CHAPTER 14

The Arbitrage Pricing Theory

Chapter Outline

14.1 Introduction

We have already made two attempts (Chapters 11–13) at asset pricing from an arbitrage perspective, i.e., without specifying a complete equilibrium structure. Here we try again from a different, more empirical angle. Before doing so, let us first collect a few thoughts as to the differences between an arbitrage approach and equilibrium modeling.

In the context of general equilibrium theory, we make hypotheses about agents—consumers, producers, investors; in particular, we start with some form of rationality hypothesis leading to the specification of maximization problems under constraints. We also make hypotheses about markets: typically, we assume that supply equals demand in all markets under consideration, and that the markets are competitive.

We have repeatedly used the fact that in general equilibrium with fully informed optimizing agents, there can be no-arbitrage opportunities—in other words, no possibilities to make money risklessly at zero cost. An arbitrage opportunity indeed implies that at least one agent can reach a higher level of utility without violating his or her budget constraint (since there is no extra cost to do so).

In particular, our assertion that one can price any asset (income stream) from the knowledge of Arrow–Debreu prices relied implicitly on a no-arbitrage hypothesis: with a complete set of Arrow–Debreu securities, some portfolio thereof will exactly replicate the asset's cash-flow stream. If there are to be no-arbitrage opportunities, the market price of the asset and the value of the replicating portfolio of Arrow–Debreu securities must therefore be the same. An arbitrageur could otherwise make arbitrarily large profits by short selling large quantities of the more expensive of the two and buying the cheaper in equivalent amounts. Such an arbitrage portfolio would have zero cost and be riskless.

While general equilibrium implies the no-arbitrage condition, it is more restrictive in the sense of imposing a heavier structure on modeling. And the reverse implication is *not* true: no-arbitrage opportunities—the fact that all arbitrage opportunities have been exploited—does not imply that a general equilibrium in all markets has been obtained.[1] Nevertheless, or precisely for that reason, it is interesting to see how far one can go in exploiting the less restrictive hypothesis that no-arbitrage opportunities are left unexploited.

The underlying logic of the arbitrage pricing theory (APT), to be reviewed in this chapter is, in a sense, very similar to the fundamental logic of the Arrow–Debreu model, and it is very much in the spirit of a complete markets structure. It distinguishes itself in two major ways: first it replaces the underlying structure based on fundamental securities paying exclusively in a given state of nature with other fundamental securities exclusively remunerating some form of risk taking. More precisely, the APT abandons the analytically powerful, but empirically cumbersome, concept of states of nature as the basis for the definition of its primitive securities. It replaces it with the hypothesis that there exists a (stable) set of factors (random quantities) that are essential and exhaustive determinants of all asset returns. Factors may be macroeconomic in nature (e.g., the random growth rate of gross domestic product, GDP), or behavioral (e.g., the momentum factor to be considered shortly). Typically factors may be separated as the return distributions on specific financial assets. A primitive security, in a factor context, will then be defined as a security whose risk is exclusively determined by its association with one specific underlying risk factor and which is totally statistically independent of (unaffected by) any other risk factor. Its expected return premium thus represents compensation to investors for bearing the factor specific risk. The second difference from the Arrow–Debreu pricing of Chapter 9 is that the prices of the primitive securities are not derived from fundamentals—supply and

[1] An arbitrage portfolio is a self-financing (zero net-investment) portfolio. An arbitrage opportunity exists if an arbitrage portfolio exists that yields nonnegative cash flows in all states of nature and positive cash flows in some states (Chapter 11).

demand, themselves resulting from agents' endowments and preferences—but will be deduced empirically from observed asset returns without attempting to identify their underlying macroeconomic or behavioral determinants. Once the price of each fundamental security has been inferred from observed return distributions, the usual arbitrage argument applied to the pricing of complex securities will be made (in the spirit of Chapter 12).[2]

14.2 Factor Models: A First Illustration

The main building block of the APT is a factor model, the notion of which we introduced back in Chapter 2. As discussed previously, this is the structure that replaces the concept of states of nature. The motivation has been evoked before: states of nature are analytically powerful abstractions for valuation. In practice, however, they are difficult to work with and, moreover, often not verifiable, implying that contracts cannot necessarily be written contingent on a specific state of nature being realized. We discussed these shortcomings of the Arrow–Debreu pricing theory in Chapter 9. The temptation is thus irresistible to attack the asset pricing problem from the opposite angle and build the concept of primitive securities on an empirically more operational notion, at a cost of abstracting from its potential theoretical credentials. This structure is what factor models accomplish.

The simplest conceivable factor model is a one-factor market model, which asserts that *ex post* returns on individual assets can be entirely ascribed either to their own specific stochastic components or to their common association with a single factor. This simple factor model can thus be summarized by following the relationship:[3]

$$\tilde{r}_j = \alpha_j + \beta_j \tilde{f}_1 + \tilde{\varepsilon}_j \tag{14.1}$$

with $E\tilde{\varepsilon}_j = 0$, $\text{cov}(\tilde{f}_1, \tilde{\varepsilon}_j) = 0$, $\forall j$, and $\text{cov}(\tilde{\varepsilon}_j, \tilde{\varepsilon}_k) = 0$, $\forall j \neq k$.

This model states that there are three components in individual returns: (1) an asset-specific constant α_j; (2) a common influence, in this case the unique factor, $\tilde{f}_1$, a random quantity, which affects all assets in varying degrees, with β_j measuring the sensitivity of asset j's return to fluctuations in the single factor; and (3) an asset-specific stochastic term $\tilde{\varepsilon}_j$ summarizing all other stochastic components of $\tilde{r}_j$ that are unique to asset j. Equation (14.1) has no bite (such an equation can always be written) until one adds the hypothesis $\text{cov}(\tilde{\varepsilon}_j, \tilde{\varepsilon}_k) = 0$, $j \neq k$, which signifies that *all* return characteristics common to different assets are subsumed in their link with the factor $\tilde{f}_1$.

2 The arbitrage pricing theory was first developed by Ross (1976) and substantially interpreted by Huberman (1982) and Connor (1984) among others. For a presentation emphasizing practical applications, see Burmeister et al. (1994).

3 Factors are frequently measured as deviations from their mean. When this is the case, α_j becomes an estimate of the mean return on asset j.

From our study of the capital asset pricing model (CAPM), it is natural to identify $\tilde{f}_1$ with $\tilde{r}_M$, the return on the market portfolio. With this identification, Eq. (14.1) is referred to as the "market model." If the market model were to be empirically verified, then the CAPM would be the undisputed endpoint of asset pricing. As the APT does not require the assumptions of the CAPM, it does not, *a priori*, identify $\tilde{f}_1$ with $\tilde{r}_M$.

From the empirical perspective, one may say that it is quite unlikely that a single-factor model will suffice.[4] But the strength of the APT is that it is agnostic as to the number of underlying factors (and to their identity). As we increase the number of factors, hoping that this will not require a number too large to be operational, a generalization of Eq. (14.1) becomes more and more plausible. For pedagogical purposes, let us for the moment maintain the hypothesis of one common factor which we identify with (some proxy for) the market portfolio.[5]

14.2.1 Using the Market Model

Besides serving as a potential basis for the APT, the market model, despite all its weaknesses, is also of interest on two other grounds. First it produces estimates for the β's that play a central role in the CAPM. Note, however, that estimating β's from past data alone is useful only to the extent that some degree of stationarity in the relationship between asset returns and the return on the market is present. Empirical observations suggest a fair amount of stationarity is plausible at the level of portfolios, but not of individual assets. On the other hand, estimating the β's does not require all the assumptions of the market model; in particular, a violation of the $\text{cov}(\tilde{\varepsilon}_i, \tilde{\varepsilon}_k) = 0,\ i \neq k$ hypothesis is not damaging.

The second source of interest in the market model, crucially dependent upon the latter hypothesis being approximately valid, is that it permits economizing on the computation of the matrix of variances and covariances of asset returns at the heart of the MPT. Indeed, under the market model hypothesis, one can write (you are invited to prove these statements):

$$\sigma_j^2 = \beta_j^2 \sigma_M^2 + \sigma_{\varepsilon_j}^2, \ \forall j$$

$$\sigma_{ij} = \beta_i \beta_j \sigma_M^2$$

This effectively means that the information requirements for the implementation of MPT can be substantially weakened. Suppose there are N risky assets under consideration.

[4] Recall the difficulty in constructing the empirical counterpart of M.

