TOPIC 2
THE LINEAR REGRESSION MODEL

1. Introduction

The linear regression model represents one of the most powerful tools used to model and test financial theories. To highlight its importance in finance, a number of applications are presented.

(a). <u>The Capital Asset pricing Model</u>

Consider the Capital Asset Pricing Model (*CAPM*) which relates the return on the *i*th asset at time *t*, $R_{i,t}$, to the return on the market portfolio $R_{m,t}$. Both rates are adjusted by some risk free rate of return $R_{f,t}$. The adjusted rates are called the excess returns.

The risk characteristics of an asset are determined by its β-coefficient

$$\beta = \frac{\operatorname{cov}(R_{i,t} - R_{f,t}, R_{m,t} - R_{f,t})}{\operatorname{var}(R_{m,t} - R_{f,t})}$$

The risk properties of the asset are summarized as follows:

1. $\beta > 1$: the asset exhibits greater risk than the market portfolio as its returns exhibit relatively greater variability. The stock in this case is commonly referred to as an aggressive stock.
2. $\beta = 1$: the assets exhibits the same risk as the market portfolio as its returns exhibit relatively the same variability.
3. $0 < \beta < 1$: the asset exhibits less risk than the market portfolio as its returns exhibit relatively less variability. The stock in this case is commonly referred to as a conservative stock.
4. $-1 < \beta < 0$: the asset returns move in the opposite direction to the returns on the market portfolio. The stock in this case represents an imperfect hedge against movements in the market portfolio.
5. $\beta = -1$: the asset returns move in the opposite direction to the returns on the market portfolio. The stock in this case represents a perfect hedge against movements in the market portfolio as down (up) movements in the market portfolio are matched on average by up (down) movements in the asset.

The *CAPM* relationship is conveniently summarized by the linear regression model

$$R_{i,t} - R_{f,t} = \alpha + \beta(R_{m,t} - R_{f,t}) + u_t$$

where u_t is a disturbance term.

A test of the *CAPM* for a given asset is: $\alpha = 0$

(b). Arbitrage Pricing Theory

A generalization of the *CAPM* model is based on Arbitrage Pricing Theory (APT). A simple form of this model, due to Chen, Roll and Ross ("Economic Forces and the Stock Market", *Journal of Business*, 1986), is to extend the *CAPM* equation by including a set of unanticipated changes, or news. The *APT* equation becomes

$$R_{i,t} - R_{f,t} = \alpha + \beta(R_{m,t} - R_{f,t}) + \gamma_1 X_{unanticipated,t} + u_t$$

where $X_{unanticipated,t}$ represents the unanticipated change at time *t* (for example, unexpected returns on some commodity or unexpected output growth), while the remaining variables are defined as above.

(c). Term Structure of Interest Rates

Consider the relationship between the return on a 3-month bond, $R_{3,t}$ and a 1-month bond, $R_{1,t}$. Under the pure expectations hypothesis

$$(1 + R_{3,t})^3 = (1 + R_{1,t})(1 + E_t R_{1,t+1})(1 + E_t R_{1,t+2})$$

where $E_t(R_{1,t+j})$ represents the conditional expectation of $R_{1,t+j}$ based on information at time *t*. Take the natural log of both sides of the above equation and use the approximation that $E_t R_{1,t+j} \approx \ln(1 + E_t R_{1,t+j})$ to get

$$R_{3,t} = \frac{R_{1,t} + E_t(R_{1,t+1}) + E_t(R_{1,t+2})}{3}$$

Suppose that $R_{1,t}$ follows a random walk

$$R_{1,t} = R_{1,t-1} + v_t$$

where v_t is a disturbance term. Then

$$E_t(R_{1,t+1}) = R_{1,t}$$
$$E_t(R_{1,t+2}) = R_{1,t}$$

Substituting into the above equation for $R_{3,t}$ gives

$$R_{3,t} = R_{1,t}$$

This suggests that the term structure of interest rates can be modeled by the following linear regression model

$$R_{3,t} = \alpha + \beta R_{1,t} + u_t$$

where u_t is a disturbance term. A test of the expectations hypothesis is a test of the following hypotheses

$$\alpha = 0,$$
$$\beta = 1.$$

(d). Present Value Model

According to the Gordon model, the price of a stock is equal to the expected discounted dividend stream

$$P_t = \sum_{j=1}^{\infty} \frac{E_t(D_{t+j})}{(1+R)^j}$$

where D_t is the dividend payment, R is the constant discount factor and $E_t(D_{t+j})$ represents the conditional expectation of D_{t+j} conditional on information at time *t*.

