25 Week 9. Term structure I Expectations Hypothesis and Bond Risk Premia – Notes

1. Background: We're back to regressions $R_{t+1} = a + bx_t + \varepsilon_{t+1}$ to forecast returns. We will connect them to portfolios, a job that is just beginning.

2. Definitions

3. Expectation hypothesis – 3 statements.

4. Bond risk premia (how expectations fails) – Fama/Bliss; Cochrane/Piazzesi.

5. Foreign exchange, expectations and carry trade. The link between forecasts and portfolios.

25.1 Definitions

1. *Discount bond*: A promise to pay \$1 at time $t + n$. (This is the same thing as a zero coupon bond.)

2. *Price*

$$P_t^{(n)} = \text{ price of } n \text{ year discount bond at time t}$$

$$\text{Example. } P_t^{(2)} = 0.9 \Longleftrightarrow 10\% \text{ interest rate}$$

   (n) distinguishes maturity from raising a number to a power.

$$p_t^{(n)} = \log \text{ price of } n\text{-year discount bond at time } t.$$

$$\text{Example: } \ln(0.9) = -0.10536 \approx -0.1 \text{ or "10\% discount}$$

3. Small letters = log of large letters. (ln). Note

$$\begin{aligned} \ln(1+x) &\approx x; \\ e^x &\approx 1 + x \text{ for small } x. \end{aligned}$$

   Thus $ln(1.1) \approx 0.1 = 10$ % return. Why logs? The two year return is $R_1 R_2$. The two year log return is $r_1 + r_2$. Adding is a lot easier than multiplying.

4. *Yield.*

   (a) Concept: a discount rate for bonds.

   (b) Definition (words) *What constant discount rate would make sense of the bond price, if we knew the future for sure and there were no default?*

   (c) Definition (formula). For any bond with CF_j = cashflow at time j, Y solves

$$P = \sum_j \frac{CF_j}{Y^j}.$$

   You have to find this numerically.

(d) For discount bonds,

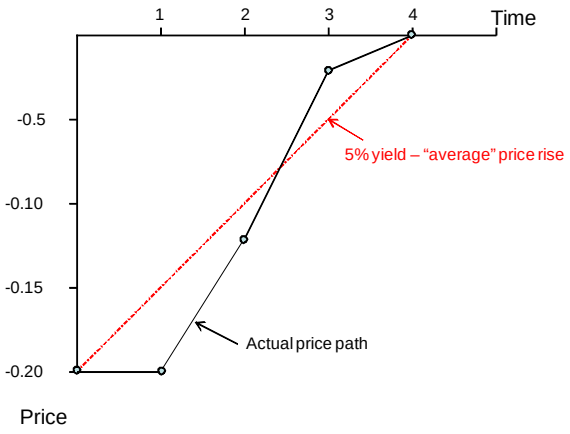
$$P_t^{(n)} = \frac{1}{\left[Y_t^{(n)}\right]^n} \leftrightarrow Y_t^{(n)} \equiv \left[P_t^{(n)}\right]^{-\frac{1}{n}}$$

$$y_t^{(n)} \equiv -\frac{1}{n} p_t^{(n)}.$$

$$\text{Example: } p_t^{(2)} = -0.1 \rightarrow y_t^{(2)} = \frac{0.1}{2} = 0.05, \text{ “5\% discount per year.”}$$

(e) *Point*: yield allows you to compare bonds of different maturity and coupon on an even basis. (It's like implied volatility for options.) *It's just another more convenient way to rewrite the price.* It is not the expected return you will earn for one year! It is the compound average yield to maturity if the bond does not default – but not the expected return.

(f)



5. *Forward rate.*

(a) Definition: The rate at which you can contract today to borrow from time $t+n-1$ and pay back at time $t+n$.

(b) Fun fact: you can synthesize this forward loan by shorting $t+n-1$ zeros and buying $t+n$ zeros[17]. Hence, we can find the forward rate from today's zero rates.

$$\begin{aligned} F_t^{(n)} &= F_t^{(n-1\rightarrow n)} \equiv \frac{P_t^{(n-1)}}{P_t^{(n)}} \\ f_t^{(n)} &= f_t^{(n-1\rightarrow n)} \equiv p_t^{(n-1)} - p_t^{(n)} \end{aligned}$$

(c) Example.

$$p_t^{(3)} = -0.15;\ p_t^{(2)} = -0.10 \rightarrow f_t^{(2\rightarrow 3)} = 0.05\ = 5\%$$

6. *Holding period return.*

[17] Move the money from $t+n-1$ to t and then back to $t+n$. \$1 at $t+n-1$ transforms to $\$P_t^{(n-1)}$ dollars at t. \$1 at t transforms to $\$1/P_t^{(n)}$ at$t+n$. Thus, $\$P_t^{(n-1)}$ transforms to $P_t^{(n-1)}/P_t^{(n)}$ at $t+n$.

(a) Definition (words). Buy an n-year bond at time t and sell it – now an $n-1$ year bond – at time $t+1$.

(b) Definition (equations)

$$R_{t+1}^{(n)} = \frac{P_{t+1}^{(n-1)}}{P_t^{(n)}}$$

$$r_{t+1}^{(n)} \equiv p_{t+1}^{(n-1)} - p_t^{(n)}.$$

(Note: *hpr* for "holding period return" in *Asset Pricing*. r is enough though.)

(c) Excess log returns (over the risk free rate)

$$rx_{t+1}^{(n)} \equiv r_{t+1}^{(n)} - y_t^{(1)}$$

Note: this is convenient but tricky. It's not really a zero cost portfolio since it's logs!

7. Why we focus on zeros. Zero coupon bonds are the a la carte menu. Coupon bonds are a bundle of zeros. CF_j = cashflow at time j. Then,

$$\text{Coupon bond } = \text{CF}_1 \times 1 \text{ year zero} + \ + \text{CF}_2 \times 2 \text{ year zero} + ...$$

Hence,

$$\text{Coupon bond price} = P^{(1)} \times \text{CF}_1 + P^{(2)} \times \text{CF}_2 + ...$$

8. Summary (you need to know these):

$$\begin{aligned} p_t^{(n)} &= \log(P_t^{(n)}) \text{ e.g. -0.1} \\ y_t^{(n)} &= -\frac{1}{n} p_t^{(n)} \\ f_t^{(n)} &= p_t^{(n-1)} - p_t^{(n)} \\ r_{t+1}^{(n)} &= p_{t+1}^{(n-1)} - p_t^{(n)} \\ rx_{t+1}^{(n)} &= r_{t+1}^{(n)} - y_t^{(1)}. \end{aligned}$$

9. Notation: (n) denotes maturity. t denotes time, when each item is observed. If you want to be really clear, use arrows for buy and sell dates. For example, buying an n period bond, holding it for a year, and selling it as an n-1 period bond gives a return

$$r_{t+1}^{(n)} = r_{t \to t+1}^{(n \to n-1)}$$

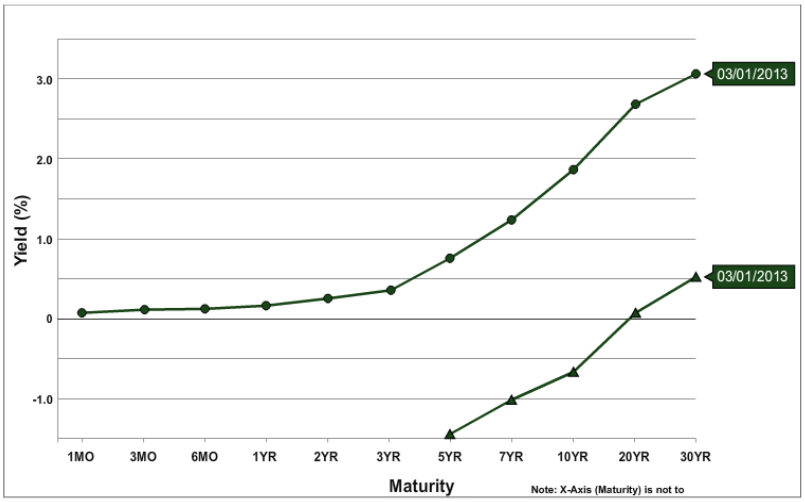
25.2 Expectations hypothesis

1. The *yield curve:* a plot of bond yields as a function of maturity. (*Forward curve:* forward rates as a function of maturity)

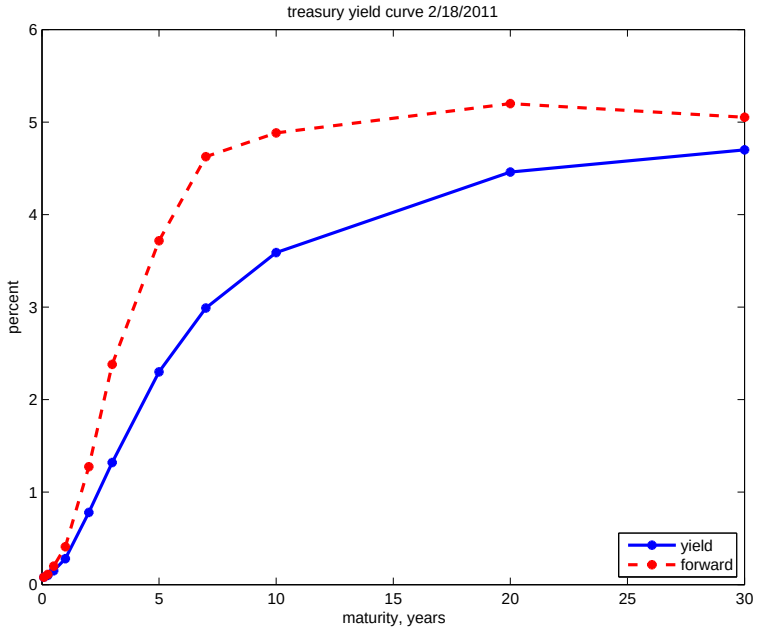
(a) Why does the yield curve sometimes slope up and sometimes slope down?

(b) Expectations hypothesis: An upward sloping curve means people think short term interest rates will rise in the future.

2. Today's yield curve (Nominal and real = TIPS)[18]



Be careful though, the x axis is stretched so it has different time intervals. This is a common presentation, and useful for many purposes, but it gives a misleading sense of upward curvature at the short end. Here is a yield curve with maturity evenly spaced, and adding forward rates



3. Why are yields different for bonds with different maturity? Three equivalent statements:

   (a) *Long maturity yield = average of expected future short rates (plus risk premium)*

   (b) *Forward rate = expected future spot rate (plus risk premium)*

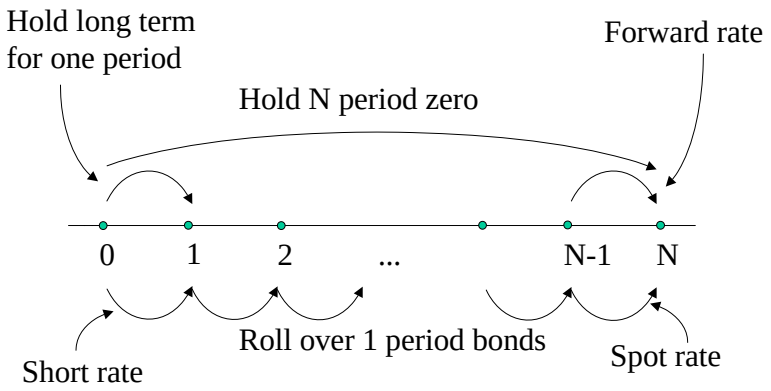
[18] http://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/Historic-Yield-Data-Visualization.aspx

(c) *Expected holding period returns should be equal across maturities (plus risk premium)*

Thus, the yield curve rises if people expect short rates to be higher in the future. The decline in future bond prices cuts off the great yield, so you don't earn any extra in long bonds for the first year.

4. Example: Today's yield curve (under EH) implies that markets expect interest rates to start rising 1 year form now, recover to about 4.8% by year 6 and then stay there.

5. Big picture: These are all relatives of the random walk for stock prices, and (approximately) risk neutral arbitrageurs. Alternative ways of getting money from one date to another must have the same expected value (plus risk premium). It's an obvious principle with many surprising implications.



25.2.1 Yields and future interest rates

1. *Two ways of getting money from now (0) to N should give same expected return.* Return from 0 to N :

Buy N year zero	Rollover
$r_{0\to N} = -p^{(N)} = Ny^{(N)}$	$y_0^{(1)} + y_1^{(1)} + y_2^{(1)} + ...y_{N-1}^{(1)}$

2. *Expected* return should be the same$\rightarrow$

$$y_0^{(N)} = \frac{1}{N} E\left(y_0^{(1)} + y_1^{(1)} + y_2^{(1)} + ...y_{N-1}^{(1)}\right) \text{ (+risk premium)}$$

3. *Long yield = average of expected future short rates (plus risk premium).*

4. Note: some people do this in levels rather than logs,

Buy N year zero	Rollover
$R_{0\to N} = \frac{1}{P^{(N)}} = \left[Y^{(N)}\right]^N$	$R_0 R_1 R_2 ... R_{N-1} = Y_0^{(1)} Y_1^{(1)} .... Y_{N-1}^{(1)}$

$$Y_0^{(N)} = E\left[\left(Y_0^{(1)} Y_1^{(1)} .... Y_{N-1}^{(1)}\right)^{\frac{1}{N}}\right] \text{ (+risk premium)}$$

$E(\ln(X)) \neq \ln(E(X))$ so the level and log statements are not exactly equivalent. They are separate embodiments of the "Expectations hypothesis." We – like all modern term structure models – use logs. (The difference is only in small $1/2\sigma^2$ terms)

5. Question: What does a yield curve that rises with maturity mean? Answer: That short rates are expected to rise in the future.

25.2.2 Forward Rate

1. *Two ways of getting money from N to N+1 should give same expected return.* Why? Lock in now, or wait and borrow/lend at the spot rate.

2. *Forward rate =* **expected** *future spot rate (plus risk premium)*

$$\begin{aligned} \text{lock in} &= \text{E(wait and use spot) (+risk premium)} \\ f_t^{(N)} &= E_t\left[y_{t+N-1}^{(1)}\right] \text{ (+risk premium)} \end{aligned}$$

3. Forward rates reveal "market's forecast of interest rates" directly.

4. What does a rising forward curve mean? That short rates will be higher in the future.

25.2.3 Holding period returns

1. *Two ways of getting money from t to t+1 should be the same. Expected holding period returns should be equal across maturities (plus risk premium)*

$$\begin{aligned} \text{hold N period bond 1 year} &= \text{hold 1 period bond for 1 year} \\ E_t\left[r_{t\to t+1}^{(N)}\right] &= E_t\left[r_{t\to t+1}^{(1)}\right] = y_t^{(1)} \text{ (+risk premium)} \\ E_t\left[rx_{t\to t+1}^{(N)}\right] &= E_t\left[rx_{t\to t+1}^{(M)}\right] = 0 \text{ (+risk premium)} \end{aligned}$$

2. This insight is again algebraically equivalent to the others.

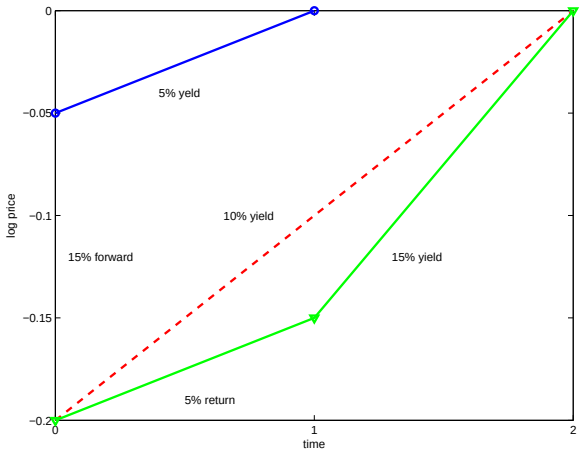
25.2.4 Example

1. Example:

   (a) One year yield = 5%. Two year yield = 10%. Prices: $p_t^{(1)} = -0.05$ $p_t^{(2)} = -0.20$.

   (b) Does this mean you expect to earn more on two year bonds? No – we should expect to earn the same *for the first year.* Thus, the EH tells us what $E_t\left(p_{t+1}^{(1)}\right)$is: $E_t(r_{t+1}^{(2)}) = E_t\left(p_{t+1}^{(1)} - p_t^{(2)}\right) = y_t^{(1)} = 0.05$. This means $E_t(p_{t+1}^{(1)}) = -0.15$ or $E_t(y_{t+1}^{(1)}) = 0.15$.

   (c) $E_t(y_{t+1}^{(1)}) = 0.15$: *We expect the short rate to rise in the future, so that you earn the same on long term bonds as short term bonds this year, despite the higher initial yield on long term bonds.*



(d) Equivalently:

i. Long yield (10%) = average of expected future short yields (5%, 15%).

ii. The expected *two* year return is the same from rolling over (5% + 15%) as from holding the two year bond (20%).

iii. The expected one-year return is the same (5%) whether you hold the one period bond or the two period bond.

iv. Forward rate $= p_t^{(1)} - p_t^{(2)} = -0.05 - (-0.20) = 0.15 =$ expected one year rate one year out (15%).

(e) Hopefully the graph emphasizes how different statements of the expectations hypothesis are equivalent. There is one thing we don't know ahead of time – where $p_{t+1}^{(1)}$ will end up (green triangle in the middle of the graph). As you move it up and down, this year's return, next year's one-year yield both change together.

(f) Note the green line is how the *expectations hypothesis* says things will work out. That's not necessarily the true *expected* path, and it certainly isn't the actual outcome!

25.2.5 Exchange rates

1. Notation:

$$\begin{aligned} S_t &= \text{Exchange rate (Euros/Dollar) at time t} \\ R_t^{US} &= \text{US interest rate (in Dollars)} \\ R_t^{F} &= \text{Foreign interest rate (in, e.g. Euros)} \end{aligned}$$

2. Should you invest in Euro bonds or dollar bonds? The realized dollar return from Euro bonds is the Euro return (interest rate) less any depreciation of the Euro

$$\begin{aligned} R_{t\to t+1}^{US} &= \frac{\$_{t+1}}{\$_t} = \frac{\text{euro}_{t+1}}{\text{euro}_t} \times \frac{\text{euro}_t/\$_t}{\text{euro}_{t+1}/\$_{t+1}} = R_{t\to t+1}^{Euro} \times \frac{S_t}{S_{t+1}} \\ r_{t\to t+1}^{US} &= r_{t\to t+1}^{Euro} + s_t - s_{t+1} \end{aligned}$$

3. The expectations hypothesis says:

(a) *The expected dollar rate of return on dollars vs. Euros should be the same (plus risk premium)*

$$E_t\left(r^{US}_{t\to t+1}\right) = E_t\left(r^{Euro}_{t\to t+1}\right) + s_t - E_t\left(s_{t+1}\right) + \text{risk premium}$$

(b) In particular, for one-period interest rates

$$y_t^{(1)US} = y_t^{(1)Eu} + s_t - E_t\left(s_{t+1}\right)$$

(c) Rearranging, *the interest spread should equal the expected depreciation*

$$y_t^{(1)Eu} - y_t^{(1)Us} = E_t\left(s_{t+1}\right) - s_t$$

If Euro rates are higher, we expect the Euro to depreciate, i.e. more Euros per dollar in the future.

Often people use i or r for interest rates, and call i^* or r^* the foreign rate.

$$\begin{aligned} i_t^* - i_t &= E_t\left(s_{t+1}\right) - s_t \\ r_t^* - r_t &= E_t\left(s_{t+1}\right) - s_t \end{aligned}$$

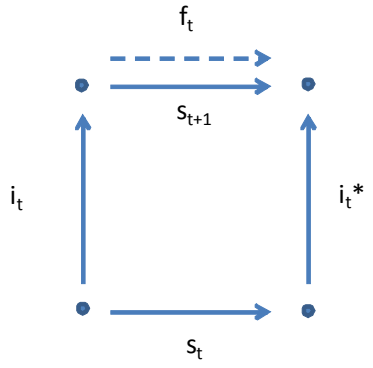
(Note: if you express exchange rates in dollars / euro, you get the opposite sign,

$$i_t^* - i_t = s_t - E_t\left(s_{t+1}\right)$$

Euros are in fact quoted in dollars per euro. Yen are quoted as yen per dollar. Keep the sign straight!)

4. Example: Just before the 1997 Asian currency crashes, crash, local rates were 20%, dollar rates are 5%. This is a sense in which high interest rates correspond to high (about to decline) exchange rates. (And have nothing to do with "speculators," "tight monetary policy," etc.)

5. Forward rate statements



(a) There are forward markets for currencies. Thus,

i. By arbitrage, going around the box gets zero.

$$i_t + f_t - i_t^* - s_t = 0$$

ii. "Covered interest parity" The forward-spot spread = the interest differential

$$i_t^* - i_t = f_t - s_t$$

you "cover" the exchange rate change and have no risk.

iii. "Uncovered interest parity." We can express the expectations hypothesis (or trade its violations!) as the forward-spot spread equals expected appreciation

$$E_t\left(s_{t+1}\right) - s_t = i_t^* - i_t = f_t - s_t$$

This is "uncovered interest parity" because you do not "cover" the exchange rate risk.

iv. Note: This is still a "hypothesis!" Yes, it's false, at least in the short run. "Covered interest parity" is arbitrage, a much stronger hypothesis.

25.3 Yield curve summary

1. Buy a long term bond, hold to n vs. rollover interest rate. *Long yield = average of expected future short rates (plus risk premium)*

$$y_0^{(n)} = \frac{1}{n} E\left(r_0 + r_1 + r_2 + ...r_{N-1}\right) + (\text{risk premium})$$

A rising yield curve means interest rates are expected to rise in the future.

2. Lock in forward rate vs. wait, borrow/lend at spot rate *Forward rate =* **expected** *future spot rate (plus risk premia)*

$$f_t^{(n \to n+1)} = E_t(y_{t+n}^{(1)}) + (\text{risk premium})$$

A rising forward curve means interest rates are expected to rise in the future.

3. Hold a short term bond for one period, vs. hold a long-term bond for one period? *Expected holding period returns should be equal across maturities (plus risk premia)*

$$E_t\left[r_{t \to t+1}^{(n)}\right] = y_t^{(1)} + (\text{risk premium})$$

4. Exchange rates *Us Rate = Foreign rate + expected depreciation of foreign currency (plus risk premium)*

$$r_t^{US} = r_t^{Euro} + s_t - E_t s_{t+1} + (\text{risk premium})$$

5. Many more. For example, expected *two-year* returns should be the same. *Two-year* forward rate ($t+3$ to $t+5$) should equal expected two-year yield in year $t+3$. The two-year Euro vs. Us yield spread equals the two-year expected depreciation.

6. Fact: 1-3 are *algebraically equivalent.* If any one holds, then all the others do too.

25.4 Risk premia

1. What risk premium do we expect? Which end of these ways to get money from x to y is riskier?

   (a) Think of one year holding period. It seems that the long term bond is riskier, long yields should be higher to compensate investors for risk. (This is the usual idea).

   (b) How about a 10 year holding period? For a 10 year horizon, it seems the 10 year bond is safer. Rolling over short bonds, interest rates could rise and decline. That suggests that 10 year yields should be *lower*–yield curves should decline on average!

   (c) An important lesson here: *for most investors, the riskless asset is a long term bond! The coupon only part of a TIP should be the bedrock riskless asset in a portfolio, NOT "cash"*

   (d) What if variation in interest rates is all due to changes in *inflation* not changes in *real interest rates*? When inflation rises, the short term rates rise 1-1. In this case, the *real* 10 year return is safer if you roll over *nominal* short term bonds (money market) rather than hold *nominal* 10 year bonds.

   (e) An interesting observation: In the 19th century, under the gold standard, there was little long-term inflation risk. Interest rate variation was real interest rates. The yield curve sloped *downward* on average, long-term bonds are safer for long-term investors.

   (f) (Long-term investors who look for quarterly performance in their bond portfolios are nuts. We *know* that bond yields rise when bond prices fall, so the long-term value is unaffected by short-term losses.)

   (g) Once it exists, the risk premium may vary over time. All risk premia go up in bad times, and the pattern of real/nominal, and demand duration/supply duration may change.

   Summary: *The sign of risk premium can go either way; it depends on investor's horizon relative to supply of bonds, and whether real interest rates or inflation are the source of interest rate risk.* Thus, we'll (for now) let the data speak on risk premium size and sign rather than impose any theory.

2. Theory II: as always,

$$\begin{aligned} E_t(R^e_{t+1}) &= cov_t(R^e_{t+1}, m_{t+1}) \\ &= \gamma_t cov_t\left(R^e_{t+1}, \Delta c_{t+1}\right) \\ &\approx -\gamma_t \; n \; cov_t\left(y^{(n)}_{t+1}, \Delta c_{t+1}\right) \end{aligned}$$

   The risk premium can go either way – do interest rate surprises come in good times or bad times?

   (In the last equation I make the approximation $R^e_{t+1} \approx rx^{(n+1)}_{t+1} = p^{(n)}_{t+1} - p^{(n+1)}_t - y^{(1)}_t = -ny^{(n)}_{t+1} + (n+1)y^{(n+1)}_t - y^{(1)}_t$ in order to express the covariance with yields)

   (a) Intuition: interest rates seem to go down in bad times, like Fall 2008, and up in good times. That means long term bonds do well in bad times. That's an extra reason they should offer lower returns than short-term bonds, and the real yield curve should slope *down*.

(b) Inflation mucks up the picture for nominal bonds. If inflation comes in good times (the usual Keynesian assumption), then again interest rates rise in good times. But if it's stagflation, inflation in bad times, then long term bonds do worse in bad times, and this could justify an upward sloping yield curve.

3. We'll just look at the risk premium empirically.

4. Definition of "expectations hypothesis"

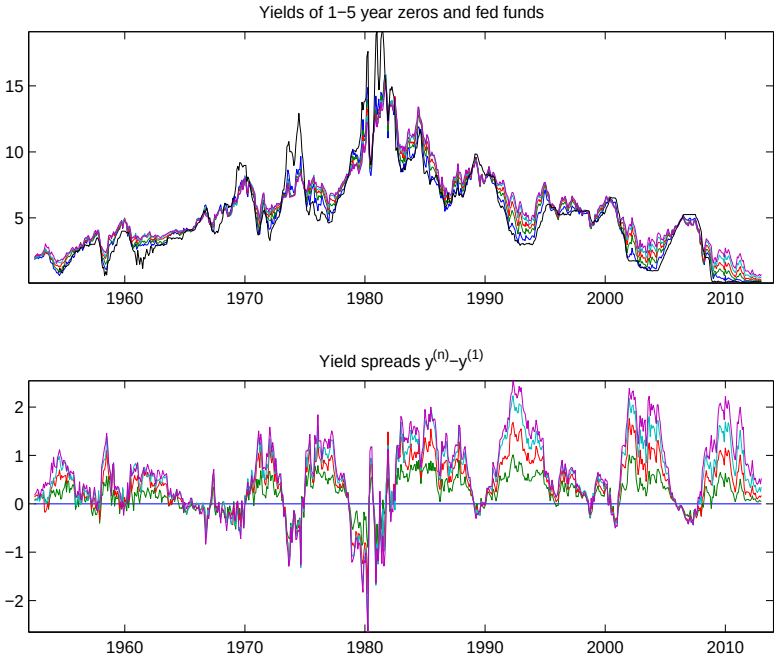
   (a) Strict definition: "Expectations hypothesis" means no risk premium.

