

CUBS MSc Finance 1999-2000

ASSET MANAGEMENT: LECTURE 4

Dr. D.N. Tambakis

ARBITRAGE PRICING THEORY (APT)

Textbook: EG16, BKM10, GT6

Readings: Shanken, J. “The Arbitrage Pricing Theory: Is it Testable?”, The Journal of Finance 37(5), December 1982, pp. 1129-1140.
Sharpe, W. “Factors in NYSE Security Returns”, Journal of Portfolio Management 8 (2), Summer 1982, pp.5-19.

4.1 Overview: From the CAPM to the APT

Arbitrage Pricing Theory (**APT**), founded upon the work of Ross (1976,1977), aims to analyze the equilibrium relationship between assets’ risk and expected return just as the CAPM does. The two key CAPM assumptions of perfectly competitive and efficient markets and homogeneous expectations are maintained. Moreover, in line with the CAPM, the APT assumes that portfolios are sufficiently diversified, so that the contribution to the total portfolio risk of assets’ unique (unsystematic) risk is approximately zero.

The APT’s two main differences from the CAPM are: (a) the explicit modeling of **several factors** affecting assets’ actual and expected returns, as opposed to the CAPM’s focus on the market portfolio, and (b) the fact that the equilibrium relationship is only approximate and is derived based on a **no-arbitrage assumption**. The two are in fact interrelated, as market equilibrium in the CAPM rests on the observability and efficiency of the market portfolio.

With regard to the first difference, recall from lecture 3 that the CAPM was essentially derived from a single-index (**single-factor**) model, i.e. from a process

generating asset returns which was only a function of returns unique to the asset (predictable and unpredictable) and returns on two factors, the market portfolio itself and the riskless asset, or the zero-beta portfolio. The sensitivity of the asset's returns to the market's was defined as the asset's beta, measuring **systematic** (market) risk, while the **unsystematic** (unique) risk of the asset (portfolio) tended to zero through diversification.

APT can then be seen as a multi-index (**multi-factor**) model, i.e. one in which the returns generating process of the portfolio is a function of several factors, generally excluding the market portfolio. The factors are not specified a priori and their choice depends on the question at hand. Possible factors may include particular sector-specific influences, such as price-dividend ratios, as well as aggregate macroeconomic variables such as inflation and interest rate spreads. The covariance of each factor with the portfolio then leads to a natural extension of measuring risk by the beta attached to each factor. In the context of arbitrage pricing models the betas are often referred to as the **factor loadings**. The fact that identification of the market portfolio is not required is a big advantage of the APT--however, its general specification is a mixed blessing because it gives no guidance as to which factors should be considered. It is, consequently, easy to misprice assets by including the wrong (i.e. irrelevant) risk factors. Moreover, in principle one always gains in explanatory power by including more factors, even if they seem economically irrelevant.

The second difference between the CAPM and the APT has to do with the equilibrium notion. In contrast to the CAPM's assumption of an efficient market portfolio which every investor desires to hold, the APT relies on the absence of free arbitrage opportunities. In particular, two portfolios with the same risk cannot offer different expected returns, because that would present an arbitrage opportunity with a net investment of zero. An investor could then guarantee a riskless positive expected return by short-selling one portfolio and holding an equal and opposite long position in the other. As such free arbitrage cannot persist, equilibrium in the APT specifies a linear relationship between expected returns and the betas of the corresponding risk factors.

The **short-selling** assumption is crucial to the equilibrium, as it constitutes one side of the arbitrage portfolio. Equally important is the requirement that the proceeds from short-selling are immediately available. However, despite its shortcomings--especially the difficulties associated with empirical testing of its validity--the APT has caught on in financial practice as it allows for a more detailed and custom-made approach to portfolio risk management than the CAPM. This has become more relevant with the wide-spread use of derivative instruments and their particular types of risk.

4.2 The Basic Analytical Framework

The multi-factor (N -factor linear) model specifies a simple linear relationship between asset i 's returns and the associated N specific factors which influence them:

$$R_i = a_i + b_{i1}I_1 + b_{i2}I_2 + \dots + b_{ij}I_j + \dots + b_{iF}I_F + e_i \quad (1)$$

, where there are j factors I_j , with j taking values from 1 to F , a_i is the predictable (constant) unsystematic component of the return which is unique to the asset, and e_i is the unpredictable (normally distributed with zero mean and constant variance) component of the return which is unsystematic, which we consider as an error term in a regression. Coefficient b_{ij} is asset i 's **factor loading** for factor j , and corresponds to the asset's beta with respect to that factor.

