Asset Pricing

Matthew Cocci

October 3, 2013

1. Introduction

If all of Asset Pricing needed to be reduced to one simple, general, yet wholly appropriate epithet, a decent starting point would be "Price equals expected discounted payoff."

To segment the discussion further, John Cochrane suggests the following two classes of pricing approaches in his Asset Pricing book:

1. *Absolute Pricing*: "Pricing assets by exposure to fundamental sources of macroeconomic risk." This approach includes general equilibrium models.

2. *Relative Pricing*: This type of pricing considers only how an asset should be priced relative to other assets that can be bought and sold in the market. Black-Scholes and any arbitrage arguments would fall under this category.

However, many asset pricing approaches are a blend of the two.

To be more specific, we can reduce asset pricing to two equations:

$$p_t = E[m_{t+1}x_{t+1}] \tag{1}$$

$$m_{t+1} = f(\text{data, parameters}) \tag{2}$$

where p_t is the asset price, x_{t+1} is the asset's payoff, and m_{t+1} is the stochastic discount factor. This approach joins together what used to be separate theories so that pricing stocks, bonds, and options now represent special cases of this more general framework summarized by the equations above.

2. Asset Pricing Facts

There are a number of empirical observations that have been made over the years, and this section will detail their conclusions and main insights.

2.1. Equity Premium and Risk

We'll beging by thinking about the *Equity Premium*, which is the returns to stocks above and beyond bonds (so if you borrow in bonds and lend in stocks): $E[R^{\text{stocks}} - R^{\text{bonds}}]$. Here are the facts:

1. The equity premium is *big*, about 7%.
2. There's a lot of risk in stocks: there's a greater standard deviation in returns (although we'll think about risk a bit differently). Stock returns are also correlated with macro variables.
3. Compared with stocks returns, GDP and Consumption growth are much more stable, though they do vary a bit (with the standard deviation being much closer to bond returns' volatility than stock returns).

2.2. Time-Varying Risk Premium

Now we want to consider what explains the Equity Premium. For that, we regress returns on some signal we hope will explain future returns. Here's a few regressions. The main result will be that *excess returns* are forecastable—that is, the time-varying reward for risk can be forecast. We're not forecasting raw rate. Now onto facts:

1. *Old View*: Regression of returns on lagged returns:

   $$R_{t+1} = a + bR_t + \varepsilon_{t+1}$$

   This regresion typically yields a value for b around 0, in which case future returns given returns today are more or less *random*. The main result:

   $$ER_{t+1} = E[a + 0 \cdot R_t + \varepsilon_{t+1}] = a \tag{3}$$

   in which case *expected returns* are large and constant.

2. *New View*: The old view, encapsulated in Equation 3, drove theory for many years. But now, we have a different picture: there are variables that forecast stock returns. Now, we run the regression

   $$R^e_{t\to t+k} = a + b\left(\frac{D_t}{P_t}\right) + \varepsilon_{t+k} \tag{4}$$

   where R^e is the excess return over bonds, D_t is current dividends, and P_t is current price. From this regression, we get a large, positive regression coefficient. So we actually *can* predict returns over the business cycle via this dividend/price ratio (but it's still tough in the short run).

3. *New View Interpretation*: The dividend yield (D_t/P_t), which is the inverse of prices, turns out to vary a lot over time. Moreover, if you buy when the dividend yield ratio is *high* (so that prices are low), you will earn high returns in expectation. Vice versa for a low dividend yield ratio (so high prices).

   Moreover, we can get the standard deviation in the expected returns over time, which is the fitted value $ER^e_{t\to t+k} = a + bE\left(D_t/P_t\right)$ from Equation 4. It turns out that expected returns *also vary*, and they vary *substatially*, at about 6%.

