Valuing and Hedging American Put Options Using Neural Networks

David L. Kelly*

Rm. 308 GSIA Bldg.
Carnegie Mellon University
Pittsburgh, PA 15213
dk34@andrew.cmu.edu

First Version: September 24, 1993
Current Version: December 15, 1994

*The idea for this paper originated from discussions with John O'Brien and Sanjay Srivastava. We would like to thank Peter Boessarts, David Goldreich, Rick Green, Jamsheed Shorish, Rajdeep Singh, and Per Stromberg for comments and suggestions.

Abstract

We use a neural network to non-parametrically estimate the market valuation function for American put options with real data. The neural network valuation function is twice differentiable, and provides an instantaneous approximation of the American put option using a set of multiple state variables. We use the neural network valuation function to form hedged portfolios for American put options against changes in the stock price, the delta of the stock price, and interest rate.

1. Introduction

One of the early successes in option valuation theory is the Black-Scholes valuation function, which gives a closed-form solution to the values of European call or put options (see Black and Scholes, 1973). This solution cannot, however, price options such as the American put option, because with the American put option exercise prior to the date of maturity may be optimal. As a result, the present discounted value of the option must be calculated at each point in time until maturity, and the decision of whether or not to exercise the option must be continuously modified. Currently, no closed-form theoretical solution exists for the valuation of the American put option.

Without a closed-form theoretical solution, various authors have devised approximation methods to value the American put option. For example, Cox, Ross, and Rubinstein (1979) and others assume that the underlying stock price follows a binomial process at discrete time intervals. This results in a tree of possible stock values. The value of the option at each node in the tree is is determined by using backward dynamic programming. The binomial process approaches the true value of the option as the discrete time intervals approach zero in length.

Brennan and Schwartz (1977 and 1978) use finite differences to value the American put option. This method is similar to the binomial approximation method in its assumptions and implementation of backward dynamic programming. Both the binomial approximation method and the finite difference method become more complex as the discrete intervals become shorter. In addition, both methods become at least geometrically more complex as more state variables are added. For practical purposes, then, these methods only value the American put option with respect to the underlying stock price. Lastly, there is no statistical measurement of error which could be used to determine the accuracy of either the binomial or the finite difference approximation method.

Geske and Johnson (1984) and Carverhill and Webber (1990) use a series of European options to approximate the American option. The compound option pricing method requires much less computation time versus the binomial approximation method. However, the compound option pricing method still suffers from computational costs, and again is derived with respect to one state variable. The compound option pricing method cannot provide any sort of statistical error measurement.

In addition to computational speed, another desirable property of a given approximation method is the ability to form portfolios hedged against changes in the underlying state variables. Such a hedge portfolio requires the partial derivatives of the approximation function with respect to the state variables which are not well-defined in the above methods.

MacMillan (1985), Barone-Adesi and Whaley (1987), and others rely upon a quadratic approximation of the American put option value. The quadratic approximation function is differentiable almost everywhere and possesses statistical properties which allow hypothesis testing.[1] Whaley (1986) forms hedging portfolios using derivatives calculated from the differentiable portions of the quadratic approximation function. The quadratic approximation method is faster than the aforementioned approximation methods, but still has computational costs arising from the numerical approximation of probability density functions.

In this paper we present an analytical tool, the neural network, a non-parametric estimator which instantaneously values the American put option with a twice differentiable function of a set of multiple state variables. Since the semi non-parametric approximation is differentiable, the partial derivatives of the valuation function exist. Thus we construct portfolios hedged against uncertain movements in the stock price, interest rate, and the delta of the stock price. One can create portfolios hedged against time to maturity and volatility analogously. Lastly, the semi non-parametric approximation is a consistent estimator. We choose neural networks as the non-parametric estimator because the consistency results extend to the first derivatives of the estimator. That is, the first derivative of the estimator converges to the true derivative of the option valuation function.

Asymptotic normality has only been established for iid environments, while here the underlying stock price follows a random walk. However, asymptotic normality can be established as the number of stocks approaches infinity, if the stocks are independent. This would allow construction of confidence intervals around the estimates of the option value or tests of the hypothesis that each state variable is relevant in determining the value of the option.

[1] The quadratic approximation function is discontinuous only at the stock price at which early exercise of the American option becomes optimal.

The neural network approximation method is similar to the approximation methods presented in Bossaerts (1989) and Hutchinson, Lo, and Poggio (1994). Bossaerts uses Simulated Method of Moments estimation to find coefficients for an option valuation function. After estimation, the option valuation function is evaluated instantaneously with respect to multiple state variables. In addition, the estimated coefficients possess well-defined statistical properties which allow for confidence intervals or hypothesis testing. However, the finite sample option valuation function is not a differentiable function of the state variables. Thus it is not possible to form a hedged portfolio from the derivatives of the option valuation function. Similarly, Hutchinson, Lo, and Poggio estimate the value of S&P call options non-parametrically, and create delta hedged portfolios. In this case, the delta hedge created from the estimator often performs better than hedge portfolios created from the Black-Scholes formula.

The structure of the paper is as follows. Section 2 develops a model that values an American put option given an optimal early exercise strategy, which can be estimated. In Section 3, we show that there exists a consistent non-parametric estimator for the option price developed in Section 2. Furthermore, we show that the derivatives of the non-parametric estimator consistently estimate the derivatives of the option price.

Section 4 applies the neural network to real American put option data, and shows that the neural network provides a close approximation of the true American put option value. Section 5 constructs a portfolio delta-hedged against changes in the stock price and the risk-free interest rate using the neural network approximation results from section 4. We also construct a portfolio that is gamma-hedged against changes in the stock price and the risk-free interest rate. Section 6 discusses some conclusions.

2. Valuation of the American Put Option Given Optimal Exercise

We first discretely model the movement of the stock price S_t and the interest rate r_t as random walks:

$$\begin{bmatrix} S_t \\ r_t \end{bmatrix} = \begin{bmatrix} S_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix} \qquad \zeta_t \sim N\left(0, \sigma^2\right) \tag{2.1}$$

Here σ^2 is the variance-covariance matrix (we allow for correlation between the stock price and interest rate). Let the expiration date be time t_1, the exercise price X, and $D = t_1 - t$ the time to maturity.

