Independent Project in Financial Engineering

Project Report: Each project report will be evaluated for both content and presentation. Therefore you should plan carefully for both the development of the financial ideas as well as careful, concise, complete communication. It should contain the following elements.

1. **Executive summary:** Write a brief executive summary of your project and findings. It should be written in plain English for non-technical senior management (not MIT finance faculty or TAs).

2. **Project definition:** State what the objective is, why it is interesting, and how it is expected to work in theory and/or practice. Where should theory and practice be expected to differ, if at all, and by how much? Explain the relevant concepts and essential mathematical background.

3. **Scope definition:** State precisely what you are exploring, analyzing, predicting, trading, or testing. (Are you studying a single derivative, a dozen, or all those within a given market? Using reinforcement learning to simulate a trading strategy or explore the impact of asymmetric information?)

4. **Data:** If your project involves data, state precisely what data you are using, why you chose it, and how you prepared it. Data may be obtained from various sources at Sloan or elsewhere, including WRDS, Bloomberg, CRSP, FactSet, FRED, or even Yahoo Finance. You should choose whichever one(s) are most appropriate or convenient, including an explanation of your choice in the final report. The report should describe any validation or data cleaning you performed so that anyone with access to the same data source could precisely reproduce your results.

5. **Methods:** Explain how you model your ideas, how you perform your analysis, how you obtain results, and how those results are evaluated. Describe the computational tools you use to explore data, visualize relationships, or compute predictive outcomes. Explain the tools used to estimate parameters, calibrate models, or evaluate significance of results.

6. **Results:** What did you find?

7. **Discussion:** Did your results match your expectations? How closely? Where might they be applied? What extensions or further analysis is suggested by your findings?

8. **Sharing permission:**

   We would like to be able to share some of your project reports with future cohorts of students as examples. If you agree to have your report shared, please include the following at the top of the front page of the report:

   "We, *Names*, consent to future use of our report for educational purposes within the scope of the MFin program. The report may be shared with MFin students, teaching assistants, and faculty."

Presentation of results: Your results should be clear, concise, and complete. By completeness, we mean that it should be possible for someone to understand everything you have done, and why you have drawn the conclusions you have. There should be sufficient detail so that a dedicated reader could replicate your results and build on your work. By concision we mean that you should carefully select the material that best makes your case, including carefully designed and labeled tables or graphs, where appropriate. Additional technical details, including computer code, can be relegated to an appendix.

Tips: Start early, and have fun. If you consult the literature or use any outside resources, be sure to cite them appropriately. Converging on models, conclusions, and presentation is often an iterative process, and you should bring up questions sooner rather than later.

You should begin writing as soon as you start your task; you will find that the very act of writing down what needs to be done can help clarify your thinking. If you run into technical problems, e.g., data or coding glitches, ask for help immediately—don't wait until the last minute, since help may not always be available when you need it.

1 Project: Option hedging under transaction costs

We are hedging a European call option, taking into account that trading in the underlying stock involves proportional transaction costs. The goal is to optimize the tracking error, which is the difference between the value of the hedging portfolio and the payoff of the option at expiration, defined as:

$$\text{tracking error} = W_T - \max(S_T - K, 0) \tag{1}$$

where W_T is the post-liquidation value of the hedging portfolio. Specifically, assume that at time T all stocks are sold and the value of the portfolio W_T is measured after the sale.

We define the objective function as a negative exponential utility, so the goal is to maximize the expected value of

$$-e^{-\gamma(W_T - \max(S_T - K, 0))} \tag{2}$$

under the Equivalent Martingale Measure (in other words, we want to maximize the risk-neutral expectation of the objective).

Assume the Black-Scholes-Merton model: the interest rate r is constant, and the price of the stock (which pays no dividends) follows a Geometric Brownian motion process under the risk-neutral measure,

$$\frac{dS_t}{S_t} = r\,dt + \sigma\,dZ_t^Q$$

- Assume the following benchmark parameter values:

$$T = 22/252, \quad r = 0.05, \quad \sigma = 0.25, \quad \gamma = 1$$

$$S_0 = 100, \quad K = 100$$

- Assume that every time you buy or sell shares of the stock, you must pay transaction fees proportional to the dollar value of the traded position. Let λ denote the proportional transaction cost rate, e.g., when you sell \$100 worth of stock shares, you collect $(1-\lambda)$\$100. When you buy \$100 worth of stock shares, you pay $(1+\lambda)$\$100. The benchmark value for λ is 0.0005.
- You start with W_0 in cash and no stock holdings.

Your goal is to solve this problem numerically using Dynamic Programming (you should recursively solve for the optimal portfolio adjustments at each time step, taking transaction costs into account) and to investigates the properties of the solution:

1. Construct a binomial return process as a discrete-time approximation to Geometric Brownian motion, consistent with the Black-Scholes framework (e.g., use the Cox-Ross-Rubinstein approximation scheme). Assume trading occurs 10 times per day, over 22 days. Additionally, assume that prices do not jump overnight, meaning that the opening price on any given day is the same as the closing price on the previous trading day. Perform your analysis under the assumption that stock returns follow this binomial distribution.

