2 Detailed notes on predictability.

This week we're going to study how returns behave *over time*. If you see a good return today, does that mean it will continue with some "momentum" tomorrow, or will it "mean-revert" in a wave of "profit-taking?" Are there other signals that can tell you whether returns will be better or worse than average going ahead?

We'll also study two questions that seem to be different, but turn out to be the same: Is there such a thing as a "bubble," and hence its opposite, a great buying opportunity? And why do prices seem to vary so much?

Both issues come down to the question, can you forecast returns ahead of time? "Forecast" or "predict" does not mean with certainty, of course. But is there a way to know that the odds are in your favor on some days and against you on others?

Recent research has really changed how we think about this question.

2.1 Method

To find out if returns are (a bit) predictable, we can simply *run a regression*

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

We run regressions of tomorrow's return R_{t+1} on any variable we can see today x_t. If we find a "big" b, or a large R^2, you might be able to make money, buying when x_t is high and vice versa. If you find a small b, and low R^2, returns are not predictable, and you can't make much money this way. Thus, the regression measures the question "*are returns (somewhat) predictable?*"

- *We use forecasting regressions such as*

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

  to see if returns are (a bit) predictable.

Equivalently, this regression model implies that the expected return *at time t* is

$$E_t(R_{t+1}) = a + bx_t.$$

This statement is a bit subtle. The "expected return" can vary over time, being higher when the signal x_t is higher and vice versa. We can state the point of the regression equivalently as

- *The return-prediction regression measures whether expected returns (risk premiums) vary over time.*

"Expected returns" generates a common confusion, which we need to address right now. A coin flip, or a sequence of unpredictable returns, has the property that the *expected* return is constant over time. That does not mean that the *return* is constant over time, just that at each point, looking forward, you don't know which way it will go, and the probabilities of each outcome are always the same.

On the other hand, it's possible for *expected* returns or other quantities to move over time. For example, the *expected* temperature one week ahead moves steadily over the season, $E_t(T_{t+7})$ changes as t moves, lower in the winter and higher in the summer. The actual temperature of course moves even more.

This is a good time to remind yourself about conditional expectations and what $E_t(R_{t+1})$ means. The *unconditionally expected* temperature tomorrow in Chicago is about 60^o, the overall average. If it's July, or if you know that today's temperature is 90 degrees, the *conditionally expected* temperature tomorrow is high, maybe 85 degrees. The *actual, or ex-post* temperature tomorrow will vary beyond this expectation.

The question for us is whether stock returns are a bit like this; whether there are times, measured by the variable x_t, when the coin is 51/49 and other times when it's 49/51. We're asking if there are "seasons" in stock returns, not whether anyone knows exactly what the return (temperature) will be tomorrow.

- *"Predictable" doesn't mean perfectly. "Expected" means conditional mean but there is a lot of variance. "Risk" includes unexpected positive returns.*

Be aware that these terms, like many you will run across, have different meanings in finance and statistics than in colloquial usage.

Most people you will talk to don't really understand this. They think "predict" or "forecast" means a soothsayer, who can tell which way the market is going. The idea that a good "prediction" is just someone who can get 55/45 odds rather than 50/50 is not what they have in mind when they hear the word "predict." Someone may say "you predicted the market to rise but it went down. You don't know anything."

The idea that "predict" is the same as "conditional mean moves over time" is even more difficult. Make sure you understand it.

At an even more basic level, "expected" to us means "conditional mean," while most people use it to mean "best case." To them, "risk" is all "downside risk. To us, "risk" includes the 50% probability that you could make *more* money than you "expect," hardly conventional usage

2.2 Classic efficient markets view

Now, what result do we expect for this regression $R_{t+1} = a + bx_t + \varepsilon_{t+1}$? In the classic "efficient-markets" view, stock prices are not predictable, (loosely, the "random walk" view) so we should see $b = 0$, $R^2 = 0$, for any variable x_t.

Its worth remembering the logic behind this classic view. If you could predict good/bad days in the stock market, what would you do? On days when a signal x_t predicts good returns, you buy, and vice versa, of course.

But, we can't all do this. If everyone sees a high x_t, they also try to buy, driving prices up until today's price is the same as the price we expect tomorrow. *Competition in stock markets should drive out any predictable movement in stock prices.* Competition means that the information in the signal x_t will quickly be impounded in today's price. Thus, "informational efficiency" – the proposition that information is already impounded in today's price – is really nothing more than the predicted *effect* of competition and free entry.

Efficiency seems so easy, but it's so easy to slip in to classic fallacies. See if you can debunk these:

1. "The market declined temporarily because of profit-taking. It will bounce back next week."
2. "The stock price rises slowly after an announcement, as new information diffuses through the market."
3. "The internet is the wave of the future. You should put your money in internet stocks."
4. "Buy stocks of strong companies, with good earnings and good earnings growth. They will be more profitable and give better returns to stockholders."
5. "The demand curve slopes down;" "Big trades have a lot of price impact." "Stock prices fell today under a lot of temporary selling pressure," "Some stocks fell too far in the crash because mutual funds and hedge funds had to unload them to meet redemptions" "Small losers fall in December as dentists harvest tax losses."
6. "Some stocks fell too far in the crash because mutual funds and hedge funds had to unload them to meet redemptions"

You may feel uncomfortable debunking all of these. Remember, efficient markets and the proposition that returns are not predictable is a *theory*, not a fact or an assumption. It makes a *prediction* about each of the above statements, and your job was to fill in why efficient markets makes the prediction it does. But we have to look at the facts in each case, and they aren't always in line with the simple theory. In each case competitive speculators should eliminate the phenomenon, see if you can tell that story. In each case, in fact, there is substantial evidence that simple "efficiency" is wrong.

We often say "prices should follow a random walk." "Returns are basically price changes, so "returns are unpredictable" is basically the same as "prices follow a random walk." When we do theory we'll be more precise about all this.

A last but very important conceptual note. Look at the logic. I said that $b > 0$ means *you* should buy when x is high and sell when x is low. Then, I said that everyone would try to do it, so $b > 0$ couldn't describe a *market in equilibrium.* Thinking about how *a market in equilibrium* looks as opposed to looking at how the market (which may not be in equilibrium) looks to an individual trader is a big shift in view.

2.3 The facts

The first generation of tests found, surprisingly to Wall Street wisdom, that basically nothing forecasts stock returns. To give you a sense of both the fact and the method, here is a regression of returns on lagged returns. A positive b coefficient means "momentum," past good returns mean higher future returns. A negative one would have meant "overreaction" or "mean reversion. What are the facts? (Your problem set 1 will update the data and you get to run this regression.)

Regression of returns on lagged returns
Annual data 1927-2008
$R_{t+1} = a + bR_t + \varepsilon_{t+1}$

	b	t(b)	R^2	E(R)	$\sigma(E_t(R_{t+1}))$
Stock	0.04	0.33	0.002	11.4	0.77
T bill	0.91	19.5	0.83	4.1	3.12
Excess	0.04	0.39	0.00	7.25	0.91

Stock returns are just basically unpredictable based on past returns. $b = 0.04$ means that if returns go up 100% this year, you expect a rise of just 4% next year, a trivial amount of "momentum." This, the "economic" significance of a coefficient, is where you should always start when reading regressions.

That coefficient is statistically insignificant. The t stat if 0.33 is well below the standard 2.0 cutoff to bless a result "significant." (Remember, "significance" just answers the question, "how likely is it we see a number this big by chance if the real number is zero?" It does not answer the question "is it economically important.")

The R^2, measuring the proportion of return variance that can be forecast one year ahead. (In this regression $R^2 = b^2 = 0.04^2 = 0.0016$.) It's tiny.

T bills provide an interesting contrast. Here the coefficient is huge, 0.91, with a t stat and R^2 as impressive as any you will ever see.

Of course we know this; just a glance at figure 3 shows the slow, predictable movements of T bill returns and the very jagged coin-flippish behavior of stock returns. I put this regression in in part to show you that b doesn't *have to* come out to zero, a series labeled "return" can be quite predictable.

And you know this about interest rates. If interest rates (and hence the one-year return on treasuries) were high last year, they are very likely to be high again this year. Most of the t bill return is known ahead of time.

In fact, you can forecast (nominal) T bill returns with 100% R^2.If you regress log nominal one year T bill returns on their current yield (rather than last year's yield), $r_{t+1} = a + by_t^{(1)} + \varepsilon_{t+1}$ you will get $b = 1$ and $R^2 = 100\%$. The return is, mechanically $r_{t+1} = -y_t^{(1)}$, that's the definition. T bill returns are *very* forecastable!

Does that mean that T bill markets are "inefficient?" No. If you know stock returns will be high next year, you can borrow money and invest in the market. If you know interest rates are high to t bill returns will be high next year, you'd have to borrow at the same high rate to invest, which doesn't do you any good.

All you can do about high interest rates is to save more and consume less, and that's a lot harder to do. We do this response of course, and especially we see international capital chasing high interest rates. But

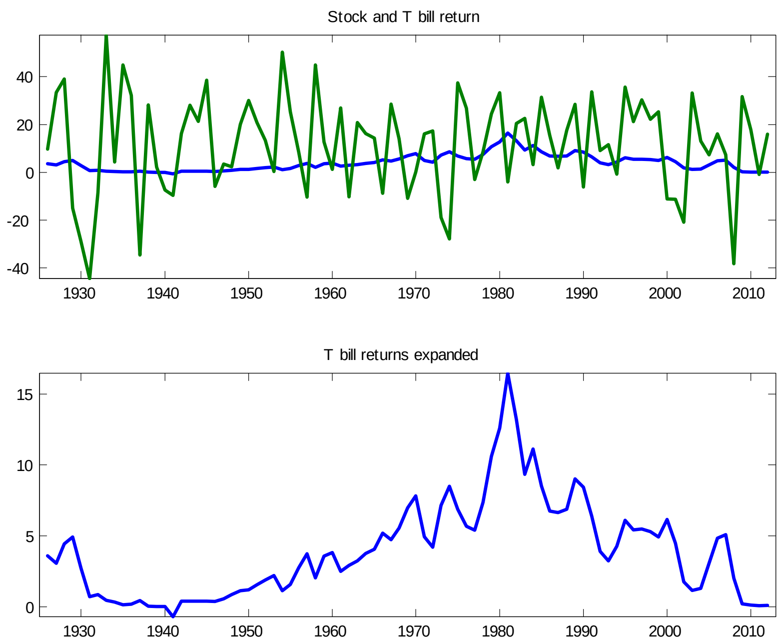


Figure 3: Time series of stock and T bill returns.

it's a lot less effective at making interest rates unpredictable than borrowing to invest is at making stock returns unpredictable.

For this reason, the "right" way to run forecasting regressions is to try to forecast the *excess* return, $R^e_{t+1} = R^{\text{stock}}_{t+1} - R^{\text{bond}}_{t+1}$, the return you can achieve by borrowing a dollar and investing. Exploiting this return takes no money out of pocket, only the willingness to bear risk. Studying it neatly separates willingness to consume less and save, "intertemporal substitution" from the willingness to bear risk.

As you see in the last line of the table, when we forecast *excess* returns we get nothing, just like forecasting stock returns, so we got the right basic answer forecasting stock returns and not worrying about interest rates. The reason is that stock returns are so much more volatile than interest rates, that the "return" and "excess return" series look much the same. Still, excess returns is the right way to do it.

This result is just a tip of an iceberg. For 40 years, an army of academic and industry analysts have been running regressions like this. Most of them are at a higher frequency, days and weeks rather than years.

Of course, when you do that you're bound to often come up with fish that seem significant, so out of sample and other statistical checks are important. "Big data" and big computers are multiplying the opportunities to fish around and find spurious forecasting regressions. But this table summarizes the conclusions of that massive academic effort pretty well: Until the 1980s, there was just not many variables that reliably forecast excess returns, at least with any economic significance.

This finding was a *revolution* in 1970, and is underappreciated to this day. Everyone thought, and many still do, that there are simple statistical systems that can be used to make money in markets. All you need is a clever statistician to run some regressions and forecast returns, some snazzy offices, and you're ready to start your hedge fund.

Logically, that proposition cannot withstand competition in financial markets. At least *ask*, "why didn't everyone else find this pattern already?"

2.4 New Facts

This view has been upended by two changes. First, researchers looked at *longer horizons.* Can we forecast returns for 5 or 10 years, rather than forecasting them for a few days or weeks? Obviously such forecasts are less interesting for high frequency trading, but it's still interesting to look. Second, researchers looked at *new variables*, and in particular *market prices.* Rather than use past returns as in the last table, what if we use prices or yields to forecast returns? After all, prices are supposed to tell us market expectations about the future.

Here is a simple regression of returns on the dividend/price ratio from "Discount rates,"

Table I

Return-forecasting Regressions

The regression equation is $R^e_{t\to t+k} = a + b \times D_t/P_t + \varepsilon_{t+k}$. The dependent variable $R^e_{t\to t+k}$ is the CRSP value-weighted return less the three-month Treasury bill return. Data are annual, 1947-2009. The five-year regression t-statistic uses the Hansen-Hodrick (1980) correction. $\sigma\left[E_t(R^e)\right]$ represents the standard deviation of the fitted value, $\sigma(\hat{b} \times D_t/P_t)$.

Horizon k	b	t(b)	R^2	$\sigma\left[E_t(R^e)\right]$	$\frac{\sigma[E_t(R^e)]}{E(R^e)}$
1 year	3.8	(2.6)	0.09	5.46	0.76
5 years	20.6	(3.4)	0.28	29.3	0.62

This is really a dramatically different result. *The return-forecasting coefficient is huge.* (Always look at the economic size of coefficients before going on to statistical tests. A statistically insignificant elephant is worth looking at, a statistically significant mouse, maybe not.)

Why do I say huge? Regression coefficients depend on units, and 0.0001 light year is a long walk. To think about this, let's think about what we expect. If the dividend yield goes up one percentage point, what does this mean? An unsophisticated investor might say "well, they're paying 1% more dividends, so I get 1% more return," a coefficient of 1.0. A sophisticated efficient-markets student would say, "no, the coefficient should be zero. Returns are not forecastable. The price is low relative to current dividends (D/P went up 1%) because everyone knows dividends will fall in the future. You forgot it's future dividends that count."

On this scale, reality is 2.8 percentage points past the unsophisticated investor! Not only do you get the current 1 percentage point of dividend yield, but either future dividends or future prices (turns out to be prices) go your way too!

In equations,

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

so, suppose dividend growth is constant $D_{t+1}/D_t = 1$ and price growth is constant $P_{t+1}/P_t = 1$, then today's dividend yield translates directly to return. If the dividend rate is 4%, $D_t/P_t = 0.04$ and $R_{t+1} = 1.04$. If the dividend yield rises to $D_t/P_t = 0.05$ then $R_{t+1} = 1.05$. So a regression of R_{t+1} on D_t/P_t would give a coefficient of 1.0.

However, if D_{t+1}/D_t or P_{t+1}/P_t decline correspondingly, then the higher dividend yield gives no extra return. For example, today's dividend yield could be high because today's dividend is \$6, and price is \$100, but everyone knows tomorrow's dividend will fall to \$5. Then the expected return is only 5%. Or, today's dividend yield could be high because everyone expects tomorrow's price to fall so the expected return is again 5%. It turns out, on average, that P_{t+1}/P_t goes 2.8 percentage points the "wrong way." A dividend yield of 6% means prices go up!

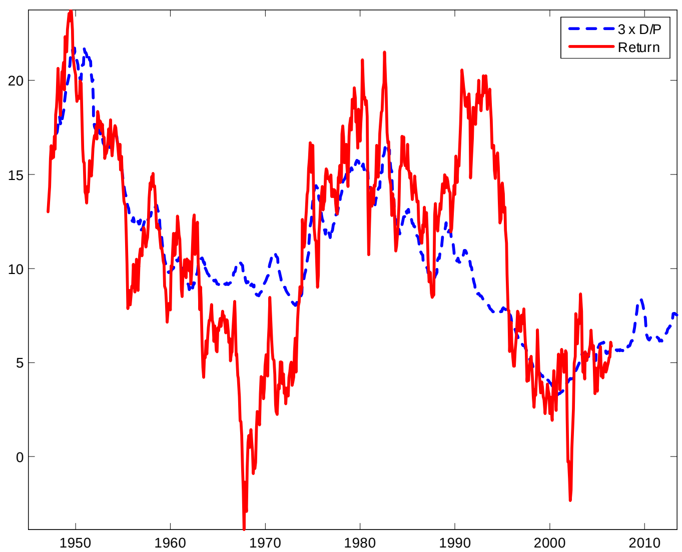
The R^2 may seem low at first blush, but remember we're forecasting returns here, and we're used to R^2 that are zero to three digits. Being able to forecast returns *at all* is big news, even if we still don't know exactly where returns are going.

The R^2 rise with horizon, becoming really interesting at the 5 year horizon. This is a way of showing the *economic* significance of forecastable returns.

The t statistic is significant, though not hugely so, and that significance doesn't really go up all that much at longer horizons – certainly not by a factor of 5. This is really a story about economic rather than statistical significance.

"Discount rates" summarized long-horizon predictability with the following graph. The top line is the dividend yield, and the bottom line is the average return for the 7 years *following* the dividend yield. (I multiplied dividend yield by 4, which is roughly the regression coefficient, to fit on the graph.) The correlation is evident: *high dividend yields (low prices) mean high subsequent returns; low dividend yields (high prices) mean low subsequent returns.*

This is a hugely important fact. It is the one central fact at the basis of all current arguments about "irrational investors," "bubbles" and so forth.



Dividend yield and following seven-year return. The dividend yield is multiplied by four. Both series use the CRSP value-weighted market index.

There is nothing really special about dividends here, so don't get too distracted by dividend questions. Price/earnings ratios, market / book (price x shares / book value of equity), etc. all work the same way. The trend to repurchases doesn't really make a difference. Dividends are just a convenient way to "detrend" prices. The key is that "low" prices – relative to dividends, earnings, book equity, etc. – signal higher future returns, and "high" prices signal lower future returns. (If you took time series, this is a classic example of a cointegration. Log prices and dividends are cointegrated, so if they are out of whack, one must revert over the long run. It turns out prices do the reverting.)

2.5 Time-varying expected returns

The regression model

$$R^e_{t+1} = a + b \times (D_t/P_t) + \varepsilon_{t+1}$$

means that at time t, the expected return is

$$E_t(R^e_{t+1}) = a + b \times D_t/P_t$$

The regression, then, means that expected returns vary over time. This is a very important concept for interpretation:

"Returns are predictable" = "Expected returns vary over time."

The coefficient 3.8 means that in order to see the expected return at any point in time, given the information provided by D/P (other variables can provide more or less information) just look at the D/P ratio and multiply by 3.8 (plus a constant). Figure 4 presents the dividend/price ratio. In the postwar era it has varied from about 1 to about 5, about 4 percentage points. That means that expected returns vary by $4 \times 3.75 = 15$ percentage points! That's huge! The long-term average expected return is only about 7 percentage points!

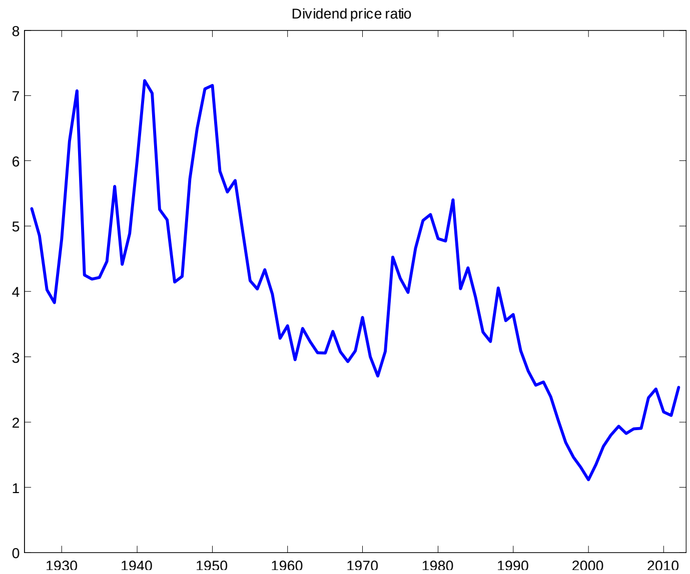


Figure 4: Dividend/Price ratio

(The dramatic 2008 price fall was matched by pretty dramatic falls in dividends, so the bump in D/P ratio is not as big as the price fall and rise. Interestingly we seem to be back at about pre-crash valuations – prices are lower, but so are dividends. Those valuations are still quite high by historical standards.)

The last two columns of the tables present this observation with some numbers, and suggest a more appropriate "R^2" calculation. From the regression model $R_{t+1} = a + bx_t + \varepsilon_{t+1}$ we can calculate the standard deviation of expected returns,

$$\sigma\left[E_t(R_{t+1})\right] = \sigma\left[bx_t\right] = b\sigma\left[x_t\right].$$

This is a tricky concept. The expected return $E_t(R_{t+1})$ changes over time, as today's forecast of tomorrow's temperature varies over time. Thus, we can ask, "how much do *expected* returns vary over time?" If you really wrap your head around the idea that expected returns $E_t\left(R_{t+1}\right)$ vary over time t, then $\sigma^2\left[E_t(R_{t+1})\right] = var(bx_t)$ will make a lot more sense.

Check: do you really understand the differences between variation in returns $\sigma(R_{t+1})$ variation in expected returns $\sigma\left[E_t(R_{t+1})\right]$? Do you really understand the difference between the mean $E(R_{t+1})$ and the conditional mean $E_t(R_{t+1})$?

The answer for D/P forecasts is $\sigma\left[E_t(R_{t+1})\right] = 5.2\%$ in the "Discount Rates" table, nearly as big as the average expected return itself! That means one-year expected returns are not a steady 7%, but are about 2% sometimes and about 12% at other times. This is huge!

Now, a and b are likely to be overfit. This measure characterizes the *estimate* of time-varying expected returns. On the other hand, at least 10 other variables also help to forecast returns. So, once we take the statistics into account, we still are likely to end up with a huge variation in expected returns.

The right two columns of the tables are an attempt to capture this idea in a measure like R^2.

The R^2 is defined as $var(bx_t)/var(R_{t+1})$. Remembering that the right hand side of the regression gives the conditionally expected returns, $E_t(R_{t+1}) = a + bx_t$, R^2 tells you how much *expected returns vary over time*, divided by the amount that actual returns vary over time.

The R^2 along with the size of b is often used as a measure of "economic significance." But, unlike regular regressions, where it makes sense to compare the variance of the "explained part" with the variance of the left hand side, that comparison makes little sense here. Nobody every claimed perfect clairvoyance, that he could *perfectly* forecast every movement.

To address that deficiency, the last two columns of the tables compare how much expected returns vary over time with the *level* of expected returns $\sigma(bx_t)/E(R_{t+1})$ rather than compare expected return variation with the *variance* of ex-post returns.

Running returns on past returns, that quantity is 0.77%, a tiny fraction of the 11.4% overall average. So even if you do believe $b = 0.04$, expected returns are very nearly constant through time. I think of this quantity as another "economic" measure of how much expected returns vary over time.

By contrast the T bill mean return is 4.1%, but the expected return varies by 3.1%, so that mean varies over time by almost as much as its average level – sometimes it's 1%, sometimes it's 7%. That's big!

And the variation of expected returns from the D/P ratio is 0.6-0.7 as large as the level of expected returns. That's big!

2.6 Long horizons

As you noticed, the R^2 was higher with longer horizons, and the 7 year return graph was really pretty. The coefficients got bigger, but the t stats didn't get any better. It seems that forecastability gets more *economically* interesting at long horizons, even if not more *statistically* significant.

But long-run forecasts are not really a different animal. They contain exactly the same information as short-run forecasts. They just present a different facet of that information that makes it look more interesting.

We'll see this in many ways later – much of "new finance" is really just old observations, but a new way of looking at them that makes us realize they're more important than we might have thought.

