Gil Graybill – BAN 525
Assignment 2 – Penalized Regression and Predicting Silver Prices
Professor Cetin Ciner
May 24, 2020

Introduction

In the movies, pirates are always searching for buried treasure of gold and silver. In the Olympics, a silver medal is second place to gold, which isn't too shabby. In "Rudolph the Red-Nosed Reindeer", Burl Ives gives silver the top billing in "Silver and Gold". Gold and silver seem to be linked together in our culture. Do the same variables predict both gold and silver prices?

More than half of the annual demand for silver are industrial uses, whereas for gold this number is 10-15% (https://www.bullionvault.com/silver-guide/silver-industrial-demand) . Silver is used for batteries, photography, medicine, solar cells, and a host of other uses. Would it be possible to create a statistical model that could predict or describe changes in the price of silver? Are there variables from currency, bond, and stock markets that can help determine the price of silver? Could the overall economy of a given country affect the price of silver? Is it correlated with gold?

The dataset we are working with consists of 10 years of weekly prices for 6 different currency exchange rates, 7 different interest rates for bonds of various terms, 12 stock market indices, and 4 miscellaneous measures, including gold (See Appendix A). Each of these variables has been normalized with a percentage change from week to week. In addition, we have created a "lag" variable so we can compare last week's changes for a given variable against this week's change in silver prices. In all, there are 58 different variables we will be testing to see if they have an influence on the price of silver.

We will be using 6 types of analysis on this dataset:

(1) OLS (Ordinary Least Squares): This type of linear regression model calculates the best fit through the data points that minimizes the squared differences between the observed values and the fitted values. All variables end up in the final equation.
(2) Lasso (Least Absolute Shrinkage and Selection Operator) is a linear regression that aims to shrink data values towards a central point by penalizing them, with the hope of eliminating variables that don't contribute a lot to the model, thereby removing unneeded detail from the model.
(3) Adaptive Lasso runs OLS behind the scenes first to get an idea of what variables appear to be the most important. Those variables are penalized less, giving them a better chance of making it into the model when Lasso is run.
(4) Adaptive Lasso Cauchy is a linear regression that behaves well even when the errors are normally distributed. Most regression models assume the errors are normally distributed, but this regression looks for a Cauchy distribution.

(5) Elastic Net combines penalties obtained from Lasso and Ridge to generate more zero valued coefficients. The square of the coefficient is used. The result can be a model where highly correlated predictors are all included in the model.
(6) Adaptive Elastic Net runs OLS first to favor more likely predictors, hence penalizing the weak performers more.

Analysis and Model Comparison

If the relationship between the response variable and the potential dependent variables is linear, OLS regression will help find those relationships. One disadvantage for OLS is that if there are many variables that are highly correlated, OLS may model random errors. OLS tries so hard to include everything that the model will become inaccurate. We will be using this as our "simple" benchmark analysis.

Penalized regressions add a penalty function to the coefficient estimates, and how that penalty function is defined determine the type of regression analysis. With Lasso regression the coefficients (weights) are penalized by using the *absolute values of the weights*. Lasso performs variable shrinking and selection at the same time. A problem with Lasso is that when there are highly correlated variables, Lasso will just pick one of them.

With Ridge regression (which we are not directly using for this assignment) the coefficients are penalized using the *square of the weights*. While it may seem that the square of the weights is bigger than the absolute value, it is actually smaller because normalization of the data makes the coefficients less than 1.

Elastic Net uses both Lasso and Ridge to determine the penalty function to capture anything that lands "in the net". With the analogy, the small weights (big penalties) go through the net, leaving only the larger ones behind, even if they are highly correlated. The advantage/disadvantage is that more variables may be included in the model, which could lead to either a more precise model or overfitting. Another possible disadvantage is more computation time, which could be an issue for large datasets.

Use of the “Adaptive” option jumpstarts the variable shrinking process by running OLS regression first and penalizing those that appear to have no predictive value. The hope is that the OLS regression was not incorrectly minimalizing those variables.

Using the “Cauchy” option with Adaptive Lasso will combat the effects of outliers, if any exist. The Cauchy distribution resembles a normal distribution, but it has a taller peak and fatter tails that decay slower. If the data has a Cauchy distribution, using this option will yield a better model.

For cross-validation we employed a 60-20-20 split for training-validation-test split. This gave us 309 rows of training data, 103 rows of validation data, and 103 rows of test data. Because this is time series data, we did not split up the data randomly, but instead used the first 60% of the data (time-wise) for training, the next 20% for validation, and the final 20% for testing. This makes sense since we are trying to build a model that could potentially be used to predict future silver prices. With time series data there is always the risk that an event or phenomena will alter the data from a certain date forward.

The price of silver was chosen as the response variable, and all of the normalized variables and their lags were tested as dependent variables using JMP Pro 15. The calculated value for the response variable, the change in the price of silver, was determined and recorded and the resulting formulas determined using all 6 methods.

To compare results between the models, we calculate the difference between each <u>observed</u> value for the change in the price of silver and the <u>calculated</u> value for the change in the price of silver. These differences can be compared in different ways. “RSquare” is a percentage that tells how much of the variability is explained by the model. “RASE” is “Root Average Standard Error”, and because this describes values of error, smaller is better. “AAE” is “Absolute Average Error” is an error term that uses the absolute value of the error to compare the size of the error without regard for direction.

