

The analysis of most influential investing factors in express and E-commerce industries

Hongyi Gou^{1,*,†}, Long Ling^{2,†}, Youyi Wu^{3,†}

¹Nankai University

²Shandong University of Finance and Economics

³University of Chicago

*Corresponding author: guanghua.ren@gecademy.cn

†These authors contributed equally.

Keywords: Value investing, Express industry, E-commerce industry, Business model Analysis.

Abstract: Based on the value investment theory, we mainly study the ranking correlation and underlying reasons of stock selection indexes of express companies and e-commerce companies. This paper mainly adopts the method of linear regression and calculating the correlation coefficient between each index. Through calculation and analysis, this paper concludes that the importance of divide yield ranks first in the stock selection indicators of the two industries.

1. Introduction

The E-commerce industry keeps growing as the internet developing to widespread application. On top of that, the express industry also reaches its peak. As those two industries are expanding hugely, they attract investors. For value investors, several distinct factors influence the value of a company or a stock. (De-Yun Cao, 2019) Thus, finding out which factor is the most influential for those two industries provides more possibilities and profitability for the investors. Moreover, as the two picked sectors are so closely related, the question that they have the same factor stands out as the most significant one needs to be answered. In this research, we will compute the multiple regression analysis to find which factor influences the stock price the most in the two industries individually. After that, we need to compare the result and find the reason behind it. To find the reason, we will discuss and reach the standard business model behind the two industries and talk about why the model influence any factor the most.

2. Method

According to the theory of value investment, we put down the factors that affect the return rate. EV/Sales (EVS), EV/EBITDA (EVE) and P/E (PE) are indicators that explain whether the price of a company's stock is cheap or not. Growth rate in EBITDA (gE), standard deviation of growth rate in EBITDA (gESD), growth rate in Sales (gS) and standard deviation of growth rate in Sales (gSSD) can reveal whether a firm keeps growing and whether it is growing steadily. GP/Assets (GPA) and dividend yield (DY) helps us to find out whether a company enjoys high profitability. By getting the data of these factors, we can calculate the correlation coefficient between return rate and them. By comparing the degree of influence of individual factors on the rate of return, the importance of each factor is sorted in order to provide some reference value for investors who invest in value.

Based on the R language, we use the method of multiple linear regression to explore the effects of various factors on returns.

The equation used is:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon \quad (1)$$

$$\text{Let } y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{bmatrix}. \quad (2)$$

We use the car package in the R language to explore the connection between the return rate(rr) and EV/Sales (EVS), EV/EBITDA(EVE), P/E(PE), growth rate in EBITDA(gE), standard deviation of growth rate in EBITDA(gESD), growth rate in Sales(gS), standard deviation of growth rate in Sales(gSSD), GP/Assets(GPA), dividend yield(DY) in e-commerce(ecom).

```
> ecom<-read.csv("ecom.csv")
> cor(ecom)
```

After filtering through the data, the chart below shows the relationship between factors and returns

Table 1. The Relationship between Factors and Returns in E-commerce Industry.

	rr	EVS	EVE	PE	gE
rr	1.0000000	-0.57393555	-0.23800228	-0.34581624	-0.72637100
EVS	-0.5739356	1.00000000	0.76405335	0.69210905	0.09042861
EVE	-0.2380023	0.76405335	1.00000000	0.91846274	-0.38704130
PE	-0.3458162	0.69210905	0.91846274	1.00000000	-0.18746007
gE	-0.7263710	0.09042861	-0.38704130	-0.18746007	1.00000000
gESD	-0.7936680	0.17735313	-0.26136744	-0.04017298	0.92027015
gS	-0.9025524	0.47192306	0.01827332	0.21488711	0.87159220
gSSD	-0.3361815	0.29057039	-0.01813169	-0.05348775	0.37597122
GPA	-0.1264965	-0.56276855	-0.65842362	-0.49285294	0.43306491
DY	0.5818771	-0.48250122	-0.25690147	-0.37784428	-0.59306387

Table 2. The Relationship between Factors and Returns in E-commerce Industry.

	gESD	gs	gSSD	GPA	DY
rr	-0.79366796	-0.90255240	-0.33618148	-0.126496	0.5818771
EVS	0.17735313	0.47192306	0.29057039	-0.5627685	-0.4825012
EVE	-0.26136744	0.01827332	-0.01813169	-0.6584236	-0.2569015
PE	-0.04017298	0.21488711	-0.0534877	-0.4928529	-0.3778443
gE	0.92027015	0.8715922	0.37597122	0.4330649	-0.5930639
gESD	1.00000000	0.92400731	0.43688392	0.2808180	-0.3887738
gS	0.92400731	1.00000000	0.32702548	0.1842821	-0.6203751
gSSD	0.43688392	0.32702548	1.00000000	-0.4402413	-0.0816121
GPA	0.28081802	0.1842821	-0.4402413	1.0000000	-0.1101008
DY	-0.38877382	-0.62037511	-0.08161210	-0.1101008	1.0000000

