20

Expected Returns in the Time Series and Cross Section

THE FIRST REVOLUTION in finance started the modern field. Peaking in the early 1970s, this revolution established the CAPM, random walk, efficient markets, portfolio-based view of the world. The pillars of this view are:

1. The CAPM is a good measure of risk and thus a good explanation why some stocks, portfolios, strategies, or funds (assets, generically) earn higher average returns than others.
2. Returns are unpredictable. In particular,
   (a) Stock returns are close to unpredictable. Prices are close to random walks; expected returns do not vary greatly through time. "Technical analysis" that tries to divine future returns from past price and volume data is nearly useless. Any apparent predictability is either a statistical artifact which will quickly vanish out of sample, or cannot be exploited after transactions costs. The near unpredictability of stock returns is simply stated, but its implications are many and subtle. (Malkiel [1990] is a classic and easily readable introduction.) It also remains widely ignored, and therefore is the source of lots of wasted trading activity.
   (b) Bond returns are nearly unpredictable. This is the expectations model of the term structure. If long-term bond yields are higher than short-term yields—if the yield curve is upward sloping—this does not mean that expected long-term bond returns are any higher than those on short-term bonds. Rather, it means that short-term interest rates are expected to rise in the future, so you expect to earn about the same amount on short-term or long-term bonds at any horizon.
   (c) Foreign exchange bets are not predictable. If a country has higher interest rates than are available in the United States for bonds of a

similar risk class, its exchange rate is expected to depreciate. After you convert your investment back to dollars, you expect to make the same amount of money holding foreign or domestic bonds.

   (d) Stock market volatility does not change much through time. Not only are returns close to unpredictable, they are nearly identically distributed as well.

3. Professional managers do not reliably outperform simple indices and passive portfolios once one corrects for risk (beta). While some do better than the market in any given year, some do worse, and the outcomes look very much like good and bad luck. Managers who do well in one year are not more likely to do better than average the next year. The average actively managed fund does about 1% *worse* than the market index. The more actively a fund trades, the lower returns to investors.

Together, these views reflected a guiding principle that asset markets are, to a good approximation, *informationally efficient* (Fama 1970, 1991). This statement means that market prices already contain most information about fundamental value. Informational efficiency in turn derives from *competition.* The business of discovering information about the value of traded assets is extremely competitive, so there are no easy quick profits to be made, as there are not in every other well-established and competitive industry. The only way to earn large returns is by taking on additional risk.

These statements are not doctrinaire beliefs. Rather, they summarize the findings of a quarter-century of extensive and careful empirical work. However, every single one of them has now been extensively revised by a new generation of empirical research. Now, it seems that:

1. There are assets, portfolios, funds, and strategies whose average returns cannot be explained by their market betas. Multifactor models dominate the empirical description, performance attribution, and explanation of average returns.
2. Returns are predictable. In particular,
   (a) Variables including the dividend/price ratio and term premium can in fact predict substantial amounts of stock return variation. This phenomenon occurs over business cycle and longer horizons. Daily, weekly, and monthly stock returns are still close to unpredictable, and "technical" systems for predicting such movements are still close to useless after transactions costs.
   (b) Bond returns are predictable. Though the expectations model works well in the long run, a steeply upward sloping yield curve means that expected returns on long-term bonds are higher than on short-term bonds for the next year.
   (c) Foreign exchange returns are predictable. If you buy bonds in a country whose interest rates are unusually higher than those in the

United States, you expect a greater return, even after converting back to dollars.

(d) Stock market volatility does in fact change through time. Conditional second moments vary through time as well as first moments. Means and variances do not seem to move in lockstep, so conditional Sharpe ratios vary through time.

3. Some funds seem to outperform simple indices, even after controlling for risk through market betas. Fund returns are also slightly predictable: past winning funds seem to do better in the future, and past losing funds seem to do worse than average in the future. For a while, this seemed to indicate that there is some persistent skill in active management. However, we now see that multifactor performance attribution models explain most fund persistence: funds earn persistent returns by following fairly mechanical "styles," not by persistent skill at stock selection (Carhart [1997]).

Again, these views summarize a large body of empirical work. The strength and interpretation of many results are hotly debated.

