# Predicting Call Option Prices Using Regression Models

Munira Shahir

July 24, 2014

**Abstract**

One method to consists of predict call option prices is the Black-Scholes equation. However, that is utilizing for predicting European call option prices not American call option prices. The work done consisted of utilitizing regression models to predict call option prices utilizing known information from the New York Stock Exchange (NYSE)

## 1 Introduction

In finance, an option is a financial contract in which the buyer has the right, but not the obligation to buy or sell an agreed upon price on or before a specific data. The contract to buy an option is a call option. The contract to sell an option is a put option. These two are related by put-call parity. The agreed upon price is the strike price. The date to decide whether or not the option by is the exercise date. There are two types different styles of options American and European. Owners of a European style options may exercise them only at expiration. Owners of American style options may exercise them at and before expiration. For predicting the price of European options, Black Scholes equation is used by some investors. However, this is only useful as an approximation for the option price. Also, there isn't an equivalent equation to the Black-Scholes equation for predicting American option prices. This than leads into developing regression models to predict call option price. This is done utilizing linear regression models and non-parametric regression models. The non-parametric regression models consist of generalized additive model (GAM), projection pursuit regression (PPR), regression trees, and random forest. The predictors consist of the strike price of the option, the current price of the stock, expiration time and the historical volatility, which is how much the option price varied over the last two months. The response is the option price. The data is taken from the New York Stock Exchange (NYSE).

## 2 Methods

### 2.1 Linear Regression Models

Linear regression models consist of modeling the response as a linear function of the predictors. $y = \beta_0 + \sum \beta_i x_i + \epsilon$. Where $y$ is the predicted response, the $\beta_i$'s are the parameters, the $x_i$'s are the predictors, and $\epsilon$ is the residual error. One of the ways to determine if a linear regression model would work is by examining a scatterplots of the predictors and the response,in addition to their log transformations.

## 2.2 GAM Models

GAM models consist of modeling the response as a linear combination of functions of a single predictor. $y = \beta_0 + \sum \beta_i f_i(x_i) + \epsilon$ This model tested all possible combination of predictors for determining the response, in addition to their log transformations.

## 2.3 PPR Models

PPR models are a generalization of GAM models. In this models functions of a linear combination of predictors utilized to predict the response, in addition to their log transformations. $y = \beta_0 + \sum \beta_i f_i(\alpha_i^T x) + \epsilon$ Where $y$ is the response,$\beta_i$'s are the parameters, $\alpha_i$'s are the projection direction vectors, and $\epsilon$ is the residual error.

## 2.4 Regression Tree Models

Regression Trees consist of partioning the data into intelligently chosen subsets, and in each subset the response will be modeled as the mean. This model tested all possible combination of predictors for determining the response, in addition to their log transformations.

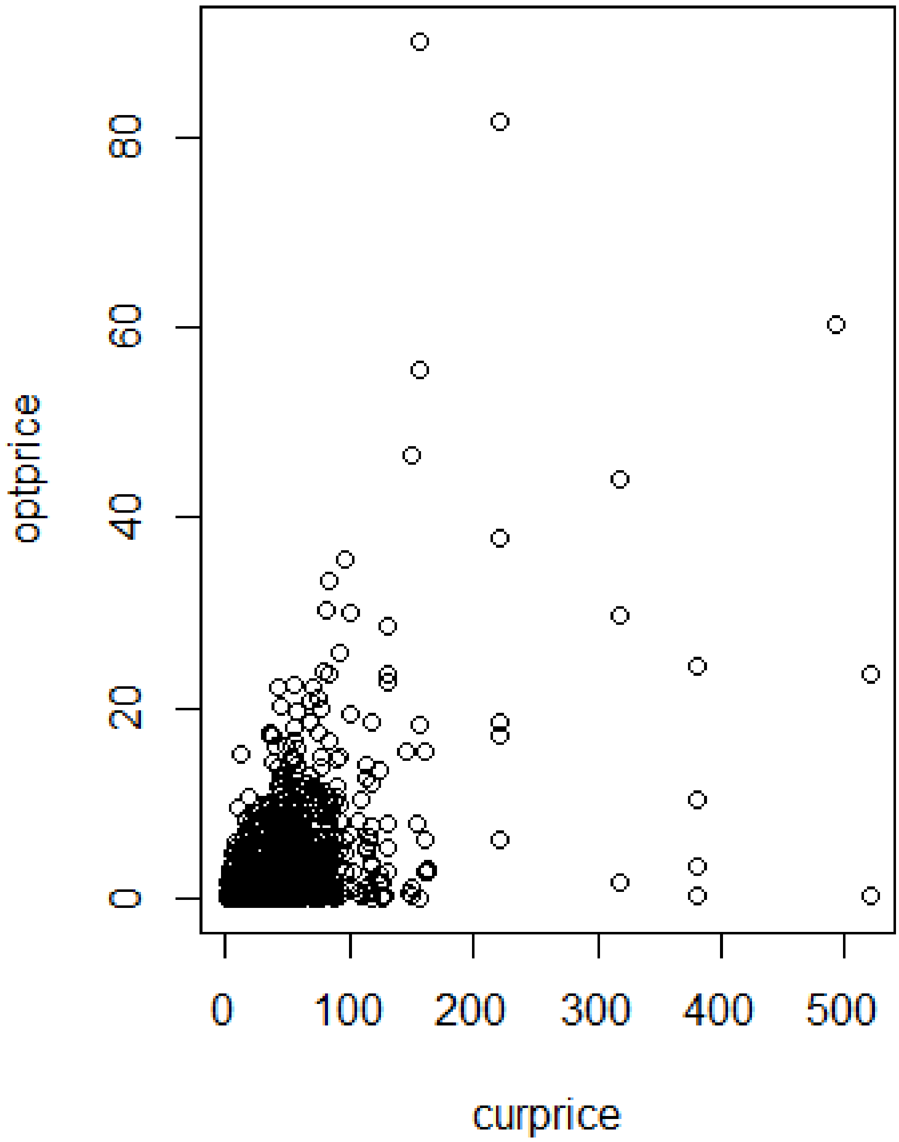
## 2.5 Random Forest Models

The Random Forest model consist of generating 500 new data sets from the original data set. Then a tree is constructed on each of these data sets. The final prediction is then obtained by averaging predictions from all 500 trees. This was tested for up to two predictors, in addition to their log transformations.

