Model Risk for European-Style Stock Index Options

Ramazan Gençay and Rajna Gibson

Abstract—**In empirical modeling, there have been two strands for pricing in the options literature, namely the parametric and nonparametric models. Often, the support for the nonparametric methods is based on a benchmark such as the Black–Scholes (BS) model with constant volatility. In this paper, we study the stochastic volatility (SV) and stochastic volatility random jump (SVJ) models as parametric benchmarks against feedforward neural network (FNN) models, a class of neural network models. Our choice for FNN models is due to their well-studied universal approximation properties of an unknown function and its partial derivatives. Since the partial derivatives of an option pricing formula are risk pricing tools, an accurate estimation of the unknown option pricing function is essential for pricing and hedging. Our findings indicate that FNN models offer themselves as robust option pricing tools, over their sophisticated parametric counterparts in predictive settings. There are two routes to explain the superiority of FNN models over the parametric models in forecast settings. These are nonnormality of return distributions and adaptive learning.**

Index Terms—**Extreme tail events, feedforward neural networks (FNNs), nonparametric methods, option pricing, risk exposure.**

I. Introduction

THERE has been two strands for pricing in empirical options literature, namely the parametric and nonparametric models. Although each has its own merit and limitations, there has not been clear evidence as to whether one should be preferred over the other based on predictive performance. Often, the evidence favored for the nonparametric methods is based on the Black–Scholes (BS) model with constant volatility. In this paper, we examine the best performing parametric models against nonparametric alternatives. In particular, we study the stochastic volatility (SV) and stochastic volatility random jump (SVJ) models as parametric benchmarks against feedforward neural network (FNNs) models, a class of neural network models.

From the parametric pricing literature, it is widely accepted that the SV, SVJ, and stochastic interest rate (SI) models improve on the basic BS model with equity indices.[1] A common motivation for these models is the evidence that the basic BS model exhibits strong bias across moneyness—the "smile effect." The pricing errors of the BS model are particularly large for the deep out-of-the-money puts and calls. How do the SV, SVJ, and SI models improve on the BS model? The SV model offers a flexible distributional structure where the correlation between volatility shocks and underlying stock returns serves to control the level of skewness and the allowance of volatility variation serves to control the excess kurtosis. On the other hand, since the SV is modeled as a diffusion with continuous sample paths, its ability to internalize short term kurtosis and, therefore, to price short term options accurately is limited. The discontinuous jumps and crashes show themselves as negative skewness and high kurtosis in the option price data. SVJ models internalize these jumps and crashes such that these negative skewness and high kurtosis are accounted for, which lead to more accurate short term to medium term option pricing models. The SI models are intended to improve on the valuation and discounting of future payoffs. The SI models do not necessarily offer a more flexible return distribution, or reduce cross-sectional pricing biases but improve pricing performance across option maturity. The comparison between these models made by Bakshi *et al.* [2] indicates that all three specifications, SV, SVJ, and SI models improve on the BS model substantially in terms of pricing accuracy for the medium to long-term options. However, all three models exhibit some U-shaped moneyness-related biases for the short-term options. This indicates the lack of robustness of these specifications for short-term options.

Overall, the empirical evidence for the parametric models indicates that SV is of the first-order importance in improving upon the BS formula. Adding the random-jump feature to the SV model further improves the pricing performance of short term options. The addition of the SI feature improves the pricing and hedging of the long-term options. In light of this evidence, one can make a case that SV with a jump component should be an essential part of an option pricing model to eliminate smile effect across moneyness. The challenge still remains though, of how to deal with the smile effects for the short-term options which cannot be cured with the SV, SVJ, or SVI specifications.

As for the nonparametric literature, there have been several attempts to demonstrate the superior performance of nonparametric models against parametric alternatives. Typically, the parametric alternatives have been limited to basic forms of parametric option pricing models that only allow for constant forms of volatility. Under such settings, the nonparametric alternatives surpass the predictive performance of a benchmark parametric alternative. Such comparisons are based on nonparametric models which do not utilize any information on conditional volatility, although they have the advantage of

Manuscript received August 9, 2005; revised March 16, 2006. This work was supported by the National Centre of Competence by the Research "Conceptual Issues in Financial Risk Management" (NCCR-FINRISK), Natural Sciences and Engineering Research Council of Canada, and the Social Sciences and Humanities Research Council of Canada.

R. Gençay is with the Department of Economics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (e-mail: gencay@sfu.ca).

R. Gibson is with the National Centre of Competence by the Research (NCCR) FINRISK, Swiss Banking Institute, University of Zurich, Zurich 8032, Switzerland (e-mail: rgibson@isb.unizh.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNN.2006.883005

[1] Bakshi *et al.* [2] provide an extensive examination of the parametric pricing models with equity indices.

a flexible functional form. The parametric benchmark has a fixed BS functional with historical volatility. Under such settings, nonparametric alternatives outplay the simple parametric benchmarks due to the restrictive cumulative normal shapes imposed by parametric benchmarks, whereas nonparametric alternatives exploit their flexibility by allowing more general cumulative distributions.[2]

In implementation and interpretation, both frameworks pose substantial difficulties. It is quite a difficult task to find starting values of SV and SVJ models with local nonlinear optimizers. Often these models explode or converge to highly undesirable local solutions. If the data set has not been studied earlier, and, hence, no guidance is available for plausible initial parameter values, it may take several weeks to find such suitable starting values for SV and SVJ models. This is often not reported and it is an overly underestimated research area that deserves careful scrutiny. Nonparametric models, such as FNNs, on the other hand, are subject to the selection of a network architecture and the selection of an appropriate training set, so that predictive performance is not compromised with in-sample fitting. These are equally difficult tasks and without any care taken, it is easy to converge to those solutions where the network does well in-sample but does not generalize in the out-of-sample. On interpretation, nonparametric models are seen as blackbox approaches due to their immediate lack of parametric transparency. Although this view has its own merits, making easily unjustifiable parametric assumptions also suffers from similar drawbacks. Certain forms of SV or SVJ, for instance, may not be easily interpreted, although they do have explicit parametric specifications.

With these difficulties in perspective, we investigate the predictive performance of the best performing parametric models versus nonparametric alternatives. The nonparametric models may have the advantage of dealing with jumps and crashes and, therefore, deal with negative skewness and kurtosis more effectively relative to the parametric alternatives. Our preferred nonparametric model is FNNs, due to its well-studied universal approximation property of an unknown function and its partial derivatives. Since the partial derivatives of an option pricing formula are risk pricing tools, an accurate estimation of the unknown option pricing function facilitates precise hedging tools within this universality principle.

To augment the FNN with volatility information, the generalized autoregressive conditional heteroskedastic (GARCH)(1,1) volatility proxy is added as an additional input. Our findings indicate that FNNs are serious contenders in a predictive setting for pricing options. In predictive performance, the FNN-G model performs the best, which is followed by the SVJ and SV models across all maturities and moneyness levels. Although FNN models are seen as blackbox approaches, they offer themselves as robust pricing tools over their sophisticated parametric counterparts.

In Section II, the parametric pricing models are briefly explained. Section III details the FNN models and their numerical implementation. The properties of the synchronous option data are presented in Section IV. Section V presents the empirical findings which is followed by conclusions.

II. Parametric Option Pricing Models

A. *Presentation of the Parametric Models*

In an influential paper, Heston [19] introduced characteristic functions to derive closed-form solutions for option valuation. Assuming that the characteristic functions for the underlying probabilities are known analytically, these probabilities can be expressed through Fourier inversion and hence closed-form solutions for option prices can be obtained. A number of authors including Bates [4] and Bakshi *et al.*[2], followed this route to develop sophisticated option pricing models with SV, SIs, and SVJ components. In this section, we follow the Bakshi *et al.* [2] framework.

