BUFN 745 Problem Set 2

Spring 2023

Professor Alex He

Due Monday, 3/6/2023 at 11:59pm EST

Please submit an electronic version on Elms

Use Python and Google Colab to answer the following questions. You may write your text responses either in the document or in the Colab (Jupyter) notebook. Please use the template Python file (Problem Set 2 template.ipynb) in the shared Google Drive folder or on Elms to load the data and guide you through the questions.

1 Earnings Surprises and Stock Prices [55 Points]

In this question you will use data to analyze the response of stock prices to earning surprises. To measure new information in the earnings announcement, we compare the earning announcement e_{tk} for company k in quarter t with the corresponding analyst forecast $\hat{e}_{tk}$. The analyst forecasts is defined as the median forecast among all the analysts that make a forecast in the last 45 (trading) days before the earning announcement. If an analyst made multiple forecasts in this time horizon, we consider the most recent one.

The measure for earnings surprise equals to:

$$s_{tk} = \frac{e_{tk} - \hat{e}_{tk}}{p_{tk}} \tag{1}$$

The difference between the earning announcement and the forecast is divided by the lagged price of a share, p_{tk}. The price of a share works as a renormalization factor: the earnings are measured as earnings in dollar per share. The division by implies that 1 is the earning surprise as fraction of the value of the company. To see this, multiply numerator and denominator of expression (1) by the number of share n_{tk}:

$$s_{tk} = \frac{e_{tk}n_{tk} - \hat{e}_{tk}n_{tk}}{p_{tk}n_{tk}}$$

In the numerator, $e_{tk}n_{tk}$ is the total profit for quarter t and $\hat{e}_{tk}n_{tk}$ is the total forecasted profit. At the denominator is the market capitalization of a company $p_{tk}n_{tk}$, The earning surprise measure, therefore, captures the unexpected profits as a share of total market value of the company. If $s_{tk} = 0.01$ it means that the company earned unexpected profits equal to 1 percent of the value of the company.

We consider the response of stock returns to earnings surprises at different horizons. To capture the immediate response, one could look at $r^{(0,0)}$ that is, the stock return the same day as the announcement (measure as price at the close on day minus the price at the close on day $t-1$). However, since announcements are often made after the markets are closed, one should look at $r^{(0,1)}$ that is the return for the same day and the next day. If one wants to look at the delayed response to the earning announcement, a typical measure is $r^{(3,75)}$ that is, the stock returns for the period $(t+3, t+75)$ where days are always meant as trading days.

Data. In the dataset *earn2.xlsx*, I have already merged for you the information from Compustat, CRSP, and IBES. The data includes earnings from 1995 on in which the Compustat and IBES announcement dates differ by no more than 5 days. I have also generated the forecast of earnings $\hat{e}$. The data set that you see includes therefore information on earnings (MEDACT, multiply by the adjustment coefficient ADJ: ADJ*MEDACT), earnings forecast in the last 60 days (MEDEST60, multiply by the adjustment coefficient ADJ: ADJ*MEDEST60), stock returns (RAWWIN*—raw returns, NETWIN*—returns net of market returns, CARWIN*—returns adjusted for correlation with market). It also contains standard deviation of earning forecast (STDEST60), company name (COMNAM), price of shares (LAGPRICE), number of shares outstanding (LAGSHR). In order to make the data set small enough, it contains only companies with name up to "M".

1. (3 points) Construct the earnings surprise s (use (ADJ*MEDACT- ADJ*MEDEST60)/LAGPRICE). Calculate the mean and standard errors.

2. (6 points) Estimate an OLS model to relate raw returns $r^{(0,1)}$ (variable "RAWWIN1") to s as a measure of surprise:

   $$r_{tk}^{(0,1)} = \alpha + \phi s_{tk} + \varepsilon_{tk}$$

   How do you interpret the coefficient ϕ? Now run the same regression restricting the sample to s_{tk} in the range $[-0.01, 0.01]$ and then in the range $[-0.001, 0.001]$. Does the coefficient $\hat{\phi}$ change? What does that suggest about the specification of this regression?

   For questions 3-7 you only need to consider cases when s_{tk} is in the range $[-0.01, 0.01]$

3. (8 points) Now we explore more directly the possible non-linearity of the relationship. To provide non-parametric evidence, do a kernel regression of stock returns $r_{tk}^{(0,1)}$ on the earnings surprise s_{tk}. You are going to have to make some choices about bandwidth and type of kernel. The type of kernel typically does not matter very much, but the bandwidth certainly does. Show the graphs for both a relatively broad and a relatively narrow bandwidth. Is the relationship between the stock returns and the earnings surprise linear? Provide at least one interpretation for the observed non-linearity.