The key observation is that D/P is also very persistent. It's a series like the interest rate, that moves slowly over time, not one like stock returns that moves quickly. Its regression is

$$DP_t = a + 0.94 \times DP_{t-1} + \varepsilon_t$$

As a result, *long horizons are not a separate phenomenon. They result mechanically from short horizon results and a persistent forecasting variable.*

To see this fact mathematically, we can figure out what the two-year regression looks like given the one-year regression and the persistence of the forecasting variable. Write $x_t = D_t/P_t$ to keep it short, and suppress the constants (treat all variables as deviations from their means),

$$\begin{aligned} r_{t+1} &= bx_t + \varepsilon_{t+1} \\ x_{t+1} &= \phi x_t + \delta_{t+1}. \end{aligned} \tag{1}$$

Now find the *two-year* regression that is *implied* by this pair of one-year regressions:

$$\begin{aligned} r_{t+1} + r_{t+2} &= (bx_t + \varepsilon_{t+1}) + (bx_{t+1} + \varepsilon_{t+2}) \\ &= bx_t + b(\phi x_t + \delta_{t+1}) + \varepsilon_{t+1} + \varepsilon_{t+2} \\ &= b(1+\phi)x_t + (b\delta_{t+1} + \varepsilon_{t+1} + \varepsilon_{t+2}) \end{aligned} \tag{2}$$

Similarly,

$$r_{t+1} + r_{t+2} + r_{t+3} = b(1 + \phi + \phi^2)x_t + (error) \tag{3}$$

Thus, we should see the coefficients rise over horizon if ϕ is a big number– if the right hand variable moves slowly over time. In fact, that's just what we see – b coefficients that rise almost linearly with horizon and then taper off a bit. In a direct estimate, we get $\phi \approx 0.94$.

It's harder to work out R^2, but if you do, you see the pattern in the data that R^2 rises with horizon. It's easy to see with an approximation, that turns out to be very good. Returns remain very nearly uncorrrelated over time, $var(r_{t+1} + r_{t+2}) = 2var(r_t)$ and $var(r_{t+1} + r_{t+2} + ... + r_{t+k}) = kvar(r_t)$. The variance of the explained part is $b^2(1 + \phi + \phi^2 + ...)^2 var(x_t)$. So when $\phi \approx 1$, the variance of explained part is growing as horizon squared; the variance of the total is growing with horizon, so the ratio, which is R^2, grows linearly with horizon. (Similarly, in (2), we have $\sigma(\varepsilon) \approx 0.20, \sigma(\delta) \approx 0.15, b \approx 0.10$, so the $b\delta$ term is tiny compared to the ε terms. The variance of these errors grows linearly with horizon, the variance of the explained part is growing with the square of horizon.)

Working out the t statistic is even less fun, but it's easy to see what's going on. As you go to two year returns, the coefficient roughly doubles. But the sample size goes down by half, so the standard error roughly doubles too, and the t statistic really doesn't change much. (My "Dog that didn't bark" shows that there is in fact considerably more statistical information in long horizons, but that's a more subtle argument than we need right now.)

The fact that predictability improves at long horizons is also equivalent to the fact that *a high D/P predicts a high return for many years in the future.* Suppose we run r_{t+2} on x_t, skipping a year between observation and forecast,

$$r_{t+2} = b_2 x_t + v_{t+2}$$

Then, again using the two-period return forecast will be

$$r_{t+1} + r_{t+2} = (b + b_2)x_t + \varepsilon_{t+1} + v_{t+2}$$

You can see that the two-year coefficient will be about twice the one-year coefficient, if and only if b_2 is pretty big. Reading this equivalence backwards, the fact that long horizon coefficients rise with horizon means that b_2 is big too. Coefficients that rise nearly linearly with horizon must mean that today's high p/d forecasts low returns many years in the future.

The structure (1) lets us figure out what b_2 should be:

$$\begin{aligned}
r_{t+2} &= bx_{t+1} + \varepsilon_{t+2} \\
&= b\left(\phi x_t + \delta_{t+1}\right) + \varepsilon_{t+2} \\
r_{t+2} &= b\phi x_t + (b\delta_{t+1} + \varepsilon_{t+2}) \\
b_2 &= b\phi
\end{aligned}$$

So the structure (1) *implies* that you should get a big coefficient b_2 forecasting returns one year out. In the same way,

$$\begin{aligned}
r_{t+3} &= bx_{t+2} + \varepsilon_{t+3} \\
&= b\left(\phi x_{t+1} + \delta_{t+2}\right) + \varepsilon_{t+3} \\
&= b\phi\left(\phi x_t + \delta_{t+1}\right) + b\delta_{t+2} + \varepsilon_{t+3} \\
r_{t+2} &= b\phi^2 x_t + (b\phi\delta_{t+1} + b\delta_{t+2} + \varepsilon_{t+2}) \\
b_3 &= b\phi^2
\end{aligned}$$

and on and on,

$$r_{t+j} = b\phi^{j-1} d_t + (\text{ big error term}) \tag{4}$$

Vector autoregression

I snuck a new tool in, and it's a very useful one, the *vector autoregression.* I started with (1),

$$\begin{aligned}
r_{t+1} &= bx_t + \varepsilon_{t+1} \\
x_{t+1} &= \phi x_t + \delta_{t+1}.
\end{aligned}$$

which expresses the fact that dividend yields x_t forecast returns one year ahead r_{t+1},and also that dividend yields are persistent – they forecast themselves (big ϕ).

This is called a "vector autoregression" because you can write it

$$\begin{aligned}\begin{bmatrix} r_{t+1} \\ x_{t+1} \end{bmatrix} &= \begin{bmatrix} 0 & b \\ 0 & \phi \end{bmatrix}\begin{bmatrix} r_t \\ x_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ \delta_{t+1} \end{bmatrix} \\ z_t &= Az_{t-1} + v_t\end{aligned}$$

This is just a regression of r and x on one lag. Then, all the things we learned to do with simple AR(1) can now apply.

More generally, of course, there won't be zeros in the A matrix – we include lagged returns too,

$$\begin{aligned}\begin{bmatrix} r_{t+1} \\ x_{t+1} \end{bmatrix} &= \begin{bmatrix} b_{rr} & b_{rx} \\ b_{xr} & \phi \end{bmatrix}\begin{bmatrix} r_t \\ x_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ \delta_{t+1} \end{bmatrix} \\ z_t &= Az_{t-1} + v_t\end{aligned}$$

but if you try it you will see that the first column is essentially zero. So for summarizing the data in this case, we really don't need it.

Summary

We *derived* from this VAR structure the proposition that long run return coefficients rise with horizon, (3)

$$r_{t+1} + r_{t+2} + r_{t+3} = b(1 + \phi + \phi^2)x_t + (error)$$

and that dividend yields x_t forecast returns more than one year ahead,

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

Thus, those three apparently dissimilar aspects of the data are in fact closely related. We'll use vector autoregressions a lot.

Figure 5 summarizes the intuition: The blue line corresponds to the first argument: a high dp forecasts a high future dp, so forecasts a high future return. The green line corresponds to the second argument: high dp forecasts future returns directly. The red bars show that *DP forecasts a big multi-year return because it forecasts many small returns for many years, but all in the same direction.*

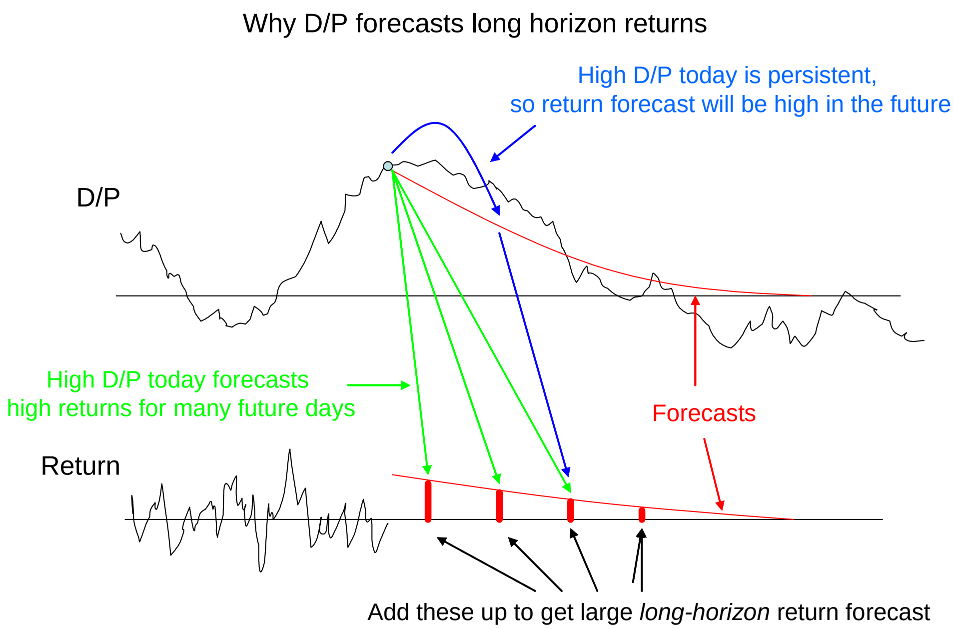


Figure 5: Intuition for the connection between persistent forecasting variables and long-run forecastability

Bottom line

1. *Forecasts from persistent variables build up over time and are more important at long horizons. Forecasts from fast-moving variables die out more quickly.*

2. *Long-horizon forecasts result when the forecasting variable (D/P) forecasts one-year returns in the same direction for many years in the future.*

3. *There is nothing special or different about long-run forecasts. They are the mechanical result of short-run forecasts and a persistent forecasting variable.*

2.7 Efficiency

We started with regressions as evidence for "efficiency." Does predictability mean that markets are "inefficient?"

No. (I.e. "not necessarily.") Efficiency always had an escape hatch, the "model of market equilibrium." Nothing in the definition of "efficiency" requires that expected returns – "required returns," "discount rates" – remain constant over time. (This is why I made such a fuss that "predictability" is the same thing as "variation in expected returns.")

So, the question is whether the forecasts we see correspond to a "rational variation in the equilibrium risk premium," or to something else, like "irrational exuberance" or "bubbles." To do that at all scientifically, you need a model and we'll look at what a "model" means soon. But we can at least think about what's plausible.

If the forecasting variable moved very quickly over time – if expected returns were 1% on Monday, Wednesday and Friday, but 13% on Tuesdays and Thursdays, we'd have a hard time cooking up a model of "equilibrium risk premium."

But that's not our situation. Dividend yields move very slowly over time. This isn't a trading strategy or an "arbitrage." Expected returns are high and low for years at a time.

Dividend yields also have a noticeable correlation with business cycles. Notice the uptick in dividend yields at the end of 2008. Imagine trying to convince an investor to buy stocks in Dec 2008: You: "Dividend yields are high, so returns going forward in the stock market look better than they have in years." Investor: "Thanks, I know that, but I'm about to lose my job, my company may not be able to roll over its debt, and my house is about to foreclose. I can't take more risk right now." In 2005, the answer to that question might have been "Sure, bring it on."

What you're feeling here is a natural, recession-related risk premium, which we might expect of an efficient market. In bad times, investors can't hold so much risk. They try to sell their stocks. They can't, collectively, so their efforts just lower prices relative to dividends. This keeps going until investors, collectively, regard stocks as enough of a better deal (high future return) to compensate for investors' lesser ability to hold risk right now. What do we see? Low prices, high D/P, followed, on average, by good returns.

This may be hard logic, and illustrates the difference between economic logic – how does a market behave in equilibrium, when everyone has bought all they want – and trader logic – how does a market look that is full of unexploited opportunities, i.e. supply has not settled down to equal demand.

Here's a test: do you think expected returns should be higher in good times, or bad times? Answer: Bad times. In good times, cashflows will be better, but you'll pay higher prices for those cashflows. Market equilibrium expected returns are higher when everyone else is scared.

I regard the discovery of a recession-related, slow time-varying risk premium as the basic lesson of predictability. And it's big! Within the general case of "efficiency," nobody thought risk premium could vary through times *so much.* Now we know they do.

However, this interpretation is not universal. To behavioral types, the fact that high prices are followed by low returns and vice versa is a sign of irrational euphoria and depression. Nothing we have said so far rules this out.

If you want to get anywhere at all in a debate like this, you have to understand the facts: *Long-run* *return predictability is the central (and pretty much only) fact behind all the arguments over "bubbles" and "irrationality" in asset prices.* People (reporters) have asked me often, "how do you explain the waves of euphoria and pessimism in the market?" Sorry, that's a story in search of a fact, it's not a fact by itself. Return predictability is a fact, and it's the only fact.

The point of economics is not to tell pleasant stories after the fact, "the market went up because risk aversion went down," "the market went up because of a wave of irrational optimism," or "sentiment increased," or "because the Gods are pleased." Our job is to make real economic theories, that tie risk aversion (or sentiment) to other facts in a quantitatively convincing way.

I have suggested here that the link between stock prices and recessions suggests some economics – people are more risk averse in bad times. We'll come back later and look at a real model that displays this characteristic. It's not perfect. But that's the goal.

2.8 Implications for valuation and portfolios

Though the theoretical possibility of time-varying expected returns has been around since about 1970, and though this fact has been around since 1986, it has yet to make its way into much of finance practice. For example, it means that the "cost of capital" in corporate finance varies tremendously over time. You were probably taught to use the CAPM to do valuation, i.e.

$$\text{value} = \frac{\text{Expected cash flow}}{\text{Expected return}} = \frac{\text{Expected cash flow}}{R^f + \beta E(R^m - R^f)}.$$

Then you spent a lot of time thinking about cash flows and betas. We just learned that the $E(R^m - R^f)$ in the denominator varies from 2% to 12%, swamping all the other terms! (In practice, people seem to jigger the formula to get the right answer, as investment rises when prices are high relative to dividends despite the formula which says it shouldn't.)

It's attractive to think about dividend yield predictability as a "trading rule" for "tactical asset allocation," but we have to think about that much harder, and we will at the end of the class. The fact that you have to wait 7 years to get returns puts a bit of a damper on it.

Also, this pattern is so obvious, and so long lasting that you really need to think "why doesn't everyone buy low and sell high, thereby removing the price and expected return variation?" Whatever is making everyone else scared in, say, December 2008, must not apply to you before you take contrarian advice. Viewed as a "market in equilibrium," phenomenon, the portfolio advice is simply to hold the market.

2.9 The nature of forecasting regressions

We're using regressions and regression statistics in very different ways from what you might have learned in typical statistics classes.

You're used to thinking of regressions as "cause" on the right and "effect" on the left. That's not what we're doing here (and it's an interpretation statistics classes emphasize far too much.) Here, we're using regressions to answer a simple forecasting question, with no causal interpretation.

For example, how do you check a weather forecaster? You could run a regression of the actual temperature on his forecasts,

$$\text{actual temperature at } t+1 = a + b \times (\text{prediction made at } t) + \text{forecast error}_{t+1}$$

If he's a good forecaster, you should see $b = 1$ (the forecast is unbiased). It would be nice if he had a good R^2– when he says it will be hot, it is, on average. However, you know that you can't make a nice weekend by forcing the forecaster to announce a better forecast. Causality runs from *left* to *right* here. When actual temperature will be higher, there is information about this that the forecaster sees, and he adjusts his forecast accordingly.

In a deep sense, then, *we use forecasting regressions in finance to understand how the* **right** *hand variable is formed, from expectations of the left hand variable!.* When we run returns and dividend growth on D/P, what we learn is that D/P is moving around, on average, in reaction to discount rate news not to cashflow news. We learn about D/P!

The statistical justification for using regressions is not "cause" on the right and "effect" on the left. The statistical assumption behind regressions is just that the right hand variable is uncorrelated with the error term. In usual regressions, you have multiple uncorrelated (you hope) causes x of an effect y, so an error term consisting of left-out causes is (you hope) uncorrelated with the right hand variables. But forecast errors are always uncorrelated with best forecasts. If not, you could exploit the correlation to make a better forecast. Thus it's right to run the regression this way, even we do *not* have cause on right and effect on left.

The basic story for our regressions is the "third variable" case you may have learned about in a regression class. Some news comes along about cashflows or discount rates, and traders agree on a higher price, changing the right-hand side of our forecasting regression. This news is (hopefully) also correlated with actual dividend growth or returns on the left hand side, otherwise it isn't news. We see what an econometrician would call a "spurious" correlation between price/dividend ratios on the right and subsequent returns on the left. He'd be right in a *causal* sense – if the government forced higher prices on us (like through a short-sale ban) we would not then get higher subsequent cashflows. But that does not mean we can't run this regression or learn anything about the world from it. It is a perfectly valid forecasting regression – with no causal interpretation – and it does let us learn about causes and effects which run the other way.

Again, many people understand the word "forecast returns" to mean you know exactly what will happen. For us, of course, any $\|b\| \neq 0$ or $R^2 > 0$ is interesting, even if small. In the financial world, just knowing when it's a 51/49 vs. 49/51 bet is extremely worthwhile. Predictability tells you when to take better chances, – not how to make more money every day.

However, be aware your investors may not understand this subtlety and they will be annoyed if you lose money on occasion! Most people don't understand that "forecast" and "predict" means a positive R^2 not an R^2 of one.

Similarly, the usual reaction to my tables is to disparage the "low" R^2. This is another bad habit from statistics classes that teach you wrongly that high R^2 is a good sign for a model. In finance, any R^2 above zero is interesting. The R^2 here says that 7% of the variance of returns is predictable one year ahead, and 23-26% of the 5 year variance is predictable ahead of time. These are *huge* R^2 in our business. If you could get those in a day you'd be fabulously rich.

At heart, we are not asking the usual "explanatory" question, "why did returns go up?" We are asking "how much of the return can you know ahead of time?" Any R^2 is important and exciting for that question. Stats classes are used to addressing the first question. We're interested in the second.

Even in statistics, the idea that a good R^2 is a sign of a "good regression" is usually wrong. For example, compare the regression

$$\text{left shoe sales} = a + \ b \text{ shoe price} + \ c \times \text{right shoe sales} + \ \text{error}$$

$$\text{left shoe sales} = a + \ b \text{ shoe price} + \ \text{error}$$

Which has higher R^2? Which is a "better regression" for understanding the effect of price changes on shoe sales? Hint: how do you interpret the coefficients in each case?

More technically, R^2 compares the variance of the right hand side, $\sigma^2\left[E_t(R_{t+1})\right] = \sigma^2(bx_t)$ to the variance of the left hand side $\sigma^2(R_{t+1})$. But we're not really interested in that question. Above, I thought it more interesting to compare the variation in expected returns $\sigma\left[E_t(R_{t+1})\right]$ to the *level* of expected returns. $E\left[E_t(R_{t+1})\right] = E(R)$. That calculation showed us that, if you believe this b, the market premium varies *dramatically* through time. R^2 didn't show us that, because that's not the question R^2 answers.

2.10 Dividends

Now, let's look at the parallel forecasts of dividend growth. Some facts from *Financial Markets and the Real Economy* (This updates *Asset Pricing* Table 20.1)

Horizon k	$R^e_{t\to t+k} = a + b\frac{D_t}{P_t} + \varepsilon_{t+k}$			$\frac{D_{t+k}}{D_t} = a + b\frac{D_t}{P_t} + \varepsilon_{t+k}$		
(years)	b	t(b)	R^2	b	t(b)	R^2
1	4.0	2.7	0.08	0.07	0.06	0.0001
2	7.9	3.0	0.12	-0.42	-0.22	0.001
3	12.6	3.0	0.20	0.16	0.13	0.0001
5	20.6	2.6	0.22	2.42	1.11	0.02

OLS regressions of excess returns (value weighted NYSE - treasury bill) and real dividend growth on the value weighted NYSE dividend-price ratio. Sample 1927-2005, annual data.

Dividend growth *should* be forecastable. When prices are high relative to current dividends, this should be a signal that investors expect dividends to be higher in the future, and they are bidding prices up on that expectation. On average, investors should be right, so on average, we should see higher dividend growth in the years following an aggressive valuation P/D. A regression of ΔD_{t+1} on P/D_t should give a positive coefficient; a regression of ΔD_{t+1} on D_t/P_t should give a negative coefficient.

Efficient markets does not mean "nothing is predictable!"

Alas, as the tables show, *this view fails miserably.* Returns, which "should not" be predictable, are predictable. Dividend growth, which *should* be predictable, is not predictable. The point estimate of the dividend prediction is slightly positive, which is the wrong sign. The t and R^2 are miserable, though, meaning this coefficient is zero for all practical purposes.

Figure 6 gives a graphical interpretation of what we expect (based on the constant expected return - efficiency view), and the way the world seems to be. When prices decline relative to current dividends, this "should be" a sign that, on average, future dividends will decline. The fact seems to be exactly the opposite. When prices decline relative to dividends, we see a higher return as prices slowly rebound, and there seems to be no expectation of changing future dividends.

(I emphasize "on average" in interpreting both the graph and the regression. The graph paints the *expected* path, but reality will often be different.)

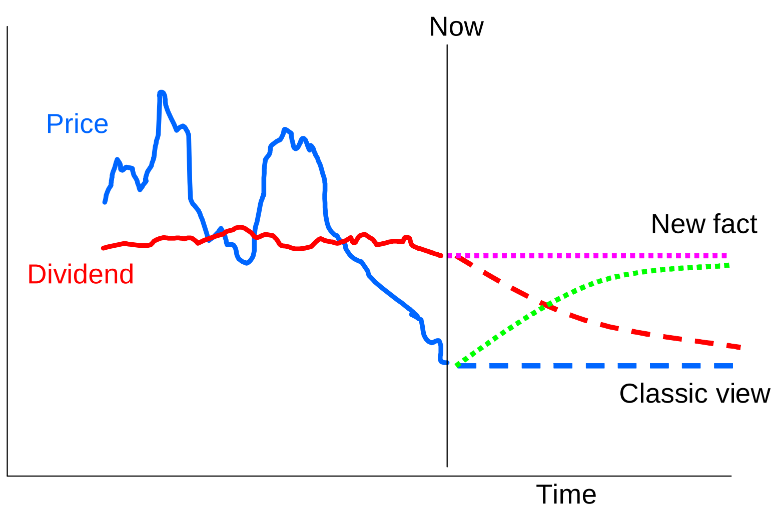


Figure 6: The new and old interpretation of a low price/dividend ratio

2.11 Present value identities and predictability – introduction

I'm telling stories that involve present value logic: People see news about future dividends or required returns; that news is impounded into current prices; the current high prices are then followed by higher dividends or poor returns whose expectation made prices low to begin with. It's also clear that returns *are* forecastable because dividends *are not* predictable. The two lines in the above graph end up in the same place. Conversely, the old constant-expected-return efficient-markets view said that in order for returns *not* to be predicable, dividends *should be* predictable. Again, both lines in the graph cannot be horizontal, or we never get back to the original P/D.

To get a handle on this kind of logic we need to write down a present value formula. This will provide clarity (I hope) and lots of unexpected insights. It's also a major and new tool.

2.11.1 One period present value model

You've seen present value formulas before. They're big ugly beasts like

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

We're going to develop a *linear* present value formula which is much simpler, but is rich enough to capture the interesting dynamics we are seeing in asset returns. For example, simplifying to a constant "required return" a

$$P_t = E_t \sum_{j=1}^{\infty} \frac{1}{R^j} D_{t+j}.$$

is a nicer (though still nonlinear) formula, but we know the required return is not constant with time (t) or horizon (j, k).

Below, we'll study proper long-horizon present value formula. But you can see the ideas (if not the numbers) with a very simply one-period model.

