II.2. Empirical Testing for Asset Pricing Models

2.1 The Capital Asset Pricing Model, CAPM

Roots of CAPM are in the Markowitz (mean variance) portfolio theory in the 1950’s.

CAPM is credited to Sharpe, Lintner and Mossin in the beginning of 1960’s, and is the first model to quantify risk and the reward for bearing it.

CAPM marks the birth of the (modern) asset pricing theory. It is widely used in applications (estimation the cost of capital, portfolio performance evaluation, etc).

However, the empirical record of the model is (very) poor.
Traditional CAPM

\( \mathbb{E}[r_i] = r_f + \beta_{im}(\mathbb{E}[r_m] - r_f), \) \hspace{1cm} (1)

where

\( \beta_{im} = \frac{\text{Cov}[r_i, r_m]}{\text{Var}[r_m]}, \) \hspace{1cm} (2)

\( r_m \) is the market return, and \( r_f \) is the risk free return.

In the excess return form

\( \mathbb{E}[z_i] = \beta_{im}\mathbb{E}[z_m], \) \hspace{1cm} (3)

where \( z_i = z_i - r_f \) and \( z_m = r_m - r_f \) are the excess returns on the stock and market, respectively.
Empirical testing relies usually to this excess return form by focusing on three implications:

(1) The intercept is zero

(2) Beta completely captures the cross-sectional variation of expected returns.

(3) The market risk premium \( \lambda = E[\tilde{z}_m] \) is positive.
Black’s Zero-Beta CAPM

Black† zero-beta generalization of the CAPM

\[ E[r_i] = E[r_{0m}] + \beta_{im}E[r_m - r_{0m}], \]

where \( r_{0m} \) is the return of the zero-beta portfolio, which is the portfolio that has minimum variance among all portfolios uncorrelated with the market portfolio \( m \).

Returns are measured in real terms and

\[ \beta_{im} = \frac{\text{Cov}[r_i, r_m]}{\text{Var}[r_m]}. \]

The Black model can be tested as a restriction of the real-return market model

\[ E[r_i] = \alpha_{im} + \beta_{im}E[r_m] \]

such that

\[ \alpha_{im} = E[r_{0m}](1 - \beta_{im}). \]

Remark 2.1:
(1) CAPM is a single period model, and does not have a time index.

(2) In econometric analysis the model must be estimated over time.

(3) Thus it is necessary to add assumptions concerning time-series behavior of the returns: The traditional assumptions are: The returns are independent and identically distributed (iid) and jointly multivariate normal.
2.2 Efficient Set Mathematics

The key testable implication of the CAPM is that the market portfolio is mean-variance efficient, i.e., is the minimum-variance portfolio (definition below) among portfolios that have expected returns larger than or equal to \( \mu_m = \mathbb{E}[r_m] \).

Suppose there are \( N \) risky assets with mean vector \( \mu \) and covariance matrix \( \Omega \).

**Definition 2.1:** A portfolio \( P \) is the minimum-variance portfolio (MVP) of all portfolios with mean return \( \mu_p \) if its portfolio weight vector \( \omega_p \) satisfies

\[
\omega'_p \Omega \omega_p = \min_{\omega} \omega' \Omega \omega \tag{8}
\]

under the restrictions

\[
\omega'_p \mu = \mu_p \tag{9}
\]

and

\[
\omega'_p \iota = 1, \tag{10}
\]

where \( \iota \) is an \( N \)-vector of ones.

**Remark 2.2:** MVP is not necessarily efficient.
The weight vector $\omega_p$ is found by using the Lagrangian method with

$$L = \omega'\Omega\omega - \delta_1(\omega'\mu - \mu_p) - \delta_2(\omega'\iota - 1).$$

\hspace{1cm} (11)

Taking partial derivatives and setting them to zero yields

$$\frac{\partial L}{\partial \omega} = 2\Omega\omega - \delta_1\mu - \delta_2\iota = 0$$
$$\frac{\partial L}{\partial \delta_1} = \omega'\mu - \mu_p = 0$$
$$\frac{\partial L}{\partial \delta_2} = \omega'\iota - 1 = 0.$$
Solving for the equations gives the $N \times 1$ weight vector $\omega_p$

$\omega_p = g + h\mu_p,$

where

$g = \frac{1}{d}\left(b\Omega^{-1} - a\Omega^{-1}\mu\right)$

$h = \frac{1}{d}\left(c\Omega^{-1} - a\Omega^{-1}\mu\right),$

and $a = \Omega^{-1}\mu$, $b = \mu'\Omega^{-1} \mu$, $c = \Omega^{-1} \mu$, and $d = bc - a^2$.

Exercise 2.1: Prove results (12)–(14). (For a solution click here.)
Results:

1. The minimum variance frontier can be generated from any two distinct MVPs.

2. Any portfolio of MVPs is also a MVP. Click here for a proof.

3. Let $p$ and $r$ be two MVPs with returns $R_p$ and $R_r$. Then

\[
\text{Cov}[r_p, r_r] = \frac{c}{d} \left( \mu_p - \frac{a}{c} \right) \left( \mu_r - \frac{a}{c} \right) + \frac{1}{c}.
\]

4. Let $g$ denote the global MVP. Then

\[
\omega_g = \frac{1}{c} \Omega^{-1} \nu,
\]

\[
\mu_g = \frac{a}{c},
\]

and

\[
\sigma_g^2 = \frac{1}{c}.
\]

Click here for a proof.

Remark 2.3: $\omega_g$ is again of the form (12) with $\mu_p = \mu_g = a/c$, which reduces the weight vector $\omega_g$ to the form given above.

Exercise 2.2: Prove results (15)–(18).
For each MVP $p$, except the global MVP $g$, there exists a unique MVP that has zero covariance with $p$. It is called the zero-beta portfolio w.r.t $p$.

If $r_q$ is the return of any asset or portfolio then

\[
\text{Cov}[r_g, r_q] = \frac{1}{c} \tag{19}
\]

where $r_g$ is the return of the global MVP.

Remark 2.4: In the absence of riskfree rate the efficient portfolios consist of those MVPs with expected returns higher than or equal to $\mu_g$.

Let $r_q$ denote the return of an asset or portfolio $q$, and let $r_p$ be any MVP (except the global MVP) and $r_{0p}$ the return of the zero-beta portfolio w.r.t $p$. Consider the regression

\[
r_q = \beta_0 + \beta_1 r_{0p} + \beta_2 r_p + \epsilon_p, \tag{20}
\]

where

\[
\mathbb{E}[\epsilon_p \mid r_p, r_{0p}] = 0. \tag{21}
\]
Then the Least Squares solution gives

\[ \beta_2 = \frac{\text{Cov}[r_q, R_p]}{\sigma_p^2} = \beta_{qp}, \]

(22)

where \( \beta_{qp} \) is the beta of asset \( q \) w.r.t portfolio \( p \),

\[ \beta_1 = \frac{\text{Cov}[r_q, r_{0p}]}{\sigma_{0p}^2} = 1 - \beta_{qp}, \]

(23)

and

\[ \beta_0 = 0. \]

(24)

Click here for a proof

5°

\[ \text{E}[r_q] = \mu_q = (1 - \beta_{qp})\mu_{0p} + \beta_{qp}\mu_p. \]

(25)

Exercise 2.3: Prove result 5.
**Portfolios with Riskfree Asset**

Relax the assumption $\omega^'\iota = 1$ and denote the fraction invested in the riskfree asset by $1 - \omega^'\iota$.

The optimization problem then is

(26) \[ \min_{\omega} \omega^'\Omega\omega \]

subject to

(27) \[ \omega^'\mu + (1 - \omega^'\iota) R_f = \mu_p, \]

where $R_f$ is the riskfree return.

The Lagrangian function is then

(28) \[ L = \omega^'\Omega\omega - \delta (\omega^'\mu + (1 - \omega^'\iota)R_f - \mu_p). \]
The partial derivatives are

\[ \frac{\partial L}{\partial \omega} = 2\Omega \omega - \delta (\mu - R_f \iota) \]  
(29)

\[ \frac{\partial L}{\partial \delta} = \mu_p - \omega' \mu - (1 - \omega' \iota) R_f. \]  
(30)

Setting these to zero and solving for the unknowns yields

\[ \omega_p = \frac{\mu_p - R_f}{(\mu - R_f \iota) \Omega^{-1}(\mu - R_f \iota)} \Omega^{-1}(\mu - R_f \iota) \]  
\[ = c_p \bar{\omega}, \]  
(31)

where

\[ c_p = \frac{\mu_p - r_f}{(\mu - r_f \iota) \Omega^{-1}(\mu - r_f \iota)} \]  
(32)

and

\[ \bar{\omega} = \Omega^{-1}(\mu - R_f \iota). \]  
(33)
Thus the optimal portfolio is a combination of riskfree asset and risky assets with weights $\omega_p$ proportional to $\bar{\omega}$.

The efficient frontier in the return standard deviation space is in this case the straight line through $(0, r_f)$ and $(\sigma_q, r_q)$. The former is obtained by selecting $\mu_p = r_f$ which implies $\omega_p = 0$. The latter is obtained by selecting

$$\omega' q = c_q \bar{\omega}' = 1,$$

so that $\omega' q = c_q \bar{\omega}' = 1$, and

$$\omega_q = \frac{1}{\mu - r_f^i} \Omega^{-1}(\mu - r_f^i).$$

Portfolio $q$ is called the *tangency portfolio*. 

