Econometría Financiera

Karoll GOMEZ
kgomezp@unal.edu.co
https://sites.google.com/unal.edu.co/karollgomezportilla

Primer semestre, 2021

UNIVERSIDAD NACIONAL DE COLOMBIA

Financial Econometrics

This part of the course attempts to provide an introductory understanding of topical issues in asset pricing and deliver econometric methods.

List of Topics:

1. The Predictability of Asset Returns
2. The Capital Asset Pricing Model
3. Multifactor Pricing Models

Additional References:

Slides are based on Oliver Linton, Empirical Finance Course at Cambridge University. January 2014.
Cochrane, J (2005). Asset pricing. Princeton University Press. 2nd Edition

1. The Predictability of Asset Returns

Predictability

A. Prices and Return

Equity capital gain over holding period $s \in \mathbb{R}$

$$R_{t,s} = \frac{P_{t+s}}{P_t} - 1$$

$$r_{t,s} \equiv \log\left(1 + R_{t,s}\right) = \log\frac{P_{t+s}}{P_t} = p_{t+s} - p_t$$

Usually take $s = 1$ and let $R_t = R_{t-1,1}$ and $r_t = r_{t-1,1}$.
Calendar time - returns are generated in calendar time, observe $P_1, P_2, P_3, P_4, P_5, P_8, P_9, \ldots$ and so Friday to Monday is a three day return
Trading time - returns are only generated when exchange is open so Friday to Monday is a one day return
For daily frequency, usually take closing price to closing price.

Predictability

Prices and returns

Index values

$$I_t = \sum_{j=1}^{J} w_{jt} P_{jt}$$

Equal weighted, price weighted (Dow Jones), value weighted (S&P500)
Dividends should be added to capital gain to make total return. For indexes, this is usually done through reinvestment.
Taxes, inflation, and exchange rates may also be relevant to investors when calculatng their return.

Notes:

- Equal weigthed: a type of weighting that gives the same weight, or importance, to each stock in a portfolio or index fund.
- Price weigthed: a type of weighting in which each asset influences the index in proportion to its price per unit.
- Value weigthed: individual components are weighted according to their market capitalization. It is defined as the total market value of all outstanding shares. To calculate a company's market cap, multiply the number of outstanding shares by the current market value of one share.

The advantages of continuously compounded returns:

Nice feature of log returns is that they can take any value, whereas actual returns are limited from below by limited liability, i.e., you cant lose more than your stake means that $R_t \geq -1$, whereas $r_t \in \mathbb{R}$. Therefore, r_t is logically consistent with a normal distribution, whereas R_t is not.
Note that if you gain 10% and lose 10%, then you are down 1% but log returns says you are evens.

$$(1+0.1)*(1-0.1) = 0.99$$

When horizon is long or per period returns are high, log returns and actual returns can be quite different.

Multiperiod returns:

For logarithmic returns

$$\begin{aligned} r_{t,H} &= \log P_{t+H} - \log P_t \\ &= \log P_{t+H} - \log P_{t+H-1} + \ldots + \log P_{t+1} - \log P_t \\ &= r_{t+H} + r_{t+H-1} + \ldots + r_{t+1} \end{aligned}$$

Weekly returns are the sum of the five daily returns
Not true for actual returns

The continuously compounded multiperiod return is simply the sum of continuously compounded single-period returns.

Multiperiod gross returns

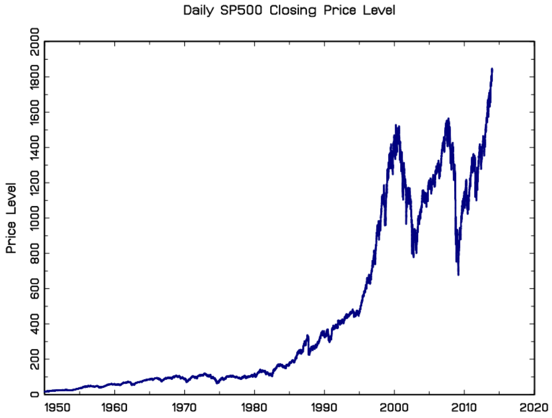
$$\begin{aligned} 1+R_{t,H} &= \mathcal{R}_{t,H} = \frac{P_{t+H}}{P_t} = \frac{P_{t+H}}{P_{t+H-1}} \times \frac{P_{t+H-1}}{P_{t+H-2}} \times \cdots \times \frac{P_{t+1}}{P_t} \\ &= \mathcal{R}_{t+H-1,1} \times \cdots \times \mathcal{R}_{t,1} \\ &= [1+R_{t+H-1,1}] \times \cdots \times [1+R_{t,1}] \end{aligned}$$

Can take geometric mean to give a per period figure $(1+R_{t,H})^{1/H}$. Note that

$$\log\left[(1+R_{t,H})^{1/H}\right] = \frac{1}{H}\sum_{s=1}^{H} \log[1+R_{t+s-1,1}]$$

For modeling of the statistical behavior of asset returns over time, it is far easier to derive the time-series properties of additive processes than of multiplicative processes.

Predictability
Prices and returns



S&P500 index was 17.03 on 10/1/1950 and was 1842.37 on 10/01/2014.
Gross return is 108.183
Annual return of 7.6 %.

Disadvantage:

Returns have property of additivity which is very convenient. However, Either time additivity:

$$r_t\left(k\right) = r_t + r_{t+1} + ... + r_{t-k+1}$$

Or portfolio additivity:

$$R_t\left(w\right) = w_1 \; R_{1t} + w_2 \; R_{2t} + \ldots + w_N \; R_{Nt}$$

but not both!

The simple return on a portfolio of assets is a weighted average of the simple returns on the assets themselves, where the weight on each asset is the share of the portfolio's value invested in that asset.

Prediction of prices simple with returns

$$E[P_{t,H}|\mathcal{F}_t] = P_t E[\mathcal{R}_{t+H}|\mathcal{F}_t] = P_t E[\mathcal{R}_{t+H}|\mathcal{F}_t]$$

But not for logarithmic returns (another approximation)

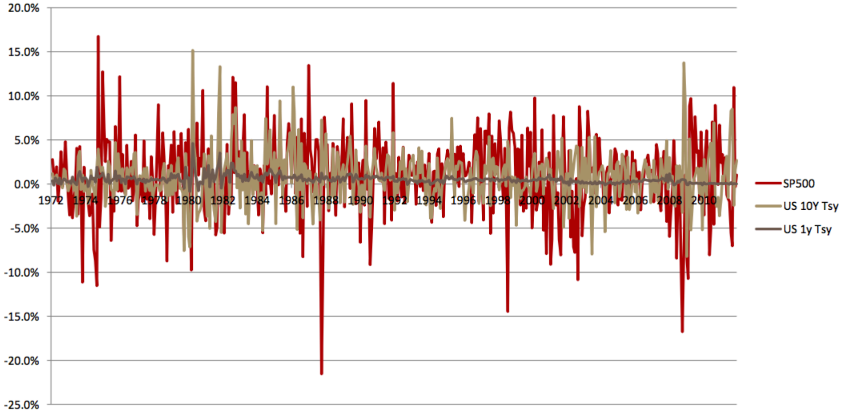
$$\begin{aligned} E[P_{t,H}|\mathcal{F}_t] &= P_t E\left[\exp\left(\log P_{t+H} - \log P_t\right)|\mathcal{F}_t\right] \\ &\geq P_t \exp\left(E\left[\log P_{t+H} - \log P_t|\mathcal{F}_t\right]\right) \end{aligned}$$

under Jensen's inequality (with strict inequality in general).
Approximately ok for small returns and short horizons

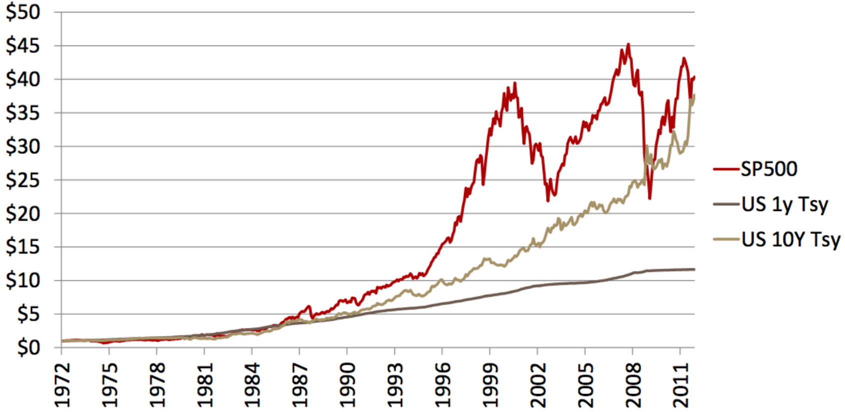
In practice:

- it is common to use simple returns when a cross-section of assets is being studied
- continuously compounded returns when the temporal behavior of returns is the focus of interest

In the time series:



In the cross-section:



Excess return:

- It is often convenient to work with an asset's excess return, defined as the difference between the asset's return and the return on some reference asset.
- The reference asset, R_{0t}, is often assumed to be riskless and in practice is usually a short-term Treasury bill return.

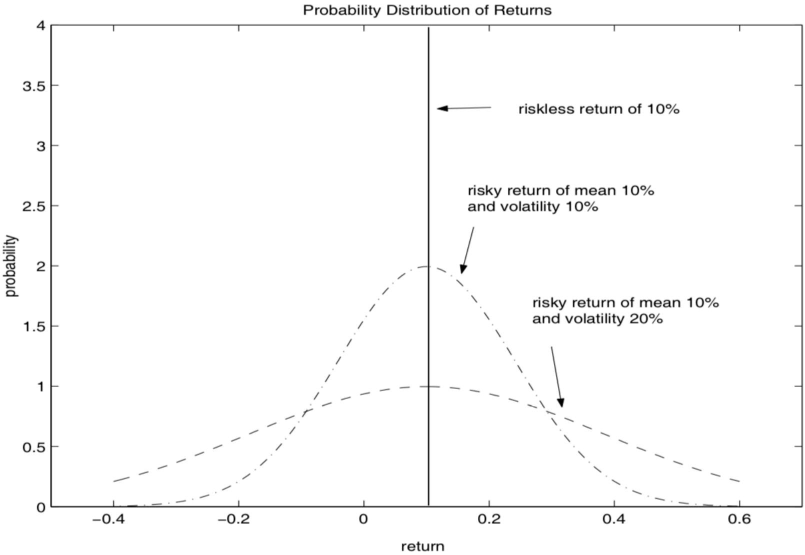
Working with simple returns:

$$RX_{it} = R_{it} - R_{0t}$$

Working with compounded returns:

$$rx_{it} = r_{it} - r_{0t}$$

Predictability



Predictability

B. The Random Walk Hypothesis

$$\epsilon_t = X_t - X_{t-1} - \mu,$$

where $X_t = P_t$ or $X_t = p_t$. Historically, μ was often assumed to be zero.
RW1:

$$\epsilon_t \sim IID; \; E\epsilon_t = 0$$

RW2:

$$\epsilon_t \; independent \; over \; time; E\epsilon_t = 0$$

RW3: For all k

$$\text{cov}\left(\epsilon_t, \epsilon_{t-k}\right) = 0$$

Then, $RW1 \implies RW2 \implies RW3$

A martingale is a time-series process X_t obeying

$$E\left[X_{t+1} \mid X_t,\ X_{t-1},\ \ldots\right] = X_t$$

or equivalently, call $\epsilon_{t+1} = X_{t+1} - X_t$ a martingale difference sequence

$$E\left[\epsilon_t \mid X_{t-1},\ X_{t-2},\ \ldots\right] = 0.$$

This corresponds with the notion of a fair game: If you toss a coin against opponent and bet successively at fair odds with initial capital X_0, current capital X_t is martingale.
This is the case that $\mu = 0$; More generally, we might assume that

$$\epsilon_{t+1} = X_{t+1} - X_t - \mu$$

is a martingale difference (MDS).

Martingale property implies that

$$\mathrm{cov}(\epsilon_t, g(X_{t-1},\ X_{t-2},\ \ldots)) = 0$$

for any (measurable) function g, in particular

$$g(X_{t-1},\ X_{t-2},\ \ldots) = X_{t-k} - X_{t-k-1} - \mu = \epsilon_{t-k},$$

so stronger condition than RW3. Call it RW2.5.

Predictability

C. Efficient Markets Hypothesis (EMH)

Samuelson (1965): informationally efficient market

Price changes must be unforecastable if they are properly anticipated, i.e., if they fully incorporate the expectations and information of all market participants.
[and if there is perfect competition in the market under conditions where all participants have free access to the information essen- tial for trading.]

<u>The expected price change based on available price information is equal to zero or the market average</u>

In this case, each spot market price will completely reflect all available information on fundamental factors affecting it: prices equal fundamental values (LeRoy, 1989).

Predictability

Fama (1970):

- He interprets the EMH as an empirically-based, falsifiable theory that could explain the actual behaviour of stock market prices.
- Fama (1965) saw efficiency as an actual outcome produced by sophisticated traders.
- Samuelson (1965) defined efficiency as a state which is reached in conditions of perfect competition, zero transaction costs and complete and freely available information.