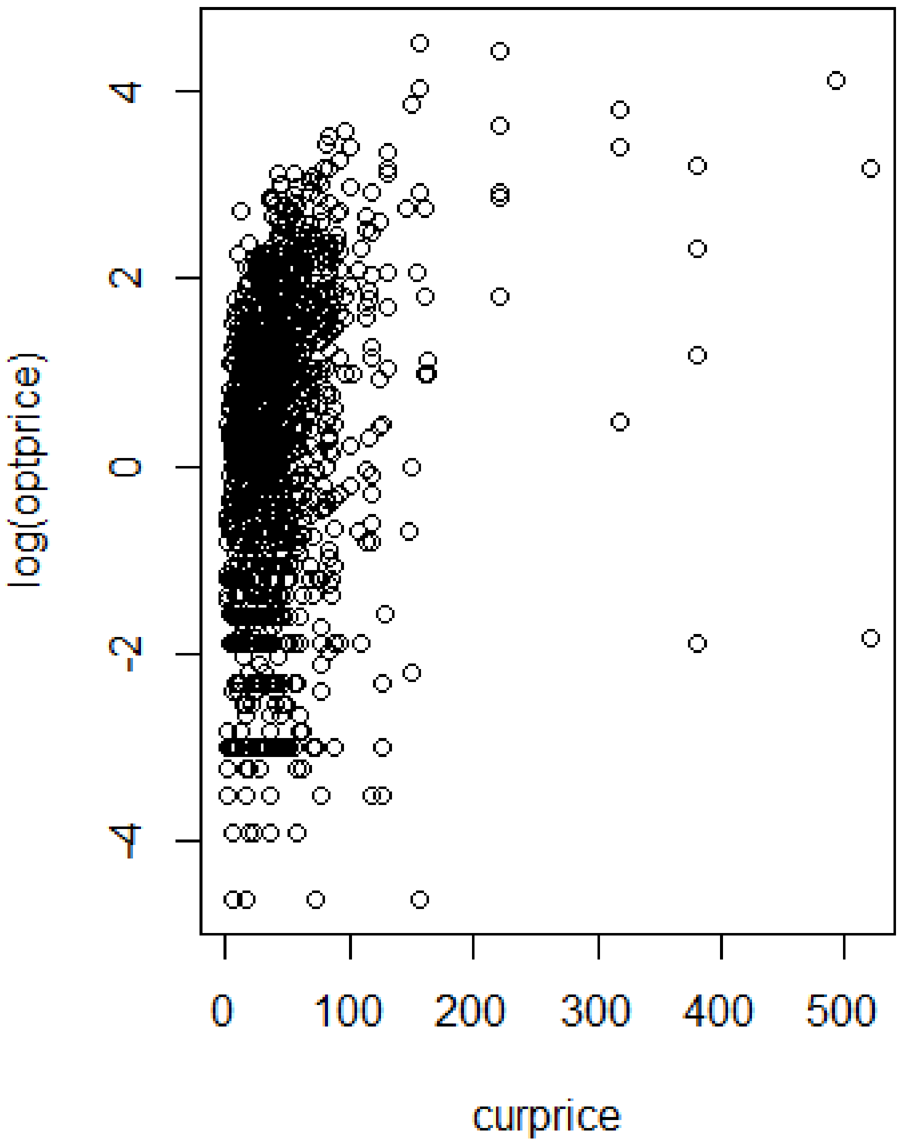
# 3 Plots

## 3.1 Scatterplots

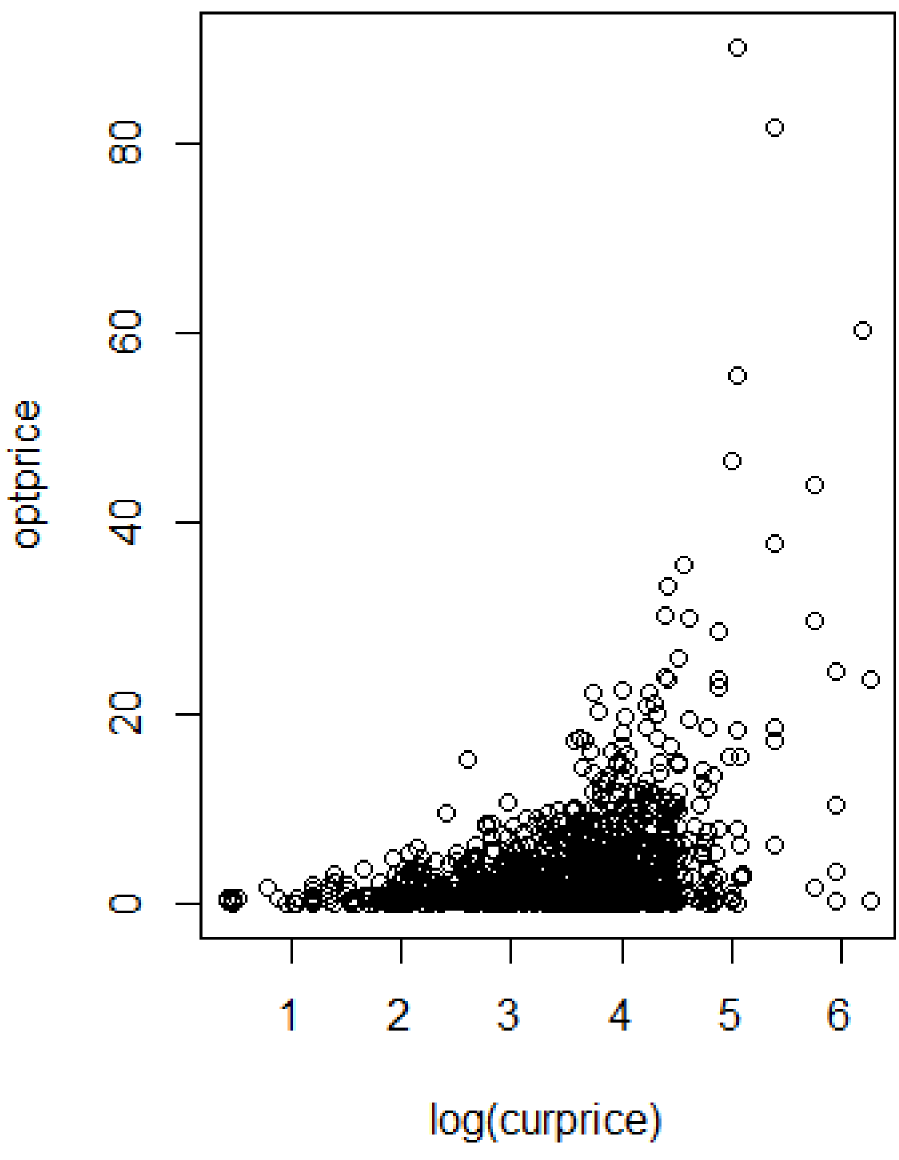
### 3.1.1 Current Price vs Option Price



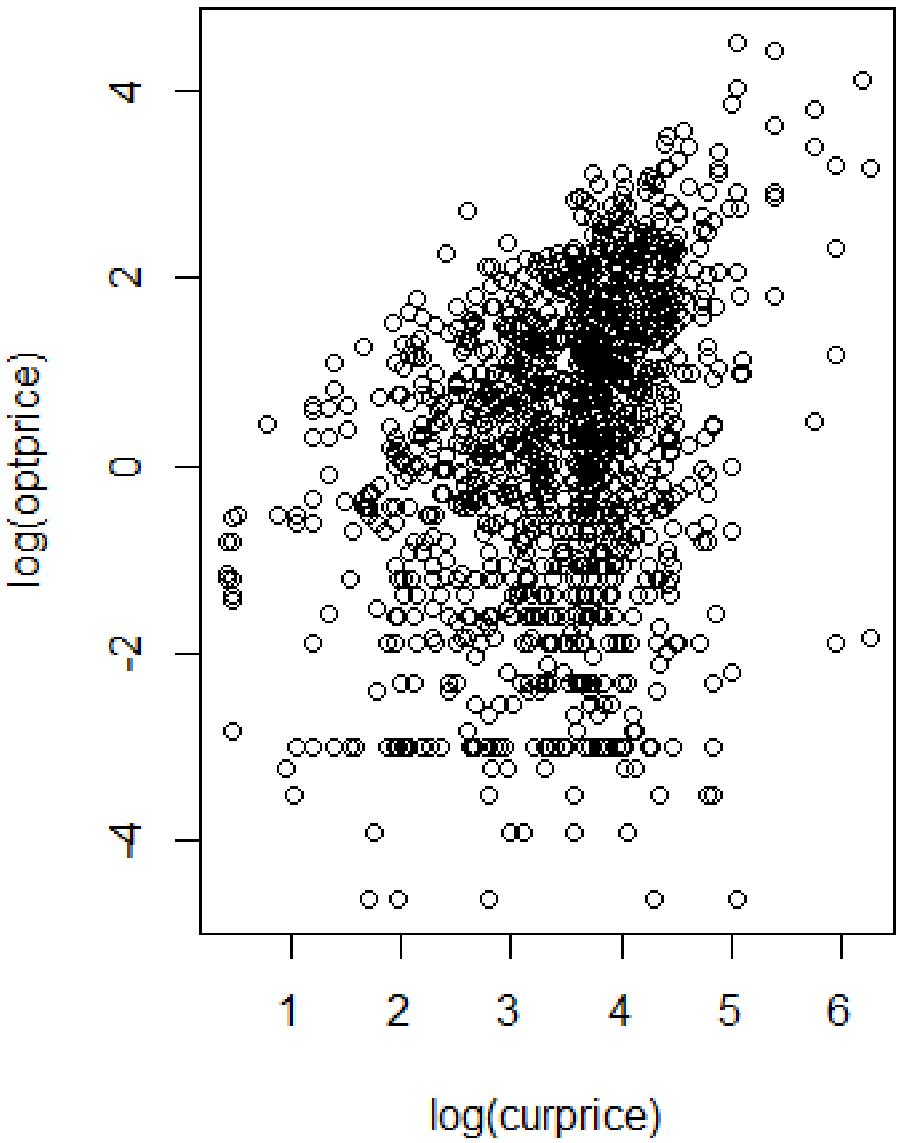
### 3.1.2 Current Price vs log(Option Price)