From the relevant coefficients in the table, we can see that the correlation coefficient between return rate and EV/Sales (EVS) is the largest in EV/Sales (EVS), EV/EBITDA (EVE) and P/E (PE), the three indicators that show whether the stock is cheap or not. This means that EV/Sales (EVS) has the greatest impact on return rate. In this case, we can only take EV/Sales (EVS) as an indicator of how cheap stocks are. Then we use the lm () function to fit the multiple linear regression model.

```
> fit<-lm (rr ~ EVS+gE+gESD+gS+gSSD+GPA+DY, data=ecom)
```

```
> Summary (fit)
```

Call:

```
Lm (formula = rr ~ EVS + gE + gESD + gS + gSSD + GPA + DY, data = ecom)
```

Table 3 The Linear Regression Model of Factors and Returns in E-commerce Industry.

Residuals	1	2	3	4	5	6	7	8	9	10
	-	0.059	0.012	0.004	-	0.002	0.047	-	0.056	0.145
	0.062	399	422	442	0.199	365	942	0.066	155	362
	197				358			532		

Coefficients:

Table 4 The Correlation Coefficient between Factors and Returns in E-commerce Industry.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.8328	1.0002	1.832	0.208
EVS	-0.1349	0.1643	-0.821	0.498
gE	0.7823	1.1838	0.661	0.577
gESD	-2.2364	3.3234	-0.673	0.570
gS	1.3316	4.7512	0.280	0.806
gSSD	-2.4505	8.3019	-0.295	0.796
GPA	-1.6498	1.4002	-1.178	0.360
DY	4.8289	7.9147	0.610	0.604

Residual standard error: 0.1979 on 2 degrees of freedom

Multiple R-squared: 0.9086, Adjusted R-squared: 0.5889

F-statistic: 2.842 on 7 and 2 DF, p-value: 0.2849

From the results given, it can be seen that none of these indicators have a significant regression coefficient of 0, indicating that their linear relationship with the return rate is not very obvious. The model can only roughly describe the effect of these indicators on the return rate, but it cannot accurately calculate the relationship between these indicators and the rate of return

Therefore, if we only do qualitative analysis, we can see that the regression coefficient of the growth rate in EBITDA (gE), growth rate in Sales (gS), and dividend yield (DY), these three indicators with the rate of return is positive, meaning that they are positively related to the return rate while other indicators are negatively related to the return rate. In this way, we investors have to select the stocks with higher growth rate in EBITDA (gE), growth rate in Sales (gS), and dividend yield (DY) and the other indicators of the stocks need to be as low as possible.

Next, we use the confint function to calculate the confidence interval of the seven indicators.

```
> confint(fit)
```

Table 5. The Confidence Interval of the Seven Indicators in E-commerce Industry.

	2.5%	97.5%
(Intercept)	-2.4708396	6.1364853
EVS	-0.8420011	0.5722698
gE	-4.3110146	5.8757089
gESD	-16.5357309	12.0629604
gS	-19.1111495	21.7743847
gSSD	-38.1705713	33.2695747
GPA	-7.6741982	4.3745741
DY	-29.2254461	38.8833236

As can be seen from the results of the confint function, the confidence intervals of the seven indicators all contain 0, which is consistent with the results given earlier by summary. This means that our calculations and conclusions above are correct.

We then plot the component residuals to compare whether the relationship between the cause variable and the argument is linear. The results in the figure show that the linear hypothesis is satisfied.

```
> Library(car)
```

```
> crPlots(fit)
```

