

Equilibrium Asset Pricing and Discount Factors: Overview and Implications for Derivatives Valuation and Risk Management

John H. Cochrane and Christopher L. Culp

Graduate School of Business, The University of Chicago

Price is expected discounted payoff. This fundamental relation underlies all asset pricing. The discount factor is an index of “bad times”. Because investors are willing to pay more for assets that do well in bad times, the risk premium on any asset is determined by how it co-varies with the discount factor. All of asset pricing comes down to techniques for measuring a discount factor in a way that is useful for specific applications. We first survey the theoretically pure consumption-based discount factor model and some of its practical counterparts, including the capital asset pricing model (CAPM) and intertemporal CAPM. We also survey recent literature that points to recession and distress-related components of a discount factor to supplement the traditional market portfolio/beta component. As an important practical implication, we show how asset pricing concepts can help investors and hedgers distinguish properly between “systematic” and “idiosyncratic” risk. We also explain how a little bit of thinking about the fundamental determinants of discount factors can go a long way towards evaluating market price of risks that bedevil practical application of derivative pricing theories. Finally, we emphasise where the pure theory requires various stages of equilibrium and market completeness, and where the theory can still be used out of such equilibrium or in the presence of incomplete markets.

INTRODUCTION

All of asset pricing comes down to one central idea: the value of an asset is equal to its expected discounted payoff. The rest is elaboration, special cases, a closet full of clever tricks, and, above all, a set of specifications for the discount factor that make this central idea useful for one or another application.

There are two fundamental and polar approaches to applying the central idea: *absolute* and *relative* asset pricing. In *absolute* pricing, we value a bundle of cashflows (dividends, coupons and principle, option payoffs, firm profits, etc) based on its exposure to fundamental sources of macroeconomic risk. Equivalently, we find a discount factor by thinking about what macroeconomic states are of particular concern to investors. The CAPM is a paradigm example of this approach. Virtually all such models are based on some notion of general equilibrium in order to use aggregate rather than individual risks, or to substitute an easily measured index such as the market return (CAPM) for poorly measured consumption. They are therefore often called equilibrium asset pricing models.

In *relative* pricing, by contrast, we ask what we can learn about one asset's value *given* the prices of some other assets. We do not ask from where the prices of the other assets came, and we use as little information about macroeconomic risk factors as possible. Modigliani and Miller (1958) pioneered this approach in valuing a firm *given* prices for its equity and debt, and Black and Scholes (1976) famously used this approach to value options and corporate liabilities *given* stock and bond prices.

Absolute pricing offers generality – it can be applied to anything – at the cost of precision in many applications. Relative pricing offers simplicity and tractability – it can be done easily, at least to a first approximation – at the cost of often-limited practical applicability.¹ Most good applications do not use one extreme or the other, but rather a blend of absolute and relative approaches appropriate for the problem at hand. Even the most die-hard applications of the CAPM, for example, usually take the market risk premium as given. Conversely, most realistic option pricing exercises implicitly use some absolute pricing model to characterise pesky “market prices of risk” that cannot be perfectly hedged.

Why should you care? Study of the “assumptions” behind theoretical models is often viewed as about the driest subject in finance.

But understanding these models and the assumptions under which they hold is vital before modern ideas can be correctly and creatively applied to actual financial problems. To value a potential investment, for example, you have to know what discount rate to use. Does the famed CAPM "beta" give the right discount rate? Even if the investor holds a portfolio different than the market portfolio, such as owning a small business? Even if other investors do not hold the market portfolio as the CAPM assumes? Or, suppose you want to set up a hedging program, but a perfect hedge is impossible. How should you value the residual risks or tracking errors? How do you know when you should take *more* of those residual risks in order to profit from their premia? Only by really understanding where absolute equilibrium asset pricing models come from can you begin to confidently answer these sorts of questions.

The theory we survey here is at the foundation of everything in asset pricing. It may look simple in hindsight, but this theory led directly to five Nobel Prize presentations in economic sciences, and to at least another five indirectly.² This chapter is longer than the others in this book, but it is so because our contribution is unifying: rather than devoting a separate chapter in this section to each application of the single fundamental value equation, we survey them all and show how they are each applications of the same simple idea.

THE ECONOMIC INTUITION FOR VALUING UNCERTAIN CASHFLOW STREAMS

The basic objective of asset pricing is to determine the value of any stream of uncertain cashflows. Consider an asset with a single cashflow or payoff x_{t+1} at time $t + 1$. (This payoff can be the price at $t + 1$ plus any dividend, so we have not abandoned the real world.) We find the value of this payoff by asking what the stream is worth to an investor.³ The answer is:

$$p_t = E_t \left[\beta \frac{u_c(c_{t+1})}{u_c(c_t)} x_{t+1} \right] \quad (1)$$

where β is the investor's discount rate, $E_t[\cdot]$ denotes an expected value conditional on information available at time t , and $u_c(c_t)$ is the

benefit to the investor of a small additional unit of consumption received at time period t (ie, the "marginal utility" of time t consumption).

This is the basic equation underlying all of asset pricing, so it is worth understanding carefully. (The appendix presents a mathematical derivation.) Investors do not value money directly. The theory adopts a more sophisticated approach and recognises that the pleasure or "utility" of the consumption that money can buy is what really matters. That is why u_c and β enter Equation (1). Specifically, people value money more if it comes sooner, and if it comes in bad times when they really need it rather than good times when they are already doing well. A β slightly less than one and a marginal utility function u_c that declines as c increases capture these important considerations. If the economist's "utility function" sounds strange, just think of $u_c(c_t)$ as an index of "bad times", or a measure of how painful it is to give up a dollar at date t .

Equation (1) then describes the investor's optimal portfolio decision as marginal cost of investment equals marginal benefit. The true cost of an extra dollar invested is the price of the asset p_t (how many dollars the investor had to give up) times the value of a dollar (utility cost to the investor) $u_c(c_t)$ at time t . The true benefit is the expectation E_t of the dollar payoff x_{t+1} times the value of a dollar $u_c(c_{t+1})$ at time $t+1$, times β which discounts future value (utility) back to time t .

The logic of Equation (1) is often confused. Equation (1) is usually used to describe a market "in equilibrium", *after* the investor has reached his or her optimum portfolio. But to get to that optimum, the investor had to know the price p_t . What's going on? In watching a market "in equilibrium", we are like scientists, watching over the market in white lab coats and trying to understand what it does. Equation (1) holds once the investor has found the optimal portfolio, but it does not describe causes on the right and effects on the left. If we observe consumption and payoff, we can use this equation to determine what the price must be. If we observe consumption and prices (a common case), we can use the equation to learn what the expected payoff (eg, expected return in the case of stocks) must be. If we observe price, consumption, and payoff (the entire distribution of payoffs, mind you, not just how it

happened to come out a few times), then we can use Equation (1) to decide that the world really does make sense after all.

We can also use Equation (1) to think of the value of payoff x_{t+1} to an investor who has *not* yet bought any of that asset – ie, who has not found the optimum portfolio or when the market is “out of equilibrium”. This interpretation is especially important in thinking about the potential value of securities that have not been created or are not traded or in giving portfolio advice. Now the value p_t need not correspond to the market price (there might not be one), and the investor need not know what the price is. Still, we can compute the value of a *small incremental investment* in this uncertain cashflow to *this particular investor* by Equation (1). If there is a market price, and the value to the investor is greater than that price, we can recommend a buy. Similarly, we can compare our private valuation with market prices of derivatives to evaluate the desirability and cost of various hedging opportunities.⁴

THE FUNDAMENTAL VALUE EQUATION

We commonly split the fundamental value Equation (1) into two parts, one that expresses price as an expected discounted payoff

$$p_t = E_t[m_{t+1} x_{t+1}] \quad (2)$$

and one that relates the *stochastic discount factor* m_{t+1} to the *intertemporal marginal rate of substitution*, the rate at which the investor is willing to substitute one unit of consumption now for one unit of consumption later:

$$m_{t+1} = \beta \frac{u_c(c_{t+1})}{u_c(c_t)} \quad (3)$$

The term m_{t+1} is called a *stochastic discount factor* by analogy with simple present value rules. We are used to discounting a payoff by some discount factor or required rate of return R – if x_{t+1} is known:

$$p_t = \frac{1}{R} x_{t+1}$$

The term m_{t+1} acts exactly as such a discount factor in Equation (2). The discount factor term m_{t+1} is *stochastic* because nobody at time t knows what consumption will be at time $t + 1$ (whether $t + 1$ will be a good or bad time), and hence nobody knows what $u_c(c_{t+1})$ will turn out to be. It is random, in the same way stock returns are random.

