

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.

© 2005 by John H. Cochrane. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Financial Markets and the Real Economy
John H. Cochrane
NBER Working Paper No. 11193
March 2005, Revised September 2006
JEL No. G1, E3

ABSTRACT

I survey work on the intersection between macroeconomics and finance. The challenge is to find the right measure of "bad times," rises in the marginal value of wealth, so that we can understand high average returns or low prices as compensation for assets' tendency to pay off poorly in "bad times." I survey the literature, covering the time-series and cross-sectional facts, the equity premium, consumption-based models, general equilibrium models, and labor income/idiosyncratic risk approaches.

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 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. I survey research on the central question: what is the nature of macroeconomic risk that drives risk premia in asset markets?

The central idea of modern finance is that prices are generated by expected discounted payoffs,

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

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

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

The first term is the risk-neutral present value. The second term is the crucial discount for risk – a large negative covariance generates a low or “discounted” price. Applied to excess returns R^{ei} (short or borrow one asset, invest in another), this statement becomes¹

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

The expected excess return or “risk premium” is higher for assets that have a large negative covariance with the discount factor.

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

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

¹From (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,

$$\begin{aligned} 0 &= E(m)E(R^e) + cov(m, R^e) \\ E(R^e) &= -R^f cov(m, R^e) \end{aligned}$$

For small time intervals $R^f \approx 1$ so we have

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

This equation holds exactly in continuous time.

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* utility, not total utility. The discount factor is high at $t + 1$ if you desperately want more wealth at $t + 1$ – and would be willing to give up a lot of wealth in other dates or states to get it.

Equation (3) thus says that the risk premium $E(R^{ei})$ is driven by the covariance of returns with the marginal value of wealth.² Given that an asset must do well sometimes and do badly at other times, investors would rather it did well when they are otherwise desperate for a little bit of extra wealth, and that it did badly when they do not particularly value extra wealth. Thus, investors want assets whose payoffs have a positive covariance with hunger, and they will avoid assets with a negative covariance. Investors will drive up the prices and drive down the average returns of assets that covary positively with hunger, and vice versa, generating the observed risk premia.

These predictions are surprising to newcomers for what they do *not* say. More volatile assets do not necessarily generate a higher risk premium. The *variance* of the return R^{ei} or payoff x^i is irrelevant per se and does not measure risk or generate a risk premium. Only the *covariance* of the return with “hunger” matters.

Also, many people do not recognize that equations (2) and (3) characterize an *equilibrium*. They describe a market after everyone has settled on their optimal portfolios. They do not generate portfolio advice. *Deviations* from (2) and (3), if you can find them, can give portfolio advice. It’s natural to think that high expected return assets 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. Since investors all want them, those assets get *lower* average returns and command higher prices in equilibrium. High average return assets are *forced* to pay those returns, or equivalently to suffer low prices, *because* they are so “bad” – because they pay off badly precisely when investors are most hungry. In the end, there is no “good” or “bad.” Equations (2) and (3) describe an equilibrium in which the quality of the asset and its price are exactly balanced.

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. For example, in the CAPM a high average return is balanced by a large tendency of an asset to fall just when the market as a whole falls – a high “beta.” In equations,

$$E_t(R_{t+1}^{ei}) = cov_t(R_{t+1}^{ei}, R_{t+1}^m) \times \gamma$$

where γ is a constant of proportionality equal to the average investor’s risk aversion. Multifactor models such as the popular Fama-French (1996) three-factor model use returns on multiple portfolios to measure the marginal value of wealth.

² m_{t+1} really measures the *growth* in marginal utility or “hunger.” However, from the perspective of time t , $V_W(t)$ is fixed, so what counts is how the realization of the return covaries with the realization of time $t + 1$ marginal value of wealth $V_W(t + 1)$.

Research connecting financial markets to the real economy – the subject of this survey – goes one step deeper. It asks what are the *fundamental, economic* determinants of the marginal value of wealth? I start with the consumption-based model,

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

which states that assets must offer high returns if they pay off badly in “bad times” as measured by consumption growth. As we will see, this simple and attractive model does not (yet) work very well. The research in this survey is aimed at improving that performance. It aims to find better measures of the marginal value of wealth, rooted in measures of economic conditions such as aggregate consumption, that explain the pattern by which mean returns $E_t(R_{t+1}^{ei})$ vary across assets i and over time t .

Who cares?

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 centerpieces of dynamic macroeconomics are the equation of savings to investment, the equation of marginal rates of substitution to marginal rates of transformation, and the allocation of consumption and investment across time and states of nature. Asset markets are the mechanism that does all this equating. 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 first made precise by the “equity premium” puzzle. 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 almost all macroeconomic models. Clearly, finance has a *lot* to say about macroeconomics, and it says that something is 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 fads and fashions disconnected from the real economy.” That might be true, but if so, by what magic are marginal rates of substitution and transformation equated? It makes no sense to say “markets are crazy” and then go right back to market-clearing models with wildly counterfactual asset-pricing implications. If asset markets are screwed up, so is the equation of marginal rates of substitution and transformation in every macroeconomic model, so are those models’ predictions for quantities, and so are their policy and welfare implications.

Finance

Many financial economists return the compliment, and dismiss macroeconomic 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 simply reflects Roll’s (1977) theorem: 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 Roll’s theorem. The instability of many “anomalies” and the ever-changing nature of factor models that “explain” them (Schwert 2003) lends some credence to this worry.

The main fishing constraint one can imagine is that the factor 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 still can’t fish without a license. For example, momentum has yet to acquire the status of a factor despite abundant empirical success, because it has been hard to come up with stories that it corresponds to some plausible measure of the marginal utility of wealth.

Second, much work in finance is framed as answering the question whether markets are “rational” and “efficient” or not. *No* amount of research using portfolios on the right hand side can *ever* address this question. The only possible 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. 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.

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 financial economics, and a crucial and unavoidable set of uncomfortable measurements and predictions 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 model of the marginal utility of wealth, then a portfolio formed by its regression on to asset returns will work just as well³. And this “mimicking portfolio”

³Start with the true model,

$$1 = E(mR)$$

where R denotes a vector of returns. Consider a regression of the discount factor on the returns, with no

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 beaten badly 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 three factor model” in pricing the Fama French 25 portfolios is rather pointless. Even if you do succeed, a “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 (market, hml, smb) are priced. 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 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)?

constant,

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

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

$$0 = E[(b'R) R]$$

Therefore, the payoff $b'R$ is a discount factor as well.

Variation over time

First, a number of variables forecast aggregate stock, bond, and foreign exchange returns. Thus, *expected returns vary over time*. The central technique is simple forecasting regression: 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. The forecasting variables x_t typically have a suggestive business cycle correlation. Expected returns are high in “bad times,” when we might well suppose people are less willing to hold risks.

For example, Table 1 reports regressions of excess returns on dividend-price ratios. A one percentage point higher dividend yield leads to a *four* percentage point higher return. This is a surprisingly large number. If there were no price adjustment, a one percentage point higher dividend yield would only lead to a one percentage point higher return. The conventional “random walk” view implies a price adjustment that takes this return away. Apparently, prices adjust in the “wrong” direction, *reinforcing* the higher dividend yield. Since the right hand variable (dividend-price ratio) is very persistent, long-horizon forecasts are even more dramatic, with larger coefficients and R^2 values.

The second set of regressions in Table 1 is just as surprising. A high dividend yield means a “low” price, and it should signal a decline in future dividends. We see tiny and completely insignificant coefficients, and tiny R^2 values. Apparently, variation in price-dividend ratios does not come from news about future dividends.

Horizon k (years)	$R_{t \rightarrow t+k}^e = a + b \frac{D_t}{P_t} + \varepsilon_{t+k}$			$\frac{D_{t+k}}{D_t} = a + b \frac{D_t}{P_t} + \varepsilon_{t+k}$		
	b	t(b)	R^2	b	t(b)	R^2
1	4.0	2.7	0.08	0.07	0.06	0.0001
2	7.9	3.0	0.12	-0.42	-0.22	0.001
3	12.6	3.0	0.20	0.16	0.13	0.0001
5	20.6	2.6	0.22	2.42	1.11	0.02

Table 1. OLS regressions of excess returns (value weighted NYSE - treasury bill) and real dividend growth on the value weighted NYSE dividend-price ratio. Sample 1927-2005, annual data. $R_{t \rightarrow t+k}^e$ denotes the total excess return from time t to time $t+k$. Standard errors use GMM (Hansen-Hodrick) to correct for heteroskedasticity and serial correlation.

This pattern is not unique to stocks. Bond and foreign exchange returns are also predictable, meaning that expected returns vary through time. The same pattern holds in each case: a “yield” or “yield spread” (dividend yield, bond yields, international interest rate differential) forecasts excess returns, it does so because something that *should* be forecastable to offset the variation in expected returns (dividend growth, short-term interest rates, exchange rates) does not move, or does not move quickly enough; and the high-expected return signal (high dividend yield, upward sloping yield curve, low interest rates relative to foreign) typically comes in bad macroeconomic times. A large number of additional variables also forecast returns.

Variation across assets

Second, expected returns vary *across assets*. Stocks earn more than bonds of course. In addition, a large number of stock characteristics are now associated with average returns. The book/market ratio is the most famous example: stocks with low prices (market value) relative to book value seem to provide higher subsequent average returns. A long list of other variables including size (market value), sales growth, past returns, past volume, accounting ratios, short-sale restrictions, and corporate actions such as investment, equity issuance and repurchases are also associated with average returns going forward. We can think of all these phenomena as similar regression forecasts applied to individual assets or characteristic-sorted portfolios: the basic finding is that there exist many variables x_{it} that give significant coefficients in

$$R_{t+1}^i - R_t^f = a + bx_{it} + \varepsilon_{it}.$$

This variation in expected returns across assets would not cause any trouble for traditional finance theory, if the characteristics associated with high average returns were also associated with large market betas. Alas, they often are not. Instead, the empirical finance literature has associated these patterns in expected returns with betas on new “factors.”

(Cochrane (1999a) is an easily accessible review paper that synthesizes current research on both the time-series and the cross-sectional 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 the facts.)

2.1 Return forecasts – variation over time

Return forecasts have a long history. 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 to say 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) added investment to the economic modeling, presaging the investment and equilibrium models we study later.

In the early 1980s, we learned that bond and foreign exchange expected excess returns vary over time – that the classic “expectations hypothesis” is false. Hansen and Hodrick (1980) and Fama (1984a) documented the predictability of foreign-exchange returns by running regressions of those returns on forward-spot spread or interest rate differentials across

countries. If the foreign interest rate is unusually higher than the domestic interest rate, it turns out that the foreign currency does not tend to depreciate and thus an adverse currency movement does not, on average, wipe out the apparently attractive return to investing abroad. (“Unusually” is an important qualifier. If you just invest in high interest rate countries, you end up investing in high inflation countries, and depreciation does wipe out any gains. The phenomenon requires you to invest in countries with a higher-than-usual interest rate spread, i.e. following a regression of returns on interest rate spreads over time, with a constant. What “usual” means, i.e. the fact of an estimated constant in these regressions, is still a bit of an open question.)

Fama (1984b) 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. Shiller, Campbell, and Schoenholtz (1983) and Campbell and Shiller (1991) analogously rejected the expectations hypothesis by regressions of future yields on current yields; their regressions imply time-varying expected returns. Campbell (1995) is an excellent summary of this line of research.

While the expectations hypothesis had been rejected before,⁴ these papers focused a lot of attention on the problem. In part, they did so by applying a simple and easily-interpretable regression methodology rather than more indirect tests: just forecast tomorrow’s excess returns from today’s yields or other forecasting variables. They also regressed *changes* in prices (returns) or yields on today’s yield or forward-rate *spreads*. The expectations hypothesis looks pretty good if you just regress (say) the ex-post spot rate on the ex-ante forward rate to test the prediction that the forward rate is equal to the expected spot rate. But this is not a very powerful test. For example, if you forecast tomorrow’s temperature by just quoting today’s temperature, you will also get a nice 1.0 coefficient and a high R^2 , as overall temperature varies over the year. To see a good weather forecaster, you have to check whether he can predict the *difference* of tomorrow’s temperature over today’s temperature. Similarly, we see the failure of the expectations hypothesis by seeing that the *difference* between the forward rate and this year’s spot rate does not forecast a *change* in the spot rate from this year to next year. Finally, when looked at this way, these papers showed the striking *magnitude* and character of expectations-hypothesis failures. If the forward rate is one percentage point higher than the spot rate, Fama and Bliss showed that expected returns rise by more than a full percentage point, and the one year short rate forecast does not change at all. Foreign exchange forecasts are even larger: a one percentage point interest differential seems to signal an increase in expected returns larger than one percentage point.

The latter findings in particular have been extended and stand up well over time. Stambaugh (1988) extended the results for short term bonds and Cochrane and Piazzesi (2005)

⁴Evidence against the expectations hypothesis of bond yields goes back at least to Macaulay (1938). Shiller, Campbell, and Schoenholtz generously say that the expectations hypothesis has been “rejected many times in careful econometric studies,” citing Hansen and Sargent (1981), Roll (1970), Sargent (1978), (1972), and Shiller (1979). Fama says that “The existing literature generally finds that forward rates...are poor forecasts of future spot rates,” and cites Hamburger and Platt (1975), Fama (1976), and Shiller, Campbell and Schoenholtz.

did so for long term bonds. Both papers ran bond returns from t to $t + 1$ on *all* forward rates available at time t , and substantially raised the forecast R^2 . The Cochrane and Piazzi bond return forecasting variable also improves on the yield spread’s ability to forecast *stock* returns, and we emphasize that a *single* “factor” seems to forecast bond returns for all maturities.

During this period, we also accumulated direct regression evidence that expected excess returns vary over time for the stock market as a whole. Rozeff (1984), Shiller (1984), Keim and Stambaugh (1986), Campbell and Shiller (1988) and Fama and French (1988b) showed that dividend/price ratios forecast stock market returns. Fama and French really dramatized the importance of the D/P effect by emphasizing long horizons, at which the R^2 rise to 60%. (The lower R^2 values in Table 1 reflect my use of both the pre-1947 and post-1988 data.) This observation emphasized that stock return forecastability is an *economically* interesting phenomenon that cannot be dismissed as another little anomaly that might be buried in transactions costs. Long horizon forecastability is not really a distinct phenomenon; it arises mechanically as the result of a small short horizon variability and a slow-moving right hand variable (D/P). It also does not generate much statistical news since standard errors grow with horizon just as fast as coefficients, as you can see by the t statistics in Table 1.

Fama and French (1989) is an excellent summary and example of the large body of work that documents variation of expected returns over time. This paper shows how dividend-price ratios term spreads (long bond yield less short bond yield) and default spreads forecast stock and bond returns. The paper emphasizes the comforting link between stock and bond markets: the term spread forecasts stock returns much as it forecasts bond returns.

If returns are predictable from variables such as dividend yields, it stands to reason that returns should also be predictable from past returns. The way the dividend yield changes after all is by having a good sequence of returns so dividends are divided by a larger price. Such “mean-reversion” in returns has the powerful implication that the variance of returns grows less than linearly with horizon, so stocks really are “safer in the long run.” Initially, this did seem to be the case. Poterba and Summers (1988) and Fama and French (1988a) documented that past stock market returns forecast subsequent returns at long horizons. However, this effect seems to have vanished, and the current consensus is that although variables such as dividend yields forecast returns, there is no univariate forecastability or mean-reversion (see, for example Cochrane 2004 p. 413-415). This is not a logical contradiction. For example the weather can be i.i.d. and thus not forecastable from its own past, yet still may be forecastable the day ahead by meteorologists who look at more data than past weather. Similarly, stock returns can be forecastable by other variables such as dividend yields, yet unforecastable by their own past.

