Algorithmic Trading & Quantitative Strategies

Lecture 2 (2/13/2024)

Giuseppe Paleologo (gardener)

Today's Remote Session

- I will upload the slides and the course notes tonight after class
- 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
- I will post a homework Thursday
- Office hours+Q&A: if it is ok with you do them over zoom, I can schedule two 1-hr sessions at 7PM on Mondays and Thursdays

Topics

- Recap from Last Class
  - GARCH
  - Kalman
- Introduction to Linear Models
  - Interpretation
  - Operations
  - Applications

Kalman

Kalman Filter

$$\begin{aligned}
\mathbf{x}_1 &\sim N(\hat{\mathbf{x}}_0, \hat{\boldsymbol{\Sigma}}_0) \\
\boldsymbol{\epsilon}_t &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon) & \boldsymbol{\epsilon}_t \perp \boldsymbol{\epsilon}_s, \boldsymbol{\epsilon}_t \perp \boldsymbol{\eta}_{s+1} && s \leq t \\
\boldsymbol{\eta}_t &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta) & \boldsymbol{\eta}_t \perp \boldsymbol{\eta}_s, \boldsymbol{\eta}_t \perp \boldsymbol{\epsilon}_{s+1} && s \leq t
\end{aligned}$$

$$\begin{aligned}
\mathbf{x}_{t+1} &= \mathbf{A}\mathbf{x}_t + \boldsymbol{\epsilon}_{t+1} \\
\mathbf{y}_{t+1} &= \mathbf{B}\mathbf{x}_{t+1} + \boldsymbol{\eta}_{t+1}
\end{aligned}$$

Define

$$\mathbf{Z}_t := \begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix} \quad \Rightarrow \quad \mathrm{cov}(\mathbf{Z}_t) = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{t|t-1} & \boldsymbol{\Sigma}_{\mathbf{x}_t}\mathbf{B}' \\ \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{x}_t} & \mathbf{B}\hat{\boldsymbol{\Sigma}}_{t|t-1}\mathbf{B}' + \boldsymbol{\Sigma}_\eta \end{bmatrix}$$

Formulas

$$\mathbf{K}_t := \hat{\boldsymbol{\Sigma}}_{t|t-1}\mathbf{B}'(\mathbf{B}\hat{\boldsymbol{\Sigma}}_{t|t-1}\mathbf{B}' + \boldsymbol{\Sigma}_{\eta})^{-1}$$

(Kalman Gain)

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{t|t} &= [\mathbf{I} - \mathbf{K}_t\mathbf{B}]\hat{\boldsymbol{\Sigma}}_{t|t-1} \\
\hat{\mathbf{x}}_{t|t} &= \boxed{(\mathbf{I} - \mathbf{K}_t\mathbf{B})\hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\mathbf{y}_t} \\
\hat{\boldsymbol{\Sigma}}_{t+1|t} &= \mathbf{A}\hat{\boldsymbol{\Sigma}}_{t|t}\mathbf{A}' + \boldsymbol{\Sigma}_{\eta} \\
\hat{\mathbf{x}}_{t+1|t} &= \mathbf{A}\hat{\mathbf{x}}_{t|t}
\end{aligned}$$

(Basic Formulas)

In steady state:

$$\mathbf{X} = \mathbf{A}\mathbf{X}\mathbf{A}' - \mathbf{A}\mathbf{X}\mathbf{B}'(\mathbf{B}\mathbf{X}\mathbf{B}' + \boldsymbol{\Sigma}_{\eta})^{-1}\mathbf{B}\mathbf{X}\mathbf{A}' + \boldsymbol{\Sigma}_{\epsilon}$$

(Riccati Equation)

