Option Return Predictability

course "Risk Management"

Prepared by:
Natalia Novosad
Oleksandr Onbysh
Natalia Rudenko
Sevil Smailova
Rostyslav Zatserkovnyi

ABSTRACT

In this project we have applied different machine learning methods in order to explain option returns. Our work is inspired by the paper of Matthias Buchner and Bryan Kelly "A Factor Model for Option Returns". We analyzed S&P Index option returns from January 2015 till May 2016.

I. DATA PREPARATION

First, let's explain here some features that will be used in our analysis.

1. The *delta* of an option is defined as the rate of change of the option price with respect to the price of the underlying asset. Suppose that the delta of a call option on a stock is 0.6. This means that when the stock price changes by a small amount, the option price changes by about 60% of that amount.

2. The *theta* of a portfolio of options is the rate of change of the value of the portfolio with respect to the passage of time with all else remaining the same.

3. The *gamma* of a portfolio of options on an underlying asset is the rate of change of the portfolio's delta with respect to the price of the underlying asset. It is the second partial derivative of the portfolio with respect to asset price.

4. The *vega* of a portfolio of derivatives, is the rate of change of the value of the portfolio with respect to the volatility of the underlying asset.

5. The *time-to-maturity* - is the remaining life of option.

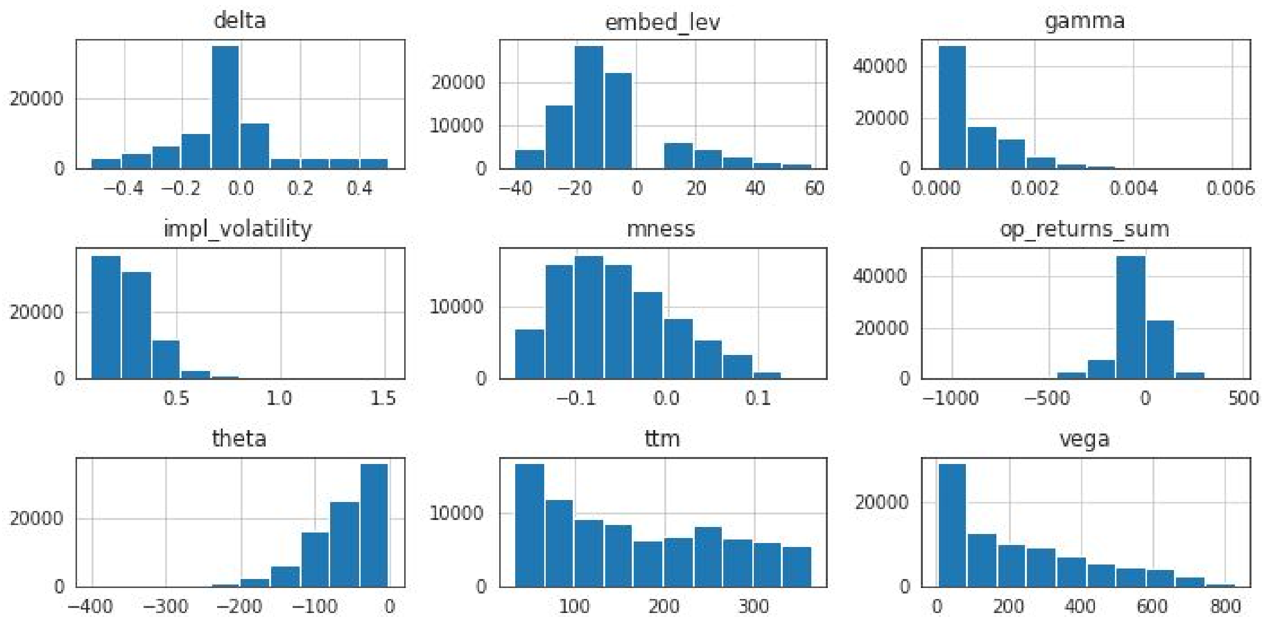
6. The *moneyness* - describes the intrinsic value of an option in its current state. The term moneyness is most commonly used with put and call options and is an indicator as to whether the option would make money if it were exercised immediately.