In an efficient market, where *prices always fully reflect available information* asset prices should be martingales, after adjusting for risk. Thus, any predictable component is due to changes in the risk premium

Predictability

Fama (1970, JoF): A market in which prices always 'fully reflect' available information is called 'efficient' $\Rightarrow$ EMH

Test 1: revealing information to market participants and measuring the reaction of security prices

- If prices are predictable $\Rightarrow$ opportunities for superior returns (free lunch) $\Rightarrow$ will be competed away immediately by a lot of hungry traders $\Rightarrow$ unpredictable random walk
  - If a security believed to be underpriced, buying pressure $\Rightarrow$ jump up to a level where no longer thought a bargain
  - If a security believed to be overpriced, (short-)selling pressure $\Rightarrow$ jump down to a level where no longer thought too expensive
- As a result, market forces respond to news quckly and make prices the best available estimates of fundamental values, i.e. values justified by likely future cash flows and preferences of investors/consumers

Test 2: ask whether hypothetical trading based on an explicitly specified information set would earn superior returns

We distinguish among three forms of market efficiency depending on the information set with respect to which efficiency is defined

- Weak form - historical prices fully reflected in current price: Can't profit from strategies based on past prices
- Semistrong form - all public information (past prices, annual reports, quality of management, earnings forecasts, macroeconomic news, etc.) fully reflected in current prices: Can't outsmart others since they read newspapers too
- Strong form - all private and public information fully reflected in current prices: Can't profit from knowledge of tomorrow's news (bonds downgraded by rating agencies, new drug failed to be approved, book-cooking by Enron, last-quarter earnings, M&A deals, etc.)

Technical analysis:

Technical analysts (chartists) try to identify regularity of some patterns in stock prices, hoping to exploit them and profit. e.g. Head and Shoulders

Believe patterns repeated in prices. Weak form EMH $\Rightarrow$ technical analysis hopeless. If everybody used profitable trading rules, wouldn't that invalidate them?

Fundamental analysis:

Fundamental analysts estimate future cash flows from securities and their riskiness, based on analysis of company-relevant data as well as the economic environment in which it operates, to determine the proper price of securities.
Analyze past earnings, balance sheets, quality of management, competitive standing within the industry, outlook of the industry as well as the entire economy. Anything relevant to the process of generating future profits.

Semistrong form $\Rightarrow$ superior returns through fundamental analysis not systematically possible

Test 1 and 2: in addition,we have to specify a model for returns, typically equilibrium models with time-varying normal security returns

- Abnormal security returns are computed as the difference between the return on a security and its normal return,
- Forecasts of the abnormal returns are constructed using the chosen information set.
- If the abnormal security return is unforecastable, and in this sense random", then the hypothesis of market efficiency is not rejected.

Critics:

- Grossman and Stiglitz (1980, AER) point out that if information collection and analysis are costly, there must be compensation for such activity in terms of extra risk-adjusted returns, otherwise rational investors would not incur such expenses. $\Rightarrow$ Markets cannot be fully informationally efficient, rather an 'equilibrium degree of disequilibrium'. Weak form may hold but semistrong harder to justify.
- Joint Hypothesis Problem. Any test of EMH must assume an equilibrium asset pricing model that defines 'normal' security returns against which investor returns are measured. If we reject the hypothesis that investors can't achieve superior risk-adjusted returns, we don't know if markets are inefficient or if the underlying model is misspecified. ($\Rightarrow$ Can never reject EMH).

- Shleifer and Vishny (1997, JF). Even arbitrage opportunities (no risk) may not be eliminated trivially. Textbook arbitrage is a costless, riskless and profitable trading opportunity; in practice usually costly and risky. Also is conducted by a small number of highly specialized professionals using other people's capital. Agency relationship $\Rightarrow$ short horizon. If misspricing temporarily worsens (the best opportunity to make long-term profit!), investors/clients may judge the manager as incompetent and not only refuse to provide additional capital (margin call) but make withdrawals, thus forcing him to liquidate positions at the worst time/terms (no performance fees, perhaps a career ender)
- Therefore, a rational specialized arbitrageur stays away. Thus, there may be no easy way to make excess profit and at the same time prices may deviate from fundamentals (no free lunch $\not\Rightarrow$ fair prices)

Thus:

- There must be a **premium**, ie higher expected return for bearing risk.
- In other words an asset's risk premium is a form of compensation for investors who tolerate the extra risk, usually compared to that of a risk-free asset, in a given investment.
- This is related with the risk-return relationship at the heart of finance.

Formally:

$$E(r_t|\mathcal{F}_{t-1}) = h(\mathrm{var}(r_t|\mathcal{F}_{t-1}))$$

for some increasing function h, where $\mathcal{F}_t$ is the investors information set at time t. If $\mathrm{var}(r_t|\mathcal{F}_{t-1})$ is time varying then so is $E(r_t|\mathcal{F}_{t-1})$.
In conclusion, we should allow

$$r_t = \mu_t + \varepsilon_t,$$

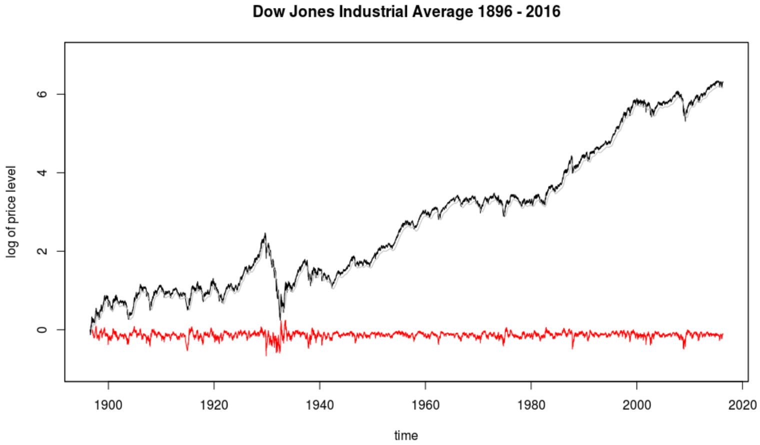
where ε_t is a martingale difference sequence with respect to some information set $\mathcal{F}_{t-1}$, i.e.,

$$E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$$

and μ_t is the risk premium that is given from some asset pricing model.
Joint hypothesis problem.

Predictability

Example:



Predictability

Example:

- If the market was truly random this line would not consistently increase like it does.
- The market goes up because investors deserve to be compensated for the risk they took when they invested in the stock market over some other investment e.g. cash.
- This return is what is called the equity risk premium and it has been approximated by a compounded 126-day rolling average return in the graph (the grey line).
- The red line represents the compounded excess / residual return of the market over our approximation of the equity risk premium.
- Assuming our approximation of the market risk premium is correct - which it isn't - the grey line represents the market and it is what you should expect to have made. It is the signal.
- The red line should just be noise or a Martingale process

Predictability

D. Statistical Tests for the EMH

In summary, EMH implies:

- competitive market for information
- rational expectations, and agreement on the implications of current information
- or no trader can consistently make better evaluations of available information
- transactors take into account all available information

$\Longrightarrow$ markets are informationaly efficient, i.e. prices contain most information about fundamental value

- weak form: $\mathcal{I}_t$ = all past public information at time t
- semi-strong form: $\mathcal{I}_t$ = all public information at time t
- strong form: $\mathcal{I}_t$ = all information at time t

Predictability

However:

- **High Frequency Traders and Arbitrageurs** argue that security prices are not random at high frequencies either because they exhibit patterns or because the law of one price is violated across geographic regions.
- **Fundamental Analysis** argues that security prices are not random at low frequencies with respect to the set of information, which contains fundamental information about the company which underlies the security.
- **Macroeconomic investors** argue that security prices are not random at low frequencies with respect to the set of information, which contains macroeconomic indicators such as business and credit cycles.
- **Technical Analysis** argues that security prices and volume data are not random at any frequency because they exhibit economically significant patterns which are identifiable and exploitable using deterministic technical indicators.

Test:

Under the weakest version of random walk hypothesis (RW3), which implies just uncorrelated increments, there are two ways to test the EHM hypothesis:

- Autocovariance functions
- AutoRegression test
- Independent Runs test

Under the strongest version of random walk hypothesis (RW1), which implies IID increments, we use:

- Variance ratios

Under the intermediate version of random walk hypothesis (RW2), we use:

- Predictive regression test

D.1 Testing RW3:

D.1.1. Autocovariance functions

One of the most direct and intuitive tests of the EMH (which is based on the random walk and martingale hypotheses) for an individual time series is to check for serial correlation, correlation between two observations of the same series at different dates.

The efficient markets hypothesis says that $\gamma_s, \rho_s = 0$ for all $s \neq 0$.

The population autocovariance and autocorrelation functions of a stationary series Y_t

$$\gamma_s = \operatorname{cov}(Y_t, Y_{t-s}) = E\left[(Y_t - EY_t)\left(Y_{t-s} - EY_{t-s}\right)\right]$$

$$\rho_s = \frac{\gamma_s}{\gamma_0}$$

for $s = 0, \pm 1, \pm 2, \ldots$ Take $Y_t = r_t$ or R_t.
Can estimate these quantities by the sample equivalents

$$\widehat{\gamma}_s = \frac{1}{T-s} \sum_{t=s+1}^{T} (Y_t - \overline{Y})(Y_{t-s} - \overline{Y})$$

$$\widehat{\rho}_s = \frac{\widehat{\gamma}_s}{\widehat{\gamma}_0}.$$

- Assume further that Y_t is i.i.d. It can be shown that for any k,

$$\sqrt{T}\widehat{\rho}_k \Longrightarrow N(0,1)$$

under the null hypothesis of no correlation.

- Therefore, you can test the null hypothesis by comparing $\widehat{\rho}_k$ with the so-called 'Bartlett intervals'

$$[-z_{\alpha/2}/\sqrt{T}, z_{\alpha/2}/\sqrt{T}],$$

where z_α are normal critical values. Values of $\widehat{\rho}_k$ lying outside this interval are inconsistent with the null hypothesis. Literally, this is testing the hypothesis that $\rho_k = 0$ versus $\rho_k \neq 0$ for a given k.

- Under the alternative hypothesis

$$\sqrt{T}\widehat{\rho}_k \xrightarrow{P} \infty$$

for at least one k.

- In fact, under this assumption we have

$$\sqrt{T}(\widehat{\rho}_1, \ldots, \widehat{\rho}_P)^{\top} \Longrightarrow N(0, I_k).$$

  The Box–Pierce Q statistic

$$Q = T\sum_{j=1}^{P}\widehat{\rho}_j^2$$

  can be used to test the joint hypothesis that $\rho_1 = 0, \ldots, \rho_P = 0$ versus the general alternative. Since $Q \Longrightarrow \chi_P^2$ under the null hypothesis, you reject when $Q > \chi_P^2(\alpha)$ for an α-level test.

- Box-Ljung version is known to have better finite sample performance

$$Q = T(T+2)\sum_{j=1}^{p}\frac{\widehat{\rho}_j^2}{T-j}$$

D.1.2. AutoRegression test

Fit the autoregression

$$Y_t = \mu + \beta_1 Y_{t-1} + \ldots + \beta_P Y_{t-P} + \varepsilon_t$$

Test the hypothesis (Standard F-test)

$$H_0 : \beta_1 = \cdots = \beta_P = 0$$

versus general alternative. When $P = 1$ this is equivalent to ACF test, but not for $P > 1$. Specifically, (setting $\mu = 0$)

$$\begin{bmatrix} \widehat{\beta}_1 \\ \vdots \\ \vdots \\ \widehat{\beta}_P \end{bmatrix} = \begin{bmatrix} 1 & \widehat{\rho}_1 & & \widehat{\rho}_{P-1} \\ & \ddots & \ddots & \\ & & \ddots & \widehat{\rho}_1 \\ & & & 1 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\rho}_1 \\ \vdots \\ \vdots \\ \widehat{\rho}_P \end{bmatrix}$$

Nonlinear functional of $\rho_1, \ldots, \rho_P$

Advantages and disadvantages of regression tests

- Advantages
  - Designed more for the conditional moment hypothesis (Martingale hypothesis)
- Disadavatages
  - If P is large, covariate matrix in OLS can be rank deficient, certainly when $P \to \infty$.
  - Not graphical or directional

D.1.3. The independent run test

- It is a non-parametric test (meaning that it does not assume much about the underlying distribution of the data) which works on binarized returns.
- Binarized returns are returns which have been converted to binary i.e. 1 or 0 depending on whether they were positive returns (+) or negative returns (-).
- A run is any consecutive sequence of either 1 (+) or 0 (-).
- Wald and Wolfowitz (1940) prove than when the number of bits in a sequence gets large, N, the conditional distribution given the number of ones, N1, and the number of zeros, N0, is approximately normal with

$$\mu = \frac{2N_+ N_-}{N} - 1$$

$$\sigma^2 = \frac{N_+ N_- (2N_+ N_- - N)}{N^2 (N-1)}$$

Then, $H0 : IID_{returns}$ vs $H1 : NoIID_{returns}$

$$z = \frac{N_{Runs} - \mu}{\sigma^2} \sim N(0,1)$$

- Note that this randomness test is conditional on the number of 1's and the number of 0's. Therefore drift, the general tendency of markets to go up over time rather than down, does NOT impact the results.
- The number of 1's in the sequence could be 90 % and the above statement would still hold true.
- Furthermore, because the runs test only deals with the sign of the return and not its magnitude, it is not affected by stochastic volatility.
- If patterns exist in the magnitude or size of returns in either direction over time, such as would be the case in a mean-reverting or momentum-driven market, the runs test will not be able to identify these.

D.2. Testing RW1:

D.2.1. Variance ratio test

An important property of all three random walk hypotheses is that the variance of random walk increments must be a linear function of the time intervals.

For example, under RW1, for log prices where continuously compounded returns $r_t \equiv \log P_t - \log P_{t-1}$ are IID, the variance of $r_t - r_{t-1}$ must be twice the variance of r_t.

Therefore, it is possible construct a statistical test of the random walk hypothesis using variance ratios, by comparing the variance of $r_t - r_{t-1}$ to twice the variance of r_t.

