4. Analysis

In this section, we present the result for each model by investigating hyper-parameters, coefficients and Mean Square Error. Lastly, we choose the best model based on Mean Square Error.

4.1 Baseline model

In order to set the baseline model, we first fit the ordinary least square regression using eleven predictors. Unlike other models, OLS does not have a tuning paramter. The p-values in coefficients indicate that Income, Limit, cards and StudentYes are statistically significant. Also, adjusted R-squared shows that this model explains well about the data.

4.2 Tuning parameter selection

As mentioned in methods section, we use 10-fold cross validation to tune hyper-paramters for each model.

Ridge regression penalizes predictors' weights by L2 norm. And lambda determines the magnitude of the penalty. Figure 1 shows that MSE increases as the lambda increments. Using cross validation, we finally obtain the minimum lambda that maximizes MSE, which is 0.01.

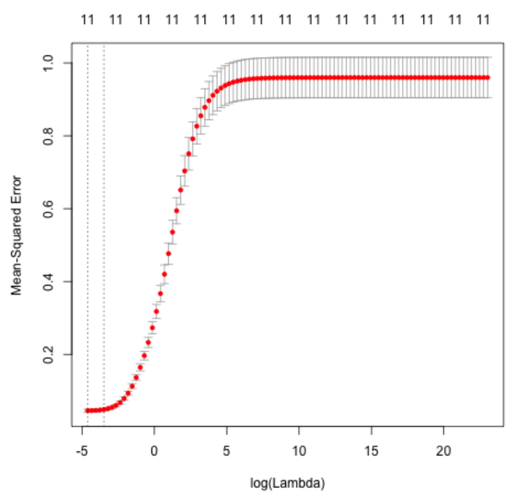


Figure 1: Lambda for Ridge

Lasso regression penalizes predictors' weights by L1 norm. Similar to Ridge regression, its lambda determines the magnitude of the penalty. Figure 2 shows that MSE increases significanly as lambda grows. Using cross validation, we finally obtain the minimum lambda that maximizes MSE, which is 0.01.

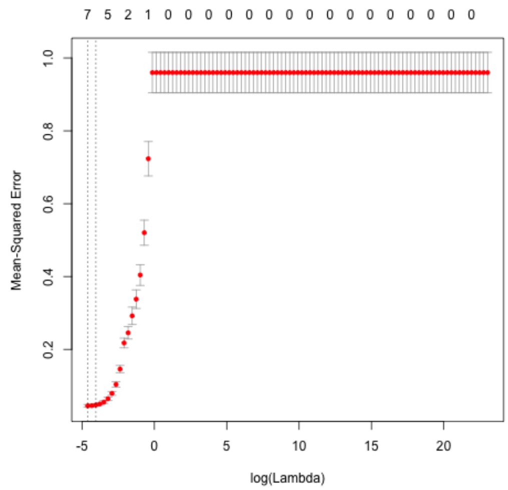


Figure 2: Lambda for Lasso

This is an interesting result in that both Ridge and Lasso have very small lambda. It means that both models end up penalizing a little bit. This makes sense because there are not many predictors and predictors are quite independant.

Principal Component Regression fits a linear regression on newly generated basis, principal components. The number of principal components is important. To obtain the best number, we use 10-fold cross validation and select one that minimizes RMSEP. Figure 3 shows that both ten and eleven PCs are very simliar and bring about the lowest RMSEP. So, we end up choosing ten pricipal components.

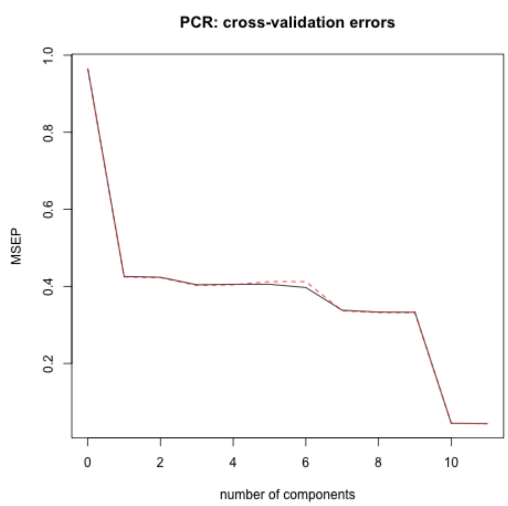


Figure 3: PCR cross validation

Partial least squares regression bears some relation to principal components regression. Instead of finding hyperplanes of maximum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space. [1] Again, The number of principal components is important. Similarly, to obtain the best number, we use 10-fold cross validation and select one that minimizes RMSEP. Figure 4 shows that four to eleven PCs bring out almost the same RMSEP. If our goal is to reduce dimensionality, we can select either four or five principal components. We end up choosing four.

