

RA Test

Liu, Jingyun

April 2020

1 Data Preprocess

	tip_ratio	pickup_count	stock_return	job_post
count	594	594	594	594
mean	0.151	174605.194	0.006	223.593
std	0.004	202276.771	0.067	297.502
min	0.139	802	-0.292	0
25%	0.149	16041.25	-0.03	36
50%	0.151	99287	0.015	106.5
75%	0.154	267664	0.044	283.75
max	0.166	793581	0.233	1901

Figure 1: Data Description

Data above was standardized first, and the interaction term is calculated after standardization to avoid being dominant by one of its components. The data set was split into train and test sets by the ratio of .8 to .2.

2 Multicollinearity Test

In order to interpret the relationship between predictors and response variable based on the coefficient in the regression models(β 's), we have to make sure there is no multicollinearity problem.

2.1 Scatter Plot

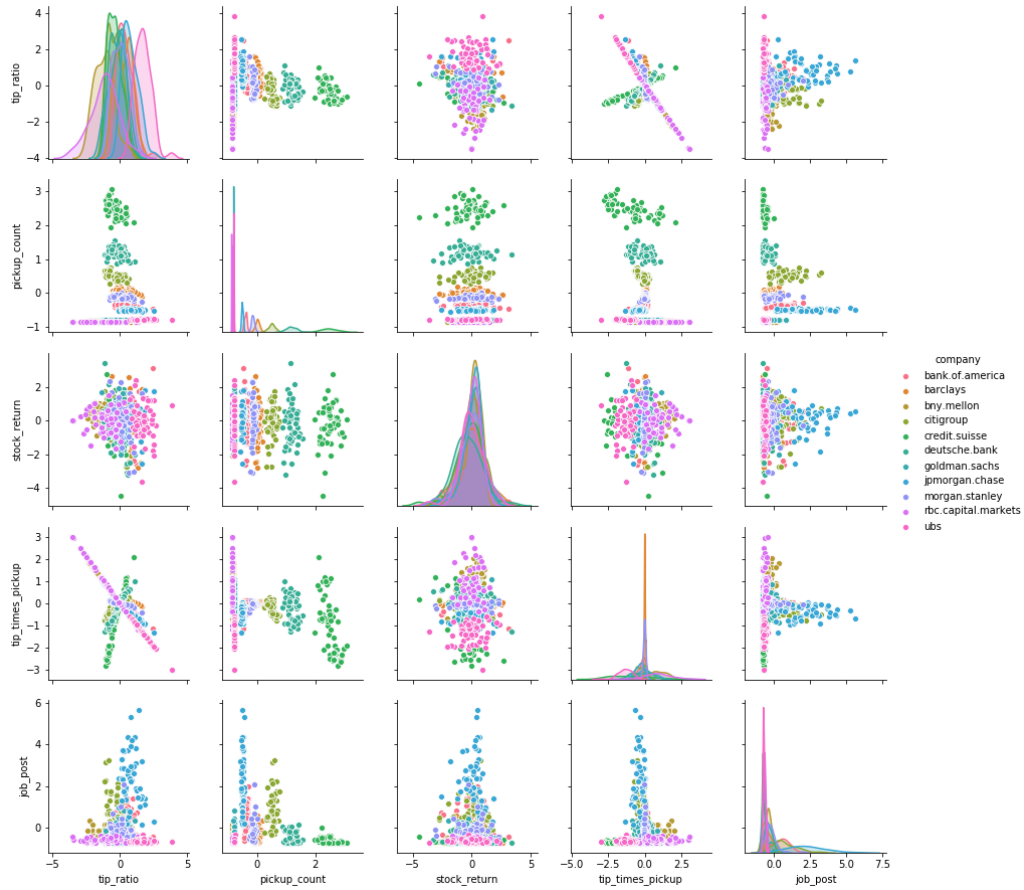


Figure 2: Scatter Matrix

Referring to the scatter plot matrix (Figure 2), it can be seen that no pair of the predictors have a significant linear relationship.

