Final exam review problems.

These are a collection of problems from old exams, which should guide your study and give you some sense of my question style and the style of answers I'm looking for. Warning, some of these cover material we didn't study. You are obviously not responsible for those questions. Warning 2: I understand things better now than when I wrote these questions, so I don't guarantee the answers any more! Warning 3: there are things we did cover that are not in these questions.

In general, you do not need to memorize formulas. You should know the 2 or 3 main points of each paper we read. "State the main point of x and how it was documented" is legitimate. You should also be able to recognize facts documented in papers we read (and when we talked about those in class).

Stock Predictability

1. (35) Suppose that returns are predictable and dividend growth is not predictable,

$$
\begin{aligned}
E_t r_{t+1} &= x_t \\
x_t &= \phi x_{t-1} + \delta_t \\
\Delta d_{t+1} &= w_{t+1},
\end{aligned}
$$

δ_t and w_t are all uncorrelated with each other and over time. When you are asked for numerical values, use parameters $\phi = 0.94$, $\rho = 0.96$, $\sigma(\delta) = 0.01$,(1%) $\sigma(w) = 0.10$.(10%). The answers can be approximate (no calculator needed), e.g. feel free to write $0.94 \times 0.96 \approx 0.9$.

(a) Find the log price-dividend ratio at time t. You're looking for a very short equation with $p_t - d_t$ on the left and variables people know at t such as x_t on the right.

(b) What does this model predict for the coefficient of log returns on the log dividend yield? I.e. what is b_r in a regression

$$r_{t+1} = a_r + b_r (d_t - p_t) + \varepsilon_{t+1}?$$

Give both the formula and the numerical value. Formula hint:

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

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

(c) Find the unexpected return at time t, $r_t - E_{t-1}(r_t)$ in terms of "structural" shocks. Hint:

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

(d) In this model, what fraction of dividend yield variance comes from expected return shocks (δ), and what fraction from expected or unexpected dividend growth shocks (w)?

(e) In this model, what fraction of the variance of unexpected returns comes from expected return shocks δ vs. what fraction comes from dividend growth shocks w? If this answer is different than in part d, how do you reconcile the two answers? (Hint: what if prices move but dividends also move so there is no change in d-p?)

(f) For the special case $\sigma_w^2 = 0$, find the univariate Wold representation of returns and the response of returns to a univariate Wold shock, $A(L)r_t = B(L)v_t;\ v_t \equiv r_t - E(r_t|r_{t-1}, r_{t-2}, ...)$ (Hint: Guess that the autoregressive root is ϕ, then look at the structural representation for $(1-\phi L)r_{t+1}$. The Wold theorem says that the Wold representation of r_t is unique and converges, so once you find any convergent ARMA model for r_t you have the Wold representation.)

(g) At the above parameter values, do returns display univariate mean reversion, momentum, or are returns uncorrelated over time? Are there reasonable parameter values that change this conclusion?

ANSWER:

(a)

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

(b) $b_r = 1-\rho\phi \approx 0.1$

(c) The return identity is[1]

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

$$\left(E_t - E_{t-1}\right)r_t = \left(E_t - E_{t-1}\right)\left(\Delta d_t\right) - \left(E_t - E_{t-1}\right)\left(\rho x_t + \rho\phi x_t + ...\right)$$

$$\left(E_t - E_{t-1}\right)r_t = w_t - \frac{\rho}{1-\rho\phi}\delta_t$$

This is the prvious version. It was wrong, because the sums have ρ^{j-1}in them not ρ^j

$$\begin{aligned} r_t - E_{t-1}r_t &\approx \left(E_t - E_{t-1}\right)\left[\Delta d_t + \sum_{j=1}^{\infty}\rho^{j-1}\Delta d_{t+j} - \sum_{j=1}^{\infty}\rho^{j-1}r_{t+j}\right] \\ &\approx \left(E_t - E_{t-1}\right)\left[\Delta d_t - \frac{1}{1-\rho\phi}x_t\right] \\ &= w_t - \frac{1}{1-\rho\phi}\delta_t \end{aligned}$$

---

[1]

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

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

$$\left(E_{t+1} - E_t\right)\left(\Delta d_{t+1} + \rho\Delta d_{t+2} + \rho^2\Delta d_{t+3} + ..\right) = \left(E_{t+1} - E_t\right)\left(r_{t+1} + \rho r_{t+2} + \rho^2 r_{t+3} + ..\right)$$

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

(d) 100% from expected returns. $p-d$ is proportional to x, so all variation comes from x variation. (We can be a bit fancier and derive the results for volatility tests, which give the same result since x and its shocks are now observed from dp observations.)

(e) *This was also wrong, now corrected to reflect the extra ρ Of course the numbers aren't so pretty – with the algebra mistake it came out 0.01 and 0.01.*

$$\begin{aligned}
\sigma^2\left(r_t - E_{t-1}r_t\right) &= \sigma_w^2 + \left(\frac{\rho}{1-\rho\phi}\right)^2 \sigma_\delta^2 \\
&= \sigma_w^2 + \left(\frac{0.96}{1-0.96\times 0.94}\right)^2 \sigma_\delta^2 \\
&\approx \sigma_w^2 + \left(\frac{0.96}{0.1}\right)^2 \sigma_\delta^2 \\
&\approx \sigma_w^2 + 92\times\sigma_\delta^2 \\
&= 0.10^2 + 92\times(0.01)^2 \\
&= 0.01 + 0.0092
\end{aligned}$$

So it's roughly half and half .Return variance also comes from current dividend growth shocks . If prices and dividends move there is no movement in pd, but there is a big return.

(f) *This part had timing wrong as well as the missing ρ.*

$$r_{t+1} = x_t + r_{t+1} - E_t r_{t+1} = x_t + w_{t+1} - \frac{\rho}{1-\rho\phi}\delta_{t+1}$$

$$\begin{aligned}
(1-\phi L)r_{t+1} &= (1-\phi L)x_t + (1-\phi L)\left(w_{t+1} - \frac{\rho}{1-\rho\phi}\delta_{t+1}\right) \\
(1-\phi L)r_{t+1} &= \delta_t + (1-\phi L)\left(w_{t+1} - \frac{\rho}{1-\rho\phi}\delta_{t+1}\right)
\end{aligned}$$

if $w_t = 0$

$$\begin{aligned}
(1-\phi L)r_{t+1} &= \delta_t - \frac{(1-\phi L)\rho}{1-\rho\phi}\delta_{t+1} \\
&= \left(1+\frac{\rho\phi}{1-\rho\phi}\right)\delta_t - \frac{\rho}{1-\rho\phi}\delta_{t+1} \\
&= \frac{1}{1-\rho\phi}\delta_t - \frac{\rho}{1-\rho\phi}\delta_{t+1} \\
&= -\frac{\rho - L}{1-\rho\phi}\delta_{t+1} \\
&= -\left(1-\frac{1}{\rho}L\right)\frac{\rho}{1-\rho\phi}\delta_{t+1}
\end{aligned}$$

This is an explosive root. The Wold decomposition for r uses the non-explosive root, so the Wold representation is

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

Note in this case I do not have to match autocovariance functions and so forth. We know the roots, so we're done. The difference between v and δ can soak up the constants (-$\rho/(1-\rho\phi)$)

You would have to match autocovariance functions (or do some more work by other means) to derive the variance of v in terms of the variance of δ, but I didn't ask for that. *In the end I had gotten the right answer despite three miraculously canceling mistakes along the way. Thank goodness I'm too old to take tests any more*

(g)

$$r_{t+1} = \frac{1-\rho L}{1-\phi L} v_{t+1} = \left(1 + \frac{(\phi - \rho) L}{1-\phi L}\right) v_{t+1} = \left(1 - \frac{0.02L}{1-0.94L}\right) v_{t+1}$$
$$r_{t+1} = v_{t+1} - 0.02 v_t - (0.94) 0.02 v_{t-1} - (0.94^2) 0.02 v_{t-2} - ...$$

$$a(1) = \frac{1-\rho}{1-\phi} = \frac{1-0.96}{1-0.94} = \frac{0.04}{0.06} = \frac{2}{3}$$

So returns have a long string of very small negative responses after a shock, and the cumulative return response function decays slowly ending up at 2/3 of its impact value. That's "some mean reversion" but of course nowhere near the total mean reversion displayed by prices or cumulative returns to the structural δ shock. .