   Recall that it was puzzling that stocks should earn 7% over bonds. Now we're saying that they earn 7% over bonds *and* vary with a standard deviation of about 6% a year. Average *changes* in $ER^e_{t\to t+k}$ are about as much as the average confounding *level*.[1]

4. *Why do prices vary?*: We know that the dividend/price ratio varies a lot, but it's mostly driven by prices. Why? Well, we used to think

$$\text{High Prices} \Rightarrow \text{High Future Growth} \Rightarrow \text{High E[Future Dividends]} \tag{5}$$

   But if we run the regression implied by Equation 5,

$$\frac{D_{t+k}}{D_t} = a + b\left(\frac{D_t}{P_t}\right) + \varepsilon_{t+k} \tag{6}$$

   we find out that the logic of 5 is **wrong**. The regression coefficient in Equation 6 is negligible and there is no explanatory power (as reflected in the R^2). So actaully

$$\text{High Prices} \Rightarrow \text{Low Future E[Returns]} \tag{7}$$

   So it's not that dividends will adjust if prices are high and low—dividends are remarkably stable relative to prices! Instead, *prices* will adjust.

   In short: high prices are correlated with a good economy, when people are willing to bear more risk. So returns will be low. Vice versa for low prices.

[1] Of course, *actual* returns vary even more than expected returns—recall about 17%.

2.3. The Cross Section of Stock Returns

We talked about returns over time in the previous sections. Now let's consider the facts about how returns vary *across assets* at a given time.

1. *Market Cap*: Small stocks earn more on average than large stocks. They are also riskier.

2. *Growth vs. Value*: Value stocks (high book to value) tend to earn more.

3. *Capital Asset Pricing Model*: This model predicts returns based on the correlation between the asset and the market portfolio.

4. *Multifactor Model*: This model was put forward by Fama and French to capture the extra variation in returns that the Capital Asset Pricing Model failed to explain. This is the modern view. Specifically, Fama and French posit

$$ER^{e,i} = \alpha_i + b_i E[R^m - R^f] + h_i E(\text{hml}) + s_i E(\text{smb}) \tag{8}$$

   where $E(\text{hml})$ represents "high minus low" (for value minus growth stocks) and $E(\text{smb})$ for "small minus big" (for market cap).

5. *Stock Market Volatility:* We used to think volatility was pretty constant over time. Now we know that the variance of returns changes over time.

3. Asset Pricing Theory Overview

This section is about understanding the fundamental formula for all of asset pricing:

$$P_t = E_t[M_{t+1}X_{t+1}] \tag{9}$$

But first, since we'll start talking about preferences below, we'll eventually need a utility function ato some point. In order to move the discussion along below, we consider that portion in the next section instead.

3.1. Utility Detour

So first, we'll consider a simple utility function for two-period consumption, which covers t and $t+1$—the relevant time periods in Equation 9. Spefically, we'll define

$$U(C_t, C_{t+1}) = U(C_t) + \beta E_t\left[U(C_{t+1})\right] \tag{10}$$

where β is the discount factor with respect to *time*, **not** risk which depends upon the shape of the utility function. Equation 10 above is a very simple form which we can generalize further if we choose.

For the *internal*, one-period utility function, $U(\cdot)$, we'll want the normal properties of utility, $U' > 0$ and $U'' < 0$, which leads us to define the simple *power function utility*

$$U(C) = \frac{C^{1-\gamma} - 1}{1-\gamma} \quad \Rightarrow \quad U'(C) = C^{-\gamma} \tag{11}$$

This utility function also has the property that as $\gamma \to 1$, $U(C) = \ln C$.

3.2. Understanding the Components

Now let's take each component of Equation 9 in turn:

1. X_{t+1}: This is the *payoff* (in gross terms, not net, percentage, etc.) at time $t+1$. Often, X_{t+1} will be random—it's a random variable.

2. P_t: To get this payoff, X_t, you have to pay some price P_t today. Note that it is possible for P_t to equal 0, like it might in a bet (nothing today, +1 tomorrow if I win, -1 if I lose). No money changes hands.

3. M_{t+1}: Recall Equation 10 above. It left us free parameters, β and γ, which allow us to tweak "impatience" and time sensitivity of consumption (β) as well as risk-aversion via curvature (γ). This gives us everything we need to charactertize the discount factor, M_{t+1}.