4. (8 points) Split the sample into a training set and a test set. Training set is all firms with names starting with "A-G", and test set is all firms with names starting with "H-L". Estimate a polynomial function with degree>3 using the training set and predict the return $r^{(0,1)}$ for the test set. What is the RMSE?

5. (8 points) Estimate a regression spline or MARS model for the training set and use the model to predict the return $r^{(0,1)}$ for the test set. What is the RMSE?

6. (8 points) Estimate kernel regression for the training set, and use the model to predict the return $r^{(0,1)}$ for the test set. (You should use cross validation to find the optimal bandwidth.) What is the RMSE?

7. (8 points) Which model from part 4 to part 6 has the best prediction in the test sample? Use that model (or the second best model) and the whole sample to predict raw returns $r^{(3,75)}$ (variable "RAWWIN375") as a function of earnings surprise and plot the relationship. What does the theory of efficient financial markets predict the relationship to be? What do you find?

8. (6 points) Companies have some discretion in the accounting procedure, so they can manipulate the earnings release at the margin. Consider the numerator of the earnings surprise,

$e_{tk} - \hat{e}_{tk}$. This is the earnings surprise per share. Plot the kernel density of this variable for values between -$0.1 and $0.1 using bandwidth of 0.0025. Is there a discontinuous jump in the density at zero? Interpret your result relating to manipulation of earnings.

2 Survival of Cryptocurrencies [45 Points]

Cryptocurrencies are decentralized payment systems in which ownership is demonstrated cryptographically. An overview of ownership of payment units is stored in a data structure called blockchain. Of the thousands of cryptocurrencies, the best known are Bitcoin, Ethereum, Ripple, Litecoin, EOS, Cardano, NEO, Dash, and Monero. Creating a new cryptocurrency is easy, but its value depends on users' willingness to pay for its units. If a cryptocurrency loses its users, it becomes worthless. In this exercise, we will analyze the survival rate of over 2,000 cryptocurrencies that are or were previously traded on cryptocurrency exchanges.

Data. In the data *crypto.xlsx*, you can find a list of cryptocurrencies. For each cryptocurrency, we have the symbol ("SYMBOL"), name ("NAME"), the price of the first week it started trading ("PRICE0"), the average daily trading volume of the first week ("VOLUME0"), the first week it started trading on exchange ("FIRSTWEEK"), and the last week it is traded on the exchanges ("LASTWEEK"). The data is from https://coinmarketcap.com. Since the data starts from April 28 2013, it is denoted as week 1, and week 2 is May 5 2013, week 3 is May 12 2013 etc. The most recent week is week 354, which is on February 2 2020.

1. (4 points) Prepare the data for survival analysis by creating a dummy variable for exit (which equals one if the cryptocurrency has been delisted from the exchange by the last week of the sample) and a time variable.

2. (8 points) Plot the survival curve for cryptocurrencies using the Kaplan-Meier method. On average what fraction of cryptocurrencies survive over a year (52 weeks)?

3. (4 points) Compare the survival curves for cryptocurrencies that are originated on or before January 1, 2017 (week 192) vs cryptocurrencies that are originated after the date. Is it riskier to buy new cryptocurrencies before or after 2017?

4. (4 points) Compare the survival curves for cryptocurrencies with a starting price of below 5 cents ($0.05) and above 5 cents. Is it riskier to buy new cryptocurrencies with a starting price of below 5 cents or above 5 cents?

5. (2 points) Merge the data with "bitcoin_price.xlsx" to get the price and supply of Bitcoins in the first week each cryptocurrency was initiated. (you should merge "FIRSTWEEK" from ctypto.xlsx to "WEEK" from bitcoin_price.xlsx)

6. Estimate a Cox proportional hazard model using the following variables to predict survival of cryptocurrencies:

   - Log starting price (log(PRICE0))
   - Log average volume in first week (log(VOLIME0))
   - Log price of bitcoin in the first week
   - A dummy variable for whether it is initiated after January 1 2017

   (a) (4 points) What is the coefficient of the log starting price? What does the coefficient mean?

(b) (4 points) Plot the baseline survival rate. On average what fraction of cryptocurrencies survive a year?

(c) (4 points) Plot the predicted survival curve for the cryptocurrency “Nexo”.

(d) (4 points) Plot the predicted survival curve for a new cryptocurrency initiated in the week of February 2, 2020 with starting price=$0.03 and first week's volume=$1 million.

7. (7 points) Use week numbers not divisible by 3 (week 1, 2, 4, 5, 7...) as training set and week numbers divisible by 3 (week 3, 6, 9...) as test set. Estimate the Cox proportional hazard model on the training set and predict the survival of cryptocurrencies in the test set. What is the concordance index? Add another variable that might predict the survival (for example, you can use characteristics about the name of cryptocurrency; be creative!). Does it improve the prediction?