2. Consider a VAR representation of returns and dividend growth with two right hand variables,

$$r_{t+1} = a_r z_t + b_r dp_t + \varepsilon_{r,t+1} \tag{13}$$
$$\Delta d_{t+1} = a_d z_t + b_d dp_t + \varepsilon_{d,t+1}$$
$$dp_t = \phi_{dp,dp} dp_{t-1} + \phi_{dp,z} z_{t-1} + \varepsilon_{dp,t+1}$$
$$z_t = \phi_{z,dp} dp_{t-1} + \phi_{z,z} z_{t-1} + \varepsilon_{z,t+1} \tag{14}$$

where as usual r represents log returns, Δd log dividend growth, $dp_t = d_t - p_t$ log dividend yield, z_t is the extra forecasting variable, and all variables are de-meaned. (Assume that r_t and Δd_t as well as further lags of dp and z have zero coefficients.)

(a) If z_t is not present, what rough values do you expect to see for b_r, b_d and $\phi_{dp,dp}$ in annual data?

(b) What restrictions on the coefficients a, b, ϕ of this VAR flow from identities?

(c) Suppose the extra variable helps to forecast dividend growth, i.e. $a_d \neq 0$. We have some intuition that if a variable helps to forecast dividend growth, it should also help to forecast returns. Is this true in your system? If $a_d \neq 0$, does that imply that $a_r \neq 0$?

(d) To investigate this issue a bit further, write a "structural" system rather than the "reduced form" VAR, in which z only exists to forecast dividend growth

$$x_{t+1} = \phi_x x_t + \varepsilon_{xt+1} \tag{15}$$
$$z_{t+1} = \phi_z z_t + \varepsilon_{zt+1}$$
$$E_t(r_{t+1}) = x_t$$
$$E_t(\Delta d_{t+1}) = z_t. \tag{16}$$

People in the economy observe both x and z; we observe z, as well as prices, dividends, and returns. Find the coefficients in the VAR representation of the form (13)-(14) that results from this "structural" system.

(e) If the VAR resulting from this system displays $a_d \neq 0$ must it also display $a_r \neq 0$?

(f) Your VAR representation from part c may have some zeros or other restrictions not present in (13)-(14). How would you modify the setup of (15)-(16) to remove those restrictions? (Just write down the system you think you need to use, don't solve it.)

Hint: In case you forgot, the return linearization (ignoring constants as usual) and associated formulas are

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

$$p_t - d_t \approx E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}; \quad \rho = \frac{1}{1 + D/P} \approx 0.96$$

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

ANSWER:

(a) $b_r \approx 0.1$, $b_d \approx 0$, $\phi_{dp,dp} \approx 0.94$

(b) The return identity is

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

Using VAR coefficients,

$$a_r z_t + b_r \left(d_t - p_t\right) = \left(1 - \phi_{dp,dp}\rho\right)\left(d_t - p_t\right) - \rho\phi_{dp,z} z_t + a_d z_t + b_d \left(d_t - p_t\right)$$

thus we must have,

$$\begin{aligned} b_r &= b_d + (1 - \phi_{dp,dp}\rho) \\ a_r &= a_d - \rho\phi_{dp,z} \end{aligned}$$

It is not an identity, but stationarity means that the eigenvalues of ϕ need to be less than one.

(c) No. we can have $a_r = 0$ if $a_d = \rho\phi_{dp,z}$. Intuitively, in this case a rise in z_t lowers $p_{t+1} - d_{t+1}$ just enough to offset the rise in Δd_{t+1} in the definition of return

(d)

$$\begin{aligned} x_{t+1} &= \phi_x x_t + \varepsilon_{xt+1} \\ z_{t+1} &= \phi_z z_t + \varepsilon_{zt+1} \\ E_t(r_{t+1}) &= x_t \\ E_t(\Delta d_{t+1}) &= z_t. \end{aligned}$$

$$\begin{aligned} d_t - p_t &= E_t \sum_{j=1}^{\infty} \rho^{j-1} \left(r_{t+j} - \Delta d_{t+j}\right) \\ d_t - p_t &= \frac{1}{1 - \rho\phi_x} x_t - \frac{1}{1 - \rho\phi_z} z_t \\ r_{t+1} &= x_t + \varepsilon_{rt+1} \\ \Delta d_{t+1} &= \delta z_t + \varepsilon_{dt+1} \end{aligned}$$

Now we have to find the VAR representation. All we need to do is manipulate until we get dp, z, on the right hand side and orthogonal error terms. (dp and z jointly reveal x and z in this system.)

$$
\begin{aligned}
r_{t+1} &= (1-\rho\phi_x)\left(\frac{1}{1-\rho\phi_x}x_t - \frac{1}{1-\rho\phi_z}z_t\right) + (1-\rho\phi_x)\left(\frac{1}{1-\rho\phi_z}z_t\right) + \varepsilon_{r,t+1} \\
r_{t+1} &= (1-\rho\phi_x)\,dp_t + \frac{1-\rho\phi_x}{1-\rho\phi_z}z_t + \varepsilon_{r,t+1}
\end{aligned}
$$

$$
\begin{aligned}
dp_{t+1} &= \frac{1}{1-\rho\phi_x}x_{t+1} - \frac{1}{1-\rho\phi_z}z_{t+1} = \frac{1}{1-\rho\phi_x}\phi_x x_t - \frac{1}{1-\rho\phi_z}\phi_z z_t \\
&= \frac{1}{1-\rho\phi_x}\phi_x x_t - \frac{1}{1-\rho\phi_z}\phi_x z_t + \frac{1}{1-\rho\phi_z}\phi_x z_t - \frac{1}{1-\rho\phi_z}\phi_z z_t + \varepsilon^{dp}_{t+1} \\
&= \phi_x dp_t + \frac{\phi_x - \phi_z}{1-\rho\phi_z}z_t + \varepsilon^{dp}_{t+1}
\end{aligned}
$$

(We can derive ε^{dp}_{t+1} in terms of structural shocks ε_x and ε_z, but it's not important here.) In sum, then, the VAR is

$$
\begin{aligned}
r_{t+1} &= [1-\rho\phi_x]\,dp_t + \left[\frac{1-\rho\phi_x}{1-\rho\phi_z}\right] z_t + \varepsilon_{r,t+1} \\
\Delta d_{t+1} &= [0]\,dp_t + [1]\,z_t + \varepsilon_{d,t+1} \\
dp_{t+1} &= [\phi_x]\,dp_t + \left[\frac{\phi_x - \phi_z}{1-\rho\phi_z}\right] z_t + \varepsilon_{dp,t+1} \\
z_{t+1} &= [\phi_z]\,z_t + [0]\,dp_t + \varepsilon_{z,t+1}
\end{aligned}
$$

(e) The coefficient of r on z on r is $[1-\rho\phi_x]\,/\,[1-\rho\phi_z]$ while the coefficient of Δd on z is 1. This satisfies the identity above,

$$
\begin{aligned}
a_r &= a_d - \rho\phi_{dp,z} \\
\frac{1-\rho\phi_x}{1-\rho\phi_z} &= 1 - \rho\frac{\phi_x - \phi_z}{1-\rho\phi_z} \\
\frac{1-\rho\phi_x}{1-\rho\phi_z} &= \frac{1-\rho\phi_z - \rho\phi_x + \rho\phi_z}{1-\rho\phi_z}
\end{aligned}
$$

Now, $a_r = 0$ means $1-\rho\phi_x = 0$ needs $\phi_x = 1/\rho \approx 1.04 > 1$ which we can't have.

(f) This system is more restricted than the original one. The problem is that z fully reveals expected dividend growth, so there can be no *marginal* forecast power for dp to forecast dividend growth. To get rid of the zeros, we need to cloud up z's ability to forecast dividends and to forecast itself. We need another state variable that agents see and we don't. For example,

$$
\begin{aligned}
x_{t+1} &= \phi_x x_t + \varepsilon_{xt+1} \\
z_{t+1} &= \phi_z z_t + \varepsilon_{zt+1} \\
w_{t+1} &= \phi_w w_t + \varepsilon_{wt+1} \\
E_t(r_{t+1}) &= x_t \\
E_t(\Delta d_{t+1}) &= z_t + w_t.
\end{aligned}
$$

would work, with agents not seeing w. This is harder to do though, because the VAR does not reveal the underlying shocks.

3. Several authors have pointed to the predictability of stock returns from price ratios as evidence that stocks are "safer in the long run," that properly-scaled return variance is lower for long-horizon returns. What is the evidence on this question?

   ANSWER: we saw that variance ratios $\sigma^2(r_{t+1} + r_{t+2} + .. + r_{t+k})/k\sigma^2(r_t)$ and long run regression coefficients do not show much evidence of univariate forecastability. We found the univariate representation implied by our standard VAR, and the univariate return response function was very close to a random walk. Thus it is possible and likely that "stocks are predictable" but "stocks are not safer in the long run." We also emphasized that the hedging + market timing demand in a Merton portfolio problem is not so easily calculated as simply looking at these ratios.

