Testing the CAPM

Multivariate GARCH models.

Financial market volatility moves together over time across assets and markets. Recognizing this commonality through a multivariate modeling framework leads to obvious gains in efficiency.

Following Bollerslev et al (1986) consider a multivariate extension of the GARCH(p,q) as follows:

Consider a system of n regression equations,

$$y_t = \mu + u_t$$

$$vech(H_t) = C + \sum_{i=1}^{p} A_i vech(u_{t-i}u'_{t-i}) + \sum_{i=1}^{q} B_i vech(H_{t-i})$$

where $u_t|I_{t-1}$~N(0,H_t)

In this formulation, $H_t = E(u_t u'_t|I_{t-1})$ is the $n \times n$ conditional variance matrix associated with the error vector u'_t and $vech(H_t)$ denotes the $(n(n+1))/2$ $\times$ 1 vector of all the unique elements of H_t obtained by stacking the lower triangle of H_t[1] .

Also

- μ is $n \times 1$
- C is $(n(n+1))/2 \times 1$
- $A_1, A_2, ...A_p, B_1, ...B_q$ are $(n(n+1))/2 \times (n(n+1))/2$

Example: A Bivariate GARCH(1,1)

Since most empirical applications of the model have restricted attention to multivariate GARCH(1,1) systems. We first consider the easiest example is the simple bivariate process which depends on its conditional variance covariance matrix (H_t).

[1] In general if

$A = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$

Then

$vech(A) = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$

For a symmetric matrix

$B = \begin{bmatrix} a & c \\ c & d \end{bmatrix}_{n \times n}$

is standard to write the vector using either the lower or the upper triangular information

$vech(B) = \begin{bmatrix} a \\ c \\ d \end{bmatrix}_{(n \times (n+1)/2) \times 1}$

$$\begin{bmatrix} x_t \\ w_t \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_w \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \nu_t \end{bmatrix}$$

where

$$\begin{aligned} \begin{bmatrix} V_{t-1}(x_t) \\ COV_{t-1}(x_t w_t) \\ V_{t-1}(w_t) \end{bmatrix} &= \begin{bmatrix} c_x \\ c_{xw} \\ c_w \end{bmatrix} + \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \begin{bmatrix} (\varepsilon_{t-1})^2 \\ (\varepsilon_{t-1}\nu_{t-1}) \\ (\nu_{t-1})^2 \end{bmatrix} \\ &+ \begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_7 & b_8 & b_9 \end{bmatrix} \begin{bmatrix} V_{t-2}(x_{t-1}) \\ COV_{t-2}(x_{t-1} w_{t-1}) \\ V_{t-2}(w_{t-1}) \end{bmatrix} \end{aligned}$$

Estimation of this model may be obtained by ML using numerical procedures. Two main problems with the estimation of these type of models should be apparent:

1. As its stands this model is very general, requiring a large number of parameters to be estimated. The "simplest" general model has 23 parameters.
2. We should ensure H_t to be positive definite. (see Bera and Higgins).

Various simplifications have been suggested to account for (1) and (2).

A Diagonal Vech Parameterization

(Bollerslev, Engle and Wooldridge (1982))

A natural simplification is to assume that each covariance depends only on its own past values and innovations, i.e., that A_i and B_i are diagonal. Then Each element of H_t follows an univariate GARCH model driven by the corresponding cross product $u_t u_t'$. The implied conditional covariance matrix is always positive definite if the matrices of parameters

$$\begin{bmatrix} c_x & c_{xw} \\ c_{xw} & c_w \end{bmatrix}, \begin{bmatrix} a_1 & a_5 \\ a_5 & a_9 \end{bmatrix}, \begin{bmatrix} b_1 & b_5 \\ b_5 & b_9 \end{bmatrix}$$

are all positive definite. This can be ensured by doing simple Cholesky transformations to each of these matrices.

A Quadratic Specification,

(Engle and Kroner(1995))

An alternative to the diagonal vech parameterization is to achieve a positive definite covariance matrix is the quadratic parameterization

$$H_t = C'C + A\prime u_{t-1}u'_{t-1}A + B'H_{t-1}B.$$

where C is a lower triangular with $n(n+1)/2$ parameters and A and B are square matrices with n^2 parameters each. Weak restrictions on A and B guarantee that the H_t is always positive definite.

A Constant-Correlation Specification,

(Bollerslev(1990))

This specification is similar to the diagonal specification, but imposes the restriction that the correlation between the assets is constant.

$$COV_{t-1}(x_t w_t) = \rho\sqrt{V_{t-1}(x_t)V_{t-1}(w_t)}$$

Where ρ is also estimated with the rest of the parameter set. The conditions to get a positive definite matrix are as in the diagonal case.