The payoff of a put option exercised at time i is $X - S_i$. We assume risk neutrality, therefore the present discounted value of the put option is then:

$$e^{-\sum_{j=t}^{i} r_j} (X - S_i) \tag{2.2}$$

The investor invests the proceeds from exercise at the risk free rate after exercising the option, which is then discounted back to the current period.

Let $Z_i = [X, D, \sigma, r_i, S_i]'$. Following for example Boessarts (1989), we define an early exercise strategy function $P : \Re^5 \to \{0, 1\}$ such that:

$$P(Z_i) = \begin{cases} 1 & \text{if exercise is optimal} \\ 0 & \text{o.w.} \end{cases} \tag{2.3}$$

We note that the function P is defined such that if $P_i = 1$, exercise prior to i is not optimal. P therefore depends also on realizations of Z between t and i. P is the optimal early exercise strategy (chosen by finding the sup of the value of the option over the space of measurable exercise strategy functions).

By combining equations (2.2) and (2.3) we obtain a discrete valuation function for an American put option at time t, given realizations $\left[\tilde{r}, \tilde{S}\right]$:

$$\tilde{V}\left(\tilde{Z}\right) = \sum_{i=t}^{t_1} e^{-\sum_{j=t}^{i} r_j} (X - S_i) P(Z_i) \tag{2.4}$$

Applying the risk neutral expectation operator gives:

$$V(Z_t) = E_t \left[\sum_{i=t}^{t_1} e^{-\sum_{j=t}^{i} r_j} (X - S_i) P(Z_i) \right] \tag{2.5}$$

Note that we can also shrink the discrete time interval to zero. Then the stock price and interest rate process (2.1) becomes a Brownian motion and we obtain a continuous time formulation of the valuation function:

$$V\left(Z\left(t\right)\right) = E_t\left[\int_t^{t_1} e^{-\int_t^i r(j)dj}\left(X - S\left(i\right)\right) P\left(Z\left(i\right)\right) di\right] \tag{2.6}$$

We assume the observed market options data fits the pricing function V plus a noise term. The noise comes from market micro structure issues (price discreteness), possible noise trading, and the bid is an imperfect measure of the price. We also assume a limited rationality on the part of the agents. Although the true optimal early exercise strategy is unknown to the agents, the market valuation is equal to the true value plus a noise term. Agents may make mistakes when estimating the optimal early exercise time, but the error in the valuation function from these mistakes is mean zero normally distributed. Let O be the observed market option price, then:

$$O_t = V\left(Z_t\right) + \epsilon_t \qquad\qquad \epsilon_t \sim N\left(0, \sigma_\epsilon^2\right) \tag{2.7}$$

We assume that ϵ is an iid random variable and is uncorrelated with ζ.

Equation (2.7) is a convenient representation of the option price. We could calculate the option price using equation (2.7), except that the valuation function V is unknown (specifically, the optimal early exercise strategy is unknown). Equation (2.7) is a standard representation of a non parametric estimation problem. The idea is to use a function approximator to approximate V, and simultaneously estimate ϵ.

We now introduce a function $\hat{V} : \Re^5 \times \left(\Theta \subset \Re^{7K+1}\right) \to \Re$ defined as:

$$\hat{V}\left(Z_t, \hat{\theta}\right) = \sum_{k=1}^{K} \hat{\beta}_k \tanh\left(\sum_{j=1}^{5} \hat{\beta}_{kj} Z_{j,t} + \hat{I}_k\right) + \hat{I}_0 \tag{2.8}$$

$$\hat{\theta} = \begin{bmatrix} \hat{\beta} \\ \hat{I} \end{bmatrix} \tag{2.9}$$

The size of θ, which is determined by the value of K is a finite number. Note that $\hat{V}$ is C^2. Finally, we assume θ is drawn from a compact set, Θ.

In this case, given a data set $i = 1 \ldots N$, we estimate θ by minimizing sum of squared errors:

$$\hat{\theta} = \min_{\theta \in \Theta} \sum_{i=1}^{n} \left[O_t - \hat{V}(Z_t) \right]^2 \tag{2.10}$$

Hence $\hat{V}_{n,K}$ is a neural network with statistical properties similar to non-linear least squares. Our choice of neural networks is motivated by the consistency results of derivatives of neural networks with respect to the true derivatives (see Gallant and White, 1992), which is important for the formation of hedged portfolios.

3. Statistical Issues.

In this section, we consider the appropriate choice for K, that is the number of parameters used for estimation, and the resulting asymptotic properties of the estimator. Hornik, Stinchombe, and White (1989) show that since V is an element of the space of functions defined by an L_p norm and tanh is an l-finite squashing function, there exists a θ_0 such that in the L_p norm:

$$\lim_{K \to \infty} \left\| \hat{V}(\theta_0) - V \right\|_p = 0 \tag{3.1}$$

Essentially, the neural network is a function approximator, like a polynomial approximation.[2] However, since we cannot estimate an infinite number of parameters with a finite sample, we let the number of parameters increase with the sample size. That is let $K = K_n$ with $K_n \to \infty$ as $n \to \infty$, then $\hat{V}(Z_t, \theta_{n,K_n})$ is a non-parametric estimator.

Next we establish consistency of the non-parametric estimator and the first derivative. Following White (1992) chapter 12 theorem 3.3, we wish to show that V is an element of a Sobolev space. We can then use the theorem to establish convergence of neural network non-parametric estimators and functionals of the estimators. Let the Sobolev norm be:

$$\|V\| = \left[\sum_{|\lambda| \leq m} \int_{\mathcal{Z}} \left| D^{\lambda} V(z) \right|^p dz \right]^p \tag{3.2}$$

[2] Neural networks use bounded functions, however, which reduces the variance of the estimate relative to using unbounded functions like a polynomial approximation.

$$D^{\lambda}V(z) = \left(\frac{\partial^{\lambda}V}{\partial z_1}\right) \cdot \ldots \cdot \left(\frac{\partial^{\lambda}V}{\partial z_5}\right) \tag{3.3}$$

The Sobolov space is then the set of all functions:

$$\mathcal{V}_{m,p,\mathcal{Z}} = \{V \,| D^m V \in L_p(\mathcal{Z})\} \tag{3.4}$$

Let $m = 1$, since we are interested in the first derivative. Then the theorem requires that we show:

$$V \in \mathcal{V}_{1+1+\left[\frac{5}{p}\right],p,\mathcal{Z}} \tag{3.5}$$

Here $[\cdot]$ represents the integer portion. Let $p > 5$ so that this term is zero. Essentially, since the Sobolev spaces are nested, the theorem shows that the error converges almost surely to zero in the norm $\|\cdot\|_{2,\infty,\mathcal{Z}}$.