7. The *embedded leverage* - that is, the amount of market exposure per unit of option.

8. The *embedded leverage* of an option contract is that value of the volatility of the underlying instrument which, when input in an option pricing model

Summary statistics (mean, median, standard deviation, number of observations) for $\mathbf{r}_{\Delta,t,Spot}$ and above mentioned predictors is provided below.

Parameter	$\mathbf{r}_{\Delta,t,Spot}$	**mness**	**ttm**	**embed_ lev**	**impl_vol**	**delta**	**theta**	**gamma**	**vega**
number of observations	87392	87392	87392	87392	87392	87392	87392	87392	87392
mean	-53.10	-0.06	168.09	-8.34	0.27	-0.04	-60.76	0.00	233.60
median	-15.95	-0.064	152.00	-11.85	0.25	-0.025	-51.82	0.00	177.60
std	130.74	0.06	98.21	19.26	0.13	0.19	48.11	0.00	209.01



Pic 1: data distributions

II. FEATURE EXPLANATION

We fitted all the data to linear regression, solved Ordinary Least Squares. Estimated panel regression model is as follows:

$$\mathbf{r}_{\Delta,t,Spot} = \mathbf{-93.45 + 0.13*ttm + 14.36*mness + 1.95*embed\_lev + 210.74*imp\_lvol + 334.56*delta - 0.2*theta - 10335.2*gamma - 0.5*vega}$$

