Kieschnick, Robert

From: Venugopalan, Rudra Vignesh
Sent: Monday, December 30, 2024 11:04 AM
To: Kieschnick, Robert
Subject: Project Status
Attachments: Fama_French.ipynb; WRDS_DB.ipynb

Dear Dr Kieschnick,

I hope this message finds you well.

I have worked on saving the CRSP and Compustat data in a SQLite database, which I have stored locally. The size of the SQLite database file is over 4.5 GB, and I'll try uploading it to the folder you shared with me.

The data spans from January 1, 1960, to December 16, 2024 and can be updated to the latest date by simply changing the end date to the current date in the provided Python code blocks before running them.
A new database automatically gets created if the name doesn't already exist in your local system.

I used my WRDS login credentials to pull the data. If someone else wants to use the attached code block, they must use their own WRDS credentials.

The database was built using SQLite in Google Colab and if someone wants to modify or build upon this database using Jupyter Notebook, they may need to adjust the syntax slightly.

The final database contains the following tables:

- factors_ff3_monthly: Monthly Fama-French 3-factor data.
- factors_ff5_monthly: Monthly Fama-French 5-factor data.
- factors_ff3_daily: Daily Fama-French 3-factor data.
- industries_ff_monthly: Industry portfolio returns (monthly).
- factors_q_monthly: Monthly Q-factor data.
- macro_predictors: Macroeconomic predictors.
- cpi_monthly: Consumer Price Index data (normalized).
- crsp_daily: CRSP daily stock-level data.
- compustat: Compustat financial statement data.
- crsp_monthly: CRSP monthly stock-level data.

Queries can be built on top of these.

I followed the steps described in the Tidy Finance book to construct this database. Since macroeconomic data is used in the analysis, I have included Fama-French factors and other related data

in the database. The Python notebook named "Fama_French" must be run first to create the database. Subsequently, WRDS data gets added sequentially through the notebook "WRDS_DB".

I am currently working on an ML model to predict prices using the data in this database. I will share the details and code once the model is completed.

Thanks and have a happy new year,
Rudra