A well known sufficient condition for V to be an element of the Sobolov space is that $(V)^p$ is differentiable of order m, in this case twice differentiable. Although time is discrete, we can assume that the elements of Z_t can take on continuous values. Given that the exercise strategy is optimal, the valuation function is twice differentiable.[3]

Next we check the assumptions on the random variables to assure that a law of large numbers exists. The error term ϵ is iid and $\{Z_t\}_n$ converges almost surely from the martingale strong law of large numbers. Hence, we have:

PROPOSITION 1 *Let the estimator $\hat{V}$ be defined by equation (2.8) and the vector $\hat{\theta}$ be defined according to equation (2.10). Let the random variables Z_t and ϵ_t be defined as above. Finally, assume that $K_n \to \infty$. Then:*

$$h\left(\hat{V}(Z_t)\right) \overset{a.s.}{\to} h(V(Z_t)) \tag{3.6}$$

[3] Let Z_t approach a critical point where the decision about exercise in the current period changes. Then the value of the option given that the option is exercised at t approaches the value of the option given that the option is not exercised at t as Z_t approaches the critical point.

We apply the theorem by considering two functional choices for h:

$$h(V) = V \tag{3.7}$$

$$h(V) = \frac{dV}{dZ} \tag{3.8}$$

The consistency results do not specify growth rates for K_n. One method is to use a deterministic rule, such as select $K_n = n^{\delta}$, $\delta \in (0,1)$, with $\delta = \frac{1}{2}$ a common choice. A value of $\delta = \frac{1}{2}$ corresponds to about $K = 5$.

Another method of determining the appropriate size for K_n is a stochastic method, such as cross validation (see White, 1992 p. 167). The procedure is to do a non-linear regression for a given K, excluding m data points. Then compute the (out of sample) error for the m data points. Repeat this procedure over all data points and select the K that minimizes the total out of sample sum of squared error.

We performed the cross validation procedure with $m = 10$, which corresponds to 136 regressions per value of K. We tried values of $K = 3$ to $K = 14$. The results are shown in Figure (1). As the value of K increases, the estimator is able to fit more complex functions. However in practice as the value of K increases, the estimator can over fit simple functions, erroneously attributing random noise as part of the function. Hence cross validation generally gives a U-shaped errors as a function of K.

The K value that minimized the cross validation statistic was $K = 5$, equal to the deterministic rule. However, we note that there is little difference over the the range $K = [4, 8]$.[4] We set K equal to 5, which also gives a low standard deviation for the hedged portfolios.

Asymptotic normality for non parametric neural networks has not been established. However, Andrews (1991) shows that for the iid case a non-parametric estimator estimated using non-linear least squares has asymptotic normality properties similar to the case when the model is fully parameterized. Therefore, we have some reason to believe that the estimator

[4]Since cross validation agreed with the deterministic rule, we did not statistically test the hypothesis that the errors are different.

has asymptotic normality properties similar to non-linear least squares.[5]

4. The Neural Network Approximation of the American Put Option

Consider an investor who wishes to value an American put option. Since there is no known closed-form valuation function for the American put option, the investor uses an approximation technique. The investor wants the approximation technique to use as little computational time as possible, since the investor incurs costs while waiting for the approximation to be calculated. To the investor, then, the approximation technique must be both accurate and fast.

Suppose the investor uses a neural network to statistically approximate the American put option. The investor periodically estimates the parameter set $\hat{\theta}$ on real data to obtain the valuation function $\hat{V}$. The evaluation of the neural network has no computational cost.

Following the investor's problem, we applied a neural network to American put option data. Since in the market the actual option price is unknown, we use the bid price of the option as a proxy. We collected closing daily bid data on American put options for four different common stocks: IBM, Chrysler, GM, and Merck. The total number of options on these stocks was 18. Options on the same stock had different exercise prices and/or time to maturity. These stocks were selected because of large trading volume when in the money and for a variety of volatility measures. The data was collected from October 1, 1993 to April 13, 1994. The data set consisted of 1369 observations.

For the risk-free interest rate, we selected the three month treasury bill. For the volatility we used the one year historical variance. We also tried other historical volatility measures, with similar results. We note that these are not exact measures of the true risk free interest rate and volatility.

For the choice of the number of parameters we selected $K = 5$, which corresponds to $\delta = .496$. The search for the error minimizing $\hat{\theta}$ was conducted using a quasi-Newton

[5]One can prove asymptotic normality for the non-parametric estimator under the condition that Z_t is m-dependent, however proof for Z_t following a random walk is more difficult. One possibility is that asymptotic normality could be done with respect to the number of stocks (panel data) as opposed to the mixture of panel and time series data examined here.

algorithm (a second derivative method). We used the Nguyen and Widrow (1993) algorithm for selection of initial conditions for the search for the error-minimizing vector $\hat{\theta}$.

Descriptive statistics summarizing the estimation are given in table (2). Figure (2) gives the predicted versus actual option prices. The estimation has a lower error when the option is out of the money.[6] However as seen from the R^2 column, out of the money options generally had less variance and were therefore easier to predict. The R^2 increases as the option becomes more in the money. Because of price discreteness, estimation is harder for options that are out of the money, regardless of what approximation technique is used. Still, the difference in the R^2 amounts to only two percent between out of the money and in the money options.

For comparison purposes we also applied the binomial approximation method to the same real data as the neural network approximation. For each data point, we used the binomial approximation method given in Cox, Ross, and Rubinstein (1979) to compute the estimated price of the option. Descriptive statistics summarizing the results of the binomial exercise are also given in table (2).

