

# Credit Risk Modelling: Shrinkage Methods and Lasso Selection in PD Modelling (Part 2)

written by Hikmat Abdulazizov Hikmät Əbdüləzizov

In the realm of statistical modeling and machine learning, regression analysis is a fundamental tool for understanding and predicting relationships between variables. Traditional linear regression, while powerful, often faces challenges when dealing with high-dimensional datasets or when multicollinearity among predictors is present. In such cases, techniques like Lasso and Ridge regression offer great solutions by introducing regularization.

Lasso (Least Absolute Shrinkage and Selection Operator) and Ridge regression are both techniques aimed at mitigating overfitting and improving the generalization of regression models. They achieve this by imposing penalties on the coefficients of the regression variables, thereby shrinking them towards zero while still maintaining their predictive power. These methods have found wide-ranging applications in fields as diverse as economics, genetics, and machine learning.

Lasso and Ridge regression are indispensable tools in the machine learning toolkit because they offer powerful mechanisms for controlling model complexity, improving generalization performance, and facilitating feature selection. By leveraging the principles of regularization, these techniques empower machine learning practitioners to build more robust, interpretable, and predictive models across a wide range of applications.

In this paper, we delve into the principles, methodologies,

and applications of Lasso and Ridge regression techniques. We explore their theoretical foundations, discuss their differences and similarities, and examine practical considerations for their implementation. Moreover, we highlight scenarios where Lasso and Ridge regression excel, and provide insights into their advantages and limitations compared to traditional linear regression methods.

Through a comprehensive examination of Lasso and Ridge regression, we aim to equip readers with a deeper understanding of regularization techniques and their role in modern data analysis. By clearing the intricacies of these methods, we seek to empower researchers, analysts, and practitioners with valuable tools for building robust and interpretable regression models in the face of complex real-world datasets.

Lasso and Ridge regression are two prominent approaches to regularization, each with distinct characteristics and advantages. Lasso regression, introduced by Robert Tibshirani in 1996, incorporates an L1 penalty term that encourages sparsity by driving some regression coefficients to exactly zero, effectively performing variable selection. On the other hand, Ridge regression, proposed by Hoerl and Kennard in 1970, employs an L2 penalty term to shrink the coefficients towards zero without eliminating them entirely, thus reducing the impact of multicollinearity and stabilizing the model.

The widespread adoption of Lasso and Ridge regression techniques across various disciplines underscores their relevance and effectiveness in modern data analysis. These methods not only offer robust solutions to common regression challenges but also provide valuable insights into feature importance and model interpretability. Furthermore, advancements in computational algorithms and optimization techniques have made the implementation of Lasso and Ridge regression more accessible and efficient than ever before.

We discuss the computational aspects of Lasso and Ridge regression, including algorithms for parameter estimation and model selection. We also highlight real-world case studies and empirical findings that showcase the efficacy of these techniques in domains such as feature selection, predictive modeling, and risk analysis.

## Ridge regression

In the simplest case, the problem of a near-singular moment matrix,

$$\mathbf{X}^T \mathbf{X}$$

is alleviated by adding positive elements to the diagonals, thereby decreasing its condition number, making it much more solvable. Analogous to the ordinary least squares estimator, the simple ridge estimator is then given by

