Big Data Asset Pricing

Lecture 3: Working with Big Asset Pricing Data
Theis Ingerslev Jensen & Marc Eskildsen

For this lecture, you should watch the videos listed below. The first video section is a quick introduction to accessing the WRDS cloud via SAS Studio. The next videos show different ways of working with asset pricing data, centering on the code and data developed in the paper Is there a Replication Crisis in Finance by Jensen, Kelly, and Pedersen (henceforth, JKP).

The videos were made prior to the stock-level data being uploaded to WRDS, meaning it is now significantly easier to download. As a result, the steps described in video 2 are no longer the easiest way to get the data, but you should still familiarize yourself with the code. The latest data can be downloaded directly using the WRDS interface at:

https://wrds-www.wharton.upenn.edu/pages/get-data/contributed-data-forms/global-factor-data

Alternatively, you can download the data programmatically, and we made a guide on how to do so in R and Python:

- How to download the JKP data from R
- How to download the JKP data from Python

On the BigFi Datahub, you can similarly find a Python tutorial with video explanations. This site also features a list of external data sources that you may find helpful in generating future research ideas:

- BigFi DataHub: How to download the JKP data from Python (with video explanations)
- BigFi DataHub: External data sources

Videos: Introduction to the WRDS cloud (made by WRDS) – 15 minutes in total

1. https://wrds-www.wharton.upenn.edu/pages/grid-items/using-sasstudio-part-1/
2. https://wrds-www.wharton.upenn.edu/pages/grid-items/sas-studio-part-2/
3. https://wrds-www.wharton.upenn.edu/pages/grid-items/sas-studio-part-3/
4. https://wrds-www.wharton.upenn.edu/pages/grid-items/sas-studio-part-4/

Videos: Working with Asset Pricing Data – a total of 120 minutes

1. Overview of JKP code and data:
   https://cbs.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=a8977cdd-42b3-4520-a65e-ae4e00d748c0

2. How to generate global stock returns and firm characteristics
   https://cbs.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=9bf2fddd-9df0-49bd-9dcc-ae4d00e5774b

3. Return Data from CRSP
   https://cbs.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=c2e9fb3b-edcc-4e15-be27-

[ae4b00db86e3](#)

4. Accounting Data from Compustat
   https://cbs.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=0c88d49d-df62-4121-b80b-ae4b00f559b6

   The SAS script discussed in the video is available here:
   https://www.dropbox.com/sh/xkcm2whxe6yih5x/AACYVxFMOypFHDiq78wwdzk4a?dl=0

5. WRDS Overview
   https://cbs.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=cf2d6ed5-86d9-4b0f-b02e-ae4d00d67f95

   The table discussed in the video is copied below.

6. Working with JKP Factors
   https://cbs.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=93025220-4dda-40e4-889a-ae4c00daef7f

   The R script discussed in the video is available here:
   https://www.dropbox.com/sh/xkcm2whxe6yih5x/AACYVxFMOypFHDiq78wwdzk4a?dl=0

7. Working with Stock-Level Global Data
   https://cbs.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=cda4fce8-4870-4a6d-b91d-ae4d01098e6b

   The R script discussed in the video is available here:
   https://www.dropbox.com/sh/xkcm2whxe6yih5x/AACYVxFMOypFHDiq78wwdzk4a?dl=0

Useful Links

- Different ways to access the data from WRDS
  https://wrds-www.wharton.upenn.edu/pages/about/3-ways-use-wrds/

- WRDS cloud via. SAS studio
  https://wrds-cloud.wharton.upenn.edu/SASStudio/index

- "Is there a Replication Crisis in Finance" by Jensen, Kelly, and Pedersen (2022)
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3774514

- Github repo for Jensen, Kelly, and Pedersen (2022)
  https://github.com/bkelly-lab/ReplicationCrisis

- Global factor returns
  https://JKPfactors.com/

Table from Video 5 ("WRDS Overview")

The table below lists some of the essential data tables in the CRSP and Compustat databases hosted by WRDS and accessible via https://wrds-cloud.wharton.upenn.edu/SASStudio/index. The first column (name) shows the table name on the WRDS servers, the next column (description) has a short description; the last column (variables) has a link with a description of the variables in the table. If you need to know the variables in other tables, follow this link: https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/ and press "data directory" for the relevant data source. CRSP is under "Center for Research in Security Prices, LLC (CRSP)" and Compustat is under "S&P Global Market Intelligence."

Name	Description	Variables
CRSP.MSF	Monthly price data for US securities	https://wrds-www.wharton.upenn.edu/data-dictionary/crsp_a_stock/msf/
CRSP.DSF	Daily price data for US securities	https://wrds-www.wharton.upenn.edu/data-dictionary/crsp_a_stock/dsf/
CRSP.CCM_LNKHIST	Historical links between CRSP and Compustat	https://wrds-www.wharton.upenn.edu/data-dictionary/crsp_a_ccm/ccmxpf_lnkhist/
CRSP.MSENAMES	Security information	https://wrds-www.wharton.upenn.edu/data-dictionary/crsp_a_stock/msenames/
CRSP.DSENAMES	Security information	https://wrds-www.wharton.upenn.edu/data-dictionary/crsp_a_stock/dsenames/
CRSP.MSEDELIST	Monthly delisting returns (includes partial monthly return)[1]	https://wrds-www.wharton.upenn.edu/data-dictionary/crsp_a_stock/msedelist/
CRSP.DSEDELIST	Daily delisting returns	https://wrds-www.wharton.upenn.edu/data-dictionary/crsp_a_stock/dsedelist/
COMP.FUNDA	Annual accounting data for North American firms	https://wrds-www.wharton.upenn.edu/data-dictionary/comp_na_daily_all/funda/
COMP.G_FUNDA	Annual accounting data for firms outside of North America.	https://wrds-www.wharton.upenn.edu/data-dictionary/comp_global_daily/g_funda/
COMP.FUNDQ	Quarterly accounting data for North American firms	https://wrds-www.wharton.upenn.edu/data-dictionary/comp_na_daily_all/fundq/

---

[1] See WRDS presentation for the difference between MSEDELIST and DSEDELIST: https://wrds-www.wharton.upenn.edu/documents/1413/CRSP_useful_variables.pptx

COMP.G_FUNDQ	Quarterly accounting data for firms outside of North America.	https://wrds-www.wharton.upenn.edu/data-dictionary/comp_global/g_funda/
COMP.SECD	Daily price data for stocks listed in North America.	https://wrds-www.wharton.upenn.edu/data-dictionary/comp_na_daily_all/secd/
COMP.G_SECD	Daily price data for stocks listed outside of North America.	https://wrds-www.wharton.upenn.edu/data-dictionary/comp_global/g_secd/
COMP.SECM	Monthly price data for North American firms	https://wrds-www.wharton.upenn.edu/data-dictionary/comp_na_daily_all/secm/
COMP.SECURITY	Security information for stocks listed in North America	https://wrds-www.wharton.upenn.edu/data-dictionary/comp_na_daily_all/security/
COMP.G_SECURITY	Security information for stocks listed outside of North America	https://wrds-www.wharton.upenn.edu/data-dictionary/comp_global/g_security/
COMP.R_EX_CODES	Reference tables for exchanges codes in Compustat.	https://wrds-www.wharton.upenn.edu/data-dictionary/comp_na_daily_all/r_ex_codes/
COMP.EXCHRT_DLY	Daily exchange rate data	https://wrds-www.wharton.upenn.edu/data-dictionary/comp_na_daily_all/exrt_dly/