In a risk neutral economy, the underlying nondividend-paying[3] stock price $S(t)$ and its components are, for any t, given by

$$\frac{dS(t)}{S(t)} = [R(t) - \lambda\mu_J] + \sqrt{V(t)}dw_S(t) + J(t)dq(t) \tag{1}$$

$$dV(t) = [\theta_v - \kappa_v V(t)]\,dt + \sigma_v\sqrt{V(t)}dw_v(t) \tag{2}$$

$$\ln[1 + J(t)] \sim N\left(\ln[1+\mu_J] - 1/2\sigma_J^2, \sigma_J^2\right) \tag{3}$$

where $R(t)$ is the time-t instantaneous spot interest rate; λ is the frequency of jumps per year; $V(t)$ is the diffusion component of stock return variance (conditional on no jump occurring); $w_s(t)$ and $w_v(t)$ are each a standard Brownian motion, with $\mathrm{Cov}_t[dw_S(t), dw_v(t)] \equiv \rho dt$; $J(s)$ is the percentage jump size (conditional on a jump occurring) that is log normally, identically, and independently distributed over time, with unconditional mean μ_J. The standard deviation of $\ln[1+J(t)]$ is σ_J, $q(t)$ is a Poisson jump counter with intensity λ where $\Pr(dq(t) = 1) = \lambda dt$ and $\Pr(dq(t) = 0) = 1 - \lambda dt$. κ_v, θ_v/κ_v, and σ_v are, respectively, the speed of adjustment, long-run mean, and variation coefficient of the diffusion volatility $V(t)$. $q(t)$ and $J(t)$ are uncorrelated with each other or with $w_s(t)$ and $w_v(t)$.

Under the assumed framework, the total return variance can be decomposed into two components

$$\frac{1}{dt}\mathrm{Var}_t\left(\frac{dS(t)}{S(t)}\right) = V(t) + V_J(t) \tag{4}$$

where $V_J(t) \equiv (1/dt)\mathrm{Var}_t[J(t)dq(t)] = \lambda[\mu_J^2 + (e^{\sigma_J^2} - 1)(1+\mu_J)^2]$ is the instantaneous variance of the jump component. The discounting of future cash flows is according to the single-factor term structure model [6]

$$dR(t) = [\theta_R - \kappa_R R(t)]\,dt + \sigma_R\sqrt{R(t)}dw_R(t) \tag{5}$$

where κ_R, θ_R/κ_R, and σ_R are, respectively, the speed of adjustment, long-run mean, and variation coefficient of the $R(t)$

[2]Examples of nonparametric alternatives can be found in [1], [3], [15], [17], [23], [24], [28], and [31].

[3]In our empirical implementation, the dividend-exclusive S&P 500 index series are calculated by first calculating the present value of daily dividends and subtracting the present value of future dividends from the current index level.

process. $w_R(t)$ is the standard Brownian motion, uncorrelated with any other process in the model.

For a European call option written on the stock with strike price K and term-to-expiration τ, its time-t price $C(t,\tau)$ must solve

$$\begin{aligned}\frac{1}{2}VS^2\frac{\partial^2 C}{\partial S^2}+[R-\lambda\mu_J]S\frac{\partial C}{\partial S}+\rho\sigma_v VS\frac{\partial^2 C}{\partial S\partial V}\\ +\frac{1}{2}\sigma_v^2 V\frac{\partial^2 C}{\partial V^2}+[\theta_v-\kappa_v V]\frac{\partial C}{\partial V}+\frac{1}{2}\sigma_R^2 R\frac{\partial^2 C}{\partial R^2}\\ +[\theta_R-\kappa_R R]\frac{\partial C}{\partial R}-\frac{\partial C}{\partial \tau}-RC\\ +\lambda E\left\{C\left(t,\tau,S(1+J),R,V\right)-C(t,\tau,S,R,V)\right\}\\ =0\end{aligned}\tag{6}$$

subject to $C(t+\tau,0)=\max\{S(t+\tau)-K,0\}$. Bakshi *et al.* [2] have shown that

$$C(t,\tau)=S(t)\Pi_1(t,\tau;S,R,V)-KB(t,\tau)\Pi_2(t,\tau;S,R,V)\tag{7}$$

where $B(t,\tau)$ is the current price of a zero-coupon bond that pays $1 in τ periods from time t. The risk-neutral probabilities Π_1 and Π_2 are recovered from inverting the respective characteristic functions

$$\begin{aligned}&\Pi_j\left(t,\tau;S(t),R(t),V(t)\right)\\ &=\frac{1}{2}+\frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{e^{-i\phi\ln[K]}f_j\left(t,\tau,S(t),R(t),V(t);\phi\right)}{i\phi}\right]d\phi\end{aligned}\tag{8}$$

for $j=1$, 2 with the characteristic functions f_j defined in the Appendix.

The European option pricing model in (7) encompasses several of the most studied cases. For example, the BS model is obtained when $\lambda=\theta_R=\kappa_R=\sigma_R=\theta_v=\kappa_v=\sigma_v=0$; the SV model when $\lambda=\theta_R=\kappa_R=\sigma_R=0$; the SI model when $\lambda=\theta_v=\kappa_v=\sigma_v=0$; the SVJ model when $\theta_R=\kappa_R=\sigma_R=0$, and stochastic volatility and stochastic interest rate model (SVSI) when $\lambda=0$.[4]

B. Estimation of SV and SVJ Models

In order to estimate the parameters of (7), we minimize the sum of squared errors (SSE) by choosing

$$\Phi\equiv\{\kappa_v,\theta_v,\sigma_v,\rho,\lambda,\mu_J,\sigma_J,\}\text{ and }V(t)$$
$$\mathrm{SSE}(t)\equiv\min_{V(t),\Phi}\sum_{n=1}^{N}\left(\epsilon_n\left[V(t),\Phi\right]\right)^2\tag{9}$$

for

$$\epsilon_n\left[V(t),\Phi\right]\equiv O_n(t,\tau_n;K_n)-C_n(t,\tau_n;K_n)\tag{10}$$

where N is the number of option prices for a given day and $O_n(t,\tau_n;K_n)$ is the market price of the nth option for day t.

The estimation is repeated for each day (t) which yields the estimated implied spot volatility $\hat{V}(t)$ and the estimated model parameters $\hat{\Phi}$. Here, a note is in order in terms of the estimation of the SV and SVJ models. The choice of the initial parameter values are crucial for these models as they exhibit inherent instability in the parameter surface and can easily explode away. If there is no proper guidance, it is possible to spend considerable time searching for a suitable set of initial parameters without reasonable success. Although nonlinear optimization which constrains the initial parameters within a band is necessary, this is not a guarantee for reaching a desirable local optima.

III. Nonparametric Option Pricing

The BS pricing formula's appeal to practitioners often originates from its analytical simplicity in determining the price of a European option on a nondividend paying asset by

$$C_t=S_tN(d_1)-Ke^{-r\tau}N(d_2)\tag{11}$$

with $d_1=[\ln(S_t/K)+(r+0.5\sigma^2)\tau]/(\sigma\sqrt{\tau})$, $d_2=d_1-\sigma\sqrt{\tau}$ where N is the cumulative normal distribution, S_t is the price of the underlying security, K is the exercise price, r is the prevailing risk-free interest rate, τ is the time-to-maturity, and σ is the volatility of the underlying asset. Equation (11) contains neither preferences of individuals nor the preferences of the aggregate market.

The BS derivation has been mostly criticized for its distributional assumptions of the underlying security.[5] Nonparametric valuation models are a natural extension, as it is easier to relax the distributional assumptions. A natural nonparametric function[6] for pricing a European call option on a nondividend paying asset will relate the price of the option to the set of variables which characterize the option

$$C_t=f(S_t,K,\sigma_t,r_t,\tau)\tag{12}$$

where S_t is the price of the underlying asset, K is the strike price, σ_t is the volatility of the underlying asset, r_t is the interest

[4]Since the models based on interest rates do not provide significant gains over the SVJ model, we only report our findings with SV and SVJ benchmarks.