If there are $i=1, \dots, N$ assets whose returns are described by equation (1), then the error terms of different assets are uncorrelated, i.e. their covariance is zero, so $E(e_i e_j) = 0$ for all i, j . This assumption (which will hold approximately in practice) implies that only the systematic risk of a portfolio is priced, since a portfolio's total (average) unsystematic risk will decline to zero as the number of assets increases. A different way to see this is that, if two or more assets' error terms were correlated, that would imply they had a common component which affected both assets. But that component would then qualify as a factor itself. Therefore, all assets' unique and unpredictable (idiosyncratic) influences are uncorrelated. Of course, this does not preclude the unique and predictable

components (the a_i) of two assets' return generating processes from being equal--for example, two stocks which are part of the same market sector.

The factors entering the RHS of (1) need not necessarily be returns on an index or asset group; they can be time-series observations of variables (e.g. interest rates, yield spreads, inflation (CPI) or other macroeconomic time series) which are thought to have an influence on the assets' returns. The factors need not have zero mean, but they should always have zero covariance with the error terms, reflecting their definition as sources of systematic risk, and that of errors as sources of unsystematic (diversifiable) risk. Therefore, whereas in the CAPM systematic risk was equivalent to market risk, under the APT it is the joint influence of all risk factors identified to be common to the assets in the portfolio.

In order to obtain a relationship between *expected*, as opposed to actual, returns and the factor loadings we need to allocate a weight x_i on each of the N assets in our portfolio. Importantly, individual asset weights are *not* constrained to be non-negative, although their algebraic sum has to equal one. In other words, asset short-selling is allowed, and this is crucial for obtaining the equilibrium. The linear combination of weights on the LHS of equation (1) then yields the portfolio's expected return, while on the RHS we have a linear combination of the products of factors and factors loadings. If each portfolio is sufficiently diversified (i.e. if N is large), then its unsystematic risk will tend to zero, so the error terms will average out. In general, as long as the number of factors is less than or equal to the number of assets ($F \leq N$), our system of N simultaneous equations and F unknowns (the coefficients on each factor loading) will have a unique and well-defined solution for any set of portfolio weights. The solution defines the portfolio's expected return as a linear combination of the factor loadings:

$$ER_i = a_i + \sum_{j=1}^F b_{ij} I_j \quad (2)$$

Equation (2) describes a single point in a space with $F+1$ dimensions, one for each of the F risk factors plus the fixed coefficient. The latter then corresponds to the

intercept in this space, i.e. the point where all factor loadings are equal to zero. In the simple case where $F=1$, the resulting two-dimensional space is just our familiar mean-beta space, and the intercept is the risk-free rate of return (the rate of return on Treasury bills, or an equivalent zero-beta rate of return).

The coefficients b_{ij} are analogous to the corresponding single beta coefficient in the CAPM, while the λ_j can be thought of as the price, or **risk premium**, of each risky factor j . Note that in the CAPM the price of market risk was always positive and had an intuitive interpretation as the slope (Sharpe ratio) of the capital market line. In contrast, in the multi-factor framework of the APT there is no such straightforward geometric intuition. Nevertheless, it is still the case that expected return as given by (2) only varies with changes in systematic risk, as measured by the factor loadings. In other words, the investor is not compensated for bearing unsystematic risk.

The APT equilibrium rests on investors' ability to construct an arbitrage portfolio by simultaneously holding a short and a long position in two different portfolios, S and L , which offers a positive expected return with zero risk and zero net investment. Thus the two portfolios have the same risk, but portfolio S offers lower expected return than portfolio L . Investing in such a portfolio guarantees that the expected returns on portfolios S and L are equalized. How does this equalization take place? As the aggregate short position on portfolio S grows, the expected return on this portfolio *rises*, while as the aggregate long position on portfolio L grows, its expected return *falls*. Thus the expected returns are driven to equality, at which point the arbitrage opportunity disappears. Note that assuming that investors have homogeneous expectations is crucial for ensuring that the expected returns converge. All investors believe that the two portfolios present an arbitrage opportunity, and they are allowed to sell one short and buy the other long. Thus, in the background there is perfect competition and full information.

