

Introduction to HPC

HPC can be well suited for some scientific workloads.
Overview + hands on session specifically for R

- Monte Carlo
- Rolling Window

Prerequisites

HPC account + Project

- Various computational projects available - MonARCH is specifically built for Monash staff/students

ssh client

- putty (Windows), terminal (macOS or Linux)

ftp client to access HPC files

- MobeXterm (Windows), Cyberduck (macOS), Nautilus/Dolphin (Linux)

Login Node Setup

- ① Login to MonARCH
 - ▶ `ssh user@monarch.erc.monash.edu`
- ② Load R modules
 - ▶ `module load R`
- ③ Setup folder for R libraries
 - ▶ `asd`
- ④ Install R libraries
 - ▶ `R`
 - ▶ `install.packages("tidyverse")`

Code Setup

Recommended to set up git repository containing all code

- `cd project/user`
- `git clone`

Job Script (Bash)

Parallelization + Splitting Workloads

HPC can be well suited for parallelized/split workloads.
This is done via the `ARRAY` environment variable

Monte Carlo Example

IV Regression Example:

$$y = \beta_0 + x_1\beta_1 + e_1, \quad e_1 \sim N(0, 1) \quad (1.1)$$

$$x_1 = \gamma z + e_2, \quad e_2 \sim N(0, 1) \quad (1.2)$$

$$\text{cor}(e_1, e_2) = \rho \quad (1.3)$$

Interested in performance of IV estimator $\widehat{\beta}_1$ across different

- Instrument strength $\gamma \in \{0, 0.25, 0.5\}$
- Endogeneity $\rho \in \{0, 0.25, 0.5\}$
- Sample size $N \in \{100, 200, 500\}$

Code

Practical Tips

Do not request too many resources - this can take a long time to be allocated

Extra Resources

Data Fluency Workshops (free for students!):

- Introduction to Bash/Shell Scripts
- Introduction to HPC

Advanced Issues

rcpp

- Compiling rcpp code on one array and asking other arrays to use this is inconsistent - no guarantee that different arrays are of same architecture
- Solution: ask explicitly for same compute instance nodes OR compile code for each array (inefficient, but not usually not prohibitively so)