MSc BUSINESS ECONOMICS WITH ANALYTICS

Athens University of Economics and Business

APPLICATIONS OF ANALYTICAL METHODS IN BUSINESS ECONOMICS & STRATEGY

ASSIGNEMENT 3

Kantali Niki

ABSTRACT

In this project, we applied Ridge and Lasso regression techniques to a monthly dataset of Italy's industrial production as the response variable and the hard and soft variables as the predictors. Ridge regression works by adding a penalty term to the OLS cost function to reduce the magnitude of the coefficient estimates, thus preventing overfitting and improving the model's generalization performance. We used cross-validation to choose the optimal value of the regularization parameter and selected an appropriate threshold for the p-values to determine the statistical significance of the coefficients. Lasso regression, on the other hand, uses an l1 penalty term to shrink some of the coefficients to exactly zero, resulting in a sparse model and potentially improved interpretability. We compared the coefficient estimates between Ridge and Lasso regression and observed any differences. Overall, this project provided an opportunity to practice and compare two commonly used regularization techniques in linear regression and to interpret their results in the context of a real-world dataset.

I. Ridge Regression

Ridge regression is a regularization technique that adds a penalty term to the OLS function. The penalty term, which is proportional to the sum of the squared values of the regression coefficients, shrinks the magnitude of the coefficients towards zero, effectively reducing the variance in the estimates. This reduction in variance comes at the expense of a slight increase in bias, which is generally acceptable as it can help prevent overfitting and improve the model's generalization performance. Overall, Ridge regression improves over simple OLS by reducing the sensitivity of the estimates to the noise in the data, leading to more stable and reliable predictions. Therefore, many of the elements of $\hat{\beta}$ will be very small (close to zero) but not zero (no model selection).

$$R(f) = \sum_{i=1}^{k} b_i^2 = \|b\|_2$$

$$\sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda R(f)$$

To obtain $\hat{\beta}^{*}_{(\lambda)} = (S + \lambda I)^{-1} X^T Y = (S + \lambda I)^{-1} S \hat{\beta}$

Where $\hat{\beta} = (X^T X)^{-1} X^T Y = S^{-1} X^T Y$ and $S = X^T X$

II. Lasso Regression

Lasso regression, like Ridge regression, adds a penalty term to the OLS cost function to prevent overfitting and improve the generalization performance of linear regression models. However, instead of the L2 penalty used in Ridge regression, Lasso regression applies an L1 penalty, which is proportional to the sum of the absolute values of the coefficients. This leads to a sparse model, where some of the coefficients are shrunk to exactly zero, resulting in variable selection and potentially improved interpretability. The amount of shrinkage is controlled by a regularization parameter, lambda, which balances the tradeoff between model complexity and goodness of fit. As with Ridge regression, a larger value of lambda leads to greater shrinkage and a simpler model, while a smaller value of lambda results in less shrinkage and a more complex model. By adding the L1 penalty term, Lasso regression encourages sparsity in the coefficient estimates, which can be particularly useful when dealing with datasets that have many features but only a small number of them are expected to be relevant to the response variable. In contrast to Ridge regression, Lasso regression can perform variable selection and provide a more interpretable model.

$$R(f) = \sum_{i=1}^{p} b_i = \|b\|_1$$

$$\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda R(f)$$

III. Differences between Lasso and Ridge Regressions

Lasso uses L1 regularization, which can shrink some coefficients to exactly zero, resulting in a sparse model with improved interpretability. Ridge, on the other hand, uses L2 regularization, which cannot force coefficients to zero, but provides better predictive performance when many variables have small or medium effects.

