Ec2723, Asset Pricing I
Class Notes, Fall 2005

Present Value Relations and Stock Return Predictability

John Y. Campbell[1]

First draft: October 20, 2003
This version: October 18, 2005

[1]Department of Economics, Littauer Center, Harvard University, Cambridge MA 02138, USA. Email john_campbell@harvard.edu.

Market efficiency

Fama (1970) defines a market as *efficient* if "prices fully reflect all available information". In practice this means $R_{i,t+1} = \Theta_{it} + U_{i,t+1}$, where Θ_{it} is the rationally expected return on asset i, so $U_{i,t+1}$ is a fair game with respect to the information set at t.

This is tautologous unless we have some way to restrict Θ_{it} using some economic model. Thus market efficiency is not testable except in combination with a model of expected returns. This is known as the *joint hypothesis problem.*

It is important to remember Jensen's Inequality. A model for $\mathrm{E}_t R_{i,t+1}$ is not necessarily consistent with the same model for $\mathrm{E}_t r_{i,t+1}$, and certainly $\mathrm{E}_t r_{i,t+1} \neq \log \mathrm{E}_t R_{i,t+1}$ when returns are random.

Even when we have a model of expected returns, we need to specify the variables that are included in the information set at t. Fama defines 3 forms of the efficient market hypothesis:

- *Weak form.* Past returns.
- *Semi-strong form.* Past publicly available information, e.g. stock splits, dividends, earnings.
- *Strong form.* Past information, even if only available to insiders.

The market efficiency literature can be divided into:

- *Cross-sectional literature.* Average returns over t and consider various i. The economic model for Θ_{it} is a cross-sectional asset pricing model. Tests of the CAPM can be thought of as joint tests of the CAPM and market efficiency.
- *Time-series literature.* Fix i, model returns over t. The simplest economic model for Θ_{it} is then $\Theta_{it} = \Theta$, a constant, but equilibrium models with time-varying expected returns can also be considered. Much of this work concentrates on the behavior of an aggregate stock index.

To devise meaningful time-series tests and interpret the results, it is helpful to have an alternative hypothesis in mind. For example:

- Market prices are contaminated by short-term noise caused by measurement errors or illiquidity (bid-ask bounce). This generates short-run reversals.
- The market reacts sluggishly to information. This generates short-run predictability of returns based on past returns or information releases.
- Market prices can deviate substantially from efficient levels, and the deviations are hard to arbitrage because they last a long time. This generates long-run reversal and predictability based on price levels, but may be hard to detect in the short run.

Short-term return predictability is easy to detect if it is present, and is hard to explain using a risk-based asset pricing model. However it has modest effects on prices, and it can disappear quickly if arbitrageurs discover the predictability or if transactions costs decline making arbitrage cheaper. Long-term return predictability can have large effects on prices, but is hard to detect without very long time series and may be explained by a more sophisticated model of risk and return.

Michael Jensen (1978): "There is no other proposition in economics which has more solid evidence supporting it than the Efficient Markets Hypothesis".

Robert Shiller (1984): "Returns on speculative assets are nearly unforecastable; this fact is the basis of the most important argument in the oral tradition against a role for mass psychology in speculative markets. One form of this argument claims that because real returns are nearly unforecastable, the real price of stocks is close to the intrinsic value, that is, the present value with constant discount rate of optimally forecasted future real dividends. This argument... is one of the most remarkable errors in the history of economic thought".

The debate continues to this day, perhaps because the EMH is a half-full (or half-empty) glass.

Tests of autocorrelation in stock returns

A direct method for testing predictability is to test whether past returns predict future returns (*weak-form market efficiency* in Fama's terminology).

The leading approach looks at the autocorrelations of stock returns. Under the null hypothesis that stock returns are iid, the standard error for any single sample autocorrelation is asymptotically given by $1/\sqrt{T}$, where T is the sample size. Unfortunately this makes it extremely hard to detect small autocorrelations since the standard error is 0.1 when $T = 100$ and is still 0.02 when $T = 2500$.