This new view of the facts need not overturn the view that markets are reasonably competitive and therefore reasonably efficient. It does substantially enlarge our view of what activities provide rewards for holding risks, and it challenges our economic understanding of those risk premia. As of the early 1970s, asset pricing theory anticipated the possibility and even probability that expected returns should vary over time and that covariances past market betas would be important for understanding cross-sectional variation in expected returns. What took another 15 to 20 years was to see how important these long-anticipated theoretical possibilities are in the data.

20.1 Time-Series Predictability

I start by looking at patterns in expected returns over time in large market indices, and then look at patterns in expected returns across stocks.

Long-Horizon Stock Return Regressions

> Dividend/price ratios forecast excess returns on stocks. Regression coefficients and R^2 rise with the forecast horizon. This is a result of the fact that the forecasting variable is persistent.

Table 20.1. *OLS regressions of percent excess returns (value weighted NYSE − treasury bill rate) and real dividend growth on the percent VW dividend/price ratio*

Horizon k	$R_{t\to t+k} = a + b(D_t/P_t)$			$D_{t+k}/D_t = a + b(D_t/P_t)$		
(years)	b	$\sigma(b)$	R^2	b	$\sigma(b)$	R^2
1	5.3	(2.0)	0.15	2.0	(1.1)	0.06
2	10	(3.1)	0.23	2.5	(2.1)	0.06
3	15	(4.0)	0.37	2.4	(2.1)	0.06
5	33	(5.8)	0.60	4.7	(2.4)	0.12

$R_{t\to t+k}$ indicates the k-year return. Standard errors in parentheses use GMM to correct for heteroskedasticity and serial correlation. Sample 1947–1996.

The left-hand regression in Table 20.1 gives a simple example of market return predictability, updating Fama and French (1988b). "Low" prices relative to dividends forecast higher subsequent returns. The one-year horizon 0.15 R^2 is not particularly remarkable. However, at longer and longer horizons larger and larger fractions of return variation are forecastable. At a five-year horizon 60% of the variation in stock returns is forecastable ahead of time from the price/dividend ratio.

One can object to dividends as the divisor for prices. However, ratios formed with just about any sensible divisor work about as well, including earnings, book value, and moving averages of past prices.

Many other variables forecast excess returns, including the term spread between long- and short-term bonds, the default spread, the T-bill rate (Fama and French [1989]), and the earnings/dividend ratio (Lamont [1998]). Macro variables forecast stock returns as well, including the investment/capital ratio (Cochrane [1991d]) and the consumption/wealth ratio (Lettau and Ludvigson [2001b]).

Most of these variables are correlated with each other and correlated with or forecast business cycles. This fact suggests a natural explanation, emphasized by Fama and French (1989): Expected returns vary over business cycles; it takes a higher risk premium to get people to hold stocks at the bottom of a recession. When expected returns go up, prices go down. We see the low prices, followed by the higher returns expected and required by the market. (Regressions do not have to have causes on the right and effects on the left. You run regressions with the variable orthogonal to the error on the right, and that is the case here since the error is a forecasting error. This is like a regression of actual weather on a weather forecast.)

Table 20.2, adapted from Lettau and Ludvigson (2001b), compares several of these variables. At a one-year horizon, both the consumption/wealth

Table 20.2. *Long-horizon return forecasts*

Horizon (years)	*cay*	$d-p$	$d-e$	*rrel*	R^2
1	6.7				0.18
1		0.14	0.08		0.04
1				−4.5	0.10
1	5.4	0.07	−0.05	−3.8	0.23
6	12.4				0.16
6		0.95	0.68		0.39
6				−5.10	0.03
6	5.9	0.89	0.65	1.36	0.42

The return variable is log excess returns on the S&P composite index. *cay* is Lettau and Ludvigson's consumption to wealth ratio. $d-p$ is the log dividend yield and $d-e$ is the log earnings yield. *rrel* is a detrended short-term interest rate. Sample 1952:4–1998:3.
Source: Lettau and Ludvigson (2001b, Table 6).

ratio and the detrended T-bill rate forecast returns, with R^2 of 0.18 and 0.10, respectively. At the one-year horizon, these variables are more important than the dividend/price and dividend/earnings ratios, and their presence cuts the dividend ratio coefficients in half. However, the d/p and d/e ratios are slower moving than the T-bill rate and consumption/wealth ratio. They track decade-to-decade movements as well as business cycle movements. This means that their importance builds with horizon. By six years, the bulk of the return forecastability again comes from the dividend ratios, and it is their turn to cut down the cay and T-bill regression coefficients. The cay and d/e variables have not been that affected by the late 1990s, while this time period has substantially cut down our estimate of dividend yield forecastability.

