Fund Performance Evaluation

Professor B. Espen Eckbo

2011

The Problem

- We would like to be able to compare portfolio managers or portfolio strategies
  - People often use simple measures
    - Benchmark portfolios, e.g. S&P 500
    - Rank relative to other portfolios/funds
  - On average, a manager can beat simple measures by increasing risk
    - Holding riskier stocks
    - Increasing leverage
  - Since investors are risk averse, we would like measures that penalize managers when they take more risk
  - Usually, simple measures are noisy; they provide little information

- *If fund managers are generating high returns simply by taking on more risk, they should not be paid for it*
  - You should only be willing to pay the manager for the extra returns that she generates in excess of what you could have generated yourself with an implementable ex ante strategy
- You can yourself
  - lever up an index fund
  - buy small stocks
  - buy high book-to-market stocks
  - buy momentum stocks

- The only way for a manager to generate <u>value</u> is by
  - market timing
  - factor timing
  - characteristics timing
  - stock selectivity
- *The past (average) return of a fund is a poor measure of performance since it does not control for risk*
- Categorizations like “Growth Funds”, “Value Funds”, “Income Funds”, etc., are also inaccurate and are usually not proper reflections or the fund’s risk

Topics

- I. Measuring investment returns
  - Dollar-weighting vs time-weighting returns
  - Arithmetic vs geometric return
- II. CAPM-based performance measures
  - Sharpe ratio and M^2
  - Jensen's alpha
  - Treynor's measure and T^2
  - Appraisal ratio
- III. Market timing
- IV. APT-based performance measures
- V. Luck versus skill
- VI. Performance attribution

I. Measuring investment returns

- I.1 Dollar- vs time-weighted returns
  - Stock prices: t_0=\$50, t_1=\$53, t_2=\$54
  - Shares bought: 1, 1, 0
  - Shares sold: 0, 0, 2
  - Investment outlay: \$50, \$53, 0
  - Dividends (2 per share): 0, \$2, \$4
  - Sales proceeds (2 shares sold): 0, 0, \$108
- Dollar-weighted (internal) rate of return
  $50+53/(1+r)=2/(1+r)+112/(1+r)^2$
  - r=7.117%

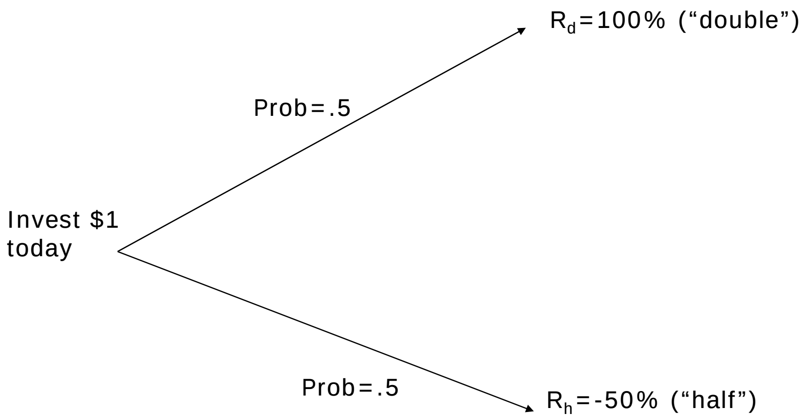
- The IRR effectively weighs each period's return with the dollar amount invested
- The average (time-weighted) return instead weighs each time period equally:
  - First period return: (53+2)/50 –1 =10%
  - Second period: (54+2)/53 –1 =5.66%
  - Average return: 7.83%
- Why is the average return > IRR here?

