Performance & Linear Factor Decomposition

FINM 36700 – Portfolio Risk Management (Lecture 2 TA Review)

Anand Nakhate

October 10, 2025

---

1 Linear Factor Decomposition (LFD)

Let $\tilde{r}_t^i$ be the *excess* return of asset/portfolio i at time t. For factor returns $\tilde{x}_t$,

$$\tilde{r}_t^i \;=\; \alpha + \boldsymbol{\beta}^\top \tilde{x}_t + \varepsilon_t, \tag{1}$$

- The goal is to *decompose* variation in $\tilde{r}^i$ into an explained component $\alpha + \boldsymbol{\beta}^\top \tilde{x}_t$ and a residual component ε_t.
- We are just trying to decompose a portfolio returns into some building blocks. It is just a linear statistical projection.
- We can also decompose on things that are not investable to understand the dynamics theoretically. Decomposing on investable assets allows us to hedge, replicate or track investments.
- The factors in the LFD should give a high R^2 in the regression if they really explain the variation of returns well.

1.1 What is LFD *not* about

- Not a no-arbitrage or *pricing* model (e.g., CAPM/Fama-French). It is a statistical *projection.*
- α is model-dependent: a large $\hat{\alpha}$ may reflect skill *or* missing betas in x_t.

1.2 Assumptions

- The only requirement for linear factor decomposition is that the factor matrix need to be a full column rank: $X = [\mathbf{1}, x]$ has rank $k+1$ (no perfect multicollinearity).
- We need X'X to be invertible.
- Even though we don't need a lot of assumptions, that doesn't mean we get precise estimates.

1.3 LFD vs. OLS / BLUE

- **Population projection** defines $\alpha, \boldsymbol{\beta}$. **OLS** provides their *sample* estimators $\hat{\alpha}, \hat{\boldsymbol{\beta}}$ by minimizing $\sum_t \hat{\varepsilon}_t^2$.
- Under Gauss–Markov (linear model with homoskedastic, uncorrelated errors), OLS is *BLUE* (Best Linear Unbiased Estimator). In financial time series, heteroskedasticity and autocorrelation are common. OLS remains unbiased under exogeneity but no longer efficient.

1.4 What are the factors x?

Organize x_t so betas are interpretable and implementable. We might want to understand how an investment behaves like.

- **Asset-class/market:** broad equity (S&P 500), small caps (Russell 2000), EAFE, EM, rates (duration), credit, commodities.
- **Region/sector/industry:** US vs. DM ex-US; tech/healthcare/energy; etc.
- **Styles/risk premia:** value, growth, size, momentum, quality, low-vol.
- **Macro/FX:** USD/EUR, carry, term structure slope, inflation proxies.

1.5 Estimation and interpretation

OLS: $\hat{\theta} = (X^\top X)^{-1} X^\top y$.

$$\begin{aligned}
\text{Alpha } \hat{\alpha} &: \text{expected non-factor component (model-dependent).} \\
\text{Betas } \hat{\boldsymbol{\beta}} &: \text{marginal sensitivities to factors (holding others fixed).} \\
\text{Residuals } \hat{\varepsilon}_t &: \text{idiosyncratic/basis component; } \hat{\sigma}_\varepsilon = \sqrt{\tfrac{1}{T-k-1} \sum \hat{\varepsilon}_t^2}. \\
R^2 &: \text{share of variance explained by } x_t.
\end{aligned}$$

Collinearity. Highly correlated factors make $X^\top X$ ill-conditioned, inflating standard errors and destabilizing $\hat{\boldsymbol{\beta}}$. Remedies: factor grouping, orthogonalization within families, ridge/lasso, constraints when implementing.

2 Use Case I — Performance Attribution and Replication

2.1 Attribution (what drove returns?)

Estimate Linear Factor decomposition with a defensible factor set. Report $(\hat{\alpha}, \hat{\boldsymbol{\beta}}, R^2, \hat{\sigma}_\varepsilon)$ and risk-adjusted statistics:

$$\text{Treynor} = \frac{\mathbb{E}[\tilde{r}^i]}{\beta_{i,m}}, \qquad \text{Information Ratio (IR)} = \frac{\hat{\alpha}}{\hat{\sigma}_\varepsilon}.$$

Interpret $\hat{\alpha}$ with caution (missing betas, small-sample luck). Use HAC t-stats and out-of-sample checks.

