Predictability notes

Introduction

1970s view:

1. Expected returns don't move much over time – stocks are unpredictable.
2. Prices move on news of cashflow (dividend).
3. Capm works pretty well.
4. Beta derives from the covariance of cashflows with market cashflows.

All are dramatically different now.

1. Expected returns move a lot over time – stocks are predictable. (Long run, business cycle correlation)
2. Prices move on news of discount rate changes.
3. We understand the cross-section with multifactor models.
   (a) A larger number of characteristics other than beta are associated with expected returns
   (b) To the extent we understand those patterns, expected returns line up with non-market betas
4. Betas derive from the covariance of discount rates with market discount rates.
5. Facts are pushing us to the "risk premium" view of the world, as opposed to the "constant expected return, cashflow" view from the 1970s.
6. These are the *facts* underlying theoretical modeling.
7. Algebra is trivial. *Why* we do what we do and *what it means* are far from trivial.

Old Facts

•1965-1985 view: Expected returns are constant over time.

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

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

•Why we look at excess returns.

•These regressions are usually higher frequency – I just present this because it's easy with the data we have. Note R^2 and $\sigma(E(R))$ typically are even smaller at high frequency.

•Many thousands of signals have been evaluated. This was a revolution in 1970 since practitioners thought it obvious you can time markets.

• Basic "efficient markets" logic, just stems from competition. At high frequency, discounting seems unlikely to matter.

New View of facts

Facts

• Here are the facts from your readings. The regression got much better with the crash of 2008.

•Table 20.1

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

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

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

•Financial markets and the real economy update:

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

Table 1. 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. $R^e_{t\to t+k}$ denotes the total excess return from time t to time $t+k$. Standard errors use GMM (Hansen-Hodrick) to correct for heteroskedasticity and serial correlation.

•Barking dog update–this one shows logs, and returns vs. excess returns

Regression	b	t	$R^2(\%)$	$\sigma(bx)(\%)$
$R_{t+1} = a + b(D_t/P_t) + \varepsilon_{t+1}$	3.39	2.28	5.8	4.9
$R_{t+1} - R^f_t = a + b(D_t/P_t) + \varepsilon_{t+1}$	3.83	2.61	7.4	5.6
$D_{t+1}/D_t = a + b(D_t/P_t) + \varepsilon_{t+1}$	0.07	0.06	0.0001	0.001
$r_{t+1} = a_r + b_r(d_t - p_t) + \varepsilon^r_{t+1}$	0.097	1.92	4.0	4.0
$\Delta d_{t+1} = a_d + b_d(d_t - p_t) + \varepsilon^{dp}_{t+1}$	0.008	0.18	0.00	0.003

• From "Discount rates"

	$R^e_{t\to t+k} = a + b\frac{D_t}{P_t} + \varepsilon_{t+k}$				
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

Table 1. Return forecasting regressions $R^e_{t\to t+k} = a + b\frac{D_t}{P_t} + \varepsilon_{t+k}$ using the dividend yield. CRSP value weighted return 1947-2009.

•Motivation: do "low" prices mean / reveal high returns?

Statistical vs. economic significance

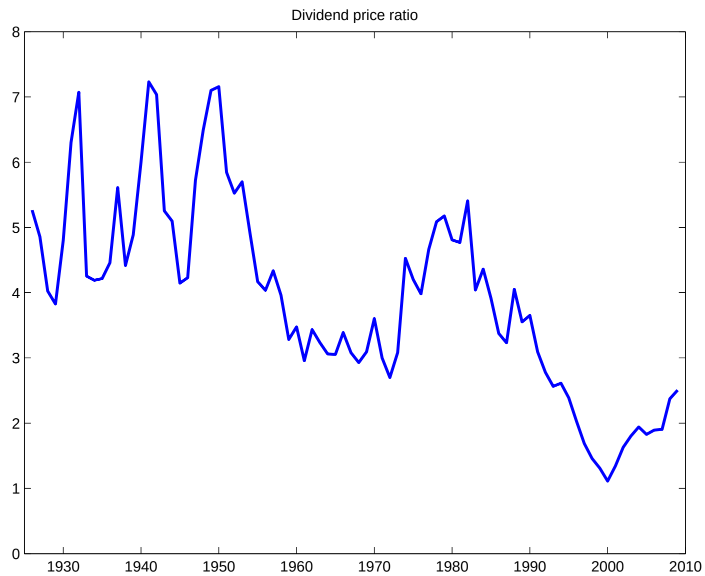
•The statistical significance is marginal and R^2 doesn't seem great. But the *economic* significance is huge, and R^2 is the wrong statistic. •One way to see the economic size of coefficients: Look at the one-year coefficient.

1. Fallacy: $b = 1$ in levels if you don't think prices adjust.

2. E(R) = constant predicts $b = 0$. (A high D/P means a low price. This should mean dividends will decline in the future so that $R_{r+1} = (P_{t+1} + D_{t+1})/P_t$ is not higher. Thus, high D/P *should* forecast low dividend growth).

3. The classic fallacy is right and more! This is one measure that the economic size of the point estimates is large.

• Another measure of economic importance. *Expected returns vary through time. A lot – **see D/P graph** and multiply by 2-5*



• "Predictability" $\leftrightarrow$ time-varying expected returns. The time varying expected return is what really matters, not how much ex-post is "predictable." We don't really care what fraction of ex-post return is predictable ex ante (R^2). We care about how much $E_t R_{t+1}$ varies over time.

•Another measure of economic size: The point estimates mean that $\sigma(E_t(R_{t+1})) = \sigma(bx) >$ is as large or more than the unconditional equity premium!

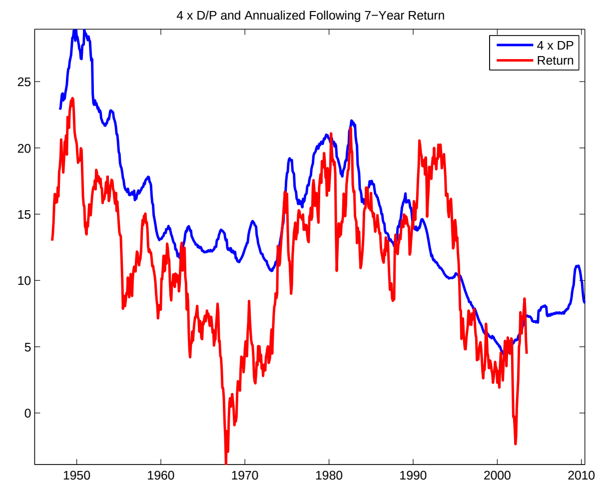
-$\sigma(E_t(R))/E(R)$ is more interesting than $R^2 = \sigma^2(E_t(R))/\sigma^2(R)$.(At any rate, it's a different question) The last table has that one (as I see the point more clearly over time.)

•The size of the coefficient again: In logs $b_r \tilde{}\, 0.10$. Keep that number in mind.

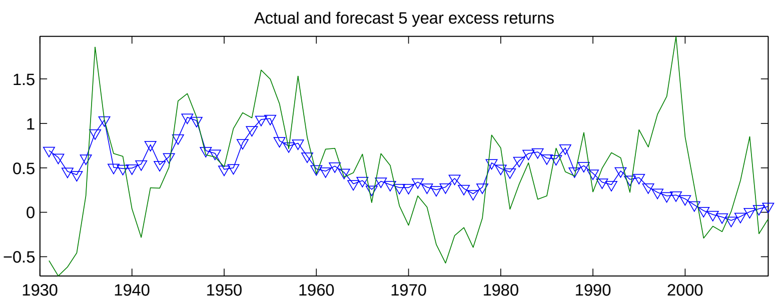
•*Excess* returns are forecastable (not shown here, but it's true). We're seeing a time-varying *risk premium.*

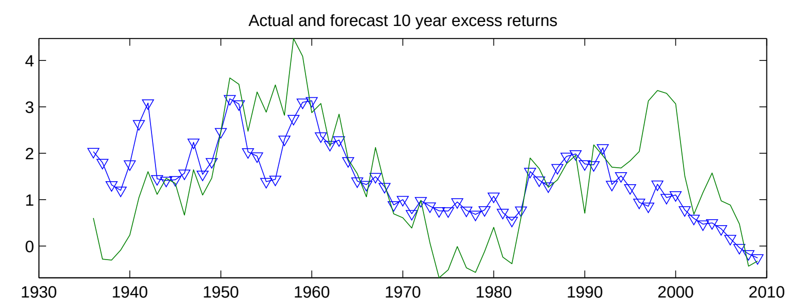
•Coefficients, R2 rise with horizon: Long-horizon R^2 is another measure of economic size. Looking at long horizons is our first "*magnifying glass*" It lets us see that a very small predictability (at say a daily horizon) is in fact very economically important. We will see that momentum, and many other effects correspond to different ways of looking at very old anomalies to recognize that they are economically bigger than we thought.

• Two graphs to show long-horizon R^2. They emphasize the fundamental fact: *High prices, relative to dividends have reliably led to many years of poor returns. Low prices have led to high returns.*



Dividend yield (multiplied by 4) and following 7 year return. CRSP VW market index.





• The t stat is not that interesting, and does not seem to grow with horizon. We lose observations as fast as the horizon increases. We'll look at this issue more carefully later. I used to think long horizons were only economically but not statistically interesting. I now think the comments on *Asset Pricing* p. 395 are wrong, and long horizons are more powerful. (This is one point of the barking dog paper.)

•Return forecastability is a robust fact, and does not depend on dividends as a divisor. The main point is price relative to anything sensible. P/E, B/M (M=P*shares), work just as well. For example

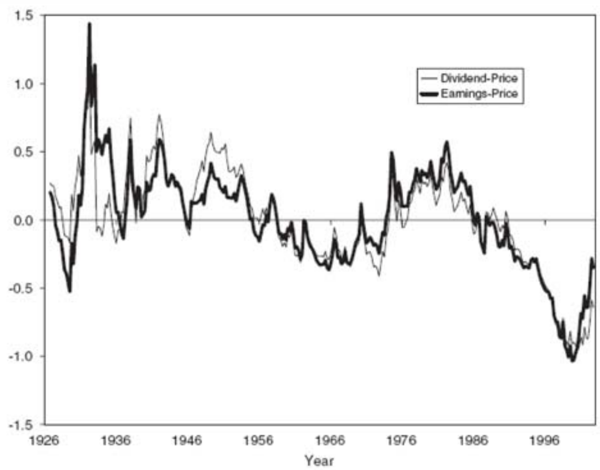


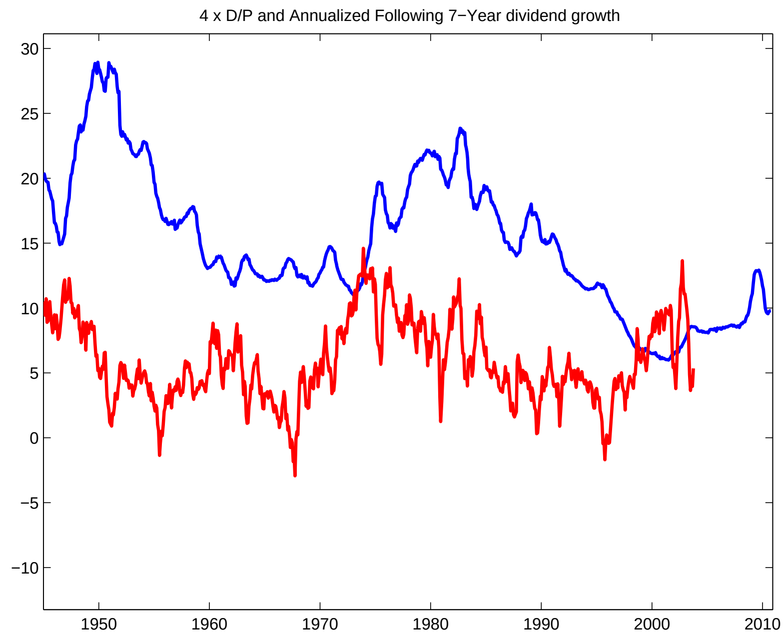
Fig. 1. Time-series plot of the valuation ratios. This figure plots the log dividend–price ratio for the CRSP value-weighted index and the log earnings–price ratio for the S&P 500. Earnings are smoothed by taking a 10-year moving average. The sample period is 1926:4–2002:4.

• I focus on D/P only for simplicity. There are *lots* of variables that forecast returns. Don't take the focus on D/P in the class as any endorsement that other forecasters are unimportant! In fact, the main area of research focus right now is on extending all this to extra forecasting variables, but without drowning in a soup of over fit multiple regressions.

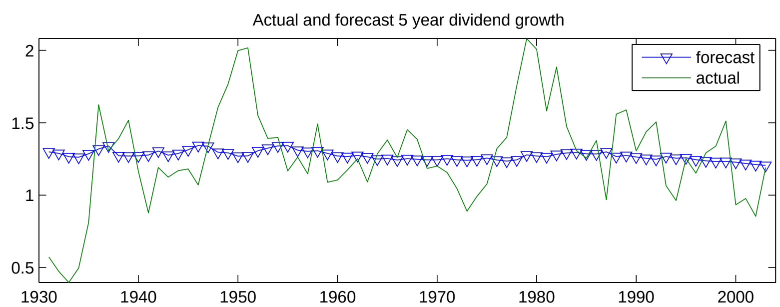
Dividend growth

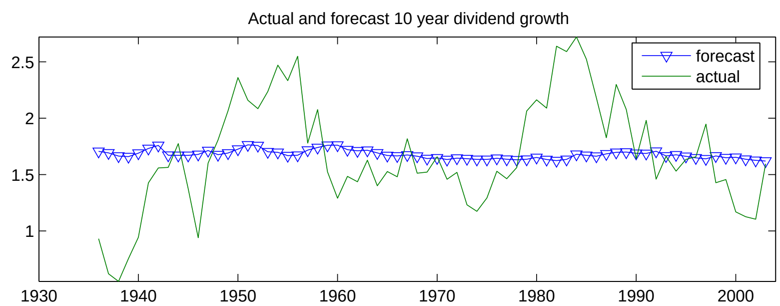
• Look at the tables. Dividend growth is not predictable! The point estimates are the "wrong" sign!

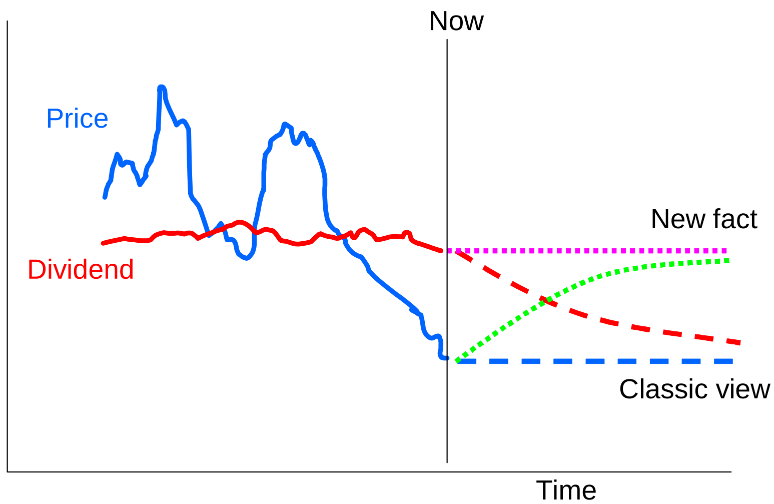
• P/D "*should*" forecast a dividend rise. Price high relative to current dividends should mean that future dividends will be higher.



DP and following 7 year dividend growth







• As you can see these are related issues – returns are forecastable because dividends are not. We will link these two forecasts in a minute.

Inefficiency?

•Does this mean markets are *"inefficient"*? Is this an invitation to "buy low and sell high?" Not necessarily. Time varying risk premia are possible.

Think like an economist, and think about market equilibrium, not trading opportunities. Prices must adjust to eliminate trading opportunities. People don't buy stocks because they're scared. Why at some times are they more scared than others?

•Are expected returns higher in good times or in bad times? (Bad, why?)

See dp plot and notice business cycle correlations. Who wanted to buy stocks in Dec 2008, despite good expected returns? (According to the Wall Street Journal, The U of C endowment sold.) Was this "irrational" fear?

So, business-cycle related time-varying risk premium is certainly possible (though not proved of course).

•This argument would be much harder to make if predictability came at high frequency, or were not so clearly associated with bad macroeconomic times.

•A *model* of time-varying risk premium / fad is the *only* way to tell the stories apart. Are the discount factors implicit in market prices connected correctly to marginal rates of transformation and substitution? This is the only answerable question. Naming residuals is fun, but gets you nowhere.

• Looking ahead, we need time-varying compensation for risk, or conditional heteroskedasticity of returns.

$$E_t(R^e_{t+1}) = R^f_t cov_t(R^e_{t+1}, m_{t+1}) = R^f_t \sigma_t(R^e_{t+1})\sigma_t(m_{t+1})\rho_t(m, R^e)$$

(The literature – i.e., Campbell and I – sort of concluded it had to be $\sigma(m)$ and $\sigma(m) \approx \gamma\sigma(\Delta c)$ means we need γ to vary over time. But the question is still open. Recently, there are signs that realized volatility or option-implied volatility does forecast returns, so perhaps $\sigma_t(R^e)$ might work again.)

TVER story to present value models.

•We started with return and dividend forecasting regressions, really thinking "does some ad hoc variable forecast returns?" The variable could have been weather or which football league won the super bowl (yes, both have been published.)

Our discussion has led to "what causes variation in *prices*? – cashflows, discount rates, or bubbles?" *Forecasting regressions are really about understanding how the right hand variable is formed.*

•The basic story is "third variable." News of E(D) or E(R) hits the market. Prices react – rising on good cashflow news or lower discount rate news. Prices *reveal* that news to us. On average, following such news, we see higher returns or dividend growth. What we learn from the regression is, which is it on average?

•Forecasting regressions do not have "cause" on the right and "effect" on the left. You can't get a sunny weekend by changing the weather forecast!

Our regressions are ok because forecast errors are orthogonal to forecasts. OLS did not assume "cause" on the right and "effect" on the left.

We often do that because we have effect = (cause x) + (other causes), and if (other causes) are orthogonal to (cause x) we have right hand variables orthogonal to errors. Certainly (cause x) = (effect) + (other cause) would be wrong, because effect is correlated with other cause. But that's not what we're doing here!

•We need present value models to make this story precise. How much do prices move when there is news of D or R?

One period "present value models"

•Identities linking returns to prices and dividends will turn out to give us a lot of intuition.

Our objective today: a quantitative tie from "expected returns rise" to price today, in order to tie empirical literature about expected returns to price variation. This present value model doesn't prove anything, but it connects ideas in subtle ways.

• In a one period model,

$$\begin{aligned} R_{t+1} &= \frac{D_{t+1}}{P_t}; E_t(R_{t+1}) = \frac{E_t(D_{t+1})}{P_t} \\ P_t &= \frac{E_t(D_{t+1})}{E_t(R_{t+1})} \\ \frac{P_t}{D_t} &= \frac{E_t(D_{t+1}/D_t)}{E_t(R_{t+1})} \end{aligned}$$

You can always discount using the ex-post return.

•Log are nicer so we can use linear time series methods,

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

$$p_t - d_t = E_t(\Delta d_{t+1}) - E_t(r_{t+1}). \tag{2}$$

This captures the idea PD is high if there is a) News about high future D b) News about low future R.

• *If p-d varies we do not live in an iid world.* No fancy statistics needed! If you look at returns or dividend growth separately, you might well conclude neither is predictable. But the fact that pd varies means this view is impossible! ("bubbles" introduce a footnote to this statement, but need an infinite period model, below.) In the 70s, the answer might have been, yes, and $E_t(\Delta d_{t+1})$ varies.

•This captures "why dividend growth should be predictable," and lets us put some equations to my story about trader information being revealed by prices: Suppose $E_t r_{t+1} = \bar{r}$. Then prices are formed by

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

But ex post,

$$\Delta d_{t+1} = E(\Delta d_{t+1}|\text{Trader information}_t) + \varepsilon_{t+1}$$

ε_{t+1} = a forecast error, which should be mean zero. Thus,

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

If agents see information which we do not see about dividend growth, we should be able to predict dividend growth with dividend yields.

• And of course the opposite works if there is a change in expected returns.

• Note forward-looking variables p *reveal agent information to us.* That's why they're so useful.

• In fact *If p-d varies, the return and dividend growth regression coefficients must add up to one.*

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

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

$$\begin{aligned} b_r &= b_d + 1 \\ 1 &= b_r - b_d \end{aligned}$$

Also

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

•This is all pretty obvious in words. All variation in return here comes from dividend growth – the only way you get a better return in a one period world is with more dividends. But it tells us important things.

•Price/dividend variation must reflect expected returns or dividend growth. Which is it? A: regressions say it's all R! In this sense return forecasts explain p-d variation . This is another measure of how "big" return forecastability is.

•Note: the identity holds expost, too, or using any information set. *It's just an identity.* You *can always discount cashflows with the asset's own return.*

• So why bother if there is no content? Identities allow us to *connect* price ideas to more conventional return ideas. You haven't really "explained" anything until you "explain" time varying expected returns. But seeing all sides of the coin via identities helps.

• A variable z_t that forecasts *both* expected returns and dividend growth equally will leave dp unchanged. So there is room for extra variables to forecast returns *and* dividend growth.

•Next Goal: do this with a real present value model.

A better present value identity.

Return identity

•Everything to follow comes from the **Campbell-Shiller linearization of the one-period return,**

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

with all symbols deviations from means.

Intuition: higher returns come from higher prices (higher valuations p-d), lower initial prices, or higher dividends.

•*Proof*

$$\begin{aligned}
R_{t+1} &= \frac{P_{t+1}+D_{t+1}}{P_t} = \left(\frac{P_{t+1}/D_{t+1}+1}{P_t/D_t}\right)\frac{D_{t+1}}{D_t} \\
r_{t+1} &= \log\left(1+\frac{P_{t+1}}{D_{t+1}}\right) - (p_t - d_t) + \Delta d_{t+1} \\
r_{t+1} &= \log\left(1+e^{p_{t+1}-d_{t+1}}\right) - (p_t - d_t) + \Delta d_{t+1} \\
r_{t+1} &\approx \log\left(1+\frac{P}{D}\right) + \frac{P/D}{(1+P/D)}\left[(p_t-d_t)-(p-d)\right] - (p_t - d_t) + \Delta d_{t+1}
\end{aligned}$$

Take out the constant, and change symbols to mean deviations from the constant.

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

$$\rho = \frac{P/D}{1+P/D} = \frac{1}{(1+D/P)} \approx \frac{1}{1.04} \approx 0.96$$

The constant can be the mean, but does not have to be the mean. This is important – we can apply the linearization cross-sectionally to different securities or portfolios which have different mean d/p.

CS PV Linearization

• **The Campbell-Shiller present value identity**

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

•*Derivation.* Iterate the return identity forward

$$r_{t+1} = \rho\left(pd_{t+1}\right) - pd_t + \Delta d_{t+1}$$

$$r_{t+1} + \rho r_{t+1} = \rho^2 pd_{t+2} - pd_t + \Delta d_{t+1} + \rho\Delta d_{t+2}$$

$$\sum_{j=1}^{k} \rho^{j-1} r_{t+j} = \rho^k\left(pd_{t+k}\right) - pd_t + \sum_{j=1}^{k}\rho^{j-1}\Delta d_{t+j}$$

$$pd_t = \sum_{j=1}^{k}\rho^{j-1}\Delta d_{t+j} - \sum_{j=1}^{k}\rho^{j-1} r_{t+j} + \rho^k\left(pd_{t+k}\right)$$

$$pd_t = \sum_{j=1}^{\infty}\rho^{j-1}\Delta d_{t+j} - \sum_{j=1}^{\infty}\rho^{j-1} r_{t+j} + \lim_{k\to\infty}\rho^k\left(pd_{t+k}\right)$$

“p “Price = dividends, discount, or bubble”

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

For now, assume that $p-d$ is bounded, so the last term converges, $\lim_{k\to\infty}\rho^k\left(pd_{t+k}\right)=0$. We'll look at that term more later.

•Again, we see high p/d if expected Δd is higher, r is lower

•*If both Δd and r are unforecastable, $p-d$ is constant.* If you look at statistics, it's easy to convince yourself that both Δd, r (and $p_{t+k}-d_{t+k}$) are unforecastable. They can't be. If p-d varies at all, something must be forecastable. *The fact that d-p varies means that we do not live in an iid world.* (Plus no bubbles, below.)