- Suppose that returns are stationary and in particular $Er_t = \mu$ and $var\, r_t = \sigma^2$.
- Look at the 2-period return

$$r_t(2) = p_{t+2} - p_t = p_{t+2} - p_{t-1} + p_{t+1} - p_t = r_{t+2} + r_{t+1}.$$

- We have

$$\text{var}(r_t(2)) = \text{var}(r_{t+2}) + \text{var}(r_{t+1}) + 2\text{cov}\left(r_{t+2}, r_{t+1}\right).$$

- Under the assumption that returns are uncorrelated (RW3) we further have $\mathrm{cov}(r_{t+2}, r_{t+1}) = 0$ and so

$$\mathrm{var}(r_t(2)) = \mathrm{var}(r_{t+2}) + \mathrm{var}(r_{t+1}) = 2\mathrm{var}(r_t).$$

  Therefore

$$VR(2) = \frac{\mathrm{var}(r_t(2))}{2\mathrm{var}(r_t)} = 1.$$

- In fact

$$VR(q) = \frac{\mathrm{var}(r_t(q))}{q\mathrm{var}(r_t)} = 1$$

  for all horizons q. What we have rediscovered here is that if the price process p_t is a unit root process, then the return variance grows linearly with the time horizon.

- If the series is actually positively (negatively) serially correlated, then

$$\begin{aligned} \mathrm{var}(r_t(2)) &> \; < \mathrm{var}(r_{t+2}) + \mathrm{var}(r_{t+1}) \\ \frac{\mathrm{var}(r_t(2))}{2\mathrm{var}(r_t)} &> \; < 1 \end{aligned}$$

$$VR(2) = \frac{\mathrm{var}(r_t) + \mathrm{var}(r_t) + 2\mathrm{cov}\left(r_{t+2}, r_{t+1}\right)}{2\mathrm{var}(r_t)} = 1 + \rho(1)$$

- Can show that for a general stationary process with ACF $\{\rho(j),\ j = 1, 2, \ldots\}$

$$VR(q) = 1 + 2\sum_{j=1}^{q-1}\left(1 - \frac{j}{q}\right)\rho(j)$$

so that the direction of the ratio depends on all the first k autocorrelations and their relative magnitudes. It is just a linear functional of the correlogram function.

D.2.2 Long Horizon Returns

- Several recent studies have focused on the properties of long-horizon returns to test the random walk hypotheses, in some cases using 5- to 10-year returns over a 65-year sample.
- For some alternatives to the random walk, long-horizon returns can be more informative than their shorter-horizon counterparts
- One motivation for using long-horizon returns is the permanent/transitory components of log prices (first proposed by Muth (1960) in a macroeconomic context)

Market inefficiencies (Fads) Model

Fads models were introduced by Shiller (1984) and Summers (1986) as plausible alternatives to the efficient markets (or constant expected) returns assumptions.

- Le Roy and Porter (1981) and Shiller (1981) have argued that the volatility observed in stock and bond markets is too high to be explained by the flow of information on fundamentals, such as dividends.
- To explain this excess of volatility, Shiller (1981) emphasized the role of investor overreactions, fashions and fads in price shocks.
- Under fads models, logarithms of asset prices embody both a martingale component, with permanent shocks, and a stationary component, with temporary shocks.

Suppose log prices have a permanent/transitory decomposition:

$$p_t = w_t + y_t$$

$$w_t = \mu + w_{t-1} + \varepsilon_t, \varepsilon_t \sim IID(0, \sigma^2)$$

where:
y_t is a stationary process and represents the fundamental value of the asset, which has permanent shocks.
w_t is a random walk plus drift and represents the current mispricing of the assets, or "fad", which has temporary shocks.

It follows that:

$$r_t = \overbrace{\mu + \varepsilon_t}^{\text{iid fundamental return}} + \overbrace{y_t - y_{t-1}}^{\text{mean zero stationary fad}}$$

Do fads exists?
To answer this question, we use the Variance Ratio.

The covariance function of the sum of independent stochastic processes is the sum of the covariance functions.

Consider the variance ratio for horizon q of the return series r_t.

$$r_t(q) = q\mu + \sum_{k=1}^{q} \varepsilon_{t-k} + y_t - y_{t-q}$$

$$\mathrm{var}(r_t(q)) = q\mathrm{var}(\varepsilon_t) + \mathrm{var}(y_t - y_{t-q})$$

Since the fad component is covariance stationary, for a large enough q the variance ratio of observed returns r is less than one. In fact, as $q \to \infty$

$$VR(q) = \frac{q\mathrm{var}(\varepsilon_t) + \mathrm{var}(y_t - y_{t-q})}{q\mathrm{var}(\varepsilon_t) + q\mathrm{var}(y_t - y_{t-1})} \to 1 - \frac{\mathrm{var}(\Delta y)}{\mathrm{var}(\Delta p)}$$

This says that if the fads model is true we should find VR less than one for long lags.

- The magnitude of the difference between the long-horizon variance ratio and one, is a kind of signal/signal+noise ratio.
- the *signal* is the transitory (short-term) component and the *noise* of the permanent (long-term) markets component.
- The fads model only has clear implications for long horizons.
- The VAR test has zero power to detect fads if fads are long-lived relative to the sample length

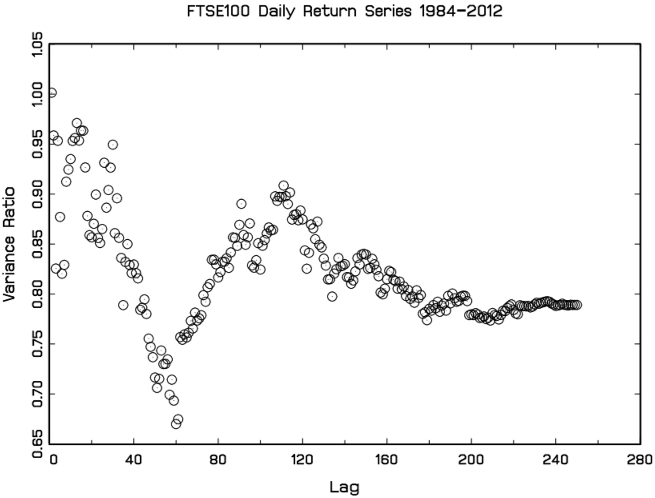
Long Horizon Variance Ratio

We can write

$$
\begin{aligned}
VR(\infty) &= 1 + 2\lim_{q\to\infty}\sum_{j=1}^{q-1}\left(1-\frac{j}{q}\right)\rho(j) \\
&= 1 + 2\sum_{j=1}^{\infty}\rho(j) - 2\lim_{q\to\infty}\frac{1}{q}\sum_{j=1}^{q} j\rho(j) \\
&= 1 + 2\sum_{j=1}^{\infty}\rho(j) = \frac{\mathrm{lrvar}(r)}{\mathrm{var}(r)}
\end{aligned}
$$

provided $\sum_{j=1}^{\infty} j|\rho(j)| < \infty$, which will be the case for many stationary and mixing processes.

The lrvar is equal to the spectral density at frequency zero and under some conditions this is consistently estimable and asymptotic normal distribution can be obtained (at rate $n^{1/2}$ not $(nq)^{1/2}$).



This figure shows VR(lag) for lag $= 2,\ldots,250$

Variance ratio tests are widely used by finance practitioners

- High frequency predictability, e.g. from market illiquidity (bid-ask bounce), or sluggish reaction to information, or disposition-effect trading by individual investors.
  - Comparatively easy to detect if present.
  - Hard to explain using a risk-based model.
  - Has small effects on prices.
  - Can disappear quickly once detected by arbitrageurs.
- Low frequency predictability, e.g. from gradually changing risk or risk aversion, or slow changes in sentiment of irrational investors.
  - Long time series are needed to detect this.
  - There may be several plausible explanations.
  - Potentially large effects on prices.
  - Hard to arbitrage away.

Predictability

D.3 Testing RW2: Regression test based

If we assume that asset prices are unforcastable given the past prices, this does not preclude then bieng forecastable given additional information.

- Consider the (predictive) regression

$$R_{t+j} = \mu + \beta^{\intercal} X_t + \varepsilon_{t+j},$$

  where X_t is observed (public information) at time t or deterministic like seasonal dummy variables. The EMH (along with constant mean or risk premium) says that $\beta = 0$. Standard regression F test for the inclusion of X_t.
- price/earnings ratio effects, dividend rate, and so on. Lots of evidence on this. Shiller website. Some econometric issues when X is very persistent process.

Final Remarks

- Tests of unit roots are not designed to detect predictability.
- Perfect efficiency is an unrealistic benchmark that is unlikely to hold in practice.
- Market efficiency is an idealization that is economically unrealizable, but that serves as a useful benchmark for measuring relative efficiency.
- The notion of relative efficiency, which is the efficiency of one market measured against another, e.g. futures markets vs. spot markets, may be a more useful concept than the traditional market-efficiency literature.

- For these reasons, modern finance theory does not take stand on market efficiency itself, but focus instead on the statistical methods that can be used to test the joint hypothesis of **market efficiency** and **market equilibrium** using models like CAPM.
- Consistent with the market efficiency hypothesis that the anomalies are chance results, apparent overreaction to information is about as common as underreaction, and post-event continuation of pre-event abnormal returns is about as frequent as post-event reversal.
- However, the field of modern financial economics assumes that people behave with extreme rationality, but they do not. Furthermore, people's deviations from rationality are often systematic.

- **Behavioral Finance** relaxes the traditional assumptions of financial economics by incorporating these observable, systematic, and very human departures from rationality into standard models of financial markets.
- Kahneman and Tversky have shown empirically that people are irrational in a consistent and correlated manner.
- However, the case for the EMH can be made even in situations where the trading strategies of investors are correlated. So long as there are some smart investors and arbitrage opportunities, they will exploit any mispricing and the irrational investors will lose money and eventually disappear from the market.

Empirical evidence in favor of predictability:

- Some strategies, portfolios, assets have returns that are not explained by their CAPM betas returns are predictable: p/d ratios, term premium predict stock returns.
- Bond returns are predictable: upward sloping yield curve implies higher expected returns for long-term bond than for short-term bonds
- Foreign exchange returns are predictable: bonds in higher interest rate countries have higher returns
- Returns are not IID: volatility changes through time, not in lockstep with the means i.e. Sharpe ratios vary over time
- Some funds outperform indices and are predictable (but Carhart (1997) argues this comes from following persistent strategies rather than skill).

Can the Efficient Markets hypothesis be saved:

- Explain as much as we can with efficiency + a model for the pricing of risk
- Analyze the market specificities, and relax constraints to allow for the observed inefficiencies:
    - transaction costs, bid-ask spreads
    - leveraging constraints, short selling constraints, arbitragers' risk bearing
    - differences in information, beliefs, noise traders
    - behavioral biases

2. The Capital Asset Pricing Model (CAPM)

A. The Capital Asset Pricing Model

Sharpe (1964) and Lintner (1965) version of the CAPM

The model assumes that:
(i) all investors act according to the $\mu - \sigma$ rule (MV rule),
(ii) face no short-selling constraints, and
(iii) exhibit perfect agreement with respect to the probability distribution of asset returns.
(iv) All investors can lend and borrow ta risk-free rate. (v) No taxes, No commissions.

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

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

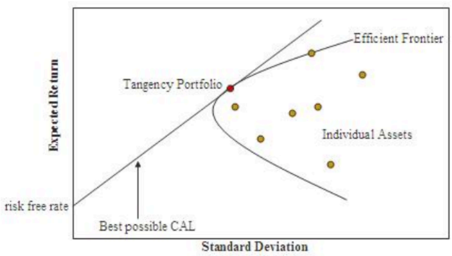
for all i. Relates three quantities

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

all of which can be estimated from time series data

where:
β_i beta risk volatility or systemic risk (how risky the assets is).
$R_i - R_f$ risk premium.
$R_m - R_f$ market price of risk.

Under these assumptions, the market portfolio is a mean-variance efficient portfolio.



The CML is a set of portfolios p such that

$$ER_p = R_f + \frac{ER_m - R_f}{\sigma_m} \cdot \sigma_p$$

All portfolios on CML have the same Sharpe ratio as the market portfolio (EHM).

Security market line (CAPM)

$$ER_i - R_f = \beta_i \cdot (ER_m - R_f)$$

slope is $ER_m - R_f$, which is also called the market price of risk, while

$$\beta_i = \frac{\mathrm{cov}(R_i, R_m)}{\mathrm{var}(R_m)}$$

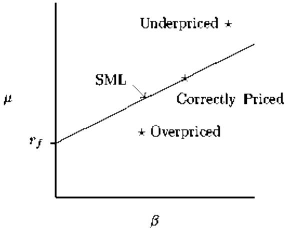
is the beta of firm i, the risk measure of firm i.

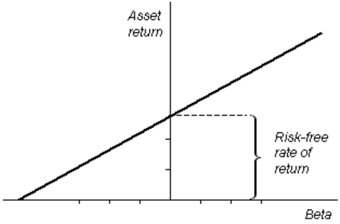
CAPM

The model

The SML defines the relation the asset premium and its market beta.

Figure 4.1: Security selection using returns.





Implications from the CAPM:

$$\text{SML} \Rightarrow \text{Empirical relationship between } ER_i \text{ and } \beta_i$$

1. Excess returns are proportional to beta risk. That implies independent term in a regression model is equal to zero.
2. Beta risk completely capture the excess returns cross-sectional variation.
3. Market risk premium (or market price of risk) is positive

In other words, the CAPM predicts that:

- All investors will hold a combination of the market portfolio of risky assets and a portfolio whose returns are uncorrelated with market returns.
- The market portfolio is mean-variance efficient implying:
  - that there exists a linear relationship between a portfolio's expected return and its market beta
  - and that no other factors are necessary to explain expected returns.
- In CAPM model systematic risk:
  (1) Cannot be diversified
  (2) Has to be hedged
  (3) In equilibrium it is compensated by a risk premium

Which means:

- time series efficiency:
    - technical analysis is pointless
    - no timing the market with available information
- cross-sectional model:
    - no way to "beat the market"
    - higher average returns are obtained only by bearing higher risk

Black's (1972) version of the CAPM

In addition to asumptiosn (i), (ii) and (iii), it is NOT assumed that they can lend and borrow at a common risk-free rate (because of inflation uncertainty, credit risk).

