2016 Morgan Stanley Prize for Excellence Project Description

Introduction

Morgan Stanley has a leading securities market-making business globally, and data analysis is an essential component in such a business.

There has been quite a buzz around the subject of big data. One ongoing challenge within the world of "big data" analytics is how to optimize large text dataset archives (ie: tweets, news articles, and email). An example of a a smaller text dataset is the Federal Reserve Bank's FOMC (Federal Open Market Committee) meeting statements, minutes, press conference transcripts, supplements to statements/minutes, etc. In this project, we will examine the FOMC meeting text dataset, and try to use it as a supplement to the price time series of traded instruments together with some macro-economic indicators.

For the 2016 Morgan Stanley Prize for Excellence Project, you will be provided with time series data of daily prices of 100 traded instruments, several non-traded macro-economic indicators, and FOMC meeting texts from 2010 to 2015. The goal of the project is to design a market making/trading strategy for the period that we provide data for, which will be tested out-of sample for the few days after the December FOMC meeting.

The deadline for submitting all codes and reports is December 6, 2015. Deliverables are described in the "Guidelines" section.

Data Description

1. The Federal Reserve Bank's FOMC meeting data are in 3 subdirectories (compressed in a zip file):

Subdirectory	File Name (sometimes file names may have "a" or "b" after the date, they mean multiple such data on the same date YYYYMMDD)	Description
Statements	Statement_YYYYMMDD.txt	Statement released on YYYYMMDD
PressConferenceTranscripts	PressConferenceTransript_YYYYMMDD.txt	Fed Chair's press conference transcripts released on YYYYMMDD

Minutes	Minutes_YYYYMMDD.txt	Meeting minutes released on YYYYMMDD

2. There are 100 tradable instruments under consideration, labeled 1-100, with daily time series in one .csv file. The instruments include:
   - 1-38: fixed-income-related instruments
   - 39-100: equity-related instruments

   They may not have the same length of historical data – empty cells mean data are not available for the date/instrument. There are 101 columns, the first column is date in YYYYMMDD format, and the remaining columns are daily dividend/split-adjusted close price time series data of the 100 instruments, respectively (with column names being their labels).

3. You may also use the macro-economic indicators. We have provided you with a few monthly macro indices from 2009-2015 in another .csv file. Those indices are non-tradable, but you can use them for your computation. The 3 columns are Date, Value, Index Name.

Guidelines

1. All submissions will need to include the following items (total size should not exceed 5MB):
   a. Summary Report: outlining approaches, methods, description of the final results, short description of how to run the code, etc - maximum 12 pages
   b. Project Code in one zip/tar archive
2. The evaluation of your submission will use updated data of the same formats as inputs:
   a. FOMC meeting text dataset will be updated with December 15-16,2015 FOMC meeting's "Statement" and "Press Conference Transcript" (if available; otherwise no such file); you should expect the input to be 3 subdirectory locations, not a zip file.
   b. Price time series updated with data on and before December 15, 2015.
   c. Macro-economic indicator time series updated with data on and before December 15, 2015.
3. Your code is expected to output static dollar positions in the given tradable instruments (can be positive/negative/0 but should not be missing; cannot involve macro-economic non-tradable indicators).
   a. Positions are put on December 16 before close and held constant in the next 5 business days, and the performance evaluation is based on these 5 days.
   b. You are encouraged to have more non-0 positions within the given instruments, to be more diversified.
   c. Output format: comma-separated .csv file (file name is your name without spaces) with 2 rows, the first row contains the instrument labels (ordered as 1-100), the second row contains the dollar positions corresponding to the labels in each column

    d. Please make sure the code works with the inputs specified above, otherwise your code will be rejected
4. The total risk limit is 10,000, namely, the sum of absolute long and short dollar positions should be smaller than or equal to this limit (and this sum cannot be 0), otherwise the results will be rejected by the performance evaluation.
5. Date convention in the data:
    a. Price time series data: the close price of each date can be only used for computation strictly after the given date
    b. Fed FOMC data and macro-economic indicators: each date is the date on which you can start using that date's data for computation
6. Since there are multiple datasets in this project, you can choose to use only part of the data for computing your output, but the allowed inputs must be as specified above.
7. Acceptable programming languages are R, Python, MATLAB. If you plan to use other programming languages, please contact us before you start.
8. Please clearly state what packages/libraries you use, and avoid using any commercial packages/libraries. In general, standard libraries are permitted, and please contact us first if you plan to use 3rd party packages/libraries.

Useful Resources

Google Scholar can be a starting point for any existing discussions about FOMC meetings, macro-economic indicators and asset prices.

Here is list of suggested references to get you started. We don't endorse any of the methodologies and you don't have to follow them Keep in mind that this is not an exhaustive list.

- http://www.federalreserve.gov/econresdata/notes/feds-notes/2015/fomc-meeting-minutes-an-assessment-of-counting-words-and-the-diversity-of-views-20150526.html
- http://stanford.edu/~rezab/papers/finlp_slides.pdf
- http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2191937
- http://www.nyfedeconomists.org/research/economists/rosenberg/Boukus_and_Rosenberg_072006.pdf
- http://www.newyorkfed.org/research/epr/2013/0913rosa.pdf

Examples

An example on 2 instruments is as follows (returns are computed from adjusted close prices):

date indices	return 1	return 2	dollar position 1	dollar position 2	total profit	return	cumulative profit	maximum drawdown
0			6000	-4000	0		0	0
1	1.00%	2.00%			-20	-0.20%	-20	20
2	-1.00%	-0.50%			-40	-0.40%	-60	60
3	1.00%	-0.50%			80	0.80%	20	60
4	0.50%	1.00%			-10	-0.10%	10	60
5	1.50%	-0.50%			110	1.10%	120	60

On day 0 (the assumed FOMC meeting date on which Statement and/or Press Conference Transcript are published), we put on the position; for every day afterwards, we have a profit (or loss if negative) on each static dollar position, e.g. on day 2, the total profit is 6000*(-1.00%)+(-4000)*(-0.50%)=-40, and return is -40/(|6000|+|-4000|)=-0.4%.

The (relative) maximum drawdown is 60/(|6000|+|-4000|)=0.6%.

The contents of the output .csv file for the above example would look like this:

```
1,2
6000,-4000
```

Evaluation Guidelines

1. Some metrics we look at are as follow (over the 5 days after the FOMC meeting):
    a. $Sharpe\ Ratio = \frac{Average\ Daily\ Returns}{Standard\ Deviation\ of\ Daily\ Returns} \times \sqrt{252}$
    b. Maximum Drawdown (absolute value), relative to the sum of absolute values of long and short dollar positions

    If any instrument stops trading at a date, its prices/returns afterwards are set to 0.

2. Please state your solution in the report in a clear manner. Better diversification, better economic rationale and innovative data analysis approaches will impact your score positively.

Attachments

Item	Description	File/directory
Prize For Excellence 2016	Problem description file	PDF file
FOMC text data	Several types of text data regarding the FOMC meeting	FOMC.zip, 3 subfolders after decompression: FOMC/Statements FOMC/PressConferenceTranscripts FOMC/Minutes
Daily price data	Adjusted price time series of 100 instruments	price.csv
Macro-economic indices	Macro-economic indicators with their values and release dates. Future release dates are also included.	macro.csv

Final Submissions / Questions

Final submissions (as well as any questions) should be e-mailed to **prizeforexcellence@morganstanley.com** by **December 6, 2015.**

NOTE: Finalists will be invited into our headquarters in NYC to present findings on January 7, 2016. If you are selected as a finalist, you will be notified no less than 1 week prior to that date and travel arrangements will be made at that time. Please keep this date available in your calendar. The first place winner will be determined and announced on this date.