• (The alternative linearized identity in "explaining the variance of price-dividend ratios" has some advantages, but this one has taken over.)

•Names. Long run returns and dividend growth are

$$\sum_{j=1}^{k} r_{t+j}\ ;\ \sum_{j=1}^{k}\Delta d_{t+j}=d_{t+k}-d_t$$

Our measures are just ρ weighted,

$$r_t^{lr}\equiv\sum_{j=1}^{\infty}\rho^{j-1}r_{t+j};\quad \Delta d_t^{lr}\equiv\sum_{j=1}^{\infty}\rho^{j-1}\Delta d_{t+j}$$

so these are a measure of long run returns, dividend growth. I date them t for the beginning year, since there is no k. I actually think the weighted ones are more interesting: *Long horizon returns were interesting because they approximated these weighted long horizon returns, which tell us about prices.*

• CS versus one period models. The iterated equations looks just like 2 period equations; long run return = long run div growth + final price - initial price.

$$\begin{aligned}\sum_{j=1}^{\infty}\rho^{j-1}r_{t+j} &= \sum_{j=1}^{\infty}\rho^{j-1}\Delta d_{t+j}-(p_t-d_t)\\ \text{two-period:}\qquad & r_{t+1}=\Delta d_{t+1}-(p_t-d_t)\end{aligned}$$

Like 2 period, all we're doing is rearranging the definition of return to give a "present value" relationship. Then, take expectations (again it also holds ex post)Now we can write

$$p_t-d_t=E_t\left(\Delta d_t^{lr}\right)-E_t\left(r_t^{lr}\right)$$

The present value identity is exactly the same as the one-period identity (2)with long-run returns and long run dividend growth in place of the one-period return and dividend growth. Mapping dynamic problems into one-period problems is the central key to making finance look simple.

•The identity is an identity; it holds ex-post as well as with any information set.

$$\begin{aligned}p_t-d_t &= \sum_{j=1}^{\infty}\rho^{j-1}\left(\Delta d_{t+j}-r_{t+j}\right)\\ p_t-d_t &= E\left[\sum_{j=1}^{\infty}\rho^{j-1}\left(\Delta d_{t+j}-r_{t+j}\right)\mid \text{Any information set including } p_t-d_t\right]\end{aligned}$$

This is no more mysterious than $1 = R_{t+1}^{-1} R_{t+1}$ hence $1 = E\left(R_{t+1}^{-1} R_{t+1}\right)$ – the logic is the same.

•Thus, *the dividend yield reveals to us a slice of the investor's information set! If you see* $p_t - d_t$ you are *looking at* investor's expectations of $\Delta d_t^{lr} - r_t^{lr}$. Now we know why it's so good at forecasting. If there is some variation in $E(r_{t+1}|\text{Investor information})$, it's likely to get revealed by $p_t - d_t$. That's not a guarantee of course – variation in $E_t(r_{t+1})$ could get offset by variation in $E_t\left(\Delta d_{t+1}\right)$ or $E_t r_{t+2})$, and we'll see important practical cases in which this happens. But at least now you know why modern forecasting expeditions so often use prices or yields on the right hand side.

• Why the CS identity is such a useful tool. Think of the alternatives

$$\frac{P_t}{D_t} = E_t \sum_{k=1}^{\infty} \frac{1}{R_{t+1} R_{t+2} R_{t+3} .. R_{t+k}} \frac{D_{t+k}}{D_t}$$

$$\frac{P_t}{D_t} = E_t \sum_{k=1}^{\infty} \frac{\frac{D_{t+1}}{D_t} \frac{D_{t+2}}{D_{t+1}} \cdots \frac{D_{t+k}}{D_{t+k-1}}}{R_{t+1} R_{t+2} R_{t+3} .. R_{t+k}}$$

$$\frac{P_t}{D_t} = E_t \sum \frac{\Lambda_{t+k}}{\Lambda_t} \frac{D_{t+k}}{D_t}$$

etc.

a) You can see how CS is a linearization of 2. What's missing of course are $cov(\Delta d, R)$ terms. I don't find them very big.

b) Plus: the CS identity lets us make contact with the huge literature on expected returns rather than say Λ to calculate present values.

c) However, it's not obvious that it is the only useful way. for example, term structure models and affine extensions use the Λ approach, not the R approach.

Volatility

Volatility question

• A Related question: why do prices vary so much?

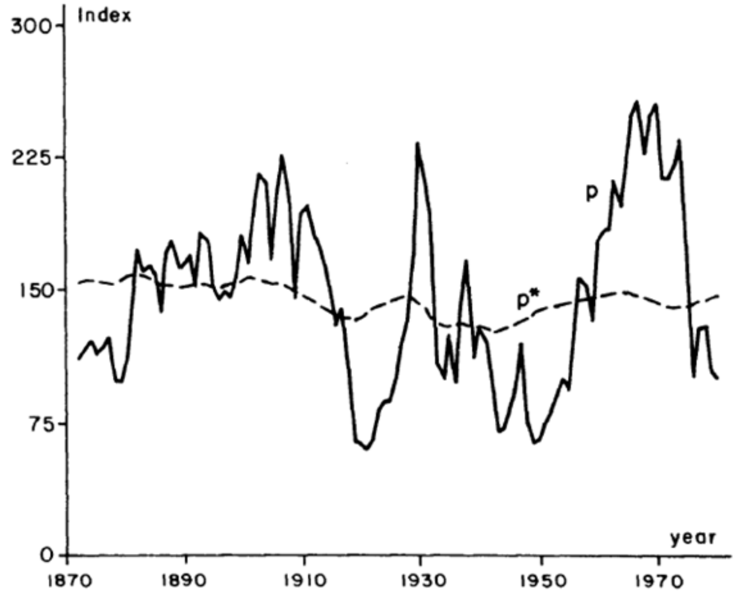


FIGURE 1
Note: Real Standard and Poor's Composite Stock Price Index (solid line p) and *ex post* rational price (dotted line p^*), 1871–1979, both detrended by dividing a long-run exponential growth factor. The variable p^* is the present value of actual subsequent real detrended dividends, subject to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

Shiller 1981 AER

• It seemed like another question, "you guys are looking in the dark. Sure, markets are not predictable, but prices vary by far more than is remotely believable that present values vary. If the market goes down 10% in a day, where is the crater into which 10% of US capital stock disappeared?"

• Related, "bubbles" are all the rage again. "price variation not related to fundamentals."

• Well, time varying discount rates help, but are they vaguely enough? Could changing discount rates possibly account for the vast variation in prices? A good lesson in writing some equations before you give up.

• Bottom line: *predictability and volatility are the same.*

• We need to do this right. Shiller took out "trends" in prices. The right answer: look at p/d (or p.e, or b.m or other ratio). Make "prices" stationary by finding a cointegrating vector. The identity:

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

Run long run regressions

$$\sum_{j=1}^{k} \rho^{j-1} r_{t+j} = a_r + b_r^{(k)} dp_t + \varepsilon_{t+k}^r$$

$$\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j} = a_d + b_d^{(k)} dp_t + \varepsilon_{t+k}^d$$

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

Plug regressions in the identity

$$dp_t = b_r^{(k)} dp_t + \varepsilon_{t+k}^r - b_d^{(k)} dp_t - \varepsilon_{t+k}^d + \rho^k \left(b_{dp}^{(k)} dp_t + \varepsilon_{t+k}^{dp} \right)$$

RESULT *The definition of return implies that coefficients and errors must obey*

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

$$0 = \varepsilon_{t+k}^r - \varepsilon_{t+k}^d + \rho^k \varepsilon_{t+k}^{dp}$$

• *This is an accounting of price volatility.* Multiply both sides by $var(dp)$ and

$$var(dp) = cov(dp_t, \sum_{j=1}^{k} \rho^{j-1} r_{t+j}) - cov(dp_t, \sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}) + cov\left(dp_t, \rho^k dp_{t+k}\right)$$

If dividend yields vary at all, then, they must forecast long-run returns, dividend growth, or a "rational bubble" of ever higher prices. These are numerators of regression coefficients. We can interpret b_d^{lr} and b_r^{lr} as "fractions of dividend yield variation accounted for by discount rate vs. expected cashflow variation"

• Your first "complementary regression" (Fama). Variation in dp is split between future dividend growth and returns. Mechanically, p-d must forecast one or the other.

•These really are the "right" coefficients with which to think about return predictability. Again, predictability regressions are really about what moves the right hand variable, not how to forecast returns, and it's long-run dividend or return that moves prices! Long horizons were not a debating point, they are the "right" way to run the regression in order to learn about price volatility.

• The units are now 1 / 0. 0.1 seemed a bit vague. Now we're looking for which b is 1 and which is 0.

• *Which is it*? Is return forecastability enough to account for price volatility?

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

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

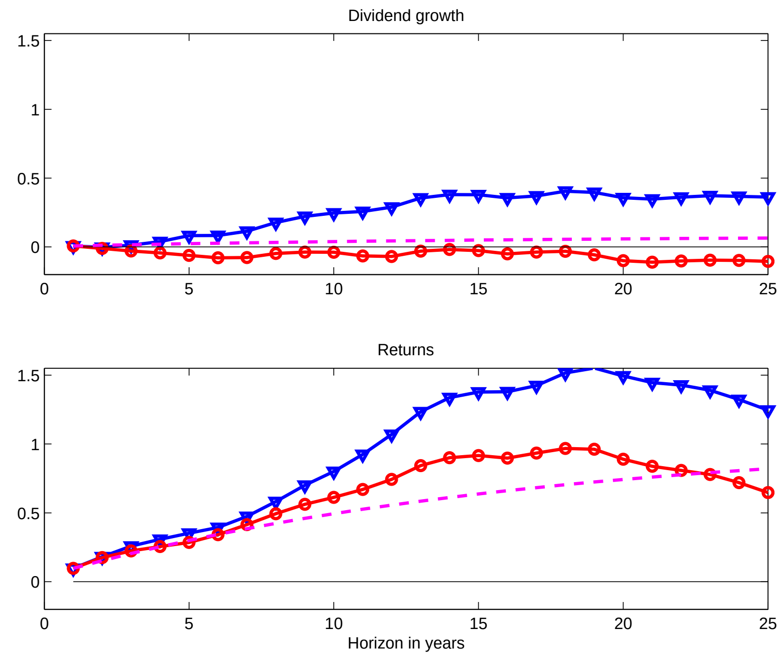
Table entries are the percent of the variance of the price/dividend ratio attributable to dividend and return forecasts, $100 \times \operatorname{cov}(p_t - d_t, \sum_{j=1}^{15} \rho^{j-1} \Delta d_{t+j}) / \operatorname{var}(p_t - d_t)$ and similarly for returns.

(Update from Cochrane RFS 1991)

AFA:

Left hand variable:	$\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$	$\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}$	$\rho^k dp_{t+k}$
Direct, $k=15$	1.01	-0.11	-0.11
VAR, $k=15$	1.05	0.27	0.22
VAR, $k=\infty$	1.35	0.35	0.00

Dog didn't bark updates



Regression forecasts of discounted dividend growth $\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}$ (top) and returns $\sum_{j=1}^{k} \rho^{j-1} r_{t+j}$ (bottom) on the log dividend yield $d_t - p_t$, as a function of the horizon k. Triangles are direct estimates: I form the weighted long-horizon returns and run them on dividend yields, e.g. $\beta\left(\sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}, d_t - p_t\right)$. Circles sum individual estimates: I run dividend growth and return at year $t+j$ on the dividend yield at t and then sum up the coefficients, e.g. $\sum_{j=1}^{k} \rho^{j-1} \beta\left(\Delta d_{t+j}, d_t - p_t\right)$. The dashed lines are the long-run coefficients implied by the VAR, e.g. $\sum_{j=1}^{k} \rho^{j-1} \phi^{j-1} b_d$.

•Summary,

-*All variation in the price dividend ratio corresponds to variation in expected returns. None corresponds to variation in expected dividend growth.*

- This finding is 100% different from classic view – 0/100 has become 100/0.

- *Magnitude question. Return predictability is "enough" to account for all var(p/d)* This is another measure of economic significance.

- How can expected-return variation possibly generate the huge price variation that we see? *Because it's so persistent. Small persistent expected return variation = large price variation.* See the identity – many little future r terms add up to a big effect on prices. This is the heart of many puzzles.

•Warning: This decomposition can be more/less than 100%. The terms are not orthogonal. It's not $dp = x + y$, $var(dp) = var(x) + var(y)$. In this case, the two terms are in fact perfectly correlated.

$$d_t - p_t = b_r^{lr} \times (d_t - p_t) - b_d^{lr} \times (d_t - p_t)$$

• What does 110% / -10% mean? Expected returns are a little too forecastable. When d-p rises, expected returns rise "too much". Hence expected dividend growth must also rise at the same time, enough to offset the "too large" rise in expected returns.

•Why three ways of computing it in the graph from "The Dog That Didn't Bark?" A direct estimate addresses, what if the VAR isn't right? - Answer: in this case there is no secret long run structure not seen in VARs. This is not always the case! Long run implications of short run VARs can be very wrong. See the "Random Walk in GNP."

• Note: alternative present value formulas. You can use real or nominal. You can separate interest rates from excess returns:

$$pd_t = \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$$

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

$$pd_t = \sum_{j=1}^{\infty} \rho^{j-1} \left(\Delta d_{t+j} - \pi_{t+j}\right) - \sum_{j=1}^{\infty} \rho^{j-1} \left(r^f_{t+j-1} - \pi_{t+j}\right) - \sum_{j=1}^{\infty} \rho^{j-1} \left(r_{t+j} - r^f_{t+j-1}\right)$$

We can apply this to real or nominal returns, and we can further divide, i.e. $r_t = rf_t + rx_t$. We can also find formulas for $p_t - x_t$ where x is earnings, book value, etc., by having $x - d$ on the right hand side.

Bubbles

• Bubbles are all the rage again. *What is a bubble?*

a) A price that went up, went down, and I wish I sold.

b) Seriously, now. "People buy because they think they can sell at a higher price." We can model this as a failure of the "transversality condition." These are "rational (almost) bubbles."

Preview: we'll find that the data are not easily consistent with rational bubbles. Proponents won't define bubble, so we'll have to interpret it as something else – discount rate variation that is not correctly connected to macroeconomics.

• A simple example. No dividends, constant expected return,

$$P_t = \frac{1}{R} E_t P_{t+1}$$

This means we expect prices to rise forever.

$$P_t = \lim_{k \to \infty} \left(\frac{1}{R^k} E_t P_{t+k} \right)$$

•A fun example:

$$P_{t+1} = \{\gamma R P_t, \text{ or } 1\}, \; \gamma > 1$$

Adjust γ and the probabilities so that $E_t P_{t+1} = RP_t$. The result, p.403.

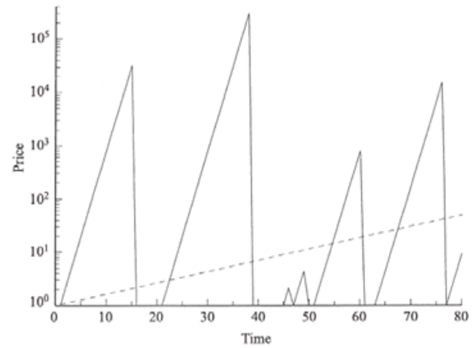


Figure 20.2. Sample path from a simple bubble process. The solid line gives a price realization. The dashed line gives the expected value of prices as of time zero, i.e., $p_0 R^t$.

Though prices are expected to grow forever, we see rises and crashes – just what bubbles advocates think. Once a bubble has crashed, then the world "starts over."

•Variation in prices without variation in fundamentals.

$$var(P_t) = var\left[\lim_{k\to\infty}\left(\frac{1}{R^k}E_t P_{t+k}\right)\right]$$

In our framework,

$$p_t - d_t = E_t \sum_{j=1}^{\infty} \rho^{j-1}\left(\Delta d_{t+j} - r_{t+j}\right) + E_t\left[\lim_{k\to\infty}\rho^k\left(p_{t+k} - d_{t+k}\right)\right]$$

$$var(p_t - d_t) = cov\left[p_t - d_t, \Delta d_t^{lr} - r_t^{lr}\right] + cov\left[p_t - d_t, \lim_{k\to\infty}\rho^k\left(p_{t+k} - d_{t+k}\right)\right]$$

It **is** *possible for prices to vary and yet not forecast returns or dividend growth, if there is a bubble. Then p-d variation all corresponds to varying expectations of terminal value.*

• In fact, we don't have to be so extreme, or worry about bubbles and whether anyone will be around to buy stocks in 4 billion years when the sun explodes.

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

Here we ask "*is the variance of dividend yields explained by its ability to forecast the first k years of dividend growth and returns?" If not, this question allows the remainder to be either "bubble" or even further-out "rational" variation.* The distinction between "expected dividends more years out than we can see" and "bubble" is philosophical anyway.

• In this context *the fact that var(p-d) is completely accounted for by the first term means there are no "rational bubbles." The first 15 years of return and dividend growth forecasts completely account for var(p-d)*

• Thus, if we're going to argue about "bubbles" and "irrational" markets, we are *only* talking about the interpretation of time-varying risk premia, not this kind of bubble. The only question is whether time varying expected returns are connected with marginal rates of transformation and substitution.

More "rational bubble" discussion:

- This isn't really right, since if there are bubbles we see we must have an explosive root in $p-d$, so it's not stationary and $var(p-d)$ doesn't exist. If $cov(p_t - d_t, p_{t+k} - d_{t+k})$ does not tend to zero, $p-d$ isn't stationary. The calculation is right under the null, but not under the alternative. One answer is to run unit root tests on p-d. They tend towards stationarity, but unit root tests are always dubious. In the second, finite-horizon decomposition,all variances and covariances are finite after conditioning on the first observation, so all terms of this are finite even under the alternative. Still, I would someday like to see the distribution theory for this bubbles test worked out right, both under null and alternative.

- There are a lot of theoretical objections to "rational bubbles" too

1) It requires E(p-d) , P to grow explosively – this violates statistics, common sense, economics. There really isn't a coherent "alternative" to work out.

2) The Transversality condition *is* a condition. A violation implies instant arbitrage in complete markets. (Short the asset, live on the dividends for free.)

3) Transversality is also a condition for optimality of a representative infinitely lived agent.

4) Thus, typically some "irrationality" needed for bubbles models. You cut off the recursion "I buy to sell in one period; he buys from me to sell in one period...." At some point the agent forgets to ask who is going to buy the asset

5) But.. OG models and some other theories do admit some "bubbles." Mostly "bubbles" theories – theories in which people buy assets above "fundamental value" – are based on some irrationality or overconfidence. (Scheinkman, Hong and Xiong are some of my favorites)

Final thoughts on volatility

• For a long time, the volatility debate was pointless: "if markets are efficient, why do they vary so much?" "Well, agents see information we don't see." "Oh that's silly."

Now we have a *test*! *If markets move on dividend information we don't see, we should still see that prices forecast dividends!* And the same for returns. We don't know what moved prices on a specific instance, but there is a testable implication of the idea that on average prices come from dividend information we don't see.

Volatility tests are not some new beast, they are the same as good old forecasting regressions.

• Result: P-d don't forecast dividend growth. But they do forecast returns – the world is consistent with the view of time-varying discount rates. And we reject the bubble view, that prices move without forecasting anything.

• In sum, the whole mess is down to *one fact: price-dividend variation forecasts long run returns and not dividend growth.*

• Why do so many people still believe in "bubbles" despite this evidence, that high prices lead to low returns?

A: They don't regard the expected return variation as "rational." In their view, var p-d *is* accounted for by variation in ER, but people *think* it comes from constant ER and prices that will rise forever. They redefine a "bubble" to mean "expected return variation that I don't agree with."

B: They willfully ignore the evidence that price-dividend variation is fully explained by return forecastability.

•In sum, we all (should) agree that high prices correspond to low returns, not to higher dividends or to a bubble; the question is whether low returns correspond to an accepted low risk premium, or whether people don't know about the low returns? *This is the central – only – fact in the "rational" vs. "irrational" "the crash shows markets are inefficient" "bubble" etc. debate.*

• And, we are back to empty residual-naming unless you have a predictive, rejectable model of the economic risk premium, or a predictive, rejectable model of when people psychologically over estimate returns and ignore the inevitability of price collapse.

Unify it all in a simple VAR

Estimates and identity

•Long-run estimates are much easier to calculate as implications from VARs than directly, i.e. by forming 15-year returns, etc. There is a danger here, that long-run implications of short-order VARs can be very wrong. If $x_{t+1} = \phi x_t + \varepsilon_{t+1}$, ϕ^{52} can be very far off. (Daily to annual). Examples I've run in to include the "Random walk in GNP," and the reconciliation of monthly and annual term-structure VARs in "Bond Risk Premia."

It turns out that *cointegrated* VARs tend to be pretty good at long-run implications, while regular ARMA processes tend to be far off (Again, that's my experience in dp/r vars as well as consumption-income VARs, see "Permanent and Transitory components."

In any case, we'll study long-run implications of short-run VARs and then check that direct estimates are at least not horribly different than the FARs.

•A basic VAR

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

(Normally you include $r_t, \Delta d_t$ and lags on the right hand side. I have found, and you'll verify in a problem set, that you can pretty much set the other coefficients to zero.)

•Motivation: This allows you to do long run calculations very easily, e.g.

$$E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} = b_r dp_t + \rho b_r \phi dp_t + \rho^2 b_r \phi^2 dp_t + ... = \frac{1}{1-\rho\phi} dp_t$$

•Estimates (problem set 2)

	Estimates		ε s. d. (diagonal,%) and correlation.		
	$\hat{b}, \hat{\phi}$	$\sigma(\hat{b})$	r	Δd	dp
r	0.108	0.050	19.8	0.67	-0.69
Δd	0.015	0.040	0.67	14.3	0.06
dp	0.937	0.042	-0.69	0.06	15.1

•Round numbers

	Estimates	ε s. d. (diagonal) and correlation.		
	$\hat{b}, \hat{\phi}$	r	Δd	dp
r	0.1	16-20	+big	-big
Δd	0		10-14	0
dp	0.94			15

-Including returns, dividend growth, further lags really don't change things much – see problem set.

-Implied dividends from the identity work very much like actual dividends.

-Including other variables *does* change things, so our results are conditioned on the $\{d-p, r, \Delta d\}$ information set. Many other variables help to forecast returns and dividend growth.

•Variance numbers

-The return shock variance is almost the same as the unconditional return variance, 16-20%. That's a lot by the way! Expected returns vary much less than the return shock, so the difference between shock and return variance is minor.

-Dividend growth variance at 14% is surprisingly large. Some of this is an artifact of data construction. I reinvest dividends paid through the year at the market return until December to make annual data. This is the right way to do it. If you just save up dividends or reinvest them at the interest rate, then the identity $R_{t+1} = (P_{t+1} + D_{t+1})/P_t$ does not hold in annual data! However, it means that dividends inherit half of a year's return volatility. Actual dividend payments are not so volatile. If you want to interpret D as actual amounts paid by firms to investors, you have to use the other definition, but then be aware that the return identity no longer works.

-The log dividend yield shock volatility at 15% implies that the variance of shocks to expected returns is about 1.5% per year.

$$\sigma\left[\left(E_{t+1} - E_t\right) r_{t+1}\right] = \sigma(b_r \varepsilon_t^{dp}) = 1.5\%$$

Identities in the first - order VAR

•It looks like we have three variables, but we really have two.

- The *variables* are linearly related

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

-Hence the *coefficients are errors* are linearly related,

$$b_r dp_t + \varepsilon^r_{t+1} \approx -\rho \left(b_d dp_t + \varepsilon^d_{t+1} \right) + dp_t + \left(b_d dp_t + \varepsilon^d_{t+1} \right)$$

Since this must be true for every value of dp_t,

$$\begin{aligned} b_r &= 1 - \rho\phi + b_d \\ \varepsilon^r &= -\rho\varepsilon^{dp} + \varepsilon^d \end{aligned}$$

-Thus, one equation is redundant – its coefficient, and error can be derived. Its *data* can be inferred.