Stationarity and Co-persistence

Stationarity

The conditions for stationarity and moment convergence for the multivariate case are similar to those discussed in the univariate case. Specially, for the multivariate vech GARCH(1,1) model defined above, the minimum square error forecast for $vech(H_t)$ as of time $s < t$ takes the form

$$E_s(vech(H_t)) = C[\sum_{k=0}^{t-s-1}(A_1 + B_1)^k] + (A_1 + B_1)^{t-s}vech(H_s)$$

where $(A_1 + B_1)^0$ is equal to the identity matrix by definition.

Let $V\Lambda V^{-1}$ denote the Jordan decomposition of the matrix $A_1 + B_1$, so that $(A_1+B_1)^{t-s} = V\Lambda^{t-s}V^{-1}$. Thus, $E_s(vech(H_t))$ converges to the unconditional covariance matrix $C(I - A_1 - B_1)^{-1}$, for $t \to \infty$, if and only if the absolute value of the largest eigen value of $A_1 + B_1$ is strictly less than one.

Results for other multivariate formulations are scarce, a possible exception is the constant conditional correlations parameterization where the conditions for the model to be covariance stationary are simply determined by the conditions of *each* of the univariate conditional variances.

Co-Persistence in Variance

The empirical estimates for univariate and multivariate ARCH models often indicate a high degree of persistence in the forecast moments of the conditional variances, i.e. $E_s(H_t)_{ii}$, $i = 1, 2, ...N$, for $t \to \infty$. At the same time, the commonality in volatility movements suggest that this persistence may be common across different series. More formally Bollerslev and Engle (1993) define the multivariate ARCH process to be *co-persistent* in variance if at least one of the elements in $E_s(H_t)$ is non-convergent for increasing forecasts, $t - s$, but there exists a linear combination $\gamma' \varepsilon_t$, such that for every forecast origin s, the forecasts of the corresponding future conditional variances $E_s(\gamma' H_t \gamma)$ converge to a finite limit, independent of time s information. The conditions for this to occur are presented in Bollerslev and Engle (1993) and are similar to the conditions for co-integration in the mean as developed by Engle and Granger (1987).

Multivariate GARCH-M models .

Engle and Bollerslev (1986) consider a multivariate extension of the gARCH-m as follows:

Consider a system of n regression equations,

$$y_t = BX_t + Dvech(H_t) + u_t$$

$$vech(H_t) = C + \sum_{i=1}^{p} A_i vech(u_{t-i}u'_{t-i}) + \sum_{i=1}^{q} B_i vech(H_{t-i})$$

where $u_t | I_{t-1} \tilde{} N(0, H_t)$

Where

- B is $n \times k$
- D is $n \times (n(n+1))/2$
- C is $(n(n+1))/2 \times 1$
- $A_1, A_2, ...A_p, B_1, ...B_q$ are $(n(n+1))/2 \times (n(n+1))/2$

A Bivariate GARCH-M(1,1) model

The easiest example is the simple bivariate process which depends on its conditional variance covariance matrix (H_t)

$$\begin{bmatrix} x_t \\ w_t \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_w \end{bmatrix} + \begin{bmatrix} d_1 & d_2 & d_3 \\ d_4 & d_5 & d_6 \end{bmatrix} \begin{bmatrix} V_{t-1}(x_t) \\ COV_{t-1}(x_t w_t) \\ V_{t-1}(w_t) \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \nu_t \end{bmatrix}$$

where

$$
\begin{bmatrix} V_{t-1}(x_t) \\ COV_{t-1}(x_t w_t) \\ V_{t-1}(w_t) \end{bmatrix} = \begin{bmatrix} c_x \\ c_{xw} \\ c_w \end{bmatrix} + \begin{bmatrix} c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \\ c_4 & c_5 & c_6 \end{bmatrix} \begin{bmatrix} (\varepsilon_{t-1})^2 \\ (\varepsilon_{t-1}\nu_{t-1}) \\ (\nu_{t-1})^2 \end{bmatrix} + \begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_4 & b_5 & b_6 \end{bmatrix} \begin{bmatrix} V_{t-2}(x_{t-1}) \\ COV_{t-2}(x_{t-1} w_{t-1}) \\ V_{t-2}(w_{t-1}) \end{bmatrix}
$$

Estimation

The estimation of all the models presented above is carried out using conditional maximum likelihood estimation. The conditional log likelihood function for a single observation can be written as

$$
L_t(\theta) = -(n/2)\log(2\pi) - (1/2)\log(H_t(\theta)) - (1/2)u_t(\theta)' H_t^{-1}(\theta) u_t(\theta)
$$

where θ represents a vector of parameters, n represents number of equations and t represents time.

Then conditional on initial values for u_0 and H_0 , the likelihood function for the sample $1, .., T$ can be written as

$$
L(\theta) = \sum_{t=1}^{T} L_t(\theta)
$$

The maximization is usually achieved through numerical methods Notice that the model is highly non-linear and very unstable.