Suppose that D_t follows a random walk,

$$D_t = D_{t-1} + v_t$$

where v_t is a disturbance term. Then

$$E_t(D_{t+j}) = D_t.$$

Substituting into the present value relationship

$$
\begin{aligned}
P_t &= D_t \left[\frac{1}{(1+R)} + \frac{1}{(1+R)^2} + \ldots \right] \\
&= \frac{D_t}{1+R} \left[1 + \frac{1}{(1+R)} + \frac{1}{(1+R)^2} + \ldots \right] \\
&= \frac{D_t}{1+R} \left[\frac{1}{1 - 1/(1+R)} \right] \\
&= \frac{D_t}{R},
\end{aligned}
$$

where the properties of a geometric progression are used

$$1 + \lambda + \lambda^2 + \lambda^3 + \ldots = \frac{1}{1-\lambda}, \qquad 0 < \lambda < 1.$$

Alternatively, the present value model can be expressed as a linear relationship by taking the natural logarithms of both sides

$$\log P_t = -\log(R) + \log(D_t)$$

This suggests that the present value model can be represented by the following linear regression model

$$\log(P_t) = \alpha + \beta \log(D_t) + u_t,$$

where u_t is a disturbance term. A test of the present value model is a test of the following hypothesis

$$\beta = 1.$$

Note that an estimate of the discount factor is obtained by noting that

$$\alpha = -\log(R).$$

Rearranging gives the discount factor

$$R = \exp(-\alpha).$$

2. Formulation and Estimation of the General Linear Regression Model

The examples above show that the relationships between financial series can be represented in general by the following linear regression model

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \ldots + \beta_K X_{K,t} + u_t, \qquad \text{(eqn.1)}$$

where the sample period is $t = 1, 2, \ldots, T$. Here Y is the dependent variable, $X_1 = 1, X_2$ to X_K is a set of explanatory variables, β_k, k=1,2,…,K, are the unknown population coefficients and u_t is a disturbance term.

The same equation in vector-matrix notation:

$$\mathbf{Y}_{T\times 1} = \mathbf{X}_{T\times(K+1)}\boldsymbol{\beta}_{(K+1)\times 1} + \mathbf{u}_{T\times 1}, \qquad \text{(eqn.1*)}$$

where bold letters represent the corresponding vectors and matrices, and the subscript indicates their dimensions. Dimensions are useful for understanding and to make sure that suggested multiplication is valid.

Note: if the model includes intercept the first column in matrix $\mathbf{X}_{T\times K}$ is the column of 1s.

The sample counterparts of (1) and (1*) are

$$Y_t = b_0 + b_1 X_{1,t} + b_2 X_{2,t} + \ldots + b_K X_{K,t} + e_t \qquad \text{(eqn.2)}$$

$$\mathbf{Y}_{T\times 1} = \mathbf{X}_{T\times(K+1)}\mathbf{b}_{(K+1)\times 1} + \mathbf{e}_{T\times 1}, \qquad \text{(eqn.2*)}$$

where b_k (vector $\mathbf{b}$) is the sample estimate of β_k (vector $\boldsymbol{\beta}$); e_t (vector $\mathbf{e}$) is known as the residual or error

$$e_t = Y_t - \hat{Y}_t, \qquad \mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$$

and the fitted values are given by

$$\hat{Y}_t = b_0 + b_1 X_{1,t} + b_2 X_{2,t} + \ldots + b_K X_{K,t},$$

$$\hat{\mathbf{Y}}_{T\times 1} = \mathbf{X}_{T\times(K+1)}\mathbf{b}_{(K+1)\times 1}$$