Under the null hypothesis that stock returns are iid, different autocorrelations are uncorrelated with one another. This suggests that we may gain by combining different autocorrelations. For example, the Q statistic of Box and Pierce calculates a sum of K squared autocorrelations:

$$Q_K = T \sum_{k=1}^{K} \widehat{\rho}_k^2,$$

and this is asymptotically distributed χ^2 with K degrees of freedom.

The Q statistic does not use the sign of the autocorrelations. Some plausible alternatives generate a large number of autocorrelations, each of which is small but which all have the same sign. To get power against such an alternative, we would like to average autocorrelations rather than squared autocorrelations. One way to do this is the variance ratio statistic,

$$\widehat{V}(K) = \frac{\widehat{\mathrm{Var}}(r_{t+1} + ... + r_{t+K})}{K\widehat{\mathrm{Var}}(r_{t+1})} = 1 + 2\sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)\widehat{\rho}_k.$$

Asymptotically, the variance of this statistic under the iid null is

$$\mathrm{Var}(\widehat{V}(K)) = \frac{4}{T}\sum_{j=1}^{K-1}\left(1 - \frac{j}{K}\right)^2 = \frac{2(2K-1)(K-1)}{3KT}.$$

As K increases, this approaches $4K/3T$.

This result can be generalized when $K \longrightarrow \infty$, $T \longrightarrow \infty$, and $K/T \longrightarrow 0$. (Priestley, *Spectral Analysis and Time Series*, 1982, p.463.) In this case the true process can be serially correlated, heteroskedastic, and nonnormal, and we still have

$$\mathrm{Var}(\widehat{V}(K)) = \frac{4KV(K)^2}{3T}.$$

Note that this is larger when the true $V(K)$ is large.

A related approach is to regress the K-period return on the lagged K-period return:

$$\beta(K) = \frac{V(2K)}{V(K)} - 1.$$

The R^2 statistic from this K-period regression is just the square of the regression coefficient:

$$R^2(K) = \beta(K)^2.$$

Results from autocorrelation tests:

- Individual stocks have negative daily autocorrelations, but broader indexes have positive autocorrelations. These effects are smaller in recent data. (Froot and Perold 1995.)

- At a weekly frequency, positive index autocorrelations are driven by small positive own autocorrelations and important cross autocorrelations from large to small stocks. (Lo and MacKinlay 1988, 1990.)

- Index autocorrelations tend to be positive for the first 6-12 months, then negative. (Poterba and Summers 1988.)

- There is a corresponding tendency for $\beta(K)$ to be negative, with maximum $R^2(K)$ of 25-45% at 3-5 years. (Fama and French 1988.)

- These results depend on the inclusion of the Great Depression and are statistically weak. The runup in stock prices in the late 1990's further weakened these findings.

- The use of asymptotics assuming $K/T \longrightarrow 0$ is dangerous when in practice K is often large relative to the sample size. Stock and Richardson (1989) develop alternative asymptotics assuming $K/T \longrightarrow \delta$, where $\delta > 0$.

To better interpret these results, we would like to have a better sense of what alternative models would imply about the pattern of autocorrelations. This requires that we develop a present value model that determines prices, and thus realized returns, given alternative models for expected returns.

Prices, dividends, and returns with constant discount rates

The net simple return on a stock is given by

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}.$$

If the expected return on the stock is constant, $\mathrm{E}_t R_{t+1} = R$, then

$$P_t = \mathrm{E}_t \left[\frac{P_{t+1} + D_{t+1}}{1+R} \right].$$

Solving forward for K periods, we get

$$P_t = \mathrm{E}_t \left[\sum_{k=1}^{K} \left(\frac{1}{1+R} \right)^k D_{t+k} \right] + \mathrm{E}_t \left[\left(\frac{1}{1+R} \right)^K P_{t+K} \right].$$

Letting $K \longrightarrow \infty$, and assuming that the last term on the right hand side converges to zero, we have the *dividend discount model* (DDM) of stock prices,

$$P_t = \mathrm{E}_t \left[\sum_{k=1}^{\infty} \left(\frac{1}{1+R} \right)^k D_{t+k} \right].$$