   (a) (5) What is a reasonable approximate value for the regression coefficient of market returns on the dividend price ratio in annual data? Answer both for log returns, log dp, and for percent returns on percent dp

   (b) (5) We said that roughly speaking 100% of the variance of market dividend yields comes from returns and 0% from dividend growth, but only roughly 60% of the variance of market returns comes from expected returns, with 40% from dividend growth. How is this possible – do dividends matter, or don't they? (Hint: What happens if Δd_t is iid?)

   (c) (5) Cochrane claims that "long run" coefficients $b_r/(1-\rho\phi)$ are more powerful tests of return predictability. But you can't beat maximum likelihood, and the ML estimate and test of b_r is just the OLS estimate. Which is right?

   ANSWER

   (a) About 2-4 in percents, about 0.1 in logs

   (b) The return decomposition is

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

   returns include *current* dividends, while dp ratios count only *future* dividend growth. If dividend growth is iid, it still is there in the return decomposition.

   (c) Cochrane is imposing $\phi < 1$. An ML test of $b_r = 0$ and $\phi < 1$ will use ϕ information too. And the theorems are all asymptotic anyway.

4. Consider the standard model of returns and expected returns,

   $$x_{t+1} = \phi x_t + \varepsilon^x_{t+1}$$
   $$r_{t+1} = x_t + \varepsilon^r_{t+1}$$

   If the shocks are uncorrelated, does this process lead to mean reversion or momentum? Since "returns are predictable" in this case, are "returns safer in the long run?" What change to the model would you have to make to reverse these results?

   ANSWER

   The univariate representation is

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

$$\left(1+\phi^2\right)\sigma_r^2+\sigma_x^2 = (1+\theta^2)\sigma_v^2$$
$$-\phi\sigma_r^2 = -\theta\sigma_v^2$$

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

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

$x+1/x$ is a declining function on (0,1), so $\theta < \phi$, $(1-\theta)/(1-\phi) > 1$ and this generates momentum; stocks are "riskier in the long run." To generate mean reversion, you need $corr(\varepsilon^x, \varepsilon^r) < 0$ and pretty big.

5. Barking dogs:

   (a) What is Goyal and Welch's main complaint about return predictability regressions using the dividend yield?

   (b) Does "the dog that did not bark" show anything wrong with Goyal and Welch's calculations, or does it admit them but counter in some other way?

   ANSWER

   (a) "Out of sample," in rolling regressions, the sample mean produces a lower mean squared error than the dividend yield.

   (b) No. It shows that under the null that all forecastability comes from time varying expected returns, we see the GW result or worse about half the time. Their finding is not inconsistent with the null that expected returns vary over time and dividend growth does not.

6. The"Dog that does not bark" claims that the "long run" coefficient

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

   is a more powerful statistic than the "short run" coefficient b_r itself.

   (a) What is the key to this result? – How can it be that data generated from the null are less likely to produce long run coefficients larger than we see than they are likely to produce short run coefficients larger than those we see in the data?

   (b) What identity does the long-run coefficient obey? What identity does the short-run coefficient obey?

   ANSWER

   (a) Under the null $b_r = 0$, $b_d = -0.1$, $\phi = 0.94$,estimates of ϕ are negatively correlated with estimates of b_r. Thus it's harder for a draw to produce *both* high b_r *and* low ϕ.

   (b)

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

. This comes from

$$r_{t+1} \approx \rho(p_{t+1} - d_{t+1}) - (p_t - d_t) + \Delta d_{t+1}.$$
$$b_r = -\rho\phi + 1 + b_d$$
$$\frac{b_r}{1-\rho\phi} = 1 + \frac{b_d}{1-\rho\phi}$$

7. A famous financial economist recently wrote "..in-sample linear predictive regressions of the realized one-year market return on the lagged log price-dividend ratio over the period 1930-2006 give statistically and economically insignificant slope coefficients and small adjusted-R2 of 2.2%." Leaving aside the question of *statistical* significance, the small adjusted R^2 certainly seems like a devastating criticism of the *economic* significance of point estimates of this phenomenon. What's your response, based on things we studied in class?

   ANSWER

   Short-horizon forecasting R^2 is a terrible measure of economic significance. We have looked at the following:

   (a) R^2 rises with horizon, since the forecasting variable is serially correlated. 5-7 year R^2 are as high as 60%

   (b) $\sigma(E_t(R_{t+1}))/E(R_{t+1})$ is more interesting; the variance of expected returns is very large relative to the level of returns. The equity premium *varies* over time by more than its *level.* This is also related to the persistence of the forecasting variable since $\sigma(E_t(R_{t+1})) = \sigma(b_r x_t) = b_r/\sqrt{(1-\phi^2)}$ A large ϕ offsets a low b_r.

   (c) The variance decomposition tells us what fraction of the (huge) variance of dividend yields comes from expected returns vs. expected dividend growth (conditioned on the $r, \Delta d, dp$ information set). The small b_r with low R^2 nonetheless is consistent with 100% of the variance of dividend yields due to expected return variation. Again, it's because ϕ is so big. $b_r^{lr} = b_r/(1-\rho\phi) = 0.1/(1-0.94\times 0.96) = 1$.

   (d) In our analysis of momentum, we saw that small R^2 can imply huge profits. In that case, $R_{t+1} = 0.1R_t + \varepsilon_{t+1}$, $R^2 = 0.01$ together with the big variance of returns so last year's winners went up 100%, gives a large 10% expected return to the winner portfolio. (This is the cross-sectional counterpart to #2)

Bond predictability

1. (5) We studied the following table, updating Fama and Bliss's results

	$r_{t+1}^{(n)} - y_t^{(1)} = a + b\left(f_t^{(n-1\to n)} - y_t^{(1)}\right) + \varepsilon_{t+1}$					$y_{t+n-1}^{(1)} - y_t^{(1)} = a + b\left(f_t^{(n-1\to n)} - y_t^{(1)}\right) + \varepsilon_{t+1}$				
n	a	b	$\sigma(a)$	$\sigma(b)$	R^2	a	b	$\sigma(a)$	$\sigma(b)$	R^2
2	0.04	0.91	0.28	0.26	0.14	-0.04	0.09	0.28	0.26	0.00
3	-0.15	1.20	0.50	0.35	0.15	-0.34	0.40	0.59	0.29	0.03
4	-0.37	1.41	0.70	0.44	0.16	-0.71	0.66	0.70	0.20	0.09
5	-0.09	1.10	0.95	0.52	0.07	-0.88	**0.85**	0.79	0.20	0.13
	forecasting *one* year returns on n-year bonds					forecasting *one* year rates n years from now				

I highlighted the number 0.85. What, *exactly*, is therefore equal to 0.15? (No need to prove, just state the answer. Be very careful where you put your ns and your ts. Feel free to draw a picture too. Clarify any notation you invent by defining it in terms of bond prices.)

ANSWER We're looking at b in $y_{t+4}^{(1)} - y_t^{(1)} = a + b\left(f_t^{(4\to5)} - y_t^{(1)}\right) + \varepsilon_{t+1}$. We're splitting the 5 year bond up to the last year's yield and the return for the first 4 years, so it's the b in

$$r_{t\to t+4}^{(5)} - r_{t\to t+4}^{(4)} = a + b\left(f_t^{(4\to5)} - y_t^{(1)}\right) + \varepsilon_{t+1}$$
$$p_{t+4}^{(1)} - p_t^{(5)} - p_t^{(4)} = a + b\left(f_t^{(4\to5)} - y_t^{(1)}\right) + \varepsilon_{t+1}$$

2. (10) Short answer

   (a) What do Cochrane and Piazzesi mean by a "one factor model of expected returns?" (A few equations are appropriate here.) How does this concept of "factor" relate to usual eigenvalue decompositions?

   (b) Do Cochrane and Piazzesi find that, statistically, bond expected returns follow a one-factor model? Explain how one might test it and whether it does or does not pass the test. (If you can't remember what we did, invent a new test. It is enough to say what regression you would run or moment condition you would look at – you do not have to derive the covariance matrix or test statistic.)

ANSWER:

(a) let rx_{t+1} denote the 4×1 vector of bond excess returns, and f_t denote the 5×1 vector of forward rates. Unrestricted regressions are of the form

$$rx_{t+1} = \beta f_t + \varepsilon_{t+1}$$

where β is 4×5. A one factor model is the restriction $\beta = b\gamma'$ where b is 4×1 and γ is 5×1. If this is true then

$$cov(E_t(rx_{t+1})) = b\left(cov\left(\gamma' f_t\right)\right)b'$$

the central item is a scalar, so this looks just like one column of $Q\Lambda Q\prime$.

