Decision Theory

A Finance Application

Ye Luo

HKU Business School

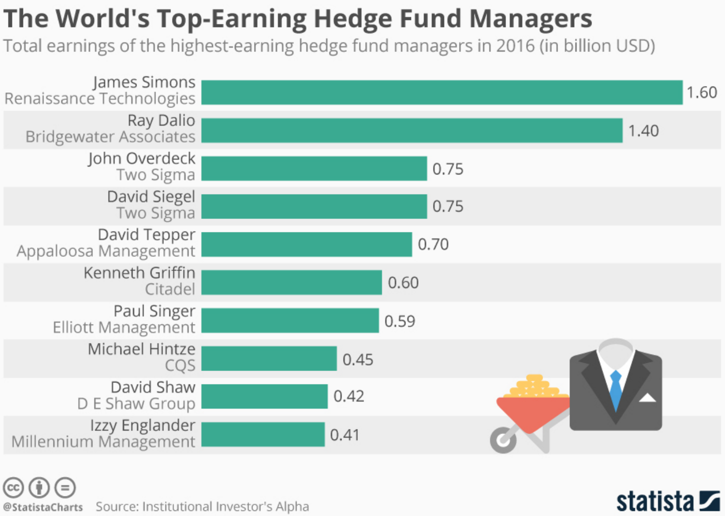
Roadmap

- Introduce Decision Theory.
- Introduce Mean-Variance Optimization Framework in Finance.
- Re-visit Fame-French's factor model.
- Discuss about methodologies, applications and others.

Why do we care about these?

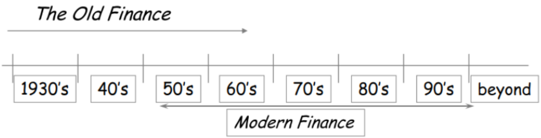
- Advances in financial asset pricing theory.
- Understanding Financial markets.
- Hedge fund industry applications.
- Modern AI application in Quantative trading.

2016 Hedge Fund Ranking



Haugen's View of Finance

Haugen's view: The Evolution of Academic Finance



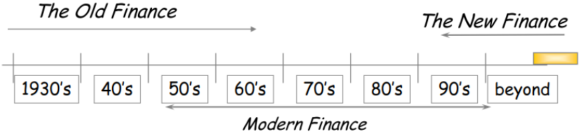
Modern Finance

Theme: Valuation Based on Rational Economic Behavior

Paradigms: Optimization *(Markowitz)* Irrelevance *(Modigliani & Miller)* CAPM *(Sharpe, Lintner & Mossen)* EMH *(Fama)*

Foundation: Financial Economics

Haugen's view: The Evolution of Academic Finance



The New Finance

Theme: Inefficient Markets

Paragidms: Inductive *ad hoc* Factor Models — Expected Return, Risk; Behavioral Models

Foundation: Statistics, Econometrics, and Psychology

Financial Economic Theory

- There is a Utility function $U(c)$. The agent (investor) is maximizing

$$\sum_{t=0}^{T} \beta^t U(c_t).$$

  c_t : consumption (or wealth).
- There is a financial market that allows us to make investments. N is number of assets. There is a risk free asset with risk free rate r or r_t.
- $\mu_t := \mathbb{E}[R_t]$ is the vector of mean return at time t. $\Sigma_t := Var(R_t)$ is the covariance matrix of returns at time t.

Markowitz's mean variance theory

- Assuming that $U(c) = \mathbb{E}[c] - \gamma/2 \times var(c)$, mean-variance utility, γ : Risk Aversion parameter.
- For a portfolio $w = (w_f, w_r)$, solve $\max_w U(c)$.
- w_f : weight of wealth invested in risk free asset. w_r weight vector $(N \times 1)$ invested in risky asset. What is the restriction for w_f and w_r?
- What's the mathematical formulation or this problem?