[5] Fama (1973), however, demonstrates that in its form (14.1) the market model is inconsistent in the following sense: the fact that the market is, by definition, the collection of all individual assets implies an exact linear relationship between the disturbances ε_j; in other words, when the single factor is interpreted to be the market, the hypothesis $\text{cov}(\tilde{\varepsilon}_j, \tilde{\varepsilon}_k) = 0, \forall j \neq k$ cannot be strictly valid. While we ignore this criticism in view of our purely pedagogical objective, it is a fact that if a single-factor model had a chance of being empirically verified (in the sense of all the assumptions in Eq. (14.1) being confirmed), the unique factor could not be the market.

In that case, the computation of the efficient frontier requires knowledge of N expected returns, N variances, and $(N^2 - N)/2$ covariance terms. (N^2 is the total number of entries in the matrix of variances and covariances; take away the N variance/diagonal terms, and divide by 2 since $\sigma_{ij} = \sigma_{ji}$, $\forall i, j$.)

Working via the market model, on the other hand, requires estimating Eq. (14.1) for the N risky returns, producing estimates for the N β_j's and the N $\sigma^2_{\tilde{\varepsilon}_j}$ and estimating the variance of the market return. This represents $2N + 1$ information items, many fewer than under the previous case.

14.3 A Second Illustration: Multifactor Models, and the CAPM

The APT approach is generalizable to any number of factors. It does not, however, provide any clue as to what these factors should be or any particular indication as to how they should be selected. This is both its strength and its weakness. Suppose we can agree on a two-factor model:

$$\tilde{r}_{j,t} = \alpha_j + b_{j,1}\tilde{f}^1_t + b_{j,2}\tilde{f}^2_t + \tilde{e}_{j,t} \tag{14.2}$$

with $E\tilde{e}_j = 0, \mathrm{cov}(\tilde{f}^1_t, \tilde{e}_j) = \mathrm{cov}(\tilde{f}^2_t, \tilde{e}_j) = 0,\ \forall j, \mathrm{cov}(\tilde{f}^1, \tilde{f}^2) = 0$ and $\mathrm{cov}(\tilde{e}_k, \tilde{e}_j) = 0,\ \forall j \neq k$.

As was the case for Eq. (14.1), Eq. (14.2) implies that one cannot reject, empirically, the hypothesis that the *ex post* return on an asset j has two stochastic components: one specific, $(\tilde{e}_j)$, and one systematic, $(b_{j,1}\tilde{f}^1_t + b_{j,2}\tilde{f}^2_t)$. What is new is that the systematic component is not viewed as the result of a single common factor influencing all assets as in the market model. Common or systematic issues may now be traced to two fundamental factors affecting, in varying degrees, the returns on individual assets (and thus on portfolios as well). Without loss of generality, we may assume that these factors are uncorrelated. As before, an expression such as Eq. (14.2) is useful only to the extent that it describes a relationship that is relatively stable over time. The two factors $\tilde{f}^1$ and $\tilde{f}^2$ must really summarize *all* that is common in individual asset returns. What could these fundamental factors be?

Much of the remainder of this chapter is devoted to presenting the answer the literature provides to this question. For the present conceptual discussion, let us assume that $\tilde{f}^1$ is the price of energy and that it is perfectly proxied by the value-weighted return on a portfolio of international oil stocks:

$$\tilde{f}^1_t = \tilde{f}^{P_1}_t$$

Let $\tilde{f}^2$ be a measure of the stock market's risk sensitivity and let us suppose that it is well proxied by a long position in a portfolio (defaultable) of Baa bonds financed by a short position in (default-free) 10-year US Treasury notes:

With these definitions, we can write Eq. (14.2) as

$$\tilde{r}_{j,t} = \alpha_j + b_{j,1}\tilde{r}_t^{P_1} + b_{j,2}\tilde{r}_t^{P_2} + \tilde{e}_{j,t} \tag{14.3}$$

Viewing Eq. (14.3) as a regression equation, the $b_{j,k}$, $k = 1, 2$ coefficients are referred to as factor loadings. Let us assume (falsely, as it turns out) that $\text{cov}\left(\tilde{r}_t^{P_1}, \tilde{r}_t^{P_2}\right) = \text{cov}\left(\tilde{r}_t^{P_i}, \tilde{e}_{j,t}\right) = \text{cov}(\tilde{e}_{j,t}, \tilde{e}_{i,t}) = 0, j \neq i, i, j \varepsilon \{1, 2\}$, and that a portfolio is also traded with zero sensitivity to either factor and zero asset-specific risk (i.e., a risk-free asset).

The APT then states that there exist scalars λ_0, λ_1, λ_2 such that

$$\overline{r}_j = \lambda_0 + \lambda_1 b_{j,1} + \lambda_2 b_{j,2} \tag{14.4}$$

That is, the expected return on an arbitrary asset j is perfectly and completely described by a linear function of asset j's factor loadings $b_{j,1}$, $b_{j,2}$. This can appropriately be viewed as a (two-factor) generalization of the security market line (SML).

Furthermore, the coefficients of the linear function are

$$\lambda_0 = r_{\text{f}}$$
$$\lambda_1 = \overline{r}_{P_1} - r_{\text{f}}$$
$$\lambda_2 = \overline{r}_{P_2} - r_{\text{f}}.$$

The reader will note the resemblance of Eqs. (14.3) and (14.4) to the results of the Fama and MacBeth (1973) methodology (Chapter 8) and herein lies the manner by which a proposed set of factors can be tested as to their efficacy.

The APT agrees with the CAPM that the risk premium on an asset, $\overline{r}_j - \lambda_0$, is not a function of its specific or diversifiable risk. It potentially disagrees with the CAPM in the identification of the systematic risk. The APT decomposes the systematic risk into elements of risk associated with a particular asset's sensitivity to a few fundamental common factors.

Note also the parallels with the Arrow–Debreu pricing approach. In both contexts, every individual asset or portfolio can be viewed as a complex security, or a combination of primitive securities: Arrow–Debreu securities in one case, the pure-factor portfolios in the other. Once the prices of the primitive securities are known, it is a simple step to compose replicating portfolios and, by a no-arbitrage argument, price complex securities and arbitrary cash flows. The difference, of course, resides in the identification of the primitive security. Although the Arrow–Debreu approach sticks to the conceptually clear notion of states of nature, the APT takes the position that there exist a few common and stable sources of risk and that they can be empirically identified. Once the corresponding risk premia are identified, by observing the market-determined premia on the primitive securities (the portfolios with unit sensitivity to a particular factor and zero sensitivity to all others), the pricing machinery can be put to work.

Let us illustrate. In our two-factor examples, a security j with, say, $b_{j,1} = 0.8$ and $b_{j,2} = 0.4$ is like a portfolio with proportions of 0.8 of the pure portfolio P_1, 0.4 of pure portfolio P_2, and consequently proportion -0.2 in the riskless asset. By our usual (no-arbitrage) argument, the expected rate of return on that security must be

$$\begin{aligned}\bar{r}_j &= -0.2r_\mathrm{f} + 0.8\bar{r}_{P_1} + 0.4\bar{r}_{P_2} \\ &= -0.2r_\mathrm{f} + 0.8r_E + 0.4r_\mathrm{f} + 0.8(\bar{r}_{P_1} - r_\mathrm{f}) + 0.4(\bar{r}_{P_2} - r_\mathrm{f}) \\ &= r_\mathrm{f} + 0.8(\bar{r}_{P_1} - r_\mathrm{f}) + 0.4(\bar{r}_{P_2} - r_\mathrm{f}) \\ &= \lambda_0 + b_{j,1}\lambda_1 + b_{j,2}\lambda_2\end{aligned}$$

The APT equation can thus be seen as the immediate consequence of the linkage between pure-factor portfolios and complex securities in an arbitrage-free context. The reasoning is directly analogous to our derivation of the value additivity theorem in Chapter 11 and leads to a similar result: Diversifiable risk is not priced in a complete (or quasi-complete) market world.