This randomness of the discount factor is crucial. As the index of good and bad times, the stochastic discount factor is high if time $t + 1$ turns out to be a bad time (ie, consumption is low). Assets that pay off well in bad times are particularly valuable, and Equation (2) will give them an appropriately high price.

In finance, we commonly do not think too much about consumption and utility functions. The stochastic discount factor remains an index of bad times (strictly speaking, *growth* in bad times). Finance theorists tend to think directly about discount factor models in which data such as the market return are used as indicators of bad vs good times.

The fundamental valuation equation in different guises

The stochastic discount factor representation of the fundamental value equation, Equation (2), can be expressed in a number of different and equivalent ways depending on the nature of the particular financial problem being solved, and the history and traditions of different fields. (As usual, unification came after the fact.)

If we use required returns to value a project or compute a capital budget, we typically use a different, risk-adjusted required return for each project. The beauty of Equations (1) or (2) is that the *same* discount factor can be used for *all* assets, simply by putting it inside the expectation – m is a *universal* discount factor. Capital budgeting thus can be undertaken using Equation (2) directly. For a project that costs I_t in initial investment and generates X_{t+1} in revenue at time $t + 1$, the usual net present value criterion tells us to undertake the investment if its value is greater than its cost,

$$E_t[m_{t+1} X_{t+1}] \geq I_t$$

In applications to stocks and portfolios, it is often convenient to think about rates of return rather than prices. To do so, we divide

both sides of Equation (2) by p_t (or recognise that a return is a payoff with price equal to one):

$$1 = E_t[m_{t+1} R_{t+1}] \quad (4)$$

where R_{t+1} is the gross return on the asset (x_{t+1} divided by p_t). Thus, simply use "return" for the payoff x and one for price to apply the equation to stocks and portfolio problems. Similarly, if we want to work with returns in excess of the risk-free rate $r_{t+1} \equiv R_{t+1} - R^f$ then Equation (4) becomes⁵

$$0 = E_t[m_{t+1} r_{t+1}] \quad (5)$$

In equity analysis, we are used to a slight transformation of Equation (5). Using a superscript i to remind us that there are many assets, writing out the definition of covariance $E(mr^i) = E(m)E(r^i) + \text{cov}(m, r^i)$ and the definition of a regression coefficient $\beta_i = \text{cov}(m, r^i) / \text{var}(m)$ and then defining $\lambda = -\text{var}(m) / E(m)$, we can write Equation (5) in classic form as

$$E_t(r^i) = \beta_i \lambda \quad (6)$$

The expected excess return of each asset should be proportional to its beta.

This is not the CAPM – it is *perfectly general*. The beta here is calculated relative to the discount factor, not the market return. All the assumptions of the CAPM (or other models) come in substituting the market portfolio return or some other index for the discount factor.

To determine the fixed price in a forward purchase contract or swap, we look for the fixed price or rate K that equates the *initial* discounted expected value of the transaction to zero. Otherwise, one of the two parties would walk away from the deal. If a forward purchase contract calls for the future delivery of one unit of an asset at time $t + 1$ whose value on that date is P_{t+1} , then the fixed price in the forward is found as⁶

$$E_t[m_{t+1} (P_{t+1} - K)] = 0 \Rightarrow K = E_t[P_{t+1}] + \text{cov}_t(m_{t+1}, P_{t+1})R^f$$

In other words, the forward price K is the expected future spot price of the asset to be purchased at time $t+1$ plus a term that reflects how the discount factor co-varies with the underlying spot price – as we shall see later, a “risk premium”. The second term is grossed up by the risk-free rate to reflect the fact that the terms of the contract are set at time t but the contract is settled at time $t+1$.

In option pricing and fixed income applications, the discount factor m_{t+1} is often used to define a set of “risk-neutral probabilities”. Equation (2) is written simply as

$$p_t = \frac{1}{R^f} E_t^* [x_{t+1}]$$

where the $*$ reminds us to take the expectation with an artificial set of probabilities that are scaled by m_{t+1} . We can either use the discount factor to make good payments in bad states more important in determining the price, or we can boost up the probabilities of those states to the same effect. Risk aversion is the same thing as overestimating the probability of bad events.

All these formulations are just different ways of writing the same thing. Continuous-time asset pricing models also all reduce to the analogue to Equation (2) in continuous time.⁷

UNDERSTANDING “SYSTEMATIC” RISK

Asset pricing is about risk and reward. Identifying risks and assessing the premium one earns for bearing risks are the central questions for asset pricing, and the point of theory is to provide necessary quantitative tools to answer these questions. Risk managers tend to classify risk as market, credit, liquidity, operational, or legal. Portfolio managers tend to think of market risk, and the risks associated with certain styles such as size, value, and growth, as well as risks associated with industry and country portfolios.

Financial economists tend instead to distinguish between “systematic” and “idiosyncratic” risk. And of course, not all risk is bad – you only earn a premium over risk-free interest rates by taking on some risk! What risks should we pay attention to? And how are these concepts related?

“Systematic” vs “Idiosyncratic” risk

A central and classic idea in asset pricing is that only systematic risk generates a premium. Idiosyncratic risks are “not priced”, meaning that you earn no more than the interest rate for holding them. That is why we employ risk managers to get rid of idiosyncratic risks.⁸ Plausible, but a theory proves its worth by helping us to understand what “systematic” and “idiosyncratic” mean in this context.

By using the definition of covariance $E(mx) = E(m)E(x) + \text{cov}(m, x)$ we can re-write the fundamental value Equation (2) as

$$p_t = \frac{E_t [x_{t+1}]}{R^f} + \text{cov}(m_{t+1}, x_{t+1}) \quad (7)$$

This equation says that asset prices are equal to the expected cashflow discounted at the risk-free rate, plus a risk premium. The risk premium depends on the *covariance* of the payoff with the discount factor. This covariance is typically a negative number, so most assets have a lower price than otherwise (or a higher average return) as compensation for risk.

Here’s why. Recall that the discount factor is an indicator of *bad* times. Most assets pay off well in *good* times. Thus, most asset returns and payoffs co-vary negatively with the discount factor. The converse case drives home the intuition. Insurance is a terrible investment. The average return is negative – you pay more in premiums than you receive, on average, in settlements. Yet people willingly buy insurance. Why? Because insurance pays off well in bad times – just as the house stops smouldering, a cheque arrives in the post. The value of insurance is *higher* than predicted by the standard present value formula, because the covariance term is *positive*. Financial assets are “anti-insurance”, and it is this feature, and *only* this feature, that generates a risk premium and allows risky assets to pay more than the interest rate.

Equation (7) has a dramatic implication: a risk may be very large in the sense of having a high *variance*, but if it is uncorrelated with the discount factor, its *covariance* is zero, and it generates no premium. Its price is just the expected payoff discounted at the risk-free rate. *The volatility of the asset’s cashflow per se is completely irrelevant to its risk premium.*

To see why in more detail, consider an investor who adds a small fraction Ψ of the asset to their portfolio. Her consumption at time $t + 1$ is now $c_{t+1} + \Psi x_{t+1}$. As always, the investor cares about the variance of *consumption* and the variance of the utility that consumption generates, not any characteristics of single assets that, in a portfolio, determine the wealth from which they draw consumption. For a small asset purchase, the variance of the investor's new time $t + 1$ consumption is

$$\text{Var}(c_{t+1}) + 2\Psi\text{Cov}(c_{t+1}, x_{t+1}) + \Psi^2 \text{Var}(x_{t+1})$$

The covariance with consumption, and hence with marginal utility m , enters with a coefficient Ψ , while the variance is a second-order effect. For a small (marginal) investment Ψ , the covariance of the cashflows on the asset with consumption is much more important to how buying the asset affects consumption – what investors care about in the end – than the volatility of the asset's cashflows.

Now we can really understand and precisely define “systematic” vs “idiosyncratic” risk. *The systematic part of any risk is that part that is perfectly correlated with the discount factor.* It is the part that generates a risk premium. The idiosyncratic part of any risk is that part that is uncorrelated with the discount factor; it generates no premium.

We can divide any payoff into systematic and idiosyncratic components by simply running a regression of the payoff on the discount factor:

$$x_{t+1} = \beta m_{t+1} + \varepsilon_{t+1} \quad (8)$$

payoff = systematic part + idiosyncratic part

Regression residuals ε_{t+1} are, by construction, uncorrelated with the right hand variable.

Once again, this modern version of the theory is perfectly general. Systematic means correlated with the investor's marginal utility – *full stop*. This is true no matter what “asset pricing model” – no matter what specification of the discount factor – is correct. The CAPM, for example is one special case of the general theory. It specifies that the discount factor is linearly related to the market

return $m_{t+1} = a - bR_{t+1}$ (more on this below). Hence, it defines systematic risk for every asset by regressions of returns on to the market portfolio return.

