Using R for Econometrics and Statistics
Take Home Exam

The due date is 2019-11-19

For the exam, you are required to submit a Rmarkdown file and a html file generated from your Rmarkdown file.

1. See "rates.doc" for a description of the data file. For all questions, use 1962: 1 through 2012: 6 as the sample period. Use the first 24 observations (1960 : 1 through 1961 : 12) for initial conditions and differencing transformations. You are to calculate the following. You should write your own code (use R), but can borrow from pre-existing code where you feel comfortable doing so.

   (a) Start by plotting the unemployment rate against time. Is the series trending? Cyclical?

   (b) Estimate an AR(4) model (always include an intercept!) by least-squares. Report coefficient estimates, robust standard errors, and a one-step point forecast for July 2012.

   (c) Estimate a set of autoregressions (always include an intercept!) by least-squares, AR (1) through AR (24). For each, calculate the Cross Validation information criterion. Also calculate the BIC, AIC, AIC^c, Mallows, Robust Mallows information criteria. Create a table for your results.

   (d) Based on the CV criteria, select an AR model.

   (e) Use this model to make a one-step point forecast for July 2012 .

   (f) Report coefficient estimates and robust standard errors.

   (g) Now consider the other variables in the data set. After making suitable transformations, include these variables in your model. Using the information criteria, select a forecasting model.

   (h) Use this forecasting model to make a one-step point forecast for July 2012 .

2. In this problem you'll use ridge regression the lasso to estimate the salary of various baseball players based on a bunch of predictor measurements. This data set is taken from the "ISLR" package, and R package that accompanies the "Introduction to Statistical Learning" textbook. You should now have the objects x, y, the former being a 263×20matrix of predictor variables, and the latter a 263 dimensional vector of salaries. (For more information, download and install the ISLR package and type

?Hitters.) Download and install the g1mnet package from the CRAN repository. We'll be using this package to perform ridge regression and the lasso. Finally, define

```
grid =10^seq(10,-2,length=100)
```

This is a large grid of λ values, and we'll eventually instruct the glmnet function to compute the ridge and lasso estimates at each one of these values of λ.

(a) The ‘glmnet‘ function, by default, internally scales the predictor variables so that they will have standard deviation 1, before solving the ridge regression or lasso problems. This is a result of its default setting ‘standardize=TRUE‘. Explain why such scaling is appropriate in our particular application.

(b) Run the following command

```
rid.mod = glmnet(x,y,lambda=grid,alpha=0)
las.mod = glmnet(x,y,lambda=grid,alpha=1)
```

This fits ridge regression and lasso estimates, over the whole sequence of λ values specified by grid. The flag "alpha=0" notifies g1mnet to perform ridge regression, and "alpha=1" notifies it to perform lasso regression. Verify that, for each model, as λ decreases, the value of the penalty term only increases. That is, for the ridge regression model, the squared ℓ_2 norm of the coefficients only gets bigger as λ decreases. And for the lasso model, the ℓ_1 norm of the coefficients only gets bigger as λ decreases. You should do this by producing a plot of λ (on the x-axis) versus) versus the penalty (on the y-axis) for each method. The plot should be on a log-log scale.

(c) Verify that, for a very small value of λ, both the ridge regression and lasso estimates are very close to the least squares estimate. Also verify that, for a very large value of λ, both the ridge regression and lasso estimates approach 0 in all components (except the intercept, which is not penalized by default).

(d) For each of the ridge regression and lasso models, perform 5 -fold cross-validation to determine the best value of λ. Report the results from both the usual rule, and the one standard error rule for choosing λ. You can either perform this cross-validation procedure manually, or use the "cv.glmnet" function. Either way, produce a plot of the cross-validation error curve as a function of λ, for both the ridge and lasso models.