Though potentially more general, the APT does not necessarily contradict the CAPM. That is, it may simply provide another, more disaggregated, way of writing the expected return premium associated with systematic risk, and thus a decomposition of the latter in terms of its fundamental elements. Clearly, the two theories have the same implications if (keeping with our two-factor model, the generalization is trivial):

$$\beta_j(\bar{r}_m - r_\mathrm{f}) = b_{j,1}(\bar{r}_{P_1} - r_\mathrm{f}) + b_{j,2}(\bar{r}_{P_2} - r_\mathrm{f}) \tag{14.5}$$

Let β_{P_1} be the (market) beta of the pure portfolio P_1 and similarly for β_{P_2}. Then if the CAPM is valid, not only is the LHS of Eq. (14.5) the expected risk premium on asset j, but we also have

$$\begin{aligned}\bar{r}_{P_1} - r_\mathrm{f} &= \beta_{P_1}(\bar{r}_M - r_\mathrm{f}) \\ \bar{r}_{P_2} - r_\mathrm{f} &= \beta_{P_2}(\bar{r}_M - r_\mathrm{f})\end{aligned}$$

Thus, the APT expected risk premium may be written as

$$b_{j,1}[\beta_{P_1}(\bar{r}_M - r_\mathrm{f})] + b_{j,2}[\beta_{P_2}(\bar{r}_M - r_\mathrm{f})] = (\beta_{j,1}\beta_{P_1} + \beta_{j,2}\beta_{P_2})(\bar{r}_M - r_\mathrm{f})$$

which is the CAPM equation provided:

$$\beta_j = b_{j,1}\beta_{P_1} + b_{j,2}\beta_{P_2}$$

In other words, CAPM and APT have identical implications if the sensitivity of an arbitrary asset j with the market portfolio fully summarizes its relationship with the two underlying common factors. In that case, the CAPM would be another, more synthetic, way of writing the APT.[6]

In reality, of course, there are reasons to think that the APT with an arbitrary number of factors will always do at least as well in identifying the sources of systematic risk as the CAPM. Indeed, this will generally be the case. But first, a formal statement of the APT.

14.4 The APT: A Formal Statement

The APT assumes a model of the financial markets that is both frictionless and competitive and in which the returns to each traded asset i are governed by a K factor structure of the following form:

i. $\tilde{R}_i = E\tilde{R}_i + \boldsymbol{b}_i^T \tilde{\boldsymbol{f}} + \tilde{\varepsilon}_i$ where (14.6i)

$\tilde{f} = [\tilde{f}^1, \tilde{f}^2, \tilde{f}^3, \ldots, \tilde{f}^K]$ is a $K \times 1$ vector of random factors, and $\boldsymbol{b}_i^T$ *is a* $K \times 1$ vector of factor sensitivities (constants) for asset i where

ii. $E\tilde{\boldsymbol{f}} = 0, \; E\tilde{f}^i \tilde{f}^j = 0, \quad$ for all i, j, (14.6ii)

$E\tilde{\varepsilon}_i \tilde{f} = 0$ and $E\tilde{\varepsilon}_i = 0$ for all i.

The APT also assumes that a very large number of such assets is traded in the sense that for well-diversified portfolios with weights $\approx (1/N)$, where N is large, $\tilde{\varepsilon}_P = \sum_{i=1}^{N} (1/N) \tilde{\varepsilon}_i \approx 0$. Although not specifically required by the APT, we will further assume the financial markets are complete and that a risk-free asset is traded.

Given the above structure, Ross (1976) demonstrates that if there are no-arbitrage opportunities, then there exists a $K \times 1$ vector of factor risk premia λ_K such that for any asset i,

$$E\tilde{R}_i - R_f \approx \mathbf{b}_i^T \boldsymbol{\lambda}_K \tag{14.7}$$

Expressions (14.6i, ii) and their consequence (14.7) constitute the APT.

Since the conclusion of the APT is an approximation, expression (14.7), it does not directly provide testable implications for asset returns.[7] Within the context of a full equilibrium model, however, Connor (1984) details additional requirements such that Eq. (14.7) can be

---

[6] The observation in footnote 5, however, suggests this could be true as an approximation only.

[7] Note that Eqs. (14.6i, ii) and (14.7) suggest the Fama and MacBeth (1973) procedure of Section 8.9.1.

expressed as an equality (exact factor pricing).[8] Alternatively, Ross (1976) demonstrates that the approximation (14.7) becomes increasingly more accurate as the number of assets increases, while Grinblatt and Titman (1985), arguing from a different angle, conclude that deviations from exact factor pricing are likely to be very small. Taken together, these studies suggest that empirical exercises based on Eq. (14.7), assumed to hold with equality, can be justified from several perspectives.

To interpret Eq. (14.7) more fully, consider a well-diversified portfolio i (well diversified in the sense of Connor (1984)), for which idiosyncratic risk has been entirely diversified away so that

$$\tilde{R}_i = E\tilde{R}_i + \mathbf{b}_i^T \tilde{\boldsymbol{f}} \tag{14.8}$$

Since we have assumed the asset market is complete we know there exists a unique SDE, $\tilde{m}$, within the space of payoffs, that allows us to price all assets.[9]

Accordingly, Eq. (14.8) allows us to write

$\tilde{m}\tilde{R}_i = \tilde{m}E\tilde{R}_i + \tilde{m}\mathbf{b}_i^T\tilde{\boldsymbol{f}}$, and, by Theorem 10.3:

$1 = E\tilde{m}\tilde{R}_i = E\tilde{R}_i E\tilde{m} + \mathbf{b}_i^T E\tilde{m}\tilde{\boldsymbol{f}}$. Thus,

$1 = \dfrac{E\tilde{R}_i}{R_{\mathrm{f}}} + \mathbf{b}_i^T E\tilde{m}\tilde{\boldsymbol{f}}$, and

$$E\tilde{R}_i - R_f = E\tilde{r}_i - r_f = \mathbf{b}_i^T[-R_{\mathrm{f}}\boldsymbol{\pi}(\tilde{\boldsymbol{f}})]$$

where we define $-R_{\mathrm{f}}\pi(\tilde{\boldsymbol{f}}) = \lambda_K$, and interpret $\pi(\tilde{\boldsymbol{f}})$ as defining the $K \times 1$ vector of factor risk premia.

As we have noted earlier, the APT is not really an economic model in the sense of our discussions in Chapters 3–10. Its only claim is that there exists a factor structure that captures all stock returns. Nothing is said as to the identity of the factors $\tilde{\boldsymbol{f}}$ nor even their number. In this sense, the APT is silent as to the fundamental underlying source of asset premia and return comovement. The conclusion to the theory does not even guarantee that all factor risk premia are strictly positive. Indeed, the APT is basically a generalization of the "market model" of Section 14.2.1 that allows a parsimonious structuring of the variance–covariance matrix of returns. Accordingly, without a theoretical discipline, there has arisen a large literature proposing a variety of factors for explaining returns.

8 Connor (1984) requires, in particular, that the universe of assets under consideration is very large and that no single asset accounts for more than a trivial proportion of the economy's total wealth.

9 See Appendix 14.1 for a graphical interpretation of the APT.

Roughly speaking these models are of two types, those that employ macroeconomic, business-cycle-related factors (e.g., the growth rate of GDP), and those whose factors are returns on the so-called factor-mimicking portfolios. In the latter case, the portfolio return patterns are presumed to capture some important and variable underlying feature of the securities markets. It may be macroeconomic or behavioral. The connection here is generally far from transparent; the important thing, most profoundly in the eyes of practitioners, is that the factors "work," meaning that they greatly assist in explaining, statistically, a wide class of portfolio returns. We explore macro factor models in Section 14.5, while Section 14.6 reviews models based on factor-mimicking portfolios. Most of the attention in the literature has focused on the latter approach because of the bountiful availability of financial return data: model performance can be assessed at monthly or even daily return frequencies. Macroeconomic data (e.g., GDP growth) is frequently available only at quarterly frequencies at best, and even so, is open to substantial subsequent revision. Of necessity, macro models therefore take a longer-term perspective.

Especially as regards the factor-mimicking portfolios, the reader should keep in mind that these portfolios' return patterns reflect a specific type of risk for which investors should, in equilibrium, receive compensating average return risk premia. Accordingly, factor-mimicking portfolios should display positive average returns appropriate to the undiversifiable risks they represent.

14.5 Macroeconomic Factor Models

A few introductory remarks are in order here in order to set the stage. First, the focus of this literature is to explain the cross section of equity returns using the Fama and MacBeth (1973) methodology (see Section 8.9.1). More specifically, it is the cross section of equity *portfolio* returns, where some *a priori* rule is adopted that assigns each individual stock to one of a limited number of well-defined portfolios.[10] Recently, the emphasis in the literature has been to explain the cross-sectional return patterns for the 25 Fama and French (1993) portfolios whose construction we detail in the next section, asking our readers' forbearance until then. A pervasive aspect to these studies is also to assess whether the addition of the "market factor," $\tilde{r}_M - r_f$, significantly improves the ability of the model under study to explain the cross section of returns beyond a corresponding model that is solely macro factor or factor-mimicking-portfolio based.

