Trading Costs, Risk Forecasts

BUSI 722: Data-Driven Finance II

Kerry Back

Transaction Costs

Categories of Transaction Costs

- **Commissions and bid-ask spread:** broker fees are near zero today, but the bid-ask spread (half-spread per trade) remains a meaningful cost, especially for illiquid stocks.
- **Market impact:** large orders move the price against you. This is typically the dominant cost for institutional strategies.
- **Opportunity cost:** trades you cancel or delay because the market moved before you could execute.

Market Impact

Market impact grows with trade size relative to the stock's liquidity:

$$\text{impact} \;\approx\; c \times \sqrt{\frac{\text{shares traded}}{\text{average daily volume}}}$$

- The square-root law is a widely used approximation.
- Impact is larger for small, illiquid stocks.
- This creates a **capacity constraint**: strategies that trade heavily in small stocks face rapidly increasing costs as the fund grows.

Estimating Transaction Costs

A simple model for backtesting:

$$\text{cost}_t = \sum_{i=1}^{n} c_i \times |w_{i,t} - w_{i,t-1}|$$

- $|w_{i,t} - w_{i,t-1}|$ = the absolute change in weight (turnover) for stock i.
- c_i = estimated one-way cost for stock i (e.g., half the bid-ask spread plus estimated market impact).
- Common simplification: use a flat cost per unit of turnover (e.g., 5–20 basis points one-way).

Short-Selling Costs

What Does Shorting Cost?

To short a stock, you borrow shares and sell them. This incurs:

- **Borrow fee:** an annualized fee paid to the lender (typically 0.3–1% for easy-to-borrow stocks, but can be 10–50%+ for hard-to-borrow names).
- **Margin and recall risk:** you must post collateral, and the lender can recall shares at any time, forcing you to close the position.
- **Short-sale constraints:** some stocks are simply unavailable to borrow.

Implications for Long-Short Strategies

- ML models often find the **short side is more profitable** than the long side — but short-selling costs disproportionately affect small, illiquid stocks where signals are strongest.
- A strategy that looks great in a frictionless backtest may be mediocre or unprofitable after accounting for borrow fees.
- **Always evaluate long-only and long-short separately** to see where the value is coming from.

Net-of-Cost Evaluation

Net-of-Cost Portfolio Returns

The net portfolio return in month t:

$$r_{p,t}^{\text{net}} = r_{p,t}^{\text{gross}} - \text{transaction costs}_t - \text{borrow costs}_t$$

- Recompute Sharpe ratios, CAPM alphas, and information ratios using $r_{p,t}^{\text{net}}$.
- This is the **real test** of a strategy. Many academic strategies fail it.

Avramov, Cheng, and Metzker (*Management Science*, 2023):

- ML profitability concentrates in **hard-to-trade** stocks (microcaps, distressed firms).
- Excluding these stocks or adding realistic trading costs significantly reduces profitability.

Rebalancing & Weighting

Rebalancing Frequency

How often should we retrain the model and rebalance the portfolio?

Frequency	Pro	Con
Monthly	tracks signal closely	high turnover, high cost
Quarterly	moderate turnover	signal may be stale
Annually	low turnover	very stale signal

- The optimal frequency depends on the **decay rate of the signal** versus the **cost of trading**.
- Momentum decays quickly (monthly rebalancing is typical); value signals are slow-moving (quarterly may suffice).
- A hybrid approach: retrain monthly but only trade when the change in target weight exceeds a threshold (**buffer rules**).

Turnover

Turnover measures the fraction of the portfolio that changes each period:

$$\text{turnover}_t = \sum_{i=1}^{n} \left| w_{i,t} - w_{i,t-1}^{+} \right|$$

where $w_{i,t-1}^{+}$ is the weight of stock i at the end of the previous month (after returns).

- A turnover of 1.0 means the entire portfolio is replaced.
- A long-short decile portfolio rebalanced monthly can have turnover > 1.0.
- Smooth weight functions (linear, power) typically generate **lower turnover** than sort-based step functions, because small changes in predicted rank produce small changes in weight.

Small-Cap Exposure

Equal-weighted portfolios tilt heavily toward small stocks, which are harder to trade, where ML signals may be spuriously strong, and not investable at scale.

