Dave Backus / NYU

International Asset Returns

October 29, 2004

Presentations:

- Outline should look like this:
  - The question: what question does this paper answer?
  - Setup: what theoretical setup? what data? ie, nuts and bolts.
  - Answer: the proposed answer to the question.
  - Future directions: what new questions does the paper suggest? how would you address them?
- Time limit: 10 minutes (15 max) + 5 for Q&A.
- Slides: prepare slides (ppt, pdf), but be ready to work with either a projector or hard-copy handouts depending on how successful we are with setting up classroom technology.
- Advice: Read Tim Kehoe's tips. His comments are directed to people presenting original papers, but they apply here as well. Save for future reference. Our business is not just about ideas: it's about communicating ideas effectively and sharing them with others. As the cliche goes: If a tree falls in the forest and no one is there, it does not make a sound.
- Scheduling. Let's pick dates asap (now?).

1 Issues

Today we turn to a new topic: asset returns. I'll focus on two issues, but the tools we develop are applicable more generally. The topics are:

- Interest rate differentials. Interest rates on (largely riskfree) instruments differ across countries, either quoted in local currency (eurocurrency rates, for example) or as local "real rates" (nominal rate minus expected inflation). Local currency differentials for major currencies are variable (std dev about 3% annually) and persistent (about 0.9 monthly). (eg, Backus, Foresi, and Telmer, JF, 2001, Table I). I haven't seen work on real interest rates, but would like to.
- Equity returns. Highly correlated across countries — in fact, less highly correlated than output growth. See Dumas, Harvey, and Ruiz (JIMF, 2003, Fig 1).

I'd love to look at interest rates in emerging markets, too, but will save this for another time. See Neumeyer and Perri (NBER 10387) if you're interested.

2 Essentials of asset pricing theory

So far we've looked at general equilibrium models. We'll do some of that shortly, but it's useful to start with much less theoretical structure, which is the norm in finance.

Basics. The fundamental theorem of asset pricing goes something like this:

Proposition. Consider a set of assets with one-period gross returns $\{r_{j,t+1}\}$. If and only if these returns are free of arbitrage opportunities, there exists an $m > 0$ satisfying

$$1 \;=\; E_t\left(m_{t+1} r_{jt+1}\right) \tag{1}$$

for all assets j and dates t.

Equation (1) says, in effect, that the field of asset pricing consists of finding a useful m and exploring its implications. Note that returns have units: say, goods later for goods now, or money later for money now. Our task is to find m's that account for observed asset prices and returns, which leads to the question: what features of m are implied by observed returns?

With complete markets m is unique, with incomplete markets not. You can see this easily if the set of states is finite ($s = 1, ..., S$) and we write out the random variables as $r(s, s')$ and $m(s, s')$. Then (1) says

$$\sum_{s'} \pi(s'|s) m(s, s') r_j(s, s') \;=\; 1$$

for all assets j. The unknowns here are $m(s, s')$ (one for each s', S in total). The equations correspond to the assets (one for each j, say J in total). Complete markets means that $J = S$, in which case there's a unique solution for m. (We also need the returns to be linearly independent.) Incomplete markets means $J < S$, in which case there are lots of solutions for m. Do we need the probs? They're simply a normalization of the state price $q(s, s') = \pi(s'|s) m(s, s')$.

Application to bond pricing. To see how this works in practice, consider a simple (but typical) bond pricing model. First some notation. Let b_t^n be the price at date t of a claim to one "dollar" (or good) at $t + n$. Yields are defined by $b_t^n = \exp(-n y_t^n)$. One-period returns are $r_{t+1}^{n+1} = b_{t+1}^n / b_t^{n+1}$, which leads to the recursive pricing relation,

$$b_t^{n+1} \;=\; E_t\left(m_{t+1} b_{t+1}^n\right)$$

starting with $b_t^0 = 1$ (a dollar today is worth one dollar). What's cool about bonds is that you can price them knowing only m.

A basic bond pricing model consists of these two equations:

$$\begin{aligned} -\log m_{t+1} &= \delta + z_t + \lambda \varepsilon_{t+1} \\ z_{t+1} &= (1 - \varphi)\mu + \varphi z_t + \sigma \varepsilon_{t+1}, \end{aligned}$$

where $\{\varepsilon_t\} \sim \text{NID}(0,1)$. Work through recursions for $\log b_t^n = A_n + B_n z_t$.

Can we add more noise to m? Eg, what if we add some random noise η_{t+1} to the m equation, where $\{\eta_t\} \sim \text{NID}(0,\tau^2)$ and is independent of ε? Typically we can project kernel on returns and go with that. Additional noise has no observational implications. (This is true, but you have to work through the model to see it.)