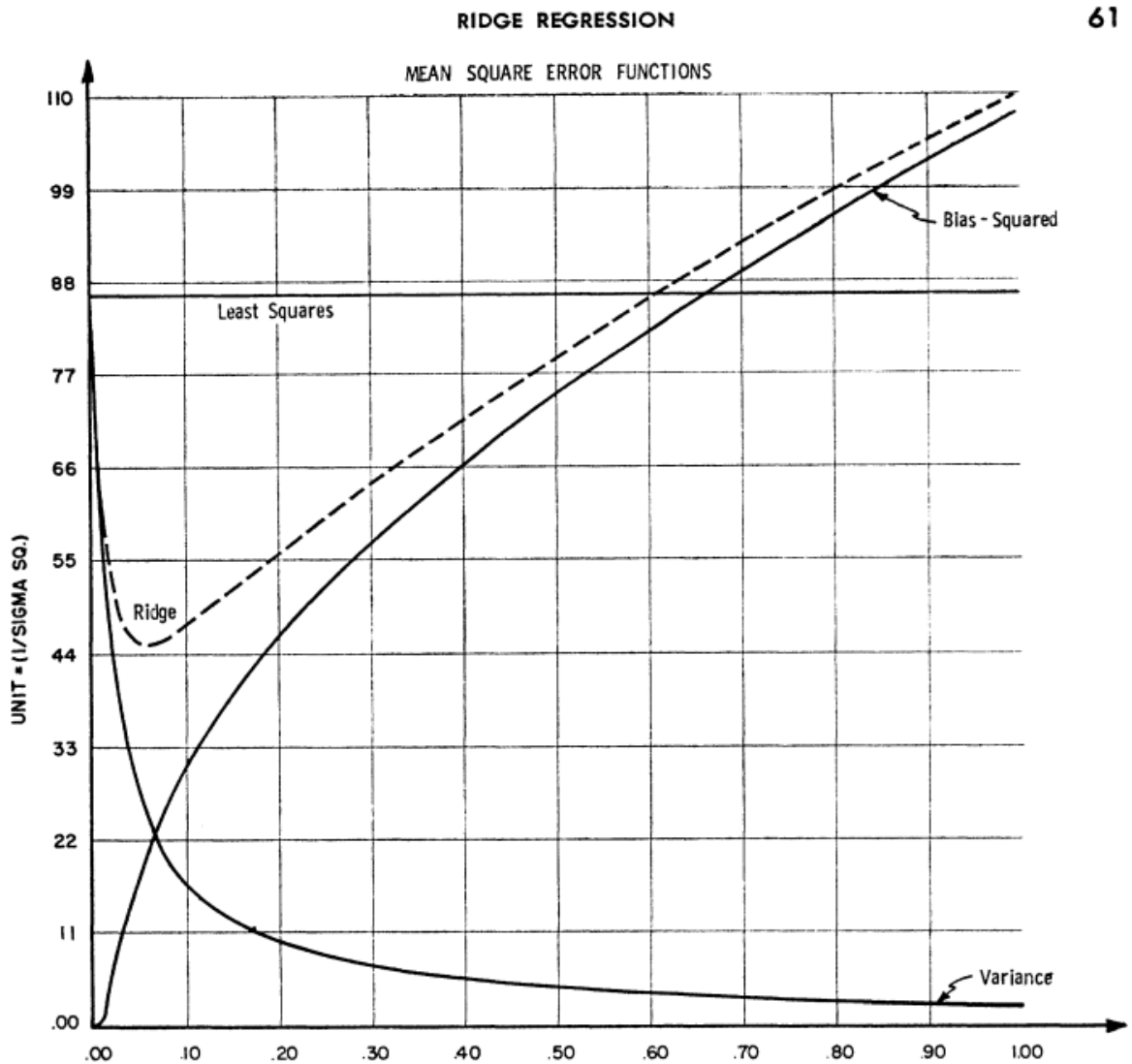
$$\hat{\beta}_R = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$$

where  $\mathbf{y}$  is regress and  $\mathbf{X}$  is the design matrix,  $\mathbf{I}$  is the identity matrix and the ridge parameter positive  $\lambda$  serves as the constant, which shifts the diagonals of the moment matrix. Note that when  $\lambda$  is zero, the equation boils down to simple OLS regression. See below the minimization problem of regularization.

$$\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \lambda (\beta^T \beta - c)$$

