Addressing blockchain privacy and efficiency challenges in mobile environments: an optimization strategy for lightweight Application of data analysis in the field of investment

Zerui Ke

Kings College London, London, WC2R 2LS, UK

Abstract. As advancing technology makes a viable option to value a stocks and also determine their economic value has become increasingly important.In the global securities market, changes in stock prices can have a significant impact on changes in the global economy. For enterprises, stock price fluctuations reflect their current financial situation, changes in industry supply and demand, and changes in the international macro environment. [1]When people interact the internet information system, large data which is based on this system will be generated by the people. To some extend, bid data can reflect the law and problems for some certain field, and make some predictions with a certain range. In order to analysis the law of stock price, we need to combine the stock itself and the changing market factors so that we can obtain accurate price forecast. When Solving nonlinear and the complex problem, the LSTM model has better performance to process the data[2]. In addition, the CNN model is also one of the most potential model which can process a large amount of data to improve the accuracy of prediction[3]. This paper will select six indicators to conduct establishment of arbitrage pricing model.Trying to optimize the model and find the factors which are more correlated with stock price in these six factors.We first set up a simple linear regression model and then use stepwise function.Secondly, we set up ridge regression and lasso regression and delete some of these factors. Then we randomly select 70percent of the from the compete data, and set up an evaluation index of the computational model. And finally we find in these data, step regression perform best comparing with the ridge regression and lasso regression, and degree of this model for the EBAY can reach over 90 percent.

Keywords: RStudio; eBay; regression; lasso; ridge.

1. Introduction

EBAY is an American online auction and shopping website. It is a formal and international e-commerce platform that allows people around the world to buy and sell items online, as long as the items do not violate the law or on eBay's prohibited sale list within that time, it can be listed for sale on eBay.

2. Pricing model set up

In this part, eBay's stock price from June 30, 2006 to June 30, 2015 will be chosen. And also the end of each quarter are selected to set up this model.

3. Factors selection

Book value per share (BVPS) is the ratio of equity available to common shareholders divided by the number of outstanding shares.

Earning per share is calculated as a company's profit divided by the outstanding shares of its common stock.

P/B ratio is price-to-book ratio. It is used to judge the rationality of the stock price and the buy point. The main significance of price-to-book ratio is that it can be used to distinguish the rationality of the book value of a company's stock price relative to the stock price, so as to facilitate comparision and judgement.

Standard&Poor's 500. Standard & Poor's 500 index, is a market capitalization weighted index of 500 major public companies in the US.

U.S. CPI month-on-month rate of change. The U.S. CPI (Consumer Price Index) reflects the price level in the United States.

Michigan Consumer Confidence Index. It was the Survey Research Center of the University of Michigan in the United States that first complied the consumer confidence index in order to study the impact of consumer demand on the economic cycle, and then some European countries began to establish and compile the consumer confidence index. The Michigan Consumer Confidence Index is a response to regular surveys conducted by researchers at the University of Michigan of consumers' views about their personal finances and the state of the national economy.

3.1 Model set up

Based on the arbitrage pricing model, we can get the relationship between adjusted price for this stock and other six factors

$$adprice = a_0 + a_1 bvpershare + a_2 EPS + a_3 pbratio + a_4 sp500 + a_5 cpi + a_6 mcsi + \varepsilon$$

Where adprice is the adjusted share price for stock, bvpershare is the BVPS, book value per share, EPS is the earning per share, pbratio is the P/B Ratio, sp500 is the S & P 500, cpi is the U.S. CPI (Consumer Price Index), mcsi is the Michigan Consumer Confidence Index, a_i(i= 1 ,2...6) is the impact if each factor on the adjusted share price, and ε is the random interference.

