



THE BIG DATA COMBINE

The first rapid-fire, live tryouts for data scientists.

ED RAMSDEN



Electronic Engineer/Inventor/Author

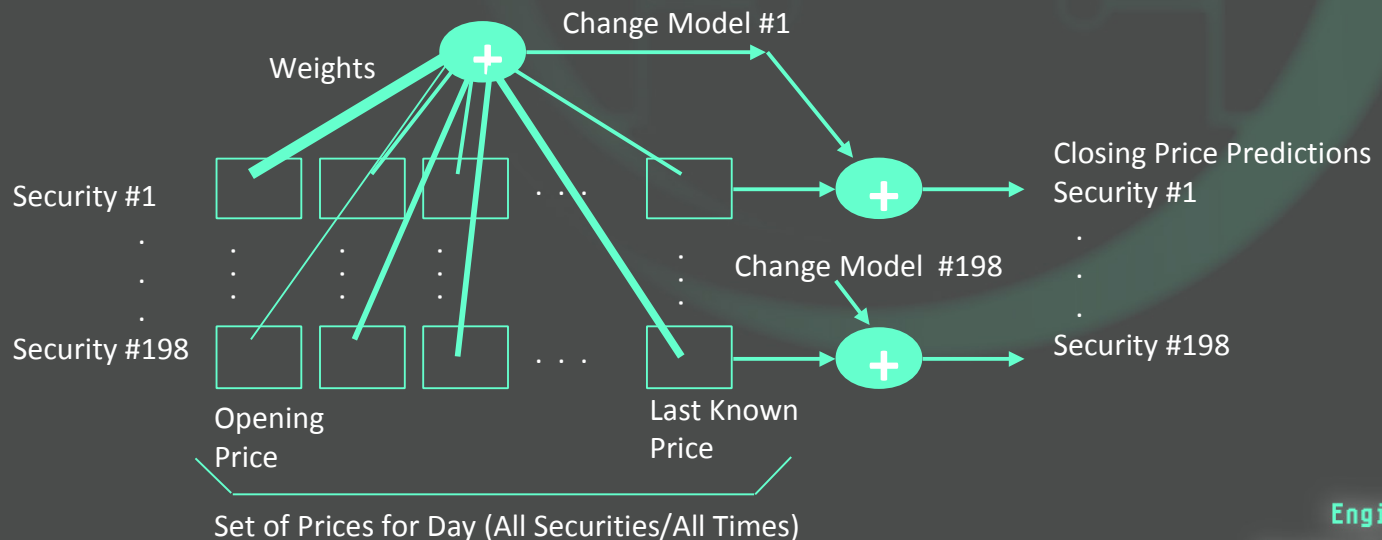
- BS Electrical Engineering
- MS Systems Science
- 11 Patents
- 'Wrote the book' on Hall-effect Magnetic Sensors.

Introduced to Analytics through Netflix Prize.

20+ Competitons since.

THE MODEL

- Predicts CHANGE between close and last known price.
- The change is the weighted sum of prices from ALL stocks/ALL times from each day – essentially a linear regression or 'line fit' using ~10,000 input variables.
- Sentiment inputs not used – resulted in worse predictions than just price info.
- Each Security gets its OWN model (set of weights).
- Weightings are determined by neural-network algorithm, not standard 'regression fit' technique.
- Slight improvement from building several models using differing subsets of data and then averaging results.



APPLYING THE MODEL - 1

LONG-ONLY STRATEGY

- Select a pool of 'securities of interest'. This pool should be chosen from those that are most predictable by the algorithm, and may (will) change over time.
- Choose an active portfolio from this pool to buy. The portfolio should be a substantial fraction of the pool and may contain hundreds or even thousands of individual securities.
- Perform predictions of changes for all securities in pool. Attempt to find pairs X (not in current portfolio) and Y (currently in portfolio) for which the change predictions differ ($\Delta_X \gg \Delta_Y$)
- If the difference exceeds the Bid-Ask spreads plus transaction costs of the two securities by some margin ($\Delta_X - \Delta_Y \gg BA_X + BA_Y + TC_X + TC_Y$), then Sell Y, Buy X.
- 'Cash' is a viable position to shift into if it looks better than an alternative security.
- Rinse, Repeat

APPLYING THE MODEL - 2

APPLICATION CONSIDERATIONS

- Primary Assumption of Strategy is that market is a near-random walk. The expectation is that buying securities expected to rise to a higher price over the short-term will continue to 'random walk' from that point.
- Large number of securities desirable for both protection against individual risk and to avoid shifting the market with large sales/purchases.
- Strategy does not protect against systematic market risk, but could be adapted to use short positions as hedges.
- Appropriate Time Horizon – The competition used 2 hours, but this may not be optimal. Longer or shorter may yield better results.
- Use of Sentiment Data – While the initial experiments with 'blind' sentiment data were not successful, the future addition of sentiment data to the model should not be ruled out.
- Strategy is short-term and greedy – it will not make long-term strategic decisions!

Comments - October 2014

- 1) Although this model ended up taking 3rd place out of over 400 competitors, it only appears to provide on the order of 10-15 basis points (0.1-0.15%) average improvement over just taking the last known price. Although this could theoretically translate to >25% a year improved returns over 'buy & hold', it does not account for commissions and slippage. For any retail trading company I am aware of, these would completely overwhelm the algorithms advantage and you would end up doing less well (probably a LOT less well) than just buy & hold.
- 2) I would like to thank Brian Tomeo, Tim Harrington, and the other people at Battlefin for sponsoring this competition and the event in Miami, which were a tremendous amount of fun. More data science competitions should be like The Big Data Combine !!
- 3) None of this presentation should be taken as financial advice. If you are looking for investment strategies, talk to a professional.