Empirical exercise 14.1 requires more programming than the other empirical exercises in the textbook.

The empirical analysis reported in the body of Chapter 14 was carried out using Matlab. You can find all of the Matlab programs we used in the replication files for the chapter. If you are familiar with Matlab, you can modify these programs to answer EE 14.1.

Stata version 16 introduced a program (*elasticnet*) for easily carrying out Ridge and Lasso regressions like those discussed in the text. (*elasticnet* uses a parameter $\alpha$, where $\alpha = 0$ produces Ridge and $\alpha = 1$ produces Lasso.) In the replication files for Chapter 14, you will also find Stata programs that carry out nearly all of the calculations reported in Chapter 14. If you are familiar with Stata, you can modify these programs to answer EE 14.1. (For the exercise, the Stata file **ch14_large.do** is the most relevant of the chapter 14 replication files.)

A few comments:

(1) Stata, Matlab, and the SW textbook use somewhat different normalizations for Ridge and Lasso, and this results in different scaling from the $\lambda$ values (see equations (14.7) and (14.9) in SW). To keep this straight, let $\lambda_{Ridge}^{SW}$ and $\lambda_{Ridge}^{SW}$ denote the Stock-Watson of the Ridge and Lasso parameters and $(\lambda_{Ridge}^{Stata}, \lambda_{Lasso}^{Stata})$ and $(\lambda_{Ridge}^{Matlab}, \lambda_{Lasso}^{Matlab})$ denote the corresponding values in Stata and Matlab. Letting $SSR = \sum_{i=1}^{n}(Y_i - b_1 X_{1i} - ... - b_k X_{ki})^2$, the table below shows the objective functions and the relationship between the $\lambda$ values.

| | Obective Function | $\lambda$ relationship |
|---|---|---|
| Ridge | | |
| Stock-Watson | $SSR + \lambda_{Ridge}^{SW} \sum_{i=1}^{n} b_i^2$ | |
| Stata | $\dfrac{1}{n} SSR + \lambda_{Ridge}^{Stata} \sum_{i=1}^{n} b_i^2$ | $\lambda_{Ridge}^{Stata} = \lambda_{Ridge}^{SW} / n$ |
| Matlab | $SSR + \lambda_{Ridge}^{Matlab} \sum_{i=1}^{n} b_i^2$ | $\lambda_{Ridge}^{Matlab} = \lambda_{Ridge}^{SW}$ |
| Lasso | | |
| Stock-Watson | $SSR + \lambda_{Lasso}^{SW} \sum_{i=1}^{n} |b_i|$ | |
| Stata | $\dfrac{1}{2n} SSR + \lambda_{Lasso}^{Stata} \sum_{i=1}^{n} |b_i|$ | $\lambda_{Lasso}^{Stata} = \lambda_{Lasso}^{SW} / (2n)$ |
| Matlab | $\dfrac{1}{2n} SSR + \lambda_{Lasso}^{Matlab} \sum_{i=1}^{n} |b_i|$ | $\lambda_{Lasso}^{Matlab} = \lambda_{Lasso}^{SW} / (2n)$ |

(2) Stata's *elasticnet* uses $k$-fold cross-validation to estimate $\lambda$ (as described in the SW text). No extra programming is required.

Finally, why it is not a programming note, we note one modification to EE 14.1. Question (a) should be:

(a)  From the 20 primitive predictors, construct squares of all the predictors, along with all of the interactions (that is, the cross products $X_{ji}X_{ki}$ for all $j$ and $k$ ). Collect the 20 primitive predictors, their squares, and all interactions into a set of $k$ predictors. Verify that you have 20 + 20 + (19×20)/2 = 230 predictors. One of the is the binary variable *charter_s*. Drop the predictor (*charter_s*)$^2$ from the list of 230 predictors. Also drop the interaction (*str_s* × *te_fte_s*) from the list of predictors, so you are now left with 228 predictors. Why should (*charter_s*)$^2$ be dropped from the original list of predictors? Why should (*str_s* × *te_fte_s*) also be dropped?