- Average return is preferable for constant-dollar portfolio strategies
  - Example: Open-end mutual fund manager who holds a fixed portfolio but has no control over the fund's inflow or outflow
  - Normally reported in fund industry
- IRR for dollar-varying portfolio strategies
  - Active management: manager varies dollar amount under investment over time
  - Need dollar-weighted performance measure to determine whether fund has ability to consistently time high future returns

- I.2 Arithmetic (simple) vs geometric (compound) average return
  - The geometric (or compound) return accounts for reinvestment of cash flows (dividends) on stock
- Generally:
  - $(1+r_G)=[(1+r_1)(1+r_2)...(1+r_T)]^{1/T}$
- In our example:
  - $(1+r_G)^2 = (1.10)(1.0566) = 1.0781$
  - r_G=7.81% (versus 7.83 arithmetic)
- Geometric < arithmetic
  - $r_G = r_A - (1/2)\sigma^2$, where σ^2 is return variance
  - Bad returns have greater influence on average

- Arithmetic or geometric?
- Geometric return is the constant return we would have needed to earn in each year to match actual performance over some period
- Thus, a good measure of past performance of a given dollar investment
- Arithmetic return better measure of expected future performance
- Example:

Investment return distribution



- Run this investment for two periods. Sequence of possible two-period returns
  - "Double, double": r_{dd} = [(2+2)-1)/1= 300%
  - "Double, halve": r_{dh} = [(2+(-1))-1]/1 = 0%
  - "Halve, double": r_{hd} = [(.5 +.5)–1]/1 = 0%
  - "Halve, halve": r_{hh} = [(.5+(-.25))-1]/1 = -75%
- $r_A=[r_{dd}+r_{dh}+r_{hd}+r_{hh}]/4=56.25\%$ over two years, or $(1.5625)^{1/2}-1=25\%$ per year
- $r_G=[(1+r_{dd})(1+r_{dh})(1+r_{hd})(1+r_{hh})]^{1/4}-1=0\%$ per year

II. CAPM-based performance measures

- Measures are motivated by the CAPM
  - Marginal investor holds the market
- Total risk perspective
  - If the fund is the only asset in our portfolio, then we care about total level of diversification
  - In this context, the measure should penalize total risk (variance)
- Marginal risk perspective
  - If the fund is one of many assets in our portfolio, then we care about what the fund adds to our total portfolio risk
  - In this context, the measure should penalize marginal risk
  - Marginal risk is defined using CAPM

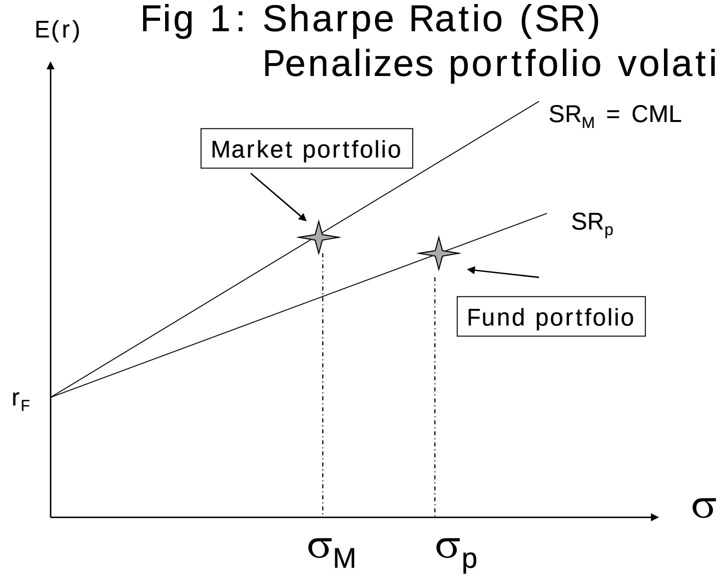
II.1 Sharpe ratio (SR)

- SR= slope of the CML:

$$SR_p = (r_p - r_F)/\sigma_p$$

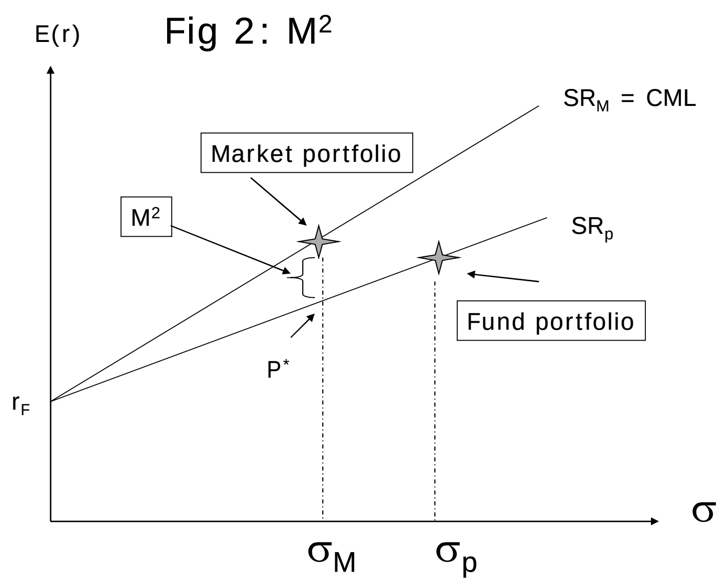
7/84 – 6/94; Monthly, r_F=0.34%	ROR %	STD %	SR_p
S&P500	1.20	4.55	0.19
Dean Witter Div Growth	1.14	3.79	0.21
Dreyfus Fund	0.87	3.69	0.14
Fidelity Magellan Fund	1.48	5.12	0.22
Janus Fund	1.22	3.99	0.22
Pioneer II	1.05	4.59	0.15
Putnam Growth & Income	1.18	3.61	0.23
Templeton World Fund	1.16	4.26	0.19
Twentieth Cent Select	1.09	5.17	0.15
Vanguard Index Tr 500	1.18	4.56	0.18
Windsor Fund	1.23	4.40	0.20

Fig 1: Sharpe Ratio (SR) Penalizes portfolio volatility



II.2: M^2 (Adjusted SR)

- Suppose your fund portfolio has a different total risk (σ_p) than the risk of the Market portfolio (σ_M)
- Define portfolio p^* as your fund portfolio leveraged or unleveraged so that $\sigma_{p*} = \sigma_M$
  - Invest (σ_p/σ_M) in p and $(1-\sigma_p/\sigma_M)$ in F
  - Ex: If $\sigma_p = (1.5)\sigma_M$: $p^* = (.67)p + (.33)F$
  - Ex: If $\sigma_p = (0.5)\sigma_M$: $p^* = (2.0)p + (-1.0)F$
- $M^2 = r_{p*} - r_M$
- Thus, measure is denoted in % return

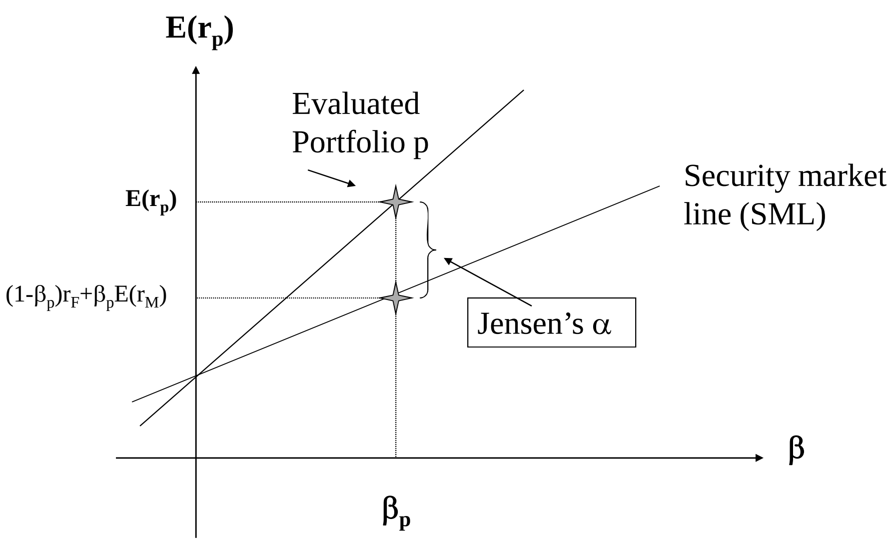


- The Sharpe Ratio is not appropriate for funds you are considering as part of a larger portfolio, or when you are deciding on how much to compensate managers
- In this case, you want to use a measure that considers the return relative to the systematic (marginal) risk of the portfolio
- These are measure based on the security market line (SML) and not the CML

- Three measures:
  - II.3 Jensen's Alpha
  - II.4 Treynor Measure
  - II.4 Appraisal Ratio
- Each of these measures asks how well the fund would have done relative to an efficient portfolio (consisting of the market and the risk-free asset) with the same systematic risk:

$$E(r_p) = r_F + \beta_p[E(r_M) - r_F]$$
$$= (1 - \beta_p) r_F + \beta_p E(r_M)$$

