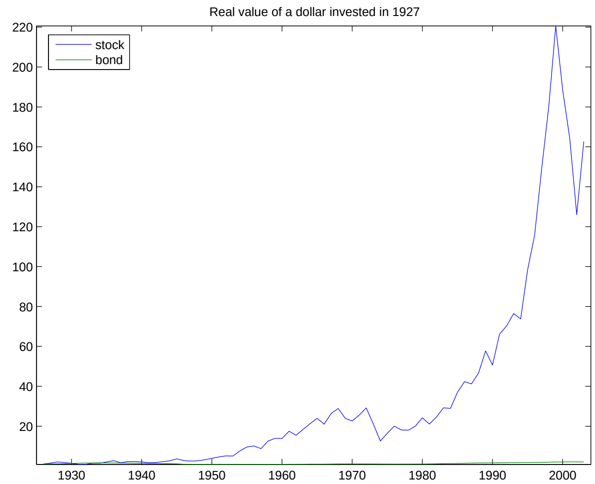
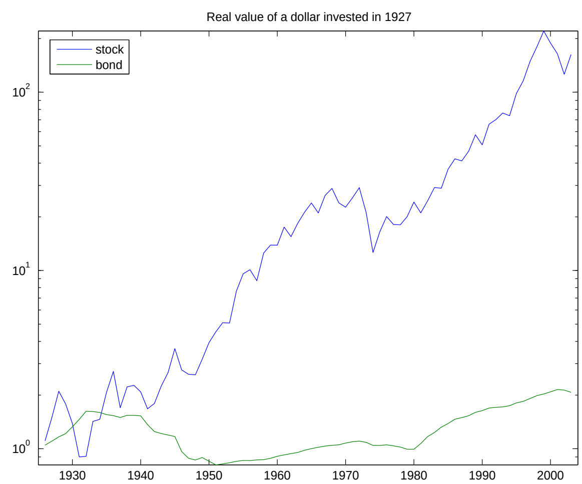
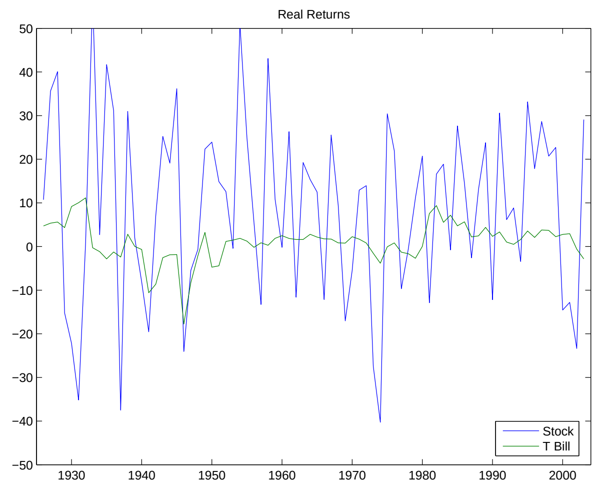
Equity premium review

Stocks have earned a tremendous amount more than bonds. However, there is real risk there – two decates (1929-1945, 1967-1982) where you earn nothing. We have certainly just started another (2000-?)



Risk!





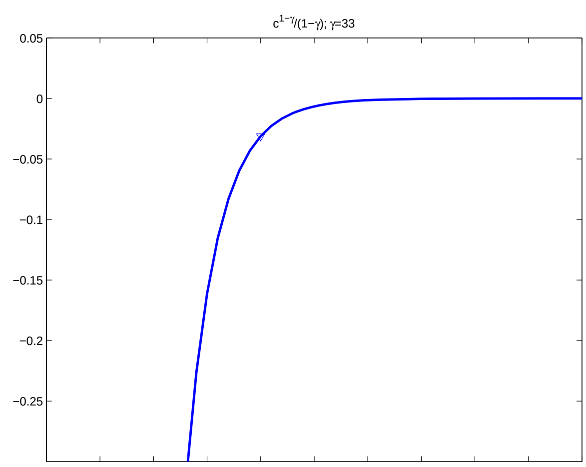
1.**Simple bound** The simplest version of the Hansen-Jagannathan bound is just a back of the envelope calculation of the basic first order condition:

$$
\begin{aligned}
0 &= E(mR^e) = E(m)E(R^e) + cov(mR^e) = E(m)E(R^e) + \rho\sigma(m)\sigma(R^e) \\
\frac{E(R^e)}{\sigma(R^e)} &= \rho\frac{\sigma(m)}{E(m)} \\
\|\rho\| &\leq 1 \\
\frac{\|E(R^e)\|}{\sigma(R^e)} &\leq \frac{\sigma(m)}{E(m)} \approx \gamma\sigma(\Delta c)
\end{aligned}
$$

Numbers in FMRA:

$$
\begin{aligned}
\frac{\|E(R^e)\|}{\sigma(R^e)} &\leq \frac{\sigma(m)}{E(m)} \approx \gamma\sigma(\Delta c) \\
0.5 &= \frac{0.08}{0.16} \leq \gamma\,(0.015) \\
\gamma &\geq 33
\end{aligned}
$$

This seems like a *lot* of risk aversion. This investor cares almost nothing for gains relative to losses.



To make the sharpest bound, you should use the largest Sharpe ratio you can find. More on this below.

More and real numbers:

Annual data 1948-2003, percent

$E(\Delta c)$	$\sigma(\Delta c)$	$E(R^e)$	$\sigma(R^e)$	$E(R^{\text{bond,real}})$	$\text{corr}(\Delta c, R^e)$	$\text{cov}(R^e, \Delta c)$
1.33	1.92	7.70	18.0	1.6	0.41	14.09

2. **Correlation puzzle.** Why ignore the information in ρ? There are good reasons – correlations are harder to measure; a one period delay in measurement will ruin them, and they're much more sensitive to the fact that consumption is time aggregated. Still, we can consider the effect of lower correlation

$$\frac{E(R^e)}{\sigma(R^e)} = \rho \frac{\sigma(m)}{E(m)} \approx \rho\gamma\sigma(\Delta c)$$

The best I could justify is $\rho = 0.4$, so we need

$$\begin{aligned} \frac{0.5}{0.4} &= \gamma \times 0.015 \\ \frac{0.5}{0.4 \times 0.015} &= \gamma = 83 \end{aligned}$$

Note, in a complete market, $\sigma(m)/E(m)$ is the slope of the mean-variance frontier. Thus $\rho < 1$ means that the market return is decidedly inside the MVF.

3. **Risk free rate puzzle.** If we accept huge γ, then $E(\beta\,(c_{t+1}/c_t)^{-\gamma}) = 1/E(m)$ goes nuts. With $m = e^{-\delta - \gamma\Delta c}$ and normality or in continuous time

$$r^f = \delta + \gamma E(\Delta c) - \frac{1}{2}\gamma(\gamma+1)\sigma^2(\Delta c)$$

(You get γ^2 with discrete time lognormal, $\gamma(\gamma+1)$ with Ito's lemma). Start with the first term:

$$\begin{aligned} r^f &= \delta + (33 \textit{ to } 83) \times 0.01 = \\ 1\% &= \delta(\%) + 33 \textit{ to } \delta(\%) + 83\% \end{aligned}$$

So we also need a huge negative (prefer the future) δ. That also seems implausible (though Kocherlakota showed it's possible.)

a) *Sensitivity* More importantly than levels, I think, high risk aversion (low intertemporal substitution) means that interest rates should be extremely sensitive to consumption growth.

$$r_t^f = \delta + \gamma E_t(\Delta c_{t+1})$$

a one percentage point change in expected consumption growth over time or across countries should accompany a 33-83 percentage point change in interest rates! These people are extremely averse to substituting consumption over time, you have to offer them extraordinary bribes to do so. (As the equity premium represents an extraordinary bribe to substitute consumption across states of nature)

b) *Precautionary saving* The second term matters and can help..

$$r^f = \delta + \gamma \times 0.01 - \frac{1}{2}\gamma(\gamma+1)\,(0.015)^2$$

$$\begin{aligned} 0.01 &= 0.01 + \gamma \times 0.01 - \frac{1}{2}\gamma(\gamma+1)\,(0.015)^2 \\ \gamma &= 87.889 \end{aligned}$$

High variance causes people to save more (precautionary). This effect can offset the fact that a high mean causes then to want to consume more now.