3.2 Result analysis

We first use R studio to test the model and plot the model
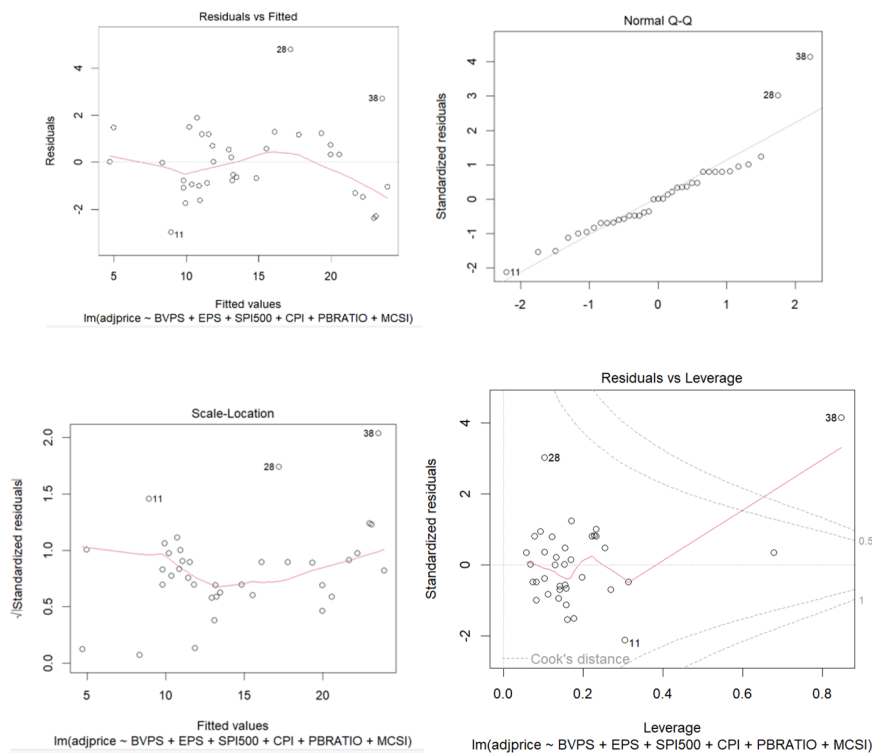
We can find there exist outliers 28th and 38th and the result for this model.

```
Call:
lm(formula = adjprice ~ BVPS + EPS + SPI500 + CPI + PBRATIO +
    MCSI, data = xianxinghuigui)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9483 -0.9908  0.0238  1.1709  4.7809

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -64.537964  13.114758  -4.921 2.91e-05 ***
BVPS          0.111168   0.127848   0.870 0.391460
EPS          -0.330867   0.613583  -0.539 0.593702
SPI500        0.003349   0.002788   1.201 0.239017
CPI           0.267546   0.061950   4.319 0.000158 ***
PBRATIO       1.862117   0.710938   2.619 0.013688 *
MCSI          0.104989   0.049332   2.128 0.041647 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.668 on 30 degrees of freedom
  (因为不存在，1个观察量被删除了)
Multiple R-squared:  0.9238,    Adjusted R-squared:  0.9086
F-statistic: 60.65 on 6 and 30 DF,  p-value: 1.976e-15
```

According to the results, we can find the economic significance of each non-systematic factor to eBay's stock price seems to be the SPI500. The price to book ratio seems to be the biggest impact on eBay's adjusted stock price. In theory, if the price-to-book ratio increase by 1 the price for the eBay' stock will increase by around 1.86 US dollars. For the other factors, BVPS increased by 1 percent then eBay's stock will increased by 0.11 US dollars, EPS increased by 1 , stock price will decrease by 0.33 US dollars, for SPI500, it increased by 1, then the price will increase by 0.003 US dollars, if CPI is increased by 1, then the price will increased by around 0.26 US dollars. If MCSI increased by 1, then the price for stock will increased by 0.105.

What is puzzling is that the Earnings per share has a negative correlation with eBay's share price. , Earnings per share represents the confidence for investors to invest on the company 's futures, so it might because of the time of interception is too short and also we just select quarterly instead of monthly.

We also find that compared to other factors, SP500 has only a small effect on the results.First of all, the market index represents the price rise and fall of most stocks in the stock market, but not every stock will rise or fall with it. Secondly, the economic significance of the S & P 500 is small or even negligible.