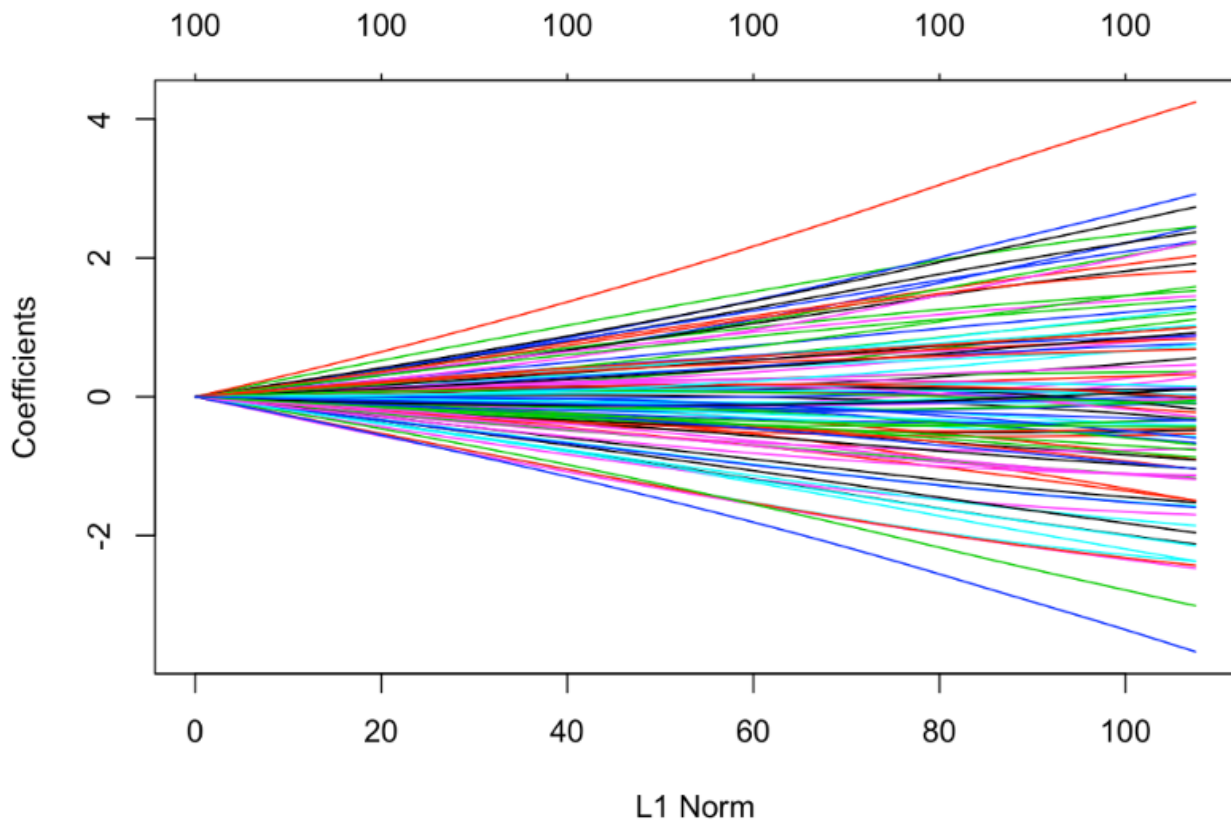
The first part is simply the standard conventional OLS part and the second part is for the ridge penalization. That is why we have smaller coefficients than OLS. The term “shrinkage model” emerges from this. The advantage of using ridge regression instead of OLS is shown in the graph below.

## Graph 1. Optimality of Shrinkage



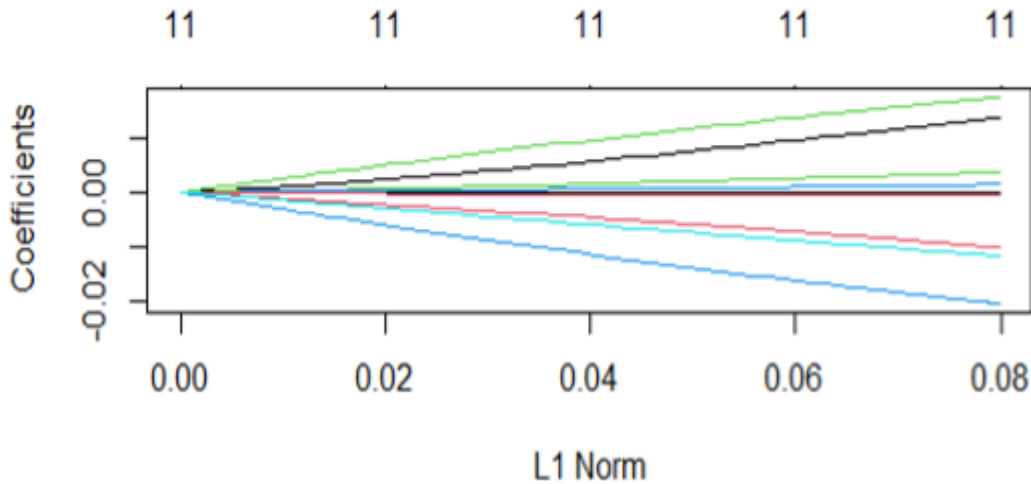
As you can see from the graph above, by introducing more bias we can decrease the variance substantially and decrease overall the mean squared error because mean squared error accounts for some of the variance and bias squared. This is very important because it means that by using ridge estimator, we can increase our predictive power and have more reliable models at the end. Now let's look at the R program outputs of the ridge model using a glmnet package with simulation and using real world credit risk data.

