F500 Empirical Finance
Lecture 5: The Capital Asset Pricing Model

Oliver Linton
obl20@cam.ac.uk

February 12, 2020

Outline

1. Mean Variance Portfolio Choice
2. The Capital Asset Pricing Model
3. Market Model
    1. Estimates for Dow
4. Maximum Likelihood Estimation and Testing
5. Cross-Sectional Regression Tests
6. Some discussion

Reading: Linton (2019), Chapter 7

We have cross section of risky assets with random return R_i, $i = 1, \ldots, N$. We suppose that

$$ER_i = \mu_i, \quad \mathrm{var}(R_i) = \sigma_{ii} > 0$$

$$\mathrm{cov}(R_i, R_j) = \sigma_{ij}$$

- Mean/Variance Portfolio choice. Choose weights w_i such that $\sum_{i=1}^{N} w_i = 1$ to find the mean variance efficient frontier (the achievable risk/return trade-off), either
  - Minimize portfolio variance for given portfolio mean, or
  - Maximize portfolio mean for given portfolio variance.
- Individual Mean Variance preferences $U(\mu(w), \sigma^2(w))$ determine which point on the frontier to choose.
- We may combine with a risk free asset with return R_f (but consider this explicitly later).
- CAPM leads to restrictions on the risk/return trade-off taking account of the market portfolio whose return is R_m

Let $R = (R_1, \ldots, R_N)^\intercal$ be the $N \times 1$ vector of returns with mean vector

$$ER = \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix}$$

and covariance matrix

$$E\left[(R-\mu)(R-\mu)^\intercal\right] = \Sigma = \begin{bmatrix} \sigma_{11} & & & \\ & \ddots & \sigma_{jk} & \\ & & \ddots & \\ & & & \sigma_{NN} \end{bmatrix}.$$

We consider the problem of mean variance portfolio choice in this general setting. Let R_w denote the random portfolio return

$$R_w = \sum_{j=1}^{N} w_j R_j = w^\intercal R,$$

where $w = (w_1, \ldots, w_N)^\intercal$ are weights with $\sum_{j=1}^{N} w_j = 1.$ The mean and variance of the portfolio are

$$\mu_w = w^\intercal \mu = \sum_{j=1}^{N} w_j \mu_j \quad ; \quad \sigma_w^2 = w^\intercal \Sigma w = \sum_{j=1}^{N} \sum_{k=1}^{N} w_j w_k \sigma_{jk}.$$

There is a trade-off between mean and variance, meaning that as we increase the portfolio mean, which is good, we end up increasing its variance, which is bad.

- To balance these two effects we choose a portfolio that minimizes the variance of the portfolio subject to the mean being a certain level.
- We first consider the **Global Minimum Variance** (GMV) portfolio.

Definition

The Global Minimum Variance portfolio w is the solution to the following minimization problem

$$\min_{w \in \mathbb{R}^N} w^\intercal \Sigma w \qquad \text{subject to } w^\intercal i = 1,$$

where $i = (1, \ldots, 1)^\intercal$.

We assume that the matrix Σ is nonsingular, so that Σ^{-1} exists with $\Sigma^{-1}\Sigma = \Sigma\Sigma^{-1} = I$ (otherwise, there exists w such that $\Sigma w = 0$).

To solve this problem we form the Lagrangian, which is the objective function plus the constraint multiplied by the Lagrange multiplier λ

$$\mathcal{L} = \frac{1}{2} w^{\intercal} \Sigma w + \lambda (1 - w^{\intercal} i).$$

This has first order condition

$$\frac{\partial \mathcal{L}}{\partial w} = \Sigma w - \lambda i = 0 \Longrightarrow w = \lambda \Sigma^{-1} i.$$

Then premultiplying by the vector i and using the constraint we have $1 = i^{\intercal} w = \lambda i^{\intercal} \Sigma^{-1} i$, so that $\lambda = 1/i^{\intercal} \Sigma^{-1} i$ and the optimal weights are

$$w_{GMV} = \frac{\Sigma^{-1} i}{i^{\intercal} \Sigma^{-1} i}.$$

This portfolio has mean and variance

$$\mu_{GMV} = \frac{i^{\intercal} \Sigma^{-1} \mu}{i^{\intercal} \Sigma^{-1} i} \quad ; \quad \sigma^2_{GMV} = \frac{1}{i^{\intercal} \Sigma^{-1} i}.$$

The global minimum variance portfolio may sacrifice more mean return than you would like so we consider the more general problem where we ask for a minimum level m of the mean return.

Definition

The portfolio that minimizes variance for a given level m of mean return solves

$$\min_{w \in \mathbb{R}^N} w^\intercal \Sigma w$$

subject to the constraints $w^\intercal i = 1$ and $w^\intercal \mu = m$.

The Lagrangian is

$$\mathcal{L} = \frac{1}{2} w^\intercal \Sigma w + \lambda(1 - w^\intercal i) + \gamma(m - w^\intercal \mu).$$

