P van Rensburg*

Investment Basics: XXXIV. The arbitrage pricing theory

1. INTRODUCTION

The APT was developed by Stephen Ross in the early 70s and first published in 'The Arbitrage Theory of Capital Asset Pricing' (1976). The APT is based on fewer and less restrictive assumptions than the more familiar capital asset pricing model (CAPM) of Sharpe (1964), Lintner (1965) and Mossin (1966). It is also a more general model than the CAPM allowing for more than one risk factor to underlie stock returns. However, the primary challenge in applying the APT is discerning the identity of the 'priced' factors operational on a particular equity market over a particular period of time.

2. THEORETICAL OVERVIEW

The APT is based on the following assumptions:

(i) Markets are perfectly competitive and frictionless
(ii) Investors prefer more wealth to less wealth and are risk averse
(iii) Individuals share the belief that (for the set of assets being considered) the stochastic process underlying the generation of security returns over time can be simplified in the form of the following linear k-factor model:

$$R_{it} = E(R_{it}) + \sum_{k=1}^{K} b_{ik} f_{kt} + \varepsilon_{it} \qquad \text{... (1)}$$

where:

R_{it} = realised returns earned by asset i in time period t, where i = 1,2 ... n and t = 1,2 ... T

$E(R_{it})$ = the expected rate of return of asset i for period t at the beginning of period t

f_{kt} = the kth risk factor that impacts on asset i's returns, where k = 1,2 ... K. All risk factors represent unexpected movements in pervasive economic forces and have an expected value of zero i.e. $E(f_{kt}) = 0$.

b_{ik} = a coefficient that measures the sensitivity of R_{it} to movements in f_{kt}

ε_{it} = a normally distributed random error term t which measures the unexplained residual return of asset i in period t, where $E(\varepsilon_{it}) = 0$; $E(\varepsilon_{it}\varepsilon_{jt})=0$ for all $i \neq j$ and $E(\varepsilon_{it} f_{kt})=0$.

In order to expand on the intuitions underlying the theory, assumption (iii) above is elaborated upon: The *ex post* return of a traded security may be divided into two components. First the 'required' or expected return, $E(R_{it})$, is that portion which shareholders anticipate or predict. It represents the embodiment of all information shareholders have access to regarding the asset in question. The uncertain or risky return (U_{it}) represents the portion arising from surprises or unanticipated information. In light of the above the realised returns of a security can be expressed quite simply as:

$$R_{it} = E(R_{it}) + U_{it} \qquad \text{... (1a)}$$

Applying the familiar distinction between systematic (undiversifiable) and unsystematic (diversifiable) risk allows the following decomposition of equation 1a:

$$R_{it} = E(R_{it}) + S_{it} + \varepsilon_{it} \qquad \text{... (1b)}$$

where:

S_{it} = the (unexpected) returns of asset i generated in period t as a result of systematic risk factors

ε_{it} = the (unexpected) returns of asset i in period t as a result of unsystematic risk factors

"At the core of the APT is the recognition that only a few systematic factors affect the long term average returns of financial assets" (Roll and Ross, 1984, 15). Such a view does not deny the existence of the myriad of interlinked forces which underlie the return realised on a financial asset. A particular security's return may have statistically significant sensitivities to a number of factors e.g. the platinum price, the status of the firm's trade union relationships, the magnitude of competitor's sales etc. However, the APT stipulates that only a few of these factors are consistently of significance to the manager of a broad multi-asset portfolio. Those sensitivities which cannot be costlessly diversified away will be 'priced' in the presence of risk averse investors i.e. investors will require a premium (in the form of higher expected returns) in order to bear this type of risk. Thus, it is assumed that the systematic portion of an asset's unexpected returns can be decomposed into the cumulative contribution of a limited number of significant factors:

$$S_{it} = \sum_{k=1}^{K} b_{ik} f_{kt} \qquad \text{... (1c)}$$

*Department of Accounting and Finance, University of Natal, King George V Avenue, Durban 4001, Republic of South Africa.