In many portfolio management applications, “systematic” and “diversifiable” components are defined with multiple regressions on style portfolios, including size, book to market value, and industry groupings as well as the market portfolio. Implicitly, these definitions correspond to models that the discount factor is a function of these portfolio groupings. These specifications are fine, but they are special cases and they reflect lots of hidden assumptions. Other specifications may be useful.

When variance does matter

The proposition that variance does not matter for risk premia does not mean “ignore variance in setting up your portfolio”. Again, Equations (2) and (7) refer only to *marginal* valuations. That is appropriate *after* the investor has already set up an optimal portfolio, or for deciding which asset to *start* buying. For very little portfolio changes, covariance matters more than variance.

For big asset purchases and sales of the sort one considers while setting up the optimal portfolio to begin with, however, variance can matter a lot. If an investor buys a big part of a payoff, the variance *will* start to affect the properties of consumption, marginal utility, and hence the investor’s discount factor. So, by all means *do* consider variance in making the big changes required to set up a portfolio!

Diversification, hedging, and special investors

We often think of “idiosyncratic” risks as those risks that affect a particular security only, leaving all others untouched. Such idiosyncratic risks include firm-specific risks like operational and liquidity risk, as well as those components of market and credit risk that are unique to the firm in question. This is a good approximation in many cases, but understanding the correct definition we can quickly see how it is only an approximation.

A risk that moves many securities, but is uncorrelated with the discount factor, is also “idiosyncratic”. The market as a whole is built of individual securities, so each “idiosyncratic” risk is in fact a small part of the “total” risk. Many apparently “firm-specific”

risks – a drop in sales, an accounting fiasco, etc – also hit many other firms in the market, and thus *become* “market risks”. We often call “idiosyncratic” risks “diversifiable” because they largely disappear in well-diversified portfolios. This too is a good approximation, but specific to the CAPM world that the market portfolio return captures the discount factor and thus defines “systematic” risk.

A good counterexample for all these cases is to think of an investor who must hold a large part of some risk. Consider, for example, the owner of a business who must hold a large amount of one company’s stock. Risks correlated with the business or company stock will be “systematic” for this investor, albeit not necessarily for the market as a whole. The investor thus must require a premium to hold such risks, even though the market as a whole would not require such a premium. The CAPM is only an appropriate cost of capital for investment decisions if the investor holds the market portfolio, and thus his or her discount factor depends only on the market portfolio. For the vast majority of investors, this is not the case.

Although idiosyncratic risk does not matter for pricing, it is not necessarily easy to avoid it. Much of the art of risk management and corporate finance consists of just how to shed oneself of idiosyncratic or residual, non-priced, risk. The fact that risk managers focus on market, credit, liquidity, operational, or legal risk rather than systematic vs unsystematic risk reflects the different techniques needed to hedge different varieties of idiosyncratic risk. Risk managers, however, need to understand the real distinction between systematic and idiosyncratic risk so that they do not hedge good risks, ones that bring rewards, as well.

THE CAPITAL ASSET PRICING MODEL AND THE CONSUMPTION CAPITAL ASSET PRICING MODEL

A critical question for practical application remains: what data do we use for the discount factor m ? The search to populate m with actual data has led to the many “named” asset pricing models. All of these models are just special cases of the fundamental value equation. They add additional structure, usually from simple general equilibrium modelling, to substitute some other variable for consumption in the discount factor.

The consumption CAPM

Armed with our presentation of the theory so far, the simple, obvious approach is to assume some reasonable utility function (power or quadratic forms are popular), use the easily available aggregate consumption data, and apply pricing Equations (2) or (7) directly. This is the famous approach of Lucas (1978) and Hansen and Singleton (1982).

A little less directly, but more popular in finance, we can linearly approximate the discount factor as a function of consumption:

$$m_{t+1} \approx a - \gamma \Delta c_{t+1}$$

where Δc_{t+1} denotes consumption growth and γ is a constant of proportionality.⁹ This specification for the discount factor together with our fundamental value equation, Equation (2), is equivalent to the statement more popular in finance in terms of average returns and betas:

$$E(R^i) = R^f + \beta_{i,\Delta c} \lambda_{\Delta c} \quad (9)$$

where $E(R^i)$ denotes the average return on the i th asset, $\beta_{i,\Delta c}$ is the regression beta of the asset return on consumption growth, and $\lambda_{\Delta c}$ is a market risk premium. Equation (9) predicts that assets with higher consumption betas will have higher average returns, with the market risk premium as constant of proportionality. This is the form of Breeden's (1979) famous statement of the consumption capital asset pricing model (CCAPM).

The basic idea is really natural. We need an indicator of bad and good times. If you really want to know how people feel, don't listen to them whine, watch where they go out to dinner. Consumption reveals everything we need to know about current wealth, future wealth, investment opportunities, and so on. Compared to many of the models discussed later, the CCAPM in all its simplicity avoids the theoretical problems associated with restrictive and unrealistic assumptions. (Historically, the CCAPM came last, in order to repair those problems.)

Although a complete answer to most absolute asset pricing questions in principle, this consumption-based approach does not (yet) work well in practice. As one might imagine from even a

rudimentary experience with the data, running regressions of returns on the Government's consumption growth numbers does not reveal much. Instead, financial economists use a wide variety of tricks of the trade that, while requiring heroic "assumptions" to derive them as perfect truth, nonetheless have great intuitive appeal and have performed well in a variety of applications. At heart, they all involve substituting or proxying some other variables for hard-to-measures consumption.

The CAPM

The simplest and probably still the most popular model is the single-factor CAPM developed by Sharpe (1964) and Lintner (1965), and later extended by Black (1972). In the usual statement of the CAPM, the expected return of asset j is higher as its beta is higher, with the expected return on the market portfolio as a constant of proportionality:

$$E(R^j) = R^f + \beta_j [E(R^m) - R^f] \quad (10)$$

where R^j and R^m denote one-period arithmetic returns on asset j and the market, R^f denotes the risk-free interest rate, and where β_j is the regression coefficient of the return on the market, or

$$\beta_j = \frac{\text{Cov}(R^j, R^m)}{\text{Var}(R^m)}$$

The CAPM is *mathematically identical* to a specification of the discount factor that is linear in the market return, rather than linear in consumption growth

$$m_{t+1} = a - bR_{t+1}^m \quad (11)$$

a and b are free parameters that can be determined by the risk-free rate and market premium which are taken as given values in Equation (10). (See Equation (6) to make the connection between Equation (10) and a discount factor.)

The CAPM was a huge empirical success for a generation. Categories of assets have now been identified, however, whose average returns bear no relation to betas calculated against traditional proxy indices for the market portfolio. (Efforts to find theoretically purer proxies for the market portfolio of world invested wealth (Roll, 1977) do not help.) In addition, the betas calculated against several new risk factors apart from the market have been empirically shown to help explain why some average returns are higher than others. Prominent examples include the “small firm” and “value” effects (Banz, 1981, and Fama and French, 1993).

The CAPM discount factor model is a sensible approximation. When the market tanks, most people are unhappy! But it is clearly only an approximation. To *derive* the CAPM formally, you need to state assumptions under which a linear function of the market return is a *completely sufficient* indicator of good and bad times. The essence of the various formal derivations of the CAPM is to get every investor’s consumption growth to depend only on the market return.¹⁰

The CAPM (and all following models) is not an *alternative* to the consumption-based model, it is a *special case*. Now, consumption surely goes down when the market return goes down, but, in the real world, other things matter as well. For the CAPM to hold, people cannot think that market fluctuations are temporary, and hence ignore a bad day and go out to dinner anyway. Instead, we must have returns that are independent over time. People cannot have jobs, houses, cars, businesses, or other sources of income that sustain them through market crashes, or these other things will start to matter to consumption and the discount factor. In addition, all investors in a CAPM world must hold the same portfolio of assets – the market portfolio.

Having seen how the sausage is made, we should be surprised, if anything, that the CAPM lasted as long as it did. We should not be surprised that it ultimately proved a first approximation.

MULTI-FACTOR MODELS

Linear factor models dominate empirical asset pricing in the post-CAPM world. Linear factor pricing models measure the discount

factor – ie, specify indices of bad times – with a model of the form

$$m_{t+1} = \beta \frac{u_c(c_{t+1})}{u_c(c_t)} = a + b_1 f_{t+1}^1 + b_2 f_{t+1}^2 + L + b_N f_{t+1}^N \quad (12)$$

where a, b_1, \dots, b_N are free parameters and where f^j is the j th “risk factor”.

What exactly does one use for factors f_{t+1}^j ? In general, factor models look for variables that are plausible proxies for aggregate consumption or marginal utility growth (measures of whether times are getting better or worse). This is just like the CAPM, with additional indicators of good and bad times.