To summarize, in practice convergence to the true θ_0 is not certain. The correct choice for K is unknown, proxies are used for the option price, risk-free rate, and volatility, there is discreteness in the observed variables, and a search algorithm is used to find the error-minimizing $\hat{\theta}$. In spite of this, the neural network estimate explains the variance of the option value well ($R^2 = .99$).

5. Formation of Hedged Portfolios Using the Neural Network

5.1. Formation of a Delta-Rho Hedged Portfolio

One application of the neural network approximation technique is the formation of a hedged portfolio. Suppose an investor desires to hedge one share of an American put option against changes in the stock price (*delta* hedging) and interest rate (*rho* hedging). The investor purchases δ shares of the stock and ρ shares of the risk-free bond B to form the hedged

[6] An option was considered out of the money if the return from exercise was less than \$-2. An option was near in the money if the return from exercise was at least \$-2 but less than \$2. Options with a higher return were considered in the money. There were 730, 303, and 336 observations, respectively.

portfolio. Suppose that the valuation function of the American put option is known with certainty. Then the hedged portfolio has a value of:

$$Q = V(X, D, \sigma, r_t, S_t) + \delta S + \rho B \tag{5.1}$$

$$\delta = -\frac{\partial V(X, D, \sigma, r_t, S_t)}{\partial S} \tag{5.2}$$

$$\rho = -\frac{\partial V(X, D, \sigma, r_t, S_t)}{\partial r}\left(\frac{dr}{dB}\right) \tag{5.3}$$

Unfortunately, the valuation function V is unknown to the investor. Suppose the investor instead uses the neural network estimation of the American put option value from Section 4. Define the neural network estimation by the function $\hat{V}(X, D, \sigma, r_t, S_t)$. Then the estimate of the hedged portfolio is:

$$\hat{Q} = V(X, D, \sigma, r_t, S_t) + \hat{\delta} S + \hat{\rho} r \tag{5.4}$$

$$\hat{\delta} = -\frac{\partial \hat{V}(X, D, \sigma, r_t, S_t)}{\partial S} \tag{5.5}$$

$$\hat{\rho} = -\frac{\partial \hat{V}(X, D, \sigma, r_t, S_t)}{\partial r}\left(\frac{dr}{dB}\right) \tag{5.6}$$

We empirically test the neural network ability to hedge by constructing a delta-rho hedged portfolio for 18 options over at most 131 days. This provides 1348 one day tests. One complication to this test is that we assume no transaction costs, allowing us to change the portfolio on a day to day basis. The actual investor cannot exactly follow this strategy because of transaction costs. Another problem is that the valuation function is non-linear. Hence the quantity of stocks and bonds required to hedge against the option changes as the option price changes. The investor must at the end of each day rebalance: sell the old portfolio, which hedged the last day against the current day. Then the investor must purchase a new portfolio hedged against the next day. By finding the percentage profits earned over each day, we obtain an overnight rate of interest.[7] Since our data is daily, there is hedging error because the investor cannot adjust the hedged portfolio during the day. The investor additionally does not hedge against changes in the number of days to maturity.

Descriptive statistics summarizing the results of our test are shown in Table (3). Table (3) compares the overnight rate of interest earned by the hedged portfolio each day to the overnight risk free rate, which is theoretically mean zero and variance zero.[8]

The hedged portfolio had a mean overnight rate of interest of -.047%, with a standard deviation of over 2%. The hedged portfolio still has substantial risk. However, the price of one share of the option has a mean *overnight* rate of interest of 1% and a standard deviation of over 27%. Thus the hedged portfolio reduces the risk of the option by over 90%, for both in the money and out of the money options. This is evident in Figure (4).

Additionally, changes in the interest rate added substantially to the changes in the variance of the option price. The risk free rate varied almost one percent over the data set and had a standard deviation of .22 percent. Hedging against changes in the interest rate was essential to obtain a good hedge.

5.2. Formation of a Gamma-Hedged Portfolio

There are no theoretical results establishing convergence of the second derivative of the neural network option valuation function to the true second derivative option valuation function. However, the neural network valuation function is twice differentiable, therefore we can form a portfolio hedged against changes in the delta of an American put option.

Suppose an investor desires to hedge against changes in the delta of an American put option. A *gamma hedged* portfolio is a portfolio hedged against changes in the stock price, and against changes in the delta of the option. This protects the investor from losses that arise from having a nearly correct, but not exact, delta during the day.

[7] Whaley (1987) calculates the rate of interest only at the expiration of the option, because the error was found to be in the same direction from one day to the next. However, this was not true generally for our data set. Additionally, our data set does not consist of enough options to calculate the interest rate at only the expiration dates.

[8] This is equivalent to a zero cost hedged portfolio: we short one share of the risk free bond to obtain a zero cost portfolio that theoretically has zero mean return and variance.

Again, since the true put option valuation function is unknown the investor uses the neural network approximation to form the following portfolio:

$$\hat{Q} = V\left(X_1, D_1, \sigma, r_t, S_t\right) + \hat{\gamma}^1 S + \hat{\gamma}^2 V\left(X_2, D_2, \sigma, r_t, S_t\right) \tag{5.7}$$

$$\hat{\gamma}^1 = -\hat{V}_s\left(X_1, D_1, \sigma, r_t, S_t\right) + \left(\frac{\hat{V}_{ss}\left(X_1, S, D_1, r, v\right)}{\hat{V}_{ss}\left(X_2, D_2, \sigma, r_t, S_t\right)}\right)\hat{V}_s\left(X_2, D_2, \sigma, r_t, S_t\right) \tag{5.8}$$

$$\hat{\gamma}^2 = -\frac{\hat{V}_{ss}\left(X_1, D_1, \sigma, r_t, S_t\right)}{\hat{V}_{ss}\left(X_2, D_2, \sigma, r_t, S_t\right)} \tag{5.9}$$

Here V_{xx} is the second partial derivative of V with respect to x.

We empirically test the ability of the neural network approximation to hedge by constructing a gamma hedged portfolio for 18 options over at most 131 days. This gives 3210 one day tests (using all possible sets of two options on the same stock). Again, we assume no transaction costs, and have the same problems of daily data, non linearity, and discreteness of the valuation function.