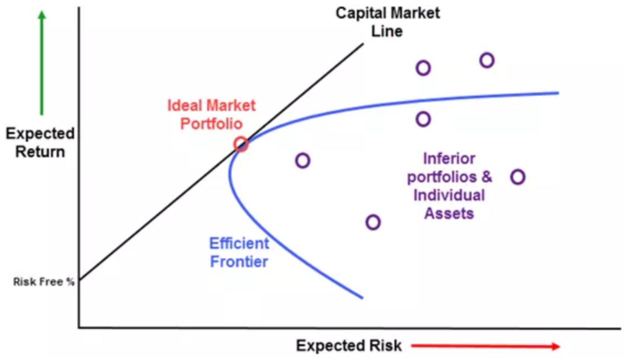
Link with portfolio optimization

- Static problem:
$$\max_{w} \mu^\intercal w - \frac{\gamma}{2} w^\intercal \Sigma w.$$
γ : risk aversion, μ : mean return, Σ : Variance-matrix.
- Ideally the optimal policy is $\frac{1}{\gamma}\Sigma^{-1}\mu$.
- Plug in *estimates*** of μ and Σ.

Discussions

- Different agents may have different utilities, e.g., risk aversion parameter γ are different.
- However, all people holds risky assets that are proportional to $\Sigma^{-1}\mu$.
- Everybody holds the same risky portfolio!
- This portfolio must be the same as the market portfolio. Why?

Capital Asset Pricing Model



Capital Asset Pricing Model

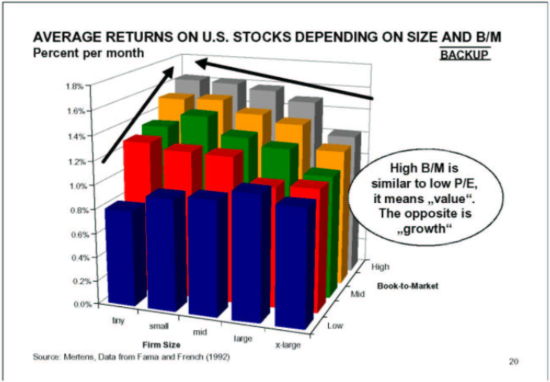
- The above MV-theory implies that there is a factor r_m : the return of the market portfolio.
- For each stock, the return $R_{it} - r = \beta(r_m - r) + \epsilon_{it}$.
- How to estimate β? Run regression of $R_{it} - r$ on $r_m - r$.
- But this is not enough for MV - need to estimate μ and Σ.

The Fame-French Factors

- Fama-French (1993)'s Nobel prize winning work.
- Three factors that governs the co-movement of financial assets:
  1. CAPM's market return (always the strongest signal!).
  2. Book to market ratio (B/M). High book to market tends to have higher returns.
  3. Firm size. Larger firms tend to have smaller returns.

Book to Market and Size

Small „value“ companies have higher returns



Dynamic portfolio optimization

- Dynamic problem:

$$\max_{w_1,\dots,w_T} \sum_{t=1}^{T} \mu_t^\intercal w_t - \frac{\gamma}{2} w_t^\intercal \Sigma_t w_t - P(w_t, w_{t-1}).$$

  $P(\cdot)$: transaction cost. For example,
  $P(w_t, w_{t-1}) = c\|w_t - w_{t-1}\|^2$.

- Portfolio position at t affects transaction cost at $t+1$.

Why dynamics?

- The industry is facing multi-period problems: hedge funds, high-frequency traders.
- Need long term "strategic" planning to reduce transaction cost.
- μ_t, Σ_t can change over time.
- Let's focus on static problem ($T = 1$) for now.

Golden formula of Markowitz's

- Golden formula: $w_r^* = \frac{1}{\gamma}\Sigma^{-1}\mu_e$. $w_f^* = 1 - w_r^* e$.
- When $T = 1$, investor should simply allocate assets according to Markowitz' golden formula.
- What is the formula of utility that we can attain?
- Any doubts or weakness?

