LASSO: Modelling an ETF's Price Movements

Bryan Chen, Cameron Fisher, Reggie Hayhurst, Finlay Webb and Andrew Wright

L1 Regularisation

$$\Sigma_{i=1}^{n}(y_i - \Sigma_j x_{ij}\beta_j)^2 + \lambda\Sigma_{j=1}^{p}|\beta_j| \qquad (1)$$

What is LASSO Regression and How Does It Work?

The Least Absolute Shrinkage and Selection Operator (LASSO) is a regression analysis method used to improve the accuracy of a statistical model by reducing the data's bias towards the training data set. This is due to the fact that the weight finding technique β_{OLS} may give a model the lowest possible variance in the training data set, but when given new data rarely gives a good estimate, due to an over fitting problem. LASSO fixes this by introducing a penalty term to our RSS, the degree of which is dictated by our regularisation parameter λ (the second term in (1)). This penalty term dampens the contribution of parameters with high *multicollinearity*, ultimately reducing over fitting.
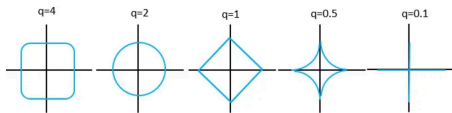

Figure 1: A geometric visualisation of our penalty term λ, where $Q = 2$ represents Ridge Regression's (see RHS of poster, and $Q = 1$ represents LASSO's. These are called constraint regions.

As we increase our hyperparameter, λ, our model becomes simpler with fewer variables, while when we decrease λ our model becomes complex with a greater number of parameters. A λ of zero will simply provide us with the OLS. The optimal level of λ is selected through k-fold cross-validation, there is information on this technique in the middle of the poster. LASSO can trace its origins back to a Geophysics paper [1] where the authors used a L1 penalty term on coefficients. However, in 1996 LASSO Regression was formally proposed, and popularised, by Robert Tibshirani [2].
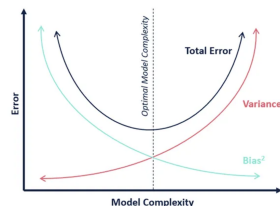

Figure 2: A visualisation of the ideal balance between bias and variance that LASSO delivers with L1 regularisation.

LASSO Implementation

- **Model Motivation** - Predict the Percentage change in price (open/close) for the iShares MSCI Japan ETF (EWJ), traded on the NYSE. With non-overlapping market hours between the NYSE & TSE, the valuation of this ETF is more complex and so sparked some curiosity. The high dimensionality of our data makes it perfect for the use of LASSO regression, which is suited to deal with large numbers of parameters. The automatic variable selection of LASSO will help reduce over fitting and the penalty terms provide a qualitative interpretation of our parameters contribution to the dependent variable.

- **Execution and Wrangling** - Importing, wrangling, and regularising our data, making it suitable for applying LASSO regression. Sourced from Yahoo Finance, we pulled 84 metrics from 478 trading days to use in training our model, summing to just over **40,000 total observations**. We used macro indicators such as the US Treasury yield and the US-JPN exchange rate, the financial metrics of the top ten holdings within the ETF, market index price data, and various metrics of volatility and market momentum. Our data was then engineered to ensure it was suitable to work with. This included time lagging in order to account for historic data and time zone differences, splitting into training/testing datasets, and handling any empty entries. After all of this the implementation into our model also required us to include an imputation step to account for n/a values. To carry these tasks out we used functions and frameworks from Sklearn, NumPy and pandas.

- **Performance and Choice of Hyperparameter** - The below figure shows how our models predictions change as we change our λ, overall we can see our model is fairly accurate in its prediction.:
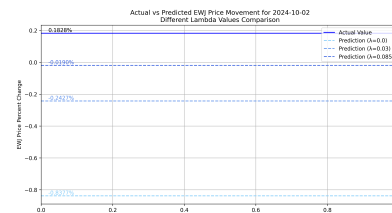

Figure 3:Actual vs Predicted ETF Price Movement for 2024-10-02, comparing different choices for λ.

The different choices of λ above demonstrates how model accuracy differs with its value. As discussed, increasing λ reduces the number of parameters in our model. For $\lambda = 0$ all 84 parameters were retained, leaving us with the β_{OLS}. At $\lambda = 0.03$, the model is reduced to fourteen parameters and for $\lambda = 0.085$, our model is reduced to a single parameter, the difference between the daily high and low of the Nikkei 225, which is given a negative coefficient. This suggests that large movements in this index, are correlated with a mean reversion in our ETF, i.e. price falling. To find the optimal value of λ one would implement a technique known as K-Fold Cross-Validation, of which information is below.

K-Fold Cross-Validation

A resampling procedure that involves randomly shuffling up our dataset and splitting it into k number of equal sized groups, which are then used as training data for the model. The optimal lambda can be found by finding the lambda that produces the best model for each group (lowest MSE) and then finding the average. This process balances the tradeoff between bias and variance, graphed in the bottom left.

Feature Selection with LASSO Regression compared to Ridge Regression

The most important difference between LASSO and other systems of Regression is that LASSO is able to reduce parameter weights to 0, on the other hand ridge regression is only able to reduce their weight to near zero. This can be shown below in Figure 2, and explained further after that.
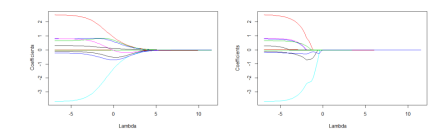

Figure 4: A comparison of the behaviour's of biases as λ varies, between Ridge Regression and LASSO

Geometrically, this is due the diamond shape produced by the constraint region of the penalty terms, which has corners on each axis line. The weight coefficients are found where the contours of the RSS first collide with the constraint region, which can be one of the axis lines in LASSO. However, due to the circular shape of the constraint in ridge regression this is impossible, so a weight coefficient can never be 0. Algebraically we can show that for Ridge Regression when we solve for our expression is asymptotic to 0 as $\lambda \rightarrow \infty$, however for brevity we leave this out. This means LASSO is much more useful when given a large amount of parameters that are known to be irrelevant to the data.
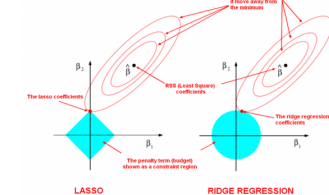

Figure 5: LASSO vs Ridge Regression geometric RSS and constraint region comparison

However, each have their benefits and drawbacks. LASSO can require considerable computational power, Ridge is extremely useful when we have a set of parameters we know are relevant. LASSO is also less reliable in some instances than Ridge Regression as it has to be optimised iteratively, meaning different optimisation techniques may produce different models.

References

[1] Santosa, Fadil, and William W. Symses. 1986. 'Linear Inversion of Band-Limited Reflection Seismograms.' *Inverse Problems 2, no. 1 (1986): 1-8.*

[2] Tilshirani, Robert. 1996. 'Regression Shrinkage and Selection via the Lasso.' *Journal of the Royal Statistical Society. Series B (Methodological) 58, no. 1 (1996): 267–88.*

[3] IBM Resources, https://www.ibm.com/topics/lasso-regression, accessed 20 November 2024.

[4] LASSO Regression, https://www.statisticshowto.com/lasso-regression/, accessed 20 November 2024.

[5] Supervised Statistical Learning Using LASSO Regression, https://rforhr.com/lassoregression.html, accessed 23 November 2024.

[6] Linear Regression, https://kitchingroup.cheme.cmu.edu/f19-06623/18-linear-regression.html, accessed 23 November 2024.