14
The **Sharpe ratio** measures the expected excess return per unit risk, and provides a useful basis for interpretation of tests of the CAPM. For any portfolio $A$ the Sharpe ratio is defined as

$$\text{SR}_A = \frac{\mu_A - r_f}{\sigma_A}. \quad (36)$$

The tangency portfolio $q$ has the maximum Sharpe ratio among the portfolios of risky assets. Consequently testing for a mean-variance efficiency of a given portfolio is equivalent to testing the portfolio has the maximum Sharpe ratio.
2.3 Statistical Framework for Estimation and Testing

In Sharpe-Lintner CAPM borrowing and lending rates are assumed to be the same.

Let $\mathbf{z}$ denote an $N \times 1$ vector of excess returns, then the market model of the returns is

$$\mathbf{z}_t = \alpha + \beta \mathbf{z}_{mt} + \epsilon_t,$$

(37)

where $\alpha = (\alpha_1, \ldots, \alpha_N)'$ and $\beta = (\beta_1, \ldots, \beta_N)'$ are $N$ vectors of intercepts and security betas, respectively, $\mathbf{z}_{mt}$ is the market excess return (scalar), $\epsilon_t$ is an $N$-vector of residuals with

$$E[\epsilon_t] = 0,$$

(38)

$$\text{Cov}[\epsilon_t] = E[\epsilon_t \epsilon_t'] = \Sigma.$$

(39)

$$E[Z_{mt}] = \mu_m, \text{ Var}[z_{mt}] = \sigma_m^2.$$  

(40)
Finally, let

\( \mathbb{E}[z_t] = \mu \) (41)

be the \( N \times 1 \) mean vector of the excess returns.

As discussed earlier the CAPM implies that \( \alpha = 0 \) in the above market model. This implies that the market portfolio is the tangency portfolio of the efficient frontier. As a consequence the test of the null hypothesis

\( H_0 : \alpha = 0 \) (42)

can be considered at the same time a test for the mean variance efficiency of the market portfolio.
The parameters of the model can be estimated by OLS that are the same as the Maximum Likelihood (ML) under the normality assumption. In this case the probability density function (pdf) of the excess return vector \( \mathbf{z}_t \) is

\[
f(\mathbf{z}_t|\mathbf{z}_{mt}) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{z}_t - \alpha - \beta \mathbf{z}_{mt})' \Sigma^{-1} (\mathbf{z}_t - \alpha - \beta \mathbf{z}_{mt}) \right].
\]

(43)

The joint pdf of the \( T \) temporally independent observations (considered as random variables) is

\[
f(\mathbf{z}_1, \ldots, \mathbf{z}_T|\mathbf{z}_{m1}, \ldots, \mathbf{z}_{mT}) = \prod_{t=1}^{T} f(\mathbf{z}_t|\mathbf{z}_{mt})
= \prod_{t=1}^{T} \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{z}_t - \alpha - \beta \mathbf{z}_{mt})' \Sigma^{-1} (\mathbf{z}_t - \alpha - \beta \mathbf{z}_{mt}) \right]
= \frac{1}{(2\pi)^{NT/2} |\Sigma|^{T/2}} \exp \left[ -\frac{1}{2} \sum_{t=1}^{T} (\mathbf{z}_t - \alpha - \beta \mathbf{z}_{mt})' \Sigma^{-1} (\mathbf{z}_t - \alpha - \beta \mathbf{z}_{mt}) \right].
\]

(44)
The **log-likelihood** function, the logarithm of the pdf viewed as a function of the parameters, is then

$$
\ell(\alpha, \beta, \Sigma) = -\frac{NT}{2} \log(2\pi) - \frac{T}{2} \log |\Sigma|
$$

$$
-\frac{1}{2} \sum_{t=1}^{T} (z_t - \alpha - \beta z_{mt})' \Sigma^{-1} (z_t - \alpha - \beta z_{mt}).
$$
The ML estimators of the parameters are those maximizing the likelihood function. These are found by setting the partial derivatives w.r.t. of the likelihood function equal to zero and solving for the unknowns. That is (c.f. the derivation rules in EIE Ch 2)

\[
\frac{\partial \ell}{\partial \alpha} = \Sigma^{-1} \left[ \sum_{t=1}^{T} (z_t - \alpha - \beta z_{mt}) \right]
\]

\[
\frac{\partial \ell}{\partial \beta} = \Sigma^{-1} \left[ \sum_{t=1}^{T} (z_t - \alpha - \beta z_{mt}) z_{mt} \right]
\]

\[
\frac{\partial \ell}{\partial \Sigma} = -\frac{T}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} \left[ \sum_{t=1}^{T} (z_t - \alpha - \beta z_{mt})(z_t - \alpha - \beta z_{mt})' \right] \Sigma^{-1}
\]
Setting these to zero and solving gives ML estimators that the same as OLS

$$\hat{\alpha} = \hat{\mu} - \hat{\beta}\hat{\mu}_m$$ \hspace{1cm} (49)

$$\hat{\beta} = \frac{\sum_{t=1}^{T}(z_t - \hat{\mu})(z_t - \hat{\mu})'}{\sum_{t=1}^{T}(z_{mt} - \hat{\mu}_m)^2}$$ \hspace{1cm} (50)

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T}(z_t - \hat{\alpha} - \hat{\beta}z_{mt})(z_t - \hat{\alpha} - \hat{\beta}z_{mt})'$$ \hspace{1cm} (51)

where

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} z_t \text{ and } \hat{\mu}_m = \frac{1}{T} \sum_{t=1}^{T} z_{mt}.$$ \hspace{1cm} (52)
In order to determine likelihood ratio tests we need to determine explicitly the value of the likelihood at the maximum. The maximum is

\[(53) \ell(\hat{\alpha}, \hat{\beta}, \hat{\Sigma}) = -\frac{NT}{2} \log(2\pi) - \frac{T}{2} \log |\hat{\Sigma}| - \frac{NT}{2}, \]

of which the essential part is just

\[(54) -\frac{T}{2} \log |\hat{\Sigma}|.\]

In (53) we have used the property \(\text{tr}(AB) = \text{tr}(BA)\), where \(A\) is an \(m \times n\) and \(B\) is an \(n \times m\) matrix, and \(\text{tr}\) is the trace operator which gives the sum of the diagonal elements of a square matrix (note especially that \(\text{tr}(I) = N\), the dimension of the \(N \times N\) identity matrix).
More precisely, the last term in (53) is obtained as

\begin{align*}
\sum_{t=1}^{T}(z_t - \hat{\alpha} - \hat{\beta}z_{mt})' \hat{\Sigma}^{-1}(z_t - \hat{\alpha} - \hat{\beta}z_{mt}) \\
= \sum_{t=1}^{T} \text{tr} \left[ (z_t - \hat{\alpha} - \hat{\beta}z_{mt})' \hat{\Sigma}^{-1}(z_t - \hat{\alpha} - \hat{\beta}z_{mt}) \right] \\
= \sum_{t=1}^{T} \text{tr} \left[ \hat{\Sigma}^{-1}(z_t - \hat{\alpha} - \hat{\beta}z_{mt})(z_t - \hat{\alpha} - \hat{\beta}z_{mt})' \right] \\
= \text{tr} \left[ \hat{\Sigma}^{-1} \sum_{t=1}^{T}(z_t - \hat{\alpha} - \hat{\beta}z_{mt})(z_t - \hat{\alpha} - \hat{\beta}z_{mt})' \right] \\
= \text{tr} \left[ \hat{\Sigma}^{-1}(T\hat{\Sigma}) \right] = T \text{tr}(I) \\
= NT,
\end{align*}
In order to make statistical inference of the parameters, we need the distributions of the estimators.

Given observations on the market excess returns $z_{m1}, \ldots, z_{mT}$ the conditional distribution of the estimators under the normality assumption are (For a proof of the distribution statement for $\hat{\alpha}$ click the formula.)

\begin{equation}
\hat{\alpha} \sim N \left( \alpha, \frac{1}{T} \left[ 1 + \frac{\mu_m^2}{\sigma_m^2} \right] \Sigma \right),
\end{equation}

\begin{equation}
\hat{\beta} \sim N \left( \beta, \frac{1}{T \hat{\sigma}_m^2} \Sigma \right),
\end{equation}

and

\begin{equation}
T \hat{\Sigma} \sim W(T - 2, \Sigma)
\end{equation}

the Wishart distribution with $T - 2$ degrees of freedom (a generalization of the chi-squared distribution to matrices).

\begin{equation}
\text{Cov}[\hat{\alpha}, \hat{\beta}] = -\frac{1}{T} \left( \frac{\mu_m}{\hat{\sigma}_m^2} \right) \Sigma.
\end{equation}
Wald Test

Wald test for testing the null hypothesis (Click here for a brief introduction to Wald-statistic)

\[(59)\quad H_0 : \alpha = 0\]

against

\[(60)\quad H_1 : \alpha \neq 0\]

is

\[(61)\quad W = \hat{\alpha}' [\text{cov}][\hat{\alpha}]^{-1} \hat{\alpha} = T \left(1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}\right)^{-1} \hat{\alpha}' \Sigma^{-1} \hat{\alpha}.\]

Under the null hypothesis \(W\) is chi-squared distributed with \(N\) (number of assets) degrees of freedom.