Volatility bounds on the kernel. Hansen and Jagannathan asked the question: what can we infer about the kernel from observed returns? Their bound is derived like this:

$$0 \;=\; E(mx) \;=\; E(m)E(x) + Cov(m,x)$$

implies

$$std(m)/E(m) \;\geq\; E(x)/std(x).$$

The rhs is called the Sharpe ratio for x and can be estimated from returns. For annual data, the mean excess return on equity (broad-based portfolio like the S&P 500) is about 0.06 and its std deviation is about 0.15, implying a Sharpe ratio of $6/15 = 0.4$. Since $E(m)$ is close to one, this gives us an approximate lower bound on the std dev of m.

In the data, the Sharpe ratio depends on the time interval. If returns were iid, then the Sharpe ratio for a time interval of h is

$$E(x)/std(x) \;=\; \frac{h\bar{x}}{h^{1/2}\sigma_x} \;=\; h^{1/2}\frac{\bar{x}}{\sigma_x}$$

For monthly data, an estimate of the lower bound on std(m) is about $(1/12)^{1/2}0.4 = 0.12$.

Alvarez and Jermann have a neat version of this based on logs. I'll run through it twice, once to give you the idea, the second time to get a more elegant result. Define "volatility" as the difference between the "log of the expectation" and the "expectation of the log":

$$L(x_{t+1}) \;\equiv\; \log Ex_{t+1} - E\log x_{t+1}.$$

To see why L works as a measure of volatility, consider Jensen's inequality, which tells us that $L(x) \geq 0$, with equality only when x is nonstochastic. Why? Because log is a concave function. In the log-normal case, with $\log x \sim N(\mu,\sigma^2)$, $L(x) = \sigma^2/2$. More generally, $L(x)$ is a function of higher-order moments of $\log x$.

First attempt. Note that

$$L(m) \;=\; \log E(m) - E\log m.$$

Jensen's inequality applied to E of (1) implies

$$E\log r \;\leq\; -E\log m.$$

Putting them together gives us

$$L(m) \;\geq\; E\log r + \log E(m).$$

What about $E(m)$? This is just the mean one-period bond price, whose log is almost minus the mean short rate. We'll fix this up shortly.

Second attempt. A conditional version of L is

$$L_t(x_{t+1}) \equiv \log E_t x_{t+1} - E_t \log x_{t+1}.$$

The conditional and unconditional versions are connected by

$$L(x_{t+1}) = E L_t(x_{t+1}) + L(E_t x_{t+1}).$$

This follows from the definitions of L and L_t, but for intuition recall that $Var(x_{t+1}) = E\,Var_t(x_{t+1}) + Var(E_t x_{t+1})$.

Now start over. Jensen's inequality applied to (1) implies:

$$E_t \log r_{t+1} \leq -E_t \log m_{t+1}$$

Same logic as before. Now subtract the one-period return, $r^1_{t+1} = 1/E_t m_{t+1}$:

$$E_t \log(r_{t+1}/r^1_{t+1}) \leq \log E_t m_{t+1} - E_t \log m_{t+1} = L_t(m_{t+1}).$$

Now take unconditional expectations of both sides:

$$E \log(r_{t+1}/r^1_{t+1}) \leq E L_t(m_{t+1}) = L(m_{t+1}) - L(E_t m_{t+1}),$$

or

$$L(m_{t+1}) \geq E \log(r_{t+1}/r^1_{t+1}) + L(b^1_t), \tag{2}$$

where $b^1_t = E_t m_{t+1}$ is the price of a one-period bond. When you apply this bound, the last term is tiny so the volatility of m is bounded by the mean excess return — in logs, which in practice doesn't mean much.

In the data, the mean of $\log r/r_1$ is about 0.06 for the equity premium with annual data and $L(b^1)$ is tiny. Therefore $L(m) \geq 0.06$. In the log-normal case with $std(\log m) = \sigma_m$, $\sigma^2_m \geq 2(0.06) = .12$ so $\sigma_m \geq 0.35$, which is roughly comparable to our earlier estimate.