Utility, investment decisions and its implications

- Assuming that we have an investment strategy (policy). This strategy can be written down as a function of μ, Σ.
- That is: $w_r^t := f(\mu_t, \Sigma_t)$.
- Assuming that there is a utility function U.
- Markowitz said: if U is mean-variance utility, then $w_r^t := \frac{1}{\gamma}\Sigma^{-1}\mu$.
- f has a simple functional form. w_r^t is independent of each other.
- We will call this situation "Markowitz" portfolio, or "Markowitz" policy.

Decision Theory-1

- Given a utility function $U(\cdot)$, consider a policy $P(\cdot)$.
- The policy is a function that maps observables (predictors, for example), into decisions.
- Choose the policy P such that $\mathbb{E}_{a\sim P}[U(a)]$ is maximized.
- Or to minimize:
  $\mathbb{E}_{a\sim P^*}[U(a)] - \mathbb{E}_{a\sim P}[U(a)]$,
  where $P^*(\cdot)$ is a benchmark policy.

- For each R_{it}, we need a model to predict μ_i. What models have we learned so far? Weighted, exponential filtering (weighting), AR models, factor models. We will call the estimator $\hat{\mu}$.
- Simple mean: $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t$.
- Exponential weighting: $\hat{\mu} = \frac{1}{T}\frac{1}{\sum_{t=1}^{T}\exp^{-\lambda(T-t)}}\sum_{t=1}^{T}\exp^{-\lambda(T-t)} R_t$.
- AR model: $R_{t+1} = a + bR_t + \epsilon_t$. $\hat{\mu} = \hat{a}/(1-\hat{b})$.
- Factor model: $R_t = \alpha_i + Bf_t + \epsilon_t$. $\mu = \alpha + B\mathbb{E}[f_t]$.

Econometrics behind Markowitz

- We need to estimate the co-variance matrix Σ. What methods do we know so far?
- Naive: $\hat{\Sigma} := \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})'$, where $\hat{\mu}$ is "some" estimator of μ, e.g., mean of R_t.
- Factor model: $\Sigma = BE[ff']B' + diag(\sigma_1^2, ..., \sigma_N^2)$.
- We plug in estimators of B and σ_i^2 from the residuals of the factor model?
- Therefore, Markowitz policy $\hat{w}_r := \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}$.

Decision Theory for mean variance problem

- Utility function U.
- Policy is $\hat{w}_r = f(\hat{\mu}, \hat{\Sigma})$.
- Regret: $\mathbb{E}U(w_r^*) - \mathbb{E}U(\hat{w}_r)$.
- w_r^* : some "ideal" target policy or strategy. Minimizing regret.
- Multi-period: $Reg := \min \sum_{t=0}^{T} \beta^t (\mathbb{E}U(w_r^{t*}) - \mathbb{E}U(\hat{w}_r^t))$.

Other Strategies beyond Markowitz

- Naive portfolio: $1/N$ on all assets.
- Zhou and Tu (2003): $1/N$ is actually doing well.
- Kan and Zhou (2007): comparison of Markowitz and other "rules"
- Ledoit's Shrinkage estimator.
- Lubos Pastor (2000): Bayesian portfolio rule.
- Summarize: Estimation of μ and Σ matters a lot!

1/N: Zhou and Tu

Table 1
Utilities in a one-factor model without mispricing.

This table reports the average utilities (annualized and in percentage points) of a mean–variance investor under various investment rules: the true optimal rule, the $1/N$ rule, the ML rule, the Jorion (1986) rule, the MacKinlay and Pástor (2000) rule, the Kan and Zhou (2007) rule, and the four combination rules, with 10,000 sets of sample size T simulated data from a one-factor model with zero mispricing alphas and with N=25 assets. Panels A and B assume that the risk aversion coefficient γ is 3 and 1, respectively.