Black version without a riskless asset find the (zero beta) portfolio return R_0 such that $R_0 = \arg\min var(R_x)$ subject to $cov(R_x, R_m) = 0$

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

for all i.

- Zero-Beta is a portfolio which returns (R_0) is uncorrelated with market portfolio (as risk-free asset).
- Zero-Beta portfolio is always in the inefficient part of the frontier.
- To obtain zero-beta portfolios we typically would have to short sell some assets.

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 $\alpha_i = 0$ for all i

Black. We have

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

and test $\alpha_i = (1 - \beta_i)E[R_0]$ for all i. Here, R_0 is the return on the (unobserved) zero beta portfolio

CAPM

B. Market Model

Tests of the CAPM are usually done in two ways:

- Time series regression test
  $E[R_i] = \alpha_i + \beta_i E[R_m] + \varepsilon_i$ and test $H_0 : \alpha_i = 0$.
- Cross-sectional regression test (two-step estimation to test SML)
  $E[R_i] = \gamma + \beta_i \lambda + \xi_i$ and test $H_0 : \lambda > 0$.

The goal is to estimate:
market risk price, beta-risk and test the CAPM model.

CAPM

However, there exits some problems:

- CAPM model is a static model. Thus, we need to assume a jointly IID return distribution.
- Expected returns are no observable. Thus, we have to assume rational expectations or assume joint normality of returns.
- Market returns are not observable. Then we need to use a proxy for it.

So, we have to impose a set of statistical assumptions on the model (returns are stationary and serially uncorrelated).

To obtain a model with observable quantities, we need to describe returns using the following market model:

Under joint normality of returns the market model holds for returns (r_{it}) or excess returns ($Z_{it} = r_{it} - r_{ft}$ or $Z_{it} = r_{it} - r_f$):

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

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

$$E\left[\epsilon_{it}\right] = 0 \qquad \mathrm{var}\left[\epsilon_{it}\right] = \sigma^2_{\epsilon i}$$

$$\mathrm{cov}\left(\epsilon_{it}, r_{mt}\right), \mathrm{cov}\left(\epsilon_{it}, \epsilon_{js}\right) = 0$$

Time series linear regression. Can estimate parameters asset by asset OLS.

Practical issues

- Which assets to include in the sample. Individual stocks, portfolios
- What sampling frequency: daily, weekly, or monthly?
- How long a time series to consider: 5 years, 10 years etc
- What market portfolio to use? CRSP indexes
- What risk free rate to use?

Normality is not necessary for the CAPM but much of the literature assumes it. 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 large negative skewness and large positive kurtosis.
Fama for example argues that monthly returns are closer to normality

Aggregation of (logarithmic) returns. Let A be the aggregation (e.g., weekly, monthly), then under RW1

$$Er_A = AEr$$

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

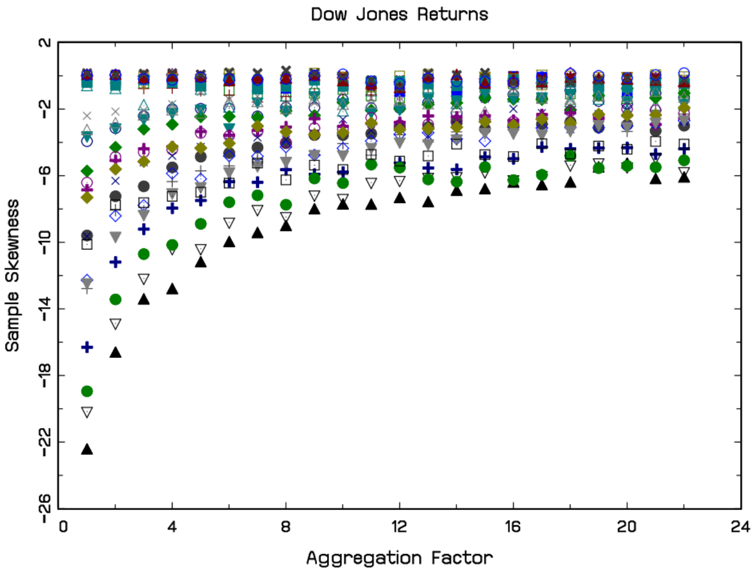
$$\kappa_3(r_A) = \frac{1}{\sqrt{A}}\kappa_3(r)$$

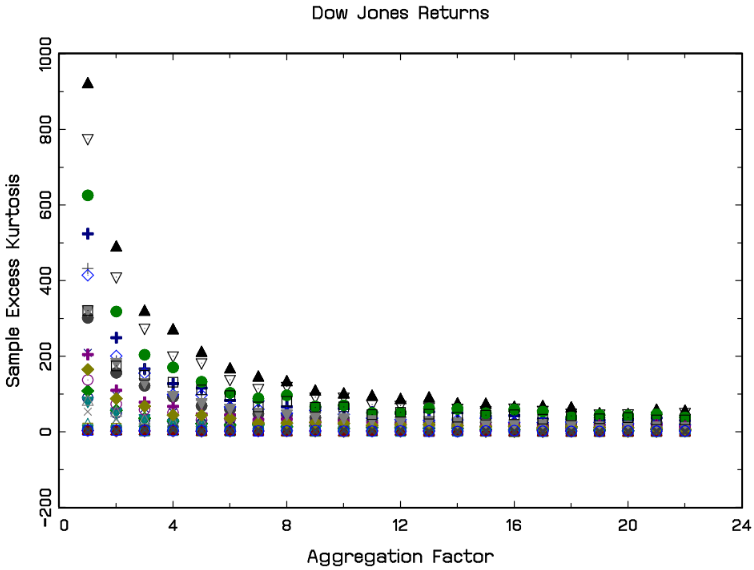
$$\kappa_4(r_A) = \frac{1}{A}\kappa_4(r)$$

Says that as you aggregate more, returns become more normal
Is this true empirically??

CAPM

Market model

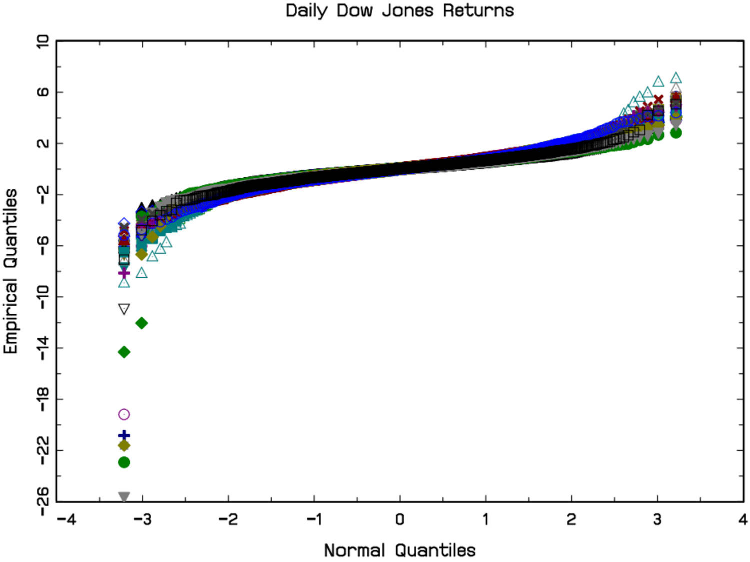




Dow Jones Returns
Sample Excess Kurtosis
Aggregation Factor
1000
800
600
400
200
0
-200
0
4
8
12
16
20
24

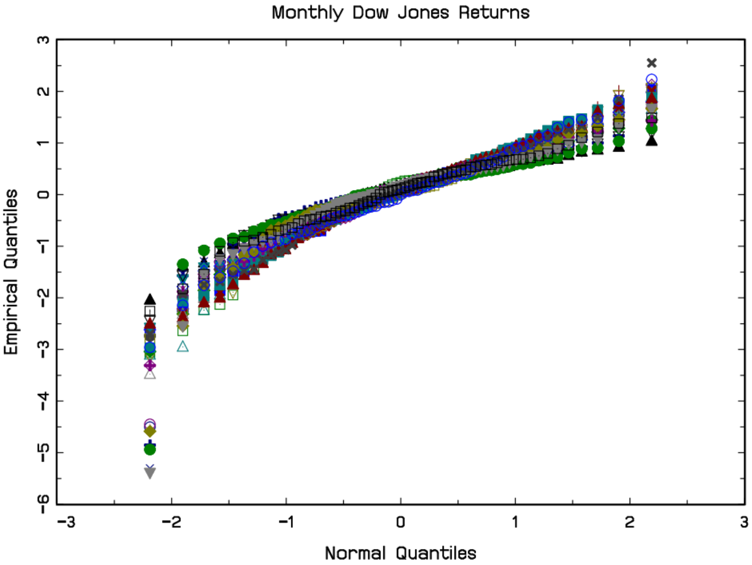
CAPM

Market model



CAPM

Market model



C. Time-series Regression Test

c.1) **OLS estimation**

$$y = X\theta + \varepsilon$$

White's standard errors (se_W) are the square root of the diagonal elements of

$$(X^\top X)^{-1} X^\top S X (X^\top X)^{-1}$$

$$S = \text{diag}(\hat{\varepsilon}_1^2, \ldots, \hat{\varepsilon}_T^2)$$

The least squares standard errors are the square root of the diagonal elements of

$$s^2 (X^\top X)^{-1}$$

$$s^2 = \frac{1}{T-2} \sum_{t=1}^{T} \hat{\varepsilon}_t^2$$

Use the "p-value": For a test statistic, the p-value is the probability of just rejecting the hypothesis. p value close to one means no evidence against the null; p-value close to zero means strong evidence against the null

Goodness of fit. Recall that

$$R^2 = 1 - \frac{\mathrm{var}(\varepsilon)}{\mathrm{var}(y)}$$

so gives the ratio of idiosyncratic variance to total variance return. Since

$$\mathrm{var}(r_i) = \beta_i^2 \mathrm{var}(r_m) + \mathrm{var}(\varepsilon_i)$$

which represents the variance decomposition of returns.

We have decomposed the total risk into two components:

- Systematic risk: related to r_m
- Idiosyncratic risk: related to ε_i

If $var(\varepsilon_i) = \sigma^2$ for all assets and we have N assets, then:

$var(r_i) = \beta_i^2 var(r_m) + (1/N)\sigma^2$

As $N \to \infty$, the variance due to unsystematic risk will disappear.
That is diversification!

So that, R^2 represents how much of the total variance return is due to idiosyncratic risk.

Results for daily returns

	α	se(α)	β	se(β)	$\mathrm{se}_W(\beta)$	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$_W$(β)	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

Results for monthly returns

	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$_W$(β)	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

	μ	σ	w_{MV}	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

c.2) ML estimation

Suppose that (where Z_{mt} are excess market returns)

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

where $\varepsilon_t \sim N(0, \Sigma)$. The Gaussian log likelihood is

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

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

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

The maximum likelihood estimate of Σ is (Provided $N < T$)

$$\widehat{\Sigma} = \frac{1}{T} \widehat{\varepsilon}\widehat{\varepsilon}^{\intercal} = \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}_{it} \widehat{\varepsilon}_{jt} \right)_{..}$$

Under the normality assumption we have, conditional on excess market returns, the exact distributions

$$\widehat{\alpha} \backsim N(\alpha, \frac{1}{T}[1 + \frac{\widehat{\mu}_m^2}{\widehat{\sigma}_m^2}]\Sigma)$$

$$\widehat{\beta} \backsim N(\beta, \frac{1}{T}\frac{1}{\widehat{\sigma}_m^2}\Sigma)$$

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

$$T\widehat{\Sigma} \backsim W(T-2, \Sigma)$$

Portfolio weights

Tangency portfolio has weights that are proportional to

$$w_{TP} \propto \Sigma^{-1}\left(ER - R_f i\right)$$

where i is the N vector of ones.
Global Minimum Variance portfolio has weights that are proportional to

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

Empirically, find many negative weights.

Annualized returns, standard deviation and portfolio weights

	μ	σ	w_{MV}	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

The Jensen's alpha

We estimate the market model:

$$r_{it} - r_f = a_i + \beta_{im}(r_{mt} - r_f) + \varepsilon_{it}$$

where:

$$a_i = (\bar{r}_i - \bar{r}_f) - \beta_{im}(\bar{r}_m - \bar{r}_f)$$

Under CAPM model

$$(\bar{r}_i - \bar{r}_f) - \beta_{im}(\bar{r}_m - \bar{r}_f) = 0$$

for $i = 1, \ldots, N$, which means that CAPM model predicts a Jensen alpha equal to zero.

Jensen's alpha represents a risk-adjusted performance measure:

- It measures the average return on a portfolio over and above that predicted by the CAPM model
- If estimated Jensen's alpha is statistically significant different from zero $\Rightarrow$ the asset return is greater than the suggested by its beta risk.

For a vector of $\hat{\alpha}_{N\times 1}$ we have the following alternative test:
(i) Wald test
(ii) Likelihood Ratio test
(iii) Lagrange multiplier test

(i) Wald Test Statistic

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

$$J_0 = \hat{\alpha}^{\intercal}[\text{var}(\hat{\alpha})^{-1}]\hat{\alpha} = T\left(1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}\right)^{-1} \hat{\alpha}^{\intercal}\Sigma^{-1}\hat{\alpha}$$

Under null hypothesis $J_0 \backsim \chi^2(N)$ given true Σ^{-1}
Can use an asymptotic approximation to replace Σ with $\hat{\Sigma}$
Under null hypothesis $J_0 \overset{a}{\backsim} \chi^2(N)$ using $\hat{\Sigma}^{-1}$ in place of Σ^{-1}

where $E[R_{mt}] = \mu_m$ and $V[R_{mt}] = \sigma_m^2$

For the empirical example:

Daily J_0 =22.943222 (p-value =0.81758677);
Monthly J_0 =33.615606 (p-value =0.29645969)

Do not reject H0.