Solutions:

- **Value-weight within groups:** sort into deciles by predicted rank, then value-weight (by market cap) within each decile.
- **Score-tilted market-cap weights** (from Session 5): $w_i \propto \text{mcap}_i \times g(u_i)$. Stays close to the market-cap benchmark.
- **Market-cap filters:** exclude micro-caps or nano-caps entirely. Evaluate only on stocks you could actually trade.

Portfolio Construction Methods

Mean-Variance Optimization

Markowitz (1952): choose weights $\mathbf{w}$ to maximize

$$\mathbf{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$$

The **tangency portfolio** maximizes the Sharpe ratio: $\mathbf{w}^* \propto \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$

Pros:

- Theoretically optimal given correct inputs. Balances return and risk.

Cons:

- Extremely sensitive to estimation errors in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
- Produces extreme, concentrated weights.
- With noisy return forecasts, often **worse** than $1/N$ out of sample.

Minimum-Variance Portfolio

Set $\boldsymbol{\mu} = \mathbf{0}$ and solve only for the lowest-risk portfolio:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \; \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \quad \text{s.t.} \quad \mathbf{w}'\mathbf{1} = 1$$

Pros:

- Requires **only the covariance matrix** — no return forecasts needed.
- Covariances are more stable and easier to estimate than expected returns.

Cons:

- Concentrates in low-volatility stocks (can be extreme without constraints).
- Still sensitive to estimation error in $\boldsymbol{\Sigma}$ — use shrinkage estimators (Ledoit-Wolf).

Risk Parity

Equalize each asset's **contribution to total portfolio risk**:

$$w_i \times (\boldsymbol{\Sigma}\mathbf{w})_i = \frac{1}{n}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \quad \text{for all } i$$

A simple approximation: $w_i \propto 1/\sigma_i$ (inverse-volatility weighting).

Pros:

- More diversified than minimum variance — no single asset dominates risk.
- Robust to estimation error; widely used in practice (e.g., Bridgewater's All Weather).

Cons:

- Ignores correlations in the simple version (inverse-volatility).
- Often requires leverage to achieve competitive expected returns.

Comparing Portfolio Construction Methods

	Tangency	Min-Variance	Risk Parity
Uses return forecasts	yes	no	no
Uses covariance	yes	yes	yes (or vol only)
Diversification	low	low	high
Est. sensitivity	very high	moderate	low
Out-of-sample Sharpe	often poor	competitive	competitive

- **If you trust your return forecasts:** tangency portfolio, but regularize heavily (shrinkage, constraints).
- **If you don't trust return forecasts:** minimum variance or risk parity.
- **Hybrid:** use predicted ranks for stock selection (which stocks to hold), and risk-based methods for position sizing (how much of each).

Covariance Estimation

Why Covariance Estimation Is Hard

All risk-based methods require a covariance matrix $\boldsymbol{\Sigma}$.

- With n stocks, $\boldsymbol{\Sigma}$ has $n(n+1)/2$ unique entries. For 500 stocks, that is **125,250 parameters** to estimate — far more than the number of monthly observations.
- The sample covariance **overfits**: its eigenvalues are too spread out, and some may be zero or negative.
- Portfolios optimized with a noisy $\boldsymbol{\Sigma}$ take extreme, unstable positions.

Ledoit-Wolf Shrinkage

Insight: the sample covariance is **unbiased but high-variance**. A structured target is **biased but low-variance**. Blending them reduces mean-squared error.

$$\boldsymbol{\Sigma}_{\text{shrunk}} = \delta\,\mathbf{F} + (1-\delta)\,\mathbf{S}$$

- $\mathbf{S}$ = sample covariance matrix (complex, noisy)
- $\mathbf{F}$ = **shrinkage target** — a simple, structured matrix
- $\delta \in [0,1]$ = **shrinkage intensity** — how much to shrink toward the target

The **bias-variance tradeoff**: $\delta = 0$ gives the noisy sample covariance; $\delta = 1$ gives the structured target; δ^* is the optimal blend minimizing estimation error.

