NBER WORKING PAPER SERIES

FINANCIAL MARKETS AND THE REAL ECONOMY

John H. Cochrane

Working Paper 11193
http://www.nber.org/papers/w11193

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2005

This review will introduce a volume by the same title in the Edward Elgar series "The International Library of Critical Writings in Financial Economics" edited by Richard Roll. I encourage comments. Please write promptly so I can include your comments in the final version. I gratefully acknowledge research support from the NSF in a grant administered by the NBER and from the CRSP. I thank Monika Piazzesi and Motohiro Yogo for comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Financial Markets and the Real Economy
John H. Cochrane
NBER Working Paper No. 11193
March 2005
JEL No. G1, E3

ABSTRACT

I survey work on the intersection between macroeconomics and finance. The challenge is to find the right measure of marginal utility of wealth, or "bad times" so that we can understand average return premia distilled in finance "factors" as compensation for assets' tendency to pay off badly in "bad times." I survey the equity premium, consumption-based models, general equilibrium models, and labor income/idiosyncratic risk approaches to this question.

John H. Cochrane
Graduate School of Business
University of Chicago
5807 S. Woodlawn
Chicago, IL 60637
and NBER
john.cochrane@gsb.uchicago.edu

1 Introduction

Risk premia

Some assets must offer higher average returns than other assets, or, equivalently, they attract lower prices. These "risk premiums" should reflect aggregate, macroeconomic risks; they should reflect the tendency of assets to do badly in bad economic times. This volume surveys current research on the central question: what *are* the sources and what is the nature of macroeconomic risk that drives risk premia in asset markets?

Assets must pay high average returns, or equivalently can only attract low prices, as a compensation for risk. The question is, how to define, measure or predict what "risk" is and how much premium it should generate. The central idea of modern finance is that prices are generated by expected discounted payoffs,

$$p_t^i = E_t(m_{t+1} x_{t+1}^i) \tag{1}$$

where x_{t+1}^i is a random payoff of a specific asset i, and m_{t+1} is a stochastic discount factor. Using the definition of covariance and the real riskfree rate $R^f = 1/E(m)$, we can write the price as

$$p_t^i = \frac{E_t(x_{t+1}^i)}{R_t^f} + cov_t(m_{t+1}, x_{t+1}^i) \tag{2}$$

The first term is the risk-neutral present value. The second term is the crucial discount for risk – a large negative covariance means the price is low. Applied to excess returns R^{ei} (short or borrow one asset, invest in another), this statement becomes[1]

$$E_t(R_{t+1}^{ei}) = -cov_t(R_{t+1}^{ei}, m_{t+1}). \tag{3}$$

The "risk premium" or expected excess return is higher for assets with a large negative covariance.

[1] From (1), we have for gross returns R,

$$1 = E(mR)$$

and for a zero-cost excess return $R^e = R^i - R^j$.

$$0 = E(mR^e).$$

Using the definition of covariance, and $1 = E(m)R^f$ for a real risk-free rate,

$$0 = E(m)E(R^e) + cov(m, R^e)$$
$$E(R^e) = -R^f cov(m, R^e)$$

For small time intervals (and exactly in continuous time), $R^f \approx 1$ so we have

$$E(R^e) = -cov(m, R^e).$$

The discount factor m_{t+1} is equal to growth in the marginal value of wealth,

$$m_{t,t+1} = \frac{V_W(t+1)}{V_W(t)}.$$

(This is a simple statement of an investor's first order conditions.) The marginal value of wealth V_W answers the question "how much happier would you be if you found a dollar on the street?" It measures "hunger" – *marginal* not total utility. Thus the discount factor is high when you desperately want more wealth – and would be willing to give up a lot of wealth in other dates/states to get it. Equation (3) says that for *each* of many assets i, or each of many time periods t, expected returns will be high when the covariance of returns with the marginal value of wealth is high.

This risk premium is thus driven by the *covariance* of returns or payoffs with the marginal value of wealth. Given that an asset must do well sometimes and do badly at other times, investors think about *when* it would be desirable for an asset to do well and when it would be tolerable for it to do poorly. Given that the asset must do poorly at some point, it would be better that it did poorly in times when investors do not particularly value extra wealth, and better that it did well in times that investors are otherwise desperate for a little bit of extra wealth. Thus, investors want assets with a *negative* covariance with hunger m, and they will avoid assets with a positive covariance. Investors will thus drive up the prices and drive down the average returns of assets that covary positively with hunger, and vice versa.

These predictions are more surprising to newcomers for what they do not say. The *variance* of the return R^{ei} or payoff is irrelevant and does not measure risk or generate a risk premium. Only the *covariance* of the return with "hunger" matters. Also, many people misunderstand finance because they do not recognize that equations (2) and (3) characterize an *equilibrium*. It's natural to think that assets with large expected returns are "good" and one should buy more of them. But the logic goes the other way: "Good" assets pay off well in bad times when investors are hungry, and have a positive covariance. Since investors all want them, they get *lower* average returns and command higher prices in equilibrium. Assets with higher average returns are forced to pay high average returns or suffer low prices because they are so "bad" – because they pay off badly precisely when investors are most hungry. In fact, there is no "good" and "bad". Equation (3) describes the equilibrium in which the good and the bad are exactly balanced, and no one wants to change portfolios. It describes an equilibrium, it does not provide portfolio advice. (*Deviations* from 3, if you can find them, give portfolio advice.)

To make these ideas operational, we need some procedure to measure the growth in the marginal value of wealth or "hunger" m_{t+1}. The traditional theories of finance, CAPM, ICAPM, and APT, measure hunger by the behavior of large portfolios of assets. In the CAPM, the marginal value of wealth is a declining linear function of the market return. When the market goes down, people are more hungry. Then, a high return must be balanced by a large tendency of an asset to fall just when the market as a whole falls – a high "beta." In equations,

$$E_t(R^{ei}_{t+1}) = cov_t(R^{ei}_{t+1}, R^m_{t+1}) \times \lambda$$

where λ is a constant of proportionality. Multifactor models such as the popular Fama-French (1996) three-factor model use returns on multiple portfolios to measure the marginal value of wealth.

Research connecting financial markets to the real economy – the subject of this volume – goes one step deeper. It asks what are the *real, economic* determinants of the marginal value of wealth? What are the *states of the economy* as a whole that investors fear? For example, I start with the consumption-based model. This model exploits the envelope theorem that at an optimum, a marginal dollar saved is in fact just as good as a marginal dollar spent, so the marginal utility of consumption equals the marginal value of wealth

$$u'(c) = V_W.$$

Hence, equation (3) implies

$$E_t(R^{ei}_{t+1}) = -cov_t\left[R^{ei}_{t+1}, \beta\frac{u'(c_{t+1})}{u'(c_t)}\right]$$

or, assuming that utility depends only on contemporaneous consumption and linearizing marginal utility

$$E_t(R^{ei}_{t+1}) = cov_t\left(R^{ei}_{t+1}, \frac{c_{t+1}}{c_t}\right) \times \gamma.$$

(γ is the coefficient of relative risk aversion $\gamma = -cu''(c)/u'(c)$, β is the subjective discount factor; the investor's willingness to trade future for current utility.) Assets must offer high returns if they pay off well in "good times" and pay off badly in "bad times" *as measured by aggregate consumption growth.*

As we will see, this simple attractive model does not (yet) work that well. The research in this volume is aimed at improving that performance. It aims to find a good measure of the marginal value of wealth, rooted in measures of economic conditions, that explains the pattern by which mean returns $E_t(R^{ei}_{t+1})$ vary across assets i and over time t.

Who cares?

With this simple version of a "real" asset pricing model in mind, we can answer questions such as: Why is this important? What do we learn by connecting asset returns to macroeconomic events in this way? Why bother, given that "reduced form" or portfolio-based models like the CAPM are guaranteed to perform better?

Macroeconomics

Understanding the marginal value of wealth that drives asset markets is most obviously important for macroeconomics. The centerpiece of dynamic macroeconomic theory is the equation of savings and investment, the equation of marginal rates of substitution with marginal rates of transformation; the forces that determine the allocation of consumption and investment across time and states of nature. Asset markets are the mechanism that does all this equating. Asset returns are the "price line" that draws together marginal rates of substitution and marginal rates of transformation. If we can learn the marginal value of wealth from asset markets, we have a powerful measurement of the key ingredient of all modern, dynamic, intertemporal macroeconomics.

In fact, the first stab at this piece of economics is a disaster, in a way made precise by the "equity premium" discussion and readings. The marginal value of wealth needed to make sense of the most basic stock market facts is orders of magnitude more volatile than that specified in all macroeconomic models. Clearly, finance has a *lot* to say about macroeconomics, and something seems desperately wrong with most macroeconomic models.

In response to this challenge, many macroeconomists simply dismiss asset market data. "Something's wacky with stocks" they say, or perhaps "stocks are driven by waves of fads and fashion disconnected from the real economy." That might be true, but if so, by what magic are marginal rates of substitution and transformation equated? It simply makes no sense to dismiss asset market data (or prices in general) but then study the solutions to planning problems as guides to the behavior of the actual, competitive, economy. Asset markets *are* the mechanism for equating marginal rates of substitution and transformation. If asset markets are screwed up, so is the equation of marginal rates of substitution and transformation in every macroeconomic model. Asset markets must have an even greater impact on macroeconomics if their economic explanation *fails* than if it succeeds! If we end up concluding that asset prices *are* disconnected from marginal rates of substitution or transformation, this disconnection will have enormous allocation and welfare implications, which macroeconomists will need to understand. Just saying "markets are crazy" and going back to frictionless models with wildly counterfactual asset-pricing implications makes no sense.

Finance

Many financial economists dismiss macro approaches to asset pricing because portfolio-based models "work better" – they provide smaller pricing errors. This dismissal of macroeconomics by financial economists is just as misguided as the dismissal of finance by macroeconomists.

First, a good part of the better performance of portfolio-based models may reflect Roll's (1977) theorem that we can *always* construct a reference portfolio that perfectly fits all asset returns–the sample mean-variance efficient portfolio. The *only* content to empirical work in asset pricing is what constraints the author put on his fishing expedition to avoid rediscovering this theorem. The instability of many "anomalies" and the changing popularity of different factor models (Schwert 2003) lends some credence to this worry.

The main fishing constraint one can imagine is that the portfolios *are* in fact mimicking portfolios for some well-understood macroeconomic risk. Fama (1991) famously labeled the ICAPM and similar theories "fishing licenses," but his comment cuts in both directions. Yes, current empirical implementations do not impose much structure from theory, but no, you can't fish without a license. For example, momentum has yet to acquire the status of a factor despite abundant empirical success, because noone can think of an even vaguely plausible story that it corresponds to some measure of the marginal utility of wealth. (Of course macro modelers need to worry about fishing biases too, but at least there is no theorem saying that a perfect in-sample fit is right in the middle of the net.)

Second, much work in finance is framed as answering the question whether markets are "rational" and "efficient" or not[2]. *No* amount of research using portfolios on the right hand side can *ever* address this question. The only content to the "rationality" question is whether the "hunger" apparent in asset prices – the discount factor, marginal value of wealth, etc. – mirrors macroeconomic conditions correctly. Markets can only be said to be "rational" if high average returns correspond to poor performance in "bad times" in the *real* economy. Absent arbitrage opportunities, markets can only be found "irrational" if we can prove that the periods of "hunger" and "fullness" mirrored in asset prices have no connection to real events in the economy. If Mars has perfectly smooth consumption growth, then prices that are perfectly "rational" on volatile Earth would be "irrational" on Mars. Price data alone *cannot* answer the question, because you can't tell from the prices which planet you're on. For example, Fama (1991, p. 1610) frames the debate, properly, in this way:

> In the end, I think we can hope for a coherent story that ...relates the behavior of expected returns to the real economy in a rather detailed way. Or we can hope to convince ourselves that no such story is possible.

In sum, the program of understanding the *real, macroeconomic* risks that drive asset prices (or the proof that they do not do so at all) is not some weird branch of finance; it is the trunk of the tree. As frustratingly slow as progress is, this is the *only* way to answer the central questions of finance, and a crucial and unavoidable set of uncomfortable measurements for macroeconomics.

The mimicking portfolio theorem and the division of labor

Portfolio-based models will always be with us. The "mimicking portfolio" theorem states that if we have the perfect, real model of the marginal utility of wealth, then a portfolio formed by its regression on to asset returns will work just as well[3]. And this "mimicking

[2]My view is that this is a pointless debate and, while very useful in the 1970s, no longer a good way to organize empirical work in asset pricing. All that matters are models, predictions and data. Physics didn't get where it is today by framing the enterprise as a "debate" between science and religion. And if 40 years haven't made one dent of difference in the debate it must not be over anything observable. But many people like to frame the investigation this way, so it's worth thinking about evidence that might conceivably bear on it.

[3]Start with the true model,

$$0 = E(mR^e).$$

Consider a regression of the discount factor on excess returns, with no constant,

$$m = b'R^e + \varepsilon.$$

By construction, $E(R^e\varepsilon) = 0$, so

$$0 = E\left[\left(b'R^e\right)R^e\right]$$

Therefore, th zero-cost portfolio $b'R^e$ is a discount factor as well.

portfolio" will have better-measured and more frequent data, so it will work better in sample and in practice. It will be the right model to recommend for many applications.

This theorem is important for doing and evaluating empirical work. First, together with the Roll theorem, it warns us that it is pointless to engage in an alpha contest between real and portfolio-based models. Ad-hoc portfolio models must always win this contest – even the *true* model would be beat by its own mimicking portfolio because of measurement issues, and it would be sorely beaten by an ad-hoc portfolio model that could slide a bit towards the sample mean-variance frontier. Thus the game "see if macro factors do better than the Fama French 3 factor model" in pricing the Fama French 25 portfolios is pointless. Even if you do succeed, the "small-growth/large-value" fourth factor (or the increasingly popular momentum factor) can always come back to trump any alpha successes.

Portfolio-based models are good for relative pricing; for describing one set of asset returns given another set. The CAPM describes average returns of stock portfolios *given* the market premium. The Fama French model describes average returns of 25 size and book/market sorted portfolios *given* the average returns of the three factor portfolios. But why is the average market return what it is? Why are the average returns of the Fama-French value and size portfolios what they are? Why does the expected market return vary over time? By their nature, portfolio models *cannot* answer these questions. Macroeconomic models are the *only* way to answer these questions.

With this insight, we can achieve a satisfying division of labor, rather than a fruitless alpha-fishing contest. Portfolio models document whether expected returns of a large number of assets or dynamic strategies can be described in terms of a few sources of common movement. Macro models try to understand why the common factors are priced. The question for a macro model is "why is HML priced?" and much less "can we beat the Fama French factors in pricing the Fama French 25?" This investigation is the only way to solve the "rational – irrational debate" or more generally (and productively, I think) it is the only way to understand the sources of macroeconomic risk. Such an understanding will of course ultimately pay off for pure portfolio questions, by helping us to understand which apparent risk premia are stable rewards for risk, and which were chimeric features of the luck in one particular sample.

2 Overview of the papers

I have included in this volume articles that exemplify recent published research in the important areas of macro/asset pricing. In most areas, I have included one classic paper that summarizes a literature in a compact way, plus one or two more recent papers that point to promising extensions.

In any such compilation, the editor is most pained by the omissions, made necessary by the limitations of space and reader patience. In many cases, I have not included the "classic" papers that justly deserve credit for first making an important point. But the point of this volume is to help the reader quickly grasp the issues as we now understand them, not to celebrate the longer and complex history of thought that brought us to where we are today, or to give just honor to people for their contributions. I also emphasize recent papers that suggest new directions. Some of these papers are a bit unpolished, circling around a point that will emerge with time rather than setting down a completely digested point for all time. This volume will be most useful if it provokes that new research rather than celebrating the old.

3 Facts: Time-variation and business cycle correlation of expected returns

We start with the facts. What is the pattern by which expected returns vary over time and across assets? What is the variation on the *left* hand side of (3) that we want to explain by understanding the marginal value of wealth on the right hand side of (3)? We are, of course especially interested in expected return variation that seems interesting for macroeconomics; variation that seems like it will reveal something about real risks.

First, a number of variables forecast aggregate stock, bond, and foreign exchange returns. Thus, *expected returns vary over time.* If we find $b > 0$ in $R_{t+1} = a + bx_t + \varepsilon_{t+1}$, then we know that $E_t(R_{t+1})$ varies over time. These variables typically have a clear and suggestive business cycle correlation. Expected returns are high in "bad times," when we might well suppose people are less willing to hold risks.

Second, a large number of stock characteristics are now associated with average returns. The book/market ratio is the most famous, but a long list of other variables including size (market value), sales growth, past returns, past volume, and accounting ratios are associated with average returns going forward. This fact alone would not cause much trouble for traditional finance theory if these characteristics were also associated with large market betas. Alas, they are not. Instead, the empirical finance literature has associated these patterns in expected returns with new "factors" – new sources of common variation in returns. As crystallized in a series of papers by Fama and French, average returns are higher for stocks that move a lot when the market moves a lot (high market beta); that move a lot when small stocks move a lot (size) and that move a lot when value stocks (stocks with low prices relative to book value) move a lot. Average returns are also higher for assets that move a lot when long-term interest rates and interest rates on default-prone bonds move. In this sense, these five indicators seem important for measuring the marginal value of wealth.

More factors are on the way. "Momentum" is the tendency for last year's winners to keep going for a few months and vice versa; this spread in expected returns is not captured by the Fama-French factors, though it is admirably captured by a "momentum" factor. The puzzles in Fama and French (1996), in particular the behavior of small growth stocks and large value stocks also corresponds to the next eigenvector of the covariance matrix of returns of their 25 portfolios. If one takes the expected return puzzle seriously, a new factor is readily at hand to capture it. However, researchers are understandably reluctant to blithely add on new factors for every anomaly (though risk models in common industry use happily use hundreds of factors), and for good reasons. While Fama and French (1996) at least had plausible words to suggest that the "value factor" was a mimicking portfolio for the marginal utility of wealth, such an argument for momentum is elusive to say the least.

New Facts in Finance

I start this section with a review paper, "New Facts in Finance," that synthesizes current research on both issues. (Chapter 20 of *Asset Pricing,* Cochrane 2004 is a somewhat expanded version, with more emphasis on the relationship between various time series representations. Campbell 2003 also has a nice summary of facts.)