## Graph 2. Coefficient Path for Simulated Ridge



Above is the output of the ridge regression showing coefficient paths by tuning the lambda parameter. Please do not forget that when lambda increases, penalization becomes harsher and the beta parameters smaller. So, the graph shows coefficient paths while the regularization parameter decreases. Coefficients start from zero when the lambda is the highest. Then the model suddenly picks up all the variables (100 variables here) and the coefficients increase. This is a very important feature of the ridge regression since it claims that we can use all variables in the model regardless of their intercorrelation. In the most radical case, I have used the same variable (thus correlation coefficient equals exactly 1) and the model still picks both of them.

**Graph 3. Coefficient Path for Real World Credit Risk Data for Ridge**



The graph above shows the result from our credit risk data. We used 11 variables for the purpose of parsimony. It results in a path similar to the simulated data.

## Lasso Regression

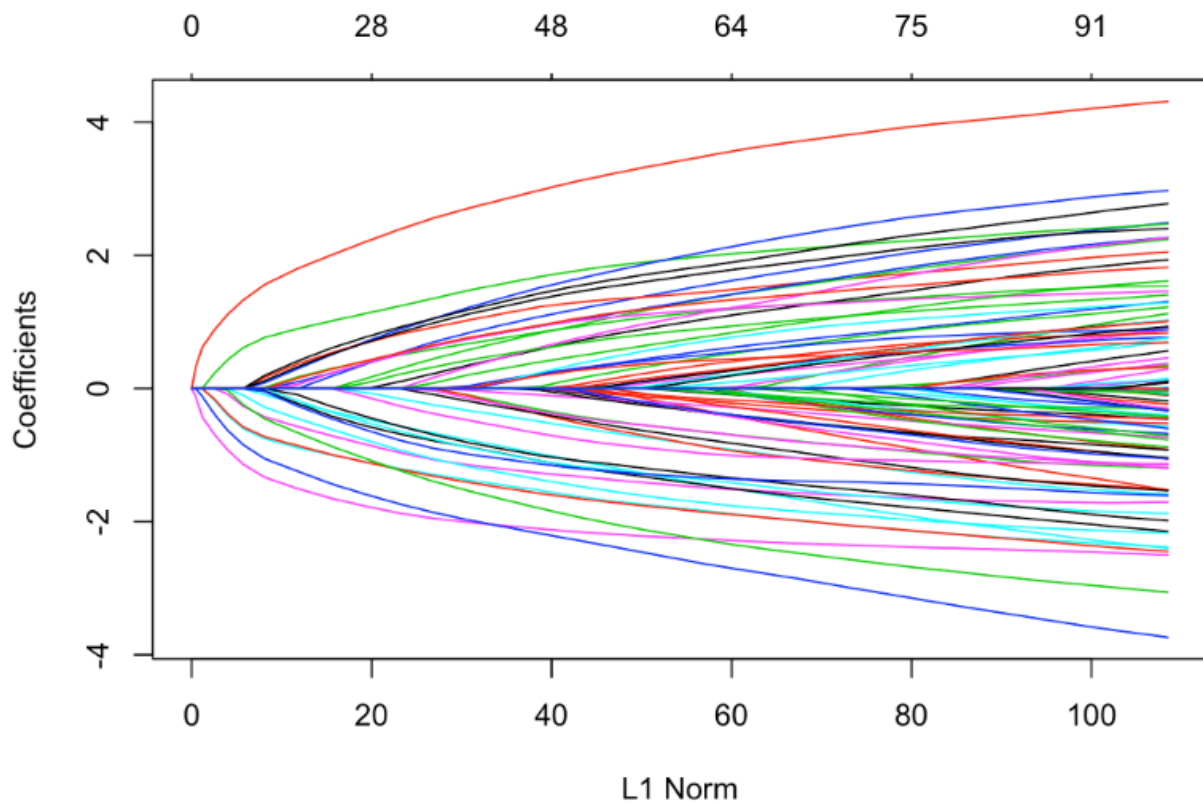
In statistics and machine learning, Lasso (least absolute shrinkage and selection operator; also lasso or LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model. The Lasso method assumes that the coefficients of the linear model are sparse, meaning that few of them are non-zero. It was originally introduced in geophysics, and later by Robert Tibshirani, who coined the term. Contrary to the Ridge regression we here use absolute values of parameters to shrink.