Wald Test Statistic (d.f correction)

We can get an exact test statistic by using an F distribution and a degrees of freedom correction

$$J_1 = \frac{(T-N-1)}{N}((1+\frac{\widehat{\mu}_m^2}{\widehat{\sigma}_m^2})^{-1}) \times \widehat{\alpha}^{\intercal}\widehat{\Sigma}^{-1}\widehat{\alpha} \sim F(N, T-N-1)$$

This is superior to the Wald test (under the assumption of normality)

(ii) Likelihood Ratio Test

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

$$J_2 = -2(\log \ell_c - \log \ell_u) = T[\log\det\widehat{\Sigma}^* - \log\det\widehat{\Sigma}] \Longrightarrow \chi^2(N)$$

$$\widehat{\Sigma}^* = \frac{1}{T}\widehat{\varepsilon}^*\widehat{\varepsilon}^{*\top} = \left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}^*_{it}\widehat{\varepsilon}^*_{jt}\right)_{i,j}$$

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

Remarks:

These tests have an exact relationship which allows us to derive a (complicated) small-sample distribution for the likelihood ratio test

$$J_1 = \frac{T-N-1}{N}(\exp(\frac{J_2}{T})-1)$$

Sharpe ratio:

Another important insight is that J_1 is proportional to the difference in the Sharpe ratios of the market portfolio m and the ex-post efficient tangency portfolio TP :

$$J_1 = \frac{T-N-1}{N}\left(\frac{\frac{\widehat{\mu}_{TP}^2}{\widehat{\sigma}_{TP}^2} - \frac{\widehat{\mu}_m^2}{\widehat{\sigma}_m^2}}{1+\frac{\widehat{\mu}_m^2}{\widehat{\sigma}_m^2}}\right)$$

This has a useful graphical interpretation, and in terms of investment theory.

Difference between market portfolio and tangency portfolio (risky asset portfolio + market portfolio = maximum SR^2) $\Rightarrow$ efficiency of market portfolio

Testing Black Version of the CAPM

- Tests for the Black version are more complicated to derive
- There is a cross-equation restriction on α under the null hypothesis,
- The model to be estimated is nonlinear in the parameters

Estimate the same unconstrained model as before using total returns instead of excess returns.

Test the constraint

$$\alpha = (i - \beta)\gamma$$

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

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

This replaces $\alpha = 0$ of the Sharpe-Lintner-based tests.

The unconstrained (market) model for the N vector of firm returns is

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

The constrained model is

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

Maximum likelihood estimation of the unconstrained model is equivalent to equation-by-equation OLS

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

To avoid non linear optimization of ML:

Useful trick: assume that the expected return on the zero-beta portfolio 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, the Black model can be estimated using the same methodology as the Sharpe-Lintner model

Then, relax the assumption that the zero-beta return is known.

For $\theta = (\gamma, \beta_1, \dots, \beta_N)^\intercal$ the (constrained) likelihood function is

$$\begin{aligned} \ell(\theta, \Sigma) &= c - \frac{T}{2} \log\det\Sigma \\ & -\frac{1}{2}\sum_{t=1}^{T}(R_t - \gamma i - \beta(R_{mt} - \gamma))^\intercal \Sigma^{-1}(R_t - \gamma i - \beta(R_{mt} - \gamma)) \end{aligned}$$

maximize with respect to θ. Profile/concentration method. Define

$$\widehat{\beta}_i^*(\gamma) = \frac{\sum_{t=1}^{T}(R_{mt} - \gamma)(R_{it} - \gamma)}{\sum_{t=1}^{T}(R_{mt} - \gamma)^2}$$

$$\widehat{\Sigma}^*(\gamma) = \frac{1}{T}\widehat{\varepsilon}^*(\gamma)\widehat{\varepsilon}^{*\intercal}(\gamma) = \left(\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_{it}^*(\gamma)\widehat{\varepsilon}_{jt}^*(\gamma)\right)_{i,j}$$

Then search the profiled likelihood over the scalar parameter γ

$$
\begin{aligned}
\ell^P(\gamma) &= c - \frac{T}{2}\log\det\widehat{\Sigma}^*(\gamma) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}(R_t-\gamma i-\widehat{\beta}^*(\gamma)(R_{mt}-\gamma))^{\intercal}\widehat{\Sigma}^*(\gamma)^{-1} \\
&\quad \times(R_t-\gamma i-\widehat{\beta}^*(\gamma)(R_{mt}-\gamma))
\end{aligned}
$$

and let $\widehat{\gamma}^*$ be the maximizing value and then let $\widehat{\beta}_i^*(\widehat{\gamma}^*)$ and $\widehat{\Sigma}^*(\widehat{\gamma}^*)$ be the corresponding estimates of β_i and Σ.

Comparing the relative fit of the constrained and unconstrained models serve to test the constraints

$$J_4 = T[\log\det\widehat{\Sigma}^* - \log\det\widehat{\Sigma}] \Longrightarrow \chi^2(N-1)$$

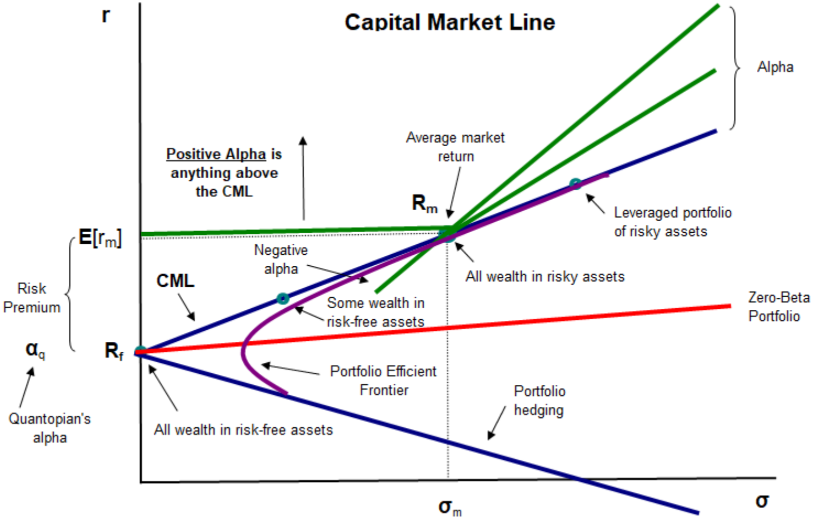
Here, $\widehat{\Sigma}^*$ is the MLE of Σ in the constrained model.
Wald, LR, and LM tests. Standard likelihood theory based on large samples.

Remarks about robustness of MLE

- Maximum likelihood estimation assumes multivariate normal returns (otherwise it is quasi-maximum likelihood).
- CAPM can hold under weaker assumptions (e.g., elliptical symmetry, which includes multivariate t-distributions with heavy tails).
- Actually, the MLE of α and β is not robust to heteroskedasticity, serial correlation, and non-normality since the estimates are just least squares.
- Need to adjust standard errors: Use, White or Newey and West standard errors.

CAPM

Market model



c.3) **GMM estimation**

When IID and normality assumption does not hold, to be able to derive test it is necessary to work under GMM setting.

GMM. Works only with conditional moment restrictions and the implied unconditional moment conditions

$$E\left[\varepsilon_t | X_t\right] = 0.$$

By the law of iterated expectations this implies for all (measurable) functions h,

$$E\left[\varepsilon_t h(X_{kt})\right] = E\left(E\left[\varepsilon_t | h(X_{kt})\right]\right) = E\left(E\left[\varepsilon_t | X_{kt}\right]\right) = 0$$

Define

$$g_T\left(\theta\right) = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t\left(\theta\right) \otimes H(X_t).$$

Then $g_T\left(\theta_0\right) \approx 0$ for large T but $g_T\left(\theta\right) \neq 0$ for $\theta \neq \theta_0$. Define the GMM estimator for a positive definite weighting matrix W_T

$$\widehat{\theta} = \arg\min g_T\left(\theta\right)^{\intercal} W_T g_T\left(\theta\right)$$

In the Sharpe Lintner case

$\theta = [\alpha, \beta]$ 2N parameters
$X_t = [1, Z_{mt}]$ 2 instruments
$\varepsilon_t = Z_t - \alpha - \beta Z_{mt}$ N variables

GMM gives $\widehat{\alpha}, \widehat{\beta} = OLS$ estimates, exactly the same as with ML.

In the Black case
$\theta = [\gamma, \beta]$ N+1 parameters
$X_t = [1, r_{mt}]$ 2 instruments
$\varepsilon_t = r_t - \gamma i - \beta(r_{mt} - \gamma)$ N variables
Robustify standard errors. Heteroskedasticity and autocorrelation.

D. Cross-Sectional Regression Tests

Two considerations:

- CAPM model implies a linear relationship between the expected returns and the beta-risk, which completely explain the cross-section of expected returns.
- CAPM is a static model.

Given this two considerations Fama and MacBeth (1973) proposed a cross-section test to validate the above implication.

They propose a two-step estimation method:

1. The first stage involves a set of regressions equal in number to the number of assets or portfolios one is testing (N regressions).
2. The second stage is a set of regressions equal in number to the number of time periods (T regressions).

First stage: use the market model to estimate $\hat{\beta}_{it}$:

$$R_{it} = \alpha_i + \beta_{im} R_{mt} + \varepsilon_{it}, \quad t = 1, \ldots, T \tag{1}$$

where ε_{it} is a iid noise. We have to run $i = 1, \ldots, N$ regressions and obtain $[\beta_{it}]_{t=1}^{T}$ for each asset.

Second stage: use the $\hat{\beta}_i$ to estimate λ_0 and λ_1:

$$\bar{R}_i = \lambda_0 + \lambda_1 \hat{\beta}_i + u_i, \quad i = 1, \ldots, N \tag{2}$$

- Example:
  - Suppose that we had a sample of 100 stocks (N=100) and their returns using five years of monthly data (T=60)
  - The first step would be to run 100 time-series regressions (one for each individual stock), the regressions being run with the 60 monthly data points
- The second stage involves a single cross-sectional regression of the average (over time) of the stock returns on a constant and the betas:
$$\bar{R}_i = \lambda_0 + \lambda_1 \beta_i + v_i$$
where $\bar{R}_i$ is the return for stock i averaged over the 60 months

Testing the CAPM:

- Essentially, the CAPM says that stocks with higher betas are more risky and therefore should command higher average returns to compensate investors for that risk
- If the CAPM is a valid model, two key predictions arise which can be tested using this second stage regression:
  1. $\lambda_0 = R_f$;
  2. $\lambda_1 = [R_m - R_f]$.

Further implications:

- Two further implications of the CAPM being valid:
  - There is a linear relationship between a stock's return and its beta
  - No other variables should help to explain the cross-sectional variation in returns
- We could run the augmented regression:

$$\bar{R}_i = \lambda_0 + \lambda_1 \beta_i + \lambda_2 \beta_i^2 + \lambda_3 \sigma_i^2 + v_i$$

  where β_i^2 is the squared beta for stock i and σ_i^2 is the variance of the residuals from the first stage regression, a measure of idiosyncratic risk
- The squared beta can capture non-linearities in the relationship between systematic risk and return
- If the CAPM is a valid and complete model, then we should see that $\lambda_2 = 0$ and $\lambda_3 = 0$.

A Different Second-Stage Regression:

- It has been found that returns are systematically higher for small capitalisation stocks and are systematically higher for 'value' stocks than the CAPM would predict.
- We can test this directly using a different augmented second stage regression:

$$\bar{R}_i = \alpha + \lambda_1 \beta_i + \lambda_2 MV_i + \lambda_3 BTM_i + v_i$$

  where MV_i is the market capitalisation for stock i and BTM_i is the ratio of its book value to its market value of equity

- Again, if the CAPM is a valid and complete model, then we should see that $\lambda_2 = 0$ and $\lambda_3 = 0$.

Problems in testing CAPM

- Heteroscedasticity
- Selection bias
- Measurement errors since beta risk is not directly observable, so β is the empirically estimated and used in a second set of regressions.
- Parameter uncertainty issue: Standard errors in the second-step are wrong, and estimated parameters could be biased and inconsistent.

Possible Solution:

- For heteroscedasticity some recent research has used GMM or another robust technique to deal with this.
- For parameter uncertainty: Testing on **portfolios** rather than individual stocks can mitigate the errors-in-variable problem as estimation errors cancel out each other.
- First step based on performing pre-ranking + estimation in different periods avoids selection bias.
- Second step based on double sorting: by beta (reduces the shrinkage in beta dispersion) + by size (takes into account correlation between size and beta).

The Jensen's alpha: cross-section version

We relate two models: the market model and CAPM model are defined, respectively, as:

$$E[Z_{it}] = \alpha_{im} + \beta_{im} E[Z_{mt}] + \varepsilon_{it}$$

$$E[Z_{it}] = \lambda_0 + \lambda_1 \beta_{im} + \varepsilon_{it}$$

where $\beta_{im} = cov(R_i, R_m)/var(R_m)$

Thus

$$\lambda_0 + \lambda_1 \beta_{im} = \alpha_{im} + \beta_{im} E[Z_{mt}]$$

and under null hypothesis of CAPM model ($E[Z_{mt}] = \lambda_0 + \lambda_1 \beta_{mm}$ where $\beta_{mm} = 1$), it has to be that:

$$\alpha_{im} = \lambda_0 (1 - \beta_{im})$$

Which can be tested by regressing:

$$\alpha_{im} = \delta_1 + \delta_2 (1 - \beta_{im}) + \varsigma_i$$

for $I = 1, ..., N$ assets.

We test $H0 : \delta_2 = 0$. If we do not reject H0, then variable $(1 - \beta_{im})$ does not explain the variability of α_{im}.