2.2 Pearson Correlation

	tip_ratio	pickup_count	tip_times_pickup	stock_return
tip_ratio	1.0000	-0.1202	-0.5759	-0.0980
pickup_count		1.0000	-0.3556	-0.0950
tip_times_pickup			1.0000	0.0181
stock_return				1.0000

Table 1: Pearson Correlation Table

2.3 VIF

In regard to the Pearson correlation table (Table 1) and VIF's calculated (Table 2), almost all the correlation's are less than 0.5 (except for a -0.58 correlation between tip ratio and its interaction term), and all the VIF is smaller than 2. Hence, we can use the coefficients (β 's) to interpret the relationship between predictors and response variable.

	VIF
tip_ratio	1.8770
pickup_count	1.4324
tip_times_pickup	2.0923
stock_return	1.037

Table 2: VIF

3 Models

3.1 Multiple Linear Regression

We start with the model provided. Tip ratio, pick-up counts, and tip ratio* pick-up counts as independent variable, and job posted as dependent variable. The firm fixed effect and year fixed effect are added to the model afterwards.

$$\begin{aligned} \text{Job Post} = & \beta_0 + \beta_1 \text{Tip Ratio} + \beta_2 \text{Pickup Count} \\ & + \beta_3 \text{Tip Ratio} * \text{Pickup Count} \end{aligned} \quad (1)$$

Equation (1) is referred as Model 1 below.

Given the fact that the job posting may be caused by company expansion. Namely, company performance could be an omitted factor affecting the number of job-posting. The monthly return of the company's stock is also selected into variable set as a proxy for company performance. The model with stock return added was tested under firm fixed effect and year fixed effect, respectively and collectively.

$$\begin{aligned} \text{Job Post} = & \beta_0 + \beta_1 \text{Tip Ratio} + \beta_2 \text{Pickup Count} \\ & + \beta_3 \text{Tip Ratio} * \text{Pickup Count} + \beta_4 \text{Stock Return} \end{aligned} \quad (2)$$

Equation (2) is referred as Model 2 below.

3.2 XGBoost

Given the limited amount of data on hand, and narrowed number of predictors, model with high flexibility like XGBoost does not yield much better result (around **26% train MSE** and some **22% test MSE**), which is similar to the result for models with both firm-fixed effect and year-fixed effect interpreted in the conclusion section below.

We can notice that *pickup count* plays an important role in terms of features selected in the model ((Figure 3)).

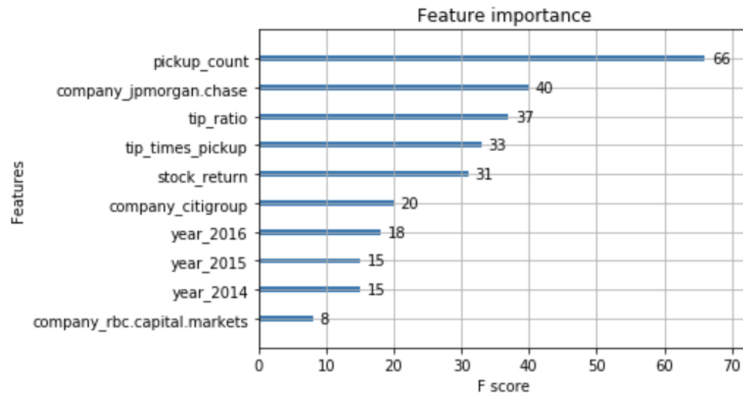


Figure 3: XGBoost - Feature Importance (Top 10)

4 Results and Conclusion

Dep. Variable: job post								
const	0.001	(0.982)	0.7212***	(0.000)	-0.3823***	(0.000)	0.2828	(0.067)
tip ratio	0.0776	(0.208)	-0.0341	(0.459)	0.1311	(0.052)	-0.0745	(0.138)
pick-up count	-0.1314*	(0.016)	0.6295	(0.052)	-0.103	(0.058)	0.3436	(0.266)
tip ratio * pickup count	-0.0335	(0.66)	0.0274	(0.599)	0.0183	(0.815)	0.0179	(0.714)
Company F.E.	N		Y		N		Y	
Year F.E.	N		N		Y		Y	
Observations	475		475		475		475	
Adj. R-squared	0.02		0.693		0.061		0.736	
F - STAT	4.2		83.45		5.421		78.62	
P-value (F-STAT)	0.00599		4.04E-112		5.40E-06		9.24E-124	
Test MSE	1.004		0.295		0.955		0.232	