2. Compute the optimal hedging strategy using Dynamic Programming. This computation should enable you to determine the hedging strategy for any initial capital W_0. Illustrate the optimal hedging strategy graphically (decide how to best visualize the results).

3. Consider a market maker with the objective defined above. Calculate the value of W_0 for which the market maker is indifferent between (i) selling the option at W_0 and dynamically hedging it until expiration, and (ii) receiving 0 at time T (not selling the option). How does W_0 compare to the no-arbitrage option price in a market with no transaction costs? Report option prices in terms of implied volatilities (treat W_0 as our estimate of the option price under trading costs).

4. Simulate the optimal hedging strategy and analyze its properties:

   (a) Plot the distribution of the hedging error (under the risk-neutral distribution) and compute its mean and standard deviation. Is the distribution symmetric around its mean? Does it appear Gaussian?

   (b) Change the risk aversion parameter γ and illustrate the resulting tradeoff between the mean and the standard deviation of the tracking error.

   (c) Under the optimal trading strategy, trades do not occur every period. What is the distribution of the (random) number of time periods between successive trades?

5. In the presence of transaction costs, consider the following heuristic strategy: calculate the hedge ratio as the standard Black-Scholes-Merton delta but rebalance at fixed intervals (e.g., once per day) instead of continuously. Simulate this hedging strategy and compare its performance to your optimal trading strategy. Determine the optimal rebalancing frequency for this heuristic strategy.

6. Investigate how your results depend on key parameters. Specifically, examine out-of-the-money options ($K > 100$) and vary the transaction cost parameter λ. Analyze how changes in each parameter affect both the characteristics of the hedging strategy and the distribution of tracking error.

7. Reflect on the effects of transaction costs on option hedging. Discuss specific trade-offs observed, such as the balance between hedging accuracy and cost minimization. Compare the performance of the optimal strategy with that of the heuristic strategy (under the optimal choices of the rebalancing frequency) across different transaction cost levels.

2 Project: Index replication

Our goal is to replicate a market index using a subset of securities.

We measure the quality of replication as follows. We rebalance monthly and compare returns on the two portfolios each month using the *tracking error*, TE. TE is the monthly difference in returns between the replicating portfolio and the market index: $TE_t = r_{port,t} - r_{mkt,t}$.

We intend our portfolio to be "green" and to be optimized to take into account rebalancing costs.

- Download the last 50 years of monthly returns on value-weighted industry portfolios. We are using 17 industry portfolios as defined by Fama and French, data is available at

  `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/17_Industry_Portfolios_CSV.zip`

  Our goal is to replicate return on a value-weighted stock market index, available from

  `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/F-F_Research_Data_Factors_CSV.zip`

  Note: Fama-French data includes excess market returns. Make sure to add back the risk-free rate to define market returns.

- We need to take into account several important constraints and frictions:

  1. We must avoid industries #2 and #3, which are "Mining and Materials" and "Oil and Petroleum Products", respectively.
  2. Portfolio weights on Transportation (Industry #13) and Utilities (Industry #14) should not exceed 1% each; Machinery and Business Equipment (#11) should not exceed 8%.
  3. Portfolio weights must be nonnegative and must add up to 100%.
  4. The weight on each industry cannot exceed 20%.
  5. Rebalancing the portfolio generates transaction costs. Assume these costs equal 0.10% of the dollar amount traded (they apply both to buy and sell trades). Portfolio must be self-financing, so trading costs need to be paid from the portfolio.

Your goal is to design a low-cost index-replicating portfolio. Formulate your optimization problem using a hierarchical objective: the primary objective is to minimize the mean-squared value of the tracking error, $E[TE^2]$, the secondary objective is to minimize transaction costs.

1. Assume that you know the covariance matrix of the joint distribution of returns on the industry portfolios and the market index, and their expected returns. Formulate the rebalancing problem and code it up using Gurobi. Make any necessary decisions, e.g., how to define the objective function for your optimization problem.

2. Next, conduct a back-test.

   (a) Label the first month of our data as $t = 1$. Starting at month $t = 120$, create a backward-looking window containing 10 years of monthly returns. For $t = 120$, this window covers $t = 1, 2, ..., 120$. To estimate the covariance matrix of returns, use the historical covariance over the 10-year window. Use historical averages to estimate expected returns.

   (b) Compute the optimal rebalancing trade at $t = 120$ and evaluate the resulting TE at $t = 121$. Slide the window forward one month and repeat until you reach the end of the sample.

3. Analyze the distribution of the monthly tracking errors and the second moment over the entire back-test period. How does the squared TE compare to its expected value implied by optimization? Do you see any systematic discrepancies? Is there a time-trend in the magnitude of the tracking error? Discuss your results.