The results of our test are shown in Table (3). The mean difference between the hedged portfolio and the risk free rate is .058% , well above zero. However, we again note that the hedged portfolio has on average 98% less risk than the option. The standard deviation of the difference between the risk free rate and the gamma hedged portfolio is 2.12%. Again, the standard deviation is substantially greater than zero, but also substantially less than 27% obtained by holding one share of the option.

6. Conclusions and Further Research

The neural network is able to approximate the American put option using real data. The neural network is more accurate than other approximation techniques and requires no computation time, providing an instantaneous value of the American put option. Since neural networks are differentiable functions of the state variables, we constructed hedged portfolios from the neural network approximation. The hedged portfolio reduced the variance of the option by 90 percent. Furthermore, the portfolio hedged against changes in interest rates, which add substantial risk to the option price. The neural network can hedge against changes in interest rates and volatility, a substantial improvement over other hedging strategies.

For further research, we hope to apply the neural network approximation method to other derivative securities. The neural network technique is useful for valuing any security where the value is difficult to compute due to large numbers of state variables or unknown exercise strategies. For example, currency options move with respect to large number of interest rate variables, so the neural network valuation technique would be especially useful. Neural networks would also be useful for valuing mortgage backed securities, which allow the mortgage holder to pay the mortgage prior to the expiration date (in essence exercising an option).

7. Appendix 1: Tables

	O	D	X	v	S	r
mean	2.36	69.28	48.75	29.79	51.34	3.22
st. dev.	2.47	42.61	9.45	14.57	10.42	0.22
max.	10.75	196.00	60.00	58.95	64.75	3.68
min.	0.06	0.00	30.00	5.82	29.12	2.92
Correlation Matrix						
	O	D	X	v	S	r
O	1.00	0.15	0.25	-0.26	-0.27	0.01
D	0.15	1.00	-0.15	-0.22	-0.15	-0.73
X	0.25	-0.15	1.00	0.62	0.82	-0.01
v	-0.26	-0.22	0.62	1.00	0.76	0.16
S	-0.27	-0.15	0.82	0.76	1.00	-0.08
r	0.01	-0.73	-0.01	0.16	-0.08	1.00

Table 1: descriptive statistics for data set. 1369 total observations.

Data Set	mae	std	sse	R^2
Network estimation: entire data set.	\$.12	.15	32.54	.996
Network estimation: out of the money.	\$.08	.11	8.47	.965
Network estimation: near in the money.	\$.14	.18	9.94	.976
Network estimation: in the money.	\$.17	.20	14.13	.988
Binomial Approx. Method	\$.99	.87	2382.74	.72

Table 2: Put option estimation results and comparison. mae: mean absolute error. std. standard deviation of error. sse: sum of squared error.

Portfolio	me	std	%
Delta-Rho Hedged portfolio	.047%	2.79%	90.0
Delta-Gamma Hedged portfolio	.058%	2.12%	92.2
One share of put option	1.078%	27.15%	

Table 3: Hedged portfolio results. me: mean error. std. standard deviation of error. %: percentage reduction in put option variance.

8. Appendix 2: Figures

Figure 1: graph of errors versus K. The number of parameters is $7K+1$.

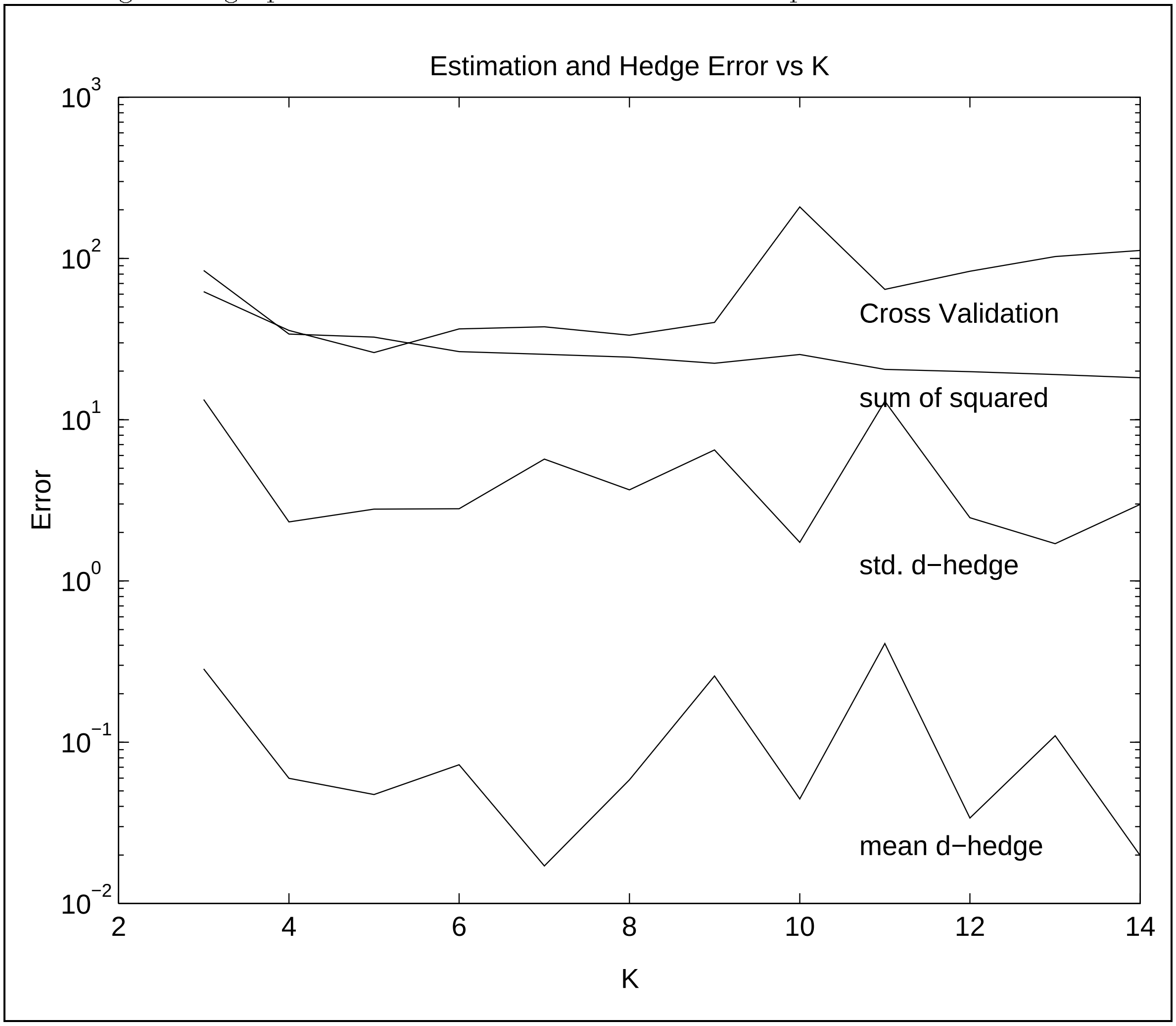


Figure 2: graph of predicted versus actual option prices. The actual option price is the dotted line.