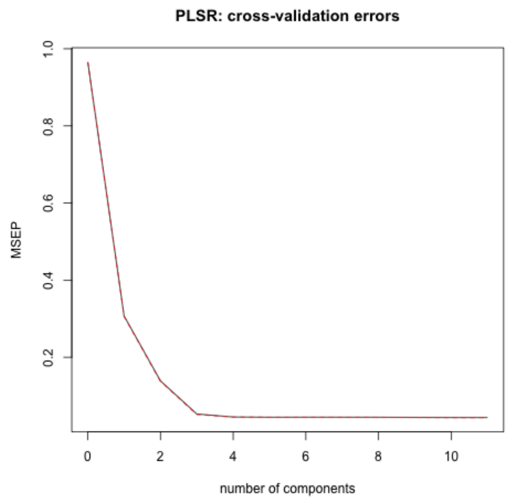


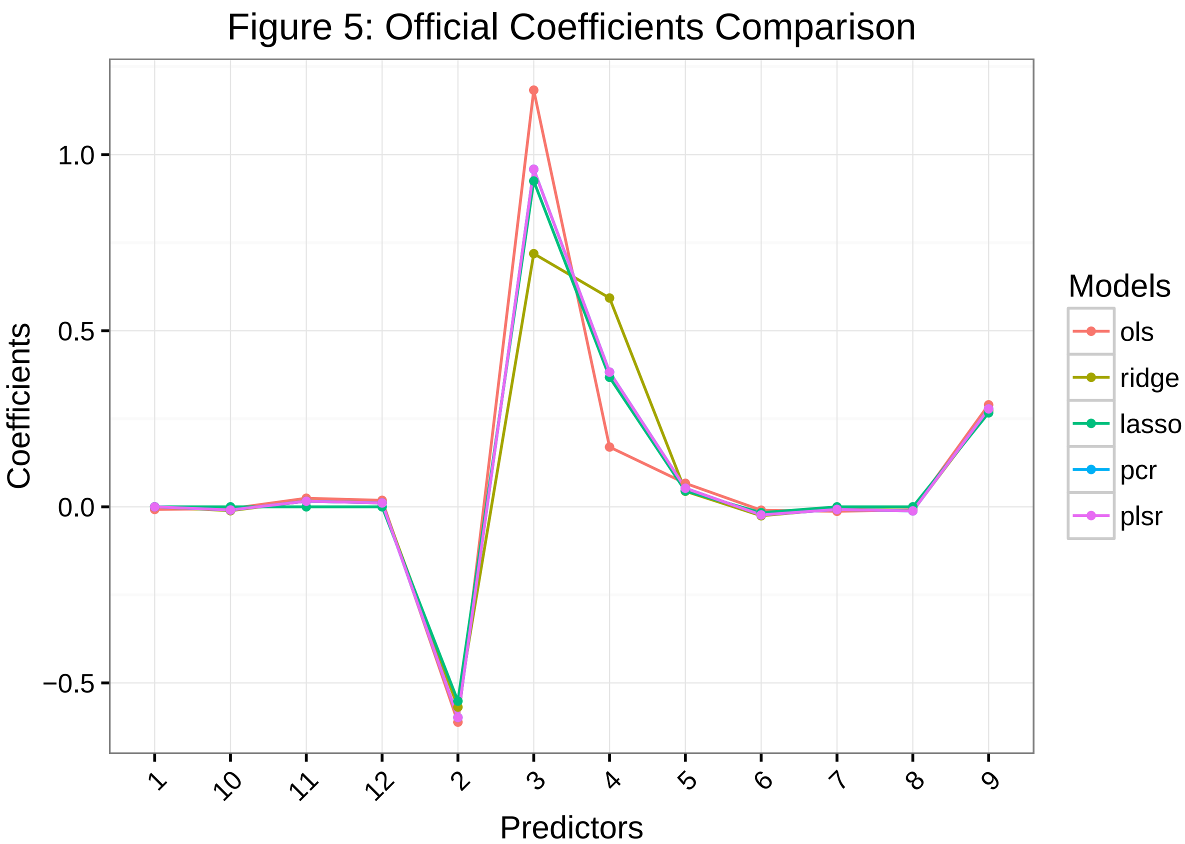
Figure 4: PLSR cross validation

4.3 Coefficients

Table 1 and Figure 5 show the coefficients for five different models. They show that cofficietns for predictors in five models are very simliar and there are slight differences in **Limit** and **Rating** predictors. It is noteworthy that there are a lot of zero weights in Lasso. So, Lasso effectively removes unnecessary predictors

Table 1: Coefficient Table

	ols	ridge	lasso	pcr	plsr
1	-0.01	0.00	0.00	0.00	0.00
2	-0.61	-0.57	-0.55	-0.60	-0.60
3	1.18	0.72	0.93	0.96	0.96
4	0.17	0.59	0.37	0.38	0.38
5	0.07	0.04	0.05	0.05	0.05
6	-0.01	-0.03	-0.02	-0.02	-0.02
7	-0.01	-0.01	0.00	-0.01	-0.01
8	-0.01	-0.01	0.00	-0.01	-0.01
9	0.29	0.27	0.27	0.28	0.28
10	-0.01	-0.01	0.00	-0.01	-0.01
11	0.02	0.02	0.00	0.02	0.02
12	0.02	0.01	0.00	0.01	0.01



4.4 Model Comparison and selection

Table 2 shows MSE for each model. This indicates that Ridge brings out the lowest MSE and Lasso and PCR give almost the same MSE and OLS gives the highest MSE. So, we decide to select Ridge regression as our best model.

Table 2: MSE Table

	mse
ols	0.065650
ridge	0.061220
lasso	0.062670
pcr	0.062090
plsr	0.063440