4. How can you improve your results?

   (a) Try alternative estimations for the covariance matrix and expected returns. The goal is to better approximate the conditional distribution of industry returns.

   (b) Consider incorporating information on relative size of different industries. The file with industry returns contains information on the number and average size of firms in each industry. You should decide how to adjust problem formulation to improve your strategy.

5. How reliable are the results of the back-test? Conduct the following cross-validation experiment – repeat the following exercise 1,000 times.

   Create an artificial series of returns on the industry portfolios and the market index. For this purpose, randomly draw with replacement 24-month long blocks of monthly data. You should choose the beginning of the block randomly and preserve the order of months within each block. You are supposed to draw all industry returns and the market return for each month as a vector, so your simulation preserves the dependencies between monthly returns within the same month. This approach is known as block-bootstrap.

   Apply your back-test code to each simulated sample. For each sample, estimate the standard deviation of the tracking error and trading costs. Analyze the distribution of your estimates across the 1,000 samples. What can you say about the accuracy of the historical back-test results?

3 Project: Optimal trading under dynamic margin constraints

In this project, we study trading under dynamic margin constraints and market impact. The objective is to design an optimal long-short trading strategy for a mean-reverting spread between two assets. The goal is to maximize a multi-period objective function, given by

$$\mathbb{E}[-e^{-\gamma W_T}],$$

where W_T represents wealth at the final period T and γ controls risk aversion.

Consider the process X_t, representing the spread between the two assets:

$$X_{t+1} = \rho X_t + \sigma \varepsilon_{t+1}, \quad 0 < \rho < 1, \quad \varepsilon_{t+1} \overset{IID}{\sim} \mathcal{N}(0, 1).$$

Each period, we choose θ_t, the position size in the spread. The evolution of wealth W_t is defined as:

$$W_{t+1} = W_t + \theta_t (X_{t+1} - X_t), \quad t = \{1, 2, \ldots, T\}.$$

If W_t ever becomes negative, your position gets liquidated and can no longer trade.

Suppose that our position is large enough relative to the depth of the relevant market to significantly impact the spread when we trade. In this case, the spread dynamics become:

$$X_t = \rho X_{t-1} + \lambda(\theta_t - \theta_{t-1}) + \sigma \varepsilon_t, \quad \theta_0 = 0, \quad \lambda > 0.$$

Additionally, our position at each time t is constrained by margin requirements, and we must reduce our position if W_t falls too low:

$$|\theta_t| \leq \frac{1}{m_t} W_t.$$

Finally, we define the dynamics of the margin constraint. It is reasonable to assume that in extreme conditions—when X_t is relatively large—margin requirements become stricter. For instance, providers of stocks for short positions may require more collateral during high-volatility periods. We model this as:

$$m_t = \overline{m} + \delta X_{t-1}^2.$$

m_t is observed before choosing θ_t.

We make the following baseline assumptions on model parameters:

$$W_0 = 1, \quad X_0 = 0, \quad \gamma = 1, \quad \overline{m} = 0.25, \quad \lambda = 0.005,$$

$$T = 60, \quad \sigma = 0.01, \quad \rho = 0.95, \quad \delta = 1$$

1. Formulate the problem as a Markov Decision Process by defining the state space, action space, transition rules, and reward function. Write down the corresponding Bellman equation.

2. Since we know the reward function and the transition probability density, we should be able to solve the Bellman equation and determine the optimal action for every state. Assume that $\lambda = 0$ and $\delta = 0$. Discretize the state space and find the optimal policy $\theta_t^*(X_t, W_t)$ via Dynamic Programming. Visualize your results by selecting specific values for t and W_t and plotting $\theta_t(X_t)$.

   Hint: Instead of choosing θ_t directly, it may be simpler to choose $a_t \in [-1, 1]$ such that $\theta_t = a_t \frac{W_t}{m_t}$.

3. Now, return to the original values of λ and δ. Approximate the solution using Reinforcement Learning(feel free to use the code for examples covered in class as your starting point).

4. Plot the training curves. Investigate and illustrate graphically how the action θ_t depends on W_t, X_{t-1}, θ_{t-1}, and m_t. How does the optimal policy depend on parameters λ and δ? Discuss your findings.

5. Consider a "spiral" scenario: starting with an aggressive position close to the margin constraint bound $\rightarrow$ an unfavorable shock causes W to drop $\rightarrow$ to meet margin constraints, the position must be unwound, which further drives X against us $\rightarrow$ W drops again, requiring further unwinding, and so on. Do you see such scenarios in simulations, how would you identify them (e.g., you may look for repeated margin-driven position cuts)? What factors increase the likelihood and severity of a spiral scenario?

6. Short squeezes and predatory trading both exploit vulnerable market positions. Link these real-world phenomena and the dynamics of your spiral scenarios. What lessons can you draw from your analysis for designing resilient trading strategies?