If an asset only lasts one period, returns are, by definition,

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

We can take expectations, and solve for price or price-dividend ratio

$$\begin{aligned} E_t\left(R_{t+1}\right) &= \frac{E_t\left(D_{t+1}\right)}{P_t} \\ P_t &= \frac{E_t(D_{t+1})}{E_t(R_{t+1})} \\ \frac{P_t}{D_t} &= \frac{E_t(D_{t+1}/D_t)}{E_t(R_{t+1})}. \end{aligned} \tag{5}$$

We can do the same thing in logs too, which makes the formula linear,

$$\begin{aligned} r_{t+1} &= d_{t+1} - p_t; \\ p_t &= d_{t+1} - r_{t+1} \\ p_t - d_t &= \Delta d_{t+1} - r_{t+1} \end{aligned}$$

This is true ex-post. A high dividend growth means a high return. Anything that's true ex post is true ex ante as well, so we can take expectations and write a present value formula.

$$p_t - d_t = E_t(\Delta d_{t+1}) - E_t\left(r_{t+1}\right) \tag{6}$$

(It's quite common in finance to use the same idea without logs. In a one-period model,

$$\begin{aligned} R_{t+1} &= \frac{D_{t+1}}{P_t} \\ E_t(R_{t+1}) &= \frac{E_t\left(D_{t+1}\right)}{P_t} \\ P_t &= \frac{E_t\left(D_{t+1}\right)}{E_t(R_{t+1})}. \end{aligned}$$

You use this idea in corporate finance classes to create discounted cash flow valuation models, and then you're taught to use the CAPM $E_t(R_{t+1}) = R^f + \beta E_t(R^m_{t+1} - R^f_t)$ to discount. This version of the algebra does not naturally extend to multiperiod models and time-varying expected returns. The log version of the same logic does, which is why I do it that way.)

These formulas allow us to be much more clear about some of the stories I've been telling.

These formulas represent the price as the discounted value of future dividends. If expected future dividends are higher, the price goes up. Duh.

If expected returns rise, the price goes down. What? If expected returns rise, shouldn't people flock to the stock and raise the price? No, and this is a good example of individual vs. equilibrium thinking. The return is gained by buying at a low price and then enjoying the dividends. If people did flock to the stock, they'd drive the price up and the return down.

Here's how to think of it: perceived risk goes up, so people try to sell the stock. That drives down the price and therefore drives up the expected returns. So higher expected return *in equilibrium* corresponds to a lower price.

2.11.2 My story for predictability

I have been telling stories about *traders* seeing information, and connecting those stories to *our* ability to forecast returns or dividend growth given prices. Now, we can make some sense out of those stories.

If traders see some information that dividend growth will be higher, they bid up prices. Similarly, if their required returns rise, they try to sell, bidding down prices.

How can *we* tell what's going on? We can't see traders' news, all we see is prices going up. We can't tell what caused an *individual* price rise, but we can tell whether prices are, on average, varying on dividend growth news or return news, because *on average traders must be right.* If traders bid prices up on dividend news, sometimes they will be right, but sometimes dividends will disappoint or do even better than forecast. But if dividends don't rise *on average* following such events, the "news" they're trading on is garbage. The weather forecaster may not be right on any individual day, but if *on average*, the weather is 50^o on days following a 60^o forecast, he's an awful forecaster.

Thus, if price variation comes from news about dividends growth, then on average dividend growth should be higher after a price rise. "On average y_{t+1} is higher when x_t is higher" is exactly the same thing as "a regression of y_{t+1} on x_t should give a positive coefficient." Thus,

- *If price variation comes from news about dividend growth, then price-dividend ratios should forecast dividend growth.* Conversely,
- *If price variation comes from news about changing discount rates, then price-dividend ratios should forecast returns.*

To make these ideas precise, suppose that expected returns are constant at $\bar{r}$, so prices move when traders see high dividend growth. Use the present value identity (6),

$$p_t - d_t = E(\Delta d_{t+1}|I_t) - \bar{r}.$$

Now *we* can't see expected dividend growth, that's in the mind of traders. Here $E(\Delta d_{t+1}|I_t)$ means trader's expectations, not anything that we can see. But we can see actual dividend growth, which is equal to expected dividend growth plus a forecast error.

$$\Delta d_{t+1} = E(\Delta d_{t+1}|I_t) + \varepsilon_{t+1}.$$

Putting the last two equations together, you have

$$\Delta d_{t+1} = \bar{r} + 1.0 \times (p_t - d_t) + \varepsilon_{t+1}.$$

Now, if traders are doing a decent job of forecasting, ε_{t+1} must be unpredictable and hence uncorrelated with anything at time t, including $p_t - d_t$. Therefore, *this is a regression forecast.* Again, if prices are moving on news of future dividend growth – even news that we cannot see – then prices should forecast high dividend growth!

Similarly, if expected returns move down, then prices rise, and prices should then forecast returns.

$$p_t - d_t = \Delta\bar{d} - E\left(r_{t+1}|I_t\right).$$

$$r_{t+1} = E\left(r_{t+1}|I_t\right) + \varepsilon_{t+1}.$$

$$r_{t+1} = \Delta\bar{d} - 1.0 \times (p_t - d_t) + \varepsilon_{t+1}.$$

You know where we're going: price-dividend ratios seem to forecast returns, but not dividend growth. We conclude that *on average, market price-dividend ratios are moving almost entirely on expected return news, not on cashflow growth news.*

Notice how I have now "reversed causality." We learn from the regression coefficients not, as usual, how the *left (y)* variable is formed from the *right* (x) variable. The forecasting regression tells us how the *right* hand side variable $(p - d, d - p)$ is formed! We run r on $d - p$ only because the forecast error is uncorrelated with $d - p$.

The temperature forecast is again a good analogy. If we want to check a forecaster, we run

$$\text{temperature}_{t+1} = a + b \times \text{forecast made at } t + \varepsilon_{t+1}.$$

In terms of *causality,* we think the forecaster gets information about future temperature, this causes him to issue a forecast. If it's a good forecast, then $b = 1$ with a good R^2. What you learn about the world from this regression is about *how the forecast is made, not what causes the weather to be good or bad* as if it were the left-hand variable in the usual (mis) use of regressions in the social sciences.

2.11.3 Volatility and predictability

One step deeper: we can tie *predictability* to the *volatility* of prices. This is important. Bob Shiller won the 2013 Nobel Prize for pointing out the "excess volatility" of stock returns. It turns out that observation is exactly the same as the presence of return predictability, and, more directly, the absence of dividend growth predictability.

Here is Shiller's Nobel-prize winning graph. P is the actual stock price. P* is the "ex-post rational" stock price, $P_t^* = \sum_{j=1}^{\infty} \frac{1}{R^j} D_{t+j}$. Here the are both detrended. Shiller said, wait a moment; if prices are *expected* discounted dividends $P_t = E_t\left(P_t^*\right) = E_t \sum_{j=1}^{\infty} \frac{1}{R^j} D_{t+j}$, then prices should surely vary less than the thing they are expecting! Instead, prices vary wildly more than they "should" even if you knew future dividends perfectly!

Proof:

$$\begin{aligned} P_t^* &= P_t + \varepsilon_t \\ \sigma^2(P_t^*) &= \sigma^2(P_t) + \sigma^2\left(\varepsilon_t\right) \\ \sigma^2(P_t^*) &> \sigma^2(P_t) \end{aligned}$$

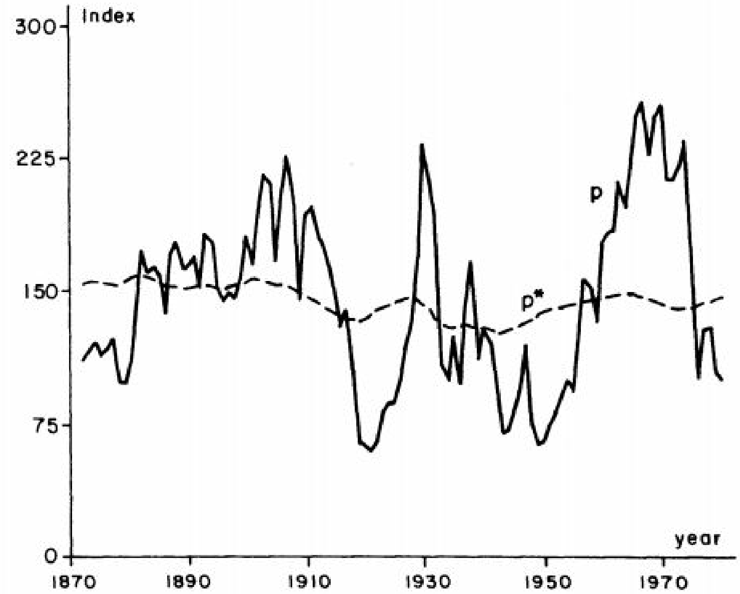


FIGURE 1
Note: Real Standard and Poor's Composite Stock Price Index (solid line p) and *ex post* rational price (dotted line p^*), 1871–1979, both detrended by dividing a long-run exponential growth factor. The variable p^* is the present value of actual subsequent real detrended dividends, subject to an assumption about the present value in 1979 of dividends thereafter. Data are from Data Set 1, Appendix.

Here is the same graph updated, from Shiller's Nobel Prize lecture

This originally seemed like a new and different kind of test. All you efficient markets types are wasting your time seeing if prices are *predictable*. You're missing the elephant in the room – prices *vary way too* *much.*

But is it really different from predictability? You have a hint in the graph – the high prices "should" be followed on average by higher dividend growth. They aren't. So, really, you can see already that *"excess" volatility is the same thing as the fact that high prices do not forecast dividend growth.*

What about returns? You will notice in my argument above that Shiller assumed constant expected returns and a constant discount rate. Well, couldn't the discount rate vary over time? Empirically Shiller and followers tried a variety of discount rate models, none of which worked. Shiller's view is intuitive: this is huge. Come on, guys, there is no way that time-varying discount rates can add up to something so huge.

But can they? We know dividend yields forecast returns, and excess returns. Do they forecast returns *enough* to account for this volatility.

The answer is yes: *price-dividend volatility is exactly and entirely accounted for by the return forecastability seen in our regressions. Return forecastability and excess volatility are* **exactly** *the same phenomenon.*

That doesn't mean we "explain" volatility. It just says it's the same phenomenon. Volatlity is another way to see the economic implications of return forecastability. This is great. We have one phenomenon, time-varying expected returns. When we understand that we understand everything!

2.11.4 Volatility and predictability in the one-period model

OK, nice words. Let's see that in equations. These are theorems, you don't chat about theorems, you prove them.

Return to

$$p_t - d_t = E_t(\Delta d_{t+1}) - E_t\left(r_{t+1}\right).$$

- *If expected dividend growth or returns were constant, neither $E_t(\Delta d_{t+1})$ nor $E_t\left(r_{t+1}\right)$ varying through time, then price-dividend ratios would be constant.*

This is cool. We know just by looking at $p - d$ that the world is not a "coin flip," with the future looking the same at every point in time. The fact that price-dividend ratios vary *at all* means that trader's expectations of dividend growth or returns *must* vary through time.

Now, we want to understand variance, so we need an equation with variance on the left. There are several ways to do it. Here's the one I like best first:

Run regressions of returns and dividend growth on $d_t - p_t$. (This just establishes notation)

$$\begin{aligned} r_{t+1} &= b_r(d_t - p_t) + \varepsilon^r_{t+1} \\ \Delta d_{t+1} &= b_d(d_t - p_t) + \varepsilon^d_{t+1} \end{aligned}$$

(As usual I'm ignoring constants.) Write the identity – the definition of return

$$d_t - p_t = r_{t+1} - \Delta d_{t+1} \tag{7}$$

Now plug the forecasting regressions in on the right hand side,

$$d_t - p_t = b_r(d_t - p_t) + \varepsilon^r_{t+1} - b_d(d_t - p_t) - \varepsilon^d_{t+1}$$

Since the terms multiplying $d_t - p_t$ must always cancel, the identity (7) implies

$$1 = b_r - b_d. \tag{8}$$

and also

$$0 = \varepsilon^r_{t+1} - \varepsilon^d_{t+1}.$$

Always stop to translate equations in to English. Here, the only way the return can be higher is if the dividend is higher or the initial price is lower. Hence, (8) the only way the *expected* return can be higher is if the *expected* dividend is higher or the initial price is lower. Hence, the only way the *unexpected* return can be higher is if the *unexpected* dividend is higher, since the initial price can't be unexpected.

A regression coefficient is covariance over variance,

$$1 = \frac{cov(d_t - p_t, r_{t+1})}{var\,(d_t - p_t)} - \frac{cov(d_t - p_t, \Delta d_{t+1})}{var\,(d_t - p_t)}.$$

so, changing sign to emphasize prices,

$$var\,(p_t - d_t) = cov(p_t - d_t, \Delta d_{t+1}) - cov(p_t - d_t, r_{t+1})$$

- *Price-dividend ratios can only vary if they forecast dividend growth or returns in regressions*
- *If price-dividend ratios vary at all, the difference between the dividend growth-forecasting regression coefficient and the return-forecasting coefficient must be one.*

The coefficients must *mechanically* add up. It's no accident that a dividend forecast coefficient is zero and a return forecast coefficient is big. In particular, both coefficients *can't* be zero!

What we learn from the regression table is *which* one is forecastable – and it's all expected returns!

2.12 Linearized present value identity

This is all fun, but not ready for the data. Our stocks last a long time. We need to do this with a "real" present value identity, not a one-period model. Here it is:

$$p_t - d_t \approx \sum_{j=1}^{\infty} \rho^{j-1} \left(\Delta d_{t+j} - r_{t+j}\right) \tag{9}$$

$p - d$ = log price/dividend ratio, Δd = log dividend growth , r = log return, and $\rho = 1/(1 + D/P) \approx$0.96.

All variables are deviations from their means. We're only interested in variation about the mean anyway. (Or add a constant to the right hand side)

This is a major new tool, and we'll use this a lot! It's often called the Campbell-Shiller formula, after its justly-famous inventors. It's also called the "dynamic Gordon growth formula," as it generalizes the formula you may have seen before

$$\begin{aligned} \frac{P}{D} &= \frac{1}{r - \Delta d} \\ p - d &= r - \Delta d \end{aligned}$$

which applies to constant returns r and dividend growth Δd (often called "g")

The formula is true ex-post. If you get a lot of dividends you get a big return. We can also take expected values and call it a present value formula,

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

The formula makes sense:

1. High dividend growth expectations $E_t \Delta d_{t+j}$ give high prices $p_t - d_t$. Low expected returns or low discount rates $E_t r_{t+j}$ also give high $p_t - d_t$.
2. The more *persistent* r or Δd, the greater their effect on p/d: More terms of $\sum_{j=1}^{\infty} \rho^{j-1}$ matter. This is pretty intuitive: if good dividend growth lasts longer, the eventual level of dividends is higher and the stock is worth more. Conversely, one period of high dividend growth or low returns don't have nearly the effect of a long-lasting shock.

The quantities on the right hand side of (9) are "long run" returns. You're used to constructing (say) the 5 year compound log return as the sum of intervening log returns $r_{t\to t+5} = r_{t+1}+r_{t+2}+r_{t+3}+r_{t+4}+r_{t+5}$. The object on the right hand side of (9) is a weighted sum of return in this same form, and gives approximately the same value.

$$r_t^{lr} \equiv \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} = \left[r_{t+1} + 0.96 r_{t+2} + 0.96^2 r_{t+3} + ...\right] \approx \left[r_{t+1} + r_{t+2} + r_{t+3} + ...\right]$$

The complete derivation, below, of the linearized identity isn't that revealing of intuition. It works exactly like the simple one period identity (6), which is just the first term of (9). The key step is a linear approximation to one period returns:

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

then take logs and approximate to get

$$r_{t+1} \approx \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} - (p_t - d_t). \tag{10}$$

Returns are high if you get an increase in prices, or if you get an increase in dividends with no change in valuation ratio. Solve this for $(p_t - d_t)$, iterate forward, and you're done.

There is a constant in the return approximation (10) and hence a constant in the approximation (9). To avoid carrying that around, think of all the variables as having had their means subtracted, i.e. x_t really means $x_t - E(x_t)$. That gets rid of the constant of approximation. Since we're only interested in variance and forecasts that makes it easier. But when you're programming, you may miss the constant._

The linear approximation in (10) and (9) is quite accurate for assets like the entire stock market, and for its purpose, which is to understand the dynamics of stock returns and price-dividend ratios with linear time series methods like regressions. It will not work for bonds (they have their own present value formulas), for options or other securities where nonlinear payoffs are really important, for individual stocks that can have dividends = 0 (there is another version for individual stocks using book/market ratios, see the appendix), or if you have in mind some very nonlinear model for dividends or expected returns.

The "present value formulas" (9) and (6) are just identities. We just rearrange the definition of return $R_{t+1} = \left(P_{t+1} + D_{t+1}\right)/P_t$. These identities just express the idea that you can always discount future dividends with returns.

The fact that these formulas are identities do not mean they aren't useful. You give them some content when you come up with a model of expected returns; the "model of market equilibrium" that Fama in 1970 taught us is needed to take "efficiency" past a tautology.

The present value identities (9) (6) let you *connect* ideas and facts about expected returns $E_t r_{t+j}$ to ideas and facts about prices p_t, and they let you do so *quantitatively.* That *is* useful, and that is their point. For example, you can answer "if expected returns rise one percentage point for 4 years, how much higher will the price-dividend ratio be?"

2.12.1 Algebra

Optional! Algebra for the return and present value identity. Start with the definition of returns. Our goal is to write returns in terms of stationary variables, price-dividend ratio and dividend growth, rather than nonstationary variables, price and dividend. So,

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

Take logs; $r_{t+1} = \log(R_{t+1}), d_{t+1} = \log(D_{t+1})$, etc.

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

Now, Taylor expand the last equation about a point $PD = e^{pd}$

$$\log\left(e^{p_{t+1}-d_{t+1}}+1\right) \approx \log\left(PD+1\right) + \frac{PD}{PD+1}(p_{t+1}-d_{t+1}-pd)$$

Define

$$\rho = \frac{PD}{PD+1}.$$

If P/D is 25, then $\rho = 25/26 \approx 0.96$ which is the value I use. Now we can write

$$r_{t+1} \approx [\log\left(PD+1\right) - \rho pd] + \rho(p_{t+1}-d_{t+1}) \ + \Delta d_{t+1} - (p_t - d_t)$$

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

Taking the mean

$$Er \approx k + \rho Epd + E\Delta d - Epd$$

and subtract the mean equation (left and right) from the return equation

$$r_{t+1} - Er \approx k - k + \rho\left[(p_{t+1}-d_{t+1}) - Epd\right] + \left[(d_{t+1}-d_t) - E\Delta d\right] - \left[(p_t - d_t) - Epd\right]$$

Now, treat every variable as a deviation from its mean, i.e. write $r_t = r_t - Er$, etc. The effect is simply to knock out the constant

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

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

To get to the preset value identity, rearrange,

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

Iterate forward.

$$p_t - d_t \approx (\Delta d_{t+1} - r_{t+1}) + \rho(\Delta d_{t+2} - r_{t+2}) + \rho^2(p_{t+2}-d_{t+2}).$$

$$p_t - d_t \approx (\Delta d_{t+1} - r_{t+1}) + \rho(\Delta d_{t+2} - r_{t+2}) + \rho^2(\Delta d_{t+2} - r_{t+2}) + \rho^3(p_{t+3}-d_{t+3})$$

$$p_t - d_t \approx \sum_{j=1}^{k} \rho^{j-1}\left(\Delta d_{t+j} - r_{t+j}\right) + \rho^k(p_{t+k}-d_{t+k}).$$

The last term is called the "rational bubble term," and causes a lot of fun. It "should" be zero. In statistical terms, $p_{t+k} - d_{t+k}$ is a stationary random variable, so it can't grow faster than ρ^{-k}. There is also a good economic argument that it should be zero. If it is not zero, then price can be positive for a security that never pays a dividend, which is a form of arbitrage opportunity. Hence, at least when we're not interested in exploring "rational bubbles," we assume that $\rho^k\left(p_{t+k} - d_{t+k}\right) \to 0$ and we have

$$p_t - d_t \approx \sum_{j=1}^{\infty} \rho^{j-1}\left(\Delta d_{t+j} - r_{t+j}\right).$$

2.13 Long run Predictability and Volatility

Now, let's make the same connection between $p-d$, volatility and predictability with this "real" present value identity that we did with the one-period model. Stare at

$$p_t - d_t \approx E_t \sum_{j=1}^{\infty} \rho^{j-1} \left(\Delta d_{t+j} - r_{t+j} \right) \tag{11}$$

Once again,

- *Price-dividend ratios can move if and only if there is news about current dividends, future dividend growth or future returns.*

That may seem pretty obvious, but maybe stating the converse will dramatize how important it is.

- *If Δd_t and r_t are totally unpredictable (*if $E_t(\Delta d_{t+j})$ and $E_t\left(r_{t+j}\right)$are the same for every time t*), then $p_t - d_t$ must be constant*!

This is a pretty stunning revelation. It's easy to come out of your first (and often second) finance course, and think that unpredictable returns and unpredictable dividend growth is a pretty decent view of the world, at least as a first approximation. Everything is like a coin flip. Returns are darn hard to predict, and who knows where dividends are going. If you do similar sorts of tests and autocorrelations, the idea that dividend growth is also unpredictable looks pretty good. But if you put these two ideas together, you see they predict a constant $p_t - d_t$, and we know that isn't true.

This isn't an observation you need any math for. Prices are about how the world looks going forward. If, looking ahead, the world always looks the same (in growth rates) then prices (relative to dividends) must always be the same. You don't need any fancy econometrics to know that our world *cannot* feature both unpredictable dividends and unpredictable returns. You *do* need to understand that there is a present value formula uniting prices, dividends, dividend growth, and returns, and then just observe that the price-dividend ratio varies.

Consider the regressions of long-run returns and dividend growth on $d_t - p_t$,

$$\sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} \equiv \Delta d_t^{lr} = b_d^{lr} \left(d_t - p_t \right) + \varepsilon^d \tag{12}$$

$$\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} \equiv r_t^{lr} = b_r^{lr} \left(d_t - p_t \right) + \varepsilon^r .$$

As before, these equations establish notation.

Now, look at the identity,

$$d_t - p_t \approx \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} - \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}.$$

Run both sides of the identity on $(d_t - p_t)$. The coefficients – terms multiplying $d_t - p_t$ must be equal, so

$$1 \approx b_r^{lr} - b_d^{lr}. \tag{13}$$

(As in the one period case, the errors must add up too,

$$0 = \varepsilon^r - \varepsilon^d.$$

This is the basis of the variance decomposition for returns, which I discuss in the Appendix.)

- *The long-run return forecasting regression coefficient and the long-run dividend growth forecasting regression coefficients must add up to one. They can't both be zero. If dividend yields vary, they must forecast long-run returns or long-run dividend growth.*

We had a "add up to one" with the simple model (8). A similar "add up to one" applies here. However, it does not apply to the simple one-period regression coefficients (a good thing, since they are 0.0 and 0.1). It applies to the *long run* regression coefficient (13).

The point of these formulas is *quantitatively* to show how *variance in dividend yields must be matched by dividend yields' ability to forecast returns and/or dividend growth.*

Now, covariances are the numerators of regression coefficients,

$$\beta = \frac{cov(x,y)}{var(x)}$$

so if you like, you can multiply both sides by $var(dp)$ and get

$$var\left(dp_t\right) \approx cov\left[dp_t, \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}\right] - cov\left[dp_t, \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}\right]. \tag{14}$$

Thus, the identity (14) tells us directly that

- *$p - d$ varies if and only if it forecasts long run dividend growth or long run returns.*

Relative to two bullets ago, *forecast* is the important word. If dividend yields vary, yes, they must correspond to changing *investor* expectations of dividend growth or returns. But we can't see investor expectations. Now we have something we *can* see. One or both of dividend growth and returns must be *forecastable,* by us in a regression sense, using dividend yields as the forecasting variable.