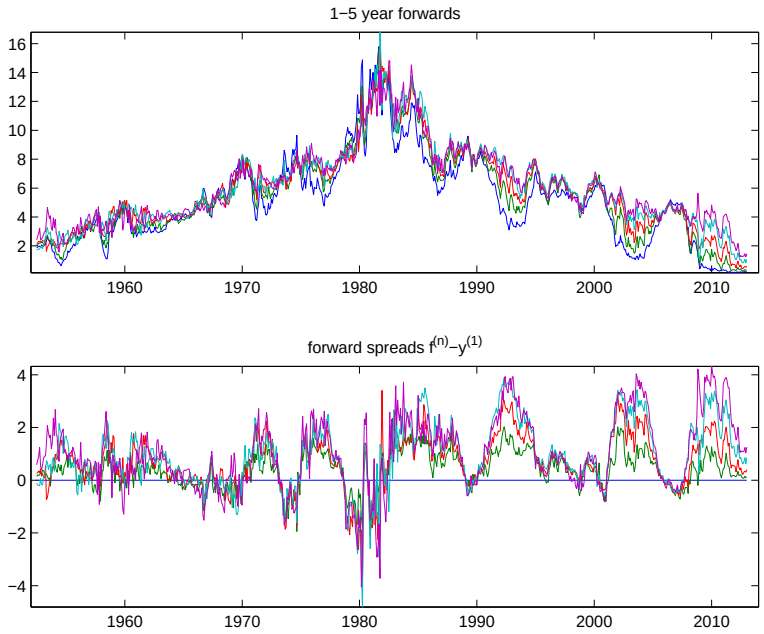
   (b) The usual definition: Expectations hypothesis means a risk premium that is small and *constant over time*. (Otherwise the theory is totally vacuous!)

   (c) "Expectations" is not quite the same as "risk neutrality" because we took logs. It's risk neutrality + $1/2\sigma^2$. Fortunately σ^2 is usually quite small.

25.5 Empirical evaluation of yield curves and risk premia

1. Facts



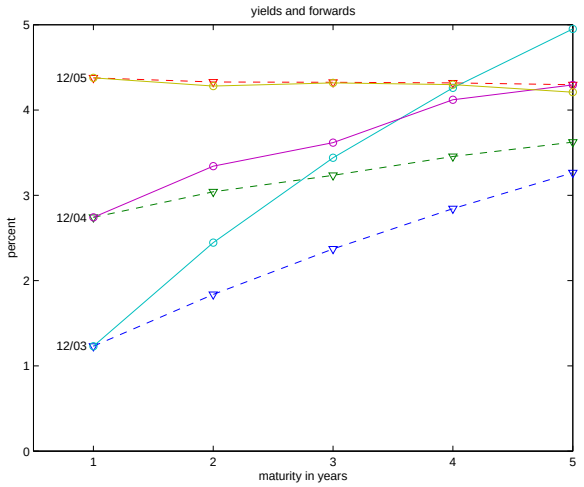


(a) The dominant movement is "level shift" – all yields up and down together.

(b) The yield curve does change shape. Sometimes rising, sometimes falling. This is the "slope" factor. There is a regular business cycle pattern: inverted at peaks, rising at troughs.

(c) When the yield/forward curve is upward sloping, interest rates subsequently rise; when the yield curve is inverted, interest rates do subsequently fall. The expectations hypothesis looks fairly reasonable.

(d) In 2003-04, the expectations hypothesis forecast big increases in rates. In 2004 and 05 the 1 year rate did rise! Actually 1 year rates rose more than forecast. The 5 year rate is dead on the forecast.



Solid = forward curve. Dash = yield curve.

(e) However, if you look closely you can see the Fama-Bliss EH failure too. The two-year rate is above the one year rate "on the way down" and through a typical regression. Rates rise eventually, but when the two year rate is higher than the one year rate, one year rates should rise *next year.*

(f) Is there a risk premium *on average?*

Interest rate data 1964:01-2012:12

Maturity n	1	2	3	4	5
$E\left[y^{(n)}\right]$	5.68	5.90	6.08	6.23	6.34
$E\left[y^{(n)}-y^{(1)}\right]$	-	0.21	0.39	0.54	0.65
$E\left[r^{(n)}-y^{(1)}\right]$	-	0.50	0.91	1.24	1.36
t statistic	-	(1.94)	(1.91)	(1.86)	(1.66)
$\sigma\left[r^{(n)}-y^{(1)}\right]$	-	1.80	3.31	4.60	5.67
"Sharpe"	-	0.28	0.28	0.27	0.24

There does seem to be a small average risk premium for long term bonds. But it seems small on average. Compare Sharpe to the 0.5 of market portfolio. This is *way* inside the Mean-Variance frontier, and $\beta \approx 0$. Expectations usually means "plus a constant (small) risk premium" and this is that constant risk premium.

Note that long term bonds look awful by one-period measures. Why do people hold them? Answer: they don't care about one year mean and variance!

2. Expectations failures–Fama Bliss. (Updated 1964-2012)

	$rx_{t+1}^{(n)} = a + b\left(f_t^{(n)} - y_t^{(1)}\right) + \varepsilon_{t+1}$			$y_{t+n-1}^{(1)} - y_t^{(1)} = a + b\left(f_t^{(n)} - y_t^{(1)}\right) + \varepsilon_{t+1}$			
n	b	$\sigma(b)$	R^2	a	b	$\sigma(b)$	R^2
2	0.83	0.27	0.11		0.17	0.27	0.01
3	1.14	0.35	0.13		0.53	0.33	0.05
4	1.38	0.43	0.15		0.84	0.26	0.14
5	1.05	0.49	0.07		0.92	0.17	0.17
	forecasting *one* year returns on n-year bonds				forecasting *one* year rates n years from now		

3. Wake up. This is *the central table.*

(a) Left hand panel: "Expected returns on all bonds should be the same" → Run a standard regression to see if anything forecasts the difference in return. Run

$$rx_{t+1}^{(n)} = r_{t+1}^{(n)} - y_t^{(1)} = a + bX_t + \varepsilon_{t+1}$$

We should see $b = 0$.

i. Over one year, expected excess returns move *one for one* with the $f - y$ spread! (The coefficient that "should be" zero.) The "fallacy of yield" is right (well, "fallacy of forward rates") at a one-year horizon.

ii. How do we reconcile this with the first table? Long bonds don't earn much more *on average*. But there are *times* when long bonds expect to earn more, and other times when they expect to earn less. $f - y$ is sometimes positive, sometimes negative.

(b) Right hand panel, row 1.

i. If $f^{(2)}$ is 1% higher than $y^{(1)}$, we should see $y_{t+1}^{(1)}$ rise 1% higher. We should see $b = 1$. Instead, we see $b = 0$! *A forward rate 1% higher than the spot rate should mean the spot rate rises. Instead, it means that the 2 year bond earns 1% more over the next year on average.*

ii. $0.83 + 0.17 = 1$ is not by chance. *Mechanically*, the two coefficients in the first row add up. If $f - y$ *does not* forecast Δy, it *must* forecast returns.[19] See the plot of bonds over time for intuition. If we move the $p_{t+1}^{(1)}$ up we increase expected returns $r_{t+1}^{(2)}$ and decrease the future yield $y_{t+1}^{(1)}$.

(c) Right hand panel, higher rows

i. At longer horizons, the $f - y$ spread *does* start to forecast n year changes in yields. (Correspondingly, it ceases to forecast n year returns – not shown.)

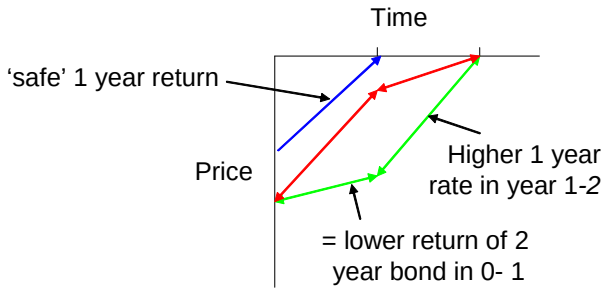
ii. Once again, expectations does seem to work "in the long run."

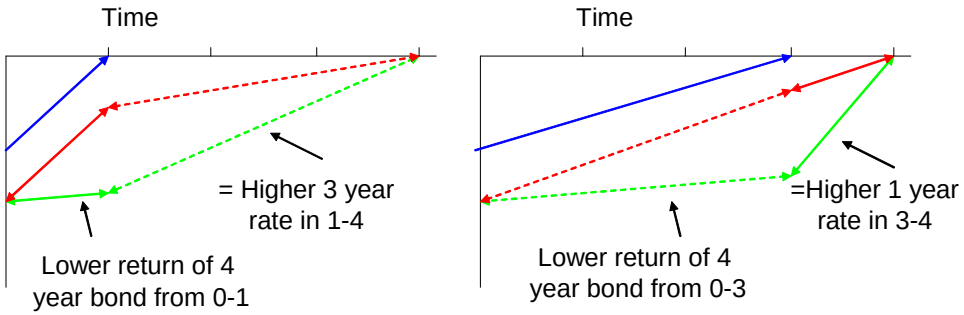
(d) Note: higher rows do not add up to 1. Why? There are "complementary" regressions that add up to 1, I just didn't show them.

[19] Why? Note that the forward-spot spread equals the change in yield plus the excess return on the two year bond.

$$\begin{aligned} f_t^{(2)} - y_t^{(1)} &= p_t^{(1)} - p_t^{(2)} + p_t^{(1)} \\ &= \left(p_{t+1}^{(1)} - p_t^{(2)} + p_t^{(1)}\right) + \left(-p_{t+1}^{(1)} + p_t^{(1)}\right) \\ &= \left(r_{t+1}^{(2)} - y_t^{(1)}\right) + \left(y_{t+1}^{(1)} - y_t^{(1)}\right) \end{aligned}$$

In this precise way, a forward rate higher than the spot rate *must* imply a high return on two year bonds or an increase in the on year rate. Now, if $F = A + B$, regress both sides on F and you get $1 =$ regression of A on F plus regression of B on F. Doesn't this remind you that dividend yields *must* forecast dividend growth or returns? It's the same idea.





4. Q: Why do we run

$$y_{t+1}^{(1)} - y_t^{(1)} = a + b\left(f_t^{(2)} - y_t^{(1)}\right) + \varepsilon_{t+1}?$$

The expectations hypothesis says $f_t^{(2)} = E_t\left[y_{t+1}^{(1)}\right]$, so why not run

$$y_{t+1}^{(1)} = a + b f_t^{(2)} + \varepsilon_{t+1}?$$

A1:That's valid but not as strong a test. An analogy: T = temperature. Just reporting $\text{Forecast}_t = \text{Temp}_t$ makes you look like a good forecaster!

$$\begin{aligned} T_{t+1} &= 0 + 1 \times T_t + \varepsilon_{t+1} \\ T_{t+1} &= 0 + 1 \times F_t + \varepsilon_{t+1} \end{aligned}$$

But this won't work for $(T_{t+1} - T_t)$. Being able to forecast *changes* is a more powerful test than being able to forecast *levels* of a slow-moving series like temp, or yield. It's much better to run

$$T_{t+1} - T_t = 0 + 1 \times (F_t - T_t) + \varepsilon_{t+1}.$$

In short, there's nothing *wrong* with doing it with levels, but differences is a more *powerful* test

A2: Does a forward *spread* forecast a rate *rise* is a different and more powerful question than, "does a high forward rate forecast a high interest rate?"

5. Interpretation.

   (a) Dividend yield "should" forecast dividend growth and not returns. Dividend yield *does* forecast excess returns and not dividend growth. The two forecasts add up mechanically.

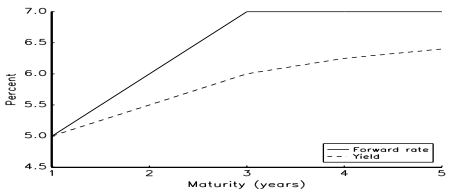
(b) Yield spread "should" forecast short rates, and not excess returns. Yield spreads *do* forecast excess returns and not yield changes. The two forecasts add up mechanically.

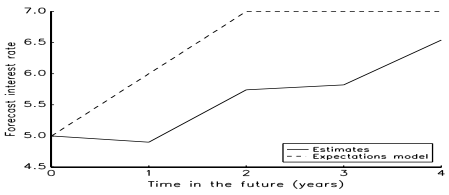
(c) The usual logic if forward rates are set so that $f_t^{(n)} = E_t y_{t+n-1}^{(1)}$ then, on average, following a time of high $f_t^{(n)}$, we should see a higher $y_{t+n-1}^{(1)}$. We can check by running this regression.

(d) "Sluggish adjustment." For stocks on D/P; bonds on F-S and exchange rates on domestic-foreign interest rate, an expected offsetting adjustment doesn't happen. (However, outcomes aren't adjusting to prices, prices reflect these expectations of outcomes, so while "sluggish adjustment" is a good way to keep track of the facts, it's not a good way to think of the mechanism.)

i. Stocks: P/D high? D should rise. It doesn't. "Buy yield" ("value").

ii. Bonds: $f - y$ high? y should rise. It doesn't (at least for a few years). "Buy yield"





If the current yield curve is as plotted in the left hand panel, the right hand panel gives the forecast of future one year interest rates. This is based on the right hand panel of the above table. The dashed line in the right hand panel gives the forecast from the expectations hypothesis, in which case forward rates today are the forecast of future spot rates.

(e) There are big risk premia and they vary over time. There are times when you expect much better returns on long term bonds, and other times when you expect much better returns on short term bonds. (An upward sloping yield curve results in rising rates, but not soon enough – you earn the yield in the meantime). High expected returns happen in bad times. Interpretation: This looks like business-cycle related variation in risk, expected returns. (γ varies over time). Who wants to hold long-term bonds in the bottom of a recession?

(f) Fama-Bliss are really about "why do forward rates move today?" As the dividend yield regression is about "why do stock prices move today?" Answer: Prices today reflect big risk premiums, not expected changes in rates. They started with the regression of yield changes on forward rates, and the returns were the complementary part. They really aren't about "how do we forecast bond returns?" which is what CP do.

25.5.1 Exchange rates

1. **Exchange rates**: If foreign interest rates are higher than US rates, the foreign currency should depreciate. It doesn't (at least for a few years). The "buy yield" advice works again. At a verbal level, this makes some sense. Domestic interest rates are low in bad times, so these are times when people are not willing to take foreign exchange risk. .

2. Numbers:

   (a) From *Asset Pricing* summary.

	DM	£	¥	SF
Mean appreciation	-1.8	3.6	-5.0	-3.0
Mean interest differential	-3.9	2.1	-3.7	-5.9
b, 1975-1989	-3.1	-2.0	-2.1	-2.6
R^2	.026	.033	.034	.033
b, 1976-1996	-0.7	-1.8	-2.4	-1.3

   i. The first row gives the average appreciation of the dollar against the indicated currency, in percent per year. The second row gives the average interest differential – foreign interest rate less domestic interest rate, measured as the forward premium – the 30 day forward rate less the spot exchange rate. On average, the expectations hypothesis seems to work; countries that have high interest differentials depreciate.

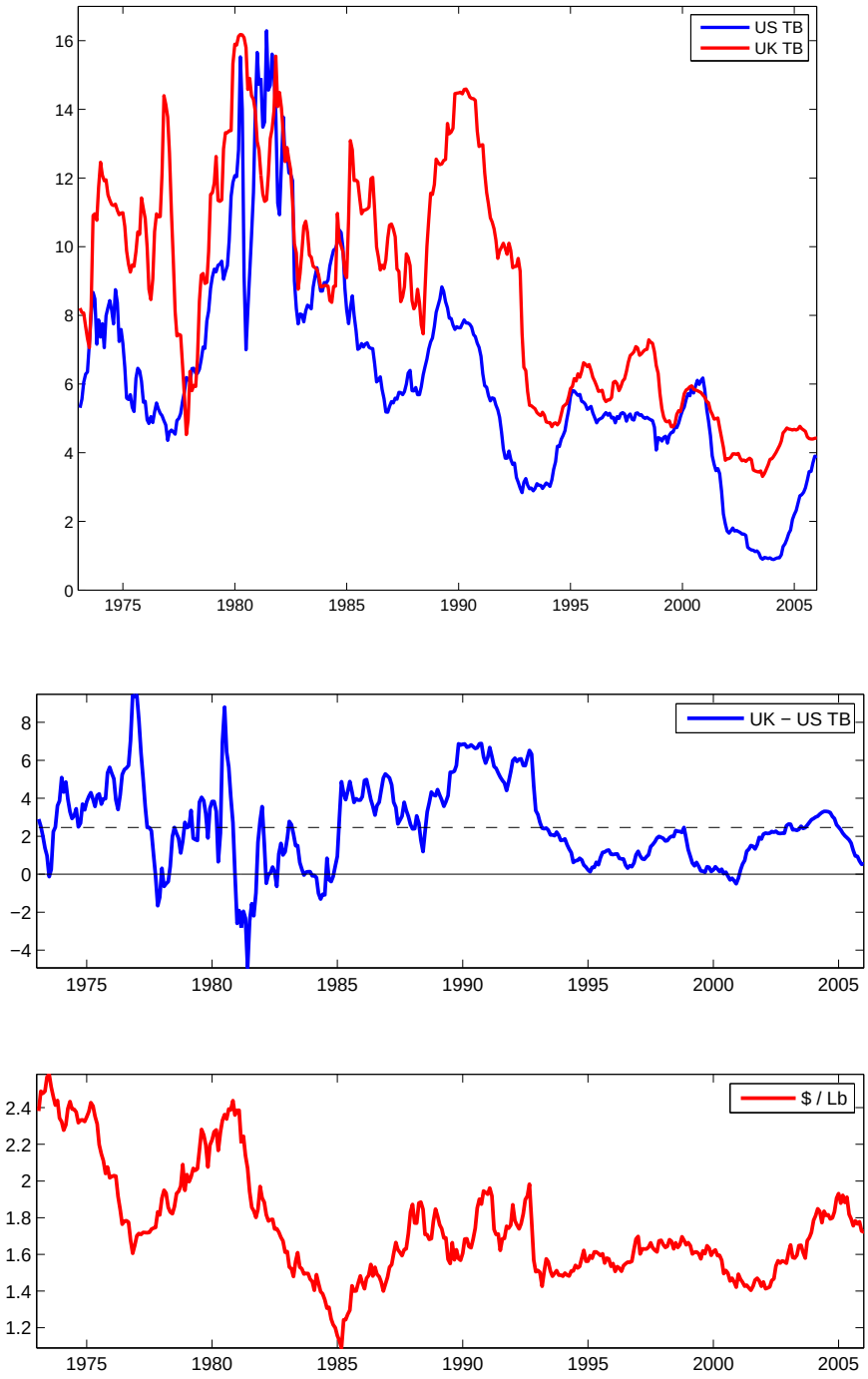
   ii. The third through fifth rows give the coefficients and R^2 in a regression of exchange rate changes on the interest differential = forward premium,

$$s_{t+1} - s_t = a + b(f_t - s_t) + \varepsilon_{t+1} = a + b(r_t^f - r_t^d) + \varepsilon_{t+1}$$

   where s = log spot exchange rate, f = forward rate, r^f = foreign interest rate, r^d = domestic interest rate. Source: Hodrick (1999) and Engel (1996). (This is one also works better in the long run, not shown). *The coefficient that should be +1 is -2 or more!*

   (b) Is the foreign rate r high? The exchange rate should fall. It doesn't – it actually goes the wrong way. This is more than sluggish adjustment.

3. Picture: as above for yields.



(a) Macro: When UK > US interest rates, we expect the Lb to be higher than the dollar, and then to fall. We do see a nice correlation between the two series – high interest rates do correspond to high exchange rates; and as interest rates mean-revert, exchange rates do fall back.

(b) When UK > US interest rates, we expect \$/Lb to fall. You can see it does in many episodes – 1975, 1981, 1990-1993, 1998 (?) 2004. Yet it takes longer than 3 months (these are 3 month rates), so there is a time of profit

(c) More recent data: See optional paper by Jurek Table 1: Regressions still are being run on a country by country basis, with similar results.

$$s_{t+1} - s_t = a + b(f_t - s_t) + \varepsilon_{t+1}$$

$b = -1$, but big standard errors *tiny* R^2. The tiny R^2 makes this look awfully thin as a trading strategy. (The other optional papers also have updated regressions. The pattern has not changed.)

25.6 Cochrane-Piazzesi update

Bottom line

1. FB run

$$rx_{t+1}^{(n)} = a_n + b_n(f_t^{(n)} - y_t^{(1)}) + \varepsilon_{t+1}^{(n)}$$

CP ask "what happens if you use all the right hand variables to improve forecasts?"

$$rx_{t+1}^{(n)} = a_n + \beta_{n,1}y_t^{(1)} + \beta_{n,2}f_t^{(2)} + \beta_{n,3}f_t^{(3)} + \beta_{n,3}f_t^{(4)} + \beta_{n,5}f_t^{(5)} + \varepsilon_{t+1}^{(n)}$$

Using a vector notation, where $f_t \equiv \left[\begin{array}{ccccc} y_t^{(1)} & f_t^{(2)} & f_t^{(3)} & f_t^{(4)} & f_t^{(5)} \end{array} \right]'$, we can write that as

$$rx_{t+1}^{(n)} = a_n + \beta_n' f_t + \varepsilon_{t+1}^{(n)}$$

2. Results:

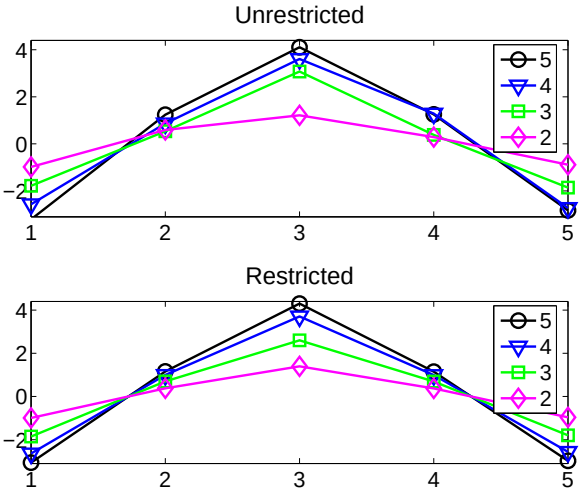
   (a) R^2 rises up to 44%, up from the Fama-Bliss / Campbell Shiller 15%

   (b) The pattern of the $\beta_{n,1}..\beta_{n,5}$ are the *same* when forecasting each maturity n.

   (c) A *single* "factor" $\gamma' f_t$ forecasts bonds of *all* maturities. We see high expected returns for all bonds in "bad times."

   (d) This factor with tent-shaped coefficients γ is correlated with slope but is not the slope of the yield curve. The improvement comes because it tells you when to bail out – when rates will rise in an upward-slope environment.

Basic regression



$$rx_{t+1}^{(n)} = a_n + \beta_{n,1} y_t^{(1)} + \beta_{n,2} f_t^{(2)} + ... + \beta_{n,5} f_t^{(5)} + \varepsilon_{t+1}^{(n)}$$

•Regressions of bond excess returns on *all* forward rates, not just matched $f - y$ as in Fama-Bliss. The plot is a plot of

•The *same* linear combination of forward rates forecasts *all* maturities' returns. Just stretch the pattern more to get longer term bonds!

• To see the point, what would Fama-Bliss coefficients look like? Answer: For each $rx^{(n)}$ you'd see a single blip up at $f^{(n)}$ and a single dip down at $y^{(1)}$. They would be different for each different n. Adapting the numbers from the "Expectations failures–Fama Bliss. (Updated 1964-2012)" table in a different format,

	$rx_{t+1}^{(n)} = a_n + b_n\left(f_t^{(n)} - y_t^{(1)}\right) + \varepsilon_{t+1}$				
n	$y_t^{(1)}$	$f_t^{(2)}$	$f_t^{(3)}$	$f_t^{(4)}$	$f_t^{(5)}$
2	-0.83	0.83			
3	-1.14		1.14		
4	-1.38			1.38	
5	-1.05				1.05

How could the coefficients be so utterly different? Answer: all forward rates are highly correlated, so huge differences in coefficients don't actually make that huge a difference in forecast. For example, if all forward rates moved exactly together then a 1.0 coefficient on the 4 year rate would give exactly the same result as a 1.0 coefficient on the 5 year rate.

A single factor for expected bond returns

We can caputre the pattern screaming at us from the data by

$$rx_{t+1}^{(n)} = b_n\left(\gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(2)} + ... + \gamma_5 f_t^{(5)}\right) + \varepsilon_{t+1}^{(n)}; \quad \frac{1}{4}\sum_{n=2}^{5} b_n = 1.$$

•*One common combination of forward rates $\gamma' f_t$ tells you where all expected returns are going at any date. Then stretch it up more for long term bonds with b_n*

•Two step estimation; first estimate γ by finding where the average (portfolio of all bonds) is going, then estimate b_n to see "if average bonds are going up 1%, how much does this maturity go up?"

$$\overline{rx}_{t+1} = \frac{1}{4}\sum_{n=2}^{5} rx_{t+1}^{(n)} = \gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(1\to 2)} + ... + \gamma_5 f_t^{(4\to 5)} + \varepsilon_{t+1} = \gamma' f_t + \varepsilon_{t+1}$$

Then

$$rx_{t+1}^{(n)} = b_n\left(\gamma' f_t\right) + \varepsilon_{t+1}^{(n)}$$

Results:

Table 1 Estimates of the single-factor model

A. Estimates of the return-forecasting factor, $\overline{rx}_{t+1} = \gamma\prime f_t + \bar{\varepsilon}_{t+1}$

	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	R^2	$\chi^2(5)$
OLS estimates	−3.24	−2.14	0.81	3.00	0.80	−2.08	0.35	105.5

B. Individual-bond regressions

	Restricted $rx_{t+1}^{(n)} = b_n\left(\gamma^\top f_t\right) + \varepsilon_{t+1}^{(n)}$		Unrestricted $rx_{t+1}^{(n)} = \beta_n' f_t + \varepsilon_{t+1}^{(n)}$
n	b_n	R^2	R^2
2	0.47	0.31	0.32
3	0.87	0.34	0.34
4	1.24	0.37	0.37
5	1.43	0.34	0.35

•γ capture tent shape.