Fama and French "Business conditions"

I include Fama and French's (1989) "Business conditions and expected returns on stocks and bonds" as a summary and exemplar of the very large body of work documenting variation of returns over time. This paper shows how dividend-price ratios and term premia (long bond yield less short bond yield) forecast stock and bond returns. The paper emphasizes the comforting link between stock and bond markets: the term premium forecasts stock returns much as it forecasts long-term bond returns. Since stock dividends can be thought of as bond coupons plus risk, so we should see any term premium reflected in stock returns. Most importantly, Fama and French show by a series of plots how the variables that forecast returns are correlated with business cycles.

This paper really is the tip of a large iceberg of return predictability, and a short comment on that history is in order. The classic view that "stocks follow a random walk," meaning that the expected return is constant over time, was first challenged in the late 1970s. Fama and Schwert (1977) found that expected stock returns did not increase one-for-one with inflation. They interpreted this result that expected returns are higher in bad economic times, since people are less willing to hold risky assets, and lower in good times. Inflation is lower in bad times and higher in good times, so lower expected returns in times of high inflation are not a result of inflation, but a coincidence.

To us, the association with inflation that motivated Fama and Schwert is less interesting, but the core finding that expected returns vary over time, and are correlated with business cycles, (high in bad times, low in good times) remains the central fact. Fama and Gibbons (1982) followed up, and added investment to the economic modeling, presaging the investment and equilibrium models we study later.

Next, we learned that foreign exchange returns and bond returns vary over time, and we accumulated direct regression evidence that expected stock returns vary over time. Hansen and Hodrick (1980) and Fama (1984) documented the predictability of exchange returns by running regressions of returns on forward-spot spread or interest rate differentials across countries. Fama (1984) documented the predictability of short-term bond returns, and Fama and Bliss (1987) the predictability of long-term bond returns, by running regressions of bond returns on forward-spot spreads or yield differentials. The latter findings in particular have been extended and stand up well over time. (Stambaugh 1988 for short term bonds and Cochrane and Piazzesi 2005 for long term bonds.) Poterba and Summers (1988) and Fama and French (1988a) documented that past stock returns forecast subsequent returns at long horizons, and Shiller, Fischer, and Friedman (1984), DeBondt and Thaler (1985), and Fama and French (1988b) showed that dividend/price ratios forecast stock returns, the latter emphasizing the long horizons at which the R^2 rise to 60%.

These papers run simple forecasting regressions of returns at time $t+1$ on variables at time t. The forecasting variables are all based on market prices, though, which seems to take us away from our macroeconomic quest. However, as emphasized by Fama and French (1989), the prices that forecast returns are *correlated* with business cycles. A number of authors including Estrella and Hardouvelis (1991) and more recently Ang and Piazzesi (2004) have also documented that the price variables that forecast returns also *forecast* economic activity. Most importantly, the term premium (long term bond yield - short term bond yield) is high in the bottoms of recessions, forecasts large stock and bond returns, and also forecasts that GDP growth will be large as we emerge from recession.

A related literature including Campbell and Shiller (1988) and Cochrane (1991a), reviewed here in "New Facts in Finance", relates the time-series predictability of stock returns to stock price volatility. Iterating the identity $1 = R_{t+1}^{-1} R_{t+1}$ we can obtain an identity that looks a lot like a present value model,

$$\frac{P_t}{D_t} = E_t \sum_{j=0}^{\infty} \left(\prod_{k=1}^{j} \frac{1}{R_{t+k}} \right) \frac{D_{t+j}}{D_t} + \lim_{j \to \infty} E_t \left(\prod_{k=1}^{j} \frac{1}{R_{t+k}} \right) \frac{D_{t+j}}{D_t} \frac{P_{t+j}}{D_{t+j}}, \tag{4}$$

If price/dividend ratios vary at all, they *must* either 1) forecast dividend growth 2) forecast returns or 3) prices must follow a "bubble" in which the price-dividend ratio rises without bound. It would make sense if variation in price-dividend ratios corresponded to dividend forecasts. Investors, knowing future dividends will be higher than they are today, bid up stock prices relative to current dividends; then the high price/dividend ratio forecasts the rise in dividends. It turns out that price dividend ratios *no not* forecast aggregate dividends at all. This is the "excess volatility" found by Shiller (1981) and LeRoy and Porter (1981). However, prices can also be high if this is a time of temporarily low expected returns; then the same dividends are discounted at a lower rate, and price-dividend ratios forecast returns. It turns out that the return forecastability we see in regressions is just enough to completely account for the volatility of price dividend ratios through (4). Thus, return forecastability and "excess" volatility are *exactly* the same phenomenon. Thankfully, the last "bubble" term seems to be absent. However, the fact that return predictability does all the work means that we have to tie stock market movements to the macroeconomy entirely through harder-to-measure time-varying *risk premia* rather than easier-to-understand cashflows.

Lettau and Ludvigson.

I include Lettau and Ludvigson's (2001) "Consumption, Aggregate Wealth and Expected Stock Returns" next. This paper is an important recent extension of stock return forecastability. Lettau and Ludvigson find that the ratio of consumption to wealth – a pure macroeconomic variable – forecasts stock returns. This allays some worry that D/P forecastability just reflects spurious mean reversion in prices. A price “too low” today also leads to a high return from today to tomorrow. This paper directly connects return forecastability to macroeconomic data. (Cochrane 1991b discussed below finds a production-side corollary: the investment/capital ratio also forecasts stock returns.)

In Cochrane (1994) I showed that consumption provides a natural “trend” for income, and so we see long-run mean reversion in income most easily by watching the consumption-income ratio. I also showed that dividends provide a natural “trend” for stock prices, so we see long-run mean-reversion in stock returns most easily by watching the dividend/price ratio. Lettau and Ludvigson nicely put the two pieces together, showing how consumption relative to income and wealth has a cross-over prediction for long run stock returns. Exploiting another stable ratio (cointegrating relation), Lamont (1998) showed that the dividend-earnings ratio forecasts returns.

Lettau and Ludvigson (2004) recently showed that the consumption-wealth ratio also forecasts *dividend* growth. This was initially surprising. So far, very little has forecast dividend growth, and it is an enduring puzzle that prices do not forecast dividend growth. And if the consumption-wealth ratio forecasts dividend growth, why is that not reflected in prices? Lettau and Ludvigson answer this puzzle by repeating that the consumption-wealth ratio forecasts returns. In the context of (4), the consumption-wealth ratio sends dividend growth and returns in the same direction, so its effects on the price/dividend ratio offset. Thus, on second thought, the observation is natural. If *anything* forecasts dividend growth it must *also* forecast returns to account for the fact that price/dividend ratios do not forecast dividend growth. And conversely, if anything has additional explanatory power for returns, it must also forecast dividend growth. And now, delightfully, we have a new variable and an opening for additional variables that forecast *both* returns and cashflows, giving stronger links from macroeconomics to finance.

Fama and French and the cross-section of returns

Currently, the size and book-to-market effects together with the momentum anomaly define the stylized facts for the cross-section of returns. I include Fama and French (1996) “Multifactor explanations of asset pricing anomalies”. This paper concisely summarizes an extensive amount of work by Fama and French (1992, 1993) and others. The basic finding is that small and value (low market/book value) stocks pay higher average returns than large and growth (high market/book value). These findings are not so surprising on their own; high expected returns *should* be revealed by low market values. The puzzle is that these high expected returns do not correspond to large market betas. Fama and French explained the average returns of 25 size and book/market portfolios by size and book/market factors; the higher returns of the 25 size and book/market portfolios line up with betas of those portfolios on a single large portfolio based on size or book/market.

If this argument seems a bit circular, realize that the central observation is that size and value stocks *move together.* If there is a spread in average returns, stocks with high average returns must move together. Otherwise, one could build a diversified portfolio of high expected return stocks, short a portfolio of low expected return stocks and make huge profits with no risk. Fama and French also address the seeming circularity by noting that size and book to market factors explain average return spreads formed when one sorts on other variables, including sales growth which is a completely non-financial variable.

Again, the important question for macro models is to explain why the market premium (CAPM), the Fama French factors (hml, smb), or a "momentum factor" are priced. These quantities are *inputs* to the portfolio-based models. Fama and French speculate suggestively on the macroeconomic foundations of the value premium (p. 77)

> One possible explanation is linked to human capital, and important asset for most investors. Consider an investor with specialized human capital tied to a growth firm (or industry or technology). A negative shock to the firm's prospects probably does not reduce the value of the investor's human capital; it may just mean that employment in the firm will expand less rapidly. In contrast, a negative shock to a distressed firm more likely implies a negative shock to the value of specialized human capital since employment in the firm is more likely to contract. Thus, workers with specialized human capital in distressed firms have an incentive to avoid holding their firms' stocks, If variation in distress is correlated across firms, workers in distressed firms have an incentive to avoid the stocks of all distressed firms. The result can be a state-variable risk premium in the expected returns of distressed stocks.

Much of the work described below tries to formalize this kind of intuition and measure the required correlations in the data.

This paper is also important because it has, for better or worse, defined the methodology for evaluating asset pricing models. A generation of papers studies the Fama-French 25 size and book to market portfolios to see whether alternative factor models can explain the average returns. More generally, empirical papers now routinely sort on other characteristics and run time-series regressions to see which factors explain the spread in average returns. Most importantly, where in the 1980s papers would focus entirely on the p value of some overall statistic, Fama and French rightly got people to focus on the spread in average returns and the spread in betas. Remarkably, this, the most successful model since the CAPM, is decisively *rejected* by formal tests. Fama and French taught us to pay attention to more important things than test statistics.

Momentum remains an anomaly to Fama and French. Portfolios formed on the basis on past returns do show a spread in average returns, but that spread is not explained by size and book/market betas. It is true that an additional "momentum" factor will "explain" the momentum portfolio returns, but we are justly nervous about adding a factor for each anomaly. The empirical literature has continued, and found a large number of additional expected return anomalies based on additional characteristics including interactions between momentum and volume, accounting ratios, managerial actions such as equity issuance and so forth. Some of these do not seem well explained by existing factor models. I do not survey this enormous literature here. Our purpose is to survey macro asset pricing, and that literature has not yet extended beyond trying to understand the equity, size, and value premiums. (And as you will see, many branches have not even gotten that far.) It's also clear that, if these additional expected return phenomena are real out of sample and survive transactions costs, they will resist macroeconomic explanation. Most are high-frequency phenomena, and macro conditions do not move that fast.

Liew and Vassalou

Naturally, a large body of empirical research asks whether the size and book to market factors do represent macroeconomic phenomena via rather a-structural methods. I include one such paper, Liew and Vassalou (2000). It is natural to suppose that value stocks – stocks with low prices relative to book value, thus stocks that have suffered a sequence of terrible shocks – should be more sensitive to recessions and "distress" than other stocks, and that the value premium should naturally emerge as a result. Initially, however, efforts to link value stocks and value premia to economic or financial trouble did not bring much success. Fama and French (1997a, 1997b) were able to link value effects to *individual* cash flows and "distress," but getting a premium requires a link to *aggregate* bad times, a link that was notoriously absent as emphasized by Lakonishok, Shleifer and Vishny (1994). However, in the 1990s and early 2000s, value stocks have moved quite reliably with the aggregate economy, so more recent estimates do show a significant and heartening link. In this connection, Liew and Vassalou (2000) show that Fama and French's size and book to market factors forecast output growth, and thus are "business cycle" variables. Alas, momentum still seems unconnected to the real economy.

4 Equity Premium

With the basic facts in mind, we are ready to see what theories can match the facts; what specifications of the marginal utility of wealth V_W can link asset prices to macroeconomics.

The most natural starting point is the classic consumption-based asset pricing model. It states that expected excess returns should be proportional to the covariance of returns with *consumption growth*, with risk aversion as the constant of proportionality. If the utility function is of the simple time-separable form

$$E_t \sum_{j=0}^{\infty} \beta^j u(c_{t+j})$$

then the marginal value of wealth is the marginal utility of consumption, and our basic asset pricing equation (3) becomes[4]

$$E_t(R^{ei}_{t+1}) = -cov_t\left(R^e_{t+1}, \beta\frac{u'(c_{t+1})}{u'(c_t)}\right), \tag{5}$$

or, with the popular power utility function $u'(c) = c^{-\gamma}$, (or using that form as a local approximation) and for short time intervals so $\beta \approx 1$,

$$E_t(R^{ei}_{t+1}) = \gamma \times cov_t\left(R^e_{t+1}, \frac{c_{t+1}}{c_t}\right). \tag{6}$$

This model is a natural first place to link asset returns to macroeconomics. It has a great economic and intuitive appeal. Assets should give a high premium if they pay off badly in "bad times." What better measure of "bad times" than consumption? People may complain, or seem to be in bad straits, but if they're going out to fancy dinners you can tell that times aren't so bad after all. More formally, consumption subsumes or reveals all we need to know about wealth, income prospects, etc. in a wide class of models. In *every* formal derivation of the CAPM, ICAPM, and any other factor model (at least all the ones I know of), the marginal utility of consumption growth is a *single* factor that should subsume all the others.

The essence of the equity premium puzzle is now simple to state. The issue is whether this consumption-based model can explain the most basic premium, that of the market portfolio over the risk free rate. (Again, notice in this exercise the proper role of macro models – the CAPM takes the mean market return as given. We are asking what are the economics behind the mean market return, and the CAPM is silent on this question.) From (6) write

$$E(R^{ei}) = \gamma\sigma(R^{ei})\sigma(\Delta c)\rho$$

so, since $\|\rho\| < 1$,

$$\frac{\|E(R^{ei})\|}{\sigma(R^{ei})} < \gamma\sigma(\Delta c) \tag{7}$$

The left hand side of (7) is the "Sharpe ratio" a common finance measure of the ratio of reward to risk in asset markets. In postwar US data, the mean return of stocks over bonds is about 8% with a standard deviation of about 16%, so the Sharpe ratio is about 0.5. Longer time series and other countries give somewhat lower values, but numbers above 0.2-0.3 are characteristic of most times and markets. Other investments (such as value stocks or some dynamic strategies in bond markets) can sometimes give much larger numbers, up to Sharpe

[4] In discrete time, the actual equation is

$$E_t(R^{ei}_{t+1}) = -\frac{1}{R^f}cov_t\left(R^e_{t+1}, \beta\frac{u'(c_{t+1})}{u'(c_t)}\right),$$

with

$$\frac{1}{R^f_t} \equiv E_t\beta\frac{u'(c_{t+1})}{u'(c_t)}.$$

The simpler form of Equation (5) results in the continuous-time limit.

ratios of 1.0.

Aggregate nondurable and services consumption volatility is much smaller, about 1.5% per year in the postwar US. To get from $\sigma(\Delta c) = 0.015$ to a Sharpe ratio of 0.5 we need a risk aversion of at least $0.5/0.015 = 33$, which seems much larger than most economists find plausible.

One might simply accept high risk aversion, but the corresponding equation for the risk free rate, from the continuous-time limit of $1 + r^f = 1/E\left(e^{-\delta}\frac{u'(c_{t+1})}{u'(c_t)}\right)$, is

$$r^f = \delta + \gamma E(\Delta c) - \frac{1}{2}\gamma(\gamma+1)\sigma^2(\Delta c). \tag{8}$$

If we accept $\gamma = 33$, with about 1% expected consumption growth $E(\Delta c) = 0.01$ and $\sigma^2(\Delta c) = 0.015^2$, we predict a risk free rate of

$$\begin{aligned} r^f &= \delta + 33 \times 0.01 - \frac{1}{2} \times 33 \times 34 \times (0.015^2) \\ &= \delta + 0.33 - 0.13 \end{aligned}$$

Thus, with $\delta = 0$, the model predicts a 20% interest rate. To generate a (say) 5% interest rate, we need a *negative* 15% discount rate. Worse, (8) with $\gamma = 33$ predicts that the interest rate will be extraordinarily sensitive to changes in expected consumption growth or consumption volatility. Therefore, the puzzle is often known as the "equity premium - risk free rate" puzzle.

The puzzle is a lower bound, and more information makes it worse. Among other observations, we do know something about the correlation of consumption and asset returns, and we know it is less than one. Using the sample correlation of 0.2 in postwar quarterly data, i.e. using the sample covariance in (6), raises the required risk aversion by a factor of 5, to 165!

It bears emphasizing that the equity premium puzzle, and the larger failure of the consumption-based model that it crystallizes, is *quantitative*, not *qualitative*. The signs are right. The stock market does covary positively with consumption growth; the correlation is about 0.41 in annual data, so the market should give a positive risk premium. The problem is that the risk premium is *quantitatively* too large to be explained given sensible risk aversion and the observed volatility of consumption growth.

Also, the puzzle necessarily unites *macroeconomic* and financial analysis. The finance profession did not notice the puzzle for many years. Finance models always had consumption hidden in them, and that consumption process had huge volatility. Consumption is proportional to wealth in the *derivation* of the CAPM, so the CAPM predicts that consumption should inherit the large 16% or so volatility of the stock market. This is a logical possibility, but not the world we live in. The crucial element of the puzzle is that in our world, consumption does *not* move one-for-one with wealth at short (even 1-5 year) horizons.

That consumption is so much smoother than wealth remains a deep insight for understanding economic dynamics, one whose implications have not been fully explored. For example, it implies that *one* of consumption or wealth must have substantial dynamics. If wealth increases 16% in a typical 1σ year and consumption moves 2% in the same 1σ year, either consumption must eventually rise 14% or wealth must eventually decline 14%, as the consumption/wealth ratio is stable in the long run. This is a powerful motivation for Lettau and Ludvigson's use of consumption/wealth as a forecasting variable.

Mehra and Prescott; Cochrane and Hansen

The ink spilled on the equity premium would sink the Titanic, so there is no way here either to do justice to all who contributed to or extended the puzzle, or even to summarize the huge literature. I include two papers, Mehra and Prescott's (1985) "Equity Premium Puzzle" and Cochrane and Hansen's (1992) review "Asset Pricing Explorations for Macroeconomics."