3.3. Deriving P=E(MX)

Let's now figure out the price an investor assigns to a risky payoff. It reduces to a utility maximization problem:

$$\max_{\xi} \quad U(C_t - \xi P_t) + \beta E_t \left[U(C_{t+1} + \xi X_{t+1}) \right]$$

4. Classic Issues in Finance

A. Discrete Time Review

A.1. Basic Building Block

The basic building block is random iid noise:

$$\varepsilon_t \sim \text{iid} \qquad E_t\varepsilon_{t+1} = 0 \qquad E_t\varepsilon_{t+1}^2 = \sigma_\varepsilon^2$$

where $E_t(\cdot)$ represents the expectation *conditional* on the information at time t. In practice, that means any variable with a t subscript inside a expectation, variance, co-variance, etc. operator can be taken out because it is *known* at time t.

Often ε_t will be draws from a normal distribution, but they don't have to be.

A.2. Canonical AR(1) Model

More complicated than simple random noise, we have the AR(1) process, which is written in one of either two standard ways, with the latter being more common as we move to continuous time:

$$\begin{aligned} x_{t+1} &= \rho x_t + \varepsilon_{t+1} \qquad E_t\varepsilon_{t+1} = 0 \qquad E_t\varepsilon_{t+1}^2 = \sigma_\varepsilon^2 \\ x_{t+1} &= \rho x_t + \sigma_\varepsilon\varepsilon_{t+1} \qquad E_t\varepsilon_{t+1} = 0 \qquad E_t\varepsilon_{t+1}^2 = 1 \end{aligned}$$

From there, it's easy to show via the linearity property of the expectation operator and by successive substitution that

$$\begin{aligned} E_t x_{t+1} &= E_t[\rho x_t + \varepsilon_{t+1}] \\ &= \rho x_t \\ E_t x_{t+k} &= \rho^k x_t \end{aligned}$$

As for the variance conditional on information at time t, denoted by Var_t, we get

$$\begin{aligned} \text{Var}_t(x_{t+1}) &= \text{Var}_t(\rho x_t + \varepsilon_{t+1}) \\ &= \sigma_\varepsilon^2 \\ \text{Var}_t(x_{t+2}) &= \text{Var}_t(\rho x_{t+1} + \varepsilon_{t+2}) \\ &= \text{Var}_t(\rho x_{t+1}) + \text{Var}_t(\varepsilon_{t+2}) + 2\,\text{Cov}_t(\rho x_{t+1},\ \varepsilon_{t+2}) \\ &= \rho^2\,\text{Var}_t(x_{t+1}) + \text{Var}_t(\varepsilon_{t+2}) + 0 = \rho^2\sigma_\varepsilon^2 + \sigma_\varepsilon^2 \\ &= (1+\rho^2)\sigma_\varepsilon^2 \\ \text{Var}_t(x_{t+k}) &= (1 + \rho^2 + \cdots + \rho^{2(k-1)})\sigma_\varepsilon^2 \end{aligned}$$

The last line we prove by induction.

Proof. In the case of $k = 3$, which we will take as the base case,

$$\begin{aligned}
\text{Var}_t(x_{t+3}) &= \text{Var}_t(\rho x_{t+2} + \varepsilon_{t+3}) \\
&= \rho^2\ \text{Var}_t(x_{t+2}) + \text{Var}_t(\varepsilon_{t+3}) + \text{Cov}_t(x_{t+2},\ \varepsilon_{t+3}) \\
\text{Var}(x_{t+2}) \text{ from above} \qquad &= \rho^2(1+\rho^2)\sigma_\varepsilon^2 + \sigma_\varepsilon^2 + 0 \\
&= (1+\rho^2+\rho^4)\sigma_\varepsilon^2
\end{aligned}$$