I emphasize that *excess* returns are forecastable. We have to understand this as time-variation in the reward for risk, not time-varying interest rates. One naturally slips in to nonrisk explanations for price variation; for example that the current stock market boom is due to life-cycle savings of the baby boomers. A factor like this does not reference *risks*; it predicts that interest rates should move just as much as stock returns.

Persistent d/p; Long Horizons Are Not A Separate Phenomenon

The results at different horizons are not separate facts, but reflections of a single underlying phenomenon. If daily returns are very slightly predictable by a slow-moving variable, that predictability adds up over long horizons. For example, you can predict that the temperature in Chicago will rise about 1/3 degree per day in the springtime. This forecast explains very little of the day-to-day variation in temperature, but tracks almost all

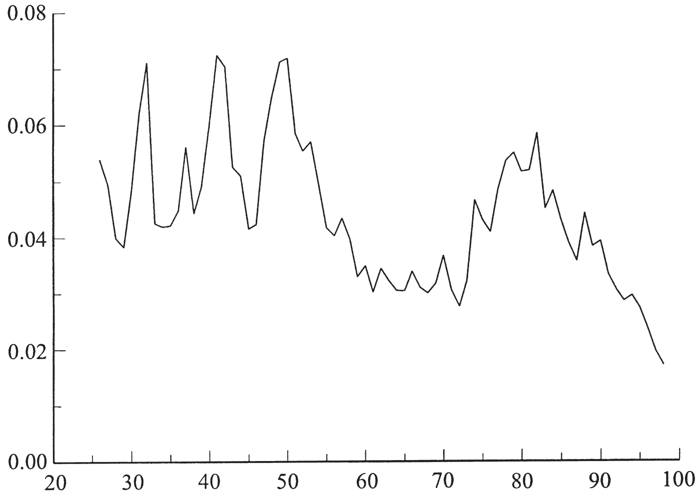

Figure 20.1. *Dividend/price ratio of value-weighted NYSE.*

of the rise in temperature from January to July. Thus, the R^2 rises with horizon.

Thus, a central fact driving the predictability of returns is that the dividend/price ratio is very persistent. Figure 20.1 plots the d/p ratio and you can see directly that it is extremely slow-moving. Below, I estimate an AR(1) coefficient around 0.9 in annual data.

To see more precisely how the results at various horizons are linked, and how they result from the persistence of the d/p ratio, suppose that we forecast returns with a forecasting variable x, according to

$$r_{t+1} = bx_t + \varepsilon_{t+1}, \tag{20.1}$$

$$x_{t+1} = \rho x_t + \delta_{t+1}. \tag{20.2}$$

(Obviously, you de-mean the variables or put constants in the regressions.) Small values of b and R^2 in (20.1) and a large coefficient ρ in (20.2) imply mathematically that the long-horizon regression has a large regression coefficient and large R^2. To see this, write

$$r_{t+1} + r_{t+2} = b(1+\rho)x_t + b\delta_{t+1} + \varepsilon_{t+1} + \varepsilon_{t+2},$$

$$r_{t+1} + r_{t+2} + r_{t+3} = b(1+\rho+\rho^2)x_t + b\rho\delta_{t+1} + b\delta_{t+2} + \varepsilon_{t+1} + \varepsilon_{t+2} + \varepsilon_{t+3}.$$

You can see that with ρ near one, the coefficients increase with horizon, almost linearly at first and then at a declining rate. The R^2 are a little messier to work out, but also rise with horizon.