The first order condition is

$$\frac{\partial \mathcal{L}}{\partial w} = \Sigma w - \lambda i - \gamma \mu = 0,$$

which yields

$$w_{opt} = \lambda \Sigma^{-1} i + \gamma \Sigma^{-1} \mu \in \mathbb{R}^N,$$

where $\lambda, \mu \in \mathbb{R}$ are the two Lagrange multipliers. Then imposing the two restrictions: $1 = i^{\intercal} w_{opt} = \lambda i^{\intercal} \Sigma^{-1} i + \gamma i^{\intercal} \Sigma^{-1} \mu$ and $m = \mu^{\intercal} w_{opt} = \lambda \mu^{\intercal} \Sigma^{-1} i + \gamma \mu^{\intercal} \Sigma^{-1} \mu$, we obtain a system of two equations in λ, γ, which can be solved exactly to yield

$$\lambda = \frac{C - Bm}{\Delta} \quad ; \quad \gamma = \frac{Am - B}{\Delta}$$

$$A = i^{\intercal} \Sigma^{-1} i,\ B = i^{\intercal} \Sigma^{-1} \mu,\ C = \mu^{\intercal} \Sigma^{-1} \mu,\ \Delta = AC - B^2,$$

provided $\Delta > 0$. This portfolio has mean m and variance

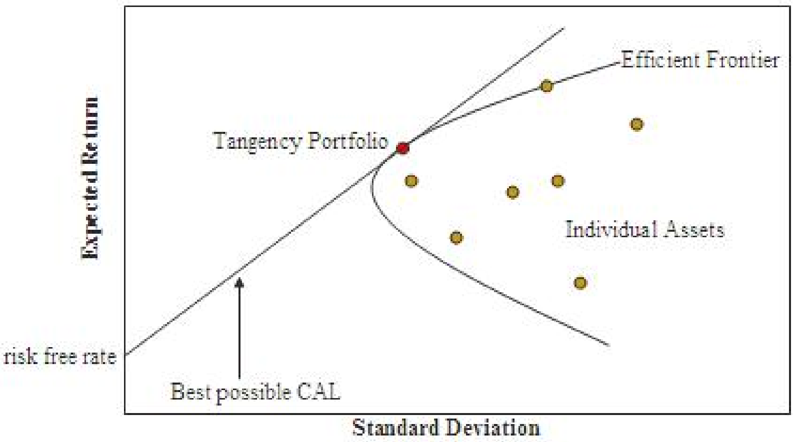
$$\sigma^2_{opt}(m) = \frac{Am^2 - 2Bm + C}{\Delta},$$

which is a quadratic function of m.

- Individual Mean Variance preferences $U(\mu(w), \sigma^2(w))$ determine which point on the frontier to choose.
- For example

$$U(\mu(w), \sigma^2(w)) = \mu(w) - \gamma\sigma^2(w),$$

  where γ is some risk aversion parameter
- Now suppose there is a risk free rate (an asset with zero variance). Can show that optimal choice involves a combination of the tangency portfolio and the riskless asset depending on risk preferences


Efficient Frontier
Tangency Portfolio
Expected Return
Individual Assets
risk free rate
Best possible CAL
Standard Deviation

Portfolio weights

- Tangency portfolio has weights that are proportional to

$$w_{TP} \propto \Sigma^{-1} (\mu - R_f i)$$

- Global Minimum Variance portfolio has weights proportional to

$$w_{GMV} \propto \Sigma^{-1} i$$

- Empirically, find many negative weights in both cases (short selling).

Annualized returns, std, and portfolio weights

	μ	σ	w_{GMV}	w_{TP}
Alcoa Inc.	-0.0665	0.2151	-0.0665	-0.0346
AmEx	0.0009	0.1851	-0.2475	-0.2482
Boeing	-0.0852	0.2350	-0.0006	0.0097
Bank of America	-0.0093	0.1540	0.0369	0.0377
Caterpillar	0.0375	0.1595	0.0715	0.0103
Cisco Systems	0.0140	0.1103	-0.0584	-0.0732
Chevron	-0.0110	0.2151	-0.1038	-0.1016
du Pont	-0.0088	0.1504	0.1374	0.1503
Walt Disney	-0.0046	0.1408	0.0820	0.0799
General Electric	-0.0782	0.2267	-0.0187	-0.0223
Home Depot	-0.0803	0.2200	0.0092	0.0220
HP	-0.0137	0.1937	-0.0978	-0.0907
IBM	-0.0296	0.2014	-0.0147	0.0065
Intel	0.0282	0.1318	0.1769	0.1463
Johnson[2]	-0.0512	0.2268	0.0706	0.0679

	μ	σ	w_{GMV}	w_{TP}
JP Morgan	0.0129	0.0991	0.1868	0.1834
Coke	-0.0577	0.2234	0.0353	0.0331
McD	-0.0188	0.1491	0.1096	0.1087
MMM	-0.0094	0.1458	0.0773	0.0903
Merck	-0.0754	0.1972	-0.0087	0.0127
MSFT	-0.0579	0.2092	-0.0332	-0.0227
Pfizer	-0.1093	0.2057	-0.0176	0.0062
Proctor & Gamble	0.0309	0.1045	0.3443	0.2999
AT&T	-0.0042	0.1327	0.0237	0.0364
Travelers	-0.0234	0.1703	-0.0043	0.0092
United Health	0.0190	0.2132	0.0342	0.0326
United Tech	-0.0054	0.1852	-0.0092	-0.0197
Verizon	0.0075	0.1280	0.0299	0.0071
Wall Mart	0.0179	0.1268	0.1877	0.1858
Exxon Mobil	0.0070	0.1470	0.0680	0.0770

The Capital Asset Pricing Model (Risk return trade-off)

- Sharpe-Lintner version with a riskless asset (borrowing or lending)

$$E[R_i] - R_f = \beta_i (E[R_m] - R_f)$$

  for all i.