(b) Many answers are ok. If the one factor model is correct,

$$rx_{t+1} = b\gamma' f_t + \varepsilon_{t+1}$$
$$rx_{t+1} - b\gamma' f_t = \varepsilon_{t+1}$$

so you can run

$$rx_{t+1} - b\gamma' f_t = \gamma^* f_t + \varepsilon_{t+1}$$

and test $\gamma^* = 0$. CP use

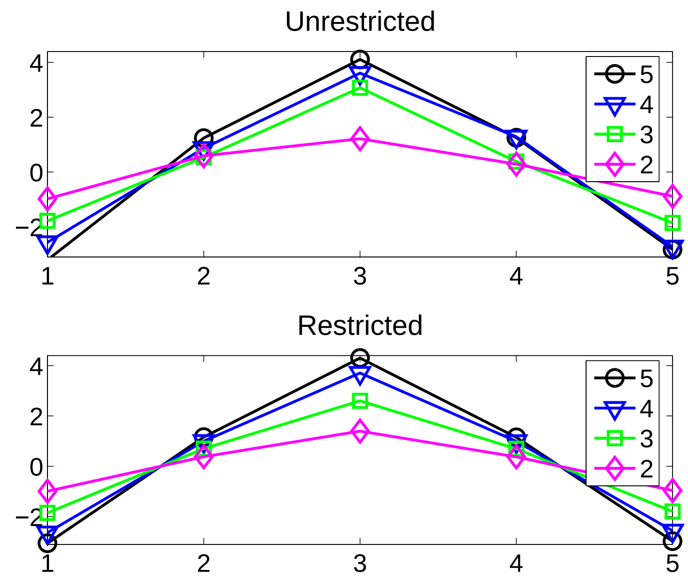
$$r\bar{x}_{t+1} = \gamma' f_t + \varepsilon_{t+1}$$

to write

$$rx_{t+1} - b\overline{rx}_{t+1} = \gamma^* f_t + \varepsilon_{t+1}$$

and again test $\gamma^* = 0$. These are equivalent to the moment conditions of the GMM test, good for you if you wrote them down.

3. (10) What does this picture represent? Be explicit, with equations.



ANSWER

These are CP regression coefficients, unrestricted and the single-factor model

$$rx_{t+1}^{(n)} = a^{(n)} + \beta_1^{(n)} y_t^{(1)} + \beta_2^{(n)} f_t^{(2)} + \beta_3^{(n)} f_t^{(3)} + ... + \beta_5^{(n)} f_t^{(5)} + \varepsilon_{t+1}$$

$$rx_{t+1}^{(n)} = a^{(n)} + b_n \left[\gamma_1 y_t^{(1)} + \gamma_2 f_t^{(2)} + \gamma_3 f_t^{(3)} + ... + \gamma_5 f_t^{(5)}\right] + \varepsilon_{t+1}$$

4. Decomposing the yield curve" uses a four factor model,

$$X_t = \left[\begin{array}{cccc} x_t & \text{level}_t & \text{slope}_t & \text{curve}_t \end{array}\right]$$

with dynamics

$$X_{t+1} = \mu + \phi X_t + v_{t+1};\ E\left(v_{t+1} v_{t+1}'\right) = V$$

to describe the yield curve. The affine model is based on

$$M_{t+1} = \exp\left(-\delta_0 - \delta_1' X_t - \frac{1}{2}\lambda_t' V \lambda_t - \lambda_t' v_{t+1}\right)$$
$$\lambda_t = \lambda_0 + \lambda_1 X_t.$$

It produces

$$E_t\left[rx_{t+1}\right] + \frac{1}{2}\sigma^2(rx_{t+1}) = cov(rx_{t+1}, v_{t+1}')\left(\lambda_0 + \lambda_1 X_t\right)$$

and

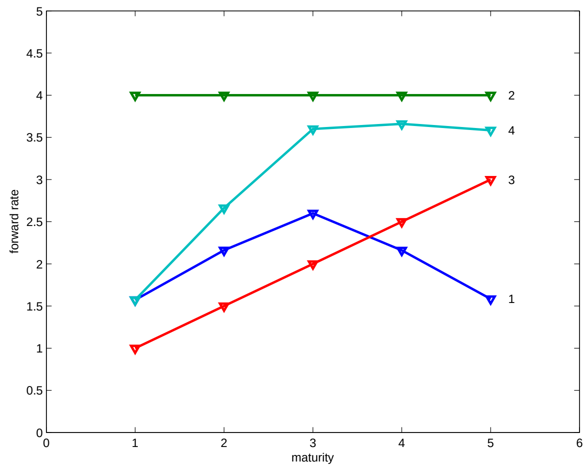
$$f_t^{(n)} = A_n^f + \delta_1' \phi^{*n-1} X_t.$$

Cochrane and Piazzesi find ϕ^* by fitting the last equation in the cross-section, essentially running nonlinear regressions of $f_t^{(n)}$ on X_t and inferring ϕ^* as coefficients. They then claim that this estimate gives us lots of information with which to make forecasts. But forecasts are made based on the true measure, $\phi \equiv \phi^* + V\lambda_1$, and λ_1 is a 4×4 matrix of unknown coefficients. Doesn't this give a huge degree of uncertainty to the resulting forecasts, based on iterating X_t forecasts?

ANSWER

The single factor model means only one column of λ_1 is non-zero and the finding that expected returns line up perfectly with level factor covariance mans that only one row of λ_1 is nonzero. Thus, there is only one free parameter linkingϕ and ϕ^*.

5. Consider the following forward curves:



Rank forward curves 1-4 by which provides the strongest signal of one year excess returns on 5 year bonds i) according to Fama and Bliss' regressions ii) according to Cochrane and Piazzesi's regressions. (There may be ties.) (A: FB look for slope, CP look for tent shapes.)

6. The current log yield on 1, 2 and 3 year bonds is 20%, 15%, 10% – an inverted yield curve

   (a) Find current log prices and forward rates.

   (b) Find the expected one year return on 2 and 3 year bonds, and the expected one and two year yields one year from now,

      1. According to the expectations hypothesis
      2. According to Fama and Bliss. Simplify their regression coefficients to 1 or 0, as appropriate.(Hint: you can figure out the expected two year yield one year from now from the expected return on the three year bond.)

   (c) Plot the expected bond prices through time in each case. (Your plot has time on the x axis and bond price on the y axis. You do not have to find the FB path for the 3 year bond past time 1).

7. (A very simple version of Cochrane/Piazzesi) Let's modify the basic Vasicek term structure model, and see if we can account for Fama-Bliss regressions. The basic model has a constant market price of risk. We need to have a time-varying price of risk. The obvious way to do that is just to make the price of risk depend on the single factor. So, let's pursue the obvious extension, in which rather than just λ we have a time-varying $\lambda_t = \lambda_0 + \lambda_1 x_t$,

$$x_{t+1} - \delta = \rho(x_t - \delta) + \varepsilon_{t+1}$$
$$\log m_{t+1} = -x_t - \frac{1}{2}(\lambda_0 + \lambda_1 x_t)^2 \sigma_\varepsilon^2 - (\lambda_0 + \lambda_1 x_t)\varepsilon_{t+1}$$

(a) Find $p_t^{(1)}$, $p_t^{(2)}$, hence $y_t^{(1)}$, $f_t^{(2)}$, $rx_{t+1}^{(2)}$, $E_t rx_{t+1}^{(2)}$ in this model. Hint: they are still linear functions (stuff)+ (stuff) x_t! You have to use $Ee^x = e^{Ex+\frac{1}{2}\sigma^2 x}$

(b) Find the predicted value of the Fama-Bliss coefficients, i.e. write $E_t rx_{t+1}^{(2)} = (\cdot)+(\cdot)(f_t^{(2)} - y_t^{(1)})$. (All you're doing here is substituting out the previous results. You had $E_t rx_{t+1}^{(2)} = a + bx_t$ and $(f_t^{(2)} - y_t^{(1)}) = c + dx_t$, so if you just write

$$E_t rx_{t+1}^{(2)} = a + b\frac{\left((f_t^{(2)} - y_t^{(1)}) - c\right)}{d}$$
$$E_t rx_{t+1}^{(2)} = a - \frac{bc}{d} + \frac{b}{d}(f_t^{(2)} - y_t^{(1)})$$

you're done.) Forget the mess in the constant, we're only interested in the coefficient, b/d. Can we find λ_0, λ_1 so that this model captures the Fama-Bliss slope coefficient of approximately 1?