The Simplest Example

$$
\begin{aligned}
x_{t+1} &= x_t + \tau_\epsilon \epsilon_{t+1} \\
y_{t+1} &= x_{t+1} + \tau_\eta \eta_{t+1}
\end{aligned}
$$

$$\kappa := \frac{\tau_\eta}{\tau_\epsilon}$$

$$\hat{\sigma}^2_{t+1|t} = \frac{1}{2}\tau_\epsilon^2\left(1 + \sqrt{(2\kappa)^2 + 1}\right)$$

$$
\boxed{
\begin{aligned}
K &= \frac{\hat{\sigma}^2_{t+1|t}}{\hat{\sigma}^2_{t+1|t} + \tau_\eta^2} \\
\hat{x}_{t+1|t} &= (1 - K)\hat{x}_{t|t-1} + K y_t
\end{aligned}
}
$$

For $\kappa \gg 1$

$$\boxed{\hat{x}_{t|t} = \frac{\kappa}{1+\kappa}\hat{x}_{t|t-1} + \frac{1}{1+\kappa} y_t}$$

A More Complex Example

A lognormal model for returns: $r_t = e^{\beta + \exp(x_t/2)\xi_t} - 1 \qquad \xi_t \sim N(0,1)$

Trick:

$$
\begin{aligned}
u_t &:= \log(1 + r_t) - \beta \\
\Rightarrow \quad u_t &= \exp(x_t/2)\xi_t \\
\Rightarrow \quad \log u_t^2 &= x_t + \log \xi_t^2 \\
&= x_t + \eta_t + \gamma
\end{aligned}
$$

Observation:

$$
\begin{aligned}
y_t &:= \log u_t^2 - \gamma \\
y_t &= x_t + \eta_t
\end{aligned}
$$

$$\boxed{\hat{x}_{t+1|t} = (1 - K)\hat{x}_{t|t-1} + K(\log(\log(1 + r_t))^2 - \gamma)}$$

Linear (Factor Models)

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t$$

- $t \in \mathbb{N}$ denotes time;
- $\boldsymbol{\alpha}$ is an n-dimensional vector;
- $\mathbf{r}_t$ is a random vector of n asset returns;
- $\mathbf{f}_t$ is the random vector of m factor returns;
- $\mathbf{B}$ is a $n \times m$ loadings matrix;
- $\boldsymbol{\epsilon}_t$ is the random vector of n idiosyncratic (or specific) returns.

$$\mathbf{r}_t = \overset{\text{alpha}}{\boldsymbol{\alpha}} + \overbrace{\underset{\text{factor loadings}}{\mathbf{B}}\,\underset{\text{factor returns}}{\mathbf{f}_t}}^{\text{systematic}} + \underset{\text{idio rets}}{\boldsymbol{\epsilon}_t}$$