•b_n increase steadily with maturity, stretch the tent shape out.

•The restricted model $b_n\gamma$ almost perfectly matches unrestricted coefficients. (This is the point of the graph)

•$R^2 = 0.34 - 0.37$ up from $0.15 - 0.17$. And we'll get to 0.44!

•There is a very significant rejection of $\gamma = 0$

•The R^2 are almost the same for the single-factor restriction as for the unrestricted regressions. The restriction looks good in the graph.

•See paper version of Table 1 for standard errors, joint tests including small sample, unit roots, etc. Bottom line: it's highly significant; Expectations is rejected, the improvement on FB/3 factor models is statistically significant.

Stock Return Forecasts

Does our bond return forecasting variable also forecast stock returns?

Table 3. Forecasts of excess stock returns (VWNYSE)

$\overline{rx}_{t+1} = a + \beta x_t + \varepsilon_{t+1}$

	$\gamma^\top f$	(t)	d/p	(t)	$y^{(5)} - y^{(1)}$	(t)	R^2
	1.73	(2.20)					0.07
			3.56	(1.80)	3.29	(1.48)	0.08
	1.87	(2.38)			−0.58	(−0.20)	0.07
	1.49	(2.17)	2.64	(1.39)			0.10
MA $\gamma^\top f$	2.11	(3.39)					0.12
MA $\gamma^\top f$	2.23	(3.86)	1.95	(1.02)	−1.41	(−0.63)	0.15

- The 5 year bond had $b = 1.43$. Thus, $1.73 - 2.11$ is what you expect for a perpetuity.
- $\gamma\prime f$ does better than D/P and term spread: It drives out spread; It survives with D/P

- A common term risk premium in stocks, bonds! Times when expected bond returns are high are also times when expected stock returns are high. That we see a pervasive change in expected returns across maturities and now across stocks and bonds is reassurance that it's real, not fads or measurement errors.
- This is the beginning of a larger project. D/P forecasts stocks, $f - y$ forecasts bonds, $i - i^*$ forecasts fx. What is the common element in the right hand side? Here we found that we don't need a separate forecaster $f_t^{(n)} - y_t^{(1)}$ for each return $rx_{t+1}^{(n)}$, there is in fact only "one" right hand variable that matters. Next, is there a similar "common factor" across stocks, bonds, foreign exchange, etc.?
- See optional notes at the end for much more on CP

25.6.1 FX Update

1. Jurek Table 1. (not assigned): This is the latest paper I could find to run the standard country by country regressions.

   (a) Regressions still are being run on a country by country basis, with similar results.

   $$s_{t+1} - s_t = a + b(f_t - s_t) + \varepsilon_{t+1}$$

   $b = -1$, but big standard errors *tiny* R^2. The tiny R^2 makes this look awfully thin as a trading strategy...

   (b) A "pooled regression." If you assume all countries have the same a and b you can put them in one big regression and get a better estimate.

   (c) "Pooled, FE" lets each *country* have its own intercept,

   $$s_{t+1}^i - s_t^i = a_i + b(f_t^i - s_t^i) + \varepsilon_{t+1}^i$$

   This is the same as the country by country regressions, forcing them only to have the same slope coefficient b. In this regression, we look only at how much variation *over time* in forward spreads or interest differentials corresponds to exchange rate changes, i.e. "if the forward spread for this country *is unusually high relative to this country's normal value* what happens to exchange rates?"

   (d) "XS" lets each *time* have its own intercept,

   $$s_{t+1}^i - s_t^i = a_t + b(f_t^i - s_t^i) + \varepsilon_{t+1}^i$$

   Then we're only looking at how much variation *across countries* in forward spreads or interest differentials corresponds to exchange rate changes, i.e. "if the forward spread for this country *is unusually high relative to the forward spreads of other countries this year* what happens to exchange rates?"

   (e) The coefficients are different. The XS R^2 is a lot higher, suggesting that portfolios will do better.

   (f) Note: I see no reason for splitting the sample, "it works" in one period (pre- or post-2000, "before the crash" and "after the crash") or set of countries vs. another. Yes, these regressions are very weak.

2. Better than regressions, let's look at returns of *portfolios* – If you put your money into all high interest rate currencies and benefit from diversification? Jurek Table II panel B. *Transforming the regressions to portfolios shows a big mean, a big t statistic, and a very attractive Sharpe ratio.* (But again, this stops before the crash)

3. Even better, why not sort into portfolios just like Fama French? See Lustig, Roussanov, Verdelhan Table 1. *Putting assets into portfolios and looking at the average return is the same thing as running forecasting regressions. However, it lets you see the economic importance of the regression in a way that forecasting regressions do not show. E(R), Sharpe = $E(R)/\sigma(R)$ can be very big across portfolios even when the forecast R^2 is very small.*

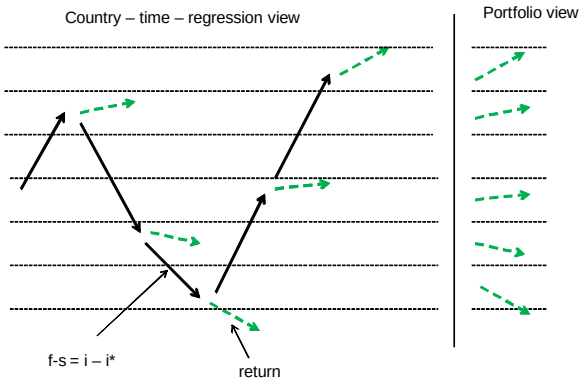


Table 1
Currency portfolios—U.S. investor

Portfolio	1	2	3	4	5	6
			Panel I: All Countries			
			Spot change: Δs^j			
Mean	−0.64	−0.92	−0.95	−2.57	−0.60	2.82
Std	8.15	7.37	7.63	7.50	8.49	9.72
			Forward Discount: $f^j - s^j$			
Mean	−2.97	−1.23	−0.09	1.00	2.67	9.01
Std	0.54	0.48	0.47	0.52	0.64	1.89
			Excess Return: rx^j (without b–a)			
Mean	−2.33	−0.31	0.86	3.57	3.27	6.20
Std	8.23	7.44	7.66	7.59	8.56	9.73
SR	−0.28	−0.04	0.11	0.47	0.38	0.64
			Net Excess Return: rx^j_{net} (with b–a)			
Mean	−1.17	−1.27	−0.39	2.26	1.74	3.38
Std	8.24	7.44	7.63	7.55	8.58	9.72
SR	−0.14	−0.17	−0.05	0.30	0.20	0.35

4. Now that they're in portfolios how about Fama and French's second idea. Do average returns line up with factor betas? Here's what they did

   (a) LRV Table 3.

Table 3
Principal components

Panel I: All Countries

Portfolio	1	2	3	4	5	6
1	0.42	0.43	0.18	−0.15	0.74	0.20
2	0.38	0.24	0.15	−0.27	−0.61	0.58
3	0.38	0.29	0.42	0.12	−0.28	−0.71
4	0.38	0.04	−0.35	0.83	−0.03	0.18
5	0.43	−0.08	−0.72	−0.44	−0.03	−0.30
6	0.45	−0.81	0.35	−0.03	0.11	0.06
% Var.	71.95	11.82	5.55	4.00	3.51	3.16

Panel II: Developed Countries

Portfolio	1	2	3	4	5
1	0.44	0.66	−0.54	−0.25	0.12
2	0.45	0.25	0.75	0.01	0.41
3	0.46	0.02	0.19	0.04	−0.86
4	0.44	−0.27	−0.29	0.78	0.20
5	0.45	−0.66	−0.14	−0.57	0.17
% Var.	78.23	10.11	4.97	3.49	3.20

This table reports the principal component coefficients of the currency portfolios presented in Table 1. In each panel, the last row reports (in %) the share of the total variance explained by each common factor. Data are monthly, from Barclays and Reuters (Datastream). The sample period is 11/1983–12/2009.

The columns here are weights by which you construct "factors." The first is "everyone against the dollar" like rmrf. The second one is "buy the high interest spread countries, short the low interest spread countries" like hml. We'll study the method for producing these numbers next week. For now

(b) See LRV Table 4.

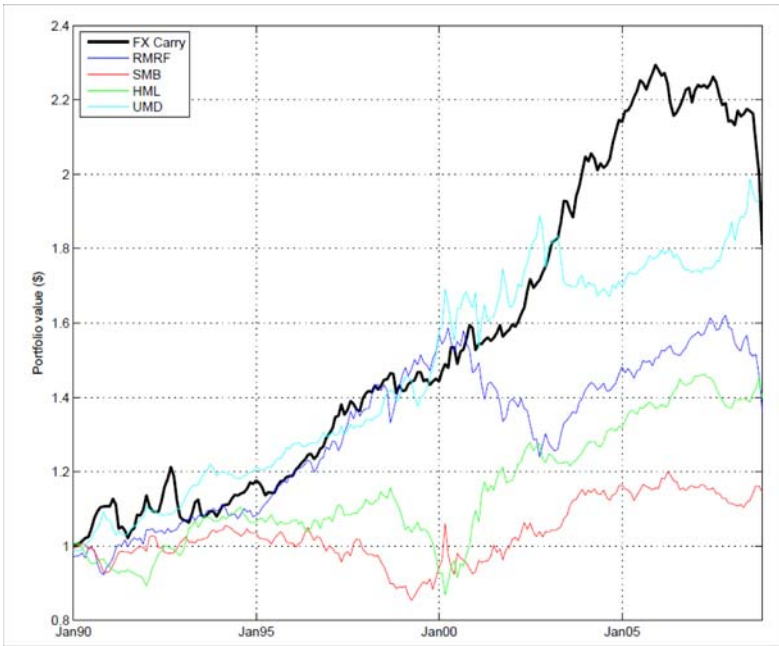
Table 4
Continued

	All Countries			
Portfolio	α_0^j	$\beta^j_{HML_{FX}}$	β^j_{RX}	R^2
1	−0.10	−0.39	1.05	91.64
	[0.50]	[0.02]	[0.03]	
2	−1.55	−0.11	0.94	77.74
	[0.73]	[0.03]	[0.04]	
3	−0.54	−0.14	0.96	76.72
	[0.74]	[0.03]	[0.04]	
4	1.51	−0.01	0.95	75.36
	[0.77]	[0.03]	[0.05]	
5	0.78	0.04	1.06	76.41
	[0.82]	[0.03]	[0.05]	
6	−0.10	0.61	1.05	93.84
	[0.50]	[0.02]	[0.03]	

Expected returns line up in betas on a high-low factor. *The high interest rate currencies all depreciate at the same time, and low interest rate currencies all appreciate at the same time.* (We'll come back to the factor stuff later) This table is a bit hard to read. The betas in the bottom half line up with average returns in Table 1. The top half gives a cross-sectional regression of average returns on betas, done three ways (the mechanics aren't that interesting).

(c) In words, *you earn returns in the carry trade by taking the risk that all the high interest rate countries will depreciate together, and all the low interest rate countries will appreciate together. You seem not to earn returns by taking the risk that all countries appreciate relative to the dollar.* But this might change...

5. Does this trade suffer Peso problems, like writing a big put? It certainly fell apart in the financial crisis, per Jurek Figure 1. Interest rates were high abroad, but the dollar gained a lot of value in the "flight to quality" However, reports are that it came back and overall is still up despite a big tail.



6. Jurek Figure 8. For a while it looked like you made money even if you buy put options to cover the crash risk. That fell apart very spectacularly in the crash.

25.7 Updates and new directions

A summary and attempt to see where we're going.

1. So far you've seen *bond-by-bond* or *country-by-country* regressions. What's going on now is an effort to "put this all together." In many ways that also amounts to "see if this is useful in investing," and if so how. Some directions

2. *More right hand variables.*

   (a) FB run

   $$rx_{t+1}^{(n)} = a + b(f_t^{(n)} - y_t^{(1)}) + \varepsilon_{t+1}^{(n)}$$

   and the FX regressions are

   $$rx_{t+1}^{(i)} = a + b(f_t^{(i)} - s_t^{(i)}) + \varepsilon_{t+1}^{(i)}$$

   (b) CP ask "what happens if you use more right hand variables to improve forecasts?"

   $$rx_{t+1}^{(n)} = a + \beta_{n,1} y_t^{(1)} + \beta_{n,2} f_t^{(2)} + \beta_{n,3} f_t^{(3)} + ...\beta_{n,5} f_t^{(5)} + ... + \varepsilon_{t+1}^{(n)}$$

   and "what if you forecast stock returns using bond variables?"

   $$R_{t+1}^{e} = a + b(DP)_t + c(\gamma' f_t) + \varepsilon_{t+1}$$

   (c) We should do this all over the place!

(d) Do country i spreads forecast country j exchange rates?

$$\Delta s_{t+1}^{j} = a + b(i_t^j - i_t^{US}) + c\left(i_t^k - i_t^{US}\right) + \varepsilon_{t+1}?$$

Maybe a CP like factor emerges across countries?

(e) Do interest spreads forecast bond returns? Does $\gamma' f$ forecast currency returns? How many common factor are there in expected returns across all asset classes?

3. *Factors* Stock average returns correspond to rmrf, hml, smb. Bond average returns correspond to level shocks. (it turns out, we didn't do this in class, but it's in "decomposing the yield curve" if you get interested) FX returns correspond to LRV's slope shock. Put all this together too!

25.8 Bond expectations / risk premia summary

1. Expectations hypothesis (in logs). Big picture: the random walk background.

   (a) Forward rate = expected future spot rate

   (b) Long term bond yield = expected future short yields

   (c) You expect to earn the same amount on bonds of any maturity over the next year

   (d) A rising term structure (yield higher with maturity, like right now) implies that interest rates will rise in the future, *not* that you make more money holding long term bonds.

2. Empirical evidence – averages

   (a) On average, the yield curve slopes slightly upwards, and returns are slightly higher on long term bonds. But poor Sharpe ratios.

3. Empirical evidence – regressions. Big picture:

$$\text{excess return}_{t \to t+1} = a + bX_t + \varepsilon_{t+1}$$

   As with stocks, are there *times* when one kind of bond is expected to do better than another?

4. Fama-Bliss

   (a) Right hand variable: For the n period bond, use the forward rate ($n-1 \to n$) less spot rate.

   (b) Fact: $b \approx 1$. At a *one* year horizon, a forward rate 1% higher than the 1 year rate means you expect to earn 1% more on long term bonds.

   (c) Bond math: $f - y$ mechanically means either return or rise in one year yield. At a one year horizon, a forward rate 1% higher than the 1 year rate *does not* signal a rise in interest rates!

   (d) At a 5 year horizon, a forward rate 1% higher than the 1 year rate *does* signal a 1% rise in interest rates. "Sluggish adjustment" Interest rates are Waiting For Godot.

5. Empirical evidence – Cochrane-Piazzesi

(a) Central innovation – forecast returns with $X = all$ forward rates, not just the matched forward spot spread.

(b) A *common* "factor" (linear combination of yields and forward rates) forecasts bond returns *of all maturities.* Long bond expected returns move more than short maturity expected returns. (This is the most important finding, and holds even if the R^2 improvement and tent are not important.)

(c) R^2 rises to 0.35-0.44 from FB 0.15.

(d) Expected bond returns are high in bad times – until the inevitable rise in interest rates happens. Then you get killed in long term bonds. Watch out!

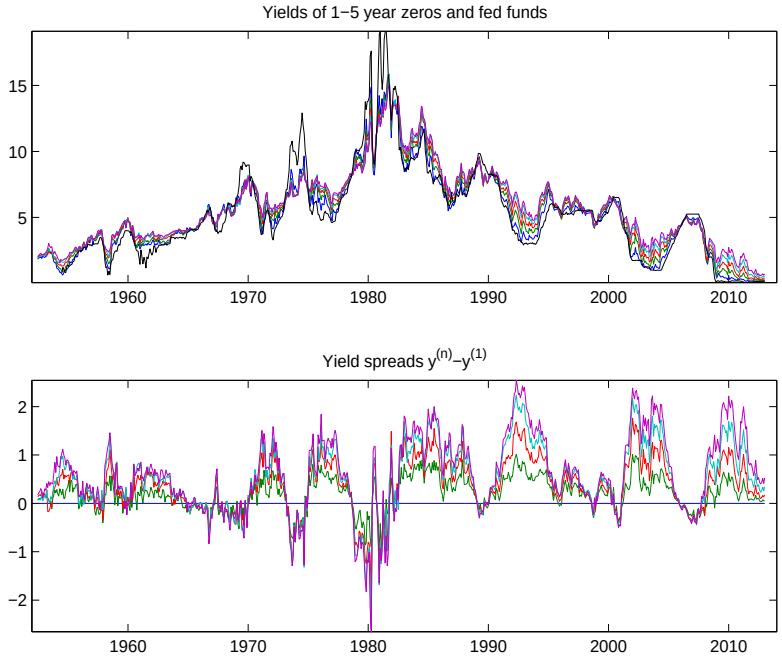
(e) "Signal" is curvature in the forward rate curve. This is not the usual "curvature" factor in the yield curve – more curved at the long end.

(f) Big question: What the heck is $\gamma' f$? What does the tent shape mean?

26 Term Structure II. Interest Rates Factor Models notes

- The "Investments notes" contain a matrix review. If you forgot how to multiply matrices you should go read that now.

26.0.1 Motivation and idea



- Look at the graph. Look at the movie.[20] Most movements in 1-5 year bonds are a) "level

[20] lecture9.m

shifts" (top graph) b) changes in slope of the term structure (bottom graph). How can we capture this behavior?

- How about
$$y_t^{(n)} = a_n + b_n \text{level}_t + c_n \text{slope}_t + (\text{other stuff}).$$
If
$$\begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \\ y_t^{(4)} \\ y_t^{(5)} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \text{level}_t + \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} \text{slope}_t + (\text{other stuff})$$
Then a 1% change in "level" moves all yields up 1%. A change in "slope" moves the short rates down and the long rates up; it changes the slope of the yield curve. *Two* "factors" describe the movements of five yields.

- Analogy: This is a lot like the FF3F model.
$$R_{t+1}^{ei} = \alpha_i + b_i rmrf_{t+1} + h_i hml_{t+1} + s_i smb_{t+1} + \varepsilon_{t+1}^i$$
We express each yield (return) as a sum of coefficients (betas) specific to that security times "factors" (level, slope; rmrf, hml, smb) common to all securities. In turn the factors are specific combinations (portfolios) of the same underlying securities. rmrf is the sum of all returns, hml is long value stocks short growth stocks etc.; level is an average of all yields, slope is long high maturity yields and short low maturity yields etc.

- We're describing *variance,* the movement of *ex-post* yields, not yet *expected returns*, etc. Now we do care about time series regressions, R^2 significance of coefficients, etc. And for now, we don't care about intercepts. We'll tie together means and variances in a bit.

26.0.2 A simple approach to a factor model

- Why don't we try exactly the FF approach? Let's define a level factor as the average of all yields (like rmrf) and the slope factor as a high-low portfolio just like hml
$$\begin{aligned} \text{level}_t &\equiv \frac{1}{5}\left[y_t^{(1)} + y_t^{(2)} + y_t^{(3)} + y_t^{(4)} + y_t^{(5)}\right] \\ \text{slope}_t &= \frac{1}{2}\left[y_t^{(4)} + y_t^{(5)}\right] - \frac{1}{2}\left[y_t^{(1)} + y_t^{(2)}\right]. \end{aligned}$$
Clearly, when all yields go up, level will rise. When the yield curve slopes up slope will be a big number, and when it slopes down, it will be a negative number.

- Now, we can just run regressions to find the "betas."
$$y_t^{(n)} = a_n + b_n \times \text{level}_t + s_n \times \text{slope}_t + \varepsilon_t^{(n)}$$

I did it,

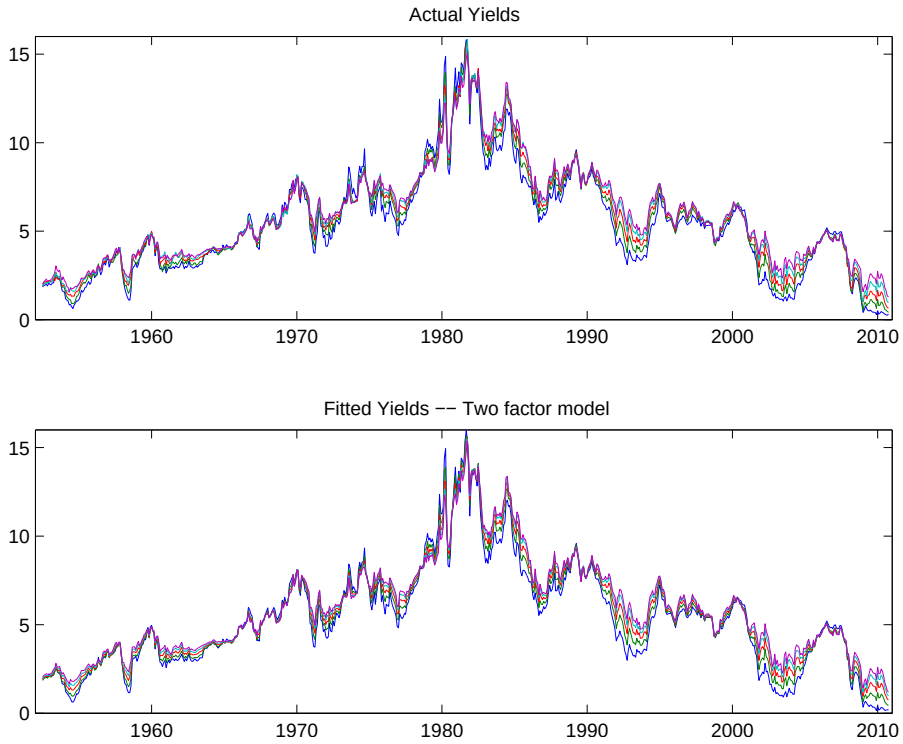
	a_n	b_n	s_n	R^2
$y^{(1)}$	0.09	0.9998	-0.77	0.97
$y^{(2)}$	-0.08	1.0003	-0.28	0.97
$y^{(3)}$	-0.04	0.9999	0.10	0.97
$y^{(4)}$	-0.02	1.0008	0.38	0.98
$y^{(5)}$	0.04	0.9993	0.56	0.98

This ought to look pretty familiar and just about what you expected. (The units are %, so the intercepts are truly tiny, i.e. 9 bp for $y^{(1)}$. These are yields not returns, so the intercepts are not "alphas.")

- The R^2 are *very* high here. The factor model is almost a perfect fit. There are really only two portfolios that change, and then everything else is just a linear combination of these. Here is a plot of actual yields and

$$\text{fit } y_t^{(n)} = b_n \times \text{level}_t + s_n \times \text{slope}_t$$

The lines are indistinguishable, meaning that the two-factor model is an excellent fit.



- The good fit means that we are very close to an "arbitrage pricing theory." Once where you know where level and slope are (or any two yields) on any date, level_t and slope_t, you know exactly where all the other yields must be! We will end up with "no-arbitrage term structure models" in which there are exactly n factors, and the rest are priced by arbitrage.

- Principal components analysis allows you to construct factors like this very easily, without having to "guess" the factors.

- Notice the two steps in constructing a factor model. Any factor model.

  1. How do you form the factors (level, slope) from yields or other observable variables? $\text{level}_t = ...$, $\text{slope}_t = ...$
  2. What are the *loadings* or *betas*; if the factor moves, how much does each bond move? $y_t^{(n)} = ...\text{level}_t + ...\text{slope}_t ...$

26.1 Factor analysis

- That's nice, but ad-hoc. Can you do better – get a closer fit – with a different construction of level and slope? What's the best way to form another factor (curvature) to get a better fit?

- Objective:

$$y_t^{(n)} = a_n + b_n\text{level}_t + s_n\text{slope}_t + c_n\text{curve}_t + (\text{other stuff, as small as possible})$$

  How do you construct factors that are uncorrelated with each other, so as to maximize R^2 / minimize the variance of the error.

- The answer is: the eigenvalue decomposition/principal components. (OLS regression picks the b, s, c to minimize the variance of the error *given* the factors. The question here is, how do you pick the *factors* so that after you have run the regressions you get the smallest possible error?)

- Here's the procedure. First a very short version for people comfortable with matrices

  1. *Form the covariance matrix of yields.* Using the notation

$$y_t = \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \\ \vdots \end{bmatrix}$$

$$\Sigma = cov(y_t, y_t')$$

     This is, for the Fama Bliss data, a 5 x 5 matrix.

  2. *Take its eigenvalue decomposition.* (The matlab function eig does this, and you don't have to know anything about how it works. If you're curious see the appendix to the notes.) This decomposition forms three matrices from Σ,

$$\Sigma = Q\Lambda Q',$$

     in which Λ is diagonal, and[21] $Q'Q = QQ' = I$.

  3. *That's it.* In matlab, there are two steps. With `yields` = a $T \times N$ matrix as usual, form `Sigma = cov(100*yields)` and then form `[Q,L] = eig(Sigma).`
     That's all you have to *do.* What does this *mean?*

[21] $QQ' = Q'Q = I$ comes from the fact that Σ is a symmetric matrix. In general, the eigenvalue decomposition is $\Sigma = Q\Lambda Q^{-1}$. Covariance matrices are symmetric.

4. *The columns of Q represent the weights by which you form factors from yields,*

$$x_t = Q' y_t$$

5. *The columns of Q also represent the loadings or betas,* which answer "how much does y move when a factor moves,"

$$y_t = Q x_t$$

Proof: x_t is defined as $x_t = Q' y_t$, so $y_t = (QQ')y_t = Q x_t$.

6. *The factors x_t are uncorrelated with each other, and the diagonals of Λ are their variances,*

$$cov(x_t, x_t') = \Lambda$$

Proof: $cov(x_t, x_t') = cov(Q' y_t y_t' Q) = Q' cov(y_t y_t') Q = Q' Q \Sigma Q' Q = \Lambda$

7. We form approximate (but often a very good approximation) factor models by leaving out factors with small variances λ, or just considering them as part of an error term.