(c) In this model, what is the "right" set of forecast variables to use – Fama-Bliss $f^{(n)} - y^{(1)}$, Cochrane-Piazzesi $\gamma' f$ or something else?

(d) (Note: this problem would be easier with $\lambda_0 = 0$)

ANSWER

(a)

$$\begin{aligned}
p_t^{(1)} &= \log E\left[e^{-x_t - \frac{1}{2}(\lambda_0+\lambda_1 x_t)^2\sigma_\varepsilon^2 - (\lambda_0+\lambda_1 x_t)\varepsilon_{t+1}}\right] = -x_t \\
p_t^{(2)} &= \log E\left[e^{-x_t - \frac{1}{2}(\lambda_0+\lambda_1 x_t)^2\sigma_\varepsilon^2 - (\lambda_0+\lambda_1 x_t)\varepsilon_{t+1}} e^{-x_{t+1}}\right] \\
&= \log E\left[e^{-x_t - \frac{1}{2}(\lambda_0+\lambda_1 x_t)^2\sigma_\varepsilon^2 - (\lambda_0+\lambda_1 x_t)\varepsilon_{t+1}} e^{-\delta-\rho(x_t-\delta)-\varepsilon_{t+1}}\right] \\
&= \log E\left[e^{- \frac{1}{2}(\lambda_0+\lambda_1 x_t)^2\sigma_\varepsilon^2 - (\lambda_0+\lambda_1 x_t)\varepsilon_{t+1}} e^{-2\delta-(1+\rho)(x_t-\delta)-\varepsilon_{t+1}}\right] \\
&= -2\delta - (1+\rho)(x_t - \delta) + \frac{1}{2}\sigma^2 + (\lambda_0 + \lambda_1 x_t)\sigma^2
\end{aligned}$$

(b)

$$\begin{aligned}
E_t\left(rx_{t+1}^{(2)}\right) &= E_t(p_{t+1}^{(1)}) - p_t^{(2)} + p_t^{(1)} \\
&= E_t(-x_{t+1}) + 2\delta + (1+\rho)(x_t-\delta) - \frac{1}{2}\sigma^2 - (\lambda_0+\lambda_1 x_t)\sigma^2 - x_t \\
&= -\delta - \rho(x_t - \delta) + 2\delta + (1+\rho)(x_t-\delta) - \frac{1}{2}\sigma^2 - (\lambda_0+\lambda_1 x_t)\sigma^2 - x_t \\
&= -\frac{1}{2}\sigma^2 - (\lambda_0+\lambda_1 x_t)\sigma^2
\end{aligned}$$

(c) As x_t varies so do expected returns. It's a single factor model, so any yield will do – no FB or CP business here. for example,

$$E_t\left(rx_{t+1}^{(2)}\right) = -\frac{1}{2}\sigma^2 - \left(\lambda_0 + \lambda_1 y_t^{(1)}\right)\sigma^2$$

8. Paper questions:

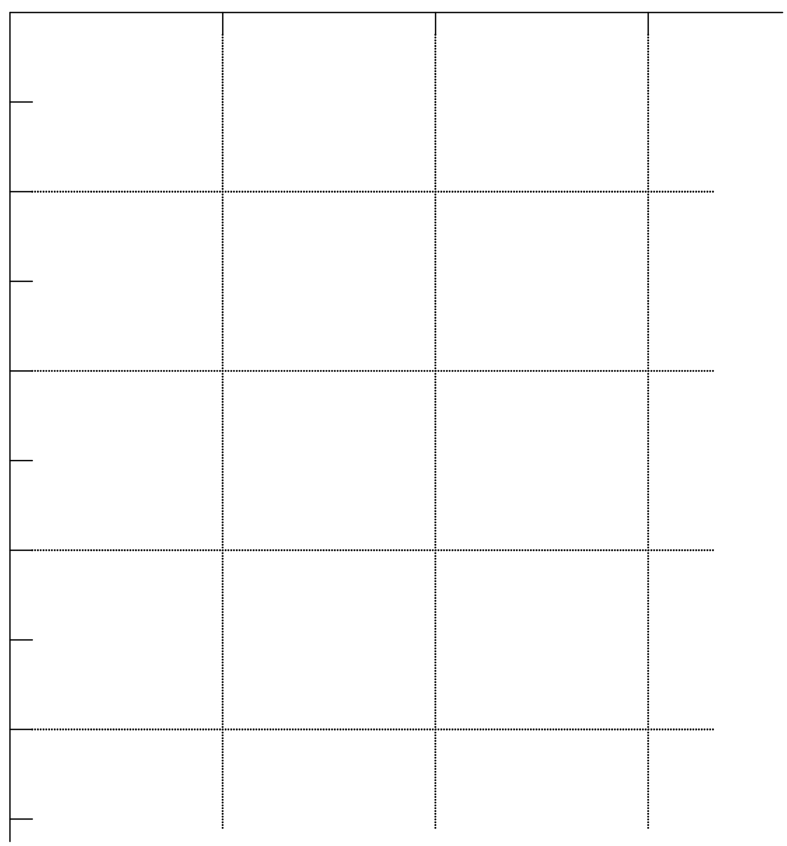
(a) Cochrane and Piazzesi AER decisively reject their single-factor model. Yet they ignore this rejection and trumpet the single factor model as a great success. Why?

(b) Expected returns are always earned for covariance of returns with shocks. According to Cochrane and Piazzesi's "Decomposing the yield curve" what is the important shock, covariance with which drives expected bond returns? Do Lustig, Verdelhan and Roussanov find that the same structure works across countries?

9. (10) The current one year log yield $y_t^{(1)}$ is 5%, and the current 2 and 3 year forward rates $f_t^{(2)}, f_t^{(3)}$ are 10% and 15% respectively.

(a) Plot the expected log bond prices through time according to the expectations hypothesis. Put numbers as well as lines on the graph below. (Connect $p_0^{(2)}$ with $p_1^{(1)}$, i.e. track a given bond's progress through time, rather than connecting $p_0^{(2)}$ with $p_1^{(2)}$, i.e. tracking bonds of given maturity through time.)

(b) Plot expected log bond prices for the first year according to Fama and Bliss' regressions, simplifying coefficients to 1 and 0.

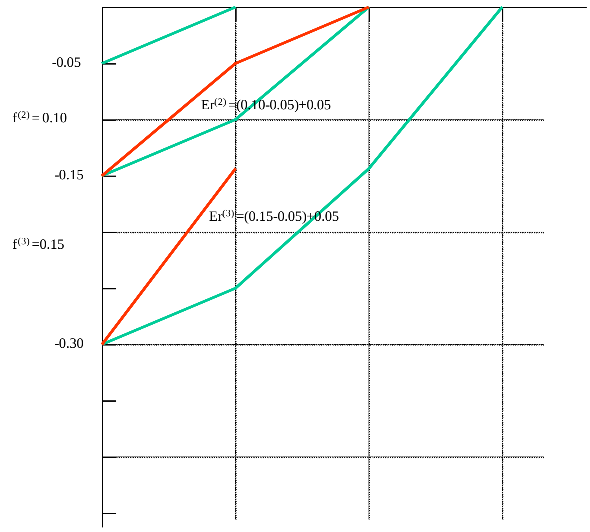


ANSWER

(a) The prices are

$$y^{(1)} = 0.05;\ f^{(2)} = 0.10;\ f^{(3)} = 0.15$$
$$p^{(1)} = -0.05;\ p^{(2)} = -0.15; p^{(3)} = -0.30$$

$f^{(2)} - y^{(1)} = 5\%$ so the 2 year bond returns $rx^{(2)} = 5\%$ or $r^{(2)} = 10\%$.$f^{(3)} - y^{(1)} = 10\%$ so $r^{(3)} = 15\%$



10. Suppose the forward curve is uniformly upward sloping. Under the expectations hypothesis, we know this means that we expect short rates $y^{(1)}_{t+n-1}$ to rise.

    (a) Does the expectations hypothesis imply that forward rates $n \geq 2$ are expected to rise, decline, or stay the same?

    (b) Same question, but suppose instead that Fama and Bliss are right, ignore the constant and set the relevant regression coefficients to one or zero to make it simple.

    Hint: Use the expectations hypothesis or FB regressions to find an expression for $E_t f^{(n)}_{t+1}$ as a function of $f^{(\cdot)}_t$

    ANSWER

    (a) Under expectations, forward rates obey

$$E_t f^{(n-1)}_{t+1} = f^{(n)}_t$$

This is a nice generalization of $f^{(n)}_t = E_t(y^{(1)}_{t+n-1})$. (Even more generally, $f^{(n)}_t = E_t\left(f^{(n-j)}_{t+j}\right)$) .Proof.

$$E_t r^{(n)}_{t+1} = y^{(1)}_t$$
$$E_t p^{(n-1)}_{t+1} = p^{(n)}_t + y^{(1)}_t$$

$$E_t f^{(n-1)}_{t+1} = E_t\left(p^{(n-2)}_{t+1} - p^{(n-1)}_{t+1}\right) = \left(p^{(n-1)}_t + y^{(1)}_t\right) - \left(p^{(n)}_t + y^{(1)}_t\right)$$
$$E_t f^{(n-1)}_{t+1} = f^{(n)}_t$$

Now, differencing,

$$E_t f^{(n-1)}_{t+1} - f^{(n-1)}_t = f^{(n)}_t - f^{(n-1)}_t$$

Thus, *if the forward curve is upward sloping we expect all forward rates to rise.*

(b) If Fama and Bliss are right, we have instead

$$E_t\left(f_{t+1}^{(n-1)}\right) = f_t^{(n-1)}$$

Proof:

$$\begin{aligned}
Er_{t+1}^{(n)} - y_t^{(1)} &= f_t^{(n)} - y_t^{(1)} \\
E_t p_{t+1}^{(n-1)} - p_t^{(n)} - y_t^{(1)} &= f_t^{(n)} - y_t^{(1)} \\
E_t p_{t+1}^{(n-2)} - p_t^{(n-1)} - y_t^{(1)} &= f_t^{(n-1)} - y_t^{(1)} \\
E_t f_{t+1}^{(n-1)} - f_t^{(n)} &= f_t^{(n-1)} - f_t^{(n)} \\
E_t f_{t+1}^{(n-1)} &= f_t^{(n-1)}
\end{aligned}$$

This nicely generalizes the FB view – everything is a random walk, not just the one year rate. Now, *forward rates are expected to stay the same, no matter what the yield curve does*

FX and international finance

1. (5) Brandt, Cochrane and Santa Clara say that discount factors are highly correlated across countries. But people exhibit a lot of "home country bias," and stock markets are not that well correlated across countries. How do we resolve this apparent contradiction? (Hint: it might be useful to think about two countries with uncorrelated stock markets and a constant exchange rate)

   ANSWER: BCSC construct discount factors that simultaneously price the different countries's returns. This has nothing to do with portfolios. In the example, a single $m = a - b_1 R^d - b_2 R^f$ would price both returns.

2. (5) The US and Europe each have stock returns with $E(R^e) = 0.08$; $\sigma(R^e) = 0.06$ and $R^f = 1.0$ (i.e., zero interest rate $r^f = 0$). The correlation between US and European stock returns is 0.40. The exchange rate between dollars and Euros has standard deviation of 20%, $\sigma(\ln(S_{t+1}/S_t)) = 0.2$. What is the correlation coefficient between the minimum-variance Euro and dollar log discount factors?

   ANSWER. This is just "did you read the paper"

$$\begin{aligned}
M_d &= M_f S_{t+1}/S_t \\
m_d &= m_f + \Delta s \\
\Delta s &= m_d - m_f \\
\sigma^2(\Delta s) &= \sigma^2(m_d) + \sigma^2(m_f) - 2\rho\sigma(m_d)\sigma(m_f) \\
.04 &= .25 + .25 - 2\rho \times 0.25 \\
0.04 &= (2 - 2\rho) \times 0.25 \\
0.02 &= (1 - \rho) \times 0.25 \\
(1 - \rho) &= \frac{0.02}{0.25} = 0.08 \\
\rho &= 0.92
\end{aligned}$$

3. Lustig and Roussanov and Verdelhan formed 6 portfolios. The mean annual returns on these 6 portfolios (percent per year) are

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ -2.92 & 0.02 & 1.40 & 3.66 & 3.54 & 5.90 \end{bmatrix}$$

They performed an eigenvalue decomposition of the covariance matrix of these returns,

$$Q\Lambda Q' = cov(R^e, R^{e\prime})$$

Here are their results:

$$Q = \begin{bmatrix} 0.43 & 0.41 & -0.18 & 0.31 & 0.72 & 0.03 \\ 0.39 & 0.26 & -0.14 & -0.02 & -0.44 & 0.75 \\ 0.39 & 0.26 & -0.46 & -0.38 & -0.31 & -0.57 \\ 0.38 & 0.05 & 0.72 & -0.56 & 0.16 & -0.01 \\ 0.42 & -0.11 & 0.38 & 0.66 & -0.37 & -0.31 \\ 0.43 & -0.82 & -0.28 & -0.10 & 0.18 & 0.11 \end{bmatrix}$$

$$100 \times \frac{\Lambda_i}{\sum \Lambda_i} = \begin{bmatrix} 70 & 12 & 6.2 & 4.5 & 3.8 & 3.2 \end{bmatrix}$$

(a) How did they construct these portfolios?

(b) What does the Q matrix mean? What do the Λ_i represent? Interpret the factors

(c) In the $E(R^e) = \beta\lambda$ (different λ) tradition, which factor accounts for most of the spread in average returns across their portfolios?

ANSWER

(a) The portfolios are formed by putting countries with high forward-spot (= interest rate) spreads over the US in portfolio 6, and low or negative forward-spot spreads in portfolio 1.

(b) Q is both the *loading* – how much each portfolio return moves when a factor moves $y_t = q_1 x_t + q_2 x_{2t}$...or $y_t = Qx_t$– and the *weight* – how to form the factor return from the 6 portfolio returns $x_t = Q' y_t$. Thus, the first column represents "level" or US dollar appreciation; the second represents "carry trade risk" or the chance that all the high spread countries (where you invested) go down and the negative spread countries (where you borrowed) go up. The Λ give fraction of variance explained and the factor covariance matrix $cov(xx')$.

Cross section of stock returns

1. (5) If you run cross-sectional Fama-MacBeth regressions of average returns on full-sample betas and loglinear functions of the characteristics size and book/market ratio, which set of variables drives the other out?

   ANSWER: Problem set. Loglinear functions of the characteristics drive out the betas.

2. (10) Start with the CAPM,

$$R_t^{ei} = \alpha_i + \beta_i R_t^{em} + \varepsilon_t^i$$

   Now, let's consider adding another factor F_t, which is also an excess return (hml or smb for example)

$$R_t^{ei} = \alpha_i + \beta_i R_t^{em} + \gamma_i F_t + \varepsilon_t^i$$

   (a) Suppose that the t statistic for γ_i is significant, for all i, the R^2 of the regression improves, and $E(F) > 0$ and also is statistically significant. Does that mean we should adopt this multifactor model, i.e. that we should describe average returns by

$$E\left(R_t^{ei}\right) = \alpha_i + \beta_i E\left(R_t^{em}\right) + \gamma_i E\left(F_t\right)?$$

(b) Suppose the GRS test rejects the second model, but does not reject the first model. Does that mean that the pricing errors of the second model are larger?

ANSWER:

(a) Yes, if the objective is to get a good factor decomposition of the return covariance matrix. But no to the main question. To answer that we need to know if the alphas got smaller. Equivalently, we need to test alpha in $F_t = \alpha_F + \beta_i R_t^{em} + \varepsilon_t$ or b_F in $m = a - bR^{em} - f_F F;\ 0 = E(mR^e)$

(b) The GRS test is based on $\alpha'\Sigma^{-1}\alpha$. α can be smaller but if Σ is also smaller the test rejects.

3. (Hint: Fama and French know the answers to these questions)

(a) Which gets better returns going forward, stocks that had great past growth in sales, or stocks that had poor past growth in sales? Is this pattern consistent with some pattern of factor exposures?

(b) If you sort stocks into "winners" that went up from year -5 to one year ago, and losers that went down from year -5 to one year ago, which ones do better for the next year? Is this consistent with some pattern of factor exposures?

(c) If we form a momentum portfolio, from stocks that did well last year, are the returns on that portfolio correlated with the returns on value stocks over the next year? I.e., if value stocks go up next year, do momentum stocks tend to go up, down, or remain the same?

(d) Here's an idea: Companies should issue stock and invest when the cost of capital is low, meaning expected returns are low. Thus, portfolios of companies that are repurchasing stock should have a lot higher returns going forward than portfolios of companies that are issuing stock. Does this idea work?

(e) Wait a minute – those issuing companies have high stock prices and the repurchasers low stock prices. Surely the big issuers are growth stocks and the repurchasers are value stocks, so we are just finding that value stocks have high average returns?

(f) Is this "net issues" effect pervasive, or confined to small stocks?

ANSWER

(a) High sales growth is low returns, and low hml betas.