E. Alternatives to Test the CAPM

- Alternative approaches:
  - Fama-MacBeth approach:
    - Fama, E. F. and MacBeth, J. D., 1973, "Risk, return and Equilibrium: Empirical Tests", *Journal of Political Economy*, 81(3), p607-636.
  - Fama-French approach:
    - Fama, E. F., and French, K. R., 1992, "The Cross-Section of Expected Stock Returns", *Journal of Finance*, 47, p427-465;
    - Fama, E. F., and French, K. R., 1993, "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics*, 33, p3-53.
  - Carhart approach:
    - Carhart, M. 1997, "On Persistence of Mutual Fund Performance", *Journal of Finance*, 52, p57-82.

F. Empirical evidence

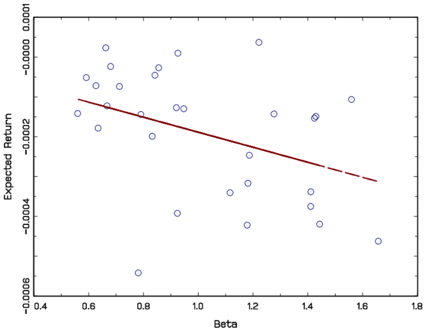
Empirical results when testing implication of the CAPM model:

Under CAPM

- $\lambda_0 = 0$ and $\lambda_1 = E[R_m - R_f]$
- We should find that there is a positive and linear relationship between beta risk and return
- in addition, a high R^2

However

Dow Jones stocks (monthly data)



Negative slope - expect market return also negative and close. Find that idiosyncratic variance strongly significant and negative.

The CAPM criticisms:

- Friedman and Savage (1948), Markowitz (1952) and Kahneman and Tversky (1979) claim that the typical preference must include risk-averse as well as **risk-seeking segments**.
- Thus, the variance is not a good measure of risk, which casts doubt on the validity of the CAPM.
- Normality (or an Elliptic distribution) is crucial to the estimation of the CAPM. The Normal distribution is statistically strongly rejected in the data.
- Omitted variable problem

CAPM

Other fundamental factors that capture risk:

Systematic Factors	What It is	Commonly Captured by
Value	➢ Captures excess returns to stocks that have low prices relative to their fundamental value	➢ Book to price, earnings to price, book value, sales, earnings, cash earnings, net profit, dividends, cash flow
Low Size (Small Cap)	➢ Captures excess returns of smaller firms (by market capitalization) relative to their larger counterparts	➢ Market capitalization (full or free float)
Momentum	➢ Reflects excess returns to stocks with stronger past performance	➢ Relative returns (3-mth, 6-mth, 12-mth, sometimes with last 1 mth excluded), historical alpha
Low Volatility	➢ Captures excess returns to stocks with lower than average volatility, beta, and/or idiosyncratic risk	➢ Standard deviation (1-yr, 2-yrs, 3-yrs), Downside standard deviation, standard deviation of idiosyncratic returns, Beta
Dividend Yield	➢ Captures excess returns to stocks that have higher-than-average dividend yields	➢ Dividend yield
Quality	➢ Captures excess returns to stocks that are characterized by low debt, stable earnings growth, and other "quality" metrics	➢ ROE, earnings stability, dividend growth stability, strength of balance sheet, financial leverage, accounting policies, strength of management, accruals, cash flows

Many assumptions that underlie the model:

- Zero transaction costs.
- Zero taxes.
- Homogeneous investor expectations.
- Available risk-free assets (Treasury bills have various risks: reinvestment risk investors may have investment horizons beyond the T-bill maturity date; inflation risk, currency risk).
- Borrowing at risk-free rates (not the case for non-institutional investors).
- Beta as full measure of risk (but inflation risk, liquidity risk)

Issues with testing

- There exist also macroeconomic factors that capture risk.
- Roll critique: Can not observe the market portfolio. So rejections of CAPM are not valid.
- Market rational bubles (Pastor and Veronesi (2007))
- Heart of CAPM is that diversiffication can eliminate idiosyncratic risk, and so only systematic risk is priced. However, diversiffication is harder to achieve nowadays, i.e., require more stocks to achieve the same risk.

3. Multifactor Pricing Models (APT Ross (1976))

A. Theoretical Background
B. The linear Factor Model
C. The Econometric model
    - Observable factors
    - Statistical factors
    - Characteristic based and macro factor models
D. The MacKinlay Critique

Multifactor pricing

A. Theoretical background

Under non arbitrage opportunities, there exists a **stochastic discount factor (SDF)**, which is a stochastic process M_{t+1} such that for any security with payoff R_{t+1} at time $t+1$, the price of that security at time t is:

$$P_t = E_t[M_{t+1}R_{t+1}]$$

where:

The payoff R_{t+1}: captures risk of the asset
The SDF M_{t+1}: capture the pricing of risk and the discount rate

Present value taking into account the aversion for risk (m represents a change of measure)

Assumptions on SDF define the different models:

Risk-free rate: $R_{ft} = E_t M_{t+1}$
CAPM - SDF: $\log M_{t+1} = a + bR_{mt}$
CCAPM - SDF: $\log M_{t+1} = \log \beta - \gamma \Delta \log C_{t+1}$
APT - SDF: $\log M_{t+1} = a + bf_1 + bf_2 + ... + bf_K$

Asset pricing theory starts with the idea that individual investment contains two types of risk:

Systematic Risk - These are market risks that cannot be diversified away (f_k). Interest rates, recessions and wars are examples of systematic risks.

Unsystematic Risk - Also known as "specific risk," this risk is specific to individual stocks and can be diversified away as the investor increases the number of stocks in his or her portfolio (ε_t). In more technical terms, it represents the component of a stock's return that is not correlated with general market moves.
Thus, the expected return generator process of an asset is:

$$E[R_t] = bf_1 + bf_2 + ... + bf_K + \varepsilon_t$$

B. The Linear Factor model

Asset returns are generated by the linear factor model (LFM). For asset $i \in \{1, ..., n\}$,

$$
\begin{aligned}
R_i &= \alpha_i + \sum_{k=1}^{K} b_{ik} f_k + \varepsilon_i \\
R &= \alpha + Bf + \varepsilon
\end{aligned}
$$

such that f_k are K common factors, b_{ik} are factor loadings (sensitivity of the return on asset i to factor k), ε_i denotes idiosyncratic risk (as opposed to systematic risk of the economy-wide factors) and

$$E\varepsilon_i = 0 \quad ; \quad \mathrm{var}(\varepsilon_i) < \infty.$$

$$\mathrm{cov}(f_k, \varepsilon_i) = 0 \quad ; \quad \mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$$

Portfolios:

A *(unit cost) portfolio* is $w_1, \dots, w_n$ such that

$$\sum_{i=1}^{n} w_i = 1$$

An (zero cost) *arbitrage portfolio* is $w_1, \dots, w_n$ such that

$$\sum_{i=1}^{n} w_i = 0$$

A *well-diversified portfolio* is such that

$$\sum_{i=1}^{n} w_i^2 \approx 0 \text{ (as } n \to \infty)$$

A portfolio that is *hedged against factor risk* (e.g., market neutral) is such that

$$\sum_{i=1}^{n} w_i b_{ik} = 0 \text{ for all } k$$

Multifactor pricing

The model

Then consider the well diversified arbitrage portfolio p that is hedged against factor risk

$$
\begin{aligned}
R_p &= \sum_{i=1}^{n} w_i R_i \\
&= \sum_{i=1}^{n} w_i \alpha_i + \sum_{k=1}^{K} \sum_{i=1}^{n} w_i b_{ik} \cdot f_k + \sum_{i=1}^{n} w_i \varepsilon_i \\
&\approx \sum_{i=1}^{n} w_i \alpha_i \\
&\approx 0
\end{aligned}
$$

otherwise you make money for nothing.

Arbitrage Hedge portfolios

Does a w exist such that

$$w^{\intercal}(i, B) = 0$$

This is just linear algebra. The vector

$$(i, B) = \begin{pmatrix} 1 & b_{11} & \cdots & b_{1K} \\ \vdots & \vdots & & \vdots \\ 1 & b_{n1} & \cdots & b_{nK} \end{pmatrix}$$

generates a (proper) subspace of $\mathbb{R}^n$ (assuming $n > K$) of dimension $K + 1$ (assuming no redundancy in B), lets call it $\mathbb{B}$.
This subspace has an orthogonal complement, denoted $\mathbb{B}^{\perp}$, of dimensions $n - K$, which contains all the vectors orthogonal to $\mathbb{B}$ and hence to (i, B) in particular.

An arbitrage portfolio p that is hedged against factor risk satisfies

$$w^{\top}(i, B) = 0,$$

i.e., w is in the null space of (i, B) ($n \times (K+1)$ matrix).
Since the vector α is orthogonal to w it must lie in the space spanned by (i, B), i.e., $(A^{\perp})^{\perp} = A$.
Therefore for some constants $\rho, \theta_1, \ldots, \theta_k$ we have

$$\alpha_i = \rho + \sum_{k=1}^{K} b_{ik}\theta_k.$$

Multifactor pricing

The model

It follows that for any asset i

$$ER_i = \alpha_i + \sum_{k=1}^{K} b_{ik} Ef_k = \rho + \sum_{k=1}^{K} b_{ik}(Ef_k + \theta_k)$$

i.e., $(ER_1, \ldots, ER_n)^{\top} \in \mathrm{span}(i, B)$ - there exists constants $\lambda_0, \lambda_1, \ldots, \lambda_K$ such that

$$ER_i = \lambda_0 + \sum_{k=1}^{K} b_{ik}\lambda_k,$$

where $\lambda_j = Ef_j + \theta_j$ are risk premia associated with the j factor.
If there is a risk free rate $\lambda_0 = R_f$ and $\lambda_k = Ef_k - R_f$

$$ER_i - R_f = \sum_{k=1}^{K} b_{ik}(Ef_k - R_f)$$

The CAPM corresponds to the case where $K = 1$ and f_1 is the return on the market portfolio.
The theory doesnt say what the factors are.

In general:

Note that any portfolio

$$R_p = \sum_{i=1}^{n} w_{pi} R_i$$

of the assets (with $\sum_{i=1}^{n} w_{pi} = 1$) would satisfy

$$ER_p = \sum_{i=1}^{n} w_{pi} ER_i = \lambda_0 + \sum_{i=1}^{n} w_{pi} \sum_{k=1}^{K} b_{ik}\lambda_k = w_p^{\intercal}(i, B)\lambda,$$

i.e., it is a known function the risk premia $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_K)^{\intercal}$, its own weights w_p, and the betas.
In some cases the factors themselves are taken to be portfolios of traded assets and so their expectation is restricted under the theory.

When does diversiffication work?

Correlation between random variables is not -1 or 1

First, averaging reduces variance provided correlation is not perfect. We have

$$
\begin{aligned}
(\mathrm{std}(X+Y))^2 &= \mathrm{var}(X+Y) \\
&= \mathrm{var}(X)+\mathrm{var}(Y)+2\mathrm{cov}(X,Y) \\
&\leq \mathrm{var}(X)+\mathrm{var}(Y)+2\mathrm{std}(X)\mathrm{std}(Y) \\
&= (\mathrm{std}(X)+\mathrm{std}(Y))^2 .
\end{aligned}
$$

with strict inequality if and only if

$$\mathrm{corr}(X,Y) \neq 1.$$

In the extreme case with $\mathrm{corr}(X,Y) = -1$, the variance of the average can be zero.
Key result: Can show that with many assets, portfolio can have zero (idiosyncratic) variance under some very weak conditions.

Upper bound on the (idiosyncratic) variance of a random variable

Let $\varepsilon \in \mathbb{R}^n$ be an idiosyncratic error term with covariance matrix $E[\varepsilon\varepsilon^{\top}] = \Omega_\varepsilon$. Let $w \in \mathbb{R}^n$ be a vector of portfolio weights. When can we say that

$$w^{\top}\varepsilon \approx 0$$

i.e., $w^{\top}\Omega_\varepsilon w \approx 0$. If $\Omega_\varepsilon = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$, then

$$\mathrm{var}(w^{\top}\varepsilon) = w^{\top}\Omega_\varepsilon w = \sum_{i=1}^{n} w_i^2\sigma_i^2$$

Suppose that $\sigma_i^2 \leq c < \infty$ and suppose that $w_i = 1/n$ for each i. Then

$$\sum_{i=1}^{n} w_i^2\sigma_i^2 \leq \frac{c}{n} \to 0$$

So if all the variances are bounded then clearly diversification works. Even works with some growth on the error variance.

Weekly correlated random variables (idiosyncratic errors)

Also works with weakly correlated idiosyncratic errors. Suppose that

$$\Omega_\varepsilon = D^{1/2}\Psi D^{1/2},$$

where $w^\top D w = \sum_{i=1}^n w_i^2\sigma_i^2 \approx 0$ and Ψ has bounded eigenvalues (as $n \longrightarrow \infty$). For example (with $\rho \in (0,1)$)

$$\Psi = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \ddots & \ddots & \\ \rho^2 & \ddots & 1 & \rho & \rho^2 \\ \vdots & \ddots & \rho & 1 & \rho \\ \rho^{n-1} & \cdots & \rho^2 & \rho & 1 \end{bmatrix}$$

Letting $\widetilde{w} = D^{1/2}w$ we have

$$w^\top \Omega_\varepsilon w = \widetilde{w}^\top \widetilde{w} \frac{\widetilde{w}^\top \Psi \widetilde{w}}{\widetilde{w}^\top \widetilde{w}} \leq \lambda_{\max}(\Psi) \sum_{i=1}^n w_i^2 \sigma_i^2 \longrightarrow 0$$

as $n \longrightarrow \infty$.

When systemic risk is NOT diversifiable?

Pervasive risk: fraud, tone at the top (atmosphere in the workplace), going concern, etc

Factors are pervasive when (for large n)

$$w^\intercal B^\intercal \Sigma_f B w \to M_{K\times K}(w) > 0,$$

where $\Sigma_f = Eff^\intercal$ is the covariance matrix of the factors. For example when $K = 1$, we require that

$$\mathrm{var}(r_{mt})\frac{1}{n}\sum_{i=1}^{n}\beta_i^2 >> 0$$

This means that for any well-diversified portfolio

$$\mathrm{cov}(w^\intercal R_t) = w^\intercal B\Sigma_f B^\intercal w + w^\intercal \Omega_\varepsilon w \simeq M(w) > 0$$

Solniks diversiffication curve

Graphical technique to describe the porportion of diversifiable and non diversifiable risk in a covariance matrix.