An application of the MGARCH-M model : Testing The cAPM

General issues

Three main difficulties associated with testing the CAPM

1. The CAPM is a statement about the relationships between ex ante risk premiums and betas, both of which there are not directly observable. This problem is usually dealt with by assuming that investors form rational expectations: the realized return on assets are drawings from the ex ante probability distribution of returns on those assets.

2. The Roll- critique; many assets are not measurable (human capital etc.) so tests of the CAPM have to be based on proxies for the market portfolio which only includes (a subset of) traded assets.

3. The CAPM is a single-period model. However, in order to test the model time series are frequently used. This is only a valid procedure if both risk premia and betas are stationary.

The Unconditional CAPM a short review of the literature.

Markowitz(1959) laid the ground work for the CAPM. He cast the investor's portfolio selection problem in terms of the expected return and variance of the return. He argued that investors would optimally hold a mean-variance efficient portfolio. Sharpe and Litener showed that if investors have homogeneous expectations and optimally hold mean-variance portfolios then, the market portfolio will also be mean-variance. They also assume the existence of lending and borrowing at a risk free rate of interest. For this version of the CAPM the expected return for asset j is[2]

$$E(r_j) = r_f + \beta_{jm}(E(r_m) - r_f)$$

Where r_j, r_f and r_m are the asset j, risk free and market rate of return respectively.

Defining excess returns $\tilde{r}$, we can rewrite the above expressions as

$$E(\tilde{r}_j) = \beta_{jm}(E\tilde{r}_m)$$

where $\tilde{r}_j = r_j - r_f, \quad \tilde{r}_m = r_m - r_f$

Since the risk free rate is assumed to be non-stochastic the two equations above are equivalent. Nevertheless in empirical applications r_f is typically stochastic and therefore the β' may differ.

Empirical tests of the Sharpe-Litener CAPM have focused on three implications of the last expression.

1. The intercept is zero
2. The parameter beta completely captures the cross sectional variation of expected excess returns
3. the market risk premium, $(E\ \tilde{r}_m)$ is positive.

Statistical Framework for the Sharpe-Lintner Version.

The CAPM is a single-period model and do not have time dimension. For econometric analysis of the model we need to add an assumption concerning the time series properties of returns over time. We assume that the returns are IID and jointly multivariate Normal.

[2] In testing the CAPM we usually do some simplifying assumptions. One is that the risk free rate is often approximated by the treasury bill lending rate.

Imposing the restriction that r_{jt} and r_{mt} are multivariate normal, it follows that

$$E(\tilde{r}_{jt}|\tilde{r}_{mt}) = \alpha + \beta\tilde{r}_{mt}$$

where
$\beta_j = cov(\tilde{r}_{jt}\tilde{r}_{mt})/var(\tilde{r}_{mt})$
and
$\alpha_j = E(\tilde{r}_{jt}) - \beta_j E(\tilde{r}_{mt})$

Therefore

$$\tilde{r}_{jt} = \alpha_j + \beta_j\tilde{r}_{mt} + \varepsilon_{jt} \tag{1}$$

Traditionally β has been estimated by OLS regression, meaning that the range of problems encountered with any linear regression model recur here as well. To test the CAPM is equivalent to test the restriction that $\alpha_j = 0$ and this can be done in this context by a simple F test. Alternatively we may consider the test in a multivariate context were

$$r_t = \alpha + \beta\tilde{r}_{mt} + \varepsilon_t$$

- $E(\varepsilon_t) = 0$
- $E(\varepsilon_t\varepsilon_t') = \Sigma$
- $E(r_{mt}) = \mu_m,\ \ E(r_{mt} - \mu_m)^2 = \sigma_m^2,\ \ Cov(r_{mt}, \varepsilon_t) = 0$

Where β is a vector of $N \times 1$ and ε_t are $N \times 1$ asset returns intercepts and disturbances.

We can estimate this model by maximum likelihood and test the N zero intercept restriction by a likelihood ratio test which is asymptotically distributed $\chi^2(N)$.

The Black (1972) version.

In the absence of a risk free asset Black (1972) derived a more general version of the CAPM. In this version the expected return of asset i en excess of the zero-beta return is linearly related to its beta.

Then

$$E(r_j) = r_{om} + \beta_{jm}(E(r_m) - r_{om})$$

where r is the return of the zero beta portfolio associated with m. This portfolio is defined to be the portfolio that has minimum variance of all the portfolios associated with m.

The econometric analysis of the Black version of the CAPM treats the zero-beta portfolio as an unobserved quantity. This version can be tested as a restriction on the real return market model

$$E(r_j) = \alpha_j + \beta_{jm} E(r_m)$$

and the implication of this version is $\alpha_{jm} = E(r_{om})(1 - \beta_{jm})$

This restrictions can also be tested by a likelihood ratio test.