A related literature including Campbell and Shiller (1988) and Cochrane (1991a) (summarized in Cochrane 1999) connects the time-series predictability of stock returns to stock price volatility. Linearizing and 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,

$$p_t - d_t = k + E_t \sum_{j=1}^{\infty} \rho^{j+1} [E_t(\Delta d_{t+j}) - E_t(r_{t+j})] + \lim_{j \rightarrow \infty} \rho^j (p_{t+j} - d_{t+j}) \quad (4)$$

where small letters are logs of capital letters, and k and $\rho = (P/D)/[1 + (P/D)] \approx 0.96$ are constants related to the point P/D about which we linearize. If price-dividend ratios vary at all, then, then either 1) price-dividend ratios forecast dividend growth 2) price-dividend ratios forecast returns or 3) prices must follow a “bubble” in which the price-dividend ratio is expected to rise without bound.

It would be lovely 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 subsequent rise in dividends. It turns out that price dividend ratios *do not* forecast aggregate dividends at all, as shown in the right hand panel of Table 1. 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 a high price-dividend ratio forecasts low returns. It turns out that the return forecastability we see in regressions such as the left hand side of Table 1 is just enough to completely account for the volatility of price dividend ratios through (4). (This is a main point of Cochrane 1991a.) Thus, return forecastability and “excess volatility” are *exactly* the same phenomenon. Since price-dividend ratios are stationary (Craine 1993) and since the return forecastability does neatly account for price-dividend volatility, we do not need to invoke the last “rational bubble” term.

Alas, the fact that almost all stock price movements are due to changing expected excess returns rather than to changing expectations of future dividend growth 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.

Macro variables and forecastability

The forecasting variables in return regressions are so far all based on market prices, though, which seems to take us away from our macroeconomic quest. However, as emphasized by Fama and French (1989) with a nice series of plots, the prices that forecast returns are correlated with business cycles, with higher expected returns in bad times. A number of authors including Estrella and Hardouvelis (1991) and more recently Ang, Piazzesi and Wei (2004) documented that the price variables that forecast returns also forecast economic activity.

One can of course run regressions of returns on macroeconomic variables, and a number of other macroeconomic variables forecast stock returns, including the investment/capital ratio (Cochrane 1991b), the dividend-earnings ratio (Lamont 1998), investment plans (Lamont 2000), the ratio of labor income to total income (Menzly, Santos and Veronesi 2004), the ratio of housing to total consumption (Piazzesi Schneider and Tuzel 2005), and an “output

gap” formed from the Federal Reserve capacity index (Cooper and Priestley 2005), and the ratio of consumption to wealth (Lettau and Ludvigson 2001a). The investment/capital ratio and consumption/wealth ratios are particularly attractive variables. The Q theory of investment says that firms will invest more when expected returns are low; the investment to capital regressions verify this fact. Similarly, optimal consumption out of wealth is smaller when expected returns are larger. In this way, both variables exploit agents quantity decisions to learn their expectations, and exploit natural cointegrating vectors to measure long-term forecasts. For example, Cochrane (1994) 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 prices 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.

Lettau and Ludvigson (2004) show that the consumption-wealth ratio also forecasts *dividend* growth. This is initially surprising. So far, very little has forecast dividend growth. And if anything does forecast dividend growth, why is a high dividend forecast not reflected in and hence forecast by higher prices? Lettau and Ludvigson answer this puzzle by noting that the consumption-wealth ratio forecasts returns, even in the presence of D/P. 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. Conversely, if anything has additional explanatory power for returns, it must also forecast dividend growth. And it makes sense. In the bottom of a recession, both returns and dividend growth will be strong as we come out of the recession. So we end up with a new variable, and an opening for additional variables, that forecast *both* returns and cashflows, giving stronger links from macroeconomics to finance.

Statistics

Return forecastability has come with a long statistical controversy. The first round of statistical investigation asked whether the initially impressive long-horizon regressions (the extra rows of Table 1) capture any information not present in one-period regressions (The first row). Given the large persistence of the dividend yield and related forecasting variables, the answer is that, by and large, they do not.

Hodrick (1992) put the point nicely: the multiyear regression amounts to a test of the moment $E[(r_{t+1} + r_{t+2})x_t] = 0$ where x is the forecasting variable and r are log returns. But this is the same moment as a *one* year regression using a moving average right hand variable, $E[r_{t+1}(x_t + x_{t-1})]$. Given the extreme persistence of the right hand variables such as dividend yield, one can naturally see that this moment is no more powerful than $E(r_{t+1}x_t) = 0$ – none would think that lags of the dividend yield have much marginal forecast power.

Campbell and Shiller (1988) also make this point by emphasizing that multiyear regres-

sions are implied by one year regressions. If

$$\begin{aligned}x_{t+1} &= \phi x_t + v_{t+1} \\ r_{t+1} &= b x_t + \varepsilon_{t+1}\end{aligned}$$

then

$$r_{t+1} + r_{t+2} = b(1 + \phi)x_t + (\varepsilon_{t+1} + b v_{t+1} + \varepsilon_{t+2}).$$

All of the information in multiyear regressions can be recovered from one year regressions, which is what maximum likelihood would have you look at anyway. Additional important contributions include

More seriously, the t statistics in Table 1 are already not that large given the long time span. In addition, the dividend yield is very persistent, and innovations in returns are highly correlated with innovations in dividend yields, since a change in prices moves both variables. As a result, the return-forecasting coefficient inherits near-unit-root properties of the dividend yield. It is biased upward, and its t-statistic is biased towards rejection. Other forecasting variables have similar characteristics. Perhaps even the forecastability as seen in the first row is really not there in the first place. Following this idea, Goetzmann and Jorion (1993) and Nelson and Kim (1993) find the distribution of the return-forecasting coefficient by simulation, and find greatly reduced evidence for return forecastability. Stambaugh (1999) derives the finite-sample properties of the return-forecasting regression, showing the bias in the return forecast coefficient and the standard errors, and shows that the apparent forecastability disappears once one takes account of the biases. More recently, Goyal and Welch (2003, 2005) show that return forecasts based on dividend yields and a menagerie of other variables do not work out of sample. They compare forecasts in which one estimates the regression using data up to time t to forecast returns at t+1 with forecasts using the sample mean in the same period. They find that the sample mean produces a better out-of-sample prediction than do the return-forecasting regressions.

Does this mean we should abandon forecastability and go back to the random walk, i.i.d. return view of the world? I think not, since there is still not a shred of evidence that price ratios forecast *dividend* (or earning or cashflow) growth. If prices vary, they must forecast *something* – we cannot hold the view that *both* returns and dividend growth are i.i.d., since in that case price dividend ratios will be constant. Thus the *lack* of dividend forecastability is important evidence *for* return forecastability, and this is ignored in the statistical studies. Cochrane (2006b) formalizes this argument and shows that return forecastability is still highly significant, including small-sample biases, when one takes into account both pieces of evidence. (The paper also contains a more complete bibliography on this statistical issue.)

2.2 The cross-section of returns – variation across assets

Fama and French (1996) is an excellent crystallization of how average returns vary across stocks. Fama and French start by summarizing for us the “size” and “value” effects; the fact that small stocks and stocks with low market values relative to book values tend to have

higher average returns than other stocks.⁵ See the average returns in their Table 1 panel A, reproduced below.

Again, this pattern is not by itself a puzzle. High expected returns *should* be revealed by low market values (see Equation (4)). The puzzle is that the value and small firms do *not* have higher market betas. As panel B of Fama and French’s Table 1 shows, all of the market betas are about one. Market betas vary across portfolios a little more in single regressions without hml and smb as additional right hand variables, but here the result is worse: the high average return “value” portfolios have *lower* market betas.

Fama and French then explain the variation in mean returns across the 25 portfolios by variation in regression slope coefficients on two new “factors,” the hml portfolio of value minus growth firms and the smb portfolio of small minus large firms. Looking across the rest of their Table 1, you see regression coefficients b, s, h rising in Panel B where expected returns rise in Panel A. Replacing the CAPM with this “three-factor model” is the central point of Fama and French’s paper. (Keep in mind, the point of the factor model is to explain the variation in *average* returns *across* the 25 portfolios. The fact that the factors “explain” a large part of the return variance – the high R^2 in the time-series regressions of Table 1 – is not the central success of an asset pricing model.)

This argument is not as circular as it sounds. Fama and French say that value stocks earn more than growth stocks not because they *are* value stocks (a characteristic) but because they all *move with* a common risk factor. This comovement is not automatic. For example, if we split stocks into 26 portfolios based on the first letter of the ticker symbol and subtracted the market return, we would not expect to see a 95% R^2 in a regression of the A portfolio on an A-L minus M-Z “factor,” because we would expect no common movement between the A, B, C, etc. portfolios.

Stocks with high average returns *should* move together. Otherwise, one could build a diversified portfolio of high expected return (value) stocks, short a portfolio of low expected return (growth) stocks and make huge profits with no risk. This strategy remains risky and does not attract massive capital, which would wipe out the anomaly, precisely because there is a common component to value stocks, captured by the Fama-French hml factor.

Fama and French go further, showing that the size and book to market factors explain average returns formed by *other* characteristics. Sales growth is an impressive example, since it is a completely non-financial variable. Stocks with high past sales growth have lower subsequent returns (“too high prices”) than stocks with low sales growth. They do not have higher market betas, but they do have higher betas on the Fama-French factors. In this sense, the Fama French 3 factor model “explains” this additional pattern in expected returns. In this kind of application, the Fama-French 3 factor model has become the standard model replacing the CAPM for risk adjusting returns.

The Fama-French paper has also, for better or worse, defined the methodology for evaluat-

⁵These expected-return findings go back a long way, including Ball (1978), Basu (1983), Banz (1981), DeBondt and Thaler (1985), and Fama and French (1992), (1993).

Table I
Summary Statistics and Three-Factor Regressions for Simple
Monthly Percent Excess Returns on 25 Portfolios Formed on Size
and BE/ME: 7/63–12/93, 366 Months

R_f is the one-month Treasury bill rate observed at the beginning of the month (from CRSP). The explanatory returns R_M , SMB, and HML are formed as follows. At the end of June of each year t (1963–1993), NYSE, AMEX, and Nasdaq stocks are allocated to two groups (small or big, S or B) based on whether their June market equity (ME, stock price times shares outstanding) is below or above the median ME for NYSE stocks. NYSE, AMEX, and Nasdaq stocks are allocated in an independent sort to three book-to-market equity (BE/ME) groups (low, medium, or high; L, M, or H) based on the breakpoints for the bottom 30 percent, middle 40 percent, and top 30 percent of the values of BE/ME for NYSE stocks. Six size-BE/ME portfolios (S/L, S/M, S/H, B/L, B/M, B/H) are defined as the intersections of the two ME and the three BE/ME groups. Value-weight monthly returns on the portfolios are calculated from July to the following June. SMB is the difference, each month, between the average of the returns on the three small-stock portfolios (S/L, S/M, and S/H) and the average of the returns on the three big-stock portfolios (B/L, B/M, and B/H). HML is the difference between the average of the returns on the two high-BE/ME portfolios (S/H and B/H) and the average of the returns on the two low-BE/ME portfolios (S/L and B/L). The 25 size-BE/ME portfolios are formed like the six size-BE/ME portfolios used to construct SMB and HML, except that quintile breakpoints for ME and BE/ME for NYSE stocks are used to allocate NYSE, AMEX, and Nasdaq stocks to the portfolios.

BE is the COMPUSTAT book value of stockholders' equity, plus balance sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock. Depending on availability, we use redemption, liquidation, or par value (in that order) to estimate the book value of preferred stock. The BE/ME ratio used to form portfolios in June of year t is then book common equity for the fiscal year ending in calendar year $t - 1$, divided by market equity at the end of December of $t - 1$. We do not use negative BE firms, which are rare prior to 1980, when calculating the breakpoints for BE/ME or when forming the size-BE/ME portfolios. Also, only firms with ordinary common equity (as classified by CRSP) are included in the tests. This means that ADR's, REIT's, and units of beneficial interest are excluded. The market return R_M is the value-weight return on all stocks in the size-BE/ME portfolios, plus the negative BE stocks excluded from the portfolios.

Size	Book-to-Market Equity (BE/ME) Quintiles									
	Low	2	3	4	High	Low	2	3	4	High
Panel A: Summary Statistics										
	Means					Standard Deviations				
Small	0.31	0.70	0.82	0.95	1.08	7.67	6.74	6.14	5.85	6.14
2	0.48	0.71	0.91	0.93	1.09	7.13	6.25	5.71	5.23	5.94
3	0.44	0.68	0.75	0.86	1.05	6.52	5.53	5.11	4.79	5.48
4	0.51	0.39	0.64	0.80	1.04	5.86	5.28	4.97	4.81	5.67
Big	0.37	0.39	0.36	0.58	0.71	4.84	4.61	4.28	4.18	4.89

Figure 1: Fama and French (1996) Table 1

ing asset pricing models for the last 10 years. A generation of papers studies the Fama-French 25 size and book to market portfolios to see whether alternative factor models can explain their average returns. Empirical papers now routinely form portfolios by sorting on other characteristics, and then run time-series regressions like Fama and French's to see which factors explain the spread in average returns, as revealed by small regression intercepts.

Most importantly, where in the 1980s papers would focus entirely on the probability value of some overall statistic, Fama and French rightly got people to focus on the spread in average returns, the spread in betas, and the economic size of the pricing errors. Remarkably, this, the most successful model since the CAPM, is decisively *rejected* by formal tests. Fama

Table 1—Continued										
Size	Book-to-Market Equity (BE/ME) Quintiles									
	Low	2	3	4	High	Low	2	3	4	High
Panel B: Regressions: $R_i - R_f = a_i + b_i(R_M - R_f) + s_iSMB + h_iHML + e_i$										
	a					t(a)				
Small	-0.45	-0.16	-0.05	0.04	0.02	-4.19	-2.04	-0.82	0.69	0.29
2	-0.07	-0.04	0.09	0.07	0.03	-0.80	-0.59	1.33	1.13	0.51
3	-0.08	0.04	-0.00	0.06	0.07	-1.07	0.47	-0.06	0.88	0.89
4	0.14	-0.19	-0.06	0.02	0.06	1.74	-2.43	-0.73	0.27	0.59
Big	0.20	-0.04	-0.10	-0.08	-0.14	3.14	-0.52	-1.23	-1.07	-1.17
	b					t(b)				
Small	1.03	1.01	0.94	0.89	0.94	39.10	50.89	59.93	58.47	57.71
2	1.10	1.04	0.99	0.97	1.08	52.94	61.14	58.17	62.97	65.58
3	1.10	1.02	0.98	0.97	1.07	57.08	55.49	53.11	55.96	52.37
4	1.07	1.07	1.05	1.03	1.18	54.77	54.48	51.79	45.76	46.27
Big	0.96	1.02	0.98	0.99	1.07	60.25	57.77	47.03	53.25	37.18
	s					t(s)				
Small	1.47	1.27	1.18	1.17	1.23	39.01	44.48	52.26	53.82	52.65
2	1.01	0.97	0.88	0.73	0.90	34.10	39.94	36.19	32.92	38.17
3	0.75	0.63	0.59	0.47	0.64	27.09	24.13	22.37	18.97	22.01
4	0.36	0.30	0.29	0.22	0.41	12.87	10.64	10.17	6.82	11.26
Big	-0.16	-0.13	-0.25	-0.16	-0.03	-6.97	-5.12	-8.45	-6.21	-0.77
	h					t(h)				
Small	-0.27	0.10	0.25	0.37	0.63	-6.28	3.03	9.74	15.16	23.62
2	-0.49	0.00	0.26	0.46	0.69	-14.66	0.34	9.21	18.14	25.59
3	-0.39	0.03	0.32	0.49	0.68	-12.56	0.89	10.73	17.45	20.43
4	-0.44	0.03	0.31	0.54	0.72	-13.98	0.97	9.45	14.70	17.34
Big	-0.47	0.00	0.20	0.56	0.82	-18.23	0.18	6.04	18.71	17.57
	R ²					s(e)				
Small	0.93	0.95	0.96	0.96	0.96	1.97	1.49	1.18	1.13	1.22
2	0.95	0.96	0.95	0.95	0.96	1.55	1.27	1.28	1.16	1.23
3	0.95	0.94	0.93	0.93	0.92	1.44	1.37	1.38	1.30	1.52
4	0.94	0.92	0.91	0.88	0.89	1.46	1.47	1.51	1.69	1.91
Big	0.94	0.92	0.87	0.89	0.81	1.19	1.32	1.55	1.39	2.15

Figure 2: Fama and French (1996) Table 1 continued.

and French taught us to pay attention to more important things than test statistics.