Lastly, it is unreasonable to presume that causality runs only from exogenous macroeconomic events to equity return patterns. As has become abundantly clear from the

[10] Recall that under the Fama and MacBeth (1973) methodology, the first pass beta regression estimates are much more precisely estimated when portfolios are the objects of discussion.

events of the recent financial crisis, financial events can have monumental consequences for the macroeconomy. In particular, the loss in the US aggregate stock market valuation, in the 2008−2009 acute phase of the crisis, undoubtedly influenced the subsequent US GDP growth rate via its effects on consumption demand (via the wealth effect) and investment demand (via increased future cash-flow uncertainty).

Since asset prices and returns are jointly determined by discount factors and expected cash-flow streams (recall the discussion in Chapter 10), each macro factor must be plausibly related to at least one of these two quantities. Chen et al. (1986) represents the first widely studied macro factor model. Its focus is to assess the ability of a class of macro factors to explain the cross sections of returns to 20 equally weighted portfolios which represent a partition of all NYSE stocks according to market value ranking. After exploring the explanatory power of a wide class of factors, Chen et al. (1986) identify the following as most significant:

cash flow related:

1. $g_t^{\mathrm{IP}} = ln\ \mathrm{IP}_t - ln(\mathrm{IP}_{t-1})$, where IP_t is an index of the level of industrial production in period t and g_t^{IP} is its (continuously compounded) growth rate,

discount factor related:

1. $\mathrm{UPR}_t =$ unanticipated risk premia in period t
   $=$ (*Baa* and lower corporate bond portfolio return in period t)
   $-\mathrm{LTG}_t$, where LTG_t is the period t return on a portfolio of (default-free) long-term US Treasury bonds.
2. $\mathrm{TS}_t =$ the slope of the term structure
   $= \mathrm{LTG}_t - \mathrm{TB}_{t-1}$, where LTG_t is measured as in (1) immediately above and TB_{t-1} is the Treasury bill (short rate) in period $t-1$.
3. a measure of unanticipated inflation,
   $\mathrm{INF}_t = \mathrm{INF}_t - E_{t-1}\mathrm{INF}_t$
   where INF_t is *ex post* inflation from period $t-1$ to t and $E_{t-1}\mathrm{INF}_t$ is expected inflation for period t conditional on information at the close of period $t-1$.

Although this latter quantity was somewhat indirectly measured in Chen et al. (1986), with the advent of Treasury Inflation Protected Securities (TIPS), it would be natural to measure $E_{t-1}\mathrm{INF}_t$ as

$$E_{t-1}\mathrm{INF}_t = r_{t-1}^{\text{1-year nominal Treasury}} - r_{t-1}^{\text{1-year TIPS bond}}$$

When added to this set of explanatory variables, Chen et al. (1996) found that the market index was statistically insignificant in its contribution to explaining the cross section of returns. In what is a somewhat surprising outcome vis-à-vis the consumption capital asset pricing model (CCAPM) theory of Chapter 10, the growth rate of real consumption per capita was similarly insignificant. The reader is referred to the article itself for the numerous details associated with the construction of the aforementioned factors.

Other macro factors have been proposed in the recent asset pricing literature. Jagannathan and Wang (1996) employ the growth rate of labor income as a proxy for the return on human capital which they propose as an important systematic risk factor. They also employ a measure of the default premium on corporate debt as a measure of the stage of the business cycle. Santos and Veronesi (2006) explore the influence of the labor income/consumption ratio, while Lettau and Ludvigson (2001) detail the consequences of including the consumption/wealth ratio.

Total factor productivity (TFP) risk (see web-related chapter) and various demographic risk measures have also been suggested as macro factors. TFP shocks directly influence the return on capital, and, because of their well-known high intertemporal persistence, can be expected to influence this return over an extended number of years. Demographic risks are potentially related to the fraction of the population in retirement in contrast to the fraction which is saving for retirement. Since retired persons are typically selling financial assets to finance their retirement consumption, as their fraction of the population increases (as is presently the case in most developed countries), lower equity prices and returns may result. See Abel (2001) and Geanakoplos et al. (2004).

14.6 Models with Factor-Mimicking Portfolios

As noted previously, the APT is silent as to the identity of the factors underlying common stock returns. From the discussion in Chapter 8, however, we are aware that (BE/ME) and "size" (market value of all common equity) are two quantities with substantial ability to explain the cross section of returns, and it is thus natural to propose "factors" related to them. Fama and French (1993) were among the first to take up this agenda.

14.6.1 The Size and Value Factors of Fama and French (1993)

To construct these factors, Fama and French (1993) first sort the universe of NYSE, AMEX, and NASDQ stocks into (first sorting) three (B/ME) value-weighted portfolios as ranked lowest to highest with (second sorting) these portfolios then subdivided into the half with the higher ME values, and the other half with the lower ME values (Table 14.1).[11]

[11] This is the same collection of stocks (and their return histories) as is used in Fama and French (1992); see Section 8.9.1.

Table 14.1: Stock sort underlying the Fama and French (1993) factor construction

Biggest (B) ½ of stocks ranked by ME	BL	BM	BH
Smallest (S) ½ of stocks ranked by ME	SL	SM	SH
	Lowest (L) 30% of stocks as ranked by (BE/ME)	Middle (M) 40% of stocks as ranked by (BE/ME)	Highest (H) 30% of stocks as ranked by (BE/ME)

The Fama–French "size" and "value" factors are then constructed from long and short positions in these six portfolios as follows:

$$\begin{aligned}\text{The ``size'' factor} &= \text{SMB}_t =_{\text{def}} ((1/3)r_t^{\text{SH}} + (1/3)r_t^{\text{SM}} + (1/3)r_t^{\text{SL}}) \\ &\qquad - ((1/3)r_t^{\text{BH}} + (1/3)r_t^{\text{BM}} + (1/3)r_t^{\text{BL}}) \\ \text{The ``value'' factor} &\equiv \text{HML}_t = \text{def}((1/2)r_t^{\text{SH}} + (1/2)r_t^{\text{BH}}) \\ &\qquad - ((1/2)r_t^{\text{SL}} + (1/2)r_t^{\text{BL}}).\end{aligned}$$

In these definitions r_t^{ij} refers to the net return on portfolio ij, where $i \in \{S, B\}, j \in \{L, M, H\}$ and the "minus" sign denotes a long position financed by a short position. The SMB and HML factor titles are acronyms for "small minus (i.e., short) big" and "high (BE/ME) minus low," respectively. More precisely, a \$100,000 SMB portfolio, for example, is composed of a \$100,000 long position in the portfolio defined by $w_{\text{SB}} = w_{\text{SM}} = w_{\text{SL}} = 1/3$, a \$100,000 short position in the portfolio defined by $w_{\text{BH}} = w_{\text{BM}} = w_{\text{BL}} = 1/3$, and \$100,000 invested in risk-free assets. The return $\text{SMB}_t - r_f$ thus represents the excess return on this portfolio; similarly for HML_t.[12] By forming long–short portfolios of these types, the small firm premium and value premium are strengthened and made more sensitive to business cycle variation as they are fundamentally highly leveraged entities. As constructed, SMB and HML are pure-factor-mimicking portfolios, although it is not clear at this point what fundamental economic determinants they represent.

Fama and French (1993) evaluate their three-factor model $\text{SMB}_t, \text{HML}_t, r_t^M - r_f$ (the market factor) as regards its ability to explain the cross-sectional pattern of returns across a universe of 25 specifically constructed, value-weighted portfolios. Since it has become commonplace in the literature to judge a model as regards its ability to explain the cross section of returns on these specific 25 portfolios, it behooves us to become acquainted with their construction.

The assignment protocol uses the same data set and is in the same spirit as their factor construction except that stocks are sorted at a higher level of refinement: at the start of a year every stock in the above universe is ranked by market value ("size") from lowest to highest, and assigned to one of five portfolios on this basis. In particular, the first quintile portfolio contains the lowest fifth of all stocks ranked by size; the top quintile contains the highest

[12] In the asset pricing literature financial factors are always constructed in this way.