The long-run regression coefficients are sums of regressions of r_{t+j} on dp_t

$$1 \approx \sum_{j=1}^{\infty} \rho^{j-1} b_d^{(j)} - \sum_{j=1}^{\infty} \rho^{j-1} b_r^{(j)}. \tag{15}$$

where $b^{(j)}$ means the j-year ahead regression coefficient, i.e.

$$r_{t+j} = a_r^{(j)} + b_r^{(j)} (p_t - d_t) + \varepsilon_{t+j}$$

Thus, the volatility of price-dividend ratios corresponds not just to its ability to forecast returns and/or dividend growth *one* year ahead, but to forecast returns and/or dividend growth many years ahead, as illustrated by the red bars in Figure 5.

2.13.1 Which is it? Facts

The identity says that the long-run return and long-run dividend growth regression coefficients must add to one. Now we can ask the question, quantitatively, *which* is it? The long-run regressions suggest it's mostly returns, but let's do it right.

Table II from "Discount Rates" puts the facts as compactly as I am able to do so. (This corresponds to *Asset Pricing* Table 20.3, p. 398 and "The Dog that Didn't Bark")

Table II

Long-run Regression Coefficients

Table entries are long-run regression coefficients, for example $b_r^{(k)}$ in $\sum_{j=1}^{k} \rho^{j-1} r_{t+j} = a + b_r^{(k)} dp_t + \varepsilon_{t+k}^r$. Annual CRSP data, 1947-2009. "Direct" regression estimates are calculated using 15-year ex-post returns, dividend growth, and dividend yields as left-hand variables. The "VAR" estimates infer long-run coefficients from one-year coefficients, using estimates in the right-hand panel of Table III. See the Appendix for details.

	Coefficient		
Method and horizon	$b_r^{(k)}$	$b_{\Delta d}^{(k)}$	$\rho^k b_{dp}^{(k)}$
Direct regression , $k = 15$	1.01	-0.11	-0.11
Implied by VAR, $k = 15$	1.05	0.27	0.22
VAR, $k = \infty$	1.35	0.35	0.00

As you can see,

- *Return forecasts – time-varying discount rates – explain virtually all the variance of market dividend yields, and dividend growth forecasts or "bubbles" – prices that keep rising forever – explain essentially none of the variance of price.*

This is a *huge* change in viewpoint from the classic efficient markets/constant returns view, circa 1980. We used to think that expected returns are constant; stocks are a random walk; there is no "good time" to invest or "bad time". Now, of course, prices move around. Isn't a low P/D a good "buying opportunity?" No, we would have said, low P/D happens when people expect declines in dividend growth. Variation in P/D occurs entirely because of cashflow news.

What we see in these results is *exactly* the opposite. Now we think that market P/D variance corresponds 100% to expected return news, and none at all to cashflow news. (Prices decline when *current* dividends decline of course.) (Things get even stronger when we add more variables; technically this result refers to forecasts using only P/D.)

In this sense, our view of the world has changed from 100% / 0 to 0 / 100%. That's another sense in which return forecastability is a "big deal" far beyond what the 2.4 t statistic and 7% R^2 might suggest.

The long-run forecasts give us a very nice sense of units. We have an identity,

$$1 = b_r^{lr} - b_d^{lr}$$

The dividend growth coefficient "should" be one – if expected returns are constant, high prices mean higher dividend growth, one for one. In fact the return coefficient is one. One and zero are good numbers to understand!

An important reminder of the basic fact.

You see P/D is high, and the salesman says "don't worry about buying apparently overpriced stocks. The market sees great growth ahead" . Can you say "No, that's all discount rates? No, alas. All you can say is, "in the past, on average, times with high P/D like this have been followed by low returns. Yes, some were followed by higher dividends, but some were followed by lower dividends." It's still perfectly possible that *this* time is different.

2.13.2 Table II construction

Now, the table in detail. We want to estimate the long-run regression coefficients in (12). But we need returns going to the infinite future to form the left hand variable. So what do we do?

The first row attacks this problem by forming only 15 years of returns. It shows the regressions

$$\begin{aligned}\sum_{j=1}^{k}\rho^{j-1}r_{t+j} &= a_r^{(k)}+b_r^{(k)}dp_t+\varepsilon_{t+k}^{r}\\ \sum_{j=1}^{k}\rho^{j-1}\Delta d_{t+j} &= a_d^{(k)}+b_d^{(k)}dp_t+\varepsilon_{t+k}^{d}\\ dp_{t+k} &= a_{dp}^{(k)}+b_{dp}^{(k)}dp_t+\varepsilon_{t+k}^{dp}\end{aligned}$$

with $k=15$. At each year t, I form the long-run return for the next 15 years.

Clearly, something is left over when you stop at any finite horizon, and it would be nice to tie this all up with an identity. To see that, the last column of the first row includes the 15 year regression of dividend yields on themselves.

$$dp_{t+k}=a_{dp}^{(k)}+b_{dp}^{(k)}dp_t+\varepsilon_{t+k}^{dp}$$

Notice that all three coefficients add up to one. The result is,

$$1\approx b_r^{(k)}-b_d^{(k)}+\rho^k b_{dp}^{(k)}.$$

Thus, at a finite horizon, it seems that mechanically all variation in dividend yields must correspond to forecastable returns, forecatsable dividend growth or forecastable future dividend yields.

Why do the coefficients add up so nicely? If you iterate the return identity (10) forward only k times instead of an infinite number of times, in place of (9) you get

$$dp_t\approx\sum_{j=1}^{k}\rho^{j-1}r_{t+j}-\sum_{j=1}^{k}\rho^{j-1}\Delta d_{t+j}+\rho^k dp_{t+k} \tag{16}$$

If you rearrange identity (16) as

$$\sum_{j=1}^{k}\rho^{j-1}r_{t+j}\approx\sum_{j=1}^{k}\rho^{j-1}\Delta d_{t+j}-\rho^k dp_{t+k}+dp_t$$

it basically says that long-run returns come from good dividends, a high final price, or a low initial price.

Now we can regress both sides of (16) on dp_t and derive the finite version of our return identity,

$$1\approx b_r^{(k)}-b_d^{(k)}+\rho^k b_{dp}^{(k)}. \tag{17}$$

This becomes (**??**) as $k\to\infty$. The point is, a high dp ratio today *must* mean either high returns in the future, low dividend growth in the future, or an even higher (growing at the rate $(1/\rho)^k$) dividend yields.

So now, the first row of Table II becomes clear: It's not just an approximation: *variation in dividend yields corresponds almost entirely to variation in subsequent 15 year returns, not to subsequent 15 year dividend growth, or dividend yields that are still high 15 years later ("bubbles").*

If you do things at shorter horizons, you will find that variation in *one week* returns is almost all due to final prices, one-week forecasts of future dp ratios, the last term. As the horizon lengthens, the return forecasts take over. Long horizons really are different, a deep and under appreciated point.

The second and third rows of Table II attack the same problem (you can't run regressions of returns 50, 100 or 10000 years out) a different way. Above I showed you how short and long run regression coefficients were linked. Let's use that to infer long run regressions from one year regressions. If we write (ignoring the constant term which doesn't matter)

$$\begin{aligned}r_{t+1} &= b_r dp_t+\varepsilon_{t+1}^{r}\\ dp_{t+1} &= \phi dp_t+\delta_{t+1}\end{aligned}$$

then look back at (4)we saw above that

$$\begin{aligned}
r_{t+2} &= b_r\phi \\
r_{t+3} &= b_r\phi^2 \\
&... \\
r_{t+j} &= b_r\phi^{j-1}
\end{aligned}$$

and thus

$$\begin{aligned}
\sum_{j=1}^{\infty}\rho^{j-1}r_{t+j} &= b_r(1+\rho\phi+\rho^2\phi^2+..)dp_t+(\text{error}) \\
\sum_{j=1}^{\infty}\rho^{j-1}r_{t+j} &= \frac{b_r}{1-\rho\phi}dp_t+(\text{error})
\end{aligned}$$

So, the third row just reports

$$\begin{aligned}
b_r^{(\infty)} &= b_r/(1-\rho\phi) \\
b_d^{(\infty)} &= b_d/(1-\rho\phi)
\end{aligned}$$

The same idea shows

$$\begin{aligned}
dp_{t+2} &= \phi^2 dp_t+(\text{error}) \\
dp_{t+3} &= \phi^3 dp_t+(\text{error}) \\
&...
\end{aligned}$$

so, since $\phi < 1/\rho$ the infinite coefficient is zero.

The middle row shows that the 15 year coefficient calculated this way gives about the same result as the 15 year coefficient calculated directly. Following the same logic,

$$\sum_{j=1}^{15}\rho^{j-1}r_{t+j}=\frac{b_r\left[1-(\rho\phi)^{15}\right]}{1-\rho\phi}dp_t+(error)$$

2.14 Volatility, volatility tests, irrational exuberance and bubbles

What we have accomplished is more important than you think. These results encapsulate *all* of the evidence on "bubbles" and "irrational exuberance" vs. "rational markets, " and the famous "volatility tests."

Here's the background. Price volatility seems puzzling. Stock prices are enormously volatile. Even in normal times, stocks move 20% up or down in a typical year, and moves of 1% in a day are common.

Isn't this by itself evidence against efficiency? What happened on Oct. 1987? What happened yesterday? What happened in Fall 2008? I can't tell you how many reporters called me in 2008-2009 to ask "doesn't the stock market crash prove that markets are inefficient?" Of course, I answer: "we promised you efficient markets, not clairvoyant markets. Efficient markets are supposed to be unpredictable and crash occasionally. If markets went up steadily and gave return without risk, now *that* would be an inefficiency." But this isn't really convincing, is. it? Yes, stock prices aren't *predictable*, but surely "unwarranted volatility" is damning evidence against something like "efficient markets?" Aren't people who look at predictability just missing the elephant in the room – the *quantity* of the US capital stock surely doesn't change 20% every year. That's like California sinking in to the ocean. Surely someone would have written about it. How can the *value* of the US capital stock change so much?

Yes, when we *do* see some information it's always quickly incorporated in prices. (Except private/inside information, which is the exception that proves the rule, and that there is some content to the theory since it can be proved wrong. Private/inside information is by and large not reflected in market prices, and thus markets are "inefficient" in this regard.) There is no "efficiency" rejection of this sort: "The government announced the new employment report, and stock prices rose slowly over the following week as traders built this news into their models."

The puzzle is that prices *also* move when we *don't see* information. As always, the puzzles in finance are neatly where they don't provide clean evidence to either side.

We seem stuck in a sterile debate: Behaviorist: "Look, prices are moving all over the place, and you have no idea why" Rationalist: "Well, traders must have all sorts of information that we don't see and doesn't get reported in the papers. That's what markets are for, to summarize vast amounts of information from all over the economy. Go read Hayek. If I could tell why the price of coffee goes up or down, communism would have worked." Behaviorist: "Come on now, there can't be *that* much information out there. If California fell in to the ocean, somebody would notice." Rationalist: "why do you think you're so smart you know everything?"

As you can see, we're getting nowhere, as all debates do that are not based on facts. But our present value formulas and predictability regressions *do* let us get somewhere in this debate. Those investors have to be right *on average.* If prices are moving up on news of higher future dividends, *on average* dividends must be higher afterwards. If prices move on dividend news, prices must forecast future dividends. We can't tell what news sparked an *individual* price movement, but we can know what kind of news sparks *typical* price movements. (In case it isn't obvious, in a predictive regression $r_{t+1} = a + bx_t + \varepsilon_{t+1}$ the coefficients $a + bx_t$ tell you what returns are "on average" after seeing a value x_t.)

This is remarkable. It didn't look like we could ever get anywhere debating about what investors might see that we can't see. Finally the rational/irrational debate can be about some facts!

Of course, as you know, this prediction is a disaster. When we look at dividend growth after high prices, there is just no tendency whatsoever for high prices to be followed by higher dividend growth. This, in a nutshell was the outcome of the "volatility tests," starting with Robert Shiller's work made famous by "Irrational exuberance."

Of course, there is a retort: If price/dividend ratios aren't moving on expected dividend growth, maybe the discount rate is changing? If the discount rate or "required return" rises, people see stocks as a bad deal. They try to sell, but of course we can't all sell, so prices must go down until the same dividend stream offers a better return. On average, after times like that, the returns will in fact be better. Low prices will then forecast higher returns instead.

We have seen, in fact, that 100% of the volatility of price-dividend ratios accounted for by changing expected returns.

You also see that, far from being something different and new,

- *"Excess" price volatility (price-dividend movement with no dividend news) is exactly the same thing as return forecastability.*

Note in (14) "variance of dp" on the left hand side and return-forecast coefficients on the right hand side. "Volatility", it turns out, is *not* a new and different test of efficiency, it's just a good old fashioned random walk test. Since it's about long run forecasts that move slowly over time, it's particularly vulnerable to the "joint hypothesis" problem that expected returns might vary for perfectly rational reasons.

Oh well. We don't "disprove efficiency" but we do learn that the huge price-dividend movement does correspond to huge expected return movement, something we didn't recognize in the 70s.

None of this was as simple at the time as it looks now. It took at least 15 years to get from the first volatility tests to the realization that in fact volatility and return forecasts are the same thing.

2.15 Bubbles

So why are people still arguing about "bubbles," "irrational exuberance," and all the rest, as if these were established facts? *You're looking at the entire sum of evidence* on these topics so you're now as well informed as anyone to join the debate.

It all depends of course on what you mean by "bubble," and bubble proponents are often frustratingly unwilling to define what they mean by the word. ("I wish I had sold before the crash" doesn't count.)

A lot of "bubble" proponents simply don't believe that expected returns should vary over time as much as they do. They implicitly assume that expected returns should be constant, so price movements that correspond to subsequent long-term returns – high prices lead to low returns, low prices lead to high returns – are ipso-facto bubbles, which must be explained by "irrational" investors.

Their *definition* of "bubble" is expected returns that vary over time. They say that yes, we the great runners of regressions understand that high prices mean low returns. But the poor deluded investors think high prices mean higher dividends. They're not buying stocks at high prices because they're really tolerant to risk right now, they're buying stocks at constant risk premiums but deluded expectations.

I think this is silly. At the peaks of booms, surely investors are willing to take risks despite mediocre prospective returns (the "hunt for yield" we read about in the financial press). In the depths of recessions, when every investor's job and business might evaporate the next day, rational investors are surely sensible in being unwilling to take stock market risks without a substantial extra return (low price) to do so.

The question must be quantitative: *How much* should risk premiums vary over time, and *when?* Give us an economic model of time-varying risk premiums that tracks the price-dividend ratio.

As you can imagine, that's pretty hard. We have some good ones (I'll discuss the Campbell-Cochrane model later), but they aren't perfect. Critics can say "see your model isn't perfect, that means markets are irrational after all." A model should take non-financial inputs (consumption, GPD, unemployment) and predict what the $p_t - d_t$ ratio is at each moment in time. That's what our habit model does, and of course it gets it wrong sometimes, though showing some promising big patterns.

Needless to say, there is not yet a *psychological* model that even attempts this sort of explanation. For the moment, psychology is at the state of naming the residual: "Markets up? It must be a wave of irrational exuberance. Markets down? It must be a wave of contagious herd pessimism." You might as well say "the Gods are angry."

As you can see, we seem to be back to the same sort of pointless argument. But it is a very different pointless argument. The *facts* are simple: high prices correspond to low returns, low prices correspond to high returns. Times of high and low expected returns are suggestively correlated to business cycles; low prices and high risk premiums come in poor macroeconomic times. The argument is now "what is the source of the slow-moving time-varying risk premium?" Are people "rationally" willing to hold risks in booms and not in recessions, or are they "irrationally" overlooking good buy and sell opportunities? But this need not be a pointless argument. You *can* answer questions like this, and there is *only* one way to answer questions like this: construct explicit views, models, theories of the risk premium, models that make predictions, that tie risk premiums together to other price or quantity data, models that can be rejected, not ex-post "explanations."

2.15.1 Rational Bubbles

There is another definition of "bubble" that people refer to occasionally: Perhaps prices move simply on expectations of further price rises, not grounded in "fundamentals." A Ponzi scheme is a good example. This is a different animal, and one we can, in fact test for.

For example, suppose there are no dividends at all, and expected returns are constant. The price should be zero. Each period prices must obey

$$\begin{aligned} E_t(R_{t+1}) &= E_t\left(\frac{P_{t+1}}{P_t}\right) = \frac{1}{R} \\ P_t &= \frac{1}{R}E_t\left(P_{t+1}\right) \\ P_{t+1} &= RP_t + \varepsilon_{t+1}. \end{aligned}$$

So the condition that expected returns are equal to R only tells us that prices should follow a slightly growing random walk.

To conclude from this observation that the price should be zero, people need to look several steps down this game, and understand that prices can't grow at R forever. If they do, the mechanics of exponential growth means that the asset will soon be worth more dollars than there are atoms in the universe. The bubble will have to burst.

Now, if you know the bubble will burst at some point in the future, the price must be zero now. Intuitively, you might decide "I think the bubble will go on another year, then I'll sell." But if you're rational, you should think "who will buy from me? Someone who thinks the bubble will go on for a year after that." And "who does he think he will sell to?" And on and on. Once you know the bubble must burst for sure eventually, this logic unwinds. (Technically, economic models rule out these paths with the "transversality condition" which is a condition for optimal behavior by investors.)

Advocates of this kind of "bubble" think that people are just a little bit irrational; they're willing to bet that they can be the smart ones to get out in time though they know the bubble must burst eventually. It's not an implausible idea.

Now, let's do it right. It turns out this is not a pointless argument, but there are *facts* to settle it as well! The finitely-iterated version of our present value identity and forecast coefficient identity (16) and (17) gives us the tools to do it.

Look again at (16), which I'll write as

$$p_t - d_t = \sum_{j=1}^{k} \rho^{j-1} \left(\Delta d_{t+j} - r_{t+j}\right) + \rho^k (p_{t+k} - d_{t+k}).$$

This formula shows us that today's price is equal to k-year discounted dividends, plus the expected price k years from now. So, this version of "bubble" says that prices (price-dividend ratio really) can be high, even with no news about dividend growth Δd or returns r if future price-dividend ratios are even higher.

Now, to take the limit as $k \to \infty$, we needed to be assured that the last term goes to zero. Again, by rational economic logic it must. $\rho = 0.96 < 1$, so to keep the last term away from zero, people must expect price-dividend ratios to explode at a rate greater than 1.04^k. Surely there is *some* upper bound – the price/dividend ratio, usually about 20, can't get bigger than 20 million. Once you understand that, the bubble pops and the last term must go to zero.

But suppose not – suppose people are only a little bit irrational, so that they think maybe 100 years ahead, but not 100,000. Then maybe prices vary, not because people expect dividends or returns, but because they're riding the bubble and expectations of the last term are changing through time.

We can test for this kind of bubble, and it turns out that Table II already has the information we need. Regressing both sides of the k period present value identity (16) on dp_t, we obtained the coefficient identity (17) which I'll repeat here

$$1 \approx b_r^{(k)} - b_d^{(k)} + \rho^k b_{pd}^{(k)}.$$

So, in fact there is another possibility at least for finite horizons:

- Dividend yields can vary if they forecast returns, $b_r^{(k)}$ dividend growth $b_d^{(k)}$ *or future dividend yields* $\rho^k b_{pd}^{(k)}$

The sum of all *three* terms must, mechanically, be one with no assumptions at all.

Now, we know that for short-run returns, the answer is almost all going to be the *last,* future price term. If dividend yields are high now, it's almost sure they will still be high in 10 minutes. Almost all daily, monthly or even annual returns come from changes in prices too. But as we go longer and longer, we expect that high prices correspond to many years of dividends or discount rates, not to ever-higher prices.

The "Discount Rates" Table II above boils it all down to the last column. If you look out 15 years, prices today correspond to 15 years of returns, and not 15 years of dividends, and not prices that are still higher 15 years from now.

This is very comforting. We could have seen that prices now correspond to prices that are higher 15 years from now, not to dividends or returns earned in the meantime. We do not. That's a fact, which rather dooms the "rational bubble" view.

- *Price-dividend ratio volatility is fully accounted for by its ability to forecast returns and dividend growth. The data say the size of any "rational bubble" is zero.*

This view of bubbles actually makes a prediction that can be falsified in the data, and that prediction is false. Many people still talk about prices rising on expectations of future prices, but they are ignoring the facts.

So, we've made some progress. The only sensible kind of bubble to discuss is the question whether time-varying expected returns correspond to real risks or waves of over-and under-enthusisasm about dividend growth.

2.16 Vector autoregression and impulse response function

Another tiny investment in technique will pay off handsomely for understanding predictability and what it means. (It's also another very useful tool in your rapidly-expanding kit.) I already used these ideas to compute long-run return regressions implied by one-period regressions.

I've been telling stories of the sort, "expected returns rise, this drives prices down, then on average we see the returns that traders expected." This technique gives us a quantitative, data-driven counterpart to these stories. That can answer questions like, "how long do changes in expected returns last?"

Example

The basic idea is really simple. Suppose you model a series as an autoregression AR(1), as we might do for dividend yields $x_t = d_t - p_t$. Ignoring the constant,

$$x_t = \phi x_{t-1} + \varepsilon_t \tag{18}$$

Now, start this up at $x_0 = 0$, and hit it with a shock $\varepsilon_1 = 1$. Set all the future $\varepsilon_2 = \varepsilon_3 = ... = 0$, and simulate forward

$$\begin{aligned} x_1 &= 1 \\ x_2 &= \phi x_1 + 0 = \phi \\ x_3 &= \phi x_2 = \phi^2 \\ x_4 &= \phi x_3 = \phi^3 \end{aligned}$$

and so forth. etc. This is pretty easy to do:

```
x = zeros(10,1);
x(1) = 1;
for i = 2:10;
    x(i) = b*x(i-1);
end;
plot(x);
```

In graphical form, it is shown in Figure 7 using $\phi = 0.95$

Interpretation

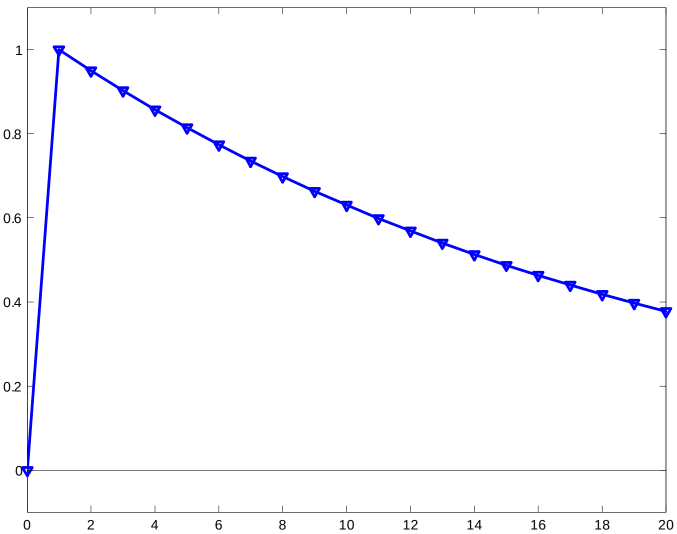


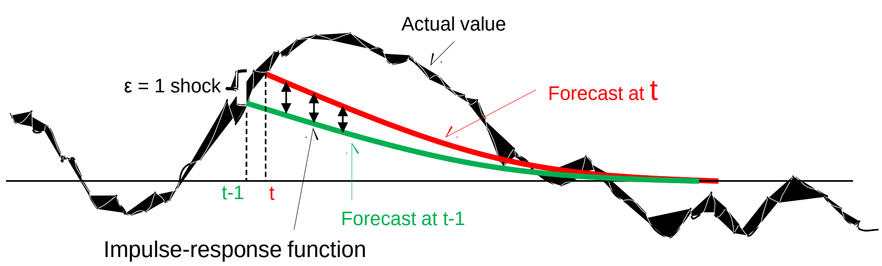
Figure 7: Impulse-response function of an AR(1).