•Conventionally, we run $\{r, dp\}$ and infer Δd. I like *thinking* of the system as $\{\Delta d, dp\}$ and inferring r. As you'll see, the Δd and dp shocks have nice economic interpretations, and they're basically uncorrelated. However, it's also a good idea when forecasting returns to really forecast returns, and not infer return forecasts from something else. The identities are not exact, and you might end up forecasting the approximation error not the actual return. (Voice of Hard Experience)

• The estimates (of course) satisfy the identities. $0.1 = 1 - 0.96 \times 0.94 + 0$. If you want to think about how the VAR could have come out, you have to specify numbers that satisfy the identities. For example, if you want $b_r = 0$, you either need $b_d = -0.1$ (dividend growth is predictable) or $\phi = 1.04$ (d-p is an explosive process, the "rational bubble" case).

• Of course we can ask, "how likely is it we see $b_r = 0.1, b_d = 0$ if the true world is $b_d = -0.1, b_r = 0$? That's "dog did not bark."

Impulse-response function in the simple VAR.

• Let's plot the responses to dividend yield and dividend growth shocks.

• It makes no sense to shock one of $\varepsilon^r, \varepsilon^{dp}, \varepsilon^d$ and leave the others alone. It violates the identity, and you can't get a return without price change and/or dividend change. And you can't change price, leave dividends alone and not change return.

• In particular, plot $\varepsilon^{dp} = 1, \varepsilon^d = 0$ and hence $\varepsilon^r = -\rho$, and then $\varepsilon^d = 1, \varepsilon^{dp} = 0$ and hence $\varepsilon^r = 1$.

$$\begin{array}{rl} \Delta d \text{ shock:} & \left[\begin{array}{ccc} \varepsilon^r & \varepsilon^d & \varepsilon^{dp} \end{array} \right] = \left[\begin{array}{ccc} 1 & 1 & 0 \end{array} \right] \\ dp \text{ shock:} & \left[\begin{array}{ccc} \varepsilon^r & \varepsilon^d & \varepsilon^{dp} \end{array} \right] = \left[\begin{array}{ccc} -\rho & 0 & 1 \end{array} \right] \end{array}$$

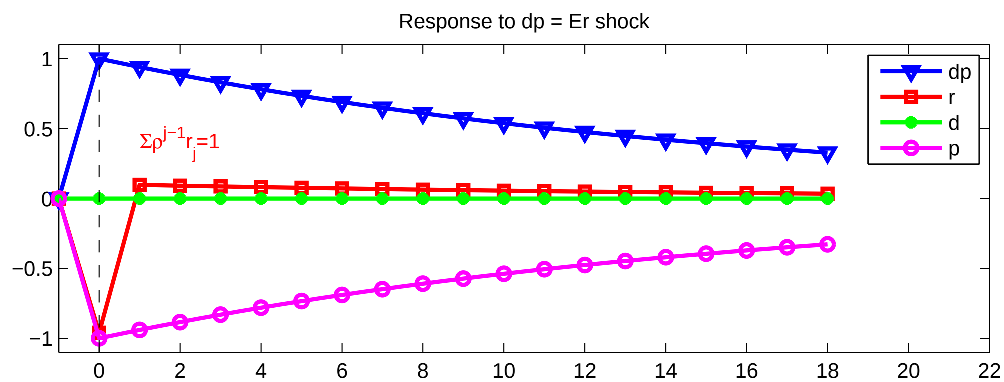
You can also define a "return shock" but then you have to say whether it came from a dd or dp change. You'll see why mine are a nice choice. Of course shock choice is arbitrary – pick what gives a nice story.

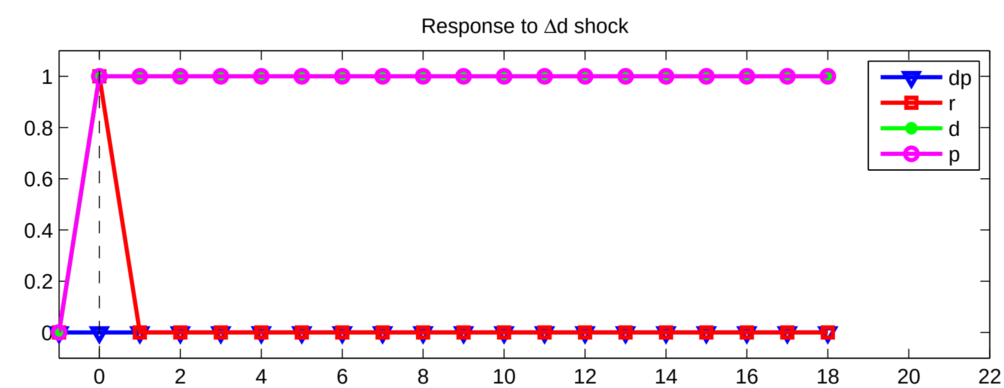
•Understanding the r response to each shock, especially the -ρ. Mechanically, a rise in dp with no change in d must mean a collapse in p. Economically, if we raise expected return without changing dividends, that means prices must collapse; we're discounting at a higher rate.

• Responses. There's no mystery, you can find responses analytically. Our system is

$$\begin{aligned}
\Delta d_{t+1} &= 0 \times dp_t + \varepsilon^d_{t+1} \\
dp_{t+1} &= 0.94 \times dp_t + \varepsilon^{dp}_{t+1} \\
r_{t+1} &= 0.1 \times dp_t + (\varepsilon^d_{t+1} - \rho \varepsilon^{dp}_{t+1})
\end{aligned}$$

The plots: ($b_d = 0$)





For intuition, I plot the response of $d_t = \sum_{j=1}^{t} \Delta d_t$. I also include "price" is computed easily as follows:

$$\begin{aligned}
\Delta p_t &= -(d_{t+1} - p_{t+1}) + (d_t - p_t) + \Delta d_{t+1} \\
b_p &= 1 - \phi + b_d \approx 1 - 0.96 = 0.04 \\
\varepsilon^p_{t+1} &= \varepsilon^d_{t+1} - \varepsilon^{dp}_{t+1}
\end{aligned}$$

$$\rightarrow \Delta p_{t+1} = 0.04 dp_t + \left(\varepsilon^d_{t+1} - \varepsilon^{dp}_{t+1}\right)$$

•The vertical line gives the period of the shock. Thus, movements on the vertical line are the other shocks that accompany the given one. Events to the right of the vertical line are changes in expectations about the future that come with the shock.

•In macro we often read impulse responses as "effects" of the shock. Technically, they only are "what revision of expectations of future variables coincides with the shock?" what $(E_{t+1} - E_t)\, x_{t+j}$ coincides with ε_{t+1}? In finance, it is useful to read a reverse causality, "what change in expectations about future returns and cashflows 'caused' price to change today?" Of course, that's dangerous too – the technique does not establish causality in either direction, and a behaviorist might want to read causality from the price shock today to the reversion (misunderstood by agents) which will follow. This interpretation is also vulnerable because agents surely see more than we do, and the interpretation conditions only on the $\{dp, r, \Delta d\}$ information set.

• Interpretation:

1. A positive dp shock with no change in dividends: (Blue) it slowly reverts, an AR(1). (Green) reflects no change in current (by assumption) nor future dividends. (Red) It comes with a big decline in *current* returns $\varepsilon^r = \varepsilon^d - \rho\varepsilon^{dp}$. With no Δd, return must come from a big change in p and hence $p - d$. Then it changes $E_t r_{t+1}$ rises by 0.1 $(= b_r)$, with slow ϕ^j decay. (The slope of p_t also shows the path of expected returns.)

   Since dividends are not forecastable, (random walk), if *current* dividends don't change, then *expected future* dividends don't change either.

   (a) A dp move *with no contemporaneous shock to dividends corresponds entirely to a change in expected returns.* We can call it an "expected return" shock. We often use present value logic and say dp moved *because* expected returns moved.

   (b) A rise in expected returns, with no change in dividends, must give a lower ex-post return, just as a rise in bond yield must mean a decline in the bond price.

   (c) p completely mean reverts. A price move with no contemporaneous move in dividends mean-reverts completely.

2. A Dividend shock with no change in dp: It is permanent, and has no effect on expected returns. Thus p must rise exactly as much as d

   (a) A dividend shock with no change in dp is a pure cashflow shock.

   (b) A price move with same move in dividends (and thus no change in d-p) is permanent

•Look at what we've done. *The VAR allows us to distinguish pure expected-return shocks and pure cashflow shocks, or more generally to say how much of each occurs at any moment. The VAR allows you to isolate changes in expected returns, as price changes with no change* *in expected future dividends. The VAR allows you to find a component of price movements that is completely transitory.*

• By contrast, if you just see a return, you don't know if dividends rose or not. Thus, you don't know if it's permanent or transitory. You have to look at dividends to know which it is. As you will see in the problem set and asset pricing, the response to a return shock (univariate) can be completely flat.

• There is nothing wrong with plotting responses to correlated shocks. It's nice that they are nearly uncorrelated but this isn't important to impulse-response functions. You need uncorrelated shocks for forecast error variance decompositions, but not for response functions. It turns out that ε^d and ε^{dp} are basically uncorrelated, so we're also plotting responses to uncorrelated shocks, but that is a minor convenience not an assumption.

• The current dividend is the Beveridge-Nelson trend or random walk component of price. It's the unforecastability of *dividends* that makes this so nice. If dividends were forecastable we could construct "expected return" and "expected cashflow" shocks, but they would not be so cleanly related to shocks with *current* zero responses.

• Above, I plotted the pure case $b_d = 0$, and no lags. Does this make a difference? is this a good approximation to the data? Here is the response with an estimated b_d.

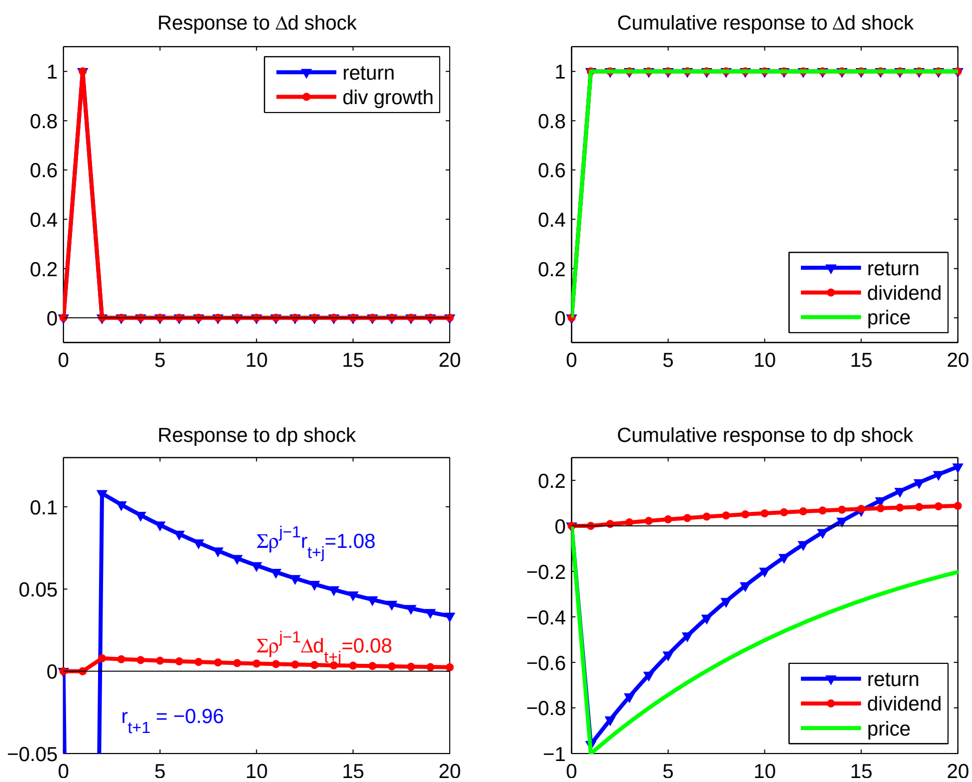


Figure 3:

As you can see, in the point estimate p-d is *almost* entirely an expected return shock. Actually returns are a little "too" predictable, requiring a little bit of dividend growth predictability in the "wrong" direction.

•Agenda. More variables? You can see that the cointegrating vector $p - d$ is central in the ability to make long-term forecasts. This suggests that in particular more cointegrating vectors will help. Consumption/income work for income, D/P works for returns, can we unite the two? Lettau and Ludvigson's cay is a start. By contrast, I suspect that the large zoo of variables that help to forecast one-period returns have very low ϕ and hence don't change the long-run picture much.

Shock correlations

• The return and dp shocks have strong negative correlation. I used to think this was weird when reading Campbell papers that focus on this correlation. Why not zero? In fact this strong negative correlation makes a lot of sense and is an important part of the worldview painted by the VAR.

•Remember the identity

$$\begin{gathered} r_{t+1} \approx -\rho dp_{t+1} + dp_t + \Delta d_{t+1} \\ \rightarrow \varepsilon^r = -\rho \varepsilon^{dp} + \varepsilon^d. \end{gathered}$$

The three shocks cannot be uncorrelated. There is nothing deep here, it's just the definition of return. For example, "correlation of return and dividend yield shocks" sounds forbidding. But if we are to get a good return, either prices have to rise, or dividends have to rise. You can't get a return without a change in either prices or dividends!

• The world we see: ε^{dp} and ε^d shocks are essentially uncorrelated. $\varepsilon^r = -\rho\varepsilon^{dp} + \varepsilon^d$ therefore means that ε^r and ε^{dp} must have a big negative correlation.

$$cov(\varepsilon^r, \varepsilon^{dp}) = -\rho\sigma^2(\varepsilon^{dp}) + cov(\varepsilon^{dp}, \varepsilon^d)$$

Thus, I like to digest the vital fact that return shocks are strongly negatively correlated with dividend yield shocks as a *consequence* of the fact that dividend growth and dividend yield shocks are essentially uncorrelated. As we saw above, ε^{dp} and ε^d are naturally "expected return" and "expected dividend growth" shocks. So we live (for once) in a nice world, one in which those "fudamental" shocks are nearly uncorrelated.

Given that fact the strong negative correlation of ε^r and ε^{dp} shocks is natural. It's just like the strong negative correlation of bond returns and bond yields. Expected returns rise, prices plummet (higher discount rates). This means the ex-post return plummets. These events are uncorrelated with dividend shocks.

•Other worlds. The only way for $\varepsilon^r, \varepsilon^{dp}$ to be uncorrelated is if $\varepsilon^{dp}, \varepsilon^d$ *are* correlated, i.e.

$$\begin{aligned} 0 &= -\rho\sigma^2(\varepsilon^{dp}) + cov(\varepsilon^{dp}, \varepsilon^d) \\ cov(\varepsilon^{dp}, \varepsilon^d) &= \rho\sigma^2(\varepsilon^{dp}) \end{aligned}$$

How could return and price-divided ratio shocks not be correlated? If lower expected return shocks (higher dp) happened to come at the same time as higher dividend growth shocks, (contemporaneously), the two effects would offset. But that's not what we see.

In the end, it's much prettier to have a world with uncorrelated dividend-growth and expected return shocks – with consequent negative correlation between dp and r VAR shocks, than the alternative.

•Similarly, the return and dividend growth shocks are strongly positively correlated. How could a big dividend growth not give you a good return? How could $cov(\varepsilon^r, \varepsilon^d) = 0$? It could only happen if expected returns rose, sending the price down at the same time, just enough to offset the increased dividend.

$$\begin{aligned} \varepsilon^r &= -\rho\varepsilon^{dp} + \varepsilon^d. \\ cov(\varepsilon^r, \varepsilon^d) &= -\rho cov\left(\varepsilon^{dp}, \varepsilon^d\right) + \sigma^2(\varepsilon^d) \end{aligned}$$

It turns out that's not the world we live in either.

•The VAR makes clear that we live in a two-shock world. Most models have one shock. For example, the Campbell-Cochrane habit model has one shock, consumption growth. Lower consumption growth raises expected returns. Thus, $corr(\varepsilon^d, \varepsilon^{dp}) = 1$ in that model. $corr(\varepsilon^d, \varepsilon^{dp}) = 0$ in the real world. A great paper topic is to extend that model to a two-shock world. There's a lot of important economics here. Risk premiums are generated by covariances. If ε^d is uncorrelated with m, then the dividend claim is risk free! $corr(\varepsilon^d, \varepsilon^{dp}) = 0$ means that dividend growth is correlated with *returns*, giving a standard cashflow-capm reason for a premium, but it means that dividend growth is uncorrelated with changes in *expected* returns. Thus, it is uncorrelated with the Merton-state variables that are also priced. Is this a key to the equity premium? ...

•In sum, as a result of the coefficient and shock identities, it's easy to remember what our world looks like with the two-variable VAR, in which you only have to remember one number,

$$\begin{aligned} \Delta d_{t+1} &= 0 \times dp_t + \varepsilon^d_{t+1} \\ dp_{t+1} &= 0.94 \times dp_t + \varepsilon^{dp}_{t+1};\ corr(\varepsilon^{dp}, \varepsilon^d) = 0. \end{aligned}$$

Then, the more complex return behavior, b_r and correlated shocks, *follows* from identities.

$$\begin{aligned} r_{t+1} &= b_r \times dp_t + \varepsilon^r_{t+1} \\ b_r &= 1 - \rho\phi + b_d = 1 - 0.94 \times 0.96 = 0.1 \\ \varepsilon^r_{t+1} &= -\rho\varepsilon^{dp}_{t+1} + \varepsilon^d_{t+1} \end{aligned}$$

Interesting alternative worlds, $b_d = -0.1$ (70s efficient markets), $\phi = 1.04$ (rational bubbles), $corr(\varepsilon^{dp}, \varepsilon^d) = -1$ (one-shock macro models such as Campbell-Cochrane) are then also easy to specify

State-space models and "structural" interpretation

•I used to be in love with these. After John Heaton has said about 15 times "I don't see what you're doing. The Wold representation is all there is," the lesson has finally sunk in – I don't think there actually is much deep in these "structural models." However, they're popular, in part from my own writing in *Asset Pricing*, so it's sensible to explain the idea and its limitations.

On the other hand, "state space" models are useful to constrain the parameters of VARs, and may give "smoothness priors" that let us incorporate the information in lags of return and dividend growth without exploding right hand variables (Koijen and van Binsbergen; Kelley)

•Here's the idea. Suppose expected returns vary slowly, and dividend growth is not predictable.

$$\begin{aligned} x_t &= \phi(=0.94)x_{t-1} + \varepsilon_t^x \\ r_{t+1} &= x_t + \varepsilon_{t+1}^r \\ \Delta d_{t+1} &= (0+)\ \varepsilon_{t+1}^d \end{aligned} \tag{3}$$

We also specify a correlation structure for the underlying ε shocks.

Look at the pretty model. x_t captures slow variation over time in expected returns. Actual returns add noise.

Agents see x_t =expected returns, and they see the structural shocks $\varepsilon^x, \varepsilon^r, \varepsilon^d$, but we don't. x_t is a "latent variable." (r_t is an ARMA(1,1) in its univariate representation, but that univariate representation does not reveal the shocks to the system – you can't recover x_t or $\varepsilon_t^x, \varepsilon_t^r$ just from a time series of $\{r_t\}$). Figure 4 illustrates. Notice how small expected return variation is compared to actual return variation – you'd have a hard time seeing the red line by filtering (taking moving averages) of returns! That's a big key in how observing dividend yields is so important.

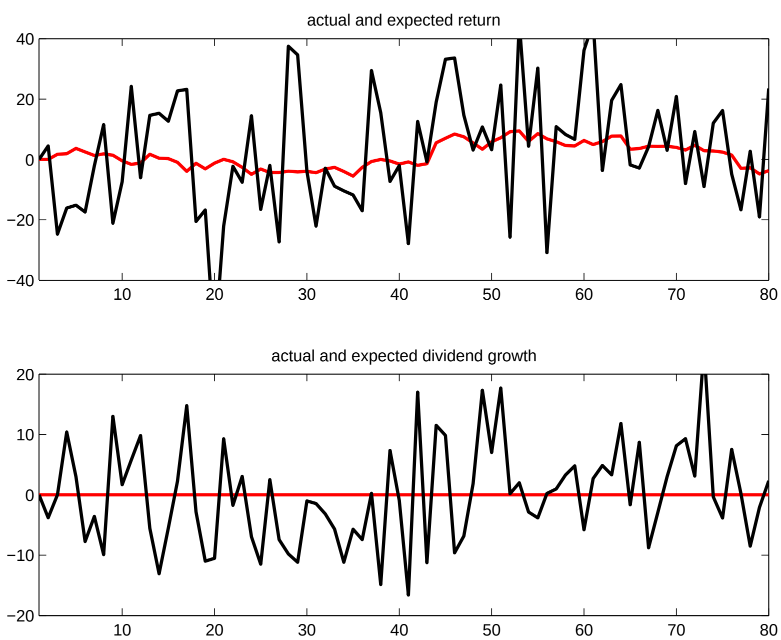


Figure 4: Actual and expected return, and dividend growth. Simulation.

•The question for us, is, what *observable* implications follow from this "structural" view of the world? We have a disadvantage – we can't see x and the shocks. But we can see dividend yields and these reveal to us a slice of investor's information.

Derive p-d from the present value formula:

$$dp_t = E_t \sum_{j=1} \rho^{j-1}(r_{t+j} - \Delta d_{t+j}) = \frac{x_t}{1-\rho\phi}$$

Here is the most important point. *p-d reveals* x, the expected return, that was previously only visible to agents.

•(This model for expected returns is common– it's the basis of all term structure models with $m_{t+1} = x_t + \varepsilon^m_{t+1}$. Then $x_t = E_t(m_{t+1})$ reveals the state variable x as the risk free rate.)

•Substituting, the *observable* dividend yield follows

$$dp_{t+1} = \phi dp_t + v^{dp}_{t+1};\; v^{dp}_{t+1} = \varepsilon^x_{t+1}/(1-\rho\phi)$$

The structural expected-return shock ε^x now is revealed to us as a regression error of dividend yield on lagged dividend yield, and the structural expected-return parameter ϕ is revealed to us as the dividend-yield autocorrelation ϕ.

•Note $1/(1-\rho\phi) = 10$. A small *persistent* ER error implies large price errors. This is a deep point. *Small but very persistent expected return variation can add up to large price variation.*

•Now, what's the observable process for returns? Substitute out x,

$$r_{t+1} = (1-\rho\phi)(d_t - p_t) + \varepsilon^r_{t+1}$$

So we predict a regression coefficient of returns on dividend yields

$$(1-\rho\phi) = (1 - 0.96 \times 0.94) \approx 0.1!$$

Hey, that's the right number! We also can now see the "structural" return forecast error ε^r_{t+1}

•Even easier, the observable dividend growth process is

$$\Delta d_{t+1} = 0 \times (d_t - p_t) + \varepsilon^d_{t+1}$$

•We have just *derived* the "reduced form regressions we expect to see given the above "structural" view of the world. The "structural" model (3) implies the "reduced form" or "observable" VAR

$$\begin{bmatrix} r_{t+1} \\ \Delta d_{t+1} \\ dp_{t+1} \end{bmatrix} = \begin{bmatrix} (1-\rho\phi) \\ 0 \\ \phi \end{bmatrix} dp_t + \begin{bmatrix} \varepsilon^r_{t+1} \\ \varepsilon^d_{t+1} \\ \varepsilon^{dp}_{t+1} \end{bmatrix}$$

and this is just about what we see!

•For example, here is a world we don't see: We could write down the "structural" model

$$\begin{aligned} r_{t+1} &= \varepsilon^r_{t+1} \\ \Delta d_{t+1} &= \varepsilon^d_{t+1} \end{aligned}$$

then,

$$d_t - p_t = E_t \sum_{j=1} \rho^{j-1}(r_{t+j} - \Delta d_{t+j}) = 0! \text{ (i.e. constant)}$$

We know this isn't true, since d-p varies. *We can't have both returns and dividend growth iid. If dp moves it must forecast something.*

•Here is another structural world we don't see: The 70's view that expected dividend growth moves

$$\begin{aligned} r_{t+1} &= \varepsilon^r_{t+1} \\ \Delta d_{t+1} &= x_t + \varepsilon^d_{t+1} \end{aligned}$$

In exactly parallel fashion this model predicts

$$\begin{bmatrix} r_{t+1} \\ \Delta d_{t+1} \\ dp_{t+1} \end{bmatrix} = \begin{bmatrix} 0 \\ -(1-\rho\phi) \\ \phi \end{bmatrix} dp_t + \begin{bmatrix} \varepsilon^r_{t+1} \\ \varepsilon^d_{t+1} \\ \varepsilon^{dp}_{t+1} \end{bmatrix}$$

•Furthermore, we now can interpret the VAR shocks as shocks to underlying, trader-information, expected returns and dividend growth.

d-p shocks are shocks to long-run expected returns

$$\varepsilon^{dp}_t = (E_t - E_{t-1}) \sum_{j=1} \rho^{j-1} r_{t+j} = \frac{1}{1-\rho\phi} \varepsilon^x_{t+1}$$

given the AR(1) structure, they are proportional to *shocks to one-period expected returns.*

ε^d *is a "cashflow" shock.* $b_d = 0 \rightarrow$ also an *"expected future cashflow" shock,*

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

• Now, the fact that VAR shocks $\varepsilon^d, \varepsilon^{dp}$ are uncorrelated is a result that comes from "structural" shocks to expected returns and dividend growth are uncorrelated. It didn't have to come out that way!