My quick overview takes Cochrane and Hansen's approach. The fundamental idea there, i.e. (7) is due to Hansen and Jagannathan (1991), which makes many deep insights into the representation of asset prices as well as give a nice treatment of the equity premium puzzle. Cochrane and Hansen (1992) discuss the bounds including correlation as above and a large number of additional extensions. Weil (1989) pointed out the risk free rate part of the puzzle. Chapters 1 and 21 of *Asset Pricing* (Cochrane 2004) also give what I think is a good and intuitive presentation of the equity premium and related puzzles. Campbell (2003) and Kocherlakota (1996) are also excellent recent reviews.

Mehra and Prescott (1985) is the paper that really brought attention to the equity premium puzzle. Mehra and Prescott take a different approach from my simple synthesis: they specify an explicit two-state-Markov process for consumption growth, they calculate the price of the consumption claim and risk free rate, and they point out that the mean stock excess return so calculated is much too low unless risk aversion is raised to apparently implausible values (55, in this case). In retrospect, I think we can see the equity premium puzzle much more clearly with the simple manipulations of first order conditions as above. However, calculating explicit asset prices in simple endowment economies should be part of every financial economist's toolkit, and Mehra and Prescott are as good a place as any to learn how to do this.

In addition, Mehra and Prescott style modeling imposes great discipline on this kind of research since the covariance of consumption with returns is generated endogenously. Seemingly normal specifications of the model can generate unexpected results. For example, positive consumption growth autocorrelation and risk aversion greater than one generates a *negative* equity premium because it generates a *negative* covariance of consumption growth with returns. In general equilibrium, you can't just take $cov(R, \Delta c)$ as given and crank up γ (see (6)) to get any premium you want.

There is some issue on who should get credit for discovering the equity premium puzzle. Grossman and Shiller (1981) reported some "preliminary results" that risk aversion estimates in consumption models seemed strangely high. The estimates were actually published in

Grossman, Melino and Shiller (1987). Hansen and Singleton (1983) show very high risk aversion estimates coming from unconditional stock and bond returns, which is the heart of the puzzle. (They give ample credit for the point to Grossman and Shiller.) It's interesting that Mehra and Prescott's more complex approach was initially more influential. In part, it is clear on rereading all these papers that Mehra and Prescott were the first to realize the importance of what they found. Grossman and Shiller dismissed their results as unbelievable and the results of puzzling recent data[5]. Hansen and Singleton simply reported high risk aversion estimates, but seemed to regard them as not particularly interesting results of inefficient (since they leave out instruments) estimates[6]. Mehra and Prescott made the case that high risk aversion was a robust feature of the data, and an important puzzle to think about. In addition, Mehra and Prescott gave a *structure* that many people found useful for thinking about variations on the puzzle. A very large number of alternative explicitly calculated endowment economies followed Mehra and Prescott. In finance as elsewhere, identifying, marketing and packaging the insight, and leaving a structure that others can play with, are justly important contributions.

My view of the literature is that "explaining the equity premium puzzle" is dying out. We have several preferences consistent with equity premium and risk free rates, including habits and Epstein-Zin preferences. No model has yet been able to account for the equity premium with low risk aversion, and Campbell and Cochrane (1999) offer some reasons why this is unlikely to be achieved. So we may have to accept high risk aversion, at least for reconciling aggregate consumption with market returns in this style of model. At the same time, many economists' beliefs about the size of the equity premium are declining from the 8% postwar average, to the 6% average in longer samples, down to 2 or 3%. The postwar US economy, and even the last century may have been lucky; people at the beginning of these periods certainly may not have confidently expect the economic growth that we experienced. Attention is moving to understanding return dynamics and the cross-section, either ignoring the equity premium or simply allowing high risk aversion to account for it. One never can tell when a striking new insight will emerge, but I can tell that new twists in the standard framework are attracting less attention.

5 Consumption models

Really, the most natural thing to do with the consumption-based model is to estimate it and test it, as one would do for any economic model. Logically, this investigation comes

[5] They report: "We have some preliminary results on the estimation of A [risk aversion] and β [discount factor]...Unfortunately, the estimates of A for the more recent sub-periods seem implausbly high." They attribute the result to "the divergence between P^* and P since the early 1950's as well as the extremely low real returns on short-term bonds in this period. There was an enormous rise in stock prices in that period..."

[6] Hansen and Singleton describe the crucial Table 5 thus: "Consistent with their [Grossman and Shiller's] results, we found $\|\hat{\alpha}\|$ $[\gamma]$ to be very large with a correspondingly large standard error when NLAG=0. Consistent with our other findings $\|\hat{\alpha}\|$ is approximately one when the serial correlation in the time-series data is taken into account in estimation. This shows the extent to which the precision and magnitude of our estimates rely on the restrictions across the serial correlation parameters of the respective time series. "

before "puzzles" which throw away information (correlation, multiple assets, time-variation of moments). The puzzles are not tests, they are useful diagnostics for why tests fail.

Hansen and Singleton

The classic consumption-based model test is due to Hansen and Singleton (1982, 1984), which I include. The influence of this paper is hard to overstate. It gives a clear exposition of the GMM methodology, which has pretty much taken over much estimation and testing. (*Asset Pricing*, Cochrane 2004 maps all standard asset pricing estimates into GMM.) Also, generalizing Hall's (1978) test for a random walk in consumption, with this work macroeconomists realized they did not need to write complete models before going to the data; they could examine the first-order conditions of investors without specifying technology, model solution, and a complete set of shocks.

Hansen and Singleton examine the discrete-time nonlinear consumption-based model with power utility,

$$E_t\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}R^i_{t+1}\right]=1 \tag{9}$$

If you like thinking in terms of expected return vs. covariances, this equation is equivalent to

$$E_t(R^i_{t+1})=-cov_t\left[R^i_{t+1},\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}\right]/E_t\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}\right]. \tag{10}$$

The method is astonishingly simple. Multiply both sides both sides of (9) by instruments – any variable z_t observed at time t – and take unconditional expectations, yielding

$$E\left\{\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}R^i_{t+1}-1\right]z_t\right\}=0 \tag{11}$$

Then, take sample averages, and search numerically for values of β, γ that make these "moment conditions" (equivalently, pricing errors) as small as possible. GMM gives a distribution theory for the parameter estimates, and a test statistic based on the idea that these pricing errors should not be too big.

Hansen and Singleton's (1984) results provide a useful baseline. If we take a single asset and multiply it by instruments (Hansen and Singleton's Table I), we are asking (see (10)) whether movements in returns predictable by some instrument z_t – as in the regressions of R_{t+1} on z_t we studied above – are matched by movements in consumption growth or by the product of consumption growth and returns underlying the conditional covariance. The results give pretty sensible parameter estimates; small coefficients of risk aversion γ and discount factors less than one. However, the standard errors on the risk aversion coefficients are pretty large, and the estimates are not that stable across specifications.

The problem – or rather the underlying fact – is that the instruments used here – lags of consumption and returns – don't forecast either consumption growth or returns very well.

Consumption and stock prices are, in fact, pretty close to random walks, especially when forecast by their own lags. To the extent that these instruments do forecast consumption and returns, they forecast them by about the same amount, leading to risk aversion coefficients near one.

Simplifying somewhat, consider the linearized risk free rate equation,

$$r_t^f = \delta + \gamma E_t\left(\Delta c_{t+1}\right) - \frac{1}{2}\gamma(\gamma+1)\sigma_t^2(\Delta c_{t+1}). \tag{12}$$

If risk premia are not well forecast by these instruments (and they aren't) and consumption is homoskedastic (pretty close) then the main thing underlying these estimates of (11) is whether predictable movements in consumption growth line up with predictable movements in interest rates. The answer for Hansen and Singleton is that they do, with a constant of proportionality (γ) near one. (Hansen and Singleton 1983 study this linearized version of the consumption based model, and their Table 4 studies this interest rate equation explicitly.)

If we take multiple assets, the picture changes however. The middle panel of Table III of Hansen and Singleton (1984) uses one stock and one bond return, and a number of instruments. It finds small, well measured, risk aversion coefficients – but the tests all decisively reject the model. Hansen and Singleton (1983) Table 5, reproduced here, makes the story clear.

Model	γ^*	β^*	Consumption Data	Lags	$\chi^{2\dagger}$	degrees of freedom
1	30.58	1.001	Nondurable	0	Just identified	
	(34.06)	(0.0462)				
2	0.205	0.999	Nondurable	4	170.25	24
					(0.9999)	
3	58.25	1.088	ND &Services	0	Just Identified	
	(66.57)	(0.0687)				
4	0.209	1.000	ND& Services	4	366.22	24
					(0.9999)	

Estimates of the consumption-based model using the value-weighted NYSE return and the Treasury bill return. Lags is the number of lags of consumption growth and returns used as instruments. Source: Hansen and Singleton (1983) Table 5. * Standard errors in parentheses. †Probability values in parentheses.

If we *just* use the unconditional moments – no instruments, the "lags = 0" rows – we find a very large value of the risk aversion coefficient. The covariance of consumption growth with stock returns is small, so it takes a very large risk aversion coefficient to explain the large mean stock excess return. This finding is the equity premium in a nutshell. (Using more recent data and the full nonlinear model, the smallest pricing error occurs around

$\gamma = 50$, but there is *no* choice of γ that sets the moment to zero, even though the model is just identified.) The β slightly greater than one is the risk free rate puzzle. The data are monthly, so even a β slightly greater than one is puzzling.

If we use instruments as well, then the estimate is torn between a small value of γ to match the roughly one-for-one movement of predicted consumption growth and returns (using past consumption growth and returns as predictors) and the very large value of γ necessary to explain the equity premium. Efficient methods weight cries from different parties by their statistical significance. Here, the moments corresponding to predictable movements are better measured, so the estimate of γ is close to those values. But the test statistic gives a huge rejection, here as in Hansen and Singleton (1984). That huge test statistic tells us that there is a tension over the value of γ. The value that makes sense of the equity premium (unconditional returns) is much larger than the value that makes sense of the conditional moments (forecasted returns vs. consumption growth), so one set of moments or pricing errors is left very large in the end.

The fact that quite high risk aversion required to digest the equity premium is robust in consumption-based model estimation. The parameter needed to understand the behavior of a single asset over time, and in particular to line up variation in expected consumption growth with variation in interest rates, is less certain. This number, (or more precisely its inverse, how much consumption growth changes when interest rates go up 1%) is usually called the *intertemporal substitution elasticity* since it captures how much people are willing to defer consumption when presented with a large return opportunity. While Hansen and Singleton found numbers near one, Hall (1988) argued the estimate should be closer to zero, i.e. a very high risk aversion coefficient here as well. Hall emphasizes the difficulties of measuring both real interest rates and especially consumption growth.

More recent literature has tended to side with Hall (Campbell 2003 gives an excellent summary with estimates.) Real interest rates have moved quite a bit and slowly over time, especially in the period since the early 1980s when Hansen and Singleton wrote. Thus, there is a good deal of predictable variation in real interest rates. After accounting for time aggregation and other problems, consumption growth is only very poorly predictable. Lining up the small movements in expected consumption growth against large movements in real interest rates, we see a small intertemporal substitution elasticity, or a large risk aversion coefficient. At least now both moments consistently demand the same puzzlingly high number!

The following 20 years have seen an enormous amount of effort aimed at the consumption-based model. There are of course all sorts of issues to address. What utility function should one use? How should one treat time aggregation and consumption data? How about multiple goods? What asset returns and instruments are informative? Asset pricing empirical work has moved from industry or beta portfolios and lagged returns and consumption growth to forecast to size, book/market and momentum portfolios and dividend price ratio and terms spreads to forecast. How does the consumption-based model fare against this higher bar?

As I see it, there were 10 years of depressing rejection after rejection, followed by 10 years of increasing success. This is heartening. At some level, the consumption-based model must be right if economics is to have any hope of describing stock markets. The data may be poor enough that practitioners will still choose “reduced form” financial models, but economic understanding of the stock market must be based on the idea that people fear stocks, and hence do not buy more despite attractive returns, because people fear that stocks will fall in “bad times” – and at some point we must be able to measure bad times by people’s decision to cut back on consumption.

New utility functions

Given problems with the consumption-based model, the most natural place to start is by questioning the utility function. Functional form is not really an issue, since linearized and nonlinear models already behave similarly. Different arguments of the utility function are a more likely source of progress. Perhaps the marginal utility of consumption today depends on variables other than today’s consumption.

To get this effect, the utility function must be non-*separable.* If a utility function is separable, $u(c,x) = v(c) + w(x)$, then $\frac{\partial u(c,x)}{\partial c} = v'(c)$ and x does not matter. This is the implicit assumption that allowed us to use only nondurable consumption rather than total consumption in the first place. To have marginal utility of consumption depend on something else, we must have a functional form that does not add up in this way, so that $\partial u(c,x)/\partial c$ is a function of x, too.

The first place to look for nonseparability is *across goods.* Perhaps the marginal utility of nondurable consumption is affected by durables, or by leisure. Also, business cycles are much clearer in durables purchases and employment, so business-cycle risk in stock returns may correlate better with these variables than with nondurable and services consumption. One problem with this generalization that we don’t have much intuition for way should the effect go. If you work harder, does that make a TV more valuable as a break from all that work, or less valuable since you have less time to enjoy it? Thus, will you believe an estimate that relies strongly on one or the other effect?

We can also consider nonseparability *over time.* This was always clear for durable goods. If you bought a car last year, it still provides utility today. One way to model this nonseparabitlity is to posit a separable utility over the services, and a durable goods stock that depreciates over time;

$$U = \sum_t \beta^t u(k_t); \;\; k_{t+1} = (1-\delta)k_t + c_{t+1}.$$

This expression is equivalent to writing down a utility function in which last year’s purchases give utility directly today,

$$U = \sum_t \beta^t u\left(\sum_{j=0}^{\infty}(1-\delta)^j c_{t-j}\right).$$

If $u(\cdot)$ is concave, this function is nonseparable, so marginal utility at t is affected by consumption (purchases) at $t-j$. At some horizon, all goods are durable. Yesterday's pizza lowers the marginal utility for another one today.

Following this line also leads us to thinking about the opposite direction: habits. If good times lead people to acquire a "taste for the good life," higher consumption in the past might *raise* rather than lower the marginal utility of consumption today. A simple formulation is to introduce the "habit level" or "subsistence level" of consumption x_t, and then let

$$U = \sum_t \beta^t u(c_t - \theta x_t);\ x_t = \phi x_{t-1} + c_t$$

or, directly,

$$U = \sum_t \beta^t u\left(c_t - \theta \sum_{j=0}^{\infty} \phi^j c_{t-j}\right).$$

Again, you see how this natural idea leads to a nonseparable utility function in which *past* consumption can affect marginal utility today.

Finally, utility could be nonseparable *across states of nature.* Epstein and Zin (1991) pioneered this idea. The expected utility function adds over states, just as separable utility adds over goods,

$$Eu(c) = \sum_s \pi(s) u\left[c(s)\right]$$

Epstein and Zin propose a recursive formulation of utility

$$U_t = \left\{\left((1-\delta)c_t^{\frac{1-\gamma}{\theta}} + \delta E_t\left(U_{t+1}^{1-\gamma}\right)^{\frac{1}{\theta}}\right)\right\}^{\frac{\theta}{1-\gamma}}. \tag{13}$$

(I use Campbell's 2003 notation.) The lack of linear aggregation makes the utility function non-state-separable.

One celebrated effect of the Epstein-Zin formulation is that it separates the coefficient of risk aversion γ from the elasticity of intertemporal substitution ψ , defined by $\theta = (1-\gamma)/(1-\frac{1}{\psi})$. It's not clear from the above discussion that this modification is vital to understanding the data, but it is nonetheless a clean way to make this distinction. Models with non-*time* separable utilities (habits, durables) also distinguish risk aversion and intertemporal substitution, but not in such a simple way.

A second celebrated effect of Epstein-Zin utility is that the first order condition can be expressed in a way that both current consumption and the wealth portfolio return enter in the marginal utility of wealth. This effect provides a route to including stock returns in the asset pricing model as well as consumption growth. However, this modification stands a bit on shaky ground: the substitution only works for the entire wealth portfolio, not the stock market return alone.

Empirics with new utility functions.

Eichenbaum, Hansen and Singleton (1988) is an early paper that combined nonseparability over time and across goods. They used a utility function (my notation)

$$U = \sum \beta^t \frac{\left(c_t^{*\theta} l_t^{*1-\theta}\right)^{1-\gamma} - 1}{1-\gamma};$$

$$c_t^* = c_t + \alpha c_{t-1}$$

$$l_t^* = l_t + b l_{t-1} \text{ or } l_t^* = l_t + b \sum_{j=0}^{\infty} \eta^j l_{t-j}$$

where l denotes leisure. However, they only test the model on the Treasury bill return, not the equity premium or certainly not the Fama-French portfolios. They also focus on parameter estimates and test statistics rather than pricing errors. Clearly, it is still an open and interesting question whether this extension of the consumption-based model can address what we now understand are the interesting questions.

Eichenbaum and Hansen (1990) investigate a similar model with nonseparability between durables and nondurables. This is harder because one needs also to model the relation between observed durable purchases and the service flow which enters the utility function. Finally, any model with multiple goods gives rise to an in*tra* temporal first order condition, for example marginal utility of nondurables / marginal utility of durables = relative price. This condition can be nonstochastic (no error term) which will cause problems in actual data in which there is no nonstochastic combination of any two economic time series. Eichenbaum and Hansen solve both problems. However, they again only look at consumption and interest rates, leaving open how well this model does at explaining our current understanding of cross-sectional risk premia.

Epstein and Zin (1991) is the classic empirical investigation of preferences that are non-separable across states. Ambitiously, for the time period, they actually have some cross section of returns: they use five industry portfolios. The instruments are lags of consumption and market returns. But industry portfolios don't show much variation in expected returns to begin with, and we now know that variables such as D/P and consumption/wealth have much more power to forecast returns. How these preferences work in a consumption-based estimation with a more modern setup has yet to be investigated.

Ferson and Constantinides (1991) took the lead in estimating a model with temporal nonseparabilities. One has to face parameter profusion in such models; they do it by limiting the nonseparability to one lag, so the utility function is

$$u(c_t - bc_{t-1}). \tag{14}$$

This is one of the first papers to include an interesting cross section of assets, including the market (equity premium) and some size portfolios, along with a modern set of instruments, including dividend/price ratio and t bill rate, that actually forecast returns. However, much of the model's apparently good performance comes down to larger standard errors rather than smaller pricing errors.

