

Introduction to Financial Databases

Sep 17, 2007

By Roger Loh & Taylor Nadauld
Department of Finance
Fisher College of Business
Ohio State University

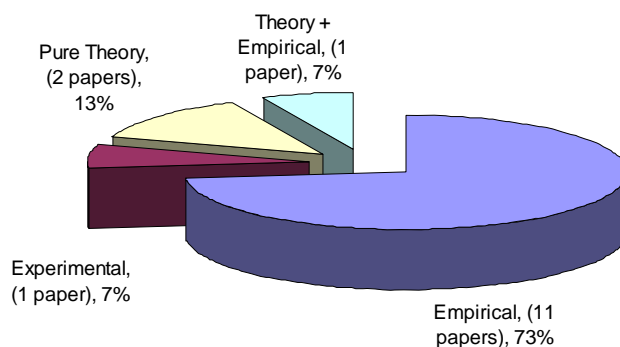
Database website http://fisher.osu.edu/~loh_26/database/

1

Importance of Data in Financial Research

Hard to write a paper that doesn't use data.

- June 2006 issue of Journal of Finance



2

[Databases platforms]

1. WRDS
 - Stock returns (CRSP)
 - Trading data (TAQ)
 - Company data (Compustat, IRRC)
 - Analyst advice (I/B/E/S)
 - Institutional holdings data (Thomson 13F), etc
2. Datastream/Worldscope
 - International stock returns, company data, economic statistics
3. SDC
 - Equity and debt issues
 - Mergers and acquisitions

3

[Outline]

Overview of some WRDS databases (Roger)

Accessing WRDS (Roger)

Datastream/SDC (Taylor)

Practice sessions in Fisher 606

4

1. CRSP

- Center for Research in Security Prices—most comprehensive US stock returns database.
- Individual stock returns and market returns (daily and monthly).
- Most used items:
 - Price
 - Return
 - Volume
 - Market-cap
 - Shares outstanding
 - SIC Industry code

5

CRSP Data E.g.

Permanent number is the CRSP security identifier

PERMNO	DATE	COMNAM	PRC	VOL	RET
11081	20050131	DELL INC	41.7600	2676671	-0.00902
11081	20050228	DELL INC	40.0900	3443626	-0.03999
11081	20050331	DELL INC	38.4200	3153971	-0.04166
11081	20050429	DELL INC	34.8300	3768089	-0.09344
11081	20050531	DELL INC	39.9300	3670710	0.14643
11081	20050630	DELL INC	39.4600	3016043	-0.01177
11081	20050729	DELL INC	40.4700	2130263	0.02560
11081	20050831	DELL INC	35.6000	5044521	-0.12034
11081	20050930	DELL INC	34.2000	3611113	-0.03933
11081	20051031	DELL INC	31.8800	4273538	-0.06784
11081	20051130	DELL INC	30.1510	5894065	-0.05423
11081	20051230	DELL INC	29.9500	4104746	-0.00667
12490	20050131	INTERNATIONAL BUSINESS MACHS COR	93.4200	1148066	-0.05234
12490	20050228	INTERNATIONAL BUSINESS MACHS COR	92.5800	802581	-0.00706
12490	20050331	INTERNATIONAL BUSINESS MACHS COR	91.3800	1075088	-0.01296
12490	20050429	INTERNATIONAL BUSINESS MACHS COR	76.3800	2175501	-0.16415
12490	20050531	INTERNATIONAL BUSINESS MACHS COR	75.5500	1389057	-0.00825
12490	20050630	INTERNATIONAL BUSINESS MACHS COR	74.2000	1377932	-0.01787
12490	20050729	INTERNATIONAL BUSINESS MACHS COR	83.4600	1571802	0.12480
12490	20050831	INTERNATIONAL BUSINESS MACHS COR	80.6200	1050638	-0.03163
12490	20050930	INTERNATIONAL BUSINESS MACHS COR	80.2200	1156901	-0.00496
12490	20051031	INTERNATIONAL BUSINESS MACHS COR	81.8800	1413377	0.02069
12490	20051130	INTERNATIONAL BUSINESS MACHS COR	88.9000	1108741	0.08818
12490	20051230	INTERNATIONAL BUSINESS MACHS COR	82.2000	1203340	-0.07537

6

[2. Compustat]

- US Company Data from financial statements
- Most commonly used Files
 - Industrial Annual
 - Industrial Quarterly
 - CRSP-Compustat Merged File
 - Executive Compensation
- Commonly used items in Annual File (Item #):
 - Long-term Debt (#9)
 - Sales (#12)
 - Earnings (income before extraordinary items, #18)
 - Book value of equity (#60)
 - Total Assets
 - R&D expenditure
 - Cash

7

[Compustat Annual Data E.g.]

GVKEY	yeara	SMBL	DATA9	DATA12	DATA18	DATA60
6066	1995	IBM	10060	71940	4178	22170.00
6066	1996	IBM	9872	75947	5429	21375.00
6066	1997	IBM	13696	78508	6093	19564.00
6066	1998	IBM	15508	81667	6328	19186.00
6066	1999	IBM	14124	87548	7712	20264.00
6066	2000	IBM	18371	88396	8093	20377.00
6066	2001	IBM	15963	85866	7723	23614.00
6066	2002	IBM	19986	81186	5334	22782.00
6066	2003	IBM	16986	89131	7613	27864.00
6066	2004	IBM	14828	96293	8448	29747.00
6066	2005	IBM	15425	91134	7994	33098.00
14489	1995	DELL	113	5296	272	973.00
14489	1996	DELL	18	7759	531	806.00
14489	1997	DELL	17	12327	944	1293.00
14489	1998	DELL	512	18243	1460	2321.00
14489	1999	DELL	508	25265	1666	5308.00
14489	2000	DELL	509	31888	2236	5622.00
14489	2001	DELL	520	31168	1246	4694.00
14489	2002	DELL	506	35404	2122	4873.00
14489	2003	DELL	505	41444	2645	6280.00
14489	2004	DELL	634	49205	3043	6485.00
14489	2005	DELL	559	55908	3572	4129.00

8

3. I/B/E/S

- Institutional Brokers Estimate System.
- Contains sell-side security analysts' earnings forecasts and stock recommendations for US firms and international firms.
- Earnings Forecasts
 - Summary File (monthly consensus forecasts)
 - Detail File (Individual analyst forecasts)
- Recommendations
 - Summary File (monthly consensus recommendations)
 - Detail File (Individual analyst recommendations)

I/B/E/S Detail Recommendations E.g.