•In the return decomposition,

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

This is the same in the AR(1) as the Campbell return decomposition,

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

Again, "all price-dividend variation is due to expected returns" and "return variation is half dividend news and half expected return news" are consistent. Expected *future* dividends have nothing to do with current returns.

What's wrong

•What's wrong with this? Well, when you think about it, we really have not gotten anywhere away from the Wold representation; the return forecasts "expected returns" "expected dividend growth" and so forth are exactly what you recover from the VAR of observables. x_t merely gives a name to $b_r \times (d_t - p_t)$. It's convenient sometimes to use a symbol x_t in place of $b_r \times (d_t - p_t)$, but that's all we do really.

In fact, agents *don't* just see x_t; surely real-world expected returns and dividend growth are much more complex, and the $dp, r, \Delta d$ VAR we see is vastly "conditioned down." So pretending x_t is an AR(1) and "structural" really doesn't make much sense.

•The "revelation" of structural shocks in my example does not generalize. For example, life is not so simple if both dividends and returns contain predictable components.

$$\begin{aligned} x_t &= \phi x_{t-1} + \varepsilon_T^x \\ y_t &= \theta y_{t-1} + \varepsilon_t^y \\ r_{t+1} &= x_t + \varepsilon_{t+1}^r \\ \Delta d_{t+1} &= y_t + \varepsilon_{t+1}^d \end{aligned}$$

Then

$$dp_t = \frac{x_t}{1-\rho\phi} - \frac{y_t}{1-\rho\theta}$$

There is no perfect revelation any more. We can't back x and y out separately. $\{dp, \Delta d, r\}$ is no longer a first-order VAR, either. Lagged dividend growth can now help d-p to forecast, but in a structured way. (Koijin and Van Binsbergen (2009) do a sophisticated estimation of this case. Chaves (2009) points out you still can recover structural shocks if $\phi = \theta$.)

•More general. What is undoubtedly true about the world is that agents' expected returns x_t and expected dividend growth y_t follow a more complex process than an AR(1), and it's codetermined with a long vector of variables, $z_t = \begin{bmatrix} z_t^u & z_t^o \end{bmatrix}$ some observable by us z_t^o and some not observed by us z_t^u. Agents' information is $\Omega_t = \begin{bmatrix} x_t & y_t & z_t \end{bmatrix}'$ and if we want to model its evolution with linear time series it's something like

$$\begin{aligned} A(L)\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} &= \varepsilon_t \\ r_{t+1} &= x_t + \varepsilon_{t+1}^r \\ \Delta d_{t+1} &= y_t + \varepsilon_{t+1}^d \end{aligned}$$

Yes, dividend yields are still generated by the identity,

$$dp_t = E\left(r_t^{lr} - \Delta d_t^{lr} | \Omega_t\right)$$

Now, we can well ask of this structure, what is the implied Wold representation for observables? What process do

$$B(L)\begin{bmatrix} r_t \\ \Delta d_t \\ d_t - p_t \\ z_t^o \end{bmatrix} = v_t$$

where v_t are regression errors? We can even ask for reduced representations (less than full Wold, which requires infinite lags). What first-order VAR does the structure predict? But again, the inverse mapping doesn't work. There are many "structures" corresponding to the "reduced form." Without assumptions, you can't tell them apart – you can't learn more than the Wold representation.

For example, we have already seen that the simple AR(1) expected-return "structure" fits the $\{r, \Delta d, dp\}$ VAR, but the whole point of the cay investigation is that this structure, with more complex expected return and dividend growth, also "conditions down" to the same $\{r, \Delta d, dp\}$ VAR.

•Con: To recover "structure" we have to add identification assumptions. My above example already did that, by assuming an AR(1) for expected returns and no variation in expected dividend growth. That would be fine if the identification assumptions came from economics, but it's not clear to me that adding statistical identification assumptions – orders of lag polynomials, AR structure, shock covariance structure – has any basis in economics. Really, the most we can hope to know from linear time series methods is the Wold representation. You just can't get around this fact.

•Pro: it's clear from these examples that the "reduced forms" imply VAR structures with lags of r, Δd that matter. However, they don't come in with arbitrary coefficients. Smooth AR structures such as my x, y example above can generate "smoothness" restrictions on the VAR and help to see the extra forecast power of many lags without adding too many right hand variables.

Rise of coefficients, R2 with horizon

•Long horizons are not a separate phenomenon. Long horizons are a mechanical result of a persistent forecasting variable

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

$$\sum_{j=1}^{k} r_{t+j} = \left(1 + \phi + \phi^2 + .. + \phi^{k-1}\right)[d_t - p_t] + \varepsilon$$

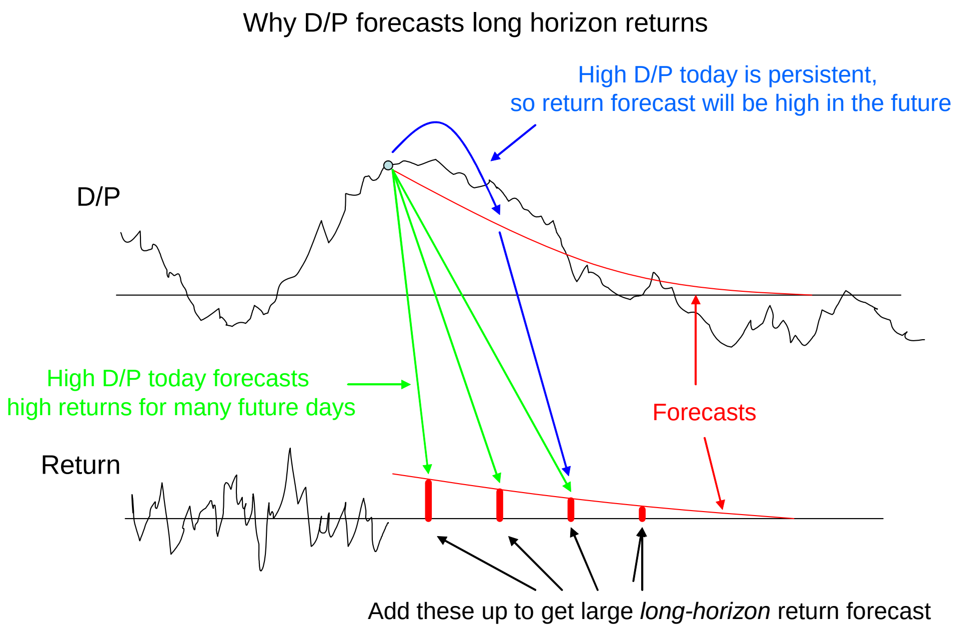
Coefficients rise over horizon if ϕ is large and near one.

• R^2 rise too in this model. It's a good exercise but too much algebra for lecture. Intuitively, the expected part $b \times dp_t$ builds linearly with horizon, so $\sigma^2(b \times dp_t)$ builds with the square of horizon. The unexpected part $\varepsilon_t + \varepsilon_{t+1} + ...$ is a sum of iid random variables, so $\sigma(\varepsilon)$ builds linearly with horizon. Thus the ratio builds with horizon. Temperature in Chicago is a good example. If you forecast temperature one day ahead based only on seasonal mean reversion, you explain next to nothing of daily variance, but a lot of seasonal variation.

Equivalently, long-horizon forecasts $\leftrightarrow$ *a low p/d predicts a high return for many periods in the future*

$$\begin{aligned} r_{t+2} &= b_{r2} dp_t + v_{t+2} \\ r_{t+1} + r_{t+2} &= (b_r + b_{r2}) dp_t + \varepsilon^r_{t+1} + v_{t+2} \end{aligned}$$

Picture:



•In sum, *long horizons are not a separate phenomenon* either. They are a way to make something we knew all long (significant t, low R2) seem much more economically important (large R^2). We will see "new telescopes" at work in many other areas, for example how portfolio formation lets us make the tiny serial correlation of returns economically significant.

• Do long horizons improve statistics? I used to think not – MLE is the one-period VAR, this is just intuition. The "dog that did not bark" argues otherwise, but mainly because at long horizons it's easier to build in the prior view $\phi < 1$.

Volatility tests in the VAR

• Recall

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

We interpreted b_d^{lr}, b_r^{lr} as "fraction of d-p variance accounted for by cashflow/expected return variation" (I standardized here on d-p rather than p-d as the right hand variable)

• In the VAR

$$b_r^{lr} = \sum_{j=1}^{\infty} \rho^{j-1}\phi^{j-1} b_r = \frac{b_r}{1-\rho\phi}$$

Thus, the volatility decomposition is simply

$$\frac{b_r}{1-\rho\phi} - \frac{b_d}{1-\rho\phi} = b_r^{lr} - b_d^{lr} = 1$$

You can get here much more quickly just from the identity

$$\begin{aligned} b_r &= 1-\rho\phi + b_d \\ 1 &= \frac{b_r}{1-\rho\phi} - \frac{b_d}{1-\rho\phi} \end{aligned}$$

But regression coefficients directly lose the $var(dp)$ interpretation which addresses Shiller, and the "forecasting long run return" implication.

• Simplified numbers

$$\begin{aligned} b_r^{lr} &= \frac{0.1}{1-0.94\times 0.96} = 1 \\ b_d^{lr} &= 0 \end{aligned}$$

Again, the return coefficient is just enough. The dividend coefficient is zero. No bubbles needed

•Real numbers (from above estimate)

$$b_r^{lr} = \frac{0.097}{1-0.96\times 0.941} = 1.0037$$

$$b_d^{lr} = \frac{0.008}{1-0.960.941} = 0.0083$$

•Intuition

1. The graph (from dog that didn't bark) above shows that first order VAR with dp is pretty good at capturing direct estimates of long-run pd forecasts.

2. The long-run return coefficient ≈ 1. The long-run dividend coefficient ≈ 0. These "should be" 0 / 1. Thus in this way we see another 0 - 100% that changed to 100% - 0% view of the world.

3. *These long run coefficients are the right way to think of return predictability.* Expected *long-run* return = 1.0 × d/p. Thus 1 , not 4.7 (R_{t+1} on D_t/P_t) or 0.1 (r_{t+1} on $d_t - p_t$) is is what I like to keep in my head. *Then*, 0.1 results from the autocorrelation of dividend yields. I think of it as

$$b_r = b_r^{lr}(1-\rho\phi)$$

not

$$b_r^{lr} = \frac{b_r}{1-\rho\phi}.$$

If, for example, dividend yields are really much more persistent, $\phi = 1$, I would expect that $b_r = 0.04$. That way the "100% variation due to dividend yields" would be preserved.

Campbell and Shiller variance decompositions.

• *Campbell variance decomposition* Campbell and Shiller use a different variance decomposition. They take variance of both sides of the present value identity, which means they have an ugly covariance term to deal with. To do the dividend yield variance decomposition, they use

$$\begin{aligned} dp_t &= E_t r_t^{lr} - E_t \Delta d_t^{lr} \\ var\left(dp_t\right) &= var\left(E_t r_t^{lr}\right) + var\left(E_t \Delta d_t^{lr}\right) - 2cov(E_t r_t^{lr}, E_t \Delta d_t^{lr}) \end{aligned}$$

They use a VAR to calculate the expectations.

•*Campbell in the Simple VAR.* In our simple VAR, we can do these analytically

$$\begin{aligned} E_t r_t^{lr} &= \frac{b_r}{1-\rho\phi} dp_t = b_r^{lr} dp_t \\ E_t \Delta d_t^{lr} &= \frac{b_d}{1-\rho\phi} dp_t = b_d^{lr} dp_t \end{aligned}$$

As you can see, the covariance term is going to be important – the two terms are perfectly correlated with each other! Thus, the decomposition is

$$var(dp_t) = \frac{b_r^2}{(1-\rho\phi)^2} var(dp_t) + \frac{b_d^2}{(1-\rho\phi)^2} var(dp_t) - 2\frac{b_r b_d}{(1-\rho\phi)^2} var(dp_t)$$

Canceling the $var(dp_t)$, we obtain a 3-way variance decomposition, which I can write.

$$1 = \frac{b_r^2}{(1-\rho\phi)^2} + \frac{b_d^2}{(1-\rho\phi)^2} - 2\frac{b_r b_d}{(1-\rho\phi)^2}$$

•*Campbell identity is also a rewriting of the VAR coefficient identity.* The coefficient identity is

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

I rewrote this as

$$1 = \frac{b_r}{1-\rho\phi} - \frac{b_d}{1-\rho\phi} = b_r^{lr} - b_d^{lr},$$

and multiplying both sides by $var(dp)$ gave it a "variance decomposition" interpretation. Campbell's identity is obviously the same thing, but he first squares both sides

$$1 = \left(b_r^{lr}\right)^2 + \left(b_d^{lr}\right)^2 - 2\left(b_r^{lr}\right)\left(b_d^{lr}\right),$$

and now multiplies by $var(dp)$.

•*Numbers and comparison.* Using typical numbers $b_r^{lr} = 1.1$, $b_d^{lr} = 0.1$, we obtain

$$\begin{aligned} 1 &= 1.1^2 + 0.1^2 - 2 \times 0.1 \times 1.1 \\ &= 1.21 + 0.01 - 0.22 \end{aligned}$$

so in the Cambpell decomposition we would say that "the variance of dividend yields is 121% explained by the variance of expected returns, 1% from the variance of expected dividend growth, and -22% from the covariance between the two." By contrast, my variance-covariance decomposition

$$1 = 1.1 - 0.1$$

would say it's "110% expected returns and -10% expected dividend growth." Neither is "right" or "wrong." They both say the same thing. Take your pick which you like.

•*Larger VARs* My variance-covariance decomposition is independent of a specific VAR – it only relates var(d-p) to the ability of d-p to forecast subsequent returns. (This will be estimated differently in sample by larger VARs, but the statistical concept is independent of which VAR you use). The variance-variance decomposition does change based on information set; it expresses (if you like that) or is affected by (if you don't) the fact that with larger information sets the variance of expected returns will grow. You can take the terms of

$$var(dp_t) = var\left(E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}\right) + var\left(E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}\right) - 2cov\left(E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j},\ E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}\right)$$

in bigger VARs. Since the left side is the same, *all* the terms of the right side increase, including the troublesome covariance term.

For this reason, I am more attracted to making this kind of calculation with *impulse-response functions*, as shown below in the context of cay.

CS return decomposition

•We can decompose the sources of *return* variation much as we decomposed the source of *price-dividend* variation. Start with the definition.

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

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

$$\left(E_{t+1} - E_t\right) r_{t+1} = \left(E_{t+1} - E_t\right) \left[\Delta d_{t+1} + \sum_{j=1}^{\infty} \rho^{j} \left(\Delta d_{t+1+j} - r_{t+1+j}\right)\right] \tag{4}$$

Return variation comes from 1) Current dividends 2) Expected future dividends 3) Expected future returns.

•The latter two effects, of course, come from their effect on future price-dividend ratios.

$$\begin{aligned} r_{t+1} &= -\rho dp_{t+1} + dp_t + \Delta d_{t+1} \\ (E_{t+1} - E_t)\, r_{t+1} &= (E_{t+1} - E_t)\left[\Delta d_{t+1} - \rho dp_{t+1}\right] \end{aligned}$$

•The negative sign. *An increase in expected return in the future means a lower ex-post return today.* Most people get this wrong; they think of an increase in expected returns as "good news." They understand that an increase in bond yields means a lower price, and hence a bad return today, but they can't apply this lesson to stocks. (This is the same effect we saw in the impulse response function. A rise in expected returns lowered today's return)

• *Chaves return variance decomposition.* We can obtain a variance decomposition in the same way as we did for d-p (Denis Chaves had this cool idea. As far as I know nobody has done this in print yet.). Let

$$\Delta E_{t+1} \equiv E_{t+1} - E_t,$$

then, multiply both sides of the return identity by $\Delta E_{t+1} r_{t+1}$ and take variances,

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

As with prices, this says "returns can vary only if there are shocks to current dividends, or if returns forecast changes in future dividends and returns (i.e. changes in dividend yields."

•These statistics are easy to calculate in a VAR. Results:

Current Dividends	Future Dividends	Future Returns
49.03	-7.81	54.80

About 50% from current dividends, 50 % from future returns, 0% from future dividends.

• People get confused. Is variation "All" expected returns or "50% returns 50% dividends?" *Price-dividend* ratios can move all on expected return news, while *returns* move on 50/50, because returns move on *current* dividend news as well. Thus, "100/0" for prices and "50/50" for returns are the same thing. It would be better to report results that separate the effect of current and future dividends, as I have.

We didn't have to get so complicated. Once you know that dividend growth is roughly iid with 14% standard deviation, you know the first term here will be big.

• Similarly

$$\Delta p_t = -dp_{t+1} + dp_t + \Delta d_{t+1}$$

so the *variance of price changes* and the *variance of price shocks* is each about half due to dividend growth, even if dp depends only on expected returns.

If you want to use language that keeps things straight, , say *valuation* depends on expected returns. This means dp.

•This decomposition depends on particular VAR we use. The last one was "conditioned down" on p-d as the only information set.

Calculation. I used the simple VAR with only $d - p$ as a state variable,

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

In the context of this VAR, (**??**) is the identity

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

The second substitution is so trivial it barely stands up on its own.

$$d_t - p_t = r^{lr}_t - \Delta d^{lr}_t$$

$$\begin{aligned} \varepsilon^{dp}_t &= (E_t - E_{t-1})\, r^{lr}_t - (E_t - E_{t-1})\, \Delta d^{lr}_t \\ &= \frac{b_r}{1-\rho\phi} \varepsilon^{dp}_t - \frac{b_d}{1-\rho\phi} \varepsilon^{dp}_t \end{aligned}$$

which is just the coefficient identity, but with a new intepretation. So, the return-innovation identity (4) reads

$$\varepsilon^r_t = \varepsilon^d_t + \frac{b_d \rho}{1-\rho\phi} \varepsilon^{dp}_t - \frac{b_r \rho}{1-\rho\phi} \varepsilon^{dp}_t$$

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 Return Decomposition.* For returns, Campbell, Campbell and Ammer express unexpected return variance in a different way, just by taking variance of both sides. Now we have three terms on the right, which potentially lead to two ugly covariance terms. Thus, customarily, they collapse the two dividend terms together,

$$var\left[\Delta E_{t+1} r_{t+1}\right] = var\left\{\Delta E_{t+1}\left[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j}\right]\right\} + var\left[\Delta E_{t+1} \sum_{j=1}^{\infty} \rho^j r_{t+1+j}\right] - 2cov(\cdot)$$

Mean Reversion

•Are stocks "Safer in the long run"? (Jeremy Siegel, http://www.jeremysiegel.com/) The fact of mean reversion certainly suggests it, no? "Temporary price variation" means "stocks are safer if you can wait"

• This is a good place to emphasize *Be very careful about information sets.* "Is there mean reversion in stocks?" "Are stocks safer in the long run?" "Can you predict stock returns?" All are very different depending on which information set you have in mind. Many authors get this completely wrong and confuse which information set they're talking about. For example, "dividends are not predictable" only refers to the $\{dp, r, \Delta d\}$ information set. Here, we compare the $\{dp, r, \Delta d\}$ information set with the $\{r\}$ information set, and get very different answers. The larger one conditions down to the smaller one. On a positive note, it's important to check that your view of the world conditions down appropriately. Looking forward, as we add new variables z, we want those results to condition down to what we've learned about $\{dp, r, \Delta d\}$

Univariate mean reversion and long run variances.

•*Definition*"Safer in the long run" means, is

$$var(r_{t+1} + r_{t+2} + ...r_{t+k}) < k \times var(r_{t+1})?$$

• *Big picture:* "Safer in the long run" (a variance question) is the same thing as "univariate mean reversion" or "univariate predictability" (an impulse-response or autocorrelation question)

• *Reminder.* You can always characterize time series equivalently by impulse-response (MA representation), forecasts from their past (AR representation), autocorrleation function, and variances (spectral density, really). Each view is connected to the others.

• *"Safety" and autocorrelations*:

$$r_t = \gamma_1 r_{t-1} + \varepsilon_t$$

$$var(r_{t+1} + r_{t+2}) = 2 \times var(r_t) + 2 \times cov(r_t, r_{t-1}) = 2 \times (1 + \gamma_1) \times var(r_t)$$

$$\frac{var(r_{t+1} + r_{t+2})}{2var(r_t)} = 1 + \gamma_1$$

Thus positive autocorrelation corresponds to variances for long run returns that grow faster than we think, and negative autocorrelation corresponds to long run variances that grow slower than we think.

Of course in regressions we found

$$r_{t+1} = \text{(very small number) } r_t + \varepsilon_{t+1}$$

This suggests essentially no mean reversion at all / no change in long vs. short run variances. . But maybe there's more structure than an AR(1)?

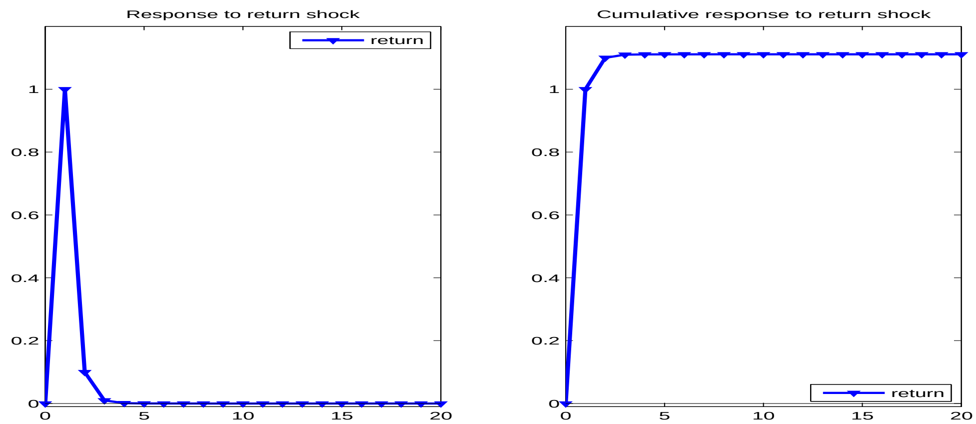


Figure 5: Response to return autoregression, $r_{t+1} = a + br_t + \varepsilon_{t+1}$

•*Long run autocorrelations and safety.* More generally, let $\gamma_j = cov(r_t, r_{t-j})$. Then,

$$\begin{aligned} var(r_t + r_{t+1} + r_{t+2} + .. + r_{t+k}) &= kvar(r) + 2(k-1)cov(r_t, r_{t-1}) + ...2cov(r_t, r_{t-k}) \\ \frac{var(r_t + r_{t+1} + r_{t+2} + .. + r_{t+k})}{k\ var(r)} &= \left[\sum_{j=-k}^{k} \frac{\|j\|}{k}\gamma_j\right] \end{aligned}$$

If returns are uncorrelated, we get the standard result, variance grows with horizon. Autocorrelation changes that; variances are proportionally lower if there is a long string of small negative autocorrelations ("mean-reversion") and larger if there is a long string of small positive autocorrelations ("momentum") An interesting number is the sum of autocorrelations.

$$\lim_{k\to\infty} \frac{var(r_t + r_{t+1} + r_{t+2} + .. + r_{t+k})}{k\ var(r)} = \sum_{j=-\infty}^{\infty} \gamma_j = S(0)$$

•*Connection to moving average representation.* In moving average form

$$r_t = a(L)\varepsilon_t$$

Recall that $a(L)$ is the univariate impulse-response function, and $\sum_{j=1}^{k} a_j$ gives the response of cumulative returns $\sum_{j=1}^{k} r_{t+j}$ to a shock ε_t. Then, a long string of small negative a_j would correspond to "mean reversion," the end of the impulse response function $a(1) = \sum_{j=0}^{\infty} a_j$ would be substantially below the impact value $a_0 = 1$. In this context, the variance of long run returns is

$$r_t + r_{t+1} + ... + r_{t+k} = \begin{array}{ccccc} & & \varepsilon_t & +a_1\varepsilon_{t-1} & +... \\ & \varepsilon_{t+1} & +a_1\varepsilon_t & +a_2\varepsilon_{t-1} & +... \\ \varepsilon_{t+2} & +a_1\varepsilon_{t+1} & +a_2\varepsilon_t & +a_3\varepsilon_{t-1} & +... \end{array}$$

$$\frac{1}{k}var(r_t + r_{t+1} + ... + r_{t+k}) \to a(1)\sigma^2(\varepsilon)$$

Thus, *"safer in the long run" variance questions are the same thing as the impulse-response question whether cumulative returns end up above or below the impact multiplier.*

• *Tests.* These equalities suggest a direct ways to measure the long-run multiplier question: look at long-run variances directly.