Heaton (1993, 1995) considers the joint effects of time aggregation, habit persistence and durability on the time series process for consumption and on consumption-based asset pricing models. The 1993 paper focuses on consumption, showing how the random walk in consumption that occurs with quadratic utility and constant real rates is replaced by interesting autocorrelation patterns with time aggregation, habit persistence, and durability. Heaton (1995) then integrates these ideas into the specification of consumption-based asset pricing models, not an easy task. In particular, Heaton gives us a set of tools with which to address time-aggregation, and Campbell and Cochrane (2000) argue in a simulation model that time-aggregation helps a lot to explain consumption-based model failures. Sensibly, Heaton finds signs of both durability and habit persistence, with durability dominating at short horizons and habit persistence at longer horizons. However, he only considers the value-weighted stock market and T-bill rate as assets.

Campbell and Cochrane

To represent the new utility function literature generally and the habit persistence (temporal nonseparability) literature specifically in this collection, I include Campbell and Cochrane (1999). We replace the utility function $u(C)$ with $u(C-X)$ where X denotes the level of habits.

$$E\sum_{t=0}^{\infty}\delta^t\frac{(C_t-X_t)^{1-\gamma}-1}{1-\gamma}.$$

Habits move slowly in response to consumption. The easiest specification would be an AR(1),

$$X_t=\phi X_{t-1}+\lambda C_t. \tag{15}$$

(Small letters denote the logs of large letters throughout this section, $c_t=\ln C_t$, etc.) This specification means that habit can act as a "trend" line for consumption; as consumption declines relative to the "trend" in a recession, people will become more risk averse, stock prices will fall, expected returns will rise, and so on.

The idea is not implausible (well, not to Campbell and myself, at least). Anyone who has had a large pizza dinner or smoked a cigarette knows that what you consumed yesterday can have an impact on how you feel about more consumption today. Might a similar mechanism apply for consumption in general and at a longer time horizon? Perhaps we get used to an accustomed standard of living, so a fall in consumption hurts after a few years of good times, even though the same level of consumption might have seemed very pleasant if it arrived after years of bad times. This thought can at least explain the perception that recessions are awful events, even though a recession year may be just the second or third best year in human history rather than the absolute best. Law, custom and social insurance also insure against *falls* in consumption as much or more than as low *levels* of consumption.

In fact, we specify a nonlinear version of (15). This nonlinear version allows us to avoid an Achilles heel of many habit models, huge variation in interest rates. When consumers have habits, they are anxious in bad times (consumption close to habit) to borrow against coming good times (consumption grows away from habit). This anxiousness results in a high interest rate, and vice versa in good times. The nonlinear version of (15) allows us to offset this "intertemporal substitution" effect with a "precautionary savings" effect. In bad times, consumers are also more risk averse, so rather than *borrow* to push consumption above habit today, they *save* to make more sure that consumption does not fall even more tomorrow. The nonlinear version of (15) allows us to control these two effects. In Campbell and Cochrane (1999) we make the interest rate constant. The working paper version (Campbell and Cochrane 1995) showed how to make interest rates vary with the state and thus create an interesting term structure model with time-varying risk premia.

Following in the tradition of Constantinides (1990), Abel (1990), and Sundaresan's (1989) habit persistence investigations, we examined the model's behavior by a combination of simulation and simple moment-matching rather than a full-blown estimation on an interesting cross-section of portfolios. We let aggregate consumption follow a random walk, calibrated the model to match sample means including the equity premium, and then compared the behavior of time-series tests in our artificial data. The model matches the time-series facts mentioned above quite well. In particular, the dividend/price ratio forecasts stock returns, and variance decompositions find all variation in stock prices is due to changing expected returns, while none is due to changing cash flow forecasts.

In this model, the marginal rate of substitution – growth in the marginal value of wealth or discount factor – between dates t and $t+k$ depends on consumption growth *and also* on the change in the ratio of consumption to habit,

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \left(\frac{S_{t+1}}{S_t} \right)^{-\gamma}. \tag{16}$$

where $S = (C - X)/C$ and X is habit. As the time period lengthens, the latter effect becomes more important. It thus answers the classic question: why do people fear stocks so much? The answer is not so much that they fear that stocks will decline when consumption is *low* in absolute terms (C); the answer is that they fear stocks will be temporarily low in future "bad times," recessions, times when consumption falls low relative to habits (S).

Additional habit and related models

Wachter (2004) extended the habit model to think seriously about the term structure of interest rates, in particular adding a second shock and making a quantitative comparison to the empirical findings of the term structure literature such as Fama and Bliss' (1987) finding that forward-spot spreads forecast excess bond returns. Buraschi and Jiltsov (2005) construct a model with monetary frictions in which there are inflation-risk premia and a difference between real and nominal bonds.

Verdelhan (2004) extends the habit model to foreign exchange premia. Here the puzzle is that high foreign interest rates relative to domestic interest rates lead to higher returns in foreign bonds, even after currency risk. This is the celebrated "foreign exchange risk premium" puzzle. His explanation is natural. First, in complete markets the exchange rate *is* simply determined by the ratio of foreign to domestic marginal utility growth. Hence, the first part of the puzzle, why should consumption covary with exchange rate risk, is perfectly natural. The second part of the puzzle is, why should this risk vary over time. Here, the habit model answers the question very simply. Recessions, times when consumption is close to habit, are times of low interest rates (as in Wachter's model and as in the data), and also times of high risk premium (people are more risk averse when consumption is near habit.) Voilá, the interest rate spread forecasts a time-varying exchange rate risk premium. More generally, these papers pave the way to go beyond equity, value, size and momentum premiums to start thinking about bond risk premia and foreign exchange risk premia.

On issue is whether one really must believe in habits, or whether other mechanisms can deliver the same slow-moving, business-cycle-related variation in risk aversion. Piazzesi, Schneider and Tuzel (2004) investigate a related model in which the share of housing consumption in total consumption plays much the same role as habits. They specify that utility is nonseparable between non-housing consumption and consumption of housing services; you need a roof to enjoy the new TV. Thus, the marginal rate of substitution is

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\sigma}} \left(\frac{\alpha_{t+1}}{\alpha_t} \right)^{\frac{\varepsilon-\sigma}{\sigma(\varepsilon-1)}}$$

Here, α is the expenditure share of housing services as a fraction of housing services plus other nondurable and services consumption. As the share of housing services varies slowly over the business cycle, you can see the analogy to the habit model (16). Furthermore, the housing share is observed, where we infer habits from the history of consumption. Housing services are part of the usual nondurable and services aggregate of course; the paper essentially questions the usual price indices used to aggregate housing services into overall services.

A nice feature of this paper is that they calibrate the elasticity of substitution between housing services and consumption ε from the behavior of the share and relative prices; the static first order condition. If $\varepsilon = 1$, the share of housing is the same for all prices. They find that $\varepsilon = 1.27$. when housing prices rise, the quantity falls enough that the share of housing expenditure actually falls slightly. For Cobb-Douglas utility, $\varepsilon = 1$, shares are constant; the relative stability of shares argues against extreme values of ε. However, As the discount factor above shows, whether the housing share enters positively or negatively in marginal utility ($\partial u/\partial c\partial$house $>$ or < 0) depends on the relative subsitutability of houses for other consumption goods ε and the substitutability of consumption over time and states, σ. Like others, they calibrate to a relatively large risk premium, hence small σ. This means that housing share enters positively in the marginal rate of substitution.

Most of Piazzesi, Schneider and Tuzel's empirical work also consists of a simulation model. They use an i.i.d. consumption growth process, and fit an AR(1) to the housing share. They then simulate artificial data on the stock price as a levered claim to consumption, and look at its properties. Expected returns are high, matching the equity premium, because investors are afraid that stocks will fall when α is low in recessions. (They also document the correlation between α and stock returns in real data). Interest rates are low, both from a precautionary savings effect due to the volatility of α and due to the mean α growth. Interest rates vary over time, since α moves slowly over time and there are periods of predictable α growth. Variation in the conditional moments of α generates a time varying risk premium. Thus, the model generates returns predictable from price-dividend ratios. The model also generates returns predictable from the housing/consumption ratio α. This turns out to work in the data. This adds to the macro variables (I/K, cay) known to forecast returns, and this one has a good economic reason. (See Table 4 and Table 5). Finally, the model generates the right, slow-moving, variation in price-dividend ratios and stock return volatility, all coming from risk premia rather than dividend growth.

Obviously, all of these models need to be evaluated in a modern cross-sectional setting, for example in the Fama French 25 size and book/market portfolios. Time aggregation makes this a bit difficult, as in other consumption-based models, but again Heaton (1995) gives us the tools to address this issue. Surprisingly, noone has tried this. The closest effort is Chen and Ludvigson (2004). They evaluate and test a habit model similar to Campbell and Cochrane's, and they examine the Fama-French 25 size and book/market portfolios. However, they use a "nonparametric" (really, highly parametric) version of the single-lag habit specification (14) rather than the slow-moving counterpart (15).

The return of consumption-based models

Recently, several researchers have gone back to the basic consumption-based model, and found that it does in fact contain some important germs of truth. Alas, much of this work is so fresh that it has not yet been published, so cannot be included in this volume.

Nobody expects the consumption-based model (and data) to work at high frequencies. Not everybody pays enough attention to calibrate purchasing an extra cup of coffee against the last hour's stock returns. Even if consumers act "perfectly" (i.e. ignoring all transaction, information, etc. costs), high-frequency *data* are unreliable. If Δc_t and r_t are perfectly correlated but independent over time, a one period timing error, in which you mistakenly line up Δc_{t-1} with r_t will show no correlation at all. The methods for collecting quantity data are just not attuned to getting high-frequency timing just right, and the fact that returns are much better correlated with macro variables one or two quarters later than contemporaneous is suggestive. In sum, at some high frequency, we expect consumption and returns to be de-linked. Conversely, at some low enough frequency, we know consumption and the value of stocks must move one for one; both must eventually track the overall level of the economy. Thus, some form of the consumption model should hold if economics has anything to say about the stock market. Following this intuition, a number of authors have found germs of truth in long-run relations between consumption and returns.

Daniel and Marshall (1997) showed that consumption growth and asset returns become more correlated at longer frequencies. They don't do a formal estimation, but they do conclude that the equity premium is less of a puzzle at longer frequencies.

Parker and Julliard (2005) examine whether size and book/market portfolios can be priced by their exposure to *long-run* consumption risk. Specifically, they examine whether a multiperiod return formed by investing in stocks for one period and then transforming to bonds for k-1 periods is priced by k period consumption growth. They study the the multiperiod moment condition

$$1 = E_t \left[\beta^k \left(\frac{C_{t+k}}{C_t} \right)^{-\gamma} R_{t+1} R^f_{t+1} R^f_{t+2}..R^f_{t+k-1} \right].$$

They argue that this moment condition is robust to measurement errors in consumption and simple "errors" by consumers. For example, they argue that if consumers adjust consumption slowly to news, this moment will work while the standard one will not. Hansen Heaton and Li (2004b) show that recursive utility of the Epstein-Zin variety produces a model in which asset returns at date $t+1$ are priced by their exposure to such "long-run" consumption risk. Parker and Julliard find that this model accounts for the value premium. Returns at date $t+1$ forecast subsequent consumption growth very slightly, and this forecastability accounts for the results. In addition to selecting one of many possible long run moment conditions, Parker and Julliard leave the moment condition for the level of the interest rate out, thus avoiding equity premium puzzles.

Bansal and Yaron (2005) also argue that average returns of value vs. growth stocks can be understood by different covariances with long-run consumption growth. This paper is very interesting because they examine long-run covariances of *earnings* with consumption, rather than *returns.* I think moving to earnings as an exogenous variable rather than tomorrow's price (β) is an important direction for asset pricing. However, Hansen, Heaton and Li (2004b) show that Bansal and Yaron's evidence that value stocks have much different long-run-consumption-betas than do growth stocks depends crucially on whether one includes a time trend in the regression of earnings on consumption. In the data, earnings and consumption move about one for one, as one might expect. With a time trend, a strong time trend and a strong opposing regression coefficient offset each other.

Yogo (2004) reconsiders nonseparability across *goods* by looking again at durable goods. He examines the utility function

$$u(C, D) = \left[(1-\alpha) C^{1-\frac{1}{\rho}} + \alpha D^{1-\frac{1}{\rho}} \right]^{\frac{1}{1-\frac{1}{\rho}}}.$$

He embeds this specification in an Epstein-Zin aggregator (13) over time. This framework allows Yogo to use quite high risk aversion without the implication of wildly varying interest rates. Following tradition in the Epstein-Zin literature, he uses the market portfolio return to proxy for the wealth portfolio or utility index which appears in the marginal rate of substitution.

Estimating the model on the Fama-French 25 size and book/market portfolios, along with the 3 month T bill rate, and including the intra-temporal first order condition for durables vs. nondurables, he estimates high ($\gamma = 191;\ 1/\gamma = 0.005$) risk aversion, as is nearly universal in models that account for the equity premium; a larger elasticity of intertamporal substitution $\sigma = 0.024$ to explain a low and relatively constant interest rate, and a modest 0.54 - 0.79

(depending on method) elasticity of substitution between durables and nondurables. As in the discussion of Piazzesi, Schneider and Tuzel above, the difference between this modest elasticity and the much smaller σ and $1/\gamma$ means that the nonseparabilities matter, and durables do affect the marginal utility of consumption.

Yogo linearizes this model giving a discount factor linear in consumption growth, durable consumption growth, and the market return

$$m_{t+1} \approx a - b_1 \Delta c_{t+1} - b_2 \Delta d_{t+1} - b_3 r_{Wt+1}$$

This linearized model prices the Fama French 25 portfolios (except the small growth portfolio, left out of many studies) with a large cross-sectional R^2 and a beautiful plot of average returns vs. model predictions The implied preference parameters (from b_i estimates) are consistent with those from the nonlinear model. By linearizing, Yogo is able to display that there is a substantial spread in betas, addressing the concern that a model prices well by an insignificant spread in betas and a huge risk premium. Yogo also shows some evidence that variation in *conditional* mean returns lines up with varying *conditional* covariances on these three factors.

Pakos (2004) also considers durables vs. nondurables, using the nonlinear specification, dealing with the intra-temporal first order condition (durable vs. nondurable and their relative price), and considering the level of the interest rate as well as the equity premium and the Fama-French 25 portfolios. Pakos needs an extreme unwillingness to substitute durable for nondurable consumption in order to make quantitatively important differences to asset pricing. To keep the durable vs. nondurable first order condition happy, given the downward trend in the ratio of durables to nondurables, he adds an income elasticity (non-homothetic preferences).

Lustig and Verdelhan (2004) do a standard consumption-beta test on foreign exchange returns, and find, suprisingly, that the standard consumption based model works quite well. One of their clever innovations is to use portfolios, formed by going in to high interest rate countries and out of low interest rate countries. As in the rest of asset pricing, portfolios can isolate the effect one is after and can offer a stable set of returns.

Most strikingly, Jagannathan and Wang (2005) have found that by using fourth quarter to fourth quarter nondurable and services consumption, the simple consumption based model (together with an admittedly rather high risk aversion coefficient and discount factor) can account for the Fama-French 25 size and book/market portfolios. This is a natural result. "Nondurables" include things like shirts. A lot of such purchases happen at Christmas, and with an annual planning horizon. And it is a stunning result: the simple power utility consumption based model *does* work quite well after all, at least for one horizon (annual).

In all these cases, I have pointed out the limitations, including specializations and linearizations of the models, and selection of which moments to look at and which to ignore. This is progress, not criticism. We've already rejected the model taken literally, i.e. using arbitrary assets, instruments, and monthly data; there is no need to do that again. But we learn something quite valuable from knowing *which* assets, horizons, specifications, and instruments *do* work.

Consumption as a factor model; Lettau and Ludvigson

As one example of a recent paper that finds some success, as well as a paper that illustrates well current trends in how we evaluate models, I include Lettau and Ludvigson's (2001) "Resurrecting the (C)CAPM."

At a modeling level, this paper follows in the tradition started by Breeden, Gibbons and Litzenberger (1989), who examine a linearized version of the consumption-based model, a form more familiar to financial economists. Breeden, Gibbons and Litzenberger ask whether average returns line up with betas computed relative to consumption growth, they correct for a number of problems with consumption data, and they use a set of industry portfolios. They find the consumption-based model does about as well as the CAPM. This work, along with Breeden (1979) and other theoretical presentations, was important in bringing the consumption-based model to the finance community. Breeden emphasized that consumption should stand in for *all* of the other factors including wealth, state variables for investment opportunities, non-traded income, and so forth that pervade finance models.

However, more recent empirical research has raised the bar somewhat: industry portfolios show much less variation in mean returns than size- and book/market portfolios that dominate cross-sectional empirical work. In addition, the fact that variables such as the dividend price ratio forecast returns much better than lagged returns, and a realization that weak instruments can cause problems, has led to a style that one lag of important variables rather than many lags of returns and growth rates as in Hansen and Singleton are used as instruments.

Lettau and Ludvigson examine a *conditional* version of the consumption CAPM. In our notation, they specify that the stochastic discount factor or growth in marginal utility of wealth is

$$m_{t+1} = a + (b_0 + b_1 z_t) \times \Delta c_{t+1}$$

They also examine a conditional CAPM,

$$m_{t+1} = a + (b_0 + b_1 z_t) \times R^w_{t+1}$$

The innovation is to allow the slope coefficient, which acts as the risk-aversion coefficient in the model, to vary over time. They use the consumption-wealth ratio to measure z_t. (Recall that the consumption/wealth ratio forecasts stock returns.)

Recast in the language of traditional finance, this specification is equivalent to a factor model in which both betas and factor risk premia vary over time,

$$E_t(R^{ei}_{t+1}) = \beta_{i,\Delta c,t} \lambda_t.$$

Though consumption is the only factor, the *unconditional* mean returns from such a model can be related to an *unconditional* multiple-factor model, in which the most important additional factor is the product of consumption growth and the forecasting variable,

$$E(R^{ei}_{t+1}) = \beta_{i,z_t}\lambda_1 + \beta_{i,\Delta c_{t+1}}\lambda_2 + \beta_{i,(z_t\times\Delta c_{t+1})}\lambda_3.$$

(See Cochrane 2004 for a derivation.) Thus, a *conditional* one-factor model may be behind empirical findings for an *unconditional* multi-factor model.