The model has extra degrees of freedom

A feature, not a bug

Interpretations of B

style

0.12	-1.91	…
-2.39	-3.00	…
-1.52	…	…
…	…	…

country

0	0	1	…
1	0	0	…
0	1	0	…
0	0	0	…
…	…	…	…

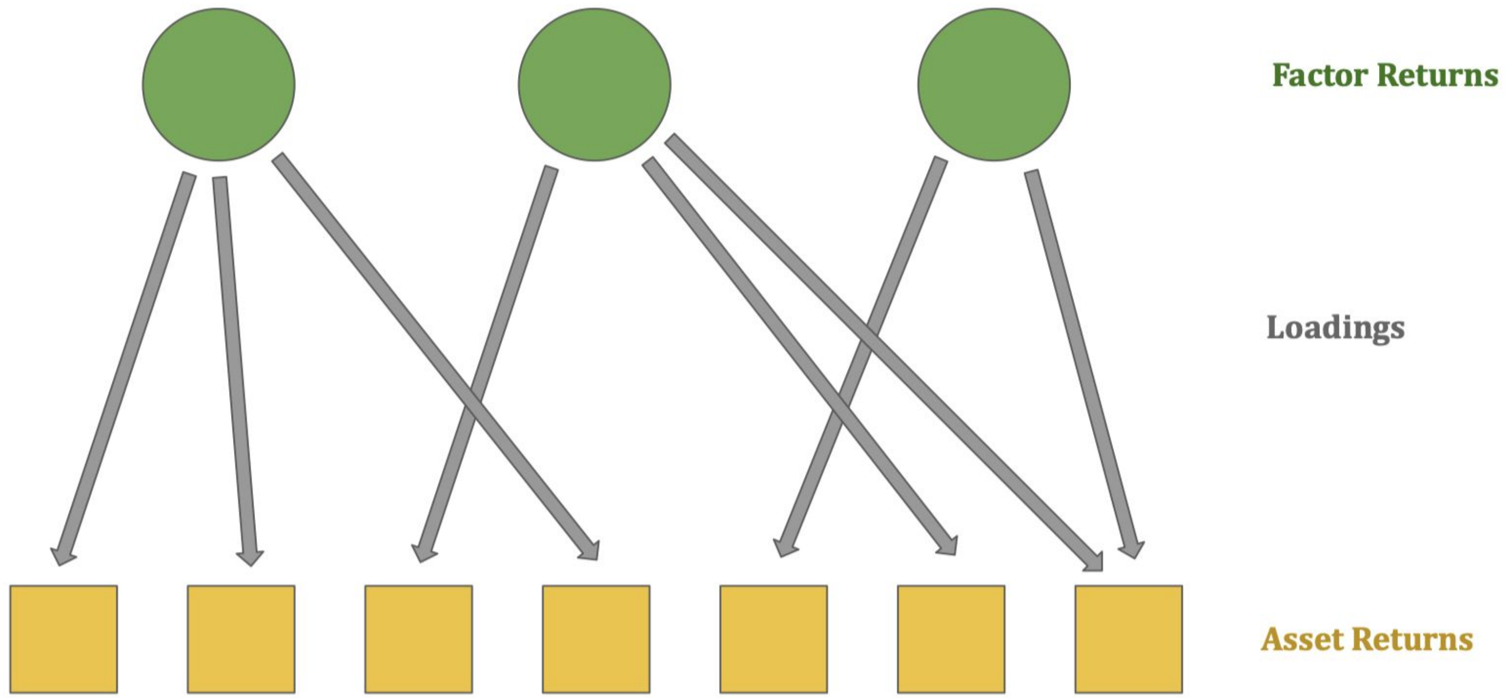
industry

0	1	0	0	…
0	0	0	1	…
1	0	0	0	…
…	…	…	…	…

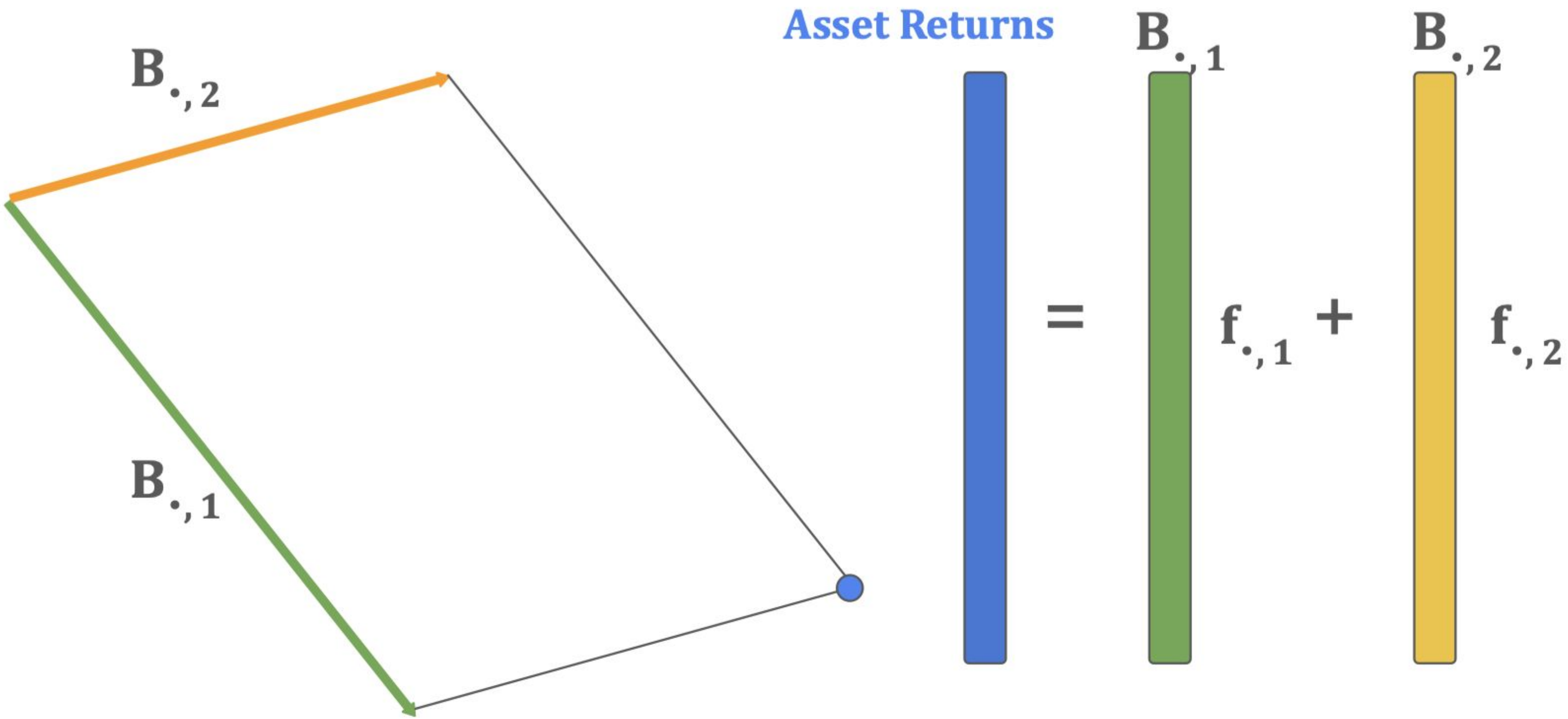
Types of Factor Models

1. **Fundamental:** you know the loadings but not the factor returns
2. **Statistical:** you don't know the loadings nor the factor returns
