Courant's TAQ data set

Lee Maclin
lee.maclin@gmail.com

Our TAQ Data Set

- **Three months of trades and quotes data for every stock**
- **Trades: Every trade print**
- **Quotes: Every change to the *national best bid and offer* (NBBO) but not the full order book**
- **About 14 gigabytes of data**
- **There are four identical databases in the Courant library for you to sign out on USB hard drives.**
- **Make sure the package says "Trades and Quotes data in Matlab format"**
- **Sign them out and copy the data to your own drive. Set aside some time to do this. It will take 15 mins to 30 mins to copy this data**
- **Copy only the Matlab data set, not the R data set. The data is the same, but the Matlab data can be copied more quickly and is the only data we use in class.**

Directory structure of our TAQ data set

- quotes
  - 20070620
    - IBM_quotes.mat
    - DELL_quotes.mat
    - ...
  - 20070621
  - ...
- trades
  - 20070620
    - IBM_trades.mat
    - ...
  - 20070621
  - ...

Loading one trades file in Matlab

load(“j:/trades/20070620/IBM_trades.mat”)

Format

- Trade files format

  e = seconds from Epoc
  n = number of records

  m = milliseconds from midnight
  s = trade print size
  p = trade price

- Quotes files format

  e = seconds from Epoc
  n = number of records

  m = milliseconds from midnight
  bs = bid size
  bp = bid price
  as = ask size
  ap = ask price