Check The Job Bulletin Before You Buy ETFs
2017 NYC Citadel Datathon

Jason Cheuk Nam Liang
Lin Shi
Chang Liu
Wenjun Wang

March 10 2017

1 Topic Question

Can we use job openings for companies in a specific industry to predict the returns of iShares U.S. sector ETFs that track this industry?

2 Non-Technical Executive Summary

2.1 General Approach

The iShares U.S. sector ETFs seeks to track the investment results of indices composed of U.S. equities in specific sectors. Instead of applying discretionary analysis to components of the ETFs, we believe that labor market dynamics within a certain industry offer valuable signals to predict returns/ price movements of the corresponding sector ETFs. In our analysis we consider the following iShares ETFs and their corresponding sectors.

Tick:	IYE	IYW	IYC	IYF	IYJ	IDU	IYH
Sector:	Energy	Technology	Consumer Services	Financials	Industrials	Utilities	Healthcare

Figure 1: iShares U.S.Sector ETFs

2.2 Modelling

Denote S as the set of all states. For each industry $i \in I = \{Energy, Technology, ConsumerServices, Financials, Industrials, Utilities, Healthcare\}$, we consider the following data within every State $s \in S$:

Number of job openings for a given i and s at time t: $n_{i,s}(t)$
Number of new graduates with degrees related to industry i: N_i.
Number of companies within industry i in State j: $m_{i,j}$
Number of companies within industry i in the U.S.: M_i

Consider the labor market for industry i and it's effect on an ETF that tracks this sector. The idea is that if labor demand in this sector is significantly greater than the supply of labor, this excessive demand signals that companies within this industry have good performances and growth potential (maybe due to global trade or consumer demand), which indicates a high probability of upward movement in ETFs that track this sector. We parameterize this effect as $f_i(t)$. Now let's first consider the labor market in industry i for state s. Define:

$$\rho_{i,s}(t) = \frac{n_{i,s}(t)}{N_i} \tag{1}$$

Moreover, since equation 1 only models the labor market dynamics in State s, and since iShare ETFs track sectors in the U.S., we aggregate the effects of supply and demand within industry i by assigning weights to all $\rho_{i,s}(t)$ and taking a sum. Let $\vec{w} = (w_1, w_2, ..., w_s)$, where

$$w_k = \frac{m_{i,j}}{M_i} \tag{2}$$

Hence we have the aggregated labor market dynamic factor in industry i:

$$f_i(t) = \sum_{k \in S} \rho_{i,s}(t) \cdot w_k \tag{3}$$

2.3 Data Analysis and Visualization

Due to space limitations, here we visualize the data for the most recent number of job postings in the Financial Industry and the trend of number of job postings of all industries in California.
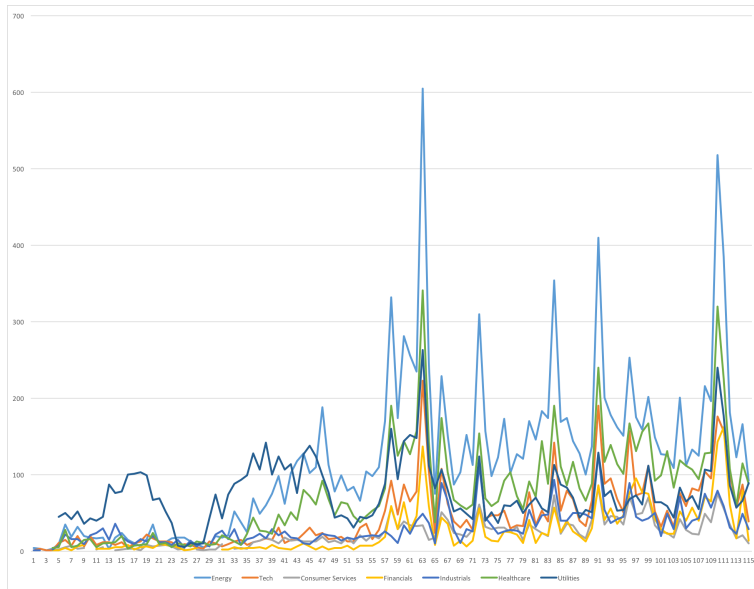

