

Example SAS Program to Extract Variables from COMPUSTAT WRDS Files

Ofira Schwartz-Soicher/Todd Hines

oschwart@princeton.edu/thines@princeton.edu

September, 2014

Notes

- The example program below may be used to extract variables from COMPUSTAT or other data files on WRDS.
- You may copy the code into SAS. Please be sure to modify the program (e.g., library names, variable names, GVKEY's) according to your needs.
- Text in green in the document below indicates comments and will not be executed by SAS.
- This program is based upon a sample code provided by WRDS :
(<http://wrds.wharton.upenn.edu/>)

```
/* tell SAS where to store the datasets on your computer  
(This should be modified according to your folder structure) */
```

```
libname myLib "H:\SAS_for_WRDS ";
```

```
/* details about the PC-connect (this is always the same), it will trigger a login  
prompt. Use your WRDS user name and password to log on*/
```

```
%let wrds = wrds.wharton.upenn.edu 4016;  
options comamid = TCP remote=WRDS;  
signon username=_prompt_;
```

```
/* everything after 'rsubmit' will be executed on WRDS!! */
```

```
rsubmit;
```

/* SQL (Structured Query Language) is a language to process data that is supported by SAS

"create table example as "

-> create a new table/data file called 'example' in the default library 'work'

"select a.* from compm.funda a"

-> select all records from compm.funda (the 'a' behind compm.funda means that a will refer to this dataset. a.* therefore selects all variables

"(keep = gvkey datadate cusip ni seq sich fyear)"

only the listed variables are included in the newly created table. Please modify the variables list according to your interests/needs.

"where gvkey in ('011636', '014489', '011456', '008606', '008479', '006266', '009411', '005047', '001447', '001690', '002285', '002663', '003144', '003170', '003735', '064768')"

-> keep only records with the listed gvkey's. GVKEY has to be 6 characters long (add as many leading zeros as necessary). GVKEY's list could be modified according to your interests/needs.

"and 2006 <= fyear <= 2008"

-> keeping only record from fiscal years 2006 through 2008. Remove this line if you are interested in all the years of data.

"and indfmt='INDL' and datafmt='STD' and popsrc='D' and consol='C'"

-> these are additional requirements that need to be included to exclude doubles*/

/* data is organized in libraries, locally, as well as on WRDS tell SAS to look in '/wrds/comp/sasdata/nam' when you use 'compm' as the library name This should be modifies if your are interested in extracting variables from a different data file */

```
libname compm '/wrds/comp/sasdata/nam';
```

PROC SQL;

```
create table example (keep = gvkey datadate cusip ni seq sich fyear) as  
select a.*  
from compm.funda a  
where gvkey in ('011636', '014489', '011456', '008606', '008479', '006266',  
'009411', '005047', '001447', '001690', '002285', '002663', '003144', '003170',  
'003735', '064768')  
and 2006 <= fyear <= 2008  
and indfmt='INDL' and datafmt='STD' and popsrc='D' and consol='C' ;  
quit;
```

```
/* The dataset is created on the WRDS server and needs to be downloaded in
order to access it locally; download the dataset work.example from the server to
your computer */
```

```
proc download data= example out=myLib.fund_exmp;
run;
```

```
/* stop processing statements on WRDS and continue working locally */
endrsubmit;
```

```
/* creating frequency tables to check the data file */
```

```
proc freq data=myLib.fund_exmp;
  tables gvkey datadate cusip ni seq sich fyear;
run;
```

```
/* examining the data as a table (similar to excel spreadsheet) */
```

```
proc print data=myLib.fund_exmp;
  var gvkey datadate cusip ni seq sich fyear;
run;
```

/* exporting the new SAS data file into .csv file so that it could be read in excel.
Please modify the folder and file names accordingly */

proc export

DATA=myLib.fund_exmp

OUTFILE="H:\SAS_for_WRDS \fund_exmp.csv"

DBMS=csv REPLACE;

run;