This model, with a constant expected stock return, is sometimes called the *random walk* or *martingale* model of stock prices. But in fact the stock price is not a martingale in this model, since

$$\mathrm{E}_t P_{t+1} = (1+R) P_t - \mathrm{E}_t D_{t+1}.$$

What is a martingale? If we reinvest all dividends in buying more shares, the number of shares we own follows

$$N_{t+1} = N_t \left(1 + \frac{D_{t+1}}{P_{t+1}} \right).$$

The discounted value of the portfolio,

$$M_t = \frac{N_t P_t}{(1+R)^t},$$

follows a martingale.

The DDM can be used to illustrate the concept of cointegration. If the dividend follows a process with a unit root, so that ΔD_t is stationary, then we have

$$P_t - \frac{D_t}{R} = \left(\frac{1}{R}\right) \mathrm{E}_t \left[\sum_{i=0}^{\infty} \left(\frac{1}{1+R}\right)^i \Delta D_{t+1+i} \right],$$

so stock prices and dividends are cointegrated.

A particularly useful special case assumes that dividends grow at a constant rate G, so that $E_t D_{t+k} = (1+G)^{k-1} E_t D_{t+1}$. Then we get the *Gordon growth model* (named after Myron Gordon, but actually due to John Burr Williams),

$$P_t = \frac{\mathrm{E}_t D_{t+1}}{R-G},$$

often written as

$$\frac{D}{P} = R - G,$$

where D denotes the next-period dividend.

Rational bubbles

So far we have assumed that

$$\lim_{K \longrightarrow \infty} \mathrm{E}_t \left[\left(\frac{1}{1+R}\right)^K P_{t+K} \right] = 0.$$

Models of rational bubbles violate this assumption. We then get an infinity of possible solutions

$$P_t = P_{Dt} + B_t,$$

where P_{Dt} is the price in the DDM, and B_t is any stochastic process satisfying

$$B_t = \mathrm{E}_t \left[\frac{B_{t+1}}{1+R} \right].$$

For example, Blanchard and Watson (1982) suggested the following bubble process: $B_{t+1} = ((1+R)/\pi)B_t + \zeta_{t+1}$ with probability π, and $B_{t+1} = \zeta_{t+1}$ with probability $1-\pi$, where $E_t\zeta_{t+1} = 0$.

The conditions for rational bubbles to exist are extremely restrictive.

- Bubbles cannot exist on finite-lived assets.
- Negative bubbles cannot exist if there is a lower bound on the asset price (e.g. zero, for assets with limited liability).
- Positive bubbles cannot exist if there is an upper bound on the asset price (e.g. a high-priced substitute in perfectly elastic supply).
- If positive bubbles can exist, but negative bubbles are ruled out, then a bubble can never start. It must exist from the beginning of trading. (Diba and Grossman 1988.)
- Bubbles cannot exist in a representative agent economy with an infinite-lived agent.
- Tirole (1985) showed that bubbles cannot exist in OLG economies that are dynamically efficient (i.e. that have an interest rate greater than the growth rate of the economy).

Nonetheless the rational bubble literature is informative because it tells us what phenomena we may observe in a world of almost rational bubbles (a small amount of persistent return predictability generating large effects on prices).

A loglinear present value model with time-varying discount rates

If the expected stock return is time-varying, then the exact present value model becomes nonlinear. Campbell and Shiller (1988) suggested an approximate loglinear present value model for use in this case. Start from the definition of the log stock return,

$$r_{t+1} = \log(1+R_{t+1}) = \log(P_{t+1}+D_{t+1}) - \log(P_t) = p_{t+1} - p_t + \log(1+\exp(d_{t+1}-p_{t+1})),$$

where p_{t+1} and d_{t+1} denote log prices and dividends. Approximate the nonlinear function

$$\log(1+\exp(d_{t+1}-p_{t+1})) = f(d_{t+1}-p_{t+1}) \approx f(\overline{d-p})+f'(\overline{d-p})(d_{t+1}-p_{t+1}-(\overline{d-p})).$$