where:

f_{kt} = the kth risk factor that impacts on assets' or portfolio returns, where k = 1,2 ...K and $E(f_{kt}) = 0$
b_{ik} = a coefficient that measures the sensitivity of R_{it} to a unit movement in f_{kt} *ceteris paribus.*

Substituting 1c into 1b allows an asset's *ex post* returns to be expressed as the linear model mentioned above:

$$R_{it} = E(R_{it}) + \sum_{k=1}^{K} b_{ik} f_{kt} + \varepsilon_{it} \quad \text{... (1)}$$

Equation 1 merely provides a simplification of the return generating process underlying equity returns and is void of economic theory. Important in its construction are the assumptions that K<n and that the relationships between the priced factors and returns can be simplified to a linear one. Note that equation 1 is not an equilibrium pricing relationship as no restrictions are placed on the constant (representing $E(R_{it})$) of this model. This allows two assets with identical sensitivities (b_{ik}s) to have differing expected returns (Sharpe, 1984, 23).

The central intuition of the APT is that, all portfolios that can be constructed from the set of assets under consideration that satisfy the conditions of (a) using no wealth and (b) having no risk, must also earn no return on average. Portfolios that satisfy conditions (a) and (b) are termed "arbitrage portfolios" and can be formally presented as having the following properties:

$$\sum_{i=1}^{n} w_i = 0 \quad \text{... (2a)}$$

where

w_i = the market value weighting of asset i

The contention that the arbitrage portfolio uses no wealth is based on the assumption that (overpriced) securities are sold short in order to obtain a cash inflow (i.e. $w_o < 0$). The arbitrageur purchases 'identical' (underpriced) securities with the proceeds of these sales (i.e. $w_u > 0$). It is assumed that investors invest all the proceeds from their sale of assets into new assets. With regard to condition (b), an arbitrage portfolio is perfectly hedged against risk due to the fact that any gains or losses realised in selling the first asset will be exactly counterbalanced by the purchase of the second asset. This is clear if it is considered that under arbitrage situations assets 1 and 2 will be perfect substitutes. This condition can be more formally stated as:

$$\sum_{k=1}^{K} w_i b_{ik} = 0 \quad \text{33... (2b)}$$

Finally, the no positive expected returns condition can be stated as:

$$E(R_{pt}) = \sum_{i=1}^{n} w_i E(R_{it}) = 0 \quad \text{... (2c)}$$

The economic reason why a portfolio described by equations 2a and 2b cannot have a positive return can be ascribed to the law of one price. If arbitrage opportunities exist, the timely action of arbitrageurs will quickly result in the price of the underpriced asset being bid up and the overpriced asset being bid down. The taking of opportunities to make riskless profits by relatively few extremely alert investors quickly precludes others from following suit. "Two portfolios with the same sensitivity to each systematic factor are very close substitutes ... Consequently, they must offer the investor the same expected return, just as ...two Treasury bills or two shares of the same stock offer the same expected return" (Roll and Ross, 1984, 15). One of the strengths of the APT is the reasonablen[1]ess of the no-arbitrage assumption. Unlike the CAPM there is no appeal to mean-variance maximising behaviour on the behalf of investors nor is there any special role for the unobservable 'market' portfolio[1]. However, in order for riskless arbitrage opportunities to exist it is also necessary that all non-factor risk can be removed from these portfolios. In his original paper Ross (1976) appealed to the law of large numbers to diversify away all non-factor risk in the arbitrage portfolio. A number of variations of the APT exist, differing primarily in the 'mechanism' employed to ensure that $\varepsilon_{pt} \approx 0$.[2] "The 'Arbitrage' in the 'Arbitrage Pricing Theory' comes from the simple proof[3] that [equation 1] implies [equation 3

---

[1]Copeland and Weston (1983) provide a list of the differences between the APT and the CAPM.

