

Chapter 10

The APT model

10.1 Introduction

Although the APT stands for 'Arbitrage Pricing Theory' the model presented here is somewhat different from the arbitrage models presented earlier. The framework is in one sense closer to CAPM in that we consider a one-period model with risky assets whose distributions may be continuous. On the other hand, there is also a clear analogue with arbitrage pricing. We will present the basic idea of the model in two steps which will illustrate how the restriction on mean return arises.

10.2 Exact APT with no noise

Consider the following model for the returns r_1, \dots, r_N of N risky assets:

$$r = \mu + Bf$$

where $r = (r_1, \dots, r_N)^\top$ is a vector of random returns, $\mu = (\mu_1, \dots, \mu_N)^\top$, and B is an $N \times K$ -matrix whose entries are real numbers, and $f = (f_1, \dots, f_K)$ is a vector of random variables (*factors*) which satisfies

$$\begin{aligned} Ef_i &= 0, & i = 1, \dots, K, \\ Cov(f) &= \Phi, & \Phi \text{ positive definite.} \end{aligned}$$

Note that this means that $Er_i = \mu_i$, $i = 1, \dots, N$. We will assume that $N > K$, and you should think of the number of assets N as being much larger than the number of factors K . The model then seeks to capture the idea that returns on assets are correlated through a common dependence on a (small) number of factors. The goal is to use the assumption of such a

common dependence to say something about the vector of mean returns μ . Assume that there is also a riskless asset with return r_0 . When we talk about a portfolio, w , we mean a vector in \mathbb{R}^N where the i th coordinate measures the relative share of total wealth invested in the i th risky asset, and the rest in the riskfree asset. (So the term 'investment strategy given by w ' would probably be better). Hence the coordinates need not sum to 1, and the expected rate of return on w is

$$\begin{aligned} E((1 - w^\top \mathbf{1})r_0 + w^\top r) &= r_0 + E(w^\top (r - r_0 \mathbf{1})) \\ &= r_0 + w^\top (\mu - r_0 \mathbf{1}), \end{aligned} \quad (10.1)$$

where as usual $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^N$.

Since $N > K$ it is possible to find a portfolio $w \in \mathbb{R}^N$ of risky assets which is orthogonal to the column space of B . This we will write as $w \in (B)^\perp$. Now the mean return is $r_0 + w^\top (\mu - r_0 \mathbf{1})$ and by using the "covariance matrix algebra rules" in the first part of Chapter 9 we see that the variance of the return on this portfolio is given as

$$\begin{aligned} V(w^\top r) &= \text{Cov}(w^\top Bf, w^\top Bf) \\ &= w^\top B \Phi B^\top w = 0. \end{aligned}$$

A reasonable no arbitrage condition to impose is that a portfolio consisting only of risky assets which has zero variance should earn the same return as the riskless asset. Hence the following implication should hold in an arbitrage free market:

$$w^\top \mathbf{1} = 1, w \in (B)^\perp : w^\top \mu = r_0 \iff w^\top (\mu - r_0 \mathbf{1}) = 0.$$

By scaling we see that

$$w^\top \mathbf{1} \neq 0, w \in (B)^\perp : w^\top (\mu - r_0 \mathbf{1}) = 0.$$

By using the same "arbitrage reasoning" on (10.1) we get that

$$w^\top \mathbf{1} = 0, w \in (B)^\perp : w^\top \mu = 0 \iff w^\top (\mu - r_0 \mathbf{1}) = 0.$$

From these two statements we see that any vector which is orthogonal to the columns of B is also orthogonal to the vector $(\mu - r_0 \mathbf{1})$, and this implies¹