Macro modelers have gotten into the habit of evaluating models on the Fama-French 25 portfolios, just as Fama and French did. I think that in retrospect, this is a misreading of the point of Fama and French's paper. The central point of the paper is that all of the important cross-sectional information in the 25 portfolios is captured by the three factor portfolios. This is true both of returns and expected returns. One could state the result that there are three dominant eigenvalues in the covariance matrix of the 25 portfolios, that explain the vast majority of the correlation structure of the portfolios, and expected returns are almost completely described by betas on these three portfolios.

To the extent that the Fama-French three-factor model is successful in describing average returns, macro-modelers need only worry about why the value (hml) and small-large (smb) portfolio have expected returns. Given these factors, the expected returns of the 25 portfolios (and any other portfolios that are explained by the three-factor model) follow automatically. The point of the 25 portfolios is to show “nonparametrically” that the three factor portfolios account for all information in stocks sorted by size and book to market. The point of the 25 portfolios is *not* to generate a good set of portfolios that captures 25 degrees of freedom in the cross section of all stocks. There are really not 25 degrees of freedom in the Fama-French portfolios, there are 3 degrees of freedom. This is very bad news for models that explain the Fama-French portfolios with 4,5, and sometimes 10 factors! This is the central point of Daniel and Titman (2005) and Lewellen, Nagel, and Shanken (2006).

The Fama-French model is rejected in the 25 portfolios, however. The rejection of the three-factor model in the 25 portfolios is caused primarily by small growth portfolios, and Fama and French’s Table 1 shows the pattern. Small growth stocks earn about the same average returns as large growth portfolios – see Table 1 “means” left column – but they have much larger slopes s . A larger slope that does not correspond to a larger average return generates a pricing error a . In addition, the R^2 are so large in these regressions, and the residuals correspondingly so small, that economically small pricing errors are statistically significant. $\alpha'\Sigma^{-1}\alpha$ is large if α is small, but Σ is even smaller. A fourth “small growth - large value” factor eliminates this pricing error as well, but I don’t think Fama and French take the anomaly that seriously.

For the division of labor and the use of 25 portfolios, however, this fact means that models which improve on the Fama-French factors in the 25 Fama-French portfolios do so by better pricing the small-growth puzzle and other very small discrepancies of the model. One must ask whether those discrepancies are at all meaningful.

The Fama-French model seems to take us away from economic explanation of risk premia. After all, hml and smb are just other portfolios of stocks. Fama and French speculate suggestively on the macroeconomic foundations of the value premium (p. 77):

One possible explanation is linked to human capital, an 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.

A large body of empirical research asks whether the size and book to market factors do in fact represent macroeconomic phenomena via rather a-structural methods. 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 Lakonishok, Shleifer and Vishny (1994) did not find. However, in the 1990s and early 2000s, value stocks have moved much more closely with the aggregate economy, so more recent estimates do show a significant and heartening link between value returns and macroeconomic conditions. In this context, 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.

The Fama-French paper closes with a puzzle. Though the three-factor model captures the expected returns from many portfolio sorts, it fails miserably on momentum. If you form portfolios of stocks that have gone up in the last year, this portfolio continues to do well in the next year and vice versa (Jegadeesh and Titman, 1993, see Fama and French’s Table VI). Again, this result by itself would not be a puzzle, *if* the “winner” portfolio had higher market, smb, or hml betas than the loser portfolios. Alas, (Fama and French Table VII) the winner portfolio actually has *lower* slopes than the loser portfolio; winners act, sensibly enough, like high-price growth stocks that should have low mean returns in the three factor model. The three factor model is worse than useless at capturing the expected returns of this “momentum” strategy, just as the CAPM is worse than useless at explaining the average returns of book-to-market portfolios.

Now, the returns of these 10 momentum-sorted portfolios *can* be explained by an additional “momentum factor” umd of winner stocks less loser stocks. You cannot form a diversified portfolio of momentum stocks and earn high returns with no risk; a common component to returns shows up once again. Yet Fama and French did not take the step of adding this fourth factor, and thus claiming a model that would explain all the known anomalies of its day.

This reluctance is understandable. First, Fama and French worry (p. 81) whether the momentum effect is real. They note that the effect is much weaker before 1963, and call for more out-of-sample verification. They may also have worried that the effect would not survive transactions costs. Exploiting the momentum anomaly requires high frequency trading, and shorting small losing stocks can be difficult. Equivalently, momentum is, like long-horizon regression, a way to enhance the *economic* size of a well-known *statistical* anomaly, as a tiny positive autocorrelation of returns can generate the observed momentum profits. Last year’s 1/10 best winners typically have gone up a tremendous amount, often 100% or more. It only takes a small, 0.1 or less autocorrelation or 0.01 forecasting R^2 to turn such past returns to

10% expected future returns. (See Cochrane 1999 for a more detailed calculation.) Can one really realize profits that result from 0.01 forecast R^2 ? Second, having just swallowed hml and smb, one might naturally be reluctant to add a new factor for every new anomaly, and to encourage others to do so. Third, and perhaps most importantly, Fama and French had at least a good story for the macroeconomic underpinnings of size and value effects, as expressed in the above quotation. They had no idea of a macroeconomic underpinning for a momentum premium, and in fact in their view (p. 81) there isn't even a coherent *behavioral* story for such a premium. They know that having some story is the only "fishing license" that keeps one from rediscovering the Roll theorem. Still, they acknowledge (p. 82) that if the effect survives scrutiny, another "factor" may soon be with us.

In the time since Fama and French wrote, many papers have examined the momentum effect in great detail. I do not survey that literature here, since it takes us away from our focus of macroeconomic understanding of premia rather than exploration of the premia themselves. However, momentum remains an anomaly.

One can begin to imagine macroeconomic stories for momentum. Good cash-flow news could bring growth-options into the money, and this event could increase the systematic risk (betas) of the winner stocks. Of course, then a good measure of "systematic risk" and good measurements of conditional betas should explain the momentum effect.

Momentum is correlated with value, so it's tempting to extend a macroeconomic interpretation of the value effect to the momentum effect. Alas, the sign is wrong. Last year's winners act like growth stocks, but they get high, not low, average returns. Hence, the component of a momentum factor orthogonal to value must have a very high risk premium, and its variation is orthogonal to whatever macroeconomic effects underlie value.

In any case, the current crop of papers that try to measure macroeconomic risks follow Fama and French by trying to explain the value and size premium, or the Fama-French 25 portfolios, and so far largely exclude the momentum effect. The momentum factor is much more commonly used in performance evaluation applications, following Carhart (1997). In order to evaluate whether, say, fund managers have stock-picking skill, it does not matter whether the factor portfolios correspond to real risks or not, and whether the average returns of the factor portfolios continue out of sample. One only wants to know whether a manager did better in a sample period than a mechanical strategy.

I suspect that if the momentum effect survives its continued scrutiny, macro-finance will add momentum to the list of facts to be explained. A large number of additional expected-return anomalies have also popped up, which will also make it to the macro-finance list of facts if they survive long enough. We are thus likely to face many new "factors." After all, each new expected-return sort *must* either fall in to one of the following categories. 1) A new expected-return sort might be explained by betas on existing factors, so once you understand the existing factors you understand the new anomaly, and it adds nothing. This is how, for example sales growth behaves for the Fama-French model. 2) The new expected-return sort might correspond to a new dimension of comovement in stock returns, and thus be "explained" (maybe "summarized" is a better word) by a new factor. 3) If a new expected-

return sort does not fall into 1 and 2, it corresponds to an arbitrage opportunity, which is most unlikely to be real, and if real to survive longer than a chicken in a crocodile pond. Thus, any expected return variation that is both real and novel must correspond to a new “factor.”

3 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 equal to the marginal utility of consumption – a marginal dollar spent gives the same utility as a marginal dollar saved – and our basic asset pricing equation (3) becomes⁶

$$E_t(R_{t+1}^{ei}) = -cov_t \left(R_{t+1}^e, \frac{u'(c_{t+1})}{u'(c_t)} \right), \quad (5)$$

or, with the popular power utility function $u'(c) = c^{-\gamma}$, (or using that form as a local approximation)

$$E_t(R_{t+1}^{ei}) = \gamma \times cov_t \left(R_{t+1}^e, \frac{c_{t+1}}{c_t} \right). \quad (6)$$

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 starting with

⁶In discrete time, the actual equation is

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

with

$$\frac{1}{R_t^f} \equiv E_t \left[\beta \frac{u'(c_{t+1})}{u'(c_t)} \right].$$

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

the Permanent Income Hypothesis. In *every* formal derivation of the CAPM, ICAPM, and every 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. They are all special cases of the consumption-based model, not alternatives to it.

The equity premium puzzle points out that this consumption-based model cannot 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 exogenously given. We are asking what are the economics behind the mean market return.) From (6) write

$$E(R^{ei}) = \gamma \sigma(R^{ei}) \sigma(\Delta c) \rho(\Delta c, R^{ei}) \quad (7)$$

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

$$\frac{\|E(R^{ei})\|}{\sigma(R^{ei})} < \gamma \sigma(\Delta c). \quad (8)$$

The left hand side of (8) is the “Sharpe ratio” a common 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 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 initial reaction is that the problem is not so much high stock average returns but low interest rates. Perhaps something is wrong with *bonds*, perhaps traceable to monetary policy, liquidity, etc. Alas, this solution does not work. The key to the calculation in (8) is the Sharpe ratio on the left hand side. There are large Sharpe ratios *between* stocks (as in the value - growth premium studied by Fama and French) ignoring bonds all together. High sample Sharpe ratios are pervasive in finance and not limited to the difference between stocks and bonds.

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

$$r^f = \delta + \gamma E(\Delta c) - \frac{1}{2} \gamma (\gamma + 1) \sigma^2(\Delta c). \quad (9)$$

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 δ . Worse, (9) 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 $\rho = 0.2$ in postwar quarterly data, i.e. using (7) or using the sample covariance in (6), raises the required risk aversion by a factor of 5, to 165! Even using $\rho = 0.41$, the largest correlation among many consumption definitions (you get this with 4th quarter to 4th quarter real chain-weighted nondurable consumption) the required risk aversion rises to $33/0.41 = 80$.

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, 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. 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. You don’t notice this prediction though unless you ask for the implicit consumption volatility and you check it against consumption data.

Equivalently, the standard optimal portfolio calculation says that the weight in risky assets should be

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

Using an 8% mean and a 16% standard deviation, this calculation predicts 100% equities ($w = 1$) at $\gamma = 0.08/0.16^2 = 3.125$, which seems like a nice sensible risk aversion. (In fact, this calculation was often cited (mis-cited, in my view) as evidence for low risk aversion.) The problem with the calculation is that the standard portfolio model also says consumption should be proportional to wealth, and thus consumption should also have a 16% standard deviation.

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. It means that time-varying expected returns, "excess" stock volatility and the equity premium puzzle are all linked in ways that are still not fully exploited.

Mehra and Prescott and the puzzle

The ink spilled on the equity premium would sink the Titanic, so there is no way here to do justice to all who contributed to or extended the puzzle, or even to summarize the huge literature.

My quick overview takes the approach of Cochrane and Hansen's (1992) review paper "Asset Pricing Explorations for Macroeconomics." The fundamental idea there, equation (8) is due to Shiller (1982) (see p. 221) and much elaborated on by Hansen and Jagannathan (1991), who also provide many deep insights into the representation of asset prices. 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) gives a review 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 calculated in this "calibrated economy" is much too low unless risk aversion is raised to apparently implausible values (55, in this case).

There is some question who should get credit for discovering the equity premium puzzle. Grossman and Shiller (1981), published in Grossman, Melino and Shiller (1987) reported that risk aversion estimates in consumption models seemed strangely high. 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 credit for the point to Grossman and Shiller.

It's interesting that Mehra and Prescott's more complex approach was so much more influential. In retrospect, I think we can see the equity premium puzzle much more clearly with the simple manipulations of first order conditions as above, which is also Shiller's (1982) early approach. In part, it seems that Mehra and Prescott were the first to realize the *importance* of what they found. Grossman and Shiller dismiss their finding of high risk aversion as "preliminary," unbelievable, and probably the spurious result of some puzzling recent data. 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 implausibly 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..." Hansen and Singleton simply report high risk aversion estimates, but seem to regard them as not particularly interesting results of inefficient (since they leave out instruments) estimates. Hansen and Singleton describe the crucial Table 5 thus:

"Consistent with their [Grossman and Shiller's] results, we found $\|\hat{\alpha}\|$ [risk aversion, γ in the above notation] 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. "

Mehra and Prescott argue instead that high risk aversion is a robust and unavoidable feature of any method for matching the model to data, and that the puzzle is important because it will require fundamental changes in macroeconomic modeling. Compare the previous quotes to these, from the first page of Mehra and Prescott:

"This result is robust to model specification and measurement problems." "The question addressed in this paper is whether this large differential in average yields can be accounted for by models that abstract from transactions costs, liquidity constraints and other frictions absent in the Arrow-Debreu setup. Our finding is that it cannot be, at least not for the class of economies considered. Our conclusion is that most likely some equilibrium model with a friction will be the one that successfully accounts for the large average equity premium. "

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. The approach I followed above, playing with simple first order conditions, similarly took off again following Hansen and Jagannathan's (1991) paper, that gave much more structure to play with than Shiller's (1982) simple calculations, though in the end we seem to have come back to that starting point. In finance as elsewhere, identifying, marketing and packaging the insight, and leaving a structure that others can play with, are justly important contributions.

Mehra and Prescott's general equilibrium modeling imposes extra discipline on this kind of research, and in many ways is also crucially important as the progenitor of the general equilibrium models described below. In a general equilibrium model, the covariance of consumption with returns is generated endogenously. You can't just take $cov(R, \Delta c)$ as given and crank up γ (see (6)) to get any premium you want. Thus, 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. Working out a general equilibrium model, one also notices that many other predictions go awry. For example, Mehra and Prescott's model does not generate nearly enough return *variance*, and measures to increase the equity premium or return variance dramatically and counterfactually increase the variation in the risk free rate over time. These basic moments remain quite difficult for general equilibrium models to capture, but you cannot notice they are a problem if you only look at first-order conditions.

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. These preferences, described in more detail below, break the link between risk aversion and intertemporal substitution, so there is no connection to a “risk-free rate” puzzle any more, and we can coherently describe the data with high risk aversion. 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 ever 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, past the 6% average in longer samples, down to 2 or 3% or less. The US economy and others with high sample equity premia may simply have been lucky. Did people in 1947 really think that the stock market would gain 8% per year more than bonds, and shy away from buying more stocks in the full knowledge of this mean, because the 16% annual standard deviation of stock returns seemed like too much risk? Or was the 8% mean return largely a surprise?

Putting the argument a little more formally, we can separate the achieved average stock return into 1) the initial dividend yield (dividend payment/initial price) 2) increases in the price/dividend ratio and 3) growth in dividends, giving growth in prices at the same price/dividend ratio. Dividend yields were about 4%, and have declined to about 2%. Dividend yields are known ahead of time, so cannot contribute to a “surprise” return. The price/dividend ratio has about doubled in the postwar era, and this increase could well be a surprise. But this doubling happened over 50 years, contributing only 1.4% (compounded; $2^{1/50} = 1.014$) to the equity return. If there is a surprise, then, the surprise is that *economic growth* was so strong in the postwar era, resulting in surprisingly strong *dividend growth*. (In the long run, *all* of the return must be dividend growth since price/dividend ratios are stationary) And of course economic growth *was* surprisingly good in the postwar era. Most people in 1947 expected a return to depression.

For these reasons, as well as perhaps simple boredom in the face of intractable questions, research attention is moving to understanding stock 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.

4 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 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.

We start here with Hansen and Singleton’s (1982, 1984) classic investigation of the consumption-based model. Alas, they decisively reject the model; among other things they find “equity premium puzzle” result that the model cannot explain the spread between stock and bond returns with low interest rates.

The following 20 years have seen an enormous 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 as instruments to the use of size, book/market and momentum portfolios, and to the dividend price ratio, term spreads and other more powerful instruments. 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.” At some point “bad times” must be mirrored in a decision to cut back on consumption.

4.1 Hansen and Singleton; power utility

The classic consumption-based model test is due to Hansen and Singleton (1982, 1984). The influence of this paper is hard to overstate. It gives a clear exposition of the GMM methodology, which has pretty much taken over estimation and testing. (At least it has for me. *Asset Pricing*, Cochrane 2004 maps all standard asset pricing estimates into GMM and shows how they can and should be easily generalized using GMM to account for heteroskedasticity and autocorrelation.) Also with this work (generalizing Hall’s 1978 test for a random walk in consumption) macroeconomists and financial economists 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,

$$E_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} R_{t+1}^i \right] = 1. \quad (10)$$

The method is astonishingly simple. Multiply both sides of (10) by instruments – any variable z_t observed at time t – and take unconditional expectations, yielding

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

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 whether movements in returns predictable by some instrument z_t – as in regressions of R_{t+1} on z_t – are matched by movements in consumption growth or by the product of consumption growth and returns as predicted by the same instrument. The results give 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 Hansen and Singleton’s instruments – 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,

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

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 estimates of (11) with a single asset and many instruments 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 Hansen and Singleton’s (1984) Table III 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 (34.06)	1.001 (0.0462)	Nondurable	0	Just identified	
2	0.205	0.999	Nondurable	4	170.25 (0.9999)	24
3	58.25 (66.57)	1.088 (0.0687)	ND & Services	0	Just Identified	
4	0.209	1.000	ND& Services	4	366.22 (0.9999)	24

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, in the lags = 4 rows, 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, as in Hansen and Singleton (1984). That huge test statistic tells us that there is a tension over the value of γ . The value of γ 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.

Risk aversion and intertemporal substitution – more recent estimates

The fact that quite high risk aversion is required to digest the equity premium is robust in consumption-based model estimation, as the equity premium discussion below makes

clear. 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.

A good deal of the more recent macro 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!

4.2 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 $\partial u(c, x)/\partial c = v'(c)$ 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 data, so business-cycle risk in stock returns may correlate better with these variables than with nondurable and services consumption.

One problem with this generalization is that we don't have much intuition for which way the effect should 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 nonseparability is to posit a separable utility over the services, and a durable goods stock that depreciates over time;

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

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 pizza 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

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

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.

A difficulty in adding multiple goods is that, if the nonseparability is strong enough to affect asset prices, it tends to affect other prices as well. People start to care a lot about the composition of their consumption stream. Therefore, if we hold quantities fixed (as in the endowment-economy GMM tradition), such models tend to predict lots of relative price and interest-rate variation; if we hold prices fixed such models tend to predict lots of quantity variation, including serial correlation in consumption growth. An investigation with multiple goods needs to include the first order condition for allocation *across* goods, and this often causes trouble.

Finally, utility could be nonseparable *across states of nature*. Epstein and Zin (1991) pioneered this idea in the asset-pricing literature. The expected utility function adds over

states, just as separable utility adds over goods,

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

Epstein and Zin propose a recursive formulation of utility

$$U_t = \left((1 - \beta) c_t^{1-\rho} + \beta [E_t (U_{t+1}^{1-\gamma})]^{\frac{1-\rho}{1-\gamma}} \right)^{\frac{1}{1-\rho}}. \quad (13)$$

(I use Hansen, Heaton and Li's (2005) notation) that among other things abandons separability across states of nature. The term $[E_t (U_{t+1}^{1-\gamma})]^{\frac{1}{1-\gamma}}$ is sometimes called a “risk adjustment” or the “certain equivalent” of future utility. The Epstein-Zin formulation separates the coefficient of risk aversion γ from the inverse of the elasticity of intertemporal substitution ρ . Equation (13) reduces to power utility for $\rho = \gamma$. Models with non-*time* separable utilities (habits, durables) also distinguish risk aversion and intertemporal substitution, but not in such a simple way.

The stochastic discount factor/marginal rate of substitution is

$$m_{t+1} = \beta \left(\frac{c_{t+1}}{c_t} \right)^{-\rho} \left(\frac{U_{t+1}}{[E_t (U_{t+1}^{1-\gamma})]^{\frac{1}{1-\gamma}}} \right)^{\rho-\gamma}. \quad (14)$$

(The appendix contains a short derivation.) If $\rho \neq \gamma$, we see a second term; expected returns will depend on covariances with changes in the utility index, capturing news about the investor's future prospects, as well as on covariances with consumption growth. As we will see, a large number of modifications to the standard setup lead to a marginal rate of substitution that is the old power formula times a multiplicative new piece.

The utility index itself is not directly measurable, so to make this formula operational we need some procedure for measurement. It turns out that the utility index is proportional to the value of the wealth portfolio (the claim to the consumption stream), so one can write the discount factor

$$m_{t+1} = \beta^{1+\gamma-\rho} (R_{t+1}^W)^{\frac{\rho-\gamma}{1-\rho}} \left(\frac{c_{t+1}}{c_t} \right)^{-\rho \left(\frac{1-\gamma}{1-\rho} \right)}. \quad (15)$$

(This formula is also derived in the appendix.) This effect provides a route to including stock returns in the asset pricing model alongside consumption growth, which of course can give a much improved fit. This was the central theoretical and empirical point of Epstein and Zin (1991). However, this modification stands a bit on shaky ground: the substitution only works for the entire wealth portfolio (claim to future consumption), including nontraded assets such as real estate and the present value of labor income, not the stock market return alone. Furthermore, wealth and consumption do not move independently; news about consumption growth moves the wealth return.

To emphasize the latter point, Restoy and Weil (1988, p. 10) show how to approximate

the wealth return by a discounted sum of future consumption, leading to a formula valid near $\rho = 1$,

$$(E_{t+1} - E_t) \log m_{t+1} \approx (E_{t+1} - E_t) \left[-\rho \Delta \log c_{t+1} + (\rho - \gamma) \sum_{j=0}^{\infty} \beta^j \Delta \log c_{t+j} \right] \quad (16)$$

where small letters denote logs of capital letters. *Future* long-horizon consumption growth enters the *current* period marginal rate of substitution. In essence, this formulation tracks the wealth-portfolio return or utility index term back to its source in future values of consumption. Thus, variables that predict future consumption growth will appear as additional risk factors even with (perfectly measured) current consumption growth.

4.3 Empirics with new utility functions

Nonseparabilities across goods

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

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

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

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

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.⁷

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. Also, any model with multiple goods gives rise to an *intra* temporal first order condition, marginal utility of nondurables / marginal utility of durables = relative price. 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

⁷Lettau (2003) footnote 2 points out that consumption and leisure are negatively correlated (people work and consume more in expansions). The product $c \times l$ and the resulting marginal rate of substitution is then typically less volatile than with c alone, making the equity premium puzzle worse. However, the greater correlation of labor with asset returns may still make asset pricing work better, especially if one admits a large risk aversion coefficient.

cross-sectional risk premia.

In the consumption-based revival, 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. He estimates a larger elasticity of intertemporal 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 . 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).

Habits

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

$$u(c_t - bc_{t-1}). \quad (17)$$

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 (even a Pizza is durable at a one-minute horizon) and habit persistence at longer horizons. However, he only considers the value-weighted stock market and T-bill rate as assets.

Campbell and Cochrane (1999) adapt a habit persistence model to generate a number of asset pricing facts. 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),

$$X_t = \phi X_{t-1} + \lambda C_t. \quad (18)$$

(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 us 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. But it seems more sensible that habits move slowly in response to consumption experience rather than with the one-period lag of many specifications. In addition, slow-moving habits will generate the slow-moving state variables we seem to see in return forecastability.

We specify a nonlinear version of (18). 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 (18) 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 (18) 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.

This sort of reverse-engineering is important in a wide variety of models. Devices that increase the volatility of the discount factor or marginal rate of substitution *across states of nature* $\sigma_t(m_{t+1})$, to generate a large equity premium, also tend to increase the volatility of the marginal rate of substitution *over time* $\sigma(E_t(m_{t+1}))$, thus generating counterfactually large interest rate variation. To be empirically plausible, it takes some care to set up a model so that it has a lot of the former variation with little of the latter.

We examine 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, as do Constantinides (1990), Abel (1990), and Sundaresan’s (1989) habit persistence investigations. We let aggregate consumption follow a random walk, we calibrate the model to match sample means including the equity premium, and we then compare the behavior of common time-series tests in our artificial data to their outcome in real 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.

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 change in the ratio of consumption to habit as well as on consumption growth,

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \left(\frac{S_{t+1}}{S_t} \right)^{-\gamma}, \quad (19)$$

where $S_t = (C_t - X_t)/C_t$ and X_t is habit. As the time period lengthens, the latter effect becomes more important. The basic question is, “why do people fear stocks so much?” This model’s 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 decline in future recessions, times when consumption falls low relative to habits (S).

A large number of models amount to something like Equation (19), in which the discount factor generalizes the power-utility case by adding another state variable. There is a danger in such models that they often work well for short run returns, but not in the long run. The trouble is that S is stationary, while consumption of course is a random walk. Now, to generate an equity premium or any large Sharpe ratio, we need a large volatility of the discount factor $\sigma(M)$, and to generate an equity premium in long-run returns we need the variance of the discount factor to increase linearly with horizon. If the second term $S^{-\gamma}$ is stationary, it may contribute a lot to the volatility of one-period discount factors, but in the long run we will be right back to the power utility model and all its problems, since the variance of a stationary variable approaches a limit while the variance of the random walk consumption component increases without bounds. This turns out not to be a problem for the Campbell-Cochrane model: while S_t is stationary, $S_t^{-\gamma}$ is not, and its conditional variance grows without bound. Thus, at *any* horizon the equity premium is generated by covariance with $S^{-\gamma}$ not so much by covariance with consumption growth. This result stems from our nonlinear habit accumulation process. It may not be there in many superficially attractive simplifications or linearizations of the habit model.

Simulation is a prequel to empirical work, not a substitute, so this sort of model needs to be evaluated in a modern cross-sectional setting, for example in the Fama French 25 size and book/market portfolios. Surprisingly, no one has tried this (including Campbell and myself). The closest effort is Chen and Ludvigson (2004). They evaluate a related habit model using the Fama-French 25 size and book/market portfolios. They use a “nonparametric” (really, highly parametric) three-lag version of the MA habit specification (17) rather than the slow-moving counterpart (18). Comparing models based on Hansen-Jagannathan (1997) distance, which is a sum of squared pricing errors weighted by the inverse of the second-moment matrix of returns, they find the resulting consumption-based model performs quite well, even better than the Fama-French three-factor model. Within this structure, they find the “internal habit” version of the model performs better than the “external habit” version in which each person’s habit is set by the consumption of his neighbors. (I add the qualifier “within this structure” because in other structures internal and external habits are observationally indistinguishable.) The “internal habit” specification may be able to exploit the correlation of returns with subsequent consumption growth, which is also the key to Parker and Julliard (2005), discussed below.

Wachter (2004) extends 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.

Verdelhan (2004) extends the habit model to foreign exchange premia. Here the puzzle

is that high foreign interest rates relative to domestic interest rates signal to higher returns in foreign bonds, even after including currency risk. His explanation is straightforward. The first part of the puzzle is, why should (say) the Euro/dollar exchange rate covary with US consumption growth, generating a risk premium? His answer is to point out that in complete markets the exchange rate is simply determined by the ratio of foreign to domestic marginal utility growth, so the correlation pops out naturally. The second part of the puzzle is, why should this risk premium vary over time? In the habit model, recessions, times when consumption is close to habit, are times of low interest rates, and also times of high risk premium (people are more risk averse when consumption is near habit.) Voilà, 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.

Related models

The essence of these models really does not hinge on habits per se, as a large number of microeconomic mechanisms can give rise to a discount factor of the form (19), where C is aggregate consumption and S is a slow moving business cycle related state variable. Constantinides and Duffie (1996), discussed below, generate a discount factor of the form (19), in a model with power utility but idiosyncratic shocks. The “S” component is generated by the cross-sectional variance of the idiosyncratic shocks.

In Piazzesi, Schneider and Tuzel (2004), the share of housing consumption in total consumption plays the role of 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

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\sigma}} \left(\frac{\alpha_{t+1}}{\alpha_t} \right)^{\frac{\varepsilon - \sigma}{\sigma(\varepsilon - 1)}}. \quad (20)$$

Here, α is the expenditure share of non-housing services, which varies slowly over the business cycle just like S in (19). Housing services are part of the usual nondurable and services aggregate of course; the paper essentially questions the accuracy of price indices used to aggregate housing services into overall services.

Does more housing raise or lower the marginal utility of other consumption, and do we trust this effect? Piazzesi, Schneider and Tuzel calibrate the elasticity of substitution ε from the behavior of the share and relative prices, exploiting 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. This does not seem like an extreme value. As (20) shows though, whether the housing share enters positively or negatively in marginal utility depends on the substitutability of consumption over time and states, σ as well as the substitutability of housing for other consumption ε . Like others, they calibrate to a relatively large risk premium, hence small σ . This calibration means that the housing share enters negatively in the marginal rate of substitution; a lower housing share makes you “hungrier” for other consumption.

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 they fit an AR(1) to the housing share. They then simulate artificial data on the stock price as a levered claim to consumption. The model works very much like the Campbell-Cochrane model. Expected returns are high, matching the equity premium, because investors are afraid that stocks will fall when the housing share 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 and from housing share ratio. They verify the latter prediction, adding to the list of macro variables that forecast returns. (See Table 4 and Table 5). Finally, the model generates slow-moving, variation in price-dividend ratios and stock return volatility, all coming from risk premia rather than dividend growth. However, the second term is stationary in their model, so it is likely that this model does not produce a long-run equity premium.

Lustig and Van Nieuwerburgh (2004a, 2004b) explore a similar model. Here, variations in housing collateral play the role of the “habit.” Consumer-investor (-homeowners) whose housing collateral declines become effectively more risk averse. Lustig and Van Nieuwerburgh show that variations in housing collateral predict stock returns in the data, as the surplus consumption ratio predicts stock returns in the Campbell-Cochrane model. They also show that a conditional consumption CAPM using housing collateral as a conditioning variable explains the value-size cross sectional effects, as implied by their model, in the same manner as with the Lettau-Ludvigson (2001a,b) cay state variable.

Raj Chetty and Adam Szeidl (2004) show how consumption commitments mimic habits. If in good times you buy a house, it is difficult to unwind that decision in bad times. Non-housing consumption must therefore decline disproportionately. Chetty and Szeidl show that this mechanism mimics habits in aggregate consumption. They also show that people who have recently moved for exogenous reasons hold a smaller proportion of stocks, acting in more risk-averse manner.

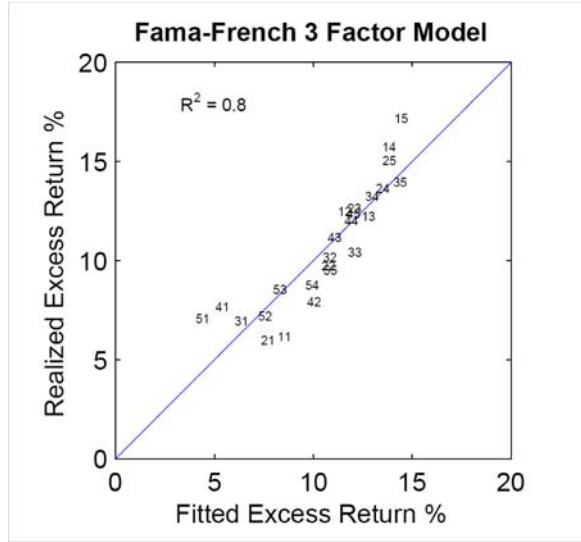
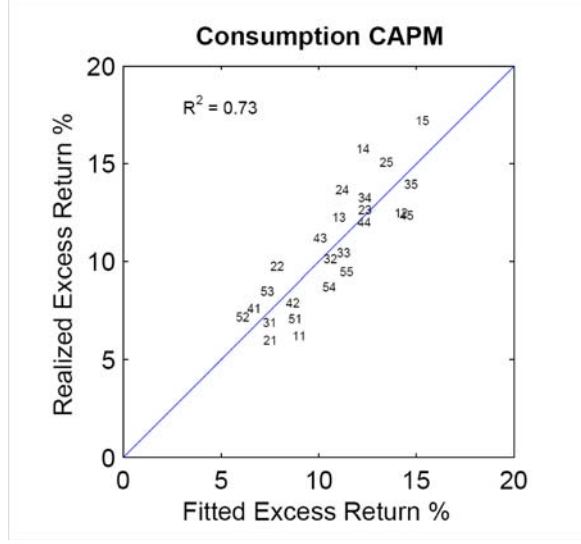
Long horizons

Nobody expects the consumption-based model (and data) to work at arbitrarily high frequencies. We do not 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 they are with contemporaneous macro variables is suggestive. The data *definitions* break down at high frequency too. Clothing is “nondurable.”

In sum, at some high frequency, we expect consumption and return data to be de-linked. Conversely, at some low enough frequency, we know consumption and stock market values must move one for one; both must eventually track the overall level of the economy and the consumption/wealth ratio will neither grow without bound nor decline to zero. Thus, some form of the consumption model may well hold at a long-enough horizon. 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 aggregate 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. Brainard, Nelson, and Shapiro (1991) show that the consumption CAPM performance gets better in some dimensions at longer horizons. However, these greater correlations do not mean the model is a total success, as other moments still do not line up. For example, Cochrane and Hansen (1992) find that long-horizon consumption performs worse in Hansen-Jagannathan bounds. There are fewer consumption declines in long-horizon data, and the observation that $(C_{t+k}/C_t)^{-\gamma}$ can enter a Hansen-Jagannathan bound at high risk aversion depends on consumption declines raised to a large power to bring up the mean discount factor and solve the risk free rate puzzle.

Most recently and most spectacularly, Jagannathan and Wang (2005) find that by using fourth quarter to fourth quarter nondurable and services consumption, the simple consumption based model can account for the Fama-French 25 size and book/market portfolios. The figure below captures this result dramatically. On reflection, this is a natural result. A lot of purchases happen at Christmas, and with an annual planning horizon. Time aggregation and seasonal adjustment alone would make it unlikely that monthly average consumption would line up with end of month returns. 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). Of course, not *everything* works. The model is linearized (Jagannathan and Wang examine average returns vs. betas on consumption growth), the slope coefficient of average returns on betas does imply an admittedly rather high risk aversion coefficient, and there are still many moments for which the model does not work. But it is a delightful sign that at least one sensible moment does work, and delightful to see an economic connection to the puzzling value premium.



Top panel: Average returns of Fama-French 25 portfolios vs. predictions of the linearized consumption-based model (essentially, consumption betas) and vs. predictions of the Fama-French 3 factor model. Fourth-quarter to fourth-quarter data, 1954-2003. Source: Jagannathan and Wang (2005) Figure 2.

Parker and Julliard (2005) similarly examine whether size and book to 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 multiperiod moment condition

$$1 = E_t \left[\beta^k \left(\frac{C_{t+k}}{C_t} \right)^{-\gamma} R_{t+1} R_{t+1}^f R_{t+2}^f \dots R_{t+k-1}^f \right]. \quad (21)$$

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. 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.

Lustig and Verdelhan (2004) do a standard consumption-beta test on foreign exchange returns at an annual horizon, and find, surprisingly, 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.

Epstein and Zin and the long run

Epstein and Zin (1991) is the classic empirical investigation of preferences that are non-separable across *states*. Ambitiously, for the time, they have some cross section of returns, 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. In essence, their empirical work, using the discount factor

$$m_{t+1} = \beta^{1+\gamma-\rho} (R_{t+1}^W)^{\frac{\rho-\gamma}{1-\rho}} \left(\frac{c_{t+1}}{c_t} \right)^{-\rho \left(\frac{1-\gamma}{1-\rho} \right)}. \quad (22)$$

amounted to showing that by using the stock market portfolio as a proxy for the utility index the consumption-based model could perform as well as the CAPM,

$$m_{t+1} = a - bR_{t+1}^W.$$

Alas, now we know the CAPM doesn’t perform that well on a more modern set of portfolios and instruments. How these preferences work in a consumption-based estimation with a more modern setup has yet to be investigated.

The Epstein-Zin framework has made a dramatic comeback along with the renewed interest in long-run phenomena. As discussed above, the Restoy and Weil (1998) expression of the discount factor ties the discount factor to news about future consumption as well as to current consumption.

$$(E_{t+1} - E_t) \log m_{t+1} \approx (E_{t+1} - E_t) \left[-\rho \Delta \log c_{t+1} + (\rho - \gamma) \sum_{j=0}^{\infty} \beta^j \Delta \log c_{t+j} \right] \quad (23)$$

Hansen, Heaton and Li (2005) point out that this expression gives another interpretation to Parker and Julliard (2005). The resulting moment condition is almost exactly the same as

that in (21); the only difference is the string of R_{t+j}^f in (21) and they are typically small and relatively constant. If the return at $t + 1$ predicts a string of small changes in consumption growth Δc_{t+j} , the finding underlying Parker and Julliard’s result, then the second term in this expression of the Epstein-Zin discount factor will pick it up.

Bansal and Yaron (2004) exploit (23) in a simulation economy context. Concentrate on the behavior of the market return, they hypothesize that consumption, rather than being a random walk, continues to grow after a shock. Together with an assumption of conditional heteroskedasticity, the second term in (23) can then act as an “extra factor” to generate a high equity premium, return volatility and the fact that returns are forecastable over time.

Bansal, Dittmar and Lundblad (2005) also argue that average returns of value vs. growth stocks can be understood by different covariances with long-run consumption growth in this framework. They examine long-run covariances of *earnings* with consumption, rather than those of *returns*. This is an interesting innovation; eventually finance must relate asset prices to the properties of cashflows rather than “explain” today’s price by the covariance of tomorrow’s price with a factor (β). Also, long-run returns must eventually converge to long-run dividend and earnings growth, since valuation ratios are stationary.

However, Hansen, Heaton and Li (2004b) show that Bansal, Dittmar, and Lundblad’s evidence that value stocks have much different long-run-consumption-betas than do growth stocks depends crucially on the inclusion of 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, leading to Bansal Dittmar and Lundblad’s finding of a strong beta to explain value premia. Without the time trend, all the betas are about one.

Piazzesi and Schneider (2006) have started to apply the framework to bonds. They generate risk premia in the term structure by the ability of state variables to forecast future consumption growth.

Questions

The central questions for the empirical importance of the Epstein-Zin framework are 1) Is the elasticity of intertemporal substitution really that different from the coefficient of risk aversion? and 2) Are there really important dynamics in consumption growth?

As discussed above, the evidence on the intertemporal substitution elasticity is not yet decisive, since there just isn’t that much time variation in real interest rates and expected consumption growth to correlate. On intuitive grounds, it’s not obvious why people would strongly resist substitution of consumption across *states of nature*, but happily accept substitution of consumption over *time*. Why would you willingly put off going out to dinner for a year in exchange for a free drink (high intertemporal elasticity), but refuse a bet of that dinner for one at the fanciest restaurant in town (high risk aversion)?

Dynamics are vital. If consumption growth is unpredictable, then Epstein-Zin utility is observationally equivalent to power utility, a point made by Kocherlakota (1990). This is

clear in (23) but it is true more generally. If there is no information about future consumption growth at $t + 1$ then U_{t+1} depends only on c_{t+1} ; there are no other state variables. Now, consumption growth is the *least* forecastable of all macroeconomic time series, for good reasons that go back to Hall’s (1978) random walk finding, especially if one takes out the effects of time aggregation, slightly durable goods, seasonal adjustment, and measurement error.

Parker and Julliard (2005) provide evidence on the central question: how much do current returns R_{t+1} forecast long-horizon future consumption growth $\sum_{j=1}^k \Delta c_{t+j}$? Alas, they include Δc_{t+1} , so we do not know from the table how important is the Epstein-Zin innovation, forecasts of $\sum_{j=2}^k \Delta c_{t+j}$, and they give unweighted truncated forecasts rather than an estimate of the weighted infinite horizon forecast $\sum_{j=2}^{\infty} \beta^j \Delta c_{t+j}$. Still, one can infer from their table the general result: the forecastability of future consumption growth by current returns is economically tiny, statistically questionable and certainly poorly measured. The returns hml_{t+1} and smb_{t+1} together generate a maximum forecast R^2 of 3.39% at a one year horizon. That R^2 is a good deal lower at longer horizons we are interested in, 1.23% at 3 years and 0.15% at nearly 4 years, and some of that predictability comes from the 1.78% R^2 from explaining Δc_{t+1} from returns at time $t + 1$.

Long-run properties of anything are hard to measure, as made clear in this context by the Hansen Heaton and Li (2004b) sensitivity analysis. Now, one may *imagine* interesting long run properties of consumption growth, and one may find that specifications within one standard error of the very boring point estimates have important asset pricing effects, which is essentially what Bansal and Yaron (2004) do. But without strong direct evidence for the required long run properties of consumption growth, the conclusions will always be a bit shaky. Without independent measurements, movements in long-run consumption growth forecasts (the second term in (23)) act like unobservable shifts in marginal utility, or shifts in “sentiment,” which are always suspicious explanations for anything. At a minimum, an explanation based difficult-to-observe shifts in long-run consumption growth should parsimoniously tie together many asset pricing phenomena.

A final doubt

An alternative strand of thought says we don’t need new utility functions at all in order to match the aggregate facts. If the conditional moments of consumption growth vary enough over time, then we can match the aggregate facts with a power utility model. Campbell and Cochrane (1999) starts with the premise that aggregate consumption is a pure random walk, so any dynamics must come from preferences. Kandel and Stambaugh (1990, 1991) construct models in which time-varying consumption moments do all the work. For example, from $E_t(R_{t+1}^e)/\sigma_t(R_{t+1}^e) \approx \gamma \sigma_t(\Delta c_{t+1})$, conditional heteroskedasticity in consumption growth can generate a time-varying Sharpe ratio. The empirical question is again whether consumption growth really is far enough from i.i.d. to generate the large variations in expected returns that we see. There isn’t much evidence *for* conditional heteroskedasticity in consumption growth, but with high risk aversion you might not need a lot, so one might be able to assume a consumption process less than one standard error from point estimates that generates all sorts of interesting asset pricing behavior.

The Epstein-Zin literature is to some extent going back to this framework. Bansal and Yaron (2004) for example, add conditional heteroskedasticity in consumption growth to generate time-varying risk premiums just as Kandel and Stambaugh do. The Epstein-Zin framework gives another tool – properties of long run consumption $E_t \sum \beta^j \Delta c_{t+j}$ to work with, but the philosophy is in many respects the same.

4.4 Consumption and factor models

A second tradition also has reemerged with some empirical success. Breeden, Gibbons and Litzenberger (1989), examine a linearized version of the consumption-based model, a form more familiar to financial economists. Breeden, Gibbons and Litzenberger just 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. More recent empirical research has raised the bar somewhat: industry portfolios show much less variation in mean returns than size and book-to-market portfolios that dominate cross-sectional empirical work. In addition, we typically use as instruments variables such as the dividend price ratio that forecast returns much better than lagged returns.

Lettau and Ludvigson (2001b) is the first modern reexamination of a consumption-based factor model, the first recent paper that finds some success in pricing the value premium from a macro-based model, and nicely illustrates current trends in how we evaluate models. Lettau and Ludvigson examine a *conditional* version of the linearized consumption-based model in this modern testing ground. 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,

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

The innovation is to allow the slope coefficient b , which acts as the risk-aversion coefficient in the model, to vary over time. They use the consumption-wealth ratio to measure z_t .

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

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

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,

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

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

Lettau and Ludvigson’s Figure 1, reproduced below, 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. This paper also set a style for many that followed: evaluate a macro-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. 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. (Cochrane (1996) also has graphs, but only uses size portfolios. Fama and French (1996) also encouraged this shift in attention by presenting average returns and pricing errors across portfolios, but in tabular rather than graphical format.)

Following Lettau and Ludvigson, 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 it is worth remembering the limitations of the technique.

Cross-sectional R^2 (average returns on predicted average returns) can be a dangerous statistic. 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. In addition, to the extent that the Fama-French three-factor model works, the information in the 25 portfolios is really all contained in the three factor portfolios, so there are really that much fewer degrees of freedom. Second, 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 (Roll and Ross 1994, Kandel and Stambaugh 1995). We can take linear combinations of the original portfolios to make the plots look as good or as bad as we want. Third, 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, and in particular for the popular time-series regression as used by Fama and French, various ways

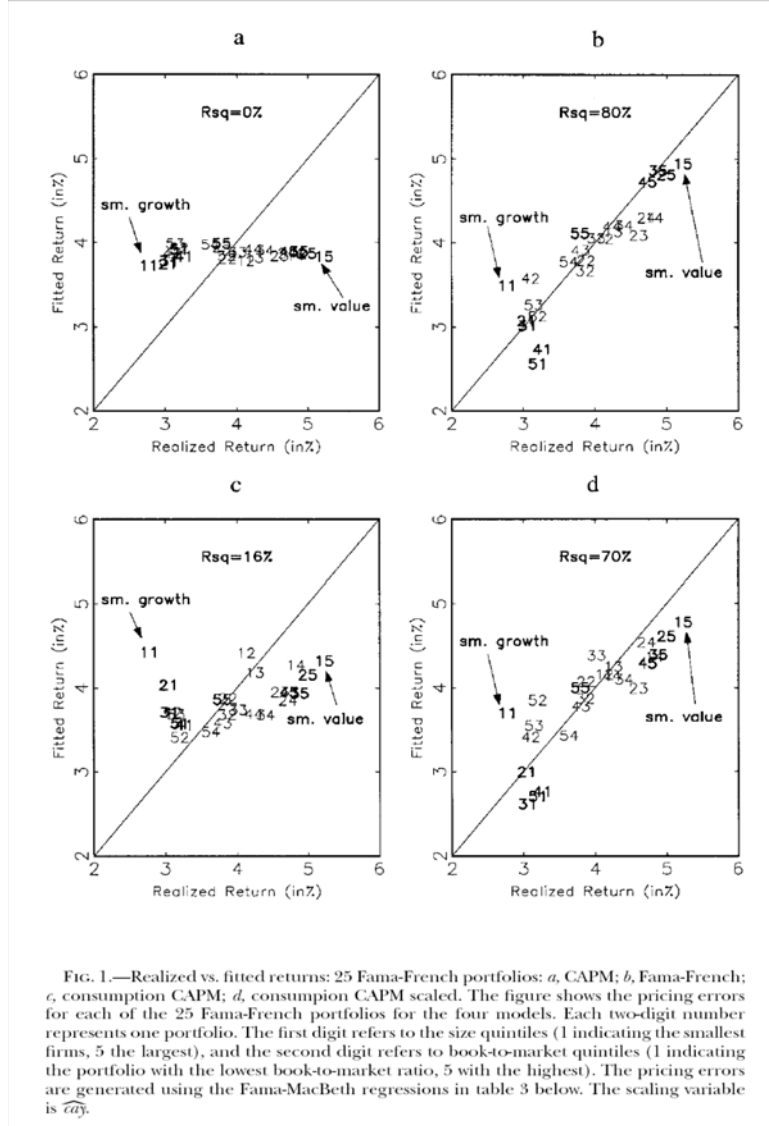


Figure 3: Lettau and Ludvigson Figure 1

of computing R^2 can give wildly different results.⁸

These criticisms are of course solved by statistical measures; test statistics based on $\alpha' cov(\alpha, \alpha')^{-1} \alpha$ where α is a vector of pricing errors are invariant to portfolio formation and take account of degrees of freedom. However, one can respond that the original portfolios are

⁸In a regression $y = a + xb + \varepsilon$, identities such as

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

only hold when b is the OLS estimate. Some of these calculations can give R^2 greater than one or less than zero when applied to other estimation techniques.

the interesting ones; the portfolios that modify R^2 a lot have unnatural and large long-short positions, and 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 carefully chosen diagnostics such as the R^2 .