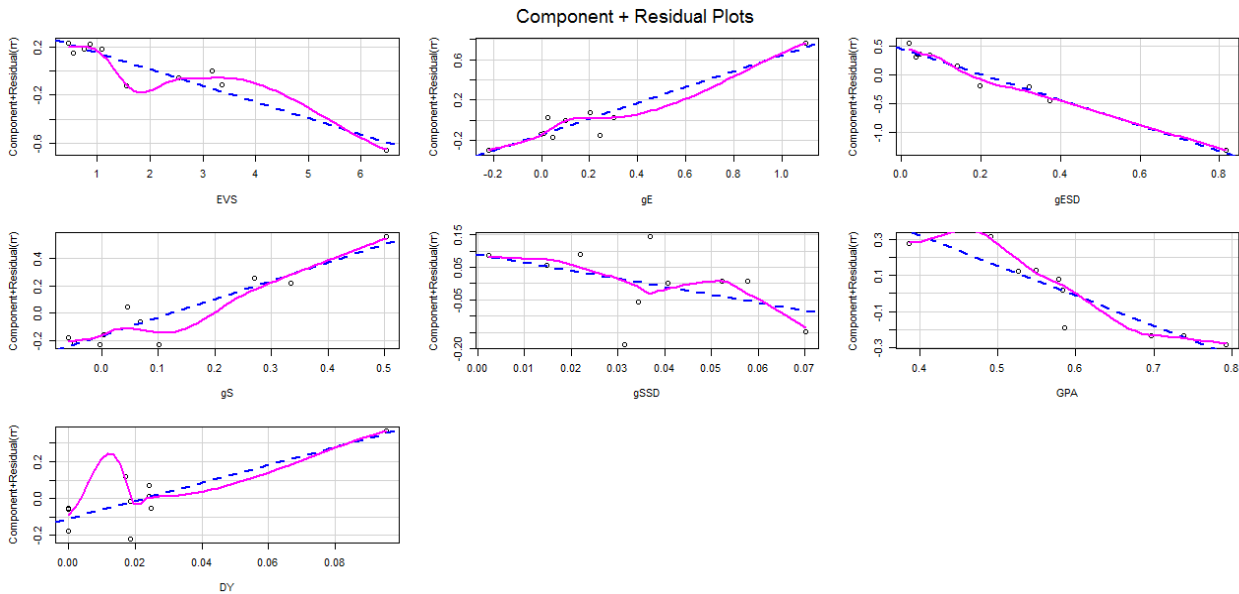


Figure 1. Component and Residual Plots of Investing Factors.

Then we can detect whether the problem of multi-collinearity exist by calculating the variance expansion factor VIF. When $\sqrt{vif} > 2$, indicates that there is a problem with multiple collinearity. (Bin-Hui Wang, 2016)The results show that there is a problem of multiple collinearity. There are two reasons for multi-collinearity, the first one is that the price of stocks and corporate profits, dividends and other indicators would have been linked. The second one is the sample data volume is small.

```
> vif(fit)
```

Table 6. The VIF of the Seven Indicators in E-commerce Industry.

EVS	gE	gESD	gS	gSSD	GPA	DY
22.037127	40.423959	155.948245	179.879757	6.568456	6.579013	11.028933

Finally, we use a standardized regression coefficient to compare the size of a standardized regression coefficient for a response variable when other variables remain unchanged in order to determine the size of the variable's fluctuations on the rate of return fluctuations and the direction of change

```
> zecom<-as.data.frame(scale(ecom))
> zfit<-lm (rr~EVS+gE+gESD+gS+gSSD+GPA+DY, data=zecom)
> coef (zfit)
```

Table 7. Standardized Regression Coefficient.

(Intercept)	EVS	gE	gESD	gS	gSSD	GPA	DY
5.741738 e-17	- 8.233483 e-01	8.980939 e-01	- 1.796090e+ 00	8.034120 e-01	- 1.616892 e-01	- 6.459656 e-01	4.330647 e-01

As can be learned from the above results, when other variables remain unchanged, a change in the standard deviation of the growth rate in EBITDA (gESD) will reduce the return rate by 1.80 standard deviations. The effect of other variables on response variables can be found in the results above.

Similarly, we use the same approach above to explore the relationship between the return rate (rr) and EV/Sales (EVS), EV/EBITDA (EVE), P/E (PE), growth rate in EBITDA (gE), standard deviation of growth rate in EBITDA (gESD), growth rate in Sales (gS), standard deviation of growth rate in Sales (gSSD), GP/Assets (GPA) and dividend yield (DY) in the express industry expressage (expr).

```
> Expr <-read.csv ("expr.csv")
> fit<-lm (rr ~ EVS+gE+gESD+gS+gSSD+GPA+DY, data=expr)
> Summary (fit)
```

Call:

Lm (formula = $rr \sim EVS + gE + gESD + gS + gSSD + GPA + DY$, data = expr)

Table 8. The Linear Regression Model of Factors and Returns in Express Industry.

