Algorithmic Trading & Quantitative Strategies

Lecture 6 (4/16/2024)

Giuseppe Paleologo (gardener)

Today's Session

- There will be a short multiple-answer quiz at the end of the lecture.
- Please send answers in the form
  - 1-a
  - 2-c
  - ...

To gp2642@nyu.edu. You will have 5 minutes

Recap on Last Quiz

1. In Fundamental Models, you can't use directly the following data
   a. Company trailing earnings
   b. Company forward earnings
   c. **Inflation rate (CPI)**
   d. Trailing 10-day returns
2. Is it better to have a peaky spectrum?
   a. Yes because there are fewer eigenvalues
   b. **Yes because they are more separated from the bulk**
   c. Not necessarily
3. Bulk eigenvalues correspond to
   a. Idiosyncratic variances
   b. **Factor variances**
   c. A combination of the two
4. Should you shrink the factor covariance matrix?
   a. Yes, you should almost shrink a covariance matrix
   b. **Yes, we have not enough observations**
   c. Yes, because of spikes
5. Should you use a non-quadratic loss
   a. Yes. It could be more robust to outliers
   b. **No. Mismatch with factor returns**
   c. Maybe. For example penalties
6. Dynamic Volatility Adjustment
   a. Only makes the model more responsive
   b. **Makes the model more predictive of risk**
   c. Makes the model more responsive but not more predictive

Topics

- Fundamental Model: residual variance estimation
    - Practical considerations: z-scoring and reducing turnover
- Advanced Portfolio Construction
- Next Time (final lecture):
    - Backtesting

Residual Volatilities

Do not reinvent the wheel (mostly). We have seen the bassic ingredients before. Steps:

1. EWMA
2. Ledoit-Wolf Shrinkage
3. **Short-Term Volatility Updating**
4. **Cluster Analysis**

Short-Term Idio Updating

- You can actually use exactly the same approach as factors, with identity correlation. It still works
- But it works much better if you include in the model the earning dates and introduce an asset weight that gets smaller the farther away from the earning date:

$$a_{i,t} = \begin{cases} 1 - |t - T_{\text{earn},i}|/\tau_{\text{earn}} & \text{if } |t - T_{\text{earn},i}| \le \tau_{\text{earn}} \\ 0 & \text{otherwise} \end{cases}$$

STIU

- And then the same machinery but easier: measure the "cross-sectional vol surprise". Easier because C=I

$$e^{\hat{x}_t/2} = \kappa_0 \sum_{s=0}^{\infty} e^{-s/\tau_0} \sqrt{\frac{\sum_i a_{i,t} (\epsilon_{i,t}/\hat{\sigma}_{i,t})^2}{\sum_i a_{i,t}}}$$

- And then update the affected vols:

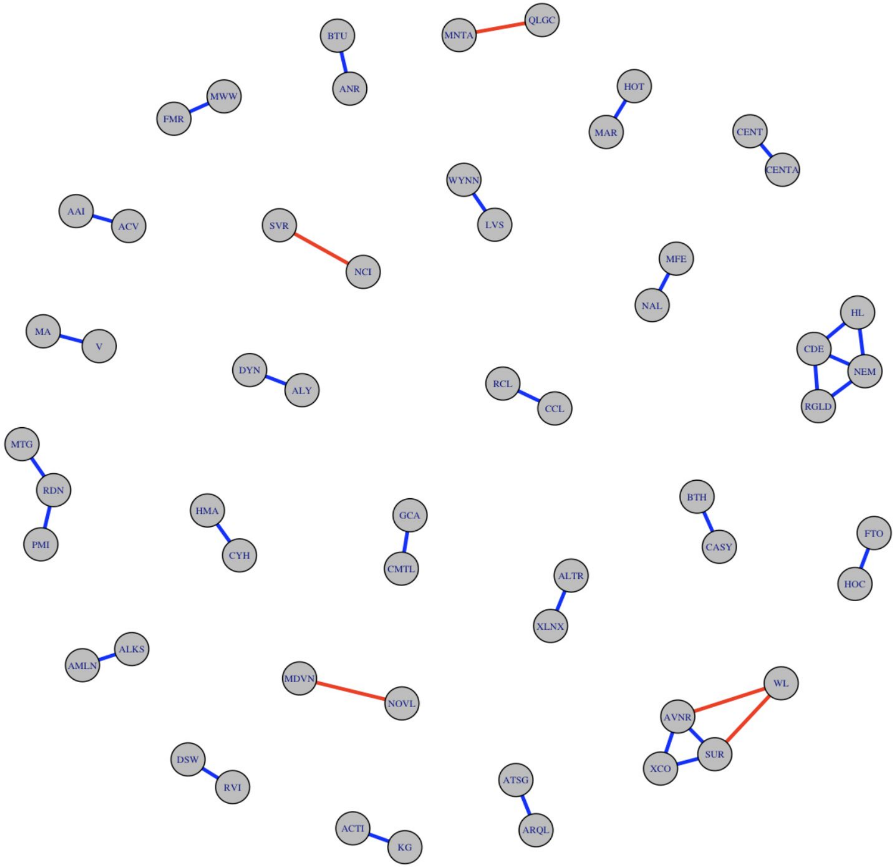
$$\hat{\sigma}_{i,t}^2 \leftarrow \left[(1 - a_{i,t}) + a_{i,t} e^{\hat{x}_t}\right] \hat{\sigma}_{i,t}^2$$