In practice \(\Sigma\) must be substituted by its estimator \(\hat{\Sigma}\) implying that the distribution property is the asymptotic.
**F-test**

The following slight modification of $W$ has better small sample properties

$$F = \frac{T - N - 1}{N} \left(1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}\right)^{-1} \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha},$$

which has the $F$-distribution with degrees of freedom $N$ and $T - N - 1$. 
Likelihood Ratio (LR) test

The Likelihood ratio test is

\[ LR = -2(\ell^* - \ell), \]

where \( \ell^* \) is the value of the maximum of the log likelihood function under the restriction of the null hypothesis, i.e., estimating the market model under the constraint \( \alpha = 0 \), and \( \ell \) is the global maximum given by (53).

The maximum of the log-likelihood under the null hypothesis, \( H_0 : \alpha = 0 \) is obtained in the same manner as in the unrestricted case.
The log-likelihood function under the null hypothesis is

\[
\ell^* (\beta, \Sigma) = -\frac{NT}{2} \log(2\pi) - \frac{T}{2} \log |\Sigma| \\
- \frac{1}{2} \sum_{t=1}^{T} (z_t - \beta z_{mt})' \Sigma^{-1} (z_t - \beta z_{mt})
\]

(64)

which gives the restricted ML-estimators

(65) \[ \hat{\beta}^* = \frac{\sum_{t=1}^{T} z_t z_{mt}}{\sum_{t=1}^{T} z_{mt}^2} \]

and

(66) \[ \hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^{T} (z_t - \hat{\beta}^* z_{mt})(z_t - \hat{\beta}^* z_{mt})' \]
In the same manner as in the unrestricted case we obtain the maximum of the log-likelihood as

\[(67) \quad \ell^* = \ell^*(\hat{\beta}^*, \hat{\Sigma}^*) = -\frac{NT}{2} - \frac{T}{2} \log |\hat{\Sigma}^*| - \frac{NT}{2}\]

Using this and (53) in (63), we obtain

\[(68) \quad LR_1 = T \left[ \log |\hat{\Sigma}^*| - \log |\hat{\Sigma}| \right].\]

Again, under the null hypothesis \( LR_1 \sim \chi^2_N \) asymptotically.

Here, however, it is pretty straightforward to show that

\[(69) \quad F_1 = \frac{T - N - 1}{N} (\exp[LR_1/T] - 1) \sim F(N, T - N - 1),\]

which is a monotonic transformation of \( LR_1 \), implying that \( F_1 \) is essentially a LR-test.
A third test is a minor modification of the original LR-statistic\textsuperscript{†}

\begin{equation}
LR_2 = \frac{\frac{T}{N} - 2}{T} LR_1 \\
= (T - \frac{N}{2} - 2) \left( \log |\hat{\Sigma}^*| - \log |\hat{\Sigma}| \right) \sim \chi^2_N
\end{equation}

asymptotically under the null hypothesis. This statistic is supposed to be more accurate in small samples.

Remark 2.5: We can write

\[ \hat{\beta}^* = \hat{\beta} + \frac{\mu_m}{\mu_m^2 + \sigma_m^2} \hat{\alpha}, \]

and,

\[ \hat{\Sigma}^* = \hat{\Sigma} + \left( \frac{\sigma_m^2}{\mu_m^2 + \sigma_m^2} \right) \hat{\alpha} \hat{\alpha}' = \hat{\Sigma} + \left( \frac{1}{1 + \frac{\mu_m^2}{\sigma_m^2}} \right) \hat{\alpha} \hat{\alpha}'. \]

From which

\[ |\hat{\Sigma}^*| = |\hat{\Sigma}| \left[ \left( \frac{1}{1 + \frac{\mu_m^2}{\sigma_m^2}} \right) \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} + 1 \right]. \]

Thus, in this special case, the likelihood ratio test statistic (68) or the small sample corrected LR-test statistic (70) can be derived without estimating the restricted model, such that, e.g., (68) becomes

\[ LR = \frac{N}{2} \log \left[ \left( \frac{1}{1 + \frac{\mu_m^2}{\sigma_m^2}} \right) \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} + 1 \right]. \]
Remark 2.6: It can be shown that

\[ \hat{\alpha}' \Sigma^{-1} \hat{\alpha} = \frac{\hat{\mu}_q^2}{\hat{\sigma}_q^2} - \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}, \]

(75)

i.e. the difference of squared Sharpe-ratios between the ex post tangency portfolio \( q \) and the market portfolio \( m \).

Exercise: Prove (75).
Using (75), test statistic $F_1$-statistic (69), which due to (73) is equivalent to $F$ in (62), has a useful economic interpretation‡

\[
F_1 = \frac{T - N - 1}{N} \left( \frac{\hat{\mu}_q^2 - \hat{\mu}_m^2}{\hat{\sigma}_q^2 - \hat{\sigma}_m^2} \right) \left( 1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right),
\]

where the portfolio denoted by $q$ represents the ex post tangency portfolio (the maximum squared Sharpe ratio) as constructed in (35) from the $N$ included assets plus the market portfolio (implying $\hat{\mu}_q^2/\hat{\sigma}_q^2 \geq \hat{\mu}_m^2/\hat{\sigma}_m^2$).

‡Details are given in Gibbons, Ross and Shanken (1989). *Econometrica*, 57, 1121–1152.
Thus when (ex post) the market portfolio will be the tangency portfolio $F_1 = 0$. On the other hand as $\hat{\mu}_m^2/\hat{\sigma}_m^2$ decreases (i.e., the market portfolio becomes increasingly inefficient) $LR_1$ increase.
Example 2.1: For illustrative purposes consider the stock return data of Example 1.1.

Unrestricted ML-estimates (EViews):

Estimation Method: Full Information Maximum Likelihood (Marquardt)
Sample: 8/22/1988 1/16/2006
Included observations: 908
Total system (balanced) observations 6356
Convergence achieved after 15 iterations

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha_mot</td>
<td>0.000344</td>
<td>0.186628</td>
<td>0.8520</td>
</tr>
<tr>
<td>beta_mot</td>
<td>1.351485</td>
<td>20.97477</td>
<td>0.0000</td>
</tr>
<tr>
<td>alpha_ms</td>
<td>0.003008</td>
<td>2.214048</td>
<td>0.0268</td>
</tr>
<tr>
<td>beta_ms</td>
<td>1.171873</td>
<td>15.88311</td>
<td>0.0000</td>
</tr>
<tr>
<td>alpha_app</td>
<td>0.000492</td>
<td>0.209641</td>
<td>0.8339</td>
</tr>
<tr>
<td>beta_app</td>
<td>1.105038</td>
<td>10.00059</td>
<td>0.0000</td>
</tr>
<tr>
<td>alpha_dell</td>
<td>0.004314</td>
<td>1.763438</td>
<td>0.0778</td>
</tr>
<tr>
<td>beta_dell</td>
<td>1.382504</td>
<td>11.12721</td>
<td>0.0000</td>
</tr>
<tr>
<td>alpha_ibm</td>
<td>-0.000158</td>
<td>-0.126520</td>
<td>0.8993</td>
</tr>
<tr>
<td>beta_ibm</td>
<td>0.948523</td>
<td>17.10562</td>
<td>0.0000</td>
</tr>
<tr>
<td>alpha_ford</td>
<td>-0.001105</td>
<td>-0.857391</td>
<td>0.3912</td>
</tr>
<tr>
<td>beta_ford</td>
<td>1.055709</td>
<td>23.42279</td>
<td>0.0000</td>
</tr>
<tr>
<td>alpha_gm</td>
<td>-0.001417</td>
<td>-1.116840</td>
<td>0.2641</td>
</tr>
<tr>
<td>beta_gm</td>
<td>1.050323</td>
<td>18.16294</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Log Likelihood 10806.38
Determinant residual covariance 1.08E-19
The restricted ML-estimates (EViews):

System: ML_CAPM_R
Estimation Method: Full Information Maximum Likelihood (Marquardt)
Sample: 8/22/1988 1/16/2006
Included observations: 908
Total system (balanced) observations 6356
Convergence achieved after 7 iterations

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta_mot</td>
<td>1.352179</td>
<td>0.063345</td>
<td>21.34643</td>
</tr>
<tr>
<td>beta_ms</td>
<td>1.178116</td>
<td>0.073836</td>
<td>15.95580</td>
</tr>
<tr>
<td>beta_app</td>
<td>1.106053</td>
<td>0.110080</td>
<td>10.04774</td>
</tr>
<tr>
<td>beta_dell</td>
<td>1.391454</td>
<td>0.123835</td>
<td>11.23635</td>
</tr>
<tr>
<td>beta.ibm</td>
<td>0.948189</td>
<td>0.054746</td>
<td>17.31973</td>
</tr>
<tr>
<td>beta.ford</td>
<td>1.053435</td>
<td>0.044900</td>
<td>23.46184</td>
</tr>
<tr>
<td>beta.gm</td>
<td>1.047396</td>
<td>0.056950</td>
<td>18.39139</td>
</tr>
</tbody>
</table>

Log Likelihood 10802.01
Determinant residual covariance 1.10E-19

The LR-test (63) is

\[ LR = -2(\ell^* - \ell) = -2 \times (10802.01 - 10806.38) = 8.74. \]

\[ df = 7, \quad p = 0.272. \] Thus the null hypothesis of zero abnormal returns is accepted.