Intertemporal capital asset pricing model

Merton’s (1973) multi-factor intertemporal CAPM (ICAPM) model was the first theoretical implementation of this idea. The ICAPM recognises that investors care about the market return, so that is the first risk factor. The extra factors are innovations to “state variables” that describe an investor’s consumption-portfolio decision. Other things being equal, investors prefer assets that pay off well when there is news that *future* returns will be bad. Such assets provide insurance; they help to reduce the risk of long-term investments. Covariance with this kind of news thus will drive risk premia as well as the market return.

In the traditional statement of the model, corresponding to the expected-return statement of the CAPM in Equation (10),

$$E(R^j) - R^f = \gamma \text{Cov}(R^j, R^m) + \lambda_z \text{Cov}(R^j, \Delta z)$$

where γ and λ_z are constants (the same for all assets), R^m denotes the market (wealth) return and Δz indicates the news or the return on a “factor-mimicking portfolio” of returns correlated with that news.¹¹ Technically, this model generalises the CAPM assumption that the market return is independent over time.

Equivalently, for a given value of the market return, investors will feel poorer and will lower consumption if there is bad news about subsequent investment opportunities.¹² News about subsequent returns thus should drive our discount factor, as well as current market returns.

The ICAPM does not tell us the precise *identity* of the state variables, and, as a result, it is only after 30 years of theoretical fame that the ICAPM has received its first serious tests – tests that do not just cite it as inspiration for ad-hoc multifactor models, but actually check whether the factors do forecast returns as the theory says they should (see Ferson and Harvey, 1999). It is not yet common in applications.

One of the most popular current multi-factor models is the Fama-French three-factor model (Fama and French, 1993, 1996). The Fama-French model includes the market portfolio, a portfolio of small minus big stocks (SMB), and a portfolio of high book/market minus low book/market stocks (HML). In expected return language, average returns on all assets are linearly related to their regression betas on these three portfolios. In discount factor language, the discount factor is a linear function of these three portfolios, as the CAPM discount factor is a linear function of the market return.

This model is popular because the betas on the additional factors *do* explain the variation of average returns across the size and book-to-market portfolios, and market betas do not. The model is not a tautology. Size and book-to-market sorted portfolios do not have to move together, and move more as their average returns rise. Size and book to market betas also explain the variation of average returns across additional portfolio sorts, beyond those that they were constructed to explain (Fama and French, 1996). This kind of more general good performance is the hallmark of an empirically useful model.

The open question for the Fama-French model is: “what are the additional sources of risk about which investors are *economically* concerned?” It is well and good to say that investors fear “value” risk uncorrelated with the market, but why? Put another way, the sales talk for “value” portfolios is that, since other investors are so afraid of this risk, you, the remaining mean-variance investor, should load up on the “value” portfolio that provides high reward for small market beta. Fine, but if the *average* investor is really scared of this value risk, maybe you should be, too. If value risk turns out to be risk of poor performance during a financial crisis, for example, are you sure you want to take that risk? To answer this question, we really need to understand what fundamental macroeconomic risks

are behind the value effect, not just understand a set of mimicking portfolios that capture them for empirical work.

An empirical counterpart to this worry is that the Fama-French model found its limitations as well. Namely, average returns on "momentum" portfolios go in the *opposite* direction predicted by Fama-French betas (see Fama and French, 1996). This can be cured by adding a fourth "momentum" factor, the return on a portfolio long recent winners and short recent losers. For all purposes except performance attribution (where it is exactly the right thing to do), however, this fix smells of such *ad-hockery* that nobody wants to take it seriously. We do not want to add a new factor for every anomaly. On the other hand, multifactor models used in many industry applications suffer no such compunction and proudly use 50 or 60 portfolios as factors, picked purely on the basis of in-sample empirical performance.

Macroeconomic multifactor models

The alternative to finding essentially ad-hoc portfolio factors that perform well for a given sample is to sit back and think about risk factors that make sense given the fundamental economic intuition. What variables are good indicators of bad vs good times for a large number of investors? The market return obviously still belongs. Following ICAPM intuition, variables that forecast future investment opportunities such as price/earnings ratios, the level of interest rates, etc, make sense. Indicators of recessions may belong as well, such as proprietary and labour income, the value of housing or other non-marketed assets, business investment, and so forth. The CAPM and ICAPM exclude these variables (ie, they derive relations in which the discount factor is *only* a function of the market return and state variables) by presuming that investors have no jobs or other assets; they simply live off their portfolios of financial assets. Because investors *do* have jobs and other assets, bad times for these will spill over into market premia.

With this intuition, a wide variety of asset pricing models have been used that tie average returns to macroeconomic risks. The grandfather of all of these is the Chen, Roll, and Ross (1986) multi-factor model. They used interest rates, industrial production, inflation, and bond spreads to measure "bad times" in the discount factor. More recently, multi-factor models have used macroeconomic

risk factors such as labour income (Jagannathan and Wang, 1996) and investment growth (Cochrane, 1991a, 1996) to explain expected stock returns.

Conditional factor models in which factors at time $t + 1$ are scaled by information variables at time t are also increasingly popular ways to allow betas and factor risk premia to vary over time (see Cochrane, 1996 and Lettau and Ludvigson, 2001).

All the special cases of the fundamental value equation that we have discussed so far impose the simplifying but very unrealistic assumption that markets are “complete”, ie, investors have insured or hedged away all personal risks, and the only risks that affect their inter-temporal marginal rates of substitution (IMRS) are aggregate, market-wide or economy-wide risks. This too is obviously an extreme simplification, so it is worth seeing if removing the simplification works. Duffie and Constantinides (1996), for example, investigate a model in which the *cross-sectional* dispersion of labour income growth matters. Given that the *overall* income is what it is, it is a “bad time” if there is a lot of cross-sectional risk; you might become very rich, but you might also become very poor.

In addition, researchers such as Pastor and Stambaugh (2001) are finally documenting the importance of liquidity. Once dismissed as an institutional friction that is “assumed away” in complete markets, it seems that assets paying off poorly in times of poor market liquidity must pay higher average returns, *viz*, the discount factor is affected by liquidity. The marginal utility of a US dollar, delivered in the middle of a market meltdown such as after the Russian bond default and LTCM collapse, may well have been very high.

The current state of affairs – better than it looks

Unfortunately, no single empirical representation “wins”, and the quest for a simple, reliable and commonly accepted implementation of the fundamental value equation continues. Models that are theoretically purer or that work over a wider range of applications tend to do worse in any given application and sample than models which are motivated by a specific application and sample. The “right” model, even if we had it, would take a long time to emerge relative to the large number of spurious fish in each pond.

This survey looks maddeningly tangled, with a long (and yet woefully incomplete!) list of approaches. But this appearance hides

an exciting common theoretical and empirical consensus that has emerged from all this work: in addition to the market return specified by the CAPM, there are a few additional important dimensions of risk that drive premia in asset markets. If you take on assets with high betas on these risks, you will get higher returns on average, but these assets will all collapse together at times that many investors find very inconvenient. Those “times” are sometimes related to recessions – when peoples’ job prospects are risky, wages are doing poorly, investment and new business formation are at a standstill – and at other times related to financial distress, poor market liquidity, and the like – when a US dollar in your pocket or an easily-liquidatable investment with a high price would be particularly convenient. The long list of empirical approaches mentioned above really only disagree about which particular data series to use to construct the best indicators of these two kinds of “bad times” unrelated to overall market returns.

FIXED INCOME AND COMMODITIES

The models used to price fixed income assets, commodities, and derivatives often simply amount to fairly *ad hoc* discount factor models. Consider first the factor models or affine models that dominate bond pricing, following Vasicek (1977), Brennan and Schwartz (1979), and Cox, Ingersoll and Ross (1985). In principle, valuing a bond is easier than valuing other securities, because the payoff is fixed. A one-year discount bond that will definitely pay one US dollar in one year’s time has a value today of simply

$$p_t^{(1)} = E_t[m_{t+1}] \quad (13)$$

where the superscript denotes maturity. To generate a bond-pricing model, you write down some model for the discount factor and then take expectations.

What makes bond pricing interesting and technically challenging is the presence of many maturities. For example, you can think of a two-period bond as a security whose payoff is a one-period bond, or as a two-period security directly. Its price is

$$p_t^{(2)} = E_t[m_{t+1} p_{t+1}^{(1)}] = E_t[m_{t+1} m_{t+2}] \quad (14)$$

Pursuing the first equality in Equation (14), you can see an interesting recursion developing, leading to differential equations for prices (across maturity). Pursuing the second equality, you can see interesting expectations or integrals to take.