This seems to solve the puzzle (except plausibility). However, this calculation is extraordinarily sensitive. Tiny variations in γ, E, σ involve huge variations in the risk free rate. So, I don't take it seriously. However, it does emphasize that precautionary saving can offset intertemporal substitution. This is a very important effect in the Campbell-Cochrane model. In HJ bound drawings, you sometimes see "model" discount factors that curve back and blast through the HJ cone. That's the same effect; as you vary γ you strongly affect the implied risk free rate or $E(m)$.

Recursive or habit (state-nonseparable or time-nonseparabe) utility separates γ risk aversion from intertemporal substitution, which lets us to some extent disentangle equity premium from risk free rate puzzles. I'm not sure the former is that plausible though. Why would people care deeply about the pattern of consumption across state, but not across time? Do people really perceive time and state that differently?

4. **Fixes that don't work**:

a) *Expected return only* Many RBC models proudly produce high expected returns with astronomical volatility. What counts is the *Sharpe ratio*. The Sharpe ratio is also independent of leverage, so stories that rely on a lot of leverage can't do much.

b) *Risk free rates.* Many early explanations (Aiyagari and Gertler were one) pointed out that T bill returns are artificially low, because there is a liquidity premium. Perhaps an excuse to raise the T bill rate rather than lower the stock return will help? Alas, Sharpe ratios are pervasive in financial markets, between different categories of bonds, HML, carry trade, etc., all of which do not involve T bills. The slope of sample mean-variance frontiers is always high.

c) *Individuals* vs. market? Using aggregate consumption and representative agent produces smooth consumption. Surely individual consumption is more volatile, so we can raise $\sigma(\Delta c)$? Alas, the first order condition holds *for each individual* as well (Everybody is "marginal"). Though individual consumption is more volatile, it's hard to think it's that much more volatile. For example, to rescue $\gamma = 2$, we need $\sigma(\Delta c) = 25\%$. That's a lot – remember this is nondurables, services and flow not durables purchases.

Worse, idiosyncratic risks are idiosyncratic. As $\sigma(\Delta c^i)$ rises due to idiosyncratic risk, ρ *must* decline. Writing (see below) $m^i = m^* + \varepsilon^i$ (see below) we see how this is exactly true. Adding ε^i does absolutely nothing to improve the equity premium. (m is nonlinear in Δc, so adding consumption volatility can add to m volatility. Constantinides and Duffie use this fact to show that idiosyncratic risk can help, though I'm not convinced it's quantitatively realistic. See the CD review.)

d) *Long run assets* -Make matters worse. $E(R^e_{t,t+k})$ is always linear in horizon, but $\sigma^2(R^e)$ rises less than linearly if there is negative serial correlation. Our estimates from predictability were, in fact, that stocks are somewhat "safer" in the long run. This means the long run Sharpe ratio is even better, and the equity premium is worse.

e) *Long run preferences.* As we'll see in the preference review, many preference specifications amount to

$$m_{t+1} = \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} X_{t+1};\ X_{t+1} \text{ stationary}$$

Since X is stationary and C_{t+1}/C_t is roughly iid, when we look at long run returns, we're back to power utility.

4. Comments and Finance

a) The puzzle is *quantitative* not qualitative. Consumption growth *is* positively correlated with stock returns.

b) The puzzle requires us to integrate financial and *macro* data. You can't see it from prices alone.

c) Related, how did traditional finance not notice? Well, traditional asset pricing was about the CAPM, $E(R^{ei}) = \beta_i E(R^{em})$. This takes (ER^{em}) for granted and doesn't ask where it came from. In application, the CAPM was more an APT. Now, the *derivation* of the CAPM has consumption implications. For example, the log utility iid CAPM predicts $C_{t+1}/C_t = R_m^{-1}$. The whole point of CAPM is to tie consumption to the market return, and substitute out for consumption growth in marginal utility. This means however, that the CAPM predicts $\sigma(\Delta c) = \sigma(R) = 18\%$. Nobody noticed or bothered to look at this or even ask if it made any sense. So much for "testing models."