¹If you prefer a mathematical statement, we are merely using the fact that

$$(B)^\perp \subset (\mu - r_0 \mathbf{1})^\perp \implies (\mu - r_0 \mathbf{1}) \in (B).$$

that $\mu - r_0\mathbf{1}$ is in the column span of B . In other words, there exists $\lambda = (\lambda_1, \dots, \lambda_K)$ such that

$$\mu - r_0\mathbf{1} = B\lambda. \quad (10.2)$$

The vector $\lambda = (\lambda_1, \dots, \lambda_K)$ is called the vector of *factor risk premia* and what the relation tells us is that the excess return is obtained by multiplying the *factor loadings* with the factor risk premia. This type of conclusion is of course very similar to the conclusion of CAPM, in which there is 'one factor' (return on the market portfolio), β_{im} plays the role of the factor loading and $Er_m - r_0$ is the factor risk premium.

10.3 Introducing noise

We continue the intuition building by considering a modification of the model above. Some of the reasoning here is heuristic - it will be made completely rigorous below.

Assume that

$$r = \mu + Bf + \epsilon$$

where r, μ, B and f are as above and $\epsilon = (\epsilon_1, \dots, \epsilon_N)$ is a vector of random variables (*noise terms* or *idiosyncratic risks*) satisfying

$$\begin{aligned} E\epsilon_i &= 0, & i &= 1, \dots, N, \\ \text{Cov}(\epsilon_i, f_j) &= 0, & i &= 1, \dots, N, \quad j = 1, \dots, K, \\ \text{Cov}(\epsilon) &= \sigma^2 I^N, & I^N & \text{is the } N \times N \text{ identity matrix.} \end{aligned}$$

Clearly, this is a more realistic model since the returns are not completely decided by the common factors but 'company specific' deviations captured by the noise terms affect the returns also. However if the variance in the noise term is not too large then we can almost eliminate the variance arising from the noise term through diversification.

Since $N > K$ it is possible to find portfolios of risky assets v_1, \dots, v_{N-K} which are orthogonal and lie in $(B)^\perp$. Let a be the maximal absolute value of the individual portfolio weights. Now consider the portfolio

$$v = \frac{1}{N-K}(v_1 + \dots + v_{N-K}).$$

The variance of the return of this portfolio is

$$\begin{aligned} \text{V}(v^\top r) &= \text{V}(v^\top Bf + v^\top \epsilon) \\ &= 0 + \frac{1}{(N-K)^2} \text{V}((v_1 + \dots + v_{N-K})^\top \epsilon) \end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{(N-K)^2} (N-K) a^2 \sigma^2 \\ &= \frac{a^2 \sigma^2}{N-K}, \end{aligned}$$

where the inequality follows from the orthogonality of the v_i 's (and the definition of a).

If we think of N as very large, this variance is very close to 0, and - entering into heuristic mode - therefore the expected return of this portfolio ought to be close to that of the riskless asset:

$$E v^\top r = v^\top \mu \approx r_0. \quad (10.3)$$

By a slight modification of this argument it is possible to construct portfolios with good diversification which span $(B)^\perp$ and for each portfolio we derive a relation of the type (10.3). This then would lead us to expect that there exists factor risk premia such that

$$\mu - r_0 \mathbf{1} \approx B \lambda.$$

The precise theorem will be given below.

10.4 Factor structure in a model with infinitely many assets

In this section we present a rigorous version of the APT.

Given is a riskless asset with return r_0 and an infinite number of risky assets with random returns (r_1, r_2, \dots) . We will use the following notation repeatedly: If $x = (x_1, x_2, \dots)$ is an infinite sequence of scalars or random variables, then x^N is the column vector consisting of the first N elements of this sequence. Hence $r^N = (r_1, \dots, r_N)^\top$.