OFTIC	TICKER	RECDATS	BROKER	AMASKCD	ANALYST	ITEXT
DELL	DELL	20060110	JYSKE	00073425	JACOBSEN	P BUY
DELL	DELL	20060110	FRIEDMAN	00080211	SUMNER	C HOLD
DELL	DELL	20060120	BAIRD	00053931	RENOUARD	D STRONG BUY
DELL	DELL	20060206	CARISCO	00001032	STAHLMAN	M HOLD
DELL	DELL	20060208	BERN	00058377	SACCONAGHI, JR	T STRONG BUY
DELL	DELL	20060217	MONTSEC	00070706	BACHMAN	K HOLD
DELL	DELL	20060219	NUTMEG	00000678	LABE	P HOLD
DELL	DELL	20060307	GRUMMAN	00065515	CHU	R STRONG BUY
DELL	DELL	20060421	SMITH	00047225	GARDNER	R SELL
DELL	DELL	20060430	NUTMEG	00000678	LABE	P BUY
DELL	DELL	20060505	GOLDMAN	00001176	CONIGLIARO	L HOLD
DELL	DELL	20060509	BAIRD	00053931	RENOUARD	D HOLD
DELL	DELL	20060509	DAKIN	00075117	SHAW	C SELL
DELL	DELL	20060519	FBOSTON	00085160	SEMPLE	R HOLD
DELL	DELL	20060519	NEEDHAM	00001047	WOLF	C HOLD
DELL	DELL	20060519	LEHMAN	00001924	BLOUNT	H HOLD
DELL	DELL	20060519	SMITH	00047225	GARDNER	R HOLD
DELL	DELL	20060519	FRIEDMAN	00080211	SUMNER	C STRONG BUY
DELL	DELL	20060522	ARGUS	00000938	ABRAMOWITZ	W HOLD
DELL	DELL	20060526	FGS	00111945	DALAL	N UNDERPERFORM
DELL	DELL	20060530	FIRSTALB	00114594	MAI	H HOLD
IBM	IBM	20060109	JPMORGAN	00072446	SHOPE	B HOLD
IBM	IBM	20060421	FGS	00111945	DALAL	N HOLD
IBM	IBM	20060505	GOLDMAN	00001176	CONIGLIARO	L BUY
IBM	IBM	20060601	CANACCOR	00072029	MISEK	P HOLD

[4. NYSE Trade and Quote (TAQ)]

- Consolidated Trades, Consolidated Quotes data.
- Used when intra-day price and quotes are needed, especially in market microstructure/liquidity research.
- Commonly used variables
 - Trade price
 - Trade size
 - Bid/offer price
 - Bid/offer size
- Managing the large size of the dataset is the challenge of using TAQ data.

[TAQ Quotes Data E.g.]

SYMBOL	DATE	TIME	BID	OFR	BIDSIZ	OFRSIZ
IBM	20050606	9:30:01	0.01	1000.00	1	1
IBM	20050607	9:30:01	0.01	1000.00	1	1
IBM	20050607	9:30:04	71.56	78.95	1	1
IBM	20050607	9:30:04	71.56	78.95	1	1
IBM	20050607	9:30:06	71.56	78.98	1	1
IBM	20050607	9:30:06	71.56	78.98	1	1
IBM	20050607	9:30:08	71.56	79.00	1	1
IBM	20050607	9:30:08	71.56	79.00	1	1
IBM	20050607	9:30:19	75.11	75.48	38	5
IBM	20050607	9:30:19	75.00	75.06	237	21
IBM	20050607	9:30:19	0.00	0.00	0	0
IBM	20050607	9:30:19	75.11	75.48	5	5
IBM	20050607	9:30:19	75.00	75.06	232	21
IBM	20050607	9:30:19	7.50	142.50	1	1
IBM	20050607	9:30:19	74.70	0.00	1	0
IBM	20050607	9:30:19	74.70	75.36	1	1
IBM	20050607	9:30:19	75.11	75.28	5	10
IBM	20050607	9:30:19	75.01	75.36	1	1
IBM	20050607	9:30:19	75.01	75.16	1	1
IBM	20050607	9:30:19	75.11	75.19	5	2
IBM	20050607	9:30:20	0.00	0.00	0	0
IBM	20050607	9:30:20	0.00	0.00	0	0
IBM	20050607	9:30:20	0.00	0.00	0	0
IBM	20050607	9:30:20	74.40	76.01	10	10
IBM	20050607	9:30:20	74.40	76.01	10	10
IBM	20050607	9:30:20	74.81	75.11	13	1
IBM	20050607	9:30:20	74.90	75.16	1	1

5. Thomson 13F

- Institutions are required to report their ownership of equities in quarterly 13F filings to the SEC.
- Aggregate holdings for the institution, regardless of the number of individual fund portfolios.
- Shows how many shares of a firm are held by each institution.
- Usage of this database among published papers increased recently.

13F Holdings Data E.g.

mgrno	rdate	mgrname	cusip	ticker	shares	shroutr
110	30JUN2005	A R ASSET MANAGEMENT, INC.	24702R10	DELL	20250	2421
110	30SEP2005	A R ASSET MANAGEMENT, INC.	24702R10	DELL	20250	2397
110	31DEC2005	A R ASSET MANAGEMENT, INC.	24702R10	DELL	20250	2354
120	31MAR2005	AAL CAPITAL MANAGEMENT CORP.	24702R10	DELL	1442976	2459
120	30JUN2005	AAL CAPITAL MANAGEMENT CORP.	24702R10	DELL	1497626	2421
120	30SEP2005	AAL CAPITAL MANAGEMENT CORP.	24702R10	DELL	1383126	2397
120	31DEC2005	AAL CAPITAL MANAGEMENT CORP.	24702R10	DELL	1103726	2354
185	31MAR2005	ASB CAPITAL MANAGEMENT, INC.	24702R10	DELL	1972001	2459
185	30JUN2005	ASB CAPITAL MANAGEMENT, INC.	24702R10	DELL	1960261	2421
185	30SEP2005	ASB CAPITAL MANAGEMENT, INC.	24702R10	DELL	2191936	2397
185	31DEC2005	ASB CAPITAL MANAGEMENT, INC.	24702R10	DELL	2088638	2354
195	31MAR2005	ABERDEEN ASSET MANAGERS LTD.	24702R10	DELL	377300	2459
195	30JUN2005	ABERDEEN ASSET MANAGERS LTD.	24702R10	DELL	368900	2421
195	30SEP2005	ABERDEEN ASSET MANAGERS LTD.	24702R10	DELL	323400	2397
195	31DEC2005	ABERDEEN ASSET MANAGERS LTD.	24702R10	DELL	323400	2354
205	31MAR2005	ABNER HERRMAN&BROCK ASSET MGMT	24702R10	DELL	103175	2459
205	30JUN2005	ABNER HERRMAN&BROCK ASSET MGMT	24702R10	DELL	94907	2421
220	30JUN2005	ACADIAN ASSET MANAGEMENT, INC.	24702R10	DELL	1200	2421
260	31MAR2005	ADAMS EXPRESS COMPANY	24702R10	DELL	400000	2459
260	30JUN2005	ADAMS EXPRESS COMPANY	24702R10	DELL	400000	2421
260	31DEC2005	ADAMS EXPRESS COMPANY	24702R10	DELL	400000	2354
350	31MAR2005	ADELL HARRIMAN& CARPENTER INC.	24702R10	DELL	86389	2459
350	30JUN2005	ADELL HARRIMAN& CARPENTER INC.	24702R10	DELL	84064	2421
350	30SEP2005	ADELL HARRIMAN& CARPENTER INC.	24702R10	DELL	85589	2397
350	31DEC2005	ADELL HARRIMAN& CARPENTER INC.	24702R10	DELL	93819	2354
440	31MAR2005	ADVANCE CAPITAL MGMT, INC.	24702R10	DELL	19200	2459

