Practical Computing for Economists

Weeks 2 & 3: A Rapid Introduction To R For Economists

Time and Location:

Harper Center, Weeks 2 & 3

Lecturer:

John Eric Humphries johnerichumphries@gmail.com

Objectives

Weeks 2 & 3 of the course will provide a rapid introduction to the R environment for statistical computing.

By the end of these two sessions, we hope to equip you with the skills to:

- Read and understand R code, understanding how R "thinks".
- Write clear, documented, reusable R code.
- Load and explore real-world data.
- Visualize and summarize results
- Use packaged functions and libraries to do work for you.
- Implement your own econometric methods to solve problems others have not already solved for you.

- Call an optimizer to maximize a function.
- Gaussian Quadrature for numerical integration (if time allows).
- Basic parallel computing (if time allows).

To accomplish the list of objectives above, we will need to review the basics quickly. If you are not familiar with R basics, I highly recommend spending a couple hours completing the online exercises available at www.codeschool.com/courses/try-r or www.datacamp.com. For additional review material, see the references at the end of this document.

Layout

The two sessions will consist of live coded examples as well as board work. The two sessions will be divided into roughly four parts:

1. Brief review of R basics (30-60 minutes)
2. Regression and simulation in R – an applied example (2-2.5 hours)
3. Working with data and producing output in R – a continuation of the applied example (1.5 - 2 hours)
4. Extensions and leveraging computational tools (as remaining time permits)

Material

Section 2.1 We will first provide a rapid overview of R. This will include a review of linear algebra in R, control-flow in R, idiosyncrasies of R, and types in R. This section will be brief, so please review the many great resources available online for R basics (see references in the next section).

Section 2.2 In the second section we will review many tools in R through an application on regression. We will simulate a synthetic data set and learn to call R's default prepackaged regression tools. Next we will program our own ordinary least squares. We will first do this with simple linear algebra. Next we will solve the same problem by numerically

minimizing the least-square problem. Finally, we will consider an implementation of Bayesian Regression. We will consider how to package these three methods into a single callable function, and then bootstrap to produce standard errors.

Section 3.1 In the third section, we will work with ACS data to practice reading-in, exploring, summarizing, and analyzing real data. To begin, we consider simple regression in R, constructing summary tables directly to latex from R, constructing regression tables directly to latex from R, merging data in R, and plotting results.

Section 3.2 As time permits, we will expand the complexity of our analysis in one of two ways: (1) A maximum likelihood example of selection correction or (2) Expanding our regression code above to account for an unobserved factor. These are both empirically more demanding and would allow us to introduce simple parallel computing or numerical integration. The goal would be to estimate on simulated data, but then to be able to apply it to the data from the section 3.

Preparation and Resources

Please install Rstudio prior to arriving. A list of packages to pre-install and instructions on how to do so will be sent out prior to the second session. Here are a list of useful R resources:

- www.codeschool.com/courses/try-r
- www.datacamp.com
- http://cran.r-project.org/doc/contrib/Farnsworth-EconometricsInR.pdf (slightly out of date, but a short useful guide
- http://link.springer.com/book/10.1007/978-0-387-77318-6 (also slightly out of date, free to download on the campus network)
- http://files.itslearning.com/data/ku/103018/teaching/lecturenotes.pdf