I make a fuss out of this because I think there is something deeply revealing in the fact that consumption *does not* fall by 40% when the market does (except for Savina). In some sense consumption and asset wealth are correlated or cointegrated in the long run. This means the average investor does regard stock market falls (W/C) as transitory and ignores them. There is something very deep in the fact that the CAPM-IID world predicts $\Delta c = R$ and we observe much, much, less.

d) Traditional portfolio theory predicts

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

Thus, you hold all equity if

$$\begin{aligned} 1 &= \frac{1}{\gamma}\frac{0.08}{0.16^2} \\ \gamma &= 3.125 \end{aligned}$$

This seems perfectly sensible, and 4 years of portfolio theory has happily used this formula. What's wrong? The portfolio theory also says $\Delta C = R$. Why take portfolio advice and not consumption advice? If you don't like consumption advice this is the wrong portfolio model!

5. **A few noteworthy trends.**

a) *Standard errors* Well, maybe 8% mean return is overstated. Surely there is the sad fact of $\sigma/\sqrt{T}$. *Stocks are so volatile that it is very hard to measure their mean.* For example, in 49 years of data, $\sigma/\sqrt{T} = 18\%/\sqrt{49} = 18\%/7 = 18/7 = 2.57$. That means a *one* standard error band is +/-2.57 percentage points, and a two standard error band is +/- 5.14 percentage points. 8% mean return could well be 3% mean return.

b) *Rare events and selection bias* That argument is symmetric however, Is there a reason to believe our sample is biased one way or another? From Goetzmann and Jorion:

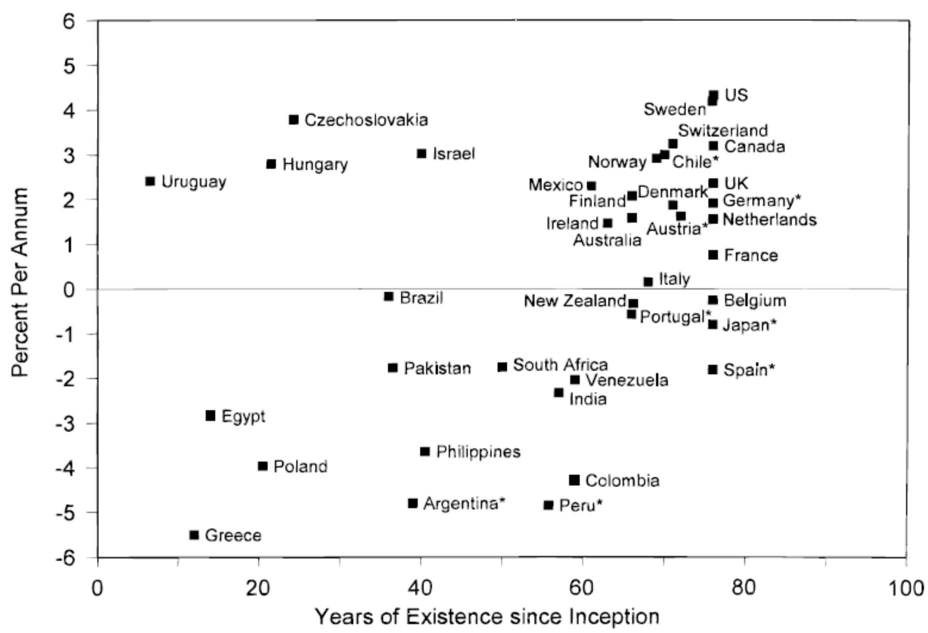


Figure 1. Real returns on global stock markets. The figure displays average real returns for 39 markets over the period 1921 to 1996. Markets are sorted by years of existence. The graph shows that markets with long histories typically have higher returns. An asterisk indicates that the market suffered a long-term break.

Hmm, it looks like the US is the best equity premium. And the US avoided major crashes. Maybe there is a reason we are all in the US studying the US equity premium and not all in Moscow studying the Russian equity premium.