Definition 41 *The returns (r_1, r_2, \dots) are said to have an approximate factor structure with factors (f_1, \dots, f_K) if for all N*

$$r^N = \mu^N + B^N f + \epsilon^N$$

where B^N is the N first rows of a matrix B with infinitely many rows and K columns, where B^N has rank K for N large,

$$\begin{aligned} E \epsilon_i &= 0, & i &= 1, 2, \dots, \\ \text{Cov}(\epsilon_i, f_j) &= 0, & i &= 1, 2, \dots, \quad j = 1, \dots, K, \\ \text{Cov}(f) &= I^K & & \text{(the } K \times K \text{ identity matrix)} \\ \text{Cov}(\epsilon^N) &= \Omega^N & & \text{(\Omega}^N \text{ is positive definite)} \end{aligned}$$

and where the eigenvalues of Ω^N are bounded uniformly in N by a constant $\bar{\lambda}$.

In other words, the same K factors are governing the returns on an infinite collection of securities except for noise terms captured by ϵ which however are uniformly of small variance. The simplest case would be where the elements of ϵ are independent² and have variance less than or equal to σ^2 in which case $\bar{\lambda} = \sigma^2$. Although our definition is slightly more general you can think of each element of ϵ as affecting only a finite number of returns and factors as affecting infinitely many of the returns.

The assumption that the covariance matrix of the factors is the identity may seem very restrictive. Note however, that if we have a structure of the form

$$r = \mu + Bf + \epsilon$$

which satisfies all the requirements of the definition of an approximate factor structure with the only exception being that

$$\text{Cov}(f) = \Phi \quad (\Phi \text{ is a positive definite, } K \times K\text{-matrix})$$

then using the representation $\Phi = CC^\top$ for some invertible $K \times K$ matrix we may choose g such that $Cg = f$. Then we have

$$r = \mu + BCg + \epsilon$$

and then this will be an approximate factor structure with g as factors and BC as factor loadings. To verify this note that

$$\text{Cov}(g) = C^{-1}CC^\top(C^{-1})^\top = I^K.$$

Hence in one sense nothing is lost by assuming the particular structure of f . We may represent the same distribution of returns in this way as if we allow a general positive definite matrix to be the covariance matrix of the factors. However, from a statistical viewpoint the fact that different choices of parameters may produce the same distributions is a cause for alarm. This means that we must be careful in saying which parameters can be identified when estimating the model: Certainly, no observations can distinguish between parameters which produce the same distribution for the returns. We will not go further into these problems and to discussions of what restrictions can be imposed on parameters to ensure identification. We now need to introduce a modified notion of arbitrage:

²In this case the returns are said to have a *strict factor structure*.

Definition 42 *An asymptotic arbitrage opportunity is a sequence of portfolios (w^N), where $w^N \in \mathbb{R}^N$, in the risky assets which satisfies*

$$\lim_{N \rightarrow \infty} E(w^N \cdot r^N) = \infty$$

and

$$\lim_{N \rightarrow \infty} V(w^N \cdot r^N) = 0.$$

The requirement that expected return goes to infinity (and not just some constant greater than the riskless return) may seem too strong, but in the models we consider this will not make any difference.

The theorem we want to show, which was first stated by Ross and later proved in the way presented here by Huberman, is the following:

Theorem 34 *Given a riskless asset with return $r_0 > -1$ and an infinite number of risky assets with random returns (r_1, r_2, \dots) . Assume that the returns have an approximate factor structure. If there is no asymptotic arbitrage then there exists a vector of factor risk premia $(\lambda_1, \dots, \lambda_K)$ such that we have*

$$\sum_{i=1}^{\infty} (\mu_i - r_0 - \lambda_1 b_{i1} - \dots - \lambda_K b_{iK})^2 < \infty. \quad (10.4)$$

This requires a few remarks: The content of the theorem is that the expected excess returns of the risky assets are in a sense close to satisfying the exact APT-relation (10.2): The sum of the squared deviations from the exact relationship is finite. Note that (unfortunately) this does not tell us much about the deviation of a particular asset. Indeed the mean return of an asset may show significant deviation from (10.2). This fact is crucial in understanding the discussion of whether the APT is a testable model.