Ex:1 - Hedge funds might invest in a lot of assets - various pods have different strategies but might be exposed to similar assets. LFD can help them get the next exposure to the factors.

Ex: 2: Some funds might label themselves following a specific style of investing. We can run an LFD using factors based on their investment thesis and understand if they are doing what market and understand how well can they select within the factors.

2.2 Replication (turn exposures into a portfolio)

Two practical modes:

1. **Unbudgeted Replication (include α in estimation).** Use $\hat{\boldsymbol{\beta}}$ as notional sleeves, and accept a drift $\hat{\alpha}$ you cannot directly buy.
2. **Budgeted TE-min Replication (enforce span, often omit α).** Given factor covariance $\Sigma_x = \mathrm{Var}(x_t)$ and $c = \mathrm{Cov}(x_t, \tilde{r}_t^i)$,
$$\min_w \mathrm{Var}(\tilde{r}_t^i - w^\top x_t) \quad \text{s.t.} \quad \mathbf{1}^\top w = 1$$
has solution
$$w^\star \;=\; \Sigma_x^{-1} c \;-\; \Sigma_x^{-1}\mathbf{1} \cdot \frac{\mathbf{1}^\top \Sigma_x^{-1} c - 1}{\mathbf{1}^\top \Sigma_x^{-1} \mathbf{1}}.$$
Track out-of-sample tracking error (TE), drift, and stability.
3. "Unbudgeted replication with intercept" imply a non-investable drift. budgeted replication without intercept enforce span and typically lower R^2 but are tradeable.

3 Use Case II — Hedging

3.1 Basis risk and optimal hedge

Long \$1 of i and sell h dollars of j. Time t exposure and risk:
$$\varepsilon_t \equiv r_t^i - h\, r_t^j, \qquad \mathrm{Var}(\varepsilon_t) = \sigma_i^2 + h^2\sigma_j^2 - 2h\,\sigma_i\sigma_j\rho_{ij}.$$
Here, ε_t is the basis and the volatility of ε_t is the *basis risk*. We can see that if $\rho_{i,j} = 1$ or $\rho_{i,j} = -1$, the basis risk is eliminated. Otherwise, we minimize the basis risk by adjusting h.

In a more general sense, basis is the difference between two (or more) instruments. It has a meaningful interpretation when the instruments are correlated. It is defined as the risk held after hedging one instrument with another.
The h that minimizes it is
$$h^* = \frac{\mathrm{Cov}(r^i, r^j)}{\mathrm{Var}(r^j)} = \frac{\sigma_i}{\sigma_j}\rho_{ij}.$$
For 2 instruments, it is just the univariate beta.

- Position is perfectly hedged if $\mathrm{P}(\varepsilon = 0) = 1$
- Higher correlation implies a larger hedge ratio h, as the assets align more, making the hedge more effective.
- Higher relative volatility of "i" implies a larger h.
- Negative correlation requires going long on the hedging security.

3.2 Hedging Multiple Factors

When we have multiple regressors, the result becomes a regression where each asset representing a risk is a regressor. $r_t^i = \sum_k \beta_{i,k} r_t^k + \varepsilon_t$

The basis becomes the net return exposure, and the absolute value of the optimal hedge is determined by the betas in the return regression above. The variances and correlations between assets change daily, so optimal hedges must adjust accordingly.

With many hedges, the OLS betas from $r_t^i = \sum_k \beta_{i,k} r_t^k + \varepsilon_t$ are the optimal hedge ratios.

3.3 Excess returns and the intercept

For excess returns, $\tilde{r}_t^i = \alpha + \sum_k \beta_{i,k} \tilde{r}_t^k + \varepsilon_t$. Including an intercept focuses betas on matching *variation*. Omitting α enforces that mean differences are absorbed by factor weights (span), but can degrade fit when sample means are noisy. Choice is dependent on the goal

If I exclude an intercept, I get a truer picture of the situation. For example, when hedging stock with the S&P, I hedge with instruments. Including α in the regression adds a part of the hedge that cannot be bought — α cannot be "bought."