Figure 1 in the paper makes a strong case for the performance of the model. Including the scaled consumption factor, they are able to explain the cross-section of 25 size and book to market portfolios about as well as does the Fama-French three-factor model. A model that uses labor income rather than consumption as a factor does almost as well.

This is a tremendous success. This was the first paper to even try to price the value effect with macroeconomic factors. To get anywhere at all would have been great; to even compete with Fama and French's ad-hoc portfolio-based factors is fantastic. This paper also set a style for many that followed: evaluate a model by pricing the Fama-French 25 size and book to market portfolios, and present the results in graphical form of actual mean returns vs. model predictions. (Cochrane 1996 also uses this kind of graph, but only tries out size portfolios.) We now are focusing on the pricing errors themselves, and less on whether a test statistic formed by a quadratic form of pricing errors is large or small by statistical standards. A "rejected" model with 0.1% pricing errors is a lot more interesting than a "non-rejected" model with 10% pricing errors, and the pattern of pricing errors across portfolios is revealing. Fama and French (for example, 1996) also encouraged this shift in attention by presenting average returns and pricing errors across portfolios, but in tabular rather than graphical format.

Limitations

By now, so many papers have found high cross-sectional R^2 in the Fama-French 25 portfolios using ad-hoc macro models (m = linear functions of macro variables with free coefficients), that pushing this line further seems to me not particularly useful.

First, the cross-sectional R^2 rises automatically as we add factors. With (say) 10 factors in 25 portfolios, a high sample R^2 is not that surprising. Also, many authors have also pointed out (Roll and Ross 1994, Kandel and Stambaugh 1995) that the cross-sectional R^2 and the corresponding visual look of plots like Lettau and Ludvigson's Figure 1 are not invariant to portfolio formation. We can take linear combinations of the original portfolios to make the plots look as good or as bad as we want. Finally, cross-sectional R^2 depends a lot on the estimation method. R^2 is only well-defined for an OLS cross-sectional regression of average returns on betas with a free intercept. For any other estimation technique – techniques that do not maximize R^2 – various ways of computing R^2 give wildly different results[7]. Both criticisms are of course solved by statistical measures; test statistics based on

[7] In a regression $y = a + xb + \varepsilon$, identities such as

$$R^2 = \frac{var(xb)}{var(y)} = 1 - \frac{var(\varepsilon)}{var(y)} = \frac{var(xb)}{var(xb) + var(\varepsilon)}$$

$\alpha' cov(\alpha, \alpha')^{-1}\alpha$ where α is a vector of pricing errors are are invariant to portfolio formation and take account of degrees of freedom. However, one can respond that the original portfolios are the interesting ones; the portfolios that modify R^2 a lot have unnatural and large long-short positions. In addition, we certainly don't want to go back to the old days of simply displaying p values and ignoring these much more revealing measures of model fit. Surely the answer is to present both formal test statistics and more revealing diagnostics.

More importantly, the core of Fama and French's finding is that the 25 size and book-market portfolios all *move together*. The time-series R^2 of the Fama French 3 factor model is very high. If any model prices the value-growth (HML) portfolio and the size (SMB) portfolio, then it must do exactly as well as the Fama-French model in pricing the cross-section of 25 portfolios. (There is little variation in market betas, so that just adds a constant to the cross-section.) So the hurdle is really only to price these *two* portfolios, not the 25. Now, macro-based models with multiple factors and free parameters start to seem a lot less exiting.

Furthermore, if we are going to go past the goal of "doing as well as the Fama-French 3 factor model in the Fama-French 25 portfolios" and moving on to the goal of "do better than Fama-French in pricing these portfolios" that means pricing Fama and French's failures. The Fama French model does not do well on small growth and large value stocks. Any model that improves on the Fama-French cross-sectional R^2 does so by better pricing the small growth/large value stocks. But is this phenomenon real? Is it interesting?

Lettau and Ludvigson and many related macro models also suffer from the fact that the factors are much less correlated with asset returns than are factor portfolios. The time-series R^2 are necessarily lower, so test results can depend on a few data points (Menzly 2001). This isn't a defect; it's exactly what we should expect from a macro model. But it does make inference less reliable. Lewellen and Nagel (2004) have also criticized the model for having too small a spread in betas; this means that the factor risk premia are unreliably large and the spread in betas may be spurious.

These comments should not be taken as criticism. Lettau and Ludvigson's paper is an excellent contribution; nobody imagined it could be done. All I'm saying is that more work just like this is probably not useful.

What next, then?

Many people have the impression that consumption-based models were tried and failed. I hope this review leaves exactly the opposite impression. Despite 20 years of effort, the consumption-based model has barely been tried. In preparing this review, I could not find any reference that systematically examined consumption and asset return data, showed which moments work (which assets, horizons, instruments, etc.) and which ones don't, even for the basic power utility model. As Parker and Julliard or Jagannathan and Wang hint, the model can do a very good job on some moments but not on others; this is still a rejection of the hypothesis that it handles all moments perfectly, but it is interesting to know that there are

only hold when b is the OLS estimate.

some successes.

The playing field for empirical work has changed since the classic investigations of non-separable utility functions. It is now routine to check a model in the size- book/market cross section. If any instruments at all are used, we focus on powerful instruments known to forecast returns. We worry about time aggregation. Above all, we focus on pricing errors rather than p values, as exemplified by Fama-French style tables of mean returns, betas, and alphas across portfolios, or by equivalent plots of actual mean returns vs. predicted mean returns.

This is part of a larger, dramatic, and quite unheralded change in the style of empirical work in finance. The contrast between, say, Hansen and Singleton (1983) and Fama and French (1996), each possibly the most important asset pricing paper of its decade, could not be starker. Both models are formally rejected. But the Fama and French paper persuasively shows the dimensions in which the model *does* work; it shows there is a substantial and credible spread in average returns to start with (not clear in many asset pricing papers), it shows how betas line up with average returns, and how the betas make the pricing errors an order of magnitude smaller than the average return spread. In the broader scheme of things, much of macroeconomics has gone from "testing" to "calibration" in which we examine *economically interesting* predictions of models that are easily statistically rejected (though the "calibration" literature's resistance to so much as displaying a standard error is a bit puzzling.)

Yet with the lone exception of Yogo (2004) and Pakos (2004) work on durables, we don't know how any of the utility function variants work in a contemporary setting. And it's time to connect the empirical success of relatively ad-hoc macro models such as Lettau and Ludvigson (as well as those on the investment side like Cochrane 1996 and Li, Vassalou, and Xing, 2003, or Vassalou's 2003 observation that mimicking portfolios that forecast GDP price the Fama-French 25) to the more specific economic foundations of the earlier models.

Though my comments may have seemed critical, they are not. We cannot expect authors 20 years ago to do things as we would today. The point of my comments is that there is still very much to do to understand where the consumption-based model works, where it doesn't, and how it might be improved.

6 Production, Investment and General Equilibrium

If we want to link asset prices to macroeconomics, *consumption* seems like the weakest link. Aggregate nondurable and services consumption is about the smoothest and least cyclical of all economic time series. Macroeconomic shocks are seen in output, investment, employment and unemployment, and so forth.

Consumers themselves are a weak link. One has to exploit consumer first order conditions judiciously. For example, a one-month delay in adjusting consumption would destroy a test in monthly data, yet would have trivial utility costs; and thus this prediction would be quite sensitive to small transaction and information costs. (See Cochrane 1989 for some calculations). This criticism is not true of the equity premium of course; failing ever to invest in stocks carries a big utility cost, but many high-frequency timing decisions are sensitive.

"Production-based asset pricing"

These thoughts led me to want to link asset prices to *production* through *firm* first-order conditions in Cochrane (1991), which I include. This approach should allow us to link stock returns to genuine business cycle variables, and firms may do a better job of optimization (more precisely, information and transactions cost frictions which we abstract from may be less important for firms). That paper essentially revives the Q theory of investment by showing that it works pretty well in first-differenced form for variation in aggregate investment over time. A production technology defines an "investment return," the (stochastic) rate of return available from investing a little more today and then investing a little less tomorrow. Furthermore, when the production function is constant returns to scale, the investment return should *equal* the stock return, data point for data point. This statement flows quickly from the Q theory. First difference the relation investment$_t = f($stock price$_t)$, and you quickly have investment *growth* proportional to stock price *growth*, and investment *return* equal to stock *return* is only a few lines of algebra away.

The major result of the paper is that investment returns – functions only of investment data – are highly correlated with stock returns. Times of high stock prices *do* correspond to times of high investment. This view bore up well even through the gyrations of the late 1990s. When internet stock prices were high, investment in internet technology boomed. Pastor and Veronesi (2004) show how the same sort of idea can account for the boom in internet IPOs as internet stock prices rose. The formation of new firms responds to market prices much as does investment by old firms.

The Q theory also says that investment should be high when expected returns (the cost of capital) are low, because stock prices are high in such times. The paper confirms this prediction: investment to capital ratios do predict stock returns. (This variable never caught on in the return - forecasting literature, alas.)

There has been a good deal of additional work on the relation between investment and stock returns since then. Lamont (2000) is one nice example. Lamont cleverly uses a survey data set on investment *plans*. The investment plan data are great forecasters of actual investment. Investment plans also can avoid some of the timing issues that make investment data hard to use. If the stock price goes up today, i.e. if the expected return or cost of capital declines, it takes time to plan a new factory, draw the plans, design the machinery, etc., so investment *expenditures* can only react with a lag. Investment *plans* can react almost instantly. Lamont finds that investment plans also forecast stock returns, even better than the investment/capital ratios in Cochrane (1991). Kogan (2004), inspired by a model with irreversible investment (an asymmetric adjustment cost, really) finds that investment forecasts the *variance* of stock returns as well.

I include a second paper in this series as well, Cochrane (1996). Here I used multiple production technologies, and I investigated the question whether the investment returns from these technologies span stock returns, i.e. whether a discount factor of the form

$$m_{t+1} = a + b_1 R^{(1)}_{t+1} + b_2 R^{(2)}_{t+1},$$

would satisfy

$$1 = E(m_{t+1} R_{t+1})$$

for a cross-section of asset returns R_{t+1}. Here $R^{(i)}_{t+1}$ denote the investment returns, functions of investment and capital only, i.e. $R^{(i)}_{t+1}(I^i_{t+1}/K^i_{t+1},\ I^i_t/K^i_t)$. The paper also explores scaled factors and returns to incorporate conditioning information, (though Cochrane 2004 does a better job of summarizing this technique) and the paper uses plots of predicted vs. actual mean returns to evaluate the model.

One important limitation is that I only considered size portfolios, not book to market or other sorts. Li, Vassalou, and Xing (2003) found that an extended version of the model with four technological factors does account for the Fama-French 25 size and book/market portfolios, extending the list of macro models that can account for the value effect.

My 1991 and 1996 papers did not quite achieve the stated goal of a "production-based asset pricing model," which links macro variables to asset returns *independently* of preferences. The trouble is that the technologies we are used to writing down allow firms to transform goods across *time*, but not across *states of nature*. We write functions like $y_{t+1}(s) = \theta_{t+1}(s) f(k_t)$ where s indexes states at time $t+1$. More k_t results in more y_{t+1} in all states, but there is no action the firm can take to increase y_{t+1} in one state, reduce it in another state, and leave k_t alone. By contrast, the usual utility function $\sum_s \pi(s) u\left[c(s)\right]$ defines marginal rates of substitution across all dates and states; $mrs_{s_1,s_2} = \left\{\pi(s_1) u'\left[c(s_1)\right]\right\} / \left\{\pi(s_2) u'\left[c(s_2)\right]\right\}$. Production functions are kinked (Leontief) across states of nature, so we cannot read contingent claim prices from outputs as we can read contingent claim prices from state-contingent consumption.

There are three ways out of this, that I hope will be investigated more fully in the future. The dynamic spanning literature in asset pricing naturally suggests two approaches: either allow continuous trading or a large number of underlying technologies. For example, with one field that does well in rainy weather and one that does well in sunshine, a farmer can span [rain, shine] contingent claims. Multiple technologies (many fields) can span more contingencies.

The third approach is that we could directly write technologies that allow marginal rates of transformation across states. Equivalently, we allow the firm to choose the distribution of its technology shock process from some set along with choosing capital (and labor). If the firm's objective is

$$\max_{\{k_t, \theta_{t+1} \in \Theta\}} E\left[m_{t+1} \theta_{t+1} f(k_t)\right] = \sum_s \pi_s m_s \theta_s f(k_t)$$

where m denote contingent claim prices, then the first order conditions with respect to θ_s

will identify m_s.

The technologies we write down, of the form $y_{t+1}(s) = \theta(s)f(k_t)$ are a historical accident. We started writing technologies for nonstochastic models and then tacked on shocks. They did not come from a detailed microeconomic investigation which persuasively argued that firms in fact have absolutely no way to transform output across states of nature, or no choice at all about the distribution of the shocks they face. Cochrane (1993) starts to think about how to represent such technologies, but gets stuck. Both a mathematical exploration of how to represent technologies with a full set of marginal rates of transformation, and a microeconomic investigation of such technologies, would bear fruit in allowing a full-fledged "production-based asset pricing" that does not require any information about consumers.

General Equilibrium

Most efforts to connect stock returns to a fuller range of macroeconomic phenomena have gone the route of constructing general equilibrium models. These models include the consumption-based first order condition but also include a full production side. In a general equilibrium model, we can go through consumers and connect returns to the determinants of consumption, basically substituting decision rules $c(I, Y, ..)$ in $m_{t+1} = \beta u'(c_{t+1})/u'(c_t)$ to link m to I, Y, etc. They also impose much greater discipline, in that quantities such as the consumption process and correlation of tomorrow's consumption with returns must be derived from the model setup rather than simply posited.

While vast numbers of general equilibrium asset pricing models have been written down, I focus here on models that have been specified and used to make a quantitative connection between asset pricing phenomena and macroeconomics, and include a serious specification of production technology. Just about any model can be quickly dressed up to be "general equilibrium," by saying "linear technology" rather than "fixed return process" or "endowment" instead of "model the consumption process."

Jermann

Urban Jermann's (1988) "Asset Pricing in Production Economies" really got this literature going. This paper starts with a standard real business cycle (one sector stochastic growth) model and verifies that its asset-pricing implications are a disaster. Capital can be instantaneously transferred to and from consumption – the technology is of the form $y_t = \theta f(k_t)$; $k_{t+1} = (1-\delta)k_t + (y_t - c_t)$. This means that the relative price of stocks – Q, or the market to book ratio – is always exactly *one*. Stock returns still vary a bit, since productivity θ is random giving random dividends, but all the stock *price* fluctuation that drives the vast majority of real-world return variation is totally absent.

Jermann therefore adds adjustment costs, the same device I used in Cochrane (1991) and that has been a mainstay of the Q theory literature. Now there is a wedge between the price of "installed" (stock market) capital and "uninstalled" (consumption) capital. That wedge increases the larger is investment. This reasonable specification leads to a good deal of equilibrium price variation. Jermann also includes habit persistence in preferences. He finds that both ingredients are necessary to give any sort of match to the data. Without habit persistence, marginal rates of substitution do not vary much at all – there is no equity premium – and expected returns do not vary over time. Without adjustment costs, the habit-persistence consumers can use the production technology to provide themselves very smooth consumption paths. In Jermann's words, "they [consumers] have to care, and they have to be prevented from doing anything [much] about it."

The challenge is to see if this kind of model can match asset pricing facts, while at the same time maintaining if not improving on the real business cycle model's ability to match quantity fluctuations. This is not a small challenge: given a production technology, consumers will try to smooth out large fluctuations in consumption used by endowment economies to generate stock price fluctuation, and the impediments to transformation across states or time necessary to give adequate stock price variation could well destroy those mechanisms' ability to generate business cycle facts such as the relative smoothness of consumption relative to investment and output.

Jermann's model makes progress on both tasks, but leaves much for the rest of us to do. He matches the equity premium and relative volatilities of consumption and output and investment. However, he does not evaluate predictability in asset returns, make a detailed comparison of correlation properties (impulse-responses) of macro time series, or begin work on the cross-section of asset returns.

Jermann also points out the volatility of the risk free rate. This is a central and important problem in this sort of model. Devices such as adjustment costs that raise the variation of marginal rates of substitution *across states*, and hence generate the equity premium, tend also to raise the variation of marginal rates of substitution *over time*, and thus give rise to excessive risk free rate variation. The nonlinear habit in Campbell and Cochrane (1999) is one device for quelling interest rate volatility with a high equity premium; a move to Epstein-Zin preferences is another common ingredient for solving this puzzle. Adding a second linear technology might work, but might give back the excessive smoothness of consumption growth. Perhaps a production analogue to Epstein-Zin preferences can give us a wide variation in marginal rates of transformation *across states of nature* but quite smooth marginal rates of transformation *over time.* In the meantime, we learn that checking interest rate volatility is an important question to ask of any general equilibrium model in finance. As usual, I point out these limitations not as a criticism, but to show how much is still left to do.

Boldrin, Christiano and Fisher

I include Boldrin, Christiano and Fisher (2001) as an excellent example of more recent work in this vein. Obviously, one task is to fit more facts with the model. Boldrin, Christiano and Fisher in particular investigate quantity dynamics. Habit persistence is a dramatic change relative to standard real business cycle models, and one would suspect that it would radically change the dynamics of output, consumption, investment and so forth.

Rather than adjustment costs, Boldrin, Christiano and Fisher have a separate capital-goods production sector with declining returns to scale. This specification has a similar effect: one cannot transform consumption costlessly to capital, so the relative prices of capital (stocks) and consumption goods can vary. They include additional frictions, in particular that labor must be fixed one period in advance. Like Jermann, they include only the one-period habit $c_t - bc_{t-1}$ rather than the more plausible autoregressive habit (15) that can generate long-run return forecastability and slow variation in the price/dividend ratio. They also replicate the equity premium, though again with a bit too much interest rate volatility. The big improvement in this paper comes on the quantity side. They find that the new model actually improves on the standard model in describing quantity dynamics.