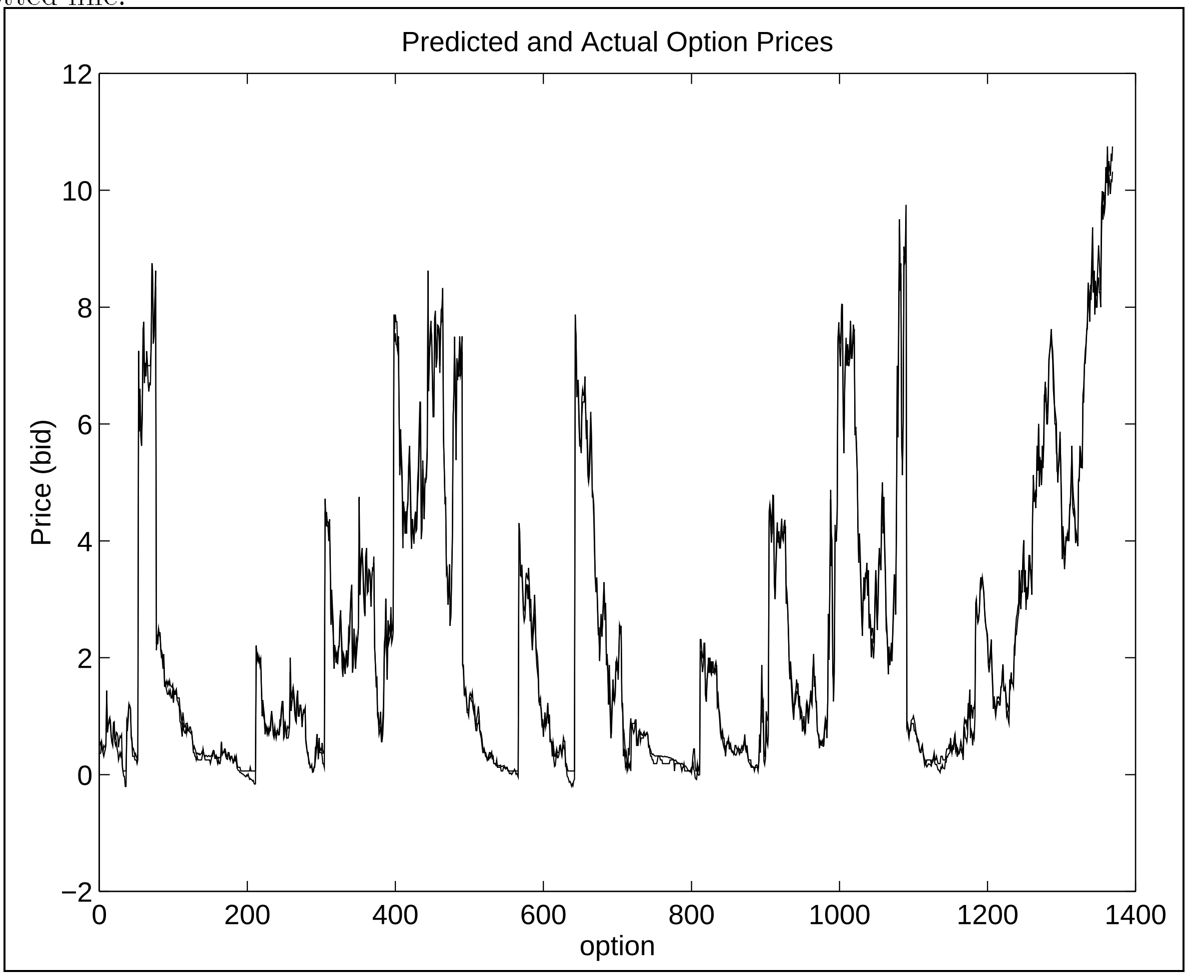


Figure 3: Graph of Residuals versus the option price.