So the base case holds. Now assume it holds for arbitrary k, and show that the condition also holds for $k+1$:

$$\begin{aligned}
\text{Var}_t(x_{t+k+1}) &= \text{Var}_t(\rho x_{t+k} + \varepsilon_{t+k+1}) \\
&= \rho^2\ \text{Var}_t(x_{t+k}) + \text{Var}_t(\varepsilon_{t+k+1}) + \text{Cov}_t(x_{t+k},\ \varepsilon_{t+k+1}) \\
&= \rho^2(1+\rho^2+\cdots+\rho^{2(k-1)})\sigma_\varepsilon^2 + \sigma_\varepsilon^2 \\
&= (1+\rho^2+\cdots+\rho^{2([k+1]-1)})\sigma_\varepsilon^2
\end{aligned}$$

so it holds for $k+1$. □

A.3. Canonical MA(1) Model

The basic form of the moving average MA(1) model, with conditional means:

$$x_{t+1} = \varepsilon_{t+1} + \theta\varepsilon_t \qquad E_t\varepsilon_{t+1} = 0, \quad E_t^2\varepsilon_{t+1} = \sigma_\varepsilon^2$$

The first moments for the process, conditional on information at time t, are as follows:

$$\begin{aligned}
\Rightarrow \qquad E_t x_{t+1} &= E_t\varepsilon_{t+1} + E_t[\theta\varepsilon_t] \\
&= \theta\varepsilon_t \\
E_t x_{t+2} &= E_t[\varepsilon_{t+2} + \theta\varepsilon_{t+1}] = E_t[\varepsilon_{t+2}] + E_t[\theta\varepsilon_{t+1}] \\
&= 0 \\
E_t x_{t+k} &= 0
\end{aligned}$$

Now for the conditional variances, dropping the covariance terms which we included above since innovations at different times are always independent:

$$\begin{aligned}
\text{Var}_t(x_{t+1}) &= \text{Var}_t(\varepsilon_{t+1} + \theta\varepsilon_t) \\
&= \text{Var}_t(\varepsilon_{t+1}) + \theta^2\ \text{Var}_t(\varepsilon_t) \\
&= \sigma_\varepsilon^2 \\
\text{Var}_t(x_{t+k}) &= \text{Var}_t(\varepsilon_{t+k} + \theta\varepsilon_{t+k-1}) \\
&= \text{Var}_t(\varepsilon_{t+k}) + \theta^2\text{Var}_t(\varepsilon_{t+k-1}) \\
&= (1+\theta^2)\sigma_\varepsilon^2
\end{aligned}$$

A.4. Unconditioanl Moments

So far, all of the means and variances above were *conditional* on information at time t. So that implies computing $E_t x_{t+1}$ and $\text{Var}_t(x_{t+1})$ via the formulas above requires knowledge of x_t.

Mean: However, suppose you wanted to generate the series from scratch. What is the *unconditional* mean and variance so that you can draw x_t *without* prior values $x_{t-1}, x_{t-2}, \ldots$?

$$\begin{aligned} Ex_t &= E[\rho x_{t-1}] + E\varepsilon_t \\ &= \rho E x_{t-1} + 0 \\ \Rightarrow \quad Ex_t &= 0 \end{aligned}$$

The last line follows because the unconditional average Ex_t must equal the unconditional average Ex_{t-1}, forcing the expecation to be zero.

Variance: Now we do the same thing for the variance. We'll also use the fact that the unconditional variance for x_t must equal that for x_{t-1}, which allows us to solve:

$$\begin{aligned} \text{Var}(x_t) &= \text{Var}(\rho x_{t-1} + \varepsilon_t) \\ &= \rho^2 \,\text{Var}(x_{t-1}) + \text{Var}(\varepsilon_t) \\ \text{Using } \text{Var}(x_{t-1}) = \text{Var}(x_t)\text{:} \quad &= \rho^2 \,\text{Var}(x_t) + \sigma_\varepsilon^2 \\ \Rightarrow \quad \text{Var}(x_t) &= \frac{\sigma_\varepsilon^2}{1-\rho^2} \end{aligned}$$