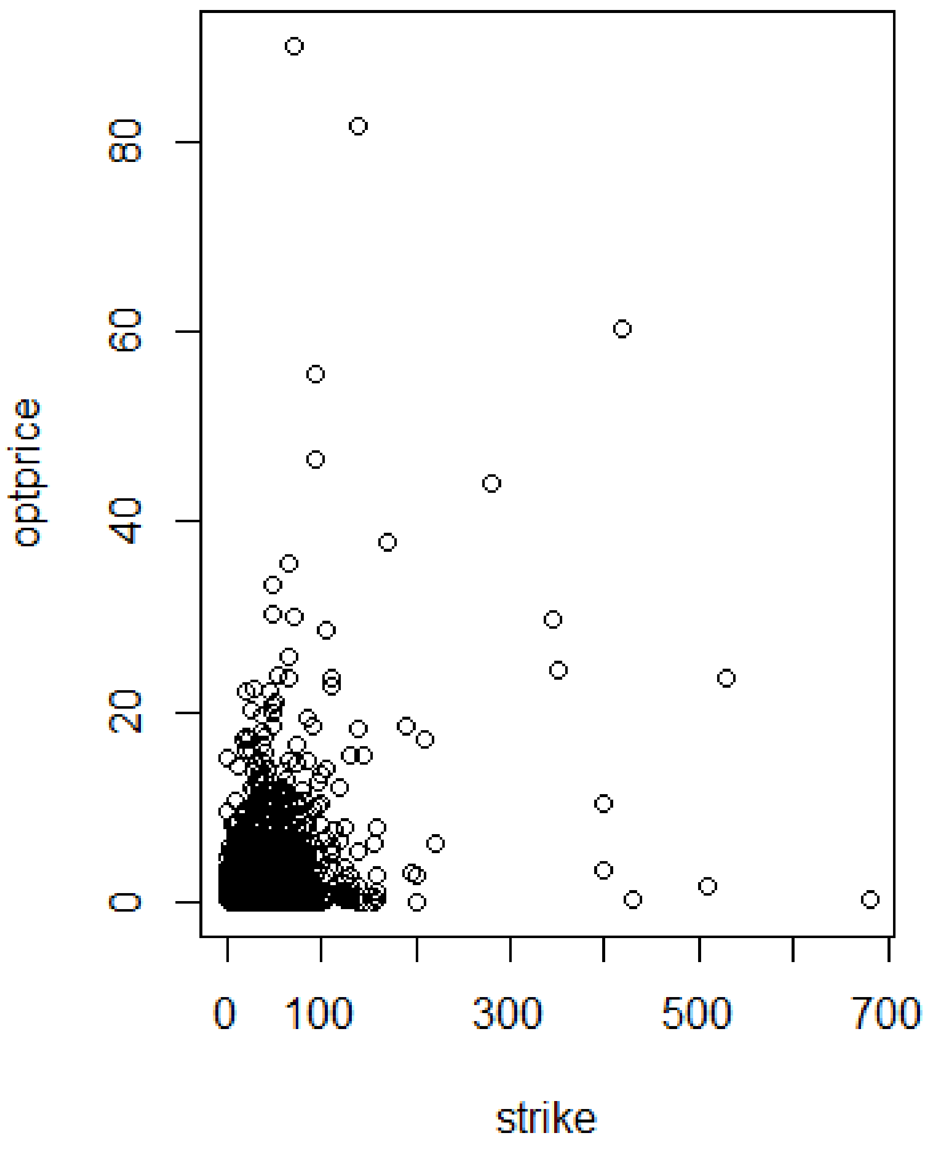
#### 3.1.3 log(Current Price) vs Option Price



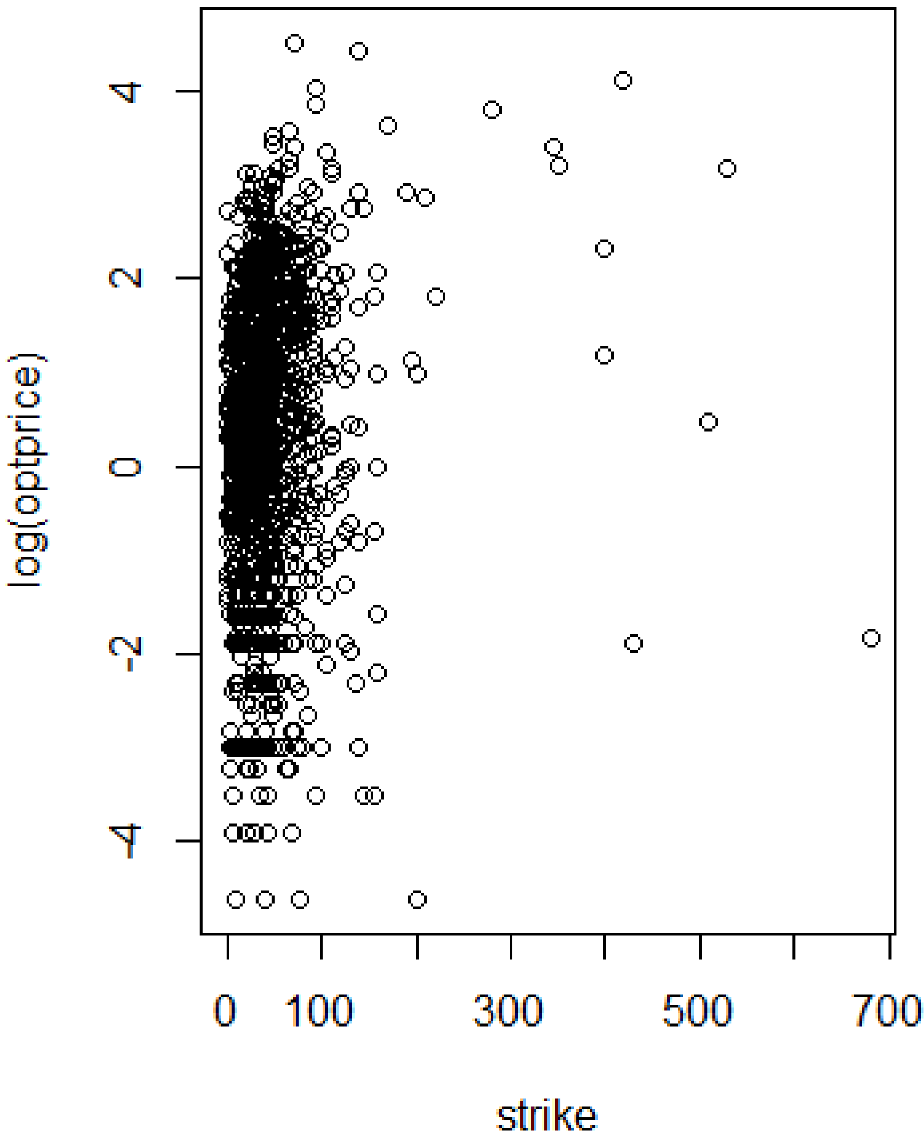
#### 3.1.4 log(Current Price) vs log(Option Price)