The proof must somehow use the same arbitrage argument as in the case with exact APT above by getting rid of the noise terms through diversification. Although this sounds easy, we discover once again that 'the devil is in the details'. To do the proof we will need the following two technical lemmas:

Lemma 35 *Let Ω be a symmetric positive definite $N \times N$ matrix and let $\bar{\lambda}$ be its largest eigenvalue. Then*

$$w^\top \Omega w \leq \bar{\lambda} \|w\|^2.$$

Proof Let (v_1, \dots, v_N) be an orthonormal set of eigenvectors of Ω and $(\lambda_1, \dots, \lambda_N)$ the corresponding eigenvalues. There exist $\alpha_1, \dots, \alpha_N$ such that $w = \sum_1^N \alpha_i v_i$ and hence

$$\begin{aligned} w^\top \Omega w &= \sum_1^N \alpha_i^2 \lambda_i v_i^\top v_i \\ &= \sum_1^N \alpha_i^2 \lambda_i \\ &\leq \bar{\lambda} \sum_1^N \alpha_i^2 = \bar{\lambda} \|w\|^2. \quad \blacksquare \end{aligned}$$

Lemma 36 *Let X be a compact set. Let $(K_i)_{i \in I}$ be a family of closed subsets of X which satisfy the finite intersection property*

$$\bigcap_{i \in I_0} K_i \neq \emptyset \quad \text{for all finite subsets } I_0 \subset I.$$

Then the intersection of all sets is in fact non-empty. i.e.

$$\bigcap_{i \in I} K_i \neq \emptyset.$$

Proof If $\bigcap_{i \in I} K_i = \emptyset$, then $\bigcup_{i \in I} K_i^c$ is an open covering of X . Since X is compact the open cover contains a finite subcover $\bigcup_{i \in I_0} K_i^c$, but then we apparently have a finite set I_0 for which $\bigcap_{i \in I_0} K_i = \emptyset$, and this violates the assumption of the theorem. \blacksquare

The proof is in two stages. First we prove the following:

Proposition 37 *Given a riskless asset with return $r_0 > -1$ and an infinite number of risky assets with random returns (r_1, r_2, \dots) . Assume that the returns have an approximate factor structure. If there is no asymptotic arbitrage then there exists a sequence of factor risk premia vectors (λ^N) , $\lambda^N \in \mathbb{R}^K$, and a constant A such that for all N*

$$\sum_{i=1}^{\infty} (\mu_i - r_0 - \lambda_1^N b_{i1} - \dots - \lambda_K^N b_{iK})^2 \leq A. \quad (10.5)$$

Proof Let us WLOG assume B^N has rank K for all N . Consider for each N the regression of the expected excess returns on the columns of B^N , i.e. the $\lambda^N \in \mathbb{R}^K$ that solves

$$\min_{\lambda} (\mu^N - r_0 \mathbf{1} - B^N \lambda^N)^\top (\mu^N - r_0 \mathbf{1} - B^N \lambda^N) = \min \|c^N\|^2$$

where the *residuals* are defined by

$$c^N = \mu^N - r_0 \mathbf{1}^N - B^N \gamma^N$$

By (matrix)-differentiating we get the first order conditions

$$(B^N)^\top c^N = 0$$

But the $K \times K$ matrix $(B^N)^\top B^N$ is invertible (by our rank K assumption), so the unique solution is

$$\lambda^N = ((B^N)^\top B^N)^{-1} (B^N)^\top (\mu^N - r_0 \mathbf{1}).$$

We also note from the first order condition that the residuals c^N are orthogonal to the columns of B^N . To reach a contradiction, assume that there is no sequence of factor risk premia for which (10.5) holds. Then we must have $\|c^N\| \rightarrow \infty$ (since $\|c^N\|^2$ is the left hand side of (10.5) with summation to N). Now consider the sequence of portfolios given by

$$w^N = \|c^N\|^{-\frac{3}{2}} c^N.$$

The expected excess return is given by

$$\begin{aligned} E(w^N \cdot (r^N - r_0 \mathbf{1}^N)) &= E(w^N \cdot (\mu^N - r_0 \mathbf{1}^N + B^N f + \epsilon^N)) \\ &= E(w^N \cdot (B^N \gamma^N + c^N + B^N f + \epsilon^N)) \\ &= \|c^N\|^{-\frac{3}{2}} c^N \cdot c^N \\ &= \|c^N\|^{\frac{1}{2}} \rightarrow \infty \text{ as } N \rightarrow \infty. \end{aligned}$$

where in the third equality we used the fact that c^N is orthogonal to the columns of B and that both factors and noise terms have expectation 0. The variance of the return on the sequence of portfolios is given by

$$\begin{aligned} V(w^N \cdot (r^N - r_0 \mathbf{1}^N)) &= V(w^N \cdot (\mu^N - r_0 \mathbf{1}^N + B^N f + \epsilon^N)) \\ &= V(w^N \cdot \epsilon^N) \\ &= \|c^N\|^{-3} (c^N)^\top \Omega^N c^N \\ &\leq \|c^N\|^{-3} \bar{\lambda} \|c^N\|^2 \rightarrow 0 \text{ as } N \rightarrow \infty \end{aligned}$$

where we have used Lemma 35 and the same orthogonality relations as in the expected return calculations. Clearly, we have constructed an asymptotic arbitrage opportunity and we conclude that there exists a constant A and a sequence of factor risk premia such that

$$\sum_{i=1}^{\infty} (\mu_i - r_0 - \lambda_1^N b_{i1} - \cdots - \lambda_K^N b_{iK})^2 \leq A. \quad \blacksquare$$

Now we are ready to finish.

Proof of Theorem 34 Let A be as in the proposition above. Consider the sequence of sets (H^N) where

$$H^N = \left\{ \lambda \in \mathbb{R}^K : \sum_{i=1}^N (\mu_i - r_0 - \lambda_1 b_{i1} - \cdots - \lambda_K b_{iK})^2 \leq A \right\}.$$

By the preceding proposition, each H^N is non-empty and clearly $H^{N+1} \subset H^N$. Define the functions $f^N : \mathbb{R}^K \mapsto \mathbb{R}$ by

$$f^N(\lambda) = (\mu - r_0 \mathbf{1} - B\lambda)^\top (\mu - r_0 \mathbf{1} - B\lambda) = \|\mu - r_0 \mathbf{1}\|^2 + (\mu - r_0 \mathbf{1})^\top B\lambda + \lambda^\top B^\top B\lambda,$$

where some of the N -superscripts have been dropped for the ease of notation. Then f^N is a convex function (because $B^\top B$ is always positive semidefinite), and we see that

$$H^N = \{ \lambda \in \mathbb{R}^K : f(\lambda) \leq A \}$$

is a closed convex set. Now pick an N so large that B has rank K . To show that H^N is then compact, it suffices (by convexity) to show that for all nonzero $\lambda \in H^N$ there exists a scaling factor (a real number) a such that $a\lambda \notin H^N$. But since B has full rank, there is no nonzero vector (in \mathbb{R}^K) that is orthogonal to all of B 's (N) rows. Hence for an arbitrary nonzero $\lambda \in H^N$ we have that $\|B\lambda\| \neq 0$ and

$$f^N(a\lambda) = \|\mu - r_0 \mathbf{1}\|^2 + a(\mu - r_0 \mathbf{1})^\top B\lambda + a^2 \|B\lambda\|^2,$$

so by choosing a large enough a we go outside H^N , so H^N is not compact. Then we may use Lemma 36 to conclude that

$$\bigcap_{N=1}^{\infty} H^N \neq \emptyset.$$

Any element $\lambda = (\lambda_1, \dots, \lambda_K)^\top$ of this non-empty intersection will satisfy 10.4. \blacksquare