The \( F_1 \)-statistic in (69) is

\[ F_1 = \frac{T - N - 1}{N} \left( \exp \left[ LR_1/T \right] - 1 \right) \approx 1.24 \]

\[ df_1 = 7 \quad (= N) \] and \( df_2 = 900 \quad (= T - N - 1), \quad p = 0.276, \]

about the same as he \( p \)-value of the LR-test.
In the absence of a riskfree return the CAPM (Black version) can be written

\[ E[R_t] = \gamma_i + \beta E[R_{mt} - \gamma] \]

\[ = (\iota - \beta)\gamma + \beta E[R_{mt}] \]

(77)

The focus is in the restriction

\[ \alpha = (\iota - \beta)\gamma. \]

(78)

Note that \( \iota \) and \( \beta \) are \( N \)-vectors and \( \gamma \) is a scalar.
The formulas of the estimators of the parameters are the same as before, except that we use the raw returns instead of the excess returns, such that

\begin{equation}
\hat{\alpha} = \hat{\mu} - \hat{\beta}\hat{\mu}_m,
\end{equation}

\begin{equation}
\hat{\beta} = \frac{\sum_{t=1}^{T}(r_t - \hat{\mu})(r_{mt} - \mu_m)}{\sum_{t=1}^{T}(r_{mt} - \hat{\mu}_m)^2},
\end{equation}

\begin{equation}
\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T}(r_t - \hat{\alpha} - \hat{\beta}r_{mt})(r_t - \hat{\alpha} - \hat{\beta}r_{mt})',
\end{equation}

where

\begin{equation}
\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} r_t \quad \text{and} \quad \hat{\mu}_m = \frac{1}{T} \sum_{t=1}^{T} r_{mt}.
\end{equation}

Again as before

\begin{equation}
\hat{\alpha} \sim N\left(\alpha, \frac{1}{T} \left[1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}\right] \Sigma\right),
\end{equation}

\begin{equation}
\hat{\beta} \sim N\left(\beta, \frac{1}{T\hat{\sigma}_m^2} \Sigma\right),
\end{equation}
\( T\hat{\Sigma} \sim W(T - 2, \Sigma), \)

and

\[
\text{Cov[}\hat{\alpha}, \hat{\beta}] = -\frac{1}{T} \left( \frac{\hat{\mu}_m}{\hat{\sigma}_m^2} \right) \Sigma.
\]

Imposing restriction (78) the (restricted) maximum likelihood estimators of the parameters are

\[
\hat{\gamma}^* = \frac{(\iota - \hat{\beta}^*)'\hat{\Sigma}^*^{-1}(\hat{\mu} - \hat{\beta}^*\hat{\mu}_m)}{(\iota - \hat{\beta}^*)'\hat{\Sigma}^*^{-1}(\iota - \hat{\beta}^*)},
\]

\[
\hat{\beta}^* = \frac{\sum_{t=1}^T (r_t - \hat{\gamma}^*\iota)(r_{mt} - \hat{\gamma}^*)}{\sum_{t=1}^T (R_{mt} - \hat{\gamma}^*)^2},
\]

and

\[
\hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^T e_t^* e_t'^*,
\]

where

\[
e_t^* = r_t - \hat{\gamma}^* (\iota - \hat{\beta}^*) - \hat{\beta}^* r_{mt}
\]

is the time \( t \) residual vector of the restricted model.
Because the estimator formulas depend on the maximization must be solved iteratively using a numerical method. A good start off point for the iteration are the unrestricted estimates.

The null hypothesis is

\( H_0 : \alpha = (\iota - \beta)\gamma \) \hspace{1cm} (91)

with the opposite hypothesis

\( H_1 : \alpha \neq (\iota - \beta)\gamma. \) \hspace{1cm} (92)

A likelihood ratio test statistic is

\( LR = T \left[ \log |\hat{\Sigma}^*| - \log |\hat{\Sigma}| \right] \sim \chi^2_{N-1}, \) \hspace{1cm} (93)

and an improved version

\( LR^* = \left( T - \frac{N}{2} - 2 \right) \left[ \log |\hat{\Sigma}^*| - \log |\hat{\Sigma}| \right] \sim \chi^2_{N-1}. \) \hspace{1cm} (94)

In both cases the distribution property holds asymptotically when the null hypothesis is true.
Remark 2.7: In the LR statistic the degrees of freedom is the difference of the number of estimated parameters under the alternative the null hypothesis (or equivalently the number of restriction in the null hypothesis).

Example 2.2: (Continued). In EViews the Black-equations can be simply defined as $r_i = (1-c(i)) \cdot c(1) + c(i) \cdot r_m$, where the estimated parameters are $\gamma = c(1)$ and $\beta_i = c(i)$.

In SAS, using PROC MODEL the parameter names are defined in the parameter list and the equations can again be written the same manner as in EViews.

SAS results for the unrestricted and restricted models are given below.
SAS-code using PROC MODEL:

```sas
proc model data = stockdata;
   Title "Unrestricted Black";
   endogenous r_mot r_ms r_app r_dell r_ibm r_ford r_gm;
   exogenous r_m;
   parameters a_mot b_mot a_ms b_ms a_app b_app a_dell b_dell
         a_ibm b_ibm a_ford b_ford a_gm b_gm;
   r_mot = a_mot + b_mot*r_m;
   r_ms = a_ms + b_ms*r_m;
   r_app = a_app + b_app*r_m;
   r_dell = a_dell + b_dell*r_m;
   r_ibm = a_ibm + b_ibm*r_m;
   r_ford = a_ford + b_ford*r_m;
   r_gm = a_gm + b_gm*r_m;

   fit r_mot r_ms r_app r_dell r_ibm r_ford r_gm/fiml;
run;
```

SAS estimates: (Note returns are here defined as per cents implying that the intercept terms are $100 \times$ the non-percent intercepts)

```plaintext
The MODEL Procedure
Nonlinear FIML Summary of Residual Errors

<table>
<thead>
<tr>
<th>Equation</th>
<th>DF</th>
<th>Model DF</th>
<th>SSE</th>
<th>MSE</th>
<th>Root MSE</th>
<th>R-Square</th>
<th>R-Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>r_mot</td>
<td>2</td>
<td>906</td>
<td>23587.4</td>
<td>5.1024</td>
<td>0.2326</td>
<td>0.2317</td>
<td></td>
</tr>
<tr>
<td>r_ms</td>
<td>2</td>
<td>906</td>
<td>14401.2</td>
<td>3.9869</td>
<td>0.2730</td>
<td>0.2722</td>
<td></td>
</tr>
<tr>
<td>r_app</td>
<td>2</td>
<td>906</td>
<td>37940.1</td>
<td>41.8765</td>
<td>6.4712</td>
<td>0.1116</td>
<td>0.1107</td>
</tr>
<tr>
<td>r_dell</td>
<td>2</td>
<td>906</td>
<td>44057.0</td>
<td>48.6280</td>
<td>6.9734</td>
<td>0.1456</td>
<td>0.1447</td>
</tr>
<tr>
<td>r_ibm</td>
<td>2</td>
<td>906</td>
<td>11306.3</td>
<td>12.4794</td>
<td>3.5326</td>
<td>0.2382</td>
<td>0.2373</td>
</tr>
<tr>
<td>r_ford</td>
<td>2</td>
<td>906</td>
<td>12563.3</td>
<td>13.8668</td>
<td>3.7238</td>
<td>0.2580</td>
<td>0.2572</td>
</tr>
<tr>
<td>r_gm</td>
<td>2</td>
<td>906</td>
<td>12130.3</td>
<td>13.3889</td>
<td>3.6591</td>
<td>0.2630</td>
<td>0.2621</td>
</tr>
</tbody>
</table>

Nonlinear FIML Parameter Estimates

| Parameter | Estimate | Approx Std Err | t Value | Pr > |t| |
|-----------|----------|----------------|---------|------|---|
| a_mot     | 0.004484 | 0.1747         | 0.03    | 0.9795 |
| b_mot     | 1.349361 | 0.0570         | 23.68   | <.0001 |
| a_ms      | 0.28563  | 0.1334         | 2.14    | 0.0325 |
| b_ms      | 1.173633 | 0.0704         | 16.67   | <.0001 |
| a_app     | 0.040676 | 0.2155         | 0.19    | 0.8503 |
| b_app     | 1.102035 | 0.1061         | 10.38   | <.0001 |
| a_dell    | 0.398368 | 0.2317         | 1.72    | 0.0858 |
| b_dell    | 1.382939 | 0.1161         | 11.91   | <.0001 |
| a_ibm     | -0.0114  | 0.1198         | -0.10   | 0.9242 |
| b_ibm     | 0.948882 | 0.0504         | 18.84   | <.0001 |
| a_ford    | -0.11519 | 0.1247         | -0.92   | 0.3557 |
| b_ford    | 1.054899 | 0.0419         | 25.20   | <.0001 |
| a_gm      | -0.14598 | 0.1226         | -1.19   | 0.2341 |
| b_gm      | 1.04993  | 0.0535         | 19.64   | <.0001 |

Number of Observations 908  Log Likelihood -18464

43
```
Imposing the Black restriction to the estimation equations

proc model data = stockdata;
Title "Restricted Black";
endogenous r_mot r_ms r_app r_dell r_ibm r_ford r_gm;
exogenous r_m;
parameters gamma b_mot b_ms b_app b_dell
     b_ibm b_ford b_gm;
   r_mot = (1-b_mot)*gamma + b_mot*r_m;
   r_ms = (1-b_ms)*gamma + b_ms*r_m;
   r_app = (1-b_app)*gamma + b_app*r_m;
   r_dell = (1-b_dell)*gamma + b_dell*r_m;
   r_ibm = (1-b_ibm)*gamma + b_ibm*r_m;
   r_ford = (1-b_ford)*gamma + b_ford*r_m;
   r_gm = (1-b_gm)*gamma + b_gm*r_m;

   fit r_mot r_ms r_app r_dell r_ibm r_ford r_gm/fiml;
run;

gives estimation results:
The MODEL Procedure
Nonlinear FIML Summary of Residual Errors