(b) This is the "reversal" effect, the winners have low returns and low hml betas

(c) Here the winners keep winning. Alas, the winners still have low hml betas. This means that when value stocks go up, momentum stocks go down. momentum returns are negatively correlated (expost) with value returns.

(d) FF dissecting anomalies showed that net issues/repurchases are strongly associated with returns in the "right" (q theory) direction'

(e) FF controlled for value with a characteristic method.

(f) Pervasive.

4. (5) Your assignment is to evaluate the CAPM using the FF 25 portfolios on postwar data. One group member uses a pure time-series regression. She reports that the CAPM is lousy; the market premium is positive, but the alphas are huge; some alphas are even bigger than the average excess returns. The other group member uses a cross-sectional approach with a free intercept. He reports

that no, the CAPM is doing fine. The alphas are reasonable, though in this sample it seems the market premium came out negative. Can both of these results happen, or did one of them make a mistake? If a mistake, who made the mistake and how can you fix it? (Illustrate your answer with an appropriate graph. Label the axes.

5. (10) You want to compare betas to characteristics, so you run Fama MacBeth regressions on the 25 FF portfolios,

$$R_t^{ei} = \lambda_{0t} + b_i\lambda_{mt} + h_i\lambda_{ht} + s_i\lambda_{st} + \delta_t \ln(bm_{it-1}) + \eta_t \ln(size_{it-1}) + \varepsilon_{it}$$

You try with and without the free intercept. The results are

	λ_0	λ_m	λ_h	λ_s	δ	η
estimate	1.8	-0.08	-0.33	-0.11	0.42	-0.13
t	5.3	-0.21	-1.66	-0.56	4.43	-2.4
estimate	-	0.85	-0.54	0.29	0.57	0.01
t	-	2.7	-2.7	1.5	5.7	0.3

(a) Which of these parameter estimates make sense, and which are unexpected?

(b) What additional regression would you run to check the idea that characteristics should not matter, given betas, but imposing "sensible" numbers on the ones that didn't make sense in a?

ANSWER This references problem set 6 2009.

(a) λ_0 is silly. $\lambda_m, \lambda_h, \lambda_l$should all be positive, so are nuts in the first row. δ an η at least have the right sign, more for b/m and less for size. λ_m is repaired in the second row by $\lambda_0 = 0$ but not λ_h

(b) As on the problem set, impose $\lambda = E(f)$ and run

$$R_t^{ei} - b_i R_t^{em} - h_i hml_t - s_i smb_t = \delta_t \ln(bm_{it-1}) + \eta_t \ln(size_{it-1}) + \varepsilon_{it}$$

6. Should you keep the smb factor? You run regressions

$$R_t^{ei} = \alpha_i + b_i rmrf_t + h_i hml_t + s_i smb_t + \varepsilon_{it} \; t = 1, 2, ..T \text{ for each } i \quad (17)$$
$$R_t^{ei} = \alpha_i + b_i rmrf_t + h_i hml_t + \varepsilon_{it} \quad t = 1, 2, ..T \text{ for each } i \quad (18)$$
$$smb_t = \alpha_s + b_s rmrf_t + h_s hml_t + \varepsilon_{st} \; t = 1, 2, ..T \quad (19)$$

(note α_i, b_i, h_i are not the same in (17) and (18).) You find that the α in (18) are about the same as in (17), and you find that $\alpha_s = 0$. On the other hand, , $E(smb)$ is quite high and statistically significant (well, suppose that is the case), the t statistics on s_i are very strong, the R^2 in (17) is much higher than in (18) and a joint test that all $s_i = 0$ decisively rejects. So, should you keep the *smb* factor or not?

ANSWER

What model you use depends on what the purpose is! Given these facts, you can drop smb for the purpose of understanding mean returns, i.e.

$$E(R^{ei}) = \alpha_i + b_i E(rmrf_t) + h_i E(hml_t) + s_i E(smb_t)$$

Even though $E(smb) \neq 0$, the results of (18) and (19) mean that the b and h will change, so the alphas do not change. For the purpose of understanding return *variance* however, the s t stats and R^2 loudly warn you not to drop *smb*. Also, including *smb* will improve standard errors, including standard errors of α_i, so it might be a good idea in any case.

7. If you want to capture the relationship between size, book/market and subsequent returns in a regression, rather than a table of numbers, what is the right functional form?

$$E(R^{ei}_{t+1}|beme_{it}, me_{it}) = ?$$

ANSWER

This is "did you do the problem set?"

(a) We found a log transformation was necessary, and something to detrend me over time. Anything that captures this is fine, mine was

$$E(R^{ei}_{t+1}) = a + b\log(beme_{it}) + c\log\left(\frac{me_{it}}{\Sigma_i me_{it}}\right)$$

(b) We also discovered that variation over time was more important than varition across assets. Fama and French's results were basically the same as a time dummy

$$E(R^{ei}_{t+1}) = a_t + b\log(beme_{it}) + c\log\left(\frac{me_{it}}{\Sigma_i me_{it}}\right)$$

but we found stronger coefficients with a firm dummy

$$E(R^{ei}_{t+1}) = a_i + b\log(beme_{it}) + c\log\left(\frac{me_{it}}{\Sigma_i me_{it}}\right)$$

You should mention that. We did not explore it, but that means a real function – which doesn't have dummies to explain things – may want something like this

$$E(R^{ei}_{t+1}) = b_1 E_T \log(beme_{it}) + b_2\left[\log(beme_{it}) + E_T\log(beme_{it})\right] + ...$$

(c) We also found that the size slope was bigger for value than growth firms, so a cross-term is necessary, somehting like

$$E(R^{ei}_{t+1}) = a + b\log(beme_{it}) + c\log\left(\frac{me_{it}}{\Sigma_i me_{it}}\right) + d\log(beme_{it}) \times \log\left(\frac{me_{it}}{\Sigma_i me_{it}}\right)$$

I ended up dropping *beme* from this specification.

Theory and m review

1. Show that a discount factor linear in the market return

$$m_{t+1} = a - bR^m_{t+1}$$

implies the CAPM

$$E(R^{ei}) = \beta_i E(R^{em})$$

ANSWER: Start with $0 = E(mR^e)$, and use the definition of covariance. $0 = E(m)E(R^e) + cov(m, R^e), E(R^e) = -cov(m, R^e)/E(m)....)$

2. Do you expect interest rates to be higher in good times or bad times? Back up your view with an equation and an explanation.

3. An investor lives for two periods, time 0 and time 1. He has a utility function over consumption c_0 in period zero and random consumption c_1 in period 1 given by

$$-\frac{1}{2}E\left[(c^*-c_0)^2+0.95\times(c^*-c_1)^2\right]$$

c^* is a parameter (number), c_0 and c_1 are consumption in the first and second periods of life. We learn from a detailed statistical analysis that his consumption follows a random walk,

$$c_1=c_0+\epsilon_1;$$

the random shock ϵ_1 is normally distributed with mean 0 and variance σ^2. (A useful preliminary: As of time zero, i.e., knowing c_0, what is the mean and variance of c_1?) We observe consumption at period 0, c_0. It is less than c^*; $c_0 < c^*$. Your answers to the following questions can contain c_0. Find the price at time 0 (i.e. knowing c_0) of the following securities.

   (a) A one period zero coupon bond. (You may assume zero inflation if this worries you.)

   (b) A "Stock," which pays a random dividend equal to c_1 and nothing thereafter.

   (c)

      1. Is the stock price greater or less than the price of a bond with c_0 face value?
      2. How and why does the price depend σ^2?
      3. How and why does the price depend on c^*? In particular, explain what happens as c_0 gets closer and closer to c^*?

Liquidity, financial crisis, mispricing

1. Paper Questions

   (a) Did Lamont and Thaler think people were irrational because they did not short Palm and buy 3 Com? If not, how can they say people were irrational?

   (b) Name at least two piece of evidence that "Stocks as money" cites for the "convenience yield" theory as opposed to the "morons" theory or the theory that "short sales constraints means that pessimists can't express their views" in explaining the high price of Palm over 3Com.

   ANSWER

   (a) The people who bought Palm should have bought 3 Com and waited 6 month for the spinoff. NOT "short palm buy 3Com." that wasn't the point of the paper.

   (b) 1) Turnover! Is higher for Palm. 2) 3Com is delinked from Palm, so you must hold Palm stock to trade Palm 3) given high turnover, the cost of overpricing is only 1% per week 4) 3Com fell drastically when Palm trading opened.