Residuals	1	2	3	4	5	6	7	8	9	10
	0.006	0.122	0.027	-	-	0.022	0.025	0.060	-	-
	854	013	860	0.042	0.086	125	009	758	0.087	0.048
				178	830				459	151

Coefficients:

Table 9. The Correlation Coefficient between Factors and Returns in Express Industry.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.04932	0.30193	-0.163	0.8853
EVS	0.02371	0.01325	1.790	0.2153
gE	1.10335	0.33316	3.312	0.0803
gESD	0.60558	1.06109	-0.571	0.6258
gS	-0.32864	0.53866	-0.610	0.6039
gSSD	0.45795	0.70844	0.646	0.5843
GPA	-0.65943	0.42737	-1.543	0.2628
DY	2.95963	5.11765	0.578	0.6215

Signif. Codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1411 on 2 degrees of freedom

Multiple R-squared: 0.8955, Adjusted R-squared: 0.5299

F-statistic: 2.449 on 7 and 2 DF, p-value: 0.3204

Plus, it can be seen from the above table that the greater the absolute value of estimate, the greater the impact of the index on the stock price. If estimate is positive, it means that the index is negatively correlated with stock price, otherwise it is positively correlated. By qualitative analysis, we can see that the regression coefficients of growth rate in Sales (gS), GP/Assets (GPA) are negative, meaning that they are negatively related to the return rate, while several other indicators are positively related to the return rate. We can see that the absolute value of divided yield is the largest, so in the express industry, divided yield has the greatest impact on the stock price. The absolute value of EV / sales is the smallest, so in the express industry, EV / sales has the least impact on the stock price.

As can be learned from the above results, when other variables remain unchanged, a change in the growth rate in EBITDA (gE) standard deviation will reduce the return rate by 0.89 standard deviations. The effect of other variables on response variables can be found in the results above.

From the data and analysis of the above calculations, in terms of e-commerce industry, we can see that divided yield has the greatest impact on the stock price, and there is a positive correlation between them, so it is the most important factor. EV/Sales (EVS) has the greatest impact on returns in the indicators that determine whether stocks are cheap or not, and that the two are positive correlations so it's the most important thing. In the indicator that determines the stability of stock prices, growth rate in EBITDA (gE) is positively related to the return on stock and the volatility of the growth rate in Sales (gS) has less effect on the volatility of stock return, so we should first consider the growth rate in Sales (gSSD).

As far as the express industry is concerned, divided yield has the greatest impact on the stock price, and there is a positive correlation between them, so it is the most important factor. As for the index to judge whether the stock price is cheap, EV / sales (EVS) has the least impact on the stock price. As far as the index to judge the stability of stock price is concerned, growth rate in EBITDA (gESD) and growth rate in Sales (gSSD) are both positively related to the return on stock. The absolute value of

estimate of growth rate in sales (gESD) is larger, so we should consider growth rate in EBITDA (gESD) first.

In summary, the stock selection index of e-commerce industry is ranked as follows: dividend yield > standard deviation of growth rate in sales > standard deviation of growth rate in sales

Growth rate in EBITDA>Gross profit/Asset>growth rate in sales> growth rate in EBITDA>EV/Sales.

The stock selection index of express industry is ranked as follows: dividend yield > growth rate in EBITDA > gross profit / asset

>standard deviation of growth rate in EBITDA>standard deviation of growth rate in sales> growth rate in sales>EV/Sales.

3. Results

Based on the multiple regression model, the highest influential index for the e-commerce and express industries is dividend yield (DY). Meanwhile, the DY is positively related to the return rate. The dividend yield indicates the profit the investors could get from each one unit of stock. The dividend yield could be obtained through the following equation,

$$\text{Dividend Yield} = \text{Annual Dividends Paid Per Share} / \text{Price per Share} \quad (2)$$

In this equation, both annual dividends paid per share and price per share could be found from the annual report of the firm. Since DY is the most influential index, it implies that the stock price would be most likely influenced by the rate of profit investors could get. While the DY is higher, the return rate increases, so the investors usually pursue stocks with high DY in the e-commerce industry and express industry.

The most direct intuition behind high DY is the company is paying most of their moneys to the investors instead of pursuing company development. Miller and Rock (1985) states that high dividend yield signaling that the firm is not expanding their capitals under asymmetric situation. Thus, if the stock prices are mainly influenced positively by the DY, the customers expected e-commerce and express industry both have less investment on future development and allocate the profit back to the investors. (Yang, 2019)

4. Discussion

We analyze the reasons why the two industries do not need a massive amount of investment, which leads to the high DY.

To start with, each industry works as coordinator instead of producer, so there is no investing fee on production. The express delivery industry is hugely depending on labor. (Hu, et.al, 2011; Su, et.al, 2019) The basic running model of the express delivery industry could be described as below, while the customer creates an order, and the express industry helps them only work as a coordinator to move the package they need and deliver them. They have nothing to do with the product itself or finding the customer since these tasks belong to the production company. Because of that, the express industry is relatively cost-efficient. (Lakew, 2014)

Similarly, the e-commerce industry is mainly a platform to connect the shops and customers. Even the delivery process is not related to the e-commerce platform. Combining the previous analysis, neither of them need to produce any type of product; instead, they are more alike to service companies who offer services. There is no fee on the production process, such as new technology, industry, and storage.