The numerator in the long-horizon regression coefficient is

$$E[(r_{t+1} + r_{t+2} + \cdots + r_{t+k})x_t], \tag{20.3}$$

where the symbols represent deviations from their means. With stationary r and x, $E(r_{t+j}x_t) = E(r_{t+1}x_{t-j})$, so this is the same moment as

$$E[r_{t+1}(x_t + x_{t-1} + x_{t-2} + \cdots)], \tag{20.4}$$

the numerator of a regression coefficient of one-year returns on many lags of price/dividend ratios. Of course, if you run a multiple regression of returns on lags of p/d, you quickly find that most lags past the first do not help the forecast power. (That statement would be exact in the AR(1) example.)

This observation shows once again that one-year and multiyear forecastability are two sides of the same coin. It also suggests that on a purely *statistical* basis, there will not be a huge difference between one-year return forecasts and multiyear return forecasts (correcting the latter for the serial correlation of the error term due to overlap). Hodrick (1992) comes to this conclusion in a careful Monte Carlo experiment, comparing moments of the form (20.3), (20.4), and $E(r_{t+1}x_t)$. Also, Jegadeesh (1991) used the equivalence between (20.3) and (20.4) to test for long-horizon predictability using one-month returns and a moving average of instruments. The direct or implied multiyear regressions are thus mostly useful for illustrating the dramatic *economic* implications of forecastability, rather than as clever statistical tools that enhance power and allow us to distinguish previously foggy hypotheses.

The slow movement of the price/dividend ratio means that on a purely statistical basis, return forecastability is an open question. What we really know (see Figure 20.1) is that low prices relative to dividends and earnings in the 1950s preceded the boom market of the early 1960s; that the high price/dividend ratios of the mid-1960s preceded the poor returns of the 1970s; that the low price ratios of the mid-1970s preceded the current boom. We really have three postwar data points: a once-per-generation change in expected returns. In addition, the last half of the 1990s has seen a historically unprecedented rise in stock prices and price/dividend ratios (or any other ratio). This rise has cut the postwar return forecasting regression coefficient in half. On the other hand, another crash or even just a decade of poor returns will restore the regression. Data back to the 1600s show the same pattern, but we are often uncomfortable making inferences from centuries-old data.

Volatility

Price/dividend ratios can only move at all if they forecast future returns, if they forecast future dividend growth, or if there is a bubble—if the price/dividend ratio is nonstationary and is expected to grow explosively. In the data, most variation in price/dividend ratios results from varying expected returns. "Excess volatility"—relative to constant discount rate present-value models—is thus exactly the same phenomenon as forecastable long-horizon returns.

I also derive the very useful price/dividend and return linearizations. Ignoring constants (means),

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

$$r_t - E_{t-1} r_t = (E_t - E_{t-1}) \left[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+j} - \sum_{j=1}^{\infty} \rho^j r_{t+j} \right],$$

$$r_{t+1} = \Delta d_{t+1} - \rho(d_{t+1} - p_{t+1}) + (d_t - p_t).$$

The volatility test literature starting with Shiller (1981) and LeRoy and Porter (1981) (see Cochrane [1991c] for a review) started out trying to make a completely different point. *Predictability* seems like a sideshow. The stunning fact about the stock market is its extraordinary *volatility*. On a typical day, the value of the U.S. capital stock changes by a full percentage point, and days of 2 or 3 percentage point changes are not uncommon. In a typical year it changes by 16 percentage points, and 30 percentage point changes are not uncommon. Worse, most of that volatility seems not to be accompanied by any important news about future returns and discount rates. Thirty percent of the capital stock of the United States vanished in a year and nobody noticed? Surely, this observation shows directly that markets are "not efficient"—that prices do not correspond to the value of capital—without worrying about predictability?

It turns out, however, that "excess volatility" is *exactly* the same thing as return predictability. Any story you tell about prices that are "too high" or "too low" necessarily implies that subsequent returns will be too low or too high as prices rebound to their correct levels.

When prices are high relative to dividends (or earnings, cashflow, book value, or some other divisor), one of three things must be true: (1) Investors expect dividends to rise in the future. (2) Investors expect returns to be low in the future. Future cashflows are discounted at a lower than usual rate, leading to higher prices. (3) Investors expect prices to rise forever, giving an adequate return even if there is no growth in dividends. This statement is not a theory, it is an identity: If the price/dividend ratio is high, either dividends must rise, prices must decline, or the price/dividend ratio must grow explosively. The open question is, which option holds for our stock market? Are prices high now because investors expect future earnings, dividends, etc. to rise, because they expect low returns in the future, or because they expect prices to go on rising forever?