Why might it be reasonable to include an intercept?

- Including an intercept lets the betas focus on matching return variation, not just the level. If α is excluded, betas also adjust the magnitude.
- Including α focuses βs on matching variation (demeans both sides), avoiding loading βs with noisy mean differences. Excluding α forces level-matching (span) and often degrades fit in short samples.
- The key point is whether we expect the difference in mean returns to persist out-of-sample.
- Historical data often provides uncertain estimates of mean returns, so excluding α may lead betas to "predict" differences in mean that aren't predictive of the future.

3.4 Hedged position properties and metrics

Hedging allows an investor to take a position on a specific thesis without exposing themselves to unwanted risks. If an investor has a view about Google and doesn't want to get exposed to the market risk, by hedging with the S&P, the investor's performance depends only on Google's relative performance. It is important to note that hedging might also introduce unwarrented risks from the hedged position.

This idea of trading on specific information while hedging out broader market movements is the origination of the term, hedge funds.

A market-hedged position satisfies $\tilde{r}_t^i - \beta_{i,m} \tilde{r}_t^m = \alpha + \varepsilon_t$: mean $= \alpha$, volatility $= \sigma_\varepsilon$. For hedges, the scorecard is *large α per unit* basis risk, i.e., a high $= \alpha/\sigma_\varepsilon$, and small factor leaks (high R^2).

4 Use Case III — Tracking

If the target i is not directly investable, estimate $\tilde{r}_t^i = \alpha + \beta\, \tilde{r}_t^j + \varepsilon_t$ (or multi-factor) and buy weights $\hat{\boldsymbol{\beta}}$ in the RHS to mimic i.

ε_t is known as the tracking error of i relative to j. R^2 measures how well j tracks i. The Information Ratio here measures the tradeoff between obtaining extra mean return at the cost of taking on tracking error from the target portfolio.

For broad market factors, mutual funds are often tracking some factor while hedge funds are trying to hedge it out (Not always true).

Metrics: Tracking error = $\hat{\sigma}_\varepsilon$ (In some use cases); R^2. Drift $\hat{\alpha}$. For tracking, aim for $\alpha \approx 0$ and small TE out-of-sample. For benchmark tracking, IR typically refers to $\frac{\mathbb{E}[r_i - r_j]}{Std(r_i - r_j)}$. With an LFD including α, the numerator equals to α

5 Hedging vs. Tracking vs. Replication

	Hedging	**Tracking**	**Replication**
Invest on LHS?	Yes	No	No
Trade the RHS?	Sell $\hat{\boldsymbol{\beta}}$	Buy $\hat{\boldsymbol{\beta}}$	Buy $\hat{\boldsymbol{\beta}}$
Primary risk metric	Basis risk σ_ε	TE σ_ε	TE σ_ε
Alpha target (OOS)	Big α, small σ_ε	$\alpha \approx 0$	Small drift/TE
R^2 role	High R^2 desired	High R^2 desired	High R^2 desired
Intercept in estimation	Usually include	Usually include	Often omit (span)

6 When to Include an Intercept

Mean returns are not very well estimated. It might not be wise to give β additional responsibility to match the level apart from variation when the mean returns are not well estimated in the first place. There is not right or wrong way to do this, but a trade off.

- **Include** α to explain *variation* (attribution/tracking) and to avoid loading betas with noisy mean differences in short samples.
- **Omit** α for a fully spanned *implementable* replication. Accept lower R^2 and potential level mismatch.

7 Evaluating Performance: Timing, Selection, Luck

Selection vs timing With a constant-beta LFD, $\hat{\alpha}$ captures average outperformance *beyond* chosen betas (selection). Allowing β_t to vary creates a timing component (covariance between time-varying beta and factor returns). In short samples, both are hard to separate—hence the need to report IR and R^2, and to validate out-of-sample.

Luck vs skill. Alpha estimates are noisy. Large in-sample $\hat{\alpha}$ can be luck. Use HAC t-stats, stability checks, and OOS IR.