

# FIN 971: Corporate finance

## Problem set 1

Due: Friday, Feb 4

This problem set asks you to dive into the Compustat/CRSP merged dataset available on WRDS. If you have not created a WRDS account, goto <https://wrds-web.wharton.upenn.edu/wrds/> and click on "Register." This will allow you to access the Compustat, North America, Monthly Updates database through their web interface.

For this problem set, we will be focusing on the following paper, which we denote as LRZ:

Lemmon, M. L., Roberts, M. R., Zender, J. F., 2008. Back to the beginning: persistence and the cross-section of corporate capital structure. *Journal of Finance* 63, 1575–1608.

This has become an important paper in the capital structure literature because the authors document significant heterogeneity in capital structure across firms that cannot be explained by easily observable characteristics, and that this unexplained heterogeneity is highly persistent.

Note that there is a data appendix in the paper that has variable definitions; use this as a guide. The definitions include Compustat data item numbers (legacy FTP style). These item numbers were deprecated several years back, and replaced with text variable names (XPF, or Xpressfeed), but many older papers reference the FTP names (e.g. **DATA41** instead of **COGS**). There is documentation on WRDS mapping to the two formats.<sup>1</sup>

For those that have not used Stata or Compustat/CRSP before, be patient and try to team up with someone who has some familiarity. You are encouraged to work together. For newbies, the commands **reg** and **areg** with the **absorb** option will be useful. Category

---

<sup>1</sup>See <https://wrds-www.wharton.upenn.edu/pages/support/manuals-and-overviews/compustat/north-america-global-bank/comparison-legacy-ftp-format-xpressfeed/> which has a description and naming information to translate between XPF and FTP.

dummy variables can be easily constructed using `tabulate varname, generate(dummy_name)`. The user-supplied function `reghdfe` is also useful when running regressions with more than one fixed effect. User-supplied functions can be installed using `ssc install` and `net install`, in this case `ssc install reghdfe`.

I encourage you to use L<sup>A</sup>T<sub>E</sub>X for writing up your results. The Stata packages `outreg2` (or `estout`) are helpful tools to export regressions into readable tables (Excel, L<sup>A</sup>T<sub>E</sub>X, etc), a useful practice for both during the exploratory phase of research and in producing a written draft.

While LRZ uses only Compustat, for this exercise use both Compustat and CRSP data because it's useful to become familiar with both at the same time. CRSP maintains a link between the two databases called the Compustat/CRSP merged (CCM) database which can be accessed through WRDS→Annual Update→CRSP/Compustat Merged. From there, you can download Compustat data with the correct link to the CRSP database, in particular the CRSP security identifier (`permno`), the date range for which this is valid (`linkdt`, `linkenddt`), and the link type (`linktype`). The CCM database does not include CRSP price data; you'll need to perform that merge yourself, as described next.

Once you've downloaded the CCM dataset with CRSP links, you'll then download the relevant CRSP data. For the exercise below, I recommend using CRSP monthly for ease of use, but daily can be used in a similar way. Merging CRSP with CCM is requires some date manipulation: CCM is annual, while CRSP is monthly. Also be aware of the difference between a company's fiscal year (`fyear`), and the calendar close of the fiscal year, or report date (`datadate`). It will also require some filtering to ensure that the link is only used during its valid period (i.e., between `linkdt` and `linkenddt`). Also, you will only want to use links with `linktype` ∈ {"LU", "LC"} and `linkprim` == "C". Note that while Compustat coverage includes North American companies, CRSP only includes stocks publicly traded on a major US exchange.

You'll want to apply this filter to Compustat data:

```
keep if indfmt=="INDL" & datafmt=="STD" & popsrc=="D" & consol=="C"
```

Please do the following analysis in Stata:

1. Replicate Table I of LRZ.
2. Replicate Table II (both Panels A and B) of LRZ.
3. Rerun Panel A of Table II where initial leverage is no longer included, but add in firm fixed effects. How do the coefficient estimates change?
4. LRZ uses Compustat data to construct the market value of equity (ME). Now construct ME using CRSP data on price and shares outstanding. Are they identical? Compare the two variables. If there are differences for some firms, see if you can figure out why for one or two specific examples.
5. Using CRSP data, construct a measure of each firm  $i$ 's equity volatility as the standard deviation of monthly returns over the last three years; call this  $\sigma_{E,i,t}$ . Unlever this quantity as follows:

$$\sigma_{A,i,t} = \frac{ME_{i,t}}{ME_{i,t} + DT_{i,t}} \sigma_{E,i,t} \quad (1)$$

where  $ME_{i,t}$  is market value of equity and  $DT_{i,t}$  is total debt as defined by LRZ. Include  $\sigma_{A,i,t}$  as a lagged regressor in Table II, Panel A. How does this coefficient estimate compare to that of cash flow volatility? What different information about the firm is being captured by cash flow volatility and asset volatility? Finally, are you concerned that the coefficient estimate on  $\sigma_{A,i,t}$  might be biased? Why or why not?