Outline

- Transaction costs in asset management
- Multi-period models
- Practical approaches
- Estimating inputs for multi-period models
A quantitative portfolio manager seeks to find the optimal trade-off among three competing concerns:

- Maximize expected portfolio return
- Minimize portfolio risk (in absolute or relative terms)
- Minimize trading costs
Mean Variance Optimization

Notation

- $n$ asset classes (or stocks),
- expected returns given by the vector $\alpha$ and covariance matrix by $\Sigma$.
- A portfolio of the available asset classes is denoted by the vector $x = (x_1, x_2, \ldots, x_n)$.

Generic MVO

Representing portfolio constraints in the generic form $x \in \mathcal{X}$, we have a simple optimization problem:

$$\max \alpha^T x \text{ s.t. } x \in \mathcal{X}, x^T \Sigma x \leq \sigma^2$$

This is one of the three alternative formulations of Markowitz’ mean-variance optimization (MVO) problem.
Trading Costs

- Processing costs (commissions and taxes)
- Bid-ask spread
- Market impact costs

The first two tend to be proportional to the size of the trade (linear growth), with the possible exception of per-trade costs. Unit market impact costs typically grow with the size of the trade which translates into superlinear growth for total market impact costs.
Processing Costs

- The order processing costs are all costs explicitly incurred to accomplish the transaction.
- Processing costs include commissions paid to brokers, as well as taxes and other fees. These costs are often the smallest component of the transactions costs and typically easy to measure.
- "Stamp taxes" in the UK and some other countries can be significant.
- For U.S. institutional investors, an average number for order processing costs is around a few pennies per share. Typically higher for individual investors.
Market Impact Costs

- The current list of purchase (bid quotes) and sale (ask quotes) offers make up the quote depth.
- Typically, these offers are made with respect to limited quantities of the stock.
- Once the quote depth is depleted, the price of the stock will move and additional transactions can not be performed at the same price. This is the *market impact* of this trade.
- Every trade alters the market. Therefore we can not know the market impact prior to the trades.
Market Impact Costs

- Market impact is a hidden cost. Since it is the difference between the transaction price and what the price would have been had there not been a transaction, it can not be directly observed or measured.
- For an institutional investor, the market impact is typically the largest component of the transactions costs and also the hardest to estimate.
- One may call the change in the stock price that is beyond the half-spread the *incremental market impact*.
- The combination of the half-spread and the incremental market impact can be several multiples of the processing costs.
Are Transaction Costs Significant?

• Assume the following scenario: We manage a portfolio with 100% annual turnover (on average, each stock will be bought and sold once each year), average share price of $50, and an average total return of 8%.

• Some optimistic estimates for transaction costs per share traded:
  - a couple of cents for processing costs
  - a couple cents for half-spread
  - 5-10 cents incremental market impact

This translates to 10-15 cents total cost per share traded. Is this significant enough to worry about?

• The average total return per share is $4, while the average transactions costs is 20-30 cents (for purchase and sale), roughly 5-7% of the total stock return!
Are Transaction Costs Significant?

- Let us consider instead a market-neutral hedge fund manager targeting 10% volatility with a Sharpe ratio of 1 (so, expected return is also 10%).
- For a market-neutral hedge fund to reach a volatility of 10% a significant level of leverage is required. For $100 investment, positions worth $500 long and $500 short is not uncommon or extreme.
- Hedge funds also turn over portfolios frequently. 1000% turnover is not unusual.
- Even if we assume transaction costs are much lower than what we assumed previously, say 3 basis points (0.03% of dollars traded), annualized transaction costs amount to:
  
  \[0.03\% \times 2 \times 10 \times 10 = 6\%\].

- This is a very significant handicap to overcome for a fund targeting 10% return.
- Conclusion: Transaction costs can potentially erode a large part of an active managers value added.
A Model for Market Impact Costs