[6. IRRC]

- Investor Responsibility Research Center
- Corporate Governance data
 - Gompers, Ishii, Metrick (2003 <http://papers.ssrn.com/id=278920>) corporate governance index. Lists the anti-takeover provisions that a firm has. More provisions=more entrenched management=poor governance. Commonly used for recent corporate governance studies.
- Directors data
 - Information on the directors of a firm—board size, age, whether they are independent, whether they hold shares in the firm.

15

[IRRC Directors Data E.g.]

TICKER	year	DID	CHAIRMAN	CEO	FNAME	LNAME	DIRTYPE	STKHOLDING
DELL	2000	31702	0	0	Michael H.	Jordan	I	0.0
DELL	2000	33232	1	1	Michael S.	Dell	E	12.4
DELL	2000	33233	0	0	Thomas W.	Luce III	L	0.0
DELL	2000	34718	1	1	Alex J.	Mandl	I	0.0
DELL	2000	34778	0	0	Michael A.	Miles	I	0.0
DELL	2000	38267	1	1	Mary Alice	Taylor	I	0.0
DELL	2000	38942	0	0	Klaus S.	Luft	I	0.0
DELL	2000	38943	1	1	Donald J.	Carty	I	0.0
DELL	2000	40255	0	0	Sam	Nunn	I	0.0
DELL	2000	38929	0	0	Morton L.	Topfer	E	0.0
DELL	2001	31702	0	0	Michael H.	Jordan	I	0.0
DELL	2001	32244	0	1	William H.	Gray III	I	0.0
DELL	2001	33232	1	1	Michael S.	Dell	E	12.0
DELL	2001	33233	0	0	Thomas W.	Luce III	L	0.0
DELL	2001	34718	0	0	Alex J.	Mandl	I	0.0
DELL	2001	34778	0	0	Michael A.	Miles	I	0.0
DELL	2001	37262	0	0	Judy C.	Lewent	I	0.0
DELL	2001	38942	0	0	Klaus S.	Luft	I	0.0
DELL	2001	38943	1	1	Donald J.	Carty	I	0.0
DELL	2001	40255	0	0	Sam	Nunn	I	0.0
DELL	2001	38929	0	0	Morton L.	Topfer	E	0.0
DELL	2002	31702	0	0	Michael H.	Jordan	I	0.0
DELL	2002	32244	0	0	William H.	Gray III	I	0.0
DELL	2002	33232	1	1	Michael S.	Dell	E	11.8
DELL	2002	33233	0	0	Thomas W.	Luce III	L	0.0
DELL	2002	34718	0	0	Alex J.	Mandl	I	0.0

16

[Outline]

Overview of some WRDS databases

Accessing WRDS

Datastream/SDC

Practice sessions in Fisher 606

17

[3 ways to access WRDS]

1. Web queries at <http://wrds.wharton.upenn.edu>
2. SAS PC Connect (need SAS installation on your PC)
3. Unix SAS (use SSH Secure Shell, download from <http://osusls.osu.edu/upgrades/stg2wnx.html>)

Write actual programs to download data. More powerful and flexible than web queries. Integrates data extraction within your research programs.

18

SAS PC Connect Vs Unix

SAS PC Connect

- Easier to edit programs and debug.
- No need to learn unix code. More user-friendly but just as powerful as the unix interface.
- Can move across windows to view program, output and log file.
- Jobs that take >1 hour to run require a SSH tunnel connection to WRDS if executed through the university's firewall.

UNIX

- Good for very large jobs (e.g. those that take many hours to execute)
- Can run more than one program concurrently.
- You can access your work anywhere since the data and programs are in the WRDS server.

Steps to use SAS PC Connect

E.g.: Compute average monthly returns of all U.S. stocks 2001-present.

Step 1: Go to <http://wrds.wharton.upenn.edu/support/dslist/dslist.shtml> and find dataset list that you need.

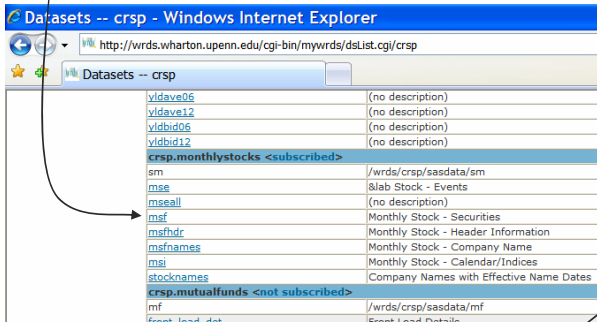
The screenshot shows the WRDS Data Files Index website. At the top, there is a search bar labeled 'datasets' with a dropdown menu. Below the search bar, there are navigation tabs: 'General', 'New User', 'Intermediate', and 'Advanced'. The main content area is titled 'Dataset List and Sample Programs' and contains a table with the following columns: Library, Contents and Subscriptions, WRDS Unix Directory, and Sample Programs. The 'CRSP' dataset is highlighted with a red box in the 'Library' column.