This graph gives you a good feeling about what the dynamics of an AR(1) look like.

Since in reality $E_t \varepsilon_{t+j} = 0$, by putting the future shocks at their expected values (0), we are plotting the *expected* path of x_t, $t = 1, 2, 3, ...$ following a shock $\varepsilon_1 = 1$ and if $x_0 = 0$.

The model is linear, which means it has another interpretation as well:

- The impulse-response calculation answers the question: *How does the expected future path of* x_t **change** *if we see an unexpected shock at time 1?*



The graph ought to make this clear. Mathematically,

$$\begin{aligned} E_{t-1}(x_{t+j}) &= \phi^{j+1} x_{t-1} \\ E_t(x_{t+j}) &= \phi^j x_t \\ E_t(x_{t+j}) - E_{t-1}(x_{t+j}) &= \phi^j (x_t - \phi x_{t-1}) = \phi^j \varepsilon_t \end{aligned}$$

Sometimes we write the last line using $(E_t - E_{t-1})$ to denote the "surprise" ("innovation") in x_{t+j}

$$(E_t - E_{t-1})(x_{t+j}) = \phi^j \varepsilon_t$$

but maybe the graph is clearer (!)

Of course, linking how prices change today (ε) to how expectations of the future have changed is exactly what we're interested in doing here.

Warning: the impulse-response function does not necessarily mean that the shock ε_t *causes* the response x_{t+j}. People often infer some causality when looking at impulse response function. The whole language of "impulse-response function" kind of sets you up for that. Chris Sims won the Nobel prize basically for popularizing the use of impulse response functions to look at the dynamics of macroeconomic models. There, there is a much greater tendency to think, for example, of the response of output (x_{t+j}) to a monetary policy shock (ε_t) as measuring cause and effect.

For us, causality is likely to go backwards. If ε is a price shock and x is the subsequent dividends or returns, we are likely to think of the response function as measuring the change in investor's expectations of future x which caused the price shock ε in the first place! Mathematically, the impulse-response function just tells us "when you see ε_t how much does $E_t x_{t+j}$ change?" That is completely a correlation, not a causation, issue.

For our case

Now, what we want to understand is, how do shocks today change (or reflect changes in) expected dividend growth, returns, etc. in the future? So we need to simulate forward the *system* of forecasts we've been running

$$\begin{aligned} r_{t+1} &= b_r dp_t + \varepsilon^r_{t+1} \\ \Delta d_{t+1} &= b_d dp_t + \varepsilon^d_{t+1} \\ dp_{t+1} &= \phi dp_t + \varepsilon^{dp}_{t+1} \end{aligned} \tag{19}$$

If you like matrices you can see that we can write this simply as

$$\begin{bmatrix} r_{t+1} \\ \Delta d_{t+1} \\ d_{t+1} - p_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & b_r \\ 0 & 0 & b_d \\ 0 & 0 & \phi \end{bmatrix} \begin{bmatrix} r_t \\ \Delta d_t \\ d_t - p_t \end{bmatrix} + \begin{bmatrix} \varepsilon^r_{t+1} \\ \varepsilon^d_{t+1} \\ \varepsilon^{dp}_{t+1} \end{bmatrix}$$

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \boldsymbol{\varepsilon}_{t+1} \tag{20}$$

Now, you can see that (20) is exactly the same thing as (18), so the simulation works exactly the same way. There are three shocks to watch of course, but conceptually we're doing nothing different. That's why we call it a "vector autoregression." (Viewed this way, you might also ask why the zeros. In fact, you can add lagged returns and dividend growth to the forecasts. Most vector-autoregressions estimate all the zero terms too. They don't make much difference which is why I left them out here.)

You can do everything we did with the simple AR(1) with this vector version. For example

$$\begin{aligned} E_t \mathbf{x}_{t+1} &= \mathbf{A}\mathbf{x}_t \\ E_t \mathbf{x}_{t+2} &= \mathbf{A}^2\mathbf{x}_t \\ E_t \mathbf{x}_{t+j} &= \mathbf{A}^j\mathbf{x}_t \end{aligned}$$

$$(E_t - E_{t-1})(\mathbf{x}_{t+j}) = \mathbf{A}^j(\mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1}) = \mathbf{A}^j\boldsymbol{\varepsilon}_t$$

If you don't like matrices, it's simple enough to simply start the system (19) with, say, $\varepsilon^{dp}_1 = 1$, simulate the responses as I simulated the AR(1), and watch the results.

Correlation between the shocks

We have three shocks, not one. You might be tempted to shock each of $\varepsilon^r, \varepsilon^d, \varepsilon^{dp}$ in turn, leaving the others alone, but this makes no sense in our case. For example, you can't shock $\varepsilon^r_1 = 1$, $\varepsilon^d_1 = 0$, $\varepsilon^{dp}_1 = 0$, because that would mean a a return with no change in dividends and no change in prices! Returns *come from* changes in dividends or prices!

Let's get this right. The surprise in returns is

$$\varepsilon^r_{t+1} = r_{t+1} - E_t(r_{t+1}).$$

We have to figure out what other surprise – to dividends or prices – must go along with the surprise in returns.

Go back to the return identity (10), which just expresses the definition $R_{t+1} = (P_{t+1} + D_{t+1})/P_t$, reproduced here

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

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

Taking surprises of both sides of this identity, the surprises must obey

$$\varepsilon^r_{t+1} \approx -\rho\varepsilon^{dp}_{t+1} + \varepsilon^d_{t+1}. \tag{21}$$

(Apply $E_{t+1} - E_t$ to both sides of the return identity. $(E_{t+1} - E_t)(dp_t) = 0$. Once we know something there can be no more surprises.)

So we can really only have *two* shocks, and the third follows along Which two shocks to look at is arbitrary in some sense, but after looking at the results every possible way for a few years (!) I find that the following choice is really pretty.

First, I look at a shock to dividend growth $\varepsilon^d_1 = 1$ with no change in dividend yield, $\varepsilon^{dp}_1 = 0$. If dividend growth rises and dividend yields don't, the returns rise too, $\varepsilon^r_1 = 1$.

Second, I look at a shock to dividend yields $\varepsilon^{dp}_1 = 1$ with no change in dividend growth $\varepsilon^d_1 = 0$. That means a plunge in prices (think about it) which means a big negative actual return $\varepsilon^r_1 = -\rho$. Alternatively, higher dividend yields mean higher expected returns, higher discount rates, and thus a plunge in actual returns on the day the expectation changes. In sum

$$\begin{array}{rl} \text{shock 1 (“}\Delta d\text{ shock”):} & \left[\begin{array}{ccc} \varepsilon^r_1 & \varepsilon^d_1 & \varepsilon^{dp}_1 \end{array}\right] = \left[\begin{array}{ccc} 1 & 1 & 0 \end{array}\right] \\ \text{shock 2 (“}dp\text{ shock” or “}Er\text{” shock)} \quad : & \left[\begin{array}{ccc} \varepsilon^r_1 & \varepsilon^d_1 & \varepsilon^{dp}_1 \end{array}\right] = \left[\begin{array}{ccc} -\rho & 0 & 1 \end{array}\right] \end{array}$$

Now, it's simple enough to simulate the *system* (19) or (20) forward through time, in response to *each* of these shocks, and see where things go.

Asset pricing Figure 20.4 and 20.5 presented one set of impulse-response results. Figure 8 gives an updated set of results with a bit more detail.

Now, what do we see in these beautiful pictures?

First, understand how they are created. Back to basics. These are just impulse-responses, like the simple AR(1) response we programmed up below.

The equations, with rough numbers, are

$$\begin{aligned} \Delta d_{t+1} &= 0.01 \times dp_t + \varepsilon^d_{t+1} \\ r_{t+1} &= 0.1 \times dp_t + \varepsilon^r_{t+1} \\ dp_{t+1} &= 0.94 \times dp_t + \varepsilon^{dp}_{t+1} \end{aligned}$$

$$\text{with } \varepsilon^r_t = \varepsilon^d_t - \rho\varepsilon^{dp}_t$$

So, top left, let's look at a dividend growth shock with no change to the dp ratio. Dividends Δd_1 go up by 1% – that's the shock. Returns r_1 also go up by 1%, following the identity or the simple intuition that if they pay you more dividends you get a return. (The return line is hidden by the dividend line.) There is no change to the dividend yield dp_1 by definition. We're asking, what happens *if* returns rise 1% because dividends rise 1% and there happens to be no change to the dividend yield. Now look at period 2, top left corner, i.e.

$$\begin{aligned} \Delta d_2 &= 0.01 \times dp_1 + 0 \\ r_2 &= 0.1 \times dp_1 + 0 \\ dp_2 &= 0.94 \times dp_1 + 0 \end{aligned} \tag{22}$$

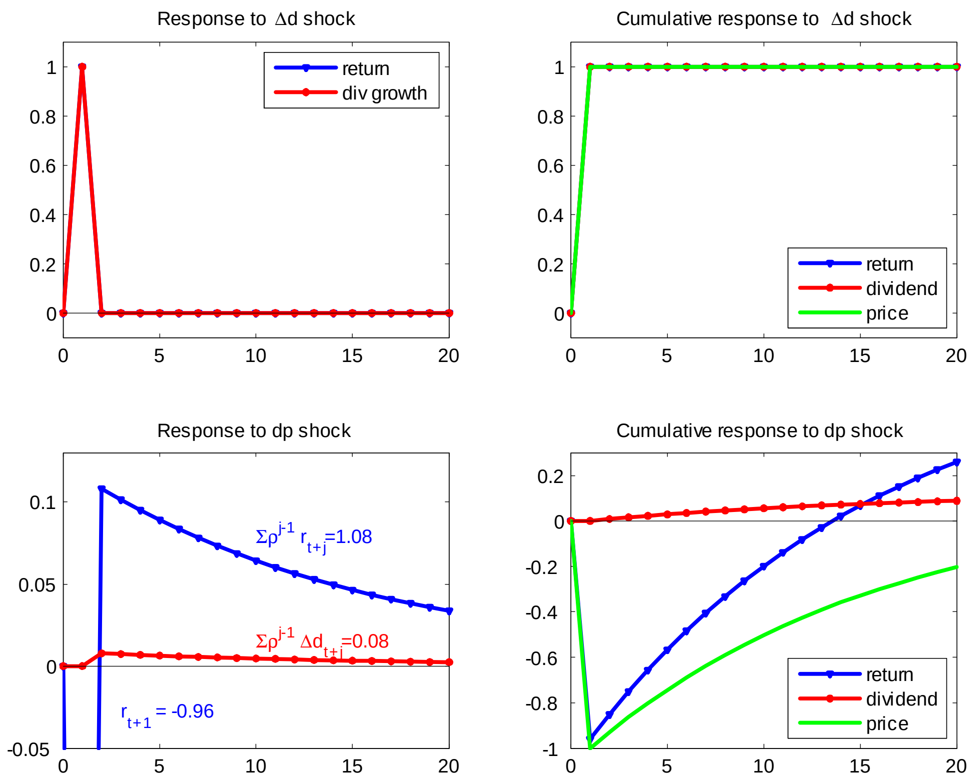


Figure 8: Impulse-response function for simple VAR

Since dp_1 didn't change – it's still zero–nothing happens here. Everything reverts to zero in one period.

Now, top right. Here I made the graphs prettier by plotting the *cumulative* response, i.e. the path of $d_t = \Delta d_1 + \Delta d_2 + ... + \Delta d_t$ following the shocks, the *cumulative return*, the path of $r_1 + r_2 + r_3 + ... + r_t$ following the shock, and where *prices* go, from

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

and then

$$p_t = \Delta p_1 + \Delta p_2 + ... + \Delta p_t.$$

Ok, what does it mean? If the *growth* of dividends follows the one-period blip in the top left graph, then the *level* of dividends goes up and stays up as in the top right graph. Same with cumulative returns and prices.

Next, the much more interesting bottom row. Here, we look what happens if there is a dividend yield shock with no change in dividend growth, $\varepsilon_1^{dp} = 1, \varepsilon_1^d = 0$. Now, if dividends don't change but $d - p$ rises, that means p went down a lot. Hence, this shock is accompanied by a huge contemporaneous return shock $\varepsilon_1^r = -\rho \approx -0.96$. It's so huge it goes off the bottom of the graph (blue line).

On to period $t = 2$. Look again at (22). This time dp_1 *has* changed – it rose 1% – so there will be ripple effects to time 2 and beyond. After the disastrous negative $r_1 = -0.96$, we have a positive $r_2 = 0.1 \times dp_1 = 0.1 \times 1$. Dividend growth goes up just a tiny amount as well. The dividend yield continues forward; it's just like the AR(1) I started with so I didn't plot it here.

The bottom right just cumulates the bottom left. The huge negative return shock brings prices way down. But the subsequent positive expected returns slowly bring cumulative returns and prices back to where they started.

2.16.1 What do the impulse-responses mean? Price formation

Now you understand the responses, i.e. where they come from mechanically given our regressions. What do they *mean*?

Again, don't let the "impulse-response" language confuse you about cause and effect. We're not seeing here the "effects" of the shocks. We are seeing "if you see one of these shocks, how does it change your guess about the future?" There's no cause and effect in that. If anything, I think it's better to reverse cause and effect: "If you see one of these shocks, what do you learn about the change in traders' expectations that caused them to change prices in the first place?" The "response" is the "cause" and the behavior of prices in ε_t^{dp} (or its absence, $\varepsilon_t^{dp} = 0$) is the effect!

Interpreted this way, the "dividend growth shock" – a movement in dividends with no change in dividend yield, and hence with prices that rise 1% exactly with the 1% rise in dividends looks like a *cashflow shock with no change in expected returns:* The rise in current dividends is expected to be permanent. If there is no change in expected returns at the same time, then by present value logic, prices just rise proportionally to dividends and sit there. This is exactly the pattern we see in the "dividend growth shock" responses.

Now, if dividend growth is unpredictable – log dividends are a random walk – and prices change with no change in *current* dividends – our ε_t^{dp} shock – then this means prices are changing with no news about *future* dividends either. The only way prices change without a change in current or expected future dividends is if expected returns (discount rates) change. So, if there were a pure "discount rate shock" – a change in prices with no change in expected dividends – we should see prices fall and then a period of high returns, with no change in dividends. This is what we do see following the ε^{dp} shock – exactly if we set $b_d = 0$ and approximately if we use the sample b_d which is slightly greater than zero. (Taking the sample $b_d \neq 0$, the ε^{dp} shock combines a little bit of dividend growth news with a lot of discount rate news.)

In sum

- *If we take $b_d = 0$, a shock to dp ratios with no current change in dividends isolates a pure discount rate movement, and is followed by a period of high expected returns. A shock to dividend growth Δd with no surprise in dp isolates a pure cashflow shock with no change in discount rates, and is followed by higher cashflows but no change in expected returns*

2.16.2 Isolating shocks

The big innovation in all this is that *you learn a tremendous amount by watching returns, dividend growth, and dividend yields together.*

A high return *with no change in dividends* isolates the event as "discount rate news" shock and is a strong signal about future returns. A high return *with no change in dividend yield* isolates the event as "cashflow news" with no implication at all about future returns.

Furthermore, look at the "price" response to the dividend yield shock.

- *Return predictability means that there is a "temporary component" to stock prices. If you see prices rise with no change in current dividends, it means expected returns are temporarily low, and this price movement will slowly but completely melt away.*

Price rises all by themselves combine "discount rate" and "cashflow" news. Now we can separate a price rise that corresponds to "cashflow news" (one that comes with a change in dividend growth but no change in dividend yield) from a price rise that corresponds to "discount rate news". We know that the latter type of price rise (or the latter component of any price rise) will eventually melt away.

To see why this is all so dramatic, let's look at the news implied by a high return all by itself – when you don't look whether dividends increased or not. You might think you'd see some of the same mean-reversion. You'd be wrong. A high *return* all by itself combines (on average) some cashflow news and some discount rate news, so is a very poor forecaster of future returns.

To see what happens, let's compute the response to a return shock taken all alone. I just ran the return autoregression $r_{t+1} = a + br_t + \varepsilon_{t+1}$, I got $b = 0.10$ and Figure 9 gives the response.

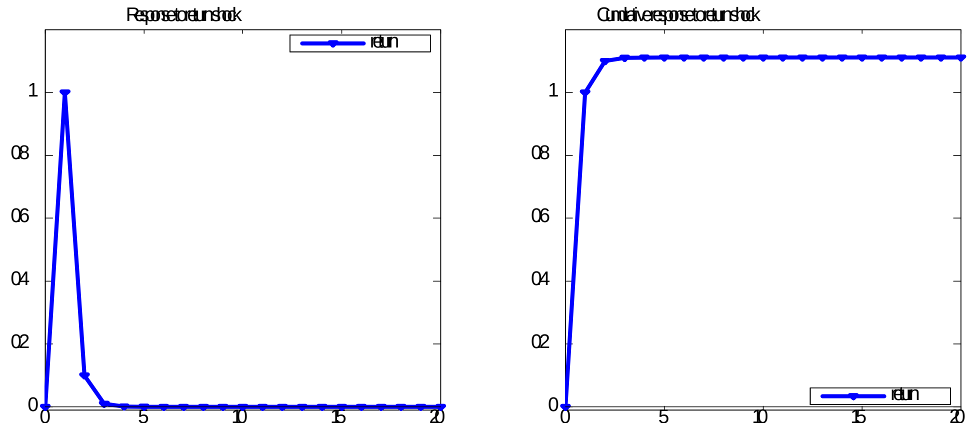


Figure 9: Response to return autoregression, $r_{t+1} = a + br_t + \varepsilon_{t+1}$

A good return by itself is a very poor signal of anything, since it's equally likely to come from "discount rate" or "cashflow" news.

The key ingredient in all this is that dividends are nearly a random walk. (When you forecast them with $[\Delta d_t, dp_t, r_t]$. Things get a lot more fun when we add more variables. "Discount rates" talks about this a lot) Thus, if prices move without *current* dividends changing, that means prices are moving without *future* dividends changing too.

If nothing else, this exploration should sell you on the power of the VAR technique. It lets you quantitatively start from a return regression and isolate the "temporary component" of prices, or the "target price" to which actual prices will revert, in a quantitative way!

(How do you add up one shock to returns that gives no response (dividend growth), another shock to returns that leads to a mean-reverting response (dp, expected return), and get a return-only shock with a positive response? This seems impossible. It isn't – the return shock is not just a sum of dividend growth and expected return shocks. The residual in the regression of returns on lagged returns is not the sum of the VAR shocks. You lose more than shocks adding up when you delete variables from a regression. You can, in fact, *derive* the return impulse response function from the VAR and show that it's almost exactly flat. It takes more algebra than I want to put you through right now, and gives basically the same answer as this graph.)

2.17 Pervasive predictability: a preview

Just to remind you, our pattern was

$$\begin{aligned} r_{t+1} &= a + 0.1 \times (d_t - p_t) + \varepsilon_{t+1} \\ \Delta d_{t+1} &= a + 0 \times (d_t - p_t) + \varepsilon_{t+1} \end{aligned}$$

or, in levels (rather than logs)

$$\begin{aligned} R_{t+1} &= a + 4 \times (D_t/P_t) + \varepsilon_{t+1} \\ D_{t+1}/D_t &= a + 0 \times (D_t/P_t) + \varepsilon_{t+1} \end{aligned}$$

We expected the opposite pattern.

A better way of putting it is in terms of long-run returns and dividend growth, since here we had an identity that the coefficients should add up to one.

$$
\begin{aligned}
r_{t+1}^{lr} &= E_t \sum_{j=1}^{\infty} \rho^j r_{t+j} = a + 1 \times (d_t - p_t) + \varepsilon_{t+1} \\
\Delta d_{t+1}^{lr} &= E_t \sum_{j=1}^{\infty} \rho^j \Delta d_{t+j} = a + 0 \times (d_t - p_t) + \varepsilon_{t+1}
\end{aligned}
$$

We expected the opposite pattern of coefficients, 0 and 1.

As a reminder, there is nothing super-special about D/P. This is just a simple precise example to look at. People use all sorts of additional variables to forecast returns. And the patterns we have seen in the stock market as a whole turn out to apply all over the place.

1. *More variables.* I emphasized D/P (or E/P, or BE/ME) as this is the most important forecasting variable. However, in practice people use lots of additional variables and multiple regressions to forecast market returns over time. The yield spread between long maturity and short maturity bonds, the spread between BAA and AAA bonds, macro variables such as investment/capital ratio, consumption/wealth ratio, and inflation, and measures of real and implied volatility are popular examples to get you thinking about these forecasts.

   $$R_{t+1} + a + b(D/P)_t + c \times \text{term}_t + d \times \text{def}_t + f \times \text{I/K}_t + g \times \text{cay}_t + h \times \pi_t + \text{volatility}_t + m \times \text{VIX}_t + n \times \text{vol}_t ... + \varepsilon_{t+1}$$

   High frequency trading strategies emphasize volatility and volume indicators. Basically, high frequency trading is built on regressions of this sort.

   Now, we have to ask how does all our variance decomposition and present value identity work when there are multiple regressions? Remember, we looked at the present value identity,

   $$dp_t \approx E_t \sum_{j=1}^{\infty} \rho^{j-1} \left(r_{t+j} - \Delta d_{t+j} \right)$$

   regressing both sides on dp_t we deduced that the long-run return forecast and the long-run dividend growth forecast coefficients had to add up to one. And the return forecast was "just enough" (1) to explain all variance in dp.

   What happens if we add another forecasting variable x_t? Well, dp_t is what it is. So *any variable x_t that forecasts long-run returns in addition to dp_t must also forecast long-run dividend growth.* This actually makes a lot of sense. In the bottom of a recession, as growth starts, expected dividend growth is high. But the risk premium is high too, so expected returns are high. These two things offset, so there is no effect on stock prices. *We expect recession indicators to help to forecast* **both** *dividend growth and returns.* Furthermore, a variable such as VIX may be able to forecast short-run returns r_{t+1} if it forecasts longer run returns r_{t+j} to go in the other direction.

   The upshot is that additional variables can make dividend growth predictable, which it surely is and which our regressions did not see, and to make returns even more predictable than they are! "Discount rates" and the first section of the Appendix below show you how this works for one popular variable, the consumption/wealth ratio.

2. *Individual stocks.* What about individual stock returns,

   $$R_{t+1}^i = a^i + b^i x_t^i + \varepsilon_{t+1}^i?$$

   This is perhaps a more interesting regression for practical purposes, since it leads you to invest in one stock vs. another, which is what half of you will do for a living. The answer is a resounding yes, returns can be forecast. We will do value, and size effects next week.

Though not presented as such, the evidence as presented in Fama/French is exactly the same as forecasting regressions. The value effect states that stocks with high book to market ratio have higher subsequent returns on average.

$$E(R^i_{t+1}) \text{ increases with } BE/ME^i_t$$

where BE/ME = book equity/market equity the Fama-French measure of value. (Actually, you should take log of the right hand side here) Though tradition states this fact as average returns in portfolios sorted by B/M ratios, this is the same evidence as a regression

$$R^i_{t+1} = a + b \times BE/ME^i_t + \varepsilon^i_{t+1}.$$

All trading strategies amount to the same thing, e.g. momentum, reversal, neural networks, etc. They amount to forecasting regressions.