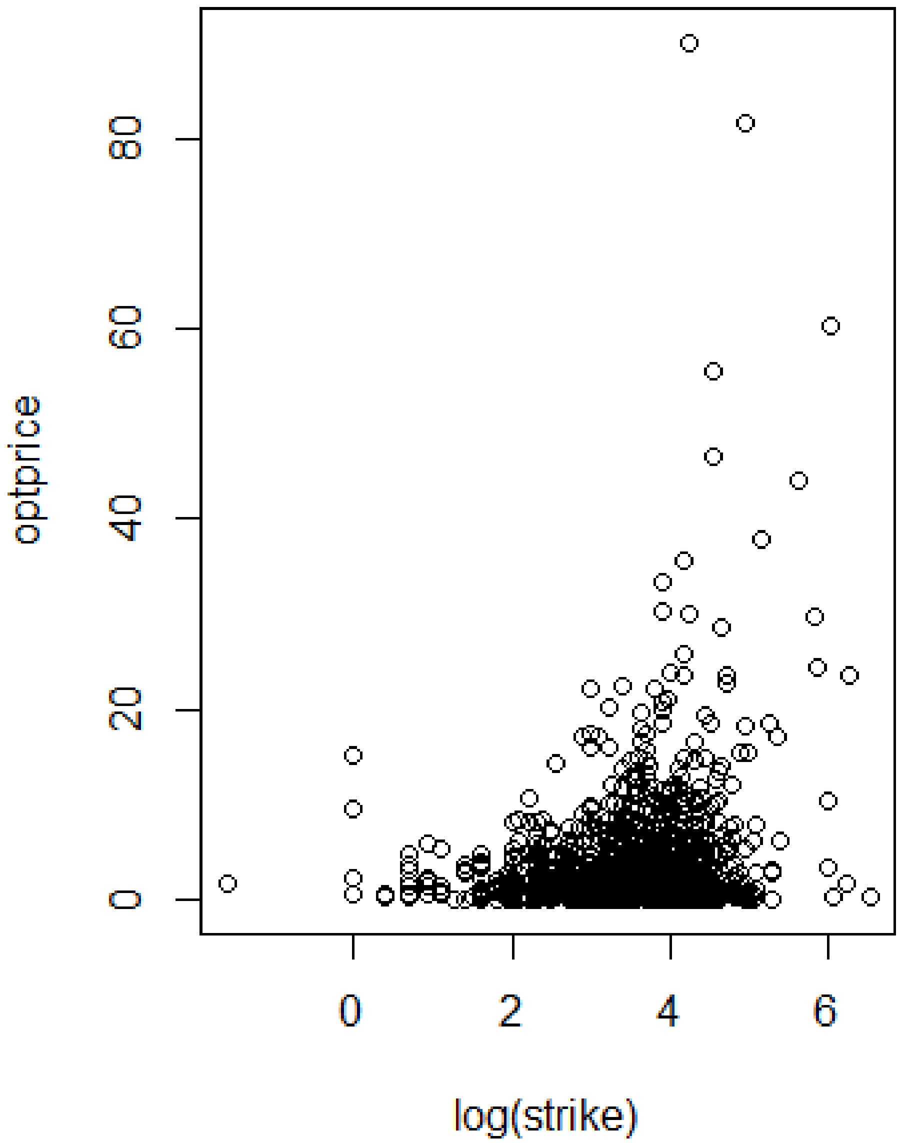
### 3.1.5 Strike Price vs Option Price



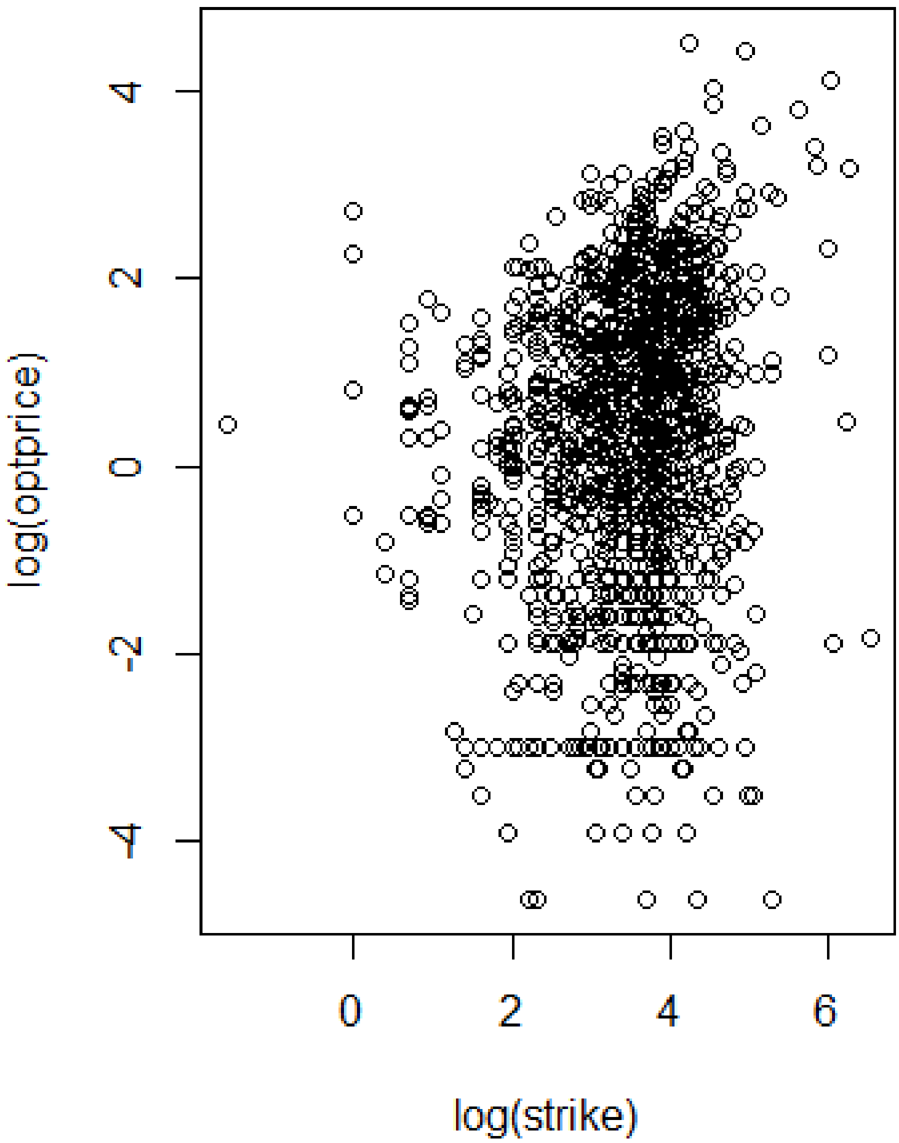
#### 3.1.6 Strike Price vs log(Option Price)



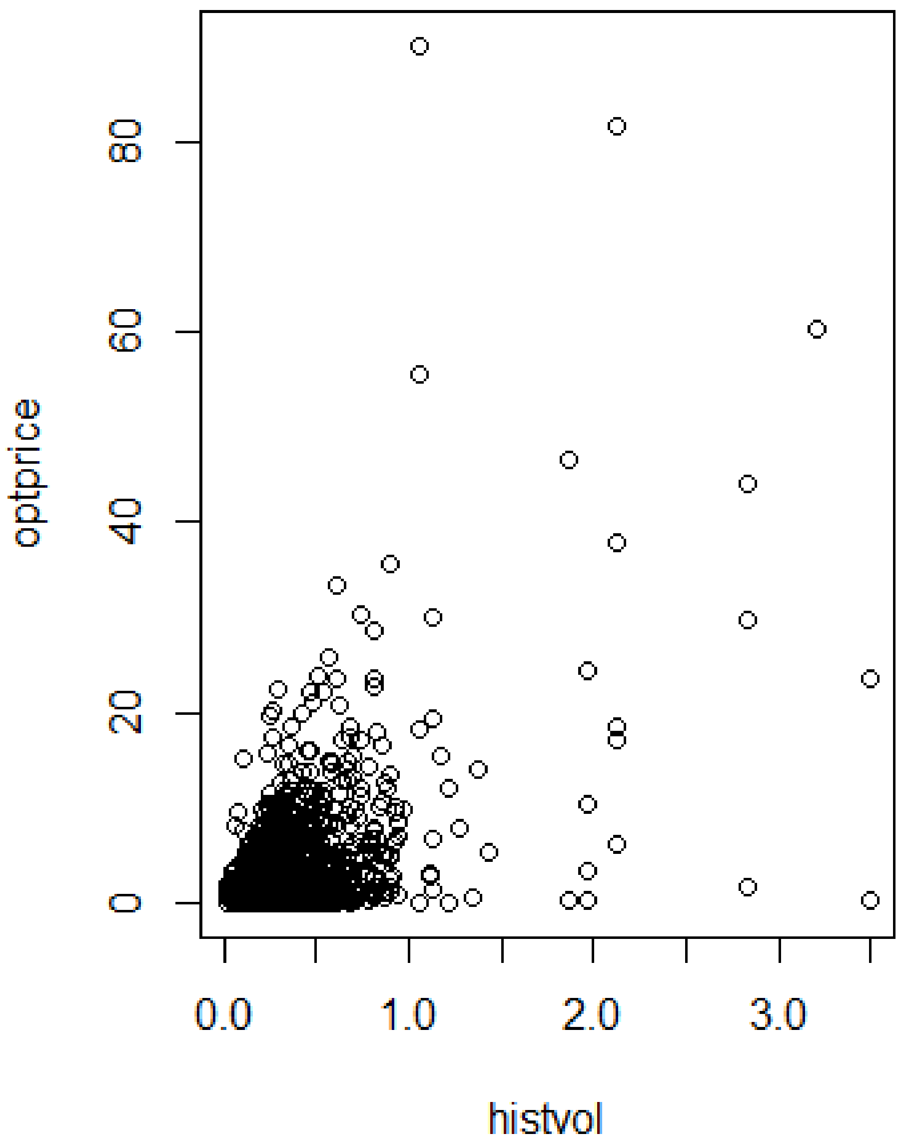
### 3.1.7 log(Strike Price) vs Option Price



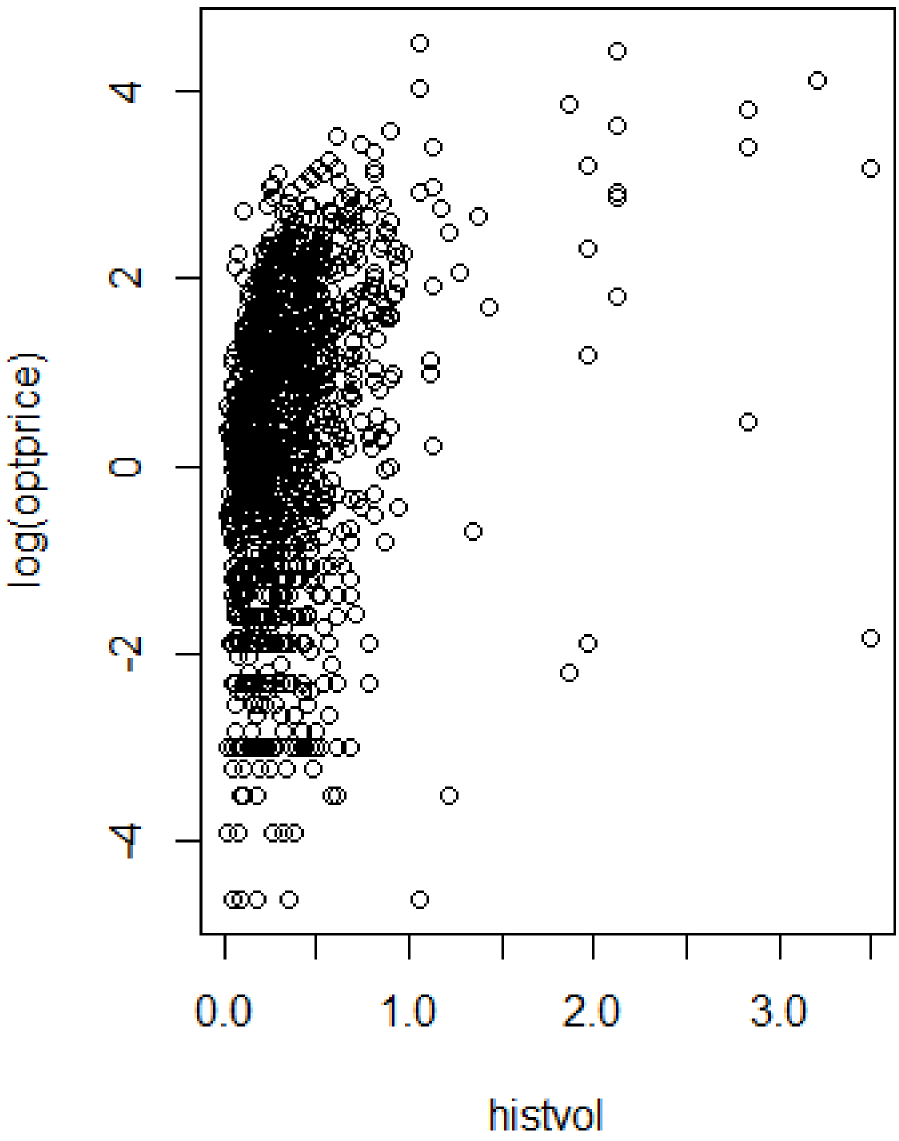
#### 3.1.8 log(Strike Price) vs log(Option Price)