Fig 3: Jensen's Alpha



- Jensen's alpha is simply the intercept term in the market model regression:

  $r_{pt}-r_{Ft}=\alpha_p+\beta_p(r_{Mt}-r_{Ft}) + \varepsilon_{pt}$

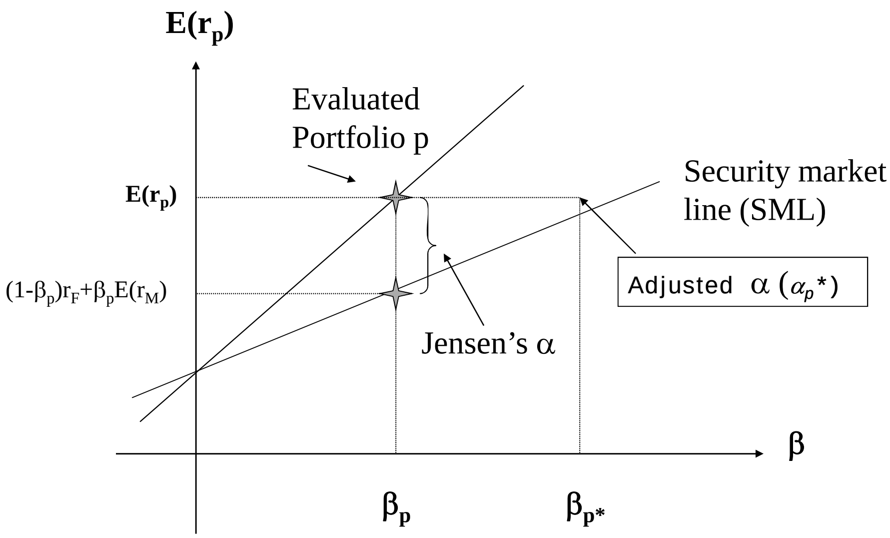
  - Alpha is in units of return. With α_p=.0015, measured using pre-expense returns, you may be willing to pay up to 0.15% per month (appr. 1.8% a year) in expenses
- If two portfolios have the same alpha but different betas, the portfolio with the lower beta is probably better
  - If you used leverage to equate the betas, the levered portfolio would have a higher alpha

Adjusting α_p for total risk

- Jensen's alpha is the maximum amount you should be willing to pay a manager
- Alpha doesn't account for the proportion of the portfolio's total risk σ_p that is non-systematic (the size of $\sigma_{\varepsilon p}$)
  - Compute the variance of the market model return on the previous slide

    $\sigma_p^2 = \beta_p^2 \sigma_M^2 + \sigma_\varepsilon^2$
  - Define portfolio p* as portfolio p with no diversifiable risk (so $\sigma_\varepsilon^2=0$)
  - $\beta_{p*}=\sigma_p/\sigma_M$. Then find α_{p*}

Fig 4: Adjusted Jensen's Alpha



The Treynor measure

- The Treynor measure is the slope of the SML for the portfolio

  $T_p = [E(r_p) – r_F]/\beta_p$

  - May be viewed as a way to control for the leverage problem in alpha
- "Treynor squared" measure:

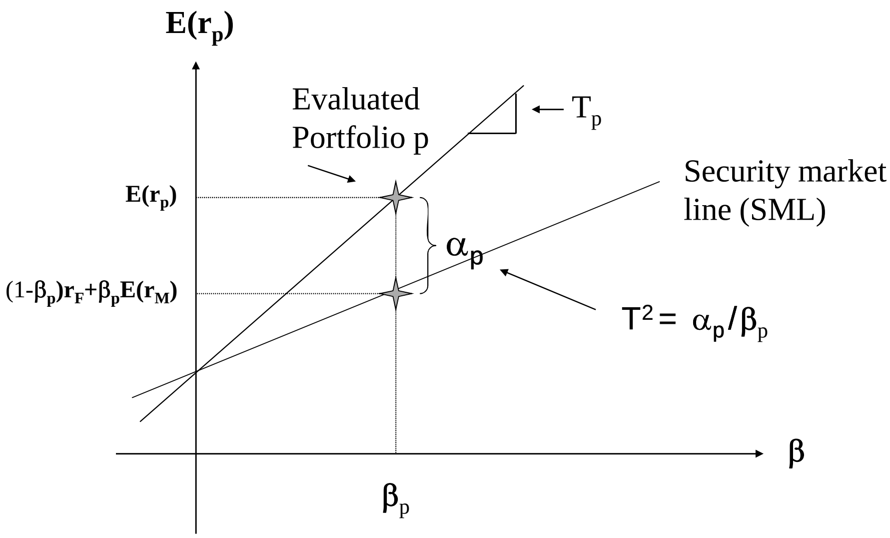
  $T^2 = T_p - T_M$

  $= [E(r_p) - r_F]/\beta_p - [E(r_M) – r_F]/\beta_M$

  $= \alpha_p/\beta_p$

  - T^2 is Jensen's alpha per unit of portfolio beta

Fig 5: Treynor Measure



The Appraisal Ratio

- The Appraisal Ratio scales Jensen's alpha with the amount of non-systematic risk:

$$AR_p = \alpha_p / \sigma_{\varepsilon p}$$

  - The appraisal ratio is like a cost/benefit ratio for a mispriced fund
  - The appraisal ratio can be used as a guide to how much of an asset/fund you want to add to your (otherwise diversified) portfolio

7/84 – 6/94; Monthly, r_F=0.34%	Beta	Alpha	T_p	T^2	$\sigma_{\varepsilon p}$	AR_p
S&P500	1.00	0.00	0.86	0.00	0.00	-
Dean Witter Div Growth	0.81	0.17	0.99	0.13	1.29	13.2
Dreyfus Fund	0.74	-0.02	0.72	-0.14	1.74	-1.2
Fidelity Magellan Fund	1.09	0.17	1.05	0.19	1.79	9.5
Janus Fund	0.80	0.26	1.10	0.24	1.88	13.8
Pioneer II	0.96	-0.11	0.74	-0.12	1.80	-6.1
Putnam Growth & Income	0.77	0.03	1.09	0.23	1.25	2.1
Templeton World Fund	0.85	0.14	0.96	0.10	2.04	6.9
Twentieth Cent Select	1.09	-0.22	0.69	-0.17	1.93	-11.4
Vanguard Index Tr 500	1.00	-0.03	0.84	-0.02	1.20	-
Windsor Fund	0.87	0.19	1.02	0.16	2.17	8.8

Information Ratio

- $IR=(r_p-r_I)/\sigma(r_p-r_I)$
  - Portfolio I is a reference portfolio or index
  - $\sigma(r_p-r_I)$ is the portfolio's "tracking error"
  - When I=M, IR relates directly to the Sharpe ratio and the appraisal ratio
  - Let I=M and use the market model to expand on both numerator and denominator in IR

Market model: $r_p-r_F=\alpha_p+\beta_p(r_M-r_F)+\varepsilon_p$

Subtract r_M-r_F: $r_p-r_M=\alpha_p+(\beta_p-1)(r_M-r_F)+\varepsilon_p$

Compute variance: $\sigma^2(r_p-r_M)=(\beta_p-1)^2\sigma^2(r_M-r_F)+\sigma^2_{\varepsilon p}$

$$IR = [\alpha_p+(\beta_p-1)(r_M-r_F)]/[(\beta_p-1)^2\sigma^2(r_M-r_F)+\sigma^2_{\varepsilon p}]^{1/2}$$

- If $\beta_p=1$ (so IR driven by nonsystematic risk):
  - $IR=\alpha_p/\sigma_{\varepsilon p}$
  - appraisal ratio
  - IR>0 driven by "stock picking"
- If β_p‡1 and $\alpha=\varepsilon=0$(so IR reflects different systematic risk:
  - $IR=(r_M-r_F)/\sigma(r_M-r_F)=(r_M-r_F)/\sigma(r_M)$
  - Sharpe ratio of M
  - IR>0 driven by "beta tilting" or "tactical allocation"
  - IR increases in beta if market risk premium positive
- So, IR>0 may be simply due to a high beta