Table 14.2: The Fama-French portfolios[i]

Highest ME quintile					(5,5)
				(4,3)	
Lowest ME quintile					
	(1,1)		(3,1)		(1,5)
	Lowest (BE/ME) quintile				Highest (BE/ME) quintile

[i]entry (i, j) denotes the ith (BE/ME) quintile with the Jth quintile ME stocks

fifth, with all other stocks assigned to the remaining three quintile portfolios in an identical manner. The stocks are then reranked into five subquintiles based on their ascending B/ME (the ratio of the book value to the market equity value), with the sorting redone annually. The net result is to assign, each year, every stock to one of the so-called "25 Fama—French value-weighted portfolios." Table 14.2 is intended to reflect this assignment.

In particular, stocks assigned to the value-weighed portfolio (1,1) are stocks with the lowest ME and the lowest (BE/ME). These firms are likely to be small and growing rapidly: most of their value is in the form of their future growth opportunities, not in their fixed assets. Small software companies come to mind. At the other end of the spectrum, stocks in the (5,5) portfolio are typically the stocks of large, well-established firms with large fixed assets, but more modest future growth opportunities. The other 23 portfolios basically reflect various degrees of these basic two distinctions.[13]

Why should this specific collection of portfolios be of interest? From one perspective, this sorting encompasses many if not most "investment strategies" that investors adopt. For example, young persons with extended future working lives may view themselves as best served by investing in a high-risk—high-reward small firm strategy somewhere in the range of (1,1)—(1,5). Alternatively, older persons who need income and wealth stability may prefer (3,5)—(5,5) portfolios. Second, the annual rebalancing of the portfolios reflects the notion of a "dynamic asset allocation" strategy typical of many hedge funds.

These factors were employed in the standard Fama and MacBeth (1973) two-step regression procedure, which we review for clarity:

First-pass regression: for all 25 Fama—French portfolios, regress

$$\tilde{r}_t^i - r_f = \hat{\alpha}_i + \hat{\beta}_M^i(\tilde{r}_t^M - r_f) + \hat{\beta}_{\mathrm{SMB}}^i(\widetilde{\mathrm{SMB}}_t - r_f) + \hat{\beta}_{\mathrm{HML}}^i(\widetilde{\mathrm{HML}}_t - r_f) + \varepsilon_t^i$$

[13] The sorting into portfolios also allows more precise estimation of market betas. However, there is something suspect in using similarly constructed portfolios to provide both the factors themselves and the portfolios whose return patterns are to be explained. See Daniel and Titman (1997).

Table 14.3: Coefficient estimates: Fama and French (1993) three-factor model

	$\hat{\gamma}_0$	$\hat{\gamma}_M$	$\hat{\gamma}_{\text{SMB}}$	$\hat{\gamma}_{\text{HML}}$
Coefficient	0.031	−0.012	0.003	0.012
t-statistic	(2.265)	(−0.753)	(0.566)	(2.034)
$R^2_{\text{Adj.}} = 0.65$				

Second-pass regression: using the $\{\bar{r}^i, \hat{\beta}^i_M, \hat{\beta}^i_{\text{SMB}}, \hat{\beta}^i_{\text{HML}}\}$, for $i = 1, 2, \ldots, 25$, regress

$$\tilde{\bar{r}}^i - r_{\text{f}} = \gamma_0 + \hat{\gamma}_M \tilde{\hat{\beta}}^i_M + \hat{\gamma}_{\text{SMB}} \tilde{\hat{\beta}}^i_{\text{SMB}} + \hat{\gamma}_{\text{HML}} \tilde{\hat{\beta}}^i_{\text{HML}} + \tilde{u}_i$$

The results are listed in Table 14.3, for one representative historical period.[14]

Unlike single-factor CAPM, the Fama and French (1992, 1993) multifactor model appears to explain the cross-sectional variation in returns to the 25 Fama−French portfolios reasonably well: all the coefficients are significant and the R^2 is reasonably high: 65% of the variation in average returns across the 25 portfolios is explained by variation in the three factors. Unfortunately, and contrary to theory, the estimate on the CAPM beta is negative and not significant.

There is no generally accepted story to explain why the SMB and HML factors work so successfully, but they must be capturing fundamental systematic macroeconomic risks or behavior biases. Systematic risk in the CAPM sense is business cycle risk: variation in the rate of growth of the underlying economy's GDP. This sort of risk must affect the profitability of all firms, though to differing degrees. Small firms, for example, are frequently credit constrained in cyclical downturns. This possibility is much less likely for large, well-established firms: established firms in slow growth industries may suffer very significant demand reductions in recessions, whereas the demand growth for small, high tech firms may not diminish at all. Unfortunately, we do not know what specific phenomena principally underlie the SMB and HML factors. It is a subject of on-going research.

In particular, Liew and Vassalou (2000) demonstrate that SMB and HML are leading indicators of future GDP growth. Vassalou (2003) goes on to show that a factor measuring news related to future GDP growth, in conjunction with the standard market factor, can explain equity returns as well as the SMB and HML factors with regard to returns on the 25 Fama and French portfolios. More recently, Zhang (2005) has proposed a model in which investment irreversibility, coupled with counter-cyclical risk aversion, serves as the underlying determinant of the HML factor. His results are derived within the framework of a dynamic general equilibrium representative agent model and, as such, his setting is much more highly structured (and therefore much more informative—recall the remarks of Lucas

[14] See the web notes to this chapter where we interpret this regression and provide comparative results in regard to the CAPM versus the FF three factor model.

in Chapter 2) than the APT. The idea is that high book to market value firms are those with high fixed assets, while growth firm assets are largely in the form of "growth options" (see Web Chapter B). Because of irreversibility, value firms find it extremely costly to reduce their capital stocks to a more efficient level in bad times while low book to market growth firms are unaffected in the same way because their capital stock is in the form of the human capital of their employees which can be reduced by discharge. In good times value firms feel less pressure to increase their capital stock as they enter these times with already excess capital. Growth firms do expand in good times, but the cost of expanding their capital stocks is comparatively low. The net effect of this phenomenon is that value firms are fundamentally more exposed to business cycle risk and, as such, require a higher equilibrium return as the HML factor presumes.

Various behavioral theories of the value premium have also been proposed. The idea in Lakonishok et al. (1994) is that the value premium centers around investors' overreaction to and over-extrapolation of recent news about firm returns. Unlike the rational economic story of Zhang (2005), value stocks are not fundamentally more risky than growth stocks, but are cheap and pay high returns because investors consistently underestimate their future prospects (with the opposite being true of growth stocks). Barberis and Huang (2001) are able to generate a value premium in the context of a model with loss aversion and mental accounting. Their idea is this: losses are very painful to loss-averse investors who, because of mental accounting (see Chapter 3), focus on losses to individual stocks rather than on their overall portfolio's return. Accordingly, value stocks are ones that suffered losses in the past and, as a result of mental accounting, investors are individually focused on these losses and are pessimistic regarding their future prospects.

As we conclude this section on the Fama and French (1993) factors, it is of interest to get some idea as to the actual return properties of the HML and SMB portfolios. These are found in Tables 14.4 and 14.5 for a prefinancial-crisis data period.

Table 14.4: International performance of the SMB factor for the period January 1997–December 2006; principal world stock markets, annualized returns

	Quarterly Rebalancing			Semiannual Rebalancing			Annual Rebalancing		
Country	Mean (%)	SD (%)	*t*-Value	Mean (%)	SD (%)	*t*-Value	Mean (%)	SD (%)	*t*-Value
Australia	6.21	15.88	1.38	2.79	16.06	0.61	5.88	19.15	1.06
Canada	4.85	10.71	2.01	6.02	10.79	2.46	5.16	10.15	2.21
France	5.22	11.70	1.66	5.46	11.42	1.79	5.40	10.49	1.92
Germany	2.07	9.69	0.68	0.82	9.63	0.27	0.46	9.94	0.14
Italy	2.19	10.50	0.67	2.92	10.32	0.89	0.59	10.29	0.18
Japan	6.78	15.27	1.81	6.92	15.37	1.82	6.57	14.05	1.87
Netherlands	2.00	12.98	0.67	1.82	12.73	0.62	2.40	11.92	0.88
Switzerland	−4.13	10.81	−1.26	−3.39	11.01	−1.02	−1.20	10.88	−0.37
UK	3.37	11.15	1.33	3.02	11.09	1.20	3.17	11.00	1.25
USA	10.73	13.65	3.48	11.46	13.93	3.62	6.45	10.57	2.65

The high returns evidenced in Tables 14.4 and 14.5 invite investors to attempt to reap the excess returns so evident in the tables, and this has the implication that over time they will be reduced in equilibrium. This has been the case most profoundly for the SMB factor. Dimson and Marsh (1999) and Fama and French (2012) both find little evidence of significant size premia internationally for recent data sets. This finding does not carry over to the HML factor, however, which seems to remain quite robust as regards the magnitude of its risk premium. There is one data set that would seem to constitute an exception to this rule, and that is found in Daniel et al. (2014). While not the focus of their study, these authors compute the average annualized returns (based on monthly rebalancing) to the market, SMB, and HML factors. The results are presented in Table 14.6.

