Final Exam Solutions, 2018

December 16, 2018

Q1.

(a)

μ_i is a firm fixed effect. We include fixed effects to remove the “between” variation (between firms) and focus on “within” variation.

(b)

$E_t\left[rv_{i,t+1} + rv_{i,t+2}\right] = 2\mu_i + \mu_i\delta_1 + \delta_1 rv_{i,t} + \delta_1^2 rv_{i,t}$

(c)

(i) and (ii)

Our signal is the spread: $E_t\left[rv_{i,t+1} + rv_{i,t+2}\right] - iv_{i,t}^2$. I call the first term RV and the second term IV for simplicity. IV and RV are not our signal individually. Option prices increase in the volatility of the underlying. Therefore, if $IV > RV$, the option is overpriced, i.e. we would like to short the option. Conversely, if $IV < RV$, we would like to go long in the option.

Each month we compute the spread and form portfolios based on the magnitude of the spread.

Grading note: I gave people who used the Fama-MacBeth credit. However, you have to think about the ability of an economic agent (not an econometrician) to implement the strategy in real time. Additionally, you must have returns on the LHS if you are planning on running FMB.

(iii) We are looking for **economic value**, not statistical significance. There are a large number of criteria one can use to determine if the strategy is economically valuable:

- Sharpe ratio
- α with respect to an asset pricing model (you have to think carefully about the type of model that would be relevant)

- Performance compared to some benchmark (e.g. a long only option portfolio)

Q2.

(a)

$$min_\beta \sum_{t=1}^{T} (y_{t+1} - \beta x_t)^2$$

Subject to:

$$\beta^2 \leq B$$

(b)

Form the Lagrangian:

$$\mathfrak{L} = (Y - X^{'}\beta)(Y - X^{'}\beta) + \lambda(B - \beta^2)$$

Simply take the FOC w.r.t. β and obtaint the following expression for β:

$$\beta = (X^{'}X + \lambda I)^{-1}X^{'}Y$$

I prefer using linear algebra but the sum notation is also completely acceptable.

(c)

In a ridge regression setting, we interested in obtaining a value for the λ parameter one of the approaches that provides us with an "optimal" value for $\lambda$is K-fold cross validation. Here is a quick description of how the procedure works:

We divide the sample into K subsamples and run a ridge regression separately for each group of K-1 folds. We use the Kth fold as an out-of-sample (OOS) test. Based on the OOS tests, we select the λ that minimizes the value of the MSE (you can also use SSE or any other appropriate decision criterion).

Cross-validation affects the β parameter through the fact that the procedure allows to choose the "optimal" λ.

(d)

Note that out prior $\bar{\beta} = 0$.

The OLS β:

$$\beta^{OLS} \sim N\left(\hat{\beta}, \left(X'X\right)^{-1} \sigma^2\right)$$

Using Bayes' rule (or the formula worked out by Professor Lochstoer in T4):

$$\beta^{posterior} \sim N\left(\frac{\left((X'X)^{-1/2}\sigma\right)^{-2}}{\left((X'X)^{-1/2}\sigma\right)^{-2} + A^{-1}}\hat{\beta}, \left(\left((X'X)^{-1/2}\sigma\right)^{-2} + A^{-1}\right)^{-1}\right)$$

(e)

All you need to do is to set the "budget" $B = 1$. The FOC is unchanged but the λ will change, which will affect the β.

(f)

LASSO uses an absolute value constraint with the following objective function:

$$min_\beta \sum_{t=1}^{T}(y_{t+1} - \beta x_t)^2$$

Subject to:

$$|\beta| \leq B$$

The LASSO constraint produces constraints that make corner solutions with zero coefficients more likely compared to a ridge regression (the likelihood of a zero coefficient with ridge is 0).

LASSO is better if we want to reduce dimensionality. Ridge is better if there is high correlation between the dependent variables.

Q3.

(a)

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 rv_{i,t-1} + \beta_2 bm_{i,t} + \sum_{j=1}^{N}\beta_j \mathbb{I}\left\{industry_k = j\right\}$$

Where $p = Prob(rv_{i,t} > 0.01)$

Note that you must have some kind of probability on the LHS.

(b)

Define a binary variable $y_{i,t} = \begin{cases} 1 & if\, rv_{i,t} > 0.01 \\ 0 & otherwise \end{cases}$

Based on this definition the likelihood function is (note that we have two dimensions, which most people ignored)

$$L(\beta) = \prod_{i=1}^{N} \prod_{t=1}^{T} p^{y_{i,t}} (1-p)^{1-y_{i,t}}$$

The **log likelihood** function is:

$$l(\beta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \{y_{i,t} log(p) + (1-y_{i,t}) log(1-p)\}$$

(c)

Definitions:

M_s: saturated model ($L = 1$), M_c: candidate (our) model, M_n: null model (all β, except for β_0 are set to 0)

Based on these definitons:

$$Null\ deviance = 2\left(logL(M_s) - logL(M_n)\right)$$

$$residual\ deviance = 2\left(logL(M_s) - logL(M_c)\right)$$

(d)

Constructing the curve:

Step 1: Set cutoffs for the series of predictions and the realized values.

Step 2: For each cutoff compute the frequency of the TPR (true positive rate) and the FPR (false positive rate) and use the two values to obtain a data point on the ROC curve

Step 3: Obtain such points for all cutoffs. These points trace out the ROC curve.

AUC is an acronym for "area under the curve" and is a single number summarizing the forecasting ability of our model. A model generating higher AUC has better forecasting ability. The AUC for the random guess model is 0.5, which is represented by the 45° line.

Q4.

(a)

Fee$< 0.9 \Rightarrow \hat{\alpha} = 0.8$

Fee$\geq 0.9 \Rightarrow \hat{\alpha} = -1.5$

(b), (c), (d)

See the 2017 final's solution on CCLE.