$$\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{N} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\}$$

As a result, we end up shrinking some parameters to zero. This is an interesting characteristics of the Lasso since it helps us choose important variables (those that are not shrunk to

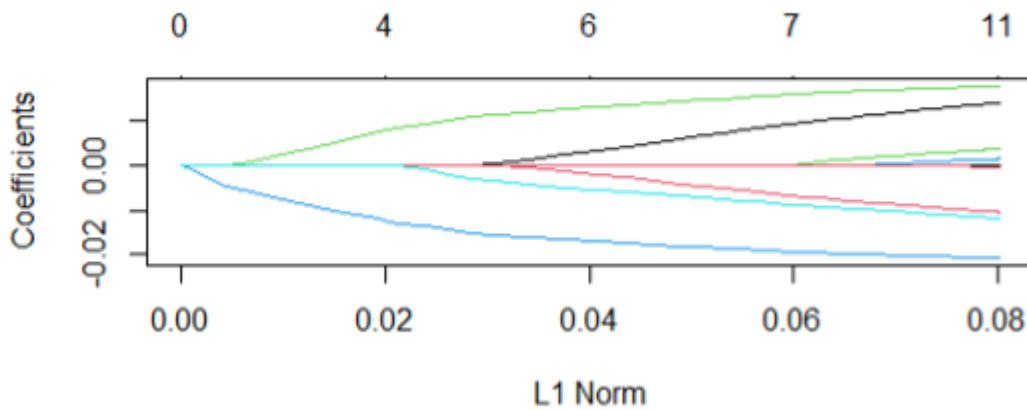
zero). It gives us a modern approach other than the p value for the significance test. We can use the bulk of variables in the model and the Lasso, depending on the lambda parameter, of course, all of which will tell us which variables we need to emphasize more. Now let's look at the coefficient paths.