In both periods, the market factor returns are robustly significant, and in neither period is this the case for SMB or the HML factors.

Table 14.5: International performance of the HML factor for the period January 1997–December 2006; principal world stock markets, annualized returns

Country	Quarterly Rebalancing			Semiannual Rebalancing			Annual Rebalancing		
	Mean (%)	SD (%)	*t*-Value	Mean (%)	SD (%)	*t*-Value	Mean (%)	SD (%)	*t*-Value
Australia	9.30	14.53	2.26	9.14	14.06	2.29	5.93	13.74	1.48
Canada	7.44	11.06	2.98	8.56	10.69	3.53	8.16	10.46	3.41
France	12.51	9.09	5.13	12.05	9.26	4.85	10.32	9.90	3.90
Germany	5.55	6.42	2.75	3.14	5.96	1.66	4.56	5.98	2.40
Italy	7.29	9.77	2.38	7.47	9.24	2.55	7.43	9.18	2.54
Japan	8.75	10.12	3.53	6.85	9.77	2.84	7.71	9.32	3.29
Netherlands	0.75	11.50	0.28	0.96	11.60	0.36	0.68	11.17	0.26
Switzerland	8.66	10.34	2.77	7.62	10.57	2.38	8.48	9.87	2.83
UK	8.33	6.09	6.06	7.45	5.90	5.56	6.91	5.84	5.14
USA	7.99	12.24	2.89	7.98	12.12	2.90	6.74	8.64	3.39

Table 14.6: Excess returns, annualized, for the SMB, HML and Market factors

February 1990 – August 2013			
	Market	SMB	HML
Average	7.01	3.04	3.75
Standard error	(2.62)	(1.70)	(1.91)
February 1990 – August 2013			
Average	7.20	2.51	2.95
Standard error	(3.37)	(2.25)	(2.60)

14.6.2 Momentum Portfolios

Momentum portfolios and the momentum factor to which they give rise are fundamentally distinct from the HML and SMB factor portfolios. The basic reference is Jegadeesh and Titman (1993), although the notion of a momentum strategy (for portfolio construction) preceded them (e.g., De Bont and Thaler, 1985, 1987). Using return data on all NYSE and AMEX stocks for the period January 1965−December 1989, Jegadeesh and Titman (1993) rank, at the start of each month, these stocks from lowest to highest on the basis of their historical returns in the $J = 3$, 6, 9, and 12-month prior periods. For each of these periods, all stocks are then assigned to one of ten portfolios again on a return ranked basis: the lowest decile portfolio contains the lowest 10% of all stocks ranked by returns, the second decile the next lowest 10%, etc. Each portfolio is equal weighted. The authors then construct a "buy−sell" portfolio where the lowest decile stock portfolio is sold short to finance an equal-value long position in the highest decile portfolio.[15] The return on this long−short portfolio for $K = 3$, 6, 9, and 12-month forward horizons is then computed. Each month the portfolios are reassembled as per above, and the indicated returns computed. The cumulative results are presented in Table 14.7.

Note that the monthly returns going forward on the buy portfolios are generally much higher relative to the sell portfolios indicating some persistence at least as regards to relative returns. For buy−sell portfolio formation based on 12 months of historical data and held for 3 months, the excess return was 1.31% per month on average which is enormous, although it does tend to peter out for longer horizons. For delayed portfolio formation (Panel B), the corresponding figure is 1.49%, or roughly 18% on an annual basis. These results are in direct contradiction to market efficiency as it is commonly understood. Furthermore, the results are driven neither by significant systematic risk differentials nor by profound differences in market capitalizations (This is confirmed in Table II of Jegadeesh and Titman, 1993). Accordingly, they also contradict the implications of the standard CAPM model (something we address in Chapter 8). Subsequently, these same authors report that similar return patterns are observed in later historical periods (Jegadeesh and Titman, 2001). Rouwenhorst (1998) finds the same patterns in European stock market data, while Asness et al. (2013) find that momentum phenomena are pervasive, being present in currency markets, commodity markets, etc. The acronym for the momentum factor is UMD, signifying "up minus down."

The momentum portfolio constructed as per above appears to represent another fundamental APT factor. Such a conclusion, however, is based on empirical evidence alone. Again,

---

15 Just as in the construction of the SMB and HML portfolios, a \$100,000 momentum portfolio is composed of a \$100,000 long position in the decile 10 portfolio, a \$100,000 short position in the decile 1 portfolio, and \$100,000 in risk-free assets. The excess return on the momentum portfolio is the return on the aforementioned portfolio less r_f.

Table 14.7: Excess returns on buy, sell, and buy−sell portfolios[a]

	Panel A				Panel B[b]			
J =	$K=3$	6	9	12	$K=3$	6	9	12
6 Sell	0.0087	0.0079	0.0072	0.0080	0.0066	0.0068	0.0067	0.0076
	(1.67)	(1.56)	(1.48)	(1.66)	(1.28)	(1.35)	(1.38)	(1.58)
6 Buy	0.0171	0.0174	0.0174	0.0166	0.0179	0.0178	0.0175	0.0166
	(4.28)	(4.33)	(4.31)	(4.13)	(4.47)	(4.41)	(4.32)	(4.13)
6 Buy-sell	0.0084	0.0095	0.0102	0.0086	0.0114	0.0110	0.0108	0.0090
	(2.44)	(3.07)	(3.76)	(3.36)	(3.37)	(3.61)	(4.01)	(3.54)
12 Sell	0.0060	0.0065	0.0075	0.0155	0.0048	0.0058	0.0070	0.0085
	(1.170	(1.29)	(1.48)	(1.74)	(0.93)	(1.15)	(1.40)	(1.71)
12 Buy	0.0192	0.0179	0.0168	0.0155	0.0196	0.0179	0.0167	0.0154
	(4.63)	(4.36)	(4.10)	(3.81)	(4.73)	(4.36)	(4.09)	(3.79)
12 Buy-sell	0.0131[c]	0.0114	0.0093	0.0068	0.0149	0.0121	0.0096	0.0069
	(3.74)	(3.40)	(2.95)	(2.25)	(4.28)	(3.65)	(3.09)	(2.31)

[a]Data from Jegadeesh and Titman (1993), Table 1. Returns are average monthly excess returns for the indicated horizons. *T*-statistics are in parentheses. Excess returns signify returns above the risk-free rate. Portfolios based on J month lagged returns and held for K months.
[b]Panel B describes the exactly analogous results except that the buy−sell portfolio is (and its buy and sell constituents) constructed 1 week after the historical returns are available. Measured returns are thus delayed by 1 week.
[c]The number 1.31 is to be interpreted as follows: Consider a portfolio composed of a long position of $100 in the buy portfolio, a $100 short position in the short portfolio, and $100 in risk-free assets. After subtracting the return on the risk-free assets, the portfolio nets $1.31 per month on average.

to date, there is no consensus as to the underlying macro (or psychological) factor for which it serves as the "factor-mimicking portfolio".

While the SMB, HML, and UMD factors have received the most attention in the literature, they are by no means the only ones to be proposed. The accrual factor, which is based on accounting data (Sloan, 1996), deserves mention. The idea here is based on the accounting distinction between the cash component of current earnings (cash actually received) and the accrual component (cash promised to the firm by customers, etc., and cash promised by the firm in the sense of delayed payments). The cash component of earnings, and its relative contribution to earnings, in particular, is viewed as more informative of future earnings potential than the accrual component. Accordingly it is possible to develop a factor formed by a long position in a portfolio of firms with low levels of accruals financed by a short position in firms' responding high levels of accruals. Let us simply denote it as the ACR factor and eschew the details of its construction.