Clustering

- Approximate and exact models. The former have diagonal idio covariances
- Sparse is fine as well. A simple thresholding (Bickel-Levina, El Karoui 2008) will take you a long way.
- The theoretical thresholds (sqrt(log n/T)) don't work well in my subjective experience. You have to do a bit of trial and error.
- Usually clusters correspond to "insensitive region" of the threshold (makes sense), and the cluster elements are close in factor loading space.

Clusters and Sparsity

ticker	name	ticker	name
AAI	AIRTRAN HOLDINGS	HOT	STRW HTL RES WRLWD
ACTI	ACTIVIDENTITY CORP	KG	KING PHARMCUTCL
ACV	ALBERTO CULVER CO NEW	LVS	LAS VEGAS SANDS
ALKS	ALKERMES INC	MA	MASTERCARD INC
ALTR	ALTERA CORP	MAR	MARRIOTT INTL INC NEW
ALY	ALLIS CHALMERS ENERGY INC	MDVN	MEDIVATION INC
AMLN	AMYLIN PHARMACEUTICALS INC	MFE	MCAFEE INC
ANR	ALPHA NATURAL RES	MNTA	MOMENTA PHARMACEUTICALS INC
ARQL	ARQULE	MTG	MGIC INVT CORP WIS
ATSG	AIR TRANSPORT SERVICES GRP I	MWW	MONSTER WORLDWIDE INC
AVNR	AVANIR PHARMACEUTICALS INC	NAL	NEWALLIANCE BANCSHARES INC
BTH	BLYTH INC	NCI	NAVIGANT CONSULTING INC
BTU	PEABODY ENERGY	NEM	NEWMONT MINING CORP
CASY	CASEYS GEN STORES INC	NOVL	NOVELL INC
CCL	CARNIVAL CORP	PMI	PMI GROUP INC
CDE	COEUR D ALENE MINES CORP IDAHO	QLGC	QLOGIC CORP
RCL	ROYAL CARIBBEAN CRUISES LTD	RDN	RADIAN GROUP INC
CMTL	COMTECH TELECOMMUNICATIONS CP	RGLD	ROYAL GOLD INC
CYH	COMMUNITY HEALTH SYS INC NEWCO	RVI	RETAIL VENTURES INC
DSW	DSW-A	SUR	CNA SURETY CORP
DYN	DYNEGY INC DEL	SVR	SYNIVERSE HLDGS INC
FMR	FIRST MERCURY FIN	V	VISA INC
FTO	FRONTIER OIL CORP	WL	WILMINGTON TRUST CORP
GCA	GCA HLDGS	WYNN	WYNN RESORTS
HL	HECLA MNG CO	XCO	EXCO RESOURCES INC
HMA	HEALTH MGMT ASSOC INC NEW	XLNX	XILINX INC
HOC	HOLLY CORP		



How Do You Use This?

- For risk, obviously
  - Risk may be under/overestimated
- And for alpha
  - Clusters are sub-sub-sub...-industry. Themes, on which people bet
  - The assets are cointegrated. Identify rich and cheap assets

Advanced Portfolio Construction

- MVO is widely used. Also widely criticized, also the criticism is often a bit vague and qualitative. "Error maximization" and the like.
- Three questions:

1. What is the impact of error on input data?
2. Are constraints good or bad for your portfolio health
3. What constraints?