Choosing the Shrinkage Target

Common choices for the target $\mathbf{F}$:

- **Scaled identity:** $\mathbf{F} = \bar{\sigma}^2 \mathbf{I}$, where $\bar{\sigma}^2$ is the average diagonal of $\mathbf{S}$. Assumes all stocks have the same variance and zero correlation. Simple but aggressive.
- **Single-factor model:** $\mathbf{F} = \hat{\beta}\hat{\beta}'\sigma_m^2 + \mathbf{D}$, where $\hat{\beta}$ are market betas. Preserves the market factor structure. This is the default in `sklearn.covariance.LedoitWolf`.
- **Constant-correlation:** all pairwise correlations equal the average sample correlation. Preserves individual variances.

Ledoit and Wolf (2004) derive a **closed-form formula** for the optimal δ^* that minimizes the expected squared Frobenius norm of the estimation error.

Effect of Shrinkage

- **Eigenvalues compress:** extreme eigenvalues of $\mathbf{S}$ are pulled toward the mean, and spurious correlations moderate. The matrix becomes better-conditioned.
- **Portfolio weights stabilize:** minimum-variance portfolios computed with $\boldsymbol{\Sigma}_{\text{shrunk}}$ have less extreme, more diversified weights.
- **Out-of-sample risk improves:** the shrunk covariance produces portfolios with **lower realized volatility** than the sample covariance.

Ledoit-Wolf in Python

`scikit-learn` provides a one-line implementation:

```
from sklearn.covariance import LedoitWolf

returns = ...  # T x n array of stock returns
lw = LedoitWolf().fit(returns)

cov_shrunk = lw.covariance_    # shrunk covariance matrix
delta = lw.shrinkage_          # optimal shrinkage intensity
```

- The `shrinkage_` attribute reports the optimal δ^* — typically 0.5–0.9 for monthly stock returns, meaning heavy shrinkage is optimal.
- Use `cov_shrunk` anywhere you would use a sample covariance: minimum-variance portfolio, risk parity, mean-variance optimization.

Other Approaches to Covariance Estimation

- **Factor models** (Barra-style): $\mathbf{\Sigma} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}' + \mathbf{D}$. Reduces dimensionality by modeling common factors. Industry standard for large-scale risk models.
- **GARCH / DCC:** capture time-varying volatility and correlations. Useful when risk regimes change.
- **Exponential weighting:** weight recent observations more heavily. Simple, effective, and captures regime changes.

Practical advice: Ledoit-Wolf shrinkage is the easiest win. It requires no modeling decisions beyond choosing a target, and the improvement over the sample covariance is reliable. Start here, then consider factor models if you need more structure.

Drawdowns

What Is a Drawdown?

A **drawdown** is the decline from a portfolio's peak value to a subsequent trough:

$$\text{drawdown}_t = \frac{V_t - V_{\max,t}}{V_{\max,t}}$$

where $V_{\max,t} = \max_{s \le t} V_s$ is the running maximum of portfolio value.

- Always negative (or zero at a new high).
- The **maximum drawdown** is the largest peak-to-trough decline over the sample.
- Drawdown **duration** measures how long it takes to recover to the previous peak.

Why Drawdowns Matter

- The Sharpe ratio treats upside and downside volatility equally. Investors don't.
- A strategy with a high Sharpe ratio but a 60% maximum drawdown is psychologically and practically devastating — investors redeem, managers get fired.
- Drawdowns capture **path-dependent risk** that summary statistics miss.

Reporting:

- Plot the **underwater chart**: drawdown over time. Shows how often and how deeply the portfolio falls below its peak.
- Report maximum drawdown and maximum drawdown duration alongside Sharpe ratios and alphas.

Drawdowns and Strategy Evaluation

- Compare drawdowns across different weight functions, rebalancing frequencies, and cost assumptions.
- A strategy that looks great on Sharpe ratio but has deeper drawdowns than a simpler approach may not be worth the complexity.
- Long-short strategies can have severe drawdowns during short squeezes; value-weighted portfolios typically have **shallower drawdowns** (less small-cap exposure).

Key takeaway: always look at the full return path, not just summary statistics. The underwater chart is one of the most important diagnostic plots for any quantitative strategy.