MECON6102 Problem Set 2

Xing Mingjie

April 25, 2024

Abstract

This report studies the relationship between monthly stock returns and a set of factors and construct an investment strategy based on various factor models. The report first uses a naive factor regression to estimate the factor loadings of the stock returns. Then, the report uses the Fama-MacBeth regression to estimate the factor risk premia. Finally, the report uses the LASSO regression to select the factors. The report also constructs a mean-variance portfolio based on the factor risk premia.

1 Data

1.1 Description

Table **??** shows the summary statistics of the data. The data set contains 13982 observations and 9 variables. The dependent variable is the default label, which is a binary variable indicating whether the individual defaults.

Figure **??** shows the heat map of the correlation matrix of the features and target. Most of the features are arguably uncorrelated. There is a high correlation between housing and age at 0.55. The correlation between income and education level is 0.51, which captures the wage premium of education.

1.2 Data Preprocessing

Figure **??** shows the distribution and the skewness of feature `income`. The distribution is right-skewed. The report uses the log transformation to reduce the skewness of the feature for better performance in models. Figure **??** shows the distribution of the feature `income` after the log transformation. The distribution is more symmetric after the transformation.

2 Model Comparison

2.1 Simple Logistic Model

3 Conclusion

Table 1: Data Description

	count	mean	std	min	max
default_label	13982.00	0.02	0.15	0.00	1.00
age	13982.00	41.66	14.56	17.00	66.00
gender	13982.00	0.46	0.50	0.00	1.00
edu	13982.00	1.69	1.10	0.00	4.00
housing	13982.00	0.63	0.48	0.00	1.00
income	13982.00	7426.48	6226.68	650.42	37515.37
job_occupation	13982.00	0.34	0.56	0.00	2.00
past_bad_credit	13982.00	0.96	0.19	0.00	1.00
married	13982.00	0.53	0.50	0.00	1.00

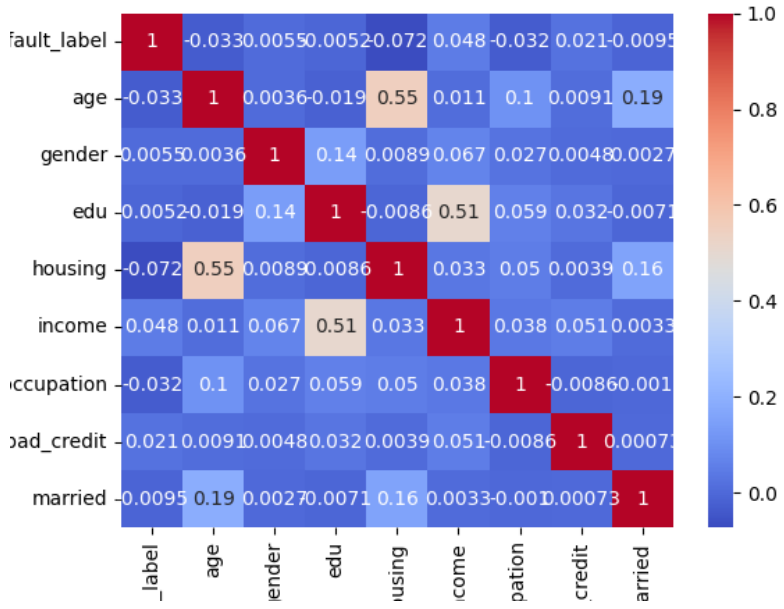

Figure 1: Heat map of the correlation matrix of the variables

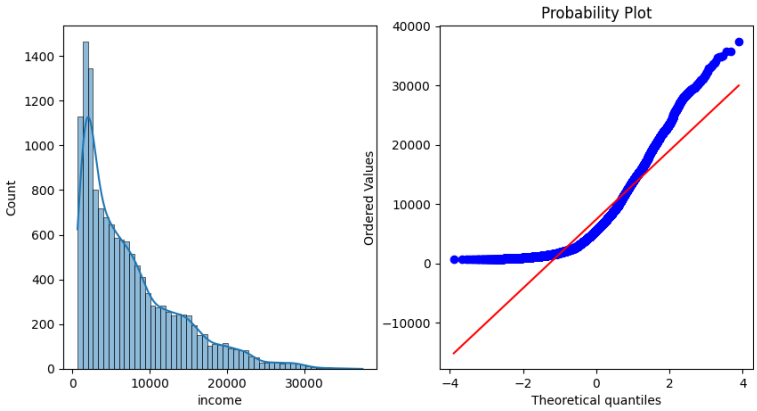

Figure 2: Income Distribution

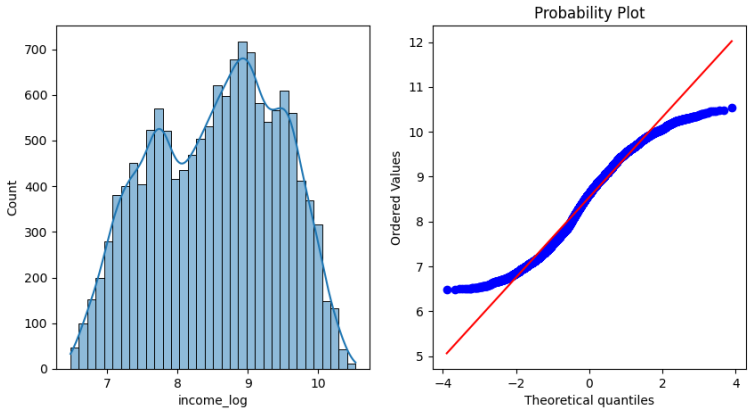

Figure 3: Income Distribution After Log Transformation