Robert Barro has been pushing *rare events* too, following an early paper by Rietz. If stocks can fall precipitously, then our samples will typically be biased up. In addition, our covariances may all be mismeasured, as the only thing that matters is covariance in the rare states. This is controversial – George Constantinides has calculated that we should see rare events more often than we do under this view of the world, and that bonds fail just as much as stocks in hyperinflations and world wars. I hope it isn't true as it makes our *empirical procedures* almost vacuous.

c) *Where is the surprise?* Fama and French advocated calculating mean returns in a way that avoids changes in valuation.

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

$$R_{t+1} \approx 1 + \Delta\left(P/D\right)_{t+1} + \frac{D_t}{P_t} + \Delta D_{t+1}$$

This makes sense. You get the dividend yield, you get price increase over current dividends, and if P/D doesn't change then ΔD translates one for one to ΔP so dividend growth is return.

Over the very long run, (where the approximation is likely to be bad, since we reinvest dividends, but leave that aside) dividend yields revert, and even in our sample the doubling of price/dividend ratios happened over 50 years so doesn't add that much to mean returns. Thus, thinking about the unconditional mean equity premium, or the premium stripped of expectations that prices will change.

$$R = D/P + \Delta D.$$

Fama and French use this to argue for a lower equity premium than conventionally measured.

It also informs my phone call to Grandpa. To be an equity premium, it must be the case that Grandpa understood mean returns were going to be 8% over bond returns, but didn't want to take the risk. He knew the dividend yield was 4%, not much more than bonds at 2-3%. Did he know that postwar dividend and earnings growth would be so spectacularly good? Or was this a surprise? This thought leads me to an equity premium (ex-ante) of 2-3%, not 8%, but that's all guessing.

2-3% not 6-7% is still a puzzle. Power utility and $\gamma = 2$ still predict $\gamma\sigma(\Delta c) = 0.03$ not $0.03/0.18 = 0.16$. But power utility and iid consumption growth also predict $\sigma(R) = \sigma(\Delta c) = 1.5\%$. *Why are stocks so volatile?* should be part of the puzzle. Well, time-varying expected returns, you say... but then power and iid are all wrong for other reasons.

d) **Boring question** In my view, "model to explain equity premium puzzle" is really boring. In 20 years we haven't gotten that far. In addition (I recognize this is a bit inconsistent) the standards are higher now that there are so many models aimed at one moment. Models need to address the equity premium, the volatility of returns, time-variation in expected returns, the low risk free rate, the stability of the risk free rate, the high autocorrelation of the risk free rate. Increasingly, models need to have an interesting cross-section of returns and describe the value premium.

A quick m* review

Understanding "international risk sharing" will be a lot easier with a quick m^* review.

1. Definition. A discount factor is a random variable such that $p = E(mx)$ for the price p of any payoff x in a set of payoffs X. The set of payoffs may be the set traded by agents, or just a subset included in a particular study. It must be *complete* in what follows, meaning that any portfolio of assets in X is also in X.

2. An agent's first order conditions are

$$p = E\left[\beta^i \frac{u'(c^i_{t+1})}{u'(c^i_t)} x_{t+1}\right] = E\left[m^i_{t+1} x_{t+1}\right]$$

Thus, each agent's marginal utility growth is a discount factor.

3. If markets are complete, then the discount factor is unique, positive, and equal to every agent's marginal utility growth

$$\beta^i \frac{u'(c^i_{t+1})}{u'(c^i_t)} = \beta^j \frac{u'(c^j_{t+1})}{u'(c^j_t)} = m_{t+1}$$

When markets are not complete (spanning, not frictions) then marginal utility growths can differ and there are many discount factors. This is the fun case...

