

# Strategy back-testing in the Cloud

Dan Wang, PhD, CFA

Chicago Trading Group

2023/5/16

# Simulation/Back-testing

- Simulations are useful tools to help analyze problems.
- Example 1: Exotic options where the underlying stock price follows a stochastic process that does not have an explicit analytic formula.
- Simulate the underlying stochastic process and aggregate the result to get an estimation of the option price.
- Example 2: Sample from a multi-dimensional distribution probability distribution.
- Gibbs sampling:
  - target to sample from a multi-dimensional distribution, it generates each coordinate from a conditional distribution given other coordinates.
  - Markov Chain Monte Carlo (MCMC)

# Simulation/Back-testing

- Example 3: a portfolio manager wants to launch a new strategy.
- **Goal of Simulation/back-testing:**
  - Approximate real-life investment process
  - Analyze portfolio returns, risk characteristics, style exposures and drawdowns etc.
  - Compare different strategies and select the optimal ones.
- **How to back-test a trading strategy?**
  - Define a set of investment rules and processes, key parameters, and apply those rules to historical data.
  - Example: when Gold price goes up, technology sector tends to outperform the rest of the market.
- **We need to answer following questions:**
  - which bonds to chose?
  - At what price shall we entry or exit?
  - What quantity?
  - How long shall we hold?
  - Etc.....

# Cloud computing

- **Cloud computing:** cloud computing is the on-demand availability of computer systems resources. Instead of buying/owning/maintaining physical data centers and serves, one can access data storage and computing power, on an as-needed basis.
- **Advantage of cloud computing:**
  - Pay-as-you-go, no fixed cost.
  - No physical on-site data center, 24-hour electricity.
  - Large clouds have functions distributed over multiple locations, huge computation power.
  - Sharing resources.
  - Cloud storage has similar structure as local one.

# Cloud computing



No



YES

# Back-testing trading strategies

- **Back testing a trading strategy:**
  - **strategy design -> historical investment simulation -> analysis of backtesting output**
  - Define a set of mechanisms of investment hypothesis/goals, rules and processes.
  - Apply the strategy or predictive model to historical data to determine its accuracy.
  - Test and compare the viability of trading strategies.



# Back-testing trading strategies

- **Performance measures for back-testing**

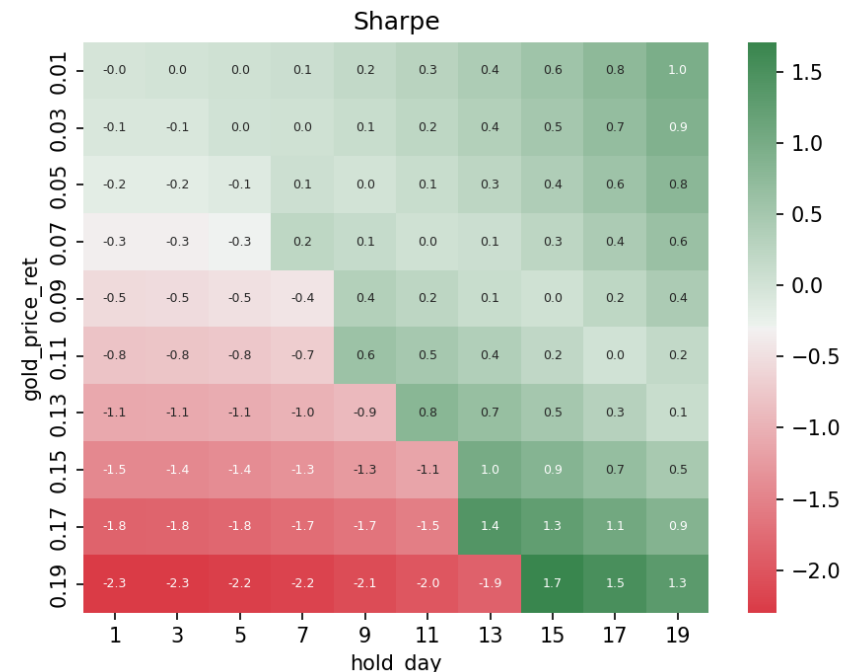
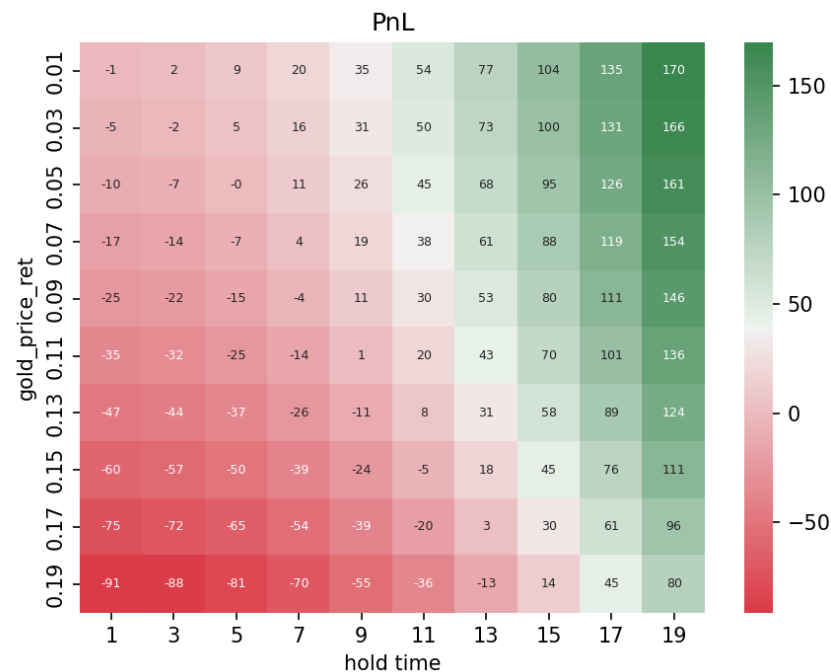
- Net PnL
- Risk adjusted returns: the return of the portfolio adjusted for a level of risk
- Market exposures: the degrees of exposure to different segment of market, different industry
- Volatility: dispersion of the portfolio returns
- Risk: exposures to risk factors
- Drawdowns