Library	Contents and Subscriptions	WRDS Unix Directory	Sample Programs
Audit Analytics	Dataset List	/wrds/audit	
Bank Regulatory	Dataset List	/wrds/bank	
Blockholders	Dataset List	/wrds/block	
Board Analyst	Dataset List	/wrds/corplib	
Bureau van Dijk	Dataset List	/wrds/bvd	
CBOE Indexes	Dataset List	/wrds/cboe	
CISDM	Dataset List	/wrds/cisdms	
COMPSTAT Global & EMDB	Dataset List	/wrds/emdb	SAS
COMPSTAT North America	Dataset List	/wrds/computat	SAS
ComScore	Dataset List	/wrds/comscore	SAS
CRSP	Dataset List	/wrds/crsp	SAS FORTRAN C
CSMAR	Dataset List	/wrds/csmar	SAS
CUSIP Master File	Dataset List	/wrds/cusip	SAS

Since we want stock returns, the dataset we need is CRSP data. Click on dataset list for list of CRSP datasets.

Steps to use SAS PC Connect

Step 2: Since we need monthly frequency, find CRSP monthly stock returns, and click on msf.



[<< Back to crsp library](#)

Contents of Dataset: crsp.msfc

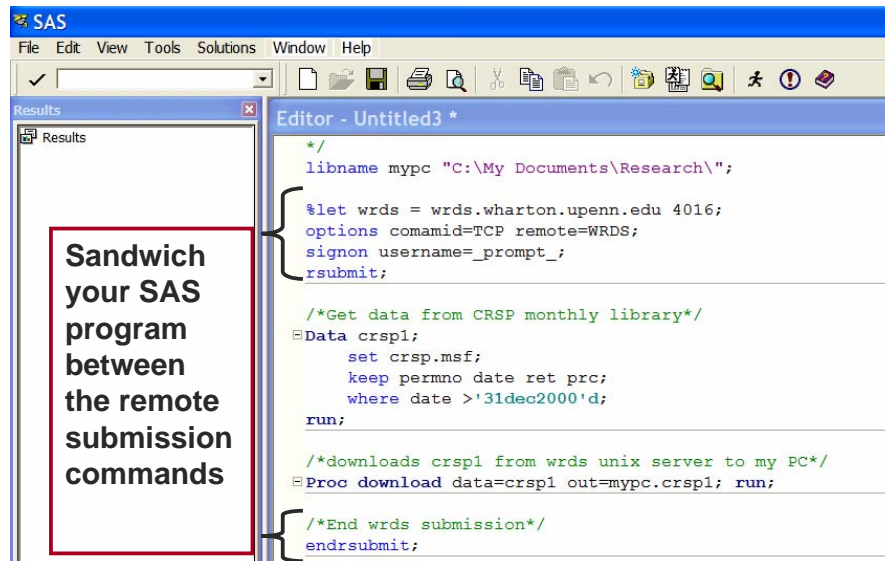
Descriptions:.

Position	Variable	Format	Length	Description
1	CUSIP	Char	8	Header CUSIP
2	PERMNO	Num	8	CRSP Permanent Number
3	PERMCO	Num	8	CRSP Permanent Company Number
4	ISSNO	Num	8	Nasdaq Issue Number
5	HEXCD	Num	8	Header Exchange Code
6	HSICCD	Num	8	Header SIC Code
7	DATE	Num	4	Date of Observation
8	BIDLO	Num	8	Bid or Low
9	ASKHI	Num	8	Ask or High
10	PRC	Num	8	Closing Price or Bid/Ask average
11	VOL	Num	8	Share Volume
12	RET	Num	8	Holding Period Return
13	BID	Num	8	Closing Bid Price
14	ASK	Num	8	Closing Ask Price
15	SHROUT	Num	8	Shares Outstanding
16	CFACPR	Num	8	Cumulative Factor to Adjust Prices
17	CFACSHR	Num	8	Cumulative Factor to Adjust Shares/Vol
18	PRC2	Num	8	Secondary Price
19	RETX	Num	8	Return Without Dividends

Step 3: Jot down library name **crsp.msfc**, and variable names needed: **permno**, **date**, **ret**, **prc**.

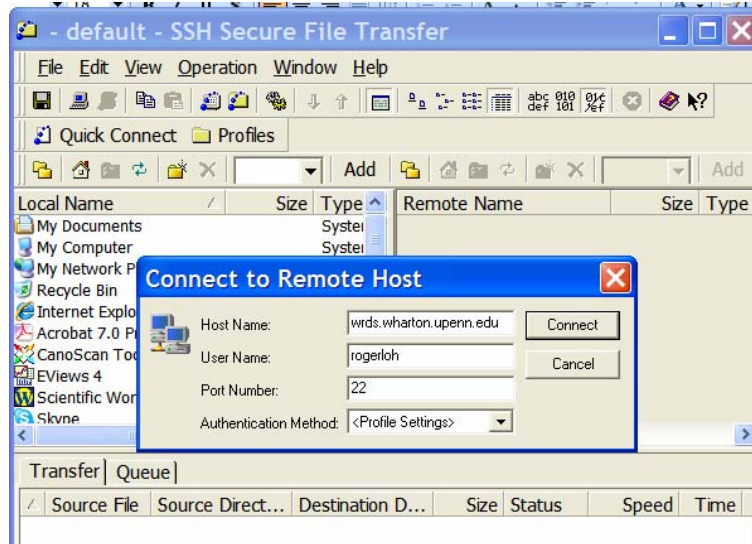
Steps to use SAS PC Connect

Step 4: Write the sas program to compute average returns and sandwich it with the PC connect commands. Then run the program.



Steps to use WRDS Unix

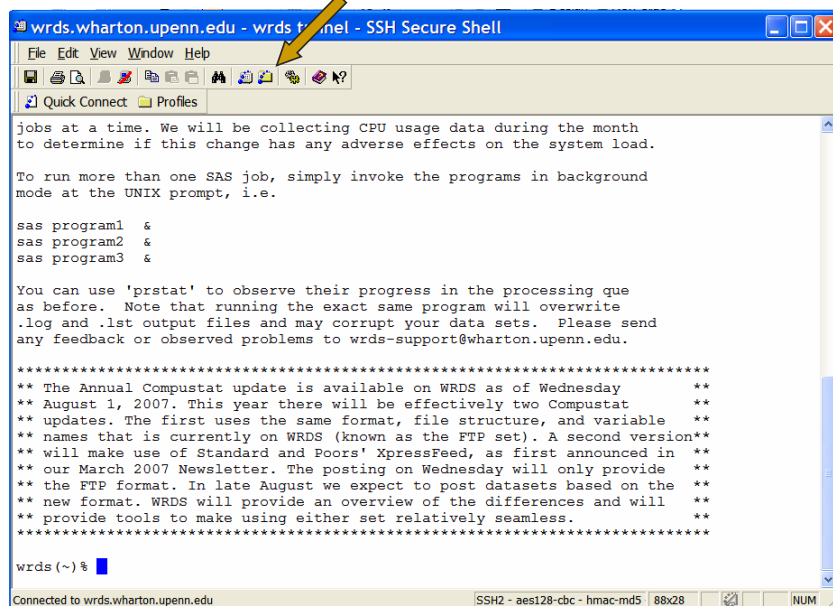
Step 1: Server name is wrds.wharton.upenn.edu. You can set up a permanent profile or use quickconnect (shown below).