Still, we need a discount factor model to value bonds; so bond-pricing models all come down to discount factor models. The simplest example with which to show this point concretely is the discrete-time Vasicek model.¹³ We write the following time series model for the discount factor:

$$\begin{aligned} \ln(m_{t+1}) &= -x_t + \varepsilon_{t+1} \\ x_{t+1} &= (1 - \phi)\mu + \phi x_t + \delta_{t+1} \end{aligned}$$

where ϕ and μ are parameters and ε and δ are shock/error terms. We then find the prices and yields of one and two period bonds from the fundamental value equation, Equation (13), and Equation (14).¹⁴ The result is a "one-factor model" for yields:

$$\begin{aligned} y_t^{(1)} &= (1 - \phi)\mu_y + y_{t-1}^{(1)} + \delta_{t+1} \\ y_t^{(2)} &= \text{const} + \frac{1}{2}(1 + \phi)y_t^{(1)} + \frac{1}{2}\text{cov}(\delta, \varepsilon) \end{aligned}$$

The short rate $y_t^{(1)}$ evolves on its own as an autoregressive order one process (AR(1)). The two-period bond yield, and all other yields, are then linear functions of the one-year yield. All yields move in lockstep. "Two-factor" and "multi-factor" models work in the same way. For example, the Brennan and Schwartz (1979) two factor model has a long rate as well as a short rate moving autonomously and then all other yields following as functions of these two.

This model is *not* composed of an AR (1) for the short rate plus "arbitrage." The fly in the ointment is the third term in the last equation $\text{cov}(\delta, \varepsilon)$; this is the "market price of interest rate risk". It is the covariance of interest rate shocks with the discount factor. From Equation (7), we recognise it as the risk premium that an asset must pay whose payoff goes up and down with the interest rate. As this term varies, the two-year bond yield can take on any value. Term structure models typically just estimate this term as a free parameter – ie, the models pick this term to fit bond yields as well as possible.

That's fine as far as it goes, but it does not obviate what we are doing here. Just as before, $\text{cov}(\delta, \varepsilon)$ specifies the "systematic" part of interest rate risk. It specifies whether the marginal utility of consumption is higher or lower when interest rates rise unexpectedly. It specifies whether higher interest rates correspond to "good times" or "bad times", and how much so. Term structure models are no more immune from assumptions about consumption, macroeconomic risks, and so forth than anywhere else. Current term structure models are much like financial multi-factor models, discussed in the last section. The discount factor depends on rather arbitrary portfolio returns, selected for empirical fit (in sample) rather than even armchair theorising about good and bad states.

Often the discount factor modelling is implicit in a transformation to "risk-neutral probabilities". Recall that we can write the fundamental value equation as

$$p_t = E_t[m_{t+1}x_{t+1}] = \frac{1}{R^f} E_t^*[x_{t+1}]$$

where the * indicates a expectation under altered "risk-neutral" probabilities. The transformation from E_t to E_t^* is called a "change of measure", and m_{t+1} is the transformation function. Algebraic manipulations can be much easier after the change of measure, but the economic content and implicit discount factor modelling are not changed. Exactly the same information must go in to forming the change of measure that goes into specifying the discount factor or market price of risk. In the same way, bond price authors often do not present the discount factor, but go directly to assumptions about the market price of risk, which we have labelled $\text{cov}(\delta, \varepsilon)$.

Similarly, the Gibson and Schwartz (1990) model for valuing long-term oil derivatives is based on oil price movements and changes in the convenience yield. Their model requires a "market price of oil price risk" and the "market price of convenience yield risk". These parameters are usually estimated to make the model fit well. Again, they are equivalent to specifying a discount factor.

The Keynesian risk premium on forward contracts is a third example of a commonly discussed "risk premium" equivalent to a statement about discount factors. In his *Treatise on Money* (Keynes, 1930), Keynes argued that speculators demand a risk premium

from hedgers in order to accept a risk that hedgers want to “insure away”. Keynes further argued that because hedgers tend to be long the underlying asset on average, they tend to be short forwards on average. The risk premium thus indicates a systematic bias of forward prices to be *above* expected future spot prices, so that for a forward purchase contract with fixed price K requiring the future delivery on an asset worth P_{t+1} ,

$$K = E_t[P_{t+1}] + \Theta$$

where Θ is the Keynesian risk premium, such that $\Theta > 0$.

Surprisingly, many people still adopt the Keynesian view that futures and forwards are biased predictors of future spot prices because speculators require a risk premium from hedgers. Yet, we saw earlier that the fixed price in a forward should just be

$$K = E_t[P_{t+1}] + \text{cov}_t(m_{t+1}, P_{t+1})R^f$$

The Keynesian risk premium amounts to an assumption about discount factors just like any other risk premium. The idea that *hedgers'* specific exposure drives the risk premium is an interesting one. It implies that both hedgers and investors at large are not perfectly diversified – that hedger's marginal utility drives risk premia in this market, even though aggregate marginal utility might produce no risk premium at all.

ARBITRAGE AND NEAR-ARBITRAGE PRICING

So far, we have only discussed “absolute” pricing methods. These methods specify a discount factor that can in principle price any asset, using only fundamental information (ie, the source of aggregate risks). In many applications, that is a far more powerful tool than is called for. Usually, we do not need to value every asset, we only want to value one asset, and we are happy to use the information about the prices of similar assets in order to do so. In this case a *relative* pricing approach is useful. To find the value of a McDonald's hamburger, absolute pricing starts thinking about how much it costs to feed a cow. Relative pricing looks at the price of a hamburger at Burger King. For many purposes, such as deciding where to eat, this is good enough. In finance, option valuation and corporate finance (the use of

comparable investments to determine required rates of return) are the prime applications of relative pricing methods.

The central question is, as always, how to construct a discount factor. The relative pricing approach uses information from *other* asset prices in order to construct a discount factor useful for pricing a *given* asset.

Arbitrage pricing

Arbitrage pricing is the purest case of relative pricing, as it makes the least assumptions about investors, utility functions, and so forth in specifying the discount factor. When it works, this approach neatly cuts short the endless discussion over what are the true risk factors, market price of risk, and so on. Black–Scholes option pricing is the canonical example. The Black–Scholes formula expresses the option price *given* the stock and bond prices.

The only assumption we need to derive an arbitrage pricing relation such as the Black–Scholes formula is that there is *some* discount factor that generates the price of the focus asset (option) and the basis assets (stock, bond). As long as there is some discount factor, then the “Law of One Price” must hold: two ways of generating the same payoff must have the same value.¹⁵ If payoffs x , y , z are related by $z = x + y$, then their prices must obey $p(z) = p(x) + p(y)$.¹⁶ The key insight in the Black–Scholes formula is that you can dynamically hedge an option with a stock and a bond. The payoff of the option is the same as the payoff of the hedge portfolio. Hence, the price of the option must be the same as the value of the hedge portfolio. (Arbitrage pricing should really be called “Law of One Price pricing”, and probably would be if the latter were not such an ugly a name.)

Another way to look at the same thing paves the way for more complex relative pricing. Because the existence of *any* discount factor implies the Law of One Price, *any* discount factor that prices the basis assets (stock and bond) must give the same result for the focus asset (option). Following this insight, we can price options by simply constructing any discount factor that prices the stock and bond. This task is easy. For example, once you know p_t , the choice

$$m_{t+1}^* = x'_{t+1} E_t(x_{t+1} x'_{t+1})^{-1} p_t$$

does a pretty good job of satisfying $p_t = E_t(x_{t+1} m_{t+1}^*)$! (The primes denote transpose and allow for vectors of prices and payoffs in the formula.) Then we can simply use the discount factor m^* to value the option.¹⁷

The discount factor m^* is not unique. There are many discount factors that price the stock and bond. For example, a new discount factor $m_{t+1} = m_{t+1}^* + \varepsilon_{t+1}$ where ε_{t+1} is any random variable uncorrelated with payoffs $E_t(\varepsilon_{t+1} x_{t+1}) = 0$ will do. But with arbitrage pricing it does not matter which one you use. They all give the same option value.

Arbitrage pricing is still technically challenging, because these trivial-sounding statements hold at every point in time, and you have to chain it all back from expiration to find the actual option price. This means solving a differential equation or an integral. But technical challenges are a lot easier to solve than economic challenges – ie, finding the right absolute asset pricing model.

You can see the attraction of arbitrage pricing. Rather than learn about discount factors from macroeconomics, introspection, philosophy, or oversimplified economic theories, we can simply construct useful discount factors from available assets and use them to price derivatives. Put another way, every asset pricing model posits that there is *some* discount factor, so implications that derive from the mere existence of a discount factor are common to every asset pricing model; we do not have to choose which asset pricing model to use if all discount factors give the same answer. Arbitrage pricing seems so pure that option pricing theorists, financial engineers, and risk managers often sneer at the models we have presented above.