Representative agent models. Can we make sense of the pricing kernel as the outcome of a general equilibrium model? We start with the representative agent model, which doesn't do very well but gets us started. Lucas suggested an exchange economy with one agent. Markov process over states with conditional probabilities $\pi(s'|s)$. All expectations below based on these probs. Lucas tree: agents choose portfolios, but in equil end up owning the tree and consuming its output. The equilibrium problem starts with portfolio choice, given returns $r'_i = r_i(s, s')$. Traditional additive model has the Bellman equation:

$$J(a, s) = \max_{c,w} u(c) + \beta E J(a', s')$$

subject to

$$a' = (a - c) \sum_i w_i r'_i.$$

and $\sum_i w_i = 1$. First-order and envelope conditions are

$$\begin{aligned} w_i: \quad \lambda &= \beta(a-c)E[J_1(a',s')r_i'] \\ c: \quad u_1(c) &= \beta E[J_1(a',s')r_p'] \\ a: \quad J_1(a,s) &= \beta E[J_1(a',s')r_p'], \end{aligned}$$

where r_p is the optimal/equilibrium return on the whole portfolio. They imply the usual "Euler equation"

$$E\left(\beta[u_1(c')/u_1(c)]r_i'\right) = 1,$$

which defines the pricing kernel as $m' = \beta u_1(c')/u_1(c)$.

With power utility, $u(c) = c^\alpha/\alpha$, $m_{t+1} = \beta g_{t+1}^{\alpha-1}$, and

$$\log m_{t+1} = \log\beta - (1-\alpha)\log g_{t+1}.$$

If $std(\log g)$ is 0.02 (I'm making this up, but it's probably not too bad for annual data), then to match the st dev of $\log m$ we inferred in the data we need $1-\alpha$ equal to 20 or so.

3 Kernels and exchange rates

Take a set of assets. Local returns (r) and foreign returns (r^*) are measured in different units but pertain to the same assets. The returns are connected by the exchange rate: e is the local price of one unit of foreign currency/good/whatever. Apply the fundamental theorem to show:

Proposition. You can choose m, m^* such that $m^*_{t+1} = m_{t+1}d_{t+1}$, where $d_{1+1} = e_{t+1}/e_t$.

Proof. Apply fundamental theorem. Premise: r and r^* tradeable, also currencies. Then there's an m satisfying $E(mr) = 1$. Ditto an m^* satisfying $E(m^*r^*) = 1$. What's the connection? Well, dollar returns on foreign assets are dr^* so we can write: $E(mdr^*) = 1$. Evidently we can choose $m^* = md$.

Consequence: The two kernels must be highly correlated. Suppose (to make this concrete) that the std dev of $\log d$ is 0.03 (a plausible monthly number for, say, the dollar-euro rate). Suppose, too, that the std deviations of $\log m$ and $\log m^*$ are at least 0.15 (plausible HJ lower bounds). What is the correlation of $\log m$ and $\log m^*$? Note:

$$\begin{aligned} Var(\log d) &= 0.03^2 \\ &= Var(\log m^*) + Var(\log m) - \rho[Var(\log m^*)\,Var(\log m)]^{1/2}. \end{aligned}$$

If the variances of the log kernels are at least 0.15^2, then ρ is at least 0.98. What if $Var(\log m^*) = Var(\log m) = 0.10$? 0.20?

An example shows how this might work. Let our pricing model have two components:

$$\begin{aligned} \log m_{t+1} &= \delta + \lambda\varepsilon_{t+1} + \lambda^*\varepsilon^*_{t+1} \\ \log m^*_{t+1} &= \delta + \lambda^*\varepsilon_{t+1} + \lambda\varepsilon^*_{t+1} \end{aligned}$$

with the ε's iid with mean zero and variance one. Suppose $Var(\log m) = Var(\log m^*) = 0.15^2$ and $Var(\log d) = 0.03^2$ [estimates for monthly data]. What are the implied parameter values? What is the correlation between the $\log m$'s? I find: $\lambda^2 = 0.0135$, $\lambda = 0.1161$ (if we take the positive root), and $\lambda^* = 0.0949$. Ie, the loadings on the two shocks have to be pretty similar. With these parameters, the correlation is 0.98.

Questions:

- Do we get any mileage by using the nonuniqueness of m? For example, what if we add noise to m? I think the answer is no: η doesn't help. In bond models, it's not identified. More generally, we don't care about variation in m that's unrelated to returns. In the log-normal setting, we can project the log returns onto the state variables and innovations and leave it at that. More formally, suppose we use $\widehat{m} = m_{t+1}e^{\eta_{t+1}}$ in place of m. Then
$$E_t\left(m_{t+1}e^{\eta_{t+1}}r_{jt+1}\right)$$
for all assets j. We now have a choice: we can add the same noise to m^*, in which case we've raised the correlation between them. Or we could violate the exchange rate relation, using $d_{t+1} = (m^*_{t+1}/\widehat{m}_{t+1})e^{\eta_{t+1}}$, which adds noise to the exchange rate. [??]