The b_k's (vector $\mathbf{b}$) are estimated by minimizing the sum of squared errors

$$\sum_{t=1}^{T} e_t^2 \qquad \text{or in vector notations, simply } \mathbf{e}'\mathbf{e}.$$

$$\sum e_t^2 = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{Xb})'(\mathbf{Y} - \mathbf{Xb}) =$$
$$= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{Xb} - \mathbf{b}'\mathbf{X}'\mathbf{Y} + \mathbf{b}'\mathbf{X}'\mathbf{Xb}$$

$$\min_{\mathbf{b}}(\mathbf{e}'\mathbf{e}) \Rightarrow \text{FOC:} \frac{\partial(\mathbf{e}'\mathbf{e})}{\partial \mathbf{b}} = 0 - \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{Xb} \equiv 0$$
$$2\mathbf{X}'\mathbf{Xb} - 2\mathbf{X}'\mathbf{Y} = 0 \Rightarrow \mathbf{X}'\mathbf{Xb} = \mathbf{X}'\mathbf{Y}$$
$$\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Xb} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \Rightarrow \mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$
$$\text{SOC:} \ \frac{\partial^2(\mathbf{e}'\mathbf{e})}{\partial \mathbf{b}\partial \mathbf{b}'} = 2\mathbf{X}'\mathbf{X} - \text{pos. definite} \ \Rightarrow \ \min$$

This rule according to which the coefficients $\boldsymbol{\beta}$ they are estimated is called the ordinary least squares (OLS) *estimator*, while an estimated numerical value for a coefficient for the observed realization of a random sample are the OLS *estimates*, while.

In vector notation OLS estimator is given by $\mathbf{b}_{(K+1)\times 1} = (\mathbf{X}'_{(K+1)\times T}\mathbf{X}_{T\times(K+1)})^{-1}\mathbf{X}'_{(K+1)\times T}\mathbf{Y}_{T\times 1}$.

Assumptions about disturbance term **u**

(i) The disturbance term has zero mean
(ii) The variance of the disturbance term is constant for all observations (Homoskedasticity assumption)
(iii) The disturbances corresponding to different observations have zero correlation (No autocorrelation)
(iv) The disturbance at time *t* is uncorrelated with the values of the explanatory variables at time *t* or, formally, $E(\mathbf{X}'\mathbf{u}) = 0$. (In this case, the explanatory variables are said to be contemporaneously exogenous). Alternatively, we could assume that **X** is non-stochastic (deterministic and can be taken outside of the expectation operator).
(v) The disturbances assumed to be normally distributed (not crucial for large *T*)
(vi) There is no perfect linear relationship between the explanatory variables (No multicollinearity).
(vii) The dependent and independent variables are stationary (that is, the variables do not contain random walk components, which we shall discuss later).

Under above assumptions, the OLS estimator is *asymptotically* (for large *T*) consistent, efficient and normally distributed. Further, the usual OLS standard errors, t-statistics, F-statistics, and LM statistics discussed further are *asymptotically* valid.

If X is non-stochastic, the OLS estimator $\mathbf{b}$ is unbiased and efficient.
Unbiasness, i.e. $E(\mathbf{b}) = \boldsymbol{\beta}$, is easy to proof by substituting $\mathbf{Y}$ from Eq. 1* in the estimator:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

$$E(\mathbf{b}) = \boldsymbol{\beta} + E\left((\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{u})\right) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}E\left(\mathbf{X}'\mathbf{u}\right) = \boldsymbol{\beta} + 0 = \boldsymbol{\beta}$$

Note: $\mathbf{b} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$

Variance-covariance matrix is given by:

$$E\left((\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})'\right) = E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} =$$

$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Note: we used here (ii) and (iii) which implies that $E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The resulting variance-covariance matrix has σ^2 on the diagonal (constant variance) and 0s on the off-diagonal elements (no serial correlations).

Variance of the disturbance term σ^2 is not observed and need to be estimated. Its estimate is given by s^2 below.

3. Diagnostics and Tests in the General Linear Regression Model

There exists a number of diagnostics which can be used to determine if the estimated model is estimated correctly. In particular, if there is no information contained in the estimated residuals, namely, in e_t, this is evidence that no information has been excluded and that the chosen model is correctly specified.