23

Steps to use WRDS Unix

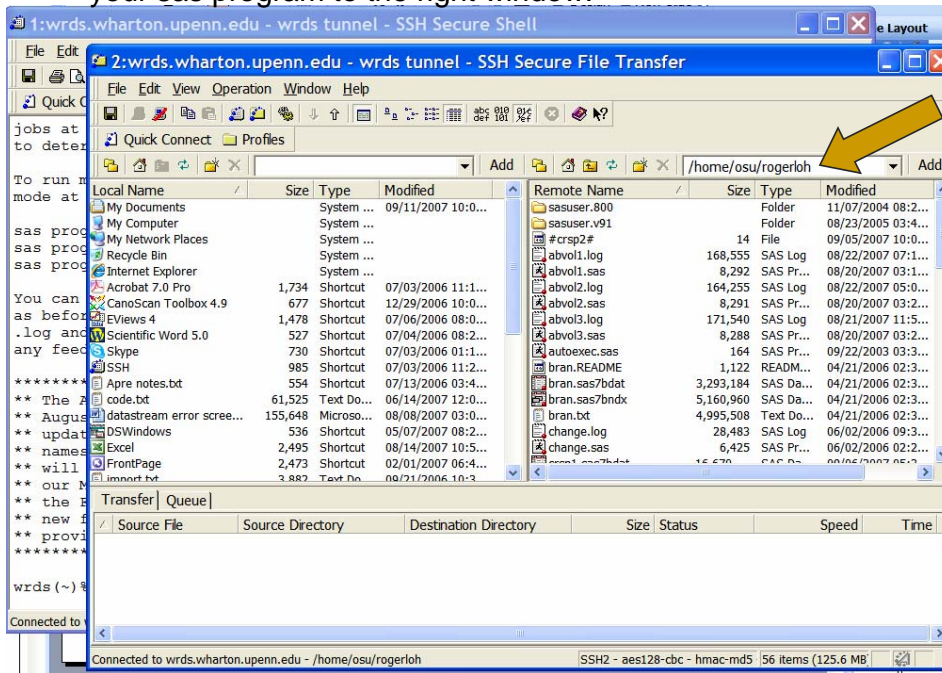
Step 2: You will see the UNIX command prompt screen once you log in, click on explorer window to upload your SAS program.



24

Steps to use WRDS Unix

Step 3: Explorer window in SSH, drag and drop capabilities. Drag your sas program to the right window.



Your personal folder on WRDS, 750 MB space.

25

Steps to use WRDS Unix

Step 4: The SAS program uploaded should look like this: This is written in a text editor and saved as, e.g., unix.sas. Note differences from PC SAS program.

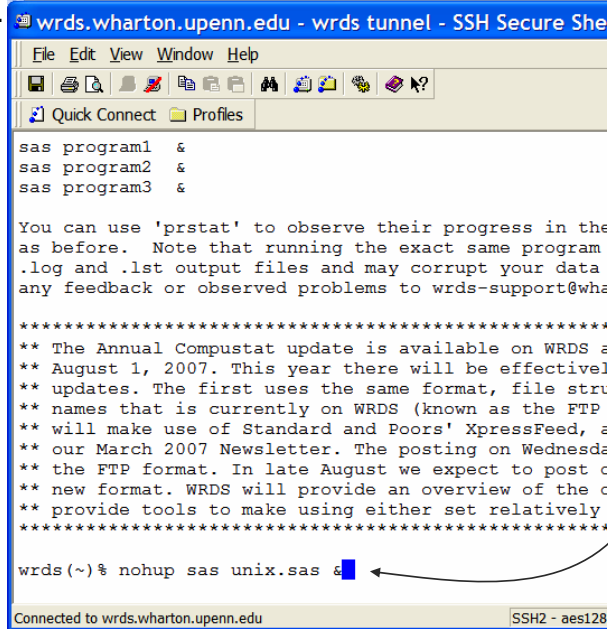
```
/*set library name*/  
libname myunix "/home/osu/rogerloh";  
  
/*Get data from CRSP monthly and save to my unix folder*/  
Data myunix.crsp1;  
set crsp.msf;  
keep permno date year ret prc;  
where date >'31dec2000'd;  
run;
```

myunix library is your personal folder.

26

Steps to use WRDS Unix

Step 5: Run SAS program in command prompt window by typing `nohup sas unix.sas &`. `nohup` allows program to run remotely even if you exit SSH.



```
wrds.wharton.upenn.edu - wrds tunnel - SSH Secure Shell
File Edit View Window Help
Quick Connect Profiles
sas program1 &
sas program2 &
sas program3 &

You can use 'prstat' to observe their progress in the
as before. Note that running the exact same program
.log and .lst output files and may corrupt your data
any feedback or observed problems to wrds-support@wha

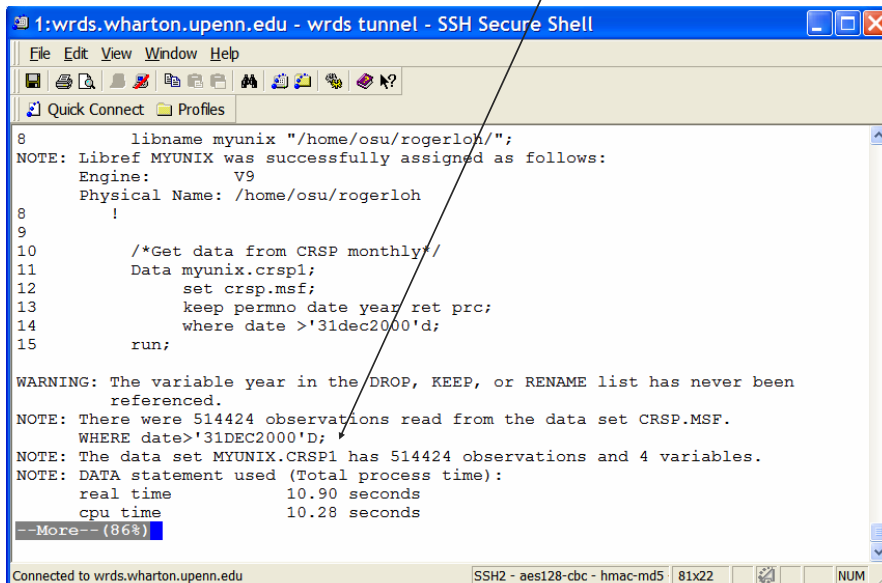
*****
** The Annual Compustat update is available on WRDS a
** August 1, 2007. This year there will be effectivel
** updates. The first uses the same format, file stru
** names that is currently on WRDS (known as the FTP
** will make use of Standard and Poors' XpressFeed, a
** our March 2007 Newsletter. The posting on Wednesda
** the FTP format. In late August we expect to post c
** new format. WRDS will provide an overview of the c
** provide tools to make using either set relatively
*****

wrds(~) % nohup sas unix.sas &
```