- Relates three quantities

$$\pi_i = E[R_i - R_f] \quad ; \quad \pi_m = E[R_m - R_f] \quad ; \quad \beta_i = \frac{\mathrm{cov}(R_i, R_m)}{\mathrm{var}(R_m)}$$

  all of which can be estimated from time series data

- Risk/return trade-off - more risk, more return. The β_i is the relevant measure of riskiness of stock i not $\mathrm{var}(R_i)$

- Fisher Black version without a riskless asset
- Find the (zero beta) portfolio with return R_0 such that

$$R_0 = R_{w_0} = \arg \min_{w:\mathrm{cov}(R_w, R_m)=0} \mathrm{var}(R_w)$$

- Then

$$E[R_i] - E[R_0] = \beta_i (E[R_m] - E[R_0])$$

  for all i.
- Since R_0 is not observed, its mean has to be treat as an unknown parameter and estimated

Testable versions embed within some class of alternatives.
Sharpe-Lintner. Letting $Z_i = R_i - R_f$ and $Z_m = R_m - R_f$

$$E[Z_i] = \alpha_i + \beta_i E[Z_m]$$

and test

$$H_0 : \alpha_i = 0 \text{ for all } i \quad vs \quad H_A : \alpha_i \neq 0 \text{ for some } i$$

Black. We have

$$E[R_i] = \alpha_i + \beta_i E[R_m]$$

and test

$$H_0 : \alpha_i = (1 - \beta_i) E[R_0] \text{ for all } i \quad vs \quad H_A : \alpha_i \neq (1 - \beta_i) E[R_0] \text{ for some } i$$

Here, R_0 is the return on the (unobserved) zero beta portfolio

Market Model

We have a time series sample on each asset, the market portfolio, and (sometimes) the risk free rate $\{R_{it}, R_{mt}, R_{ft},\ i = 1, \ldots, N,\ t = 1, \ldots, T\}$.

Definition

For $Z_{it} = R_{it} - R_{ft}$ or $Z_{it} = R_{it}$ and $Z_{mt} = R_{mt} - R_{ft}$ or $Z_{mt} = R_{mt}$:

$$Z_{it} = \alpha_i + \beta_i Z_{mt} + \varepsilon_{it},$$

$$E(\varepsilon_{it} | Z_{m1}, \ldots, Z_{mT}) = 0 \ ; \ E(\varepsilon_{it}\varepsilon_{js} | Z_{m1}, \ldots, Z_{mT}) = \begin{cases} \sigma_{ij} & \text{if } t = s \\ 0 & \text{else} \end{cases} .$$

- Convention here is to allow for cross-sectional correlation in the idiosyncratic terms.
- We either
  - assume that $\varepsilon_{1t}, \ldots, \varepsilon_{Nt}$ are normally distributed
  - or we assume that the sample size T is large

Evidence for Normality

- Early proofs of the CAPM often assumed joint normality of returns, but later it was shown that it can hold under weaker distributional assumptions. Nevertheless, much of the literature uses exact tests based on assumption of normality.
- Measures of non-normality: skewness and excess kurtosis

$$\kappa_3 \equiv E\left[\frac{(r-\mu)^3}{\sigma^3}\right]$$

$$\kappa_4 \equiv E\left[\frac{(r-\mu)^4}{\sigma^4}\right] - 3$$

  For a normal distribution $\kappa_3, \kappa_4 = 0.$ Daily stock returns typically have negative skewness and large positive kurtosis.
- Fama for example argues that monthly returns are closer to normality

Theorem

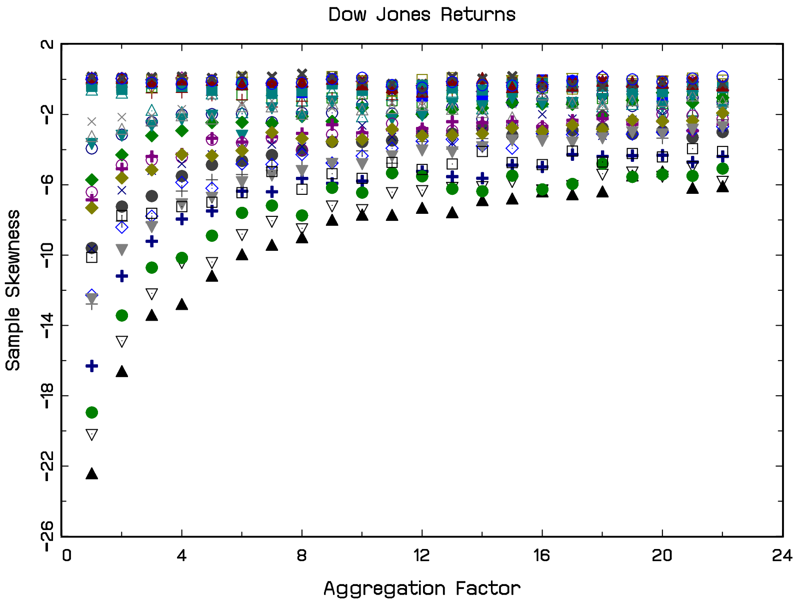
(Aggregation of (logarithmic) returns). Let A be the aggregation (e.g., weekly, monthly) level such that $r_A = r_1 + \cdots + r_A$. Then under RW1