Seen in another way, the above assumptions combined with no restrictions on short-selling imply that the **law of one price** holds: two assets (portfolios) with the same risk cannot sell at different prices, so their expected returns have to be equal. We have

seen that, in the case of a single risky factor ($F=1$), equation (2) amounts to the CAPM and defines equilibrium as the security market line in expected return-beta space. The risky factor is then identified with the market portfolio. Similarly, when there are two risky factors ($F=2$), their linear combination with the zero-beta intercept defines a plane in 3-dimensional space, with the portfolio's expected return on the vertical axis and the combinations of the two factors' betas on the horizontal plane. Equation (2) then becomes:

$$ER_i = I_0 + I_1 b_{i1} + I_2 b_{i2} \quad (3)$$

It is easy to derive the coefficients λ_i in the RHS of equation (3) as a function of the assets' observed characteristics, expected return in particular. Clearly, if we set both factors' beta to be zero for each asset in the portfolio, then equation (3) determines the expected return on a risk-free portfolio to be λ_0 . If a riskless asset exists in practice, then we can set λ_0 to equal its rate of return R_F , while if it does not exist then λ_0 is just the expected return on the (risky) zero-beta portfolio ER_Z . In either case, the expected return on the zero-beta portfolio is equal to the vertical intercept.

Substituting R_F or ER_Z for λ_0 in equation (3), we set factor 1's betas with each asset to one and factor 2's betas with each asset to zero. The last term and its coefficient then cancel, and we are left with the value of the coefficient λ_1 . Assuming that a riskless asset exists, that value is simply:

$$I_1 = ER_1 - R_F \quad (4)$$

The risk premium associated with factor 1 is then just equal to the excess expected return of a portfolio which is only affected by factor 1-type risk over the riskless rate of return. Similarly, setting set factor 2's betas for all assets to one and factor 1's betas to zero, we obtain that the risk premium associated with factor 2 is equal to the excess expected return of a portfolio which is only affected by factor 2-type risk over the riskless rate of return. The price of each risk factor--its risk premium--can therefore be

identified with the excess expected return required of a portfolio which is uniquely exposed to that risk factor.

This risk valuation principle generalizes to the multi-index model with any finite number of risk factors: the price (risk premium) associated with each factor equals the excess return over the riskless rate of a portfolio which only covaries with that particular risk factor. This implies that a relatively high (low) exposure is compensated by a relatively high (low) expected return. It follows that the inclusion of relevant risk factors in pricing assets and/or projects using the APT is key for determining the expected excess return required of that asset/project. Moreover, once the risk premia have been identified via the aforementioned technique, it is essential to be able to estimate the betas (factor loadings) of each factor in order to price portfolios which are simultaneously exposed to more than one risk factor, i.e. portfolios all of whose factor loadings are non-zero. Practical problems related to these two issues underlie a substantial part of the empirical literature on the APT.

4.3 Econometric Testing and Applications of APT

4.3.1 *Factor Identification*

In practice, the issue of the identification of risk factors is addressed by means of three general categories of factors which may be selected in the multi-index model. First, there are factors pertaining to particular stock (firm) characteristics, for example its beta with a broad market index such as the FTSE-All Share or a narrower sector index, the market value of its equity, its sales figures and/or its size or, in the case of corporate bonds, their beta with a long-term bond index. Also, recall from our earlier discussion that the choice of characteristics has to be such that the changes in the corresponding factors affect all assets under consideration. An important study of this category is Fama and McBeth (1973).

Second, macroeconomic factors may be identified based on the criterion that they should affect the generating process of future cash flows of the asset and/or the valuation

of those cash flows by investors. The latter relates directly to interest rates and their term structure, which influences both stocks and fixed-income markets. The former is up to the discretion of the analyst; provided it has an impact on all possible asset subgroups. Typically, such variables include a variety of macroeconomic indices, such as CPI, ULC, industrial production, the import/export balance, as well as forward-looking indices which are considered as leading indicators of economic activity, such as consumer confidence, investment in R&D, etc.

Finally, the third category includes the returns on relevant portfolios which themselves capture the broader influences on the assets under consideration, as in the study of Fama and French (1992). The upshot is that factor identification is a highly judgmental process. In principle, factors are continuously dropped or reintroduced as their influence on the particular asset/portfolio group is seen to strengthen or weaken. This flexibility and the capacity to adapt to the evolution of specific observable sources of risk is an important advantage of the APT over the CAPM.

4.3.2 *Joint Estimation*

In the absence of clearly identified factors and their associated factor loadings, the general APT pricing formula given by equation (2) above requires the simultaneous estimation of all b_{ij} and λ_j for each factor and all assets. Notice that, in principle, each λ_j enters in all N equations as it is the price of a systematic risk factor affecting more than one asset, while each b_{ij} is a unique characteristic describing asset i 's sensitivity to that factor. Thus, each beta coefficient is a particular stock (firm) attribute, and as such it can be estimated from a time-series regression of each asset's realized (actual) returns on all factors' generating processes. This regression, using OLS so that the errors are assumed to be uncorrelated and normally distributed, amounts to the following:

$$R_i = a_i + \sum_{j=1}^F b_{ij} I_j + e_i \quad (5)$$

The I_j are time-series observations of the particular factors chosen. Equation (5) thus provides j OLS estimates of b_{ij} for each i . Once these have been determined, we can substitute their values as known parameters in equation (2) and run a **cross-sectional** (i.e. across stocks/assets) OLS regression to estimate all the λ_j , i.e. estimate the value of the risk premium of each factor. We have then completed estimation of our multi-factor model and can apply the results to all portfolios of the N assets. Notice that this two-stage estimation procedure is necessary because our model involves the simultaneous identification of factor loadings (betas) and risk premia. However, testing the more general structure of the APT is no different from the simpler structure of the CAPM: recall that there also we had to use a time-series regression of each asset's actual returns on the market's to obtain an estimate of the asset's beta, and we then had to perform a cross-sectional regression across assets in order to estimate the coefficient of beta determining each asset's expected return. The only difference was that, in the single-index framework of the CAPM, the coefficient of beta in the cross-section regression was defined as the price of market risk (the **market risk premium**).

It should be clear from the above that the two-stage estimation procedure of simultaneous determination of factors and factor loadings amounts to a *joint* test of the APT and the validity of the set of risk factors included in the assets' returns generating processes. In particular, the first stage (time-series) regression is testing the relevance of the common risk factors included, while the second (cross-section) regression is testing the equilibrium risk-expected return relationship of the APT. In practice, different simultaneous determination techniques are sometimes collectively referred to as **factor analysis**.

Recall from the presentation of the multi-index model in section 2 that the correlation of the error terms (the assets' unique risk) had to be zero in order for diversification to reduce a portfolio's unsystematic risk to zero. In practice, this implies that the covariances of the error terms in the time-series regression should be close to zero, i.e. the $N \times N$ covariance matrix of the error terms should be as close as possible to being diagonal (the non-zero diagonal elements corresponding to the errors' own

variances). This suggests that the actual identification process consists of successive examination of different factors and the resulting error covariances.

The resulting trade-off from including more factors plays a significant role in hypothesis testing. On the one hand, including more factors is bound to improve the R^2 of the regression, i.e. it explains a higher percentage of the variance of the dependent variable (actual returns). On the other hand, extra risk factors increase the likelihood that some covariances of error terms will be significantly non-zero. Therefore, a compromise is made which is reflected in the final number of factors included. The percentage of the variance of actual returns which is left unexplained is attributed to the unsystematic risk of each asset. Note, also, that in section 4.2 the number of factors was already restricted to be at most equal to the number of assets in the model. That restriction imposes an upper bound on the number of possible factors in any given financial application.

Once the betas (factor loadings) of each factor have been estimated, they can be used in the second-stage (cross-section) regression in order to estimate the risk premia (λ_j) for each factor. This necessarily introduces some errors in the significance of the risk premia, reflecting the size of the confidence intervals associated with the estimated betas. In addition, in contrast to tests of the CAPM which always yield a positive market risk premium, in the multi-factor framework of the APT there is no such guarantee. Since it is the *product* of the betas and the risk premia which enters in the cross-section regression (2), their signs can in fact be reversed with no effect on the regression results. Moreover, their scaling can be changed arbitrarily provided that their product stays constant. Such issues of non-uniqueness of the estimates are implicit in factor analysis and pose a serious problem for an unambiguous test of the theory. For example, performing the two-stage testing procedure on different subgroups of stocks may yield different results depending on the order in which different risk factors are included in the first-stage regression.

Overall, however, such econometric ambiguities reflect the ambitious objective of simultaneously determining the risk factors and their associated risk premia. In financial practice, the complications implicit in this dual objective are often circumscribed in favor

of the approach outlined in section 4.3.1. In other words, if the analyst is constrained to follow a particular limited set of risk factors, then the betas are already defined and the testing reduces to the second-stage regression. In this restricted sense, then, the APT has proved to be of great value in accounting for variations in expected returns. For example, in lecture 3 we suggested that, over long time intervals and especially since the 1970s, the consensus predictions of the standard CAPM have yielded a smaller sensitivity of assets' expected returns to market risk than is actually seen in the data. This may be due to particular risk factors which are subsumed in the (unobservable) market portfolio.

A typical such illustration of APT's strength lies in explaining the differences in average returns between small and large stocks. If small stocks have a smaller beta, then--according to the CAPM--they should also have a lower expected return. However, in several markets, including Japan, we observe that small stocks and their indexes outperform large stocks on average, thus contradicting the CAPM prediction. The explanation lies in particular risk factors which affect small stocks but not large stocks, and which therefore are compensated with higher expected return. Such influences can be easily tracked and evaluated in the APT framework.