27

Steps to use WRDS Unix

- Step 6: See log or output by typing `more unix.log` or `more unix.lst` at the command prompt. Final file `crsp1` was saved in your WRDS personal folder.



```
1:wrds.wharton.upenn.edu - wrds tunnel - SSH Secure Shell
File Edit View Window Help
Quick Connect Profiles

8      libname myunix "/home/osu/rogerloh/";
NOTE: Libref MYUNIX was successfully assigned as follows:
      Engine:          V9
      Physical Name:  /home/osu/rogerloh
9
10     /*Get data from CRSP monthly*/
11     Data myunix.crsp1;
12     set crsp.msf;
13     keep permno date year ret prc;
14     where date >'31dec2000'd;
15     run;

WARNING: The variable year in the DROP, KEEP, or RENAME list has never been
referenced.
NOTE: There were 514424 observations read from the data set CRSP.MSF.
      WHERE date>'31DEC2000'D;
NOTE: The data set MYUNIX.CRSP1 has 514424 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time          10.90 seconds
      cpu time           10.28 seconds

--More-- (86%)
```

28

[Common Unix commands on wrds]

- See wrds.wharton.upenn.edu/support/docs/UNIXIntro.pdf for more unix commands.
 - nohup sas program.sas &: to submit jobs
 - ls: list all files in the folder
 - more filename.ext: to read text file
 - pico filename.ext: to edit file
 - emacs filename.ext: to edit file
 - cd /folder/: to access folder
 - ps -fu username: to check status of jobs
 - kill #jobnumber: to kill a program

29

[Fisher unix server]

- Fisher's own unix server. Programs SAS, Matlab, R, etc are available. See <http://fisher.osu.edu/services/computing-services/knowledge-base/unix-research/>
- Advantage, less competition for processing power, more user space (4GB in /home/student/username/). Disadvantages, no financial databases there, need to upload your own data.
- Apply for unix account (I will give you the form)
- Log in to research1.cob.ohio-state.edu (for SAS) using your email user and password.
- Works the same as WRDS Unix except the "nohup" doesn't work. Instead do these 2 steps.
 - sas program.sas &
 - disown #jobnumber

30

[Outline]

Overview of some WRDS databases

Accessing WRDS

Datastream/SDC (Taylor)

Practice sessions in Fisher 606

31

[Datastream/Worldscope]

- Datastream is a comprehensive international database:
 - Stock returns (>50,000 stocks in >60 countries)
 - Sompany financial data (provided within Datastream by Worldscope)
 - Bond returns
 - Stock and bond indices
 - Foreign exchange, commodity prices, and economic data.
- Indispensable for studies needing international data.
- However, the data interface is not user-friendly and there are many errors in the dataset. See for eg., Ince and Porter (2004). <http://papers.ssrn.com/id=486523> or appendix of Griffin, Nadari, and Kelly (2006).

32

Datastream/Worldscope

- Requires a software installation on your desktop. Important to log out after use because only one person can be logged in.
- 2 ways of access:
 - Datastream Advance Excel add-in (simple but limited usefulness)
 - DS Windows macros
 - Cross-sectional data (use 900A macro)
 - Time-series data (use 900B macro)
- If you have many macros to run, DS Agenda can help you to run them sequentially.

33

Datastream/Worldscope

- Categories of Data that will be useful
 - Equities
 - Equity Lists
 - Equity & Misc Indices
 - Exchange Rates
 - Economic Series
 - Interest Rates
- Variables available
 - Datatypes
 - Worldscope Data Items

	Base-Date	Mnemonic	Code	A/C	MSCI	WS	Mkt
	09-11-03		27564U	N			IR
	06-25-03		27211D	N			SD
	12-22-05		32594W	N			UK
	06-02-05		31162L	N			HN
	07-17-00	D:EN2X	296896	N			BD
	07-16-04	D:PLXX	29245T	Y			
	07-16-04	D:PMKX	29245X	Y			
	05-03-05		30925R	N			HN
	03-23-05		30727F	N			UK
	08-09-04		29326J	N			IR
	02-22-05	800A	30379R	N			US
	04-13-06	D:BSW	35703C	Y			BD
	02-07-06	D:BQP	32794V	Y			BD
	02-08-06	D:EXQ	32801N	Y			BD
	02-09-06	D:FJQ	32803F	Y			BD

34

Eg. Get prices for all 30 Dow Jones Index stocks

1. Open DS Windows
2. Get dscodes (identifier codes) for all 30 firms. Under “equity lists” category, search for Dow Jones Industrial and we find the code LDJINDUS.
3. Enter LDJINDUS into the Codes: section of the 900A macro (cross-section macro).
4. Run 900A macro to get dscodes.
5. Collect the output dscodes and paste them onto the 900B macro.
6. Run 900B macro to get the monthly price series for each of the 30 firms.

35

900A macro example

```
STARTDC(CSVFILE, "C:\dataosu\listfirm.csv", overwrite)
OpenData Codes
Loop:
If &endOfData = FALSE Then
Input code

>900A
Send(CODE)
>NAME, P
>12-31-05
>[CLEAR]

      Goto Loop
EndIf
ENDDC
End

Codes:
DATA

"LDJINDUS"

ENDDATA
```

36

900B – Time-series macro

```

STARTDC(CSVFILE,"C:\dataosu\900bdata.csv")
OpenData Codes
Loop:
If &endOfData = FALSE Then
Input code

Send( "900B " + code + "(P), 1-1-2002, 12-31-2003, D//C" )

Send("[CLEAR]")
Goto Loop

EndIf
ENDDC
End

Codes:
DATA

"902172"
"945388"
"905113"
"904853"
"906156"
"916305"
"904818"

ENDDATA

```

Making downloaded Datastream data usable

- Data extracted is in a panel format.
- But we need it in a stacked format for analysis. Need to write a program (e.g. in SAS) to transpose the data.