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         op_returns_sum   R-squared:                       0.435
Model:                            OLS   Adj. R-squared:                  0.435
Method:                 Least Squares   F-statistic:                     8413.
Date:                Thu, 04 Jun 2020   Prob (F-statistic):               0.00
Time:                        18:02:50   Log-Likelihood:            -5.2493e+05
No. Observations:               87392   AIC:                         1.050e+06
Df Residuals:                   87383   BIC:                         1.050e+06
Df Model:                           8
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const             -93.4541      1.867    -50.046      0.000     -97.114     -89.794
ttm                 0.1305      0.006     22.667      0.000       0.119       0.142
mness              14.3606     20.382      0.705      0.481     -25.587      54.309
embed_lev           1.9445      0.054     36.121      0.000       1.839       2.050
impl_volatility   210.7380      5.494     38.359      0.000     199.970     221.506
delta             334.5569      2.545    131.472      0.000     329.569     339.545
theta              -0.1963      0.016    -12.283      0.000      -0.228      -0.165
gamma           -1.034e+04    879.815    -11.747      0.000   -1.21e+04   -8610.803
vega               -0.0501      0.003    -16.025      0.000      -0.056      -0.044
==============================================================================
Omnibus:                    29448.362   Durbin-Watson:                   0.194
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           215745.499
Skew:                          -1.430   Prob(JB):                         0.00
Kurtosis:                      10.146   Cond. No.                     9.49e+05
==============================================================================
```

Pic 2: OLS summary statistics

R^2 is 43%, which means that the linear model explains only 43% of variation among the sample. Almost all features (except "moneyness") are significant on 0.01 significance level. Negative intercept for this regression model has no real meaning. It only shows that all predictors could not be simultaneously equal to zero. Interpretation of other coefficients is as follows:

"time-to-maturity" - if time-to-maturity is longer for day, option spot return increases by 0.13.

"moneyness" - if moneyness increases by one unit, option spot return will increase by 14.36. It seems quite intuitive, as intrinsic value increases.

"embedded leverage" - if embedded leverage increases by one unit, option spot return will increase by 1.94.

"implied volatility" - if implied volatility increases by one unit, option spot return will increase by 210.74.

"delta" - if delta increases by one unit, option spot return will increase by 334.56.

“theta” - if theta increases by one unit, option spot return will decrease by 0.2.

“gamma” - if gamma increases by one unit, option spot return will decrease by 10 335.2.

“vega” - if vega increases by one unit, option spot return will decrease by 0.5.

III. MODELING

For further steps we have normalized predictors. We trained several models and provided the results in three metrics: R^2 score on the train data, out-of-sample Mean Squared Prediction Error (MSPE) and R^2_{oos} - out-of-sample R^2 score on test data. The options are daily S&P500 stock index. The Risk Free Rate was taken as a LIBOR rates.

A. Linear models

The performance of the models is summarized below:

1. Linear Regression Model: MSPE of 9 835, R^2 of 43% and R^2_{oos} of 44%.
2. Ridge Regression Model: MSPE of 9 835, R^2 of 43% and R^2_{oos} of 44%.
3. Lasso Regression Model: MSPE of 9 836, R^2 of 43% and R^2_{oos} of 44%.
4. ElasticNet Regression Model: MSPE of 9 835, R^2 of 43% and R^2_{oos} of 44%.

As we see from the statistics above all the models have the same R^2_{oos} and MSPE. This happens because there are little features and all of them are needed. Ridge, Lasso, ElasticNet is used to prevent overfitting, but here the model is not overtitted, it is underfitted. We should try bigger models.

B. Dimensionality Reduction

PCA, PLS used to decrease the dimensionality and extract the most relevant information. But it shows worse performance compared to penalized regressions on the previous step. With roughly the same R^2, MSPE is bigger.

1. PCA with 3 components: MSPE of 10 615, R^2 of 39% and R^2_{oos} of 40%.
2. PCA with 5 components: MSPE of 10 074, R^2 of 42% and R^2_{oos} of 43%.
3. PLS with 3 components: MSPE of 10 046, R^2 of 42% and R^2_{oos} of 43%.
4. PLS with 5 components: MSPE of 9 848, R^2 of 43% and R^2_{oos} of 44%.

The explanation for the results is the following: all the features that we used to predict the option revenues are relevant and explain the revenue to some extent. What is more, the dimension is not that big to cause a curse of dimensionality. So the attempts to reduce the feature space introduced the loss of information and metric decrease.

C. Forests

The performance of the tree-based classifiers is summarized below:

1. Random Forest Regressor: MSPE of 7 272, R^2 of 59% and R^2_{oos} of 59%.
2. Gradient Boosting Regressor: MSPE of 8 733, R^2 of 51% and R^2_{oos} of 51%.
3. “Extra Trees” Regressor: MSPE of 9 587, R^2 of 45% and R^2_{oos} of 46%.

4. Ensemble of all three models: MSPE of 8 307, R^2 of 53% and R^2_{oos} of 53%.

Here, the random forest regressor shows lower MSPE compared to the models above. The extremely randomized tree regressor and gradient boosting regressor use different modifications to the basic random forest algorithm to reduce overfitting, but this leads to underfitting on the test set and all-around worse performance. Building a voting ensemble from the three tree-based models leads to performance "averaging" between three of them.

D. Neural networks

Neural networks show higher R^2. MSPE is also lower compared to previous models. Results are as follows:

With 1 hidden layer : MSPE of 7 033, R^2 of 60% and R^2_{oos} of 60%.
With 2 hidden layers : MSPE of 7 828, R^2 of 55% and R^2_{oos} of 56%.

With application of with stochastic gradient descent:

With 1 hidden layer: MSPE of 6 738, R^2 of 62% and R^2_{oos} of 62%.
With 4 hidden layers : MSPE of 6 258, R^2 of 65% and R^2_{oos} of 65%.

CONCLUSION:

Machine learning methods could be applied for option returns explanation, but the dataset must be bigger. Despite small dataset neural networks showed best performance for explaining variation and also the lowest values of Mean Squared Prediction Error compared to other models. Random Forest showed good MSPE, but explained only 59% of variation in our sample. Penalized regressions did not show better performance compared to linear models. Application of PSL with 5 components has not improved the results. For further engineering we would like to combine methods in order to obtain better performance.