

How to Update the Equity Risk Premium Model

Eric Huang and (more recently) Charlie Smith

February 7, 2020

1 ERP Background

The equity risk premium (ERP) is an important aspect in asset pricing. It is the difference between the expected returns on equities and the risk free rate. We estimate the ERP by combining data from twenty models that have been proposed in the literature. Check out `2015_EPR_equity-risk-premium.pdf` for an overview of this project.

2 Repository Location

The bare repository for the Equity Risk Premium (ERP) Model is located in `\\ranhpc01-smb.ny.frbres.org\data\cmwbb\GitBareERP\ERP_Article.git`. Clone the repository to your RAN directory (rce) and make all changes there. When naming the local Git repository cloned from the remote `EPR_article.git`, please follow the proper program naming routine (no spaces, no rare characters, etc.) so that the program does not crash due to invalid directory.

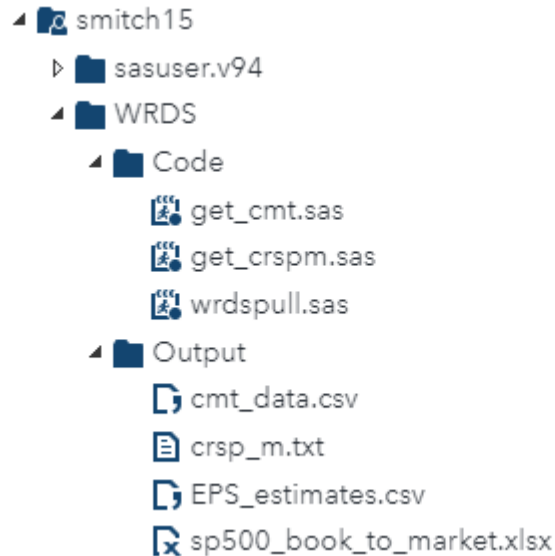
3 Running ERP

3.1 Tools Required

1. Anaconda for Python 2.7, available [here](#).
2. Matlab 2019b, with the Statistics and Machine Learning Toolbox.
3. Stata version 16 or greater.
4. WRDS account, available [here](#).
5. For FredFetch, you may need an api key, which is stored in the readme folder under `api.txt`. The FredFetch toolbox is saved in the `code/Matlabsubfunctions` folder. To get a new FRED api key, create an account with FRED and then get an API key [here](#).
6. This project is run on the RAN cluster with 25 GB of memory.

3.2 Before Updating

1. Change the variable “root_dir” in the `run_all.m` to the location of the directory.
2. After creating a WRDS account, go to the WRDS web portal [here](#), and create a folder structure that matches the following image:



3.
 - (a) Inside your folder in WRDS, you need to create a folder called “WRDS” and within that folder two folders, one called “Code” and one called “Output.”
 - (b) After doing so, go to `Code\project_scripts\download_data.m` and update the variables “wrds_code_dir” and “wrds_out_dir” to match the location of those two folders. In addition, change “wrds_username” and “wrds_password” to match your username and password.
 - (c) When you eventually run `Code\project_scripts\download_data.m`, it puts the .sas files from `Code\wrds_scripts` into the code folder, runs the sas files in your WRDS folder, and then downloads the data.
4. Update the files that require manual updating, specifically `input\cfo_data.csv` and `input\tb1m.csv`.
 - (a) `tb1m.csv` - is a HAVER download that is transferred from the SAN to the RAN. The code to download this data is located in `code\san_scripts\haver_data.do`. Run this do file in Stata locally on the SAN. Then load up WINSXP on your computer, navigate to the folder where the data was downloaded on the SAN and copy that data to `\Input` in ERP.

- (b) `cfo_data.csv` - requires manually updating the data using the CFO survey website, located [here](#). Go to the current year and quarter and select US high level results. Then find the part in the pdf where they report the expected S&P500 growth over the next 1 and 10 years. The precise questions are: “Over the next 10 years, I expect the average annual S&P 500 return will be:” and “Over the next year, I expect the average annual S&P 500 return will be:.” Add the mean prediction and the number of predictors to the `cfo_data.csv` file for those two questions.
- 5. In `Code\project_scripts\cay.m` the url in line 35 should be updated. To do so, go to the website [here](#). Select the most recent version and then right click on section two. Select “copy link address” and then paste that in place of the url in line 35.

