2017 Citadel Datathon

Team: #9

Members: Zhen Hu, Tian Liu, Anton Petukhov, Tessa(Yuqing) Xie

Job postings and the Economy

1. Topic question

The main question we set up to answer is how job posting activity is related to Monetary Policy actions taken by the Federal Reserve and how Financial Markets have responded to those. Most of the publicly available measures of economic activity have monthly or lower frequency. Job postings data collected from the LinkUp can be aggregated at daily frequency to represent daily labor market activity. The obtained measure could be used by Federal Reserve officials to make better or more up to date policy decisions. Financial Markets participants can use that measure to execute trades at daily frequency and moreover predict Federal Reserve actions on the days when monetary policy decisions become public (such as FOMC meetings).

In the current report we are trying to tackle several questions.
1) Can LinkUp based measure of labor market activity forecast more conventional measures of economic activity (such as unemployment rate, index of industrial production, GDP, etc.)?
2) Can market participants exploit this daily measure of job posting activity to forecast Federal Reserve decisions on FOMC The Federal Open Market Committee?

The main dataset we used provided by the organizers was jobs.csv. On top of that we used monthly data on unemployment rate, index of industrial production, daily data on Fed Funds rate, 3-Month Treasury Bill rate, 10-Year Treasury Bond rate, Sp500 returns all downloaded from https://fred.stlouisfed.org/. For high frequency analysis of returns and job posting activity around The Federal Open Market Committee meetings we download dates of the meetings from the Federal Reserve Board website.

Non-technical summary

The main variable that we extract from the jobs.csv dataset is aggregate number of new job postings per day. This measure aggregated at quarterly frequency appears to be a sound forecaster of the future unemployment. Figure 3.1 summarizes results of the predictive regressions of unemployment changes by lagged unemployment changes and lagged measure of job openings aggregated at 1, 3, 6, 12 months frequency. Regardless of the availability of a very short data sample (2007--2016 at monthly frequency) the coefficient for 3-month aggregate measure of job openings is significant.

In the second part of the project we tested if job postings observed prior to the FOMC meetings aggregated across 1, 5, 10, 20, 100 days is a helpful predictor of the changes in Fed Funds rate determined during the meeting and helpful for predicting change in yield of 3-Month bills and 10-Year bonds. Key findings are: higher job postings prior to the day of the meeting significantly predicts higher change in Federal Fund rate (see table 3.5). However, it has no significant predictive power for longer yields.

2. Data processing

2.1 Group by time and industry

We manipulated the job posting data set to get the numbers of new job postings on each day from August 2007 with the help of Python.

For this purpose we created a dictionary whose keys are the dates and values are initialized as zero. Next, when scanning through the data set (jobs.csv), we increment the value of the corresponding date by one when reading a valid entry of job posting on a given date.

An even more detailed work involves grouping both by date and by industry, which helps to analyze in an industrial level. Here the python package pandas is of great help and we used the dataframe structure and the .groupby function in pandas. These together generated a table with double indices, one for date and another for industry.

Because of the frequency of other data sets like real estate, which is monthly, we then aggregate the daily data into monthly ones.

2.2 Combining with external financial data sets

Besides utilizing the data sets provided by Datathon, we also incorporate other data sets due to the fact that there are several macroeconomic variables involved in our hypothesis, including the Federal fund rate, unemployment rate, and FOMC meeting dates.

3. Time Series Analysis

We take the sum of number of job postings in the previous 1,2,3,6 and 12 months and try to use these variables to predict the change of unemployment rate. The results of regression are shown below.