8. Once you've formed the factors x_t, A regression of y_t on x_t also gives the loadings Q. Since the factors are uncorrelated with each other, if you leave out some factors, regressions with the remaining factors also recover the remaining columns of Q.

- Now, a more detailed version of the same thing. This just amounts to writing out the matrices and looking at them.

1. `Sigma = cov(yields);` Form the covariance matrix of yields

$$\begin{aligned} y_t &= \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \end{bmatrix} \\ \Sigma &\equiv cov(y_t, y_t') \end{aligned}$$

I'll do three yields in the examples, so Σ is a 3×3 matrix. The FB data has 5 yields so this is a 5×5 matrix. In general you'd do this with many, many maturities.

2. `[Q,L] = eig(Sigma);` Use a computer to find the *eigenvalue decomposition* of Σ,

$$\Sigma = Q \Lambda Q'.$$

Λ is a *diagonal matrix*

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

and write the Q matrix in terms of its columns as

$$Q = \begin{bmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{bmatrix}$$

You do *not* need to know what an eigenvalue is, or how to compute this decomposition by hand (essentially impossible). Just trust and verify that the computer does it right.

3. The *eigenvalue decomposition* looks like this:

$$\begin{aligned}\Sigma &= Q\Lambda Q' \\ &= \begin{bmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} - & q_1' & - \\ - & q_2' & - \\ - & q_3' & - \end{bmatrix}\end{aligned}$$

The property of the Q matrix

$$Q'Q = QQ' = I$$

means

$$\begin{bmatrix} - & q_1' & - \\ - & q_2' & - \\ - & q_3' & - \end{bmatrix} \begin{bmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

4. Now, *the columns of Q tell you how to construct the factors x_t from the yields y_t.*You can form the factors by

$$x_t = Q'y_t.$$

This means

$$\begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \end{bmatrix} = \begin{bmatrix} - & q_1' & - \\ - & q_2' & - \\ - & q_3' & - \end{bmatrix} \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \end{bmatrix}$$

for instance,

$$x_t^{(1)} = \begin{bmatrix} - & q_1' & - \end{bmatrix} \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \end{bmatrix}.$$

For example: if

$$q_1' = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$$

then this operation constructs the "level" factor.

$$x_t^{(1)} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \end{bmatrix} = \frac{1}{3}\left(y_t^{(1)} + y_t^{(2)} + y_t^{(3)}\right).$$

5. *The factors are uncorrelated with each other, and the Λ diagonals tell you the variance of each factor*

$$cov(x_t, x_t') = \Lambda$$

this means

$$cov\left(\begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \end{bmatrix} \begin{bmatrix} x_t^{(1)} & x_t^{(2)} & x_t^{(3)} \end{bmatrix}\right) = \begin{bmatrix} \sigma^2(x^{(1)}) & cov(x^{(1)}, x^{(2)}) & cov(x^{(1)}, x^{(3)}) \\ \cdot & \sigma^2(x^{(2)}) & cov(x^{(2)}, x^{(3)}) \\ \cdot & \cdot & \sigma^2(x^{(3)}) \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

6. *The columns of Q are also the* **loadings**, *they tell you how much each y moves when a factor x moves.*

$$y_t = Qx_t.$$

This means

$$\begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \end{bmatrix} = \begin{bmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \end{bmatrix} = \begin{bmatrix} | \\ q_1 \\ | \end{bmatrix} x_t^{(1)} + \begin{bmatrix} | \\ q_2 \\ | \end{bmatrix} x_t^{(2)} + \begin{bmatrix} | \\ q_3 \\ | \end{bmatrix} x_t^{(2)}$$

or

$$y_t^{(n)} = q_1^{(n)} x_t^{(1)} + q_2^{(n)} x_t^{(2)} + q_3^{(n)} x_t^{(3)}$$

$y = Qx$ is a factor model–the columns of Q are the "loadings" or the "betas"; the b_n and c_n of the first equation.

7. To form a factor model, *leave out the factors with really small eigenvalues*, either ignoring them or treating them as small errors For example,

$$\begin{aligned} y_t^{(n)} &= q_1^{(n)} x_t^{(1)} + q_2^{(n)} x_t^{(2)} + \varepsilon_t \\ y_t^{(n)} &\doteq q_1^{(n)} x_t^{(1)} + q_2^{(n)} x_t^{(2)} \end{aligned}$$

Since the factors are uncorrelated with each other, ε_t is uncorrelated with $x_t^{(1)}$ and $x_t^{(2)}$ so q are still regression coefficients.

- *A simple example*: Suppose the covariance matrix is

$$cov(y_t, y_t') = \begin{bmatrix} \sigma^2\left(y_t^{(1)}\right) & cov\left(y_t^{(1)}, y_t^{(2)}\right) \\ cov\left(y_t^{(1)}, y_t^{(2)}\right) & \sigma^2\left(y_t^{(2)}\right) \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Now, find its eigenvalue decomposition:

```
>> Sigma = [2 1 ; 1 2];
>> [Q,L] = eig(Sigma);
>> disp(Q);
-0.71 0.71
 0.71 0.71
>> disp(L);
 1.00  0
  0   3.00
```

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} -0.71 & 0.71 \\ 0.71 & 0.71 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} -0.71 & 0.71 \\ 0.71 & 0.71 \end{bmatrix}$$

(note $0.71 = 1/\sqrt{2}$) Now check that this worked.

```
>> disp(Q*L*Q');
 2.00 1.00
 1.00 2.00
>>disp(Q'*Q);
 1.00 0
  0  1.00
>> disp(Q*Q');
 1.00 0
  0  1.00
```

- *Example continued:* We interpret the result as a factor model.

  1. The factors are two random variables $x^{(1)}$ and $x^{(2)}$ with

$$cov(x_t, x_t') = \begin{bmatrix} \sigma^2\left(x_t^{(1)}\right) & cov\left(x_t^{(1)}, x_t^{(2)}\right) \\ cov\left(x_t^{(1)}, x_t^2\right) & \sigma^2\left(x_t^{(2)}\right) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

     $x^{(2)}$ is more volatile, and the two x's are uncorrelated with each other.

  2. We think of y as generated from the factors by

$$\begin{aligned} y_t &= Qx_t \\ \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \end{bmatrix} &= \begin{bmatrix} -0.71 & 0.71 \\ 0.71 & 0.71 \end{bmatrix} \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix} \\ \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \end{bmatrix} &= \begin{bmatrix} -0.71 \\ 0.71 \end{bmatrix} x_t^{(1)} + \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix} x_t^{(2)} \end{aligned}$$

  3. *What does this mean?* $x^{(2)}$ is the more volatile factor, with $\sigma^2\left[x^{(2)}\right] = 3$. When it moves, both y move together ("level"). $x^{(1)}$ is a less volatile factor with $\sigma^2\left[x^{(1)}\right] = 1$. When it moves, $y^{(1)}$ goes down and $y^{(2)}$ goes up ("slope"). The y were correlated with each other. We capture this correlation here with a larger "common component" $x^{(2)}$ and a smaller component that sends them in different directions.

  4. *Language* The x are *factors* and the columns q are the *factor loadings* since they tell you how much each y moves when an x moves. The q are also "*betas*", regression coefficients of y on the factors x. The first, smaller factor is a "slope factor" and the second larger factor is a "level factor." This statement just describes the loadings.

  5. *Constructing a time-series of the factors.* The x are constructed from the y by .

$$\begin{aligned} y_t &= Qx_t \\ x_t &= Q^{-1}y_t = Q'y_t \\ \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix} &= \begin{bmatrix} -- & q_1' & -- \\ -- & q_2' & -- \end{bmatrix} \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \end{bmatrix} \\ \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix} &= \begin{bmatrix} -0.71 & 0.71 \\ 0.71 & 0.71 \end{bmatrix} \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \end{bmatrix} = \begin{matrix} -0.71y_t^{(1)} + 0.71y_t^{(2)} \\ 0.71y_t^{(1)} + 0.71y_t^{(2)} \end{matrix} \end{aligned}$$

(Q is usually not symmetric!) Note that *the same columns of Q that describe how y responds to each of the x, also describe how to construct factors x from the data y* We recover the "level" factor from an average of the two yields. We recover the "slope" factor from long-short yield spread!

26.1.1 Unit variance factors – Optional

- *A refinement: unit variance factors.* Now we have

$$\begin{aligned} \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \end{bmatrix} &= \begin{bmatrix} -0.71 \\ 0.71 \end{bmatrix} x_t^{(1)} + \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix} x_t^{(2)}; \\ \sigma^2\left(x_t^{(1)}\right) &= 1;\ \sigma^2\left(x_t^{(2)}\right) = 3 \end{aligned}$$

  As you can see, the second factor accounts for more of the movement of y. Rather than express this fact by remembering that the second x is more volatile, let's express it as bigger loadings on the second factor. *Let's just rescale the factors by their standard deviations to express the model in terms of factors with unit variance.*

- Define new factors $z^{(i)}$ where $x^{(i)} = \sigma\left[x^{(i)}\right] z^{(i)}$. The z will have the same (one) variance.

$$\begin{aligned} \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \end{bmatrix} &= \begin{bmatrix} -0.71 \\ 0.71 \end{bmatrix} \left(\sigma\left[x_t^{(1)}\right] z_t^{(1)}\right) + \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix} \left(\sigma\left[x_t^{(2)}\right] z_t^{(2)}\right); \\ \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \end{bmatrix} &= \left(\begin{bmatrix} -0.71 \\ 0.71 \end{bmatrix} \sigma\left[x_t^{(1)}\right]\right) z_t^{(1)} + \left(\begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix} \sigma\left[x_t^{(2)}\right]\right) z_t^{(2)}; \\ \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \end{bmatrix} &= \begin{bmatrix} -0.71 \\ 0.71 \end{bmatrix} z_t^{(1)} + \begin{bmatrix} 1.22 \\ 1.22 \end{bmatrix} z_t^{(2)}; \\ \sigma\left(z_t^{(1)}\right) &= \sigma\left(z_t^{(2)}\right) = 1;\ cov\left(z_t^{(1)}, z_t^{(2)}\right) = 0 \end{aligned}$$

  (I used $\sigma\left[x^{(1)}\right] = 1,\ \sigma\left[x^{(2)}\right] = \sqrt{3}$ and $\sqrt{3}/\sqrt{2} = 1.22$)

- *In this way you can see more easily that the second factor is "bigger."* A typical unit movement in $z^{(2)}$ moves both $y^{(1)}$ and $y^{(2)}$ up by 1.22. A typical unit movement in $z^{(1)}$ moves $y^{(1)}$ up by 0.71 and $y^{(2)}$down by 0.71.

- The matrix version of this calculation:

$$\begin{aligned} y &= Qx;\ cov(xx') = \Lambda \\ x &= \Lambda^{\frac{1}{2}} z \\ y &= \left(Q\Lambda^{\frac{1}{2}}\right) z;\ cov(zz') = I \end{aligned}$$

  Thus, loadings are the columns of $Q\Lambda^{\frac{1}{2}}$ rather than just Q. Powers of a diagonal matrix are easy,

$$\Lambda^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix}$$

and multiplying a matrix by a diagonal is easy,

$$Q\Lambda^{\frac{1}{2}} = \begin{bmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{bmatrix} = \begin{bmatrix} | & | & | \\ q_1\sqrt{\lambda_1} & q_2\sqrt{\lambda_2} & q_3\sqrt{\lambda_3} \\ | & | & | \end{bmatrix}.$$

We can recover the z with

$$z = \Lambda^{-\frac{1}{2}} Q' y.$$

- Another view, "normalization." If you have a factor model

$$y_t = q_1 x_t^{(1)} + q_2 x_t^{(2)}$$

there is always a bit of arbitrariness. You can always make the factor bigger if you make the loading smaller. This is the same as

$$y_t = (2 \times q_1)\left(\frac{1}{2} \times x_t^{(1)}\right) + q_2 x_t^{(2)}$$

$$y_t = (a \times q_1)\left(\frac{1}{a} \times x_t^{(1)}\right) + q_2 x_t^{(2)}$$

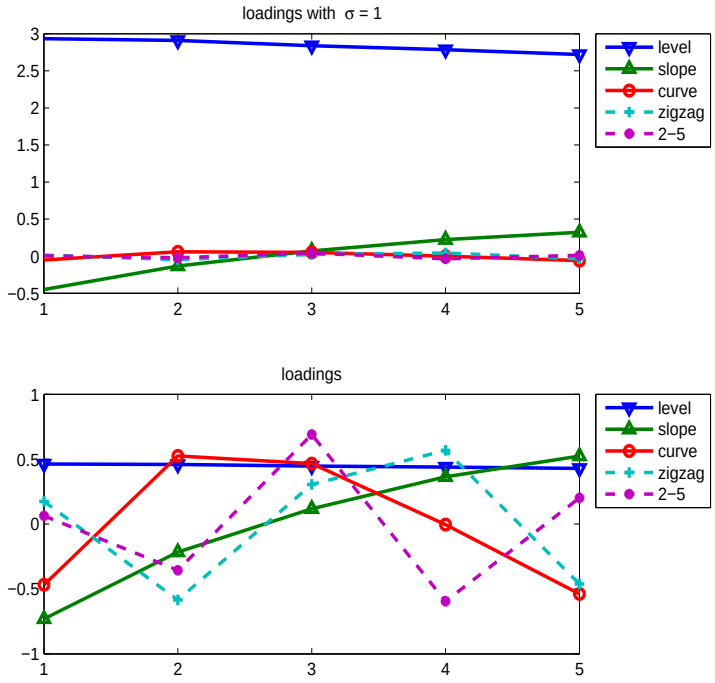
So what's the right scale?

1. The original eigenvalue decomposition picks the "size" from $q_1' q_1 = 1$ meaning that the sum of squared loadings is one, $\sum_{n=1}^{N} \left(q_1^{(n)}\right)^2 = 1$.
2. That's nice, but not particularly intuitive. This new idea chooses a scaling factor so that the variance of the factors is one. By choosing $a = \sigma(x_t^{(1)})$ we have $\sigma\left(\frac{1}{a} \times x_t^{(1)}\right) = 1$.
3. You're free to choose other scaling factors if they are more convenient.

26.1.2 A real example: yields

- A real example. Yields! This also shows you that all this talk is only about 2-3 lines of matlab

```
Sigma = cov(100*yields);
[Q,L] = eig(Sigma);
loads = Q*L^0.5;
disp('standard deviation of factors');
disp(diag(L)'.^0.5);
(my names)2-5           zigzag        curve         slope         level
          0.06          0.07          0.10          0.58          5.80
disp(Q)
          0.06          0.15         -0.47         -0.74          0.46
         -0.35         -0.55          0.56         -0.21          0.46
          0.70          0.32          0.44          0.12          0.45
         -0.59          0.57         -0.03          0.36          0.44
          0.19         -0.49         -0.52          0.51          0.43
plot(loads)
plot(Q)
```



1. Graph interpretation: *If factor x moves, how much do yields of maturity 1,2,3,4,5 move?* Almost all variation in yields is due to "level" shifts (5.80% σ). Slope and curve are almost all the rest.

2. Graph interpretation: *What combination of $y^{(1)}...y^{(n)}$ produces each factor?* Sensibly, "level" is an average of all yields, slope is long - short, etc.

$$\text{Interpretation 1: } \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \end{bmatrix} = \begin{bmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \end{bmatrix}$$

$$\text{Interpretation 2: } \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \end{bmatrix} = \begin{bmatrix} - & q_1 & - \\ - & q_2 & - \\ - & q_3 & - \end{bmatrix} \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \end{bmatrix}$$

   The names "level" "slope" and "curvature" come entirely as verbal descriptions of the shapes of these graphs.

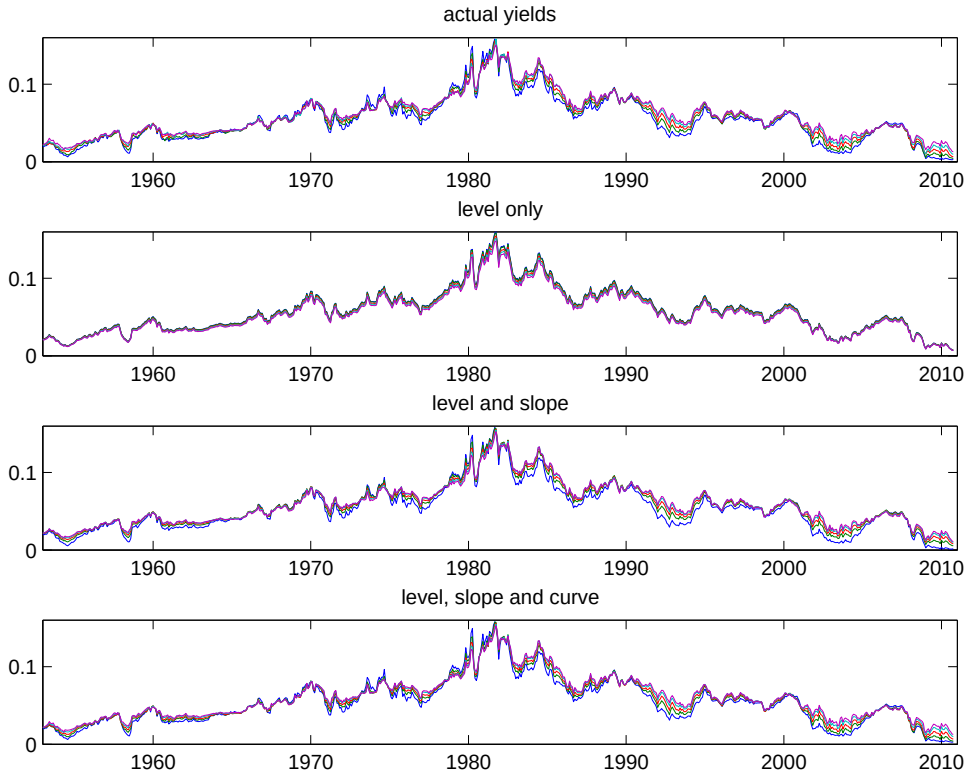
3. The standard deviations of the factors are 5.8%, 0.58% and then less than 10 basis points. The factors past the first two are really small!

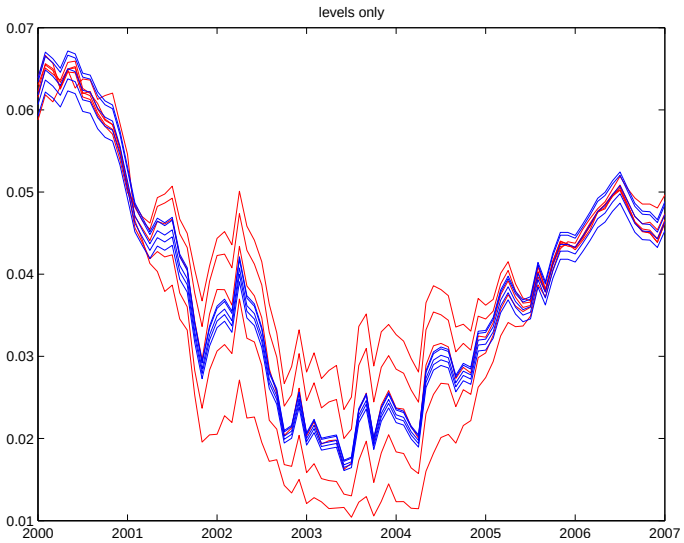
4. The top graph $Q\Lambda^{\frac{1}{2}}$ tells us "what does the typical (1 σ) movement of each factor do to the yield curve?" The bottom graph tells us "What does a unit movement of each factor (very unlikely for the small factors) do to the yield curve?" The bottom graph (Q) is better for eyeballing the *shape* of the factor loadings without worrying about which is more important than the other. Use it only in conjunction with the λ. The top graph reminds us that the smaller factors are tiny.

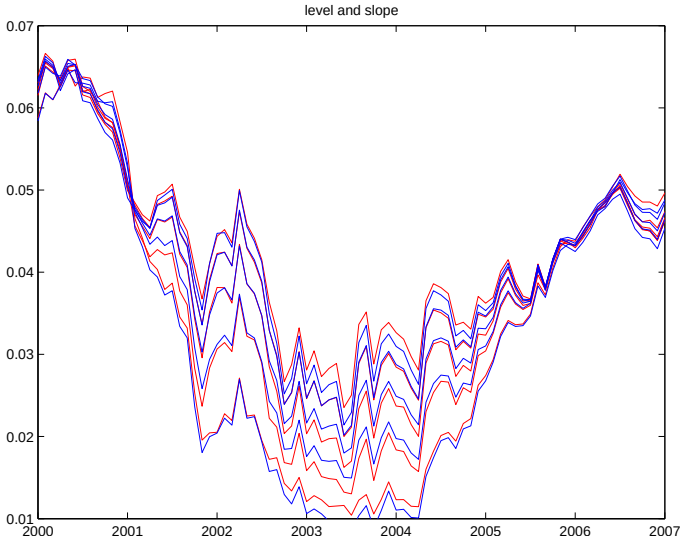
26.1.3 Dropping factors

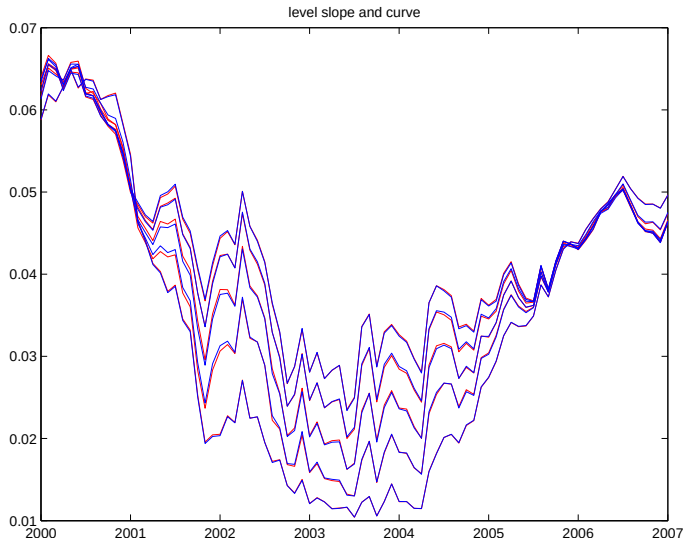
- *Dropping some factors.* As you can see, some of the factors are very small. You capture almost all movements in yields by keeping just the first few factors.

$$\begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \\ y_t^{(4)} \\ y_t^{(5)} \end{bmatrix} \approx q_1 \times \text{level}_t + q_2 \times \text{slope}_t + q_3 \times \text{curve}_t$$




levels only
0.07
0.06
0.05
0.04
0.03
0.02
0.01
2000
2001
2002
2003
2004
2005
2006
2007


level and slope
0.07
0.06
0.05
0.04
0.03
0.02
0.01
2000
2001
2002
2003
2004
2005
2006
2007



Movements in yields can be captured very well by movements in the first two - three factors alone.

- Dropping factors II. Note that *the factors are uncorrelated with each other.* $cov(xx') = \Lambda$. Thus, the left out factors are uncorrelated with the factors you keep in.

$$y_t^{(n)} \approx q_1^{(n)} \times \text{level}_t + q_2^{(n)} \times \text{slope}_t + q_3^{(n)} \times \text{curve}_t + (\text{left out}_t)$$

  Therefore, this is a regression equation! *This is a way of finding a regression model like FF3F when you don't know what to use on the right hand side.*

- Notice the analogy to FF3F: three factors (market, hml, smb) account for almost all return variation (R^2 above 90%). The factors are constructed as weighted combinations of the same securities.

- R^2 and variance. Look, for example at $y^{(1)}$. Now, the factors are uncorrelated so

$$\begin{aligned} \sigma^2(y^{(1)}) &= q_1^{(1)2}\sigma^2\left(x^{(1)}\right) + q_2^{(1)2}\sigma^2(x^{(2)}) + q_3^{(1)}\sigma^2(x^{(3)}) + .. \\ &= q_1^{(1)2}\lambda_1 + q_2^{(1)2}\lambda_2 + q_3^{(1)}\lambda_3 + ... \end{aligned}$$

  ($q_2^{(1)}$ is the first or top element of the second column of Q and so forth.) The fact that the λ get small after 3 or so means that the error term from leaving out 4 and 5 will be very small – high R^2

- We often say that $\lambda_i / \sum \lambda_i$ gives the "percent of the variance explained by the ith factor."

```
Ld= diag(L);
disp(Ld');
disp(Ld'/sum(Ld));
fractions of variance
         0.00          0.01          0.01          0.33         33.62
         0.00          0.00          0.00          0.01          0.99
```

26.1.4 Some uses of a factor model

- We're talking about bonds and yields here, but you can just as easily do the same thing for stocks and for returns.

$$\begin{aligned} \Sigma &= cov(R, R') \\ \Sigma &= Q\Lambda Q' \\ x &= Q'R \end{aligned}$$

now we have the factors

$$R_t = Qx_t \approx Q(:, 1:3)x_t + \text{small error}$$

This is the heart of Barra, commercial risk management models.

- Often

$$E(R_t) = Q(:, 1:3)E\left(x_t\right) + E(\text{small error}) = = Q(:, 1:3)E\left(x_t\right) + small \text{ alpha}$$

and the model ends up describing means as well.

- *Uses* 1: multifactor immunization.

  1. Recall the basic duration hedge: You choose 1 security so that the portfolio value doesn't change if level changes. The problem with this is, if the slope changes, you are exposed to risk.
  2. Multifactor immunization: You can instead choose 3 securities so that the portfolio value doesn't change if level, slope and curvature change. Since there are so few additional changes, the portfolio is very well immunized.

- *Uses* 2: hedging/manager evaluation. Like FF3F for bonds. We now have a 3 factor model for yields, use just like FF3F 3 factor model. If your portfolio has no q on anything, then it will have very little variance, as no b, h, s gives no variance. This lets you separate "directional bet" on factors from "mispricing" (α) of bonds.

- As an analogy, back when we did

$$R^{ei}_{t+1} = \alpha_i + b_i rmrf_{t+1} + h_i hml_{t+1} + s_i smb_{t+1} + \varepsilon^i_{t+1}$$

we were focused on means and alphas, but there was an important lesson for variance and risk management too.

  1. For example, we could think of the variance (risk) of the portfolio as coming from its various betas (assuming for simplicity that the factors are uncorrelated)

$$\sigma^2(R^{ei}) = b_i^2\sigma^2(rmrf) + h_i^2\sigma^2(hml) + s_i^2\sigma^2(smb) + \sigma^2(\varepsilon)$$

  2. This "factor model" was also useful for risk management. For example, if we want to hold an asset R^{ei}, the model can tell us "what will happen to the asset if the market goes down another 10%."