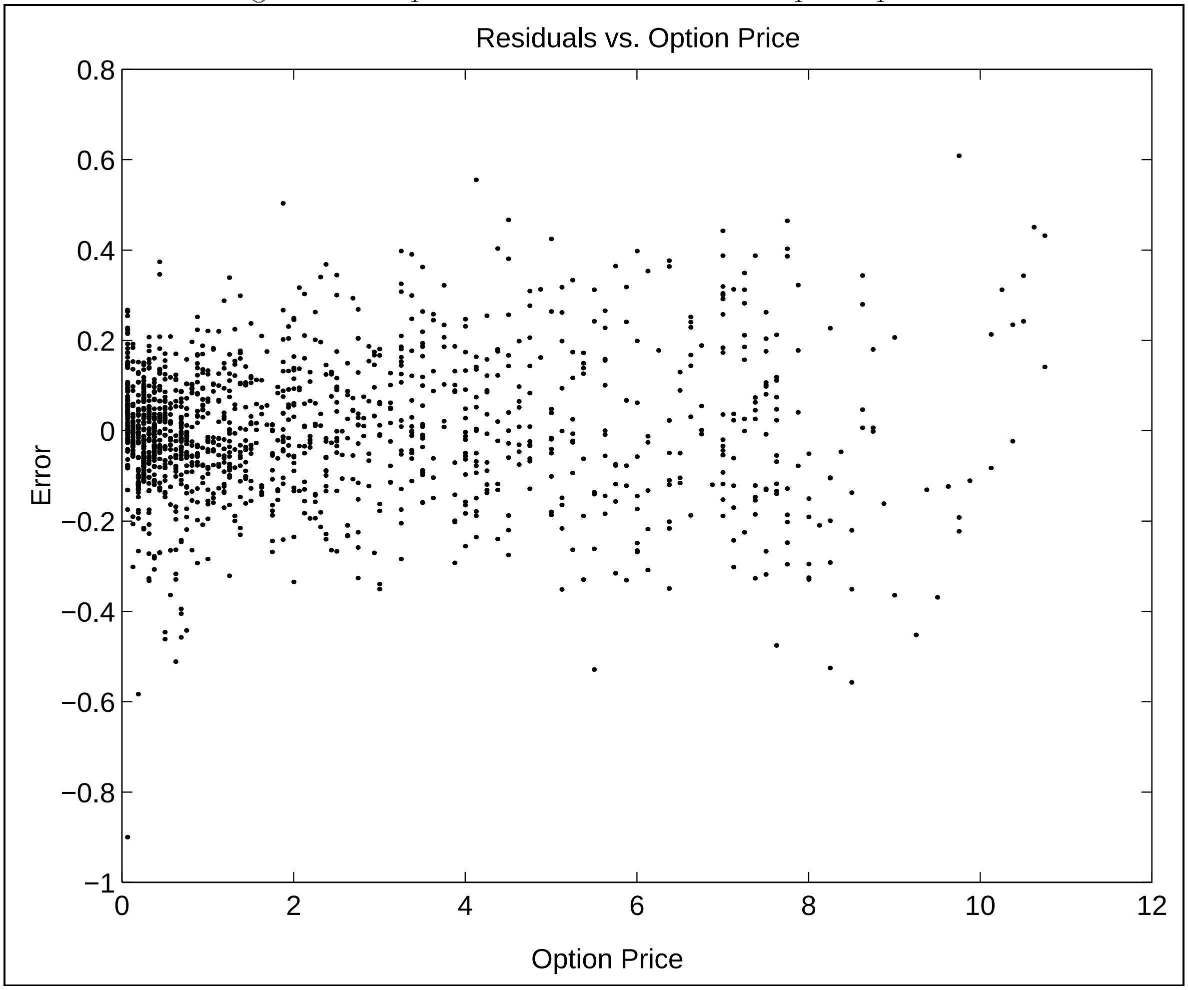


Figure 4: graph of delta-rho hedged portfolio (solid line) versus buying and holding the option (dotted line). The option exhibits a higher return and greater variance.

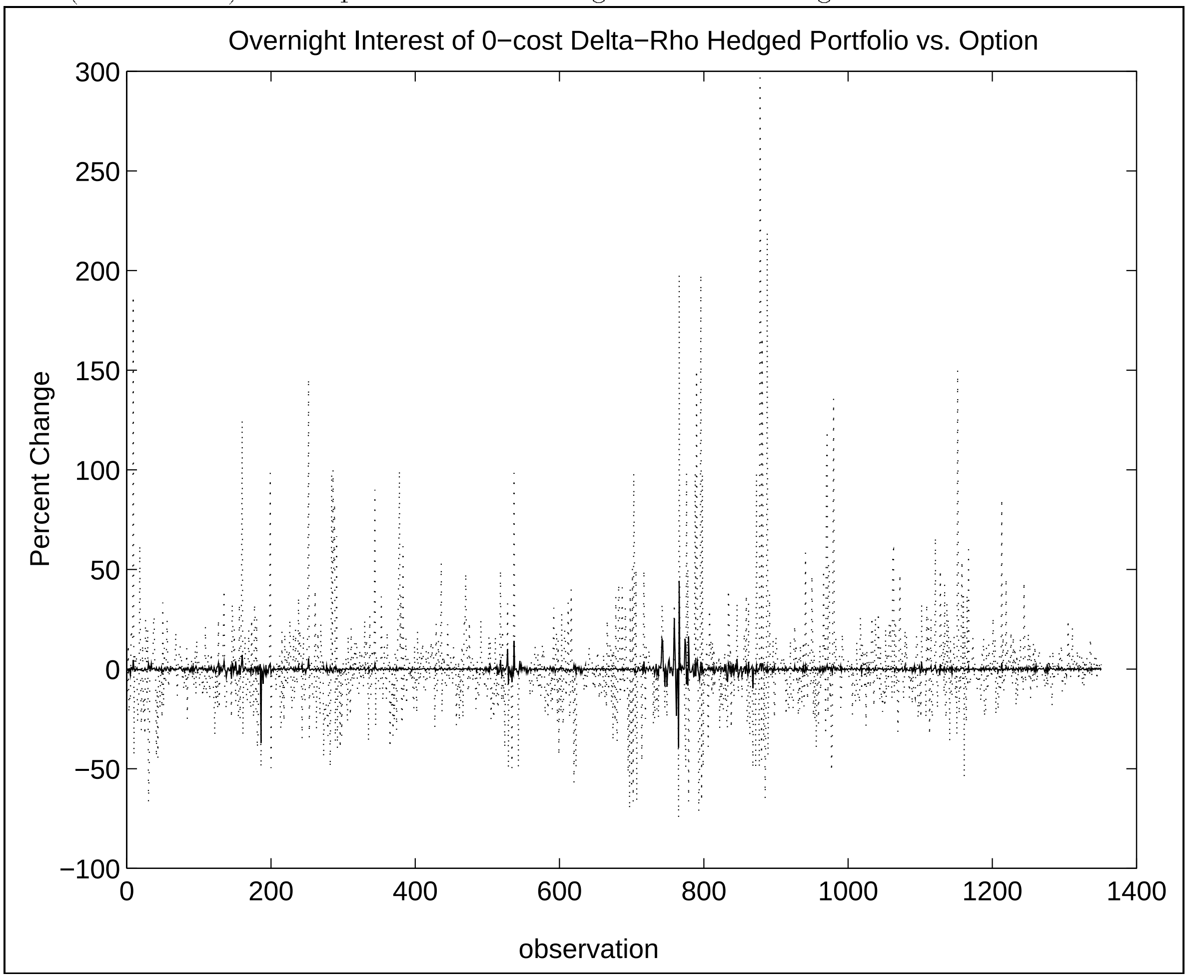
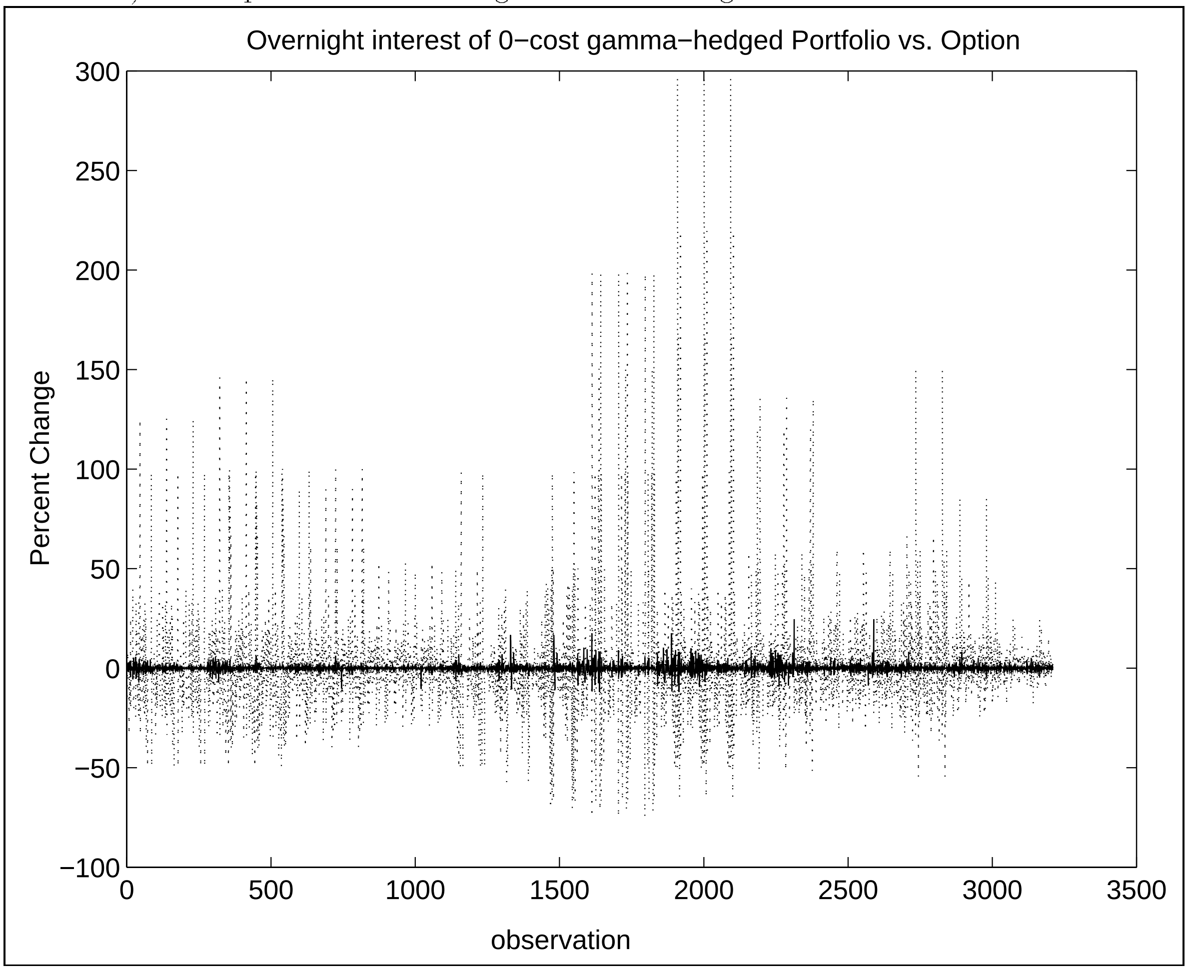