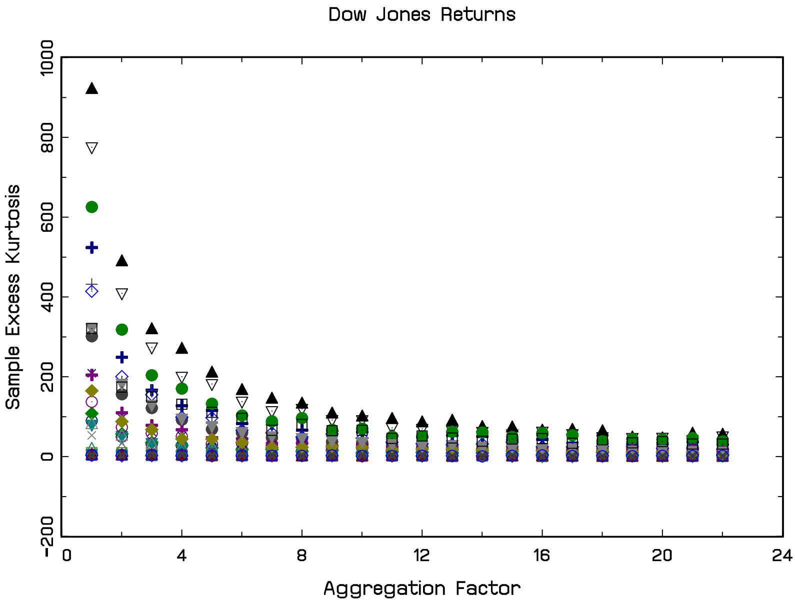
$$Er_A = AEr \quad \mathrm{var}(r_A) = A\mathrm{var}(r)$$

$$\kappa_3(r_A) = \frac{1}{\sqrt{A}}\kappa_3(r)$$

$$\kappa_4(r_A) = \frac{1}{A}\kappa_4(r).$$

This says that as you aggregate more ($A \to \infty$), returns become more normal, i.e., $\kappa_3(r_A) \to 0$ and $\kappa_4(r_A) \to 0$ as $A \to \infty$. Similar result for martingale difference sequence case.


Dow Jones Returns
Sample Skewness
Aggregation Factor
2
-2
-6
-10
-14
-18
-22
-26
0
4
8
12
16
20
24


Dow Jones Returns
Sample Excess Kurtosis
1000
800
600
400
200
0
-200
0
4
8
12
16
20
24
Aggregation Factor

Market Model (S&P500-tbill) Daily estimates 1990-2013

	α	se(α)	β	se(β)	$se_W(\beta)$	R^2
Alcoa Inc.	-0.0531	0.0480	1.3598	0.0305	0.0400	0.3929
AmEx	0.0170	0.0332	1.4543	0.0211	0.0317	0.6073
Boeing	-0.0644	0.0492	1.6177	0.0312	0.0667	0.4661
Bank of America	-0.0066	0.0341	0.9847	0.0217	0.0301	0.4020
Caterpillar	0.0433	0.0335	1.0950	0.0213	0.0263	0.4635
Cisco Systems	0.0060	0.0264	0.6098	0.0168	0.0272	0.3005
Chevron	0.0017	0.0486	1.3360	0.0308	0.0401	0.3793
du Pont	-0.0101	0.0358	0.8422	0.0228	0.0314	0.3084
Walt Disney	-0.0009	0.0281	1.0198	0.0178	0.0241	0.5159
General Electric	-0.0732	0.0575	1.0650	0.0365	0.0296	0.2169
Home Depot	-0.0720	0.0534	1.1815	0.0339	0.0418	0.2835
HP	-0.0083	0.0462	1.0797	0.0294	0.0315	0.3055
IBM	-0.0232	0.0483	1.1107	0.0307	0.0349	0.2992
Intel	0.0282	0.0280	0.8903	0.0178	0.0242	0.4490
Johnson[2]	-0.0400	0.0539	1.2803	0.0342	0.0360	0.3132

	α	se(α)	β	se(β)	$se_W(\beta)$	R^2
JP Morgan	0.0041	0.0231	0.5810	0.0147	0.0226	0.3382
Coke	-0.0380	0.0456	1.5811	0.0290	0.0595	0.4927
McD	-0.0270	0.0392	0.6012	0.0249	0.0237	0.1598
MMM	-0.0117	0.0349	0.8093	0.0221	0.0228	0.3032
Merck	-0.0783	0.0519	0.7893	0.0329	0.0268	0.1575
MSFT	-0.0536	0.0521	1.0427	0.0331	0.0408	0.2444
Pfizer	-0.1121	0.0545	0.7912	0.0346	0.0256	0.1455
Proctor & Gamble	0.0221	0.0250	0.5804	0.0159	0.0230	0.3036
AT&T	-0.0065	0.0303	0.8076	0.0193	0.0264	0.3643
Travelers	-0.0209	0.0402	0.9750	0.0255	0.0410	0.3224
United Health	0.0173	0.0564	0.8278	0.0358	0.0563	0.1482
United Tech	-0.0029	0.0452	0.9779	0.0287	0.0298	0.2742
Verizon	0.0039	0.0296	0.7606	0.0188	0.0238	0.3472
Wall Mart	0.0141	0.0293	0.7555	0.0186	0.0249	0.3495
Exxon Mobil	0.0053	0.0349	0.8290	0.0222	0.0292	0.3128

Market Model (S&P500-tbill) Monthly estimates with standard errors

	α	se(α)	β	se(β)	$se_W(\beta)$	R^2
Alcoa Inc.	-0.0528	0.0499	1.4107	0.1759	0.2185	0.3179
AmEx	0.0193	0.0254	1.5594	0.0895	0.1045	0.6875
Boeing	-0.0675	0.0499	1.6575	0.1759	0.2298	0.3915
Bank of America	0.0018	0.0311	1.2767	0.1097	0.1232	0.4955
Caterpillar	0.0454	0.0306	1.2212	0.1078	0.1300	0.4819
Cisco Systems	0.0134	0.0243	0.8418	0.0856	0.1114	0.4122
Chevron	0.0034	0.0472	1.4240	0.1663	0.1505	0.3468
du Pont	-0.0191	0.0318	0.5596	0.1120	0.1082	0.1532
Walt Disney	-0.0048	0.0258	0.9200	0.0908	0.0867	0.4264
General Electric	-0.0630	0.0568	1.4430	0.2002	0.2678	0.2734
Home Depot	-0.0715	0.0532	1.1787	0.1877	0.0992	0.2222
HP	0.0048	0.0435	1.4292	0.1535	0.1651	0.3858
IBM	-0.0270	0.0451	1.1867	0.1592	0.1274	0.2872
Intel	0.0299	0.0247	0.9255	0.0872	0.0935	0.4492
Johnson[2]	-0.0449	0.0550	1.1815	0.1940	0.1281	0.2118