Multiple managers

- IR often used with the tracking error $\sigma(r_p-r_I)$ to control performance of individual portfolio managers
- A manager's "alpha contribution": $\alpha_p= \sigma_{\varepsilon p}IR$
- Comparison across managers presume their unsystematic risks are uncorrelated
- A positive IR can also reduce the overall portfolio's Sharpe ratio, if IR is small and tracking error large

Statistical significance of IR

- Define $t(IR) = [\text{Average}(r_p - r_M)] / s(r_p - r_M)$
  - $s(r_p - r_M)$ is the standard error of the average difference return
  - If the individual difference returns are independent and drawn from a stationary distribution, then

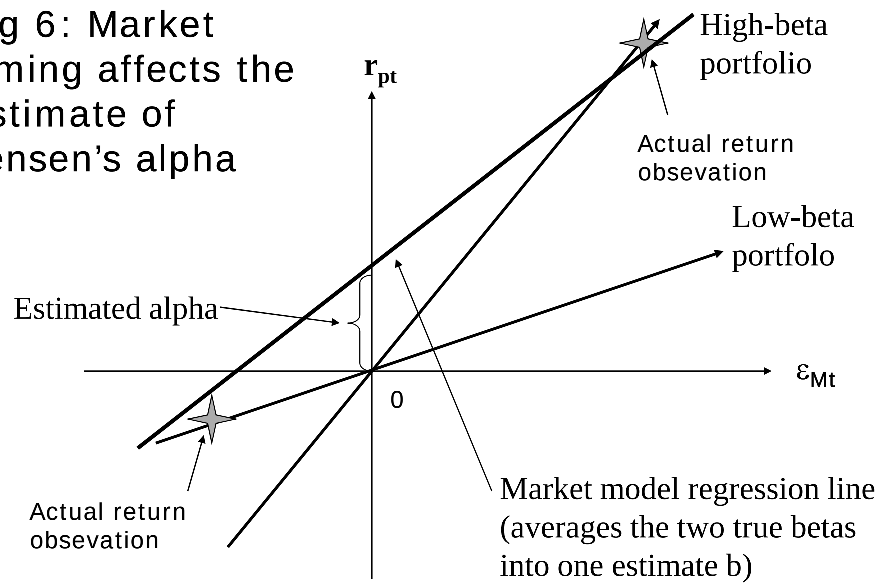
    $s(r_p - r_M) = \sigma(r_p - r_M)/\sqrt{T}$

    where $\sigma(r_p - r_M)$ is the standard deviation of each observation on $r_p - r_M$
  - In this case, $t'(IR) = t(IR)\sqrt{T}$ which has a student t distribution with T-1 degrees of freedom

III. Market timing

- You believe you can forecast the market and that other investors are not forecasting correctly
  - You may be using variables like the dividend yield, business cycle indicators and macroeconomic analysis to forecast returns
- You shift funds between a market index portfolio and the riskfree asset based on your forecasts
- Successful timing changes the estimate of Jensen's alpha

Fig 6: Market timing affects the estimate of Jensen's alpha



- When using CAPM as your benchmark:

$$r_p - r_F = \beta_{Mt}([E(r_M) + \varepsilon_{Mt}] - r_F) + \varepsilon_p$$

- Even if ε_{Mt} is on average zero, the market timer's average return would be

$$E(r_p - r_F) = \beta^*_M [E(r_M) - r_F] + cov(\varepsilon_{Mt}, \beta_{Mt})$$

- Here, β^*_M is the time series average portfolio beta. You would earn superior returns if $cov(\varepsilon_{Mt}, \beta_{Mt})$ is positive

- Treynor and Mazuy suggest the following regression:

$r_{pt}-r_{Ft} = \alpha_p+\beta_p(r_{Mt}-r_{Ft}) + c_p(r_{Mt}-r_{Ft})^2 + \varepsilon_{pt}$

- Here, alpha is unbiased for selectivity
- The value of timing equals

$cov(\varepsilon_{Mt}, \beta_{Mt}) = c_p\sigma^2_M$

- Problem: If the funds holds option-like securities, then c_p may be non-zero even without timing ability
- Alternative: Estimate beta for periods of increasing and decreasing markets separately