Historically, we find that *virtually all variation in price/dividend ratios has reflected varying expected excess returns.*

Exact Present-Value Identity

To document this statement, we need to relate current prices to future dividends and returns. Start with the identity

$$1 = R_{t+1}^{-1} R_{t+1} = R_{t+1}^{-1} \frac{P_{t+1} + D_{t+1}}{P_t} \tag{20.5}$$

and hence

$$\frac{P_t}{D_t} = R_{t+1}^{-1} \left(1 + \frac{P_{t+1}}{D_{t+1}}\right) \frac{D_{t+1}}{D_t}.$$

We can iterate this identity forward and take conditional expectations to obtain the identity

$$\frac{P_t}{D_t} = E_t \sum_{j=1}^{\infty} \left(\prod_{k=1}^{j} R_{t+k}^{-1} \Delta D_{t+k} \right), \tag{20.6}$$

where $\Delta D_t \equiv D_t / D_{t-1}$. (We could iterate (20.5) forward to

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

but prices are not stationary, so we cannot find the variance of prices from a time-series average. Much of the early volatility test controversy centered on stationarity problems. Equation (20.6) also requires a limiting condition that the price/dividend ratio cannot explode faster than returns, $\lim_{j\to\infty} E_t \left(\prod_{k=1}^{j} R_{t+k}^{-1} \right) P_{t+j}/D_{t+j}$. I come back to this condition below.)

Equation (20.6) shows that high prices must, mechanically, come from high future dividend growth or low future returns.

Approximate Identity

The nonlinearity of (20.6) makes it hard to handle, and means that we cannot use simple time-series tools. You can linearize (20.6) directly with a

Taylor expansion. (Cochrane [1991a] takes this approach.) Campbell and Shiller (1988a) approximate the one-period return identity before iterating, which is algebraically simpler. Start again from the obvious,

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

Multiplying both sides by P_t/D_t and massaging the result,

$$\frac{P_t}{D_t} = R_{t+1}^{-1} \left(1 + \frac{P_{t+1}}{D_{t+1}}\right) \frac{D_{t+1}}{D_t}.$$

Taking logs, and with lowercase letters denoting logs of uppercase letters,

$$p_t - d_t = -r_{t+1} + \Delta d_{t+1} + \ln\left(1 + e^{p_{t+1} - d_{t+1}}\right).$$

Taking a Taylor expansion of the last term about a point $P/D = e^{p-d}$,

$$\begin{aligned} p_t - d_t &= -r_{t+1} + \Delta d_{t+1} + \ln\left(1 + \frac{P}{D}\right) \\ &\quad + \frac{P/D}{1 + P/D}[p_{t+1} - d_{t+1} - (p - d)] \\ &= -r_{t+1} + \Delta d_{t+1} + k + \rho(p_{t+1} - d_{t+1}) \end{aligned} \qquad (20.7)$$

where

$$k = \ln\left(1 + \frac{P}{D}\right) - \rho(p - d).$$

Since the average dividend yield is about 4% and average price/dividend ratio is about 25, ρ is a number very near one. I will use $\rho = 0.96$ for calculations,

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

Without the constant k, the equation can also apply to deviations from means or any other point.

Now, iterating forward is easy, and results in the approximate identity

$$p_t - d_t = \text{const.} + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}). \qquad (20.8)$$

(Again, we need a condition that $p_t - d_t$ does not explode faster than ρ^{-t}, $\lim_{j\to\infty} \rho^j (p_{t+j} - d_{t+j}) = 0$. I return to this condition below.)

Since (20.8) holds *ex post*, we can take conditional expectations and relate price/dividend ratios to *ex ante* dividend growth and return forecasts,

$$p_t - d_t = \text{const.} + E_t \sum_{j=1}^{\infty} \rho^{j-1}(\Delta d_{t+j} - r_{t+j}). \tag{20.9}$$

Now it is really easy to see that a high price/dividend ratio *must* be followed by high dividend growth Δd, or low returns r. Which is it?