Daniel and Titman (2006) propose a composite share issuance variable, ISU_t which measures "the amount of equity a firm issues (or retires) in exchange for cash or services. Thus seasoned issues, employee stock option plans and share based acquisitions increase the issuance measure while repurchases, dividends, and other actions that take cash out of the firm reduce the issuance measure" (Daniel and Titman, 2006, p. 1608). The economics

Table 14.8: *Ex post* Sharpe ratios: various combinations of the listed factor-mimicking portfolio[a]

Portfolio Proportions in Percent						
Market	SMB	HML	UMD	ISU	ACR	*Ex post* Sharpe Ratio
100						0.31
75.07	24.93					0.32
28.19	14.63	57.18				0.80
21.13	10.16	41.92	26.79			1.18
18.82	15.33	13.67	9.55	42.44		1.55
17.35	14.47	12.32	8.18	36.44	11.04	1.60

Data period: July 1963—December 2012 portfolios annually rebalanced.
[a]From Daniel (2012).

behind this measure is the observation that firms tend to issue shares when management receives favorable information regarding investment opportunities not previously reported and tend to repurchase shares in the absence of such information.

Given these new factors, we conclude this section with a comment regarding the relative importance of the many factor-mimicking portfolios. This is provided by Daniel (2012) who, in the tradition of MPT, treats each of the factors, $\tilde{r}^M - r_f$ (the market), $\widetilde{SMB}, \widetilde{HML}, \widetilde{ISU}$ and $\widetilde{ACR}$ as a distinct portfolio, and computes the *ex post* maximum Sharpe ratio realizable as progressively more assets are added. The results of this exercise are presented in Table 14.8.

Note that the largest proportional boosts come from the addition of the HML and UMD factors, while the SMB factor's contribution is slight. Note also the power of the ISU_t factor.

14.7 Advantage of the APT for Stock or Portfolio Selection

The APT helps to identify the sources of systematic risk, and to split systematic risk into its fundamental components. It can thus serve as a tool for helping the portfolio manager modulate his risk exposure. For example, studies show that, among US stocks, the stocks of chemical companies are much more sensitive to short-term inflation risk than stocks of electrical equipment companies. This would be compatible with both having the same exposure to variations in the market return (same beta). Such information can be useful in at least two ways. When managing the portfolio of an economic agent whose natural position is very sensitive to short-term inflation risk, chemical stocks may be a lot less attractive than electricals, all other things equal (even though they may both have the same market beta). Second, conditional expectations, or accurate predictions, on short-term inflation may be a lot easier to achieve than predictions of the market's return. Such a refining of the information requirements needed to take aggressive positions can, in that context, be of great use. To summarize these thoughts more succinctly, the APT reminds investors that when they construct their preferred equity portfolios, they are in reality constructing portfolios of factors. Accordingly, the portfolio risk premia they earn will be determined not so much by the risk premia of the individual assets in their portfolios, but the individual risk premia of the assorted factors they have thereby implicitly elected to hold.

14.8 Conclusions

We have now completed our review of asset pricing theories. At this stage, it may be useful to draw a final distinction between the equilibrium theories covered in Chapters 8–10 and the theories based on arbitrage such as the Martingale pricing theory and the APT. Equilibrium theories aim at providing a complete theory of value on the basis of *primitives*: preferences, technology, and market structure. They are inevitably *heavier*, but their weight is proportional to their ambition. By contrast, arbitrage-based theories can only provide a relative theory of value. With what may be viewed as a minimum of assumptions, they

- offer bounds on option values as a function of the price of the underlying asset, the stochastic behavior of the latter being taken as given (and unexplained);
- permit estimating the value of arbitrary cash flows or securities using risk-neutral measures extracted from the market prices of a set of fundamental securities, or in the same vein, using Arrow–Debreu prices extracted from a complete set of complex securities prices;
- explain expected returns on any asset or cash-flow stream once the price of risk associated with pure-factor portfolios has been estimated from market data on the basis of a postulated return-generating process.

Arbitrage-based theories currently have the upper hand in practitioner circles where their popularity far outstrips the degree of acceptance of equilibrium theories. This, possibly temporary, state of affairs may be interpreted as a measure of our ignorance and the resulting needs to restrain our ambitions.

References

Abel, A., 2001. Will requests attenuate the predicated meltdown in stock prices when baby boomers retire? Rev. Econ. Stat. 83, 589–595.
Asness, C., Moskowitz, T., Pedersen, L., 2013. Value and momentum everywhere. J. Finan. 68, 929–985.
Barberis, N., Huang, M., 2001. Mental accounting, loss aversion, and individual stock returns. J. Finan. 56, 1247–1292.
Burmeister, E., Roll, R., Ross, S.A., 1994. A practitioner's guide to arbitrage pricing theory. A Practitioner's Guide to Factor Models. Research Foundation of the Institute of Chartered Financial Analysts, Charlottesville, VA.
Chen, N.F., Roll, R., Ross, S.A., 1986. Economic forces and the stock market. J. Bus. 59 (3), 383–404.
Connor, G., 1984. A unified beta pricing theory. J. Econ. Theory. 34 (1), 13–31.

Daniel, K., 2012. Behavioral finance: a selective survey for owners of money. Presentation, JOIM Fall Conference.
Daniel, K., Hodrick, R., Lu, Z., 2014. The carry trade: risks and drawdowns. Working Paper, Columbia Business School.
Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. J. Finan. 52 (1), 1–33.
Daniel, K., Titman, S., 2006. Market reactions to tangible and intangible information. J. Finan. 61 (4), 1605–1643.
De Bont, W., Thaler, R., 1985. Does the stock market overreact? J. Finan. 40, 793–805.
De Bont, W., Thaler, R., 1987. Further evidence on investor overreaction and stock market seasonality. J. Finan. 42, 557–581.
Dimson, E., Marsh, P., 1999. Murphy's law and market anomalies. J. Portf. Manage. 25, 53–69.
Fama, E.F., 1973. A note on the market model and the two-parameter model. J. Finan. 28 (5), 1181–1185.
Fama, E., French, K., 1992. The cross section of expected stock returns. J. Finan. 47, 427–465.
Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. J. Finan. Econ. 33, 3–56.
Fama, E., French, K., 2012. Size, value and momentum in international stock returns. J. Financ. Econ. 105, 457–472.
Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: empirical tests. J. Polit. Econ. 81, 607–636.
Geonakoplos, J. Magill, Quinzii, M., M., 2004. Demography and the long – run predictability of the stock market. Brookings Pap. Econ. Act. 1, 241–307.
Grinblatt, M., Titman, S., 1985. Approximate factor structures: interpretations and implications for empirical tests. J. Finan. 15 (5), 1367–1373.
Huberman, G., 1982. A simple approach to arbitrage pricing. J. Econ. Theory. 28, 183–191.
Jagannathan, R., Wang, Z., 1996. The conditional capm and the cross-section of expected returns. J. Finan. 51, 3–53.
Jegadeesh, N., Titman, S., 1993. Returns lo buying winners and selling losers: implications for stock market efficiency. J. Finan. 48, 65–91.
Jegadeesh, N., Titman, S., 2001. Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. The Journal of Finance. 56, 699–720.
Lakoneshok, J., Schleefer, A., Visblny, R., 1994. Contrarian investment, extrapolation, and risk. J. Finan. 44, 154–1578.
Lettav, M., Ludvigson, S., 2001. Consumption, aggregated wealth and expected stock returns. J. Finan. 55, 815–849.
Liew, J., Vassalou, M., 2000. Can book to market, size and momentum be risk factors that predict economic growth? J. Finan. Econ. 57, 221–245.
Ross, S.A., 1976. The arbitrage pricing theory. J. Econ. Theory. 1, 341–360.
Rouwenhorst, G., 1998. International momentum strategies. J. Finan. 53, 267–284.
Santos, T., Veronesi, P., 2006. Labor income and predictable stock returns. Rev. Finan. Stud. 19 (1), 1–44.
Santos, T. Veronesi, P. Labor income and predictable stock returns. Rev. Financ. Stud.. 19, 1–44.
Sloan, R., 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? Account. Rev. 71 (3), 289–315.
Vassalou, M., 2003. News related to future GDP growth as a risk factor in equity returns. J. Finan. Econ. 68, 47–73.
Zhang, L., 2005. The value premium. J. Finan. 60 (1), 67–103.

Appendix A.14.1: A Graphical Interpretation of the APT

For illustrative purposes, we assume one common factor. As noted in Connor (1984), the APT requires existence of a *rich* market structure with a large number of assets with different characteristics and a minimum number of trading restrictions. Such a market structure, in particular, makes it possible to form a portfolio P with the following three properties:

Property 1 P has zero cost; in other words, it requires no investment. This is the first requirement of an arbitrage portfolio.

Let us denote w_i as the ***value*** of the position in the ith asset in portfolio P. Portfolio P is then fully described by the vector $\boldsymbol{w}^T = (w_1, w_2, \ldots, w_N)$, and the zero cost condition becomes

$$\sum_{i=1}^{N} \mathbf{w}_i = 0 = \mathbf{w}^T \cdot \mathbf{1}$$

with $\mathbf{1}$ the (column) vector of 1's. (Positive positions in some assets must be financed by short sales of others.)