3. That information is especially useful in forming a *portfolio.* For example, if you have two securities R^{ei} and R^{ej}, the variance of an equally-weighted portfolio is

$$var(R^{ei} + R^{ej}) = \sigma^2(R^{ei}) + \sigma^2(R^{ej}) + 2cov(R^{ei}, R^{ej})$$

Well, modeling covariances is really tough. Instead, though, what if you use the factor model? Then

$$var(R^{ei}+R^{ej}) = (b_i + b_j)^2 \sigma^2(rmrf)+(h_i+h_j)^2\sigma^2(hml)+(s_i + s_j)^2 \sigma^2(smb)+\sigma^2(\varepsilon^i)+\sigma^2(\varepsilon^j)+2cov(\varepsilon^i\varepsilon^j)$$

Once you know what each beta is (covariance with the factors) then you can quickly figure out the variance of the portfolio, especially if the ε are small and uncorrelated with each other. You can also quickly figure out how much of j to buy in order to hedge i and make this portfolio variance small. *This is why risk management is practically all conducted in the context of factor models. (e.g. Barra)*

4. All of this is much easier if the errors are small. For bonds, the errors are essentially zero, which makes these uses of a factor model even more compelling.

26.1.5 Summary

1. Find the eigenvalue decomposition of yield covariance matrix,

$$Q\Lambda Q' = cov(y, y')$$

2. The columns of Q tell you how much yields of each maturity move when that factor moves – they are "factor loadings". Example: "First factor" is typically a "level factor" so $Q_1 = \begin{bmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{bmatrix}'$. This means "if factor 1 moves 1, all yields go up by 1/5." These are also regression coefficients of yields on factors.

3. The columns of Q also tell you how much to weight each yield when you construct the *factor* from the *yields.* Example: you find the first factor at any point in time by taking the average of the 5 yields.

4. The elements of diagonal Λ are the variance of each of the (uncorrelated with each other) factors.

5. A "factor model" comes from ignoring the very small factors – small Λ.

6. A factor model for *yields* also implies a factor model for *prices* or *forward rates.* You can also directly factor-analyze prices, forward rates, returns, changes in yields, prices, forward rates or returns. The results are very similar.

26.2 Simplest term structure model – expectations hypothesis

- Now, a third approach to a factor model. In this case we will "derive" a factor model. Here, I'll use the expectations hypothesis to connect bonds of various maturity. That's a shortcut, real term structure models start with $P = E(mx)$, but that adds a bit of algebra and not much clarity. The logic is the same.

- Here is the simplest model, which shows you the logic. We specify a "short rate process" and use the expectations hypothesis to get other maturities.
- Start with an AR(1) for the short rate,

$$\left(y_{t+1}^{(1)} - \delta\right) = \rho(y_t^{(1)} - \delta) + \varepsilon_{t+1}.$$

- Now, let's find all the other forward rates assuming the expectations hypothesis. It's easy to calculate forecasts of an AR(1)!:

$$f_t^{(2)} = E_t(y_{t+1}^{(1)}) = \delta + \rho(y_t^{(1)} - \delta)$$

$$f_t^{(3)} = E_t(y_{t+2}^{(1)}) = \delta + \rho^2(y_t^{(1)} - \delta)$$

$$f_t^{(N)} = E_t(y_{t+N-1}^{(1)}) = \delta + \rho^{N-1}(y_t^{(1)} - \delta)$$

Yields are just a little more complicated

$$\begin{aligned}
y_t^{(2)} &= \frac{1}{2}\left[E_t(y_{t+1}^{(1)}) + y_t^{(1)}\right] \\
&= \frac{1}{2}\left[\delta + \rho(y_t^{(1)} - \delta) + y_t^{(1)}\right] \\
y_t^{(2)} - \delta &= \frac{1+\rho}{2}\left(y_t^{(1)} - \delta\right) \\
&\cdots \\
y_t^{(N)} - \delta &= \frac{1+\rho+\rho^2+..}{N}\left(y_t^{(1)} - \delta\right) = \frac{1-\rho^N}{N(1-\rho)}\left(y_t^{(1)} - \delta\right)
\end{aligned}$$

Alternatively, calculate yields from the forward rates

$$\begin{aligned}
y_t^{(2)} &= -\frac{1}{2}p_t^{(2)} = -\frac{1}{2}\left(p_t^{(2)} - p_t^{(1)} + p_t^{(1)}\right) \\
y_t^{(2)} &= \frac{1}{2}\left(y_t^{(1)} + f_t^{(2)}\right) \\
y_t^{(2)} &= \frac{1}{2}\left(y_t^{(1)} + \delta + \rho(y_t^{(1)} - \delta)\right) \\
y_t^{(2)} - \delta &= \frac{1+\rho}{2}\left(y_t^{(1)} - \delta\right)
\end{aligned}$$

and so on.

- Look what we have
  1. *This is a one-factor model with the one-year rate as the single factor.*
  2. The model can produce different shapes of the forward and yield curve, including upward and downward sloping. The curve is upward sloping when short rates lower than average ($y_t^{(1)} < \delta$) and thus are expected to rise. This captures a lot of reality!

3. You want more complex shapes or more factors? Move past the simple AR(1)! An example:

$$\begin{aligned} z_t &= z_{t-1} + \varepsilon_t \\ x_t &= \phi x_{t-1} + \delta_t \end{aligned}$$

$$\begin{aligned} y_t^{(1)} &= z_t + x_t \\ f_t^{(2)} &= E_t y_{t+1}^{(1)} = z_t + \phi x_t \\ f_t^{(n)} &= E_t y_{t+n-1}^{(1)} = z_t + \phi^{n-1} x_t \end{aligned}$$

z_t is a "level" factor and x_t is a "slope" factor.

$$f_t^{(n)} - y_t^{(1)} = \left(\phi^{n-1} - 1\right) x_t$$

- How is this different from a real term structure model? Real models start with $P_t^{(n)} = E_t(m_{t+1}m_{t+2}...m_{t+n} \times 1)$, but otherwise follow the same logic. The expectations hypothesis is only an approximation. Doing it right, real models add extra $1/2\sigma^2$ terms. For example, the "single-factor Vasicek" model looks just like our model but with some extra terms.

$$\left(y_{t+1}^{(1)} - \delta\right) = \rho\left(y_t^{(1)} - \delta\right) + \varepsilon_{t+1}.$$

$$\begin{aligned} \left(f_t^{(2)} - \delta\right) &= \rho\left(y_t^{(1)} - \delta\right) - \left[\frac{1}{2} + \lambda\right]\sigma_\varepsilon^2 \\ \left(f_t^{(3)} - \delta\right) &= \rho^2\left(y_t^{(1)} - \delta\right) - \left[\frac{1}{2}(1+\rho)^2 + \lambda(1+\rho)\right]\sigma_\varepsilon^2 \\ &.... \end{aligned}$$

- What do we use models like this for? When you do it right, you are sure there are no arbitrage opportunities. With a single factor, *every* term structure derivative depends only on the single factor. Then you can price omitted maturities or term structure options. The idea is exactly the same as Black and Scholes' idea to price stock options from the "single factor" of the underlying stock, by arbitrage.

26.3 What Is and Is not Important about CP

- What is important, and very robust: *A single factor model for expected returns.* What is not important: tent vs slope, or the ability of additional variables to forecast returns

- OK. Now, with multiple securities and forecasts *expected returns move around over time too.* $E_t rx_{t+1}^{(n)}$ is a random variable *at time t.* So, how correlated is variation in *expected* return of one bond with another, what is $cov(E_t(rx_{t+1}^{(2)}), E_t(rx_{t+1}^{(3)}))$? Stop and wrap you your mind around the concept that expected returns vary over time and can be correlated with each other.

  For example, can the expected return of a two year bond go up and a three year bond go down? Can the expected return on stocks rise and the expected return on bonds fall?

1. The real, end (I think) enduring point of the CP model is that *there is a near-perfect single-factor model of all bond expected returns. Expected returns on all bond maturities rise and fall together. It does not happen that expected return on one bond goes up while the expected return on another bond goes down.* In equations

$$E_t(rx_{t+1}^{(n)}) = b_n(\text{factor}_t)$$

$$cov\left[E_t(rx_{t+1}^{(n)}), E_t(rx_{t+1}^{(m)})\right] = b_n b_m \times \sigma^2(\text{factor}_t)$$

   it's like a sigle colum of q a single λ, and zeros everywhere else.

2. For counterexample, the Fama-Bliss regressions suggest that there are four factors in the four expected returns. If

$$\begin{aligned} E_t\left(rx_{t+1}^{(2)}\right) &= a_2 + b_2(f_t^{(2)} - y_t^{(1)}) \\ E_t\left(rx_{t+1}^{(3)}\right) &= a_3 + b_3(f_t^{(3)} - y_t^{(1)}) \\ &\quad etc. \end{aligned}$$

   Then it is perfectly possible for $E_t\left(rx_{t+1}^{(2)}\right)$ and $E_t\left(rx_{t+1}^{(3)}\right)$ to go their separate ways.

3. Now, (a little more subtle) even the FB regressions could have a one-factor model of expected returns if $(f_t^{(2)} - y_t^{(1)})$ and $(f_t^{(3)} - y_t^{(1)})$ were perfectly correlated: If there were a one-factor model of *forward rates* then FB regressions would imply a one-factor model of *expected returns*. But the fact is that's not true. There are at least three forward-rate factors (level, slope, and curvature) so FB regressions imply a three-factor model of expected returns.

- *What's not important about CP:* The tent vs. the slope. It's too easy to get swept up in the R^2 contest, and whether out of sample the average FB forecast does better or worse than CP. Suppose when all is said and done, we decide that the CP factor really is slope, i.e. something like

$$E_t(rx_{t+1}^{(n)}) = a_n + b_n\left(-2y_t^{(1)} - 1f_t^{(2)} + f_t^{(4)} + 2f_t^{(5)}\right)$$

  rather than

$$E_t(rx_{t+1}^{(n)}) = a_n + b_n\left(-2y_t^{(1)} + 1f_t^{(2)} + 2f_t^{(3)} + 1f_t^{(4)} - 2f_t^{(5)}\right).$$

  We will be perfectly happy. The exact shape of the coefficients that recover the factor is not well identified.

- Similarly, it is likely that in the end other variables will help to forecast returns. *The interesting question* in the end is whether those variables leave the single factor structure intact, i.e., say

$$E_t(rx_{t+1}^{(n)}) = a_n + b_n\left(-2y_t^{(1)} - 1f_t^{(2)} + f_t^{(4)} + 2f_t^{(5)} + \beta x_t\right)$$

  The interesting question *is not* whether such variables exist, i.e. if $\beta = 0$. We will be perfectly happy with $\beta \neq 0$

- Why we're excited about CP. What is the *factor structure of expected returns* more broadly? Is there some combination of DP and CP that forecasts both stocks and bonds? Is it true that

$$\begin{aligned} \text{stocks:} \qquad & E_t R^{em}_{t+1} = b_m \left(\beta \times CP_t + \gamma \times DP_t \right) \\ \text{bonds:} \qquad & E_t rx^{(n)}_{t+1} = b_n \left(\beta \times CP_t + \gamma \times DP_t \right) \end{aligned}$$

  If not, is there at least "factor structure" whereby a few factors describe most of the variance of expected returns across asset classes? The pictures in "Discount rates" of lots of prices moving together, and my interpretation of a big common discount rate shock, certainly suggests it!

- *Uniting the models.* So far, we have

$$E_t(rx^{(n)}_{t+1}) = b_n \times (\text{cp factor}_t)$$

  (So far, this is all just "describe," not "explain") In "Decomposing the yield curve," Piazzesi and I look for factors to account for expected returns. Expected returns should correspond to covariance with something, no? The answer is

$$E_t(rx^{(n)}_{t+1}) = cov(rx^{(n)}_{t+1}, \text{level}_{t+1}) \times (\text{cp factor}_t)$$

  The CP factor describes how expected returns vary *through time*. The b_n coefficients that tell you how much each bond expected return loads on the factor is given by the covariance of return with the level factor!

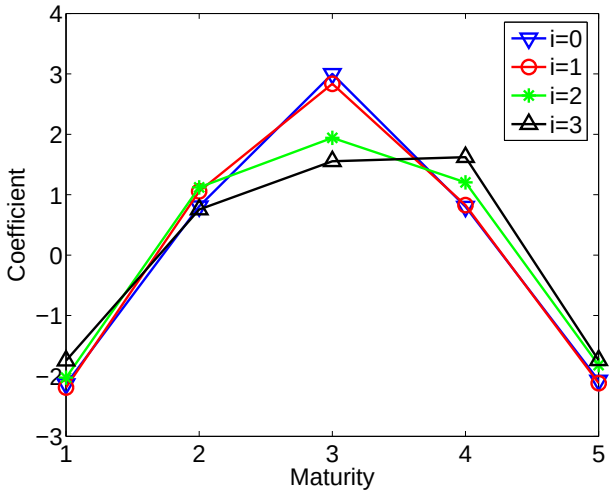
27 Term structure extras

The following is *not required material either in class problem set or final.* In previous years I did two lectures on interest rates, and some of this is the second lecture that you're missing. If you're interested at some point, you might find it useful. If you're having enough problems keeping up with what we actually did, skip it!

27.1 CP extras

We do a lot of extra work to convince you it's real and robust. This is the "notes" version of material in the paper, and some extensions. I won't cover this in class unless by miracle I have extra time.

27.1.1 More lags



$$
\begin{aligned}
rx_{t+1}^{(n)} &= a_n + \beta_n' f_{t-i} + \varepsilon_{t+1}^{(n)} \\
rx_{t+1}^{(n)} &= a_n + \sum_{i=1}^{3} \beta_n^{i\prime} f_{t-i} + \varepsilon_{t+1}^{(n)}
\end{aligned}
$$

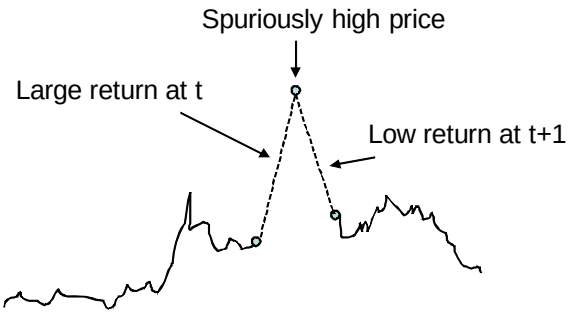
- More lags are significant, with the same pattern.
- Checking individual lags reassures us it's not just measurement error, i.e.

$$p_{t+1} - p_t = a + bp_t + \varepsilon_{t+1}$$

if p_t is measured with error, you'll see something. But

$$p_{t+1} - p_t = a + bp_{t-1/12} + \varepsilon_{t+1}$$

fixes this problem.



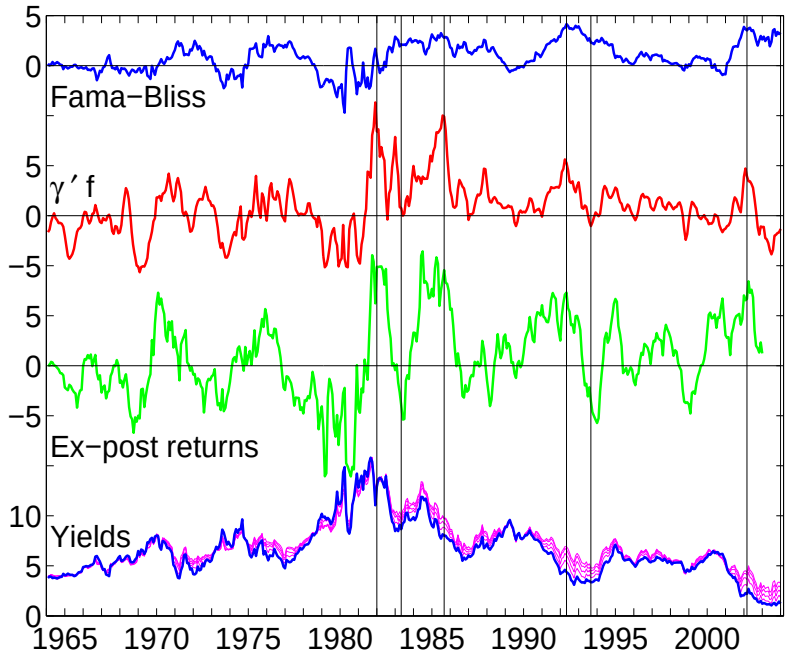
•The pattern suggests moving averages

$$rx_{t+1} = a + \gamma' \left(f_t + f_{t-1} + f_{t-2} + ... \right) \varepsilon_{t+1}$$

k	1	2	3	4	6
R^2	0.35	0.41	0.43	0.44	0.43

•Interpretation: Yields or forwards at t *should* should carry all information about the future. If the lags enter, there must be a little measurement error. f change slowly over time, so $f_{t-1/12}$ is informative about the true f_t. Moving averages are a good way to enhance a slow-moving "signal" buried in high-frequency "noise."

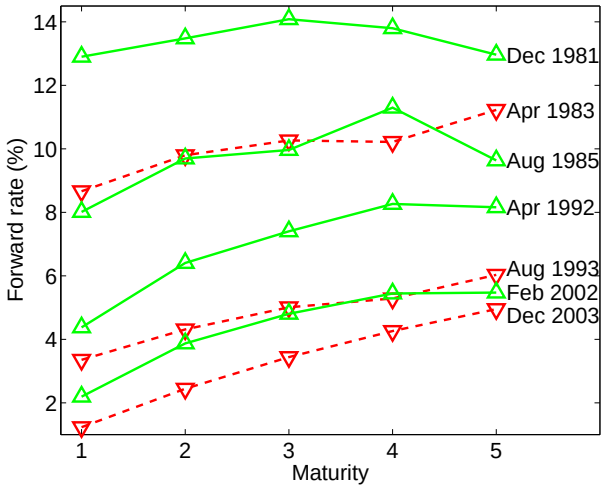
27.1.2 History



•$\gamma' f$ predictions are consistent in many episodes

•$\gamma' f$ and slope are correlated. Both show a rising yield curve but no rate rise

•$\gamma' f$ improvement in many episodes. $\gamma' f$ says get out in 1984, 1987, 1994, 2004. What's the signal....?

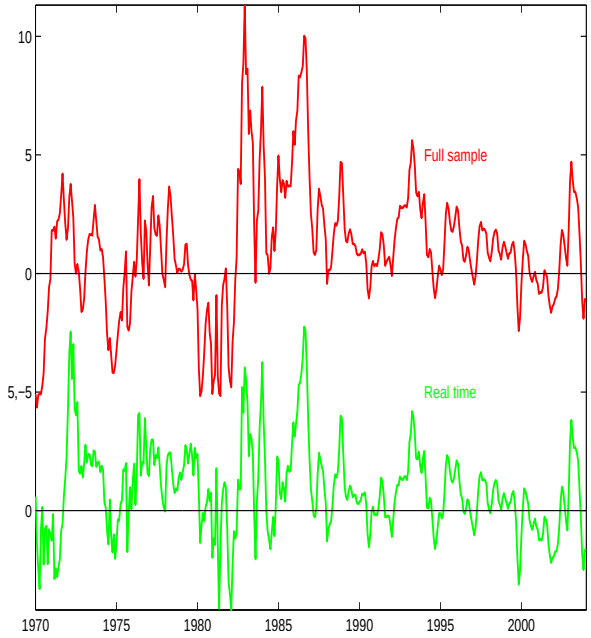


•Green: CP say go. Red: FB say go, CP say no.

•Tent-shaped coefficients interact with tent-shaped forward curve to produce the signal.

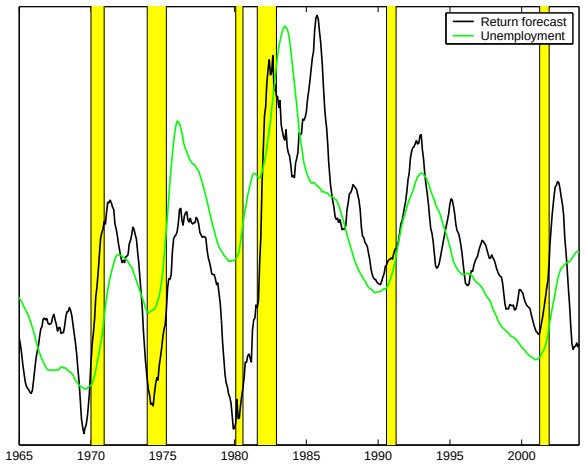
•CP: in the past, tent-shape often came with upward slope. Others saw upward slope, thought that was the signal. But an upward slope without a tent does not work. The tent is the real signal.

27.1.3 Real time



Regression forecasts $\hat{\gamma}^\top f_t$. "Real-time" re-estimates the regression at each t from 1965 to t. (Note: Just because we don't have data before 1964 doesn't mean people don't know what's going on. Out of sample is not crucial, but it is interesting.)

27.1.4 Macro



•Is it real, a time-varying risk premium? Or is it some new psychological "effect," an unexploited profit opportunity?

• Here, $\gamma' f$ is correlated with business cycles, and lower frequency. (Level, not growth.) Suggests a "business cycle related risk premium."

• It's also significant that the same signal predicts all bonds, and predicts stocks. If "overlooked" it is common to a lot of markets!

•**The Point** of history, real time and macro: *this is the kind of analysis you do to make sure you're not finding a fish. 1/20 t statistics look good. You need to persuade yourself.*

27.1.5 Failures and spread trades

•What this *is* about (so far): when, overall is there a risk premium (high expected returns) in long term vs. short term bonds. "Trade" is just betting on long vs. short maturity, "betting on interest rate movements."

•What this is *not* about (so far). Much fixed income "arbitrage" involves relative pricing, small deviations from the yield curve. "Trade" might be short 30 year, long 29.5 year.

A hint of spread trades

If the one-factor model is exactly right, then *deviations from the single-factor model should not be predictable.*

$$rx_{t+1}^{(2)} - b_2\overline{rx}_{t+1} = a^{(2)} + 0'f_t + \varepsilon_{t+1} = a^{(2)} + 0'y_t + \varepsilon_{t+1}$$

(Why?

$$\begin{aligned} rx_{t+1}^{(2)} &= \alpha^{(2)} + b_2\left(\gamma' f_t\right) + \varepsilon_{t+1}^{(2)} \\ \overline{rx}_{t+1} &= \alpha + \gamma' f_t + \varepsilon_{t+1}^{(2)} \end{aligned}$$

multiply the second by b_2 and subtract.)

Table 7. Forecasting the failures of the single-factor model

A. Coefficients and t-statistics

		Right hand variable				
Left hand var.	const.	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(3)}$	$y_t^{(4)}$	$y_t^{(5)}$
$rx_{t+1}^{(2)} - b_2\overline{rx}_{t+1}$	−0.11	−0.20	**0.80**	−0.30	−0.66	0.40
(t-stat)	(−0.75)	(−1.43)	**(2.19)**	(−0.90)	(−1.94)	(1.68)
$rx_{t+1}^{(3)} - b_3\overline{rx}_{t+1}$	0.14	0.23	−1.28	**2.36**	−1.01	−0.30
(t-stat)	(1.62)	(2.22)	(−5.29)	**(11.24)**	(−4.97)	(−2.26)
$rx_{t+1}^{(4)} - b_4\overline{rx}_{t+1}$	0.21	0.20	−0.06	−1.18	**1.84**	−0.82
(t-stat)	(2.33)	(2.39)	(−0.33)	(−8.45)	**(9.13)**	(−5.48)
$rx_{t+1}^{(5)} - b_5\overline{rx}_{t+1}$	−0.24	−0.23	0.55	−0.88	−0.17	**0.72**
(t-stat)	(−1.14)	(−1.06)	(1.14)	(−2.01)	(−0.42)	**(2.61)**

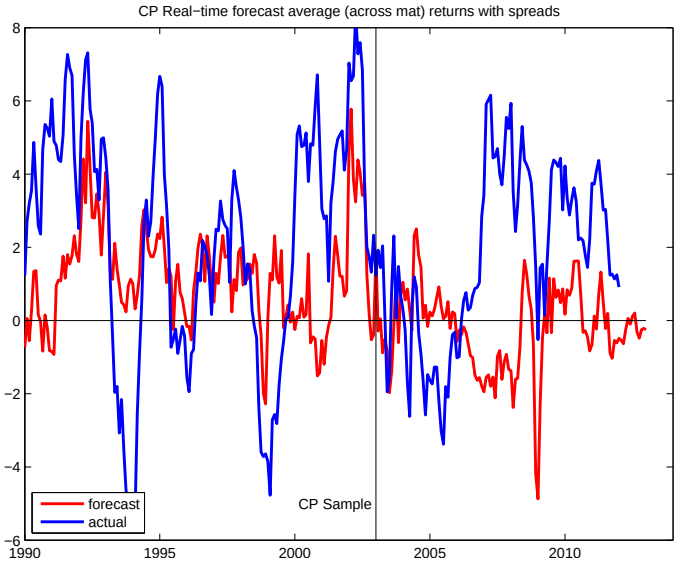
B. Regression statistics

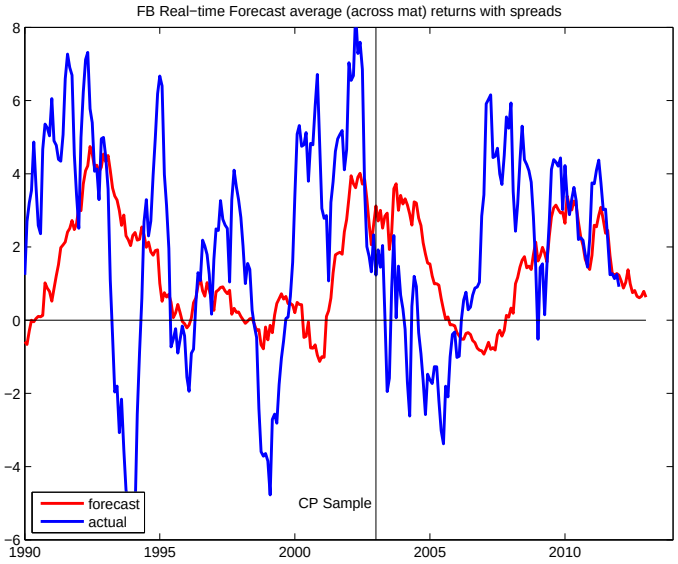
Left hand var.	R^2	$\chi^2(5)$	$\sigma(\tilde{\gamma}^\top y)$	σ(lhs)	$\sigma(b^{(n)}\gamma^\top y)$	$\sigma(rx_{t+1}^{(n)})$
$rx_{t+1}^{(2)} - b_2\overline{rx}_{t+1}$	0.15	41	0.17	0.43	1.12	1.93
$rx_{t+1}^{(3)} - b_3\overline{rx}_{t+1}$	0.37	151	0.21	0.34	2.09	3.53
$rx_{t+1}^{(4)} - b_4\overline{rx}_{t+1}$	0.33	193	0.18	0.30	2.98	4.90
$rx_{t+1}^{(5)} - b_5\overline{rx}_{t+1}$	0.12	32	0.21	0.61	3.45	6.00

• Pattern: if $y^{(n)}$ is a little out of line with the others (low price), then $r^{(n)}$ is good *relative* to all the others.

- No common factor. Bond-specific mean-reversion.
- This is *tiny*. 17-21 bp, compare to 200-600 bp returns.
- The single-factor $\gamma' f$ accounts for all the *economically important* variation in expected returns
- But the left hand side is tiny too, so tiny/tiny = good R^2
- Tiny isn't so tiny if you leverage up like crazy!
- But... measurement error looks the same.

27.1.6 Latest data



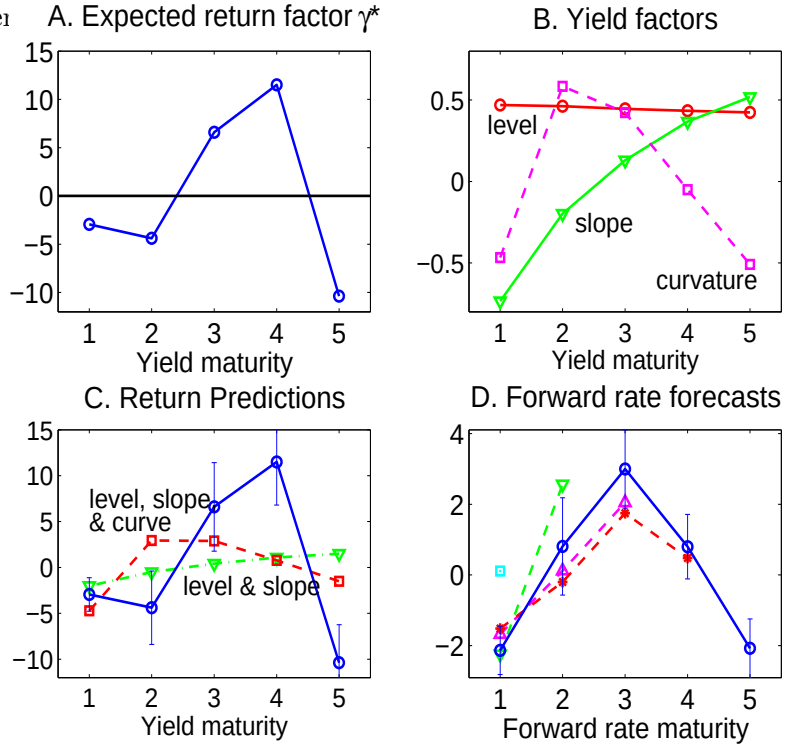


Out of sample looked pretty good until the financial crisis. CP got the low level right and matched the ups and downs a lot better than the FB forecast. We both missed the great returns to long-term bond holders before the financial crisis (big negative spike). We got the big negative spike in returns with the crisis. We've been matching the ups and downs pretty well since then, though off on the overall level. Of coure, the Fama-Bliss forecast has been doing spectacularly well since the crisis, so the correlation of yield spread with tent is the big story, and in fact tent is hurting relative to yield spread.

However, none of this bother us that much. *The main point of the CP model is that there is a single-factor model of expected bond returns. It does not really matter whether that single factor is recovered by a "slope" operator, a "tent" operator, or includes additional variables.* More on this point, with equations, after we discuss a bit what "factor models" means.

27.1.7 CP factor vs. level slope and curvature

The CP return-forecasting variable $\gamma' f$ is a "factor" – it's a linear combination of forward rates or yields. Here, we express it as a function of yields to compare it to the level slope and curvature factors. This is her



Panel A: Panels A and B are the *loadings*, how you construct $\gamma' f$ and yield factors from underlying yields. This is the tent, expressed as a function of yields. We can express $\gamma' f$ as a function of yields too. $\gamma' f = \gamma^{*\prime} y$; What *yield* curve signals high returns on long term bonds (not *forward* curve)? A: $\gamma^* \approx$ Slope plus 4-5 spread.

Panel B: $\gamma' f$ has *nothing* to do with level slope and curvature slope and curvature ($\gamma' f$ is curved at the long, not short end).

Panel C: We can forecast returns using *factors* at time t, to try to summarize the information in the entire yield curve. $rx_{t+1} = a + b \times level_t + c \times slope_t + \varepsilon_{t+1}$. Then, from the construction $level_t = q' y_t$, etc, you can figure out the implied regression coefficients of rx_{t+1} on y_t. The graph plots those implied coefficients. Moral: You can't approximate the forecasting power of $\gamma' f$ well by running rx_{t+1} on level, slope, and curvature factors.

•Moral 1 Term structure models need L, S, C to get yield behavior *and* $\gamma' f$ to capture the information in the yield curve about expected returns.

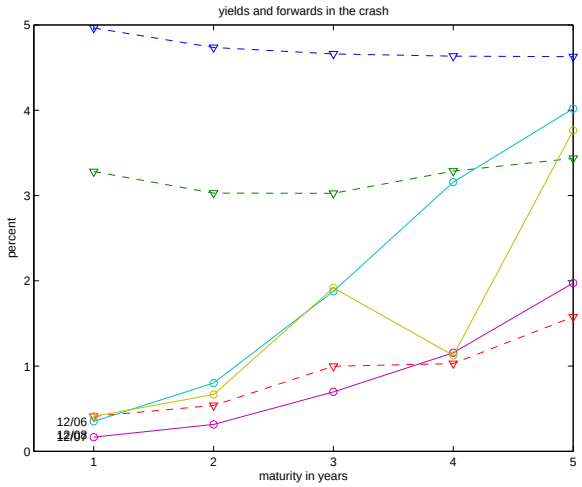
•Adding $\gamma' f$ will not help much to hit yields (pricing errors) but it will help to get *forecasts* right

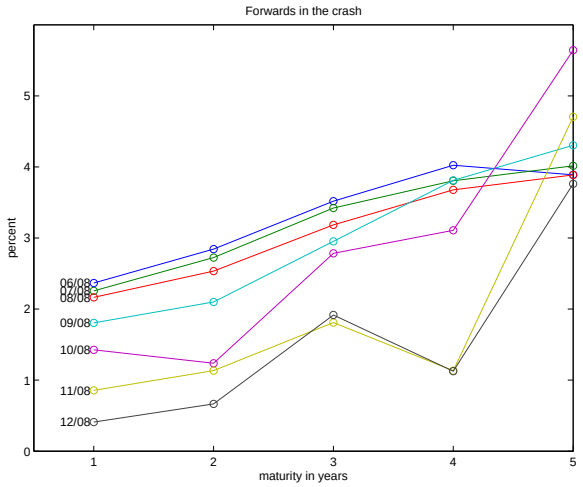
•Moral 2. You can't *first* reduce to L, S, C, *then* examine $E_t rx_{t+1}$ →Reason #1 this was missed.

Panel D: The tent pattern is stable as we add forward rates, a guard against multicollinearity. (Not really related, but we had a blank panel to fill!)

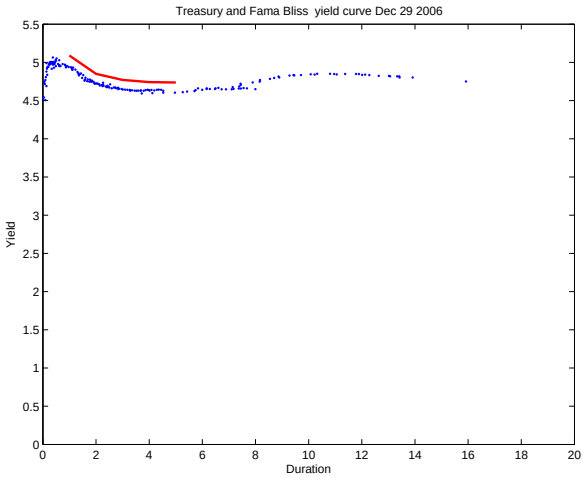
27.1.8 Treasury curves during the crash

Some of this comes down to weirdness in the treasury yield curve during the crisis. In Dec 2008 there was a huge demand for treasuries that could be repoed to finance other positions. There was also a separaton into two yield curves, one for treasuries with 6% couponds and one for the others, documented in the Pancost extra reading below. This is leading to very weird treasury yield curves, which I think are beyond the FB data construction technique to make sense of .

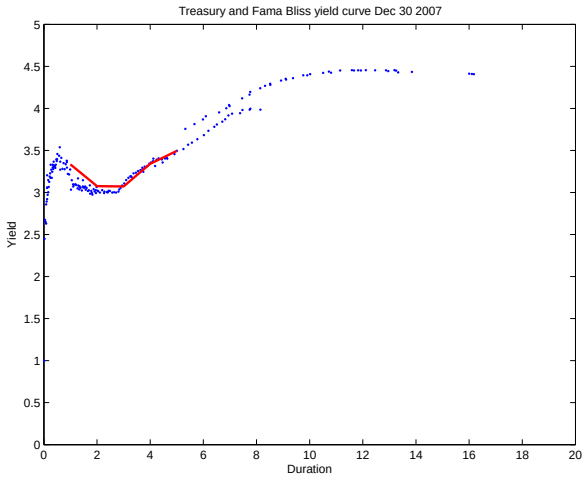




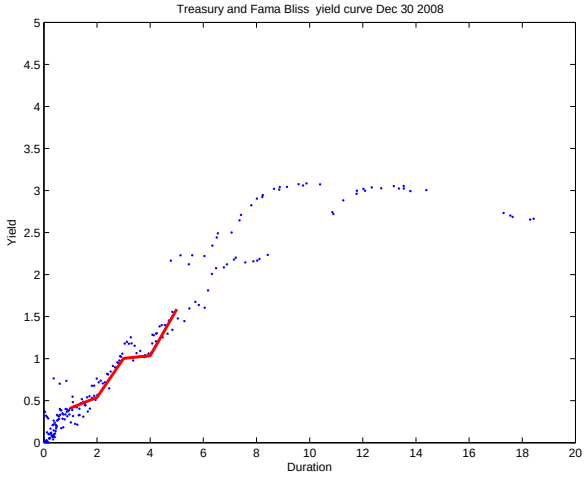
Here are the underlying data. Each bond is a blue dot. The red line is the FB fit to the yield curve. You can see it's pretty hopeless lately because there's so much noise in blue dots In 06 it was all sensible



In Dec 07 things were getting worse. The massive upward curve causes the CP factor to say returns should be extremely negative.

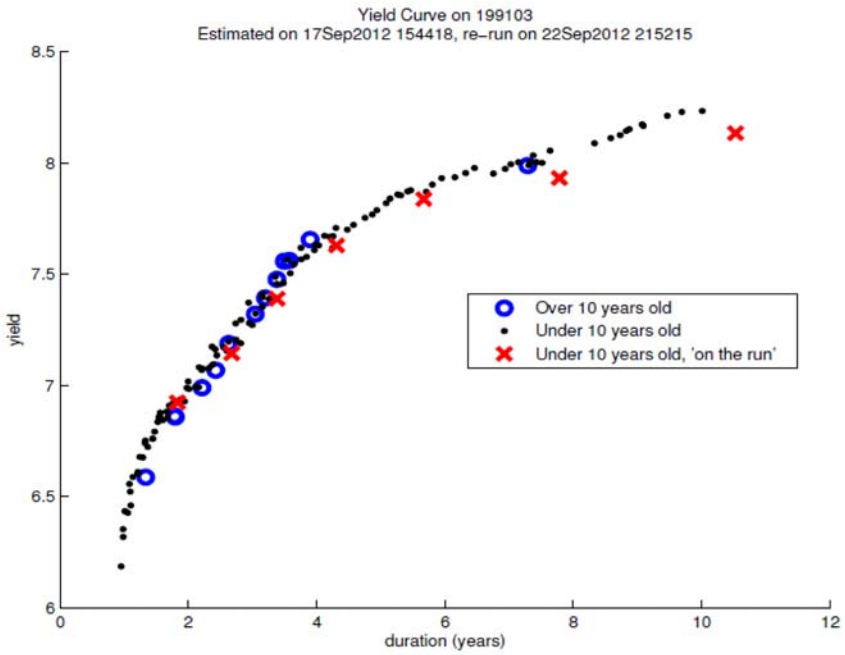


Now it's just a mess. The zig zag in yields gives rise to the weird forward rate pattern. Look at these huge spreads for bonds of the same duration! I think that's the on the run (and deliverable to short/repo) vs. off the run spread.



Moral: In future research, I need to find better ways to extract the yield curve from data like this.

Aaron Pancost (optional reading) dug in to treasury curves in Dec 2008, and found another nifty "arbitrage" in the financial crisis



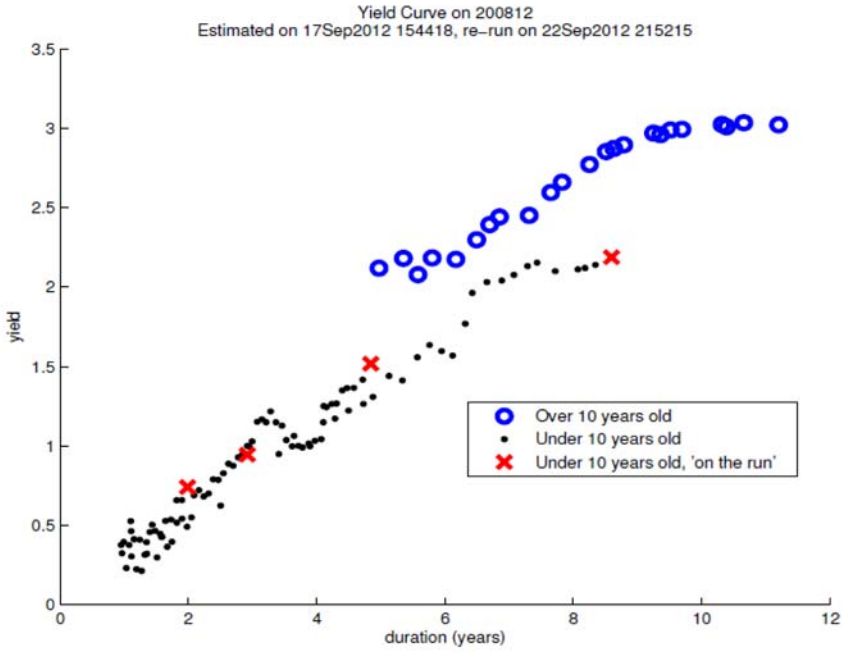
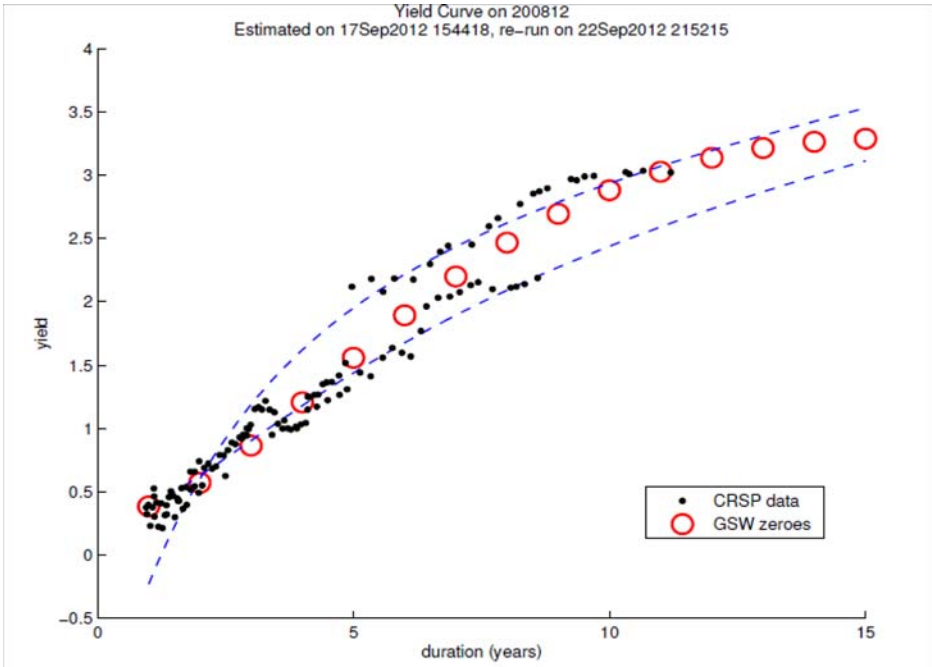


Figure 6: Yield Curves in 1991 and 2008



(b) Actual CRSP Yields in December 2008

Figure 1: Estimated Yield Curves In Late 2008

According to Aaron, UBS had a huge portfolio of old high coupon bonds and dumped them in Dec 2008. It's an an interesting "downward sloping demand" "apparent arbitrage" "no, a stop loss order is not a put option," "fire sale," etc. story. Quotes from an un-named source:

> Since I'm sitting on the xxx fixed income trading floor today, I walked over and talked with one of the main Treasury bond traders who was here during the 2008 crisis period. Here's his take.
>
> 1. The numbers are real. There really was a large disconnect in the market during this period, although a 40 basis point mispricing was pretty small relative to some of the other strange stuff we were seeing back then.
>
> 2. The proximate cause for the strange pricing here is UBS. UBS was the largest dealer in the Treasury market at that time given their willingness to commit balance sheet. When UBS along with everybody else started losing money, they pulled back their balance sheet commitments and started selling big chunks of their Treasury portfolio.
>
> 3. Given that these older bonds had always been a bit on the cheap side given their illiquidity, UBS had put together a large position in these older bonds. Typically, these bonds were about 5bps cheap with gusts up to 15bps. When UBS suddenly had to start unwinding their balance sheet commitments, these bonds were differentially affected since UBS was selling them much more than they were the notes simply because that's what their inventory was. [JC: I.e. following my advice that the trader who doesn't want liquidity should buy the cheap illiquid asset, but forgetting my advice about the "average investor," and they might indeed be the trader who wants liquidity! You can't leverage a liquidity-providing position!]

4. Also, these longer bonds were traded primarily on the STRIPS desk which isn't the same as the usual Treasury desk. This opened the door to some disconnects between the markets.

5. Anecdotally, I'm told that UBS lost upwards of 200MM on the unwind of these long bond positions.

6. Academics will ask, why wasn't this stuff arbed? But you have to remember the environment when nobody was willing to commit balance sheet or take the mark to market risk. There were far larger arbs in the market than this 40 basis point mispricing. There wasn't any risk capital for something this small.

27.2 Factor models in FX

- We return to Lustig Roussanov and Verhdelahn and review the factor models in the 1-6 portfolios.
- (Possibly) we return to Brant and Kavajecz and review the factor model in daily yield changes.

28 Term Structure III Term structure models notes

- *Reminders 1* Two steps in everything we're doing
  1. Form the factors (level, slope) from yields (returns, etc) $\text{level}_t = ...$, $\text{slope}_t = ...$
  2. *Loadings* or *betas*; $y_t^{(n)} = ...\text{level}_t + ...\text{slope}_t...$
- *Reminder 2*
  1. Our first attempt: a) Form factors from intuition, b) Run regressions to find loadings

$$\begin{aligned} \text{level}_t &\equiv \frac{1}{5}\left[y_t^{(1)} + y_t^{(2)} + y_t^{(3)} + y_t^{(4)} + y_t^{(5)}\right] \\ \text{slope}_t &= \frac{1}{2}\left[y_t^{(4)} + y_t^{(5)}\right] - \frac{1}{2}\left[y_t^{(1)} + y_t^{(2)}\right]. \end{aligned}$$

$$y_t^{(n)} = a_n + b_n \times \text{level}_t + s_n \times \text{slope}_t + \varepsilon_t^{(n)}$$

  2. Our second attempt. Use principal components to maximize R^2. Q answers both questions.

$$\begin{aligned} Q\Lambda Q' &= cov(y_t, y_t') \\ x_t &= Q'y_t \\ y_t &= Qx_t \end{aligned}$$

     Leave out columns of Q corresponding to small eigenvalues – small factors.
  3. Now: More ways to refine the same two steps.
- *Purpose:*

1. The factor model was nice, but it could only describe bonds that we included from the beginning.
2. How do you *interpolate the yield curve* across maturities? What should the yield be of a 2.5 year bond? What should be the yield be of a 6 year bond? How do we find a good set of $b^{(n)}, c^{(n)}, d^{(n)}$ to calculate, for example,

$$y_t^{(2.5)} = b^{(2.5)} \times \text{level}_t + c^{(2.5)} \times \text{slope}_t + d^{(2.5)} \times \text{curve}_t$$

3. How do you price term structure options, other derivatives? How do you "Interpolate" (or extrapolate) *across instruments*, as Black and Scholes do for stock and bond? Example: what price should we charge for special products – ARM mortgages with an interest rate cap? This is a big business.
4. In both cases, we want to extend in a way that doesn't introduce *arbitrage opportunities* or *huge Sharpe ratios.*
5. Example: if $y^{(2)} = 2\%$ and $y^{(3)} = 3\%$, maybe you fit $y^{(2.5)} = 2.5\%$. But are we sure this price doesn't give an arbitrage opportunity or a huge Sharpe ratio to a hedge fund if they could do fancy long-short strategies between 2, 2.5 and 3 year bonds? If 2.53% was right, they'll bury you.

28.1 Real, "arbitrage free" term structure models

- Still to do:

  1. This model has no average slope, and no risk premium – $E(y^{(i)})$ are all the same. That isn't a surprise, we imposed the expectations hypothesis! A better model has to incorporate risk premiums.
  2. One nagging problem – are we sure we haven't let in arbitrage opportunities that clever options traders might exploit?

- *Approach:* $P = E(mx)$ of course. This is equivalent to "risk-neutral" "arbitrage-free" approaches.

- *Idea:*

  1. No arbitrage opportunities $\Longleftrightarrow$ we can generate prices with some marginal utility $m > 0$.
  2. m allows you to introduce risk premia, i.e. recall $E(R^e) = cov(R^e, m)/E(m)$
  3. The ideal approach would be to write, say,

$$m_{t+1} = \beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma}$$

  Then

$$\begin{aligned} P_t^{(1)} &= E_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} \times 1 \right] \\ P_t^{(2)} &= E_t \left[\beta^2 \left(\frac{c_{t+2}}{c_t} \right)^{-\gamma} \times 1 \right] \end{aligned}$$

etc. This would tie bond prices, returns to macroeconomic events more than our hand waving so far. Alas, these don't work in practically useful ways yet.

4. What current term structure models do:
   (a) Write a *statistical* model for m, for example,
   $$m_{t+1} = \mu + \rho m_t + \varepsilon_{t+1}.$$
   (b) Find bond prices for any maturity from
   $$\begin{aligned} P_t^{(1)} &= E_t\left(m_{t+1} \times 1\right) = \mu + \rho m_t \\ P_t^{(2)} &= E_t\left(m_{t+1} m_{t+2} \times 1\right) = E_t(m_{t+1} P_{t+1}^{(1)}) = E_t(m_{t+1}\left(\mu + \rho m_{t+1}\right))... \end{aligned}$$
   and so forth. (This is equivalent to "short rate process" plus "market price of risk." $cov(m, \varepsilon)$ generates "market price of risk").
   (c) Calibrate parameters $\mu, \rho, \sigma^2(\varepsilon)$ to fit bond prices (or other securities) in hand (for example, choose a few maturities to match, or better, match "level" "slope" and "curvature" factors).
   (d) Use the m to find prices of all other bonds, swaps, options, etc.
5. This should remind you of the problem set in which you found an m to price stock and bond, and then used it to value an option. We're doing the same thing – finding an m that prices a few bonds, and then using that m to try to price all bonds and bond options.

28.1.1 Simplest term structure model, "discrete time Vasicek"

- It's the simplest example and shows the ideas.
- **Where we're going**: (the answer when we're all done) A "short rate equation" plus "one-factor arbitrage-free model of the term structure"
$$\left(y_{t+1}^{(1)} - \delta\right) = \rho\left(y_t^{(1)} - \delta\right) + \varepsilon_{t+1}.$$
$$\begin{aligned} \left(f_t^{(2)} - \delta\right) &= \rho\left(y_t^{(1)} - \delta\right) - \left[\frac{1}{2} + \lambda\right]\sigma_\varepsilon^2 \\ \left(f_t^{(3)} - \delta\right) &= \rho^2\left(y_t^{(1)} - \delta\right) - \left[\frac{1}{2}(1+\rho)^2 + \lambda(1+\rho)\right]\sigma_\varepsilon^2 \\ &.... \end{aligned}$$
$$\begin{aligned} \left(y_t^{(2)} - \delta\right) &= \frac{(1+\rho)}{2}\left(y_t^{(1)} - \delta\right) - \frac{1}{2}\left(\frac{1}{2} + \lambda\right)\sigma_\varepsilon^2 \\ \left(y_t^{(3)} - \delta\right) &= \frac{(1+\rho+\rho^2)}{3}\left(y_t^{(1)} - \delta\right) - \frac{1}{3}\left\{\frac{1}{2}\left[1 + (1+\rho)^2\right] + \lambda\left[1 + (1+\rho)\right]\right\}\sigma_\varepsilon^2 \\ &.... \end{aligned}$$
Intuition: the first equation tells you where interest rates are going over *time*. The second and third sets of equations tell you where each forward rate is *at any date*, depending only on where the short rate is on that date. This is just like the last one-factor model with a bit more complex constant term.

- **Construction**

1. Construction. Suppose m follows the time series model

$$\begin{aligned}(x_{t+1}-\delta) &= \rho(x_t-\delta)+\varepsilon_{t+1}\\ \ln m_{t+1} &= -x_t-\frac{1}{2}\lambda^2\sigma_\varepsilon^2-\lambda\varepsilon_{t+1}\end{aligned}$$

To be specific, ε are iid Normal. σ_ε^2, δ,ρ,λ are free parameters; we'll pick these to make the model fit as well as possible. (More intuition on this model below.) The idea is that we can't see x_t but market participants can; it carries the information about the conditional mean of their discount factors/marginal utility growth. (Again, view this as analogous to constructing a discount factor to price stock and bond, rather than as a deeply-founded idea.)