(a). <u>Sum of Squared Residuals</u>

The objective of OLS is to minimize the sum of squared residuals. The sum of squares can be used to compute the variance of the residuals

$$s^2 = \frac{1}{T - (K+1)}\mathbf{e}'\mathbf{e} = \frac{1}{T - (K+1)}\sum_{t=1}^{T} e_t^2$$

Note: need to divide by $T - (K+1)$ (but not T as in sample mean estimator) to obtain *unbiased* estimator of variance σ^2. This accounts for the fact that K+1 parameters are estimated in this regression.

The standard error of the regression is given by

$$s = \sqrt{\frac{1}{T - (K + 1)} \sum_{t=1}^{T} e_t^2}$$

Relatively large values of s indicate that a substantial amount of change in the dependent variable cannot be explained by changes in the independent variables.

(b). The Coefficient of Determination

The coefficient of determination is a measure of the goodness of fit of the model. It measures the proportion of variation in the dependent variable Y that is explained by the regression equation. It is computed as

$$R^2 = \frac{\text{Explained sum of squares}}{\text{Total sum of squares}} = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T} (Y_t - \bar{Y})^2}$$

where:

$\text{Explained sum of squares} = \sum_{t=1}^{T} (\hat{Y}_t - \bar{Y})^2$

This is the sum of squared deviations of the regression values of Y, $\hat{Y}$ about the mean of Y.

$\text{Total sum of squares} = \sum_{t=1}^{T} (Y_t - \bar{Y})^2$

This is the total sum of squared deviations of the sample values of Y about the mean of Y.

Interpretation

If the regression equation contains a constant term, R^2 is between zero and one. The closer is R^2 to one, the better the fit. For example, an $R^2 = 0.9162$ means 91.62% of variation in the dependent variable is explained by the regression equation. This is considered to be a good fit. On the other hand, an $R^2 = 0.21$ means that only 21% of the variation in the dependent variable is explained by the regression equation. The fit is not particularly good and suggests that the regression equation has excluded important explanatory variables.

It can be shown that R^2 will never decrease when another variable is added to the regression equation. Hence there may be a tendency to keep adding explanatory variables into the regression equation so as to increase R^2 without reference to any underlying economic theory. To circumvent this problem, the adjusted R^2 is computed as

$$\bar{R}^2 = 1 - (1 - R^2)\frac{T - 1}{T - (K + 1)}$$

(c). Test of Coefficients (t-tests)

To test the importance of an explanatory variable in the regression equation, the associated parameter estimate can be tested to see if it is zero. A t-test is used to do this. The null and alternative hypotheses are, respectively,

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

The test statistic is

$$\text{t-statistic: } t = \frac{b_k}{SE(b_k)}$$

where b_k is the OLS estimated coefficient of β_k and $SE(b_k)$ is the corresponding OLS standard error.

The t-test is distributed as Student t with *T-K* degrees of freedom. For large *T*, for 2-sided test values of the t-test in the range of -2 to 2, represent a failure to reject the null hypothesis at approximately the 5% level. Alternatively, p-values less than $\alpha = 0.05$ constitute rejection of the null hypothesis at the 5% level.

(d). Robust standard errors

The OLS standard error of b_k (denoted $SE(b_k)$) is not valid if the errors in the regression model are heteroscedastic and\or serially correlated. White (1980) derived the correct formula for the standard error of b_k when the errors are heteroscedastic of unknown form and are not autocorrelated. These standard errors are known as White or heteroscedastic-consistent standard errors. Denoted the White or heteroscedastic-consistent standard error of b_k as $SE^W(b_k)$. If heteroscdasticity but not autocorrelation is present in the estimated residuals, a t-test of the significant of b_k should be undertaken using the statistic

$$\text{t-statistic: } t = \frac{b_k}{SE^W(b_k)}$$

Newey and West (1987) generalized the formula of White to cover both the case of heteroscedasticity and serial correlation of unknown form in the residuals. Denote the Newey-West or heteroscedastic and autocorrelation consistent standard error of b_k as $SE^{NW}(b_k)$. If heteroscdasticity and autocorrelation is present in the estimated residuals, a t-test of the significant of b_k should be undertaken using the statistic