3. **Macroeconomic:** you know the factor returns but not the loadings

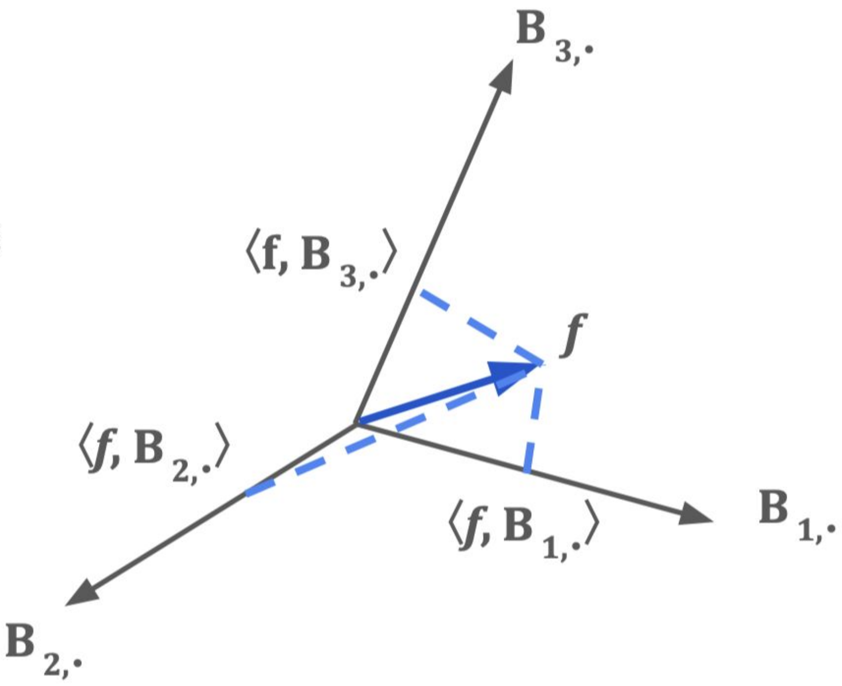
$$E(\mathbf{r} - \boldsymbol{\alpha}|\mathbf{f}) = \sum_j [\mathbf{B}]_{\cdot,j} f_j$$



$$E(r_i - \alpha_i | \mathbf{f}) = \langle [\mathbf{B}]_{i,\cdot}, \mathbf{f} \rangle$$



$$E\left(\mathbf{w}'\mathbf{r}\middle|\,\mathbf{f}\right) = E\left(\sum_i w_i r_i \middle|\, \mathbf{f}\right)$$
$$= \sum_i \left[\alpha_i + \langle [\mathbf{B}]_{i,\cdot}\mathbf{f}\rangle\right] w_i$$