The Simplest Example

Two assets, Sharpes s_1, s_2, and correlation ρ

$$\mathbf{C}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$$

$$v_1^\star = \frac{\kappa}{1-\rho^2}(s_1 - \rho s_2)$$

$$v_2^\star = \frac{\kappa}{1-\rho^2}(s_2 - \rho s_1)$$

So simple and yet so rich:

1. Second asset has no alpha and positive correlation? It's a hedge
2. Second asset has alpha, but is positively correlated? Relative Value

Let's Introduce Errors

True Sharpe is s tilde. Assume $||\tilde{\mathbf{s}} - \mathbf{s}|| \leq \epsilon$

How much do we loss in *realized* sharpe?

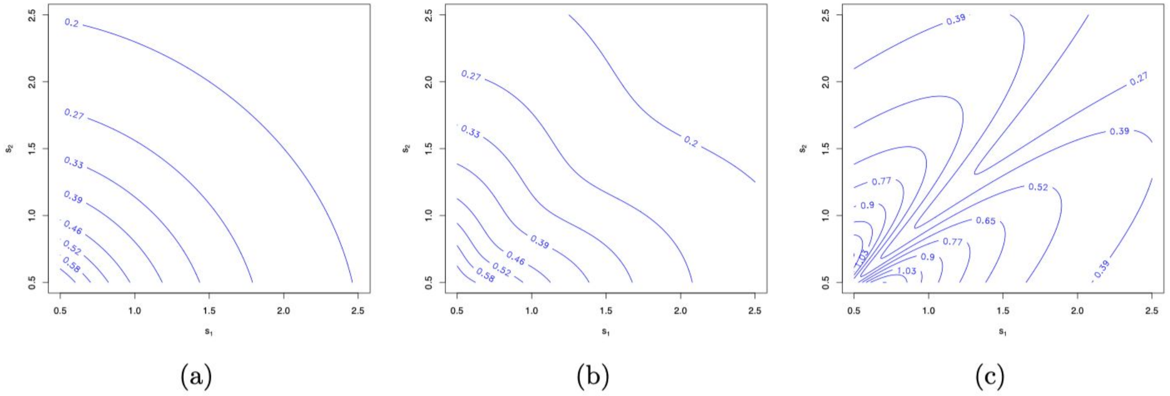


Figure 5.1: Level plots of the loss of PnL (and Sharpe Ratio) as a function of the Sharpe Ratio of two assets, assuming a maximum error ϵ in the Sharpe Ratio norm. Parameters: $\epsilon = 0.5$; Correlation: (a) $\rho = 0.1$, (a) $\rho = 0.5$, (c) $\rho = 0.9$.

Assume error in correlation. Below, and error of +/-0.1:

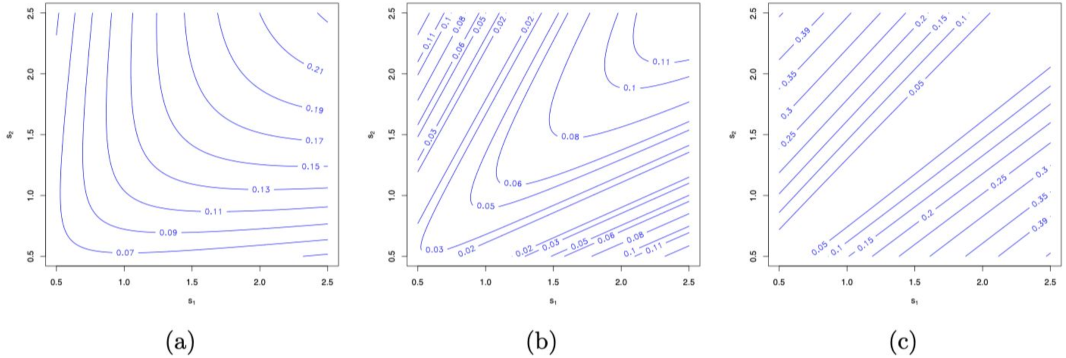


Figure 5.2: Level plots of the loss of PnL (and Sharpe Ratio) as a function of the Sharpe Ratio of two assets, assuming a maximum error ϵ in the correlation. Parameters: $\epsilon = 0.1$; Correlation: (a) $\rho = 0.1$, (a) $\rho = 0.5$, (c) $\rho = 0.9$.

Constraints

Sweeping statement: Everything in life has a) an objective; b) some data; c) some structure. Constraints are the structure. But why?