	A	B	C	D	E	F	G
1	Name	3M	AT&T	ALCOA	ALTRIA	AMERICA	AMERICA
2	Code	902172(P)	945388(P)	905113(P)	904853(P)	906156(P)	916305(P)
3	Currency	US	US	US	US	US	US
4	1/1/2002	59.105	39.17	35.55	45.85	31.2412	79.4
5	1/2/2002	58.57	39.9	35.68	46.64	31.4075	78.75
6	1/3/2002	58.375	40.17	36.14	46.59	31.9064	78.58
7	1/4/2002	58.55	39.99	37.3	46.09	33.0181	77.8
8	1/7/2002	57.85	39.91	38.16	46.58	32.9131	76.8

	A	B	C	D
1	Date	Name	dscode	Price
2	1/1/2002	3M	902172	59.105
3	1/2/2002	3M	902172	58.57
4	1/3/2002	3M	902172	58.375
5	1/4/2002	3M	902172	58.55
6	1/7/2002	3M	902172	57.85
7	1/8/2002	3M	902172	57.525
8	1/9/2002	3M	902172	57.325
9	1/10/2002	3M	902172	56.6
10	1/1/2002	AT&T	945388	39.17
11	1/2/2002	AT&T	945388	39.9
12	1/3/2002	AT&T	945388	40.17
13	1/4/2002	AT&T	945388	39.99
14	1/7/2002	AT&T	945388	39.91
15	1/8/2002	AT&T	945388	39.74
16	1/9/2002	AT&T	945388	38.17
17	1/10/2002	AT&T	945388	38.5

[Finding DS Codes]

- Steps to search through Datastream's menu of codes
- 1. Type "help" in DS Window program number prompt.
- 2. Select "codes" (#4 on the menu)
- 3. Search through Datastream code menu to find relevant code.

Note: Datastream has many data items that Worldscope does not. In general, Worldscope has annual firm-level accounting data. Datastream will have daily and monthly stock returns as well as international stock market indices, industry identifiers, and other equity specific data items (e.g. volume, dividends, total return, etc..)

39

[Downloading Individual DS Time Series]

- Steps to download individual time series in DS using 900B
- 1. Type 900B into Program Number prompt.
- 2. Click on furthest right icon to create a blank csv file to which the data will be sent. Follow the prompt to create a file name.
- 3. The prompt will ask for a code or expression. Enter mnemonic of the data series you want with the data type in parenthesis.
 - Example: MSUEIMV(RI)
- 4. Enter start date of desired data.
- 5. Enter end date of desired data.
- 6. Enter data frequency (e.g., d(daily), w(weekly), m(monthly))
- 7. Hit return until download is complete.
- 8. Go to file – end data channel/fundline.

40

Downloading Individual DS Time Series

```
DATASTREAM: 900B
PLEASE ENTER CODE OR EXPRESSION : MSUEIMV(RI)
PLEASE ENTER START DATE : 1-1-2000
PLEASE ENTER END DATE : 1-1-2007
PLEASE ENTER FREQUENCY : |
```

41

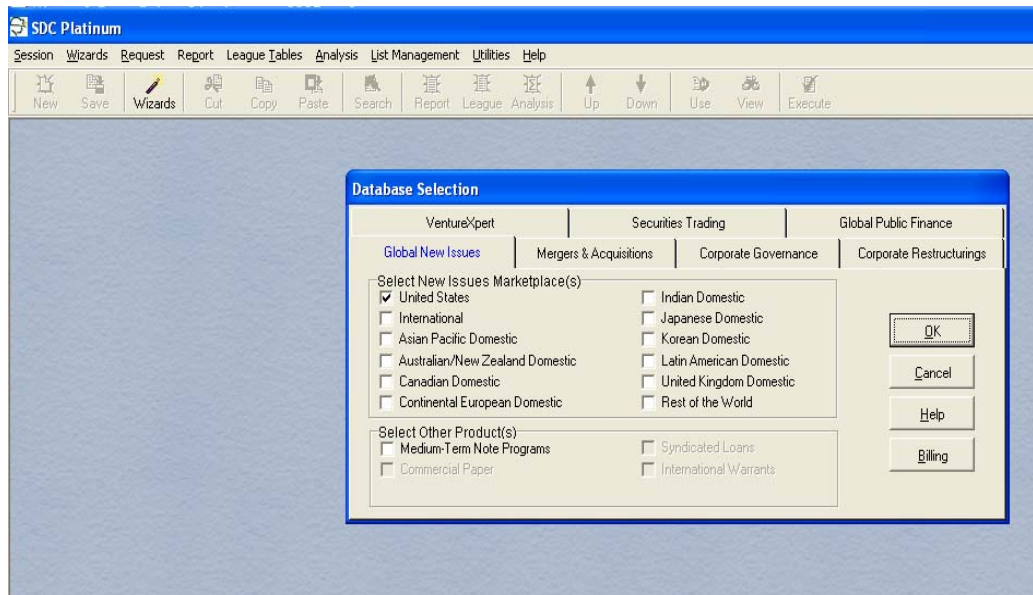
SDC

- Thomson's Securities Data Company.
- Corporate issuance database
 - New Issues (US and International)
 - IPOs and secondary issues
 - Bond offerings
 - Rights offerings
 - Mergers and Acquisitions
 - Hostile/friendly takeovers
 - Successful and unsuccessful deals
 - US and non-US targets
- SDC also contains some errors. See:
 - <http://pages.stern.nyu.edu/~aljungqv/research.htm> (Alexander Ljungqvist's website)
 - <http://bear.cba.ufl.edu/ritter/ipodata.htm> (Jay Ritter's website)