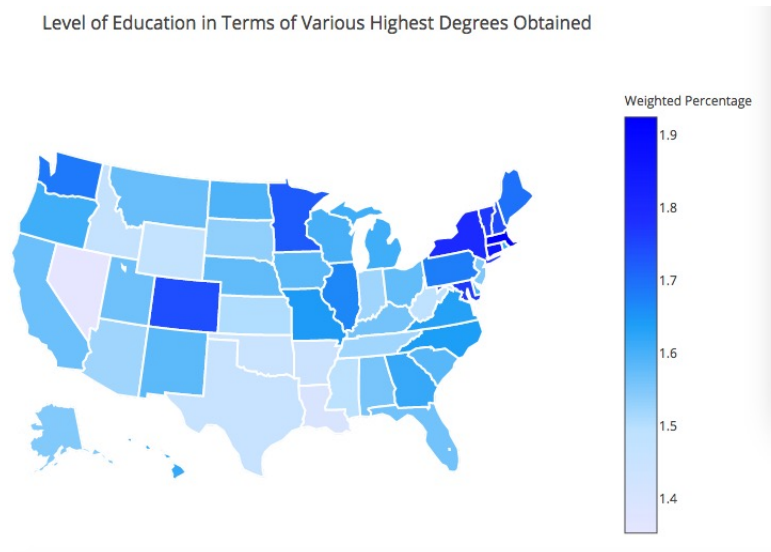Figure 2: Job Opening Trends in California for all industries

Level of Education in Terms of Various Highest Degrees Obtained

Figure 3: Job Openings in Financial Industry
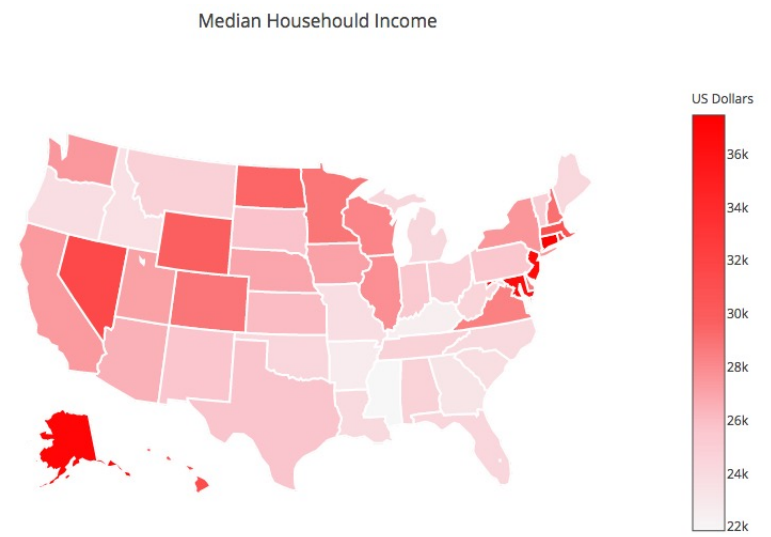
Median Housechould Income

Figure 4: Job Openings in Technology Industry

2.4   Variable Selection Using LASSO

The main purpose of analyzing labor market dynamics is to predict the return for ETFs R_i that track a specific sector i. However, not all labor factors $f_i(t)$ may have a significant relation in the prediction of R_i, so it is necessary to

conduct variable selection. We will apply **LASSO** regression to select suitable variables as input factors to our model. Define $\overrightarrow{f(t)} = (f_i(t))_{i \in I}$. The **LASSO** regression model is given as followed:

Select α_i and $a_{i,j}$ such that we obtain

$$min\left(R_i(t) = \alpha_i + \sum_{j \in I} a_{i,j} f_j(t) + \lambda ||\overrightarrow{f(t)}||\right) \tag{4}$$

where $\lambda > 0$.