	T					
Rules	120	240	480	960	3000	6000
Panel A: $\gamma = 3$						
True	4.17	4.17	4.17	4.17	4.17	4.17
$1/N$	3.89	3.89	3.89	3.89	3.89	3.89
ML	−85.72	−25.81	−8.35	−1.61	2.42	3.30
Jorion	−12.85	−3.79	−0.18	1.55	2.98	3.47
MacKinlay-Pástor	2.11	3.00	3.44	3.65	3.79	3.83
Kan-Zhou	−2.15	−0.00	1.13	1.90	2.97	3.47
$\hat{w}^{CML}$	1.68	2.95	3.42	3.60	3.81	3.90
$\hat{w}^{CPJ}$	1.42	2.93	3.46	3.71	3.88	3.86
$\hat{w}^{CMP}$	2.19	3.05	3.48	3.67	3.80	3.83
$\hat{w}^{CKZ}$	3.71	3.77	3.81	3.85	3.91	3.95
Panel B: $\gamma = 1$						
True	12.50	12.50	12.50	12.50	12.50	12.50
$1/N$	6.63	6.63	6.63	6.63	6.63	6.63
ML	−257.16	−77.42	−25.05	−4.83	7.25	9.91
Jorion	−38.55	−11.38	−0.55	4.66	8.95	10.42
MacKinlay-Pástor	6.33	9.00	10.31	10.94	11.37	11.48
Kan-Zhou	−6.44	−0.01	3.38	5.69	8.92	10.40
$\hat{w}^{CML}$	1.14	4.79	6.39	7.47	9.50	10.62
$\hat{w}^{CPJ}$	1.28	5.68	6.97	7.11	7.46	10.34
$\hat{w}^{CMP}$	6.57	9.16	10.49	11.09	10.95	11.43
$\hat{w}^{CKZ}$	6.36	6.70	6.99	7.41	8.78	9.97

1/N: Zhou and Tu, factor model

Table 2
Utilities in factor models with mispricing.

This table reports the average utilities (annualized and in percentage points) of a mean–variance investor under various investment rules: the true optimal rule, the $1/N$ rule, the ML rule, the Jorion (1986) rule, the MacKinlay and Pástor (2000) rule, the Kan and Zhou (2007) rule, and the four combination rules, with N=25 assets for 10,000 sets of sample size T simulated data from a one-factor model (Panel A) and a three-factor model (Panel B), respectively. The annualized mispricing α's are assumed to spread evenly between −2% to 2%. The risk aversion coefficient γ is 3.

			T			
Rules	120	240	480	960	3000	6000
Panel A: One-factor model						
True	6.50	6.50	6.50	6.50	6.50	6.50
$1/N$	3.89	3.89	3.89	3.89	3.89	3.89
ML	−84.75	−23.84	−6.18	0.65	4.73	5.62
Jorion	−12.36	−2.99	0.95	3.09	5.06	5.71
MacKinlay-Pástor	2.34	3.23	3.67	3.88	4.02	4.06
Kan-Zhou	−2.35	0.02	1.64	3.14	5.06	5.71
$\hat{w}^{CML}$	2.02	3.32	3.91	4.43	5.38	5.82
$\hat{w}^{CPJ}$	2.27	3.70	4.02	3.92	4.83	5.72
$\hat{w}^{CMP}$	2.41	3.27	3.71	3.90	4.02	4.04
$\hat{w}^{CKZ}$	3.84	3.95	4.12	4.41	5.14	5.62
Panel B: Three-factor model						
True	14.60	14.60	14.60	14.60	14.60	14.60
$1/N$	3.85	3.85	3.85	3.85	3.85	3.85
ML	−81.09	−17.11	1.39	8.52	12.76	13.69
Jorion	−7.85	2.84	7.65	10.45	12.99	13.75
MacKinlay-Pástor	1.78	2.66	3.09	3.30	3.44	3.48
Kan-Zhou	1.61	5.12	7.96	10.45	12.99	13.75
$\hat{w}^{CML}$	3.84	6.15	8.44	10.63	13.02	13.76
$\hat{w}^{CPJ}$	5.79	5.36	4.17	9.67	13.02	13.76
$\hat{w}^{CMP}$	1.86	2.73	3.12	3.30	3.45	3.48
$\hat{w}^{CKZ}$	5.09	6.06	7.57	9.59	12.56	13.58

1/N: Zhou and Tu, factor model, Sharpe Ratios

Table 3
Sharpe ratios in a one-factor model.

