Next: Purpose of this document

SAS for Finance PhDs

André de Souza

- Purpose of this document
- Where to get help for SAS problems
- Some preliminaries you should know
  - SAS basics
  - The logic of the DATA step
    - The OUTPUT statement
    - The counter _N_
  - Some CRSP variable names
  - Macro basics

- Move to SAS on Unix
  - Cygwin and X-Windows

- Reading in datasets
  - Quick and dirty: convert proc import log output to an input statement
  - guessingrows
  - Why does proc import hang on Unix?

- Working with WRDS datasets
  - CRSP: DSF/MSF
  - Names files in CRSP: stocknames and fundnames
  - CRSP: ncusip and cusip
  - Adjusting shares in holdings data
  - Merging CRSP and Compustat

- Working with time-series data and dates
  - Moving dates around
  - Avoiding the lag() function
  - Rolling regressions without macros

- Missing values
- Code debugging
  - Run on a small subset of observations, making sure your code doesn't overwrite anything
  - See what the macro sees
  - Set msglevel=i
  - Print variables to the log within the data step
  - Using the DEBUG option

- Macros
  - Macro variables
    - Subsetting observations from large datasets efficiently
  - Loops
    - Counters are dates
    - %FOR variable in &LIST
    - Subsetting observations from large datasets efficiently
  - Writing frequently-used code as macros
    - Merging macros
    - Hash-merging without having to type
    - Checking for duplicate by-variable combinations
    - Sundry macros
    - Regressions

- The DOW loop
- Merges
  - Merging a dataset with a single observation into another dataset
  - Merging without sorting: the DATA step HASH object
  - One-to-many with the DOW loop
  - Many to many
    - Using proc sql
    - Using hashes with multidata
  - Post-merge processing: merge and then calculate summary statistics

- Statistics
  - Weighted means with negative weights
  - Robust standard errors
  - Fixed effects
    - Brute-force
    - An efficient way with one-way fixed effects
    - Two-way fixed effects
  - Pairwise statistics with large datasets

- Efficiency
  - Reading into the PDV: where, keep, and in
  - Never use proc sql
  - Don't sort more than strictly necessary
  - Macro loops make things possible, they don't make them fast

- SAS on Unix
  - Transferring files from DOS: line endings
  - Locating your work directory
  - Relocating your work directory
  - Relocating your log and lst files
  - Memory issues
  - Why does proc import hang on Unix?
  - Searching for files on Unix
  - Utilities

        - Parsing the SAS log: logsasparse
        - Looking at SAS datasets: sasprint
        - Killing all sas processes: saskill
    - The autoexec.sas file
    - The .sasv9.cfg file
    - SAS with X-Windows

- Further interesting topics
    - What really happens in the merge statement
    - The fascinating Do-loop

- Bibliography
- About this document ...

Andre de Souza 2012-11-19