$$\text{t-statistic: } t = \frac{b_k}{SE^{NW}(b_k)}$$

To calculate the Newey-West standard error of b_k (that is, $SE^{NW}(b_k)$), a lag truncation parameter, which represents the number of autocorrelations used in accounting for the persistence in the OLS residuals, must be chosen. Newey and West suggest taking the lag truncation parameter as the integer part of $4(T/100)^{2/9}$. Eviews adopts this suggestion. Others have suggested $T^{1/4}$. Note that if the lag truncation parameter is chosen to be zero, $SE^{NW}(b_k)$ corrects only for heteroscedasticity and is identical to $SE^{W}(b_k)$.

(e). <u>F-test</u>

A joint test of all the explanatory variables is determined by the F-test. For the case where there is an intercept, the null and alternative hypotheses are respectively,

$$H_0 : \beta_1 = \ldots = \beta_K = 0$$
$$H_1 : \text{ at least one } \beta_k \text{ is not zero}$$

The F-statistic is computed as

$$F = \frac{R^2 / K}{(1 - R^2)/(T - (K + 1))}$$

This is distributed as $F_{K,T-(K+1)}(\alpha)$. Large values of F constitute acceptance of the alternative hypothesis. Alternatively, p-values less than $\alpha = 0.05$ constitute rejection of the null hypothesis.

(f). <u>Testing Linear Restrictions</u>

A special case of the F-test discussed immediately above is when it is necessary to test subsets of parameters. In the case of testing *APT*, the restrictions are

$$H_0 : \beta_i, \beta_j = 0$$
$$H_1 : \beta_i, \beta_j \neq 0$$

where β_i and β_j are the coefficients associated with the unanticipated variables.

To perform the test,

1. Estimate the APT model and retrieve the unrestricted sum of squared residuals SSU.

2. Estimate the CAPM and retrieve the restricted sum of squared residuals SSR. (Note that the CAPM model is the restricted model since it is a special case of the APT model where the coefficients on the unanticipated variables are zero).

3. Compute the F-statistic

$$F = \frac{(SSR - SSU)/R}{SSU/(T-K)},$$

where R is the number of restrictions. The statistic is distributed as $F_{R,T-K}$.

4. A large value of the statistic (larger than critical values) constitutes rejection of the null hypothesis that the restrictions are valid.

(g). Durbin-Watson Test of Autocorrelation

One way of testing the adequacy of the regression specification is to examine if there are any patterns in the residuals. A common statistic used for this purpose is the Durbin-Watson (*DW*) statistic. The null and alternative hypotheses are respectively:

H_0: No autocorrelation

H_1: Autocorrelation (positive)

The DW statistic is given by:

$$DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

Values of *DW* around 2 constitute acceptance of the null hypothesis. As a broad rule of thumb, values of *DW*<1.5 suggest (positive) first order autocorrelation.

(h). LM test of Autocorrelation

This is a more general test of autocorrelation than the DW test as it allows for higher order autocorrelation. This test can also be used with and without lagged dependent variables. Suppose the OLS estimated regression model is

$$Y_t = b_0 + b_1 X_{1,t} + b_2 X_{2,t} + e_t$$

The null and alternative hypotheses are, respectively,

H_0: No autocorrelation

H_1: Autocorrelation of order q

The LM test is as follows. First estimate the auxiliary regression

$$e_t = \gamma_0 + \gamma_1 X_{1,t} + \gamma_2 X_{2,t} + \delta_1 e_{t-1} + \ldots + \delta_q e_{t-q} + error_t$$

Denote the R-squared from this auxiliary regression as R_a^2. Then the LM test statistic is

$$LM = (T - q)R_a^2 \sim \chi^2(q) \text{ under the null}.$$

This is usually called the Breusch-Godfrey test for serial correlation of order q. An alternative is to compute the F test for the joint significance of the coefficients on the lagged residuals in the auxiliary regression.