Arbitrage pricing, however, is not completely assumption-free. It assumes that there is *some* discount factor. In turn, this assumption requires that there is at least one unconstrained investor out there forming an optimal portfolio. We need to know nothing about the utility function and consumption stream (ie, what states of nature the investor fears) – we will learn all we need to know about that from stock and bond prices – but we do need something. The Law of One Price is routinely violated in retail stores – the price of a 907.2 gram bottle of ketchup is *not* twice the price of a 453.6 gram bottle – so it does reflect some assumptions. It is not a law of nature.

Much more seriously, application of relative pricing techniques to real-world problems is not nearly as straightforward as the Black-Scholes example suggests. The lost car keys are usually not right under the streetlight. In practically every interesting application, even a textbook-perfect hedge is exposed to *some* risk arising from institutional frictions, transactions costs, illiquidity, and so forth. More often, there is no textbook-perfect hedge, due to non-marketed risks such as changing volatility, shifting interest rates, asset specific liquidity premia, non-marketed fundamental securities, etc. And when there is no way to *perfectly* replicate the payoff of the focus asset with some portfolio of other assets, there is no way to perfectly "arbitrage price" the focus asset. The premia for market prices of the unavoidable basis risks will matter. Different, and apparently arbitrary, choices among the many discount factors that price hedge assets (different choices of ε in $m = m^* + \varepsilon$) produce different valuations for the derivatives. We cannot avoid the questions: "how big are the extra risks" and "what is the premium for those extra risks?"

At this point, we could simply return to the beginning and try to answer these questions using some implementation of the fundamental value equation. But we usually want to avoid an extended discussion of CAPM versus ICAPM and consumption in every little application. Many option-pricing exercises leave "market price of risk" assumptions as free parameters, as we discussed above for term structure and commodity models. But in many cases, the market price of risk assumptions matter a lot, and pulling them out of thin air is very unsatisfactory. Instead, we can add a little "absolute" information to winnow down the range of possible discount factors, while still using the information in hedge assets (stock and bond) as much as possible. This works because in most options pricing applications, the residual or tracking error risks are *small*. It only takes a little discount factor economics to make sure that small residuals have small effects on values.

Arbitrage bounds

What can we say about discount factors? The weakest thing we can say in general (beyond existence) is that investors always like more over less.¹⁸ Marginal utility is positive, and this implies that the discount factor is positive. Keep in mind that the discount factor is

random, so "positive" means "positive in every state of the world at time $t + 1$, no matter what happens".

As the existence of a discount factor implies the Law of One Price, a positive discount factor has a nice portfolio interpretation: the "principle of no arbitrage". If a payoff x cannot be negative and might be positive, then it gets a positive price. (Multiplying two positive things in $E_t(m_{t+1} x_{t+1})$ and taking the average, we must get a positive result. In finance terminology, this stronger property is called "arbitrage". Colloquial use of "arbitrage" usually refers to the Law of One Price).

Positive discount factors lead to "arbitrage bounds" on option prices when we cannot completely hedge the option payoff. We solve the problem: "what are the largest and smallest option prices we can generate, searching over all positive discount factors that price stock and bond?" More formally, we solve

$$\begin{aligned}
 & \max_{\{m_{t+1}\}} E_t(m_{t+1} x_{t+1}^{\text{option}}) \\
 & \text{s.t. } p_t^{\text{stock}} = E_t(m_{t+1} x_{t+1}^{\text{stock}}) \\
 & \text{s.t. } p_t^{\text{bond}} = E_t(m_{t+1} x_{t+1}^{\text{bond}}) \\
 & \text{s.t. } m_{t+1} > 0
 \end{aligned} \tag{15}$$

Most textbooks solve this problem more simply for a simple European call option. For example, we note that the call option payoff is always positive, so its price must always be positive. In more complex situations, this discount factor search (a linear program) is the only way to make sure you have not forgotten some clever dominating portfolio, and it can provide useful arbitrage bounds even in dynamic applications (see Ritchken, 1985).

BEYOND ARBITRAGE: A LITTLE BIT OF ABSOLUTE PRICING GOES A LONG WAY

Alas, arbitrage bounds are too wide for many applications. They still allow us to generate weird option prices, because there are weird positive discount factors that nonetheless price the stock and bond. For example, we generate the lower arbitrage bound on a European call option $C(t) > 0$ by imagining a discount factor

arbitrarily close to zero any time the option finishes in-the-money. Now we are all happy when the stock market goes up, but so wildly happy that more money has become *worthless*? Surely we can sensibly restrict the discount factor more than that without getting back into the messy model business.

Following this insight, a number of approaches suggest how to use a *little* absolute pricing even in traditional relative pricing situations in order to at least bound the effects of un-hedgeable residual risks. Equivalently, we can combine information about the discount factor from basis assets (stock, bond) whose prices we do not want to question with relatively weak but hence robust information about the discount factor available from economic theory and practical experience in many markets.

The arbitrage pricing theory

Ross's (1976b) arbitrage pricing theory (APT) is the first such mixture of a little absolute pricing into a fundamentally relative-pricing problem. His APT is also a second source of inspiration for the factors in multi-factor models.

Ross pointed out that many *portfolios* of stocks can be reasonably approximated as linear combinations of the return on a few basic or "factor" portfolios. In equations, we can run a regression of the focus portfolio on factor portfolios,

$$R_{t+1}^i = a + \beta_1^i f_{t+1}^1 + \beta_2^i f_{t+1}^2 + \dots + \varepsilon_{t+1}^i$$

and the error term will be small. For example, most of the thousands of mutual funds' returns can be quite well approximated once we know the funds' style in terms of market, size, value, and a few industry groupings. If the actual, ex-post, returns on these portfolios can be approximated in this way, it stands to reason that the expected returns can also be so approximated. If not, one could buy the focus portfolio, short the right hand side combination of the factor portfolio returns, and earn a high return with little risk. As a result we derive a multi-factor representation in which average returns depend on the betas on the factor portfolios.

$$E(R_{t+1}^i) = R^f + \beta_1^i \lambda_1 + \beta_2^i \lambda_2 + \dots \quad (16)$$

Equivalently, a discount factor that is a linear function of the factor returns will do a good job of pricing the focus portfolios. Of course, like any relative pricing approach, the APT's applicability is limited. You cannot apply it to assets whose returns are poorly approximated by the returns of the few basic portfolios – assets with large residuals ε_{t+1}^i .

Unfortunately, if you are willing to say nothing at all about absolute pricing – utility functions, risk aversion, macroeconomic states, etc. – then *any* residual risk ε_{t+1}^i can have an *arbitrarily* large price or risk premium, and the hoped for APT approximation can be arbitrarily wrong. With any error, the Law of One Price alone says nothing about the focus portfolio. For stock portfolios, arbitrage (positive discount factors) does not help, as no portfolio of stocks does better than another portfolio *always*.

Ross realised a way out of this dilemma. If the risk premium associated with a residual ε_{t+1}^i were very large, it would be a very attractive investment in terms of its Sharpe ratio (ratio of mean return to standard deviation). A high Sharpe ratio is not an arbitrage opportunity or a violation of the Law of One Price, but extremely high Sharpe ratios are nonetheless unlikely to persist. If we rule out very high Sharpe ratios *in addition to* the Law of One Price and principle of no arbitrage, we do obtain a well-behaved approximate APT. Small errors ε_{t+1}^i now must mean small risk premia, so Equation (16) will hold as a good approximation.

Ruling out high Sharpe ratios is another little bit of absolute pricing. It is equivalent to the assumption that discount factors are not too *volatile*. Precisely, Hansen and Jagannathan (1991) show that the maximum possible Sharpe ratio attained by all assets priced by a particular discount factor is given by ¹⁹

$$\sigma(m) = \frac{1}{R^f} \max_{\{R:s.t.1=E(mR)\}} \left\{ \frac{E(R) - R^f}{\sigma(R)} \right\}$$

Limiting volatility is an additional but plausible and mild restriction on marginal utility. Does marginal utility growth – growth in the pleasure we get from an extra US dollar's consumption – really vary by more than, say, 50% per year? The historical market Sharpe ratio is 0.5 (8% mean, 16% standard deviation), so even such a high

bound on discount factor volatility is enough to restrict market prices of risk to CAPM values.

In more economic terms, the volatility of the discount factor is given by the volatility of consumption growth times the risk aversion coefficient. If we know that at least one marginal investors' consumption growth varies less than, say, 5% per year, and risk aversion is sensible – say, less than 10 – then we know that the discount factor varies by less than 0.5 per year, and the maximum Sharpe ratio should be less than 0.5. Equivalently, we might be willing to assume that “traders will take any Sharpe ratio more than twice the Sharpe ratio of the market as a whole” and impose that maximum Sharpe ratio in evaluating the premium for residual risks.