This is called the "variance ratio" test for long-run univariate mean reversion. These long-run autocorrelations (lots of small $a(L)$ for large L) are often captured poorly by low-order ARMA models. Those models are fit to capture one step ahead forecast error, not long-run properties of the series. (See "radom walk in GNP.")

- Fama and French long horizon regressions ($r_{t+1} + r_{t+2} + ...r_{t+k}$ on $r_t + r_{t-1} + r_{t-2} + ...$) are very similar, and another way to detect a string of long-horizon autocorrelations. See *Asset Pricing*

-Sharpe ratios. A fun way to scale the question. If returns are uncorrelated over time, i.e. no horizon effects, then mean and variance scale with horizons; standard deviation and Sharpe ratio scale with square root of horizon. This is a nice way to get directly at the Jeremy Siegel question, "are stocks safer for long-run investors" is to present the annualized Sharpe ratio not just the annualized volatility as above.

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

•Facts, Table 20.5, 20.6, 20.7 present these long-run estimates, which you now understand are all getting at the same thing.

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.

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

Logs: the formulas work better. Levels: One-period mean-variance investors care about arithmetic, not geometric returns. Surprisingly, there is *not* much consistent evidence that stocks are "safer in the long run." If they are, the magnitude of such stabilization is much lower than you might have thought given the multivariate evidence.

•We really should be looking at $\sum \rho^j r_{t+j}$, i.e. $a(\rho)$ not $a(1)$. Nobody has really investigated this question.

Solving a puzzle

• *The Puzzle.* How is it possible for d/p to forecast returns, but returns are not "safer in the long run?" How is it possible that we find a *completely* mean-reverting component to prices in response to a dp shock, but no mean reversion here? How is it possible that a rise in price *relative to dividend* is followed by low returns, but large *past returns* are not followed by low returns?

Conversely, perhaps we should be expecting "momentum." If we pass into a generation-long period of high expected returns, surely after a few years we should see an average of high actual returns that would tell us we're in a high expected-return regime. How could that fail to work?

Intuition 1: a high return is half likely to be a dp shock, which mean reverts, but also half likely to be a dividend growth shock which does not revert.

Intuition 2: It's not quite that easy. Univariate shocks (a return not predicted from past returns) are not the sum of multivariate shocks (a return not predicted by past dp, Δd, and past returns). This intuition would predict dampened mean-reversion, but would not let you see the correct possibility that the VAR is consistent with a pure random walk in returns, or even momentum behavior in the univariate representation. The other half of the intuition is, if you see a high return you don't know if it was expected or unexpected by dp.

We need to put equations to this intuition.

•*It's possible to have multivariate predictability and a univariate iid return.* A weather forecasting example: Suppose temperature is uncorrelated over time, but Minneapolis gets our temperature a day before us. Thus, you can't forecast temperature by looking at Chicago's temperature history (returns) but you *can* perfectly forecast temperature in Chicago by looking at temperature in Minneapolis. *Weather can be uncorrelated over time, yet forecastable.*

In equations: Is it possible for $b_r > 0$ but $\{b_j = 0\}$? A: Yes. Here is an example:

$$\begin{bmatrix} r_{t+1} \\ dp_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_t \\ dp_t \end{bmatrix} + \begin{bmatrix} 0 \\ \varepsilon^{dp}_{t+1} \end{bmatrix}.$$

This is just my weather forecasting story. It's clear but not our VAR.

•*Information sets.* This is about information sets. $E(\cdot|r_t, r_{t-1}...)$ is not the same as $E(\cdot|r_t, dp_t, r_{t-1}, dp_{t-1}, ...$ We are asking for the *univariate representation* of returns rather than the response to return shocks in a *multivariate* Wold representation.

• *Technique:* Given a multivariate view

$$\begin{bmatrix} r_{t+1} \\ dp_{t+1} \end{bmatrix} = \begin{bmatrix} . & b_r \\ . & \phi \end{bmatrix} \begin{bmatrix} r_t \\ dp_t \end{bmatrix} + \begin{bmatrix} \varepsilon^r \\ \varepsilon^{dp} \end{bmatrix}$$

Find the implied univariate Wold representation; predict what the VAR implies for

$$\begin{aligned} r_{t+1} &= \sum_{j=0}^{\infty} b_j r_{t-j} + v_{t+1}. \\ r_t &= a(L)v_t \end{aligned}$$

Put another way, suppose you simulate a long data series from the VAR, and then run the regression of returns on past returns. What would you get? You can always do this by simulation, but you can also do it analytically. This is not as easy as it looks, because v is not a linear combination of ε^r and ε^{dp}.

Univariate return process implied by our VAR

• *Univariate question.* Is this true of our VAR? Finding the univariate Wold representation implied by a VAR is in general a pain, because the univariate shocks are not linear combinations of the multivariate shocks. (The multivariate shocks are not in the space $v_{t+1} = r_{t+1} - a_0 r_t - a_1 r_{t-1}$...for some $\{a_j\}$.) *Asset pricing* shows how to do it analytically by factoring spectral density matrices. It's easy to do by simulation, of course: simulate a long history from the VAR and run a long autoregression.

The restricted VAR can be analyzed with pretty simple methods, and *it is very close to the case that the VAR implies a pure random walk in prices/returns unpredictable given their own past.* This is where we're going, and our point was to get the intuition behind this puzzle anyway.

• *VAR reminder* Our simplified VAR, as a reminder:

$$\begin{aligned}
\begin{bmatrix} r_{t+1} \\ \Delta d_{t+1} \\ dp_{t+1} \end{bmatrix} &= \begin{bmatrix} b_r \\ 0 \\ \phi \end{bmatrix} dp_t + \begin{bmatrix} \varepsilon^r_{t+1} \\ \varepsilon^d_{t+1} \\ \varepsilon^{dp}_{t+1} \end{bmatrix}. \\
b_r &= 1 - \rho\phi \\
\varepsilon^r_{t+1} &= -\rho\varepsilon^{dp}_{t+1} + \varepsilon^d_{t+1}; cov(\varepsilon^d, \varepsilon^{dp}) \approx 0
\end{aligned}$$

•*In this limited VAR, the univariate return process follows an ARMA(1,1).* Guess that the AR root is also ϕ. Then

$$(1-\phi L)r_{t+1} = b_r(1-\phi L)dp_t + \varepsilon^r_{t+1} - \phi\varepsilon^r_t = b_r\varepsilon^{dp}_t + \varepsilon^r_{t+1} - \phi\varepsilon^r_t$$

Since the right hand side is also an MA(1) (has autocorrelations of order 0, 1 and no more), we know that the univariate return process must be of the form

$$(1-\phi L)r_{t+1} = (1-\theta L)v_{t+1}$$

We can find θ and (if need be) σ_v^2 by matching the first two autocorrelations. Take the variance and autocovariance of both sides of this,

$$\varepsilon^r_{t+1} - \phi\varepsilon^r_t + b_r\varepsilon^{dp}_t = v_{t+1} - \theta v_t.$$

Our baseline VAR has a strong correlation $cov(\varepsilon^r, \varepsilon^{dp}) < 0$, and $cov(\varepsilon^d, \varepsilon^{dp}) \approx 0$.Thus, it's easier to parameterize this in terms of dp and d shocks,

$$\begin{aligned}
\left(-\rho\varepsilon^{dp}_{t+1} + \varepsilon^d_{t+1}\right) - \phi\left(-\rho\varepsilon^{dp}_t + \varepsilon^d_t\right) + (1-\rho\phi)\varepsilon^{dp}_t &= v_{t+1} - \theta v_t. \\
\varepsilon^d_{t+1} - \rho\varepsilon^{dp}_{t+1} + \varepsilon^{dp}_t - \phi\varepsilon^d_t &= v_{t+1} - \theta v_t.
\end{aligned}$$

•Special case for lecture: $\rho = \phi$, $cov(\varepsilon^d, \varepsilon^{dp}) = 0$

$$\varepsilon^d_{t+1} - \rho\varepsilon^{dp}_{t+1} + \varepsilon^{dp}_t - \rho\varepsilon^d_t = v_{t+1} - \theta v_t$$

$$\begin{aligned}
(1+\rho^2)(\sigma^2_{dp} + \sigma^2_d) &= (1+\theta^2)\sigma^2_v \\
-\rho\left(\sigma^2_{dp} + \sigma^2_d\right) &= -\theta\sigma^2_v
\end{aligned}$$

$$\rho = \phi = \theta.$$

The return is iid in its uivariate representation, exactly!

•It is *not* true that $v_t = \varepsilon_t^{dp} - \rho\varepsilon_t^d$. or $v_{t+1} = \varepsilon_{t+1}^d - \rho\varepsilon_{t+1}^{dp}$. The ρ goes in a different place. There's a "noninvertible" between univariate and multivariate shocks.

•*Admire* Stop and admire this result! There is a *completely* mean-reverting component to prices; a price move with no dividend change is *completely* transitory. Yet in the univariate representation, when we integrate over dividends as if we could not see them, returns are *completely* unforecastable, and changes in wealth are completely permanent! *Multivariate forecasts typically show much more structure than univariate ones.*

•*More general result*, still keeping $cov(\varepsilon^d, \varepsilon^{dp}) = 0$: The autocorrelations are

$$\begin{aligned} \left(1+\rho^2\right)\sigma_{dp}^2 + \left(1+\phi^2\right)\sigma_d^2 &= (1+\theta^2)\sigma_v^2 \\ \rho\sigma_{dp}^2 + \theta\sigma_d^2 &= \theta\sigma_v^2 \end{aligned}$$

we don't really care about σ_v^2, so we can focus on θ,

$$\frac{\left(1+\rho^2\right)\sigma_{dp}^2 + \left(1+\phi^2\right)\sigma_d^2}{\rho\sigma_{dp}^2 + \theta\sigma_d^2} = \frac{1+\theta^2}{\theta} \tag{5}$$

You can see right away that we're going to get a large θ, midway between $\rho = 0.96$ and $\phi = 0.94$, thus putting us squarely in the extreme trouble range for ARMA(1,1) that the roots nearly cancel.

Unfortunately, at this point you have to find θ and σ_v^2 by solving a quadratic, which I leave to a problem set. (We'll look at some very intuition-producing special cases in a minute.) Well, this is better than factoring a spectral density matrix! Figure 6 gives the answer for our baseline VAR, $\theta = 0.9510$, very slightly larger than $\phi = 0.94$.

•As you can see, *there is very little univariate mean-reversion despite the strong multivariate mean-reversion.* I will argue in fact that the estimates are very close to the case of uncorrelated returns, with $a(1) = 1$, but we'll get there slowly. For now think of this as an example of how ARMA(1,1) behave

•If $\phi = \rho = 0.96$, we then have $\theta = \phi = \rho$ and we have exactly the case of a pure random walk in returns. If $\phi > \rho$, then θ will be between ρ and ϕ, and slightly less than ϕ, actually a little bit of univariate "momentum." Estimates of ϕ in small (80 year) samples are quite downward-biased, so in fact bias-corrected estimates of ϕ are 0.96 or even larger. Thus, I think our world is actually even closer to the case that there is no univariate mean-reversion at all, and possibly a little bit of "momentum" as well, than the baseline VAR estimates I have been using imply. Finally, $cov(\varepsilon^d, \varepsilon^{dp}) = 0$ is not written in stone. Adding small amounts of correlation (such as the 7% observed in sample) can also push the above result up to the random walk or momentum camp.

•(Pastor and Stambaugh argue that Bayesian considerations move you even more to the "momentum" camp. Seeing a string of past positive returns changes your Baysian guess about mean returns, and that is just like momentum. Here I'm arguing you don't even need

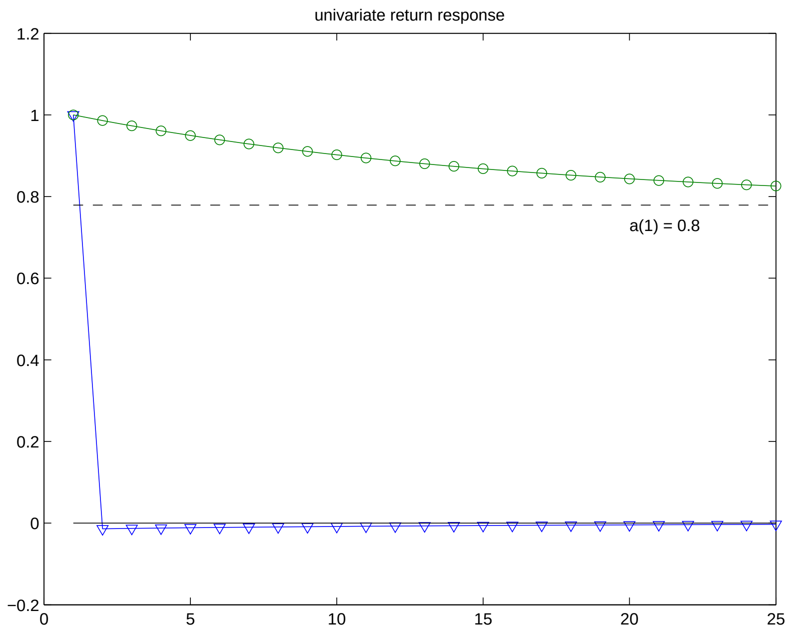


Figure 6: Univariate impulse response function for $(1 - 0.94L)r_{t+1} = (1 - 0.951L)v_{t+1}$, as implied by the restricted VAR. The blue line gives the response of returns, the green line gives the response of cumulative returns. a(1) gives the limiting value of the cumuative response.

that complication; pure old fashioned statistics say that the VAR implies close to random-walk return behavior.)

•Of course, we're here *understand* this result, so let's plunge in to a little deeper exploration *Objective: understand that near-random walk is a robust result of the basic facts, not a knife edge*

Digression on the ARMA (1,1)

• *State space model behind ARMA(1,1).* This is an instance of a more general and pervasive fact which you should know. *Suppose expected returns follow an AR(1),*

$$\begin{aligned} x_{t+1} &= \phi x_t + \varepsilon^x_{t+1} \\ r_{t+1} &= x_t + \varepsilon^r_{t+1} \end{aligned}$$

Then the univariate representation of r_t is an ARMA(1,1), with autoregressive root ϕ. This holds for $cov(\varepsilon^x_{t+1}, \varepsilon^r_{t+1}) \neq 0$. Strong negative correlation is in fact very important, as you have seen.

Proof. As before, match autocorrelations

$$(1 - \phi L)r_{t+1} = \varepsilon^x_t + \varepsilon^r_{t+1} - \varepsilon^r_t = v_{t+1} - \theta v_t$$

•*Filtering.* Since the ARMA(1,1) answers the question, "what is $E_t(r_{t+1}|r_t, r_{t-1}, r_{t-2}, ...)$?" it also answers the question "What is $E(x_t|r_t, r_{t-1}, r_{t-2}, ...)$? " we can view the whole question as, what can we learn about x (dp) by looking only at the history of r? You can view it as a filtering exercise

• *The ARMA(1,1,)* with θ near ϕ is a process worth looking at and understanding. The moving average representation is

$$\begin{aligned}
(1-\phi L)r_t &= (1-\theta L)v_t \\
r_t &= \frac{1-\theta L}{1-\phi L}v_t \\
r_t &= 1+\frac{(\phi-\theta)\,L}{1-\phi L}v_t \\
r_{t+1} &= v_{t+1}+(\phi-\theta)\left(v_t+\phi v_{t-1}+\phi^2 v_{t-2}+...\right)
\end{aligned}$$

Thus, the impact response is one, but then there is a very long tail of potentially very small following responses, all slightly positive or all slightly negative.

-The response of cumulative returns $\sum_{j=1}^{k} r_{t+j}$ is the cumulative sum of these responses. The limiting cumulative response is a good characterization of "mean-reversion" vs. "momentum,"

$$1+\frac{\phi-\theta}{1-\phi}=\frac{1-\theta}{1-\phi}$$

It is either less than one (mean reversion) or greater than one (momentum) depending on whether θ is greater or less than ϕ.

-(This kind of process also shows up in discussions of permanent vs. transitory movements, i.e. in GDP. If Δy follows an ARMA(1,1), then y may have undiscovered transitory components. In that context as well, I now think multivariate estimates far dominate univariate approaches.)

•*A particularly difficult process.* The ARMA(1,1) process is very hard to approximate with short-order AR models. If it's mean-reverting, an AR(1) for returns would give a slightly negative coefficient, and thus predict an oscillatory response function and also miss the mean-reversion in cumulative returns. An AR(1) for cumulative returns would fit the first few responses well, but then would decay to zero, predicting full mean reversion rather than stop at 0.8. An AR(1) for $(1-L)r_{t+1}$ would impose a long-run response of 1. Higher order ARs would not do any better

Anyway reality is not likely to be a pure ARMA(1,1) anyway. If reality has a lot of small positive or negative responses, adding up to substantial momentum or mean reversion, but that do not follow a simple low-order ARMA process, then you're really stuck. You can try a return to ML estimation and MA terms, now out of fashion in the VAR literature, but you're still stuck with the fact that ML wants to minimize one-step ahead forecast errors, and will happily sacrifice long-run forecasts to do so. You can try a return to "nonparametric" techniques like variance ratios.

•*Moving averages, filters.* If you must forecast an ARMA(1,1) or similar process from its own past, a very good approximation to the optimal forecast is to use a long moving average.

The actual forecast is this:

$$
\begin{aligned}
(1-\phi L)r_{t+1} &= (1-\theta L)v_{t+1} \\
\frac{(1-\phi L)}{(1-\theta L)}r_{t+1} &= v_{t+1} \\
\left[1+\frac{(\theta-\phi)L}{(1-\theta L)}\right]r_{t+1} &= v_{t+1} \\
r_{t+1} &= \frac{(\phi-\theta)}{(1-\theta L)}r_t + v_{t+1} \\
r_{t+1} &= (\phi-\theta)\left[r_t+\theta r_{t-2}+\theta^2 r_{t-2}\right]+v_{t+1} \\
E_t\left(r_{t+1}\right) &= (\phi-\theta)\left[r_t+\theta r_{t-2}+\theta^2 r_{t-2}\right] \\
(&= E(x_t|r_t, r_{t-2}, ...) \text{ for all you filtering fans})
\end{aligned}
$$

The long moving average of past returns captures well their information about future returns. This is the heart of Fama and French's idea to use long-run regressions in a univariate context. It also gives you useful (soon) intuition about these processes: in one with "momentum," a long stretch of above average returns forecasts that this pattern will continue. In one with "mean reversion," a long stretch of positive past returns means future returns will be lower

• *VARs are better.* But I think the best answer is to estimate the two variable VAR – which turns out to suffer from none of these problems – and find the implied univariate process. dp just has much much more information about future returns. Why would you ever throw out all that information? I'm conjecturing here that the implied univariate estimate from a VAR is much more accurate than a direct univariate estimate, especially about long-horizon features of the data – a(1). Moreover, why would any investor throw out information? To what question is "look at the univariate response function" the answer, when the multivariate data is available? Ask yourself why you care about the univariate process in the first place – as I argue later you don't.

More intuition for our case, and other possibilities

• Here is some intuition why we are very near $\theta = \phi$, and what the whole business means.

1. *Three limits in our dp, Δd VAR*

- Look at the $\sigma^2_{dp} = 0$ and $\sigma^2_d = 0$ limits. Naturally, these do not make assumptions about $corr(\varepsilon^d, \varepsilon^{dp})$.

$$
\begin{aligned}
\varepsilon^d_{t+1} - \rho\varepsilon^{dp}_{t+1} + \varepsilon^{dp}_t - \phi\varepsilon^d_t &= v_{t+1} - \theta v_t \\
\frac{\left(1+\rho^2\right)\sigma^2_{dp} + \left(1+\phi^2\right)\sigma^2_d}{\rho\sigma^2_{dp} + \theta\sigma^2_d} &= \frac{1+\theta^2}{\theta}
\end{aligned}
$$

a) *Only dividend growth shocks.* Suppose $\sigma^2_{dp} = 0$ – no expected return shocks. Then of course $\theta = \phi$, the roots cancel, and returns are just a pure random walk. It's even more obvious in the context of the original VAR; if we have $\varepsilon^r = \varepsilon^d$ and delete the dp row, then returns and dividend growth are both random walks. In this case we even learn the multivariate Δd shock by observing the univariate r shock. Of course, the potential for mean-reversion comes from the presence of varying expected returns.

However, we have $b_r \sigma_{dp} << \sigma_r^2$, the variance of expected returns is much smaller than the variance of returns. This is why we have $\theta \approx \phi$, the nasty case of nearly-canceling roots.

b) *Only expected-return shocks* Suppose instead $\sigma_d^2 = 0$. In this case stocks really are like bonds; there is no "cashflow" risk, there is only "expected return" risk. Thus, we should be able to see expected return shocks! Now we have

$$-\rho \varepsilon_{t+1}^{dp} + \varepsilon_t^{dp} = v_{t+1} - \theta v_t$$

Matching autocorrelations,

$$\begin{aligned}\left(1+\rho^2\right)\sigma_{dp}^2 &= (1+\theta^2)\sigma_v^2 \\ \rho\sigma_{dp}^2 &= \theta\sigma_v^2\end{aligned}$$

Thus, $\theta = \rho$, and returns follow

$$(1-\phi L)\, r_{t+1} = (1-\rho L)\, v_t$$

with $\rho = 0.96, \phi = 0.94$, this implies

$$a(1) = \frac{1-\rho}{1-\phi} = \frac{0.04}{0.06} = \frac{2}{3}$$

This is a substantial amount of mean-reversion; A return shock is 1/3 temporary.

Notice in this example that

$$corr(\varepsilon_t^r, \varepsilon_t^{dp}) = corr(\varepsilon_t^r, E_t r_{t+1} - E_{t-1} r_{t+1}) = -1$$

return shocks and expected return shocks are perfectly negatively correlated. I advertised this correlation is important.

However, this example is still puzzling. Why aren't return shocks completely transitory? There is no cashflow news to confound with discount rate news! Worse, if ϕ is larger, $\phi = \rho$, we still are back to the case that returns are a random walk, and with $\phi > \rho$, we have momentum in returns, even though there are no cashflow shocks; the *only* multivariate response is the above response to a dp shock. How can this possibly be?

Answer: *the univariate shocks are not the multivariate shocks*, even in this case that there is only one multivariate shock. This case makes it easy to see what I mean by that statement.

There are two ways to see this. First, use the ρ solution we just arrived at. Then, univariate and multivariate shocks must be related by

$$\begin{aligned}-\rho\varepsilon_{t+1}^{dp} + \varepsilon_t^{dp} &= v_{t+1} - \rho v_t \\ (-\rho + L)\,\varepsilon_{t+1}^{dp} &= (1-\rho L) v_{t+1} \\ \varepsilon_{t+1}^{dp} &= \frac{(1-\rho L)}{(-\rho+L)} v_{t+1} \\ \varepsilon_{t+1}^{dp} &= \frac{(1-\rho L)}{-\rho L\,(1-\rho L^{-1})} v_{t+1}\end{aligned}$$

the multivariate dividend yield shock is a function of future univariate return shocks. The dividend yield shock "reveals the future" of return shocks.