	α	se(α)	β	se(β)	$se_W(\beta)$	R^2
JP Morgan	0.0046	0.0201	0.5912	0.0708	0.0738	0.3358
Coke	-0.0435	0.0428	1.4111	0.1508	0.1590	0.3880
McD	-0.0262	0.0364	0.6342	0.1284	0.1218	0.1501
MMM	-0.0129	0.0317	0.7906	0.1116	0.0814	0.2665
Merck	-0.0716	0.0506	0.9235	0.1784	0.2365	0.1627
MSFT	-0.0529	0.0571	1.1163	0.2012	0.1862	0.1824
Pfizer	-0.1135	0.0586	0.7811	0.2066	0.1804	0.0938
Proctor & Gamble	0.0255	0.0223	0.6626	0.0786	0.0838	0.3401
AT&T	-0.0111	0.0305	0.6668	0.1075	0.1069	0.2182
Travelers	-0.0255	0.0361	0.8321	0.1272	0.0977	0.2367
United Health	0.0186	0.0566	0.8556	0.1997	0.1751	0.1174
United Tech	-0.0047	0.0455	0.9470	0.1606	0.1620	0.2012
Verizon	0.0004	0.0277	0.6265	0.0976	0.1200	0.2301
Wall Mart	0.0142	0.0255	0.6802	0.0899	0.1026	0.2933
Exxon Mobil	0.0025	0.0339	0.7124	0.1196	0.1020	0.2044

Maximum Likelihood Estimation and Testing of Sharpe Lintner

- Suppose that with $Z_t = (Z_{1t}, \ldots, Z_{Nt})^\intercal$, $Z_{it} = R_{it} - R_{ft}$

$$Z_t = \alpha + \beta Z_{mt} + \varepsilon_t,$$

$$\varepsilon_t \sim N(0, \Omega_\varepsilon).$$

  Do not restrict Ω_ε to be diagonal.
- We require $N << T$. This essentially requires to work with portfolios rather than individual stocks. Recent work on large matrices attempts to address this.
- The Gaussian log likelihood is

$$\ell(\alpha, \beta, \Omega_\varepsilon) = c - \frac{T}{2} \log \det \Omega - \frac{1}{2} \sum_{t=1}^{T} (Z_t - \alpha - \beta Z_{mt})^\intercal \Omega_\varepsilon^{-1} (Z_t - \alpha - \beta Z_{mt})$$

The maximum likelihood estimates $\widehat{\beta}$ are the equation-by-equation time-series OLS estimates

$$\widehat{\beta}_i = \frac{\sum_{t=1}^{T}(Z_{mt} - \widehat{\mu}_m)(Z_{it} - \widehat{\mu}_i)}{\sum_{t=1}^{T}(Z_{mt} - \widehat{\mu}_m)^2} = \frac{1}{\widehat{\sigma}_m^2}\frac{1}{T}\sum_{t=1}^{T}(Z_{mt} - \widehat{\mu}_m)(Z_{it} - \widehat{\mu}_i)$$

$$\widehat{\mu}_m = \frac{1}{T}\sum_{t=1}^{T} Z_{mt} \quad ; \quad \widehat{\sigma}_m^2 = \frac{1}{T}\sum_{t=1}^{T}(Z_{mt} - \widehat{\mu}_m)^2, \qquad \widehat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T} Z_{it}$$

The maximum likelihood estimates of $\widehat{\alpha}$ are

$$\widehat{\alpha}_i = \widehat{\mu}_i - \widehat{\beta}_i \widehat{\mu}_m$$

The maximum likelihood estimate of Ω_ε is

$$\widehat{\Omega}_\varepsilon = \left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_{it}\widehat{\varepsilon}_{jt}\right)_{i,j}, \ \widehat{\varepsilon}_{it} = Z_{it} - \widehat{\alpha}_i - \widehat{\beta}_i Z_{mt}$$

Maximum Likelihood Estimation and Testing

Under the normality assumption we have, conditional on excess market returns $Z_{m1}, \ldots, Z_{mT}$, the exact distribution

$$\widehat{\alpha} \backsim N\left(\alpha, \frac{1}{T}\left(1 + \frac{\widehat{\mu}_m^2}{\widehat{\sigma}_m^2}\right)\Omega_\varepsilon\right)$$

Without normality (but under iid) we have for large T (taking $\mu_m = E(Z_{mt})$ and $\sigma_m^2 = \mathrm{var}(Z_{mt})$)

$$\sqrt{T}\left(\widehat{\alpha} - \alpha\right) \Longrightarrow N\left(0, \left(1 + \frac{\mu_m^2}{\sigma_m^2}\right)\Omega_\varepsilon\right)$$

Exact Test Statistic

Under normality, we can use the F-statistic

$$F = \frac{(T - N - 1)}{N} ((1 + \frac{\widehat{\mu}_m^2}{\widehat{\sigma}_m^2})^{-1}) \times \widehat{\alpha}^\intercal \widehat{\Omega}_\varepsilon^{-1} \widehat{\alpha}$$

whose null distribution is known exactly

$$F \sim F(N, T - N - 1)$$

Can use the $F(N, T - N - 1)$ critical values when normality holds and have an exact test.