Important: Alpha Spanned and Alpha Orthogonal

Write $\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\lambda} + \boldsymbol{\alpha}_\perp$

$$\mathbf{r}_t = \boldsymbol{\alpha}_\perp + \mathbf{B}[\boldsymbol{\lambda} + E(\mathbf{f}_t)] + \mathbf{B}[\mathbf{f}_t - E(\mathbf{f}_t)] + \boldsymbol{\epsilon}_t$$

Choose $\mathbf{w} = \boldsymbol{\alpha}_\perp / \|\boldsymbol{\alpha}_\perp\|$

$$E(\mathbf{w}'\mathbf{r}_t) = \|\boldsymbol{\alpha}_\perp\|$$

$$\mathrm{var}(\mathbf{w}'\mathbf{r}_t) = \frac{\boldsymbol{\alpha}_\perp' \boldsymbol{\Omega}_\epsilon \boldsymbol{\alpha}_\perp}{\|\boldsymbol{\alpha}_\perp\|}$$

$$\mathrm{SR}(\mathbf{w}) \geq \frac{\|\boldsymbol{\alpha}_\perp\|}{\|\boldsymbol{\Omega}_\epsilon\|_{\mathrm{op}}} \geq \sqrt{n}\frac{\mu}{\|\boldsymbol{\Omega}_\epsilon\|_{\mathrm{op}}}$$

Alpha orthogonal is extremely valuable. Very high SR

Transformations #1: Rotations

Replace loadings and returns with a "rotated" version:

$$\tilde{\mathbf{B}} = \mathbf{B}\mathbf{C}^{-1}$$
$$\tilde{\mathbf{f}} = \mathbf{C}\mathbf{f}$$

You get an equivalent model:

$$\mathbf{r} = \boldsymbol{\alpha} + \tilde{\mathbf{B}}\tilde{\mathbf{f}} + \boldsymbol{\epsilon}$$

Examples:

1. Uncorrelated, unit vol factor returns
2. Orthogonal loadings
3. Can we center loadings and get an equivalent model?

Transformation #2: Projections

Sometimes, the model is just too big. Why?

1. Maybe it's good, but hard to interpret
2. Maybe it's not even that good (=commercial models) and it's just better to have a simpler model?

From $\mathbf{r} = \boldsymbol{\alpha} + \mathbf{Bf} + \boldsymbol{\epsilon}$ To $\mathbf{r} = \boldsymbol{\alpha} + \mathbf{Ag} + \boldsymbol{\eta}$

Theorem 3.1. *Let the distance between the original factor returns* $\mathbf{f}$ *and the approximate factor returns* $\mathbf{g}$ *be* $E\|\mathbf{Bf} - \mathbf{Ag}\|^2$. *The distance-minimizing approximate factor returns are* $\mathbf{g} = \mathbf{Hf}$, *where* $\mathbf{H} := (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{B}$. *The corresponding value of* $\boldsymbol{\Omega}_g$ *is*

$$\boldsymbol{\Omega}_g := \mathbf{H}\boldsymbol{\Omega}_f\mathbf{H}'$$

Transformation #3: Push-outs

- This transformation increases the dimensionality of the model
- Posit that the residuals are not uncorrelated and/or can be predicted

$$\boldsymbol{\epsilon} = \mathbf{A}\mathbf{g} + \boldsymbol{\eta} \qquad\qquad \mathbf{r} = \mathbf{B}\mathbf{f} + \mathbf{A}\mathbf{g} + \boldsymbol{\eta}$$

- The essential step is "orthogonalization" of **A** with respect to **B**: regress **A** on **B** and keep the residuals, so that **B'A**=0
- This is the same as "Multivariate regression as a sequence of univariate regressions" (read Friedman-Hastie-Tibshirani)
- And the same as Cholesky
- Read the Appendix: Frisch-Waugh-Lovell is essential