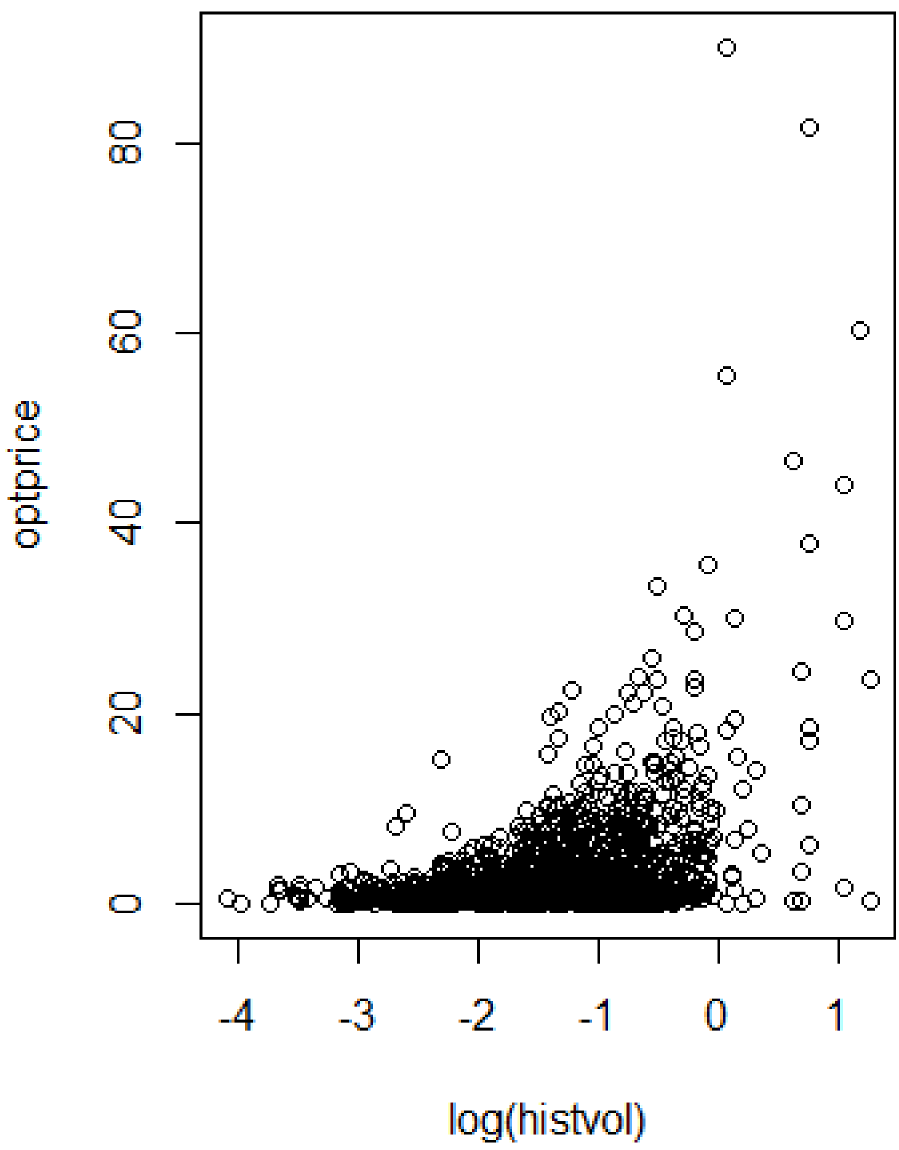
#### 3.1.9 Historical Volatility vs Option Price



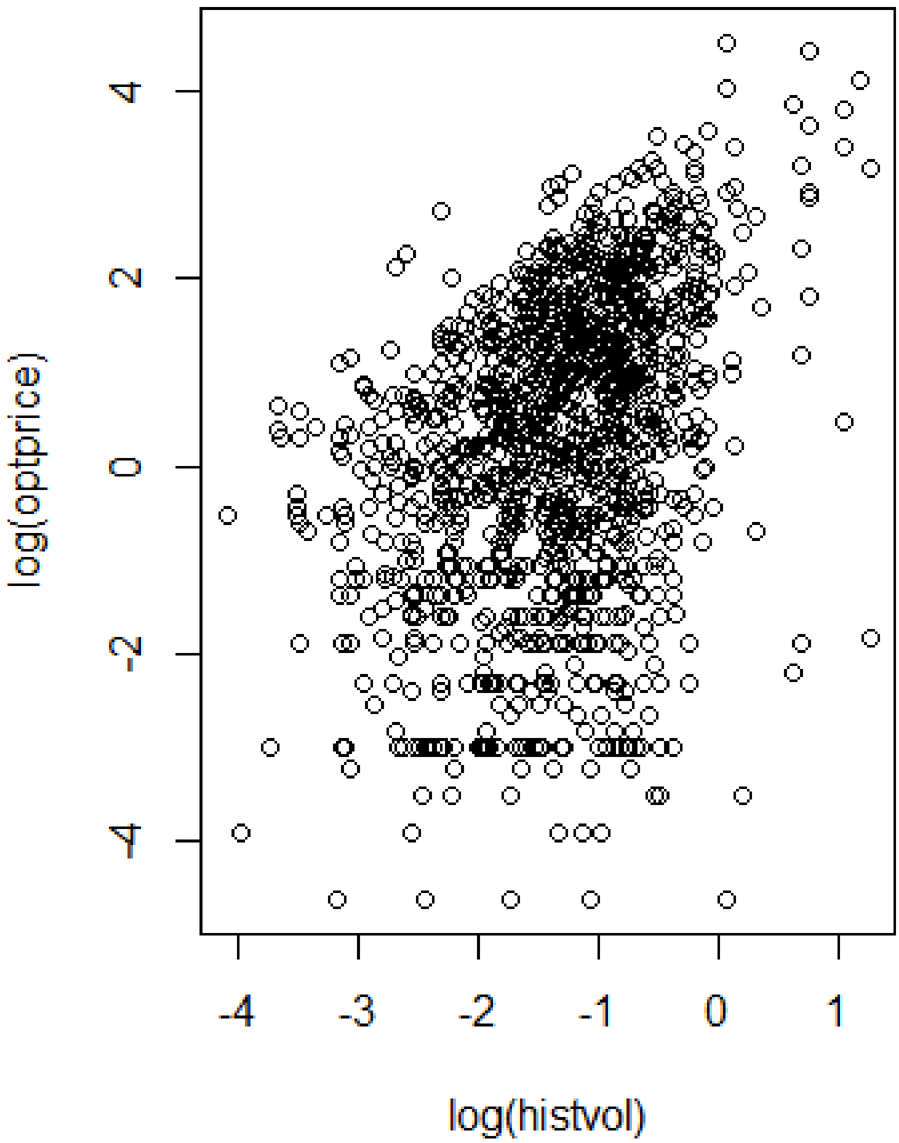
#### 3.1.10 Historical Volatility vs log(Option Price)