2. First bond price.

$$P_t^{(1)}=E_t(m_{t+1})=E_t\left(e^{\ln m_{t+1}}\right)$$

(a) Math fact[22]: If x is normal with mean μ and variance σ^2, then

$$E(e^x)=e^{\mu+\frac{1}{2}\sigma^2}. \tag{30}$$

Intuition: e^x is curved up, so $E(e^x)>e^{E(x)}$ (like risk aversion but backwards).

(b) Applying that fact (watch this line carefully!)

$$P_t^{(1)}=e^{E_t[\ln(m_{t+1})]+\frac{1}{2}\sigma_t^2[\ln(m_{t+1})]}=e^{-x_t-\frac{1}{2}\lambda^2\sigma_\varepsilon^2+\frac{1}{2}\lambda^2\sigma_\varepsilon^2}=e^{-x_t}$$

Take logs,

$$\begin{aligned}p_t^{(1)} &= \ln E_t(m_{t+1})=-x_t\\ y_t^{(1)} &= x_t\end{aligned}$$

Now you see why I set up the problem with the $1/2\lambda^2\sigma_\varepsilon^2$ to begin with!

(c) *The one year interest rate "reveals" the latent (not observed) state variable x_t.* So, we can write the first equation

$$\left(y_{t+1}^{(1)}-\delta\right)=\rho\left(y_t^{(1)}-\delta\right)+\varepsilon_{t+1}$$

just as we did with the expectations-hypothesis version above.

3. On to the next price. *Pay attention to this algebra*

$$\begin{aligned}P_t^{(2)} &= E_t\left(m_{t+1}P_{t+1}^{(1)}\right)\\ &= E_t\left(e^{\ln(m_{t+1})}e^{p_{t+1}^{(1)}}\right)=E_t\left(e^{\ln(m_{t+1})+p_{t+1}^{(1)}}\right)\\ &= E_t\left(e^{-\frac{1}{2}\lambda^2\sigma_\varepsilon^2-x_t-\lambda\varepsilon_{t+1}-x_{t+1}}\right)\end{aligned}$$

[22]Where does this come from?

$$E(e^x)=\frac{1}{2\pi\sigma}\int_{-\infty}^{\infty}e^x e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}dx$$

Complete the square and you can reexpress this as $e^{\mu+\frac{1}{2}\sigma^2 x}$ times the integral of a normal, which is 1.

Now we use the model for x,

$$x_{t+1} = \delta + \rho\left(x_t - \delta\right) + \varepsilon_{t+1},$$

$$x_{t+1} + x_t = 2\delta + (1+\rho)\left(x_t - \delta\right) + \varepsilon_{t+1},$$

substituting,

$$\begin{aligned} P_t^{(2)} &= E_t\left(e^{-\frac{1}{2}\lambda^2\sigma_\varepsilon^2 - \lambda\varepsilon_{t+1} - 2\delta - (1+\rho)(x_t-\delta) - \varepsilon_{t+1}}\right) \\ P_t^{(2)} &= E_t\left(e^{-2\delta-(1+\rho)(x_t-\delta) - \frac{1}{2}\lambda^2\sigma_\varepsilon^2 - (1+\lambda)\varepsilon_{t+1}}\right) \end{aligned}$$

Using (30) again,

$$\begin{aligned} P_t^{(2)} &= e^{-2\delta-(1+\rho)(x_t-\delta) - \frac{1}{2}\lambda^2\sigma_\varepsilon^2 + \frac{1}{2}(1+\lambda)^2\sigma_\varepsilon^2} \\ P_t^{(2)} &= e^{-2\delta-(1+\rho)(x_t-\delta) + (\frac{1}{2}+\lambda)\sigma_\varepsilon^2} \\ p_t^{(2)} &= -2\delta - (1+\rho)\left(x_t - \delta\right) + (\frac{1}{2}+\lambda)\sigma_\varepsilon^2 \end{aligned}$$

4. From prices, we find yields and forwards,

$$\begin{aligned} y_t^{(2)} &= -\frac{1}{2}p_t^{(2)} = \delta + \frac{(1+\rho)}{2}\left(x_t - \delta\right) - \frac{1}{2}\left(\frac{1}{2}+\lambda\right)\sigma_\varepsilon^2 \\ \left(y_t^{(2)} - \delta\right) &= \frac{(1+\rho)}{2}\left(y_t^{(1)} - \delta\right) - \frac{1}{2}\left(\frac{1}{2}+\lambda\right)\sigma_\varepsilon^2 \end{aligned}$$

$$\begin{aligned} f_t^{(2)} &= p_t^{(1)} - p_t^{(2)} \\ &= -\delta - \left(x_t - \delta\right) + 2\delta + (1+\rho)\left(x_t - \delta\right) - (\frac{1}{2}+\lambda)\sigma_\varepsilon^2 \\ f_t^{(2)} &= \delta + \rho\left(x_t^{(1)} - \delta\right) - (\frac{1}{2}+\lambda)\sigma_\varepsilon^2 \\ \left(f_t^{(2)} - \delta\right) &= \rho\left(y_t^{(1)} - \delta\right) - (\frac{1}{2}+\lambda)\sigma_\varepsilon^2 \end{aligned}$$

- Next price

$$P_t^{(3)} = E_t\left(m_{t+1}P_{t+1}^{(2)}\right).....$$

You don't really want to do it, right?

- Result:

$$\left(y_{t+1}^{(1)} - \delta\right) = \rho\left(y_t^{(1)} - \delta\right) + \varepsilon_{t+1}.$$

$$\begin{aligned} \left(f_t^{(2)} - \delta\right) &= \rho\left(y_t^{(1)} - \delta\right) - (\frac{1}{2}+\lambda)\sigma_\varepsilon^2 \\ \left(f_t^{(3)} - \delta\right) &= \rho^2\left(y_t^{(1)} - \delta\right) - \left[\frac{1}{2}(1+\rho)^2 + \lambda(1+\rho)\right]\sigma_\varepsilon^2 \\ \left(f_t^{(4)} - \delta\right) &= \rho^3\left(y_t^{(1)} - \delta\right) - \left[\frac{1}{2}\left(1+\rho+\rho^2\right)^2 + \lambda(1+\rho+\rho^2)\right]\sigma_\varepsilon^2 \\ &... \end{aligned}$$

$$\begin{aligned}
\left(y_t^{(2)} - \delta\right) &= \frac{(1+\rho)}{2}\left(y_t^{(1)} - \delta\right) - \frac{1}{2}\left(\frac{1}{2} + \lambda\right)\sigma_\varepsilon^2 \\
\left(y_t^{(3)} - \delta\right) &= \frac{(1+\rho+\rho^2)}{3}\left(y_t^{(1)} - \delta\right) - \frac{1}{3}\left\{\frac{1}{2}\left[1 + (1+\rho)^2\right] + \lambda\left[1 + (1+\rho)\right]\right\}\sigma_\varepsilon^2 \\
&\ldots
\end{aligned}$$

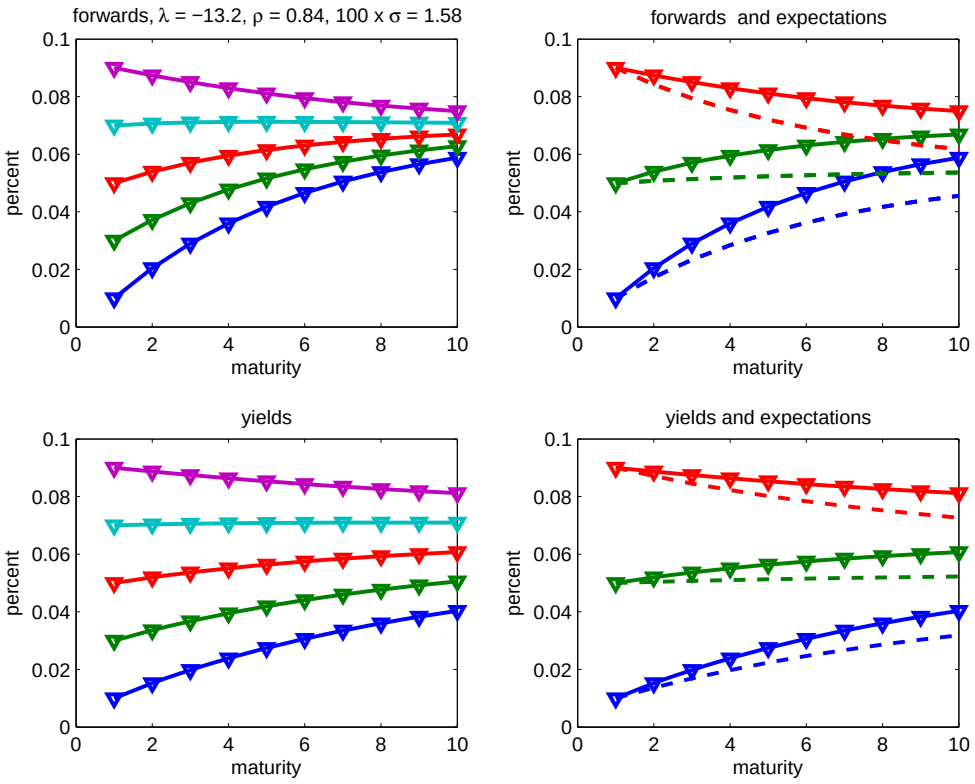
1. As before, we have a "short rate process" plus a "one factor model" When $y_t^{(1)}$ moves, all the other yields move together, longer ones move less than shorter ones.
2. This is just like the EH model, with a different set of constants. Since the $E_t y_{t+j}^{(1)}$ are the same as in my expectations model, the new set of constants *are* a risk premium in the term structure – If $f_t^{(n)} = E_t y_{t+n-1}^{(1)}$+ stuff, well, stuff is a risk premium!

- Let's look at the results.

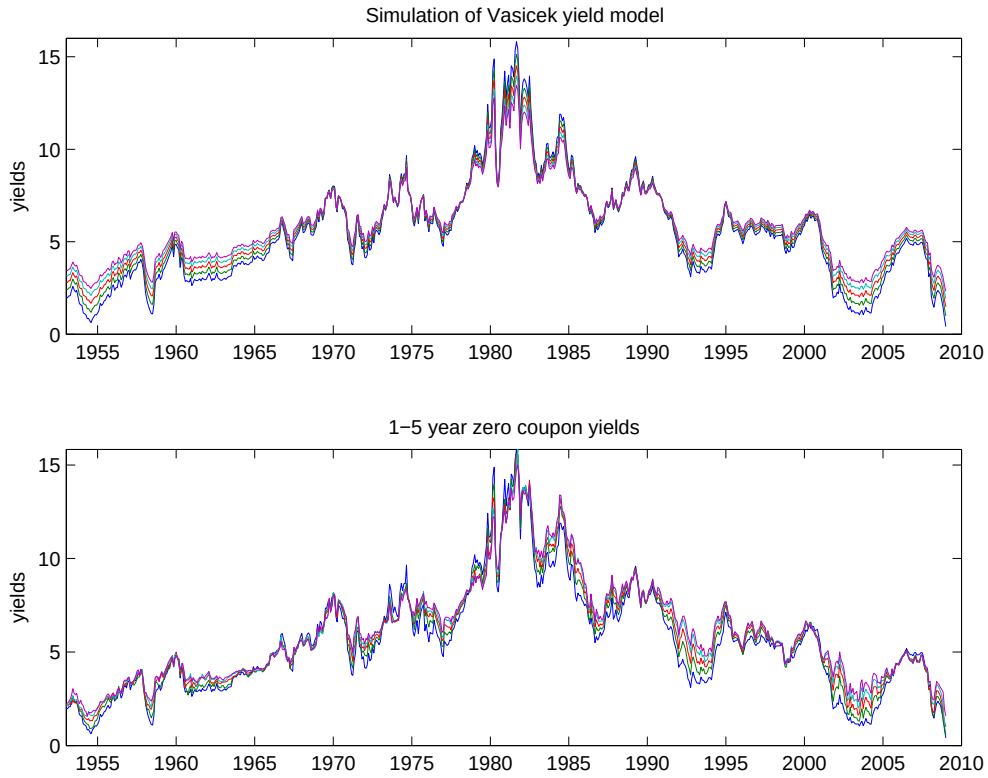
1. I chose some parameters to fit the FB zero coupon bond data[23]
2. I plot $y_t^{(n)}$ for a bunch of assumed values for $y_t^{(1)}$ These are the possible yield curves at any date according to the model. The dashed lines represent the expectations hypothesis, $E_t(y_{t+n}^{(1)})$, calculated using the AR(1) model, and therefore also the forward rates that were produced by my simple expectations model.

[23] I ran a regression of $y_{t+1}^{(1)}$ on $y_t^{(1)}$ to get ρ; I took the variance of errors from that regression to get σ_ε; I took the mean $\delta = E\left(y_t^{(1)}\right)$. Finally, I picked the market price of risk λ to fit the average 5 year forward spread:

$$\begin{aligned}
f_t^{(5)} &= \delta + \rho^4\left(y_t^{(1)} - \delta\right) - \left[\frac{1}{2}\left(1+\rho+\rho^2+\rho^3\right)^2 + \lambda(1+\rho+\rho^2+\rho^3)\right]\sigma_\varepsilon^2 \\
\lambda &= -\frac{E\left(f_t^{(5)}\right) - \delta}{(1+\rho+\rho^2+\rho^3)\sigma_\varepsilon^2} - \frac{1}{2}\left(1+\rho+\rho^2+\rho^3\right)
\end{aligned}$$

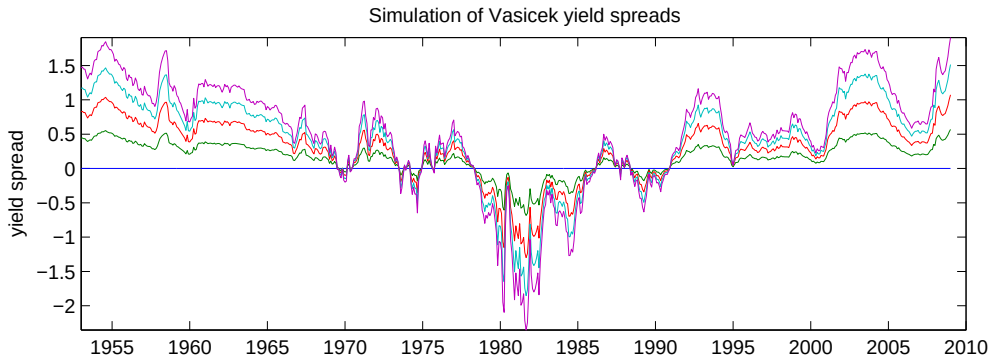


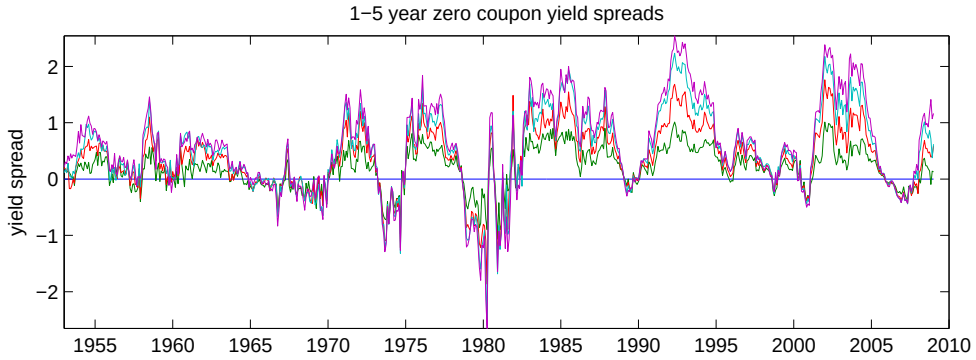
3. Cool! This captures some basic patterns; yields are upward sloping when lower, downward sloping when higher. The substantial risk premium I estimated to match the average upward slope does introduce a substantial deviation of the model from expectations at the long end.
4. I use the history of $y_t^{(1)}$, and find the model-implied $y_t^{(n)}$



It looks pretty good! Again, the spread is higher when yield is lower.

5. But you can see it's not perfect. In the model, the spread reflects how far yields are from the *overall* mean of about 6%. In the data, the spread reflects how far yields are from some sort of "nearby" mean. For example, the spread is off in 1973, 1981, and 1994. To see this more clearly,





We really need a two, or three-factor model.

- A two-factor Vasicek model: Make one factor move quickly (slope, business cycle), and make the other one move slowly (level, inflation)

$$\begin{aligned}
(x_{t+1} - \mu_x) &= \rho_x (x_t - \mu_x) + \varepsilon^x_{t+1} \\
(z_{t+1} - \mu_z) &= \rho_z (z_t - \mu_z) + \varepsilon^z_{t+1} \\
\ln m_{t+1} &= -\frac{1}{2}\lambda_x^2 \sigma^2_{\varepsilon^x} - \frac{1}{2}\lambda_z^2 \sigma^2_{\varepsilon^z} - (\delta_x x_t + \delta_z z_t) - \lambda_x \varepsilon^x_{t+1} - \lambda_z \varepsilon^z_{t+1}
\end{aligned}$$

$$y_t^{(1)} = -\log E_t (m_{t+1}) = (\delta_x x_t + \delta_z z_t)$$

...here we go! This will give "level" and "slope" factors. (Or just interpret "x" above as a vector and "ρ" as a matrix.)

- *What about returns*? We can find the model for returns directly from the model for yields and prices

$$\begin{aligned}
r^{(2)}_{t+1} &= p^{(1)}_{t+1} - p^{(2)}_t \\
r^{(3)}_{t+1} &= p^{(2)}_{t+1} - p^{(3)}_t
\end{aligned}$$

After some algebra[24], we can write the answer as

$$\begin{aligned} E_t\left(rx_{t+1}^{(2)}\right) &= -\left(\frac{1}{2}+\lambda\right)\sigma_\varepsilon^2 \\ E_t\left(rx_{t+1}^{(3)}\right) &= -\left[\frac{1}{2}(1+\rho)^2+\lambda(1+\rho)\right]\sigma_\varepsilon^2 \\ E_t\left(rx_{t+1}^{(4)}\right) &= -\left[\frac{1}{2}\left(1+\rho+\rho^2\right)^2+\lambda(1+\rho+\rho^2)\right]\sigma_\varepsilon^2 \end{aligned} \tag{31}$$

$$\begin{aligned} rx_{t+1}^{(2)} &= E_t\left(rx_{t+1}^{(2)}\right)-\varepsilon_{t+1} \\ rx_{t+1}^{(3)} &= E_t\left(rx_{t+1}^{(2)}\right)-(1+\rho)\varepsilon_{t+1} \\ rx_{t+1}^{(4)} &= E_t\left(rx_{t+1}^{(4)}\right)-(1+\rho+\rho^2)\varepsilon_{t+1} \end{aligned} \tag{32}$$

What does this mean?

1. *Expected returns are constant over time. There is a constant risk premium which increases for longer-term bonds. See the risk premiums in yields, forwards and returns.*
2. *The sign and size of the risk premium* depend on λ. $\lambda > 0$ means long term bonds earn less than short term bonds, as above, and vice versa. The rather complex σ^2 terms are just numbers – they don't vary over time. That's the point. They are the same as the σ^2 terms in forward rates, which are also controlled by λ.
3. *Returns follow a one-factor model.* (Equations (32).) When there is a positive shock to interest rates $\varepsilon_{t+1} > 0$, returns on all bonds suffer. Long term bond returns suffer more, by the $(1+\rho+..)$ factor.

- Comments: risk premiums, returns, risk neutral probabilities.

[24]

$$\begin{aligned} p_t^{(1)} &= -\delta-(1)\left(y_t^{(1)}-\delta\right) \\ p_t^{(2)} &= -2\delta-(1+\rho)\left(y_t^{(1)}-\delta\right)+\left(\frac{1}{2}+\lambda\right)\sigma_\varepsilon^2 \\ p_t^{(3)} &= -3\delta-\left(1+\rho+\rho^2\right)\left(y_t^{(1)}-\delta\right)+\left\{\frac{1}{2}\left(1+(1+\rho)^2\right)+\lambda(1+(1+\rho))\right\}\sigma_\varepsilon^2 \end{aligned}$$

$$\begin{aligned} r_{t+1}^{(2)} &= p_{t+1}^{(1)}-p_t^{(2)} \\ r_{t+1}^{(2)} &= -\delta-\left(y_{t+1}^{(1)}-\delta\right)-\left[-2\delta-(1+\rho)\left(y_t^{(1)}-\delta\right)+\left(\frac{1}{2}+\lambda\right)\sigma_\varepsilon^2\right] \\ r_{t+1}^{(2)} &= \delta-\left(y_{t+1}^{(1)}-\delta\right)+(1+\rho)\left(y_t^{(1)}-\delta\right)-\left(\frac{1}{2}+\lambda\right)\sigma_\varepsilon^2 \\ E_t r_{t+1}^{(2)} &= \delta-\rho\left(y_t^{(1)}-\delta\right)+(1+\rho)\left(y_t^{(1)}-\delta\right)-\left(\frac{1}{2}+\lambda\right)\sigma_\varepsilon^2 \\ E_t r_{t+1}^{(2)} &= \delta+y_t^{(1)}-\delta-\left(\frac{1}{2}+\lambda\right)\sigma_\varepsilon^2 \\ E_t rx_{t+1}^{(2)} &= E_t r_{t+1}^{(2)}-y_t^{(1)}=-\left(\frac{1}{2}+\lambda\right)\sigma_\varepsilon^2 \end{aligned}$$

and so forth. Isn't this fun?

1. *Expected returns and betas are still here*[25]!

$$E_t\left(rx_{t+1}^{(n)}\right) + \frac{1}{2}\sigma^2(rx_{t+1}^{(n)}) = -cov_t(rx_{t+1}^{(n)}, \varepsilon_{t+1})\lambda = -cov_t(rx_{t+1}^{(n)}, y_{t+1}^{(1)})\lambda$$

   Expected return is earned as a compensation for exposure to interest rate shocks. When interest rates rise, bond returns go down. The more a return is negative when interest rates rise, the greater the bond must pay in expected return, and the larger λ, the greater this effect. λ is the "*market price of interest rate risk*". (The annoying $1/2\sigma^2$ term comes from levels vs. logs and is usually quite small.)

2. We can write the risk premium in forward rates similarly

$$\left(f_t^{(n)} - \delta\right) + \frac{1}{2}\sigma^2(rx_{t+1}^{(n)}) = E_t\left(y_{t+n-1}^{(1)} - \delta\right) - cov_t(rx_{t+1}^{(n)}, y_{t+1}^{(1)})\lambda$$

   If expected returns are higher on long-term bonds, $\lambda < 0$, then forward rates are higher than expected future spot rates (plus the usual annoying $1/2\sigma^2$ term.)

3. What does λ mean? How much risk premium should we expect? What sign should it be? What economic forces drive the risk premium? This model helps us to organize our discussion about safety of long vs. short bonds. We started with

$$\ln m_{t+1} = -\gamma\Delta c_{t+1} = -x_t - \frac{1}{2}\lambda^2\sigma_\varepsilon^2 - \lambda\varepsilon_{t+1} = ... - \lambda\left[y_{t+1}^{(1)} - E_t y_{t+1}^{(1)}\right]$$

   λ tells you whether people are happier or unhappier when interest rates rise. If $\lambda > 0$, that means people are happier (consumption rises) with higher interest rates. $\lambda < 0$ means that people are unhappy (consumption declines) with higher interest rates.

4. Which is it? .
   (a) My first guess: a lower interest rate $y_{t+1}^{(1)}$means higher m (bad state), which means $\lambda > 0$ sign and negative premium. This is a typical result – models like this tend to predict downward sloping yield curves, since "long term bonds are safer for long term investors." Real yield curves (TIPS) do show this pattern, and yield curves did slope downward under the Gold standard.
   (b) Nominal yield curve data so far show the opposite pattern, with typically upward-sloping curves. $\lambda < 0$ might make sense if we view high nominal rates as a sign of inflation – if the "level" factor corresponds to expected inflation, since inflation occurs in "bad times."

[25] Why? Just plug in from (32),

$$\begin{aligned}
\sigma^2\left(rx_{t+1}^{(2)}\right) &= \sigma_\varepsilon^2 \\
\sigma^2\left(rx_{t+1}^{(3)}\right) &= (1+\rho)^2\sigma_\varepsilon^2 \\
\sigma^2\left(rx_{t+1}^{(4)}\right) &= (1+\rho+\rho^2)^2\sigma_\varepsilon^2
\end{aligned}$$

$$\begin{aligned}
cov\left(rx_{t+1}^{(2)}, \varepsilon_{t+1}\right) &= -1\sigma_\varepsilon^2 \\
cov\left(rx_{t+1}^{(3)}, \varepsilon_{t+1}\right) &= -(1+\rho)\sigma_\varepsilon^2 \\
cov\left(rx_{t+1}^{(4)}, \varepsilon_{t+1}\right) &= -(1+\rho+\rho^2)\sigma_\varepsilon^2.
\end{aligned}$$

Then look at the two terms on the right of (31).

(c) News reports seem to think that lowering rates is good for the economy. However, there is supply and demand. Higher rates are a sign of a humming economy (demand); lower rates might help an otherwise sluggish economy (supply). News reports always forget there is both supply and demand.

(d) λ is a free parameter, so you can set any risk premium you want. Practitioner uses of these models never even stop a moment to think if the parameters make economic sense.

5. Part of the risk premium stays even with $\lambda = 0$, risk neutrality This is "the effect of convexity," the fact that "Expectations hypothesis is not the same as risk neutrality." It stems from $E(R) = E(e^r) = e^{E(r)+\frac{1}{2}\sigma^2}$. You care about actual returns R, not log returns r. This is another force for typical downward slope of many term structure models. However, with $\lambda = 0$ the $1/2\sigma^2$ terms are typically very small, so there is in practice little difference between expectations and risk neutrality. (At most bond returns have 10% annual volatility, so $1/2\sigma^2 = 0.10^2/2 = 0.005 = 0.5\%$)

6. The risk premium is constant over time – not like Fama-Bliss or Cochrane Piazzesi. To get a time-varying risk premium, we have to add time-varying market price of risk, λ_t. Then you would have $E_t\left(rx_{t+1}^{(n)}\right)$ varying over time with λ_t. Cochrane and Piazzesi "Decomposing the yield curve" shows you how to do this.