- **Back-testing as industry standard**

- *In a CFA Institute survey of nearly **250** analysts, portfolio managers, and private wealth managers on quantitative investment techniques, **50%** of respondents reported that they had conducted back testing of an investment strategy within the past 12 months of the survey date. This result underscores the importance of back testing (and other simulation techniques) for investors in practice.*

# Back-testing examples

- Using Examples 3 we have earlier, we want to visualize the performance of a strategy that enters a position when the gold\_price\_ret is x, and holds the position for n-days





# Back-testing pitfalls

- There are some subtle things need to consider during the back-testing:
  - **In-sample testing:**
    - train and test data, it will inflate the performance. (over-fitting)
  - **Survival bias:**
    - index content is changing over time. If we fail to consider the excluded names, we will introduce the ‘picking the winners’ bias
  - **Look-ahead bias:**
    - future data “sneak in” into the back-tester.
    - Eg: reporting lags for major economic data, revisions, index additions.

# Back-testing pitfalls (cont.)

- **Market-regime change:**
  - the market “parameters” are not stationary.
  - The recession period
  - The crisis period: 2008-global financial crisis; 2020 Covid
  - Periods of varying interest rates
  - Business cycles
  - Major technological changes: dot-com bubble, AI
- **Trading cost:**
  - fees/commission/borrow-cost.
- **Price impact:**
  - We need to take into account the price impact from the strategy’s own trades

# Other issues

- Trading opportunity: back testing is easy when using 'infinite' paper money.
- Need to consider:
  - The fact that 'real-life' strategy capital is constrained
  - Total strategy trading volume
  - Drawdown limit
  - Risk tolerance
- Seasonality:
  - Is the trade flow at "end of month" or "year end" different than the rest of time?

# Other issues (cont.)

- Robustness of back-testing:
  - Varying the starting time of the strategy may change the result
  - Does it matter if the strategy starts on Monday or Tuesday
  - Varying the strategy parameters
  - Varying the pricing data sources
- Bias-variance tradeoff:
  - the choice is art.

# Practical considerations when running in the cloud

- **Data:**
  - Data need to be stored in the cloud, local data cannot be accessed by the cloud.
  - To optimize the run on the cloud, one need to determine the best format to store the data.
  - Speed of loading/accessing the data need to be considered.
  - Sometimes, preprocessing the data could help optimize the usage of cloud.



# Practical considerations when running in the cloud (cont.)

- **Cluster/workers:**

- With given/limited number of clusters/workers, consider ways to maximize the total speed.
- An estimation of the data size and total memory used for each run.

- **Smooth process:**

- When running in the cloud, there is risk of losing connectivity (though most cloud platform have automatically reconnect)
- Design and run the process step by step and save the intermediate steps.
- Label the runs in a meaningful way.

# Questions

# Disclaimer

- Views and opinions expressed in this presentations are those of the speaker and not necessarily reflect the views or opinions of CTC.