3. *Bonds.* Term spreads forecast returns, not interest rates. The "expectations hypothesis" states that a rising yield curve should signal higher future interest rates, not a chance to earn money on long term bonds. In fact, the exact reverse is true. We'll forecast returns and interest rate changes, just as we forecast returns and dividend growth, and we'll find roughly speaking

$$\begin{aligned} R^{bond}_{t+1} - R^f_t &= a + 1.0(y^{long}_t - y^{short}_t) + \varepsilon_{t+1} \\ R^f_{t+1} - R^f_t &= a^f + 0.0(y^{long}_t - y^{short}_t) + \varepsilon^f_{t+1} \end{aligned}$$

We'll find that at a one year horizon, the term spread is 100% a sign of higher risk premiums in long term bonds and 0% a sign of interest rate rises. The expectations hypothesis is exactly wrong. Once again, the term spread is a classic business cycle indicator. (For a preview, see *Asset Pricing* Table 20.9)

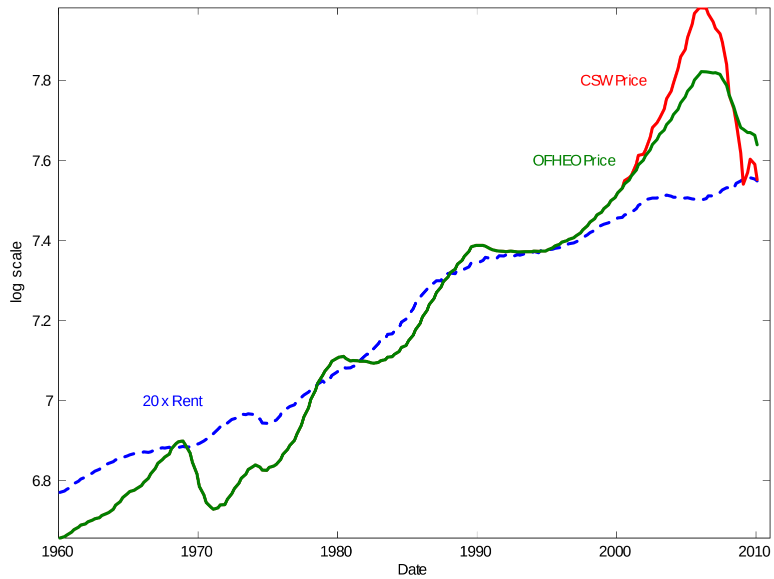
4. *Foreign exchange.* Exchange rates forecast returns, not depreciation. If Euro interest rates are 5%, dollar interest rates are 1%, we "should" see the Euro depreciate on average 4% next year. We'll run regressions and once again find exactly the same (actually worse) pattern: the Euro seems to appreciate on average, and you earn a risk premium by buying Euros.

$$\begin{aligned} \text{excess return} &= a + 1 \times (r^{Eu}_t - r^{\$}_t) + \varepsilon_{t+1} \\ \Delta e^{Eu/US}_{t+1} &= a + 0 \times (r^{Eu}_t - r^{\$}_t) + \varepsilon_{t+1} \end{aligned}$$

Excess return = the return to holding Euro bonds for a year, unhedged, financed at the US rate, i.e. $r^{Eu} - r^{\$} - \Delta e^{Eu/\$}$, r = interest rate, e = exchange rate. (Actually this number is >1 and <0. See *Asset Pricing* Table 20.11 for a preview.)

5. *Houses.* A high price/rent ratio does not mean rents will rise in the future. It means what I will charitably call "low expected returns to housing." Amazingly the coefficients in a price-rent regression

look almost exactly the same as those in the stock regression! (This is from "Discount Rates")



Houses:	b	t	R^2	Stocks:	b	t	R^2
r_{t+1}	0.12	(2.52)	0.15		0.13	(2.61)	0.10
Δd_{t+1}	0.03	(2.22)	0.07		0.04	(0.92)	0.02
dp_{t+1}	0.90	(16.2)	0.90		0.94	(23.8)	0.91

6. *Corporate bonds.* When corporate bond yields are higher than Treasuries, that shouldn't mean greater return, it should simply reflect greater default probability. Instead a huge portion of corporate spreads reflect expected returns not default probabilities. December 2008 was the buying opportunity of a lifetime for BAA bonds.

It's amazing to see the same phenomenon show up all over the place. We don't know yet: are there important cross effects? Does stock DP forecast bond returns? Etc. More importantly, do all these things happen at the same time? Are there one or two fundamental indicators of a global market time-varying risk premium that is common to all these securities? Or to stock and bond risk premiums go their own way? In finance lingo "what is the factor structure of expected returns?"

2.18 Warnings and interpretation

2.18.1 Statistics

The statistics here are very slippery.

Fishing. Hundreds and hundreds of papers on stock market predictability have been written. There's a lot of fishing in these regressions, and more in the forecast models used by hedge funds. 1000 assistant professors x 5 regressions per day x 20 years, we are bound to find a few t statistics greater than 2. Add 10,000 hedge fund quants x 1000 regressions per day and the problem gets worse!

That's why some sort of economic story is important. For academics it is one small way of overcoming overly fished regressions. Good hedge funds don't trade on pure statistical relationships, either, they want to see some sensible economic story why the pattern will persist.

Peso problems. Many statistical relationships hide a small chance of a huge crash not seen in sample. This is in some sense the Lesson Of The Decade. The historical returns of subprime MBS looked great in 2006.

Slow movers. The statistics of forecasting regressions with very slow moving right hand variables are subtle. Conventional t statistics can be quite wrong. "The dog that didn't bark" (reading list) is my bottom line: I think D/P forecastability really is there, though it may not be as useful in forming portfolios as the regressions seem to indicate. But conventional t statistics *are* wrong, and you need to be careful.

Sampling uncertainty. Before you go off making forecasts based on D/P, remember that the size of the coefficients is quite uncertain. So, if you properly include estimation uncertainty in your forecasts, you'll find we're still quite uncertain about how much expected returns really do vary over time.

Structural shifts. So, D/P is still low by historical standards. Does this mean expected returns are low? Or maybe the relation between D/P and returns shifts over time? For example, the shift to repurchases might mean measured dividends aren't a good divisor any more (I look at this in more detail below). I'm dubious about structural shift stories, but when the full sample t statistic is barely above 2, it's hard to find stark evidence that there haven't been any shifts either.

As a result, D/P predictability had been under a pretty heavy academic attack. I had to apologize for years, I kept saying "we'll look like geniuses as soon as the market crashed." Well, maybe not geniuses, but the t statistics look a *lot* better now than they did in 1999 or 2008!

2.18.2 Practical use

Trading

Return predictability of the sort uncovered here is not as immediately useful as you might think. Practically by definition, a signal that changes once a decade at most is not useful for high speed trading! (Return forecasts based on higher frequency signals are by definition more useful – but surprise, surprise, the high-frequency signals are a lot weaker.

Following the D/P or similar signals also requires a contrarian spirit. You would have missed the 90's bull market– but done well in the bust. You had to buy in the middle of the crash, in Winter 2008-2009. You have to be unpopular, be able to sit by and watch your buddies get rich, buy when they are all selling. You have to buy in the middle of depressions and sell in hot booms. We'll study portfolio implications in detail later, but I just want to point out now that this is not a quick and easy trading system.

At a fundamental level, the average investor must hold the market portfolio without even rebalancing. (This is the very important *average investor theorem.* Pay attention.) For everyone who says "DP is high, let's buy," someone else must sell. If we all say "wow, DP is high, let's buy," we just drive the price up until DP isn't high any more. If this regression is the basis of a trading strategy, you must think everyone else is being really dumb. And has been, for 400 years, in every market we've examined.

The other possibility is that the predictability represents a time-varying risk premium. In winter 2008, the average investor can't buy because he is afraid of losing his job. Now it makes sense for you to buy – *if* you are less worried about losing your job than the average investor. Now you are writing insurance, helping your neighbor by buying at his fire sale, which is all well and good. But are you really sure you're not the half who is more worried about his job than average?

If this were a quick and easy trading system, it would be a violation of efficiency. If it is, instead, a new dimension of risk and return, then it takes a lot more thought to decide what if anything you do with it.

Worldview

The statistics are marginal, but the point estimates are really big, and worth thinking about. It's best to think of the phenomenon not as "returns are predictable" but "all market P/X variation comes from changing return forecasts, and none from changing dividend / cash flow / earning forecasts." Our view of the world is 100% different than a generation ago. Return news, discount rate news, risk premium news drives markets, not cashflow news. This is the measure of how important the whole business is.

This fact has many implications beyond portfolios, and means most of finance (and half of accounting) must be rewritten. This is where I think the real "practical use" comes, not in some quick and easy trading strategy. For example,

1. The CAPM assumes constant expected returns. Now we know the CAPM must be wrong!
2. Betas reflect covariation between firm and factor (e.g., market) discount rates rather than reflecting the covariation between firm and market cash flows. Betas are therefore different at different horizons.
3. A change in prices driven by discount rate changes does not change how close the firm is to bankruptcy. Corporate finance ties stock price changes to the bankruptcy market (Debt/Equity) question. We see in fact very inertial capital structures, and this makes sense now. All corporate finance assumes that stock price changes come from news about cashflows!
4. Cost-of-capital calculations

$$V = E(CF)/E(R) = E(CF)/\left[R^f + \beta E(R^m - R^f)\right]$$

featuring the CAPM and a steady 6% market premium need to be rewritten, at least recognizing the dramatic variation of the initial premium, and more deeply recognizing likely changes in that premium over the lifespan of a project and the multiple pricing factors that predictability implies.

5. Marking to market assumes a random walk. "Accounting for future value" makes sense.

3 Time Series Predictability Appendix

These are some extensions and interesting issues that I will not pursue in class for lack of time. If you're interested in pursuing these issues further, this is the place to start.

3.1 More variables

Important: *I only emphasized D/P regressions for simplicity. Lots of other variables help to forecast returns.* Among these, the Term premium, Level of interest rates, Consumption/wealth ratio ("cay"), Investment , Share issuance, volatility, the realized-implied vol spread, and many other variables all help to forecast stock returns. Remember the fact as

$$R_{t+1} = a + b_1(D/P)_t + b_2 x_t + ... + \varepsilon_{t+1}$$

How will our picture change as we add variables past DP? In "Discount Rates" I looked at one very nice variable, Martin Lettau and Syndey Ludvigson's "cay" or consumption to wealth ratio as an example. Let's look

Table IV

Forecasting Regressions with the Consumption-wealth Ratio

Annual data 1952-2009. Long-run coefficients in the last two rows of the table are computed using a first-order VAR with dp_t and cay_t as state variables. Each regression includes a constant. *Cay* is rescaled so $\sigma(cay) = 1$. For reference, $\sigma(dp) = 0.42$.

	Coefficients		t-statistics		Other statistics		
Left-hand Variable	dp_t	cay_t	dp_t	cay_t	R^2	$\sigma\left[E_t(y_{t+1})\right]\%$	$\frac{\sigma[E_t(y_{t+1})]}{E(y_{t+1})}$
r_{t+1}	0.12	0.071	(2.14)	(3.19)	0.26	8.99	0.91
Δd_{t+1}	0.024	0.025	(0.46)	(1.69)	0.05	2.80	0.12
dp_{t+1}	0.94	-0.047	(20.4)	(-3.05)	0.91		
cay_{t+1}	0.15	0.65	(0.63)	(5.95)	0.43		
$r_t^{lr} = \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$	1.29	0.033				0.51	
$\Delta d_t^{lr} = \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}$	0.29	0.033				0.12	

First row: *Cay* helps to forecast one-period returns. The t-statistic (3.19) is large, and it raises the variation of expected returns substantially. 26% is a huge R^2! It is an *additional* forecast variable. Adding *cay* doesn't change the *dp* coefficient much at all.

Second row: *Cay* only marginally helps to forecast dividend growth.

Notice also that the *cay* signal decays much more quickly than *dp*, with a 0.65 coefficient rather than 0.94. That's a big difference.

Figure 10 graphs the one-year return forecast using *dp* alone, the one-year return forecast using *dp* and *cay* together, and the actual ex-post return. Adding *cay* lets us forecast business-cycle frequency "wiggles" while not affecting the "trend." That's why *both* cay and *dp* help each other to forecast returns.

The long run?

So, it looks like we have a potent new forecaster, eh? But now, how does adding *cay* affect all our long-run forecasts, our understanding of where price variance comes from, and our impulse-response functions?

Long-run return forecasts are quite different. The bottom rows of the table show long run forecasts. The dp coefficient is about the same – around one – but the cay coefficient is now tiny! Figure 11 contrasts long-run return forecast with and without *cay*. Though *cay* has a dramatic effect on *one-period* return r_{t+1} forecasts, *cay* has almost no effect at all on *long-run* return $\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$ forecasts!

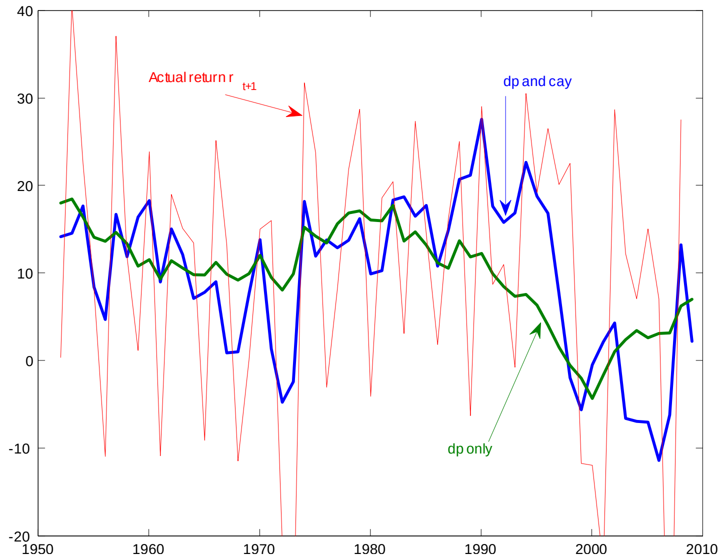


Figure 10: **Forecast and actual one-year returns.** The forecasts are fitted values of regressions of returns on dividend yield and *cay*. Actual returns r_{t+1} are plotted on the same date as their forecast, $a + b \times dp_t$.

Figure 11 includes the actual dividend yield, to show how dividend yields break into long-run return vs. dividend growth forecasts. The forecast based on dp still accounts for all the dividend yield variation!

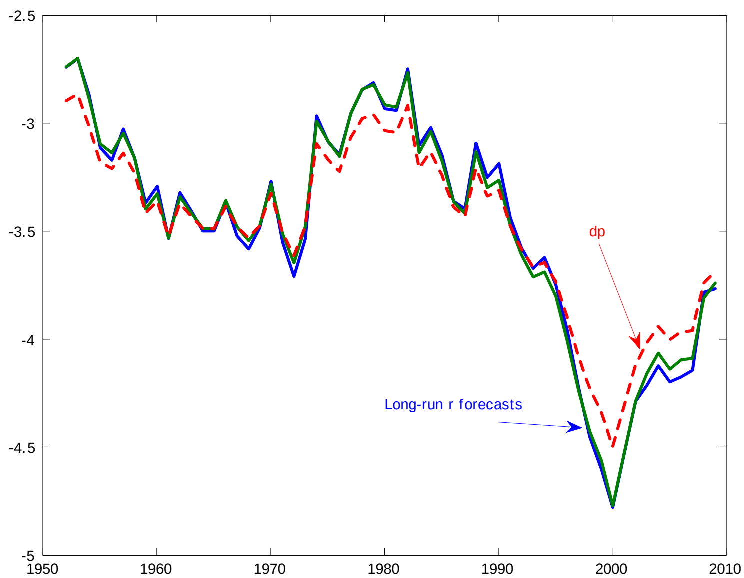


Figure 11: **Log dividend yield dp and forecasts of long-run returns** $\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$. Return forecasts are computed from a VAR including dp, and a VAR including dp and cay.

How can *cay* forecast one-year returns so strongly, but have such a small effect on the terms of the dividend-yield present value identity?

Look again at the present value identity,

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

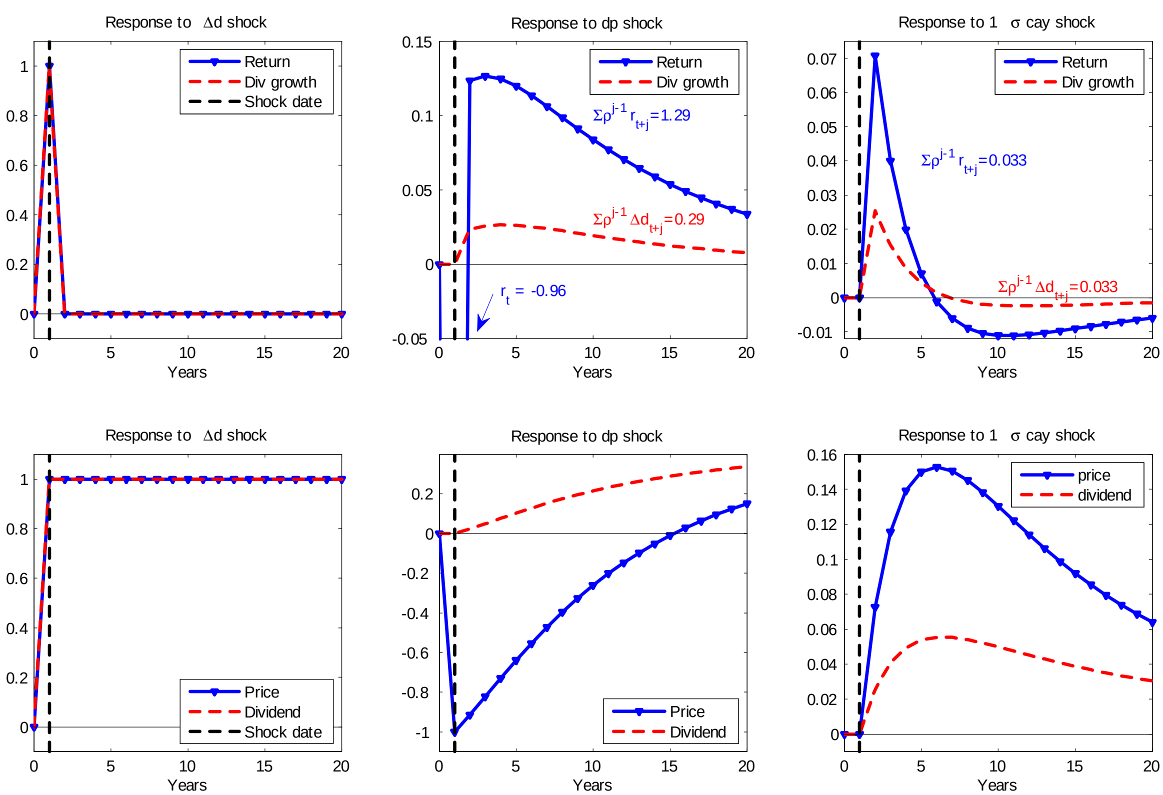
Now, if a variable (cay) changes expected return forecasts $E_t r_{t+1}$ *without affecting* $d_t - p_t$ (remember, this is a multiple regression, so the whole point is that cay can forecast returns while $d - p$ doesn't move), then it must *also* affect some of the other terms on the right hand side.

For example, it's possible that cay can forecast one-year returns r_{t+1} and also help to forecast long-run returns $\sum \rho^{j-1} r_{t+j}$ if cay *also* forecasts dividend growth. This is plausible: in the bottom of a recession, people expect strong growth, but they are also very risk averse: high $E_t(r_{t+j})$ and also high $E_t(\Delta d_{t+j})$. But in the present value identity, moving r and Δd up together doesn't change prices – the two terms offset. So there is the strong possibility that a variable like *cay* can increase return foreastability *and* dividend forecastability at the same time.

This is an important point. A low valuation can mean high expected returns *or* low future earnings. By finding other variables that help to forecast earnings/dividends, those variables will "clean up" valuation ratios and help us to forecast returns. This *is* what happens later in the Fama-French readings.

There's another possibility. the new variable could raise $E_t r_{t+1}$ but *lower* $E_t r_{t+j}$ so that $d_t - p_t$ is not affected. The variable can affect the *term structure of expected returns.* That turns out to be the case.

We can display this behavior with impulse-response functions. Figure 3.1 plots responses to a dividend growth shock, a dividend yield shock, and a *cay* shock. In each case, I include a contemporaneous return response to satisfy the return identity $r_{t+1} = \Delta d_{t+1} - \rho dp_{t+1} + dp_t$.



Impulse-response Functions. Response functions to dividend growth, dividend yield, and *cay* shocks. Calculations are based on the VAR of Table IV. Each shock changes the indicated variable without changing the others, and includes a contemporaneous return shock from the identity $r_{t+1} = \Delta d_{t+1} - \rho dp_{t+1} + dp_t$. The vertical dashed line indicates the period of the shock.

The left and middle panels look just like what we saw above. The last panel says, what do we learn from a shock to *cay* , with no contemporaneous movement of dp or of Δd? In the top right panel, you see it raises expected returns – that's the great one-year return forecast we saw in the table. But then this expected return decays rapidly, much faster than the expected returns that follow a dp shock. Eventually it even reverses and goes down, which is the offsetting effect we were looking for.

So cay turns out to be a "short run signal." It does help to forecast one-year returns. Hedge funds love signals like that! But it turns out not to change our picture of long run or where the bulk of *price* variation comes from.

And, by going one step further I hope you see the value of impulse response functions!

3.2 More Present value formulas

In (9) we studied the linearized present value identity

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

Here I give some more background and link it to some other useful present value formulas

3.2.1 General identity

You may ask, why do we need an approximation? Why not use the real, exact, present value formula? If you follow the above idea without doing any approximation, the corresponding formula is this[1]:

$$P_t = E_t \sum_{j=1}^{\infty} \left(\frac{1}{R_{t+1}} \frac{1}{R_{t+2}} ... \frac{1}{R_{t+j}} \right) D_{t+j}$$

$$\frac{P_t}{D_t} = E_t \sum_{j=1}^{\infty} \left(\frac{1}{R_{t+1}} \frac{1}{R_{t+2}} ... \frac{1}{R_{t+j}} \right) \frac{D_{t+j}}{D_t} \tag{23}$$

You do see that higher future dividends or lower future returns mean higher prices, so the qualitative features of the previous formulas are all there. You can also see that the approximate formula (9) is basically a Taylor expansion of this one, using sums in the place of products.

The problem is, it's is going to be a bear to get any *numbers* out of this formula, to connect it *quantitatively* to our forecasting regressions. Products of things are hard to deal with. Sums are much easier!

[1]Derivation: Start with the definition of return

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

then

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

Now use the same equation for P_{t+1} to substitute

$$\begin{aligned} P_t &= \frac{1}{R_{t+1}} \left(\frac{1}{R_{t+2}} \left(P_{t+2} + D_{t+2} \right) + D_{t+1} \right) \\ &= \frac{1}{R_{t+1}} D_{t+1} + \frac{1}{R_{t+1}} \frac{1}{R_{t+2}} D_{t+2} + \frac{1}{R_{t+1}} \frac{1}{R_{t+2}} P_{t+2} \end{aligned}$$

Keep going this way, (yes the last term goes to zero)

$$P_t = \sum_{j=1}^{\infty} \left(\frac{1}{R_{t+1}} \frac{1}{R_{t+2}} ... \frac{1}{R_{t+j}} \right) D_{t+j}$$

Take E_t of both sides, divide by D_t and you get the formula in the text.

You might worry about the accuracy of the approximation (9). It is leaving out all sorts of effects that (23) will include, such as the fact that $E(\frac{1}{R_1}\frac{1}{R_2})$ or $E(\frac{1}{R_1}\frac{D_1}{D_2})$ will reflect *correlations* between the terms not present in the linear model, or the evident nonlinearity. However, exploration over a wide range of parameters applicable to stocks shows that the approximation (9) is typically excellent for all the uses here. For some big nonlinear econometric model, or for a security like options with a weird payoff structure, the linear formula would not work.