Derivatives valuation bounds

Cochrane and Saá-Requejo (2001) apply Ross’s idea to option pricing, when either market frictions (eg, you cannot trade continuously) or non-marketed risks (eg, stochastic volatility or interest rates) break simple arbitrage pricing and require us to evaluate market prices of risks. Cochrane and Saá-Requejo find that the upper and lower bounds on option prices (searching over all discount factors that price stock and bond) are positive *and* have limited volatility. This amounts to adding an additional restriction to arbitrage bound Equation (15):

$$\begin{aligned}
 & \max_{\{m_{t+1}\}} E_t \left(m_{t+1} x_{t+1}^{\text{option}} \right) \\
 & \text{s.t. } p_t^{\text{stock}} = E_t \left(m_{t+1} x_{t+1}^{\text{stock}} \right) \\
 & \text{s.t. } p_t^{\text{bond}} = E_t \left(m_{t+1} x_{t+1}^{\text{bond}} \right) \\
 & \text{s.t. } m_{t+1} > 0 \\
 & \text{s.t. } \sigma(m) \leq R^f h
 \end{aligned}
 \tag{17}$$

The last restriction is the novelty over arbitrage bounds. *h* is the upper limit on discount factor volatility; the extra assumption is that investors would want to take any bet with a Sharpe ratio greater than *h*. Cochrane and Saá-Requejo (2001) find that the resulting bounds on option prices are much tighter than the arbitrage bounds that result from ignoring the last term – ie, the bounds that result from arbitrary assignment of the market price of residual risk.

For option pricing, both positive discount factors and a limit on discount factor volatility are important. So far, the discount factor interpretation has been a matter of aesthetics. You could solve problems (for example) imposing a positive discount factor or by checking that all positive payoffs had positive prices. The restrictions in Equation (17) have no pure portfolio interpretation. The only way to put all these ideas together is to add restrictions on the discount factor, tightening the bounds on option pricing. They have to be posed and solved with discount factor methods.

This is only the beginning. Bernardo and Ledoit (2000) add the restriction that discount factors cannot be too small or too large, $a \leq m \leq b$. This is a sensible tightening of the arbitrage restriction $m \geq 0$. It also produces usefully tight option pricing bounds. Constantinides and Zariphopoulou (1999) consider the sensible restriction that higher index values must make investors happier. The discount factor m thus must be a monotonically decreasing function of the stock index. The beauty of a discount factor framework is that it is easy to add all these restrictions and more together.

CONCLUSION

All of asset pricing comes down to one simple idea: price equals expected discounted payoff. The art is in what one should use for a discount factor.

We surveyed "absolute" approaches such as the CAPM and consumption CAPM that infer the discount factor from measures of macroeconomic or financial "bad times". We emphasised that a sound grasp of asset pricing theory is required to define systematic risk and thus to identify those remaining risks on which investors and firms should focus their risk management efforts, and to define "good risks" which give high returns to investors who take them. We surveyed term structure and derivatives models that specify much simpler and more ad hoc discount factor models, resulting in free parameters for market prices of risk. We surveyed "relative pricing" approaches that learn about the discount factor for pricing one asset from information in other assets. We argued that a little bit of economics, a little bit of absolute pricing (such as a limit on discount factor volatility) can help to plug the market price of risk holes in much applied option pricing and resulting risk management.

The fundamental value equation has significant implications for finance in general and for portfolio selection, capital budgeting, hedge evaluation, and derivatives valuation applications in particular. Far from being distantly related to risk management, the consumption model and all its special cases and empirical representations are, in fact, as central to risk management as they are to all finance.

APPENDIX: DERIVATION OF THE FUNDAMENTAL VALUE EQUATION

To describe what an investor likes and dislikes, and hence to think about how the investor values an asset, we employ the standard economist's model that investors want to find the highest value of a "utility function":²⁰

$$U(c_t, c_{t+1}) = u(c_t) + \beta E_t[u(c_{t+1})]$$

$u(c_t)$ describes how more consumption at time t makes the investor happier. We typically assume that investors always prefer more to less ($u_c(c) > 0$), and each incremental unit of consumption brings slightly less happiness than the unit before it ($u_{cc}(c) < 0$).

Now think of a financial asset whose price at time t is p_t and whose payoff (total value) at time $t + 1$ is x_{t+1} . The investor can freely buy or sell as much of this asset as they like at time t . How much will they buy or sell? To find the answer, denote $e(t)$ as their consumption level before they buy any of the asset, and denote by ξ the amount of the asset they choose to buy. Then, their problem is to choose the ξ that solves

$$\begin{aligned} \max_{\xi} & u(c_t) + E_t [\beta u(c_{t+1})] \\ \text{s.t.} & c_t = e_t - \xi p_t \\ \text{s.t.} & c_{t+1} = e_{t+1} + \xi x_{t+1} \end{aligned}$$

Substitute the two constraints into the objective, and take a derivative with respect to the ξ . Set the derivative equal to zero to characterise the maximum. The result is that the investor's optimal consumption-investment choice satisfies

$$p_t u'(c_t) = E_t [\beta u'(c_{t+1}) x_{t+1}]$$

or, rearranging with price on the left and the rest on the right,

$$p_t = E_t \left[\beta \frac{u_c(c_{t+1})}{u_c(c_t)} x_{t+1} \right]$$

- 1 Culp (2003) explores the relative-pricing approach in the subsequent chapter of this volume.
- 2 Winners of the Nobel Prize in Economic Sciences for their specific contributions to the theory of finance are Harry M. Markowitz (1990), Merton H. Miller (1990), William F. Sharpe (1990), Robert C. Merton (1997), and Myron S. Scholes (1997). Other Nobel laureates whose work had a critical impact on finance include John R. Hicks (1972), Kenneth J. Arrow (1972), Gerard Debreu (1983), Franco Modigliani (1985), and Robert E. Lucas, Jr. (1995).
- 3 The discussion here and in the next two sections is adapted from Cochrane (2001).
- 4 As a matter of theory, this only really works for investors. Corporations do not have utility functions and thus engage in a different type of analysis of hedging opportunities – see Culp (2001, 2002a).
- 5 Derivation: For any two returns R^j and R^k , $E_t[m_{t+1}(R^j - R^k)] = E_t[m_{t+1}R^j] - E_t[m_{t+1}R^k] = 0$, where $R^k = R^j$ is a special case.
- 6 Derivation: Note that $R^j = 1/E(m)$ and $E(mP) = E(m)E(P) + \text{cov}(m, P)$ and solve.
- 7 See, for example, Ross (1976b) and Cox, Ross, and Rubinstein (1979). Cochrane (2001) discusses the relations between discrete-time and continuous-time formulations of the problem and how to implement discount factors in continuous time.
- 8 Alternatively, that is why investors diversify their holdings to get rid of these risks on their own. See Culp (2003).
- 9 The parameter γ can be interpreted as the degree of risk aversion.
- 10 Cochrane (2001) shows different derivations of the CAPM under alternative assumptions and provides a comparison of the multiple approaches.
- 11 z is usually a vector of multiple state variables and factor-mimicking portfolios.
- 12 Technically, this statement requires a risk aversion coefficient greater than one, but that is the usual case.
- 13 This treatment is adapted from Campell, Lo, and MacKinlay (1997).
- 14 Derivation: The price and yield of a one-period bond are

$$p_t^{(1)} = E_t[m_{t+1}] = E_t[e^{\ln(m_{t+1})}] = E_t[e^{-x_t + \varepsilon_{t+1}}] = e^{-x_t + \frac{1}{2}\sigma_\varepsilon^2}$$

$$y_t^{(1)} = -\ln(p_t) = x_t - \frac{1}{2}\sigma_\varepsilon^2$$

With an adjustment to the constant μ , the state variable x thus is the short rate $y_t^{(1)}$. Things get more interesting with a two-year bond:

$$y_t^{(2)} = -\frac{1}{2} \ln E_t[m_{t+1} m_{t+2}] = -\frac{1}{2} \ln E_t[e^{-x_t - x_{t+1} + \varepsilon_{t+1} + \varepsilon_{t+2}}] = -\frac{1}{2} \ln E_t[e^{-(1+\phi)x_t - (1-\phi)\mu - \delta_{t+1} + \varepsilon_{t+1} + \varepsilon_{t+2}}]$$

$$y_t^{(2)} = \frac{1}{2}(1+\phi)x_t + \frac{1}{2}(1-\phi)\mu + \frac{1}{2}\text{cov}(\delta, \varepsilon) - \frac{1}{4}\sigma_\delta^2 - \frac{1}{2}\sigma_\varepsilon^2$$

$$y_t^{(2)} = \frac{1}{2}(1-\phi)\mu + \frac{1}{2}(1+\phi)y_t^{(1)} + \frac{1}{2}\text{cov}(\delta, \varepsilon) - \frac{1}{4}\{\sigma_\delta^2 + (1-\phi)(\sigma_\varepsilon^2)\}$$

- 15 The converse statement is one of the most famous founding theorems of finance. If the Law of One Price holds, then there exists a discount factor. See Ross (1976), Harrison and Kreps (1979), and Hansen and Richard (1987).