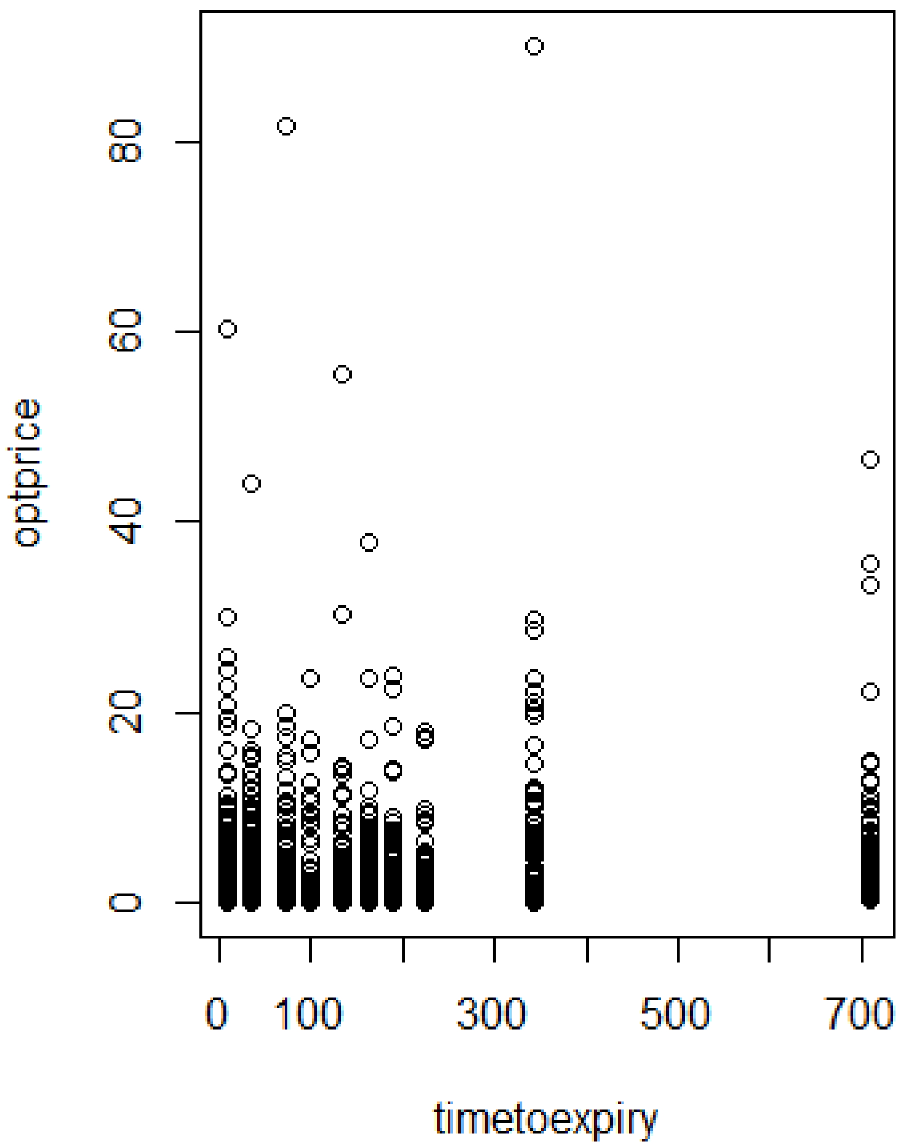
#### 3.1.11 log(Historical Volatility) vs Option Price



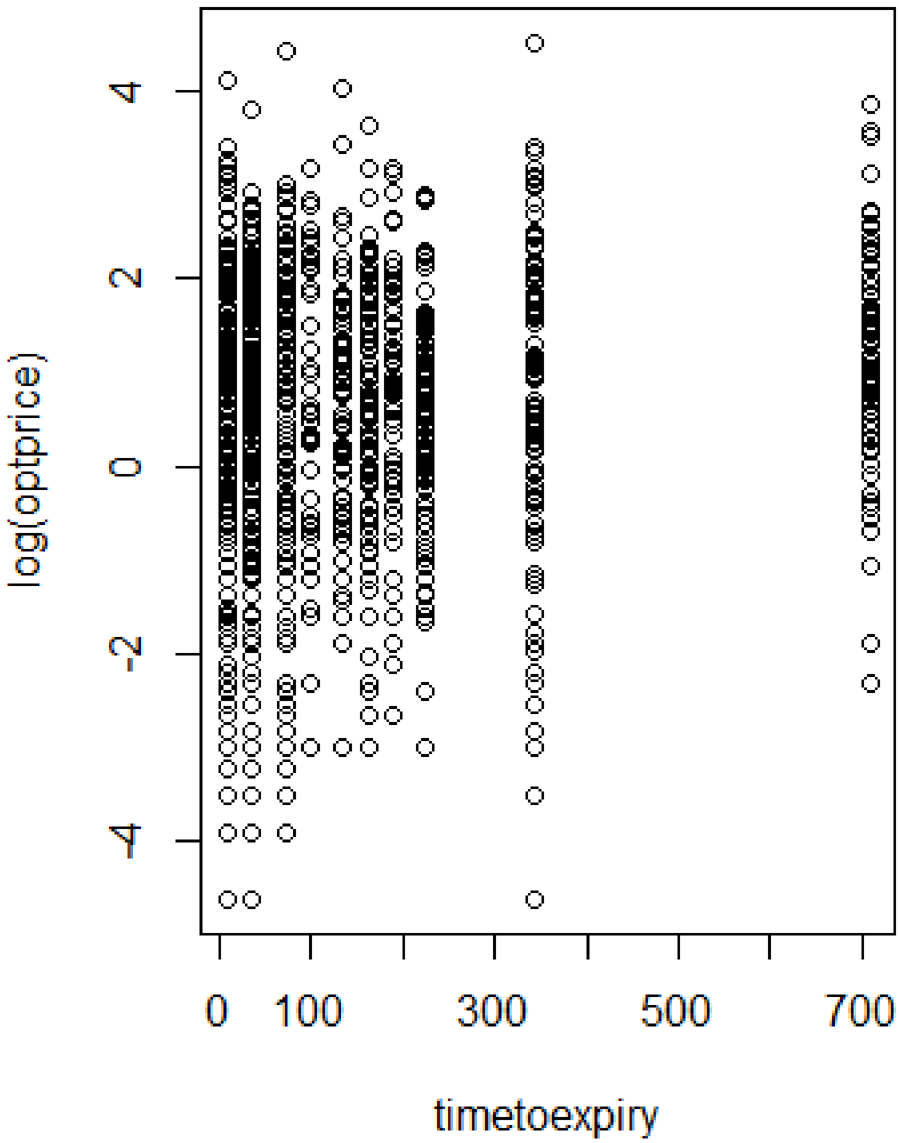
### 3.1.12 log(Historical Volatility) vs log(Option Price)



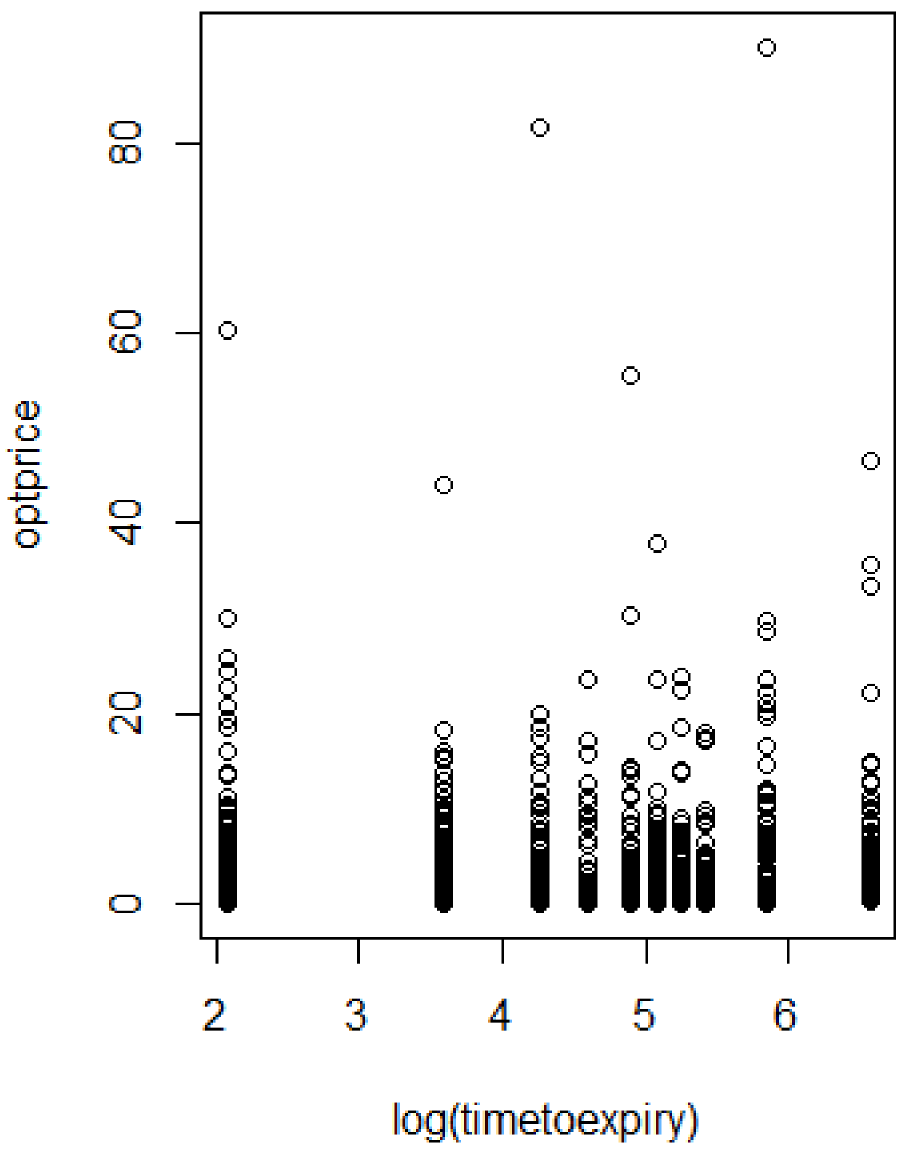
#### 3.1.13 Expiration Time vs Option Price