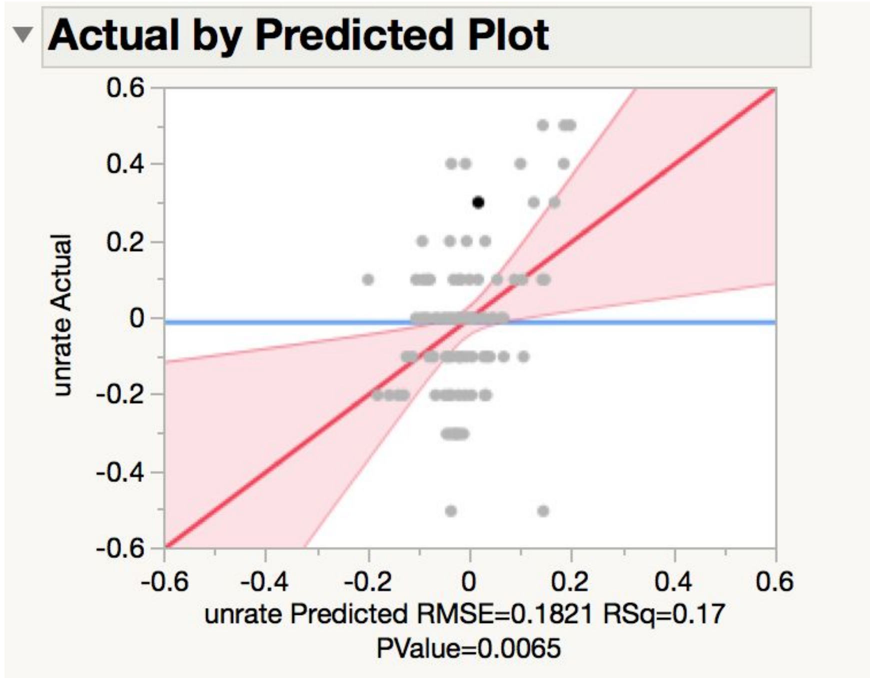

Figure 3.1

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.127365	0.125443	-1.02	0.3126
1m	-6.302e-7	1.895e-6	-0.33	0.7402
2m	-2.491e-6	2.454e-6	-1.02	0.3125
3m	3.4869e-6	1.857e-6	1.88	0.0636
6m	-6.65e-7	7.155e-7	-0.93	0.3550
12m	1.7267e-7	3.075e-7	0.56	0.5757
unrate_lag1	0.3202839	0.095997	3.34	0.0012*

Table 3.1

Summary of Fit

RSquare	0.170207
RSquare Adj	0.117242
Root Mean Square Error	0.182138
Mean of Response	-0.01089
Observations (or Sum Wgts)	101

Table 3.2

Since we have relatively short periods of observations, the aggregated 3 months of number of job posting is considered significant with t-stat of 1.88.

In addition to linear regression model, we applied Neural Network method. We partition the dataset and set the observations from 2007-08 to 2014-12 as the training set and the rest as the validation set. After several trainings of different models, we have come up with the optimal one shown as following with 15 first and 5 second TanH layers:

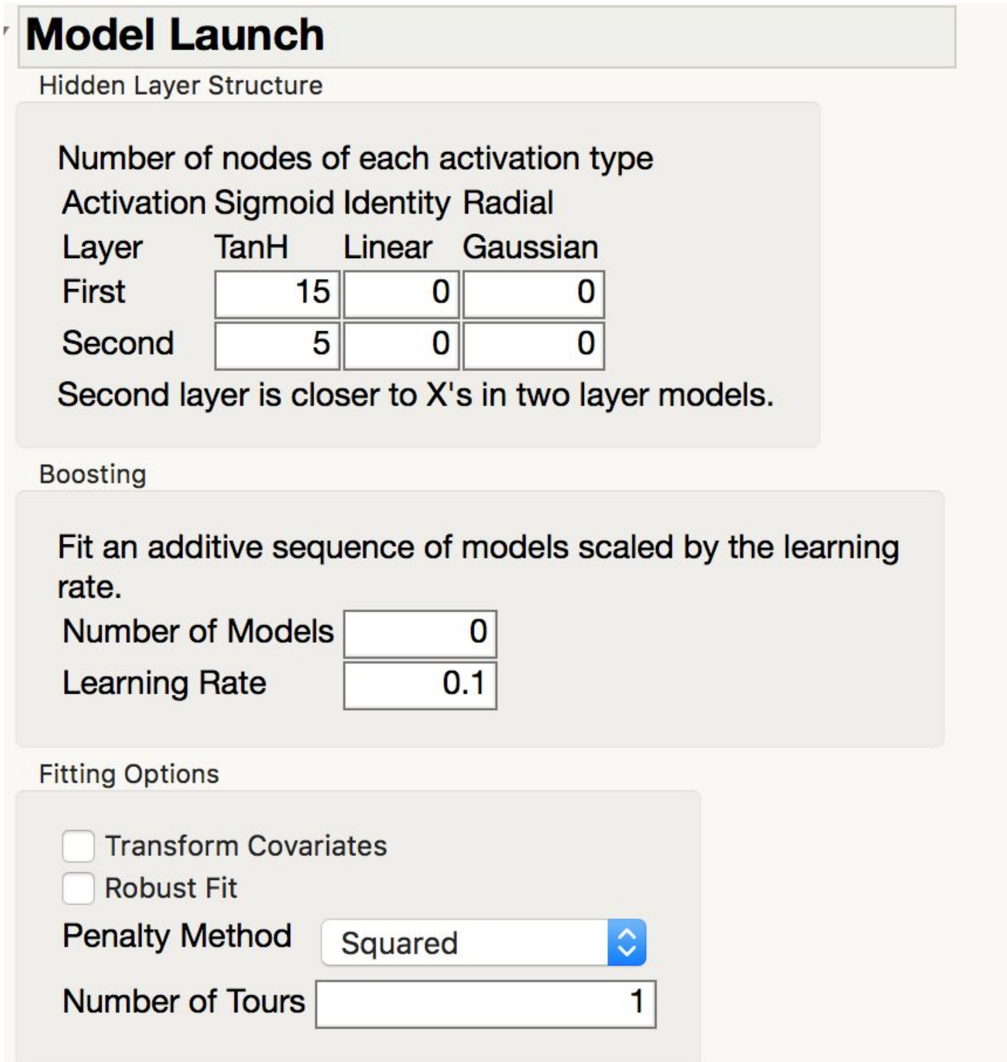

Table 3.3

The result of the model prediction is shown below. Given the small monthly dataset, the RMSE is roughly 0.16 for the training set but it decreases to 0.11 for the validation set. This indicates that It is a promising promising since we can look at the number of job postings 3 months before the FOMC meeting to forecast the unemployment rate and so forth predict the level of Fed Fund rate and trade on this signal.

Model NTanH(15)NTanH2(5)

Training: unrate Measures	Value	Validation: unrate Measures	Value
RSquare	0.3983997	RSquare	0.0958363
RMSE	0.1635707	RMSE	0.1111508
Mean Abs Dev	0.1213851	Mean Abs Dev	0.0914166
-LogLikelihood	-29.75942	-LogLikelihood	-19.44823
SSE	2.0334088	SSE	0.3088623
Sum Freq	76	Sum Freq	25

Table 3.4

Horizon	t-stat	RSquare
1	2.02	3%
5	0.3	0.4%
10	0.1	0.02%
20	-0.1	0.02%
100	0.2	0.5%

Table 3.5

4. Alternative model

4.1 From frequency to length

An additional question we tried to answer is: within each state, whether there is a correlation between the average length of "description" of job postings and real estate price change of that state. Our hypothesis is: longer job posting with more detailed descriptions prelude that companies are increasing their hiring standards, which might be connected to a decrease of hiring of the company. This might be caused by financial downturn of the area or the financial difficulties of the specific company. Thus we hypothesize that there's a correlation between the length of company's job description and a decreasing of local housing prices, both connected to financial hardship in the area.

4.2 Data processing

We split the description of each job posting into lists of words and further counted the number of words in each posting. Then we joined that table with Jobs.csv file by 'hash" to acquire the information of the company that made the posting. Using the company's information, we can aggregate the average length of the postings by state and by month.

The next step we attempted was to aggregate average change of housing price by month and state. By joining the housing price change and length of job postings and sectioning by states and month, we will be able to further explore the correlation relationship.