4. So long as the Law of One price (price is a linear function of payoff) holds, we can always construct a unique discount factor m^* that is itself a portfolio ($\in X$)

Example: With a vector of returns R, we can just construct it

$$m^*_t = \frac{1}{R^f} - \frac{1}{R^f} E(R^e)'\Sigma^{-1}(R^e - E(R^e))$$

a) You can check that this discount factor solves

$$\begin{aligned} E(mR^e) &= 0 \\ E(mR^f) &= 1. \end{aligned}$$

How?

$$\begin{aligned} E\left(m^*_t R^f\right) &= 1 - E(R^e)'\Sigma^{-1} E(R^e - E(R^e)) = 1 \\ E(m^*_t R^{e\prime}_t) &= \frac{1}{R^f} E(R^{e\prime}) - \frac{1}{R^f} E(R^e)' = 0 \end{aligned}$$

b) The discount factor is a linear function of excess returns R^e and a constant. Thus it is in the space of payoffs.

c) It is the unique discount factor in the space of payoffs. If you look for $m = a \times R + b' \times R^e$, you can find this a and b uniquely.

d) Notation: The payoff space X is the linear span of our basis payoffs, $\{X = a \times R + b' \times R^e\}$. So "in X" means "a linear function of basis payoffs." The weights are *not* forced to sum to one, and X is not limited to returns (price 1) or excess returns (price zero). Hence m^* is typically neither a return nor an excess return; its price $p(m^*) = E(m^{*2})$ is not 1 or zero.

e) m^* is also not any agent's portfolio. This also holds in complete markets. $m = m^*$ is marginal utility growth, not the agent's portfolio. It is a special portfolio equal to marginal utility growth. That's interesting, but that is not the portfolio that the agent is holding.

Example: In continuous time, if a vector of stock returns and bond return follow

$$\begin{aligned} \frac{dS}{S} &= \mu dt + \sigma dz \\ \frac{dB}{B} &= r dt \end{aligned}$$

then we can construct the discount factor by the analogous formula.

$$\frac{d\Lambda^*}{\Lambda^*} = -rdt - (\mu - r)'\Sigma^{-1}\sigma dB;\ \Sigma = \sigma\sigma'$$

a) This formula satisfies the continuous time pricing conditions

$$\begin{aligned} 0 &= E(d\Lambda S) \\ 0 &= E(d\Lambda B) \end{aligned}$$

or, more precisely

$$E\left(\frac{dB}{B}\right) = rdt = -E\left(\frac{d\Lambda^*}{\Lambda^*}\right)$$

$$\begin{aligned} E\left(\frac{dS}{S}\right) - rdt &= -E\left(\frac{d\Lambda}{\Lambda}\frac{dS}{S}\right) \\ (\mu - r)\, dt &= (\mu - r)'\Sigma^{-1}\sigma\sigma' dt \end{aligned}$$

by construction as well.

Note These are "finite-dimensional" payoff spaces, since they are spanned by a finite set of basis assets. The language and theorems allow infinite-dimensional as well. For example, the full set of options on a stock is an "infinite-dimensional" payoff space.

5. For every discount factor m its mimicking payoff prices just as well.

$$E(mx) = E\left[(proj(m|X) + \varepsilon)\, x\right] = E\left[proj(m|X)x\right] = E\left[m^* x\right]$$

The first equality is just a regression. Define it by $m = b'x + \varepsilon$, and $proj(m|x) = b'x$. The defining property of regression is $E(\varepsilon x) = 0$, so the second equality holds as well. Now, $proj(m|X)$ is acting as a discount factor in the third equality (see definition) and it's in the payoff space

6. Every discount factor m^i can be expressed as m^* plus idiosyncratic noise

$$m^i = m^* + \varepsilon^i;\ \ E(\varepsilon^i x) = 0$$

If the payoff space is generated by a riskfree rate R^f and excess returns R^e, this means that $E(\varepsilon) = 0$ (since $E(\varepsilon R^f) = E(\varepsilon)R^f = 0$) , $E(\varepsilon R^e) = 0$ and $cov(\varepsilon R^e) = 0$. (Picture)

In continuous time,

$$\frac{d\Lambda}{\Lambda} = -rdt - (\mu - r)'\Sigma^{-1}\sigma dB + dw; E(dwdB) = 0$$

7. m^* is the minimum second-moment discount factor

$$\begin{aligned} m^i &= m^* + \varepsilon^i \\ E(m^{i2}) &= E(m^{*2}) + E(\varepsilon^{i2}) + E(m^* \varepsilon^i) \end{aligned}$$

but of course the last term is zero. *If there is a risk free rate,m^* is the minimum-variance discount factor.*