[5]Empirical studies of stock prices find too many outliers for a simple constant variance log-normal distribution [26]. Bakshi *et al.* [2] provide closed form solutions for valuing options under SV and SIs using Heston's [19] Fourier inversion method to calculate volatility and interest rate market risk premiums. Their results document that SV and SI models are structurally misspecified. However, adding the SV/SVJ features to the BS model improves out-of-sample pricing and hedging performance of the model. In [27], Sarwar and Krehbiel report that the BS model calculated with daily revised implied volatilities performs as well as the SV model for European currency call options. Derman and Kani [8], [9] and Dupire [11] develop a deterministic volatility function (DVF) option valuation model in an attempt to exactly explain the observed cross-section of option prices. However, Dumas *et al.* [10] report that the DVF option valuation model's fit is no better than an ad hoc procedure that merely smooths BS implied volatilities across exercise prices and time-to-maturity.

[6]In [15], [17], and [31], the homogeneity hint was used to construct a more parsimonious model and to bound the stock price process. In this paper, our goal is to identify the performance of the FNNs with GARCH volatility relative to sophisticated SV models. Addition of the homogeneity hint makes the identification of gains more difficult whether the gains are originating from the GARCH effect or the homogeneity hint. For the nonstationarity without the homogeneity hint, this is not a major constraint because our goal is to examine forecasting performance rather than inference.

rate, and τ is the time-to-maturity. The unknown function f is estimated by a single layer FNN as described later.

A. FNNs

An artificial neural network is a parallel distributed statistical model made up of simple data processing units, which processes information in currently available data, and makes generalizations for future events. The results in [7], [13], [21], [22], and [20] indicate that FNNs with sufficiently many hidden units and properly adjusted parameters can approximate an arbitrary function arbitrarily well. Hornik *et al.* [22] and Hornik [20] further show that the FNNs can also approximate the derivatives of an arbitrary function. The universal approximation property in which both the unknown function and its derivatives can be uncovered from the data is an important result theoretically and has immediate implications for financial and economic modeling.

In options pricing, for instance, Hutchinson *et al.* [23] and Garcia and Gençay [15] demonstrate that FNNs can be used successfully to estimate a pricing formula for options, with good out-of-sample pricing and delta-hedging performance. In the option pricing framework, it is crucial to approximate both the function and the derivatives of the function accurately as the derivatives of the option pricing formula are the risk management tools (e.g., delta, gamma of an option). A small function approximation error may lead to larger errors in the derivatives of the function and, therefore, poorly approximated risk management tools. Garcia and Gençay [15] and Gençay and Qi [17] show that FNNs provide great enhancements over the BS model in terms of providing more accurate pricing and hedging performances with equity indices.

B. Network Architecture

Let $\mathbf{x}_t$ and y_t be the input (regressors) and the target (regressand) vectors with dimensions $1 \times n$ and $1 \times w$ with t indicating the time index.[7] The observations for a sample size N are denoted by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ and $y_1, y_2, \ldots, y_N$. Given inputs $\mathbf{x}_t = (x_{1,t}, \ldots, x_{n,t})$, a single layer FNN regression model with q hidden units is written as

$$\begin{aligned} y_t &= s\left(\beta_0 + \sum_{i=1}^{q} \beta_i h_{i,t}\right) + \epsilon_t \\ h_{i,t} &= g\left(\alpha_{i0} + \sum_{j=1}^{n} \alpha_{ij} x_{j,t}\right) \end{aligned} \quad (13)$$

for $i = 1, \ldots, q$ or

$$\begin{aligned} y_t &= s\left[\beta_0 + \sum_{i=1}^{q} \beta_i g\left(\alpha_{i0} + \sum_{j=1}^{n} \alpha_{ij} x_{j,t}\right)\right] + \epsilon_t \\ &= f(x_t, \theta) + \epsilon_t \end{aligned} \quad (14)$$

where s and g are known activation functions; ϵ_t is an error term distributed with zero mean and variance σ_t^2, and the parameters to be estimated are $\theta = (\beta_0, \ldots, \beta_q, \alpha_1, \ldots, \alpha_q)'$ and $\alpha_j = (\alpha_{j,0}, \ldots, \alpha_{j,n})$. The range of the output values of the FNN model is controlled by s such that if the output takes discrete values, then s can be chosen to be a threshold function, piecewise linear function, or a signum function. If the range of the output function is not restricted to a particular interval, then it can simply be set to an identity function, where $s(x) = x$. In a typical neural network model, s is normally an identity function.

[7]For simplicity, we will assume that $w = 1$ here.

This paper uses the nonlinear least squares (NLS) estimator which minimizes

$$\min_{\theta} \ L(\theta) = \sum_{t=1}^{N} \left[y_t - f(\mathbf{x}_t, \theta)\right]^2. \quad (15)$$

Here, the goal is to choose the parameter vector θ such that the sum of squared errors is minimized as much as possible. Since the function f is nonlinear (a neural network model) and it is a nonlinear function of θ, this procedure is named as nonlinear least squares or nonlinear regression. This is a multivariate minimization problem and conjugate gradient routines studied in [16] work very well for this problem. In [14], it is shown that the least squares method can consistently estimate a function and its derivatives from an FNN model, provided that the number of hidden units increases with the size of the data set. This would mean that a larger number of data points would require a larger number of hidden units to avoid overfitting in noisy environments.

C. Network Selection

The specification of an FNN model requires the choice of the type of inputs, the number of hidden units, the number of hidden layers, and the connection structure between the inputs and the output layers. The common choice for this specification design is to adopt the model-selection approach. Information based criteria such as the Schwarz information criterion (SIC) and the Akaike information criterion (AIC) are used widely. Swanson and White [29] report that the SIC fails to select sufficiently parsimonious models in terms of being a reliable guide to the out-of-sample performance. Since the SIC imposes the more severe penalty than the AIC, the results with AIC would lead to poorer out-of-sample predictions. Gençay and Qi [17] advocated bagging as a parsimonious method of model selection for FNN models. This method is used in this study.

1) Bagging: In bagging (or bootstrap aggregating), multiple versions of a predictor is generated and they are used to get an aggregated predictor. The multiple versions are formed by making bootstrap replicates of the training set and using these as new training sets. When predicting a numerical outcome, the aggregation takes the average over the multiple versions that are generated from bootstrapping. According to Breiman [5], both theoretical and empirical evidence suggests that bagging can greatly improve the forecasting performance of neural network models.

As proposed by Breiman [5], when the neural network output y is numerical, bagging works as follows. Let L represent the training set that consists of data $\{(y_i, X_i), i = 1, \ldots, N_L\}$, where N_L is the number of observations in the training set.