Wald Test Statistic

- Wald test statistic for null hypothesis that $\alpha = 0$

$$W = T\left(1 + \frac{\widehat{\mu}_m^2}{\widehat{\sigma}_m^2}\right)^{-1} \widehat{\alpha}^\intercal \widehat{\Omega}_\varepsilon^{-1} \widehat{\alpha} = \frac{NT}{T-N-1} F$$

- Under the null hypothesis for large T

$$W \Longrightarrow \chi^2(N)$$

provided $N < T$. **The asymptotic approximation is valid regardless of whether the errors are normally distributed or not.**

- For the Dow stocks:
  - Daily W =22.943222 (p-value =0.81758677);
  - Monthly W =33.615606 (p-value =0.29645969)

Likelihood Ratio Test

- The likelihood ratio test is a natural alternative to a Wald test

$$LR = -2(\log \ell_c - \log \ell_u) = T[\log \det \widetilde{\Omega}_\varepsilon - \log \det \widehat{\Omega}_\varepsilon] \Longrightarrow \chi^2(N)$$

$$\widetilde{\Omega}_\varepsilon = \left(\frac{1}{T} \sum_{t=1}^{T} \widetilde{\varepsilon}_{it} \widetilde{\varepsilon}_{jt} \right)_{i,j}$$

where $\widehat{\varepsilon}_{it}^*$ are the constrained residuals ie the no intercept residuals.

$$\widetilde{\beta}_i = \frac{\sum_{t=1}^{T} Z_{mt} Z_{it}}{\sum_{t=1}^{T} Z_{mt}^2}, \quad \widetilde{\varepsilon}_{it} = Z_{it} - \widetilde{\beta}_i Z_{mt}$$

- The Wald, F, and LR tests have an exact relationship which allows us to derive the exact distribution for the likelihood ratio test under normality and large sample valid test otherwise.

CLM Table 5.3

Data 1965-1994. Monthly returns on ten value weighted portfolios based on size. CRSP value weighted index, 1 month tbill rate.

- Results for full period show rejection at 5% but not at 1% level.
- Five year subperiods: some rejections at 5% some not.
- Aggregate test statistics across subperiods assuming independence, ie if $T_1 \sim \chi^2(i)$ and $T_2 \sim \chi^2(j)$ (and mutually independent), then

$$T_1 + T_2 \sim \chi^2(j + i).$$

  Obtain very strong rejections. Likewise for ten year subperiods.
- The aggregated results allow for different parameter values across the subperiods but hide whether the evidence is getting stronger against the CAPM or not

Testing Black Version of the CAPM

Tests for the Black version are more complicated to derive.

- Estimate the same unconstrained model as before using total returns instead of excess returns, that is,

$$R_t = \alpha + \beta R_{mt} + \varepsilon_t$$

- The constrained model is

$$R_t = (i - \beta)\gamma + \beta R_{mt} + \varepsilon_t$$

  for some scalar unknown γ, where i is the N vector of ones. There are $N-1$ **nonlinear cross-equation** restrictions

$$\frac{\alpha_1}{1-\beta_1} = \cdots = \frac{\alpha_N}{1-\beta_N} = \gamma$$

- Estimating the constrained model requires numerical maximization of the nonlinear (in parameters) system of equations.

- Useful computational trick (**profiling** or **concentration**): assume that the expected return on the zero-beta portfolio, γ, is known exactly (use a noisy estimate as proxy)

$$R_t - \gamma i = \beta(R_{mt} - \gamma) + \varepsilon_t$$

so that conditionally on γ the model is linear in β.

- With the zero-beta return known, β can be estimated by OLS asset by asset
- Then, relax the assumption that the zero-beta return is known and search over the profiled likeliood for the best choice of γ

Testing Black Version of the CAPM

- We can show that the restricted estimates $\widetilde{\theta} = (\widetilde{\gamma}, \widetilde{\beta}^{\intercal})^{\intercal}$ satisfy $\sqrt{T}(\widetilde{\theta} - \theta)$ is asymptotically normal with mean zero and some variance matrix.
- The LR statistic compares the relative fit of the constrained and unconstrained models. As $T \to \infty$

$$LR = T[\log \det \widetilde{\Omega}_{\varepsilon} - \log \det \widehat{\Omega}_{\varepsilon}] \Longrightarrow \chi^2(N-1)$$

  where $\widetilde{\Omega}_{\varepsilon}$ is the MLE of Ω_{ε} in the constrained model. Note only $N-1$ degrees of freedom
- Exact theory for F statistic much more tricksy, require simulation methods.

Robustness of MLE and tests to Heteroskedasticity (RW2.5)

- Maximum likelihood estimation apparently assumes multivariate normal returns. CAPM can hold under weaker distributional assumptions (e.g., elliptical symmetry, which includes multivariate t-distributions with heavy tails).
- Actually, the Gaussian MLE of α, β is robust to heteroskedasticity, serial correlation, and non-normality since the estimates are just least squares.
- The asymptotic test statistics are robust to normality of the errors, but they are not robust to heteroskedasticity or serial correlation, and in that case we need to adjust the standard errors

- Suppose that

$$\begin{aligned} E(\varepsilon_t | Z_{mt}) &= 0 \\ E(\varepsilon_t \varepsilon_t^\intercal | Z_{mt}) &= \Omega_t, \end{aligned}$$

  where Ω_t is a potentially random time varying covariance matrix. This is quite a general assumption, but as we shall see below it is quite natural to allow for dynamic heteroskedasticity for stock return data.
- In this case, OLS is consistent but it is not possible to perform an exact test and the tests we already defined are unfortunately not properly sized in this case.
- However, we can construct robust Wald tests based on large sample approximations à la White (1980).