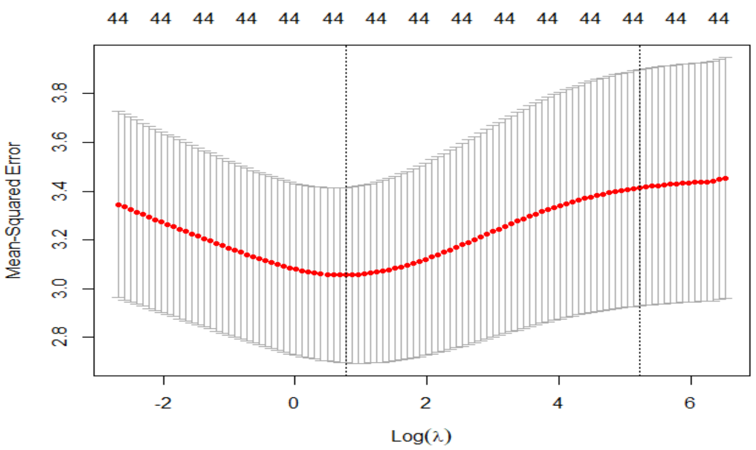
IV. Methodology

Our first step was to load our data. Our data is the Italy's industrial production. The dependent variable Y is the industrial production, and we have X1 which includes all the hard variables and the X2 that includes the soft variables. We combine to a general X both X1 and X2.

<u>Ridge</u>

We start with the first shrinkage method, the Ridge regression. In Step 1, **cv.glmnet()** function from the **glmnet** package is used to perform k-fold cross-validation (with k equal to the number of rows in Y) on the data, where X is the matrix of predictors and Y is the response variable. The **alpha** argument is set to 0 to specify Ridge regression and **standardize** and **intercept** arguments are set to TRUE to standardize the predictors and fit an intercept. In Step 2, the value of lambda that results in the smallest mean-square error is identified using **outcv$lambda.min**. The image 1 shows the value of log(l) that minimize the mean-square error that is log(lamda)=0.7648. Additionally, a plot of the coefficient paths as log(lambda) increases is displayed in image 2. We can see that for log(lamda)=0.7648 the coefficients are shrunken.

Image 1 Ridge – MSE over log (λ)



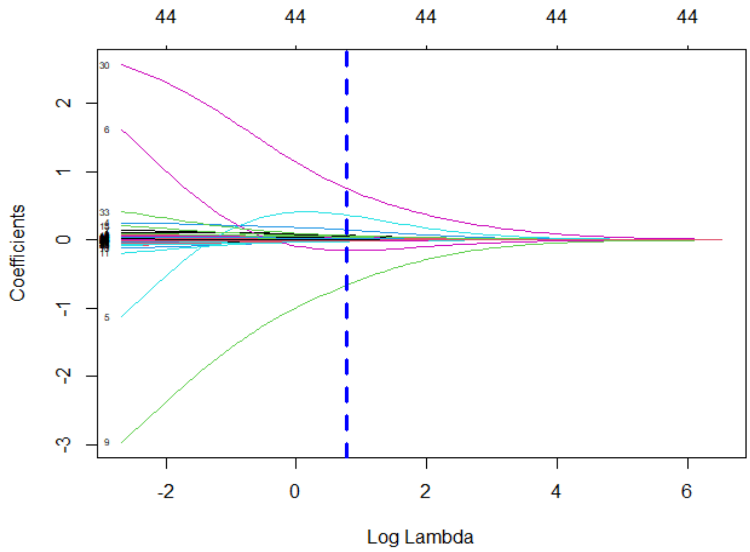


Image 2 Ridge-Coefficients Path

Finally, the coefficients at the optimal value of lambda were extracted and they are displayed in the table 2. In Ridge regression the model selection isn't automatic like in the Lasso. The threshold we used was 0.03 in order to sustain an appropriate variable number.

Table 1 Coefficients from Ridge

	V1		V1
alpha	0	SD14	0
HD2	0.065	SD15	0
HD3	0.040	SD16	0
HD4	0	SD17	0
HD5	0.146	SD18	0
HD6	0.363	SD19	0
HD7	-0.156	SD20	0
HD8	0.056	SD21	0.751
HD9	0.000	SD22	0
HD10	-0.666	SD23	0
SD1	0	SD24	0.063
SD2	0	SD25	0
SD3	0	SD26	0
SD4	0	SD27	0
SD5	0	SD28	0
SD6	0	SD29	0
SD7	0	SD30	0
SD8	0	SD31	0
SD9	0	SD32	0
SD10	0	SD33	0
SD11	0	SD34	0
SD12	0	SD35	0
SD13	0		

Lasso

In this shrinkage method in the function cv.glmnet we altered the a to 1 (a=1) instead of alpha=0 as we had in Ridge.