<table>
<thead>
<tr>
<th>Equation</th>
<th>DF</th>
<th>DF Adj</th>
<th>SSE</th>
<th>MSE</th>
<th>Root MSE</th>
<th>R-Square</th>
<th>R-Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>r_mot</td>
<td>1.143</td>
<td>906.9</td>
<td>23622.4</td>
<td>26.0486</td>
<td>5.1038</td>
<td>0.2314</td>
<td>0.2313</td>
</tr>
<tr>
<td>r_ms</td>
<td>1.143</td>
<td>906.9</td>
<td>14426.7</td>
<td>15.9084</td>
<td>3.9885</td>
<td>0.2717</td>
<td>0.2716</td>
</tr>
<tr>
<td>r_app</td>
<td>1.143</td>
<td>906.9</td>
<td>37940.5</td>
<td>41.8373</td>
<td>6.4682</td>
<td>0.1116</td>
<td>0.1115</td>
</tr>
<tr>
<td>r_dell</td>
<td>1.143</td>
<td>906.9</td>
<td>44078.2</td>
<td>48.6054</td>
<td>6.9718</td>
<td>0.1452</td>
<td>0.1451</td>
</tr>
<tr>
<td>r.ibm</td>
<td>1.143</td>
<td>906.9</td>
<td>11306.6</td>
<td>12.4679</td>
<td>3.5310</td>
<td>0.2382</td>
<td>0.2380</td>
</tr>
<tr>
<td>r_ford</td>
<td>1.143</td>
<td>906.9</td>
<td>12580.9</td>
<td>13.8731</td>
<td>3.7247</td>
<td>0.2570</td>
<td>0.2569</td>
</tr>
<tr>
<td>r_gm</td>
<td>1.143</td>
<td>906.9</td>
<td>12155.1</td>
<td>13.4035</td>
<td>3.6611</td>
<td>0.2614</td>
<td>0.2613</td>
</tr>
</tbody>
</table>

Nonlinear FIML Parameter Estimates

| Parameter | Estimate | Approx Std Err | t Value | Approx Pr > |t| |
|-----------|----------|----------------|---------|-------------|-------|
| gamma     | -0.61366 | 0.3972         | -1.54   | 0.1227      |       |
| b_mot     | 1.316296 | 0.0545         | 24.17   | <.0001      |       |
| b_ms      | 1.201841 | 0.0657         | 18.30   | <.0001      |       |
| b_app     | 1.098579 | 0.1029         | 10.68   | <.0001      |       |
| b_dell    | 1.408673 | 0.1121         | 12.56   | <.0001      |       |
| b.ibm     | 0.952027 | 0.0463         | 20.56   | <.0001      |       |
| b_ford    | 1.031447 | 0.0409         | 25.19   | <.0001      |       |
| b_gm      | 1.022109 | 0.0495         | 20.67   | <.0001      |       |

Number of observations 908  Log Likelihood -18467
Almost all $\alpha$ estimates are statistically zero in the unrestricted case. Thus imposing the Black restriction yields non-significant estimate for the $\gamma$-parameter. As a result there is noting much to test null hypothesis $\alpha = (1 - \beta) \gamma$. Calculating it for illustrative purposes yields

$$LR = -1867 - (-1864) = 3,$$

$df = 6$, yielding $p$-value $= 0.809$. Thus the restrictions implied by the Black model would be accepted.
GMM-estimation

Stylized facts in return time series are among others that they are non-normal, volatility is clustering, and they are slightly negatively autocorrelated. Thus the vanilla ML and OLS estimation may be unreliable.

GMM is supposed to be more robust, because the variance-covariance matrix of the residuals can be estimated by allowing autocorrelation and cross-sectional (and to some extend time series) heteroscedasticity (ML-estimation and GLS account for the cross-sectional correlations and heteroscedasticity in the multivariate setup).
The moment conditions in the unrestricted case are simply

\[(95) \quad u_t(b) = (r_t - \alpha - \beta r_{mt}) \otimes \begin{pmatrix} 1 \\ r_{mt} \end{pmatrix}\]

with \(E[u_t(b)] = 0\), where \(b = (\alpha', \beta')'\) is the \((2N \times 1)\) parameter vector \((N = \text{number of shares})\).

Thus the moment conditions for return \(i\) in (95) are

\[(96) \quad E \begin{bmatrix} r_{it} - \alpha_i - \beta_i r_{mt} \\ (r_{it} - \alpha_i - \beta_i r_{mt}) r_{mt} \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad i = 1, \ldots, N.\]
The value added of GMM over OLS, GLS, and (normal distribution) ML lies in the selection of the weighting matrix of the object function

\[ g_T(b)' W g_T(b), \]

where

\[ g_T(b) = \frac{1}{T} \sum_{t=1}^{T} u_t(b), \]

the sample average of \( u_t(b) \)s defined by (95).

\( W = I \) yields OLS estimators, and assumption of zero autocorrelation in the residuals yields GLS or (normal theory) ML estimator.
Use of an estimate of $S^{-1}$ as $W$, where

\[(99) \quad S = \sum_{j=-\infty}^{\infty} E[u_t(b)u'_{t+j}(b)],\]

[see (1.15)–(1.19)] accounts for both autocorrelations and heteroscedasticity across cross-sectional returns.
Example 2.4 (Example 2.3 continued) GMM estimates for the unrestricted Black CAPM are (produced by SAS proc model). The unrestricted model is exactly identified, implying that the estimates are the same as OLS or ML. However, the standard errors should be more robust.

```
proc model data = stockdata;
  Title "Unrestricted Black GMM";
  endogenous r_mot r_ms r_app r_dell r_ibm r_ford r_gm;
  exogenous r_m;
  parameters a_mot b_mot a_ms b_ms a_app b_app a_dell b_dell
      a_ibm b_ibm a_ford b_ford a_gm b_gm;
  r_mot = a_mot + b_mot*r_m;
  r_ms = a_ms + b_ms*r_m;
  r_app = a_app + b_app*r_m;
  r_dell = a_dell + b_dell*r_m;
  r_ibm = a_ibm + b_ibm*r_m;
  r_ford = a_ford + b_ford*r_m;
  r_gm = a_gm + b_gm*r_m;

  fit r_mot r_ms r_app r_dell r_ibm r_ford r_gm/itgmm;
  instruments _exog_;
run;
```
Nonlinear ITGMM Parameter Estimates

| Parameter | Estimate | Approx Std Err | t Value | Approx Pr > |t| |
|-----------|----------|---------------|---------|-------------|---------|
| a_mot     | 0.004484 | 0.1495        | 0.03    | 0.9761      |
| b_mot     | 1.349361 | 0.1308        | 10.32   | <.0001      |
| a_ms      | 0.28563  | 0.1307        | 2.18    | 0.0292      |
| b_ms      | 1.173633 | 0.0595        | 19.74   | <.0001      |
| a_app     | 0.040676 | 0.2275        | 0.18    | 0.8581      |
| b_app     | 1.102035 | 0.1016        | 10.84   | <.0001      |
| a_dell    | 0.398368 | 0.2177        | 1.83    | 0.0676      |
| b_dell    | 1.382939 | 0.1096        | 12.62   | <.0001      |
| a_ibm     | -0.0114  | 0.1159        | -0.10   | 0.9217      |
| b_ibm     | 0.948882 | 0.0709        | 13.38   | <.0001      |
| a_ford    | -0.11519 | 0.1217        | -0.95   | 0.3440      |
| b_ford    | 1.054899 | 0.0917        | 11.50   | <.0001      |
| a_gm      | -0.14598 | 0.1221        | -1.20   | 0.2321      |
| b_gm      | 1.04993  | 0.0711        | 14.77   | <.0001      |

Number of Observations Statistics for System

Used 908 Objective 1.274E-30
Missing 1 Objective*N 1.157E-27
Restricted GMM estimation (iterative GMM)

```sas
proc model data = stockdata;
Title "Restricted Black: GMM";
endogenous r_mot r_ms r_app r_dell r_ibm r_ford r_gm;
exogenous r_m;
parameters gamma b_mot b_ms b_app b_dell
        b_ibm b_ford b_gm;
r_mot = (1-b_mot)*gamma + b_mot*r_m;
r_ms = (1-b_ms)*gamma + b_ms*r_m;
r_app = (1-b_app)*gamma + b_app*r_m;
r_dell = (1-b_dell)*gamma + b_dell*r_m;
r_ibm = (1-b_ibm)*gamma + b_ibm*r_m;
r_ford = (1-b_ford)*gamma + b_ford*r_m;
r_gm = (1-b_gm)*gamma + b_gm*r_m;

fit r_mot r_ms r_app r_dell r_ibm r_ford r_gm/itgmm;
instruments _exog_;
run;
```
### Nonlinear ITGMM Parameter Estimates

| Parameter | Estimate | Std Err | t Value | Pr > |t| |
|-----------|----------|---------|---------|------|---|
| gamma     | -0.40284 | 0.3375  | -1.19   | 0.2329 |
| b_mot     | 1.369831 | 0.1169  | 11.72   | <.0001 |
| b_ms      | 1.194299 | 0.0574  | 20.82   | <.0001 |
| b_app     | 1.099028 | 0.0939  | 11.70   | <.0001 |
| b_dell    | 1.391683 | 0.1061  | 13.11   | <.0001 |
| b_ibm     | 0.949755 | 0.0672  | 14.13   | <.0001 |
| b_ford    | 1.033148 | 0.0785  | 13.16   | <.0001 |
| b_gm      | 1.033162 | 0.0644  | 16.03   | <.0001 |