This table reports in percentage points the average Sharpe ratios of a mean–variance investor under various investment rules: the true optimal rule, the 1/*N* rule, the ML rule, the Jorion (1986) rule, the MacKinlay and Pástor (2000) rule, the Kan and Zhou (2007) rule, and the four combination rules, with 10,000 sets of sample size *T* simulated data from a one-factor model with *N*=25 assets. Panels A and B assume that the annualized mispricing α's are zeros or between −2% to 2%, respectively.

	T					
Rules	120	240	480	960	3000	6000
Panel A: $\alpha = 0$						
True	14.43	14.43	14.43	14.43	14.43	14.43
1/*N*	13.95	13.95	13.95	13.95	13.95	13.95
ML	3.88	5.59	7.54	9.54	12.19	13.18
Jorion	4.54	6.46	8.40	10.18	12.38	13.24
MacKinlay-Pástor	12.19	13.51	13.86	13.89	13.89	13.89
Kan-Zhou	4.97	7.03	8.80	10.27	12.34	13.24
$\hat{w}^{CML}$	12.04	12.88	13.34	13.53	13.83	13.98
$\hat{w}^{CPJ}$	10.40	12.36	13.22	13.67	13.94	13.90
$\hat{w}^{CMP}$	12.07	13.44	13.87	13.90	13.89	13.89
$\hat{w}^{CKZ}$	13.70	13.79	13.86	13.91	14.00	14.07
Panel B: α *in* [−2%, 2%]						
True	18.02	18.02	18.02	18.02	18.02	18.02
1/*N*	13.95	13.95	13.95	13.95	13.95	13.95
ML	5.92	8.34	10.94	13.32	16.06	16.97
Jorion	5.61	8.03	10.69	13.16	16.03	16.95
MacKinlay-Pástor	12.70	13.98	14.28	14.30	14.31	14.31
Kan-Zhou	4.77	7.15	10.09	12.97	16.02	16.95
$\hat{w}^{CML}$	12.81	13.69	14.30	15.02	16.45	17.09
$\hat{w}^{CPJ}$	11.64	13.73	14.31	14.12	15.60	16.96
$\hat{w}^{CMP}$	12.52	13.89	14.26	14.28	14.27	14.25
$\hat{w}^{CKZ}$	14.02	14.23	14.54	15.04	16.21	16.91

loss of utility: Kan and Zhou

TABLE 1

Percentage Loss of Expected Out-of-Sample Performance Due to Estimation Errors in the Means and Covariance Matrix of Returns

Table 1 presents the percentage loss of expected out-of-sample performance from holding a sample tangency portfolio of N risky assets with the parameters estimated using T periods of historical returns instead of using the true parameters. The first column reports the percentage loss due to the use of the sample average returns $\hat{\mu}$ instead of true expected returns. The second column reports the percentage loss due to the use of the sample covariance matrix $\hat{\Sigma}$ instead of the true covariance matrix. The third column reports the interactive effect from using $\hat{\mu}$ and $\hat{\Sigma}$. The fourth column reports the total percentage loss of expected out-of-sample performance from using $\hat{\mu}$ and $\hat{\Sigma}$. Panel A assumes the Sharpe ratio (θ) of the N risky assets is 0.2 and Panel B assumes $\theta = 0.4$.