8. The Hansen-Jagannathan bound and Sharpe ratios. Therefore, we know that any admissible discount factor must have variance greater than m^*. (Given a riskfree rate, or if there is none, for any fixed value of $E(m)$ which then implies a riskfree rate)

$$\sigma(m) \geq \sigma(m^*)$$
$$\sigma^2(m^*) = \frac{1}{R^{f2}} E(R^e)'\Sigma^{-1}E(R^e)$$

a) Hansen and Jagannathan took a vector of returns R, and for each hypothesized value of $R^f = 1/E(m)$ they plotted

$$\sigma^2(m^*) = \frac{1}{R^{f2}} E(R - R^f)'\Sigma^{-1}E(R - R^f)$$

This is a parabola.

b) The minimum variance discount factor equals the maximum Sharpe ratio

$$\frac{\sigma^2(m^*)}{E(m^*)^2} = R^{f2}\sigma^2(m^*) = E(R^e)'\Sigma^{-1}E(R^e)$$

This is the formula for the maximum Sharpe ratio you can obtain from simple methods:

$$\begin{aligned} R^p &= w'R^e \\ \min \sigma^2(R^p) \text{ s.t. } E(R^p) &= \mu \\ & \min w'\Sigma w - \lambda w'E \\ \Sigma w &= \lambda E \\ w &= \lambda \Sigma^{-1} E \\ \sigma^2(R^p) &= \lambda^2 E'\Sigma^{-1}E \\ E(R^p) &= \lambda E'\Sigma^{-1}E \\ \frac{E(R^p)^2}{\sigma^2(R^p)} &= E'\Sigma^{-1}E \end{aligned}$$

9. Define R^* as the return corresponding to m^*,

$$R^* = m^*/p(m^*) = m^*/E(m^{*2})$$

R^* is the minimum second-moment return, and therefore lies on the bottom of the mean-variance frontier. This fact is useful to drive home that R^* and m^* are *not* portfolios that anyone holds, they are simply portfolios that reveal something about marginal utility growth. The minimum second moment return is on the bottom of the mean-variance frontier. Quadratic utility investors hold a market portfolio on the top of the mean-variance frontier, not R^*.

R^* is not the minimum variance return, even if there is a risk-free rate. This fact is easiest to show with the decomposition $R^i = R^* + w_i R^{e*} + \eta_i$ which I won't develop here.

10. These mean-variance properties have nothing to do with quadratic utility or normality, as a mean-variance efficient portfolio always prices asset returns. Even though m or m^* has a mean-variance property, the investor's portfolio and consumption growth in general do not. Again, m^* is not the portfolio.

11. In incomplete markets, marginal utility growths may differ due to idiosyncratic risk. However, people will adjust their consumption paths so that the projection of marginal utility growth on the set of asset payoffs is the same for everybody. In this sense, they will use existing markets to share risk as well as possible.

This statement is just a reinterpretation of the above formulas.

$$m^i_{t+1} = \beta^i \frac{u'(c^i_{t+1})}{u'(c^i_t)} = m^*_{t+1} + \varepsilon^i_{t+1} = proj(m^i_{t+1}|X) + \varepsilon^i_{t+1}$$

Let's be clear what this does and does not mean.

a) The ε may be correlated across people. They are uncorrelated with asset payoffs, but that's all.

b) Suppose there is a shock, like the market goes down 40% and credit spread go through the roof, but no idiosyncratic shocks ε. This *event* must have the same effect on your and my marginal utility growth. That does not mean we hold the same portfolio.

-For example, I may own a privately-held piano company, and therefore my asset portfolio is a bit short the piano industry, so my overall exposure to piano stocks equals the market. You hold the market portfolio, including piano stocks. When piano stocks go up, my asset portfolio goes down, my company goes up, and on net my consumption changes exactly as much as yours does

-For example, as in complete markets, people may differ in risk aversion, so a shock to the market affects one's consumption more than another's, but has the same effect on marginal utility.

c) This is a useful way to think about international finance. We typically think that countries can trade in limited securities like stocks and bonds, but not complete claims.