Once the game goes past “do as well as the Fama-French three factor model in the Fama-French 25 portfolios” and moves on to “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? As above, I think it would be better for macro models to focus on pricing the three Fama-French factors rather than the highly cross-correlated 25 portfolios, which really add no more credible information.

Macro models also suffer from the fact that real factors are much less correlated with asset returns than are portfolio-based factors. 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 macro models 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. Presumably, correctly-done standard errors should reveal this problem.

Finally, these linearized macro models almost always leave as free parameters the betas, factor risk premia and (equivalently) the coefficients linking the discount factor to data, hiding the economic interpretation of these parameters. This observation also applies to current models on the investment side such as Cochrane (1996) and Li, Vassalou and Ying (2003) and to most ICAPM style work such as Vassalou (2003), who shows that variables which forecast GDP growth can price the Fama-French 25 portfolios. Let's not repeat the mistake of the CAPM that hid the implied 16% volatility of consumption growth or extraordinary risk aversion for so many years.

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 and its variants has barely been tried.

The playing field for empirical work has changed since the classic investigations of the consumption-based model and its extension to non-separable utility functions. We now routinely check any model in the size and book-market (and, increasingly, momentum) cross section rather than industry or beta portfolios, since the former show much more variation in average returns. When we use instruments, we use a few lags of powerful instruments known to forecast returns rather than many lags of returns or consumption growth, which are very weak instruments. We worry about time aggregation (or at least we should!) 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. We are interested when models capture some moments quite well, even admitting that they fail on others. We recognize that simulation models, in which artificial data display many patterns of real data are interesting, even though those models may miss other patterns in the data (such as predicting perfect correlations) that they are easily rejected by formal statistical tests.

This change is part of a larger, dramatic, and 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.)

Of course, we cannot expect authors of 20 years ago to do things as we would today. But it remains true that we are only beginning to know how the standard consumption-based model and its extensions to simple nonseparability across time, goods, and states behaves in this modern testing ground. There is still very much to do to understand where the consumption-based model works, where it doesn’t work, and how it might be improved.

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, and it is gratifying to know that there are some.

5 Production, Investment and General Equilibrium

If we want to link asset prices to macroeconomics, consumption seems like a weak 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; we have to think about which predictions of the model are robust to small costs of information, transaction or attention. For example, a one-month delay in adjusting consumption would destroy a test in monthly data, yet it would have trivial utility costs, or equivalently it could result from perfect optimization with trivially small transaction and information costs (Cochrane 1989).

5.1 “Production-based asset pricing”

These thoughts led me to want to link asset prices to *production* through *firm* first-order conditions in Cochrane (1991b). This approach should allow us to link stock returns to genuine business cycle variables, and firms may do a better job of optimization, i.e., small information and transactions cost frictions from which our models abstract may be less important for firms.

Time series tests

A production technology defines an “investment return,” the (stochastic) rate of return that results from investing a little more today and then investing a little less tomorrow. With a constant returns to scale production function, the investment return should *equal* the stock return, data point for data point. The major result is that investment returns – functions only of investment data – are highly correlated with stock returns.

The prediction is essentially a first-differenced version of the Q theory of investment. The stock return is pretty much the change in stock price or Q, and the investment return is pretty much the change in investment/capital ratio. Thus, the finding is essentially a first-differenced version of the Q theory prediction that investment should be high when stock prices are high. 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. Cochrane (1991b) confirms this prediction: investment to capital ratios predict stock returns.

There has been a good deal of additional work on the relation between investment and stock returns. Lamont (2000) cleverly uses a survey data set on investment *plans*. Investment plans data are great forecasters of actual investment. Investment plans also can avoid some of the timing issues that make investment expenditures data hard to use. If the stock price goes up today, it takes time to plan a new factory, draw the plans, design the machinery, issue stock, 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.

Zhang (2004) uses the Q theory to “explain” many cross-sectional asset pricing anomalies. Firms with high prices (low expected returns or cost of capital) will invest more, issue more stock, and go public; firms with low prices (high expected returns) will repurchase stock. We see the events, followed by low or high returns, which constitutes the “anomaly.”

Mertz and Yashiv (2005) extend the Q theory to include adjustment costs to labor as well as to capital. Hiring lots of employees takes time and effort, and gets in the way of production and investment. This fact means that gross labor flows and their interaction with investment should also enter into the Q-theory prediction for stock prices and stock returns. Mertz and Yashiv find that the extended model substantially improves the fit; the labor flow and in particular the interaction of labor and investment correlate well with aggregate stock market variations. The model matches slow movements in the *level* of stock prices, such as the events of the late 1990s, not just the *returns* or first differences on which my 1991 paper focused (precisely because it could not match the slow movements of the level). Mertz and Yashiv’s Figure 2 summarizes this central finding well.

Cross-sectional tests

Cochrane (1996) is an attempt to extend the “production-based” ideas to describe a cross-section of returns rather than a single (market) return. I use multiple production technologies, and I investigate the question whether the investment returns from these technologies span stock returns, i.e. whether a discount factor of the form

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

satisfies

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

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

I only considered size portfolios, not the now-standard size and book-to-market or other portfolio sorts. Li, Vassalou, and Xing (2003) find 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.

Really “production-based” asset pricing

These papers do not achieve the 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 output y_{t+1} in one state and reduce it in another state. By contrast, the usual utility function $E[u(c)] = \sum_s \pi(s)u[c(s)]$ defines marginal rates of substitution across all dates and states; $mrs_{s_1, s_2} = \{\pi(s_1)u'[c(s_1)]\} / \{\pi(s_2)u'[c(s_2)]\}$. 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.

Cochrane (1993) explains the issue and suggests three ways to put marginal rates of transformation into economic models. The dynamic spanning literature in asset pricing naturally suggests the first two approaches: 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 all [rain, shine] contingent claims. Jermann (2005) pursues the idea of spanning across two states of nature with two technologies, and constructs a simulation model that reproduces the equity premium based on output data.

Third, we can directly write technologies that allow marginal rates of transformation across states. Equivalently, we can allow the firm to choose the distribution of its technology shock process as it chooses capital and labor. If the firm's objective is

$$\max_{\{k_t, \varepsilon_{t+1} \in \Theta\}} E[m_{t+1} \varepsilon_{t+1} f(k_t)] - k_t = \sum_s \pi_s m_s \varepsilon_s f(k_t) - k_t$$

where m denotes contingent claim prices, then the first order conditions with respect to ε_s identify m_s in strict analogy to the consumption-based model. For example, we can use the standard CES aggregator,

$$\Theta : \left[E \left(\frac{\varepsilon_{t+1}}{\theta_{t+1}} \right)^\alpha \right]^{\frac{1}{\alpha}} = \left[\sum_s \pi_s \left(\frac{\varepsilon_s}{\theta_s} \right)^\alpha \right]^{\frac{1}{\alpha}} = 1 \quad (24)$$

where θ_{t+1} is an exogenously given shock. As an interpretation, nature hands the firm a production shock θ_{t+1} , but the firm can take actions to increase production in one state relative to another from this baseline. Then, the firm's first order conditions with respect to ε_s give

$$m_s f(k_t) = \lambda \frac{\varepsilon_s^{\alpha-1}}{\theta_s^\alpha}$$

or

$$m_{t+1} = \lambda \frac{y_{t+1}^{a-1}}{\theta_{t+1}^\alpha f(k_t)^\alpha}. \quad (25)$$

Naturally, the first order conditions say that the firm should arrange its technology shocks to produce more in high-contingent-claim-price states of nature, and produce less in states of nature for which its output is less valuable.

This extension of standard theory is not that strange. The technologies we write down, of the form $y_{t+1}(s) = \varepsilon(s)f(k_t)$ 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. Putting the choice of the shock distribution back into production theory, restoring its symmetry with utility theory, will give us marginal rates of transformation that we can compare to asset prices.

Belo (2005) takes a crucial step to making this approach work, by proposing a solution to

the problem of identifying θ_{t+1} in (25). He imposes a restriction that the sets Θ from which firms can choose their technology shocks are related. Belo shows that the resulting form of the production-based model for pricing excess returns is the same as a standard linear macro-factor model,

$$m_{t+1} = 1 + \sum_i b_i \Delta y_{i,t+1}$$

where y denotes output. The derivation produces the typical result in the data that the b_i have large magnitudes and opposing sign. Thus, the standard relative success of macro-factor models in explaining the Fama-French 25 can be claimed as a success for a truly “production-based” model as well.

5.2 General Equilibrium

Most efforts to connect stock returns to a fuller range of macroeconomic phenomena instead construct 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. The consumption model predictions are still there, but if we throw them out, perhaps citing measurement issues, we are left with interesting links between asset returns and business cycle variables.

While vast numbers of general equilibrium asset pricing models have been written down, I focus here on a few models that make quantitative connections between asset pricing phenomena and macroeconomics.

Market returns and macroeconomics

Urban Jermann’s (1998) “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 $y_t = \theta_t f(k_t)$; $k_{t+1} = (1 - \delta)k_t + (y_t - c_t)$. This feature 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_t is random giving random dividends, but all the stock price fluctuation that drives the vast majority of real-world return variation is absent.

Jermann therefore adds adjustment costs, as in the Q theory. Now there is a wedge between the price of “installed” (stock market) capital and “uninstalled” (consumption) capital. That wedge is larger when investment is larger. This 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 and habits 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. On the preference side, 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. Production technologies such as (24) may allow us to separately control the variability of marginal rates of transformation across states and 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.

Boldrin, Christiano and Fisher (2001) is a good example of more recent work in this area. Obviously, one task is to fit more facts with the model. Boldrin, Christiano and Fisher focus on quantity dynamics. Habit persistence and adjustment costs or other frictions to investment constitute a dramatic change relative to standard real business cycle models, and one would suspect that they would radically change the dynamics of output, consumption, investment and so forth. Boldrin, Christiano and Fisher’s major result is the opposite: the frictions they introduce actually improve on the standard model’s description of quantity dynamics, in particular the model’s ability to replicate hump-shaped dynamics rather than simple exponential decay.

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 autoregressive habit (18). They 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.

The next obvious step in this program is to unite the relative success of the Campbell-Cochrane (1999) habit specification with a fleshed-out production technology, in the style of Jermann (1998) or Boldrin, Christiano and Fisher (1999). Such a paper would present a full set of quantity dynamics as it matches the equity premium, a relatively stable risk-free rate, and time-varying expected returns and return predictability. As far as I know, nobody has put these elements together yet.

Does the divorce make sense?

Tallarini (2000) goes after a deep puzzle in this attempt to unite general equilibrium macroeconomics and asset pricing. 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 modeling innovations are pretty extreme compared to the improvement in quantity dynamics.

Tallarini explores 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 short-sighted after all. (At least leaving out welfare questions, in which case models with identical dynamics can make wildly different predictions.) Tallarini adapts Epstein-Zin preferences to a standard RBC model; utility is

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

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

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

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 able to account for the market price of risk or Sharpe ratio of the stock market (mean stock-bond return / standard deviation), but the quantity dynamics remain almost unchanged. In Tallarini's world, macroeconomists might well not have noticed the need for

large risk aversion.

There is a strong intuition for Tallarini's result. In the real business cycle model without adjustment costs, risk comes entirely from the technology shock, and there is nothing anyone can do about it, since as above, production sets are Leontief across states of nature. The production function allows relatively easy transformation over time, however, with a little bit of interest rate variation as $\partial f(K, N)/\partial K$ varies a small amount. Thus, if you raise the intertemporal substitution elasticity, you can get quite different business cycle dynamics as agents choose more or less smooth consumption paths. But if you raise the risk aversion coefficient without changing intertemporal substitution, saving, dissaving or working can do nothing to mitigate the now frightful technology shocks, so quantity dynamics are largely unaffected. The real business cycle model is essentially an endowment economy across states of nature.

With this intuition we can see that 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, and Lucas's model with power utility and low risk aversion is a disaster on asset pricing facts including the equity premium and return volatility. Tallarini's observational equivalence cuts both ways: *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 an explicit calculation along these lines.)

Second, the equity premium is Tallarini's only asset pricing fact. In particular, with no adjustment costs, he still has $Q=1$ at all times, so there is no stock price variation. Even when there is a high Sharpe ratio, both the mean stock return and its standard deviation are low. Papers that want to match more facts, including the mean and standard deviation of returns separately, price-dividend ratio variation, return predictability and cross-sectional value / growth effects, are driven to add habits and adjustment costs or the more complex ingredients. In these models, higher risk premia may well affect investment/consumption decisions and business cycle dynamics, 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.

Intangible capital

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 this view. 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 mul-

tiplied 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.

The stock market values profit streams, however, 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, and nor does the output of “intangible goods” that are accumulated to “intangible capital.” 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 the asset pricing implications of this view. (This is the tip of an iceberg of work in macroeconomics and accounting on the effects of potential intangible capital.) Hall 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. 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.

This is a provocative paper, throwing in to question much of the measurement underlying all of the macroeconomic models so far. It has its difficulties – it’s hard to account for the large stock market *declines* as loss of “organizational capital,” – but it bears thinking about.

The cross-section of returns

Obviously, the range of asset pricing phenomena addressed by this sort of model needs to be expanded, in particular to address cross-sectional results such as the value and growth effects.

Menzly, Santos and Veronesi (2004) approach the question through 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 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 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 generate value and growth effects in cross-sectional average returns from the 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 since more of the expected cashflows are pushed out to the future. This makes the cashflow more exposed to the aggregate risk premium, giving it a higher expected return and a lower price.

The obvious next step is to amplify its underpinnings to multiple production functions, allowing us understand the joint determination of asset prices with output, investment, labor, etc., moving from a “multiple-endowment” economy to “multiple production” economies just as the single representative firm literature did in moving from Mehra and Prescott’s endowment model to the production models discussed above. Berk, Green and Naik (1999),

Gomes, Kogan and Zhang (2003) derive size and book/ market effects in general equilibrium models with a bit more explicit, but also fairly stylized, technologies. For example, Gomes, Kogan and Zhang envision “projects” that arrive continuously; firms can decide to undertake a project by paying a cost, but then the scale of the project is fixed forever. Zhang (2005) uses a multiple-sector technology of the usual $y = \theta f(k)$ form with adjustment costs and both aggregate and idiosyncratic shocks, but specifies the discount factor process exogenously, rather than via a utility function and consumption that is driven by the output of the firms in his model. Gourio (2004) generates book/market effects in an economy with relatively standard adjustment-cost technology and finds some interesting confirmation in the data.

Gala (2006) is the latest addition to this line of research. This is a full general equilibrium model – the discount factor comes from consumption via a utility function – with a relatively standard production function. He includes adjustment costs and irreversibilities. The model produces value and growth effects. Fast-growing firms are investing and so on the positive, adjustment-cost side of the investment function. Value firms are shrinking and up against irreversibility constraints. Thus, when a shock comes, the growth firms can adjust production plans more than value firms can, so value firms are more affected by the shocks. Gala has one non-standard element; there is an “externality” in that investment is easier (lower adjustment costs) for small firms. This solves a technical aggregation problem, and also produces size effects that would be absent in a completely homogenous model.