- Is the high correlation between pricing kernels an indication of the effectiveness of risk-sharing? My take: absolutely not! The relation between m and m^* is simply a units correction and need not have anything to do with risk-sharing. Example 1: take a one-agent economy with two goods, apples and bananas. Let m be the apple kernel and m^* the banana kernel. By adjusting the supplies of the two goods, we can generate changes in the relative price of bananas to apples and drive m and m^* apart. Does this suggest lack of risk-sharing? No, we have a one-agent economy with (presumably) as good risk-sharing as is possible. Example 2: Take a two-agent version of the same economy where the two agents have identical homothetic preferences. Consumption growth is perfectly correlated. But the apple and banana kernels won't be perfectly correlated. In this case, the consumption growth correlation is higher than the kernel correlation. Example 3: Take a two-country model with nontraded goods. Consumption and consumption growth won't be perfectly correlated, but with complete markets we get as good risk-sharing as is possible. There will be, however, a tight link between the real exchange rate and consumption growth. Evidence against this is suggestive that the complete markets model is inadequate.

4 Exotic preferences

I wanted to give you a quick overview of exotic preferences. Originally this was to be able to work through the Dumas-Harvey-Ruiz paper, which uses the version of Kreps-Porteus preferences developed and popularized by Epstein and Zin (Econometrica 1989, JPE 1991). I've now decided to skip the DHR paper to save time, but I want to give you an overview anyway because NYU has so many people with expertise on the subject: Benhabib, Bisin, Caplin, Eliaz, Leahy, Ok, Sargent, Schneider, and probably others that temporarily slipped my mind. You might think about working out an application with the help of the local team. If you'd like to read more, see my survey with Routledge and Zin: "Exotic preferences for macroeconomists," available on my web site.

The traditional additive utility function,

$$U(c) \;=\; \sum_t \beta^t \sum_{s^t \in S^t} \pi(s^t) u[c(s^t)] \tag{3}$$

underlies most work in economics. It's tight and regular structure is extremely helpful – recall the solution to the risk-sharing problem that serves as the foundation of this course. But it's not the only legitimate preference relation we might use. Here's a quick summary:

- Time. Note that (3) is recursive. If U_t is utility from date t on, then we could express it as

  $$U_t \;=\; u_t + \beta U_{t+1},$$

  a relation that will look familiar to anyone who has done dynamic programming. Koopmans noted that we could retain recursivity with a more general time aggregator V:

  $$U_t \;=\; V(u_t, U_{t+1}).$$

  The properties of V (notably V_2) govern discounting. What's changed is that the discount factor is endogenous: it changes with u and U.

- Uncertainty. There's lots of work on departures from expected utility. Rather than review it, let me simply note that a convenient way to summarize attitudes toward risk is with a certainty equivalent function. In the case of expected utility, we write

  $$\mu(c) \;=\; u^{-1}\left[Eu(c)\right] \tag{4}$$

  For our purposes, μ is simply a convenient way to summarize risk preference.

- Time and uncertainty. The combination of the two is tricky. Epstein and Zin suggested a combination of the two preceding ideas:

  $$U_t \;=\; V[u_t, \mu_t(U_{t+1})].$$

  The idea here is that we use the certainty-equivalent of future utility, the subscript t meaning that this is based on the conditional probabilities at date t (think E_t). Kreps-Porteus utility is a special case with the expected utility certainty-equivalent described above.

- Risk-sensitive control. My interpretation of risk-sensitive control is that it's Kreps-Porteus with a particular functional form for the function u in the certainty-equivalent function (4): $u(x) = -\exp(-\alpha x)$.

- Ambiguity. Some of the most interesting work is based on incomplete knowledge of the probabilities. It's not strictly necessary, but we typically assume in applied work that the probabilities in preferences are the same as those that generate the data (rational expectations). Gilboa and Schmeidler suggested an alternative in which there are multiple sets of probabilities and you choose the one that gives you the lowest expected utility:

$$U(c) \;=\; \min_{\pi \in \Pi} \; E_\pi u(c).$$

There are a couple ways to attack this in dynamic settings: robust control (Hansen, Sargent, others) and recursive multiple priors (Epstein and Schneider). The choice between them involves some subtle issues. More important, I think, is that both allow interesting behavior. I urge you to ask them about it.

5 Equities

I changed my mind — we don't have the time to do this, and I've come to believe Lasse Pedersen that the result (higher correlation of returns than outputs) is obvious. Why? They share the same pricing kernel, which can easily have the effect of changing returns up and down together.