TABLE I
SAMPLE PROPERTIES OF S&P 500 INDEX CALL OPTIONS. THE REPORTED NUMBERS ARE THE AVERAGE QUOTED BID-ASK MIDPOINT PRICE, THE STANDARD DEVIATION OF THE BID-ASK MIDPOINT PRICES (SHOWN IN PARENTHESES), AND THE TOTAL NUMBER OF OBSERVATIONS FOR EACH MONEYNESS-MATURITY CATEGORY. DAILY INFORMATION FROM THE LAST QUOTE (PRIOR TO 3:00 p.m. CST) OF EACH OPTION CONTRACT IS USED TO OBTAIN THE SUMMARY STATISTICS. S DENOTES THE SPOT S&P 500 INDEX LEVEL AND K IS THE EXERCISE PRICE. OTM, ATM, AND ITM DENOTE OUT-OF-THE-MONEY, AT-THE-MONEY, AND IN-THE-MONEY OPTIONS, RESPECTIVELY. THE SAMPLE PERIOD EXTENDS FROM JANUARY 3, 1989 THROUGH DECEMBER 31, 1991, FOR A TOTAL OF 40 719 CALLS

		Days-to-Expiration			
	Moneyness (S/K)	<60	60-180	≥ 180	Subtotal
OTM	< 0.94	\$1.55	\$5.14	\$10.61	
		(0.93)	(2.99)	(4.65)	
		561	2927	2732	6220
	0.94 – 0.97	\$2.54	\$9.69	\$18.76	
		(1.58)	(3.86)	(3.94)	
		2046	2118	1184	5348
ATM	0.97 – 1.00	\$5.29	\$15.29	\$25.40	
		(2.80)	(4.30)	(4.38)	
		2795	2064	1129	5988
	1.00 – 1.03	\$11.38	\$22.01	\$31.96	
		(3.44)	(4.35)	(4.32)	
		2643	1907	965	5515
ITM	1.03 – 1.06	\$19.17	\$28.81	\$38.42	
		(3.48)	(4.32)	(4.50)	
		2438	1688	810	4936
	≥ 1.06	\$41.03	\$50.61	\$64.14	
		(16.62)	(16.11)	(15.02)	
		4999	4932	2781	12712
Subtotal		15482	15636	9601	40719

The neural network model represented by (15) is fitted to the training set and this generates a predictor $f(X, L)$: If the input is X, we predict y by $f(X, L)$. Now, suppose we have a sequence of training sets $\{L_k, k = 1, \ldots, K\}$ each consisting of N_L independent observations from the same underlying distribution as L. We can use the $\{L_k\}$ to get a better predictor than the single learning set predictor $f(X, L)$ by working with the sequence of predictors $\{f(X, L_k)\}$. An obvious procedure is to replace $f(X, L)$ by the average of $f(X, L_k)$ over k, i.e., by $f_A(X) = \sum_{k=1}^{K} f(X, L_k)$. However, usually there is only a single training set L without the luxury of replicates of L. In this case, repeated bootstrap samples $L^{(b)} = \{(y_i^{(b)}, X_i^{(b)}), i = 1, \ldots, N_L\}$ can be drawn from $L = \{(y_i, X_i), i = 1, \ldots, N_L\}$. Each $\{(y_i^{(b)}, X_i^{(b)})\}$ is a random pick from the original training set $\{(y_i), X_i), i = 1, \ldots, N_L\}$ with replacement. The bootstrap samples $L^{(b)}$ are used to form predictors $\{f(X, L^{(b)}\}$. The bagging predictor f_B can thus be calculated as

$$f_B(x) = \sum_{b=1}^{B} f\left(X, L^{(b)}\right) \tag{16}$$

where B represents the total number of bootstrap replicates of the training set.

First, the initial history is divided into the training and validation subsets to predict the first available option price. Second, a bootstrap sample is selected from the training set. The bootstrap sample is then used to train the neural network with 1–10 hidden layer units. The validation set is used to select the best neural network that has the optimal number of hidden layer units, and the best model is used to generate one prediction on the prediction set. This is repeated 25 times giving 25 sets of predictions ($B = 25$).[8] Third, the bagging prediction is calculated as the average across the 25 sets of predictions, and the prediction error is computed as the difference between the actual and the bagging prediction values. Estimations are carried out with a moving sample of 1000 observations. These steps are repeated for every predicted option price, a total of 38 055 predictions.

IV. DATA DESCRIPTION

The data is the intraday bid-ask quotes for S&P 500 index[9] options from the Berkeley Option Database for the period January 3, 1989 to December 31, 1991.[10] The S&P 500 index option market is extremely liquid and it is one of the most active options markets in the U.S. This market is the closest to the theoretical setting of the BS model. The options contracts on this index trade on the Chicago Board Options Exchange; they are actively traded European-style options and the settlements are always in cash. S&P 500 index options are very popular among institutional investors as portfolio insurance instruments.

In the empirical calculations, the last reported bid-ask quote prior to 3:00 pm central standard time (CET) for each day is used. The S&P 500 index values correspond to the index levels at the time of the option data is recorded and there are no nonsynchronous price issues in the data construction. The interest rate data is calculated from the Treasury bill bid-ask discounts with maturities up to one year.[11] In the calculation of the annualized interest rates, the average of the bid-ask quotes are used. For each option written on the S&P 500 index, the data set contains the date of the transaction, time-to-maturity in days, strike price, option price, dividend adjusted S&P 500 index price, and annualized interest rate that matches the maturity of the option.

The statistical properties of the data set are presented in Table I. The data is divided into several categories in terms of moneyness and time-to-maturity. A call option is defined to be out-of-the money (OTM) if $S/K \leq 0.97$, at-the-money (ATM) if $0.97 < S/K < 1.03$ and in-the-money (ITM) if $S/K \geq 1.03$. An option is classified as short-term if $\tau < 60$ days, medium-term if $60 \leq \tau < 180$ days and long-term if

[8]The choice of ($B = 25$) is due to the length of the computational time. The higher number of bootstrap repetitions can be selected if computational resources permit. Further information on bagging can be found in [17].

[9]S&P 500 stock index represents the market value of all outstanding common shares of 500 firms selected by Standard and Poor.

[10]This is the raw data set used in [2] and kindly provided to us by C. Cao. The provided data set is such that the dividend-exclusive S&P 500 index series are calculated by first calculating the present value of daily dividends and subtracting the present value of future dividends from the current index level. Our data set is for the period January 3, 1989 to December 31, 1991 whereas in [2], it is from June 1, 1988 to May 31, 1991.

[11]This data is kindly provided to us by C. Cao and originally hand-collected from the *Wall Street Journal* by H. Choe and S. Freund.

TABLE II
STRUCTURAL PARAMETER ESTIMATES. FOR EACH DAY IN THE SAMPLE, THE STRUCTURAL PARAMETERS OF A GIVEN MODEL ARE ESTIMATED BY MINIMIZING THE SUM OF SQUARED ERRORS BETWEEN THE MARKET PRICE AND THE MODEL. THE REPORTED PARAMETERS ARE THE DAILY AVERAGES WITH THEIR STANDARD ERRORS IN PARENTHESES. SSE REFERS TO THE SUM OF SQUARED ERROR. THE SAMPLE PERIOD EXTENDS FROM JANUARY 3, 1989 THROUGH DECEMBER 31, 1991, FOR A TOTAL OF 40 719 CALLS

Parameters	Stochastic Volatility	Stochastic Volatility (with Jump)
κ_v	1.02 (0.13)	2.13 (0.22)
θ_v	0.04 (0.01)	0.03 (0.02)
σ_v	0.41 (0.08)	0.39 (0.08)
ρ	-0.67 (0.12)	-0.62 (0.11)
λ		0.71 (0.18)
μ_J		-0.06 (0.04)
σ_J		0.02 (0.01)
SSE	21.15	11.01

$\tau \geq 180$ days. The reported numbers are the average quoted bid-ask midpoint price, the standard deviation of the bid-ask midpoint prices (shown in parentheses) and the total number of observations for each moneyness-maturity category. OTM, ATM, and ITM options take approximately 28%, 28%, and 43% of the total sample, respectively. The average prices of call options range from \$1.55 to \$64.14 with standard deviations as large as 15.02. In the OTM category, average option prices range from \$1.55 (short-term) to \$10.61 (long-term).