Finally we can find the adjusted r squared is 0.9086 which indicates that over 90percent of the variance of the dependent variable being studied is explained by the variance of the independent variable.

```
Call:
lm(formula = adjprice ~ BVPS + EPS + SPI500 + MCSI, data = xianxinghuig
ui)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2724 -0.9995 -0.4133  0.5870  6.3429

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.920525   3.047661  -2.927  0.00625 **
BVPS         0.338438   0.123909   2.731  0.01018 *
EPS          0.374252   0.728230   0.514  0.61084
SPI500       0.012480   0.001903   6.558 2.18e-07 ***
MCSI         0.016669   0.054929   0.303  0.76350
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.057 on 32 degrees of freedom
  (因为不存在，1个观察量被删除了)
Multiple R-squared:  0.8765,    Adjusted R-squared:  0.861
F-statistic: 56.76 on 4 and 32 DF,  p-value: 4.422e-14
```

To consider whether EPS really has a negative correlation with eBay's share price, we try to set up the new model by deleting other factor. PB ratio is the ratio of share price per share to net asset per share which is shown as share price divided by BPS, and return on equity is PB divided by PE or EPS divided by BPS which means PB ratio is share price*Earning per share/return on equity , so we guess there has some linear relation between EPS and PB ratio. CPI as the Consumer Price Index which can impact on people consuming. The rise of the CPI is bound to make people make positive feedback on shopping, so when regression CPI could dominate and influence other data. When we delete PB ratio and CPI to set up the new model, according to the picture, we find now the Earning per share has a positive correlation with the adjusted price which indicates EPS is increased by 1 dollar, then the adjusted price will be increased by 0.374252. BVPS and EPS are under the significance level of 10 percent. This solve problems above. Earning per share is positive correlated, if we set up the model without factor PBRATIO and CPI.

```
adjprice ~ BVPS + EPS + SPI500 + CPI + PBRATIO + MCSI

          Df Sum of Sq     RSS    AIC
- EPS      1     0.809  84.319 42.477
- BVPS     1     2.105  85.615 43.041
- SPI500   1     4.017  87.527 43.858
<none>                   83.510 44.120
- MCSI     1    12.608  96.118 47.322
- PBRATIO  1    19.097 102.607 49.740
- CPI      1    51.920 135.430 60.009

Step:  AIC=42.48
adjprice ~ BVPS + SPI500 + CPI + PBRATIO + MCSI

          Df Sum of Sq     RSS    AIC
- BVPS     1     2.073  86.393 41.375
<none>                   84.319 42.477
- SPI500   1     5.153  89.472 42.671
+ EPS      1     0.809  83.510 44.120
- MCSI     1    12.657  96.976 45.651
- PBRATIO  1    18.288 102.607 47.740
- CPI      1    52.207 136.526 58.307

Step:  AIC=41.38
adjprice ~ SPI500 + CPI + PBRATIO + MCSI

          Df Sum of Sq     RSS    AIC
<none>                   86.393 41.375
- SPI500   1     5.958  92.350 41.843
+ BVPS     1     2.073  84.319 42.477
+ EPS      1     0.778  85.615 43.041
- MCSI     1    12.346  98.738 44.318
- PBRATIO  1    17.107 103.499 46.060
- CPI      1    76.498 162.891 62.840
> |
```

Then we use stepwise function in R studio to select the relevant factors among these factors. According to the table, comparing with the first model with factor BVPS, EPS, SPI500, CPI, PBRATIO, MCSI, we delete the factor BVPS and EPS which also indicates that PBRATIO has impact on EPS when set up the model. BVPS might contain the data of EPS.

```
x=model.matrix(adjprice~.,data=data)[,-1]
y=data$adjprice
ridge=glmnet(x=model.matrix(adjprice~.,data=data)[,-1],y=data$adjpri
plot(ridge,xvar='lambda')
```