**Graph 4. Coefficient Paths for Simulated Lasso**



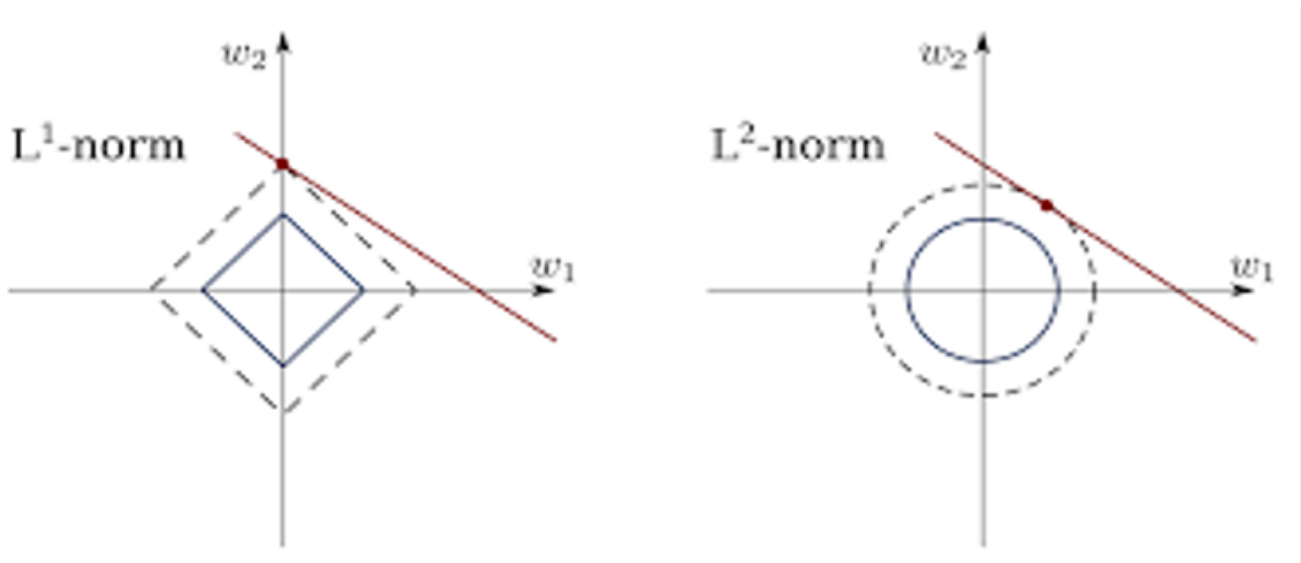
Above we see the coefficient paths from the simulated 100 variables. As you see, it is visibly very different than the Ridge coefficients paths. As per lambda the model chooses 28 variables first, and then the model chooses 48 variables to be important and not shrunk to zero. If we did variable selection with p values, it would consume extraordinary time and effort to end up with a model with 28 variables out of 100. Now we do it with one command, and this is the efficacy of the Lasso modelling.

**Graph 5. Coefficient Path for Credit Risk Data for Lasso**



Above is the result from our real-world credit risk data. It is evident from the graph that Lasso first chooses 4 variables and then the count of variables increased as we change tuning parameter lambda. The model is compatible with the simulated data as well.

### Graph 6. Regularization Mechanisms



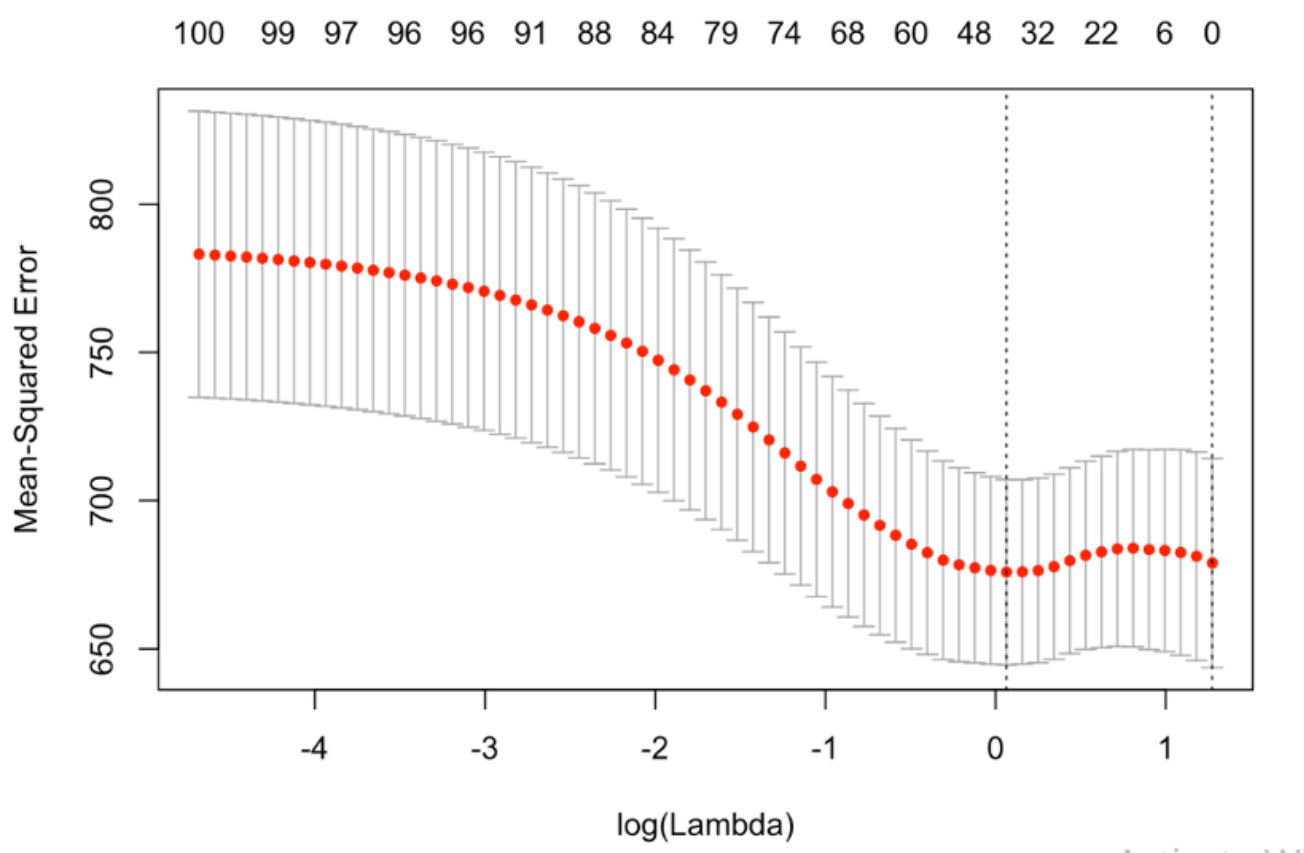
Above are graphs that show the difference between the Lasso and Ridge regressions. The first is Lasso ( $L^1$  norm) and the second is Ridge ( $L^2$  norm). It is clear from the graphs why Lasso chooses one parameter to be zero and Ridge chooses both in the model .