V. EMPIRICAL RESULTS

A. Parameter Estimation

Before we begin the analysis of call index options, a discussion on the numerical implementation of econometric methods is appropriate. Potential issues are with the choice of initial values for the SV and the SVJ models. In the presence of four parameters for SV and seven parameters for SVJ, finding the appropriate initial parameter values can be a daunting task. If the initial parameters are not chosen appropriately, these models can easily explode. Similarly, the choice of the initial parameters, as well as the length of the training sample are important determinants of the FNN models. Longer training samples often lead to overfitting and poor out-of-sample generalizations. Amongst the SV, SVJ, and FNN models, each has its own difficulty in its numerical implementation and, there is no obvious guidance as to which model could easily be implemented based on numerical stability. Numerical implementation and stability of these models should be subject to further extensive research. Here, the initial parameter values for the optimization are those of Bakshi *et al.* [2]. These initial values quickly lead to convergence, as our data set is a subset of theirs. Otherwise, slight variations from these initial parameter values lead to explosive nonconvergent behavior.

Estimations of the structural parameters and daily spot volatility for the SV and SVJ models are based on all available call options in a given day across all moneyness and maturities. The daily average of these parameters together with their standard errors are reported in Table II. The estimated spot volatility parameter does not differ substantially between the SV and SVJ models (0.41 versus 0.39) with identical standard errors. The variation coefficient for the diffusion volatility σ_v and the correlation coefficient ρ are slightly lower for the SVJ model indicating that the SVJ model relies less on the volatility process $V(t)$ relative to the SV model. The average frequency of jumps in the SVJ model is 0.71 per year, average jump size is -6% with a jump volatility of 2%. The sum of squared errors of the SV and SVJ models are 21.15 and 11.01, respectively. The relative better fit of the SVJ model is attributable to the additional freedom led by the jump component which tackles the skewness and excess kurtosis in the data. Overall, the SVJ model provides a better in-sample fit relative to the SV model and, these results are consistent with Bakshi *et al.* [2].

The out-of-sample predictions for the SV and SVJ models are calculated in two stages. In the first stage, the structural parameters and volatility values are obtained from the data available for day t. These parameter values are, in turn, used to predict the option prices for day $t+1$. This procedure is repeated for every day in the sample in order to calculate the predicted prices and their forecast errors. There are a total of 38 055 predictions from February 16, 1989 through December 31, 1991.

TABLE III
OUT-OF-SAMPLE PERFORMANCE. THE PREDICTIONS ARE BASED ON 38 055 CALL PRICES FROM FEBRUARY 16, 1989 THROUGH DECEMBER 31, 1991

	Absolute Pricing Errors			Squared Pricing Errors		
Maturity	<60	60-180	≥ 180	<60	60-180	≥ 180
			$S/K < 0.94$			
BS-H	0.98	1.47	2.55	1.58	3.89	10.28
BS-G	0.97	1.60	1.80	1.47	4.42	5.94
SV	0.80	0.63	0.85	0.97	1.10	2.10
SVJ	0.53	0.57	0.80	0.49	1.00	2.03
FNN-G	0.35	0.32	0.44	0.25	0.18	0.38
			$0.94 < S/K < 0.97$			
BS-H	0.97	2.17	3.44	1.74	6.80	15.26
BS-G	0.83	1.67	1.68	1.28	5.01	5.56
SV	0.68	0.67	0.82	0.80	0.70	1.00
SVJ	0.49	0.65	0.80	0.44	0.67	1.03
FNN-G	0.29	0.31	0.43	0.16	0.17	0.36
			$0.97 < S/K < 1.00$			
BS-H	1.28	2.91	3.99	2.65	11.24	20.01
BS-G	0.79	1.72	2.11	1.41	6.40	8.41
SV	0.64	0.66	0.80	0.66	0.68	0.97
SVJ	0.51	0.67	0.76	0.40	0.70	0.94
FNN-G	0.30	0.35	0.46	0.17	0.22	0.40
			$1.00 < S/K < 1.03$			
BS-H	1.60	3.48	4.45	3.76	14.67	24.40
BS-G	1.02	2.04	2.78	2.17	8.32	12.05
SV	0.51	0.64	0.78	0.43	0.64	0.88
SVJ	0.45	0.63	0.73	0.34	0.64	0.85
FNN-G	0.35	0.36	0.46	0.22	0.23	0.39
			$1.03 < S/K < 1.06$			
BS-H	1.52	3.53	4.10	3.39	14.65	20.50
BS-G	1.22	2.34	2.58	2.62	9.43	10.18
SV	0.46	0.62	0.71	0.36	0.60	0.81
SVJ	0.40	0.56	0.69	0.28	0.51	0.78
FNN-G	0.36	0.37	0.44	0.23	0.25	0.37
			$S/K > 1.06$			
BS-H	0.99	2.32	3.35	1.66	7.40	14.73
BS-G	0.96	1.91	2.64	1.63	5.92	10.25
SV	0.54	0.56	0.70	0.47	0.52	0.91
SVJ	0.36	0.45	0.60	0.24	0.35	0.60
FNN-G	0.40	0.39	0.50	0.31	0.27	0.48

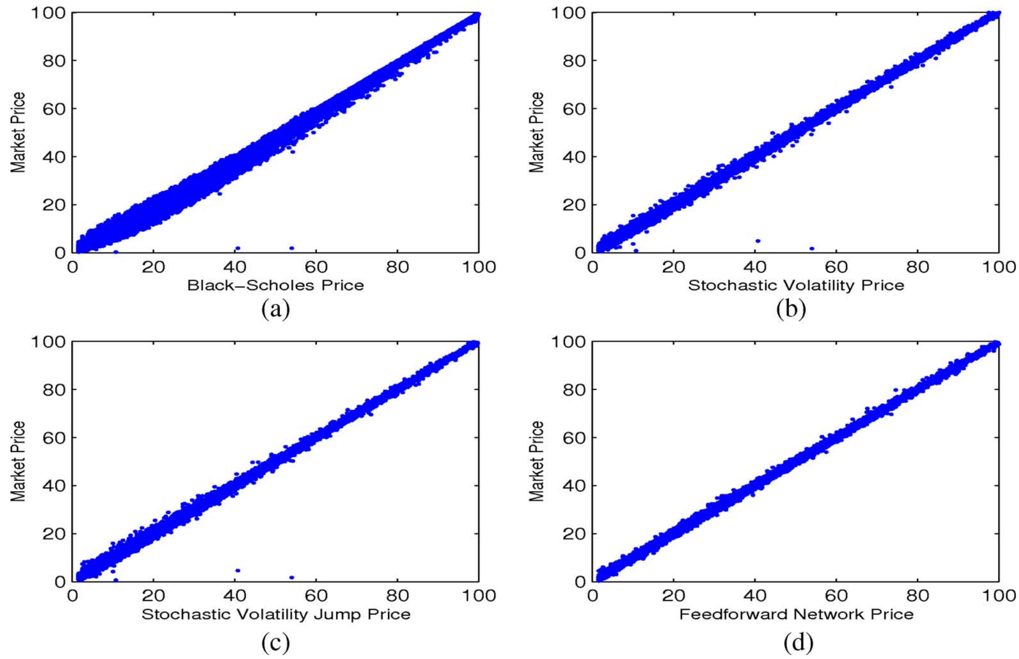

Fig. 1. S&P 500 call prices versus model price predictions for all maturity and moneyness levels. (a) Price predictions based on BS model with GARCH(1,1) volatility. (b) Price predictions based on SV model. (c) Price predictions based on SVJ model. (d) Price predictions based on FNN with GARCH(1,1) volatility. Absolute and squared pricing errors are provided in Table III. The predictions are based on 38 055 call prices from February 16, 1989 through December 31, 1991.