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>Statistics for System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used 908</td>
<td>Objective 0.006948</td>
</tr>
<tr>
<td>Missing 1</td>
<td>Objective*N 6.3083</td>
</tr>
</tbody>
</table>

The estimation results are about the same as in the ML estimation. That is again most of the $\alpha$ estimates in the unrestricted model are non-significant and the $\hat{\gamma}$-estimate in the restricted case is non-significant. There are 6 overidentification conditions (14 equations minus 8 estimated coefficients). The test statistic for the over-identification conditions is $T$ times the objective function, such that

$$\chi^2 = 6.3083.$$  

With $df = 6$ the $p$-value is 0.390. Thus, formally, again there is no statistical evidence against the Black CAPM restrictions.
Cross-Sectional Regression Tests

In addition to mean-variance efficiency of the market portfolio, CAPM implies a linear relation between expected returns and market betas which completely explain the cross section of expected returns.

Assuming betas are known then the excess return vector of the $N$ assets obeys the regression equation

\[
  z_t = \gamma_0 t + \gamma_m t \beta + \eta_t,
\]

where $z_t$ is the excess return vector, $\gamma_0 t$ is the time $t$ intercept term which should be zero if CAPM holds, $\gamma_m t$ is the market risk premium. If CAPM holds $\gamma_m t > 0$. 

(100)
Fama and MacBeth (1973) test

In practice the $\beta$ is unknown. So it must be estimated. The Fama-MacBeth procedure involves three steps. First given $T$ period time series, estimate the betas with OLS. In the second step use the estimated betas and estimate at each time period $\gamma_0t$ and $\gamma_{mt}$. In the third steps analyze the estimated $\gamma_0t$ and $\gamma_{mt}$ series.

The null hypotheses to be tested are

(101) \[ H_0 : \gamma_j = 0, \]

where $\gamma_j = E[\tilde{\gamma}_{jt}]$ with $j = 0$ or $j = m$.

The alternative hypothesis for $\gamma_0$ is $H_1 : \gamma_0 \neq 0$ and for $\gamma_m$ is $H_1 : \gamma_m > 0$. These can be tested using the $t$-test

(102) \[ t_j = \frac{\hat{\gamma}_j}{\hat{\sigma}\tilde{\gamma}_j}, \]
where

\begin{equation}
\hat{\gamma}_j = \frac{1}{T} \sum_{t=1}^{T} \hat{\gamma}_{jt},
\end{equation}

and

\begin{equation}
\hat{\sigma}_{\hat{\gamma}_j}^2 = \frac{1}{T(T-1)} \sum_{t=1}^{T} (\hat{\gamma}_{jt} - \hat{\gamma}_j)^2
\end{equation}

is the square of the standard error of \( \hat{\gamma}_j \).

A problem with this approach is that estimating \( \beta \)s in the first round leads in the second round estimation into the errors-in-variables problem. Ramaswamy (1979) and Shanken (1992) have suggested adjustments to alleviate this problem.

The cross-sectional testing can be extended to test whether there are other factors (standard error, firms size, etc.) that affect the returns. If so the they are against what CAPM implies.
2.3 Arbitrage Pricing Model, APM

Empirical evidence indicates that the CAPM beta does not completely explain the cross section of expected asset returns. This suggests that additional factors may be required.

Ross (1976)† introduced the Arbitrage Pricing Theory (APT) as an alternative to the CAPM.

The basic assumption is that there are a number of, say $K$, common risk factors generating the returns so that

$$R_i = a_i + b_i'f + \epsilon_i,$$

with

$$\mathbb{E}[\epsilon_i|f] = 0$$

$$\mathbb{E}[\epsilon_i^2] = \sigma_i^2 \leq \sigma^2 < \infty,$$

inside and

$$\mathbb{E}[\epsilon_i\epsilon_j] = 0, \text{ whenever } i \neq j,$$

where $i = 1, \ldots, N$, the number of assets, $a_i$ is the intercept term of the factor model, $b_i$ is a $K \times 1$ coefficient vector of factor sensitivities (loadings) for asset $i$, $f$ is a $K \times 1$ vector of common factors, and $\epsilon_i$ is the disturbance term.
Without loss of generality we may assume that the common factors have zero mean, i.e., $\mathbb{E}[f] = 0$, which implies that $a_i = \mathbb{E}[r_i] = \mu_i$ are the mean returns.

In the matrix form the return generating model is

(108) \[ r = \mu + Bf + \epsilon, \]

(109) \[ \mathbb{E}[\epsilon | f] = 0, \]

and

(110) \[ \mathbb{E}[\epsilon \epsilon' | f] = \Sigma, \]
where \( r = (r_1, r_2, \ldots, r_N)' \), \( \mu = (\mu_1, \ldots, \mu_N)' \), 
\( B = (b_1, \ldots, b_N)' \) is an \( N \times K \) factor loading matrix, 
\( \epsilon = (\epsilon_1, \ldots, \epsilon_N)' \), and \( \Sigma \) is a \( N \times N \) matrix (assumed diagonal in the original Ross model). Furthermore, it is assumed that \( K \ll N \).
The derivation of the APM relies on the no arbitrage assumption.

Let \( w = (w_1, \ldots, w_N)' \) be an arbitrage strategy. Then

\[
\mathbf{w}'_i = \sum_{i=1}^{N} w_i = 0,
\]

(111) and the implied portfolio should be riskfree (or more precisely its starting value should be zero, non-negative in the meantime with probability one, and have a strictly positive expected end value).

The return of the portfolio is

\[
\mathbf{r}_p = \mathbf{w}' \mathbf{r} = \mathbf{w}' \mathbf{a} + \mathbf{w}' \mathbf{Bf} + \mathbf{w}' \mathbf{\epsilon}.
\]

(112)
In order to make the portfolio riskfree, the market risk $\mathbf{w}'\mathbf{Bf}$ and unsystematic risk $\mathbf{w}'\mathbf{\epsilon}$ must be eliminated. The unsystematic risk can be eliminated by selecting $N$ large.

Now $\text{Var}(\mathbf{w}'\mathbf{\epsilon}) = \mathbf{w}'\mathbf{\Sigma w} = \sum_{i=1}^{N} w_i^2 \sigma_i^2$. The weights $w_i$ are of order $1/N$, so $\sum_{i=1}^{N} w_i^2 \sigma_i^2 \to 0$ as $N \to \infty$. 
In order to eliminate the market risk the weights must be selected such that

$$(113) \quad w'B = 0'. \quad \tag{113}$$

In the language of linear algebra the columns of the expanded matrix $\tilde{B} = (\iota, B)$ spans a $K + 1$-dimensional linear subspace, call it $V$, in $\mathbb{R}^N$. Because $N > K + 1$ all vectors lying in the orthogonal complement, $V^\perp$, of $V$ are valid candidates for $w$ to satisfy conditions (111) and (113).†

†More precisely, $V = \{y \in \mathbb{R}^N : y = \tilde{B}x, \ x \in \mathbb{R}^{K+1}\}$ and $V^\perp = \{z \in \mathbb{R}^N : z'y = 0 \ \forall \ y \in V\}$. Note further that $\mathbb{R}^N = V \cup V^\perp$, actually $\mathbb{R}^N = V \oplus V^\perp$, the direct sum of $V$ and $V^\perp$. 
Given an arbitrage strategy $\mathbf{w}$ that satisfies (113), we get with large $N$ approximately (because $\mathbf{w}'\epsilon \approx 0$)

$$\left(114\right) \quad r_p \approx \sum_{i=1}^{N} w_i \mu_i = \mathbf{w}'\mu,$$

which is a riskfree return.

The absence of arbitrage implies that any arbitrage portfolio must have a zero return. In other words

$$\left(115\right) \quad r_p \approx \mathbf{w}'\mu = 0,$$

which implies that the expected return vector $\mu$ is orthogonal to $\mathbf{w}$. 
But then it is a vector in the linear space $V$ and hence of the form

$$\mu = \tilde{B}\lambda,$$

where $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_K)' \in \mathbb{R}^{K+1}$.

In other words, the expected returns of the single assets are of the form

$$\mu_i = \lambda_0 + b_{i1}\lambda_1 + \ldots + b_{iK}\lambda_K.$$
If there is a riskfree asset with return $r_f = \mu_0$, it has by definition, zero exposure on the common market risk factors. That is $b_{0j} = 0, j = 1, \ldots, K$. Then from (117) with $b_{0j} = 0$, we get

\begin{equation}
(118) \quad r_f = \lambda_0,
\end{equation}

and we can rewrite (117) as

\begin{equation}
(119) \quad \mu_i = r_f + \lambda_1 b_{i1} + \cdots + \lambda_K b_{iK},
\end{equation}

This is the APT equilibrium model for the expected asset returns.