Procedure:

- Randomly selected equally weighted portfolio of m assets
- Take the sample variance $S(m)$
- Repeat the process for a large number of assets for $m = m+1, m+2, ...m+, n$ in the equally weighted portfolio
- Graph the $m+n$ sample variance: diversification curve

Formally:

Sample variance $S(m)$ of a randomly selected equally weighted portfolio of m assets for $m = 1, 2, \ldots, n$

Can show that

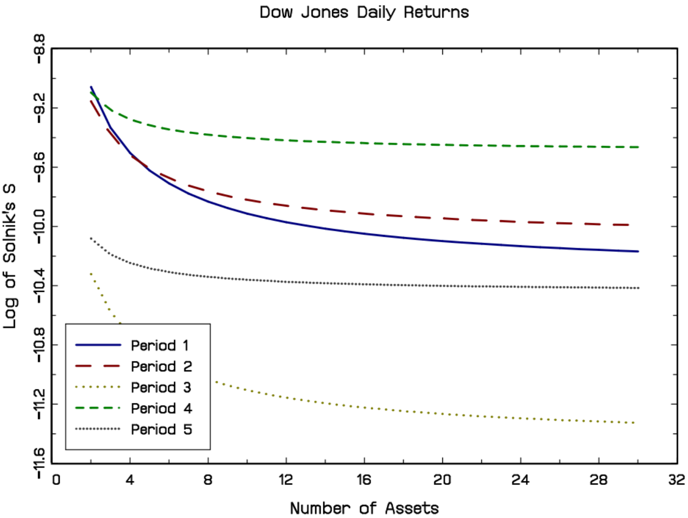
$$\begin{aligned} S(m) &= \frac{1}{m}\overline{\sigma}_i^2 + \left(1 - \frac{1}{m}\right)\overline{\sigma}_{ij} \\ &\to \overline{\sigma}_{ij} \text{ as } m \to \infty \end{aligned}$$

$$\begin{aligned} \overline{\sigma}_i^2 &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{var}(R_i) \\ \overline{\sigma}_{ij} &= \frac{2}{n(n-1)} \sum_{j=i+1}^{n} \sum_{i=1}^{n-1} \mathrm{cov}(R_i, R_j) \end{aligned}$$

Under the market model

$$\overline{\sigma}_{ij} = \mathrm{var}(R_m)\mathrm{ave}_{i \neq j}(\beta_i \beta_j)$$

Figure: Log Solniks curve for five periods



- The curve is steeply downward sloping for small m and then flattens
- Risk-reduction benefits of diversification damper after a relative small number of assets are included
- This motivates the popular *twenty-stock rule* for near complete diversification
- However, Campbell et. al (2001) show that rule does not work under increase in asset specific variance.
- Empirical evidence show that the same level of diversification achievable by twenty stocks portfolio in 1960s require fifty stocks during 1990s.
- This methods is subjected by simulation noise (computer random generator system)

Alternative method: Global Minimum Variance portfolio

For weights w_{mv} with $i = (1, \dots, 1)^{\top} \in \mathbb{R}^m$

$$w_{mv} = \frac{\Sigma^{-1} i}{i^{\top} \Sigma^{-1} i}$$

we achieve variance

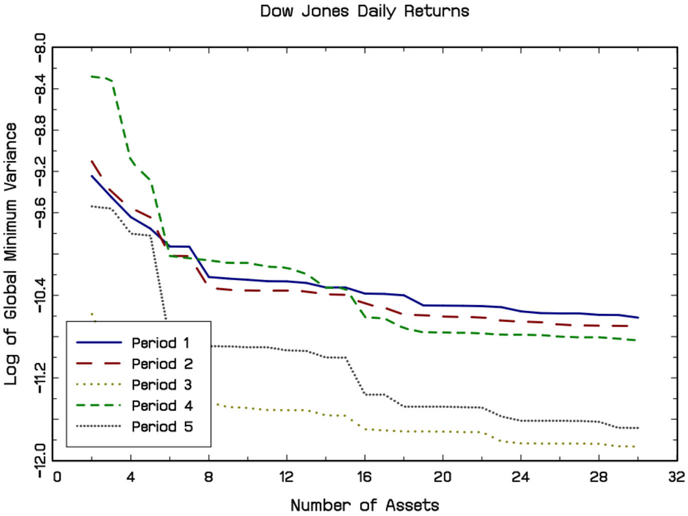
$$\sigma^2_{mv}(m) = \frac{1}{i^{\top} \Sigma^{-1} i}$$

Compute this for assets $R_1, \dots, R_m$ with $m = 2, \dots, n$

Multifactor pricing

The model

Figure: Log of σ^2_{mv} for five different subperiods



C. The Econometric Model

The K-factor model (for returns or excess returns) can be written

$$Z_{it} = \mu_i + \sum_{j=1}^{K} b_{ij} f_{jt} + \varepsilon_{it}$$

$$Z_t = \mu + B f_t + \varepsilon_t,$$

where ε_{it} is an (iid over t) idiosyncratic error term with

$$\begin{aligned} \operatorname{cov}(\varepsilon_t, f_t) &= 0 \\ E(\varepsilon_t \varepsilon_t^{\top}) &= \Sigma \end{aligned}$$

The equivalent of market model. Can be justified from multivariate normality of returns and the factors. We have

$$B = \operatorname{cov}(Z_t, f_t^{\top}) \operatorname{cov}(f_t)^{-1}$$

Different types of Factor Models

There are three different types of factor models

1. Observable factor models (b_{ij} are unknown quantities)
    1. The factors are returns to traded portfolios (specifically, the returns on portfolios formed on the basis of security characteristic such as size, B/M)
    2. The factors are macro variables such as yield spread etc.
2. Statistical factor models (f_{jt} and b_{ij} are unknown quantities)
3. Characteristic based models (b_{ij} are observed characteristics such as industry or country and f_{jt} are unknown quantities)

In each case there are slight differences in cases where there is a risk free asset and in cases where there are not.

C.1 Observable factors: Multivariate test

Risk Free Asset

First suppose that there is a risk-free asset and that the factor returns F are observable

$$Z_t = a + BZ_{Kt} + \varepsilon_t$$

The multivariate tests of $a = 0$ are the same as those for the CAPM; the best performing (as for the CAPM) is J_1 under normality of errors

$$J_1 = \frac{(T-n-K)}{n}(1+\widehat{\mu}_K^{\top}\widehat{\Omega}_K^{-1}\widehat{\mu}_K)^{-1}\widehat{a}^{\top}\widehat{\Sigma}^{-1}\widehat{a} \sim F(n, T-n-K)$$

Without Risk Free Asset and Portfolio return factors

If there is no risk-free asset and the factor returns are returns to observable portfolios then the tests are completely analogous to those for the zero-beta CAPM.

$$R_t = a + BZ_{Kt} + \varepsilon_t$$

$$a = (\iota - B\iota)\gamma_0$$

Macro Factors and Risk Free Asset

Suppose instead that the observed factors are macroeconomic shocks rather than portfolio returns. Then the expected returns associated with the factors have to be estimated as additional free parameters.
We have

$$R = a + Bf + \varepsilon$$

$$E[f] = \mu_K$$

It follows that

$$a = \iota\lambda_0 + B(\lambda_K - \mu_K) = \iota\gamma_0 + B\gamma_1$$

Substituting a into R gives the same estimable model as before except for the additional K parameters to estimate γ_1.

C.2 Statistical factors: Identifying the Factors in Asset Returns

Two methods:

i) Maximum Likelihood factor analysis: estimates factor betas

ii) Asymptotic principal component analysis (APC): estimates factor returns

Both methods use an identification strategy and assume small n and large T (implies not large covariance matrix).

i) Maximum Likelihood factor analysis

$$R_t = \mu + Bf_{Kt} + \varepsilon_t$$

Strict factor structure - idiosyncratic error is cross-sectionally uncorrelated so

$$E\varepsilon_t\varepsilon_t^{\top} = D,$$

where D is a diagonal matrix (containing the idiosyncratic variances). Then

$$\mathrm{cov}(R_t) = E\left[(R_t - E(R_t))(R_t - E(R_t))^{\top}\right] = \Omega_{n\times n} = B\Sigma_f B^{\top} + D,$$

where $\Sigma_f = Ef_tf_t^{\top}$ or $\mathrm{cov}(f_t)$ or $\mathrm{cov}(Z_{Kt})$.
The RHS has $nK + K(K+1)/2 + n$ parameters which is less than $n(n+1)/2$ on LHS.

Identification issue:

When factors are unknown there is an identification issue. Can't uniquely identify (B, Σ_f) or (B, f_t)
One can write for any nonsingular matrix L,

$$f_t^* = Lf_t$$

$$\Sigma_{f^*} = L\Sigma_f L^{\intercal}$$

so that

$$Bf_t = BL^{-1}Lf_t = B^* f_t^*$$

$$B\Sigma_f B^{\intercal} = B^* \Sigma_{f^*} B^{*\intercal}$$

Possible solution:

One solution is to restrict $\Sigma_f = I_K$. In this case,

$$\Omega = BB^\intercal + D.$$

Then B, D are unique (B upto sign)

Maximun Likelihood estimation (Ω)

Can be estimated by maximum likelihood factor analysis provided $n << T$ (*small n and large T*).

$$\ell(\mu, B, D) = c - \frac{T}{2}\log\det\Omega - \frac{1}{2}\sum_{t=1}^{T}(R_t - \mu)^{\top}\Omega^{-1}(R_t - \mu).$$

We have

$$\widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t$$

Total of $nK + n$ parameters in Ω. Iterative nonlinear procedure for B, D, quite time consuming. Solve first order conditions (for $i = 1, \ldots, n$ and $k = 1, \ldots, K$)

$$\begin{aligned} \frac{\partial \ell}{\partial b_{ik}}(\widehat{\mu}, \widehat{B}, \widehat{D}) &= 0 \\ \frac{\partial \ell}{\partial \sigma_i^2}(\widehat{\mu}, \widehat{B}, \widehat{D}) &= 0 \end{aligned}$$

Minimum distance estimation (Ω))

Minimum Distance Estimation. Also iterative nonlinear procedure
Sample covariance matrix $n(n+1)/2$ free parameters

$$\widehat{\Omega} = \frac{1}{T}\sum_{t=1}^{T}\left(R_t - \overline{R}\right)\left(R_t - \overline{R}\right)^{\mathsf{T}} = \left(\frac{1}{T}\sum_{t=1}^{T}\left(R_{it} - \overline{R}_i\right)\left(R_{jt} - \overline{R}_j\right)\right)_{i,j=1}^{n}$$

Find $\widehat{\theta} = (\mathrm{vec}(\widehat{B}), \mathrm{diag}(\widehat{D}))$ to minimize the quadratic form

$$\left(\widehat{\omega} - q(\widehat{\theta})\right)^{\mathsf{T}} W \left(\widehat{\omega} - q(\widehat{\theta})\right)$$

with weighting matrix W and function q such that

$$q(\theta) = \mathrm{vech}(BB^{\mathsf{T}} + D),$$

where $\widehat{\omega} = \mathrm{vech}(\widehat{\Omega})$.

Estimating factors using $\hat{\Omega} \Rightarrow \hat{B}$

Given $\widehat{B}$, the factor returns can be estimated by cross-sectional regression, i.e., OLS or GLS

$$\widehat{f}_t = (\widehat{B}^\intercal \widehat{B})^{-1} \widehat{B}^\intercal (R_t - \widehat{\mu})$$

$$\widehat{f}_t = (\widehat{B}^\intercal \widehat{D}^{-1} \widehat{B})^{-1} \widehat{B}^\intercal \widehat{D}^{-1} (R_t - \widehat{\mu})$$

Consistency requires $\widehat{B}^\intercal \widehat{B} \to \infty$ (pervasive factors).
Estimated factors. Replacing B with $\widehat{B}$ creates an errors in variables problem, affects standard errors at least

Thus we have $\hat{f}_t$

Factor Mimicking Portfolios

- Portfolios that **hedge or mimic factors** are the basic components of various portfolio strategies.
- The mimicking portfolio for a given factor, is the portfolio with the maximum correlation with the factor
- If all assets are correctly priced, then each investor's portfolio should be some combination of cash and the mimicking portfolios
- Other portfolios have the same level of expected return and sensitivities to the factor, but greater variance.
- Interpretation of Cross-sectional GLS/OLS estimates of factor returns as the returns to factor-mimicking portfolios

Multifactor pricing

Multivariate test

Formally,

We can write

$$\widehat{f}_{jt} = \widehat{w}_j^{\top} \left(R_t - \widehat{\mu} \right),$$

where $\widehat{w}_j = ((\widehat{B}^{\top}\widehat{B})^{-1}\widehat{B}^{\top})_j$ or $\widehat{w}_j = ((\widehat{B}^{\top}\widehat{D}^{-1}\widehat{B})^{-1}\widehat{B}^{\top}\widehat{D}^{-1})_j$.