Notes: 1. number in parentheses stands for the p-value for each coefficient

2. significant level: *** 0.001

** 0.01

* 0.05

Figure 4: Model (1) with Fixed Effects

Dep. Variable: job post								
const	0.0011	(0.981)	0.7315***	(0.000)	-0.3943***	(0.000)	0.2903	(0.062)
tip ratio	0.0782	(0.214)	-0.0524	(0.261)	0.1413*	(0.039)	-0.0779	(0.125)
pick-up count	-0.1309*	(0.017)	0.6356*	(0.049)	-0.0949	(0.084)	0.3471	(0.262)
tip ratio * pickup count	-0.0329	(0.668)	-0.0558*	(0.028)	0.0416	(0.363)	-0.0122	(0.615)
stock return	0.0024	(0.958)	0.0177	(0.734)	0.0272	(0.730)	0.0156	(0.750)
Company F.E.	N		Y		N		Y	
Year F.E.	N		N		Y		Y	
Observations	475		475		475		475	
Adj. R-squared	0.018		0.696		0.061		0.735	
F - STAT	3.144		77.48		4.845		74.15	
P-value (F-STAT)	0.0144		3.47E-112		9.44E-06		7.35E-123	
Test MSE	1.004		0.293		0.944		0.233	

Notes: 1. number in parentheses stands for the p-value for each coefficient

2. significant level: *** 0.001

** 0.01

* 0.05

Figure 5: Model (2) with Fixed Effects

For both model 1 (Figure 4) and model 2 (Figure 5), the model with both firm-fixed effect and year-fixed effect pass the global F-test, yield the highest

adjusted R square (around 74%) and lowest test MSE (around 23%). However, the coefficient for all the predictors are not significant.

For both model 1 and model 2 with firm-fixed effect, the model can also pass the global F-test, and yield a relatively high adjusted R square (around 70%) and low test MSE (around 0.3). In model 2, given a 95% confidence level, the coefficient for *pick-up count* is 0.6356, and the coefficient for *tip ratio* pick-up* is -0.0558. The positive coefficient for *pick-up count* indicates the more intense overtime work, the more likely people will resign. The negative coefficient for the interaction term indicates that the good mood after work can somehow mitigate the influence of overwork on resignation. Nonetheless, we can not conclude on a causal relationship without further analysis.

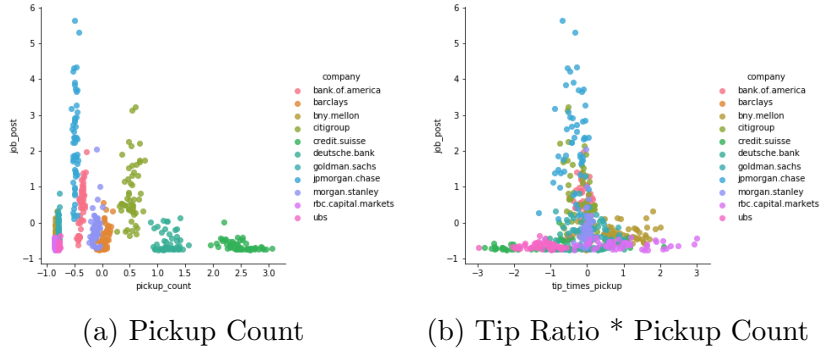


Figure 6: scatter plot - predictors & job posting

This negative relationship between *pickup count* and *job post* can be witnessed in the scatter plot Figure 6a. However, the relationship within each firm is not obvious enough. This may partially be caused by sampling error, given the fact that only four-and-a-half-year's monthly data was obtained for each company.