Based on Almgren, Thum, Hauptmann and Li (2005) whose main feature is that it explicitly and separately estimates the permanent \( I_{t}^{perm} \) and temporary \( I_{t}^{temp} \) market impacts for each order of \( x_{i} \) shares of stock \( i \):

\[
I_{t}^{perm}(x_{i}) = \gamma \cdot T \cdot \sigma_{i} \cdot \text{sign}(x_{i}) \cdot \left| \frac{x_{i}}{V_{i} \cdot T} \right|^{\alpha} \cdot \left( \frac{\Theta_{i}}{V_{i}} \right)^{\delta} + \varepsilon_{i}^{perm}
\]

\[
I_{t}^{temp}(x_{i}) = \eta \cdot \sigma_{i} \cdot \text{sign}(x_{i}) \cdot \left| \frac{x_{i}}{V_{i} \cdot T} \right|^{\beta} + \varepsilon_{i}^{temp}
\]

where

- \( V_{i} \) is the stocks average daily volume,
- \( \sigma_{i} \) is the one-day standard deviation of the stocks return,
- \( \Theta_{i} \) is the number of outstanding shares of stock \( i \),
- \( T \) is the fraction of the day over which the trade is executed, and
- \( \varepsilon_{i}^{perm} \) and \( \varepsilon_{i}^{temp} \) are unexplained residual terms.

The dimensionless term \( \frac{\Theta_{i}}{V_{i}} \) in the formulation of the permanent impact costs measure the fraction of the companys value traded each day and, as such, is a measure of relative liquidity of the stock.
A Model for Market Impact Costs

Using a large set of trades, the cross-sectional model parameters \( \alpha, \beta, \gamma, \delta, \) and \( \eta \) can be estimated, giving the following qualitative results.

- First, permanent impact cost is linear \((\alpha \approx 1)\) in trade size.
- Second, \( \beta \approx 1/2 \) meaning that the temporary impact cost is roughly proportional to the square root of the fraction of volume represented by one’s own trading during the period of execution. Hence, for a given rate of trading, a less volatile stock with large average daily volume has the lowest temporary impact costs.

Using these observations and after including linear costs (commissions and bid-ask spread) we arrive at a transaction cost function of the following form:

\[
TC_i(x_i) = a_i \cdot |x_i| + b_i \cdot |x_i|^{3/2} + c_i \cdot x_i^2.
\]
Transaction Cost Functions

Nonlinear Market Impact Costs

Unit transaction cost (in basis points)

Trade Size as a Percentage of Average Daily Volume
Incorporating the transaction cost function into the objective using a transaction cost aversion coefficient $\phi$, we arrive at the following generalization of MVO:

$$\max \alpha^\top x - \phi \cdot \sum_{i} TC_i(x_i - x_i^0) \text{ s.t. } x \in \mathcal{X}, x^\top \Sigma x \leq \sigma^2.$$ 

Above, $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$ is the vector of initial holdings. Or, if there is a benchmark with weights $x_B$, we may instead solve

$$\max \alpha^\top x - \phi \cdot \sum_{i} TC_i(x_i - x_i^0) \text{ s.t. } x \in \mathcal{X}, (x - x_B)^\top \Sigma (x - x_B) \leq \sigma^2$$
Problem with Single Period Models

• Consider a very simple but extreme example: We plan to find an optimal portfolio of two stocks for the next several investment periods.
• Our risk and transaction cost estimates are identical for these two stocks and we expect stock 1 to over-perform slightly in odd-numbered periods and stock 2 to over-perform slightly in even-numbered periods.
• Starting from a cash-only portfolio, for a low level of risk-aversion, an optimal solution that ignores future transactions costs would allocate 100% to stock 1 in period 1.
• Period 2 solution would depend on the transaction cost aversion. With low aversion, we may switch the portfolio to 100% stock 2, with 100% turnover.
• In contrast, a high transaction cost aversion would cause us not to trade and hold the (suboptimal) stock 1.
• However, inclusion of future transactions costs in a multi-period model would allow us discover the stable optimal allocation of 50-50 in each stock.
Multi-Period Models

Some of the multi-period models come from optimal control theory, like the *linear-quadratic regulator*:

\[
J(s_0) := \min_{u_0, \ldots, u_{N-1}} \left\{ \sum_{t=0}^{N-1} (s_t^T Q s_t + u_t^T R u_t) + s_N^T Q s_N \right\}.
\]

And we get a ”nice” solution that may be hard to implement:

Thus, the optimal control at stage \( t \) is

\[
u^*_t = -(R + B^T K_{t+1} B)^{-1} B^T K_{t+1} A s_t = L_t s_t,
\]

where

\[
L_t = -(R + B^T K_{t+1} B)^{-1} B^T K_{t+1} A.
\]

Plugging this value of \( u^*_t \) in the above expression for \( J_t(s_t) \) we get

\[
J_t(s_t) = s_t^T Q s_t + s_t^T A^T K_{t+1} A s_t - s_t^T A^T K_{t+1} B (R + B^T K_{t+1} B)^{-1} B^T K_{t+1} A s_t = s_t^T K_t s_t,
\]

where

\[
K_t = Q + A^T (K_{t+1} - K_{t+1} B (R + B^T K_{t+1} B)^{-1} B^T K_{t+1} B) A.
\]
Practical Multi-Period Models