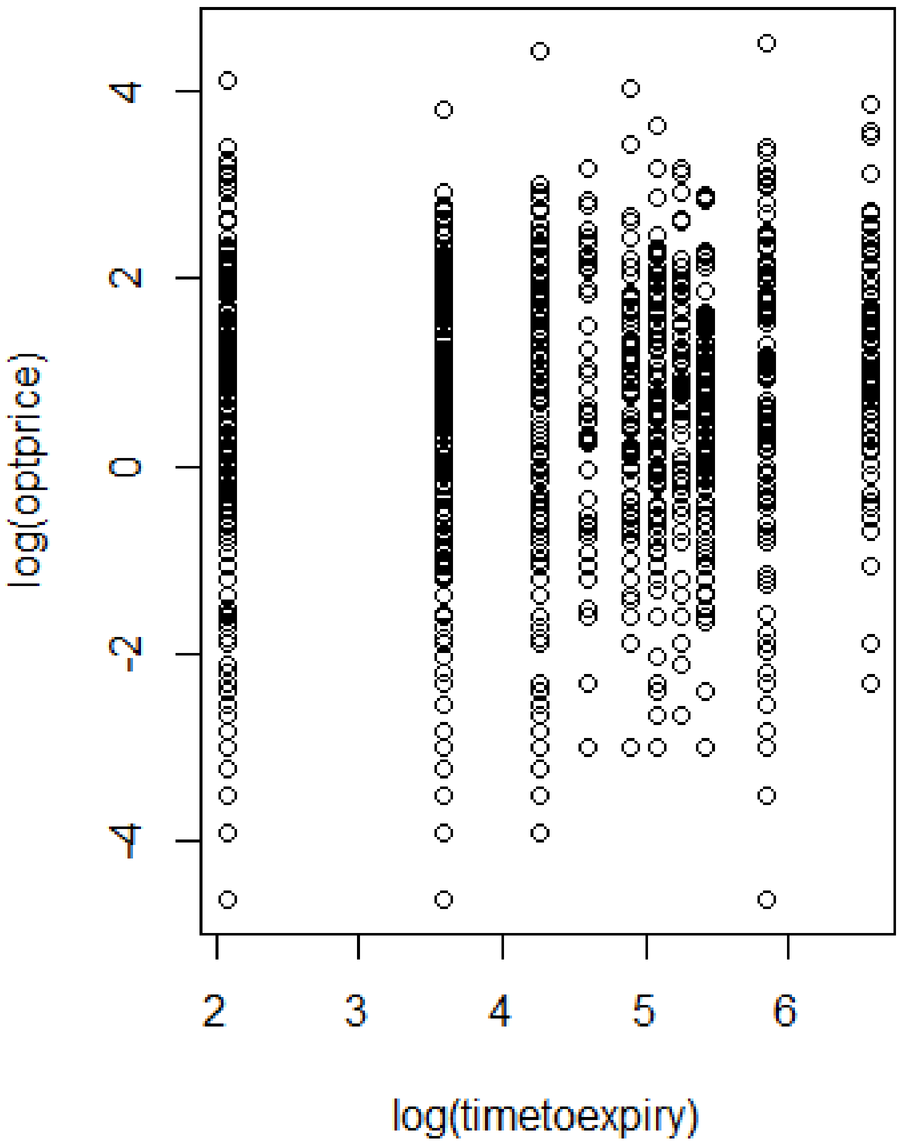
#### 3.1.14 Expiration Time vs log(Option Price)