In the matrix form the APM for the expected equilibrium returns is

\begin{equation}
(120) \quad \mathbb{E}[r_t] = \mu = \nu \lambda_0 + B\lambda,
\end{equation}

where $\lambda_0 = r_f$ is the riskfree return if it exists, and $\lambda = (\lambda_1, \ldots, \lambda_K)'$. 

67
Remark 2.8: Because $b_{ij}$ is the sensitivity of the return on the $j$th common risk factors, it is natural to interpret that $\lambda_j$ represents the risk premium (the price of risk) for factor $j$ in the equilibrium.

Remark 2.9: Generally, if there is no riskfree return, $\lambda_0$ can be interpreted as the zero-beta return.

Remark 2.10: If the APT hold the discount factor in $1 = \mathbb{E}[m r_i]$ is of the form

$$m = c_0 + c_1 f_1 + \cdots + c_K f_K.$$  \hfill (121)

Exercise 2.4: Prove (119) $\iff$ (121). (Hint. See Cochrane Ch 6.)
Estimation and Testing of APT

Assumption: Returns are normally and temporally independently distributed.

APT does neither specify the factors nor the number of factors. We consider four versions:

Factors are

(1) portfolios of traded assets and a riskfree asset exists;
(2) portfolios of traded assets and a riskfree asset does not exist;
(3) not portfolios of traded assets;
(4) portfolios of traded assets and the factor portfolios span the mean-variance frontier of risky assets.
The derivation of the test statistics is analogous to the CAPM case. Relying on normality the LR test statistic is of the form

\[
LR = - \left( T - \frac{N}{2} - K - 1 \right) \left( \log |\hat{\Sigma}| - \log |\hat{\Sigma}^*| \right),
\]

where $|\hat{\Sigma}|$ and $|\hat{\Sigma}^*|$ are the unconstrained and constrained ML-estimators, respectively.

Again, as before, the asymptotic null distribution of $LR$ is chi-square with degrees of freedom equal to the number of restrictions imposed by the null hypothesis.
(1) *Portfolios as Factors with a Riskfree Asset*

Denote the unconstrained form of the factor model (108) in this case as

\[ z_t = a + Bz_{Kt} + \epsilon_t \]  

with

\[ \mathbb{E}[\epsilon_t] = 0, \]  
\[ \mathbb{E}[\epsilon_t \epsilon_t'] = \Sigma, \]  
\[ \mathbb{E}[Z_{Kt}] = \mu_K, \]  
\[ \mathbb{E}[z_{Kt} - \mu_K)(z_{Kt} - \mu_K)'] = \Omega_K, \]  
and

\[ \text{Cov}[Z_{Kt}, \epsilon_t] = 0, \]
where $\mathbf{B}$ is the $N \times K$ matrix of factor sensitivities, $\mathbf{z}_{Kt}$ is the $K \times 1$ vector of factor portfolio excess returns, and $\mathbf{a}$ and $\epsilon_t$ are $N \times 1$ vectors of intercepts and error terms, respectively.

The APM implies that $\mathbf{a} = 0$. In order to test this with the LR-test we need to estimate the unconstrained and constrained model.
Model (123) is a *seemingly unrelated regression* (SUR) case, but because each regression equation has the same explanatory variables the ML-estimators are just the OLS estimators

\[
\hat{a} = \hat{\mu} - \hat{B}\hat{\mu}_K,
\]

\(\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} z_t,\)

\(\hat{\mu}_K = \frac{1}{T} \sum_{t=1}^{T} z_{Kt}.\)
The constrained estimators are

\[ \hat{B}^* = \left[ \sum_{t=1}^{T} z_t z'_K \right] \left[ \sum_{t=1}^{T} z_K t z'_K \right]^{-1} \tag{134} \]

and

\[ \hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^{T} (z_t - \hat{B}^* z_K)(z_t - \hat{B}^* z_K)' \tag{135} \]

The exact multivariate \( F \)-test becomes then

\[ F = \frac{T - N - K}{N} \left( 1 + \hat{\mu}' K \hat{\Omega}^{-1} K \hat{\mu} \right)^{-1} \hat{a}' \hat{\Sigma}^{-1} \hat{a}, \tag{136} \]

where

\[ \hat{\Omega}_K = \frac{1}{T} \sum_{t=1}^{T} (z_K - \hat{\mu}_K)(z_K - \hat{\mu}_K)' \tag{137} \]

Under the null hypothesis \( F \sim F(N, T - N - K) \).
(2) Portfolios as Factors without a Riskfree Asset

Let \( R_{Kt} = (R_{1t}, \ldots, R_{Kt})' \) be portfolios that are factors of the APT model, and denote the related factor model (108) as

\[
(138) \quad r_t = a + B r_{Kt} + \epsilon_t.
\]

If there does not exist a riskfree asset, then as in the CAPM there exist a portfolio which is uncorrelated with the portfolios in \( r_{Kt} \). Let \( \gamma_0 \) denote the expected return of this portfolio.
In the APM $\lambda_0 = \gamma_0$ and the constrained factor model is

$$
\begin{align*}
  r_t &= \nu \gamma_0 + B(r_K^t - \nu \gamma_0) + \epsilon_t \\
  &= (\nu - B \nu) \gamma_0 + Br_K^t + \epsilon_t,
\end{align*}
$$

(139)

so that

$$
  a = (\nu - B \nu) \gamma_0.
$$

(140)
The constrained ML-estimators are

\[(141)\]
\[\hat{B}^* = \left[ \sum_{t=1}^{T} (r_t - \hat{\gamma}_0)(r_{Kt} - \hat{\gamma}_0)' \right] \left[ \sum_{t=1}^{T} (r_{Kt} - \hat{\gamma}_0)(r_{Kt} - \hat{\gamma}_0)' \right]^{-1},\]

\[(142)\]
\[\hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^{T} [r_t - \hat{\gamma}_0 - \hat{B}^*(r_{Kt} - \hat{\gamma}_0)] [r_t - \hat{\gamma}_0 - \hat{B}^*(r_{Kt} - \hat{\gamma}_0)]',\]

and

\[(143)\]
\[\hat{\gamma}_0 = \left[ (\hat{v} - \hat{B}^*\hat{\mu})'\hat{\Sigma}^{*-1}(\hat{v} - \hat{B}^*\hat{\mu}) \right]^{-1} \left[ (\hat{v} - \hat{B}^*\hat{\mu})'\hat{\Sigma}^{*-1}(\hat{v} - \hat{B}^*\hat{\mu}) \right].\]

Again a suitable iterative procedure must be implemented in the estimation.
The null hypothesis

(144) \( H_0 : a = (\iota - B\iota)\gamma_0 \)

can be tested again with the likelihood ratio test \( LR \) of the form (122), which is asymptotically chi-squared distributed with \( N - 1 \) degrees of freedom under the null hypothesis.
(3) Macroeconomic Variables as Factors

Let $f_K$ be macroeconomic factors of the APM, with $E[f_{Kt}] = \mu_{fK}$. Then (108) becomes

$$r_t = a + Bf_{Kt} + \epsilon_t$$

with parameter structure and unconstrained ML estimators similar as derived in the case of (123).
In order to formulate the constrained parameters consider the expected value of (145)

\begin{equation}
\mu = a + B \mu f_K.
\end{equation}

If the APM holds then this should be equal to (120). Equating these we get for \( a \)

\begin{equation}
a = \nu \lambda_0 + B (\lambda - \mu f_K).
\end{equation}

Denoting \( \gamma_0 = \lambda_0 \) and \( \gamma_1 = \lambda - \mu f_K \), an \( K \times 1 \) vector, the restricted model is

\begin{equation}
r_t = \nu \lambda_0 + B \gamma_1 + B f_K t + \epsilon_t.
\end{equation}
Again estimating the parameters, the APM implied null hypothesis

\[ H_0 : a = \nu \lambda_0 + B \gamma_1 \]  

(149)

can be tested with the likelihood ratio test of the form (122) which in this case is under the null hypothesis asymptotically chi squared distributed with degrees of freedom \( N - K - 1 \).
(4) Factor Portfolios Spanning the Mean-Variance Frontier

Consider again the factor portfolio model (138). If the factor portfolios in $R_{Kt}$ span the mean-variance frontier then in (120) $\lambda_0$ is zero (c.f. the zero-beta case considered earlier).

This form of APM imposes the restriction

(150) $H_0: a = 0$ and $B_i = i$

on (138).
Example: Zero-beta version of CAPM (a two-factor model). Let $r_{mt}$ denote the market portfolio (a MVP) and $r_{0t}$ the associated zero-beta portfolio. Then $K = 2$, $B = (\beta_{0m}, \beta_m)$: $(N \times 2)$, and $r_{Kt} = (r_{0t}, r_{mt})'$: $(2 \times 1)$, so that

$$r_t = a + \beta_{0m} r_{mt} + \beta_m r_{mt} + \epsilon_t.$$  

As found earlier $a = 0$, and $\beta_{0m} + \beta_m = \iota$.

Again estimating the constrained and unconstrained model, we can use the LR-statistic (122) to test the null hypothesis (150).

The degrees of freedom are $2N$. These come from $N$ restrictions in $a = 0$ and $N$ restrictions in $B$ to satisfy $B\iota = \iota$.  