(i). White test of Heteroscedasticity

This tests the constancy of the error variance. Suppose the OLS estimated regression model is

$$Y_t = b_0 + b_1 X_{1,t} + b_2 X_{2,t} + e_t$$

The null and alternative hypotheses are, respectively,

H_0 : No heteroscedasity

H_1 : Heteroscedasity

White's test is as follows. First estimate the auxiliary regression

$$e_t^2 = \gamma_0 + \gamma_1 X_{1,t} + \gamma_2 X_{2,t} + \gamma_{11} X_{1,t}^2 + \gamma_{22} X_{2,t}^2 + \gamma_{12} X_{1,t} X_{2,t} + error_t$$

Denote the R-squared from this auxiliary as R_a^2. Then the LM test statistic is

$$LM = TR_a^2 \sim \chi^2(q) \text{ under the null}$$

where q is the number of variables (excluding the constant!) in the auxiliary regression (in this case q=5).

(j). Normality test

The Jacque-Bera test of normality can be applied to the OLS residuals. The null and alternative hypotheses are respectively:

$H_0 : u_t$ Normal

$H_1 : u_t$ Nonnormal

(k). Residual Plots

If the regression equation explains all of the movements in the dependent variables, there should be no pattern in the residuals e_t. Otherwise, if there is a pattern in the residuals, this pattern can (hopefully) be modeled thereby improving the predictions of the model. This suggests that a good way to see if there are patterns in the regression residuals is to plot the residuals over time.

4. An Application: Mobil CAPM

In this application the beta coefficient for the U.S. petroleum firm *Mobil* is estimated. The data are monthly starting in January 1978 and ending in December 1987. (The data are from the now classic applied econometrics textbook by Berndt).

The excess returns variables are calculated as

$$E\_MOBIL_t = MOBIL_t - RISKFREE_t$$
$$E\_MARKET_t = MARKET_t - RISKFREE_t, \quad \backslash$$

where *MOBIL* are the returns on Mobil stock, *MARKET* is the market return and *RISKFREE* is the risk free rate of interest.

The following linear regression model is estimated by *OLS*

$$E\_MOBIL_t = \alpha + \beta E\_MARKET_t + u_t$$

The results from estimating this equation are given in the following table and figure. The key points are:

(i) The estimate of β is 0.715, shows that the stock is a defensive stock in the portfolio (moves up/down slower than the market).

(ii) Excess market returns (*E_MARKET*) is an important explanatory variable: it is significant at the 1% level (p-value<0.01).

(iii) The intercept term is insignificant (p-value>0.05) so that the CAPM model holds for this stock.

(iv) R^2 and $\bar{R}^2$ show less than 40% of variation in Mobil excess returns are explained by market excess returns.

(v) The F-statistic is 69.685 is highly significant (p-value<0.01).

(vi) The Durbin-Watson statistic of 2.087 is near 2, shows that there is no evidence of autocorrelation.

(vii) Inspection of the plots of the residuals reveals a large positive residual in 1980:2.

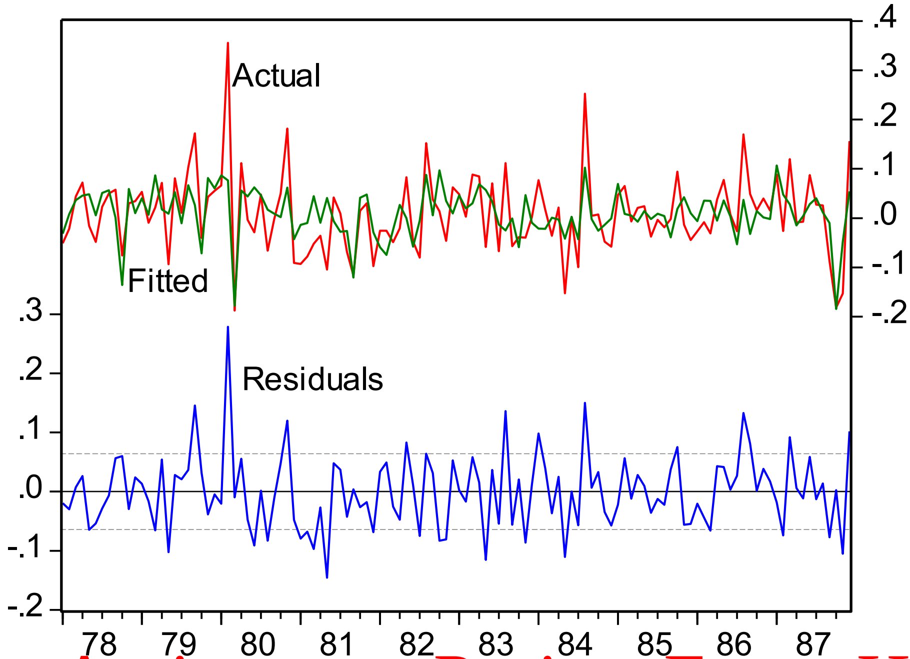
Table 1: OLS Estimates of the Mobil CAPM Model