Menzly, Santos and Veronesi

Obviously, the range of asset pricing phenomena addressed by this sort of model needs to be expanded. In the aggregate, we need to see time-varying, business-cycle-related expected stock and bond returns, and the equivalent stock price volatility. In the cross-section, we need models that generate size, value (and momentum?) effects in expected returns, the failure of the CAPM, and the comovement of size-and book/market stocks that generates the Fama-French factor model. We need to model the suggestive relation between size, value, and other cross-sectional relations with macroeconomic events. The latter goals require a serious cross-section; we need to get beyond pricing the aggregate consumption stream and model what a "firm" or "industry" is, and what it means technologically to be a "value" vs. a "growth" firm.

I include Menzly, Santos and Veronesi (2004) as an excellent example of this recent line of work. Menzly, Santos and Veronesi specify a long-lived autoregressive habit, which can generate long-horizon return predictability and slow movement of the price/dividend ratio as in Campbell and Cochrane (1999). They then specify multiple technologies. This is a multiple-endowment economy; they model the cashflows of the multiple technologies but not the investment and labor decisions that go behind these cashflows. They specify a very clever model for the *shares* of each cashflow in consumption so that the shares add up to one and the model is easy to solve for equilibrium prices. They generate value and growth effects in cross-sectional average returns *interaction* between the changes in aggregate risk premium and the variation in shares. When a cashflow is temporarily low, the duration of that cashflow is longer than in the opposite case. This makes the cashflow more exposed to the aggregate risk premium, giving it a higher expected return and a lower price.

More models

This is an excellent "multiple-endowment" economy. The obvious next step is to amplify its technological underpinnings, bringing it more in line with the RBC literature and allowing us better to compare quantity predictions including investment with economic data. Berk, Green and Naik (1999), Gomes, Kogan and Zhang (2003) also derive size and book/market effects in general equilibrium models with a bit more explicit, but also fairly stylized, technologies. Zhang (2005) uses a multiple-sector technology of the usual $y = \theta f(k)$ form with adjustment costs and both aggregate and idiosyncratic shocks. Zhang (2004) extends the q theory to cross sections to relate many asset pricing anomalies to firm investment decisions. Gourio (2004) generates book/market effects in an economy with relatively standard technology and finds some interesting confirmation in the data. The technical achievements of these papers should not be overlooked. The minute you have an interesting cross-section of firms, the state-space explodes as each firm's capital stock or investment/capital ratio can now become a state variable.

Many of these papers model the discount factor directly as a function of shocks rather than specify any preferences and thus derive the discount factor from the equilibrium consumption process, in a sort of symmetry with Menzly, Santos and Veronesi's exogenous treatment of investment. All of these models price the cross-section with a conditional CAPM or consumption-CAPM, and none of them yet accounts quantitatively for the failure of the unconditional CAPM nor the emergence of other factors in the cross-section, i.e. the *comovement* of value and small stocks.

Clearly, the integration of asset pricing and macroeconomics is not complete. A model with a fully-fleshed out technology and preferences that reproduces the full range of aggregate and cross-sectional asset pricing facts, while fitting the (joint) dynamics of output, consumption, investment, and labor supply, as well as the joint dynamics of these variables with output has yet to be developed. This project is at a minimum a huge piece of reverse-engineering.

Tallarini

I interpret Tallarini (2000) to go after a deep puzzle. If asset pricing phenomena require such a complete overhaul of equilibrium business cycle models, why did nobody notice all the missing pieces before? Why did a generation of macroeconomists trying to match quantity dynamics alone not find themselves forced to adopt fairly extreme and long-lasting habit persistence in preferences and adjustment costs or other frictions in technology? Of course, one answer, implicit in Boldrin, Christiano and Fisher (2001) is that they should have; that these ingredients help the standard model to match the hump-shaped dynamics of impulse-response functions that real business cycle models have so far failed to match well. But the innovations are pretty extreme compared to the improvement in quantity dynamics.

Tallarini goes after a different possibility, one that I think we should keep in mind; that maybe the divorce between real business cycle macroeconomics and finance isn't that stupid after all. Tallarini adapts Epstein-Zin preferences to a standard RBC model; utility is

$$U_t = \log C_t + \theta \log L_t + \frac{\beta}{\sigma} \log \left[E_t \left(e^{\sigma U_{t+1}} \right) \right]$$

where L denotes leisure. Output is a standard production function with no adjustment costs,

$$Y_t = X_t^{\alpha} K_{t-1}^{1-\alpha} N_t^{\alpha}$$
$$K_{t+1} = (1-\delta) K_t + I_t$$

where X is stochastic productivity and N is labor. The Epstein-Zin preferences allow him to raise *risk aversion* while keeping *intertemporal substitution* constant. As he does so, he is better and better able to account for the equity premium, but the quantity dynamics remain almost unchanged! In Tallarini's world macroeconomists might well not have noticed the need for large risk aversion.

On reflection, this is a natural result. The technology of the standard real business cycle model is almost risk-free. While productivity and hence profits do vary over time a bit, q is always equal to one, so there are no stock price swings. Thus, the decision to consume vs. invest, which drives business cycle dynamics is almost entirely driven by interest rates, or *intertemporal substitution*, not by risk aversion.

Tallarini does not quite establish that "macroeconomists safely go on ignoring finance." First of all, the *welfare costs* of fluctuations rise with risk aversion. Lucas' famous calculation that welfare costs of fluctuations are small depends on small risk aversion. Tallarini's observational equivalence cuts both ways: once risk aversion and intertemporal substitution are separated, *business cycle* facts tell you nothing about *risk aversion.* You have to look to prices for risk aversion, and they say risk aversion, and hence the cost of fluctuations, is large. (See Alvarez and Jermann 2004 for a calculation.)

Second, the equity premium is Tallarini's only asset pricing fact. In particular, he still has a q=1 model, so there is no price variation. Papers that want to match more facts, such as market/book variation, return predictability and cross-sectional value / growth effects are driven to add habits and adjustment costs. Tallarini's technology is nearly riskless, so investment decisions really only depend on interest rates. "Real world" technologies like those in adjustment-cost models are much riskier; if risk premia rise and stock market prices decline, less investment gets done. In this case, investment/consumption decisions will be affected by higher risk premia in the models, and higher risk premia again will affect business cycle facts, as suggested by Boldrin, Christiano and Fisher.

For these reasons, I think that we will not end up with a pure "separation theorem" of quantity and price dynamics. I certainly hope not! But the simple form of the observation given by Tallarini is worth keeping in mind. The spillovers may not be as strong as we think, and we may well be able to excuse macroeconomists for not noticing the quantity implications of ingredients we need to add to understand asset prices and the joint evolution of asset prices and quantities.

Hall

If prices and quantities in standard models and using standard measurement conventions resist lining up, perhaps those models or measurements are at fault. Hall (2001) is a provocative paper suggesting one such line. In thinking about the extraordinary rise of stock values in the late 1990s, we so far have thought of a fairly stable *quantity* of capital multiplied by a large change in the relative *price* of (installed) capital. Yes, there was a surge of measured investment, but the resulting increase in the quantity of capital did not come close to accounting for the large increase in stock market valuations.

Hall pursues a different line. The stock market values profit streams, not just physical capital. A firm is bricks and mortar to be sure, but it is also ideas, organizations, corporate culture and so on. All of these elements of "intangible capital" are crucial to profits, yet they do not show up on the books. Neither does the investment in these forms of "capital" nor the output of the invested goods. Could the explosion of stock values in the late 1990s reflect a much more normal *valuation* of a huge, unmeasured stock of "intangible capital," accumulated from unmeasured "output of intangibles?" Hall pursues this line of thought. He allows for adjustment costs and some variation in the price of installed vs. uninstalled capital, and backs out the size of those costs from investment data and reasonable assumptions for the size of adjustment costs. Needless to say, these are not sufficient, so he finds that the bulk of stock market values in the late 1990s came from a large *quantity* of intangible capital.

Obviously, this is a provocative paper, throwing in to question much of the measurement underlying all of the macroeconomic models so far. It has its troubles – it's hard to account for the large stock market *declines* as loss of "organizational capital," but it bears thinking about. Perhaps the ultimate resolution of the asset pricing/economics puzzle will not lie in the direction of just the right combination of temporal and state nonseparabilities in preferences, and adjustment costs, lags, and irreversibilities in capital accumulation, but in a totally new concept such as "intangible capital" or in much larger than anticipated effects of market frictions.

Final comments

As this survey makes clear, we have only begun to scratch the surface of explicit general equilibrium models – models that start with preferences, technology, shocks, market structure – that can address basic asset pricing and macroeconomic facts. Asset prices measure marginal rates of substitution and/or transformation so they should be very informative about such models, though the reverse-engineering problem is challenging. My remarks are not meant to criticize existing papers, but to show how wide open the field still is for substantial progress.

I remain a bit worried about the accuracy of approximations. Most papers solve their models by making a linear-quadratic approximation about a nonstochastic steady state. But we examine parameterizations with very high risk aversion. In the land of facts, the equity premium of 8% is much larger than the interest rate of 1%, so second moment terms are by no means second order. One would want to expand such an economy around a 9% rate of return, not a 1% rate of return! There is an alternative but less popular approach, exemplified by Hansen (1987). Rather than specify a nonlinear and unsolvable model, and then find a solution by linear-quadratic approximation, it shows you how to write down a linear-quadratic (approximate) model, and then quickly find an exact solution. This technique, emphasized in a large number of papers by Hansen and Sargent, might avoid many approximation and computation issues. Hansen (1987) is also is a wonderful exposition of how general equilibrium asset-pricing economies work, and well worth reading on those grounds alone.

Explicit general equilibrium models place far more discipline on the researcher than is commonly realized. First, you cannot simply take covariances as given. The beta or covariance of tomorrow's price with a state variable is endogenous to a general equilibrium model and often works out in unexpected ways. For example, in the Mehra-Prescott economy a positive autocorrelation of consumption growth and risk premium greater than one conspire to deliver a *negative* equity premium, since they deliver a negative correlation between stock returns and consumption growth. General equilibrium models force us (finally) to stop treating tomorrow's price as an exogenous variable; to focus on *pricing* rather than one period returns.

Second, general equilibrium models require you to answer "where are the shocks?" There is a bit of an unstated schizophrenia, in that q type models essentially posit stable technology, which means shocks to preferences, while consumption models posit stable preferences, which means shocks to technology. The standard q model says q is high if expected future cashflows are high, which does seem consistent with technology shocks. But the fact is that aggregate market q is high when expected returns are low, as q does not forecast aggregate dividend growth or cashflows nearly as well as it forecasts returns. So where did the variation in expected returns come from, preference shocks? One can envision a model in which technology shocks are the ultimate source; news of good technology shocks tomorrow induces consumers to adjust consumption today in a way that, interacted with habits or something similar, lowers risk aversion and hence the equity premium. But I'm envisioning something a long way from anything that has been written down here. And that's the point; it needs to be written down.

7 Labor income and idiosyncratic risk

The basic economics we are chasing is the idea that assets must pay a higher average return if they do badly in "bad times," and we are searching for the right macroeconomic measure of "bad times." A natural idea in this context is to include labor income risks in our measure "bad times." Surely people will avoid stocks that do badly when they have just lost their jobs, or are at great risk for doing so. Here, I study models that emphasize first *overall* employment as a state variable ("labor income") and second increases in individual *risk* from non-market sources ("idiosyncratic risk").

7.1 Labor income

The economics of labor income as a state variable are a little tricky. If utility is separable between consumption and leisure, then consumption should summarize *all* economically relevant risks. If someone loses their job and this is bad news, they should consume less as well, and consumption therefore reveals all we need to know about the risk.

Leisure (labor) can also enter, as above, if utility is non-separable between consumption and leisure. However, current labor income work does not stress this possibility, perhaps again because we don't have much information about the cross-elasticity. It's possible that people fear stocks because people fear that when stocks down, they will lose their jobs, and *in addition* to increased "hunger" because they have to reduce their consumption, the are *even hungrier* (larger effect on marginal utility of consumption) because of all their new-found leisure; they would not fear the event so much if they were just as poor but still working. It's possible (maybe the unemployed exercise more, or *really* wish they could buy a new set of golf clubs), but this is not the channel that current work has pursued, perhaps since it seems pretty quantitatively implausible.

A better motivation for labor income risk, as for most factor models in finance, is the suspicion that consumption data are poorly measured or otherwise correspond poorly to the constructs of the model. The theory of finance from the CAPM on downward consists of various tricks for using *determinants* of consumption such as wealth (CAPM) or news about future investment opportunities (ICAPM) in place of consumption itself; not because anything is wrong with the consumption-based model in the theory, but on the supposition that it is poorly measured in practice.

With that motivation, labor income is one big determinant of consumption that is not included in stock market indices. Thus a multifactor model that includes labor income makes sense.

Measurement is tricky, however. The *present value* of labor income, or the value of "human capital," belongs most properly in asset pricing theory. Consumption does not decline (marginal utility of wealth does not rise) if you lose your job and know you can quickly get a better one. One can certainly cook up a theory in which labor income itself tells us a lot about the present value of labor income. An AR(1) time series model and constant discount rates are the standard assumptions, but they are obviously implausible. The same model applied to stocks says that today's dividend (equivalent to today's labor income) tells us all we need to know about stock prices; that a beta on dividends would give the same answer as a beta on prices, that price-dividend ratios are constant through time. We would laugh at any paper that did this for stocks, yet it is standard practice for labor income.

Still, the intuition for the importance of labor income risk is strong. The paragraph from Fama and French (1996, p. 77) quoted above combines some of the "labor income" risk here and the "idiosyncratic risk" that follows. What remains is to find evidence in the data for these mechanisms.

Jagannathan and Wang

Jagannathan and Wang (1996) is so far the most celebrated recent model that includes a labor income variable. (See also the quite successful extension in Jagannathan, Kubota, and Takehara, 1988.) The main model is a three factor model,

$$E(R^i) = c_0 + c_{vw}\beta_i^{VW} + c_{prem}\beta_i^{prem} + c_{labor}\beta_i^{labor}$$

where the betas are defined as usual from time series regressions

$$R_t^i = a + \beta_i^{VW} VW_t + \beta_i^{prem} prem_t + \beta_{i,labor} labor_t + \varepsilon_t^i;$$

VW is the value weighted market return, $prem$ is the previous month's BAA-AAA yield spread and labor is the previous month's growth in a two-month moving average of labor income. *prem* is included as a conditioning variable; this is a restricted specification of a conditional CAPM. ("Restricted" because in general one would include $prem \times VW$ and $prem \times labor$ as factors, as in Lettau and Ludvigson's 2001 conditional CAPM.)

With VW and prem alone, Jagannathan and Wang report only 30% cross-sectional R^2 (expected return on betas), presumably because the yield spread does not forecast returns as well as the cay variable used in a similar fashion by Lettau and Ludvigson (2001). Adding labor income, however, they obtain up to 55% cross-sectional[8] R^2 .

Alas, the testing ground is not portfolios sorted by book to market ratio, but 100 portfolios sorted by beta and size. The do check (Table VI) that the Fama French 3 factor model does no better (55% cross-sectional R^2) on their portfolios, but we don't know from them if labor income does well on book/market sorted portfolios. Furthermore, the paper makes the usual assumption that labor income is a random walk and is valued with a constant discount rate so that the current change in labor *income* measures the change in its *present value* (p. 14 "we assume that the return on human capital is an exact linear function of the growth rate in per capita labor income") Finally, the one-month $R_t^{labor} = [L_{t-1} + L_{t-2}]/[L_{t-2} - L_{t-3}]$ means that the factor is really *news about aggregate labor income*, since L_{t-1} data is released at time t rather than *actual labor income as experienced by workers.*

I do not include Jagannathan and Wang in the readings, because most of the empirical point can be seen in Table of Lettau and Ludvigson (2001) which is included. The rows that include factors Δy use labor income as a factor, and this time measured contemporaneously. Lettau and Ludvigson use the consumption to wealth ratio rather than the bond premium as the conditioning variable, which seems to give better results. They also examine the Fama-French 25 size and book/market portfolios which allows us better to compare across models in this standard playground. They actually find reasonable performance (58% R^2) in an *unconditional* model that includes only the market return and labor income growth as factors. Adding the scaled factors of the conditional model, i.e.

$$m_{t+1} = a + b_1 R_{t+1}^{VW} + b_2 \Delta y_{t+1} + b_3 cay_t + b_4 \left(cay_t \times R_{t+1}^{VW} \right) + b_5 \left(cay_t \times \Delta y_{t+1} \right)$$

they achieve essentially the same R^2 as the Fama - French 3 factor model.

Campbell

Campbell (1996) uses labor income in a three-factor model. His factors are 1) the market return 2) innovations in variables that help to forecast future market returns 3) innovations in variables that help to forecast future labor income. The analysis starts from a vector autoregression including the market return, real labor income growth, and as forecasting variables the dividend/price ratio, a detrended interest rate and a credit spread.

[8] Again, I pass on these numbers with some hesitation – unless the model is fit by an OLS cross-sectional regression, which maximizes R^2, the R^2 depends on technique and even on how you calculate it. Only under OLS is $var(x\beta)/var(y) = 1 - var(\varepsilon)/var(y)$. Yet cross-sectional R^2 is a popular statistic to report, even for models not fit by OLS cross-sectional reguresion.

This paper has many novel and distinguishing features. First, despite the nearly 40 years that have passed since Merton's (1973) theoretical presentation of the ICAPM only a very small number of empirical papers have ever checked that their proposed factors do, in fact, forecast market returns. This is one of the rare exceptions. (Ferson and Harvey 1999 and Brennan, Xia and Wang 2004 are the only other ones I know of.) Campbell's factors also forecast current and future labor income, again taking one big step closer to innovations in human capital rather than just the flow of labor income. Finally, parameters are tied to estimates of fundamental parameters such as risk aversion, unlike the case in practically all other macro asset pricing papers (In the derivation of almost all models, the factor risk premium is tied to the degree of risk and risk aversion, yet this implication is hardly ever checked.)