When comparing the models we will be using our cross-validation split. Because we are looking for a predictor of future silver prices we will put more emphasis on the

RSquare/RASE/AAE values for the test data (later chronologically) than the results for the training and validation. The results are as follows:

Cross-Validation	Predictor	RSquare	RASE	AAE
Training	OLS	0.5098	0.0326	0.0242
Training	Lasso	0.3427	0.0377	0.0264
Training	Adaptive Lasso	0.3111	0.0386	0.0273
Training	Adaptive Lasso Cauchy	0.2896	0.0392	0.0268
Training	Elastic Net	0.3426	0.0377	0.0264
Training	Adaptive Elastic Net	0.3107	0.0386	0.0273
Validation	OLS	-0.188	0.0369	0.0298
Validation	Lasso	0.1285	0.0316	0.0238
Validation	Adaptive Lasso	0.1423	0.0314	0.0242
Validation	Adaptive Lasso Cauchy	0.2503	0.0294	0.0227
Validation	Elastic Net	0.1287	0.0316	0.0238
Validation	Adaptive Elastic Net	0.1426	0.0314	0.0242
Test	OLS	-0.465	0.0297	0.0234
Test	Lasso	***0.4085***	***0.0189***	***0.0148***
Test	Adaptive Lasso	0.2783	0.0209	0.0164
Test	Adaptive Lasso Cauchy	0.3337	0.0201	0.0156
Test	Elastic Net	***0.4085***	***0.0189***	***0.0148***
Test	Adaptive Elastic Net	0.2779	0.0209	0.0164

While the Training data yielded a 51% value for RSquare, it looks like a lot of that was due to overfitting since this model was deemed almost worthless with a negative value (indicating the regression line is worse than the mean…ouch) for the Validation and Test data. Lasso and Elastic Net actually tied with RSquare of 0.4085, RASE of 0.0189, and AAE of 0.0148. The use of the Adaptive option did not help, probably because the OLS model that it's based on was so bad. Because Lasso is easier to compute, we'll choose that one as our winning model. The value of RSquare being 41% means that 41% of the variance in the data can be explained by the model, which is not bad.

Interpretation

The formula that was determined using Lasso is:

Predicted silver price = (-0.000294466466159384) + 0.0684078484214725 * RFXS + 0.293780750863955 * RFXA + 0.281222228335729 * RFXC + 0.293845831475944 * RFXF + 0.176033181611256 * RUSO + 0.590468922558416 * RTIP + 0.253715849089776 * RXLB + 0.0341556415487817 * LRFXY

All chosen variables affect the price of silver in a positive way. When running variable importance with these variables against this model, we see the following:

Variable	Main Effect	Total Effect
RFXF	0.356	0.368
RUSO	0.217	0.229
RXLB	0.192	0.205
RTIP	0.076	0.088
RFXA	0.06	0.072
RFXC	0.021	0.031

RFXF is the Swiss Franc, which accounts for 37% of the effect on the model. When investors want stability, they look to Swiss Francs. Because silver is used both as money and as an industrial base metal, it's not surprising that stability and silver go hand-in-hand.

RUSO represents the price of oil, which accounts for 23% of the effect on the model. Oil and silver are both needed by industry and manufacturing, so it makes sense that their fortunes can rise and fall together.

RXLB is a stock index representing the materials sector and it accounts for 21% of the effect of the model. Because silver's use is so prevalent in materials and industry, it reasonable to assume there is a correlation there.

RTIP represents inflation (8% of effect), indicating that silver is used as a hedge against inflation. RFXA represents Australian dollars (7% effect), which is not surprising since Australia (https://en.wikipedia.org/wiki/List_of_countries_by_silver_production) is the world's 8th largest silver producing country. RFXC is Canadian dollars (3% effect), and Canada usually tracks well with commodities.

What's not in the model? RGLD, or Gold. As much as our culture ties gold and silver together, there does not appear to be a strong correlation between their prices. This confirms the work of C. Ciner (2001 Global Finance Journal 12). If there ever was a price relationship between the two before the discovery of silver's industrial, it is certainly gone now.

Appendix A: Normalized Variables used in analysis

FXB: British pound
FXS: Swedish krona
FXY: Japanese yen
FXA: Australian dollar
FXC: Canadian dollar
FXF: Swiss franc
LQD- Investment grade corporate bonds
SHY- 1 to 3 year US treasury bonds
EMB- International (Emerging market) bonds
TLT- Long term (20+) US treasury bonds
IEI- 3 to 7 year US treasury bonds
TLH- 10 to 20 year US treasury bonds
HYG- High yield (risky) US corporate bonds
SPY: S&P 500 (large company) index
SLY: S&P 600 (small company) index
XLB: Material sector index
XLE: Energy sector index
XLF: Financial services sector index
XLI: Industrials sector index
XLK: Technology sector index
XLP: Consumer staples sector index
XLU: Utilities sector index
XLV: Healthcare sector index
XLY: Consumer discretionary sector index
IYR: Real estate sector index
USO: oil prices
VIX: stock market volatility
TIP: inflation measure
GLD: Gold prices