We set up two sets for ridge regression x contains the data for all factors, y contains the data for adjusted price. Then we perform the ridge regression on the data and plot the model.
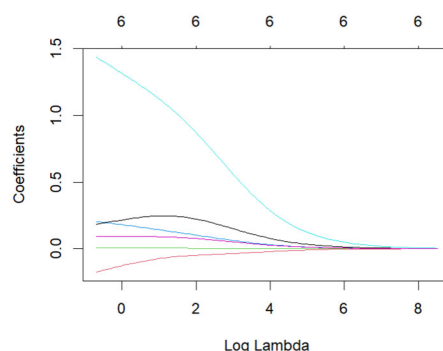
According to the graph, we can see with log Lambda increase, the coefficient of all variables tend to zero.
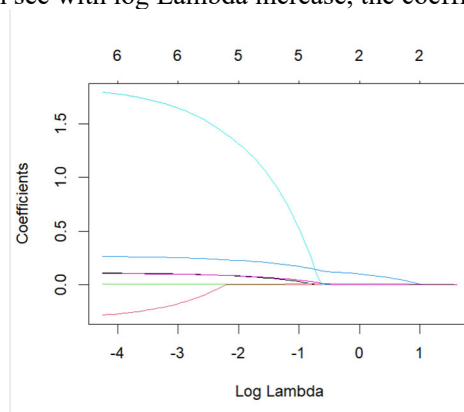
Using lasso function and plot the lasso regression on the data.Same as the graph for ridge regression all variable tend to be 0 when log lambda increases. But some variables go to zero quicker than other variable.
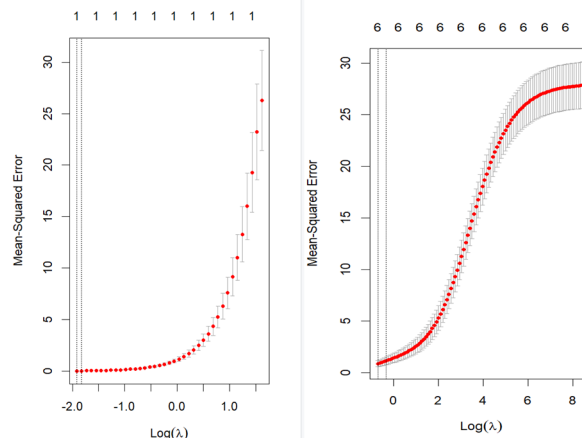
Lambda selection is performed by using cv.ridge function for cross-validation. According to the graph, we can see for the ridge regression, when log lambda is around-2, the average error for the cross-validation is minimal. For lasso regression, when log lambda is around-1, the average error for the cross-validation is minimal.

```
> reg_metric=function(true,predict)
+    {RMSE=sqrt(mean((true-predict)^2))
+            MAE=mean(abs(true-predict))
+            MAPE=mean(abs((true-predict)/true))
+            R2=1-sum((true-predict)^2/sum((true-m
ean(true))^2))
+            return=(list(RMSE=RMSE,MAE=MAE,MAPE=M
APE,R2=R2))
+ }
> |
```

Then set up an evaluation index of the computational mode.

lasso_metric	list [4]	List of lengt.
RMSE	double [1]	0.1988973
MAE	double [1]	0.1667981
MAPE	double [1]	0.01748281
R2	double [1]	0.9989512

ridge_metric	list [4]	List of lengt...
RMSE	double [1]	0.5834361
MAE	double [1]	0.4828045
MAPE	double [1]	0.04013662
R2	double [1]	0.9909751

step_metric	list [4]	List of lengt...
RMSE	double [1]	7.349928
MAE	double [1]	5.871419
MAPE	double [1]	0.5112555
R2	double [1]	-1.98387

We can see for these four evaluation index, step regression perform best comparing with ridge regression and lasso regression, it might we have only 6 factor when we using regression . However ridge regression perform a bit better than the lasso regression.

4. Conclusion

In our study, we first make simple linear regression analysis on EBAY stock price and our arbitrage pricing model is constructed by making further stepwise selection, ridge regression, lasso regression and model comparison. Generally speaking, the relationship between the adjusted share price and other factors confirm to common sense, but there still exist some exceptions through this study, such as when PB ratio and CPI exist, EPS which has negative correlation between the share price, and also for some special cases, the model can not reasonably predict .For this

study, we randomly select 70percent of the from the compete data as the training data and rest of thirty percent as the test data. In the future study, i hope we can improve this arbitrage pricing model by adding other systematic factors or adding other period of time to make the analysis of this model more comprehensive so that we can predict the result more accurately. By further analysis of both eBay or other similar applications such as Amazon market positions and industrial features which indicates that the these systemic factors has influence on stock price for the eBay. Based on the arbitrage pricing mode, the correlation between systematic factors and the market on eBay can be described, which also indicates that this model can be applied in the US stock market. This meas that the US stock market has developed quite well to reach the requirement of full competition and there are still few opportunities in the market.

References

[1] J. Chandler, D. Shapiro Conducting clinical research using crowdsourced convenience samples Annual Rev Clin Psychol, 12 (1) (2016), pp. 53-81

[2] W. Li, D.M. Becker Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling Energy, 237 (2021), p. 121543

[3] B. Wu, L. Wang, S. Wang, Y.R. Zeng, Forecasting the US oil markets based on social media information during the COVID-19 pandemic Energy, 226 (2021), Article 120403