Dependent Variable: E_MOBIL
Method: Least Squares
Date: 08/03/04 Time: 23:52
Sample: 1978M01 1987M12
Included observations: 120

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004241	0.005881	0.721087	0.4723
E_MARKET	0.714695	0.085615	8.347761	0.0000

R-squared	0.371287	Mean dependent var	0.009353
Adjusted R-squared	0.365959	S.D. dependent var	0.080468
S.E. of regression	0.064074	Akaike info criterion	-2.641019
Sum squared resid	0.484452	Schwarz criterion	-2.594561
Log likelihood	160.4612	F-statistic	69.68511
Durbin-Watson stat	2.087124	Prob(F-statistic)	0.000000

Figure 1: Plots of actual, fitted and residuals of the Mobil CAPM Model



5. Extensions

(a). Dynamics

The discussion so far has concentrated on relationships between variables at the same point in time. However, dynamics can be incorporated into the model by including lag variables. As an example, the regression can be specified as

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \ldots + \beta_K X_{K,t} + \lambda Y_{t-1} + u_t$$

where the dependent variable Y_t is a function of the independent variables given by $X_{i,t}$ and the lagged dependent variable Y_{t-1} with parameter λ.

(b). Dummy Variables

Dummy variables are used to model qualitative changes in finance variables and\or relationships between financial variables. Some examples are:

1. Stock market crash
   Consider the present value model
   $$PRICE_t = \alpha + \beta DIV_t + \gamma CRASH_t + u_t$$

where $PRICE_t$ is the stock market price, DIV_t is the dividend payment, u_t is a disturbance term and

$$CRASH_t = \begin{cases} 0: \text{pre-crash period} \\ 1: \text{post-crash period} \end{cases}$$

is a dummy variable. The dummy variable $CRASH_t$ has the effect of changing the intercept in the regression equation.

$$PRICE_t = \alpha + \beta DIV_t + u_t \qquad : \text{pre-crash period}$$
$$PRICE_t = (\alpha + \gamma) + \beta DIV_t + u_t \qquad : \text{post-crash period}$$

2. Day-of-the-Week Effects
Sometimes share prices exhibit greater movement on Monday than during the week. One reason is that the extra movement is the result of the build up of information over the weekend when the stock market is closed. To capture this behaviour consider the regression model

$$R_t = \alpha + \beta_1 MONDAY_t + \beta_2 TUESDAY_t + \beta_3 WEDNESDAY_t + \beta_4 THURSDAY_t + u_t$$

where the data are daily. The dummy variables are defined as follows

$$MONDAY_t = \begin{cases} 0: \text{not Monday} \\ 1: \text{Monday} \end{cases}$$

$$TUESDAY_t = \begin{cases} 0: \text{not Tuesday} \\ 1: \text{Tuesday} \end{cases}$$

$$WEDNESDAY_t = \begin{cases} 0: \text{not Wednesday} \\ 1: \text{Wednesday} \end{cases}$$

$$THURSDAY_t = \begin{cases} 0: \text{not Thursday} \\ 1: \text{Thursday} \end{cases}$$

A statistically significant estimate of β_1 for example, would imply that returns are different on Monday.

Note: Friday is not included to avoid dummy variable trap (perfect multicollinearity). Alternatively we could drop the intercept and include all week dummies.