3.2.2 Gordon growth formula.

The *Gordon growth* formula is an even easier back-of-the envelope present value formula that applies to long-lived assets like stocks, and a nice addition to your toolkit. Use small letters for net or log growth rates, $D_{t+1}/D_t = 1 + \Delta d_{t+1}$ or $\Delta d_{t+1} = \log(D_{t+1}/D_t)$ and similarly $E(R) = 1 + r$,

- *If Δd and r are constant forever, then the general present value formula simplifies to*[2]

$$\frac{P}{D} = \frac{1}{r-g}$$

Look at some implications

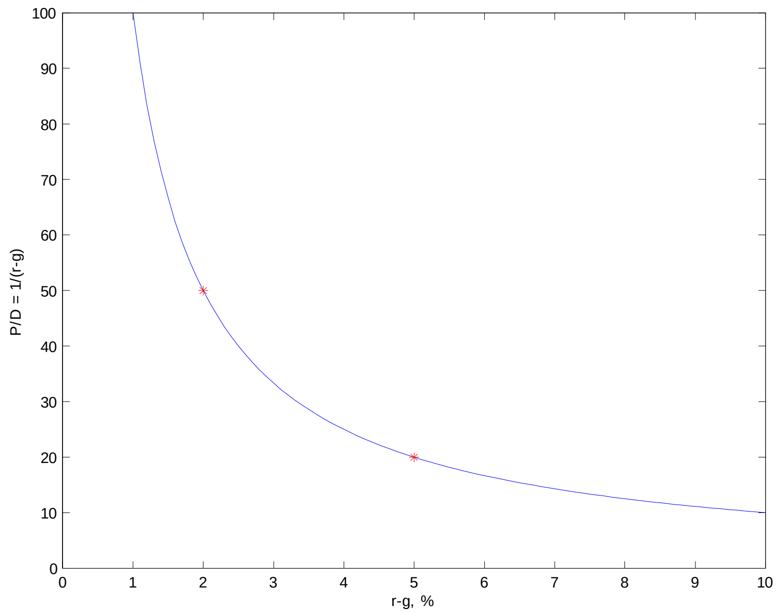
1. Low r implies high P/D. Lower discount rates (expected returns) means higher prices. Of course, higher prices for the same dividend stream corresponds to lower returns.
2. Duh, but this is quantitative. For us,
$$\frac{D}{P} = r - g$$
suggests a slope of about 1 in the relation between returns and D/P. Ok, it's really 3.25, but we're in the ballpark
3. The formula is nonlinear: P/D is very sensitive to r and g when $r-g$ near 0. This is an advantage over the approximate formula (9), which doesn't let us think about nonlinearities. This fact illuminates some controversies and recent events. For example,

[2] Derivation:

$$\frac{P}{D} = \int_{t=0}^{\infty} e^{-rt}\frac{D(t)}{D}dt = \int_{t=0}^{\infty} e^{-rt}e^{gt}dt = \int_{t=0}^{\infty} e^{-(r-g)t}dt = -\frac{1}{r-g}e^{-(r-g)t}\Big|_0^{\infty} = \frac{1}{r-g}$$

. We can also get there as a special version of the discrete-time formula

$$\begin{aligned} \frac{P_t}{D_{t+1}} &= \sum_{j=1}^{\infty}\left(\frac{1}{R_{t+1}}\frac{1}{R_{t+2}}...\frac{1}{R_{t+j}}\right)\frac{D_{t+j}}{D_{t+1}} \\ &= \sum_{j=1}^{\infty}\frac{(1+g)^{j-1}}{(1+r)^j} = \frac{\frac{1}{1+r}}{1-\frac{1+g}{1+r}} = \frac{1}{r-g} \end{aligned}$$

4. 

   (a) Suppose a stock has $P/D = 50$. This means $r - g = 0.02$. Now a 1% decline in r or 1% rise in g means prices double to $P/D = 100$!
   (b) But suppose instead a stock has $P/D = 20$. now $r - g = 0.05$. A 1% decline in r or 1% rise in g means $P/D = 25$, a much smaller increase. The slope of the above graph is much higher at $r - g = 0.02$ than it is at $r - g = 0.05$. This means...
   (c) Growth stocks or high markets should be more volatile (if hit by the same kinds of g, r news)

5. The formula is also a handy way to demonstrate how very small discount rate movements, if persistent, can have huge effects on prices. Are markets "close to efficient"? It depends on how you look at it. At a $P/D = 50$, $r - g = 0.02$, a 10bp (0.1%) return anomaly means a 5% price rise. Is the anomaly a "trivial" 10 bp return, or a substantial 5% error in the level of prices? Short costs are about 1% per year, so a "trivial," unexploitable 1% per year return error can translate into doubling the price!

The Gordon growth formula improves on our linearized identity, in that it is nonlinear. As you can see it captures some nonlinear effects that are important for big changes in dividend growth and returns. Yes, $1/(r - g) \approx 1 - r + g$, but the latter has its limits.

The Gordon formula is not enough however. This model assumes r and g are constants forever – it predicts P/D never changes!. People often think of $E(r)$ and $E(g)$ as varying through time, and then use a formula for P/D. But that's inconsistent, because the formula assumes that r and g never change. It also doesn't allow us to evaluate what happens to prices if there is a long-lived but ultimately temporary fluctuation in expected return or dividend growth.

3.2.3 Other versions of the linearized present value identity

Present value formulas are of course incredibly useful in finance. In some sense that's all finance is about. The linearized present value formula is a major step forward our tool kit. Here are some other versions of it.

1. Interest rates and excess returns. Sometimes you want to think about how lower interest rates affect stock prices, or how credit spreads might do so. Decompose the stock return into a risk free rate and a risk premium (this is just a definition of rx) $r_t = rf_t + rx_t$, and then

$$p_t - d_t \approx E_t \sum_{j=1}^{\infty} \rho^{j-1} \left(\Delta d_{t+j} - rf_{t+j} - rx_{t+j} \right)$$

We can ask if pd forecasts dividend growth, interest rates, or excess returns. Empirically, the answer is, it's all excess returns. Prices seem to move on *risk premia*, not on interest rate news.

2. Real and nominal returns. You can write the same formula in terms of real or nominal dividend growth and returns. This just shifts expected inflation from one column to the other. Real dividend growth is $\Delta d_t^r = \Delta d_t - \pi_t$; real returns are $r_t^r = r_t - \pi_t$. the π cancel, so

$$p_t - d_t \approx E_t \sum_{j=1}^{\infty} \rho^{j-1} \left(\Delta d_{t+j}^r - r_{t+j}^r \right).$$

3. We don't need to use dividends. A formula without dividends is more useful for thinking about individual stocks, because they often don't pay dividends. Tuomo Vuolteenaho developed an alternative

$$\log \left(\frac{\text{market}_t}{\text{book}_t} \right) \tilde{=} E_t \sum \rho^j \left[\log \left(1 + \frac{\text{earnings}_{t+j}}{\text{book}_{t+j}} \right) - r_{t+j} \right]$$

3.3 The return identity

We have focused on the information in *prices*, how price-dividend ratios reflect information about expected cashflows and discount rates. It's interesting to do the same thing for *returns*, to understand how returns reflect similar information. Of course once you have one, you have the other, since returns come from dividends and changes in prices. Still, looking explicitly at returns is important, and you get further by looking at a few equations.

You get returns from price changes or dividends. Thus, we expect that the "return" version of our present value identity will look something like a first-difference of that identity, plus a term reflecting changes in current dividends, which change current prices but leave the P/D ratio unaffected.

Here it is

$$r_t - E_{t-1} r_t = (E_t - E_{t-1}) \left(\Delta d_t + \sum_{j=1}^{\infty} \rho^j \Delta d_{t+j} - \sum_{j=1}^{\infty} \rho^j r_{t+j} \right) \tag{24}$$

"$(E_t - E_{t-1})(x)$" means $E_t(x) - E_{t-1}(x)$. It's the "surprise" in x, the change in the expected value of x. For example, if Wednesday's forecast for Sunday is 60^o but Thursday's forecast is 70^o then

$$E_{\text{thurs}}(T_{\text{sun}}) - E_{\text{wed}}(T_{\text{sun}}) = (E_{\text{thurs}} - E_{\text{wed}})(T_{\text{sun}}) = +10^o.$$

Thus, (24) tells us that

- We see an *unexpected* return, when there is news about current dividends, future dividends, and/or future returns. Good news about current or future dividends raises today's return, higher future expected returns ("required returns," "discount rates") lowers today's return.

The derivation basically involves taking surprises of $p - d$ formula

Derivation: Take $(E_{t+1} - E_t)$ of the identity

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

$(E_{t+1} - E_t)(p_t - d_t) = 0$ of course, so

$$\begin{aligned} 0 &= (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}) \\ (E_{t+1} - E_t) r_{t+1} &= (E_{t+1} - E_t) \Delta d_{t+1} + \sum_{j=2}^{\infty} \rho^{j-1} (\Delta d_{t+1+j} - r_{t+1+j}) \end{aligned}$$

Shifting time index back one, we get (24). We also had this back when we ran regressions (12)-(13). Runing both sides of the identity on $(d_t - p_t)$, we deduced

$$1 \approx b_r^{lr} - b_d^{lr}.$$

But, as in the one period model, we also have

$$0 = \varepsilon^r - \varepsilon^d.$$

which is our return identity.

One of the most (initially) counterintuitive aspects of formula (24) (and, implicitly, the dividend-yield identity (9)) is that

- A decline in *expected* return ($E_t r_{t+j}$) leads to high *ex-post* return (r_t). Conversely, a *rise* in expected returns $E_t(r_{t+j})$ *lowers* prices and current returns r_t.

You probably thought, "if expected returns rise, that's good news for investors, so that means prices must rise." That's wrong. If expected returns rise, holding dividends constant, the only way to achieve those returns is for current prices to fall, so we see a negative current return. This is clear for bonds: if interest rates rise, prices fall. It's less clear for stocks (and lots of people get this wrong), but the same principle holds. Equivalently, you can think, "if required returns and discount rates rise, future dividends are discounted at a higher rate, so prices must fall."

If you're still scratching your head, you're probably confusing expected *dividends and cashflows* with expected financial *returns*. If expected dividends rise, sure prices rise and you see a good return. But if expected returns rise, prices must fall. Or, perhaps you're making a common mistake among investors, assuming there is a lot of "momentum" in stock returns. A string of good past returns means returns will be better going forward. Examine the first table. In fact, good returns without a change in dividends means *lower* future returns.

For example, why did stocks boom in the 1990s? Here is one good story. Expected returns $E(r)$ got progressively *lower* (yes, lower). Boomers got into stocks, 401(k) and mutual funds made it easier, economic boom made people more willing to take risk. As $E(r)$ declines *actual* (ex-post) r is high, and $p - d$ rises. Eventually, Er stops declining, and the party is over. Then Er is low for the foreseeable future (us).

One sum of what we learned in both price and returns here is that *Stocks are a little like bonds.* For bonds, we understand well that price rises, and ex-post returns are good when yields (subsequent returns) decline. Most people see a good stock return and think expected future returns have *risen.* They're wrong.

Equation (24) allows us to make a variance decomposition of returns as we did for prices – multiply both sides of (24) by unexpected return and take expectations. To simplify the formula a bit, write

$$\Delta E_t(r_t) \equiv E_t(r_t) - E_{t-1}(r_t).$$

then, proceeding as before[3],

$$\begin{aligned} var\left[\Delta E_t(r_t)\right] &= cov\left[\Delta E_t(r_t), \Delta E_t(\Delta d_t)\right] \\ &+ cov\left[\Delta E_t(r_t), \Delta E_t\left(\sum_{j=1}^{\infty}\rho^j \Delta d_{t+j}\right)\right] - cov\left[\Delta E_t(r_t), \Delta E_t\left(\sum_{j=1}^{\infty}\rho^j r_{t+j}\right)\right] \end{aligned}$$

The left is the variance of unexpected returns, which (since returns are nearly unpredictable) is nearly the same thing as the variance of returns. The formula says that

- *Returns can only vary unexpectedly, if return shocks covary with current dividend growth shocks, or with shocks to future long-run returns or dividend growth.*

[3] This idea is from Denis Chaves' (2009) thesis.

Once again, which is it? The answer is that about half of the variance of returns comes from *current* dividends, about half comes from *future* returns (changes in discount rates) and about none from future dividend growth.

Current Dividends	Future Dividends	Future Returns
49.03	-7.81	54.80

Variance decomposition of returns. The three terms are the contributions of the three terms on the right hand side of (24) to return variance.

Now, you might think (and plenty of academics make this mistake) that there is big news here: *half* of return variance comes from dividend growth, where *non* of pd variance came from dividend growth. But this is not new at all. When *current* dividends change and prices change the same amount, *returns* change, but the price-dividend ratio does not. Dividend yields only respond to variation in the last two terms on the right hand side.

Algebra. The calculation behind the return decomposition is not hard but it does take a few equations. I start with regressions,

$$\begin{aligned}
dp_{t+1} &= \phi dp_t + \varepsilon^{dp}_{t+1} \\
r_{t+1} &= b_r dp_t + \varepsilon^{r}_{t+1} \\
\Delta d_{t+1} &= b_d dp_t + \varepsilon^{dp}_{t+1}
\end{aligned}$$

The expected future return is

$$\begin{aligned}
E_t\left(r_{t+j}\right) &= b_r \phi^{j-1} dp_t \\
E_t\left(\Delta d_{t+j}\right) &= b_d \phi^{j-1} dp_t
\end{aligned}$$

Thus, the unexpected future return and dividend growth are

$$\begin{aligned}
\left(E_t - E_{t-1}\right)\left(r_{t+j}\right) &= b_r \phi^{j-1}\left(E_t - E_{t-1}\right) dp_t = b_r \phi^{j-1} \varepsilon^{dp}_t \\
\left(E_t - E_{t-1}\right)\left(\Delta d_{t+j}\right) &= b_d \phi^{j-1} \varepsilon^{dp}_t
\end{aligned}$$

Thus, (24) reads

$$\begin{aligned}
\varepsilon^r_t &= \varepsilon^d_t + \sum_{j=1}^{\infty} \rho^j b_d \phi^{j-1} \varepsilon^{dp}_t - \sum_{j=1}^{\infty} \rho^j b_r \phi^{j-1} \varepsilon^{dp}_t \\
\varepsilon^r_t &= \varepsilon^d_t + \frac{b_d \rho}{1-\rho\phi} \varepsilon^{dp}_t - \frac{b_r \rho}{1-\rho\phi} \varepsilon^{dp}_t
\end{aligned}$$

Now, do the same variance decomposition trick we did for prices,

$$var(\varepsilon^r_t) = cov(\varepsilon^r_t, \varepsilon^d_t) + \frac{b_d \rho}{1-\rho\phi} cov(\varepsilon^r_t, \varepsilon^{dp}_t) - \frac{b_r \rho}{1-\rho\phi} \varepsilon^{dp}_t cov(\varepsilon^r_t, \varepsilon^{dp}_t)$$

The table just computes these three terms.

Campbell does a similar return decomposition, but takes the *variance* of the right hand side, which means troublesome covariance terms show up between the right hand side components.

We can also compute the terms using the simplified VAR given in the notes.

$$\begin{aligned}
\left(E_{t+1} - E_t\right) r_{t+1} &= \varepsilon^r_{t+1} \\
\left(E_{t+1} - E_t\right) r_{t+2} &= b_r dp_{t+1} - b_r \phi dp_t = b_r \left(\phi dp_t + \varepsilon^{dp}_{t+1}\right) - b_r \phi dp_t = b_r \varepsilon^{dp}_{t+1} \\
\left(E_{t+1} - E_t\right) r_{t+3} &= b_r \phi dp_{t+1} - b_r \phi^2 dp_t = b_r \phi \varepsilon^{dp}_{t+1}
\end{aligned}$$

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

$$\varepsilon^r_{t+1}=\varepsilon^d_{t+1}+b_d\rho(1+\rho\phi+\rho^2\phi^2+...)\varepsilon^{dp}_{t+1}-b_r\rho(1+\rho\phi+\rho^2\phi^2+...)\varepsilon^{dp}_{t+1}$$

$$\varepsilon^r_{t+1}=\varepsilon^d_{t+1}+\rho\frac{b_d}{1-\rho\phi}\varepsilon^{dp}_{t+1}-\rho\frac{b_r}{1-\rho\phi}\varepsilon^{dp}_{t+1}$$

recall

$$b_r-b_d=1-\rho\phi$$

$$\varepsilon^r_{t+1}\approx\varepsilon^d_{t+1}-\rho\varepsilon^{dp}_{t+1}$$

so we're right back at the return identity. But now we have interesting names and numbers for the three terms.

$$\begin{aligned}
\varepsilon^r_{t+1} &= \varepsilon^d_{t+1}+0\times\varepsilon^{dp}_{t+1}-0.96\frac{0.1}{1-0.96\times 0.94}\varepsilon^{dp}_{t+1}\\
\varepsilon^r_{t+1} &= \varepsilon^d_{t+1}-0.98\varepsilon^{dp}_{t+1}\\
\sigma^2\left(\varepsilon^r_{t+1}\right) &= \sigma^2\left(\varepsilon^d_{t+1}\right)+0.98^2\sigma^2\left(\varepsilon^{dp}_{t+1}\right)\\
0.20^2 &= 0.14^2+0.98^2\times 0.15^2
\end{aligned}$$

As advertised, about half and half.

3.4 Are stocks "Safer in the long run"?

You see in our impulse-response functions that predictability means that prices have a transitory component. "Low prices" mean "high returns" and vice versa. There *is* in fact some meaning to the standard interpretation that there are "short run price movements that will melt away." It seems the market really is "safer" for "long run investors" who can "wait out temporary price movements."

"Safety" and "Predictability" are linked of course. If "temporary" price movements will "revert", then the market is "safer" for investors who can "wait out" those temporary price movements. *Asset Pricing* uses some models to link the two concepts, but you don't need that to understand the basic point. Problem set 1 asked you to investigate this issue by looking at risks over various horizons induced by persistent processes.

By analogy, 10 year indexed zero coupon bonds are riskless for 10 year investor. Long-term bonds *are* safer in the long run, and "short-term" price movements *do* melt away. Yet 10 year bonds have huge price swings at a one-year horizon. Why? Because a good *past* return comes with a reduction in *future* yield. Well, we see much the same pattern in stocks. Can we arrive at the same conclusion?

Is it true? It turns out, amazingly, that the answer is "not necessarily." This case is a good one for the advantages of writing out a few equations, because many Sages of Finance have gotten it wrong.

Already you saw in Figure 9 that returns, taken by themselves, do not seem safer at all. There is, in the full sample, no tendency for a high return to be reversed in the long run. If anything, return shocks seem to build up a bit!

How is this possible? How is this consistent with our dividend yield regressions? What are the facts about "safety in the long run" more generally? That's what we're here for.

3.4.1 Horizon effects

First, we need to understand how returns behave over different horizons.

As a benchmark, if returns are uncorrelated over time, there are no horizon effects. If you're looking at a coin flip it does *not* get safer by flipping many times. (This is the "fallacy of time-diviersification," see *Investments Notes*)

Now if returns are uncorrelated over time, then mean and variance scale with horizons; standard deviation and Sharpe ratio scale with square root of horizon.[4]

$$\begin{aligned} r_{t\to t+2} &= r_{t+1} + r_{t+2} \\ E(r_{t\to t+2}) &= 2E(r) \\ \sigma^2(r_{t\to t+2}) &= 2\sigma^2(r) \\ \sigma(r_{t\to t+2}) &= \sqrt{2}\sigma(r) \\ \frac{E(r_{t\to t+2})}{\sigma(r_{t\to t+2})} &= \sqrt{2}\frac{E(r)}{\sigma(r)} \end{aligned}$$

3.4.2 Fact

With this in mind, let's look at the Facts. From *Asset Pricing* Table 20.5, 20.6, 20.7.

Table 20.5. *Mean-reversion using logs, 1926–1996*

	Horizon k (years)					
	1	2	3	5	7	10
$\sigma(r_k)/\sqrt{k}$	19.8	20.6	19.7	18.2	16.5	16.3
β_k	0.08	−0.15	−0.22	−0.04	0.24	0.08
Sharpe$/\sqrt{k}$	0.31	0.30	0.30	0.31	0.36	0.39

r denotes the difference between the log value-weighted NYSE return and the log treasury bill return. $\sigma(r_k) = \sigma(r_{t\to t+k})$ is the variance of long-horizon returns. β_k is the long-horizon regression coefficient in $r_{t\to t+k} = \alpha + \beta_k r_{t-k\to t} + \varepsilon_{t+k}$. The Sharpe ratio is $E(r_{t\to t+k})/\sigma(r_{t\to t+k})$.

Table 20.6. *Mean-reversion using gross returns, 1926–1996*

	Horizon k (years)					
	1	2	3	5	7	10
$\sigma(r_k)/\sqrt{k}$	20.6	22.3	22.5	24.9	28.9	39.5
β_k	0.02	−0.21	−0.22	−0.03	0.22	−0.63
Sharpe$/\sqrt{k}$	0.41	0.41	0.41	0.40	0.40	0.38

r denotes the difference between the gross (not log) long-horizon value-weighted NYSE return and the gross treasury bill return.

[4] Reminder: This is a classic application of hte formulas for mean and variance of a sum

$$\begin{aligned} E(x+y) &= E(x) + E(y) \\ var(x+y) &= var(x) + var(y) + 2cov(x,y). \end{aligned}$$

Table 20.7. *Mean-reversion in postwar data*

	Horizon k (years)					
1947–1996 logs	1	2	3	5	7	10
$\sigma(r_k)/\sqrt{k}$	15.6	14.9	13.0	13.9	15.0	15.6
β_k	−0.10	−0.29*	0.30*	0.30	0.17	−0.18
Sharpe/$\sqrt{k}$	0.44	0.46	0.51	0.46	0.41	0.36
1947–1996 levels	1	2	3	5	7	10
$\sigma(r_k)/\sqrt{k}$	17.1	17.9	16.8	21.9	29.3	39.8
β_k	−0.13	−0.33*	0.30	0.25	0.13	−0.25
Sharpe/$\sqrt{k}$	0.50	0.51	0.55	0.48	0.41	0.37

Using log returns (top) you see that the standard deviation, corrected for horizon, decreases very slightly from about 20% to about 16%, and the Sharpe ratio rises from about 0.31 to 0.39. However, using gross returns in the middle table, you see the opposite pattern, a slight increase in risk with horizon, and using postwar data the risk is absolutely flat, independent of horizon. In sum,

- There is *not* much consistent evidence that stocks are "safer in the long run" What evidence we might see is much weaker than the dp forecasts suggested.

How is this possible? How can it possibly be true that d/p forecasts returns, but stocks are not "safer in the long run?"

The fact is that high *price-dividend* is followed by low returns, but large *past returns* are not followed by low returns. Consider a weather forecasting story. Suppose that each day's temperature is truly random, so that today's temperature tells you nothing about tomorrow's. However, the weather forecaster looks at more than just today's weather. He looks at the satellite photo and the computer forecast. These include data beyond just today's temperature. By looking at other data, he might be able to tell you the temperature a day in advance.

There you have it: temperature could be quite forecastable from *other information*, but perfectly unforecastable *looking only at past temperature*. That's pretty much what we have. Returns are forecastable, and we can identify temporary movements in prices, using *other information* (dividends) though today's weather forecast (incorporating d news) can tell you a lot. Weather can be uncorrelated over time, yet forecastable.

Implication: *market timing* may work a bit (for patient investors) – buy when D/P is high. It is not clear that "long run investors should hold more stocks *on average*."

3.5 A simple proof that stocks can be predictable and uncorrelated

If returns are independent over time, both means and variances of returns grow with horizon, so stocks are not at all "safer" for long-horizon investors. For example, consider two-period log returns

$$
\begin{aligned}
E\left(r_{t+1}+r_{t+2}\right) &= 2E(r) \\
\sigma^2\left(r_{t+1}+r_{t+2}\right) &= 2\sigma^2(r)+2cov(r_{t+1},r_{t+2}).
\end{aligned}
$$

If returns are independent over time, the covariance term is zero, and both mean and variance scale linearly with horizon. If the covariance term is negative, stocks bounce back after declines, so the variance scales less than linearly with horizon, and stocks are safer for long-run investors.