IV. APT-based measures

- The measures extend from CAPM to APT
- The Sharpe measure is obviously the same since the definition of total risk has not changed
- The intuition for Jensen's alpha is the same, but now alpha is the constant term in the APT regression
- The Appraisal Ratio is as before
- We now get a Treynor measure for each risk factor in the APT regression

- There is now also a timing coefficient for each factor in the APT regression
- Two-factor example:

$$r^e_{pt} = \alpha_p + \beta^e_{p1} r^e_{1t} + \beta^e_{p2} r^e_{2t} + c_p (r^e_{1t})^2 + c_p (r^e_{2t})^2 + \varepsilon_{pt}$$

- In this model, the maximum a manager should be compensated is equal to:

  Max compensation: $\alpha_p + c_{p1}\sigma^2_1 + c_{p2}\sigma^2_2$
- These measures ask the question of whether managers beat the well-diversified benchmark portfolio with the same factor loadings

Characteristics-based measures

- Evidence that firm-specific characteristics such as size, book-to-market ratio, and momentum help determine expected return (in addition to the market factor)
- Thus, one may compare fund performance to a benchmark portfolio consisting of a random selection of stocks with the same value for these characteristics
- One difficulty with this approach is that it requires that you have the portfolio weights of the fund on (ideally) a monthly basis

- Note (again) that characteristics based performance measures imply that fund managers should not be rewarded for taking on value or momentum stocks
- Daniel-Grinblatt-Titman-Wermers (JF 1997) estimate characteristics-based selectivity measures
- They find that a number of managers appear to have a positive alpha when using CAPM, but that this is due to the funds buying high momentum stocks
- Little or no selectivity after controlling for momentum

Evaluating Fidelity's Magellan Fund

- Use the Fama-French three-factor model (06/77-12/99)
  $R_{magellan,t}-r_{F,t}=a+1.11(r_{M,t}r_{F,t})+0.5SMB_t+0.05HML_t+e_t$
- Factor-contribution to expected return
  - Market risk: 1.11(0.42)=0.47%
  - Size risk (SMB): 0.50(.027)=0.14%
  - Distress risk (HML): 0.05(0.44)=0.02%
  - Sum (Predicted monthly excess return): 0.63%
- Compare actual to predicted return
  - Actual monthly return: 1.36%
  - Predicted monthly return: 0.63%
  - Difference (Jensen's alpha): 0.73% per month (9% annually)

Magellan Fund regressions

	alpha	b(M)	b(SMB)	b(HML)	R^2
Total	0.38	1.11	0.10	0.03	.89
77-04	(.10)	(.02)	(.03)	(.04)	
	3.80				
P Lynch	0.81	1.13	0.50	-0.02	.94
77-90	(.13)	(.03)	(.05)	(.06)	
	6.22				
After L	0.002	1.01	-0.08	-0.03	.93
90-04	(.10)	(.02)	(.03)	(.03)	
	0.02				

V. Luck versus skill

- Generally, given the variability of stock returns, you need a lot of data to be able to decide whether good (or bad) portfolio performance is luck or skill
- Illustration using Jensen's alpha
  - Suppose we estimate α_p by regressing fund excess returns on the market:

    $r_p - r_F = \alpha_p + \beta_p(r_M - r_F) + e_p$

- Standard error of the alpha-estimate:
  - $SE(\alpha_p) \approx \sigma(e_p)/\sqrt{T}$
    where T is the number of time periods

- The t-statistic is
  - $t(\alpha_p) = \alpha_p\sqrt{T}/\sigma(e_p)$

- Q: How much data (T) do we need to conclude whether measured superior (or inferior) performance is luck or skill?

- A: A LOT!