(e) From the last part, you should have computed 4 values of the tuning parameter:

$$\lambda_{\text{min}}^{\text{ridge}}, \lambda_{\text{1se}}^{\text{ridge}}, \lambda_{\text{min}}^{\text{lasso}}, \lambda_{\text{1se}}^{\text{lasso}}$$

These are the results of running 5-fold cross-validation on each of the ridge and lasso models, and using the usual rule (min) or the one standard error rule (1se) to select λ. Now, using the predict function, with type: "coef", and the ridge and lasso models fit in part (b), report the coefficient estimates at the appropriate values of λ. That is, you will report two coefficient vectors coming from ridge regression with $\lambda = \lambda_{\text{min}}^{\text{rige}}$ and $\lambda = \lambda_{\text{1se}}^{\text{rige}}$, and likewise for the lasso. How do the coefficient estimates from the usual rule compare to those from the one standard error rule? How do the ridge estimates compare to those from the lasso?

(f) Suppose that you were coaching a young baseball player who wanted to strike it rich in the major leagues. What handful of attributes would you tell this player to focus on? (That is, how to measure variable importance?)

3. Value at Risk (VaR) is a statistical measure of downsiden current position, It estimates how much a set of investments might lose given normal market conditions in a set time period. A vaR statistic has three components.al time period b) confidence level. c) loss ammount (or loss percentage). For 95% confidence level, we can say that the worst daily loss will not exceed VaR estimation. If we use historical data, we can estimate vaR by the quantile value. For our data this estimation is:

```
quantile(sp500ret , 0.05)
```

Delta-normal approach assumes that all stock returns are normally distributed. This method consists of going back in time and computing the variance of returns. Value at Risk can be defined as:

$$VaR(\alpha) = \mu + \sigma * N^{-1}(\alpha)$$

where μ is the mean stock return, σ is the standard deviation of returns, α is the selected confidence level and N^{-1} is the inverse CDF function, generating the corresponding quantile of a normal distribution given α.

The results of such a simple model often disappointing and are rarely used in practice today. The assumption of normality and constant daily variance is usually wrong and that is the case for data as well.

Previously we observed that returns exhibit time-varying volatility. Hence for the estimation of VaR we use the conditional variance given by GARCH model. For the underlined asset's distribution properties we can use the student't-distribution. For this method Value at Risk is expressed as:

$$\mathrm{VaR}_t(\alpha) = \mu + \hat{\sigma}_{t|t-1} * F^{-1}(\alpha)$$

where $\hat{\sigma}_{t|t-1}$ is the conditional standard deviation given the information at $t-1$ and F^{-1} is the inverse CDF function of $t-$distribution.

(a) For the "sp500price" data set we have used in class, you need to specify an appropriate form of GARCH model for the return of 'sp500' with information criterion method and rolling estimation method. For rolling method, the window size is 2500 and the estimation is implemented every 100 observations (To save the computation cost, moving window method is recommended.)

(b) Forward looking risk management uses the predicted quantiles from the GARCH estimation. Using " ugarchroll" method to estimate the best GARCH model you have obtained from (a) by rolling method to compute $\mathrm{VaR}_t(\alpha)$ for $\alpha = 0.05$. The setting of rolling estimation is the same as (a).

(c) A VaR exceedance occurs when the actual return is less than the predicted value-at-risk: $R_t < VaR_t$. Plot the scattered points of actual returns, the predicted $\mathrm{VaR}_t(\alpha)$, and highlight the VaR exceedance points by different color.

(d) The frequency of VaR exceedances is called the VaR coverage. A valid prediction model has a coverage that is close to the probability level α used. If coverage $> \alpha$: too many exceedances: the predicted quantile should be more negative. Risk of losing money has been underestimated.If coverage $< \alpha$: too few exceedances, the predicted quantile was too negative. Risk of losing money has been overestimated. Compute the VaR coverage of the AR(1)-GJR-GARCH with skew-t distribution, AR(1)-GJR-GARCH with t distribution, AR(1)-GARCH with t distribution, and AR(1)-GJR-GARCH with skew-t distribution with rolling estimation implemented every 1000 observations instead of 100. You can create a table to display your results.