Alas, again this paper came out before that much attention was lavished on the book/market effect, so the test portfolios are an intersection of size and industry portfolios. Size really does little more than sort on market beta, and industry portfolios give little variation in expected returns, as seen in Campbell's table 5. As one might suspect, most variation in the present value of labor income and return comes not from current labor income or changing forecasts of future labor income, but from a changing discount rate applied to labor income. However, the discount rate here is the same as the stock market discount rate. On one hand we expect discount rate variation to dominate as it does in stock prices. But it's not obvious that the stock discount rate should apply to labor income, and at a data level it means that labor income is really not a new factor. The bottom line is on p. 336: The CAPM is pretty good on size portfolios, and other factors seem not that important.

Current work

The next step is a current working paper by Malloy, Moskowitz, and Vissing-Jorgenson (2005). Among other refinements, these authors check whether their model explains portfolios sorted on book/market, size and momentum as well as individual stocks; they use measures of hiring and firing rather than the quite smooth average earnings data; and they measure the permanent component of labor income which at least gets one step closer to the present value or human capital that should matter in theory. They find good performance of the model in book/market sorted portfolios, suggesting that labor income risk (or associated macroeconomic risk) really is behind the "value effect"

Santos and Veronesi (2005) study a two-sector version of the model in Menzly, Santos and Veronesi (2004) and related to Lettau and Ludvigson (2001). They think of the two sectors as labor income (human capital) vs. market or dividend income, corresponding to physical capital. A conditional CAPM holds in the model in which the ratio of labor income to total income is a conditioning variable – expected returns etc. vary as this ratio varies. In addition the relevant market return is the total wealth portfolio including human capital, and so shocks to the value of labor income are priced as well. This is a very nice completely-solved model that shows the potential effects of labor income on asset pricing.

One part of the empirical work checks that the ratio of labor to total income forecasts aggregate returns; it does and better than the dividend price ratio, adding to evidence that macro variables (cay, I/K) forecast stock returns. The second part of the empirical work checks whether the factors can account for the average returns of the 25 Fama-French size and book/market portfolios (Table 6). Here, adding the ratio of labor to total income as a *conditioning variable* helps a lot, raising the cross-sectional R^2 from nearly zero for the CAPM to 50% for this conditional CAPM, in line with Lettau and Ludvigson's (2001) findings. Alas, adding shocks to the present value of labor income (measured here by changes in wages, with all the usual warnings) does not help much, either alone or in combination with the conditioning variables. The major success with this specification comes then as a conditioning variable rather than as a risk-factor.

7.2 Idiosyncratic risk

In most of our thinking about macroeconomics and finance, we use a "representative consumer." We analyze economy-wide aggregates, making a first approximation that the *distribution* across consumers, while important and interesting, does not affect the evolution of aggregates. We say that a "tax cut" or "interest rate reduction" may increase "consumption" or "saving" affecting "employment" and "output;" in doing so we are treating economy-wide averages as if they are generated by the actions of a single consumer; we are assuming that *distributional* effects while present do not affect *averages* and *market prices* in important ways. Of course the theory needed to justify this simplification is extreme: either we need to believe there is full insurance, so everybody's consumption does move in lockstep, or we need to believe severe restrictions on preferences.

Macroeconomics and finance are thus full of investigations whether cross-sectional distributions matter. Two particular strains of this investigation are important for us. First, most people don't hold any stocks at all. Therefore, their consumption may be de-linked from the stock market, and models that connect the stock market only to those who actually hold stocks might be more successful. Second, perhaps idiosyncratic risk matters. Perhaps people fear stocks not because they might fall at a time when *total* employment or labor income falls, but because they might fall at a time when the *cross-sectional risk* of unemployment or labor income increases.

While plausible on an intuitive level, both possibilities are not obviously quantitatively important on second look. In considering limited participation, are people who hold no stocks really not "marginal?" The costs of joining the stock market are trivial; just turn off your spam filter for a moment and that becomes obvious. Thus, people who do not invest at all *choose* not to do so in the face of trivial fixed costs. This choice must reflect the attractiveness of a price ratio relative to the consumer's marginal rate of substitution; they really are "marginal" or closer to "marginal" than most theories (with large "participation costs") are forced to assume. Also, we are all more linked to the stock market than we think. Most data on assets excludes defined-contribution pension plans, most of contain stock market investments. Even employees with a defined benefit plan should at least watch the stock of their own company when making consumption plans, as employees of United Airlines recently found out to their dismay. Finally, while there are a lot of people with little stock holding, they also have little consumption and little effect on market prices. Aggregates weight by dollars, not people, and many more dollars of consumption are enjoyed by rich people who own stocks than the *numbers* of such people suggests.

In considering idiosyncratic risk, note that its effects are based entirely on nonlinearities in marginal utility. The individual first order conditions still hold,

$$1 = E\left(\beta \frac{u'(c^i_{t+1})}{u'(c^i_t)} R_{t+1}\right). \tag{17}$$

we can always "aggregate" by averaging *marginal utilities*

$$1 = E\left(\left[\frac{1}{N}\sum_i \beta \frac{u'(c^i_{t+1})}{u'(c^i_t)}\right] R_{t+1}\right) \tag{18}$$

The point is that we cannot in general aggregate by averaging *consumption*

$$1 \neq E\left(\beta \frac{u'(\frac{1}{N}\sum_i c^i_{t+1})}{u'(\frac{1}{N}\sum_i c^i_t)} R_{t+1}\right). \tag{19}$$

Now, if marginal utility were linear, as it is under quadratic utility, then of course averaging consumption *would* work, and would give the same answer as aggregating marginal utility, even in incomplete markets (See Hansen 1987 for a very nice treatment). Thus, if we are investigating predictions of (18) when (19) fails, the nonlinearities in marginal utility are crucially important.

The idiosyncratic risk view is challenging on a quantitative level as well. We need idiosyncratic risk – the cross-sectional distribution of individual income shocks, say – to vary over time; the distribution needs to widen when high-average-return securities (stocks vs. bonds, value stocks vs. growth stocks) decline. And, if we are to avoid the usual high risk aversion traps it needs to widen *a lot* at such times.

Finally, both idiosyncratic risk and limited participation views still maintain the basic consumption-based prediction for individuals, (17). Equity premium problems are just as difficult for (correctly measured) individual consumption as for aggregate consumption. For example, the Hansen-Jagannathan bound says that the volatility of marginal utility growth must exceed 50% per year (and more, to explain the value premium). For log utility, that means consumption growth must vary by 50 percentage points per year! (Keep in mind that this is nondurable consumption, and the flow of durables services, not durables purchases. Buying a house once in 10 years or a car once in three does not count towards this volatility.) Furthermore, what counts is the portion of consumption volatility correlated with the stock market. Purely idiosyncratic volatility (due to individual job loss, illness, divorce, etc.) does not count.

Based on both of the last points, the reviews in Cochrane (1997) and Cochrane (2004) suggest that idiosyncratic risk and limited participation views, while they may well improve our understanding of the connection between asset markets and macroeconomics, will not be able to avoid the requirement for very high risk aversion. Heaton and Lucas (1996) consider some of the same issues in a carefully-calibrated model. They find that they need a very large transaction cost to generate the observed equity premium.

Constantinides and Duffie

I include here Constantinides and Duffie (1996). This is a classic and very elegant paper. Basically, they prove a constructive existence theorem: there *is* a specification of idiosyncratic income risk that can explain *any* premium, using only power (constant relative risk aversion, time-separable) utility, and Constantinides and Duffie show you how to construct that process. The key is that the cross-sectional variance of labor income growth must change as the discount factor that generates asset returns changes. Then, the nonlinearity of marginal utility in the power specification means that the Jensen's inequality term $\frac{1}{2}\sigma^2$ shows up in the mean (across consumers) marginal utility and hence in asset prices. This is a brilliant contribution as a decade of research into idiosyncratic risk had stumbled against one after another difficulty, and had not been able even to demonstrate the theoretical possibility of this effect.

Empirical work; Brav, Constantinides and Geczy

Empirical work on whether variation in the cross-sectional distribution of income and consumption is important for asset pricing is just beginning. I include Brav, Constantinides and Geczy (2002) as an example.

Most investigations find some support for the effect – consumption and income *do* become more volatile across people in recessions and at times when the stock market declines. However, they typically find that the *magnitudes* are not large enough to explain the equity premium without high risk aversion, as my (1997) back of the envelope calculation suggested. And nobody has even started to talk about the value premium. Storesletten, Telmer and Yaron (2005) document greater dispersion in labor income across households in PSID in recessions, but they do not connect that greater dispersion to asset pricing. Cogley (1998) examines the cross-sectional properties of consumption from the consumer expenditure survey. He finds that "cross- sectional factors" – higher moments of the cross-sectional distribution of consumption growth – "are indeed weakly correlated with stock returns, and they generate equity premia of 2 percent or less when the coefficient of relative risk aversion is below 5." Even ignoring the distinction between consumption and income, Lettau (2002) finds that the cross-sectional distribution of idiosyncratic income does not vary enough to explain the equity premium puzzle without quite high risk aversion. Constantinides and Duffie's model also requires a substantial permanent component to idiosyncratic labor income, in order to keep consumers from smoothing it by saving and dissaving. Yet standard calibrations such as in Heaton and Lucas (1996) don't find enough persistence in the data. If labor income were very volatile and persistent, then of course the distribution of income would quickly fan out over time.

In contrast, Brav, Constantinides and Geczy (2002) do report some asset-pricing success.

They use household consumption data from the consumer expenditure survey and consider measurement error extensively. They examine one implication only, whether by aggregating marginal utility rather than aggregating consumption, they can explain the equity premium $0 = E(mR^e)$. The point estimates do so at a coefficient of between 2 and 3. In aggregate consumption data, there is no risk aversion parameter γ that satisfies this moment.

I hope that future work will analyze this result more fully. What are the time-varying cross-sectional moments that drive the result, and why did Brav Constantinides and Geczy find them where Cogley and Lettau did not, and my (1997), (2004) calculations suggest that the required properties are extreme? And of course, pricing the Fama French 25 portfolios (or really, just pricing the value portfolio) is the current challenge for macro models, not just the equity premium.

Of course, another approach to this question is that *individuals* still price assets exactly as before, $1 = E\left[\beta \left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} R_{t+1}\right]$ holds for each *individual's* consumption in all these models. So, once we have opened the CES or PSID databases, we could simply test whether asset returns are correctly related to household level consumption and forget about aggregation either of consumption or of marginal utility. Alas, this approach is not so easy either: individual consumption data is full of measurement error as well as idiosyncratic risk, and raising measurement error to the γ power can swamp the signal (See Brav, Constantinides and Geczy for an extended discussion.) In addition *individuals* may not be stable over time, where *aggregates* are. For just this reason, we do not test asset pricing models on the individual stock level, but instead form characteristic-sorted portfolios. It makes sense to aggregate the m in $1 = E(mR)$ just as we aggregate the R into portfolios. I won't attempt an extended review of this literature, (Vissing-Jorgensen 2002 is a good recent example) other than to say a success in the standard exercises of explaining the equity premium, market predictability, the value premium or the 25 book/market - size portfolios still awaits us.

One approach is simply to ignore *why* few people invest in stocks at all, and to examine the consumption of stockholders or wealthy individuals presumed to be stockholders directly. Mankiw and Zeldes (1991) is the classic first investigation of this effect. They found that stockholder's consumption is more volatile and more correlated with the stock market than that of nonstockholders, which helps to explain the equity premium. Most recently Ait-Sahalia, Parker and Yogo (2004) find similarly that consumption of "luxury goods" explains the equity premium with much less risk aversion than that of normal goods. Still, the properties of consumption required to explain the equity or value premium with power utility are so extreme that this line is not a full explanation.

8 Challenges for the future

Though this review may seem extensive and exhausting, it is clear at the end that work has barely begun. The challenge is straightforward: we need to understand what macroeconomic risks underlie the "factor risk premia", the average returns on special portfolios, that finance research uses to crystallize the cross section of assets. A current list might include the equity premium, the value and size premiums, the momentum premium, term premia in bond markets and the foreign exchange risk premium revealed by interest differentials. And more premia will certainly emerge through time. In each case, these premiums should line up with a tendency of the factor portfolios to decline in "bad times".

On the empirical side, we are really only starting to understand how the simplest models such as the power utility model do and do not address these premiums, looking across data issues, horizons, time aggregation and so forth. Really understanding whether nonseparable utility helps awaits. The success of ad-hoc macro factor models in explaining the Fama-French 25 is suggestive, but their performance still needs careful evaluation and they need connection to economic theory.

The general equilibrium approach is a vast and largely unexplored new land. The papers covered here are like Columbus' reports that the land is there. The pressing challenge is to develop a general equilibrium model with an interesting cross-section. The model needs to have multiple "firms"; it needs to generate the fact that low-price "value" firms have higher returns than high price "growth firms"; it needs to generate the failure of the CAPM to account for these returns, and it needs to generate the *comovement* of value firms that underlies Fama and French's factor model. An ideal model would do this with relatively standard technology specification, consistent with microeconomic investigation. The papers surveyed here, while path-breaking advances along these lines, do not come close to the full list of desiderata.

More generally, the eventual goal of macro / asset pricing research must take the form of a general equilibrium model. General equilibrium models generate covariances *endogenously.* You don't get to say "we explain expected returns because betas are high"; the models generate the betas too, and hence give us an understanding of where betas come from. We really don't have an economic explanation of asset markets until we understand betas as well as average returns. General equilibrium models force us to start from cash flows, and deeper: the economic sources of cash flows. They force us to think about prices as the endogenous variable, not just another characteristic for sorting returns. We really cannot say we understand asset prices until we can construct a general equilibrium model that generates artificial data, and that artificial data resembles actual data.

Having said "macroeconomics," "risk" and "asset prices," the reader will quickly spot a missing ingredient: money. In macroeconomics, monetary shocks and monetary frictions are an essential ingredient of business cycles, and should certainly matter at least for bond risk premia. More generally, understanding asset prices may require us to look past the simple frictionless models of the real business cycle tradition. Coming from the other direction, there is now a lot of evidence for "liquidity" effects in bond and stock markets (see Cochrane 2005 for a review), and perhaps both sorts of frictions are related.

9 Papers

9.1 Facts: time-variation and business cycle correlation of returns

Cochrane, John H., 1999, “New Facts in Finance” *Economic Perspectives* Federal Reserve Bank of Chicago 23, 36-58. (1999)

Fama, Eugene F. and Kenneth R. French, 1989, “Business Conditions and Expected Returns on Stocks and Bonds,” *Journal of Financial Economics*, 25 (1), 23-49.

Lettau, Martin, and Sydney Ludvigson, 2001, “Consumption, Aggregate Wealth, and Expected Stock Returns,” *Journal of Finance* 56 (3), 815-850.

Fama, Eugene F., and Kenneth R. French, 1996, “Multifactor Explanations of Asset-Pricing Anomalies,” *Journal of Finance* 51 (1), 55-84.

Liew, Jimmy, and Maria Vassalou, 2000, “Can Book-to-Market, Size and Momentum be Risk Factors that Predict Economic Growth?” *Journal of Financial Economics* 57, 221-45

9.2 Equity Premium

Mehra, Rajnish and Edward Prescott. 1985. “The Equity Premium: A Puzzle,” *Journal of Monetary Economics* 15 (2), march 145-161.

Cochrane, John H. and Lars Peter Hansen, 1992, “Asset Pricing Explorations for Macroeconomics” In Olivier Blanchard and Stanley Fisher, Eds.,*NBER Macroeconomics Annual 1992,* 115-165.

9.3 Consumption models

Hansen, Lars Peter and Kenneth J. Singleton, 1982, “Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models,” *Econometrica* 50, 1269-1288.

Hansen, Lars Peter and Kenneth J. Singleton 1984, “Errata” *Econometrica* 52, 267-268

Lettau, Martin, and Sydney Ludvigson, 2001, “Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying,” *Journal of Political Economy* 109, 1238-87

Campbell, John Y. and Cochrane, John H. (1999) “By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior” *Journal of Political Economy*, 107, 205-251.

9.4 Production, Investment and General Equilibrium models

Cochrane, John H., 1991b, “Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations” *Journal of Finance* 46 207-234.

Cochrane, John H., 1996, “A Cross-Sectional Test of an Investment-Based Asset Pricing Model” *Journal of Political Economy*, 104,572-621

Jermann, Urban, 1998 “Asset Pricing in Production Economies” *Journal of Monetary Economics* 41, 257-275

Boldrin, Michele, Lawrence J. Christiano and Jonas Fisher, 2001 “Habit Persistence, Asset Returns, and the Business Cycle,” *American Economic Review* 91 149-166

Menzly, Lior, Tano Santos and Pietro Veronesi, 2004 “Understanding Predictability,” *Journal of Political Economy* 112, 1-47

Tallarini, Thomas D. Jr., 2000, “Risk-Sensitive Real Business Cycles,” *Journal of Monetary Economics,* 45, 507-532

Hall, Robert E., 2001, “The Stock Market and Capital Accumulation,” *American Economic Review,* 91, 1185 - 1202.

9.5 Labor income and idiosyncratic risk

Campbell, John Y., 1996, Understanding risk and return,” *Journal of Political Economy* 104, 298–345.

Constantinides, George M. and Darrell Duffie, 1996, “Asset Pricing with Heterogeneous Consumers,” *Journal of Political Economy* 104, 219–240.

Brav, Alon, George Constantinides and Christopher Geczy, 2002, “Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence.” *Journal of Political Economy 110, 793-824*

10 References

Abel, Andrew B., 1990, "Asset Prices Under Habit Formation and Catching Up With The Joneses," *American Economic Review* 80, 38–42.

Ait-Sahalia, Yacine, Jonathan Parker and Motohiro Yogo, 2004, "Luxury Goods and the Equity Premium," *Journal of Finance*, 59 2959-3004.

Alvarez, Fernando and Urban J. Jermann, 2004, "Using Asset Prices to Measure the Cost of Business Cycles," *Journal of Political Economy,* 112, 1223-1256

Ang, Andrew, Monika Piazzesi, and Min Wei, 2004, "What Does the Yield Curve Tell us About GDP Growth?" *Journal of Econometrics*, forthcoming.

Bansal, Ravi, and Amir Yaron, 2005, "Risks For the Long Run: A Potential Resolution of Asset Pricing Puzzles" forthcoming, *Journal of Finance*

Berk, Jonathan B, Richard C. Green and Vasant Naik, 1999, "Optimal Investment, Growth Options and Security Returns," *Journal of Finance*, 54, 1153 - 1607.