Property 2 P has zero sensitivity (zero beta) to the common factor:[16]

$$\sum_{i}^{N} w_i \beta_i = 0 = w^T \cdot \beta$$

Property 3 P is a well-diversified portfolio. The specific risk of P is (almost) totally eliminated:

$$\sum_{i}^{N} w_i^2 \sigma_{\varepsilon_i}^2 \cong 0$$

The APT builds on the assumed existence of such a portfolio, which requires a rich market structure.

Statement and Proof of the APT

The APT relationship is the direct consequence of the factor structure hypothesis, the existence of a portfolio P satisfying these conditions, and the no-arbitrage assumption. Given that returns have the structure of Eq. (14.1), Properties 2 and 3 imply that P is riskless. The fact that P has zero cost (Property 1) then entails that an arbitrage opportunity will exist unless:

$$\bar{r}_P = 0 = \mathbf{w}^T \cdot \bar{r} \tag{A.14.1}$$

[16] Remember that the beta of a portfolio is the weighted sum of the betas of the assets in the portfolio.

The APT theorem states, as a consequence of this succession of statements, that there must exist scalars λ_0, λ_1, such that

$$\begin{aligned} \bar{r} &= \lambda_0 \cdot \mathbf{1} + \lambda_1 \beta, \text{or} \\ \bar{r}_i &= \lambda_0 + \lambda_1 \beta_i \quad \text{for all assets}, i \end{aligned} \tag{A.14.2}$$

which is the main equation of the APT.

Equation (A.14.2) and Properties 1 and 2 are statements about four vectors: $\boldsymbol{w}$, β, $\mathbf{1}$, and $\bar{r}$.

Property 1 states that w is orthogonal to **1.** Property 2 asserts that w is orthogonal to β. Together these statements imply a geometric configuration that we can easily visualize if we fix the number of risky assets at $N = 3$, which implies that all the vectors have dimension 3. This is illustrated in Figure 14.1.

Equation (A.14.2)—no arbitrage—implies that $\boldsymbol{w}$ and $\bar{r}$ are orthogonal. But this means that the vector $\bar{r}$ must lie in the plane formed by $\mathbf{1}$ and β, or that $\bar{r}$ can be written as a linear combination of $\mathbf{1}$ and β, as Eq. (A.14.2) asserts.

More generally, one can deduce from the triplet

$$\sum_{I}^{N} w_i = \sum_{I}^{N} w_i \beta_i = \sum_{I}^{N} w_i \bar{r}_i = 0$$

that there exist scalars λ_0, λ_1 such that

$$\bar{r}_i = \lambda_0 + \lambda_1 \beta_i \text{ for all } i$$

This is a consequence of the orthonormal projection of the vector $\bar{r}_i$ into the subspace spanned by the other two.

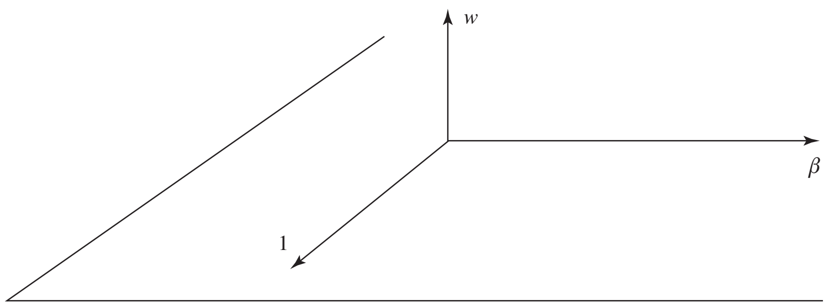

Figure 14.1
Geometric representation: w orthogonal to 1 and β.

The CAPM and the APT

Suppose that there exists a risk-free asset or, alternatively, that the sufficiently rich market structure hypothesis permits constructing a fully diversified portfolio with zero sensitivity to the common factor (but positive investment). Then

$$\bar{r}_f = r_f = \lambda_0$$

That is, λ_0 is the return on the risk-free asset or the risk-free portfolio.

Now let us compose a portfolio Q with unitary sensitivity to the common factor: $\beta = 1$. Then applying the APT relation, one gets

$$\bar{r}_Q = r_f + \lambda_1 \cdot \mathbf{1}$$

Thus, $\lambda_1 = \bar{r}_Q - r_f$, excess return on the pure-factor portfolio Q. It is now possible to rewrite Eq. (A.14.2) as

$$\bar{r}_i = r_{\mathrm{f}} + \beta_i(\bar{r}_Q - r_f) \tag{A.14.3}$$

If, as we have assumed, the unique common factor is the return on the market portfolio, in which case $Q = M$ and $\bar{r}_Q \equiv \bar{r}_M$, then Eq. (13.4) is simply the CAPM equation:

$$\bar{r}_i = r_f + \beta_i(\bar{r}_M - r_f)$$

Appendix 14.2: Capital Budgeting

To illustrate an example of the use of the Fama and French (1992, 1993) three factor model for a capital budgeting cost of capital exercise, let us estimate Microsoft Corporation's cost-of-capital. We contract this method with the CAPM. The estimates are derived from 5 years of monthly data from the period 1996.1 to 2000.12.

A. CAPM results
   1. Standard regression

$$\tilde{r}_t^{\text{micro}} - r_{\mathrm{f}} = \hat{\alpha}^{\text{micro}} + \hat{\beta}\left[\tilde{r}_t^M - r_f\right] + \tilde{\varepsilon}_t^{\text{micro}}$$

Estimate	0.994	1.71
Standard error	(1.41)	(0.289)
t-statistic	704	5.91

2. For this period $= [\bar{r}_M - r_f] = 0.0617, \sigma_M = 0.1666,$ and $r_f = 0.06.$

$$\sigma^{\text{micro}} = 0.4644$$

$$(\sigma^{\text{micro}})^2 = (0.4644)^2 = (\hat{\beta}^{\text{micro}})^2 \sigma_M^2 + (\sigma_\varepsilon^{\text{micro}})^2$$

$$(0.4644)^2 = (1.71)^2 (0.1666)^2 + (\sigma_\varepsilon^{\text{micro}})^2$$

$$\sigma_\varepsilon^{\text{micro}} = 0.3668 \text{ or } 36.68\%$$

$$\begin{aligned} E\tilde{r}^{\text{micro}} &= r_f + \hat{\beta}^{\text{micro}} [\bar{r}_M - r_f] \\ &= 0.06 + (1.71)[0.0617] \\ &= 0.166 \text{ or } 16.6\% \end{aligned}$$

This is the estimate of Microsoft's cost of capital using the CAPM.

B. Fama–French results

$\tilde{r}_t^{\text{micro}} - r_f$	$= \hat{\alpha}^{\text{micro}}$	$+ \hat{\beta}_M^{\text{micro}} [\tilde{r}_t^M - r_f]$	$+ \hat{\beta}_{\text{SMB}}^{\text{micro}} [\widetilde{\text{SMB}}_t - r_f]$	$+ \hat{\beta}_{\text{HML}}^{\text{micro}} [\widetilde{\text{HML}}_t - r_f]$	$+ \tilde{\varepsilon}_t^{\text{micro}}$
Estimate	0.6335	1.096	−1.374	−1.389	
Standard error	(1.25	(0.303)	(0.381)	(0.333)	
t-statistic	0.505	3.62	−3.599	−4.165	

$$(\bar{r}_M - r_f) = .0617$$

$$\overline{SMB}^A - r_f = .0212$$

$$\overline{HML}^A - r_f = .0379$$

$$\begin{aligned} E\tilde{r}^{\text{micro}} &= r_{\text{f}} + \hat{\beta}_M^{\text{micro}} [\bar{r}_M^A - r_f] + \hat{\beta}_{\text{SMB}}^{\text{micro}} \left(\overline{\text{SMB}}^A - r_f\right) + \hat{\beta}\left(\overline{\text{HML}}^A - r_f\right) \\ &= 0.06 + (1.1)[0.0617] + (-1.374)[0.0212] + (-1.39)(0.0379) \\ &= 0.0461, \text{ or } 4.61\% \end{aligned}$$

$E\tilde{r}^{micro}$ is the estimate of Microsoft's cost of capital using the Fama-French three factor model. There is a substantial difference between the CAPM and the Fama-French estimates.