The lmin is 0.2050999 and the log(lmin)= -1.584258. In image 3 we see the values of log(lamda) that minimize the MSE.

In image 4 we see the coefficient path. Finally, the table 2 shows the coefficients.

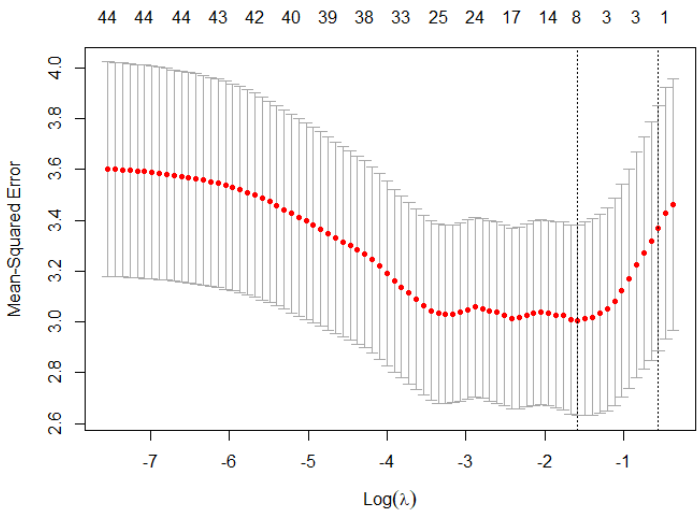


Image 3 Lasso - MSE over log(λ)

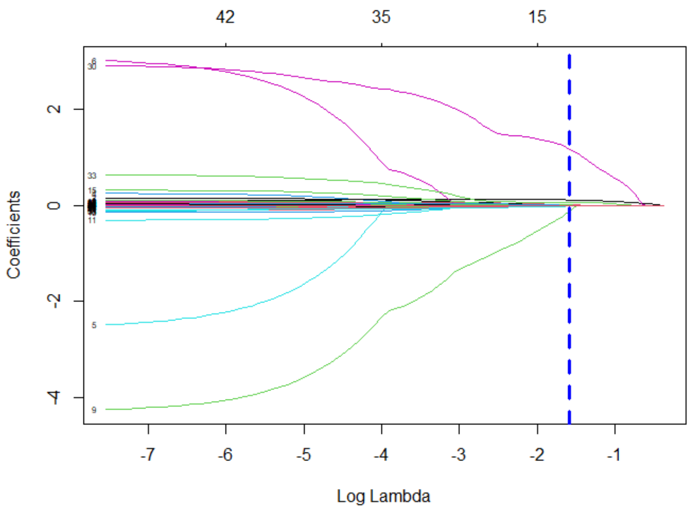


Figure 1 Lasso - Coefficients Path

	V1		V1
alpha	-0.0254	SD14	0
HD2	0.1100	SD15	0
HD3	0.0100	SD16	0
HD4	0	SD17	0
HD5	0	SD18	0
HD6	0	SD19	0
HD7	0	SD20	0
HD8	0	SD21	1.1661
HD9	0	SD22	0
HD10	-0.1232	SD23	0
SD1	0	SD24	0.0639
SD2	0	SD25	0
SD3	0	SD26	0
SD4	0	SD27	0
SD5	0	SD28	0
SD6	0	SD29	0
SD7	0	SD30	0.0039
SD8	0	SD31	0
SD9	0.0004	SD32	-0.0170
SD10	0	SD33	0
SD11	0	SD34	0
SD12	0	SD35	0
SD13	0		

Table 2 Coefficients from Lasso

V. Results

As we can see, when we compare the results of the Ridge and the Lasso (table 1 and table 2), we notice that only a small number of common variables are important in both shrinkage methods.