Applied work

Most tests of the CAPM have used a time-series of monthly rates of return on common stocks listed in the New York Stock Exchange as a proxy of the market portfolio. Suppose that the objective is to explain the (excess) return of the j^{th} portfolio of assets r_{jt} . If there is only one asset per portfolio this will simply be the (excess) return on that asset. "Portfolios" can be interpreted in diverse ways. For example Harvey(1989) has r_{jt} as the (excess) return on equity in the j^{th} country. In the market model r_{jt} is related to the return on the market or aggregate portfolio, r_{mt} . The latter may be a simple average of the returns to all stocks in the economy or perhaps is a weighted average, with weights depending on the value of the portfolio accounted for each stock.

b) The second approach interprets the tests conditional on the investors' information set which is assumed to be jointly stationary and with multivariate normally distributed asset returns.

$$E(\tilde{r}_{jt}|I_{t-1}) = \beta_{jt} E(\tilde{r}_{mt}|I_{t-1}) \tag{2}$$

The Conditional CAPM

In (1) β_j was a constant equal to the ratio $cov(r_{jt}r_{mt})/var(r_{mt})$. However, in line with the distinction between conditional and unconditional moments, one might wish to consider models for r_{jt} in which the conditional density rather than the unconditional density returns is used. Let I_{t-1} be a set of conditioning variables including the past history of r_{jt} and r_{mt} . Then the conditional asset pricing model has

$$E(\tilde{r}_{jt}|I_{t-1}) = \beta_{jt} E(\tilde{r}_{mt}|I_{t-1}) \tag{2}$$

where $\beta_{jt} = cov(\tilde{r}_{jt}\tilde{r}_{mt}|I_{t-1})/var(\tilde{r}_{mt}|I_{t-1})$.

Because the coefficients of the conditional market model are functions of the conditional moments for r_{jt} and r_{mt} it is necessary to model this in some way.

A natural way to proceed is to allow the conditional mean for r_{jt} to depend on its conditional variance, as in GARCH-M models, and to subsequently model $cov(r_t|I_{t-1})$ by a multivariate GARCH.

Multivariate GARCH-M models

A CAPM with time-varying covariances

As discussed above the conditional CAPM can be written as

$$E(r_{jt}|I_{t-1}) - r_{ft-1} = \beta_{jt}[E(r_{mt}|I_{t-1}) - r_{ft-1}]$$

where
$\beta_{jt} = cov(r_{jt}r_{mt}|I_{t-1})/var(r_{mt}|I_{t-1})$.
where, because we now allow the covariance matrix of returns

$$H = \begin{bmatrix} var(r_{jt}|I_{t-1}) & cov(r_{jt}r_{mt}|I_{t-1}) \\ cov(r_{jt}r_{mt}|I_{t-1}) & var(r_{mt}|I_{t-1}) \end{bmatrix}$$

to vary over time, both the expected returns and the betas will, in general, be time varying.

This formulation of the CAPM is, however, non-operational because of the lack of an observed series for the expected market returns.

If we assume that the "market price of risk", λ is constant, where

$$\lambda = (E(r_{mt}|I_{t-1}) - r_{ft-1})/var(r_{mt}|I_{t-1}).$$

So that

$$[E(r_{jt}|I_{t-1}) - r_{ft-1}] = \lambda cov(r_{jt}, r_{mt}|I_{t-1})$$

then we can write

$$r_{jt} = r_{ft-1} + \lambda cov(r_{jt}, r_{mt}|I_{t-1}) + u_{jt}$$

Also nothing that

$$E(r_{mt}|I_{t-1}) - r_{ft-1} = \lambda var(r_{mt}|I_{t-1})$$

(since $\lambda = (E(r_{mt}|I_{t-1}) - r_{ft-1})/var(r_{mt}|I_{t-1})$)

$$r_{mt} = r_{ft-1} + \lambda var(r_{mt}|I_{t-1}) + u_{mt}$$

Where u_{jt} and u_{mt} are the innovations.

This time varying CAPM can be put into multivariate GARCH-M form as

$$y_t = b + d vech(H_t) + u_t$$

where $y_t = (r_{jt} - r_{ft-1}, r_{mt} - r_{ft-1})'$, $vech(H_t) = (var(r_{jt}|I_{t-1}), cov(r_{jt}r_{mt}|I_{t-1}), var(r_{mt}|I_{t-1}))'$, $u_t = (u_{jt}, u_{mt})'$ and

$$d = \lambda \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The zero restrictions implied by the theory may be tested (against a general model) by a likelihood ratio test which is asymptotically distributed $\chi^2(5)$. A by-product of this methodology is that plot of conditional time varying betas can easily been obtained. This plot is very informative since it may provide as with time varying beta ranking and also show how different stocks react to shocks to the economy.