Figure 5: graph of gamma-hedged portfolio (solid line) versus buying and holding the option (dotted line). The option exhibits a higher return and greater variance.



Reference

Andrews, D. K. (1990). Asymptotic Normality of Series Estimators for NonParametric and Semiparametric Regression Models. *Econometrica, 59.*

Barone-Adesi, G. and Whaley, R. (1987). Efficient Analytic Approximation of American Option Values. *Journal of Finance*, *42*, 301–320.

Black, F. and Scholes, M. S. (1973). The Pricing of options and corporate liabilities. *Journal of Political Economy*, *7*, 637–54.

Bossaerts, P. (1989). Simulation Estimators of Optimal Early Exercise. Graduate School of Industrial Administration.

Brennan, M. and Schwartz, E. (1977). The Valuation of American Put Options. *Journal of Finance*, *32*, 449–462.

Brennen, M. J. and Schwartz, E. S. (1978). Finite Difference Methods and Jump Processes Arising in the Pricing of Contingent Claims: A Synthesis. *Journal of Financial and Quantitative Analysis*, *13.*

Carverhill, A. and Webber, N. (1990). American Options: Theory and Numerical Analysis. In Hodges, S. (Ed.), *Options: Recent Advances in Theory and Practice*. New York: Manchester UP, 1990.

Cox, J., Ross, S., and Rubinstein, M. (1979). Option Pricing: A simplified approach. *Journal of Financial Economics*, *7*, 229–263.

Gallant, A. R. and White, H. (1992). On learning the Derivatives of an Unknown Mapping With Multilayer Feedforward Networks. *Neural Networks*, *5*(1).

Geske, R. and Johnson, H. (1984). The American Put Option Valued Analytically. *Journal of Finance*, *39*, 1511–1524.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multi-Layer Feedforward Networks are Universal Approximators. Tech. rep. 88-45R, University of California, San Diego Dept. of Economics.

Huang, C. and Litzenberger, R. H. (1988). *Foundations for financial economics.* North–Holland.

MacMillan, L. (1985). Analytical Approximation for the American Put Option. *Advances in Futures and Options Research*, *5*, 119–139.

Nguyen, D. and Widrow, B. (1990). Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights. In *International Joint Conference on Neural Networks*, Vol. 3.

Whaley, R. (1986). Valuation of American Futures Options: Theory and Empirical Tests. *Journal of Finance*, *41*, 127–150.

White, H. (1989). Learning in Artificial Neural Networks: A Statistical Perspective. University of California San Diego.

White, H. (1992). *Artificial Neural Networks: Approximation and Learning Theory.* Blackwell Publishers.