Challenges for general equilibrium models of the cross-section

Bringing multiple firms in at all is the first challenge for a general equilibrium model that want to address the cross section of returns. Since the extra technologies represent nonzero net supply assets, each “firm” adds another state variable to the equilibrium. Some papers circumvent this problem by modeling the discount factor directly as a function of shocks rather than specify preferences and derive the discount factor from the equilibrium consumption process. Then each firm can be valued in isolation. This is a fine shortcut in order to learn about useful specifications of technology, but in the end of course we don’t really understand risk premia until they come from the equilibrium consumption process fed through a utility function. Other papers are able cleverly to prune the state space or find sufficient statistics for the entire distribution of firms in order to make the models tractable.

The second challenge is to produce “value” and “growth” firms that have low and high valuations. Furthermore, the low valuations of “value” firms must correspond to high expected returns, not entirely low cashflow prospects, and vice versa for growth. This challenge has largely been met too.

The third challenge is to reproduce the failures of the CAPM, as in the data. Again, the puzzle is not so much the *existence* of value and growth firms but the fact that these characteristics do not correspond to betas. A model in which some firms have high-beta cashflows and some firms have low-beta cashflows will generate a spread in expected returns, and prices will be lower for the high expected-return firms so we will see value and growth effects. But these effects will be explained by the betas. Few of the current models really achieve this step. Most models price assets by a conditional CAPM or a conditional

consumption-based model; the “value” firms do have higher conditional betas. Any failures of the CAPM in the models are due to omitting conditioning information or the fact that the stock market is imperfectly correlated with consumption. In most models, these features do not account quantitatively for the failures of the CAPM or consumption-based model in the data.

Fourth, a model must produce the comovement of value and growth firm returns that lies behind the Fama-French factors. Most models still have a single aggregate shock. And we haven’t started talking about momentum or other anomalies.

Finally, let us not forget the full range of aggregate asset pricing facts including equity premium, low and smooth risk free rate, return predictability, price-dividend ratio volatility and so forth, along with quantity dynamics that are at least as good as the standard real business cycle model.

I remain a bit worried about the accuracy of approximations in general equilibrium model solutions. Most papers solve their models by making a linear-quadratic approximation about a nonstochastic steady state. But the central fact of life that makes financial economics interesting is that risk premia are not at all second order. The equity premium of 8% is much larger than the interest rate of 1%. Thinking of risk as a “second - order” effect, expanding around a 1% interest rate in a perfect foresight model, seems very dangerous. 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, Hansen writes down a linear-quadratic (approximate) model, and then quickly finds an exact solution. This technique, emphasized in a large number of papers by Hansen and Sargent, might avoid many approximation and computation issues, especially as the state space expands with multiple firms. Hansen (1987) is also a very nice exposition of how general equilibrium asset-pricing economies work, and well worth reading on those grounds alone.

Clearly, there is much to do in the integration of asset pricing and macroeconomics. It’s tempting to throw up one’s hands and go back to factor fishing, or partial equilibrium economic models. They are however only steps on the way. We will not be able to say we understand the economics of asset prices until we have a complete model that generates artificial time series that look like those in the data.

What does it mean to say that we “explain” a high expected return $E_t(R_{t+1})$ “because” the return covaries strongly with consumption growth or the market return $cov_t(R_{t+1}\Delta c_{t+1})$ or $cov_t(R_{t+1}R_{t+1}^m)$? Isn’t the covariance of the return, formed from the covariance of *tomorrow’s* price with a state variable, every bit as much an endogenous variable as the expected return, formed from the level of *today’s* price? I think we got into this habit by historical accident. In a one-period model, the covariance is driven by the exogenous liquidating dividend, so it makes a bit more sense to treat the covariance as exogenous and today’s price or expected return as endogenous. If the world had constant expected returns, so that innovations in tomorrow’s price were simple reflections of tomorrow’s dividend news, it’s almost as excusable. But given that so much price variation is driven by expected return variation,

reading the standard one-period first order condition as a causal relation from covariance or betas to expected returns makes no sense at all.

General equilibrium models force us to avoid this sophistry. They force us to generate the covariance of returns with state variables endogenously along with all asset prices; they force us to tie asset prices, returns, expected returns, and covariances all back to the behavior of fundamental cash flows and consumption, and they even force us to trace those “fundamentals” back to truly exogenous shocks that propagate through technology and utility by optimal decisions. 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.

This feature provides great discipline to the general equilibrium modeler, and it makes reverse-engineering a desired result much harder, perhaps accounting for slow progress and technically demanding papers. As a simple example, think about raising the equity premium in the Mehra-Prescott economy. This seems simple enough; the first order condition is $E_t(R_{t+1}^e) \approx \gamma \text{cov}_t(R_{t+1}^e, \Delta c_{t+1})$, so just raise the risk aversion coefficient γ . If you try this, in a sensible calibration that mimics the slight positive autocorrelation of consumption growth in postwar data, you get a large *negative* equity premium. The problem is that the covariance is endogenous in this model; it does not sit still as you change assumptions. With positive serial correlation of consumption growth, good news about today’s consumption growth implies good news to future consumption growth. With a large risk aversion coefficient, good news about future consumption growth *lowers* the stock price, since the “discount rate” effect is larger than the “wealth” effect.⁹ In this way, the model endogenously generates a negative covariance term. To boost the equity premium, you have also to change assumptions on the consumption process (or the nature of preferences) to raise the risk aversion coefficient without destroying the covariance.

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 including the equity premium, predictable returns, and value, size, and similar effects in the cross-section of returns.

⁹The price of a consumption claim is

$$P_t = E_t \sum_{j=1}^{\infty} \beta^j \left(\frac{C_{t+j}}{C_t} \right)^{-\gamma} C_{t+j}$$

or, dividing by current consumption,

$$\frac{P_t}{C_t} = E_t \sum_{j=1}^{\infty} \beta^j \left(\frac{C_{t+j}}{C_t} \right)^{1-\gamma}$$

With $\gamma > 1$, a rise in C_{t+j}/C_t lowers P_t/C_t .

6 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 of “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 survey models that emphasize *overall* employment as a state variable (“labor income”) and then models that emphasize increases in individual *risk* from non-market sources (“idiosyncratic risk”).

6.1 Labor and outside 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 labor income information as it summarizes all other economically relevant risks. If someone loses their job and this is bad news, they should consume less as well, and consumption should therefore reveal all we need to know about the risk.

Labor hours can also enter, as above, if utility is non-separable between consumption and leisure. However, current work on labor income work does not stress this possibility, perhaps again because we don’t have much information about the cross-elasticity. Does more leisure make you hungrier, or does it substitute for other goods?

A better motivation for labor income risk, as for most traditional 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 or one big source of wealth that is not included in stock market indices. Many investors also have privately-held businesses, and the income from those businesses affects their asset demands exactly as does labor income, so we can think of the two issues simultaneously.

Measurement is still tricky. 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 you know you can quickly get a better one. Now, 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. For example, the same procedure applied to stocks says that today’s dividend tells us all we need to know about stock prices; that a beta on dividend growth would give the same answer as a beta on returns, that price-dividend ratios are exact functions of each period’s

dividend growth. 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.

Labor income growth in linear discount factor models

Jagannathan and Wang (1996) is so far the most celebrated recent model that includes a labor income variable. (See also the successful extension in Jagannathan, Kubota, and Takehara, 1998.) 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 2001b conditional CAPM.)

With VW and $prem$ alone, Jagannathan and Wang report only 30% cross-sectional R^2 (average 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 (2001b). Adding labor income, they obtain up to 55% cross-sectional¹⁰ R^2 .

Alas, the testing ground is not portfolios sorted by book to market ratio, but 100 portfolios sorted by beta and size. Jagannathan and Wang 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 the paper if labor income prices the 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 labor income factor $labor_t = [L_{t-1} + L_{t-2}] / [L_{t-2} - L_{t-3}]$ 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.

¹⁰Again, I pass on these numbers with some hesitation – unless the model is fit by an OLS cross-sectional regression, 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 regression.

Much of Jagannathan and Wang's empirical point can be seen in Table 1 of Lettau and Ludvigson (2001b), reproduced below. Δy is labor income growth, this time measured contemporaneously. Lettau and Ludvigson use the consumption to wealth ratio cay rather than the bond premium as the conditioning variable, which may account for the better results. Most importantly, 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.

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

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

TABLE 1 FAMA-MACBETH REGRESSIONS USING 25 FAMA-FRENCH PORTFOLIOS: λ_j COEFFICIENT ESTIMATES ON BETAS IN CROSS-SECTIONAL REGRESSION									
ROW	CONSTANT	\widehat{cay}_t	Factors _{t+1}				$\widehat{cay}_t \cdot \text{Factors}_{t+1}$		R^2 (\bar{R}^2)
			R_{vw}	Δy	SMB	HML	R_{vw}	Δy	
1	4.18		-.32						.01
	(4.47)		(-.27)						-.03
	(4.45)		(-.27)						
2	3.21		-1.41	1.26					.58
	(3.37)		(-1.20)	(3.42)					.54
	(1.87)		(-.67)	(1.90)					
3	1.87		1.33		.47	1.46			.80
	(1.31)		(.83)		(.94)	(3.24)			.77
	(1.21)		(.76)		(.86)	(2.98)			
4	3.70	-.52	-.06				1.14		.31
	(3.88)	(-.22)	(-.05)				(3.59)		.21
	(2.61)	(-.15)	(-.03)				(2.41)		
5	3.70		-.08				1.16		.31
	(3.86)		(-.07)				(3.58)		.25
	(2.60)		(-.44)				(2.41)		
6	5.18	-.44	-1.99	.56			.34	-.17	.77
	(5.59)	(-1.60)	(-1.73)	(2.12)			(1.67)	(-2.40)	.71
	(3.32)	(-.95)	(-1.02)	(1.26)			(.99)	(-1.42)	
7	3.81		-2.22	.59			.63	-.08	.75
	(4.02)		(-1.88)	(2.20)			(2.79)	(-2.52)	.70
	(2.80)		(-1.31)	(1.53)			(1.94)	(-1.75)	

NOTE.—The table presents λ estimates from cross-sectional Fama-MacBeth regressions using returns of 25 Fama-French portfolios: $E[R_{i,t+1}] = E[R_{0,t}] + \beta\lambda$. The individual λ_j estimates (from the second-pass cross-sectional regression) for the beta of the factor listed in the column heading are reported. In the first stage, the time-series betas β are computed in one multiple regression of the portfolio returns on the factors. The term R_{vw} is the return of the value-weighted CRSP index, Δy_{t+1} is labor income growth, and SMB and HML are the Fama-French mimicking portfolios related to size and book-market equity ratios. The scaling variable is \widehat{cay}_t . The table reports the Fama-MacBeth cross-sectional regression coefficient; in parentheses are two t -statistics for each coefficient estimate. The top statistic uses uncorrected Fama-MacBeth standard errors; the bottom statistic uses the Shanken (1992) correction. The term R^2 denotes the unadjusted cross-sectional R^2 statistic, and \bar{R}^2 adjusts for the degrees of freedom.

Figure 4: Lettau and Ludvigson Table 1

The take-away point, then, is that a large number of macroeconomic variables can be added to ad-hoc linear factor models ($m_{t+1} = a - bf_{t+1}$) to price the Fama-French 25 portfolios, including consumption, investment, and now labor income. Of course the usual caveat

applies that there are really only three independent assets in the Fama-French 25 portfolios (market, hml, smb), so one should be cautious about models with many factors.

Explicit modeling of labor income in a VAR framework.

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.

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, Brennan, Xia and Wang 2005 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, rather than being left unexamined as is the usual practice.

Alas, 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 the present value of labor income, as it does in stock prices. This model serves as a good warning to the vast majority of researchers who blithely use current labor income to proxy for the present value of future labor income. On the other hand, though, 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 do not seem that important.

Campbell and Vuolteenaho (2004) follow on the ICAPM component of Campbell (1996). They break the standard CAPM beta into two components, a "bad" cashflow beta that measures how much an asset return declines if expected future market cashflows decline, and "good" return beta that measures how much an asset return declines if a rise in future expected returns lowers prices today. The latter beta is "good" because in an ICAPM world (long-lived investors) it should have a lower risk premium. Ignoring the troubling small-growth portfolio, the improvement of the two-beta model over the CAPM on the Fama French 25 portfolios can be seen quickly in their Figure 3. Petkova (2006) also estimates an ICAPM-like model on the Fama-French 25 portfolios, finding that innovations to the dividend yield, term spread, default spread and level of the interest rate, all variables known to forecast

the market return, can account for the average returns of the Fama-French 25. Ultimately, ICAPM models should be part of macro-finance as well, since the “state variables” must forecast consumption as well as the market return in order to influence prices.

Proprietary income

Heaton and Lucas (2000) note that proprietary income – the income from non-marketed businesses – should be as, if not more, important to asset pricing than labor income as measured by Jagannathan and Wang. For rich people who own stocks, fluctuations in proprietary income are undoubtedly a larger concern than are fluctuations in wages. They find that individuals with more and more volatile proprietary income in fact hold less stocks. They also replicate Jagannathan and Wang’s investigation (using the same 100 industry/beta portfolios) using proprietary income. Using Jagannathan and Wang’s timing, they find that proprietary income is important, but more importantly the proprietary income series still works using “normal” timing rather than the one-period lag in Jagannathan and Wang.

Micro data

Malloy, Moskowitz, and Vissing-Jorgenson (2005) take another big step in the labor income direction. Among other refinements, they 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”

A model

Santos and Veronesi (2005) study a two-sector version of the model in Menzly, Santos and Veronesi (2004). 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 completely-solved model nicely shows the potential effects of labor income on asset pricing.

One part of Santos and Veronesi’s 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 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) conditional labor-income model that uses *cay* as a conditioning variable. Alas, adding shocks to the present value of labor income (measured here by changes in wages, with all the usual

warnings) as a *factor* 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.

6.2 Idiosyncratic risk, stockholding and microdata

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 aggregate prices or quantities. We say that a “tax cut” or “interest rate reduction” may increase “consumption” or “savings,” thereby affecting “employment” and “output,” but we ignore the possibility that the effect is different if it hits people differently. Of course the theory needed to justify perfectly this simplification is extreme, but seems a quite sensible first-approximation.

Macroeconomics and finance are thus full of investigations whether cross-sectional distributions matter. Two particular strains of this investigation are important for us. 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. Second, 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. Both considerations suggest examining our central asset pricing conditions using individual household data rather than aggregate consumption data.

Constantinides and Duffie and idiosyncratic risk

Basically, Constantinides and Duffie (1996) 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 they show you how to construct that process. This is a brilliant contribution as a decade of research into idiosyncratic risk had stumbled against one after another difficulty, and had great trouble demonstrate even the possibility of substantial effects.

Constantinides and Duffie’s Equation (11) gives the central result, which I reproduce with a slight change of notation,

$$E_t \left\{ \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \exp \left[\frac{\gamma(\gamma+1)}{2} y_{t+1}^2 \right] R_{t+1} \right\} = 1. \quad (26)$$

Here, y_{t+1}^2 is the *cross-sectional* variance of individual log consumption growth taken after aggregates at time $t+1$ are known. Equation (26) adds the exponential term to the standard consumption-based asset pricing equation. Since you can construct a discount factor (term before R_{t+1}) to represent any asset-pricing anomaly, you can construct a idiosyncratic risk

process y_{t+1}^2 to rationalize any asset-pricing anomaly. For example, DeSantis (2005) constructs a model in which the conditional variance of y_{t+1}^2 varies slowly over time, acting in many ways like the Campbell-Cochrane surplus consumption ratio (19) and generating the same facts in a simulation economy.

