Prof. James MacKinnon   *QUEEN'S UNIVERSITY*   *Department of Economics*

Economics 882   Winter, 2025

Assignment 3

Due: March 20, 2025

This assignment uses a partly real and partly simulated dataset called `wealth.csv`. The variables are `wealth` (in dollars), `age` (in years), `educ` (five levels), and `marry` (binary; 1 if married). There are 40,000 observations.

1. Using the entire sample estimate a regression tree to explain `wealth` using all the inputs. Then create the log of wealth, say `logwlth`, and estimate a tree to explain it. Plot both trees.
[15]

2. Randomly split the sample into a training sample with 70% of the observations and a test sample with 30%. Do this in such a way that you obtain the same split every time. The rest of the models you estimate should use the training sample for estimation and the test sample for prediction.
[5]

3. Using the training sample, estimate a linear regression model to explain `logwlth` that includes a spline on `age`. Calculate a measure of how well this model performs on the test sample.
[10]

4. Using the training sample, estimate at least two generalized additive models for `logwlth`, where `age` is treated as continuous and education is treated as discrete. Calculate a measure of how well these models perform on the test sample.
[15]

5. Using the training sample, estimate at least two random forest models to explain `logwlth`. Different models might, for example, use different values of `mtry`, the number of variables to split on, or different numbers of trees. Check the package documentation. Calculate measures of how well your random forests perform on the test sample.
[20]

6. Using the training sample, estimate at least three gradient boosted tree models to explain `logwlth`. These should vary in what you think are important tuning parameters. Calculate measures of how well these models perform on the test sample.
[20]

7. What do you conclude? What model(s) seem to work best? Have you learned anything interesting about the relationship between wealth, age, marital status, and education?
[15]