[2]Dybvig and Ross (1985) provide an overview of the alternatives developed up to that time.

[3]As the derivation uses a result in linear algebra it is relegated to a footnote. Converting equations 2a to 2c into matrix notation:

$w'I = 0$ (2a')

where:

$w = (w_1, \dots, w_n)'$ and I is a n dimensional column vector of ones

$w'b = 0$ (2b')

where:

$b = (b_1, \dots b_n)'$ and $b_i = (b_{i1}, \dots, b_{ik})$ for i = 1, ... ,n

$w'E = 0$ (2c')

where:

$E = (E(R_1), \dots, E(R_n)'$

The no-arbitrage condition is that the validity of equations 2a' and 2b' imply the validity of 2c'. This is the algebraic statement that if a vector of asset weightings is orthogonal to a vector of ones and K vectors of sensitivity coefficients, this implies that the vector of weightings is also orthogonal to the vector of expected returns. There is a theorem in linear algebra that states that if the fact that a vector is orthogonal to N-1 vectors implies it is orthogonal to the Nth vector, then the Nth vector must a linear combination of the N-1 vectors. In other words there exist constants λ_0 and λ such that below] using the absence of arbitrage alone" (Dybvig and Ross, 1985:1174, see also Roll and Ross, 1980).

$$E(R_{it}) = \lambda_{0t} + \sum_{K=1}^{K} b_{ik}\lambda_k \qquad \dots(3)$$

A natural interpretation of equation 3 is that λ_{0t} = the return on a risk free asset (R_{ft}) typically proxied by the 3 month treasury bill. If a true risk free asset does not exist λ_{0t} = the return on a hypothetical asset with a zero sensitivity to all factors (i.e. $b_{i1} = b_{i2} = \dots b_{ik} = 0$). λ_k represents the risk premium required in order to induce investors to bear a unit sensitivity to factor k. The intuition behind equation 3 is relatively clear. Each factor represents an independent source of risk that cannot be diversified away. The expected or required return of an asset is equal to the risk free rate plus the compensation for each type of (undiversifiable) risk that the security bears. Substituting the expression for $E(R_{it})$ from the fundamental APT pricing relation (equation 3) into the linear factor model (equation 1) we obtain what Berry, Burmeister and McElroy, 1988:31 call the "full APT":

$$R_{it} = \lambda_{0t} + \sum_{k=1}^{K} b_{ik}\lambda_k + \sum_{k=1}^{K} b_{ik} f_{kt} + \varepsilon_{it} \qquad \dots(4)$$

When compared to equation 1, it can be seen that restrictions have been placed on the intercept term of the linear factor model. These restrictions embody and, indeed, are a direct algebraic consequence of, the no-arbitrage conditions. Note that econometric estimation of equation 1 would not be the estimation of an APT model because the absence of arbitrage profits constraints are not applied to the linear factor model.

3. GRAPHICAL DEPICTION

Having algebraically derived the fundamental APT pricing relationship (equation 3), it can be graphically depicted in risk-(expected) return space. For simplicity, it is initially assumed that only a single priced factor underlies an asset's returns. As can be seen in figure 1 the slope of the arbitrage pricing line represents the risk premium associated with a particular factor and the intercept, R_f, is the risk free rate.

$$E(R_i) = \lambda_0 + b_i\lambda$$

For each asset i where:

$$\lambda = (\lambda_1, \dots, \lambda_k)'$$

Which, restated in scalar form is:

$$E(R_{it}) = \lambda_{0t} + \sum_{k=1}^{K} b_{ik}\lambda_k$$

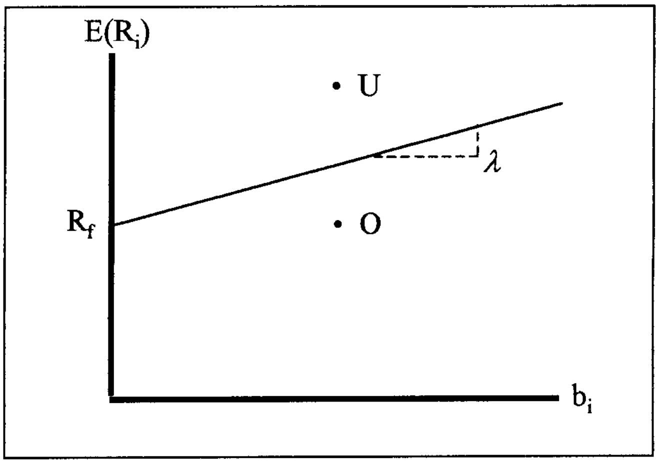

Figure 1: The arbitrage pricing line

It can be re-emphasized that it is the absence of arbitrage opportunities which is responsible for ensuring that all assets fall on this line. For example, should there be an overvalued security (labelled O) and an undervalued security (labelled U), this situation could be exploited by an investor holding an arbitrage portfolio. He or she would do so by selling asset O short and simultaneously buying a long position of equal monetary value in U. Note that no new investment would be necessary as funds are merely 'translocated' between the two assets. In addition, such a transaction bears no risk and gains an engineered profit of $E(R_u) - E(R_o)$. The increased demand for asset U and the desire to sell asset O embodied in such transactions will serve to bring both prices back to equilibrium (or, graphically, back on to the arbitrage pricing line).

The graphical depiction of a multi-factor arbitrage pricing 'line' is similar in concept to the single factor model considered above. In figure 2 the arbitrage pricing plane for a two factor model is portrayed. The sensitivity of asset i's returns to factor 1 and 2 are represented by b_{i1} and b_{i2} along the z and x axes respectively. Figure 2 may be extended to depict APT specifications employing 3 or more factors. Such specifications will be characterised by a pricing 'hyper-plane'. However, graphical depictions of such models in two dimensional space are not easily constructed or interpreted.

4. THE IDENTITY OF THE APT FACTORS

Unfortunately the APT, itself, does not reveal the identity of its priced factors. Indeed, there are good reasons to suggest that the number and nature of these factors is likely to change over time and between economies. As a result, this issue is essentially empirical in nature. However, the following *a priori* guidelines as to the characteristics required of potential APT factors can be deduced:

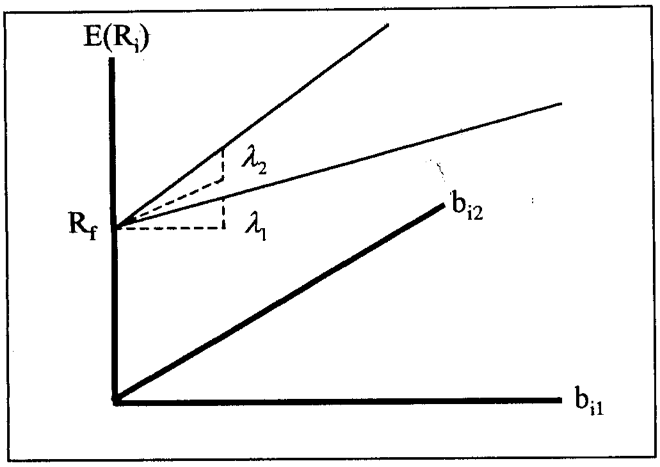

Figure 2: The two factor arbitrage pricing plane

The defining characteristic of a priced factor is its association with a significantly non-zero risk premium. This is the issue that is focussed on by empiricists aiming to identify priced factors. Characteristics that should be exhibited by 'candidate' factors are that they should represent: (i) unexpected movements in (ii) undiversifiable influences (which are, hence, more likely to be macroeconomic rather than firm specific in nature (see Chen, Roll and Ross, 1986). For practical purposes (iii) timeous and accurate information on these variables is required. In addition, (iv) the variable should be theoretically justifiable on economic grounds (Berry, Burmeister and McElroy, 1988:29 and Alexander, Sharpe and Bailey, 1990:252-253). It is also relevant to note Berry, Burmeister and McElroy, (1988:30-31) "... there is no one 'correct' set of factors; there are many equivalent sets of correct factors, all of which give rise to equivalent empirical results. Intuitively a factor such as 'unexpected changes in the money supply' might work as well as the factor 'unexpected inflation'; a set of factors with unexpected change in the money supply substituted for unexpected inflation would give equivalent results. The choice of a set of 'correct' factors can be made on empirical grounds: The factors should adequately explain asset returns; they should pass the statistical tests necessary to qualify as legitimate APT factors; the actual asset returns should exhibit plausible sensitivities to the realisations of these factors; and the factors should have non zero APT prices...".

One manner in which to frame the theoretical justification of a 'candidate' factor is within the Gordon-Shapiro constant growth dividend discount model:

$$P_0 = \frac{D_1}{k-g} \quad \ldots (5)$$

where:

P_0 = share price at time 0

D_1 = expected dividend at time 1 = $D_0(1+g)$

k = an appropriate discount rate

g = the expected (constant) growth rate of dividends, where $g < k$ else $P_0 = \infty$

It can be argued that any variable that influences the magnitude of D_0, k or g is instrumental in explaining price levels (see also Chen, Roll and Ross, 1986, 384-385 and Burmeister and Wall, 1986, 1-7). It is reasonable to conjecture that the level of current dividends is related to measures of the magnitude of current earnings and broad measures of economic output. As financial securities are claims against future output, expectations regarding future levels of GNP will be of concern to the investor. These expectations are manifested in g. Unfortunately it is difficult to measure perceptions directly. However, there are *a priori* reasons to believe that those macroeconomic variables that reflect perceptions (such as other 'market prices' e.g. bond rates, the term structure of interest rates, foreign exchange rates and international equity prices) are likely to be related to equity values. Also note that the dividend payout ratio is likely to bridge a close inverse relationship between D_0 and g, where lower current dividend payout ratios increase the portion of earnings retained to finance the future growth in dividends.The discount rate k is likely to be intimately related to interest rate levels. However, unlike government bonds, in the case of equity securities the timing and magnitude of the (nominal) cash-flows associated with ownership of the asset are not certain. Consequently, risk perceptions regarding the frequency and magnitude of future dividend payments and the degree of risk aversion exhibited by investors are also likely to be determinants of k. However, notwithstanding the value of the above theoretical speculation, the idenfication of those sources of risk that explain the cross-section of returns is an empirical issue.

REFERENCES

Alexander GJ, Sharpe WF and Bailey JV. 1993. *Fundamentals of Investments*. Prentice Hall.

Burmeister E and Wall KD. 1986. The arbitrage pricing theory and macroeconomic factor measures. *The Financial Review*, 21:1-20.

Chen N, Roll R and Ross SA. 1986. Economic forces and the stock market. *Journal of Business*, 59:383-403.

Copeland TE and Weston JF. 1983. *Financial Theory and Corporate Policy*. Addison Wesley.

Dybvig PH and Ross SA. 1985. Yes, the APT is testable. *Journal of Finance*, 40:1173-1188.

Francis JC. 1986. *Investments: Analysis and Management*, (fourth ed.), McGraw Hill.

Lintner J. 1965. The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47:13-47

Mossin J. 1966. Equilibrium in a capital asset market. *Econometrica*, 34:768-783.

Roll R and Ross SA. 1980. An empirical investigation of the arbitrage pricing theory. *Journal of Finance*, 35:121-130.

Roll R and Ross SA. 1984. The arbitrage pricing theory approach to strategic portfolio planning. *Financial Analysts Journal*, May-Jun:14-26.

Ross SA. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13:341-360.

Ross SA. 1977. Return, risk and arbitrage. *Risk and Return in Finance*. Friend I and Bicksher J (ed.). Balleringer.

Sharpe WF. 1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, 19:425-442.