Stock returns are, in fact, predictable. But that does not mean they are any safer for long-run investors. This is a nice little paradox.

Sum up the simplest version of stock predictability with a vector autoregression

$$\begin{aligned} r_{t+1} &= b_r dp_t + \varepsilon^r_{t+1} \\ \Delta d_{t+1} &= b_d dp_t + \varepsilon^d_{t+1} \\ dp_{t+1} &= \phi dp_t + \varepsilon^{dp}_{t+1} \end{aligned}$$

Here dp is the log dividend yield and Δd is log dividend growth. The return coefficient is about $b_r \approx 0.1$ and the dividend growth coefficient is about zero $b_d \approx 0$. So, low prices mean high subsequent returns, and high prices (relative to dividends) mean low subsequent returns.

It would seem that stocks are indeed much safer for long-run investors, as there really is a sense that low prices are "temporarily" low and will revert if you can wait long enough. But this conclusion is wrong.

I've been through three versions of showing how this paradox works. In *Asset Pricing* the best I could come up with was a complex factorization of the spectral density matrix in order to derive the univariate process for returns implied by the VAR. In later Ph.D. classes, I found a way to do it much more simply, by seeing that returns have to follow an ARMA(1,1), and then matching coefficients. This year, I found a way to show it even more simply and intuitively. Here goes.

Use the VAR to write

$$\begin{aligned} r_{t+1} &= b_r dp_t + \varepsilon^r_{t+1} \\ r_{t+2} &= b_r \phi dp_t + b_r \varepsilon^{dp}_{t+1} + \varepsilon^r_{t+2} \end{aligned}$$

so

$$\begin{aligned} cov(r_{t+1}, r_{t+2}) &= cov\left[b_r dp_t + \varepsilon^r_{t+1}, b_r\left(\phi dp_t + \varepsilon^{dp}_{t+1}\right) + \varepsilon^r_{t+2}\right] \\ cov(r_{t+1}, r_{t+2}) &= b_r^2 \phi \sigma^2(dp_t) + b_r cov(\varepsilon^r_{t+1}, \varepsilon^{dp}_{t+1}) \end{aligned} \tag{25}$$

The first term $b_r^2\phi\sigma^2(dp)$ induces a positive autocorrelation or "momentum," making stocks actually riskier for long term investors. dp_t moves slowly over time, so if returns r_{t+1} are higher than usual, then returns r_{t+2} are likely to be higher than usual as well.

The second term $b_r cov(\varepsilon^r_{t+1}, \varepsilon^{dp}_{t+1})$ is strongly negative. If there is a positive shock to *expected* returns, this sends current prices and hence current returns down. In this way stocks are like bonds: if yields rise, prices fall and current returns fall. This is the "safer in the long-run" term.

So, we have two offsetting effects with natural interpretations as "momentum" due to slow-moving expected returns, and "mean-reversion" of price declines that correspond to rising expected returns. The fun part: In the standard parameterization, these effects almost exactly offset.

Use Campbell and Shiller's linearized return identity,

$$r_{t+1} \approx -\rho dp_{t+1} + dp_t + \Delta d_{t+1};\ \rho \approx 0.96$$

Plugging the left hand terms of the VAR into the return identity, we have

$$\begin{aligned} b_r &\approx 1 - \rho\phi + b_d \approx 1 - \rho\phi \\ \varepsilon^r_{t+1} &\approx -\rho\varepsilon^{dp}_{t+1} + \varepsilon^d_{t+1} \end{aligned} \tag{26}$$

Now, empirically, dividend growth shocks and dividend yield shocks are just about uncorrelated, $cov(\varepsilon^d_{t+1}, \varepsilon^{dp}_{t+1}) \approx 0$. So, multiplying the shock identity by ε^{dp}_{t+1} and taking expectations, we have

$$cov(\varepsilon^r_{t+1}, \varepsilon^{dp}_{t+1}) = -\rho\sigma^2\left(\varepsilon^{dp}_{t+1}\right). \tag{27}$$

Minor digression: I always found it confusing to specify directly a strong negative correlation between return and dividend yield shocks. I find it much easier to remember that dividend yield ("expected return") and dividend growth ("cashflow") shocks are, empirically, nearly uncorrelated. The strong negative correlation of dividend yield and return shocks then follows naturally.

Now, plug (27) and (26) into (25), to obtain

$$\begin{aligned} cov(r_{t+1}, r_{t+2}) &= b_r^2 \phi \sigma^2(dp_t) - b_r \rho \sigma^2 \left(\varepsilon_{t+1}^{dp}\right) \\ &= b_r \left(\phi \frac{1-\rho\phi}{1-\phi^2} - \rho\right) \sigma^2 \left(\varepsilon_{t+1}^{dp}\right) \end{aligned}$$

If we had $\phi = \rho = 0.96 = 0$, we get the result,

$$cov(r_{t+1}, r_{t+2}) = 0.$$

$\phi \approx 0.94$ is the usual estimate. But ϕ is an OLS estimate of a very persistent series, so biased down. $\phi = 0.96$ is not that far off.

Predictability does not affect the safety of stocks in the long run!

In our simple analysis we linked the autocorrelation of returns to the question whether stocks are safer for long run investors. Since stocks are essentially uncorrelated, it would seem that despite predictability they are not safer for long run investors. However a properly done (Merton) portfolio theory isn't as simple as what we did in our simple excercises. It turns out there is a "market timing" demand and a "hedging demand," and the hedging demand can be positive in our case. So, it's still possible that long-run investors should put more into stocks, even though simple Sharpe ratios are not better at long horizons. Those portfolio problems go past fixed horizon mean and variance, and those investors are actively market-timing as well as judging long-run mean and variance. I haven't yet found a simple way to calculate hedging demands however.

3.6 Data, dividends, and divisors

The use of dividends in the price/dividend ratio has attracted lots of attention. Here are two issues

3.6.1 Dividends and other divisors

As with the forecast of returns based on lagged returns, this is the tip of another iceberg. *Lots* of variables help to forecast stock returns including book/market, price/earnings, and other valuation ratios, bond yield spreads, credit spreads, and macroeconomic variables such as the consumption/wealth and investment/capital ratios. I use dividend yields here just to illustrate the idea in the simplest possible specification. We'll look at some of these other variables below.

What about dividends, and the fact that may firms don't pay them? First, keep in mind that *this fact is about the market as a whole*, not (yet) individual stocks. "Dividends" includes liquidations, all cash payouts to firms. And, as above, predictability works about as well for any other divisor. We will come back to look at individual stocks later. (Hint: the "value" effect that B/M ratios forecast returns is essentially the same thing, applied to individual stocks.)

The heart of D/P forecasts is about *prices*, not dividends. "Low" prices seem to slowly bounce back. "High" prices seem to slowly melt away (or not grow as fast as dividends for a while). "Duh," you might say, but that shouldn't happen in an efficient market with constant expected returns. Low prices should mean poor cashflows; there shouldn't be such "temporary" price movements.

Dividends are just a convenient measure of the "trend" in prices, and don't contribute any real information to the return forecasts. You get about the same results if you use price/earnings, market value / book value, price/moving average, or other divisors. Adjusting dividends for repurchases or other accounting measures doesn't make a bit of difference. You need to do something since the level of prices trends up, but anything sensible you do to remove the trend gives the same result.

For example, Figure 3.6.1 presents P/E and D/P. (Graph from Campbell and Thompson, Journal of Financial Economics 2006). Rather obviously, these will forecast returns in exactly the same way, so whether firms are paying earnings as dividends right away or not doesn't matter.

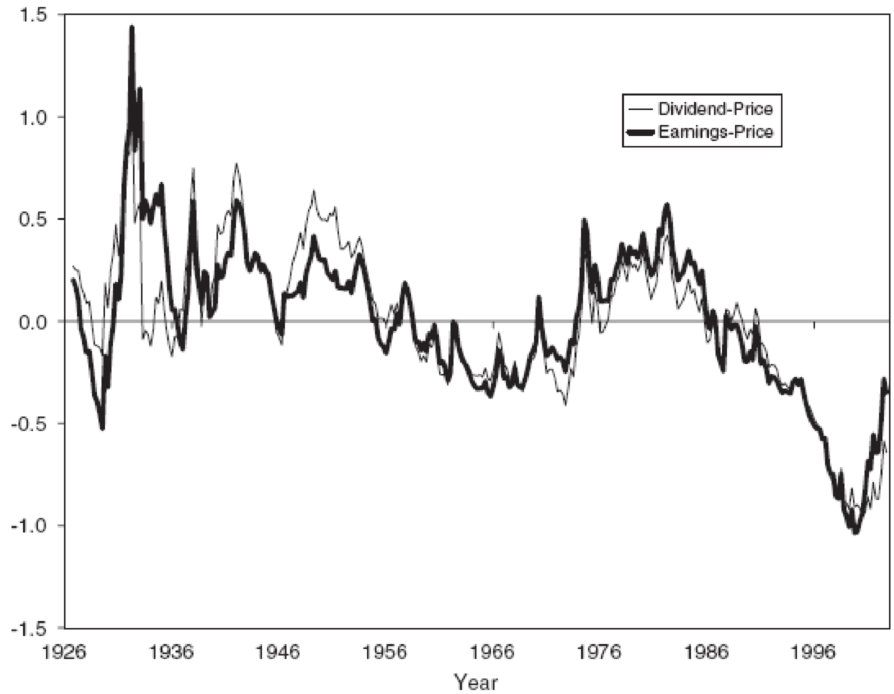


Fig. 1. Time-series plot of the valuation ratios. This figure plots the log dividend–price ratio for the CRSP value-weighted index and the log earnings–price ratio for the S&P 500. Earnings are smoothed by taking a 10-year moving average. The sample period is 1926:4–2002:4.

3.7 Repurchases

What about repurchases? Many companies are not paying dividends at all, choosing instead to occasionally repurchase shares. I show this as a more detailed case of "what happens when you add other divisors." This one attracted my attention because it seemed to make a dramatic difference. It turned out to be, I think, a spurious result that adds a little bit to the R^2 but not the promised drama. My treatment of this follows Boudoukh et al.[5]

Now, repurchases are not as big a problem as you may think. The "dividends" series we use includes all cash payments to shareholders. For example, if you run a small company, pay no dividends, and then Microsoft buys it, giving the shareholders cash, that is counted as a "dividend." Any liquidation value is also counted as "dividend." Eventually, there must be some cash payout to shareholders, or the stock is worthless!

Price still is the present value of dividends, even if firms choose to repurchase shares. Thus there is nothing "wrong" with using dividends. If a company buys back half of its shares, the remaining half are now still claims to the eventual dividend stream.

What the shift from dividend payment to repurchases might do, however, is to delay eventual payment, just like a refusal to pay dividends at all. Thus, prices might be moving on dividend news, but the dividends are so postponed that we never see them. However, in this case prices would *not* be moving on news of returns, and we see the long-run return forecastability that is just enough to explain price forecastability.

Also, these regressions are being applied to the market as a whole. If you want to look at individual companies, yes, you really do need to address the fact that dividends are zero. However, for the market as a whole, we add up over all the big companies that still are paying dividends, as well as adding up all those

[5]Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael Roberts, 2007, On the Importance of Measuring Dividend Yields: Implications for Empirical Asset Pricing," *Journal of Finance* 62, 877-915.

big one-time cash payments when Microsoft buys you out.

Still, like all questions, this should be answered in the data, not by a-priori theorizing. Maybe the very low recent D/P just reflects a change to repurchases rather than low expected returns. Perhaps this observation can help, especially in the 90s, when we saw high returns despite very low D/P, nearly killing the regression until the 2000 crash mad it look good again.

To address this question Boudoukh et al construct two measures:

(Dividends+ Repurchases)/Price

and

(Dividends+Repurchases-Issues) /Price.

Already you can see some of the accounting problems with thinking about repurchases as "payments to shareholders in lieu of dividends." But you can also see why it might make sense – this might be a better "divisor" or "trend" for prices, a better way to see "low" prices induced by high discount rates. They construct two measures, CF= based on cash flows, TS = based on treasury stock data.

Let's look at their results. Here's Table 2:

Panel A: Univariate Predictive Regressions

	Log(Dividend Yield)	Log(Payout (CF) Yield)	Log(Payout (TS) Yield)	Log(0.1 + Net Payout Yield
Full Sample: 1926 - 2003				
Coefficient	0.116	0.209	0.172	0.759
SE	0.052	0.062	0.060	0.143
t-statistic	2.240	3.396	2.854	5.311
P-Value	0.014	0.001	0.003	0.000
Sim Pval	0.170	0.045	0.080	0.000
R^2	0.055	0.091	0.080	0.262

Wow! When we use repurchase-adjusted dividends, the t jumps from 2.2 to 3.4-5.3; and the R^2 jumps from 0.055 to 0.08-*0.262!!! 0.262 is absolutely huge. This seems to make an enormous improvement.*

What about multiple regressions?

Panel B: Multivariate Predictive Regressions

	Log(Dividend Yield)	Log(Payout (CF) Yield)	Log(Payout (TS) Yield)	Log(0.1 + Net Payout Yield	R^2
Coefficient	-0.088	0.318			0.098
SE	0.111	0.129			
Coefficient	-0.394		0.641		0.112
SE	0.216		0.251		
Coefficient	-0.042			0.830	0.267
SE	0.064			0.108	

Payout yields drive out DP in multiple regressions! And the R^2 is up to a near-miraculous 0.267!

Can it really be that good? I downloaded the data (It's on Robert Whitelaw's website) and here is a plot of the various D/P measures and subsequent returns. The green line is higher than the blue line. You see how much (or little) adding payouts does to the dividend yield series that we use to forecast returns. . Payouts didn't happen before 1972, which is why the lines are the same there. The red line subtracts issues

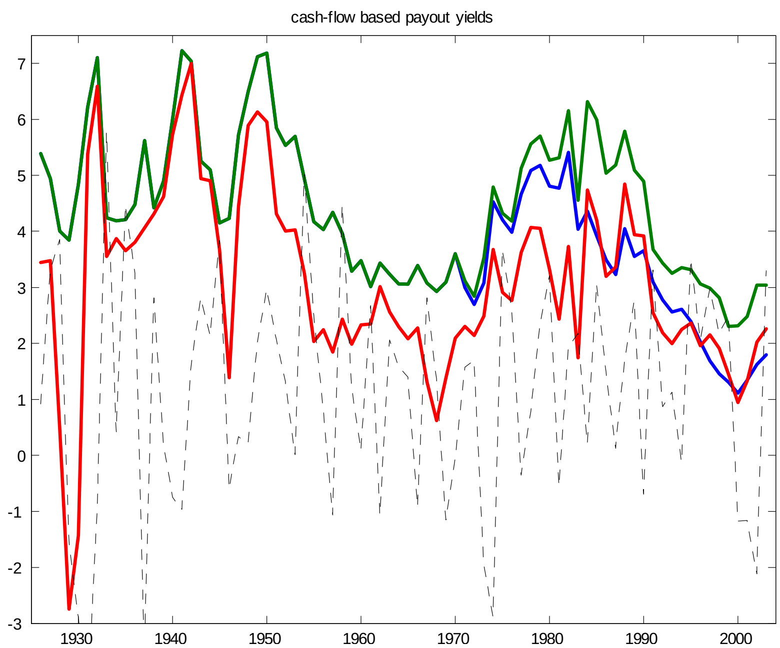


Figure 12: Boudoukh, Richardson and Whitelaw payout-adjusted dividend yields and supbsequent returns. Blue: D/P. Green: Payout/P. Red: (Net Payout)/P. Dashed: Return/10

as well. The green and red lines don't fall as much as the blue line in the 1990s, so the great 90s return is not so much of an embarrassment. That's what we came in thinking we'd see.

But the eyeball regression doesn't suggest that the green line is *that* much better than the blue line. And the huge R^2 came from the net payout yield, the red line. But that doesn't look much different either... except that one data point in the 1930s. Wait a minute, that one huge spike in the 1930s (a huge drop in share issues, followed by the -44% return in 1931) might really be helping the regression. What if we leave it out?

variable	b	t	R^2
1926-2003			
DP	4.11	2.70	0.08
payout	5.25	3.46	0.10
net	5.88	5.05	0.22(!)
1931-2003			
DP	4.04	2.69	0.09
payout	4.91	3.23	0.11
net	4.57	3.25	0.12

Aha! the huge increase in R^2 *came from this one data point.*

Morals.

- Payouts vs. dividends don't make a huge different after all. The huge increase in R^2 comes from one huge outlier data point.
- Don't believe everything you read.
- Always graph your data, before you publish the paper

Despite these critical comments, I do think there is something of value here. Here is my interpretation of the results.

- Share issuance does help to forecast returns, above and beyond D/P. Companies issue a lot of shares when prices are "high" and cost of capital (expected returns) are "low." They see the "real" divisor better than is measured by our D series. (This is the standard "Q theory" of investment) By combining share issuance with D/P regressions, this measure improves a bit. It would be better to simply include share issuance (and cay) in a multiple-regression forecast of returns.

3.8 Week 1 Review and interpretation

1. Big picture: expected returns vary *over time* in ways not described by the classic random walk/CAPM. (Later, *across assets.)*

2. Tool 1: *Forecasting regression*
$$R_{t+1} = a + bx_t + \varepsilon_{t+1}$$
These are quite different in spirit from the causal or explanatory regressions you focused on in stats class. They're a great use of the regression tool however.

3. 1st generation view: $b \approx 0$, R^2 small. "Can't beat the market" "Random walks"

4. New view: you can forecast returns.
   (a) Evidence: Forecasting regressions $R_{t+1} = a + b(D/P)_t + \varepsilon_{t+1}$.
   (b) Innovation: "1/p" variables, long horizons
   (c) $b \approx 2-5$; b, R^2 grow with horizon.

5. Inefficiency/quick profit? No, alas
   (a) R^2 is only large at long horizons; x_t moves slowly
   (b) To profit, you have to buy in bad times. (We'll look more closely at portfolio implications in last week)
   (c) Bottom line, what do we learn? *The risk premium varies over time.*

6. Long horizon predictability.
   (a) Wow!
   (b) It is not a separate phenomenon. Long horizon predictability is the same as small short horizon predictability plus persistent forecasters x_t.

7. Tool 2: A Useful linearized version of
$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}$$
$$r_{t+1} \approx \rho(p_{t+1} - d_{t+1}) + (d_{t+1} - d_t) - (p_t - d_t)$$
Use it to link $p-d$, r, Δd forecasts.

8. Tool 3: linearized present value formula
$$p_t - d_t \approx \kappa + E_t \sum_{j=1}^{\infty} \rho^{j-1} \left(\Delta d_{t+j} - r_{t+j} \right)$$
$$dp_t \approx -\kappa + E_t \sum_{j=1}^{\infty} \rho^{j-1} \left(r_{t+j} - \Delta d_{t+j} \right)$$
p-d *reveals to us* changes in market risk premia. Story: bad times come, Er rises, "selling pressure" drives down prices. We see: low prices, followed (on average) by high returns.

9. Prices (p-d) can only vary if there is news of future dividend growth or returns. Variance decomposition
$$\begin{aligned} 1 &= b_r^{lr} - b_d^{lr} \\ var\left(dp_t\right) &= var\left(p_t - d_t\right) \approx \sum_{j=1}^{\infty} \rho^{j-1} cov\left[dp_t, r_{t+j}\right] - \sum_{j=1}^{\infty} \rho^{j-1} cov\left[dp_t, \Delta d_{t+j}\right] \end{aligned}$$

(a) Results: $E_t r_{t+j}$ is the dominant cause of market price movements. (Alas).

(b) D/P does *not* forecast Δd. It "should."

(c) If D/P varies, it *must* forecast one of dividend growth or returns. Long-run regression coefficients *must* add to one

(d) Evidence on "bubbles." Long run expected return variation *is* enough. Really

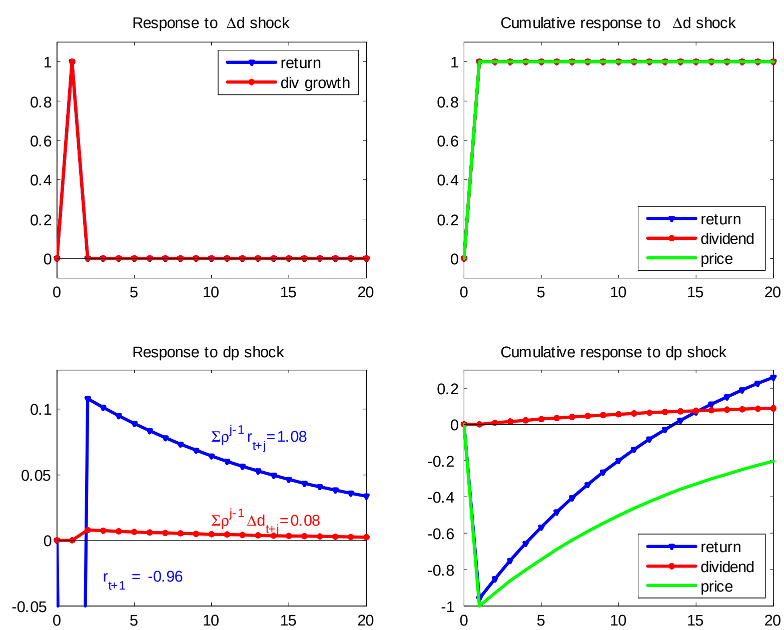
$$p_t - d_t \approx \kappa + E_t \sum_{j=1}^{k} \rho^{j-1} \left(\Delta d_{t+j} - r_{t+j} \right) + \rho^k (p_{t+k} - d_{t+k})$$

$$dp_t \approx -\kappa + E_t \sum_{j=1}^{k} \rho^{j-1} \left(r_{t+j} - \Delta d_{t+j} \right) + \rho^k (dp_{t+k})$$

$$\begin{aligned} 1 &= b_r^{(k)} - b_d^{(k)} + \rho^k b_{dp}^{(k)} \\ var\left(dp_t\right) &= var\left(p_t - d_t\right) \approx \sum_{j=1}^{k} \rho^{j-1} cov\left[dp_t, r_{t+j}\right] - \sum_{j=1}^{k} \rho^{j-1} cov\left[dp_t, \Delta d_{t+j}\right] + \rho^k cov(dp_t, dp_{t+k}) \end{aligned}$$

and the r terms are enough; no "rational bubbles."

10. Tool 4: VAR, impulse response, and "expected return" vs. "cashflow" shocks



11. Many other markets show the same pattern. A few:

    (a) Yield spread forecasts returns, not future rates; coefficients add to one.

    (b) Interest rate spreads forecast fx returns, not exchange rates; coefficients add to one. Yield spread forecasts returns, interest rate spread forecast fx returns.

    (c) Bond variables also forecast stocks. There seems to be a common time-varying risk premium.

12. Real world: Uses many more variables. Example: cay

$$R_{t+1} = a + b \times (D/P)_t + c \times cay_t + \varepsilon_{t+1}$$

Present value identity: variables help to forecast returns if they *also* help to forecast dividends or term structure of risk premia.

13. Warnings/Practical use
    (a) Fishing; Peso problems
    (b) Imprecise estimates.
    (c) Not useful for high speed trading!
    (d) "Buy in bad times," and miss bull markets.
    (e) *But* the point estimates are big.
    (f) View of the world is 100% different – *Expected return news not dividend news drives markets.*
14. Many other implications
    (a) CAPM. Assumes iid returns.
    (b) Cost-of-capital. $V = E(CF)/E(R) = E(CF)/\left[R^f + \beta E(R^m - R^f)\right]$
    (c) Marking to market assumes a random walk. "Accounting for future value" makes sense.
15. Some concept benchmarks
    (a) *Definition of Market Efficiency*
    (b) *Average investor theorem*
    (c) *Time-varying market risk premium – thinking about a market in equilibrium*
    (d) "*Forecast*" and use of regressions to forecast without causal interpretation