In the opposite direction, you can write

$$\begin{aligned}(-\rho + L)\,\varepsilon^{dp}_{t+1} &= (1-\rho L)v_{t+1}\\ \frac{(-\rho + L)}{(1-\rho L)}\varepsilon^{dp}_{t+1} &= v_{t+1}\end{aligned}$$

so v_{t+1} is a combination of current and past ε^{dp} shocks. But it's not an *invertible* function of these shocks, you can't undo this relationship and get the ε^{dp} shocks out.

Alternatively, go back to this point

$$-\rho\varepsilon^{dp}_{t+1} + \varepsilon^{dp}_{t} = v_{t+1} - \theta v_t$$

It's tempting to solve this with $v_t = -\rho\varepsilon^{dp}_t$ and $\theta = 1/\rho$ – and thus to argue that univariate and multivariate shocks are the same. But you can't do that – the Wold representation needs convergent sums in both AR and MA representations. The shock v_t needs to be a sum of current and past returns, not current and future returns.

Why are univariate and multivariate shocks different? Because even here, you can't tell whether a return was expected or unexpected in the multivariate representation. Even though d is constant, you don't see p the price. Thus, if you observe a 10% return, you don't know if this is a good shock, or if in fact we have a 20% dividend yield, and this is a bad shock. You could learn that if you could see future returns, but not just by seeing past returns.

You might object, "think of a bond. You can know the price just by looking at past returns." But this is a perpetuity, you never saw the "initial" price, and it's been there forever. So, while you can back out a bond price from its return data and knowledge of the initial price, that does not mean you can back out p here from the history of returns.

The second case below adds even more intuition for this specification

2. *Modeling returns and expected returns.*

You can get even more intuition by thinking of the univariate system as driven by an unobserved expected return x_t. Of course with $x_t = (1-\rho\phi)dp_t$ this is just notation, but it's useful to think about "the correlation of ex-post return and expected return shocks."

So, think about

$$\begin{aligned}x_{t+1} &= \phi x_t + \varepsilon^x_{t+1}\\ r_{t+1} &= x_t + \varepsilon^r_{t+1}\end{aligned} \qquad (6)$$

This is a very important class of process, and we'll see it over and over again, so it's important to understand how it works.

Proceeding as before, the univariate representation is

$$\begin{aligned}(1-\phi L)r_{t+1} &= (1-\theta L)v_{t+1}\\ \varepsilon^r_{t+1} - \phi\varepsilon^r_t + \varepsilon^x_t &= v_{t+1} - \theta v_t\end{aligned}$$

$$\begin{aligned}\left(1+\phi^2\right)\sigma_r^2+\sigma_x^2-2\phi\sigma_{rx} &= (1+\theta^2)\sigma_v^2\\ -\phi\sigma_r^2+\sigma_{rx} &= -\theta\sigma_v^2\end{aligned}$$

$$\frac{\left(1+\phi^2\right)+\frac{\sigma_x^2}{\sigma_r^2}-2\phi\frac{\sigma_{rx}}{\sigma_r^2}}{\phi-\frac{\sigma_{rx}}{\sigma_r^2}}=\frac{1+\theta^2}{\theta}$$

a) Suppose $corr(\varepsilon_t^r, \varepsilon_t^{dp}) = 0$, what if return and expected return shocks were uncorrelated? If you were new to this game or not paying attention, you might write this down as a place to start. This is a nice case because it displays *momentum.*

$$\frac{1+\phi^2+\frac{\sigma_x^2}{\sigma_r^2}}{\phi}=\frac{1+\theta^2}{\theta}$$

$$\frac{1}{\phi}+\phi+\frac{1}{\phi}\frac{\sigma_x^2}{\sigma_r^2}=\frac{1}{\theta}+\theta$$

$x+1/x$ is a declining function on (0,1), so $\theta<\phi$, $(1-\theta)/(1-\phi)>1$.

This is really nice, because it illustrates my intuition from above that we really expected momentum here, not mean-reversion. If we enter a period of high expected returns, you should see high actual returns after a while, and your experience of high actual returns should tell you about the high expected returns. Many papers have tried to filter return series and come up with such estimates, especially with a Bayesian flavor. We know it's fundamentally misguided, but the key will be to understand why.

The top left corner of Figure 7 plots r and x from this process. You should really understand how (6) works, and how it's a natural model for returns. You should also intuit that this one displays "momentum." If you see a period of high returns, it's clear that expected returns continue to be high. You can guess a lot about the red line without seeing it. (This is all about $E(x_t|r_t, r_{t-1}, r_{t-2}, ...)$) The puzzle really is how do we get *mean reversion* ever! How do we escape the logic that a period of high expected returns should be revealed by the experience of high ex-post returns?

b) Well, one aspect I've been stressing in our data is that *return shocks are very negatively correlated with expected-return shocks.* This is equivalent to the near zero correlation of expected-return shocks with dividend-growth shocks, via the identity. A rise in expected return, since it comes with no news about cashflows, must send prices down, $cov(\varepsilon^r, \varepsilon^x) << 0$.

So, let's explore the opposite possibility, that return shocks are perfectly negatively correlated with expected return shocks, $\sigma_{rx}=-\sigma_r\sigma_x$. This is also the $\sigma_d^2=0$ limit we explored a bit above.

Intuitively, you see how it has the potential to overturn the last result.

>>If expected returns rise, you get a huge negative shock to ex-post returns. Then you get many years of slightly higher (on average) ex-post returns. If the initial negative shock is large, and the subsequent expected returns not so large or long-lasting, then the average ex-post return during a period of high expected return will be negative, mean reversion. If the

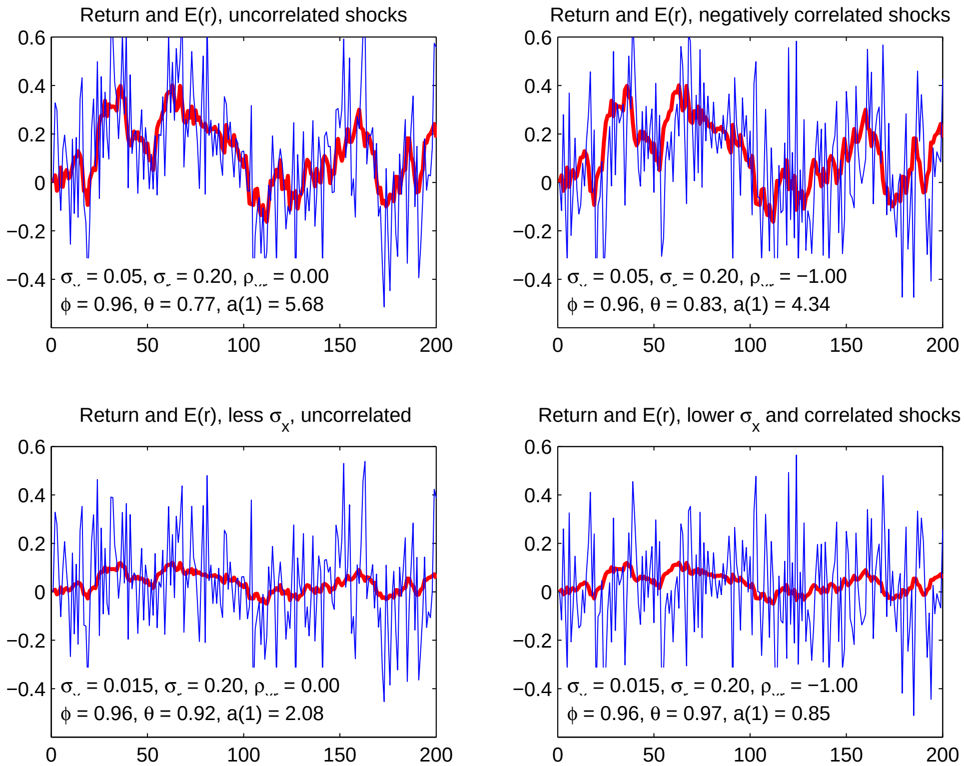


Figure 7: Simulations from $x_{t+1} = \phi x_t + \varepsilon^x_{t+1}$, $r_{t+1} = x_t + \varepsilon^r_{t+1}$. x is the red line, r is the blue line.

initial negative shock is medium, and the subsequent expected returns medium size and length, the average ex-post return during a period of high expected return will be zero – uncorrelated. If the initial negative shock is small or absent, and the subsequent expected returns are large and long-lasting, then the average ex-post return during a period of high expected return will be positive – momentum.

Let's put that in equations and pictures. Mechanically, we have now

$$\frac{\left(1+\phi^2\right) + \frac{\sigma_x^2}{\sigma_r^2} + 2\phi\frac{\sigma_x}{\sigma_r}}{\phi + \frac{\sigma_x}{\sigma_r}} = \frac{1+\theta^2}{\theta}$$

$$\frac{1+\left(\phi + \frac{\sigma_x}{\sigma_r}\right)^2}{\phi + \frac{\sigma_x}{\sigma_r}} = \frac{1+\theta^2}{\theta}$$

If $\phi + \frac{\sigma_x}{\sigma_r} < 1$ the solution is obvious

$$\phi + \frac{\sigma_x}{\sigma_r} < 1 : \theta = \phi + \frac{\sigma_x}{\sigma_r}$$

Again, this is $\theta > \phi$ so $(1-\theta)/(1-\phi) < 1$ or mean-reversion.

In fact, it doesn't take much σ_x/σ_r (volatility of expected returns relative to return shocks) to induce complete mean reversion! If $\sigma_x/\sigma_r = (1-\phi)$ (i.e. 0.06) then $\theta = 1$ and we have $a(1) = 0$!

For bigger σ_x/σ_r you need to be a bit careful. There is always a solution with root greater than one and a solution with root less than one. Note

$$\frac{1+\theta^2}{\theta} = \frac{1}{\theta} + \theta$$

so we always have two solutions. We need to pick the solution less than one, so if we get a really big σ_x, the answer is

$$\phi + \frac{\sigma_x}{\sigma_r} > 1 : \theta = \frac{1}{\phi + \frac{\sigma_x}{\sigma_r}}.$$

Now as σ_x/σ_r increases we get steadily less mean reversion. For $\sigma_x/\sigma_r > 2(1-\phi)$ we go back to uncorrelated returns and even momentum.

In the top left corner of Figure 7 I used a very large value of σ_x/σ_r so you could see the variation in expected return more clearly, and it would not get buried in an avalanche of ex-post returns. In the top right corner, I put the correlation of expected returns with ex-post returns to -1, as indicated. Surprisingly (to me) that's not enough – we still have momentum. Why? Well, we still have a very high σ_x/σ_r. The negative return shock can be no larger than σ_r/σ_x times the expected return-shock (correlations can't be greater than one), so though you can see rises in expected returns coming with declines in ex-post returns, those declines in ex-post returns aren't enough to offset the higher subsequent ex-post returns when expected returns rise. In the context of our formula, $\sigma_x/\sigma_r = 0.05/0.20 = 0.25$ which puts us back into momentum territory.

Clearly, to see mean-reversion we have to lower σ_x/σ_r. The variance of ex-post returns must be higher, to better obscure the information ex-post returns give you about expected returns. The bottom left panel of Figure 7 lowers σ_x/σ_r back to the value we have in the VAR, but returns to the uncorrelated-error specification. You can see clearly how having a larger variance of ex-post returns makes it harder to see variation in expected returns from the history of returns alone. But as shown, and as from the above equation, we're always in momentum-land with uncorrelated errors.

To get mean-reversion or uncorrelated returns, we need both a limited amount of σ_x/σ_r*, so that it is hard to learn about* x *from* r *before running out of* $\sigma/\sqrt{T}$*, and we need negative correlation of return and expected-return shocks.* I plot that case in the bottom right. Here I generate slight mean-reversion (I still have perfectly negatively correlated shocks). Obviously, we can get a pure random walk out of this setup with

$$\begin{aligned}
\phi &= \theta = \frac{1}{\phi + \frac{\sigma_x}{\sigma_r}} \\
\frac{\sigma_x}{\sigma_r} &= \frac{1}{\phi} - \phi = \frac{1}{0.94} - 0.94 = 0.12
\end{aligned}$$

At least visually, the bottom right panel is pretty clear – *average ex-post returns are no higher during periods of truly high expected returns than they are during periods of truly low expected returns.*

Our VAR

Now, why is it so robust that our VAR generates a near-random walk in the univariate representation? This seems like a knife-edge case when we look at the last set of results. Yet, when we look at the formula for our VAR

$$\frac{\left(1+\rho^2\right)\sigma_{dp}^2+\left(1+\phi^2\right)\sigma_d^2}{\rho\sigma_{dp}^2+\theta\sigma_d^2}=\frac{1+\theta^2}{\theta}$$

there is no wiggle room at all. θ has to be somewhere in between $\phi = 0.94$ (and downward biased) and $\rho = 0.96$.

1. The near-zero correlation of dividend growth and dp shocks means that *return and dp shocks are strongly negatively correlated.*

2.

$$\sigma_x^2 = (1-\rho\phi)^2\,\sigma^2(dp) = \frac{(1-\rho\phi)^2}{\left(1-\phi^2\right)}\sigma_{dp}^2.$$

$$\sigma_r^2 = \sigma^2\left(-\rho\varepsilon^{dp}+\varepsilon^d\right) = \rho^2\sigma_{dp}^2+\sigma_d^2$$

$$\frac{\sigma_x^2}{\sigma_r^2} = \frac{\frac{(1-\rho\phi)^2}{\left(1-\phi^2\right)}\sigma_{dp}^2}{\rho^2\sigma_{dp}^2+\sigma_d^2} = \frac{\frac{(1-\rho\phi)^2}{\left(1-\phi^2\right)}}{\rho^2+\frac{\sigma_d^2}{\sigma_{dp}^2}}$$

But $\frac{(1-\rho\phi)^2}{\left(1-\phi^2\right)}$ is a small number. *Even if dividend growth is constant, expected return shocks are small compared to actual return shocks.* Why? *because ϕ is a large number – because expected return shocks are long lasting.*

$$\begin{aligned}
\varepsilon_{t+1}^r &= \left(E_{t+1}-E_t\right)\left(\Delta d_{t+1}+\sum\rho^{j-1}\Delta d_{t+j}-\sum\rho^{j-1}r_{t+j}\right)\\
\varepsilon_{t+1}^x &= \left(E_{t+1}-E_t\right)r_{t+1}
\end{aligned}$$

Far from a knife-edge, these are very robust features of the data!

Multivariate

The "guess the AR root" trick can be extended to a multivariate system. Suppose we have

$$\begin{aligned}
r_{t+1} &= b_r' x_t + \varepsilon_{t+1}^r\\
x_{t+1} &= \Phi x_t + \varepsilon_{t+1}^x
\end{aligned}$$

we can eigenvalue decompose x_t, e.g.

$$\Phi = Q\begin{bmatrix}\phi_1 & 0\\ 0 & \phi_2\end{bmatrix}Q^{-1}$$

and thus wolg write the system in terms of diagonal x

$$\begin{aligned} r_{t+1} &= b_{rx}x_t + b_{ry}y_t + \varepsilon^r_{t+1} \\ \begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} &= \begin{bmatrix} \phi_x & 0 \\ 0 & \phi_y \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \begin{bmatrix} \varepsilon^x_{t+1} \\ \varepsilon^y_{t+1} \end{bmatrix} \end{aligned}$$

Now we have

$$r_{t+1} = b_{rx}\left(1-\phi_x L\right)^{-1}\varepsilon^x_t + b_{ry}\left(1-\phi_y L\right)^{-1}\varepsilon^y_t + \varepsilon^r_{t+1}$$

Hence

$$\left(1-\phi_x L\right)\left(1-\phi_y L\right) r_{t+1} = b_{rx}\left(1-\phi_y L\right)\varepsilon^x_t + b_{ry}\left(1-\phi_x L\right)\varepsilon^y_t + \left(1-\phi_x L\right)\left(1-\phi_y L\right)\varepsilon^r_{t+1}$$

Aha, r_t follows an ARMA(2,2)!

I don't know how to generalize this to the case that r itself is a state variable. The other project is, how to find the implied $\{r_t, x_t\}$ representation so you can see that multivariate VARs condition down to something like the restricted VAR.

Summary

-Multivariate predictability / mean reversion is strong. If you see a price rise or return *with no change in current dividend*, you know that price rise will completely melt away.

-Univariate predictability is weak or absent. If you see a price rise or return *and don't look at current dividends*, your best forecast is that this rise is essentially permanent.

-No, these two statements are not inconsistent. It is possible for stocks to be predictable and have strong mean-reverting components, yet display no univariate mean-reversion nor decline in variance with horizon. In fact, it's easy for dp predictability to be consistent with "momentum," returns that are positively autocorrelated.

- Our baseline VAR implies very slight univariate mean-reversion. It is attractively close to a parameter configuration that implies uncorrelated returns over time.

- The two key features of our data that drive this result: 1) return shocks are negatively correlated with expected-return or dp shocks. When dp rises unexpectedly, returns fall. This lower ex-post return plus the period of higher ex-post returns after the expected return rise offset, so that average ex-post returns are no higher nor lower than usual during a period of high expected returns. 2) expected return variation is quite small compared to ex-post return variation. Thus, it is very hard to see a period of high expected returns by looking at ex-post returns.

-WARNING: There is no interesting economic question to which this is the answer. You are often told that if "stocks are safer in the long run" then "investors with long horizons should hold more stocks." This certainly looks sensible – $E(R^e)/\gamma\sigma^2(R^e)$, the standard formula for portfolio shares, is better at long horizons. But this formula assumes iid, and the whole point is that returns are not iid. If you do an optimal portfolio theory with non-iid returns, you not blindly allocate more to stocks when univariate mean-reversion is higher. You have to do the whole state-variable thing. So this is an interesting academic exercise, uniting the univariate and multivariate representations, but it does not have the practical relevance you might suspect.

-WARNING 2: There is no interesting statistical question to which this is the answer either. Why would you care about the univariate wold representation when you have more information? Why would you forecast weather using a week of lagged temperature data when you have the NWS forecast?

What's new

More variables

For stocks, I emphasized D/P for simplicity, but many other variables forecast stock returns, and putting them together you get stronger forecasts. Modern research does not stop at D/P, we run regressions with lots of extra variables

$$R_{t+1} = a + b_1(D/P)_t + b_2x_t + b_3y_t... + \varepsilon_{t+1}$$

Among others:

1. The term premium (bond yield spread) forecasts stocks as well as bonds. (The CP bond-return forecasting factor is even better)
2. The consumption/wealth ratio ("cay"),
3. Investment/capital ratios (bad economic times mean higher returns to the few who hold risk) and numerous other economic variables.
4. Share issuance (firms issue shares when prices are high, i.e. cost of capital = future returns are low)
5. Volatility and implied volatility
6. Other cointegrating vectors / divisors. D/E (or P/E), B/M (Or B/D) etc.
7. Even more variables forecast individual stock returns, B/M, size, past returns (momentum), accruals, etc. Since dividends can be zero and earnings can be negative book/market is often favored in place of D/P for individual firms, but gives similar results.

This is very important. D/P work (mine in particular) has often been misquoted as "dividends are not predictable." No, "dividends are not predictable *using only past dividends, returns and dividend yields.*" If you use other variables, all bets are off.

To show you what happens with more variables, I'll show you what happens with a particular favorite, Lettau and Ludvigson's cay variable. The idea is simple and natural. Consumers look forward. If a price movement is "temporary," they don't lower consumption because they know the market will "bounce back." Well, maybe yes, maybe no, let's look at the data.

Cay helps to forecast returns

Here is a reproduction from Lettau-Ludvigson's Table III. It's also useful to give you a sense that yes, people are using all sorts of other variables to forecast returns. For the moment, most papers are one-variable papers; there is not much work yet really sorting out which variables survive in multiple regressions

cay	d-p	d-e	rrel	trm	def	R^2
2.17						0.09
(3.23)						
	0.025					0.00
	(0.97)					
2.27	-0.011					0.09
(4.12)	(-0.40)					
1.91	0.011	-0.004	-1.38	-0.08	-0.88	0.10
(3.20)	(0.10)	(0.27)	(2.44)	(-0.12)	(-0.54)	

d-e=dividend/earnings; rrel = tbill rate/MA; trm=10yr-3 mo yield; def=credit spread

Lettau/Ludvigson Table III. Quarterly Regressions forecasting S&P500 index excess returns 1953-1998.

Another LL table:. This one shows you that cay is helpful at a 1 year horizon, but does not help so much at long horizons. If you therefore suspect that cay is a more quickly mean-reverting variable, you guess right.

Table 20.2. *Long-horizon return forecasts*

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

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

Note, D/P looks terrible here. That's because of the 52 (starts late) - 98 (just before the big crash) sample. Always a skeptic, I reran the regression with more recent data (This data is on Sydney Ludvigson's website.)

	Regression	cay	t	dp	t	R^2
Excess Return	cay only	5.0	(3.5)			0.15
	dp only			4.5	(2.0)	0.08
	cay, dp	4.7	(3.1)	3.9	(1.8)	0.21
Return	cay only	5.0	(3.6)			0.15
	dp only			5.5	(2.6)	0.12
	cay, dp	4.6	(3.1)	4.9	(2.3)	0.25

$R_{t+1} = a + b \times cay_t + c \times dp_t = \varepsilon_{t+1}$; VW returns (-TB), 1952-2008

First row: Yes, cay still forecasts nicely including post-1998 data. Second row: our familiar D/P regression. The coefficient and R^2 are about the same as above, but the t is smaller since the sample is shorter. Note *cay* has a bigger R^2 and t, at least for one period returns. (cay is much less persistent, so is much less important for long-horizon returns, not shown.) Third row: putting them both together, both dp and cay coefficients and t statistics are only slightly lower in multiple regressions. DP coefficients of 4-5 are economically huge, so I wouldn't drop it.

In sum, I conclude that *both* cay and D/P are, together, useful for forecasting returns. cay contributes another independent (orthogonal) dimension of return forecastability.

However, at long horizons, the improvement is much less significant. cay must be a more quickly mean-reverting variable.

Is there one big data point or other skulduggery? Always plot your data. I plot the *cay* forecast, the D/P_t forecast, and the actual R_{t+1} together so we can see how they work. In the second graph, I plot the multiple regression forecast using both cay and D/P together.

As you can see, cay seems to capture the "wiggles" in returns better than the slow-moving D/P It also helps a lot in the late 1990s, when prices were high, and D/P along with Bob Shiller and Alan Greenspan were issuing dire forecasts year after year and getting it wrong. But we still need D/P to capture the slower-moving trend. D/P helps to get the boom of the 80s and early 90s right (when P/D was low) and the poor returns of the post 2000 period. You see in the graphs why the regressions seemed to want them both; one tells you about wiggles, the other about trends.

The combined forecast based on Dec 2008 data pointed to a big rebound which we know happened in 2009. Wealth fell a lot more than consumption in the crash.

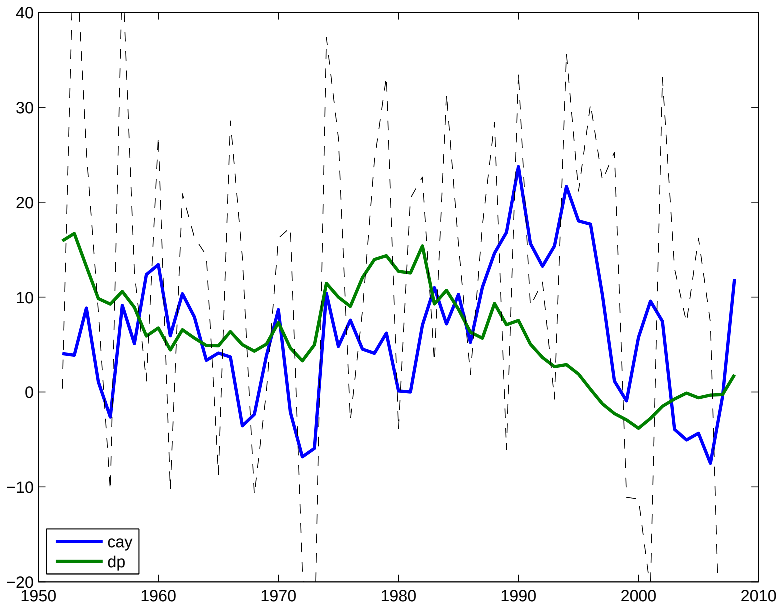


Figure 8: Green: Exess return forecast from D/P. Blue: Excess return forecast from cay. Dashed: Actual excess return. Note R_{t+1} is plotted at the same date as $a + bx_t$.

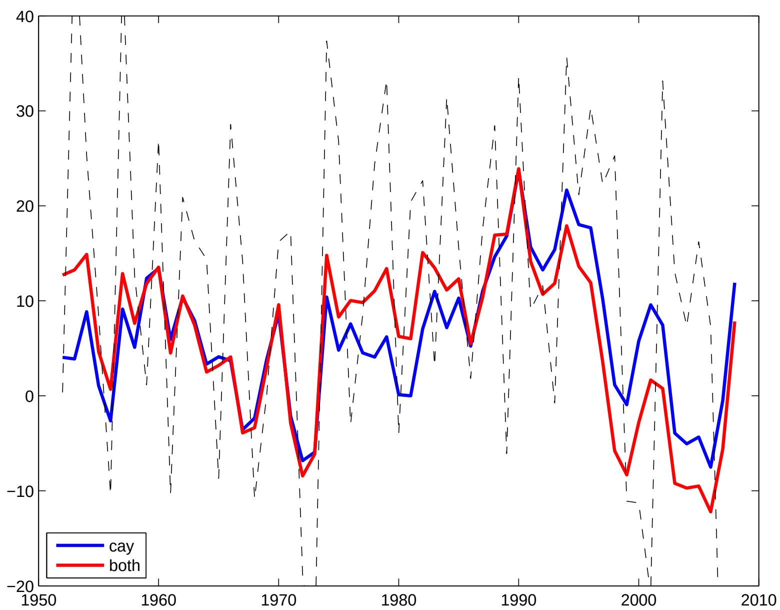
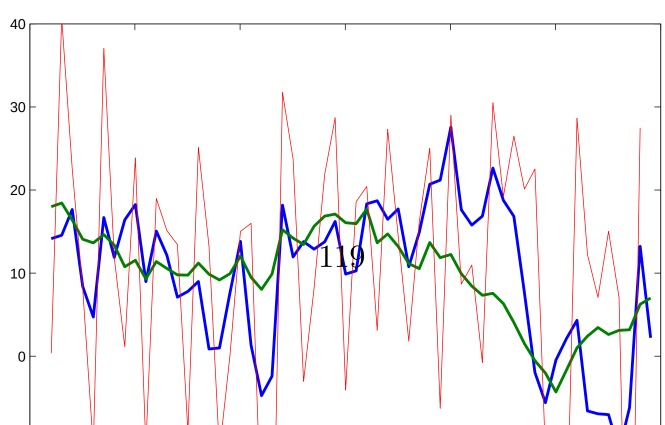


Figure 9: Red: excess return forecast using *both* cay and D/P.Blue: forecast using only cay. Dashed: actual excess return. R_{t+1} is plotted on the same date as $a + bx_t$.



Cay, and identities in larger VARs.

Puzzle. "Wait a minute," I hope you're saying. "We found that return predictability was more than enough to account for all the variance of dividend yields. Now you're telling me that returns are *even more* predictable than we thought. But we can't account for more than 100% of dividend yield volatility. What gives?"

Forecasting both. Well, the present value identity still holds,

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

Thus, *if another variable helps $d-p$ to forecast r_{t+1} it must also help to forecast dividend growth Δd_{t+j} or long-run returns r_{t+j}. These forecasts must offset, so that given $d_t - p_t$, the forecast of the entire right hand side does not change.*

Another way of putting the issue: Price-dividend ratios are attractive forecasters because they reveal to us market expectations of future returns. However, they mix that information with expectations of future returns many years ahead, and information about dividend growth. If we can find other variables that forecast long run returns or dividend growth, we can use them to "clean up" d-p, so that d-p more accurately forecasts one-period returns. (We see the same pattern in the cross section, e.g. Fama and French's "dissecting anomalies." Variables that forecast cash flows help to "clean up" B/M of cashflow-forecast effects and thus help to forecast returns.)

Identities in larger systems. This intuition just reflects how our identities – all of which stem from the return identity – generalize for larger systems. If we project (run a multiple regression) of the present value identity on $d_t - p_t$, *and a new variable* z_t we obtain

$$\begin{aligned} dp_t &= r_t^{lr} - \Delta d_t^{lr} \\ dp_t &= \left(\beta_{\Delta d_t^{lr}, dp_t} - \beta_{r_t^{lr}, dp_t} \right) dp_t + \left(\beta_{\Delta d_t^{lr}, z_t} - \beta_{r_t^{lr}, z_t} \right) z_t \end{aligned}$$

Thus,

• *The regression coefficients (direct or implied by VARs) of long-run forecasting regressions* (this defines β)

$$\begin{aligned} r_t^{lr} &= \beta_{r_t^{lr}, dp_t}(dp_t) + \beta_{r_t^{lr}, z_t} z_t + \varepsilon_t^r \\ \Delta d_t^{lr} &= \beta_{\Delta d_t^{lr}, dp_t}(dp_t) + \beta_{\Delta d_t^{lr}, z_t} z_t + \varepsilon_t^r \end{aligned}$$

must obey the identities

$$\begin{aligned} 1 &= \beta_{\Delta d_t^{lr}, dp_t} - \beta_{r_t^{lr}, dp_t} \\ 0 &= \beta_{\Delta d_t^{lr}, z_t} - \beta_{r_t^{lr}, z_t} \end{aligned}$$

We already had the first equality; this was my "variance decomposition." It now applies to multiple regression coefficients. The news is the second equality:

$\rightarrow$*If another variable z helps d-p to forecast long-run dividend growth* $\beta_{\Delta d_t^{lr}, z_t} \neq 0$ *given d-p, it must help to forecast long-run returns* $\beta_{r_t^{lr}, z_t} \neq 0$ *and vice versa.*

These "long-run" coefficients are the sum of one-period coefficients. Thus, the extra variable z_t can forecast one-year returns by forecasting a different pattern, yet leaving long-run returns r_t^{lr} alone. Denote multiple regression coefficients

$$r_{t+j} = \beta_{r,dp}^{(j)}(dp_t) + \beta_{r,z}^{(j)} z_t + \varepsilon_{t+j}^r$$

then

$$\beta_{r_t^{lr}, z_t} = \sum_{j=1}^{\infty} \rho^{j-1} \beta_{r,z}^{(j)}$$

etc. Thus, the "long run" identity above implies that

- *A variable z may help d-p to forecast one-year returns if it alters the temporal pattern of return forecasts.*

In sum,

- *If* r_{t+1} *becomes more predictable, then* r_{t+j} $j \geq 2$ *must become more predictable, or* $\Delta d_{t+j}, j \geq 1$ *must become more predictable, and the forecasts must perfectly offset.*

-The shock identity

$$\varepsilon_{t+1}^r = -\rho \varepsilon_{t+1}^{dp} + \varepsilon_{t+1}^d$$

Is unchanged. The shock ε_{t+1}^z need not be reflected in any other shock.

-*Impulse-response coefficient.* The impulse-response functions

$$(E_t - E_{t-1})\, r_{t+j} = e_{dp \rightarrow r}^{(j)} \varepsilon_t^{dp} + e_{z \rightarrow r}^{(j)} \varepsilon_t^z$$

where $E_t \equiv E(\cdot | dp_t, z_t)$ must obey the same relation,

$$\begin{aligned} 1 &= \sum_{j=1}^{\infty} \rho^{j-1} e_{dp \rightarrow r}^{(j)} - \sum_{j=1}^{\infty} \rho^{j-1} e_{dp \rightarrow \Delta d}^{(j)} \\ 0 &= \sum_{j=1}^{\infty} \rho^{j-1} e_{z \rightarrow r}^{(j)} - \sum_{j=1}^{\infty} \rho^{j-1} e_{z \rightarrow \Delta d}^{(j)}. \end{aligned}$$

Just take innovations $E_{t+1} - E_t$ of the basic present value identity

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

, in response to a single shock to dp or z.

• *Generalization* You can write the identities down for arbitrary VARs too, e.g.

$$\begin{bmatrix} r_{t+1} \\ \Delta d_{t+1} \\ dp_{t+1} \\ z_{t+1} \end{bmatrix} = A \begin{bmatrix} r_t \\ \Delta d_t \\ dp_t \\ z_t \end{bmatrix} + \begin{bmatrix} \varepsilon^r_{t+1} \\ \varepsilon^d_{t+1} \\ \varepsilon^{dp}_{t+1} \\ \varepsilon^z_{t+1} \end{bmatrix}$$

Then the identity

$$r_{t+1} = \Delta d_{t+1} - \rho dp_{t+1} + dp_t$$

implies

$$A_r = A_d - \rho A_{dp} + \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$$

where A_r denotes the r row of A, and we still have

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

• *Which is it?* Does cay help to forecast one period returns because it changes the *time path* of expected returns? Or does it forecast higher dividend growth at the same time as higher returns, in a way that offsets so that $d_t - p_t$ is unaffected?

To answer these questions I generalized the VAR we studied above to include cay. Here is a table from "Discount rates." Here I rescaled cay to have variance 1 so it would be about on the same size as dp. This accounts for the lower numerical value of the coefficients

	Coefficients		t-statistics		Other statistics		
	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		
$r^{lr} = \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$	1.29	0.033					
$\Delta d^{lr} = \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}$	0.29	0.033					

Table 5. Forecasting regressions using dividend yield and consumption-wealth ratio, 1952-2009, annual data. Long-run coefficients are computed using a first-order VAR with dp_t and cay_t as state variables. Each regression includes a constant. The long-run regression coefficients are implied by the VAR.

(Each regression also has constants.) This is much prettier in vector notation, with $z = cay$,

$$\begin{bmatrix} r_{t+1} \\ \Delta d_{t+1} \\ dp_{t+1} \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & b_r & c_r \\ 0 & 0 & b_d & c_d \\ 0 & 0 & \phi_{dp} & \theta_{dp} \\ 0 & 0 & \phi_z & \theta_z \end{bmatrix} \begin{bmatrix} r_t \\ \Delta d_t \\ dp_t \\ z_t \end{bmatrix} + \begin{bmatrix} \varepsilon^r_{t+1} \\ \varepsilon^d_{t+1} \\ \varepsilon^{dp}_{t+1} \\ \varepsilon^z_{t+1} \end{bmatrix}$$

This is exactly what we did before, just with one more variable. (Again, you can painlessly estimate numbers in place of the zeros, but they don't make much difference.)

Cay obviously is helping a lot to forecast returns. Cay doesn't look like it's doing much to forecast dividend growth. We may want to think of a simplified VAR with both b_d and $c_d = 0$ in this case. Cay mean-reverts much quicker than dp, with an autoregressive coefficient of 0.65 rather than 0.94. This is a central part of "wiggles" vs. "trends" in my graphs. cay and dp don't seem to have all that much to do with each other. A high cay forecasts a little bit lower dp (t=-3.05), but not the other way around.

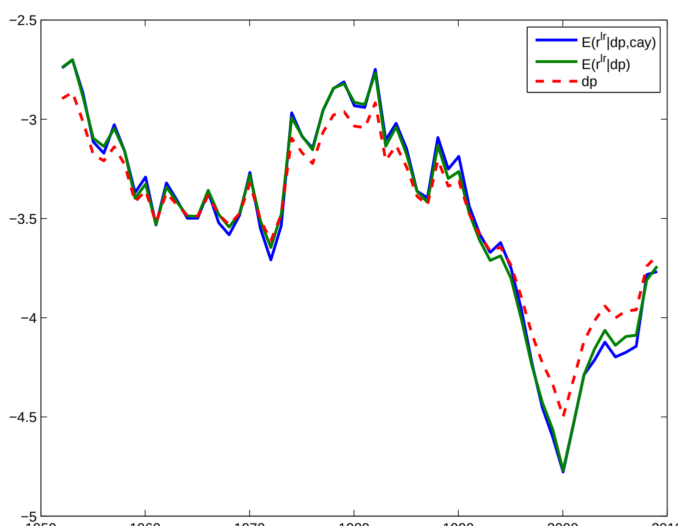
As above, these one-period regression estimates obey the identities

$$
\begin{aligned}
b_r &= 1 - \rho\phi_{dp} + b_d \\
c_r &= c_d - \rho\theta_{dp}.
\end{aligned}
$$

Again, the first one is as before, now applying to multiple regression coefficients. The second one is new. Obviously, how a variable helps to forecast returns must be linked to its ability to forecast prices or dividends! Duh, but as you've seen before these identities can yield a lot of intuition.

The above estimates obey the identities of course. b_r has risen to 0.12 with a rise in b_d to 0.024. This is an effect of the sample – post 1953 and including the crash of 2008. (It is not a difference between single and multiple regression coefficients in this sample; they are nearly identical.) Relative to the 1926-2008 sample, returns are "even more" predictable, there is no change in ϕ and so b_d must rise in the "wrong" direction (high prices mean lower dividends).

Here are long run forecasts, from the cay dp VAR. There's almost no difference! How can this be? Let's track it down.



Dividend yield and expected long horizon returns, $E_t r_t^{lr} = E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$.

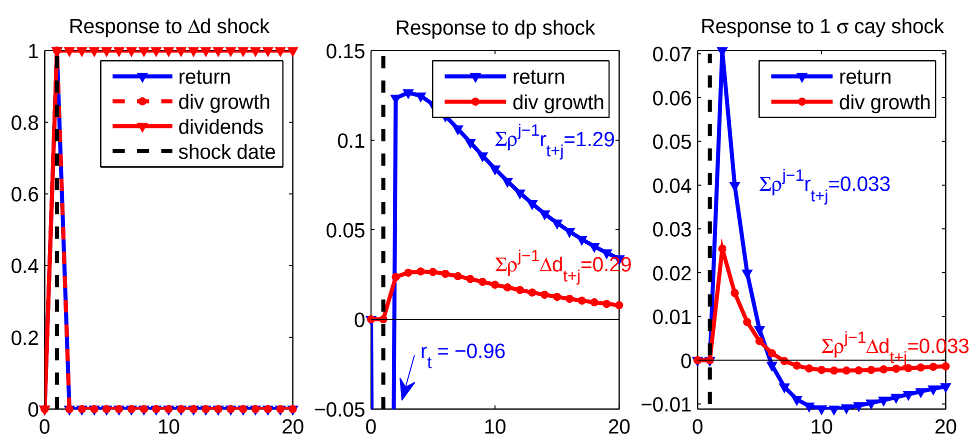
Figure **??** presents the response functions. The natural shocks (consistent with the identity $\varepsilon^r = -\rho\varepsilon^{dp} + \varepsilon^d$) are now

$$
\begin{array}{lc}
\Delta d \text{ shock:} & \left[\begin{array}{cccc} \varepsilon^r & \varepsilon^d & \varepsilon^{dp} & \varepsilon^z \end{array}\right] = \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \end{array}\right] \\
dp \text{ shock:} & \left[\begin{array}{cccc} \varepsilon^r & \varepsilon^d & \varepsilon^{dp} & \varepsilon^z \end{array}\right] = \left[\begin{array}{cccc} -\rho & 0 & 1 & 0 \end{array}\right] \\
cay \text{ shock } : & \left[\begin{array}{cccc} \varepsilon^r & \varepsilon^d & \varepsilon^{dp} & \varepsilon^z \end{array}\right] = \left[\begin{array}{cccc} 0 & 0 & 0 & 1 \end{array}\right]
\end{array}
$$

The response to a dividend growth shock looks almost the same as it did in Figure 3, so there is no point in repeating it and I don't graph it. It's still a pure "shock to current and expected cashflows with no change in expected returns."

Next, let's look at a dividend-yield shock ε_t^{dp}. This response looks almost exactly as before in Figure 3. b_d and b_r are a very little bit larger in this sample (this is an effect of leaving out 1926-1951), but that's it, and the b_d response is still statistically insignificant. Thus, if we see dp change with no current change in cay or Δd, this is still primarily an "expected return shock."

The right hand panel of Figure **??** gives the news. Here, I plot the result of a shock to cay $\varepsilon_t^{cay} = 1$ with no change in dp, $\varepsilon_t^{dp} = 0$ or dividend growth $\varepsilon_t^d = 0$. As a result there is no contemporaneous change in return either. As you can see, the rise in cay with no change in dp corresponds to a sharp rise in expected returns at time $t+1$. That's the finding of our regressions above. However, the expected-return movement is short-lived and even reverses after 6 years. Because the sample coefficient of dividend growth on cay is positive (as we had $b_d > 0$), a shock to cay also corresponds to a smaller and similar short-term rise in expected dividend growth. However, this is also statistically insignificant.



The regression-coefficient identities also apply to implulse-response coefficients, as above.

$$
(\text{dp shock})\ 1 = \sum_{j=1}^{\infty} \rho^{j-1} \left(\varepsilon_t^{dp} \to \Delta d_{t+j} \right) - \sum_{j=1}^{\infty} \rho^{j-1} \left(\varepsilon_t^{dp} \to r_{t+j} \right)
$$

and

$$
(\text{cay shock})\ 0 = \sum_{j=1}^{\infty} \rho^{j-1} \left(\varepsilon_t^{cay} \to \Delta d_{t+j} \right) - \sum_{j=1}^{\infty} \rho^{j-1} \left(\varepsilon_t^{cay} \to r_{t+j} \right)
$$

The responses to the dividend yield shock must add up to one; the responses to the cay shock must add up to zero.

Equivalently, define the "long run" response

$$\left(\varepsilon_t^{dp} \rightarrow r_t^{lr}\right) = \sum_{j=1}^{\infty} \rho^{j-1} \left(\varepsilon_t^{dp} \rightarrow r_{t+j}\right)$$

These identities are exactly the same as the identities for regression coefficients,

$$1 = \left(\varepsilon_t^{dp} \rightarrow \Delta d_t^{lr}\right) - \left(\varepsilon_t^{dp} \rightarrow r_t^{lr}\right)$$

$$0 = \left(\varepsilon_t^{cay} \rightarrow \Delta d_t^{lr}\right) - \left(\varepsilon_t^{cay} \rightarrow r_{t+j}^{lr}\right).$$

At last we can answer the puzzle: how can cay give us *even more* expected returns without explaining "too much" dividend yield volatility. If *cay* rises, returns r_{t+1} rise 0.07 (this is the first response on the right hand side, $\varepsilon_t^{cay} \rightarrow r_{t+1}$). What is less? The graph (blue) shows that the *whole* term $\left(\varepsilon_t^{cay} \rightarrow r_{t+j}^{lr}\right)$ is $\sum_{j=1}^{\infty} \rho^{j-1} \left(\varepsilon_t^{cay} \rightarrow r_{t+j}\right) = 0.033$. Thus, all those negative terms on the right hand side of the blue line add up to the difference

- *The main reason cay can help to forecast positive near-term returns is that it forecasts a long steam of negative long term returns.*

If the dividends did not move at all, this would be the whole story. The point estimates say that the small dividend forecast offsets the remaining 0.0033 return forecast:

- *Higher cay means* **both** *higher returns and higher dividend growth. These offset so that d-p doesn't change.*

Now, cay's dividend growth forecasts are not statistically significant, so we should probably simplify our picture of the world by setting the coefficient to zero. Then the entire ε^{dp} shock corresponds to expected returns, and the entire *cay* shock corresponds to a variation in the *time-path* of expected returns.

However, I left the point estimate dividend coefficients in the graph **??** to emphasize an important point. Dividend growth *is* forecastable by many other time series. Many *other* forecasting variables help dp to forecast returns *by* forecasting dividend growth, in such a way that the dividend growth and return forecasts offset. Don't leave thinking that dividend growth is a pure random walk just because dp and cay can't forecast it!

Variance decompositions

How does adding more variables change our variance decompositions? How does it change our answer to the question "does price (/dividend) volatility come from cash flow forecasts, discount rates, or bubbles?"

Again, we started from the dividend yield identity,

$$pd_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}.$$

Our first "variance decomposition" multiplied both sides by $p_t - d_t$ and took expectations, relating $var(p-d)$ to the ability of $p-d$ to forecast dividend growth and returns.

- *This variance decomposition is completely unaffected by the presence of cay or other forecasting variables.*

In this calculation we related dividend yield variance to the ability of *dividend yields* to forecast returns and dividend growth, with no help from cay or other variables.

The Campbell variance decomposition is affected by extra variables. Here we use E_t from the whole VAR, and simply take variances of both sides,

$$var(pd_t) = var\left[E_t\left(d_t^{lr}\right)\right] + var\left[E_t\left(r_t^{lr}\right)\right] - 2cov\left[E_t\left(d_t^{lr}\right), E_t\left(r_t^{lr}\right)\right]$$

where I denote the "long run return"

$$\sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} = r_t^{lr}$$

This decomposition can reflect the fact that $E_t\left(r^{lr}\right)$ varies *even more* than it does using only d-p to forecast, though we know that such greater forecastability will have to be met by a larger value of the troublesome covariance term.

I just calculate the E_t terms using both dp_t and cay_t as forecasters and take the variance. Here is the result. The left panel shows what happens without cay. The right panel shows how that changes when we add cay to the system.

	Only dp			dp and cay		
	σ^2	σ	% σ^2(dp)	σ^2	σ	% σ^2(dp)
var(dp)	0.123	0.351	100	0.124	0.352	100
var$\left[E_t(r_t^{lr})\right]$	0.223	0.472	180	0.224	0.473	181
var$\left[E_t(\Delta d_t^{lr})\right]$	0.015	0.120	12	0.015	0.122	12
-2*cov$\left[E_t(r_t^{lr}), E_t(\Delta d_t^{lr})\right]$	-0.114		-92	-0.115		-92

Variance decomposition, 1951-2008

Start with the results using only dp. You can see some of the same points that we saw in the variance decomposition based on covariances. The variance of expected long-run returns is in fact almost double what is needed on its own to explain dividend yield volatility. The standard deviation of expected returns is about one and a half times as much as needed. Expected long-run dividend growth contributes very little. (The larger b_d in this sample accounts for the larger numbers.)

Alas, this decomposition requires the covariance term, and it is large. Expected dividend growth and expected returns *are* correlated. In fact, in this case, they are perfectly correlated. Each forecast is based on $d_t - p_t$ after all, and $b_r(d_t - p_t)$ is perfectly correlated with $b_d(d_t - p_t)$. Now, in this sample b_d is not exactly zero. b_r is a little "too big," so b_d must also be a little "too big" to offset it. When *dp* changes, expected returns rise more than they need to, so expected dividend growth must move at the same time to forestall a larger dp change. So, it makes sense, but it's not very satisfactory to say "92 percent of the variance of price-dividend growth comes from the covariance of expected return and dividend forecasts."

However, despite this disadvantage (which isn't so bad once you understand it), this variance decomposition has some advantages. You can directly compute the contribution of varying expected long run returns and long-run dividend growth. to dividend yields; you can *look at* $var(E_t \sum \rho^{j-1} r_{t+j})$ and $var(E_t \sum \rho^{j-1} \Delta d_{t+j})$ and see how big each is. And, you can add extra forecasters to E_t and see what happens in larger systems.

Now, we can look at what happens when we add *cay*. The result is almost exactly the same! How is that possible? Cay helps to predict *one-period* returns, so you'd think the variance of expected returns must rise when we add cay to the forecast. However, the variance of expected *long-run* returns cannot rise unless cay helps to forecast long-run dividend growth. Cay mostly helps to predict one period returns by changing the *time-path* of expected returns, in a way that leaves the long-run return completely unchanged. *To see a change in variance decomposition, you need to find a variable that predicts long run returns; to do so it must predict* ***long-run*** *dividend growth as well (since it can't change p-d). We would see both terms rise, and a larger negative covariance term. A variable such as cay that only changes the time-path of expected returns has no effect on volatility tests.*

In sum,

- cay helps to predict one-period returns largely by changing the time-path of predicted returns. When cay rises, but dp stays the same, near-term expected returns rise, long-term expected returns fall, and the long-run expected return is largely unchanged.
- As a result, adding cay to the regressions does not make much difference at all to the price-dividend variance decomposition, the question "does price variation come from cashflows or discount rates?"
- Other variables may help to predict one-year returns by also helping to predict dividend growth. If, in addition, they help to predict long-run dividend growth, they will change the answer to the variance decomposition of price-dividend ratios, by raising the variance of long-run expected returns, the variance of long-run expected dividend growth *and* the covariance, since such changes must come together and offset each other.
- Cay is a good example of my disenchantment with the "structural" interpretation above. In the information set including cay, "true" expected return variation is not the AR(1) that we used to digest the dp-r system.