Decomposing The Variance of Price/Dividend Ratios
To address this issue, equation (20.8) implies

$$\text{var}(p_t - d_t) = \text{cov}\left(p_t - d_t, \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}\right) - \text{cov}\left(p_t - d_t, \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}\right). \tag{20.10}$$

In words, price/dividend ratios can *only* vary if they forecast changing dividend growth or if they forecast changing returns. (To derive (20.10) from (20.8), multiply both sides by $(p_t - d_t) - E(p_t - d_t)$ and take expectations.) Notice that both terms on the right-hand side of (20.10) are the numerators of exponentially weighted long-run regression coefficients.

This is a powerful equation. At first glance, it would seem a reasonable approximation that returns are unforecastable (the "random walk" hypothesis) and that dividend growth is not forecastable either. But if this were the case, the price/dividend ratio would have to be a *constant*. Thus the fact that the price/dividend ratio varies *at all* means that either dividend growth or returns must be forecastable—that the world is not i.i.d.

At a simple level, Table 20.1 includes regressions of long-horizon dividend growth on dividend/price ratios to match the return regressions. The coefficients in the dividend growth case are much smaller, typically one standard error from zero, and the R^2 are tiny. Worse, the *signs* are wrong in Table 20.1. To the extent that a high price/dividend ratio forecasts any change in dividends, it seems to forecast a small *decline* in dividends!

Having seen equation (20.10), one is hungry for estimates. Table 20.3 presents some, taken from Cochrane (1991a). As one might suspect from Table 20.1, Table 20.3 shows that in the past *almost all variation in price/dividend ratios is due to changing return forecasts.*

The elements of the decomposition in (20.10) do not have to be between 0 and 100%. For example, −34, 138 occurs because high prices seem to forecast lower real dividend growth (though this number is not statistically

Table 20.3. *Variance decomposition of value-weighted NYSE price/dividend ratio*

	Dividends	Returns
Real	−34	138
Std. error	10	32
Nominal	30	85
Std. error	41	19

Table entries are the percent of the variance of the price/dividend ratio attributable to dividend and return forecasts, $100 \times \text{cov}(p_t - d_t, \sum_{j=1}^{15} \rho^{j-1} \Delta d_{t+j})/\text{var}(p_t - d_t)$ and similarly for returns.

significant). Therefore they must and do forecast really low returns, and returns must account for more than 100% of price/dividend variation.

This observation solidifies one's belief in price/dividend ratio forecasts of returns. Yes, the statistical evidence that price/dividend ratios forecast returns is weak, and many return forecasting variables have been tried and discarded, so selection bias is a big worry in forecasting regressions. But the price/dividend ratio (or price/earning, market book, etc.) has a special status since it must forecast something. To believe that the price/dividend ratio is stationary and varies, but does not forecast returns, you have to believe that the price/dividend ratio does forecast dividends. Given this choice and Table 20.1, it seems a much firmer conclusion that it forecasts returns.

It is nonetheless an uncomfortable fact that almost all variation in price/dividend ratios is due to variation in expected excess returns. How nice it would be if high prices reflected expectations of higher future cash-flows. Alas, that seems not to be the case. If not, it would be nice if high prices reflected lower interest rates. Again, that seems not to be the case. High prices reflect low risk premia, lower expected *excess* returns.

Campbell's Return Decomposition

Campbell (1991) provides a similar decomposition for unexpected returns,

$$r_t - E_{t-1} r_t = (E_t - E_{t-1}) \left[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+j} - \sum_{j=1}^{\infty} \rho^j r_{t+j} \right]. \tag{20.11}$$

A positive shock to returns must come from a positive shock to forecast dividend growth, or from a negative shock to forecast returns.

Since a positive shock to time-t dividends is directly paid as a return (the first sum starts at $j = 0$), Campbell finds some fraction of return variation is due to current dividends. However, once again, the bulk of index return variation comes from shocks to future returns, i.e., discount rates.

To derive (20.11), start with the approximate identity (20.8), and move it back one period,

$$p_{t-1} - d_{t-1} = \text{const.} + \sum_{j=0}^{\infty} \rho^j (\Delta d_{t+j} - r_{t+j}).$$

Now take innovations of both sides,

$$0 = (E_t - E_{t-1}) \sum_{j=0}^{\infty} \rho^j (\Delta d_{t+j} - r_{t+j}).$$

Pulling r_t over to the left-hand side, you obtain (20.11). (Problem 3 at the end of the chapter guides you through an alternative and more constructive derivation.)