6. Maximum Likelihood Estimation of the Simple Linear Regression Model

Consider the simple linear regression model

$$Y_t = \alpha + \beta X_t + u_t \text{ where } u_t \sim iid\, N(0, \sigma^2)$$

It follows that $Y_t \sim N(\alpha + \beta X_t, \sigma^2)$ so that Y_t has probability density function

$$f(Y_t \mid \alpha + \beta X_t, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(Y_t - \alpha - \beta X_t)^2}{\sigma^2}\right\}$$

Under the *iid* assumption on the error term, the Y_t's are also *independent*. Thus the joint probability density function for the Y_t's is

$$f(Y_1, Y_2, \ldots, Y_T \mid X_1, X_2, \ldots, X_T, \alpha, \beta, \sigma^2) = \prod_{t=1}^{T} f(Y_t \mid \alpha + \beta X_t, \sigma^2)$$
$$= \frac{1}{\sigma^T(\sqrt{2\pi})^T} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\frac{(Y_t - \alpha - \beta X_t)^2}{\sigma^2}\right\}$$

(eqn.3)

The LHS of this expression (that is, the joint density) can be interpreted as the probability[1] to observe a given set of Y values, conditional on the set of X values and the parameters α, β and σ^2. Alternatively, it may be interpreted as a function of α, β and σ^2 conditional on the sample outcomes $Y_1, Y_2, \ldots, Y_T$ and $X_1, X_2, \ldots, X_T$. In the latter interpretation it is referred to as a likelihood function and written

$$\text{Likelihood function } = L(\alpha, \beta, \sigma^2; Y_t, X_t, t = 1, \ldots, T)$$

with the order of the symbols in brackets reflecting the emphasis on the parameters being conditioned on the observations. Maximizing the likelihood function with respect to the three parameters ($\beta_0, \beta_1, \sigma^2$) gives estimators of the parameters ($\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2$) which maximize the probability of obtaining the sample values that have actually been observed.

---

[1] We talk about continuous random variables here and, therefore, technically the probability to observe any specific value is equal to zero. Instead, we need to talk about the probability of observing a range of values around Y_i:

$$P\left(Y_t - \frac{\delta}{2} \le Y \le Y_t + \frac{\delta}{2}\right) \approx \delta f(Y_t), \text{where } \delta \text{ is a small constant}$$

This point is often omitted since δ is a constant, which does not affect the optimization procedure.

In most applications, it is computationally easier to maximize the log of the likelihood function rather than maximizing the likelihood function itself. It makes no difference since the resulting expressions will yield estimators of the parameters that maximize both the likelihood and the log of the likelihood since the log is a monotonic transformation that does not alter the location of the maximum. The log-likelihood corresponding to equation (3) is

$$l(\alpha, \beta, \sigma^2 \mid Y_t, X_t \; t = 1, \ldots, T) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(Y_t - \alpha - \beta X_t)^2$$

(eqn. 4)

The first-order conditions for a maximum are

$$\frac{\partial l}{\partial \alpha} = 0, \frac{\partial l}{\partial \beta} = 0, \frac{\partial l}{\partial \sigma^2} = 0.$$

Solving these three equations yields the maximum likelihood estimators of the parameters. They are

$$a = \bar{Y} - b\bar{X}$$

$$b = \frac{\sum_{t=1}^{T} Y_t X_t - T\bar{Y}\bar{X}}{\sum_{t=1}^{T} X_t^2 - T\bar{X}^2}$$

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(Y_t - a - bX_t)^2$$

Note the following. First the maximum likelihood estimators *a* and *b* are simply the OLS estimators. The reason is that the values that maximize the log likelihood also minimize the sum of squared residuals $\sum_{t=1}^{T}(Y_t - a - bX_t)^2$. Second, the first order conditions are easily solved because they are *linear*. As we will see latter, this is not the case with respect to ARCH\GARCH models. Third, the major properties of maximum likelihood estimators (MLEs) are *large sample* or *asymptotic* ones. MLEs are consistent and asymptotically normal. They are also asymptotically efficient in the sense that no other consistent and asymptotically normal estimator can have a smaller asymptotic variance.

We can extend this easily to the linear regression with multiple coefficients. Vector notation will be handy in this case.