1. Investors' preferences
2. Tactical considerations
3. Regulatory considerations
4. Fiduciary considerations
5. Implementation considerations

Linear Constraints

$$\begin{aligned} \mathbf{A}'\mathbf{w} &\leq \mathbf{c} && \textit{(Inequality constraints)} \\ \mathbf{A}'\mathbf{w} &= \mathbf{c} && \textit{(Equality constraints)} \end{aligned}$$

1. Long-only $\mathbf{w} \geq 0$
2. GMV constraint. Not linear but linearized with a trick. $\sum_i |w_i| \leq G.$

$$\begin{aligned} \mathbf{x} &\geq 0 && \textit{(GMV constraint)} \\ \mathbf{y} &\geq 0 \\ \mathbf{w} &= \mathbf{x} - \mathbf{y} \\ \sum (x_i + y_i) &\leq G \end{aligned}$$

3. Long-Short $\sum_i x_i = K \sum_i y_i$
4. Exposure constraint $\sum_i \beta_i w_i = b_0$
5. Constant unit trading costs. Also linearizable $\sum_i c_i |w_i - w_i^{\text{start}}|^\gamma \leq C$

Quadratic Constraints

1. Factor/Style/Industry constraints

$$\begin{aligned} \mathbf{b}^{\text{style}} &= (\mathbf{B}^{\text{style}})'\mathbf{w} \\ (\mathbf{b}^{\text{style}})'\boldsymbol{\Omega}_{\mathbf{f}}^{\text{style}}\mathbf{b}^{\text{style}} &\le \sigma^2_{\text{style}} \qquad \textit{(Style factor vol constraint)} \end{aligned}$$

2. Tracking Error constraints. Define active holdings

$$\mathbf{w}^{\mathrm{a}} = \mathbf{w} - \mathbf{w}^{\text{bench}}.$$

$$(\mathbf{w}^{a})'\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{w}^{a} \le \sigma_a^2$$

$$\mathbf{w}'\boldsymbol{\Omega}_{\epsilon}\mathbf{w} \ge p_{\text{idio}}\mathbf{w}'\boldsymbol{\Omega}_{\mathbf{r}}\mathbf{w}$$

Non-Convex Constraints

1. Discrete number of assets, number of stocks. **Usually you don't need those**
2. Minimum percentage idio variance. Non-pd quadratic. Don't do it.

$$\mathbf{w}'\boldsymbol{\Omega}_\epsilon\mathbf{w} \geq p_{\text{idio}}\mathbf{w}'\boldsymbol{\Omega}_\mathbf{r}\mathbf{w}$$

i.e. $$\mathbf{w}'[p_{\text{idio}}\mathbf{B}\boldsymbol{\Omega}_\mathbf{f}\mathbf{B}' - (1 - p_{\text{idio}})\boldsymbol{\Omega}_\epsilon]\mathbf{w} \leq 0$$

3. Also, maximum drawdown.
4. Usually having these constraints reflect bad modeling practice. Even if you can find the global minimum, don't

Do Constraints Improve or Worsen Performance?

- Received wisdom: yes. You're reducing the feasible region
- True only if input parameters (risk and alpha) are correct
- In practice, constraints act as regularization terms that correct input error
- Constraints can be priced out as penalties, and reinterpreted as controlling for input error, in expectation or minimax

Approach	Penalty	Parameter Interpretation
Uncertain Alpha	$\tau^2 \left\|\mathbf{w}\right\|^2$	std.error of $\hat{\boldsymbol{\alpha}}$
Robust Alpha	$d \left\|\mathbf{w}\right\|$	max distance $\left\|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}\right\|$
Robust Factor	$\lambda\tau^2 \left\|\mathbf{w}\right\|^2$	volatility of a missing factor
Robust Correlations	$\lambda d \left\|\Lambda\mathbf{w}\right\|_1^2$	max distance $\lvert\rho_{i,j} - \hat{\rho}_{i,j}\rvert$
Robust Covariance	$\lambda d^2 \left\|\mathbf{w}\right\|^2$	max distance $\left\|\boldsymbol{\Omega}_{\mathbf{r}} - \hat{\boldsymbol{\Omega}}_{\mathbf{r}}\right\|$