B. Out-of-Sample Evidence of FNNs

The FNN model utilizes the following inputs:

$$C_t = f\left(S_t, K, \hat{\sigma}_{t,\mathrm{GARCH}(1,1)}, r_t, \tau\right) \tag{17}$$

where the estimated volatility is based on the GARCH(1,1) model[12] where the estimate at day t is based on the data available up to day $t-1$ such that days for estimation match the number of days to maturity. S_t is the price of the underlying asset, K is the strike price, r_t is the interest rate, and τ is the time-to-maturity. The unknown function f is estimated by a single layer FNN (FNN-G).[13]

The out-of-sample evidence is reported in terms of average absolute (AAE) and average squared errors (ASE) across all levels of moneyness and maturity categories. The average squared errors magnify larger errors and the impact of these errors is more pronounced in this measure. Our findings are summarized in Table III where the out-of-sample forecast performance of the BS model with historical volatility (BS-H),[14] the BS model with GARCH(1,1) volatility (BS-G), SV, SVJ, and FNN model with the GARCH(1,1) volatility (FNN-G) are reported. The predictions are based on 38 055 call prices from February 16, 1989 through December 31, 1991.

The overall evaluation is that the FNN-G model does uniformly better than all other models across all maturities and moneyness levels. There are a few striking performance differences. Without any surprise, the BS-G performs better than the BS-H model, the gain in performance is small in shorter maturities; however, the discrepancy gets larger in longer maturities. For longer maturities (>180), the AAE performance difference between BS-H and BS-G is as large as 90%. Between the SV and SVJ models, the predictive performance differences are smaller, although the SVJ models do well over the SV models for shorter maturities (<60). The performance improvement for these maturities are as large as 50% between the SV and SVJ models.

When the FNN-G model is compared with the BS-G model, the FNN-G model does well consistently across all maturities and moneyness levels. In particular, the performance difference is striking for longer maturities where the differential in AAE is as large as 300% (for $S/K > 1.06$ and $\tau > 180$). This performance difference is widely reported in the literature. For instance, Hutchinson *et al.* [23] and Garcia and Gençay [15] reported that artificial neural networks outperform the BS parametric benchmark with constant volatility. Although our findings confirm these earlier findings, we also put forward out-of-sample evidence that the FNN-G model does better than the SVJ model. In fact, the predictive performance of the FNN-G model can be up to five times better than that of the SVJ model. In light of this evidence, we conclude that the importance of FNNs in option pricing should not be underestimated.

In Fig. 1, the market call price is plotted against the BS-G, SV, SVJ, and FNN-G price predictions for all moneyness and for

[12]GARCH(1,1) is a parsimonious volatility model. There is common consensus in finance literature of its ability to capture conditional volatility amongst other less parsimonious parametric conditional volatility models. Our goal was to relax constant volatility but at the same time avoid introducing many more implicit parameters to the FNN model. Introduction of additional implicit parameters may be an unfair advantage to the FNN model and by using GARCH(1,1), we maintain a fair comparison relative to the SV models.

[13]This is the FNN model which uses GARCH(1,1) volatility as an input.

[14]For BS price calculations, historical volatilities are calculated using the daily S&P 500 returns. If an option has less than 22 days to expiration, historical volatility is calculated using the last 22 days daily returns. If an option has more than 22 days to maturity then the historical volatility is calculated using the historical returns that match the exact number of days to maturity.

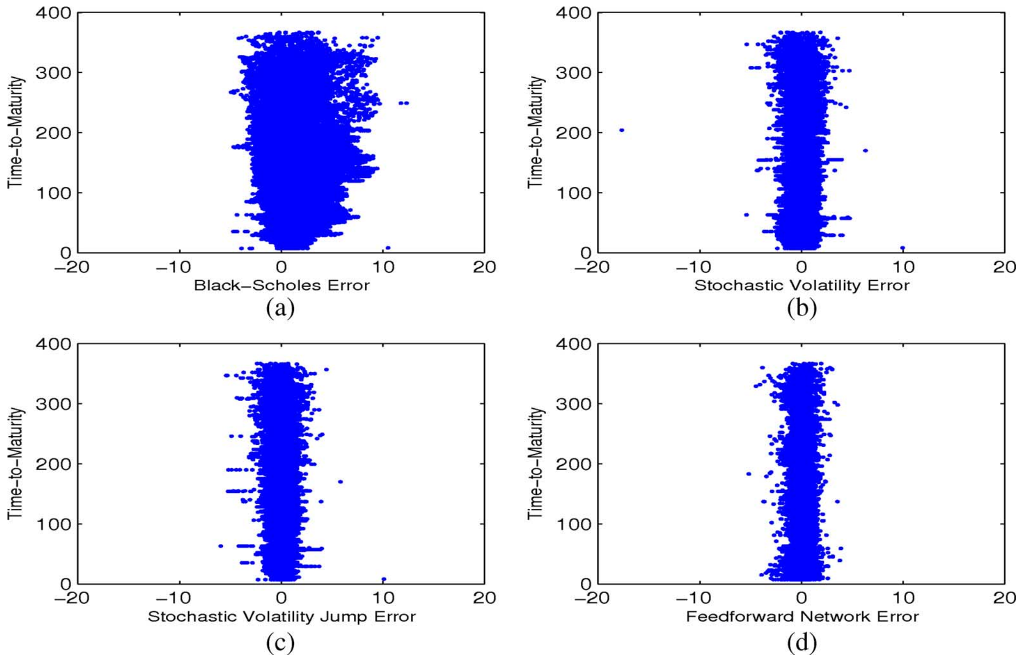

Fig. 2. Time-to-maturity versus model price prediction errors for all maturity and moneyness levels. (a) Price prediction errors based on BS model with GARCH(1,1) volatility. (b) Price prediction errors based on SV model. (c) Price prediction errors based on SVJ model. (d) Price prediction errors based on FNN with GARCH(1,1) volatility.

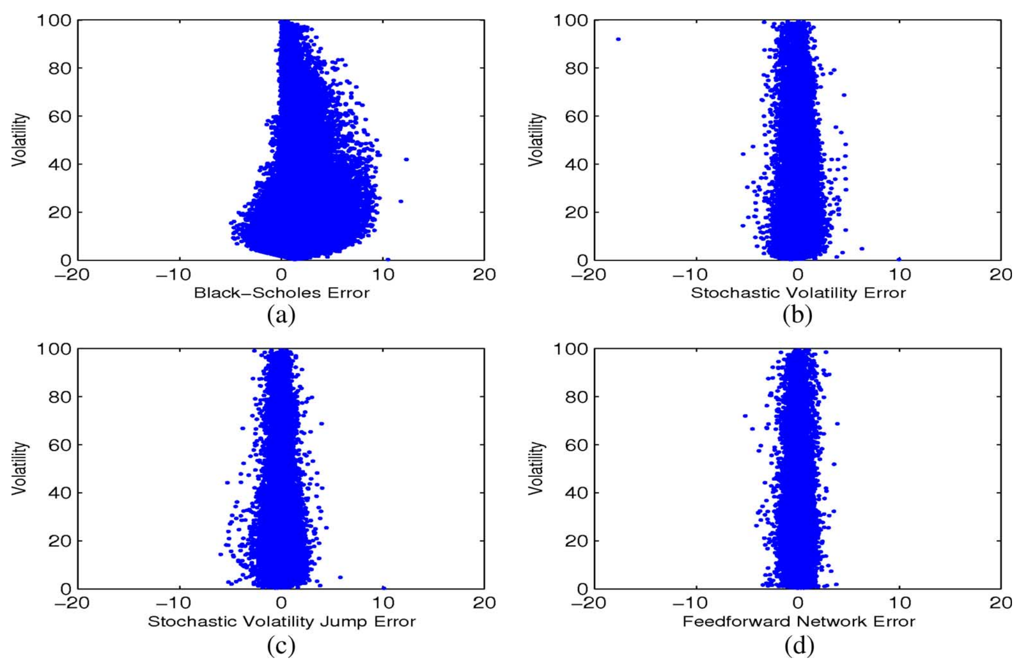

Fig. 3. Volatility versus model price prediction errors for all maturity and moneyness levels. (a) Price prediction errors based on BS model with GARCH(1,1) volatility. (b) Price prediction errors based on SV model. (c) Price prediction errors based on SVJ model. (d) Price prediction errors based on FNN with GARCH(1,1) volatility.