More variables still

What's left to do? A lot! Most current research still proceeds on a variable by variable basis. First we need to sort out which variables really do help, and at which horizons. You can't just stuff them all in a big multiple regression because there are more candidate variables than data points.

Second, we have proceeded on an asset by asset basis; we found stock return forecasters, bond return forecasters, etc. The next question is, which of these are really common across assets, and which are truly idiosyncratic to an asset class? As a first step, Cochrane and Piazzesi found some big surprises in finding the one variable that forecast all bond returns. Now, can we do something like that across asset classes, not just across maturities?

And finally, don't forget the identities. At a minium it helps the intuition and economic believability a lot if you understand the *other* thing a variable must forecast to help forecast returns.

What about Repurchases?

(Boudoukh et al. Journal of finance). This is a cautionary tale about plotting your data before you publish a paper.

Note *there is nothing wrong with using CRSP dividends. P really is the present value of future D, not of "repurchase adjusted" future D.* You can always hold your stock when they repurchase.

However, if firms stop paying regular dividends, the time-series process of dividends could change, resulting in much delayed big lumps.

Anyway, you can look at price/x and see if it forecasts returns for any x.

Data definitions:

Gross = (Dividends + Repurchases)/Price;

Net: (Dividends + Repurchases - Issues) / Price (net).

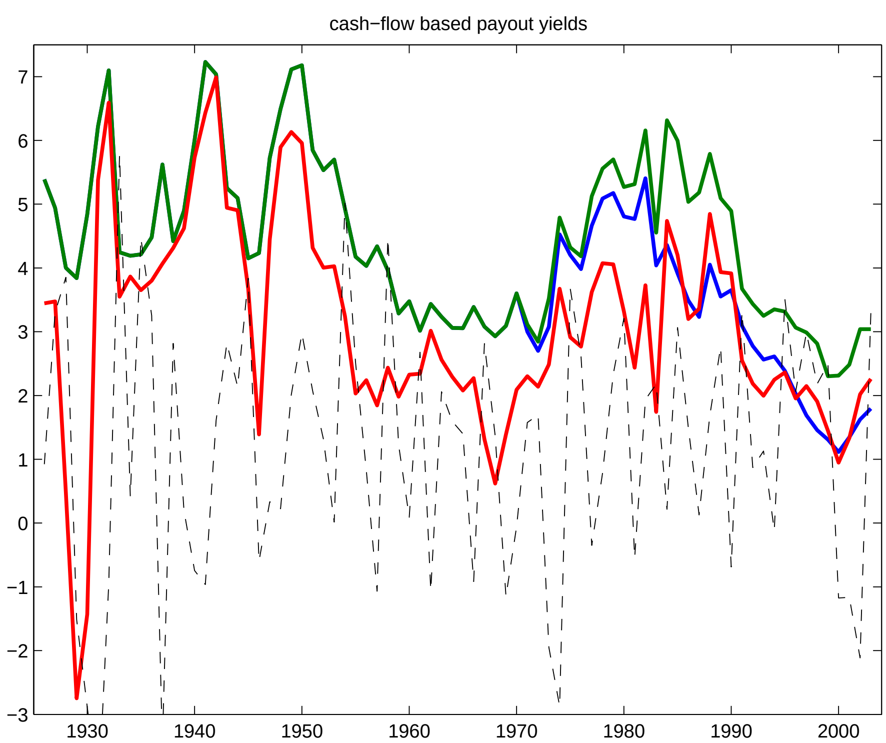
CF= based on cash flows, TS = based on treasury stock data.

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

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	

0.267 R^2!!!Can it be that good?



Blue: D/P. Green: Payout/P. Red: (Net Payout)/P. Dashed: Return/10

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

Appendix (not in lecture).

These are some issues I have talked about in the past, but decided to demote from lectures and required reading this year.

Other present value identities and models

CS variations

There are a lot of variations on the Campbell-Shiller identity

- Real and nominal. It doesn't make any statistically significant difference.

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

- Dividend growth, interest rates, and risk premiums. No surprise, it's all excess returns.

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

- Vuolteenaho: a similar market/book expansion useful for individual stocks, which may have $D = 0$. .

$$\log \left(\frac{market_t}{book_t} \right) \tilde{=} E_t \sum \rho^j \left[\log \left(1 + \frac{earnings_{t+j}}{book_{t+j}} \right) - r_{t+j} \right]$$

- An example of how to use other units: an identity for price-earnings ratios rather than dividend. This method lets you use any other divisor

$$(p_t - e_t) = \sum_{j=1}^{\infty} \rho^{j-1} \left[\Delta e_{t+j} - r_{t+j} - (1 - \rho) \left(e_{t+j} - d_{t+j} \right) \right]$$

Derivation:

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

$$r_{t+1} = \rho\left(p_{t+1} - e_{t+1}\right) - \left(p_t - e_t\right) + \rho\left(e_{t+1} - d_{t+1}\right) - \left(e_t - d_t\right) + \Delta d_{t+1}$$

$$r_{t+1} = \rho\left(p_{t+1} - e_{t+1}\right) - \left(p_t - e_t\right) + \rho\left(e_{t+1} - d_{t+1}\right) - \left(e_t - d_t\right) + \Delta e_{t+1} + \left(d_{t+1} - e_{t+1}\right) - \left(d_t - e_t\right)$$

$$r_{t+1} = \rho\left(p_{t+1} - e_{t+1}\right) - \left(p_t - e_t\right) - (1-\rho)\left(e_{t+1} - d_{t+1}\right) + \Delta e_{t+1}$$

$$\left(p_t - e_t\right) = \Delta e_{t+1} - r_{t+1} - (1-\rho)\left(e_{t+1} - d_{t+1}\right) + \rho\left(p_{t+1} - e_{t+1}\right)$$

- The same idea has bee used all over the place. For example **??** express that the net debt position of a country in terms of future trade surpluses and future returns on holders of the debt. Surprise surprise countries tend to "default" (offer low returns) on debt rather than run export surpluses.

- I still like the identity in Cochrane (1991). It lets you talk about why the mean pd and mean return is higher in some securities than others. But I have to admit CS won the race. Theirs is tightly linked to the return identity, which is another plus.

An exact return-based present value relation

You can do the same iteration without linearization

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

Similarly

$$\begin{aligned}
P_t &= \sum_{j=1}^{\infty}\left(\prod_{k=1}^{j} R_{t+k}^{-1}\right) D_{t+j} \\
\frac{P_t}{D_t} &= \sum_{j=1}^{\infty}\left(\prod_{k=1}^{j} R_{t+k}^{-1} \frac{D_{t+k}}{D_{t+k-1}}\right)
\end{aligned}$$

- Again, this is not a model, it's an identity. It has to work. You can always "discount" with returns.

- Again this works ex post as well as ex ante. (You can always add E)

- It makes the general point, *High price/dividend ratios, must mechanically come from high future dividend growth or low future returns (or PD that rise forever)*

- But this is not very useful, because it's complex. The linearized formula lets you connect with linear time series tools, and see the first order effects of "expected return" thinking on prices. The linearized formula is leaving out some other effects of course, but is usually pretty accurate, at least for the stock market as a whole. Don't use it for securities with wild payoffs like options.

Gordon Growth

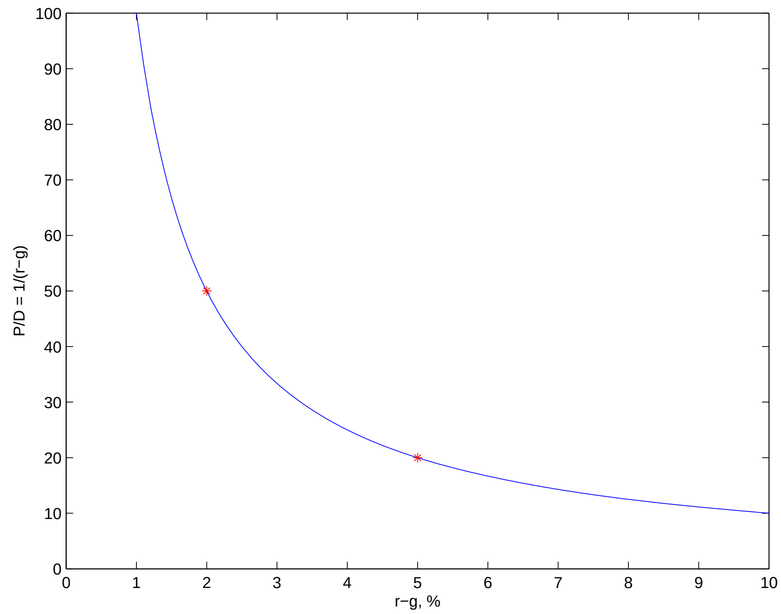
This is a great back of the envelope model

$\Delta d = g$ constant. $r =$ constant

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

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

Cool points you can see with this model:



a) How big are anomalies? *Small ER changes mean big price changes*

- Example. A bias/friction raises price 10%. (Socially responsible investing, HSBC) How much does this affect returns? Say D/P = 2.5% (common) then

$$\frac{D}{P} = r - g = 0.0250 = 2.5\%$$

$$\frac{P}{D} = \frac{1}{0.025} = 40$$

$$\frac{1}{44} = 0.0227 = 2.27\%$$

a *10 % irrational price = 13 bp of return/year = 1bp per month.* Standard tests argue about 1% per month!

b) Growth stocks/high markets should be more volatile (Assuming the same volatility in r,g)

c) Volatility in g is good for prices, since it's convex. (Pastor and Veronesi explanation of the bubble)

d) The last two examples shows an important difference between GG and CS. GG captures the nonlinearities that we left out with CS. And, for small $r-g$, this is a decently nonlinear formula.

e) If Er rises is this good or bad news for stock prices? Most beginners will say "good" and get it wrong. Think like an economist!

What's wrong? This formula assumes that r and g last forever. We often use it and then let r and g vary, but that's not correct. The linearized CS model lets us add time series models for r and g.

What about m?

Wait, you should be saying. Real present value models should look like

$$P_t = E_t \sum m_{t,t+j} D_{t+j}$$

or

$$P_t = E_t \sum \beta^j \frac{u'(C_{t+j})}{u'(C_t)} D_{t+j}.$$

What's going on?

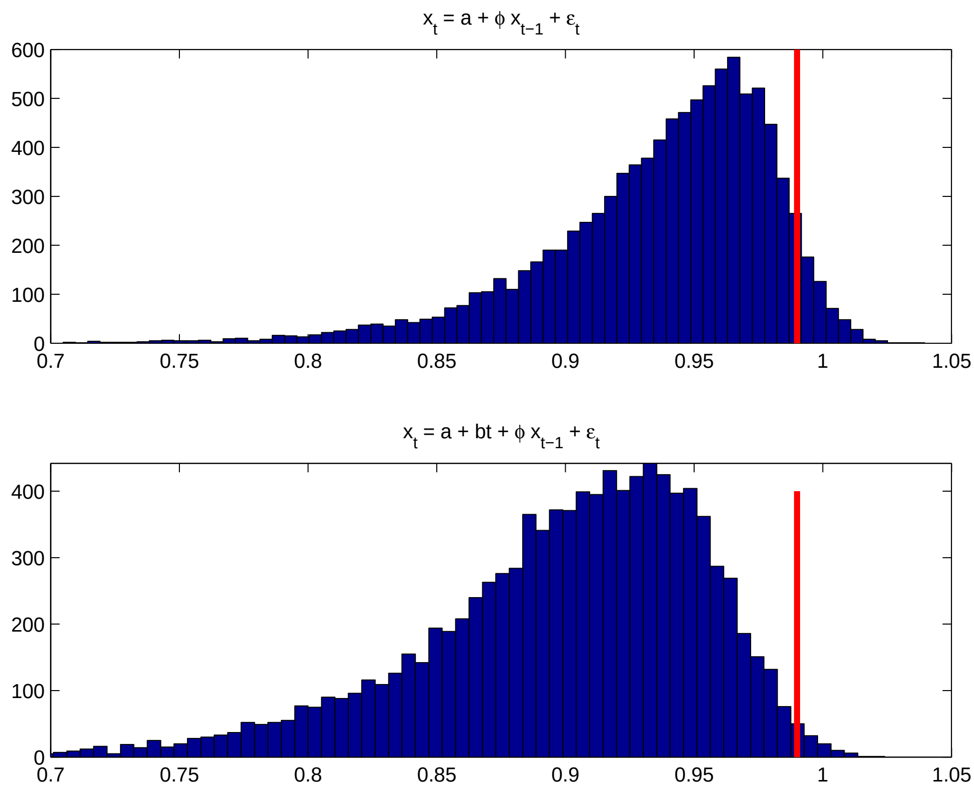
The models I'm studying here are *identities*, not models. The point is to tie ideas and facts about returns to ideas and facts about prices. We still have to explain returns, i.e. with

$$1 = E_t \left[\beta \frac{u'(C_{t+1})}{u'(C_t)} R_{t+1} \right]$$

and a real present value model is still our goal.

Statistics and the dog

Statistical problem 1: Unit roots and Stambaugh bias:



•Distribution of estimates from an AR(1), with true coefficient = 0.99. T = 100. Note it's even worse if you allow a time trend. Basically, the mean and the trend soak up low-frequency variation that is really stochastic.

	mean	median
b1	0.94	0.95
b2	0.90	0.91

Thus, in

$$dp_{t+1} = a + \phi dp_t + \varepsilon_{t+1}^{dp}$$

ϕ is biased down, because of overfitting of the sample mean.

•In the extreme case $\phi = 1$, there is no true mean. (Show a time series graph showing "overfitting" of sample mean from a case with pure random walk and time trend fitted.)

•Bias = "correlation of errors with right hand variable." Return regression

$$r_{t+1} = a + b_r dp_t + \varepsilon_{t+1}^r$$

The errors ε^r and ε^{dp} are strongly negatively correlated. So downward bias in ϕ which reflects correlation of dp with ε^{dp}, means *br is biased up.*

- Also

$$
\begin{aligned}
b_r - b_d &= 1 - \rho\phi \\
\hat{b}_r - \hat{b}_d &= 1 - \rho\hat{\phi}
\end{aligned}
$$

$$
\hat{b} = b + (x'x)^{-1}x'\varepsilon
$$

is a linear function of errors. Thus, $\hat{b}_r$ and $\hat{\phi}$ inherit the correlation of ε^r and ε^{dp} (and b_d inherits the lack of correlation)

-There is no bias if $\varepsilon^r, \varepsilon^{dp}$ errors are not strongly correlated, i.e. Δd.regressions. *There's nothing wrong in principle with regressions that have serially correlated right hand variables.*

$\rightarrow$ *Return coefficient is biased up. Return t stats are biased up.*

Stambaugh: OLS p-values of 6% (1927-1996) and 2% (1952-1996), correct: p-values are 17% and 15%.

Statistical question 2: Are long horizons more powerful or not?

Intuition: long horizons are just a function of 1 period results. The 1 period regression is MLE. What could be more powerful? Intuition 2: as we saw, long horizon standard errors grew just as fast as means, so there seems to be no greater information in long horizon regressions.

Answer: No. Intuition is wrong. (And so was I for many years, including statements in Asset Pricing)

Statistical question 3:

Goyal-Welch systematically demolish the out-of sample forecastability of estimators. These are useless in real time. (Show GW pictures) (Answer coming: They focus on "usefulness to investor," who presumably wants to market time. I agree. It's not useful to investors. In fact, it can't be – the average investor is holding the market portfolio. But we can still radically change our "view of the world." R^2 is not the measure of success!)

Answer

- But what about Dividend forecasts? You can't have both dividend and return forecasts = 0! (They might answer, we can't tell dividend forecast from $b_d = -0.1$, no extra information here)

- I might also complain that point estimates rather than tests are what's important. Viewed as % of pd variance the point estimates are still huge and 100%/0%. But let's not give up yet.

Dog null:

$$\begin{aligned} r_{t+1} &= a_r + b_r(d_t - p_t) + \varepsilon^r_{t+1} \\ \Delta d_{t+1} &= a_d + b_d(d_t - p_t) + \varepsilon^d_{t+1} \\ d_{t+1} - p_{t+1} &= a_{dp} + \phi(d_t - p_t) + \varepsilon^{dp}_{t+1}. \end{aligned}$$

$$r_{t+1} = -\rho dp_{t+1} + \Delta d_{t+1} + dp_t$$

Again, the identities imply

$$\begin{aligned} b_r &= 1 - \rho\phi + b_d. \\ \varepsilon_r &= -\rho\varepsilon_{dp} + \varepsilon_d \end{aligned}$$

The identities are the key to everything here:

1. $b_r = 1 - \rho\phi + b_d$ expresses possibilities. $b_r = 0$ must imply much higher $\phi = 1.04$ (bubbles) or $b_d = -0.1$.(dividend growth forecastable)

2. $\hat{b}_r = 1 - \rho\hat{\phi} + \hat{b}_d$ holds in each sample too

3. The correlation of errors. Regression coefficients are linear function of errors, so $\hat{b}_r$ and $\hat{\phi}$ are negatively correlated, $\hat{b}_r$ and $\hat{b}_d$ are positively correlated, $\hat{b}_d$ and $\hat{\phi}$ are basically uncorrelated.

Null:

$$\begin{aligned} b_r &= 0,\ \phi = 0.94,\ \ b_d = -(1 - \rho\phi) \approx -0.1, \\ b_r &= 0,\ \ \phi = 0.99, b_d = -(1 - \rho\phi) \approx -0.05 \end{aligned}$$

The latter is, as we'll see, a decent "bias-adjusted" version.

Alternative/data:

$$b_r = 0.1,\ b_d = 0,\ \ \phi = 0.94$$

How often do we see observations "more extreme" than the alternative, when data is generated by the null?

(Note the paper is still guilty of "how I thought of it" organization. Do better in your papers!)

-FIGURE 1: Joint distribution of $\hat{b}_r, \hat{b}_d$

-Yes, b_r is insignificant.

-b_d is significant! It seems it does make a difference whether you test $b_r = 0$ or whether you test $b_d = -0.1$.

-Notice the upward sweep, resulting from the positive correlation of $\hat{b}_r, \hat{b}_d$

-This is uncomfortable. Why do we have to choose one test or the other?

FIGURE 2 long run coefficients.

$$\begin{aligned}\frac{b_r}{1-\rho\phi}-\frac{b_d}{1-\rho\phi} &= 1\\ b_r^{lr}-b_d^{lr} &= 1.\\ 1-0 &= 1\end{aligned}$$

-This is attractive because there are no multidimensional issues. This holds *in each sample.* If you look at one b_r^{lr} by the identity you know exactly what the other one does, so the test statistic is the same either way you look at it.

-Also, these are more interesting (I think so). They express fractions of dp variance explained. I think the interesting question about the world is the value of b_r^{lr}, b_d^{lr} not the underlying b_r, b_d, ϕ. Result: significant. Table 4 numbers

FIGURE 3 Joint distribution of b_r, ϕ;;

Identities – let's look at b_r, ϕ leaving out b_d. This is the conventional way to look at it.

Notice the negative correlation between $\hat{b}_r$ and $\hat{\phi}$, coming from negative correlation in the errors.

$b_r = 1-\rho\phi+b_d$ thus diagonal lines show us the $\{b_r, \phi\}$ corresponding to sample b_d,and greater and less – NE of the diagonal line is above the b_d, b_r Figure 1 horizontal line

Similarly, holding $b_r/(1-\rho\phi)$ constant shows where the long run coefficients reject.

Notice *the rejecting b_r all come with very low ϕ.* A "joint" test, do you see high b_r *and* ϕ no lower than 0.94 really rejects.

This explains the b_r, b_d results. The rejecting b_r came with low ϕ so b_d was unaffected.

b_r and ϕ are negatively correlated. It's much harder to produce a high b_r without a low ϕ, That's why $b^{lr}=b_r/(1-\rho\phi)$ is less likely to reject and why $b_d = b_r+\rho\phi-1$ is less likely to reject.

You can't just pick arbitrarily small regions and then reject.

1) Economics. What's interesting? What is "more extreme" than data? Most $\hat{b}_r > 0.1$ have low $\hat{\phi}$. Thus, $\frac{b_r}{1-\rho\phi} < 1$, $b_d < 0$, (See Figure 3) so dividends are predictable, and volatility tests are "some of each".

Plea that the Economically interesting result is b^{lr}, "how often is $b_r^{lr} > 1$" *The combination of small and persistent, so adds up to lots of price volatility, is the puzzle.* Thus, I would count

$$\hat{\phi}=1, \hat{b}_r = 0.04,\ \hat{b}_d = 0, b_r^{lr} = 0.04/(1-0.96) = 1$$

"just as surprising" as what we see, but

$$\phi = 0.64, b_r = 0.2, b_r^{lr} \approx 0.2/(1-0.96\times 0.64) = 0.2/0.4 = 0.5$$

i.e. 50/50 as "closer to the null".

2) Statistics. why does testing b_r involve other paramters? Answer: because the null is not $\{b_r = 0, \phi = \text{anything}\}$ $\{b_r = 0, \|\phi\| < 1\}$. We're adding prior information. We're adding prior information that $\phi < \bar{\phi}$. If you move phi up, things get worse. The likelyhood region is then negatively sloped in b_r, ϕ space, as the b_d and b_{lr} tests do (Figure 3).

Higher phi Table. This Emphasizes how upper limits on prior for phi are crucial.

If I allowed $\phi = 1.04$, we could have both $b_r = 0$ and $b_d = 0$.

LONG HORIZON FORECASTS

Table 6, Figure 5. Previous long run results were not long enough.

But long horizon is the same as bd. It's not about long horizons really, it's about economic distance from the null. b_d is not correlated with phi.

Warning: a) forecasting long horizons from a VAR is dangerous b) it's much less dangerous with cointegrating vectors, since these dominate the long run

GOYAL AND WELCH

"Out of sample" Goyal welch statistic. See Figure 6 – this is expected to happen! GW is not a "test." It is a warning against using this in real time for market timing. There is no fight. We can decide that $b_r^{lr} = 1$,$b_{lr}^d = 0$ is our view of the world with high probability, yet, GW is very likely to be right about "out of sample" one period return forecasting, since we don't know b_r and ϕ *separately* very well.

Bias (hidden point in table 7 and last paragraph of 1571) A good "bias corrected" estimate is $\phi = 0.99$, $b_r = 0.05$, $b_d = 0$. But this still has $b_r^{lr} = 1$!

The downward bias in phi is exactly the same pheonmenon as the upward bias in br.

In SUM. (Introduction) Key features are 1) the identity 2) the correlation structure of the errors (hard to avoid) 3) Hence the correlation structure of the b, ϕ 4) the (well, my) prior that $\phi < 1$

Read the last three paragraphs! The implications of return forecastability have only begun to be appreciated in finance, portfolio theory (add this fall's merton state variables, WSJ article), corporate finance, accounting, ...

Thoughts

- You can see how useful the linearized present value model is. Big picture: stock analysis is stuck in one period thinking; bonds are moving to analyzing prices (affine models). Stocks need that too, but the models need to be tractable.

Why not..

What about the 1 period model

$$
\begin{aligned}
P_t &= E_t(M_{t+1}D_{t+1}) \\
\frac{P_t}{D_t} &= E_t(M_{t+1}\frac{D_{t+1}}{D_t})
\end{aligned}
$$

or multiperiod

$$
\begin{aligned}
P_t &= E_t\sum_{j=1}^{\infty} M_{t,t+j}D_{t+j} \\
\frac{P_t}{D_t} &= E_t\sum_{j=1}^{\infty} M_{t,t+j}\frac{D_{t+j}}{D_t} = E_t\sum_{j=1}^{\infty}\left(\prod_{k=1}^{j} M_{t+k-1,t+k}\frac{D_{t+k}}{D_{t+k-1}}\right)
\end{aligned}
$$

Isn't that what the author of *Asset Pricing* should recommend?

-Answer 1: the linearized identity helps us to relate prices, dividends and the vast empirical work on one period returns. The linearized model lets us relate P/D, ER, ED. Why ER varies will still be a mystery (Need M, 1=E(MR) for what is ER needs M) This is the final goal though.

- Answer 2: Eventually, we need to get the whole business away from one period returns. The covariance of daily returns with daily consumption growth is not the Rosetta stone for asset pricing. The covariance of the *stream of dividends* my portfolio gives me with the *stream of consumption flows* I support with it; that's what matters.