- When decisions are spread over several periods, there is more potential to think ahead and develop portfolios that will reduce future turnover and transactions cost.
- In general, among otherwise similar portfolio mixes, the optimizer will pick those that are more likely to be turned into an optimal mix for future periods.
- In this sense, multi-period models with transactions costs have similarities to robust optimization models.
- One of the most important aspects of building multi-period models is understanding how the return models will evolve in the future periods.
Dynamics of Return Models

- Expected returns are typically a composite of return predicting signals (multi-factor return models).
- These signals lose their value through time, some slowly (e.g., value signals), some faster (e.g., reversal signals).
- Expected return of a portfolio estimated at a rebalance point stays relevant only for a certain period and will likely be inaccurate once the information decays.
- Between rebalances, portfolios cease to be optimal and can often become severely sub-optimal. Portfolios must be rebalanced frequently to keep them close to being optimal.
- On the other hand, rebalancing portfolios frequently leads to higher turnover and incurs higher t-costs.
Information Decay
Garleanu & Pedersen


\[ r_{t+1} = Bf_t + u_{t+1} \]

where \( r \) is a vector of excess returns, \( B \) is a matrix of factor loadings (exposures), \( f \) is a vector of factor returns, and \( u \) is a vector of white noise.

Information decay (mean reversion):

\[ \Delta f_{t+1} = f_{t+1} - f_t = -\Phi f_t + \varepsilon_{t+1} \]

where \( \Phi \) is a matrix (typically diagonal) of mean-reversion coefficients.
Garleanu & Pedersen

Additional assumptions:

- Quadratic transaction costs (no linear term): \( TC(\Delta x_t) = \frac{1}{2} \Delta x_t^\top \Gamma \Delta x_t \)
- Transaction costs are proportional to risk: \( \Gamma = \lambda \Sigma \).
- No constraints.
- Utility function is a sum of discounted future period utilities.

Under these assumptions, a closed form solution is available. The solution has an intuitive interpretation:

- An “aim portfolio” is a combination of the current (t-cost unaware) optimal portfolio and the expected future such portfolios
- T-cost aware optimal portfolio is a combination of the current portfolio and the aim portfolio.
Panel A: Construction of Current Optimal Trade

- Old position $x_{t-1}$
- New position $x_t$
- Aim $t$ $E_t(aim_{t+1})$
- Markowitz $t$
They build a model using intuition developed from three hypothetical traders:

- The **ideal** trader: Optimizes utility with no transaction costs
- The **optimal** trader: Optimizes utility with transaction costs
  - The optimal trader tracks the ideal trader in a cost efficient manner.
- The **random** trader: Chooses a trading path randomly. Probability of a path is an increasing function of its utility.
  - For the random trader, the unknown portfolios in the future, $x_t$ are random variables.
  - Their distributions are determined by the previous state as well as the cost of transition.
  - The most likely course of action is to match the optimal trader.
Kolm & Ritter vs. Garleanu & Pedersen

• Removes the assumption on quadratic (and proportional to risk) transaction cost function. Can handle all convex and separable t-cost functions.
• Allows constraints on single asset positions and trades (but general constraints are not handled directly)
• No assumptions on the modeling of future expected returns, can vary in an arbitrary fashion
• Allows a time-varying term structure for covariance and t-cost as well
Modeling the Random Trader

The model for the random trader is a Hidden Markow Model:

- Coupled stochastic processes \((X_t, Y_t)\).
- \(X_t\) is Markov but unobservable (corresponds to the true optimal portfolio)
- \(Y_t\) is observable, contemporaneously coupled to \(X_t\)
  (corresponds to the ideal portfolio)

\[
p(x \mid y) = \prod_t p(y_t \mid x_t) p(x_t \mid x_{t-1})
\]
An Equivalence

Theorem

Given a utility function of the form

\[ u(x) = \sum_t \left[ x_t^T r_{t+1} - \frac{\gamma}{2} x_t^T \Sigma_t x_t - c_t(\Delta x_t) \right] \]  \hspace{1cm} (1)

there exists a Hidden Markov Model with observation sequence \( y_t \) such that

\[ \log[p(y|x) \cdot p(x)] = \kappa \cdot u(x). \]