all maturities. Ideal predictions would line up with the market prices on the 45° line. The BS-G model consistently overpredicts, whereas SV, SVJ, and FNN-G predictions are centered around the 45° line. Fig. 2 presents the model prediction errors across maturity. Ideally, the model prediction errors should be centered around zero across all maturities. Although there is no systematic pattern across maturities, the BS-G model prediction errors are large and exhibit excessive tail coverage relative to a normal distribution. All three remaining models limit these tail events to a certain degree. The best performing model is the FNN-G model without excessive tail coverage. In Fig. 3, the model prediction errors are plotted across volatility. The BS-G prediction errors are mainly positive (excess price prediction). Interestingly, BS-G prediction errors are larger for lower levels of volatility and the prediction errors are substantially smaller at large levels of volatility. The most desirable prediction errors are displayed by the FNN-G model with a minimum number of prediction error outliers across volatility.

$$f_1(t,\tau) = \exp\Bigg\{-\frac{\theta_R}{\sigma_R^2}\left[2\ln\left(1-\frac{[\xi_R-\kappa_R](1-e^{-\xi R\tau})}{2\xi_R}\right)+[\xi_R-\kappa_R]\tau\right]-\frac{\theta_v}{\sigma_v^2}\left[2\ln\left(1-\frac{[\xi_v-\kappa_v+(1+i\phi)\rho\sigma_v](1-e^{-\xi v\tau})}{2\xi_v}\right)\right]$$
$$-\frac{\theta_v}{\sigma_v^2}[\xi_v-\kappa_v+(1+i\phi)\rho\sigma_v]\tau+i\phi\ln[S(t)]+\frac{2i\phi(1-e^{-\xi R\tau})}{2\xi_R-[\xi_R-\kappa_R](1-e^{-\xi R\tau})}R(t)+\lambda(1+\mu_J)\tau$$
$$\times\left[(1+\mu_J)^{i\phi}e^{(i\phi/2)(1+i\phi)\sigma_J^2}-1\right]-\lambda i\phi\mu_J\tau+\frac{i\phi(i\phi+1)(1-e^{-\xi v\tau})}{2\xi_v-[\xi_v-\kappa_v+(1+i\phi)\rho\sigma_v](1-e^{-\xi v\tau})}V(t)\Bigg\} \quad \text{(A1)}$$

$$f_2(t,\tau) = \exp\Bigg\{-\frac{\theta_R}{\sigma_R^2}\left[2\ln\left(1-\frac{[\xi_R^*-\kappa_R](1-e^{-\xi^* R\tau})}{2\xi_R^*}\right)+[\xi_R^*-\kappa_R]\tau\right]$$
$$-\frac{\theta_v}{\sigma_v^2}\left[2\ln\left(1-\frac{[\xi_v^*-\kappa_v+i\phi\rho\sigma_v](1-e^{-\xi^* v\tau})}{2\xi_v^*}\right)+[\xi_v^*-\kappa_v+i\phi\rho\sigma_v]\tau\right]+i\phi\ln[S(t)]-\ln[B(t,\tau)]$$
$$+\frac{2(i\phi-1)(1-e^{-\xi^* R\tau})}{2\xi_R^*-[\xi_R^*-\kappa_R](1-e^{-\xi^* R\tau})}R(t)+\lambda\tau\left[(1+\mu_J)^{i\phi}e^{(i\phi/2)(i\phi-1)\sigma_J^2}-1\right]$$
$$-\lambda i\phi\mu_J\tau+\frac{i\phi(i\phi-1)(1-e^{-\xi^* v\tau})}{2\xi_v^*-[\xi_v^*-\kappa_v+i\phi\rho\sigma_v](1-e^{-\xi^* v\tau})}V(t)\Bigg\} \quad \text{(A2)}$$

$$\hat{f}_1 = \exp\Bigg\{-i\phi\ln[B(t,\tau)]-\frac{\theta_v}{\sigma_v^2}\left[2\ln\left(1-\frac{[\xi_v-\kappa_v+(1+i\phi)\rho\sigma_v](1-e^{-\xi v\tau})}{2\xi_v}\right)\right]-\frac{\theta_v}{\sigma_v^2}[\xi_v-\kappa_v+(1+i\phi)\rho\sigma_v]\tau$$
$$+i\phi\ln[S(t)]+\lambda(1+\mu_J)\tau\left[(1+\mu_J)^{i\phi}e^{(i\phi/2)(1+i\phi)\sigma_J^2}-1\right]-\lambda i\phi\mu_J\tau$$
$$+\frac{i\phi(i\phi+1)(1-e^{-\xi v\tau})}{2\xi_v-[\xi_v-\kappa_v+(1+i\phi)\rho\sigma_v](1-e^{-\xi v\tau})}V(t)\Bigg\} \quad \text{(A3)}$$

$$\hat{f}_2 = \exp\Bigg\{-i\phi\ln[B(t,\tau)]-\frac{\theta_v}{\sigma_v^2}\left[2\ln\left(1-\frac{[\xi_v^*-\kappa_v+i\phi\rho\sigma_v](1-e^{-\xi^* v\tau})}{2\xi_v^*}\right)\right]-\frac{\theta_v}{\sigma_v^2}[\xi_v^*-\kappa_v+i\phi\rho\sigma_v]\tau+i\phi\ln[S(t)]$$
$$+\lambda\tau\left[(1+\mu_J)^{i\phi}e^{(i\phi/2)(i\phi-1)\sigma_J^2}-1\right]-\lambda i\phi\mu_J\tau+\frac{i\phi(i\phi-1)(1-e^{-\xi^* v\tau})}{2\xi_v^*-[\xi_v^*-\kappa_v+i\phi\rho\sigma_v](1-e^{-\xi^* v\tau})}V(t)\Bigg\} \quad \text{(A4)}$$

VI. Conclusion

There are two routes to explain the superiority of FNN models over the parametric models in forecast settings. These are nonnormality of return distributions and adaptive learning. We know that the BS model assumes normality but asset returns tend to be nonnormal. It is also well known that time-varying volatility (e.g., GARCH models) can account for some but not all of the nonnormality in returns. FNN models capture this nonnormality in returns better than SV or SVJ models.

The other route of improvement is the accounting for the systematic errors of the BS model. SV and SVJ estimate the day t parameters and use them to forecast option prices in day $t+1$. Thus, if there is a systematic error in SV and SVJ over time, this information will not be incorporated into the forecast. In comparison, FNN uses all data up to day t, then forecasts option prices in day $t+1$. Therefore, if it sees a dynamic change, it learns and self-corrects. To put it differently, the typical SV and SVJ environments assume that parameters of the SV and SVJ models are random walks. If they are not, the forecasts will have predictable errors. The learning capabilities of FNN allows us to look back in time and evaluate its forecast errors, and take appropriate action to account for them in the next round leading to more precise forecasts.

Appendix

The characteristic functions are as in (A1) and (A2), shown at the top of the page, where

$$\xi_R = \sqrt{\kappa_R^2 - 2\sigma_R^2 i\phi}$$
$$\xi_v = \sqrt{[\kappa_v - (1+i\phi)\rho\sigma_v]^2 - i\phi(i\phi+1)\sigma_v^2}$$
$$\xi_R^* = \sqrt{\kappa_R^2 - 2\sigma_R^2(i\phi-1)}$$
$$\xi_v^* = \sqrt{[\kappa_v - i\phi\rho\sigma_v]^2 - i\phi(i\phi-1)\sigma_v^2}.$$

The SI, SV, SVSI, and SVJ models are all nested within the general formula in (7). In the case of SVJ, $R(t) = R$ is a constant, $B(t,\tau) = e^{-R\tau}$, and the characteristic functions are as in (A3) and (A4), shown at the top of the page.

The characteristic functions for the SV model can be obtained by setting $\lambda = 0$ in (A3) and (A4).

Acknowledgment

The first author, R. Gençay, would like to thank C. Cao for providing the synchronous option pricing data and for his generous assistance in the implementation of the parametric options pricing models used in this paper.

REFERENCES

[1] P. Andreou, S. Martzoukos, and C. Charalambous, "Option pricing and trading with ANNs, advanced parametric models and implied parameters," in *Proc. IEEE Int. Joint Conf. Neural Netw.* , 2004, vol. 4, pp. 2741–2746.

[2] G. Bakshi, C. Cao, and Z. Chen, "Empirical performance of alternative option pricing models," *J. Finance*, vol. 52, pp. 2003–2049, 1997.

[3] J. Bennell and C. Sutcliffe, "Black-scholes versus artificial neural networks in pricing FTSE 100 options," *Intell. Syst. Accounting, Finance Manage.*, vol. 12, pp. 243–260, 2004.

[4] D. Bates, "Testing option pricing models," in *Handbook of Statistics*, G. S. Maddala and C. R. Rao, Eds. Amsterdam, The Netherlands: North Holland, 1996, vol. 15, Statistical Methods in Finance, pp. 567–611.

[5] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, pp. 123–140, 1996.

[6] J. C. Cox, J. E. Ingersoll, and S. A. Ross, "A theory of the term structure of interest rates," *Econometrica*, vol. 53, pp. 385–407, 1985.

[7] G. Cybenko, "Approximation by superposition of a sigmoidal function," *Math. Control, Signals, Syst.*, vol. 2, pp. 303–314, 1989.

[8] E. Derman and I. Kani, "The volatility smile and its implied tree," in *Quantitative Strategies Research Notes*. New York: Goldman Sachs, 1994.

[9] ——, "Riding on the smile," *Risk*, vol. 7, pp. 32–39, 1994.

[10] B. Dumas, J. Fleming, and R. E. Whaley, "Implied volatility functions: empirical tests," *J. Finance*, vol. 53, pp. 2059–2106, 1998.

[11] B. Dupire, "Pricing with a smile," *Risk*, vol. 7, pp. 18–20, 1994.

[12] F. D. Foresee and M. T. Hagan, "Gauss-Newton approximation to Bayesian learning," in *Proc. IEEE Int. Conf. Neural Netw.*, 1997, vol. 3, pp. 1930–1935.

[13] K. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Netw.*, vol. 2, pp. 183–192, 1989.

[14] A. R. Gallant and H. White, "On learning the derivatives of an unknown mapping with multilayer feedforward networks," *Neural Netw.*, vol. 5, pp. 129–138, 1992.

[15] R. Garcia and R. Gençay, "Pricing and hedging derivative securities with neural networks and a homogeneity hint," *J. Econometrics*, vol. 94, pp. 93–115, 2000.

[16] R. Gençay and W. D. Dechert, "An algorithm for the n Lyapunov exponents of an n-dimensional unknown dynamical system," *Physica D*, vol. 59, pp. 142–157, 1992.

[17] R. Gençay and M. Qi, "Pricing and hedging derivative securities with neural networks and bayesian regularization, early stopping and bagging," *IEEE Trans. Neural Netw.*, vol. 12, no. 4, pp. 726–734, Jul. 2001.

[18] S. Haykin, *Neural Networks*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[19] S. Heston, "A closed form solution for options with stochastic volatility with applications to bond and currency options," *Rev. Financial Studies*, vol. 6, pp. 327–343, 1993.

[20] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Netw.*, vol. 4, pp. 251–257, 1991.

[21] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, pp. 359–366, 1989.

[22] ——, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks," *Neural Netw.*, vol. 3, pp. 551–560, 1990.

[23] J. M. Hutchinson, A. Lo, and A. W. Poggio, "A nonparametric approach to pricing and hedging derivative securities via learning networks," *J. Finance*, vol. 31, pp. 851–889, 1994.

[24] P. R. Lajbcygier and J. T. Connor, "Improved option pricing using artificial neural networks and bootstrap methods," *Int. J. Neural Syst.*, vol. 8, pp. 457–471, 1997.

[25] D. J. C. MacKay, "Bayesian interpolation," *Neural Comput.*, vol. 4, pp. 415–447, 1992.

[26] R. C. Merton, "Option pricing when underlying stock returns are discontinous," *J. Financial Econom.*, vol. 3, pp. 125–144, 1976.

[27] G. Sarwar and T. Krehbiel, "Empirical performance of alternative pricing models of currency options," *J. Futures Markets*, vol. 20, pp. 265–291, 2000.

[28] C. Schittenkopf and G. Dorffner, "Risk-neutral density extraction from option prices: improved pricing with mixture density networks," *IEEE Trans. Neural Netw.*, vol. 12, no. 4, pp. 716–725, Jul. 2001.

[29] N. Swanson and H. White, "A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks," *J. Business Econom. Stat.*, vol. 13, pp. 265–275, 1995.

[30] H. White, *Artificial Neural Networks: Approximation and Learning*. Oxford, U.K.: Blackwell, 1992.

[31] R. Gençay and A. Salih, "Degree of mispricing with the black-scholes model and nonparametric cures," *Ann. Econom. Finance*, vol. 4, pp. 73–101, 2003.

Ramazan Gençay is a Professor of Economics at the Department of Economics, Simon Fraser University, Burnaby, BC, Canada.

His broad specializations are financial econometrics, nonlinear time series, nonparametric econometrics, and chaotic dynamics. His scientific publications have appeared in finance, economics, engineering, statistics, and physics journals. Some of his publications are published in the *Journal of the American Statistical Association*, *Journal of Econometrics*, *Journal of International Economics*, *International Economic Review*, *Journal of Nonparametric Statistics*, *Journal of Empirical Finance*, *Journal of Economic Dynamics and Control*, *Journal of Applied Econometrics*, *European Economic Review*, *Journal of Forecasting*, IEEE TRANSACTIONS ON NEURAL NETWORKS, *Physica A*, *Physica D*, and *Physics Letters A*. He is the coauthor of two books: *An Introduction to High-Frequency Finance and Wavelets* and *Other Filtering Methods in Finance and Economics* (Academic Press and the founding editor of *Finance Research Letters*).

Rajna Gibson received the Ph.D. degree in economic and social sciences and a Graduate degree in business administration from the University of Geneva, Geneva, Switzerland.

She has been a Professor of Finance at the University of Zurich, Zurich, Switzerland, since March, 2000. She is also an Honorary Professor of Finance at the University of Lausanne, Switzerland. She was a Professor of Finance at Lausanne University from October 1991 to February 2000 and directed its M.S. program in banking and finance. She was the Assistant Professor of Finance at Groupe HEC, Paris, and a Visiting Scholar at the New York University, New York, and John Anderson's Graduate School of Management at University of California at Los Angeles. Her papers "Stochastic convenience yield and the pricing of oil contingent claims (coauthored with Eduardo Schwartz, UCLA) in the *Journal of Finance* in 1990 and "Valuing Swiss default-free callable bonds: theory and empirical evidence," *Journal of Banking and Finance*, 1990, were awarded the Greenwich Capital Markets 1990 Investments Prize and the Iddo Sarnat 1990 Prize, respectively.

Dr. Gibson is a member of the Board of Directors at Swiss Re and Director of Research of the Swiss Finance Institute.