Here $f(x) = \log(1+\exp(x))$ and $f'(x) = \exp(x)/(1+\exp(x))$. The resulting approximation for the log return is

$$r_{t+1} \approx k + \rho p_{t+1} + (1-\rho)d_{t+1},$$

where

$$\rho = \frac{1}{1+\exp(\overline{d-p})},$$

and

$$k = -\log(\rho) - (1-\rho)\log(1/\rho - 1).$$

This approximation replaces the log of a sum with an average of logs, where the relative weights depend on the average relative magnitudes of dividend and price.

The approximate expression for the log stock return is a difference equation in log price, dividend, and return. Solving forward and imposing the terminal condition that

$$\lim_{j\longrightarrow\infty} \rho^j p_{t+j} = 0,$$

we get

$$p_t = \frac{k}{1-\rho} + \sum_{j=0}^{\infty} \rho^j[(1-\rho)d_{t+1+j} - r_{t+1+j}].$$

This equation holds ex post, as an accounting identity. It should therefore hold ex ante, not only for rational expectations but for any expectations that satisfy identities. We have

$$p_t = \frac{k}{1-\rho} + \mathrm{E}_t \sum_{j=0}^{\infty} \rho^j[(1-\rho)d_{t+1+j} - r_{t+1+j}] = \frac{k}{1-\rho} + p_{CF,t} - p_{DR,t},$$

where $p_{CF,t}$ and $p_{DR,t}$ are the components of the price driven by cash flow (dividend) expectations and discount rate (return) expectations respectively.

If log dividends follow a unit root process, then log dividends and log prices are cointegrated with a known cointegrating vector. The log dividend-price ratio is stationary:

$$d_t - p_t = \frac{-k}{1-\rho} + \mathrm{E}_t \sum_{j=0}^{\infty} \rho^j [-\Delta d_{t+1+j} + r_{t+1+j}].$$

Campbell (1991) decomposes the variation in stock returns into revisions in expectations of dividend growth and future returns:

$$r_{t+1} - \mathrm{E}_t r_{t+1} = (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} = N_{CF,t+1} - N_{DR,t+1}.$$

Here N denotes news, CF denotes cash flows (dividends), and DR denotes discount rates (expected future returns). Note that the time index in the cash flow sum starts at zero, whereas the time index in the discount rate sum starts at one.

An interesting implication of this formula is that better information about future dividends reduces the volatility of returns. The reason is that news about dividends must enter prices at some point; the earlier it does, the more heavily the effect is discounted. West (1988) emphasizes this point, which is often misunderstood.

Present value models with earnings

So far we have treated dividends and their growth rate as exogenous. Alternatively, we may ask how dividends are determined by a firm's payout policy and profitability. Write earnings as X_t and the book equity of the firm as B_t. Then we have

$$B_t = B_{t-1} + X_t - D_t.$$

(This is exactly true under *clean-surplus accounting*, and approximately true under real-world accounting.)

Define *return on equity* (ROE) as earnings divided by lagged book equity, $ROE_t = X_t / B_{t-1}$.

Define the *retention* or *plowback ratio* λ_t as the fraction of earnings that is retained for reinvestment. Then

$$D_t = (1 - \lambda_t) X_t.$$

In the steady state of the Gordon growth model, book equity, earnings, and dividends all grow at the common rate G. Thus we have

$$G = \frac{B_t - B_{t-1}}{B_{t-1}} = \frac{X_t - D_t}{B_{t-1}} = \lambda \frac{X_t}{B_{t-1}} = \lambda ROE.$$

Substituting these expressions into the Gordon growth model, we have

$$P_t = \frac{(1-\lambda)\mathrm{E}_t X_{t+1}}{R - \lambda ROE}$$

or

$$\frac{X}{P} = \frac{R - \lambda ROE}{1-\lambda},$$

where X denotes next-period earnings. This shows that stock prices increase with the retention ratio when $ROE > R$, and decline with the retention ratio when $ROE < R$.