Applications #1: Performance Attribution

Gives you an understanding of where the PnL comes from

$$
\begin{aligned}
\textit{(portfolio PnL}_t\textit{)} &= \mathbf{w}_t'\mathbf{r}_t \\
&= \mathbf{w}_t'\mathbf{B}\mathbf{f}_t + \mathbf{w}_t'(\boldsymbol{\alpha}_\perp + \boldsymbol{\epsilon}_t) \\
&= \mathbf{b}_t'\mathbf{f}_t + \mathbf{w}_t'(\boldsymbol{\alpha}_\perp + \boldsymbol{\epsilon}_t) \qquad (\mathbf{b}_t := \mathbf{B}'\mathbf{w}_t) \quad \text{"Factor exposures"}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{PnL} &= \sum_{t=1}^{T} \textit{(Factor PnL}_t\textit{)} + \textit{(Residual PnL}_t\textit{)} \\
&= \sum_{t=1}^{T}\sum_{j=1}^{m} [\mathbf{b}_t]_j [\mathbf{f}_t]_j + \sum_{t=1}^{T}\sum_{i=1}^{n} [\mathbf{w}_t]_i (\alpha_{\perp,i} + [\boldsymbol{\epsilon}_t]_i) \\
&= \sum_{j=1}^{m}\sum_{t=1}^{T} [\mathbf{b}_t]_j [\mathbf{f}_t]_j + \sum_{i=1}^{n}\sum_{t=1}^{T} [\mathbf{w}_t]_i (\alpha_{\perp,i} + [\boldsymbol{\epsilon}_t]_i) \\
&= \sum_{j=1}^{m} \textit{(Factor } j \textit{ PnL)} + \sum_{i=1}^{n} \textit{(Stock } i \textit{ Residual PnL)}
\end{aligned}
$$

Application #2: Risk Management

Start with variance prediction and decomposition:

$$\begin{aligned}\mathrm{var}(\mathbf{r}'\mathbf{w}) =& \mathbf{w}'(\mathbf{B}\boldsymbol{\Omega}_{\mathbf{f}}\mathbf{B}' + \boldsymbol{\Omega}_{\epsilon})\mathbf{w} \\ =& \mathbf{b}'\boldsymbol{\Omega}_{\mathbf{f}}\mathbf{b} + \mathbf{w}'\boldsymbol{\Omega}_{\epsilon}\mathbf{w}\end{aligned}$$

But then you can answer almost any question:

1. First order of concern: take first derivatives of everything
2. Strategic risk management: create scenarios where you
   a. Stress parameters
   b. Stress future returns

Application #3: Alpha Research

1. Linear models are powerful
2. They are not only linear. Nonlinearity is hidden inside the factor loadings. What 99% of research think of as non-linear is just linear models with arbitrary loadings
3. A better description is "*shallow models*". Shallow models are great, except when dating, probably
4. Recent research on "benign overfitting" extends models to very large-dimensional models

Application #4: Portfolio Management

1. Factor models fit like a glove with mean-variance optimization
2. Example: it's easier to address data errors with factor models
3. And it's easier to interpret and solve these models

Quiz

1. Can you model autoregressive processes order greater than 1 with Kalman Filters?
    a. Yes, they are just bigger
    b. No, because it's only for linear processes
    c. No
2. What types of alphas is related to factor returns?
    a. Alpha orthogonal
    b. Alpha Spanned
    c. Both
3. In a factor rotation the matrix C must be
    a. Unitary
    b. Nonsingular
    c. Symmetric positive definite
4. When pushing out a model, you build a model using the
    a. Residual returns of the existing model
    b. Total returns
    c. Either works
5. Performance attribution decomposes
    a. The PnL of a strategy
    b. The return of a strategy
    c. Both
6. What can't you use a factor model for?
    a. Alpha research
    b. Market impact
    c. Risk management
7. What model is used most often in practice?
    a. Macroeconomic
    b. Fundamental
    c. Statistical