Portfolio with the smallest idiosyncratic variance

Can show that these portfolio weights solve the following problem (taking $\Omega = \widehat{D}$ or $\Omega = I$)

$$\min_{w_j} w_j^{\mathsf{T}} \; \Omega \; w_j$$

subject to

$$w_j^{\mathsf{T}} \widehat{b}_h = 0 \quad h \neq j \quad ; \quad w_j^{\mathsf{T}} \widehat{b}_j = 1$$

The set of portfolios that are hedged against factors h, $h \neq j$ and have unit exposure to factor j is of dimension $n - K$. We are finding the one with smallest idiosyncratic variance. Can normalize the weights $\widehat{w}_j$ to sum to one, so that they are portfolio weights.

ii) Asymptotic Principal Components

An alternative to maximum likelihood factor analysis is asymptotic principal components (*small* T *and large* n). In the population we can write the $T \times 1$ excess return vector as

$$R_i - \mu_i i_T = Fb_i + \varepsilon_i,$$

where F is $T \times K$. Assume (for interpretation) that b_i, σ_i^2 are random variables from a common distribution with $E(b_i b_i^{\top}) = \Sigma_b$ and $E\sigma_i^2 = \sigma^2$(i.e., $\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sigma_i^2$).
Then

$$\Psi_{T\times T} = E\left[(R_i - ER_i)(R_i - ER_i)^{\top}\right] = F\Sigma_b F^{\top} + \sigma^2 I_T,$$

The right hand side has $TK + K(K+1)/2 + 1$ parameters which is less than LHS which has $T(T+1)/2$.

Identification issue:

There is an identification problem: for any nonsingular matrix L

$$Fb_i = FLL^{-1}b_i = F^*b_i^*$$

In this case, assume that (with $\gamma_1 \geq \ldots \geq \gamma_K$)

$$\Sigma_b = \mathrm{diag}\{\gamma_1, \ldots, \gamma_K\} = \Gamma \quad ; \quad F^{\top}F = I_K.$$

$$\Psi = F\Gamma F^{\top} + \sigma^2 I_T,$$

Then F, Γ, σ^2 are unique (F upto sign)

Solution: eigenvalue decomposition

Recall that for any symmetric $T \times T$ matrix Ψ we have the unique eigendecomposition

$$\Psi = Q \Lambda Q^{\top} = \sum_{t=1}^{T} \lambda_t q_t q_t^{\top}$$

where eigenvectors Q satisfy

$$QQ^{\top} = Q^{\top} Q = I_T$$

and eigenvalues

$$\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_T\}$$

ordered from largest to smallest.

Let G be a $T \times T - K$ matrix such that $Q = (F_{T\times K}, G_{T\times T-K})$ satisfies

$$QQ^{\top} = Q^{\top}Q = I_T.$$

Provided F has full rank (K) this Q exists and is unique.
We have

$$\begin{aligned}
\Psi &= F\Gamma F^{\top} + \sigma^2 I_T \\
&= (F, G)\begin{bmatrix} \Gamma & 0 \\ 0 & 0 \end{bmatrix}(F, G)^{\top} + \sigma^2 I_T \\
&= Q\begin{bmatrix} \Gamma + \sigma^2 I_K & 0 \\ 0 & \sigma^2 I_{T-K} \end{bmatrix}Q^{\top},
\end{aligned}$$

It follows that F are the eigenvectors corresponding to the K largest eigenvalues of Ψ

$$F = \mathrm{eigvec}_K[\Psi]$$

This shows unique identification of F. The Γ, σ^2 are also uniquely identified by this.

Procedure:

The sample covariance matrix

$$\widehat{\Psi} = \frac{1}{n} Z^{\intercal} Z = \left(\frac{1}{n} \sum_{i=1}^{n} (Z_{it} - \overline{Z}_t)(Z_{is} - \overline{Z}_s) \right)_{s,t=1}^{T},$$

which is a $T \times T$ matrix of excess return cross-products. Do the empirical eigendecomposition and take

$$\widehat{F} = \mathrm{eigvec}_K [\widehat{\Psi}]$$

Given this estimate of F, the factor betas can be estimated by time-series OLS regression

$$b_i = (\widehat{F}^{\intercal} \widehat{F})^{-1} \widehat{F}^{\intercal} (R_i - \widehat{\mu}_i i_T)$$

APC issue:

The main problem with this approach is that it **assumes time series homoskedasticity** for the idiosyncratic error, which is not a good assumption.

Jones' Solution: Iterative application of APC

Jones (2001, JFE) has extended the estimation problem to allow for time varying average idiosnycratic variance.

$$\Psi = F\Gamma F^{\top} + D,$$

where

$$D = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_T^2\}$$

Iterative application of APC. That is, given first round estimates, calculate the time series heterosekdasticity

$$\widehat{D} = \mathrm{diag}\{\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_T^2\}$$

and then recompute the factors

$$\widehat{F} = \mathrm{eigvec}_K[\widehat{\Psi} - \widehat{D}]$$

and loadings likewise and iterate

Large n and T (large covariance matrix)

Bai and Ng (2002) consider the (constrained) least squares procedure for estimation simultaneously of $b_1, \ldots, b_n, f_1, \ldots, f_T$

$$\sum_{t=1}^{T} \sum_{i=1}^{n} \left\{ r_{it} - b_i^{\mathsf{T}} f_t \right\}^2$$

subject to the identification constraint either that $B^{\mathsf{T}} B / n = I_K$ or $F^{\mathsf{T}} F = I_K$

- The procedure can be understood as iterative least squares (cross-section regression then time series regression then etc)
- They show consistency of this procedure when n and T are large
- They also propose model selection method to determine the number of factors K
- BUT, curse of dimensionality problem

C.3 Characteristics based and macro factor models

Rosenberg (1974): Explicit characteristics

Commonly used in asset management industry. Rosenberg (1974)

$$R_t = Bf_t + \varepsilon_t$$

Relate B to observable stock (time invariant) characteristics such as: industry/country (dummy variables). Other characterstics: size, value, and momentum characteristics (normalized to have mean zero and one cross sectionally) so $B_{ij} = X_{ij}$ are observed

$$R_{it} = \sum_{j=1}^{J} X_{ij} f_{jt} + \varepsilon_{it}$$

Linear cross sectional regression at each time point

$$\widehat{f}_t = (X^\top X)^{-1} X^\top R_t$$

Fama and French (1993):

- Fama and French (1993) Construct (6) double sorted portfolios formed on 2 size and 3 book to market
    - The size factor return is proxied by the difference in return between a portfolio of low-capitalization stocks and a portfolio of high-capitalization stocks, adjusted to have roughly equal book-to-price ratios (SMB)
    - The value factor is proxied by the difference in return between a portfolio of high book-to-price stocks and a portfolio of low book-to-price stocks, adjusted to have roughly equal capitalization (HML)
    - The market factor return is proxied by the excess return to a value-weighted market index (MKT)

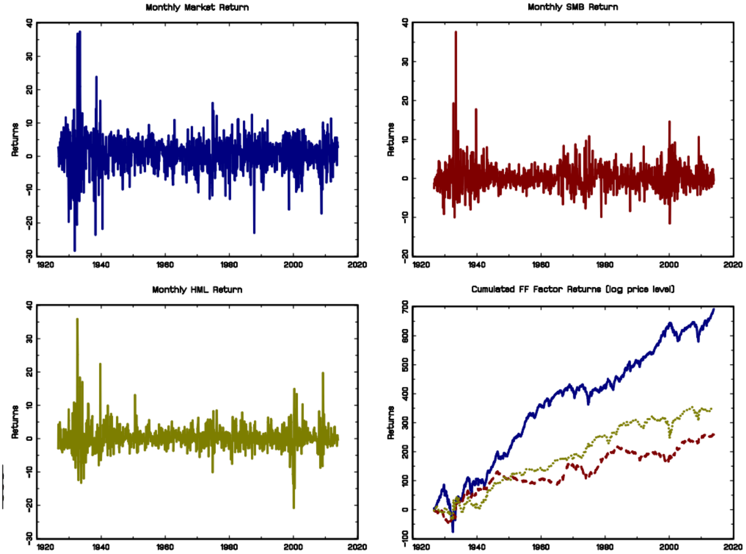
$$SMB = \frac{1}{3}\left[\begin{array}{c} \left(\frac{\text{large size}}{\text{high BTP}} - \frac{\text{small size}}{\text{high BTP}}\right) \\ + \left(\frac{\text{large size}}{\text{medium BTP}} - \frac{\text{small size}}{\text{medium BTP}}\right) \\ + \left(\frac{\text{large size}}{\text{lowl BTP}} - \frac{\text{small size}}{\text{low BTP}}\right) \end{array}\right]$$

$$HML = \frac{1}{2}\left[\begin{array}{c} \left(\frac{\text{high BTP}}{\text{large size}} - \frac{\text{low BTP}}{\text{large size}}\right) \\ + \left(\frac{\text{high BTP}}{\text{small size}} - \frac{\text{low BTP}}{\text{small size}}\right) \end{array}\right]$$

Data on the factors is available from Kenneth French web site.

Multifactor pricing

Characteristics



Multifactor pricing

Characteristics

- Then form 25 portfolios (on size and value characteristics too!)

$$R_{pt} = a_p + b_{p,SMB} SMB_t + b_{p,HML} HML_t + b_{p,MKT} MKT_t + \varepsilon_{pt}$$

  They do not reject the APT restrictions in their sample (although they do reject CAPM where SMB and HML are dropped). They explain well the size and value anomalies of the CAPM.

- Additional factors have been proposed in the literature
  - Momentum factor, Carhart (1997).
  - Own-volatility factor, Goyal and Santa Clara (2003).
  - Liquidity, Amihud and Mendelson (1986), Pastor and Stambaugh (2003)

Non parametric estimation:

Connor, Hagmann, and Linton (2012, Econometrica)

$$R_{it} = f_{ut} + \sum_{j=1}^{J} \beta_j(X_{ji}) f_{jt} + \varepsilon_{it}$$

Nonparametric method for estimating the beta functions and then cross-sectional regression to get the factors.
Find nonlinear shapes in β_j and good out of sample fit.

Macroeconomics factors (Chan et al. (1985) and Chen et al. (1986)).

$$R_{it} = a_i + b_i^\top f_t + \varepsilon_{it}$$

$$f_t = m_t - E_{t-1} m_t$$

where m_t are (nonstationary) macroeconomic variables. Monthly data. 1958-1984

1. The percentage change in industrial production (led by one period)
2. A measure of unexpected inflation
3. The change in expected inflation
4. The difference in returns on low-grade (Baa and under) corporate bonds and long term government bonds (junk spread)
5. The difference in returns on long-term government bonds and short term Tbills (Term spread)

Procedure: they estimate b_i by time series regressions and then do cross sectional regressions on $\hat{b}_i$ to estimate factor risk premia

- 20 portfolios are used on the basis of form size at the beginning of the period.
- They find that average factor risk premia are statistically significant over the entire sample period for industrial production, unexpected inflation, and junk
- They also include a market return, but find it is not significant
- Recent developments: Macroeconomic policy endogeneity to level of asset prices. Event study approach.

Number of estimated parameters: comparison

A statistical factor model requires estimation of

$$\overbrace{nK}^{B} + \overbrace{kT}^{F} + \overbrace{n}^{D}$$

parameters in total using nT observations on returns
A macroeconomic factor model requires estimation of

$$\overbrace{nK}^{B} + \overbrace{K(K+1)/2}^{\Sigma_f} + \overbrace{n}^{D}$$

parameters in total using nT observations on returns and KT observations on macro factors
A characteristic based factor model requires estimation of

$$\overbrace{KT}^{F} + \overbrace{K(K+1)/2}^{\Sigma_f} + \overbrace{n}^{D}$$

parameters in total using nT observations on returns and the nK set of security characteristics. For large n, this works better.

Concluding empirical remarks: rational risk

- The studies of Fama and French (1993, 1996) show that there exists covariation in returns related to size and value, which are captured by loadings on SMB and HML factors, and beyond the covariation is explained by the market return.
- It suggests that the three factors in the model capture much of the common variation in portfolio returns that is missed by univariate risk factor (betas) in the CAPM.
- The CAPM anomalies reflect the fact that size and value are proxies for **distress**.
- Small stocks and high value stocks have high average returns because they are risky, for which investors require a positive risk premium.

Concluding empirical remarks: irrational pricing?

- In contrast to the risk-based explanation, behavioralists explain the size and value anomalies as a result of irrational pricing.
- Lakonishok, Shleifer and Vishny (1994) suggest that higher returns associated with small stocks and value stocks are due to mispricing.
- They interpret the value anomaly as investors tend to extrapolate firms' past performance into the future.
- Therefore, prices of growth stocks (low value) are usually too high as a result of over-optimistic expectations.
- Nevertheless, these pricing errors will eventually be corrected, resulting low returns for growth stocks.

Critical issue:

Determine whether size and value are proxies for common risk factors, or irrational mispricing (Barber and Lyon, 1997)

F. The MacKinlay (1995) Critique

Criticism of the three-factor model has centered on data mining (or data snooping) and survivorship bias (sample-selection bias).

- The data mining story predicts that the size and value effects would disappear in out-of-sample tests.
- In other words, when analyze another time period or another data source, the three-factor model will reduce to the CAPM, and the three factors will be completed explained by CAPM betas if the data mining story holds.
- Results change using annual rather than monthly data.
- They attribute the contradictions to the survivorship bias: they include distressed firms that have survived and exclude distressed firms that have failed.

Alternative Explanations of Observed Premia

- Lo and MacKinlay and others note the pretest bias of multifactor models, based on size and value-related factors, that improve upon the pricing performance of the CAPM.
- Fama and French claim that there are pervasive factors associated with size and value characteristics and that is why they carry risk premia
- Some other researchers claim that it is irrationality, not risk premia, which best explain the return premia associated with these characteristics
- Mackinlay uses the notion of approximate arbitrage to differentiate between risk-factor based and non-risk-factor based explanations

More in Empirical Asset Pricing

Further topics in Empirical Asset Pricing:

- Intertemporal Equilibrium pricing models (I-CAPM)
- Consumption based Equilibrium pricing models (C-CAPM)
- Event study analysis (Fama, Fisher, Jensen, Roll (1969))
- Present-value relations (fundamentals vs Bubles)
- Habits models (Constantinides (1990))
- Long run risks model Bansal and Yaron (2004)
- Pricing models for: derivatives, fixed income securities and term structure models