Secondly, the marketing cost is relatively low. Both industries indeed rely on a large amount of customer pool, but the pool is relatively easy to approach. (Global E-Commerce Market 2021-2025, 2021) As online purchase develops, people are more get used to it, especially during the Post-COVID era. For the sake of convenience, though some brands have an online shopping website, most customers still choose the assembled ones, like Amazon. Further, if they choose to shop online, it means that they must use delivery services, which directly connects to the turnover of the express

industry. Also, express delivery is the only way people send packages to people far. Thus, to conclude, nowadays, there is no way that people could escape from the two industries just mentioned.

Also, since the market is relatively huge, mainly, there will be huge amounts of orders devoted to the existed huge companies. (GAO, Mao, 2021) Further, since most online shopping platforms or delivery companies work as the same, there is small differentiation among, the orders will relatively be devoted averagely to the companies. To sum up, there is relatively less need for the two industries to put on a huge amount of advertisement since they are both needed massively. (Shi-yi, 2021)

Lastly, both industries have relatively low profits, so the DY is a way to show their ability to earn profit. Working as a coordinator, most of the profits go to the production companies, so the left profits are relatively small. Thus, high DY gives the investor confidence in this company's potential for return.

From the previous analysis, we know that why the DY is crucial for the e-commerce industry and express industry. The reason why they have this same property is complex, but one of the reasons is the express industry is downstream of the e-commerce industry on the chain. When the e-commerce platform runs, the express industry must jump in. Thus, both industries are connected to each other closely, and somehow, they develop through the same stage. It could also be said that the investors started to look forward to profiting from these two low-profit industries.

5. Conclusion

As there are several factors in value investing, to figure out the out influential one is crucial for value investors.

To analyzing the most effective factors in the two close sectors, the e-commerce industry and the express industry, the multiple regression analysis is utilized. Through the analysis, it could be told that the highest influential index is the dividend yield (DY), which indicating that the companies pay their money to the investors instead of pursuing further investment or development. The reasons behind found by comparisons are: (1) both industry work as coordinators instead of producers. (2) Both industry has low marketing cost. (3) DY is a way to show their abilities of earning with relatively low profit.

References

- [1] Bin-Hui Wang. (2016). Multivariate statistical analysis and R language modeling.
- [2] De-Yun Cao. (2019). Adhere to the concept of value investment and be the forerunner of long-term funds in the capital market. 31-33.
- [3] GAO, J., & Mao, H.-J. (2021). Forecasting China's Express Delivery based on R using ARIMA Model. 2021 International Conference on E-Commerce and E-Management (ICECEM), E-Commerce and E-Management (ICECEM), 2021 International Conference on, ICECEM, 197–200.
- [4] Global E-Commerce Market 2021-2025: Size, Trends and Forecast with Impact Analysis of COVID-19. (2021). M2PressWIRE.
- [5] Hu, Q., Schaufeli, W.B., & Taris, T.W. (2011). The Job Demands-Resources Model: An Analysis of Additive and Joint Effects of Demands and Resources. *Journal of Vocational Behavior*, 79, 181-190.
- [6] Lakew, P. A. (2014). Economies of traffic density and scale in the integrated air cargo industry: The cost structures of FedEx Express and UPS Airlines. *Journal of Air Transport Management*, 35, 29–38.
- [7] Merton H. Miller, & Kevin Rock. (1985). Dividend Policy under Asymmetric Information. *The Journal of Finance*, 40(4), 1031–1051

- [8] Shu-Fen Yang. (2019). Long-term Investment Research on Companies with High Dividend Yields and Low P/E Ratios - Evidence from Taiwan Listed Companies. Master's in-service special class thesis of Institute of finance, Chenggong University, 1.
- [9] Sih-Yi You. (2021). Research on the Network Marketing Strategy of the Cultural and Creative Industries during the COVID-19 Epidemic: Take the Cultural and Creative E-commerce Platform of the National Palace Museum Shop as an Example. *Journal of Cultural and Creative Industries Research*, 1, 27.
- [10] Su, F., Lei, Y., He, Y., & Luo, H. (2019). Perceived High-performance Work Systems and Work Well-being in the Express Industry: A Moderated Mediation Model. 2019 16th International Conference on Service Systems and Service Management (ICSSSM), Service Systems and Service Management (ICSSSM), 2019 16th International Conference On, 1–6.