Define the Heteroskedastic Consistent standard errors

$$\widehat{\widehat{\Omega}}_T = \frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t\widehat{\varepsilon}_t^{\intercal} \; ; \; \widehat{\Psi}_T = \frac{1}{T}\sum_{t=1}^{T}\left(Z_{mt} - \widehat{\mu}_m\right)^2\widehat{\varepsilon}_t\widehat{\varepsilon}_t^{\intercal}$$

$$\widehat{V} = \widehat{\widehat{\Omega}}_T + \frac{\widehat{\mu}_m^2}{\widehat{\sigma}_m^4}\widehat{\Psi}_T$$

$$W_H = T\widehat{\alpha}^{\intercal}\widehat{V}^{-1}\widehat{\alpha}$$

Then, under the null hypothesis as $T \to \infty$

$$W_H \Longrightarrow \chi^2(N).$$

Cross-Sectional Regression Tests

- The CAPM says that

$$\mu_i = \beta_i \mu_m,$$

where $\mu_i = E\left(R_i - R_f\right)$ and $\mu_m = E\left(R_m - R_f\right)$.

- Fama and MacBeth (1973) embed this in a richer cross-sectional relationship

$$\mu_i = \gamma_0 + \beta_i \gamma_1.$$

We should find $\gamma_0 = 0$ and $\gamma_1 > 0$ with $\gamma_1 = \mu_m = E\left(R_m - R_f\right)$.

- In fact we should find $\gamma_0, \gamma_2, \gamma_3 = 0$ and $\gamma_1 > 0$ in

$$\mu_i = \gamma_0 + \beta_i \gamma_1 + \beta_i^2 \gamma_2 + \sigma_{\epsilon i}^2 \gamma_3$$

Problem: We don't observe β_i (or $\sigma^2_{\epsilon i}$). Solution

- First estimate β_i for each stock or portfolio using time series data.
- Then estimate the cross-sectional regression by OLS (or GLS)

$$\widehat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T}(R_{it} - R_{ft}) = \gamma_0 + \widehat{\beta}_i\gamma_1 + u_i$$

$$\widehat{\gamma}_1 = \frac{\sum_{i=1}^{N}\left(\widehat{\beta}_i - \overline{\widehat{\beta}}\right)\widehat{\mu}_i}{\sum_{i=1}^{N}\left(\widehat{\beta}_i - \overline{\widehat{\beta}}\right)^2}, \quad \widehat{\gamma}_0 = \overline{\widehat{\mu}} - \overline{\widehat{\beta}}\widehat{\gamma}_1$$

- Under the CAPM, $\gamma_0 = 0$ and $\gamma_1 = E(R_m - R_f)$. Test $\gamma_0 = 0$ by t-test
- In fact, they include additional variables, e.g.,

$$\frac{1}{T}\sum_{t=1}^{T}(R_{it} - R_{ft}) = \gamma_0 + \widehat{\beta}_i\gamma_1 + \widehat{\beta}_i^2\gamma_2 + \widehat{\sigma}^2_{\varepsilon i}\gamma_3 + u_i$$

  where $\widehat{\sigma}^2_{\varepsilon i}$ is an estimate of the idiosyncratic error variance. Test also $\gamma_j = 0,\ j = 2, 3$ using t-tests or Wald statistic.
- Standard errors?

Fama-MacBeth Standard errors

- Estimate the cross-sectional regressions at each time period to give a time series of risk premia

$$\widehat{\gamma}_{1t} = \frac{\sum_{i=1}^{N} \left(\widehat{\beta}_i - \overline{\widehat{\beta}}\right)(R_{it} - R_{ft})}{\sum_{i=1}^{N} \left(\widehat{\beta}_i - \overline{\widehat{\beta}}\right)^2}, \quad \widehat{\gamma}_{0t} = \widehat{\mu}_t - \overline{\widehat{\beta}}\widehat{\gamma}_1$$

$$\widehat{\mu}_t = \frac{1}{N} \sum_{i=1}^{N} (R_{it} - R_{ft})$$

- Then average the estimates over time

$$\widehat{\gamma} = \begin{bmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_1 \end{bmatrix} = \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} \widehat{\gamma}_{0t} \\ \widehat{\gamma}_{1t} \end{bmatrix}$$

- They estimate the asymptotic variance matrix of $\widehat{\gamma}$ by

$$\widehat{V} = \frac{1}{T} \sum_{t=1}^{T} \left(\begin{bmatrix} \widehat{\gamma}_{0t} \\ \widehat{\gamma}_{1t} \end{bmatrix} - \begin{bmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_1 \end{bmatrix} \right) \left(\begin{bmatrix} \widehat{\gamma}_{0t} \\ \widehat{\gamma}_{1t} \end{bmatrix} - \begin{bmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_1 \end{bmatrix} \right)^{\intercal}$$

Actual methodology even more complicated. There are four stages:

1. Time series estimation of individual stock betas in period A
2. Portfolio formation based on estimated double sorted individual stocks on beta and size
3. Estimate portfolio betas from the time series in period B
4. Estimate γ by the cross-sectional regression of the mean excess returns on the betas using data in period B. Standard errors are obtained from the cross-sectional regressions for each time period in B. Test hypothesis on average of time series of risk premia using standard errors from above

- Standard errors are widely used, but wrong unless combined with the portfolio grouping of a large original set of assets. Errors in variables/generated regressor issue. Shanken (1992) suggests an analytical correction necessary for individual stocks.
- Testing on portfolios (composed of a large number of individual stocks) rather than individual stocks can mitigate the errors-in-variable problem as estimation errors cancel out each other.
- Sorting by beta reduces the shrinkage in beta dispersion and statistical power; sorting by size takes into account correlation between size and beta;
- Performing pre-ranking and estimation in different periods avoids selection bias.