Cross Section

So far, we have concentrated on the index. One can apply the same analysis to firms. What causes the variation in price/dividend ratios, or, better book/market ratios (since dividends can be zero) across firms, or over time for a given firm? Vuolteenaho (1999) applies the same sort of analysis to individual stock data. He finds that as much as half of the variation in individual firm book/market ratios reflects expectations of future cashflows. Much of the expected cashflow variation is idiosyncratic, while the expected return variation is common, which is why variation in the index book/market ratio, like variation in the index dividend/price ratio, is almost all due to varying expected excess returns.

Bubbles

In deriving the exact and linearized present-value identities, I assumed an extra condition that the price/dividend ratio does not explode. Without that condition, and taking expectations of both sides, the exact identity reads

$$\frac{P_t}{D_t} = E_t \sum_{j=1}^{\infty} \left(\prod_{k=1}^{j} R_{t+k}^{-1} \Delta D_{t+k} \right) + \lim_{j\to\infty} E_t \left(\prod_{k=1}^{j} R_{t+k}^{-1} \Delta D_{t+k} \right) \frac{P_{t+j}}{D_{t+j}}, \tag{20.12}$$

and the linearized identity reads

$$p_t - d_t = \text{const.} + E_t \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}) + E_t \lim_{j\to\infty} \rho^j (p_{t+j} - d_{t+j}). \tag{20.13}$$

As you can see, the limits in the right-hand sides of (20.12) and (20.13) are zero if the price/dividend ratio is stationary, or even bounded. For these terms not to be zero, the price/dividend ratio must be expected to grow explosively, and faster than R or ρ^{-1}. Especially in the linearized form (20.13) you can see that stationary r, Δd *implies* stationary $p - d$ if the last term is zero, and $p - d$ is not stationary if the last term is not zero. Thus, you might want to rule out these terms just based on the view that price/dividend ratios do not and are not expected to explode in this way. You can also invoke economic theory to rule them out. The last terms must be zero in an equilibrium of infinitely lived agents or altruistically linked generations. If wealth explodes, optimizing long-lived agents will consume more. Technically, this limiting condition is a first-order condition for optimality just like the period-to-period first-order condition. The presence of the last term also presents an arbitrage opportunity in complete markets, as you can short a security whose price contains the last term, buy the dividends separately, and eat the difference right away.

On the other hand, there are economic theories that permit the limiting terms—overlapping generations models, and they capture the interesting possibility of "rational bubbles" that many observers think they see in markets, and that have sparked a huge literature and a lot of controversy.

An investor holds a security with a rational bubble not for any dividends, but on the expectation that someone else will pay even more for that security in the future. This does seem to capture the psychology of some investors from the (alleged, see Garber [2000]) tulip bubble of 17th century Holland to the dot-com bubble of the millennial United States—why else would anyone buy Cisco Systems at a price/earnings ratio of 217 and market capitalization 10 times that of General Motors in early 2000?

A "rational bubble" imposes a little discipline on this centuries-old description, however, by insisting that the person who is expected to buy the security in the future also makes the same calculation. He must expect the price to rise even further. Continuing recursively, the price in a rational bubble must be expected to rise forever. A Ponzi scheme, in which everyone knows the game will end at some time, cannot rationally get off the ground.

The *expectation* that prices will grow at more than a required rate of return forever does not mean that sample paths do so. For example, consider the bubble process

$$P_{t+1} = \begin{cases} \gamma R P_t, & \text{prob} = \frac{P_t R - 1}{\gamma P_t R - 1}, \\ 1, & \text{prob} = \frac{P_t R(\gamma - 1)}{\gamma P_t R - 1}. \end{cases}$$

Figure 20.2 plots a realization of this process with $\gamma = 1.2$. This process yields an expected return R, and the dashed line graphs this expectation as of the first date. Its price is positive though it never pays dividends. It

