Experimentation in trading
For: Robo-Advisors & Systematic Investing / Vasant Dhar
Introductions

• Me
  • Quant trader
  • Market making, stocks
  • Experimental optimization

• You
  • Studying
  • Trading
  • Modeling
  • Backtesting
Quant trading

Execution

• Agency: no position risk

• Big customer order ==> many small trades

• Thousands of trades / day
Quant trading

High-frequency market making (HFMM)

• Principal: Small positions for minutes
• Continuous quoting
• Thousands of trades / day
Quant trading

Statistical arbitrage

• Principal: Larger positions for hours to many days
• Opportunistic trade selection
• Tens - hundreds of trades / day
Experimentation

- Compare performance in **live trading**
- Returns predictions incomplete
  - risk, liquidity, capital, preferences
- Simulation (aka backtest) too hard
  - Market reaction / impact
  - Latencies
  - Complexity
Experimentation: Complexity

- Example: US stock market
  - 13 lit exchanges
  - continuous book, auctions, blind bidding, block trading, internalization
  - orders: limit, market, IOC, FOK, AON, ISO, hide & slide, hidden, post-only
Experimental methods

- Experimental methods b/c
  - Evaluation is expensive: $$/time/risk
  - Evaluation is uncertain (noisy)
- Experimental methods
  - Minimize expense
  - Minimize uncertainty
Experimental methods

• Without experimental method
  • Deploy new model, make money
  • ==> “My new model is great!”

• With experimental method
  • Run new and old models side-by-side
  • \( P\{\text{new beats old}\} = 0.53, \text{deltaPnL} = 700 \pm 2200 \)
  • ==> “My new model isn’t \textit{bad}.”
Experimentation pervasive

- Medicine
- Psychology
- Behavioral Economics
- Web search
- Online advertising
- Social media
- Food engineering
- Materials science
- Social science
- Manufacturing
- Consumer product design
Example: HFMM A/B test

- Scenario: HTF strategy
  - A: Existing model, across 100 stocks
  - B: Your new model
- Model B looks better than Model A
  - Lower out-of-sample RMSE
  - Higher PnL in simulations
- Next: A/B test in live trading
Example: HFMM A/B test

Design
Determine number of measurements to take

Measure
Trade and measure PnL

Analyze
Decide whether to accept or reject B version
Example: HFMM A/B test

- Design:
  - Which stocks will run Model A, Model B?
  - How many days will this run?
- Measure:
  - Trade and record pnl by stock, day: $p_{s,d}$
- Analyze:
  - Which is better A or B?
  - Significantly better?
Design: Randomization

- Randomly assign 50 stocks to Model B, $\chi_B = 1$
- Assign other 50 to Model A, $\chi_B = 0$
- Experiment measures $corr(p_{s,d}, \chi_B)$
Design: Randomization

- Correlation w/anything else is confounder bias
  - Ex: All tech stocks in A, all energy stocks in B
  - Ex: High-liquidity stocks in A, low-liquidity stocks in B
  - Ex: …

- Randomization removes* confounder bias
  - Even if you don’t know the confounders

*“removes” == “aims to minimize”
Design: Replication

- PnL is noisy
- Replication: Avg. over multiple days

\[ \mu = \frac{\sum d P_d}{N} \]

\[ se = \sqrt{\frac{\sum (p_d - \mu_d)^2}{N}} = \frac{\sigma}{\sqrt{N}} \]

- Replication decreases “noise” as \(1/\sqrt{N}\)
Design: Replication

• Before experiment

1. Estimate $se$ ← Data
2. Specify minimum interesting $\delta_{min}$ ← Subjective
3. Calculate N
Design: Find N

• Some analysis:
  • Limit false positives (5%)
  • Limit false negatives (20%)

\[ N \geq \left( \frac{2.8se}{\delta_{\text{min}}} \right)^2 \]

NB: Quadratic

• Typically 1-2 weeks for real experiments
Measure

- Trade!
- Start small for safety
- Stop if *any* metrics look very bad / different
  - PnL terrible — or wonderful!
  - Trading way too little / too much
  - Sending too many / too few orders
- Log everything, record $p_{s,d}$
Analyze

• Which model earned more PnL?

\[ \Delta_{PnL} = \mu_B - \mu_A \]

• Enough to care?

\[ \Delta_{PnL} > \delta_{min} \]

• Statistically significantly more?

\[ t = \frac{\Delta_{PnL}}{se_{PnL}} > 1.64 \]
Measure: Warning

- Plan: Wait $N$ days, then ask “Is $t > 1.64$ ?”

- Why not just check every day?
  - Stop when $t > 1.64$?  ➡️ Don’t do this!
  - Many false positives!

- You can monitor $t$ — Just don’t stop based on $t$
Bayesian optimization

• Modern, flexible, efficient method(s)
• A.K.A.
  • Adaptive experimentation
  • Black box optimization
  • Surrogate optimization
  • Model-based optimization
Bayesian optimization

- Can compare A, B
- Also: A, B, C, D, ...
- Also: 1, 2, 3, ...
- Also: $[0,1], [0,1]^D$
- Also: $\{A, B, C, \ldots\} \otimes \{a, b, c, \ldots\} \otimes \{1, 2, 3, \ldots\} \otimes [0,1]^D$
- IOW: BO can optimize your strategy’s parameters.
Bayesian optimization

- Other uses:
  - Hyperparameter optimization (HPO) for supervised learning models
    - NNs, trees
  - Optimize parameters of strategy in simulation
- BO takes longer to calculate a design …
  - Compared to, ex., BFGS, CMA-ES
- … but requires fewer experiments overall
Bayesian optimization: Design

- Fit a surrogate:
  \[ y(x) = \text{PnL(parameters)} \]

- Usually Gaussian process regression (GPR)

- Then, maximize \( y(x) \) over \( x \):
  \[ x = \arg\max_x E[f(x)] \]

- **BUT**: Surrogate is poor b/c so few measurements
Bayesian optimization: Design

- Instead:

\[
x = \arg\max_x \left[ E[f(x)] + \sqrt{\text{VAR}[f(x)]] \right]
\]

- GPR outputs \( E[f(x)] \) and \( \text{VAR}[f(x)] \)

- Acquisition function: \( E[f(x)] + \sqrt{\text{VAR}[f(x)]} \)

- Exploration improves surrogate
Bayesian optimization: Measurement

- Same as before
- Go trade
Bayesian optimization: Analysis

- Have you exhausted your budget for experimentation?
- If not, rebuild GPR, design again
Bayesian optimization

- **Mixed variable types**: Continuous, ordinal (integer), and categorical (boolean)
- **Multiple metrics**: PnL, risk, volume, order rate, … simultaneously
- **Multiple fidelities**: Combine simulator results w/live results
- **Constraints**: Limit risk, capital, market participation
- **Arbitrary measurements**: Build surrogate from all available measurements

- You build your surrogate as a model of your whole trade.
Bayesian optimization: Tools

- Open source: botorch.org
  - SKLearn: Supervised learning :: BoTorch: Bayesian optimization
  - Flexible, powerful modeling toolkit
- cogneato.xyz
  - User-friendly
  - Works with a spreadsheet or Pandas
  - Simple, but produces good designs
Bayesian optimization: Summary

- Surrogate models trade
- Acquisition optimization designs experiments
- Iterate until done