Example: Inferring luck v. skill

- Suppose
  - The monthly return distribution of your portfolio has a constant mean, beta and alpha
  - $\alpha_p = .2\%$, $\beta_p = 1.2$, $\sigma(e_p) = 2\%$
  - $\sigma(r_M) = 6.5\%$, $\beta_p^2\sigma^2(r_M) = 60.84$. so the correlation coefficient with r_M is:
    $\{\beta_p^2\sigma^2(r_M)/[\beta_p^2\sigma^2(r_M) + \sigma^2(e_p)]\}^{1/2} = 0.97$
  - Thus, the portfolio is highly diversified

Example…

- We want to infer the portfolio (Jensen's) alpha by regressing portfolio excess return on market excess return.
- Suppose there is NO estimation error in alpha and beta (a fact not known to the evaluator)
- Evaluator computes the t-value to infer significance of the estimated alpha of .2%, and requires a significance level of 5% (two-sided test):

  $t(\alpha_p) = \alpha_p \sqrt{T} / \sigma(e_p)$

  $1.96 = .2\sqrt{T}/2$

  T=384 months or 32 years (!)

  The average tenure of a fund manager is 4.5 years

Example…

- The example is biased in favor of finding the positive (true) alpha as we have assumed away such statistical complications as parameter non-stationarity and that the regression estimates are perfect (no sampling error).
- What is the probability that the estimated alpha of .2% per month is due to luck of the draw and that the true alpha is zero?
- The alpha exceeds zero by .2/2=.1 standard deviations. From the t-distribution table, the probability of such an event (if random) is 46%.

Example with parameter estimation risk

- The estimated standard deviation of the $s(e_p)$, is a function of the standard deviation of the market return (s_M), the portfolio estimated beta (b_p), and the R^2 from the market model:

  $R^2 = b_p^2 s_M^2 / s_p^2$

  $\quad = b_p^2 s_M^2 / [b_p^2 s_M^2 + s^2(e_p)]$

Or: $R^2[b_p^2 s_M^2 + s^2(e_p)] = b_p^2 s_M^2$

Or: $s^2(e_p) = b_p^2 s_M^2 (1-R^2)/R^2$

- The t-statistic is
  - $t(\alpha_p) = \alpha_p \sqrt{T} / \sigma(e_p)$

    $\quad = \alpha_p \sqrt{T} / \beta_p \sigma_M \sqrt{(1-R^2)/R^2}$

  Or,

  $T = t(\alpha_p)^2 \beta^2_p \sigma^2_M (1-R^2) / \sigma^2_p R^2$

- Suppose the true (annual) alpha is 3%
  - Annual standard deviation of the market is 15%
  - How many years of data do we need so the (expected) t-statistic is greater than 2?

Correlation between R_p and R_M						
Beta	0.10	0.25	0.50	0.75	0.90	0.95
0.5	2475	375	75	19	6	3
1.0	9900	1500	300	78	24	11
1.5	22,275	3375	675	175	53	24

VII. Performance attribution

- What decisions led to superior or inferior performance?
- What caused deviations from the benchmark,
  - the allocation across asset classes, or
  - the selection of securities within classes?
- Partition securities into N asset classes
  - Each class has a weight w_{Bi} in the benchmark and w_{pi} in the portfolio
  - Each class has a return r_{Bi} in the benchmark and r_{pi} in the portfolio

- Returns on benchmark and portfolio:
  - $r_B = \Sigma(w_{Bi} r_{Bi})$
  - $r_p = \Sigma(w_{pi} r_{pi})$
- Return difference:
  - $r_p - r_B = \Sigma(w_{pi} r_{pi} - w_{Bi} r_{Bi})$

$$= \Sigma(w_{pi} - w_{Bi}) r_{Bi} \qquad \text{(Asset allocation)}$$

$$+ \Sigma w_{Bi}(r_{pi} - r_{Bi}) \qquad \text{(Security selection)}$$

- Eckbo-Smith (1998) tests for significance of the asset allocation component using insider trades on the Oslo Stock Exchange (discussed tomorrow)

Summary

- Performance measurement tries to determine how well the fund manager did relative to a “comparable” benchmark portfolio
- The definition of “comparable” depends on the model you subscribe to
- With the CAPM, the comparable portfolio is the market levered up or down to get the same beta as the fund. Jensen’s alpha is the average difference between the return on the fund and the return on this portfolio

- Under the APT, the “comparable” portfolio is a passive portfolio with exactly the same loadings on all factors as the fund
- If you believe characteristics determine expected returns, then the comparable portfolio is a passive (randomly selected) portfolio whose holdings have the same size, book-to-market ratios, and momenta
- Finally, the performance measure that you use depend on the use to which your managed portfolio will be put