In the long run, we might expect investments to continue until ROE is driven to equality with R. In this case the earnings-price ratio equals the discount rate, regardless of payout policy:

$$\frac{X}{P} = R.$$

Vuolteenaho (2002) works out a loglinear approximation for the dynamic version of this model:

$$b_t - v_t = \mu + E_t \sum_{j=0}^{\infty} \rho^j [-roe_{t+1+j} + r_{t+1+j}],$$

where v_t is the log market value of the firm and $roe_t = \log(1 + ROE_t)$.

Volatility and valuation

Pastor and Veronesi (2003, 2005) point out that uncertainty about growth rates increases firm value. One way to understand this is from the Gordon growth model:

$$\frac{P}{D} = \mathrm{E}\left[\frac{1}{R-G}\right] > \frac{1}{R - \mathrm{E}[G]}$$

by Jensen's Inequality. Another way to understand it is that the returns that appear in the Campbell-Shiller dividend growth formula or the Vuolteenaho ROE formula are log returns. Uncertainty lowers the geometric average return corresponding to any arithmetic average return. Thus if uncertainty does not increase the arithmetic average return too much (for example, if it is idiosyncratic), then it lowers the geometric average return and increases the price. This effect is helpful in understanding high prices for volatile technology stocks in the late 1990's.

An illustrative model with return predictability

Much recent work considers a specific model of time-varying expected returns, in which the expected return is an AR(1) process:

$$E_t r_{t+1} = r + x_t,$$

$$x_{t+1} = \phi x_t + \xi_{t+1.}$$

The realized return equals the expected return plus noise:

$$r_{t+1} = r + x_t + u_{t+1}.$$

The process for x_t implies that $p_{DR,t}$, the component of the price that is driven by changes in future expected returns, is

$$p_{DR,t} = E_t \sum_{j=0}^{\infty} \rho^j r_{t+1+j} = \frac{r}{1-\rho} + \frac{x_t}{1-\rho\phi}.$$

The variance of this component is

$$\mathrm{Var}(p_{DR,t}) = \frac{\sigma_x^2}{(1-\rho\phi)^2},$$

so the expected return may have a very small volatility yet may still have a very large effect on the stock price if it is highly persistent.

We get a similar insight if we calculate the news about future returns,

$$N_{DR,t+1} = (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} = \frac{\rho\xi_{t+1}}{1-\rho\phi} \approx \frac{\xi_{t+1}}{1-\phi}.$$

A 1% increase in the expected return today is associated with a capital loss of about 2% if the AR coefficient is 0.5, a loss of about 4% if the AR coefficient is 0.75, and a loss of about 10% if the AR coefficient is 0.9.

This example can be used to calculate the implied autocovariances of returns:

$$\gamma_i = \text{Cov}(r_{t+1}, r_{t+1+i}) = \phi^{i-1}\left[C + \left(\frac{\phi}{1-\phi^2} - \frac{\rho}{1-\rho\phi}\right)\sigma_\xi^2\right],$$

where

$$C = \text{Cov}(\xi_{t+1}, N_{CF,t+1}).$$

- Autocovariances are all of the same sign and die off at rate ϕ.
- Sign of autocovariances depends on three terms:
  - Positive effect of covariance between dividend news and revisions in expected returns
  - Direct positive effect of autocorrelated expected returns
  - Negative effect from capital loss that occurs when expected returns increase.
- Autocovariances can all be zero, even if expected returns vary through time. This shows that prices can be weak-form efficient even if they are not semi-strong form efficient.
- However for reasonable parameter values (C not strongly positive, ϕ not too large), autocorrelations will tend to be negative.

Predictive return regressions

The example can also be used to see what happens if we regress the stock return onto the predictor variable x_t. For simplicity, assume $C = 0$. Then the variance of the stock return is

$$\text{Var}(r_{t+1}) = \sigma_{CF}^2 + \sigma_x^2\frac{1+\rho^2-2\rho\phi}{(1-\rho\phi)^2} \approx \sigma_{CF}^2 + \frac{2\sigma_x^2}{1-\phi},$$

where $\sigma^2_{CF} = \text{Var}(N_{CF})$.

