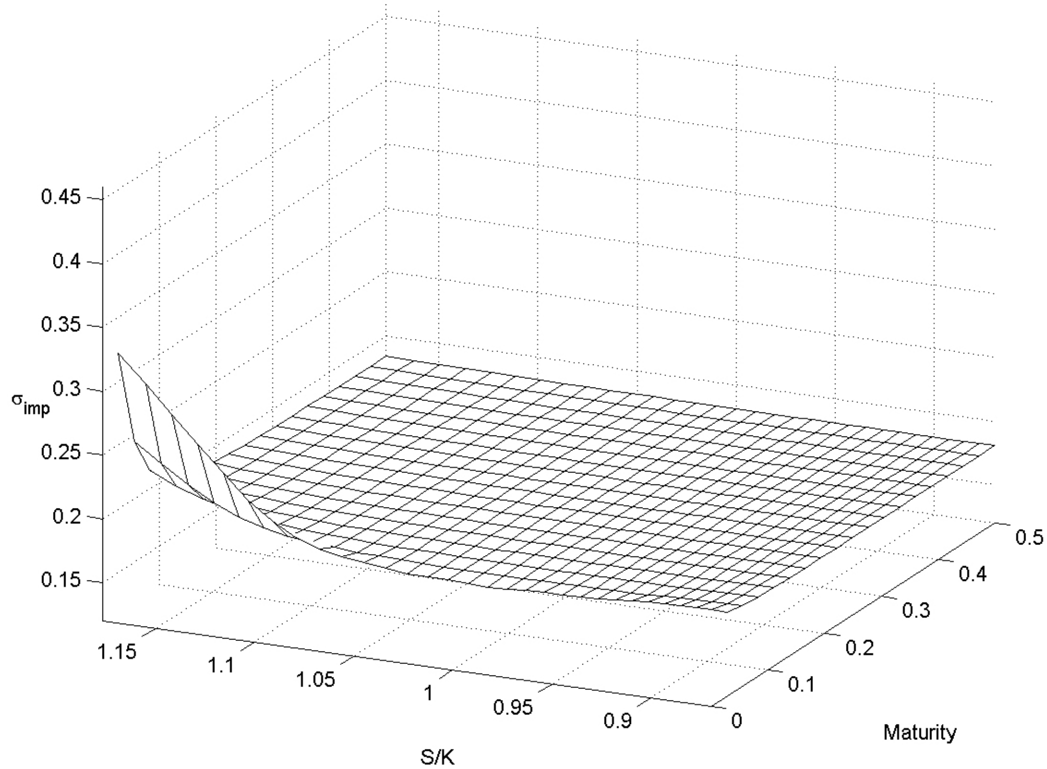

Figure 3. The implied volatility surface of the European call option values computed with the nonparametric pricing method under the dynamic model assumption using the initial volatility of 0.22237484. The data set contains daily S&P 500 index excess returns from the first trading day of 1996 to the last trading day of 2000 totaling 1263 data points. Maturity is stated in fractions of one year and stock-to-strike price ratio is used to represent moneyness.

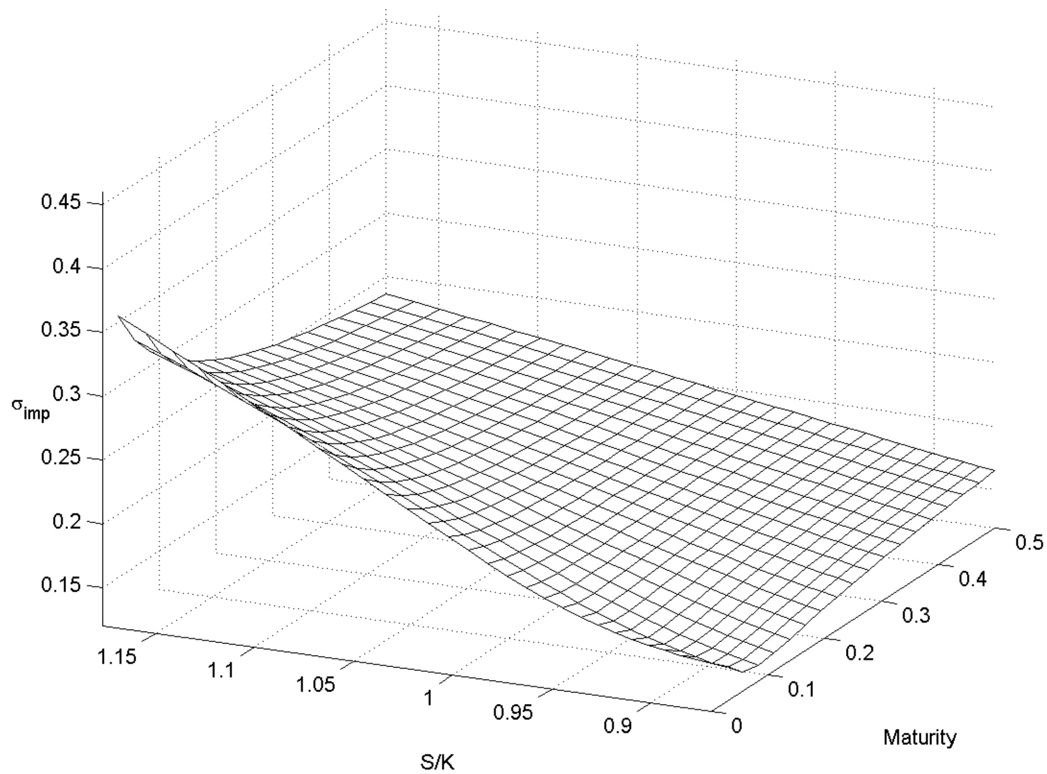

Figure 4. The implied volatility surface of the European call option values computed with the nonparametric pricing method under the dynamic model assumption using the initial volatility of 0.35018012. The data set contains daily S&P 500 index excess returns from the first trading day of 1996 to the last trading day of 2000 totaling 1263 data points. Maturity is stated in fractions of one year and stock-to-strike price ratio is used to represent moneyness.