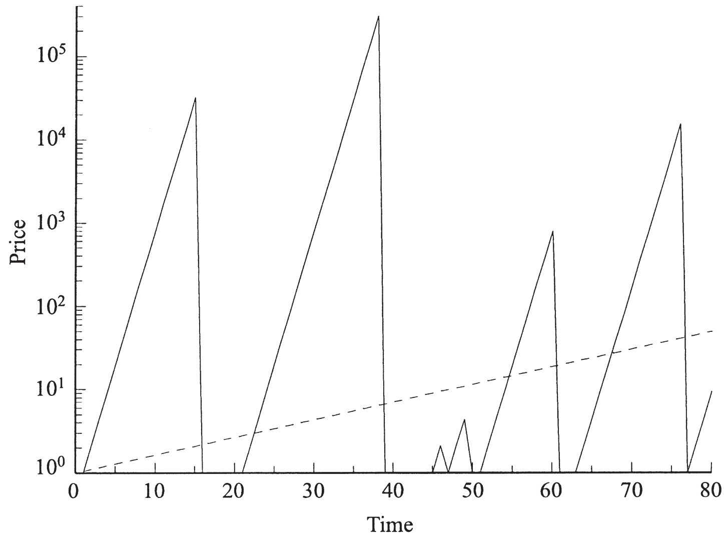

Figure 20.2. *Sample path from a simple bubble process. The solid line gives a price realization. The dashed line gives the expected value of prices as of time zero, i.e., $p_0 R^t$.*

repeatedly grows with a high return γR for a while and then bursts back to one. The *expected* price always grows, though almost all sample paths do not do so.

Infinity is a long time. It is really hard to believe that prices will rise *forever.* The solar system will end at some point; any look at the geological and evolutionary history of the earth suggests that our species will die out a lot sooner than that. Thus, the infinity in the bubble must really be a parable for "a really long time." But then the "rational" part of the bubble pops—it must hinge on the expectation that someone will be around to hold the bag; to buy a security without the expectation of dividends or further price increases. (The forever part of usual present-value formulas is not similarly worrying because 99.99% of the value comes from the first few hundred years of dividends.)

Empirically, bubbles do not appear to be the reason for historical price/dividend ratio variation. First, price/dividend ratios do seem stationary. (Craine [1993] runs a unit root test with this conclusion.) Even if statistical tests are not decisive, as is expected for a slow-moving series or a series such as that plotted in Figure 20.2, it is hard to believe that price/dividend ratios can explode rather than revert back to their four-century average level of about 20 to 25. Second, Table 20.3 shows that return and dividend forecastability terms add up to 100% of the variance of price/dividend ratios. In a bubble, we would expect price variation not matched by any variation in expected returns or dividends, as is the case in Figure 20.2.

I close with a warning: The word "bubble" is widely used to mean very different things. Some people seem to mean any large movement in prices. Others mean large movements in prices that do correspond to low or perhaps negative expected excess returns (I think this is what Shiller [2000] has in mind), rather than a violation of the terminal condition, but these expected returns are somehow disconnected from the rest of the economy.

A Simple Model for Digesting Predictability

> To unite the various predictability and return observations, I construct a simple VAR representation for returns, price growth, dividend growth, dividend/price ratio. I start only with a slow-moving expected return and unforecastable dividends.
>
> This specification implies that d/p ratios reveal expected returns.
>
> This specification implies return forecastability. To believe in a lower predictability of returns, you must either believe that dividend growth really is predictable, or that the d/p ratio is really much more persistent than it appears to be.
>
> This specification shows that small but persistent changes in expected returns add up to large price changes.

We have isolated two important features of the long-horizon forecast phenomenon: dividend/price ratios are highly persistent, and dividend growth is essentially unforecastable. Starting with these two facts, a simple VAR representation can tie together many of the predictability and volatility phenomena.

Start by specifying a slow-moving state variable x_t that drives expected returns, and unforecastable dividend growth,

$$x_t = bx_{t-1} + \delta_t, \tag{20.14}$$

$$r_{t+1} = x_t + \varepsilon_{r\,t+1}, \tag{20.15}$$

$$\Delta d_{t+1} = \varepsilon_{d\,t+1}. \tag{20.16}$$

All variables are de-meaned logs. (The term structure models of Chapter 19 were of this form.)

From this specification, using the linearized present-value identity and return, we can *derive* a VAR representation for prices, returns, dividends,