2.5   ETF Return Prediction Using Vector Auto-Regression

After choosing the labor factors, we then use the **VAR (Vector Auto-regression) Model** to predict $R_i(t)$. For $R_i(t)$, we first train a VAR model using the data from **August 2006** to **October 2016** and then make a one-step prediction.

3   Technical Executive Summary
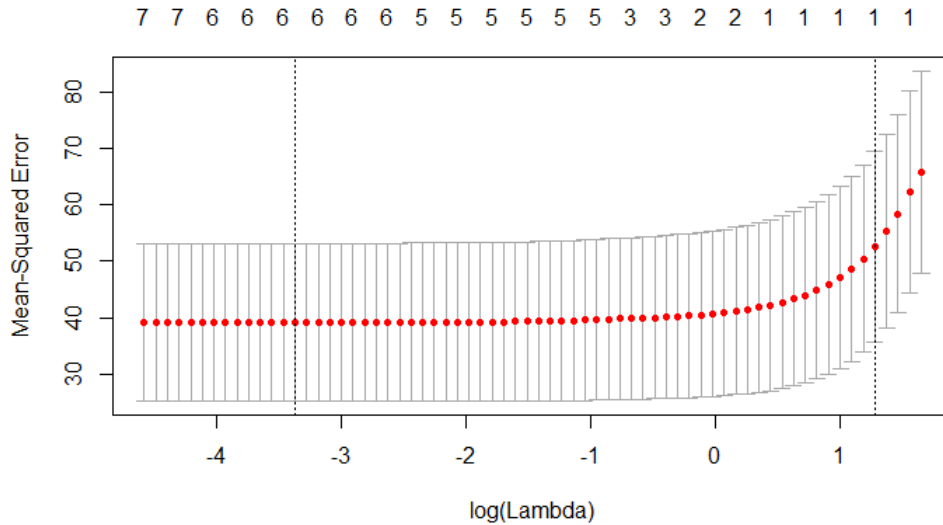
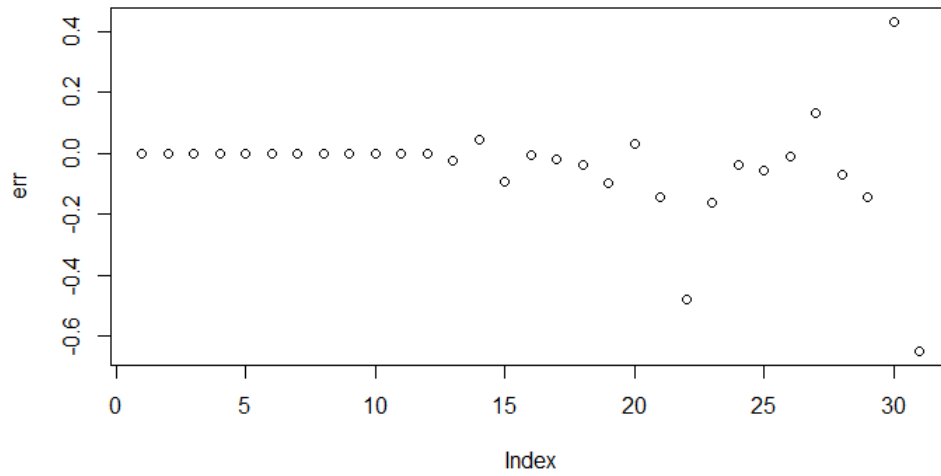## 3.1   LASSO Regression Results

Figure 5: LASSO

3.2   VAR Results

Figure 6: LASSO

4   Conclusion

From the results above we can see that by using **VAR** we can precisely predict the returns for iShares sector ETFs. These models and techniques hence offer significant signals to predict the price dynamics of iShares ETFs.