42

SDC Example

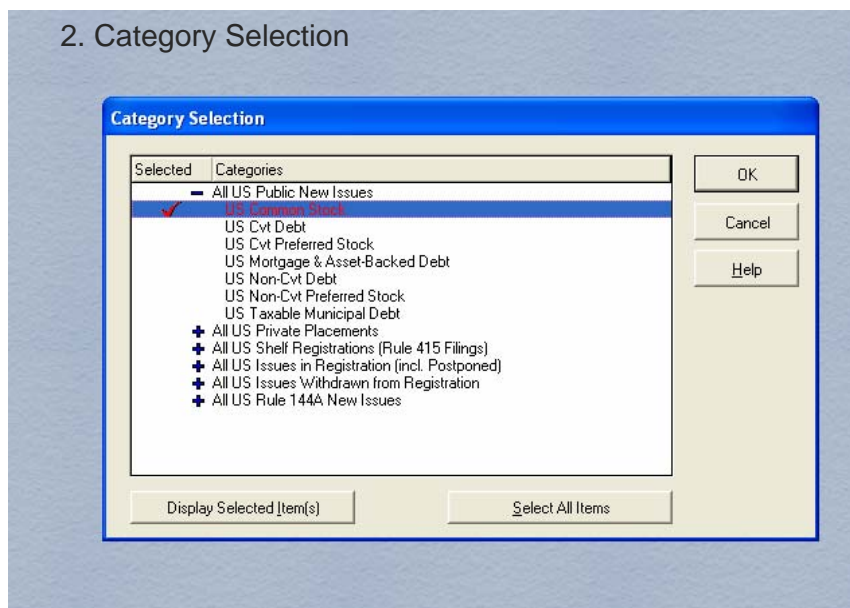
1. Select Database



43

SDC Example

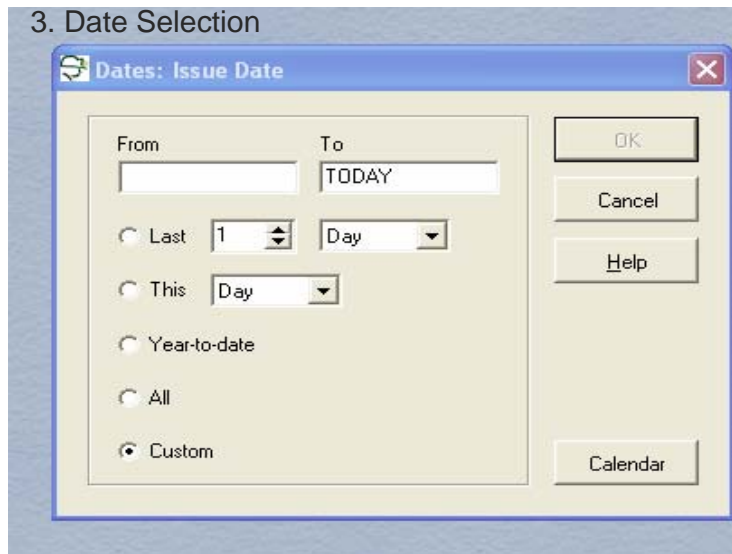
2. Category Selection



44

SDC Example

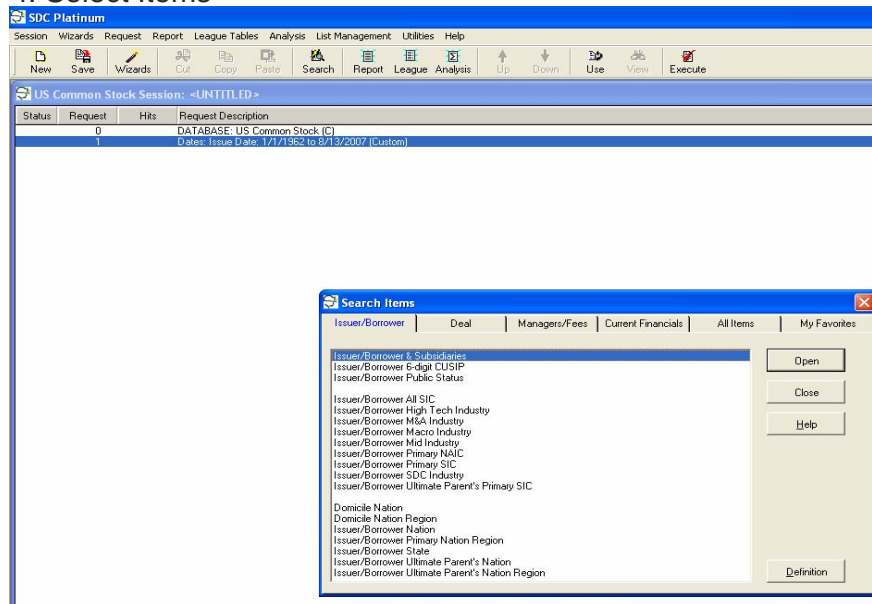
3. Date Selection



45

SDC Example

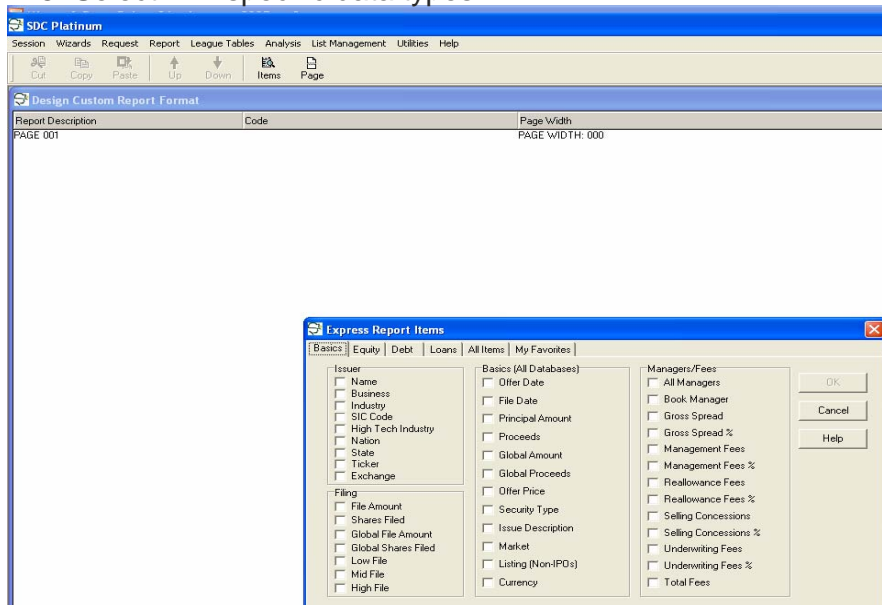
4. Select Items



46

SDC Example

5. Select firm-specific data types



47

Outline

Overview of some WRDS databases

Accessing WRDS

Datstream/SDC

Practice sessions in Fisher 606

48