#### 3.1.15 log(Expiration Time) vs Option Price

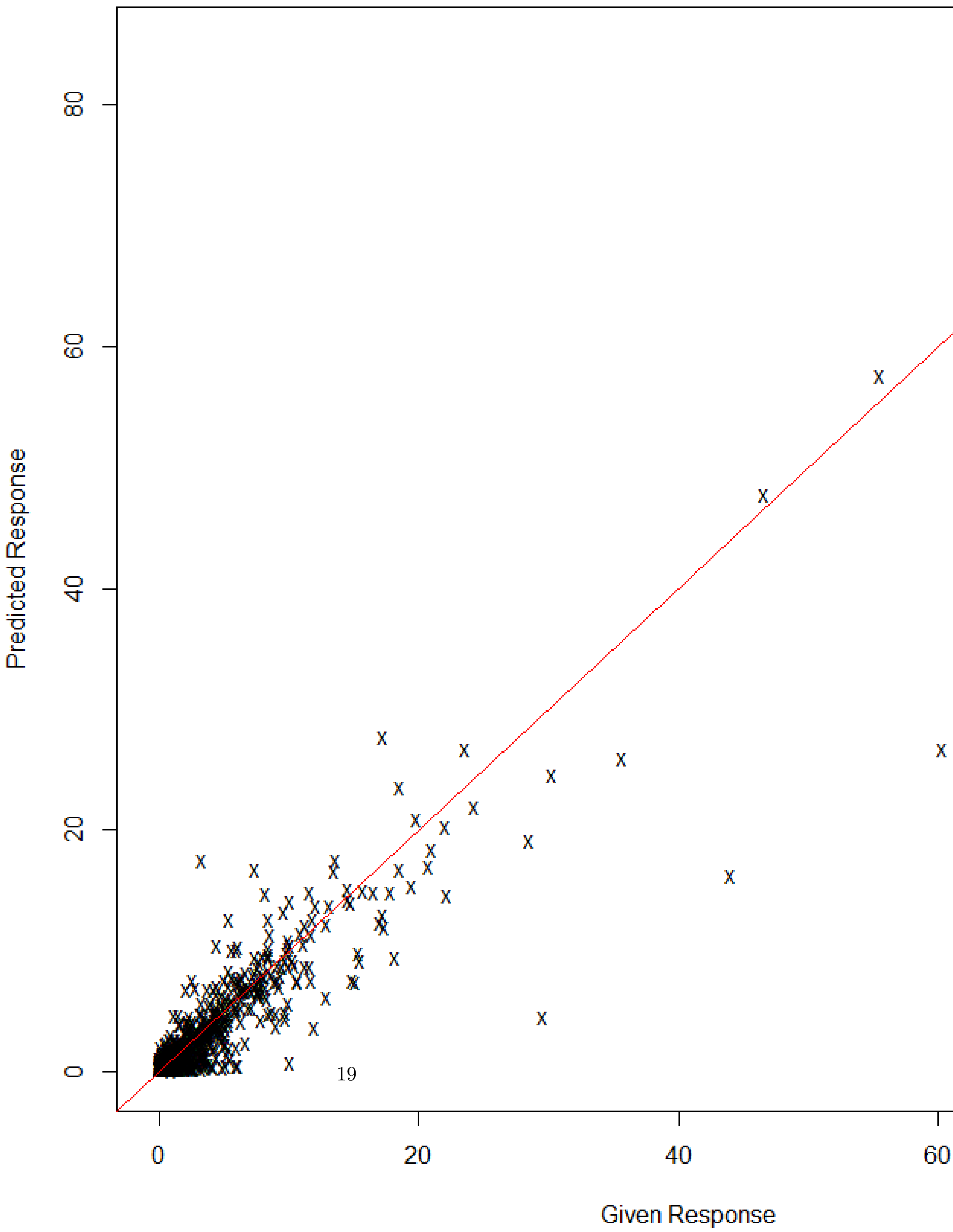


#### 3.1.16 log(Expiration Time) vs log(Option Price)

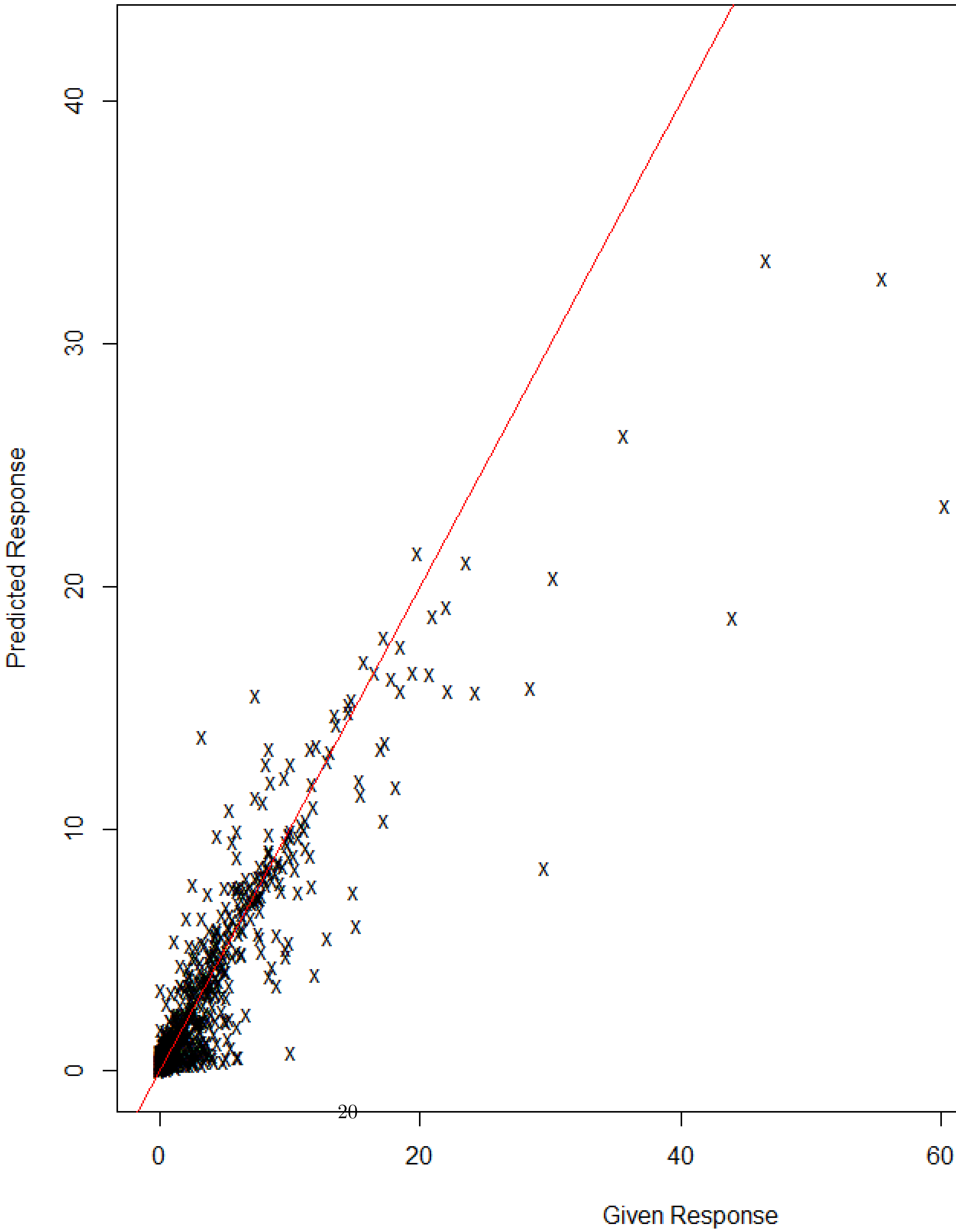


## 3.2 PPR Models

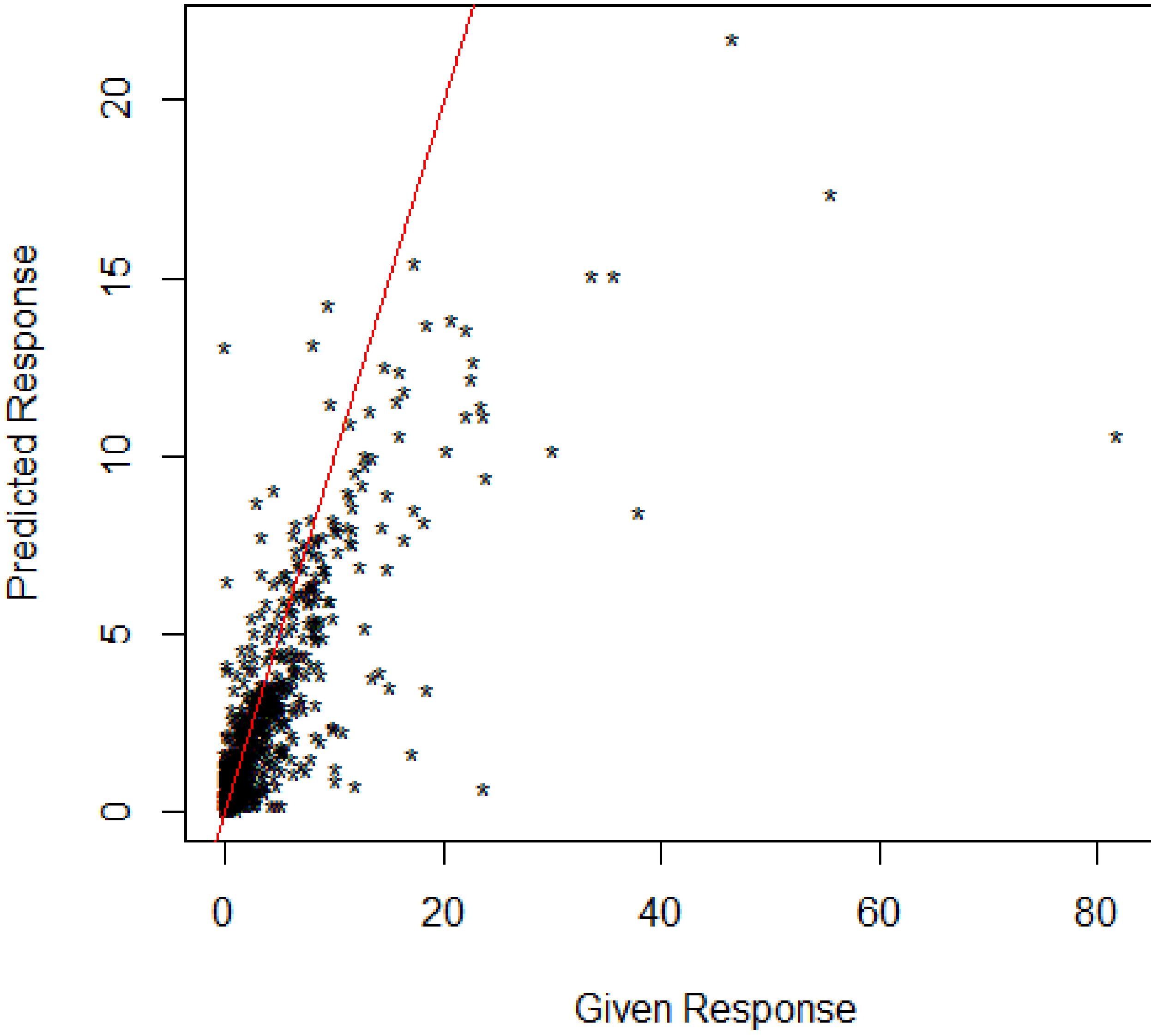
### 3.2.1 log(option price) current price + strike price



### 3.2.2 log(option price) log(current price)+log(strike price)



### 3.3 Random Forest Models



# 4 Results

## 4.1 Linear Regression Models

It's apparent from the scatterplots that there isn't a linear relationship between the expiration time and the option price. A lack of a linear relation ship between the historical volatility and the option price. With regard to the current strike and strike price, they both have a slight linear relationship with the option price. However, it's apparent that both of them don't have a strong enough linear relationship with the option price. As a result, linear regression models aren't a good choice for modeling option prices.

## 4.2 GAM Models

A number of the tested models failed at predicting the response. There were some models that initially appeared to work, however, there when tested on another data set the model failed at predicting the response. This was do to the model being heaviliy overfitted.

## 4.3 Projection Pursuit Regression Models

The "good" models worked best for predicting smaller option prices, This is due to the data posssesing more information for the smaller option prices. The two models both consist of first predicting the log of the option prices. The predictors used in the first model are the current price and the strike price. The predictors used in the second model consists of the log transformations of both the current price and the strike price of the stock in question.

## 4.4 Regression Tree Models

A number of the tested models failed at predicting the response. There were some models that initially appeared to work, however, there when tested on another data set the model failed at predicting the response. This was do to the model being heaviliy overfitted.

## 4.5 Random Forest Models

The model that worked tended to underestimate the option price. This model consisted of predicting the log transformation of the option price based on the log transformations of both the current price and strike price of the stock in question

# 5 Conclusions

In conclusion, the best models for predict call option prices are PPR and Random Forest. Future work would consist of improving the PPR models, using different non-parametric models, and factoring the volume of options sold from the amount avalable.