83
Again, assuming that the multivariate normality holds, there exist an exact test

\[ F = \frac{T - N - K}{N} \left( \frac{\left| \hat{\Sigma}^* \right|}{\left| \hat{\Sigma} \right|} - 1 \right), \]

which is \( F[2N, 2(T - N - K)] \)-distributed under the null hypothesis.
Estimation of Risk Premia and Expected Returns

As given in (120) the expected return of the assets are under the APT

$$\mu = \lambda_0 \iota + B\lambda.$$  

In order to make the model operational one needs to estimate the riskfree or zero beta return $\lambda_0$, factor sensitivities $B$, and the factor risk premia $\lambda$.

The appropriate estimation procedure varies across the four cases, considered above.

The principle is that one uses the appropriate restricted estimators in each case to estimate the parameters of (120).
For example, in the excess return case of (123), $\mathbf{B}$ is estimated by (134), and

\[
\hat{\lambda} = \hat{\mu}_K = \frac{1}{T} \sum_{t=1}^{T} z_{Kt},
\]

which is a vector of the average excess return of the market factors.

A question of interest then is whether the factors are jointly priced. The null hypothesis is

\[
H_0 : \lambda = 0.
\]
An appropriate test procedure is the multivariate mean test (known as Hotelling $T^2$ test)

\begin{equation}
H = \frac{T - K}{TK} \hat{\lambda}' [\hat{\text{Var}}[\hat{\lambda}]]^{-1} \hat{\lambda}.
\end{equation}

where

\begin{equation}
\hat{\text{Var}}[\hat{\lambda}] = \frac{1}{T} \hat{\Omega}_K = \frac{1}{T^2} \sum_{t=1}^{T} (Z_{Kt} - \hat{\mu}_{Kt})(Z_{Kt} - \hat{\mu}_{Kt})'.
\end{equation}

Asymptotically $J_3 \sim F(K, T - K)$ under the null hypothesis (153).
Significance of individual factors, i.e., null hypotheses

(156) \[ H_0 : \lambda_j = 0 \]

can be tested with the \( t \)-test

(157) \[ t = \frac{\hat{\lambda}_j}{\sqrt{v_{jj}}} \]

which is asymptotically \( N(0, 1) \)-distributed under the null hypothesis, and where \( v_{jj} \) is the \( j \)th diagonal element of \( \hat{\text{Var}}[\hat{\lambda}] \), \( j = 1, \ldots, K \).
Remark 2.11: Test of individual factors is sensible only if the factors are theoretically specified. If they are empirically specified they do not have clear-cut economic interpretations.

Remark 2.12: Another way to estimate factor risk premia is to use a two-pass cross-sectional regression approach. In the first pass the factor sensitivities ($B$) are estimated and in the second pass the premia parameters of the regression

$$
Z_t = \lambda_{0t} + \hat{B}\lambda + \eta_t
$$

(158)

can be estimated time-period-by-time-period. Here $\hat{B}$ is identical to the unconstrained estimator of $B$. 
Selection of Factors

The factors of APM must be specified. This can be on \textit{statistical} or \textit{theoretical} basis.

Statistical Approaches

Linear factor model as given in (108), in general form is

\begin{equation}
\mathbf{r}_t = \mathbf{a} + \mathbf{B} \mathbf{f}_t + \epsilon_t
\end{equation}

(159)

with

\begin{equation}
E[\epsilon_t \epsilon_t'|\mathbf{f}_t] = \Sigma.
\end{equation}

(160)

In order to find the factors there are two primary statistical approaches \textit{factor analysis} and \textit{principal component analysis}.
Factor Analysis

Using statistical factor analysis the assumption is that there are $K$ common latent (not directly observable) common factors that affect the stock prices. Especially it is assumed that the common factors capture the cross-sectional covariances between the asset returns.

The factor model is of the form (159), which with the above additional assumption implies the following structure to the covariance matrix of the returns

$$\Omega = \mathbf{B}\Phi\mathbf{B}' + \mathbf{D},$$

where $\Phi = \text{Cov}(f_t)$ is a $K \times K$ covariance matrix of the common factors, and $\mathbf{D} = \text{Cov}(\epsilon)$ is an $N \times N$ diagonal matrix of the residuals (unique factors).
All the parameter matrices on the right hand side of the decomposition (161) (altogether \((NK + K(K + 1)/2 + N\) parameters) are unknown. Furthermore the decomposition is not unique, because (given that \(\Phi\) is positive definite) we can always write

\[
\Phi = GG' 
\]

so that redefining \(B\) as \(BG\)

\(\Omega = BB' + D\).

Even in this case \(B\) is not generally unique, because again defining \(C = BT\) where \(T\) is an arbitrary \(K \times K\) matrix such that \(TT' = I\), we get \(\Omega = CC' + D\). Matrix \(T\) is called a rotation matrix.

The APT does not give the number of common factors \(K\). Thus the first task is to determine the number of factors. Empirically this can be done with various statistical criteria using factor analysis packages.
A popular method is to select as many factors as there are eigenvectors larger than one computed from the return correlation matrix. Modern statistical packages provide also explicit statistical tests as well as various criterion functions for the purpose.
Principal Component

Principal component analysis is another tool for deriving common factors. This however is a more technical approach. Furthermore usually different components are obtained from correlation matrix than from covariance matrix. Nevertheless, there is no clear-cut research results which one, PCA or FA, should be a better choice in APT analysis.
Theoretical Approaches

One approach is to *macroeconomic* and *financial* market variables that are thought to capture systematic risk of the economy.

Chen, Roll and Ross (1986) *Journal of Business*, 31, 1067–1083, use five factors:
(1) long and short government bond yield spread (maturity premium), expected inflation,
(2) unexpected inflation,
(3) industrial production growth, yield spread between high- and low-grade bonds (default premium),
(4) aggregate consumption growth and
(5) oil prices.
"Empirical factors" (Fama-French factors)

The central prediction of the CAPM besides that the market portfolio is mean-variance efficient is that the beta explains completely cross-sectional variation in returns.

Fama and MacBeth (1973) rolling beta method.

Fama and French (1992) sorted portfolio method.

Fama and French (1993) factor mimicking portfolios as a tree-factor-model: market factor (market return minus risk-free return) HML (high minus low, book-to-market factor), SMB (small-minus-large, size factor)

A summary of the three factor model is given in Fama and French (1996).
Earlier findings suggest that firm size (ME, market value of equity) helps explain variation in cross-sectional returns (Banz 1981, JFE). Also book value of equity (BE) to ME seems to have a relation to returns.

Fama and French (1992) investigate these matters in detail.

Fama-French methodology:

**Size portfolios:**

(1) 10 decile portfolios according to the ME (updated July every year)
(2) Monthly returns for the portfolios for the next 12 months
Results:

(1) Strong negative (cor)relation between size and returns
(2) Strong positive relations between $\beta$ and returns $\Rightarrow$ support CAPM
(3) However, size and beta are strongly negatively related!

Situation?

\[
\begin{array}{c}
\text{size} \\
/ \quad / \\
\text{beta} \\
\text{return}
\end{array}
\]

Figure 2.1: Size-beta-return relation.

I.e., when size is controlled there is no relation between beta and return?
In order to control the "size effect", Fama and French subdivide each decile into 10 portfolios on the basis of pre-ranking $\beta$s for individual stocks every June updated every year (FF estimate the pre-ranking $\beta$s from approximately 60 month period prior to July each year).

Next Fama and French estimate betas for the portfolios.

Result: Almost no relation within the size deciles between the average returns and betas!

Thus Fama and French conclude that size is an important factor explaining cross-sectional return variation.
With similar technique Fama and French observe that book-to-market equity (BE/ME) helps explaining variation in returns (at the cost of beta).

On the basis of these results Fama and French (1995) form factor mimicking portfolios SMB (small minus big) and HML (high minus low) for proxies of the size and book-to-market factors.

The SMB portfolio is the difference of return on a portfolio of small stocks and the return of a portfolio of large stocks.

Similarly the HML portfolio is the difference of return on a portfolio of high book-to-market stocks and the return on a portfolio of low book-to-market stocks.
Accordingly thy suggest a three factor model:

\[ E[r_i] - r_f = b_i (E[r_m] - r_f) + s_i E[SMB] + h_i E[HML] \]

(163)

which results to regression

\[ r_i - r_f = b_i(r_m - r_f) + s_i SMB + h_i HML + u_i, \]

(164)

where \( u_i \) is the regression residual.

In addition to the papers a detailed description of the factors can be found at Professor French’s web-site.

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/
Exercise 2.1: Prove results (12)–(14).

Exercise 2.2: Prove results (15)–(18).

Exercise 2.3: Prove MVP result 5⁰.

Exercise 2.4: Work out the Consumption CAPM example found at SAS Web-Site (http://support.sas.com/rnd/app/examples/ets/harvey/index.htm)

Exercise 2.5: Download from French’s web-site monthly data series for the market return, SMB, and HML factors. Download further monthly data for the 25 size portfolios (use the value weighted series and form excess returns using the one month risk-free rate).

a) Estimate the factor loadings of the portfolios on the factors.

a.1) Discuss the model referring the APM

a.2) Test APT-implied hypothesis empirically

b) Estimate the corresponding factor model for the discount factor. Discuss the estimation results.

c) Using the relationship between the discount factor and factor risk-premiums, find out the premiums.

d) Compare the premiums with the time series estimates (i.e., factor averages).