Breeden, Douglas, Michael Gibbons, and Robert Litzenberger, 1989, "Empirical Tests of the Consumption-Oriented CAPM," *Journal of Finance*, 44, 231 - 262.

Breeden, Douglas T., 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities "*Journal of Financial Economics* 7, 265-96

Brennan, Michael J., Yihong Xia, and Ashley Wang, 2004, "Estimation and Test of a Simple Model of Intertemporal Asset Pricing," *Journal of Finance* forthcoming.

Buraschi, Andrea and Alexei Jiltsov, 2005, "Time-Varying Inflation Risk Premia and the Expectations Hypothesis: A Monetary Model of the Treasury Yield Curve," *Journal of Financial Economics*, Forthcoming.

Campbell, John Y., 2000, "Asset Pricing at the Millennium," *Journal of Finance* 55, 1515-67

Campbell, John Y., 2003, "Consumption-Based Asset Pricing", Chapter 13 in George Constantinides, Milton Harris, and René Stulz, eds. *Handbook of the Economics of Finance* Vol. IB, North-Holland, Amsterdam, 803-887. (Available online as NBER Working Paper 9845, 1998)

Campbell, John Y. and Cochrane, John H., 1995, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior" *NBER Working Paper* 4995..

Campbell, John Y., and John H. Cochrane, 2000, "Explaining the Poor Performance of Consumption Based Asset Pricing Models," *Journal of Finance*.55, 2863 - 2878.

Campbell, John Y. and Robert J. Shiller, 1988, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors" *Review of Financial Studies* 1, 195–228.

Chan, Lewis, and Leonid Kogan, 2001, "Catching Up with the Jones: Heterogeneous Preferences and the Dynamics of Asset Prices, *Journal of Political Economy* 110, 1255-85

Chen, Nai-Fu, 1991, "Financial Investment Opportunities and the Macroeconomy," *Journal of Finance* 46, 529 - 554.

Chen, Nai-Fu, Richard Roll and Steven Stephen A. Ross, 1986, "Economic Forces and the Stock Market," *Journal of Business* 59, 383-403.

Chen, Xiaohong, and Sydney Ludvigson, 2004, "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models" Manuscript, New York University

Cochrane, John H., 1989, "The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives" *American Economic Review* 79, 319-337.

Cochrane, John H., 1991a, "Explaining the Variance of Price-Dividend Ratios" *Review of Financial Studies* 5, 243-280.

Cochrane, John H., 1993, "Rethinking Production Under Uncertainty" Manuscript, University of Chicago.

Cochrane, John H., 1994, "Permanent and Transitory Components of GNP and Stock Prices" *Quarterly Journal of Economics* 109, 241-266.

Cochrane, John H., 1997,"Where is the Market Going? Uncertain Facts and Novel Theories" *Economic Perspectives* Federal Reserve Bank of Chicago, 21: 6 (November/December 1997) also NBER Working paper 6207

Cochrane, John H., 1999, "New Facts in Finance" *Economic Perspectives* Federal Reserve Bank of Chicago 23 (3) 36-58.

Cochrane, John H., 1999, "Portfolio Advice for a Multifactor World" *Economic Perspectives* Federal Reserve Bank of Chicago 23 (3) 59-78.

Cochrane, John H. 2004, *Asset Pricing*, Princeton: Princeton University Press, 2nd Edition

Cochrane, John H. and Monika Piazzesi, 2002, "The Fed and Interest Rates: A High-frequency Identification" *American Economic Review* 92, 90-95

Cochrane, John H. and Monika Piazzesi, 2005, "Bond Risk Premia," forthcoming *American Economic Review*.

Cochrane, John H., 2005, "Liquidity Trading and Asset Prices" NBER Reporter, National Bureau of Economic Research www.nber.org/reporter.

Constantinides, George, 1990, "Habit formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy* 98, 519–543.

Cogley, Timothy, 2002, "Idiosyncratic risk and the equity premium: evidence from the consumer expenditure survey," *Journal of Monetary Economics*, 49 (2), 309-334.

Constantinides, George, and Darrell Duffie, 1996, "Asset Pricing With Heterogeneous Consumers," *Journal of Political Economy* 104, 219–240.

Daniel, Kent, and David Marshall, 1997, "Equity-Premium and Risk-Free-Rate Puzzles at Long Horizons" *Macroeconomic Dynamics* 1, 452-84.

De Bondt, Werner F. M. and Richard Thaler,1985, "Does the Stock Market Overreact?" *Journal of Finance* 40, 793-805.

Eichenbaum, Martin, and Lars Peter Hansen, 1990, "Estimating Models With Intertemporal Substitution Using Aggregate Time Series Data," *Journal of Business & Economic Statistics* 8, 53–69.

Eichenbaum, Martin, Lars Peter Hansen and Kenneth Singleton, 1988, "A Time-Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics* 103, 51-78.

Engel, Charles, 1996, "The Forward Discount Anomaly and the Risk Premium: a Survey of Recent Evidence," *Journal of Empirical Finance* 3, 123-192.

Epstein, Larry G. and Stanley E. Zin. 1989. "Substitution, Risk Aversion and the Temporal Behavior of Asset Returns." *Journal of Political Economy* 99: 263-286.

Estrella, Arturo, and Gikas Hardouvelis, 1991, "The Term Structure as a Predictor of Real Economic Activity ," *Journal of Finance* 46, 555-76.

Fama, Eugene F., 1975, "Short-term Interest Rates as Predictors of Inflation," *American Economic Review* 65, 269–282.

Fama, Eugene F. 1984, "The Information in the Term Structure," *Journal of Financial Economics*, 13, 509-28

Fama, Eugene F., 1984, "Forward and Spot Exchange Rates," *Journal of Monetary Economics* 14, 319-338.

Fama, Eugene, F., 1990, "Stock Returns, Expected Returns, and Real Activity," *Journal of Finance* 45, 1089 - 1108.

Fama, Eugene F. and G. William Schwert, 1977,"Asset Returns and Inflation," *Journal of Financial Economics* 5, 115-46.

Fama, Eugene F. and Michael R. Gibbons, 1982, "Inflation, Real Returns and Capital Investment," *Journal of Monetary Economics* 9, 297-323.

Fama, Eugene F. and Robert R. Bliss, 1987, "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680-92.

Fama, Eugene F. and Kenneth R. French, 1988a, "Permanent and Temporary Components of Stock Prices." *Journal of Political Economy* 96, 246-273.

Fama, Eugene F. and Kenneth R. French, 1988b, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics* 22, 3-27.

Fama, Eugene F., and Kenneth R. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance* 47, 427–465.

Fama, Eugene F. and Kenneth R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* 33, 3-56.

Fama, Eugene F., and Kenneth R. French, 1997a, "Size and Book-to-Market Factors in Earnings and Returns,"*Journal of Finance* 50, 131-55.

Fama, Eugene F., and Kenneth R. French, 1997b, "Industry Costs of Equity," *Journal of Financial Economics* 43, 153-193.

Ferson, Wayne E. and George Constantinides, 1991, "Habit Persistence and Durability in Aggregate Consumption: Empirical Tests." *Journal of Financial Economics* 29, 199–240.,

Ferson, Wayne E. and Campbell R. Harvey (1999)"Conditioning Variables and the Cross Section of Stock Returns" *The Journal of Finance* 54,1325-1360

Gomes, Francisco, and Alexander Michaelides, 2004, "Asset Pricing with Limited Risk Sharing and Heterogeneous Agents," *Journal of Finance*, forthcoming.

Gomes, Joao F., Leonid Kogan, and Lu Zhang, 2003, "Equilibrium Cross-Section of Returns," *Journal of Political Economy* 111, 693-732.

Gourio, Francois, 2004, "Operating Leverage, Stock Market Cyclicality and the Cross-Section of Returns," Manuscript, University of Chicago.

Grossman, Sanford J., and Robert J. Shiller, 1981, "The Determinants of the Variability of Stock Market Prices," *American Economic Review* 71, 222–227.

Grossman, Sanford, Angelo Melino, and Robert J. Shiller, 1987, "Estimating the Continuous-Time Consumption-Based Asset-Pricing Model," *Journal of Business & Economic Statistics*, 5, 315-28

Hall, Robert E., 1978, "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy* 86, 971-87

Hansen, Lars Peter, 1987, "Calculating Asset Prices in Three Example Economies," in T.F. Bewley, Advances in Econometrics, Fifth World Congress, Cambridge University Press.

Hansen, Lars Peter and Robert J. Hodrick, 1980, "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy* 88, 829-53.

Hansen, Lars P., Thomas J. Sargent, and Thomas Tallarini, 1998, "Robust Permanent Income and Pricing," *Review of Economic Studies* 66, 873-907

Hansen, Lars Peter and Ravi Jagannathan, 1991, "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy* 99, 225-62.

Hansen, Lars Peter and Kenneth J. Singleton, 1983, "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy* 91, 249-268.

Hansen, Lars Peter, John C. Heaton and Nan Li, 2004a, "Intangible Risk?" Manuscript, University of Chicago.

Hansen, Lars Peter, John C. Heaton and Nan Li, 2004b "Consumption Strikes Back?" Manuscript, University of Chicago.

Heaton, John C., 1993, "The Interaction Between Time-Nonseparable Preferences and Time Aggregation," *Econometrica*, Vol. 61, No. 2. 353-385.

Heaton, John C., 1995, "An empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications," *Econometrica* 63, 681–717.

Heaton J. and D. Lucas, 1992, "The Effects of Incomplete Insurance Markets and Trading Costs in a Consumption-Based Asset Pricing Model" *Journal of Economic Dynamics and Control* 16, 601-20

Heaton, John C. and Deborah Lucas, 1995, "The Importance of Investor Heterogeneity and Financial Market Imperfections for the Behavior of Asset Prices," *Carnegie-Rochester Conference Series on Public Policy* 42, 1-32.

Heaton, John C. and Deborah Lucas, 1996, "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy* 104, 443-87.

Heaton, John C., and Deborah Lucas, 2000, "Stock Prices and Fundamentals," *NBER Macroeconomics Annual 1999,* 213-42

Jagannathan, Ravi and Zhenyu Wang, 1996, "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance* 51, 3-53

Jagannathan, Ravi, and Keiichi Kubota, and Hotoshi Takehara, 1988, "Relationship between Labor-Income Risk and Average Return: Empirical Evidence from the Japanese Stock Market," *Journal of Business* 71, 319-47

Jagannathan, Ravi and Yong Wang, 2005, "Consumption Risk and the Cost of Equity Capital" NBER working paper 11026

Kandel, Shmuel and Robert F. Stambaugh. 1990, "Expectations and Volatility of Consumption and Asset Returns," *Review of Financial Studies* 3, 207-232.

Kandel, Shmuel and Robert F. Stambaugh. 1991, "Asset Returns and Intertemporal Preferences," *Journal of Monetary Economics* 27, 39-71.

Kandel, Shmuel and Robert F. Stambaugh, 1995 "Portfolio Inefficiency and the Cross-Section of Expected returns" *Journal of Finance* 50, 157-184.

Kocherlakota, Narayanna, 1996, "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature* 34, 42-71.

Kogan, Leonid, 2004, "Asset prices and real investment," *Journal of Financial Economics* 73, 411-431.

Krusell, Per, and Anthony A. Smith, Jr., 1997, "Income and Wealth Heterogeneity, Portfolio Choice, and Equilibrium Asset Returns," *Macroeconomic Dynamics* 1, 387-422.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, "Contrarian Investment, Extrapolation, and Risk," *The Journal of Finance,* 49, 1541-1578.

Lamont, Owen, 1998, "Earnings and Expected Returns," *Journal of Finance* 53, 1563 - 1587.

Lamont, Owen A., 2000, Investment Plans and Stock Returns, *Journal of Finance*, 55,2719 - 2745.

LeRoy, Stephen F., 1973, Risk Aversion and the Martingale Property of Stock Prices, *International Economic Review* 14, 436-46.

LeRoy, Stephen F., and Richard D. Porter, 1981, "The Present-Value Relation: Tests Based on Implied Variance Bounds," *Econometrica* 49, 555-74.

Lettau, Martin, 2002, "Idiosyncratic Risk and Volatility Bounds, or, Can Models with Idiosyncratic Risk Solve the Equity Premium Puzzle?" *Review of Economics and Statistics*, 84 (2, May), 376-380.

Lettau, Martin, and Sydney Ludvigson, 2002, "Time-Varying Risk Premia and the Cost of Capital: An Alternative Implication of the Q Theory of Investment," *Journal of Monetary Economics*, 49, 31 - 66.

Lettau, Martin, and Sydney Ludvigson, 2001, "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy* 109, 1238-87

Lettau, Martin, and Sydney Ludvigson, 2001, "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance* 56, 815-49.

Lettau, Martin, and Sydney Ludvigson, 2004, "Expected Returns and Expected Dividend Growth," *Journal of Financial Economics,* forthcoming.

Lewellen, Jonathan and Stefan Nagel, 2004, "The conditional CAPM does not explain asset-pricing anomalies," Manuscript, MIT.

Li, Qing, Maria Vassalou, and Yuhang Xing, 2003, "Investment Growth Rates and the Cross-Section of Equity Returns," Manuscript, Columbia University

Lucas, Robert E., Jr., 1978, "Asset Prices in and Exchange Economy," *Econometrica* 46, 1429-1446.

Lucas, Deborah J.,2001, Asset Pricing with Undiversifiable Risk and Short Sales Constraints: Deepening the Equity Premium Puzzle," *Journal of Monetary Economics* 34, 325-41.

Lustig, Hanno and Adrien Verdelhan, 2004, "The Cross-Section of Foreign Currency Risk Premia and US Consumption Growth Risk" Manuscript, University of Chicago and UCLA.

Mankiw, N. Gregory. 1986. "The Equity Premium and the Concentration of Aggregate Shocks," *Journal of Financial Economics* 17, 211-219.

Mankiw, N. Gregory and Stephen Zeldes, 1991, "The Consumption of Stockholders and Non-Stockholders," *Journal of Financial Economics*, 29, 97-112.

Mehra, Rajnish, and Edward Prescott, 1985, "The Equity Premium: A Puzzle," *Journal of Monetary Economics* 15, 145–161.

Merton , Robert C., 1973, "An Intertemporal Capital Asset Pricing Model." Econometrica, 41, 867-887.

Menzly, Lior, 2001, "Influential Observations in Cross-Sectional Asset Pricing Tests" manuscript, University of Chicago.

Malloy, Christopher, Tobias Moskowitz, and Annette Vissing-Jorgenson 2005 "Job Risk and Asset Returns" Manuscript, University of Chicago

Ogaki, Masao, and Carmen M. Reinhart, 1998, "Measuring Intertemporal Substitution: The Role of Durable Goods" *Journal of Political Economy.* 106, 1078-1098.

Pakos, Michal, 2004, "Asset Pricing with Durable Goods and Non-Homothetic Preferences," Manuscript, University of Chicago.

Parker, Jonathan, and Christian Julliard, 2005, "Consumption Risk and the Cross-Section of Expected Returns," *Journal of Political Economy,* 113, in press.

Pastor, Lubos, and Pietro Veronesi, 2004, "Rational IPO Waves," *Journal of Finance*, forthcoming

Piazzesi, Monika, 2001, "An Econometric Model of the Yield Curve with Macroeconomic Jump Effects," NBER Working Paper 8246.

Piazzesi, Monika, 2003, "Bond Yields and the Federal Reserve," *Journal of Political Economy* forthcoming.

Piazzesi, Monika, Martin Schneider and Selale Tuzel, 2004, "Housing, Consumption, and Asset Pricing," Manuscript, University of Chicago, NYU and UCLA

Poterba, James, and Lawrence H. Summers, 1988, "Mean Reversion in Stock Returns: Evidence and Implications," *Journal of Financial Economics* 22, 27–60.

Roll, Richard 1977, "A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory" *Journal of Financial Economics,* 4 (2), 129-176.

Roll, Richard and Stephen A. Ross, 1994, "On the Cross-sectional Relation between Expected Returns and Betas" *Journal of Finance* 49, 101-121.

Schwert, G. William, 2003, "Anomalies and Market Efficiency," Chapter 15, *Handbook of the Economics of Finance*, eds. George Constantinides, Milton Harris, and René Stulz, North-Holland, pp. 937-972.

Shiller, Robert J., 1981 "Do Stock Prices Move too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review* 71, 421-436.

Stambaugh, Robert F., 1988, "The Information in Forward Rates: Implications for Models of the Term Structure,"*Journal of Financial Economics* 21, 41-70

Sundaresan, Suresh M., 1989, "Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth," *Review of Financial Studies* 2, 73–88.

Storesletten, Kjetil, Chris Telmer and Amir Yaron, 2005 "Cyclical Dynamics of Idiosyncratic Labor Market Risk," Manuscript University of Pennsylvania, Forthcoming, *Journal of Political Economy*

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron, 2000, "Asset Pricing with Idiosyncratic Risk and Overlapping Generations." Manuscript, Carnegie Mellon University.

Telmer Chris, 1993, "Asset Pricing Puzzles and Incomplete Markets," *Journal of Finance*, 48,1803-1832.

Santos, Tano and Pietro Veronesi, 2004, "Labor Income and Predictable Stock Returns," Manuscript, University of Chicago, forthcoming *Review of Financial Studies.*

Vassalou, Maria, 2003, "News Related to Future GDP Growth as a Risk Factor in Equity Returns," *Journal of Financial Economics* 68, 47-73.

Verdelhan, Adrien, 2004, "A Habit-Based Explanation of the Exchange Rate Risk Premium", Manuscript, University of Chicago.

Vissing-Jorgensen, Annette, 2002, "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy* 110, 825-853.

Wachter, Jessica, 2004, "A Consumption-Based Model of the Term Structure of Interest Rates," Manuscript, University of Pennsylvania.

Weil, Philippe, 1989, "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics* 24, 401-21.

Yogo, Motohiro, 2003, "A Consumption-Based Explanation of Expected Stock Returns," Manuscript, Wharton School, University of Pennsylvania.

Zhang, Lu, 2005, "The Value Premium," *Journal of Finance,* 60 (1), 67-104.

Zhang, Lu, 2004, "Anomalies," Manuscript, University of Rochester.