The R^2 of a single-period return regression onto x_t is

$$R^2(1) = \frac{\text{Var}(E_t r_{t+1})}{\text{Var}(r_{t+1})} \approx \frac{\sigma_x^2}{\sigma^2_{CF} + 2\sigma_x^2/(1-\phi)} = \left(\frac{\sigma^2_{CF}}{\sigma_x^2} + \frac{2}{1-\phi}\right)^{-1} \leq \frac{1-\phi}{2}.$$

When x_t is extremely persistent, the one-period return regression must have a low R^2, even if there is no cash flow news at all!

Now consider a long-horizon regression of the K-period return onto the predictor variable x_t:

$$r_{t+1} + ... + r_{t+K} = \beta(K)x_t,$$

where

$$\beta(K) = 1 + \phi + ... + \phi^{K-1} = \frac{1-\phi^K}{1-\phi}.$$

The ratio of the K-period R^2 to the 1-period R^2 is

$$\begin{aligned} \frac{R^2(K)}{R^2(1)} &= \left[\frac{\text{Var}(E_t r_{t+1} + ... + E_t r_{t+K})}{\text{Var}(r_{t+1} + ... + r_{t+K})}\right] \Big/ \left[\frac{\text{Var}(E_t r_{t+1})}{\text{Var}(r_{t+1})}\right] \\ &= \frac{\beta(K)^2}{\beta(1)^2}\frac{1}{KV(K)} = \left(\frac{1-\phi^K}{1-\phi}\right)^2 \frac{1}{KV(K)}. \end{aligned}$$

This grows at first with K if ϕ is large, then eventually dies away to zero.

There has been great interest in long-horizon return regressions because they have much higher R^2 statistics than short-horizon regressions, and the usual asymptotic t-statistics (calculated allowing for overlap in the residuals following Hansen-Hodrick or Newey-West) deliver stronger rejections. However these t-statistics tend to have size distortions when the overlap is large relative to the sample size, so the long-horizon regression evidence is tenuous statistically.

Even the single-period regression test can be problematic when the predictor variable is highly persistent. Stambaugh (1999) shows that the finite-sample bias of the one-period regression coefficient is

$$\text{E}[\widehat{\beta}(1) - \beta(1)] = -\left(\frac{1+3\phi}{T}\right)\frac{\sigma_{u\xi}}{\sigma_\xi^2} = \frac{\rho(1+3\phi)}{(1-\rho\phi)T},$$

where the second equality holds for the case where $C = 0$. There are similar problems with the distribution of the t-statistic when ϕ is close to one. Lewellen (2003), Campbell and Yogo (2005), and Polk, Thompson, and Vuolteenaho (2005) suggest alternative methods to handle this problem.

VAR methodology

If one is willing to assume that a vector autoregression (VAR) describes the data, then long-horizon behavior can be imputed from short-horizon behavior. Define x_t as a vector of state variables, and assume that

$$x_{t+1} = Ax_t + \epsilon_{t+1}.$$

The assumption that x_t follows a first-order VAR is not restrictive because one can always rewrite a higher-order VAR as a first-order VAR with an expanded state vector and a singular variance-covariance matrix of innovations. This model implies that

$$\mathrm{E}_t x_{t+1+j} = A^{j+1} x_t.$$

Suppose that the stock return is the first element of the VAR, and the other variables help to predict returns. Then

$$r_{t+1} - \mathrm{E}_t r_{t+1} = e1' \epsilon_{t+1},$$

where $e1' = [10...0]$, a vector with first element one and all other elements zero. We have

$$N_{DR,t+1} = e1' \sum_{j=1}^{\infty} \rho^j A^j \epsilon_{t+1} = e1' \rho A (I - \rho A)^{-1} \epsilon_{t+1},$$

and

$$N_{CF,t+1} = r_{t+1} - E_t r_{t+1} + N_{DR,t+1} = e1'(I + \rho A(I - \rho A)^{-1})\epsilon_{t+1}.$$