A.5. Translating Between AR and MA Models

AR to MA: First, you can transform an AR(1) model into a MA(∞) model by "solving" the AR model through successive substitution:

$$\begin{aligned} x_t &= \rho x_{t-1} + \varepsilon_t \\ &= \rho(\rho x_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \rho^2 x_{t-2} + \rho\varepsilon_{t-1}) + \varepsilon_t \\ &= \quad \vdots \qquad \vdots \\ &= \sum_{j=1}^{\infty} \rho^j \varepsilon_{t-j} \end{aligned}$$

MA to AR: Similarly, you can transform an MA(1) model to an AR(∞) model as follows:

B. Continuous Time

B.1. Brownian Motion

We define Brownian Motion, the continuous version of the random walk, as follows:

1. Normal Increments for all $\Delta \in \mathbb{R}$:

$$z_{t+\Delta} - z_t \sim \mathrm{N}(0,\ \Delta)$$

2. Independence of non-overlapping increments.

3. Non-overlapping increments of the same length are identically (and normally) distributed.

B.2. Fundamental Building Block, dz_t

Just as random noise, ε_t was the basic building block for discrete time, the analogous building block in continuous time is

$$dz_t = \lim_{\Delta \to 0} z_{t+\Delta} - z_t \tag{12}$$

Note, d is *forward* difference operator.

We also have the result that dz is of size (or order) $\sqrt{dt}$ (which is equivalent to the limit of Δ), since recall

$$\begin{aligned} \mathrm{Var}(z_{t+\Delta} - z_t) &= \Delta \\ \Leftrightarrow \quad \mathrm{Sd}(z_{t+\Delta} - z_t) &= \sqrt{\Delta} \end{aligned}$$

So now we can see why z_t is not differentiable. Namely, if we tried to consider dz_t/dt, we know that dz_t is of order $\sqrt{dt}$; therefore, the ratio dz_t/dt diverges to $\pm\infty$ since dt is close to zero, while the numerator, dz_t of order $\sqrt{dt}$, is larger. So we can talk about *the change*, dz_t, because that change is going to be very small over short time horizons. But we can't talk about the *rate of change*, because compared to "the change" in the process, the change in time is smaller still, such that the ratio diverges.

Moments: Next, we have

$$\begin{aligned} E_t dz_t &= 0 \\ \mathrm{Var}_t(dz_t) &= dt \\ \mathrm{Cov}(dz_t,\ dz_s) &= 0 \qquad s \neq t \end{aligned}$$

C. Price-Dividend and Return Linearizations

C.1. The Identity

We'll derive an approximation to an identity that says the following: there are three potential reasons the price-dividend ratio might be high:

1. Investors expect dividends to rise.
2. Investors expect low future returns, so future cashflows are discounted at a lower than usual rate. This leads to higher prices.
3. Investors expect prices to rise forever, giving an adequate return even if there are no dividends.

Now, we asserted above that Option 2 is the correct observation. But how can we test that? We'll, let's derive the identity that lays out the theory behind Options 1-3.

Start with an identity, and do some rearranging, letting R equal gross returns (price increases plus dividends), D be dividends, and P be price.

$$
\begin{aligned}
1 &= R_{t+1}^{-1} R_{t+1} = R_{t+1}^{-1} \frac{P_{t+1} + D_{t+1}}{P_t} \\
\Leftrightarrow \quad \frac{P_t}{D_t} &= R_{t+1}^{-1} \left(1 + \frac{P_{t+1}}{D_{t+1}}\right) \frac{D_{t+1}}{D_t} && (13) \\
\Leftrightarrow \quad P_t &= R_{t+1}^{-1} \left(D_{t+1} + P_{t+1}\right) && (14)
\end{aligned}
$$

Now, we can iterate forward Equation 14 and take a conditional expectation to get

$$
P_t = E_t \sum_{j=1}^{\infty} \left(\prod_{k=1}^{j} R_{t+k}^{-1} \right) D_{t+j} \tag{15}
$$