Utility functions and applying the consumption model

1. (30) (This is from 35904, a bit harder than I am heading towards on this final, but we did talk about nonseparable utility, so it's useful) Let's think about how our asset pricing formulas would change if

we recognize that the consumption series we're using is durable. Assume a single durable good, so the representative investor objective is

$$E\sum_{j=0}^{\infty}\beta^j u(k_{t+j})\text{s.t. } k_t = (1-\delta)\,k_{t-1} + c_t$$

c_t now represents durable good purchases. General hint: This problem does *not* require lots of algebra. I used no more than 3 lines for each part. I strongly advise you to work it out on the scratch paper at the end before answering it here!

(a) (5) State the investor's first order conditions for buying an asset with price p_t and payoff x_{t+1}. How is this equation different from the standard nondurable case $pu'(c_t) = E_t\left[\beta u'(c_{t+1})x_{t+1}\right]$?

(b) (10) Now assume a constant riskfree rate $R^f = 1/\beta$. Use the equation for pricing the risk free rate to collapse the new terms, so you have an asset pricing equation $p = E(mx)$ expressed in terms of $u'(k_t)$ and $u'(k_{t+1})$.

(Hints: 1) Do the $\delta = 1$ case first, then show this solution works for the $\delta < 1$ case. You do not have to prove this is the only solution. 2) If you're having trouble, start with quadratic utility, and then generalize to arbitrary $u(k)$.)

(c) (7.5) In the case of power utility, express the discount factor in terms of c_{t+1}/c_t and a purchases/stock c_t/k_t ratio. Suppose as in the Campbell/Cochrane model that the variance of purchases growth $\sigma(c_{t+1}/c_t)$is constant over time, When does this model generate high risk premia – in booms when purchases are high relative to the stock of durables or in recessions when purchases are low relative to the stock?

(d) (7.5) Express the model in continuous time,

$$\max E_0 \int_0^{\infty} e^{-\rho t}u(k_t)dt$$
$$dk_t = -\delta k_t dt + c_t dt$$

assume c follows a diffusion process and power utility $u'(k) = k^{-\gamma}$. Assume your results from part a, b go through so $\Lambda_t = e^{-\rho t}u'(k_t)$. By characterizing this discount factor, do risk premia increase or decrease in this model relative to the nondurable model? How might you modify this continuous-time setup to generate the opposite result (one sentence)?

2. Show that the asset pricing predictions of internal vs. external habit models are the same for power utility, an AR(1) habit and linear technology.

3. Paper questions

(a) Campbell and Cochrane claim that they produce imperfect correlation between consumption growth and stock returns. Yet their model has a single shock – doesn't this mean every variable has to be perfectly correlated?

4. (5) In a complete market, all marginal utilities move together. What is the generalization of this proposition if markets are incomplete. ("Incomplete" here means that some states of nature are not spanned by asset payoffs, not that there are frictions such as transactions costs for existing assets.)

ANSWER

$$p = E(m^i x) = E(proj(m^i|X)x) = E(m^* x)$$

thus the *projection* (fitted value of regression) of each of our marginal utilities on the common set of asset payoffs should be the same. In this sense we should use existing asset markets as much as possible to share risks. (this is also a n "international" question as it is used in "international diversification is better than you think"

5. Short answers

   (a) T/F/U and why. "By force of habit" is a one-shock model, so all series must be perfectly correlated. Therefore, this model cannot address the "correlation puzzle" in the data, that consumption growth and asset returns are poorly correlated with each other.

   (b) T/F/U and why. "By force of habit" ends up with a marginal utility process

$$m_{t,t+k} = \beta \left(\frac{S_{t+k}}{S_t}\right)^{-\gamma} \left(\frac{C_{t+k}}{C_t}\right)^{-\gamma}$$

   where $S_t = \frac{C_t - X_t}{C_t}$ is the "surplus consumption ratio" and is a stationary variable. Alas, this fact means that while the conditional variance of consumption growth increases linearly with horizon, the conditional variance of S growth must approach a limit, the unconditional variance of S. Therefore, this model cannot generate a "long run" equity premium, matching the Sharpe ratio of "long run" returns $R_{t,t+k}$

ANSWER

   (a) It's all conditionally perfectly correlated, but the slope of that correlation changes over time. Time aggregation also helps.

   (b) $S^{-\gamma}$ is not stationary, and its variance explodes.

   (c) Add: at some point, how does CC compare with the 2-shock VAR?

Production and general equilibrium

1. Suppose the technology side of a model is

$$y_t = \alpha k_t - \frac{1}{2}\eta\left(\frac{i_t}{k_t}\right) i_t$$
$$k_{t+1} = (1-\delta)k_t + i_t$$

What is the investment return – the one period return available from investing a little more at t and disinvesting at t+1?. (You do not need to be formal. It's enough to get the right answer by marginal change logic.

ANSWER

The marginal cost of a unit of capital is

$$1 + c_i(t) = 1 + \eta\left(\frac{i_t}{k_t}\right)$$

The marginal benefit at time $t+1$ is

$$a - c_k(t+1) = \alpha + \frac{1}{2}\eta\left(\frac{i_{t+1}^2}{k_{t+1}^2}\right)$$

so the investment return is

$$\frac{(1-\delta)\left[1+\eta\left(\frac{i_t}{k_t}\right)\right]+\alpha+\frac{1}{2}\eta\left(\frac{i_{t+1}^2}{k_{t+1}^2}\right)}{1+\eta\left(\frac{i_t}{k_t}\right)}$$

Portfolio Theory

1. (5) The standard portfolio allocation rule is

$$w=\frac{1}{\gamma}\frac{E(R^e)}{\sigma^2(R^e)}.$$

Thus, using standard numbers, it makes sense to hold 100% equity if

$$1=\frac{1}{\gamma}\frac{0.08}{0.16^2};\quad \gamma=3.125$$

Yet the equity premium literature complains that at $\mu-r=8\%,\sigma=16\%$, investors with risk aversion below 10 should want to buy much more stock. How do we reconcile these two calculations? (Note: the "standard portfolio allocation rule" is the solution to $\max_{\{c_t,w_t\}} E\int e^{-\delta t}\frac{c_t^{1-\gamma}}{1-\gamma}dt$ st. $\frac{dW_t}{W_t}=r^f+w_t'\left(\frac{dS}{S}-r^f\right)dt-c_t dt;\ \frac{dS}{S}=\mu dt+\sigma dB$)

ANSWER

The standard method also predicts $c=kW$ so $\sigma(\Delta c)=\sigma(R)=18\%$. If consumption had 18% volatility, all would be well with the world.

2. An investor with log utility $u(W_{t+1})=\ln(W_{t+1})$ can invest in a stock which currently has price \$100. It will either go up to \$130 or down to \$90, with probability 1/2 of each event. (Call the two states u and d.) He can also invest in a bond, which pays zero interest–A \$100 investment gives \$100 for sure.

   (a) Find a discount factor m_{t+1} that prices stock and bond.

   (b) Find the return of the investor's optimal portfolio in the two states.

ANSWER

(a)

$$100=E(mS)=\frac{1}{2}\times m_u\times 130+\frac{1}{2}\times m_d\times 90$$
$$100=E(m\times 100)=\frac{1}{2}\times m_u\times 100+\frac{1}{2}\times m_d\times 100$$

$$200=m_u\times 130+m_d\times 90$$
$$2=m_u+m_d$$

$$200=m_u\times 130+(2-m_u)\times 90$$
$$200=m_u\times 130+180-90\times m_u$$
$$20=m_u\times 40$$
$$m_u=1/2;\ m_d=3/2$$

(b)

$$\begin{aligned} u'(W) &= \lambda m \\ \frac{1}{W} &= \lambda m \\ W_0 &= E(mW) = E(\frac{m}{\lambda m}) = 1/\lambda \\ \frac{1}{W} &= \frac{1}{W_0} m \\ R^p &= \frac{1}{m} \end{aligned}$$

The return is 2 in the up state and 2/3 in the down state.

(c) I did not ask, what portfolio of stock and bond gets this. The answer:

$$W_{t+1,u} = \$100 \times 2 = \$200;\ W_{t+1,d} = \$100 \times \frac{2}{3} = \$66.67$$

What portfolio gets this?

$$\begin{aligned} W^u &= 200 = hS_u + kB_u = h \times 130 + k \times 100 \\ W^d &= 66.67 = hS_d + kB_d = h \times 90 + k \times 100 \end{aligned}$$

$$\begin{aligned} 200 - h \times 130 &= k \times 100 \\ \frac{200}{3} &= h \times 90 + 200 - h \times 130 \\ 200\frac{2}{3} &= h \times 40 \\ h &= \frac{400}{120} = \frac{10}{3} = 3.333 \\ k &= 2 - h \times 1.3 = 2 - \frac{10}{3} \times 1.3 = -2.333 \end{aligned}$$