		Percentage Loss of Expected Out-of-Sample Performance			
N	T	$\hat{\mu}$	$\hat{\Sigma}$	Interaction	$\hat{\mu}$ and $\hat{\Sigma}$
Panel A. $\theta = 0.2$					
1	60	41.67	4.31	6.18	52.15
	120	20.83	1.90	1.46	24.19
	240	10.42	0.89	0.36	11.66
	360	6.94	0.58	0.16	7.68
	480	5.21	0.43	0.09	5.73
2	60	83.33	6.85	17.61	107.80
	120	41.67	2.93	4.09	48.69
	240	20.83	1.35	0.99	23.17
	360	13.89	0.88	0.43	15.20
	480	10.42	0.65	0.24	11.31
5	60	208.33	16.64	89.69	314.66
	120	104.17	6.44	19.62	130.23
	240	52.08	2.84	4.61	59.53
	360	34.72	1.81	2.01	38.54
	480	26.04	1.33	1.12	28.49
10	60	416.67	42.99	387.46	847.12
	120	208.33	13.95	75.36	297.64
	240	104.17	5.65	16.85	126.67
	360	69.44	3.51	7.23	80.19
	480	52.08	2.54	4.00	58.62
25	60	1041.67	336.67	5211.57	6589.91
	120	520.83	55.53	591.64	1168.01
	240	260.42	17.18	110.77	388.37
	360	173.61	9.81	45.19	228.61
	480	130.21	6.81	24.39	161.42

loss of utility2: Kan and Zhou

TABLE 2

Expected Out-of-Sample Performance of Various Portfolio Rules with 10 Risky Assets When Returns Follow a Multivariate Normal Distribution

Table 2 reports the expected out-of-sample performance (in percentages per month) of 13 portfolio rules that choose an optimal portfolio of 10 risky assets and a riskless asset for different lengths of the estimation period (T). The excess returns of the 10 risky assets are assumed to be generated from a multivariate normal distribution with the mean and covariance matrix chosen based on the sample estimates of 10 size-ranked NYSE portfolios. The investor is assumed to have a risk aversion coefficient of three. The expected out-of-sample performance of the first eight rules and the global minimum-variance rule are obtained analytically. For the other four rules, the expected out-of-sample performances are approximated using 100,000 simulations.

Portfolio Rule	$T = 60$	$T = 120$	$T = 180$	$T = 240$
Parameter certainty optimal	0.419	0.419	0.419	0.419
Theoretical optimal two-fund	0.044	0.088	0.122	0.150
Theoretical optimal three-fund	0.133	0.168	0.191	0.209
1st Plug-in, $\hat{\Sigma}$	−5.122	−1.531	−0.748	−0.411
2nd Plug-in, $\tilde{\Sigma} = T\hat{\Sigma}/(T-1)$	−4.936	−1.498	−0.735	−0.404
3rd Plug-in, $\breve{\Sigma} = T\hat{\Sigma}/(T-N-2)$	−3.110	−1.156	−0.596	−0.329
Bayesian (diffuse prior)	−2.996	−1.130	−0.584	−0.323
Parameter-free optimal two-fund	−1.910	−0.879	−0.476	−0.263
Estimated optimal two-fund	−0.185	−0.007	0.060	0.102
Uncertainty aversion two-fund	−0.001	0.004	0.007	0.012
Global minimum-variance	−0.152	−0.010	0.040	0.064
Jorion's shrinkage	−0.899	−0.220	−0.030	0.062
Estimated optimal three-fund	−0.343	−0.053	0.051	0.107
Portfolio Rule	$T = 300$	$T = 360$	$T = 420$	$T = 480$
Parameter certainty optimal	0.419	0.419	0.419	0.419
Theoretical optimal two-fund	0.173	0.193	0.210	0.224
Theoretical optimal three-fund	0.224	0.237	0.248	0.258
1st Plug-in, $\hat{\Sigma}$	−0.225	−0.107	−0.025	0.034
2nd Plug-in, $\tilde{\Sigma} = T\hat{\Sigma}/(T-1)$	−0.221	−0.104	−0.023	0.036
3rd Plug-in, $\breve{\Sigma} = T\hat{\Sigma}/(T-N-2)$	−0.174	−0.072	0.000	0.054
Bayesian (diffuse prior)	−0.170	−0.069	0.002	0.055
Parameter-free optimal two-fund	−0.132	−0.043	0.022	0.070
Estimated optimal two-fund	0.133	0.157	0.177	0.194
Uncertainty aversion two-fund	0.017	0.024	0.032	0.040
Global minimum-variance	0.079	0.089	0.096	0.101
Jorion's shrinkage	0.117	0.155	0.182	0.203
Estimated optimal three-fund	0.143	0.169	0.189	0.206