3.3 Running ERP

1. Open `run_all.m` in the root folder of the ERP repository and run the file.
2. Check to make sure the output seems reasonable (compare to previous vintages).
3. If everything works, make a commit.

3.4 Areas where things break

1. Shiller data has broken (code stops running) due to changes in the websites data format. If this occurs, the code will break in `NY_Fed_CM_DAPM.m`.
2. The WRDS download can sometimes break when downloading `crsp_m.txt`. I put it in a try catch command to keep the code running. If it is not downloading, you can download it manually To do so:
 - (a) Log into WRDS SAS Studio.
 - (b) Find the file in your folder structure.
 - (c) Right click on the file and download it.
 - (d) Transfer it to the `\Input`.

4 Project/Folder Outline

4.1 `run_all.m`

This file runs the entire ERP project. Ideally, we will reach the situation where all that needs to happen is click run on this file. It takes around 20 minutes to run in Matlab2019b interactive session, with `Code\project_scripts\download_data.m` occupying around 15 minutes.

4.2 Code

This folder contains all the code to run the project. The code is sorted into a variety of folders. The scripts that are called by `run_all.m` are stored in `code\project_scripts`. Functions are stored in three separate folders, specifically `code\functions`, `code\Matlabsubfunctions`, and `code\NYFed_functions`. There are a few python (stata) scripts that are called by the .m files in `code\project_scripts`, that are stored in `code\python_scripts` (`code\stata_scripts`). Finally, the scripts that are run in WRDS are stored in `code\wrds_scripts`. `run_all.m` calls files from `code\project_scripts`, which use functions from the various function folders and calls stata, python, and wrds.sas files occasionally.

4.3 Input

This folder contains all the raw data that is needed to run the ERP project.

4.4 Temp

This folder contains data at various stages of the project. Each folder maps to a script in the `code\project_scripts` folder. Each folder contains an input folder and an output folder. The input is the data that enters that script and the output is the data produced by the script.

4.5 Output

This folder contains all the output. The most important files are:

1. `Original_data.csv` contains information on the input data, its sources, and the dates it spans. Use this to identify which datasets are updating and which are not.
2. `Constructed_variables.csv` contains information on the variables that this project constructs. Similarly, use this to identify which variables are updating and which are not.
3. `ERP_1yr_ahead.csv` contains time series on the 20 models this project computes and uses. Similarly, use this to identify which models are updating and which are not.
4. `ERP_principal_components.csv` contains the ERP estimates at different time horizons. This is the main output.

4.6 BW

This folder, located in the code folder, includes data and files necessary to replicate the Baker Wurgler investor sentiment measure. I have almost replicated the measure, we are just missing two data-series necessary to construct it, specifically the dividend premium and the closed end fund discount. The code is

written in such a way that once are those completed, it just has to be included in to the `collect_data.m` script where bw data is currently accessed.

In addition, two of the data-series for this measure necessitate downloading them on the san. They are the issuance of equity and debt. To do so, run `code\san_scripts\issuance_download.m` on the SAN and then transfer the two csvs to the RAN input folder.

You can check their website [here](#) to see if they have updated their data. The most recent version runs until December 2018.

5 State of Project February 2020

All data in `Output\original_data` is updated until late 2019, except for sentiment, equity issuance, and debt issuance. Those variables are all part of Baker Wurgler.

In `Output\Constructed_variables` the same holds true. Here all variables are updated until at least September 2019, except the Baker Wurgler variables listed above.

In `Output\ERP_1yr_ahead` all models are updated until September 2019 except, Model 8, Model 9 and 10 (which are annual), Model 18, Model 19 (which is the Baker Wurgler model), and model 20 (which is quarterly). These exceptions have all been updated between December 2018 and July 2019.

6 Next Steps

1. Replicate BW and include automatically updating BW in the project.
2. Automate `tblm.csv`
3. Automate `cfo_data.csv` but most likely is not worth it.