- Empirically they find that there is a positive and linear relationship between beta risk and return with a high R^2, but $\gamma_0 > 0$ and γ_1 significantly lower than market excess return.
- FM do not reject null

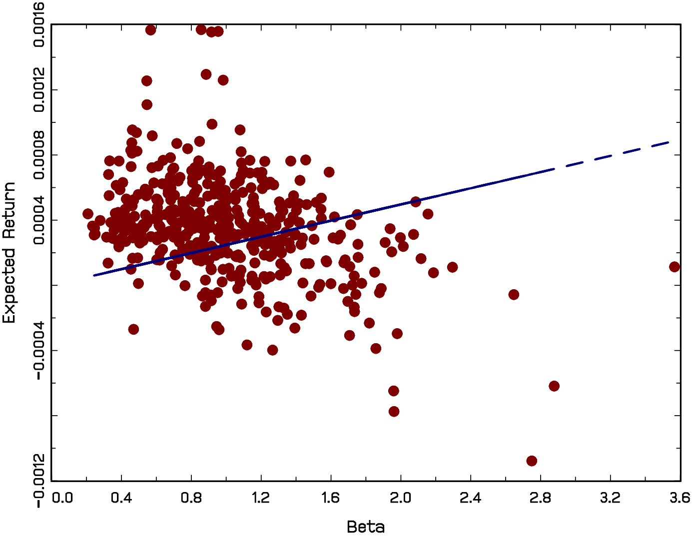


Figure: Risk Return Relation

Empirical Evidence in the Literature

Many tests and many rejections of the CAPM!!

- Size Effect. Market capitalization
  - Firms with a low market capitalization seem to earn positive abnormal returns ($\alpha > 0$), while large firms earn negative abnormal returns ($\alpha < 0$)
- Value effect. Dividend to price ratio (D/P) and book to market ratio (B/M).
  - Value firms (low value metrics relative to market value) have $\alpha > 0$ while growth stocks (high value metrics relative to market value) have $\alpha < 0$
- Momentum effect.
  - Winner portfolios outperform loser portfolios over medium term.

Time Varying Parameters

- The starting point of the market model and CAPM testing was that we have a sample of observations independent and identically distributed from a fixed population. This setting was convivial for the development of statistical inference. However, much of the practical implementations acknowledge time variation by working with short, say 5 year or 10 year windows.
- A number of authors have pointed out the variation of estimated betas over time. We show estimated betas for IBM (against the SP500) computed from daily stock returns using a five year window over the period 1962-2017. We present the rolling window estimates.

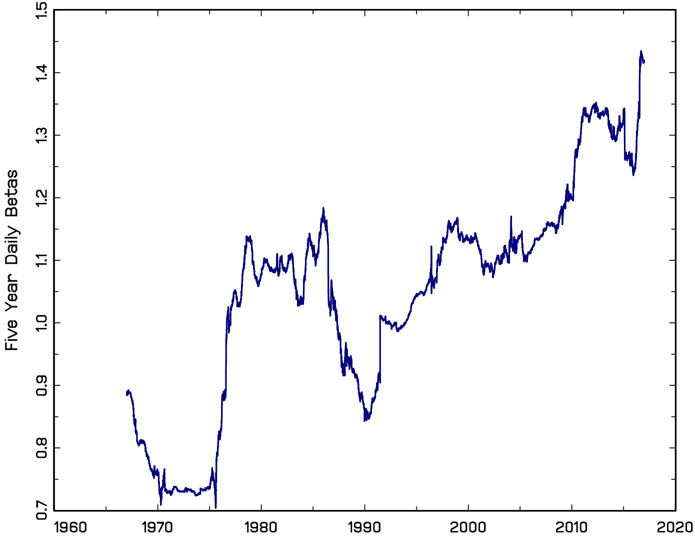


Figure: Time Varying Betas

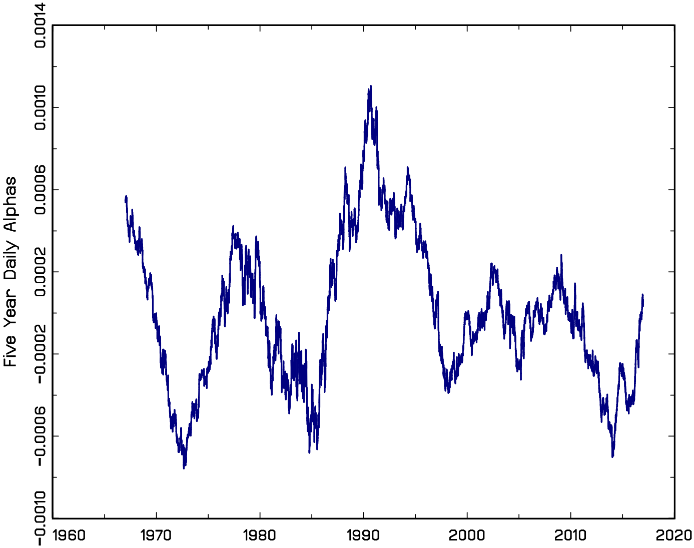


Figure: Time Varying Alphas

Criticisms of Mean-Variance analysis and the CAPM

- Roll critique. Cant observe the market portfolio. So rejections of CAPM are not valid
- Normality (or an Elliptic distribution) is crucial to the derivation of the CAPM. The Normal distribution is statistically strongly rejected in the data.
- Furthermore, the CAPM has only negligible explanatory power.
- Ex ante versus ex post betas. Conditional CAPM. Time varying risk premia. For example recession indicators. Will discuss later.