Now look at the previous line, and do the reverse of what we did in going from Equation 13 to Equation 14 (defining $\Delta D_t := D_t / D_{t-1}$:

$$
\begin{aligned}
\Rightarrow \quad \frac{P_t}{D_t} &= E_t \sum_{j=1}^{\infty} \left(\prod_{k=1}^{j} R_{t+k}^{-1} \right) \frac{D_{t+j}}{D_t} \\
\Leftrightarrow \quad \frac{P_t}{D_t} &= E_t \sum_{j=1}^{\infty} \left(\prod_{k=1}^{j} R_{t+k}^{-1} \right) \left(\prod_{k=1}^{j} \frac{D_{t+k}}{D_{t+k-1}} \right) \quad \text{Telescoping product trick!} \\
\Leftrightarrow \quad \frac{P_t}{D_t} &= E_t \sum_{j=1}^{\infty} \left(\prod_{k=1}^{j} R_{t+k}^{-1} \Delta D_{t+k} \right) && (16)
\end{aligned}
$$

So, what do we have? The identitiy in Equation 15 tells us that prices will increase if the discount rate *falls* or if Expected future dividends rise. *However*, prices aren't stationary, which is why we went to the trouble of deriving the identity in Equation 16, which *is* a stationary variable that captures the same logic as 15, and can be estimated properly via traditional time series approaches.

C.2. The Linearization

To make the identities derived above easier to handle, we take logs to linearize! Letting lowercase letters represent the logs of uppercase letters, we get from Equation 13

$$\begin{aligned} \ln \frac{P_t}{D_t} &= \ln \left\{ R_{t+1}^{-1} \left(1 + \frac{P_{t+1}}{D_{t+1}} \right) \frac{D_{t+1}}{D_t} \right\} \\ p_t - d_t &= -r_{t+1} + \ln(1 + e^{p_{t+1} - d_{t+1}}) + \Delta d_{t+1} \end{aligned} \tag{17}$$

Now do Taylor Expansion of the last term about a point $P/D = e^{p-d}$ (we'll use both the left- and righthand side of that point to simplify, so watch out for that). Also implicit in the deriviation, I'll be evaluating the derivative in the second term at $P/D = e^{d-d}$, so watch out for that too:

$$\begin{aligned} f(p_t - d_t) &= \ln(1 + e^{p_{t+1} - d_{t+1}}) \\ &\approx \ln(1 + e^{p-d}) + \frac{d\left[\ln(1 + e^{p_{t+1} - d_{t+1}})\right]}{d(p_t - d_t)} \cdot \{(p_{t+1} - d_{t+1}) - (p - d)\} \\ &= \ln(1 + e^{p-d}) + \frac{e^{p_{t+1} - d_{t+1}}}{1 + e^{p_{t+1} - d_{t+1}}} \cdot \{(p_{t+1} - d_{t+1}) - (p - d)\} \\ &= \ln(1 + P/D) + \frac{P/D}{1 + P/D} \cdot \{(p_{t+1} - d_{t+1}) - (p - d)\} \\ &= \ln(1 + P/D) + \rho \{(p_{t+1} - d_{t+1}) - (p - d)\} \end{aligned}$$

So now let's just plug this log-linear approximation into Equation 17:

$$p_t - d_t \approx -r_{t+1} + \Delta d_{t+1} + k + \rho \{(p_{t+1} - d_{t+1}) - (p - d)\} \tag{18}$$

where k equals $\ln(1 + P/D)$, which is a constant.[2] Now iterating forward is even more straightforward:

$$p_t - d_t = C + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}) \tag{19}$$

From this, it's clear that an ex-post high dividend-price ratio will result only if three was high dividend growth or low subsequent returns. To turn Equation 19 into an ex ante price-dividend ratio, take expectations of everything on the right.

[2]Note, ρ is just a constant of approximation that can be simplified using the fact that dividend yields roughly 4% on average, so the Price/Dividend ratio is about 25:

$$\rho = \frac{P/D}{1 + P/D} = \frac{1}{1 + D/P} \approx 1 - D/P = 0.96$$

C.3. Decomposing the Variance

To answer the question of what drives a high dividend-price ratio, we'll decompose the variance in Equation 19 as follows: