Statistical Arbitrage 237M – Section 1
Professor Ledoit – Anderson Graduate School of Management - UCLA

Problem Set #1 : Universe

WARNING: This problem set is very long. You have one week to do it. It is due on Tuesday 16 October 2012 before class (10am). If you start two days before you'll never make it. You should start today.

1) We will work with the following list of 13 European stock market indices:

- Belgium: BEL-20
- Denmark: OMX Copenhagen 20
- Finland: HEX-25
- France: SBF-120
- Germany: DAX-100
- Holland: AEX-25 and AMX-25
- Italy: MIB-30 and MIDEX
- Norway: OBX-25
- Spain: IBEX-35
- Sweden: OMX Stockholm 30
- Switzerland: SMI-20

For each one of these 13 stock market indices, retrieve the list of the DataSteam codes (dscodes) of its constituents on each of these 5 dates:

- January 1st, 1998
- January 1st, 1999
- January 1st, 2000
- January 1st, 2001
- January 1st, 2002

If historical constituents for a given index are unavailable prior to, say, January 1st 2000, then you are allowed to backfill the January 1st 1998 and 1999 data with the January 1st 2000 list. We will ignore the forward-looking bias that it induces. But you must go as far back into the past as you possibly can.

HINTS:

- It is possible to get this data from Bloomberg. But then you need to figure out a way to map whatever you're getting out of Bloomberg into DataStream and specifically DScodes.
- We highly recommend that you install the DataStream Excel add-in on your account when you use DataStream in the library lab – rather than use the stand-alone DataStream front-end.

2) Form a merged list of all unique dscodes. Let `n` denote the length of the merged list.

3) Create a Matlab structured variable of dimension `(1×n)` called `allstocks` such that `allstocks(i).dscode` is a string equal to the dscode of the `i`th stock on the merged list (for `i=1,...,n`).

4) Within the variable `allstocks`, create a sub-structure called `namelist`: `allstocks(i).namelist`

Set `allstocks(i).namelist(1).date = '01-Jan-1998'`
Set `allstocks(i).namelist(1).name` equal to the name of the `i`th stock on January 1st, 1998.

Check whether stock `i` had changed name by January 1st, 1999.

- If so, set `allstocks(i).namelist(2).date = '01-Jan-1999' and set allstocks(i).namelist(2).name` equal to the name of the `i`th stock on January 1st, 1999.
- If not, do nothing.

Reiterate for all the years from 2000 to 2002. Once you are done, the dimension of `allstocks(i).namelist` must be equal to the number of different names the company has had over these 5 years. In general this will be strictly less than 5. For many stocks, it will be just one.

5) In a similar way, populate `allstocks(i).industrylist` using the 5-letters DataStream level-4 industry mnemonic.

6) In a similar way, populate `allstocks(i).ibeslist` using the I/B/E/S ticker. There should be only one I/B/E/S ticker per dscode, but we have to be ready in case this rule is violated.

7) In a similar way, populate `allstocks(i).indexlist` using stock index memberships. If a stock does not belong to any index at the beginning of year `j`, then set `allstocks(i).indexlist(j).index=''`

8) In a similar way, populate `allstocks(i).bblist` using Bloomberg tickers.

9) In a similar way, populate `allstocks(i).isinlist` using 12-character alphanumeric ISIN codes.

Problem Set output: You must e-mail T.A. Kyle Matoba a Matlab database containing the variable `allstocks`.

Grading: You will be judged by how complete and correct `allstocks` is.

VERY IMPORTANT: Due to the limited time that the T.A. can devote to this class, every student is allowed at most one e-mail question to the T.A. – so use it wisely! In addition, the T.A. will hold one office hour on Wednesday which will be devoted to this problem set. That's another reason to get started today or tomorrow at the very latest.