- Read the Notes for the rigorous derivation of these

The Impact of Error on Performance

- It is an old topic
    - Michaud (1989): The Markowitz Optimization Enigma: Is `Optimized' Optimal?
    - Ziemba and Chopra (1992): The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice
    - Shepard (2009): Second Order Risk
    - Kan and Zhou (2007): Optimal Portfolio Choice with Parameter Uncertainty
    - Several papers, starting with Jagannathan and Ma (2003) interpret constraints as parameter regularization terms
- It is an important topic
    - See the two-asset portfolio examples. These are massive reductions in profitability!
    - The relative importance of alpha and risk is usually reported 90/10. It's based on no evidence whatsoever; really bad research and hot takes

Realized vs Predicted Sharpe

Realized Sharpe: $\mathrm{SR}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Omega}}_{\mathrm{r}}) = \dfrac{\boldsymbol{\alpha}'(\hat{\boldsymbol{\Omega}}_{\mathrm{r}}^{-1}\hat{\boldsymbol{\alpha}})}{\sqrt{(\hat{\boldsymbol{\Omega}}_{\mathrm{r}}^{-1}\hat{\boldsymbol{\alpha}})'\boldsymbol{\Omega}_{\mathrm{r}}(\hat{\boldsymbol{\Omega}}_{\mathrm{r}}^{-1}\hat{\boldsymbol{\alpha}})}}$

Sharpe Ratio Efficiency (SRE): $\dfrac{\mathrm{SR}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Omega}}_{\mathrm{r}})}{\mathrm{SR}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_{\mathrm{r}})}$

Fact: SRE ≤ 1 and =1 iff parameters are correct

Proof: Cauchy-Schwartz

$$\mathbf{a} := \boldsymbol{\Omega}_{\mathrm{r}}^{-1/2}\boldsymbol{\alpha}$$

$$\mathbf{b} := \boldsymbol{\Omega}_{\mathrm{r}}^{1/2}\hat{\boldsymbol{\Omega}}_{\mathrm{r}}^{-1}\hat{\boldsymbol{\alpha}}$$

$$\frac{\mathrm{SR}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Omega}}_{\mathrm{r}})}{\mathrm{SR}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_{\mathrm{r}})} = \frac{\mathbf{a}'\mathbf{b}}{\|\mathbf{a}\|\,\|\mathbf{b}\|}$$

Impact of Alpha Error

Fact: If

$$\left\| \frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|} - \frac{\hat{\boldsymbol{\alpha}}}{\|\hat{\boldsymbol{\alpha}}\|} \right\| \leq \delta_{\text{alpha}}$$

Then

$$\boxed{\frac{\mathrm{SR}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Omega}}_{\mathbf{r}})}{\mathrm{SR}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_{\mathbf{r}})} \geq 1 - \left\|\boldsymbol{\Omega}_{\mathbf{r}}^{-1}\right\|_2 \left\|\boldsymbol{\Omega}_{\mathbf{r}}\right\|_2 \delta_{\text{alpha}}^2}$$

Proof: complicated but not complex. See the Appendix in the Notes

Alpha Error: Implications

- You can estimate the angle between alphas by cross-sectional regression
- The result is an upper bound on SRE loss, but there is a matching lower bound, up to a constant! So, this is kind of tight.
- If you have an ill-conditioned (idio) true covariance matrix, the alpha error can be very damaging (see the 2-asset example)
- Intuition: I am wrong on a low-vol stock => I am going to be very very wrong in PnL

Impact of Risk Error

Fact: if $\left\|\boldsymbol{\Omega}_{\mathbf{r}}^{1/2}\hat{\boldsymbol{\Omega}}_{\mathbf{r}}^{-1}\boldsymbol{\Omega}_{\mathbf{r}}^{1/2} - \kappa\mathbf{I}\right\|_2 \leq \delta$

then $$\boxed{\frac{\mathrm{SR}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Omega}}_{\mathbf{r}})}{\mathrm{SR}(\boldsymbol{\alpha}, \boldsymbol{\Omega}_{\mathbf{r}})} \geq 1 - \frac{2\delta}{\kappa + \delta}}$$

Risk Error: Implications

- The bound is **not** tight, in the sense that: a) you can be very wrong on the risk of a subset of your universe; b) you can alpha alpha on the complement set; and c) experience no loss in Sharpe
- Still: if your cov mat is off by a constant, that is ok and the bound gets it right
- And assume that your get the eigenportfolios right. Then there is a nice formula:

$$\nu_i := \lambda_i / \hat{\lambda}_i \qquad \boxed{\mathrm{SRE} \geq 1 - \frac{\max_i \nu_i - \min_i \nu_i}{\max_i \nu_i} = \frac{\min_i \nu_i}{\max_i \nu_i}}$$