This result indicates that the utility function of the form (1) is the log-posterior of some probability distribution up to a scalar.
Recall

\[ p(x|y) = \prod_t p(y_t|x_t)p(x_t|x_{t-1}), \]

or

\[ \log p(x|y) = \sum_t [\log p(y_t|x_t) + \log p(x_t|x_{t-1})]. \]

So, letting

\[ \alpha_t = \mathbf{E}[r_{t+1}], y_t = (\gamma \Sigma_t)^{-1} \alpha_t \]

and setting

\[ b(x_t, y_t) = \frac{\gamma}{2} (y_t - x_t)^\top \Sigma_t (y_t - x_t), \]

\[ c(x_{t-1}, x_t) = c_t(\Delta x_t) \]

produces the equivalence.

"In summary, mean-variance-cost optimization reduces to tracking the ideal sequence \( y_t = (\gamma \Sigma_t)^{-1} \alpha_t. "$
An Important Simplification

Aggregating the variables over time into \( x = (x_1, x_2, \ldots, x_T) \) and \( y = (y_1, y_2, \ldots, y_T) \), we can write the negative of the utility function as

\[
f(x) = b(y - x) + c(x)
\]

where \( b \) is nice and quadratic, and \( c \) is convex but possibly non-differentiable.

**Assumption:** The trading cost function \( c \) is separable across assets, that is \( c(x) = \sum_i c^i(x^i) \), where the terms in the summation are the total cost of asset \( i \)'s trading path.

This assumption allows one to solve the problem iterating over assets using the **blockwise coordinate descent** (BCD) algorithm:

1. Optimize \( f(x) \) over \( x^i \), holding the remaining variables fixed. Let \( \hat{x}^i \) denote the solution.
2. Update \( x \) using \( \hat{x}^i \).
3. Set \( i = i + 1 \) (or to 1, if \( i = N \)).
Multiperiod Optimization for a Single Asset

Separability assumption on the non-differentiable cost function ensures that a limit point of the BCD iterations is an optimal solution (not true in general without the assumption). This means that if we can solve the multi-period problem efficiently for a single asset, we can use the BCD algorithm for solving the multi-asset multi-period problem.

To solve the single asset problem, one can iterate over $\Delta x_t$, the trades in time period $t$. The utility function for the single asset is also a combination of a differentiable term and a separable (across time) term, and therefore the coordinate descent iterations converge to a solution:

- Optimize the utility for trade at time $t$, holding all other trades constant
- Future positions (and therefore the $b(y - x)$ term) depend on $\Delta t$ in a convex, differentiable way
- We can loop over $t$ until convergence.
The Algorithm

- An outer loop over the assets
- An inner loop over time periods
- Each optimization subproblem of the inner loop is of the form:

\[ u(x_{t,i}) = a_{t,i}x_{t,i} - b_{t,i}x_{t,i}^2 + c(x_{t,i} - x_{t-1,i}) \]  

Note that, single-variable convex optimization problem of the form (2):

- is very easy to solve (bisection, Newton’s method, Brent’s method, etc.)
- can handle rich family of cost functions \( c \) (e.g., a combination of non-smooth and nonlinear terms)
- can also handle position and trade limits directly
Is the algorithm efficient? We have a double loop over assets and time periods, so each subproblem has to be solved very efficiently for the overall method to be efficient.

- Given $a_{t,i}, b_{t,i}$ optimizing $a_{t,i}x_{t,i} - b_{t,i}x_{t,i}^2 + c(x_{t,i} - x_{t-1,i})$ is ”easy”
- The coefficients for period $t$ problem need to be computed from the original coefficients for periods $t$ through $T$. One can do this recursively in an efficient way.
- Also, one needs to be careful about platforms that perform poorly on loops.

A more general-purpose implementation is based on the Hidden Markov Model interpretation of the utility function and relies on Viterbi’s algorithm for finding most likely state sequence in a finite HMM.
Estimating Multiperiod Model Inputs

- A typical idea is to use the mean reverting factor return assumption in Garleanu & Pedersen. Then, the main task is estimating the factor decay (mean reversion) coefficients.

- One way to do this is to build a factor portfolio today, and measure its returns at different lags and compute the decay in the returns.

- Another approach is to build factor portfolios daily and look at the correlations between today’s factor portfolio and yesterday’s (or last week’s, etc.) factor portfolio.

- A special case: "2-12 momentum factors." A typical method of building momentum signals is to look at security returns over the last twelve months, excluding the last month to eliminate reversal effects.

  - In this case, next month’s 2-12 momentum signal is completely predictable today.