- 16 Proof: $E[m(x + y)] = E(mx) + E(my)$.
 17 See Cochrane (2001) for the algebra.
 18 You can always burn or give away what you do not want, so less is never preferred to more.
 19 Proof: $0 = E(m(R - R^f)) = E(m)E(R - R^f) + \sigma(m)\sigma(R)\rho(m, R)$

$$E(m) \frac{E(R - R^f)}{\sigma(R)\rho(m, R)} = \sigma(m)$$

Correlations are less than one.

- 20 This simple formulation treats all consumers as alike and presumes that utility is "additively separable" across time. Numerous alternatives to this simple set-up have been proposed. See, for example, Constantinides (1990). Many alternatives are reviewed in Cochrane (1997, 1999a, 1999b).

BIBLIOGRAPHY

- Banz, R. W.**, 1981, "The Relationship Between Return and Market Value of Common Stocks", *Journal of Financial Economics* 9, pp. 3-18.
- Black, F.**, 1972, "Capital Market Equilibrium with Restricted Borrowing", *Journal of Business* 45, pp. 444-55.
- Black, F., and M. Scholes**, 1973, "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy* 81, pp. 637-54.
- Bernardo, A. E., and O. Ledoit**, 2000, "Gain, Loss, and Asset Pricing", *Journal of Political Economy* 108, pp. 144-72.
- Billingsley, P.**, 1995, *Probability and Measure* (New York: Wiley Interscience).
- Breeden, D.**, 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities", *Journal of Financial Economics* 7, pp. 265-96.
- Brennan, M. J., and E. S. Schwartz**, 1979, "A Continuous-Time Approach to the Pricing of Bonds", *Journal of Banking and Finance* 3, pp. 133-55.
- Campbell, J., A. Lo, and C. MacKinlay**, 1996, *The Econometrics of Financial Markets* (Princeton University Press).
- Chen, N.-F., R. Roll, and S. A. Ross**, 1986, "Economic Forces and the Stock Market", *Journal of Business* 59, pp. 383-403.
- Cochrane, J. H.**, 1991a, "Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations", *Journal of Finance* 46, pp. 207-34.
- Cochrane, J. H.**, 1991b, "Volatility Tests and Efficient Markets: A Review Essay", *Journal of Monetary Economics* 27, pp. 463-85.
- Cochrane, J. H.**, 1996, "A Cross-Sectional Test of an Investment-Based Asset Pricing Model", *Journal of Political Economy* 104, pp. 572-621.
- Cochrane, J. H.**, 1997, "Where is the Market Going? Uncertain Facts and Novel Theories", *Federal Reserve Bank of Chicago Economic Perspectives* 21, pp. 3-37.
- Cochrane, J. H.**, 1999a, "New Facts in Finance", *Federal Reserve Bank of Chicago Economic Perspectives* 23, pp. 36-58.

- Cochrane, J. H., 1999b, "Portfolio Advice for a Multi-factor World", *Federal Reserve Bank of Chicago Economic Perspectives*, 23, pp. 59–78.
- Cochrane, J. H., 2001, *Asset Pricing* (Princeton University Press).
- Cochrane, J. H., and J. Saá-Requejo, 2000, "Beyond Arbitrage: Good Deal Asset Price Bounds in Incomplete Markets", *Journal of Political Economy* 108, pp. 79–119.
- Constantinides, G. M., 1990, "Habit Formation: A Resolution of the Equity Premium Puzzle", *Journal of Political Economy* 98, pp. 519–43.
- Constantinides, G. M., and D. Duffie, 1996, "Asset Pricing with Heterogeneous Consumers", *Journal of Political Economy* 104, pp. 219–40.
- Constantinides, G. M., and T. Zariphopoulou, 1999, "Bounds on Prices of Contingent Claims in an Intertemporal Economy with Proportional Transaction Costs and General Preferences", *Finance and Stochastics* 3, pp. 345–69.
- Cox, J. C., S. A. Ross, and M. Rubinstein, 1979, "Option Pricing: A Simplified Approach", *Journal of Financial Economics* 7, pp. 229–63.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, "A Theory of the Term Structure of Interest Rates", *Econometrica* 53, pp. 385–408.
- Culp, C. L., 2001, *The Risk Management Process: Business Strategy and Tactics* (New York: John Wiley and Sons).
- Culp, C. L., 2002a, *The ART of Risk Management: Alternative Risk Transfer, Capital Structure, and the Convergence of Insurance and Capital Markets* (New York: John Wiley and Sons).
- Culp, C. L., 2002b, "Normal Backwardation (Theory of)", in *Encyclopedia of Financial Engineering and Risk Management* (London: Fitzroy-Dearborn Publishers, forthcoming).
- Culp, C. L., 2003, "Modigliani-Miller Propositions", in *Modern Risk Management: A History* (London: Risk Books, forthcoming), in this volume.
- Fama, E. F., 1991, "Efficient Markets II", *Journal of Finance* 46, pp. 1575–1618.
- Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics* 33, pp. 3–56.
- Fama, E. F., and K. R. French, 1996, "Multi-factor Explanations of Asset-Pricing Anomalies", *Journal of Finance* 47, pp. 426–65.
- Ferson, W., and C. R. Harvey, 1999, "Conditioning Variables and the Cross-Section of Stock Returns", *Journal of Finance* 54, pp. 1325–60.
- Gibson, R., and E. S. Schwartz, 1990, "Stochastic Convenience Yield and the Pricing of Oil Contingent Claims", *Journal of Finance* 45, pp. 959–76.
- Hansen, L. P., and R. Jagannathan, 1991, "Implications of Security Market Data for Models of Dynamic Economies", *Journal of Political Economy* 99, pp. 225–62.
- Hansen, L. P., and S. F. Richard, 1987, "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models", *Econometrica* 55, pp. 557–90.
- Hansen, L. P., and K. J. Singleton, 1982, "Generalised Instrumental Variables Estimation of Nonlinear Rational Expectations Models", *Econometrica* 50, pp. 1269–86.
- Harrison, J. M., and D. M. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory* 20, pp. 381–408.

- Ho, T. S. Y., and S.-B. Lee, 1983, "Term Structure Movements and Pricing Interest Rate Contingent Claims", *Journal of Finance* 41, pp. 1011-29.
- Jagannathan, R., and Z. Wang, 1996, "The Conditional CAPM and the Cross-Section of Expected Returns", *Journal of Finance* 51, pp. 3-53.
- Keynes, J. M., 1930, "A Treatise on Money", (Macmillan).
- Lettau, M., and S. C. Ludvigson, 2001, "Consumption, Aggregate Wealth, and Expected Stock Returns", *Journal of Finance* 56, pp. 815-49.
- 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, pp. 13-37.
- Lucas, R. E., 1978, "Asset Prices in an Exchange Economy", *Econometrica* 46, pp. 1429-46.
- Merton, R. W., 1973, "An Intertemporal Capital Asset Pricing Model", *Econometrica* 41, pp. 867-87.
- Modigliani, F., and M. H. Miller, 1958, "The Cost of Capital, Corporation Finance, and the Theory of Investment", *American Economic Review* 48(2), pp. 261-97.
- Pastor, L., and R. F. Stambaugh, 2001, "Liquidity Risk and Expected Stock Returns", NBER Working Paper No. w8462 (September). Forthcoming *Journal of Political Economy*.
- Ritchken, P. H., 1985, "On Option Pricing Bounds", *Journal of Finance* 40, pp. 1219-33.
- Roll, R., 1977, "A Critique of the Asset Pricing Theory's Tests' Part I: On Past and Potential Testability of the Theory", *Journal of Financial Economics* 4, pp. 129-76.
- Ross, S. A., 1976a, "The Arbitrage Theory of Capital Asset Pricing", *Journal of Economic Theory* 13, pp. 341-60.
- Ross, S. A., 1976b, "Risk, Return and Arbitrage", in I. Friend and J. Bicksler (eds), *Risk and Return in Finance: Volume I*, pp. 189-218 (Cambridge: Ballinger).
- Rubinstein, M., 1976, "The Valuation of Uncertain Income Streams and the Price of Options", *Bell Journal of Economics* 7, pp. 407-25.
- Sharpe, W., 1964, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk", *Journal of Finance* 19, pp. 425-42.
- Vasicek, O., 1977, "An Equilibrium Characterization of the Term Structure", *Journal of Financial Economics* 5, pp. 177-88.