Below is the output for the Lasso model showing mean squared error per lambda. It is very nice to see that how parameters



decrease via lambda increases and how mean squared error is minimized.

### Graph 7. Error Minimization and Variable Selection



In conclusion, the exploration of Lasso and Ridge regression techniques illuminates their significance as powerful tools in the realm of statistical modeling and machine learning. Throughout this paper, we have delved into the theoretical foundations, methodologies, and practical applications of Lasso and Ridge regression, shedding light on their roles in enhancing model robustness, interpretability, and predictive performance.

Lasso and Ridge regression offer elegant solutions to the challenges inherent in regression analysis, including overfitting, multicollinearity, and high-dimensional data. By introducing regularization penalties that penalize the magnitude of regression coefficients, these techniques strike a delicate balance between bias and variance, leading to models that generalize well to unseen data while maintaining

interpretability.

From a practical standpoint, Lasso and Ridge regression present valuable advantages for feature selection, variable importance assessment, and model stability. The L1 penalty of Lasso regression promotes sparsity by driving certain coefficients to zero, facilitating automatic feature selection and enhancing model interpretability. Meanwhile, the L2 penalty of Ridge regression mitigates the effects of multicollinearity, stabilizing the estimation process and improving model performance in correlated predictor variable scenarios.

The widespread adoption of Lasso and Ridge regression techniques across various domains underscores their versatility and efficacy in addressing real-world regression challenges. From finance and economics to healthcare and engineering, these techniques find applications in diverse fields where data-driven insights and predictive modeling are paramount.

Looking ahead, continued research and innovation in regularization techniques promise to further enhance the capabilities of Lasso and Ridge regression. Advancements in computational algorithms, optimization methods, and model interpretability tools will continue to expand the applicability and accessibility of these techniques, empowering researchers, analysts, and practitioners to extract meaningful insights from complex datasets.

In summary, the study of Lasso and Ridge regression is an important part of contemporary data science because the two offer a sophisticated framework for building robust, interpretable, and predictive regression models. By understanding the principles and applications of these techniques, we can leverage their power to tackle the complexities of modern data analysis and unlock new avenues for knowledge discovery and decision-making.

## References:

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55–67.  
<https://doi.org/10.2307/1267351>

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288.  
<http://www.jstor.org/stable/2346178>

[Credit Risk Modelling: Shrinkage Methods and Lasso Selection in PD Modelling \(Part 1\)](#)