7. *"Risk neutral probabilities"* You will hear about these in fixed income and options classes. Here's the idea. Look at the one-factor model we wound up with,

$$\left(y_{t+1}^{(1)} - \delta\right) = \rho\left(y_t^{(1)} - \delta\right) + \varepsilon_{t+1}.$$

$$f_t^{(2)} = \delta + \rho\left(y_t^{(1)} - \delta\right) - \left(\frac{1}{2} + \lambda\right)\sigma_\varepsilon^2.$$

Now, suppose you write a new model with a different mean

$$\left(y_{t+1}^{(1)} - \delta^*\right) = \rho\left(y_t^{(1)} - \delta^*\right) + \varepsilon_{t+1}.$$

$$\delta^* = \delta - \frac{\sigma_\varepsilon^2}{1-\rho}\lambda$$

but you set $\lambda = 0$. You'd get *exactly the same forward rate, prices, yields.*

$$\left(f_t^{(2)} - \delta^*\right) = \rho\left(y_t^{(1)} - \delta^*\right) - (\frac{1}{2} + 0)\sigma_\varepsilon^2$$

$$\begin{aligned}
f_t^{(2)} - \left(\delta - \frac{\sigma_\varepsilon^2}{1-\rho}\lambda\right) &= \rho\left(y_t^{(1)} - \left(\delta - \frac{\sigma_\varepsilon^2}{1-\rho}\lambda\right)\right) - \left(\frac{1}{2}\right)\sigma_\varepsilon^2 \\
f_t^{(2)} - \delta &= \rho\left(y_t^{(1)} - \delta\right) - \frac{\sigma_\varepsilon^2}{1-\rho}\lambda + \rho\frac{\sigma_\varepsilon^2}{1-\rho}\lambda - \left(\frac{1}{2}\right)\sigma_\varepsilon^2 \\
f_t^{(2)} - \delta &= \rho\left(y_t^{(1)} - \delta\right) - \left(\frac{1}{2} + \lambda\right)\sigma_\varepsilon^2
\end{aligned}$$

Thus we get the same answer if we "transform to risk-neutral probabilities which alter the drift, and then use risk-neutral valuation formula" We can write the whole model

this way.

$$\begin{aligned}
\left(y_{t+1}^{(1)} - \delta^*\right) &= \rho\left(y_t^{(1)} - \delta^*\right) + \varepsilon_{t+1}. \\
\left(f_t^{(2)} - \delta^*\right) &= \rho\left(y_t^{(1)} - \delta^*\right) - (\frac{1}{2})\sigma_\varepsilon^2 \\
\left(f_t^{(3)} - \delta^*\right) &= \rho^2\left(y_t^{(1)} - \delta^*\right) - \left[\frac{1}{2}(1+\rho)^2\right]\sigma_\varepsilon^2 \\
\left(f_t^{(4)} - \delta^*\right) &= \rho^3\left(y_t^{(1)} - \delta^*\right) - \left[\frac{1}{2}\left(1+\rho+\rho^2\right)^2\right]\sigma_\varepsilon^2
\end{aligned}$$

The advantage of this approach is that we get prettier formulas that hide λ. It's also algebraically easier once you find λ. The disadvantage: we can't directly describe the actual dynamics. We can't fit the first equation by running a regression of $y_{t+1}^{(1)}$ on $y_t^{(1)}$. This approach is common in interpolation or option pricing exercises in which you don't particularly care about the dynamics, or in which you brazenly fit different (inconsistent) dynamics at different dates in order to fit the cross section better.

- Additional comments on model construction .

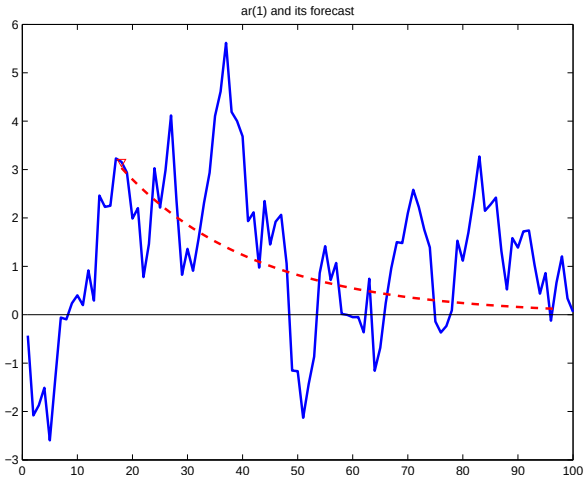
1. Why do we need a more complex model? Why can't we just use an AR(1) for m? If we need more, why not an ARMA model? It turns out this time series model is an extremely useful model for many uses in finance and a natural for when an AR(1) isn't enough. If you have an AR(1),

$$x_{t+1} = \rho x_t + \varepsilon_{t+1}$$

then you know how to find its forecasts,

$$E_t x_{t+j} = \rho^j x_t$$

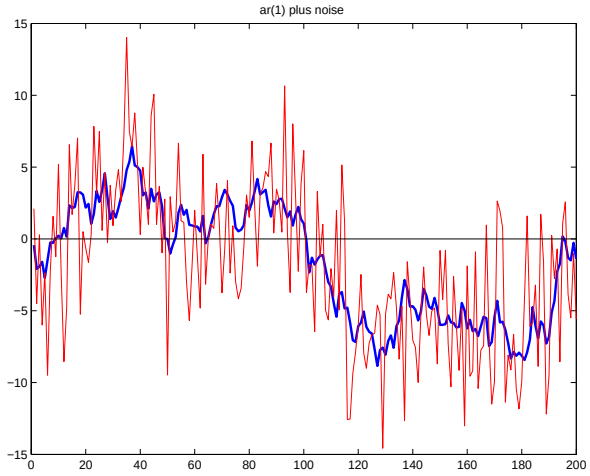
Here's what an AR(1) and its forecast look like



You can think of our model as an AR(1) plus some noise,

$$\begin{aligned}
x_{t+1} &= \rho x_t + \varepsilon_{t+1} \\
y_{t+1} &= x_t + v_{t+1}
\end{aligned}$$

Here's what that looks like



This captures a lot of things we do. x captures an expected value that moves slowly over time, and then y is the outcome. For example, we thought that expected returns move slowly over time in a way captured by d/p,

$$\begin{aligned} r_{t+1} &= (d/p)_t + v_{t+1} \\ (d/p)_{t+1} &= \rho(d/p)_t + \varepsilon_{t+1} \end{aligned}$$

Here's the key: we can still easily compute expected values in this model.

$$\begin{aligned} x_{t+1} &= \rho x_t + \varepsilon_{t+1} \\ y_{t+1} &= x_t + v_{t+1} \end{aligned}$$

$$\begin{aligned} E_t(y_{t+1}) &= x_t \\ E_t(y_{t+2}) &= \rho x_t \\ E_t(y_{t+3}) &= \rho^2 x_t \end{aligned}$$

and so on. As in the expectations model, we're going to need to calculate these. Note that we don't have $E_t y_{t+j} = \rho^j y_t$. The y series is very jumpy. We could not model a series that is both very jumpy and has a slow-moving mean with an AR(1). This model turns out to be equivalent to an ARMA(1,1) with roots that nearly cancel, but this representation of it (in terms of x and y) is much easier to deal with .

2. Why did we need logs? If we use an AR(1) or other time series model for m_t, then we see $m_t < 0$ sometimes, and we know $(c_{t+1}/c_t)^{-\gamma} > 0$. Equivalently, the point is to keep out arbitrage opportunities "arbitrage-free term structure model." There is a deep theorem: *A set of prices p and payoffs x don't have arbitrage opportunities if and only if they can be represented by some positive $m_{t+1} > 0$* and $p_t = E_t(m_{t+1}x_{t+1})$. Thus, we model *log* m, this keeps $m > 0$. It also anticipates that we want answers in terms of *log* prices, yields, etc.

3. I could have written the model without the mysterious xs as

$$\begin{aligned} \left(y_{t+1}^{(1)} - \delta\right) &= \rho\left(y_t^{(1)} - \delta\right) + \varepsilon_{t+1} \\ \ln m_{t+1} &= -\frac{1}{2}\lambda^2\sigma_\varepsilon^2 - y_t^{(1)} - \lambda\varepsilon_{t+1} \end{aligned}$$

"a short rate process plus a market price of risk." Then, I would have checked that the $y_t^{(1)}$ the model *produces* is the same $y_t^{(1)}$ I *started* with, i.e.

$$y_t^{(1)} = -\ln E_t\left(m_{t+1}\right) = -\ln E_t\left(e^{-\frac{1}{2}\lambda^2\sigma_\varepsilon^2 - y_t^{(1)} - \lambda\varepsilon_{t+1}}\right) = y_t^{(1)}$$

Take your pick – is this more or less confusing than starting with an x_t you "can't see" and then showing that it turns out to be $y_t^{(1)}$?

28.1.2 A tour of term structure models.

1. What do we do next? Just more complex time-series model for m_{t+1} in order to fit the data better! *The algebra gets worse, but you don't do anything at all different conceptually.*

2. *Multi-factor models*, with "level shifts" that move the overall mean (inflation?), plus "slope shifts" when yield is temporarily above or below that overall mean. Something like

$$X_t = \begin{bmatrix} level_t \\ slope_t \\ curve_t \end{bmatrix}$$

$$X_{t+1} = \mu + \phi X_t + \varepsilon_{t+1};\ \varepsilon_{t+1} \sim \mathcal{N}(0, V).$$

The discount factor is related to these state variables by

$$m_{t+1} = \exp\left(-\delta' X_t - \frac{1}{2}\lambda' V \lambda - \lambda'\varepsilon_{t+1}\right)$$

$$P_t^{(n)} = E_t\left(m_{t+1} P_{t+1}^{(n-1)}\right)$$

This is *exactly* what we just did, but with vectors and matrices.

3. *Time-varying risk premia.* How do we match Fama-Bliss, CP evidence for time-varying risk premiums? Well, we saw

$$E_t rx_{t+1}^{(n)} + \frac{1}{2}\sigma^2(rx_{t+1}^{(n)}) = -cov(rx_{t+1}^{(n)}, y_{t+1}^{(1)})\lambda$$

If you want $E_t rx_{t+1}^{(n)}$ to vary over time, you need λ_t to vary over time. An example (this is a scalar version of the CP "decomposing the yield curve.")

$$y_{t+1}^{(1)} = \mu + \phi y_t^{(1)} + \varepsilon_{t+1};\ \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma_\varepsilon^2). \tag{33}$$

$$\begin{aligned} \ln m_{t+1} &= -y_t^{(1)} - \frac{1}{2}\lambda_t^2\sigma_\varepsilon^2 - \lambda_t\varepsilon_{t+1} \\ \lambda_t &= \lambda_0 + \lambda_1 y_t^{(1)}. \end{aligned} \tag{34}$$

If $y_t^{(1)}$ rises, then λ_t rises, and then ε_{t+1} is multiplied by a bigger number. We find bond prices as usual:

$$\begin{aligned} P_t^{(1)} &= E_t(m_{t+1}) \\ P_t^{(n)} &= E_t\left(m_{t+1} P_{t+1}^{(n-1)}\right) \end{aligned}$$

4. *Time-varying volatility* matches the fact that interest rates are sometimes more, and sometimes less volatile. It lets you capture changing option prices when volatility rises and falls

$$\begin{aligned} x_t &= \rho x_{t-1} + \sqrt{x_t}\varepsilon_t \\ \ln m_{t+1} &= -\frac{1}{2}\sigma_\varepsilon^2 - x_t - \lambda\sqrt{x_t}\varepsilon_{t+1} \end{aligned}$$

   Now, when x varies, interest rates also become more or less volatile. ("Discrete time Cox Ingersoll Ross")

5. *Continuous time.* It's usually easier to write all this stuff in continuous time.

$$\begin{aligned} dx &= \phi(\mu - x)dt + \sigma dz \\ \frac{dM}{M} &= -xdt - \sigma_M dz \\ P_t^{(n)} &= E_t\left(\frac{M_{t+n}}{M_t}\right) \end{aligned}$$

   But it's the same idea.

6. Don't lose the big picture: Again, it's just a time series model for m then take expectations.

   (a) Art: what are good models that fit the data well and are used in practice? Algebra: cranking out these $E_t's$ (it's easier in continuous time)

   (b) ....See Veronesi/Heaton!

28.1.3 How would you use these models?

1. Price excluded maturities, "interpolate the yield curve." Fit the model (δ, σ^2, ρ) to data on $y^{(1)}, y^{(3)}, y^{(4)}$, figure out what $y_t^{(2)}$ *should* be. Is $y^{(2)}$ underpriced? Hence, price coupon bonds.

2. Price an option.

   (a) Example: A cap. Pays a stream of "dividends" $Y_t^{(1)} - 1$, except it pays only $\bar{Y} - 1$ if $Y_t^{(1)} > \bar{Y}$.

$$P_t = E_t\left[m_{t+1}\max\left(Y_t^{(1)} - 1, \bar{Y} - 1\right) + m_{t+1}m_{t+2}\max\left(Y_{t+1}^{(1)} - 1, \bar{Y} - 1\right)\right] + ...$$

   (Interest rates are lognormal, so each one is a mini Black-Scholes call option.)

   (b) Example 2: What is the option to refinance a 30 year fixed mortgage worth? How much lower would the rate be if you didn't have the option to refinance? You often can't find pretty formulas for this sort of stuff, but it's easy enough to program a computer.

3. Risk management, hedge movements in yields. My name is China. I hold 800 billion long U.S. Treasuries. How many interest rate swaps do I need to avoid interest rate risk? Multifactor models are better than simple duration or convexity hedges.

28.2 Factor model review.

1. Bond yields, changes in yields, returns, etc. are often summarized by *factor models*, in which movements in all n bonds are summarized by movements in a few factors,

$$y_t^{(n)} = b_1^{(n)} \times \text{level}_t + b_2^{(n)} \times \text{slope}_t + b_3^{(n)} \times \text{curve}_t \ \ (+\text{error}_t)$$

2. Level, slope and curve are themselves special linear combinations of yields.

3. The eigenvalue decomposition of the yield covariance matrix gives you a very easy way to fit a factor model. If $y = \left[\begin{array}{ccc} y_t^{(1)} & y_t^{(2)} & ... \end{array} \right]'$,

$$\Sigma = cov(y, y'); \ \Sigma = Q\Lambda Q'$$

   Then

$$x_t = Q'y_t$$

   lets you construct factors, and

$$y_t = Qx_t$$

   is the factor model! Directions: 1) Sigma = cov(yields) 2) [Q,L]=eig(Sigma) 3) Plot Q 4) interpret.

4. This is just like FF3F regressions for stocks, $Q'y$ produces the orthogonalized factors (hml, smb) for you.

5. "Term structure models" CIR, etc. are just a specification of a time series process for a discount factor, like

$$m_t = \rho m_{t-1} + \varepsilon_t$$

   and then

$$P_t^{(1)} = E_t(m_{t+1}); P_t^{(2)} = E_t(m_{t+1}P_{t+1}^{(1)})...$$

   etc. (This is the same thing as a "short rate process" plus a "market price of risk.") Since the time series process for the discount factor runs off a few "factors" x_t, they *produce* factor models for yields.

6. The "single-factor Vasicek model" is

$$\left(y_{t+1}^{(1)} - \delta\right) = \rho\left(y_t^{(1)} - \delta\right) + \varepsilon_{t+1}.$$

$$\begin{aligned} \left(f_t^{(2)} - \delta\right) &= \rho\left(y_t^{(1)} - \delta\right) - \left[\frac{1}{2} + \lambda\right]\sigma_\varepsilon^2 \\ \left(f_t^{(3)} - \delta\right) &= \rho^2\left(y_t^{(1)} - \delta\right) - \left[\frac{1}{2}(1+\rho)^2 + \lambda(1+\rho)\right]\sigma_\varepsilon^2 \\ &.... \end{aligned}$$

   (a) $\left(y_t^{(1)} - \delta\right)$ is the "single factor," $\left[\begin{array}{cccc} 1 & \rho & \rho^2 & ... \end{array} \right]$ are the "loadings" which measure how much each forward moves when the factor moves.

   (b) $\rho^2\left(y_t^{(1)} - \delta\right) = E_t y_{t+n-1}^{(1)}$ is the expectations hypothesis component, and the rest is the risk premium.

(c) This model is formed from the assumption

$$\begin{aligned} x_{t+1} - \delta &= \rho\left(x_t - \delta\right) + \varepsilon_{t+1} \\ \ln m_{t+1} &= -\frac{1}{2}\lambda^2\sigma_\varepsilon^2 - x_t - \lambda\varepsilon_{t+1}. \\ P_t^{(n)} &= E_t\left[m_{t+1}P_{t+1}^{(n-1)}\right] \end{aligned}$$

7. Returns and yields in the single-factor Vasicek model also follow "one-factor" models

(a) Yields

$$\begin{aligned} y_t^{(2)} - \delta &= \frac{(1+\rho)}{2}\left(y_t^{(1)} - \delta\right) - \frac{1}{2}\left(\frac{1}{2} + \lambda\right)\sigma_\varepsilon^2 \\ y_t^{(3)} - \delta &= \frac{(1+\rho+\rho^2)}{3}\left(y_t^{(1)} - \delta\right) - \frac{1}{3}\left\{\frac{1}{2}\left[1 + (1+\rho)^2\right] + \lambda\left[1 + (1+\rho)\right]\right\}\sigma_\varepsilon^2 \end{aligned}$$

(b) Returns

$$\begin{aligned} E_t\left(rx_{t+1}^{(2)}\right) &= -\left(\frac{1}{2} + \lambda\right)\sigma_\varepsilon^2 \\ E_t\left(rx_{t+1}^{(3)}\right) &= -\left[\frac{1}{2}(1+\rho)^2 + \lambda(1+\rho)\right]\sigma_\varepsilon^2 \\ E_t\left(rx_{t+1}^{(3)}\right) &= -\left[\frac{1}{2}\left(1+\rho+\rho^2\right)^2 + \lambda(1+\rho+\rho^2)\right]\sigma_\varepsilon^2 \end{aligned}$$

$$\begin{aligned} rx_{t+1}^{(2)} &= E_t\left(rx_{t+1}^{(2)}\right) - \varepsilon_{t+1} \\ rx_{t+1}^{(3)} &= E_t\left(rx_{t+1}^{(2)}\right) - (1+\rho)\varepsilon_{t+1} \\ rx_{t+1}^{(4)} &= E_t\left(rx_{t+1}^{(4)}\right) - (1+\rho+\rho^2)\varepsilon_{t+1} \end{aligned}$$

28.3 Appendix to factor model notes

(This is highly optional and pretty heavy algebra. It's just a reference for the curious).

28.3.1 Eigenvalue-based factor/principal components models

(This is a background reference only, for people who know the math and want to dig deeper. You have to know what an eigenvalue is)

Summary of the procedure

Given an $N \times 1$ vector of random variables y, with $cov(y, y') = \Sigma$, we form $Q\Lambda Q' = cov(y)$ by the eigenvalue decomposition. If Y is a $T \times N$ matrix of data on y, then [Q,L] = eig(cov(y)) in matlab. Λ is diagonal and Q is orthonormal, $QQ' = Q'Q = I$.

We form "factors" by $x = Q'y$.The columns of Q thus express how to construct factors x from the data on y.

We can then write $y = Qx$, i.e. $y_t = q_1 x_t^{(1)} + q_2 x_t^{(2)} + ...$where $q_1 = Q(:, 1)$ denotes the first column of Q. The columns of y thus also give "loadings" that describe how each y moves if one of the factors x moves.

We have $cov(x, x') = Q'Q\Lambda Q'Q = \Lambda$, i.e. the x are uncorrelated with each other.

If some of the diagonals Λ are zero, then we express all movements in y by reference to only a few underlying factors. For example, if only the first Λ is nonzero, then we can express $x = q_1 z_1$ In practice, we often find that many of the diagonals of Λ are very small, so setting them to zero and fitting y with only a few factors leads to an excellent approximation.

Since the factors are uncorrelated, if we ignore some factors, the loadings on the remaining ones are the same as if we ran regressions,

$$\begin{aligned} y_t &= q_1 x_t^{(1)} + q_2 x_t^{(2)} + [q_3 x_t^{(3)}] \\ y_t &= q_1 x_t^{(1)} + q_2 x_t^{(2)} + \varepsilon_t \end{aligned}$$

Derivation

Factors constructed in this way solve in turn the question "what linear combinations of y has maximum variance, subject to the constraint that the sum of squared weights is one and each linear combination is orthogonal to the previous ones?" In equations, each column q_i of Q satisfies

$$\max \left[var \left(q_i' y \right) \right] \text{ s.t. } q_i' q_i = 1,\ q_i' q_j = 0,\ j < 1$$

Why is this an interesting question? "What linear combination $q'y_t$ of the y_t has the highest variance?" would be too easy a question. Just make q big. The constraint $q'q = 1$ means the sum of squared q must be equal to one, so you can't boost variance by making q big. You have to find the right *pattern* of q across the elements of y.

Why is this the answer? Let's look at the first maximization – what linear combination of the y has the largest variance, if you constrain the sum of squared weights to be less than one?

$$\max_{\{q\}} var(q'y_t) \text{ s.t. } q'q = 1$$

Forming a Lagrangian,

$$L = q'\Sigma q - \lambda q'q$$

The first order condition $(\partial / \partial q)$ is

$$\Sigma q = \lambda q$$

This is an eigenvalue problem! The answer to this question is then, *choose q as an eigenvector of the matrix Σ*. Now, which one? Let's see what variance of x we get out of all this

$$var(x_t) = var(q'y) = q'\Sigma q = q'\lambda q = \lambda q'q\lambda$$

Aha! *The eigenvalue gives us the the variance of x*. So, the answer to our maximization is, *choose the eigenvector q corresponding to the largest eigvenvalue λ*.

In sum, to find the linear combination of y with largest variance, and sum of squared weights equal one, we choose as weights q the eigenvector corresponding to the largest eigenvalue of the covariance matrix of y. Q is a matrix of eigenvectors, so you're done. Eigenvectors are orthogonal

$q_i' q_j = 0$ so the eigenvectors corresponding to successively smaller eigenvalues answer the question for the remaining factors.

Now, what does this have to do with R^2? Suppose we leave out some factors

$$y_t = q_1 x_t^{(1)} + q_2 x_t^{(2)} + [q_3 x_t^{(3)}]$$

Since $x^{(1)}$ and $x^{(2)}$ have *maximum* variance, this means $x^{(3)}$ and beyond have *minimum* variance. In short *we have found a factor model that for each choice of how many factors to use maximizes the R^2 in these regressions.*

Rotation

There are lots of equivalent ways to write any factor model. If we have

$$y_t = Q x_t$$

then if R is any orthogonal (rotation) matrix, i.e. any matrix with $RR' = R'R = I$, we can define new factors $z_t = R' x_t$. Then $x_t = R z_t$

$$y_t = Q(R z_t) = (QR) z_t$$

Now the columns of QR give us new loadings on the new factors, which are constructed by $z_t = R' x_t = R' Q' y_t = (QR)' y_t$

This works even better if we use unit variance shocks.

$$\begin{aligned} y_t &= \left(Q \Lambda^{\frac{1}{2}}\right)\left(\Lambda^{-\frac{1}{2}} x_t\right) = \left(Q \Lambda^{\frac{1}{2}}\right) z_t \\ cov(z_t, z_t') &= \Lambda^{-\frac{1}{2}} \Lambda \Lambda^{-\frac{1}{2}} = I \end{aligned}$$

Now if we rotate,

$$w_t = R z_t, z_t = R' w_t$$

we have

$$y_t = \left(Q \Lambda^{\frac{1}{2}}\right) z_t = \left(Q \Lambda^{\frac{1}{2}}\right) R' w_t$$

$$cov(w_t, w_t') = cov(R z_t, z_t' R') = RR' = I$$

In sum, *if we rotate unit-variance factors, they are still uncorrelated with each other and still have unit variance.* You're free to recombine factors in any way you want to make them look pretty.

28.3.2 How do you really solve the one-factor Vasicek model?

Guess

$$P_t^{(n)} = A_n - B_n (x_t - \delta)$$

then

$$\begin{aligned}
P_t^{(n)} &= E_t\left(M_{t+1}P_{t+1}^{(n-1)}\right) \\
A_n - B_n\left(x_t - \delta\right) &= \log E_t\left(\exp\left(-\frac{1}{2}\lambda^2\sigma_\varepsilon^2 - x_t - \lambda\varepsilon_{t+1}\right)\exp\left(A_{n-1} - B_{n-1}\left(x_{t+1} - \delta\right)\right)\right) \\
&= \log E_t\left(\exp\left(-\frac{1}{2}\lambda^2\sigma_\varepsilon^2 - x_t - \lambda\varepsilon_{t+1} + A_{n-1} - B_{n-1}\rho\left(x_t - \delta\right) - B_{n-1}\varepsilon_{t+1}\right)\right) \\
&= \log E_t\left(\exp\left(-\delta + A_{n-1} - \left(1 + B_{n-1}\rho\right)\left(x_t - \delta\right) - \frac{1}{2}\lambda^2\sigma_\varepsilon^2 - \lambda\varepsilon_{t+1} - B_{n-1}\varepsilon_{t+1}\right)\right) \\
A_n - B_n\left(x_t - \delta\right) &= -\delta + A_{n-1} - \left(1 + B_{n-1}\rho\right)\left(x_t - \delta\right) + \left(B_{n-1}\lambda + \frac{1}{2}B_{n-1}^2\right)\sigma_\varepsilon^2
\end{aligned}$$

The constant and the term multiplying x_t must separately be equal. Thus,

$$B_n = 1 + B_{n-1}\rho$$

$$A_n = -\delta + A_{n-1} + \left(B_{n-1}\lambda + \frac{1}{2}B_{n-1}^2\right)\sigma_\varepsilon^2$$

That's easy to solve

$$\begin{aligned}
B_0 &= 0 \\
B_1 &= 1 \\
B_2 &= 1 + \rho \\
B_3 &= 1 + \rho\ + \rho^{\,2} \\
B_n &= \sum_{j=0}^{n-1}\rho^{\,j} = \frac{1 - \rho^{\,n}}{1 - \rho}
\end{aligned}$$

$$\begin{aligned}
A_0 &= 0 \\
A_1 &= -\delta \\
A_2 &= -2\delta + \left(\lambda + \frac{1}{2}\right)\sigma_\varepsilon^2 \\
A_3 &= -3\delta + \left[\left(1 + \rho\ \right)\lambda + \frac{1}{2}\left(1 + \rho\ \right)^2\right]\sigma_\varepsilon^2 \\
A_4 &= -4\delta + \left[\left(1 + \rho\ + \rho^{\,2}\right)\lambda + \frac{1}{2}\left(1 + \rho\ + \rho^{\,2}\right)^2\right]\sigma_\varepsilon^2
\end{aligned}$$

Or just

$$A_n = -n\delta + \left[B_{n-1}\lambda + \frac{1}{2}B_{n-1}^2\right]\sigma_\varepsilon^2$$

You see the pattern from here