The nonlinearity of marginal utility is the key to the Constantinides-Duffie result. You might have thought that idiosyncratic risk cannot matter. Anything idiosyncratic must be orthogonal to aggregates, including the market return, so $E(m_{t+1} + \varepsilon_{t+1}^i, R_{t+1}) = E(m_{t+1}, R_{t+1})$. But the shocks should be to *consumption* or *income*, not to *marginal utility*, and marginal utility is a nonlinear function of consumption. Examining $E(m_{t+1}^i R_{t+1}) = E[E(m_{t+1}^i | R_{t+1}) R_{t+1}]$ we see that a nonlinear m will lead to a Jensen's inequality $1/2\sigma^2$ term, which is exactly the exponential term in (26). Thus, if the *cross-sectional* variance of idiosyncratic shocks is higher when returns R_{t+1} are higher, we will see a premium that does not make sense from aggregate consumption. The derivation of (26) follows exactly this logic and doesn't take much extra algebra¹¹.

Idiosyncratic consumption-growth risk y_{t+1} plays the part of consumption growth in the standard models. In order to generate risk premia, then, we need the distribution of idiosyncratic risk to vary over time; it must widen when high-average-return securities (stocks vs. bonds, value stocks vs. growth stocks) decline. It needs to widen *unexpectedly*,

¹¹Individual consumption is generated from $N(0, 1)$ idiosyncratic shocks $\eta_{i,t+1}$ by

$$\ln\left(\frac{C_{t+1}^i}{C_t^i}\right) = \ln\left(\frac{C_{t+1}}{C_t}\right) + \eta_{i,t+1}y_{t+1} - \frac{1}{2}y_{t+1}^2. \quad (27)$$

You can see by inspection that y_{t+1} is the cross-sectional variance of individual log consumption growth. Aggregate consumption really is the sum of individual consumption – the $-\frac{1}{2}y_{t+1}^2$ term is there exactly for this reason:

$$E\left(\frac{C_{t+1}^i}{C_t^i} \middle| \frac{C_{t+1}}{C_t}\right) = \frac{C_{t+1}}{C_t} E\left(e^{\eta_{i,t+1}y_{t+1} - \frac{1}{2}y_{t+1}^2}\right) = \frac{C_{t+1}}{C_t}.$$

Now, start with the individual's first order conditions,

$$\begin{aligned} 1 &= E_t \left[\beta \left(\frac{C_{t+1}^i}{C_t^i} \right)^{-\gamma} R_{t+1} \right] \\ &= E_t \left\{ \beta E \left[\left(\frac{C_{t+1}^i}{C_t^i} \right)^{-\gamma} \middle| \frac{C_{t+1}}{C_t} \right] R_{t+1} \right\} \\ &= E_t \left\{ \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} E \left[\left(\frac{C_{t+1}^i/C_{t+1}}{C_t^i/C_t} \right)^{-\gamma} \middle| \frac{C_{t+1}}{C_t} \right] R_{t+1} \right\} \\ &= E \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} e^{-\gamma(\eta_{i,t+1}y_{t+1} - \frac{1}{2}y_{t+1}^2)} R_{t+1} \right] \\ &= E \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} e^{\frac{1}{2}\gamma y_{t+1}^2 + \frac{1}{2}\gamma^2 y_{t+1}^2} R_{t+1} \right] \\ &= E \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} e^{\frac{\gamma(\gamma+1)}{2} y_{t+1}^2} R_{t+1} \right]. \end{aligned}$$

to generate a covariance with returns, and so as not to generate a lot of variation in interest rates. And, if we are to avoid high risk aversion, it needs to widen *a lot*.

As with the equity premium, the challenge for the idiosyncratic risk view is about quantities, not about signs. The usual Hansen-Jagannathan calculation

$$\frac{\sigma(m)}{E(m)} \geq \frac{E(R^e)}{\sigma(R^e)}$$

means that the discount factor m must vary by 50% or so. ($E(R^e) \approx 8\%$, $\sigma(R^e) \approx 16\%$, $R^f = 1/E(m) \approx 1.01$.) We can make some back of the envelope calculations with the approximation

$$\sigma \left\{ \exp \left[\frac{\gamma(\gamma+1)}{2} y_{t+1}^2 \right] \right\} \approx \frac{\gamma(\gamma+1)}{2} \sigma(y_{t+1}^2). \quad (28)$$

With $\gamma = 1$, then, we need $\sigma(y_{t+1}^2) = 0.5$. Now, if the *level* of the cross-sectional variance were 0.5, that would mean a cross-sectional standard deviation of $\sqrt{0.5} = 0.71$. This number seems much too large. Can it be true that if aggregate consumption growth is 2%, the typical person you meet either has +73% or -69% consumption growth? But the problem is worse than this, because 0.71 does not describe the *level* of idiosyncratic consumption growth, it must represent the *unexpected increase or decrease* in idiosyncratic risk in a typical year. Slow, business-cycle related variation in idiosyncratic risk y_{t+1}^2 will give rise to changes in interest rates, not a risk premium. Based on this sort of simple calculation, the reviews in Cochrane (1997) and Cochrane (2004) suggest that idiosyncratic risk model will have to rely on high risk aversion, just like the standard consumption model, to fit the standard asset-pricing facts.

Again, I am not criticizing the basic mechanism or the plausibility of the signs. My only point is that in order to get anything like plausible magnitudes, idiosyncratic risk models seem destined to need high risk aversion just like standard models.

The situation gets worse as we think about different time horizons. The required volatility of individual consumption growth, and the size of unexpected changes in that volatility $\sigma_t(y_{t+h}^2)$ must explode as the horizon shrinks. The Sharpe ratio $E_t(R^e)/\sigma_t(R^e)$ declines with the square root of horizon, so $\sigma_t(m_{t,t+h})$ must decline with the square root of horizon h . But y_{t+h}^2 governs the *variance* of individual consumption growth, not its *standard deviation*, and variances usually decline linearly with horizon. If $\sigma_t(y_{t+h}^2)$ declines only with the square root of horizon, then typical values of the *level* of y_{t+h}^2 must also decline only with the square root of horizon, since y_{t+h}^2 must remain positive. That fact means that the annualized variance of individual consumption growth must rise *unboundedly* as the observation interval shrinks. In sum, neither consumption nor the conditional variance of consumption growth y_t^2 can follow diffusion (random walk-like) processes. Both must instead follow a jump process in order to allow enormous variance at short horizons. (Of course, they may do so. We are used to using diffusions, but the sharp breaks in individual income and consumption on rare big events like being fired may well be better modeled by a jump process.)

In a sense, we knew that individual consumption would have to have extreme variance

at short horizons to get this mechanism to work. Grossman and Shiller (1982) showed that marginal utility is linear in continuous-time models when consumption and asset prices follow diffusions; it's as if utility were quadratic. The basic pricing equation is, in continuous time

$$E_t(dR_t) - r_t^f dt = \gamma E_t \left(dR_t \frac{dC_t^i}{C_t^i} \right) \quad (29)$$

where $dR_t = dP_t/P_t + D_t/P_t dt$ is the instantaneous total return. The average of dC^i/C^i across people must equal the aggregate, dC/C , so we have

$$E_t(dR_t) - r_t^f dt = \gamma E_t \left(dR_t \frac{dC_t}{C_t} \right).$$

Aggregation holds even with incomplete markets and nonlinear utility, and the Constantinides-Duffie effect has disappeared. It has disappeared into terms of order $dzdt$ and higher of course. To keep the Constantinides-Duffie effect, one must suppose that dC^i/C^i has variance larger than order dz , i.e. that it does not follow a diffusion¹².

Conversely, we may anticipate the same generic problem that many models have at long horizons. Like many models (See the Campbell-Cochrane discussion above), the Constantinides-Duffie model (26) adds a multiplicative term to the standard power-utility discount factor. To generate an equity premium at long-horizons, the extra term must also have a variance that grows linearly with time, as does the variance of consumption growth, and functions of stationary variables such as the cross-sectional variance idiosyncratic shocks usually does not growth with horizon, leaving us back to the power utility model at long horizons.

Empirical work

Of course, empirical arguments should be made with data, not on the backs of envelopes. Empirical work on whether variation in the cross-sectional distribution of income and consumption is important for asset pricing is just beginning.

Most investigations find some support for the basic effect – consumption and income *do* become more volatile across people in recessions and at times when the stock market declines. However, they confirm that the *magnitudes* are not large enough to explain the equity or value premia without high risk aversion. Heaton and Lucas (1996) calibrate an income process from the PSID and find it does not have the required volatility or correlation with stock market declines. Cogley (2002) 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

¹²There is another logical possibility. $E_t(dR_t) = r_t^f dt$ does not imply $E_t(R_{t+1}) = R_t^f$ if interest rates vary strongly over time, so one could construct a Constantinides-Duffie discrete time model with consumption that follows a diffusion, and hence no infinitesimal risk premium, but instead strong instantaneous interest rate variation. I don't think anyone would want to do so.

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. 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. 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. Of course, abundant measurement error in micro data will give the false appearance of mean-reversion, but if labor income were really very volatile and persistent, then the distribution of income would fan out quickly and counterfactually over time.

In contrast, Brav, Constantinides and Geczy (2002) report some asset-pricing success. They use household consumption data from the consumer expenditure survey and consider measurement error extensively. They examine one central implication, whether by aggregating marginal utility rather than aggregating consumption, they can explain the equity premium and (separately) the value premium, $0 = E(mR^e)$. Specifically, remember that the individual first order conditions still hold,

$$1 = E \left(\beta \frac{u'(C_{t+1}^i)}{u'(C_t^i)} R_{t+1} \right). \quad (30)$$

We therefore can always “aggregate” by averaging *marginal utilities*

$$1 = E \left(\left[\frac{1}{N} \sum_i \beta \frac{u'(C_{t+1}^i)}{u'(C_t^i)} \right] R_{t+1} \right). \quad (31)$$

We cannot in general aggregate by averaging *consumption*

$$1 \neq E \left(\beta \frac{u'(\frac{1}{N} \sum_i C_{t+1}^i)}{u'(\frac{1}{N} \sum_i C_t^i)} R_{t+1} \right). \quad (32)$$

Brav, Constantinides and Geczy contrast calculations of (31) with those of (32). This analysis also shows again how important nonlinearities in marginal utility are to generating an effect: If marginal utility were linear, as it is under quadratic utility or in continuous time, then of course averaging consumption *would* work, and would give the same answer as aggregating marginal utility.

This estimation is exactly identified; one moment $E(mR)$ and one parameter γ . Brav, Constantinides and Geczy find that aggregating *marginal utilities*, $E(mR) = 1$ they are able to find a γ between 2 and 5 that matches the equity premium, i.e. satisfies the single moment restriction. By contrast, using aggregate *consumption* data, the best fit requires very high risk aversion, and there is no risk aversion parameter γ that satisfies this single moment for the equity premium. (One equation and one unknown does not guarantee a solution.)

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 back-of the envelope calculations suggest that the required properties are extreme? How will this approach work as we extend the number of assets to be priced, and to be priced simultaneously?

Jacobs and Wang (2004) take a good step in this direction. They use the Fama French 25 size and book to market portfolios as well as some bond returns, and they look at the performance of a two-factor model that includes aggregate consumption plus the cross-sectional variance of consumption, constructed from consumer expenditure survey data. They find that the cross-sectional variance factor is important (i.e. should be included in the discount factor), and the two consumption factors improve on the (disastrous, in this data) CAPM. Not surprisingly, of course, the Fama-French ad-hoc factors are not driven out, and the overall pricing errors remain large however.

Microdata

Of course, *individuals* still price assets exactly as before. The equation (30) still 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 with (30) and forget about aggregation either of consumption (32) or of marginal utility (31). With micro data, we can also isolate stockholders or households more likely to own stocks (older, wealthier) and see if the model works better among these.

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 a large $-\gamma$ power can swamp the signal (See Brav, Constantinides and Geczy for an extended discussion.) In addition *individual* behavior may not be stationary over time, where *aggregates* are. For just this reason (betas vary over time), we use characteristic-sorted portfolios rather than individual stock data to test asset pricing models. It may make sense to aggregate the m in $1 = E(mR)$ just as we aggregate the R into portfolios. Also, typical data sets are short and do not include a long panel dimension; we do not track individual households over long periods of time. Finally, 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. 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, only the portion of consumption (really marginal utility) volatility correlated with the stock market counts. Purely idiosyncratic volatility (due to individual job loss, illness, divorce, etc.) does not count.

Despite these problems, there are some empirical successes in micro data. Mankiw and Zeldes (1991) find that stockholder's consumption is more volatile and more correlated with the stock market than that of nonstockholders, a conclusion reinforced by Attanasio, Banks and Tanner (2002). Ait-Sahalia, Parker and Yogo (2004) find that consumption of "luxury

goods,” presumably enjoyed by stockholders, fits the equity premium with less risk aversion than that of normal goods. Vissing-Jorgensen (2002) is a good recent example of the large literature that actually estimates the first order condition (30), though only for a single asset over time rather than for the spread between stocks and bonds. Thus, we are a long way from a full estimate that accounts for the market as well as the size and value premia (say, the Fama French 25) and other effects.

Must we use microdata? While initially appealing, however, it’s not clear that the stockholder/nonstockholder distinction is vital. 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 assume. More formally, Heaton and Lucas (1996) examine a carefully-calibrated portfolio model and find they need a very large transaction cost to generate the observed equity premium. Even non-stockholders are linked to the stock market in various ways. Most data on household asset holdings excludes defined-contribution pension plans, most of which contain stock market investments. Even employees with a defined-benefit plan should watch the stock market 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 sum, while there is nothing wrong with looking at stockholder data to see if their consumption really does line up better with stock returns, it is not so obvious that there is something terribly wrong with continuing to use aggregates, even though few households directly hold stock.

7 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, and its variation over time underlying return forecastability and volatility, the value and size premiums, the momentum premium, and the time-varying term premia in bond foreign exchange markets. More premia will certainly emerge through time.

On the empirical side, we are really only starting to understand how the simplest power utility models do and do not address these premiums, looking across data issues, horizons, time aggregation and so forth. The success of ad-hoc macro factor and “production” 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' report 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, all this with preference and technology specifications that are at least not wildly inconsistent with microeconomic investigation. The papers surveyed here, while path-breaking advances in that direction, do not come close to the full list of desiderata.

Having said "macroeconomics," "risk" and "asset prices," the reader will quickly spot a missing ingredient: money. In macroeconomics, monetary shocks and monetary frictions are considered by many to be an essential ingredient of business cycles. They should certainly matter at least for bond risk premia. (See Piazzesi 2005 for the state of the art on this question.) Coming from the other direction, there is now a lot of evidence for "liquidity" effects in bond and stock markets (see Cochrane 2005a for a review), and perhaps both sorts of frictions are related.

8 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.
- Attanasio, Orazio P., James Banks, and Sarah Tanner 2002, "Asset Holding and Consumption Volatility," *Journal of Political Economy* 110, 771-792.
- Ball, Ray, 1978, Anomalies in Relationships Between Securities' Yields and Yield-Surrogates," *Journal of Financial Economics* 6, 103-126.
- Bansal, Ravi, Robert F. Dittmar and Christian Lundblad, 2005, "Consumption, Dividends, and the Cross-Section of Equity Returns." *Journal of Finance* 60, 1639-1672.
- Bansal, Ravi, and Amir Yaron, 2004, "Risks For the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance* 59:4, 1481-1509.
- Banz, Rolf W. 1981, "The Relationship Between Return and Market Value of Common Stocks," *Journal of Financial Economics* 9, 3-18.
- Basu, Sanjoy, 1983, "The Relationship Between Earnings Yield, Market Value, and Return for NYSE Common Stocks: Further Evidence," *Journal of Financial Economics* 12, 129-156.
- Belo, Frederico, 2005, "A Pure Production-Based Asset Pricing Model," Manuscript, University of Chicago.
- Berk, Jonathan B, Richard C. Green and Vasant Naik, 1999, "Optimal Investment, Growth Options and Security Returns," *Journal of Finance* 54, 1153 - 1607.
- Boldrin, Michele, Lawrence J. Christiano and Jonas Fisher, 2001, "Habit Persistence, Asset Returns, and the. Business Cycle," *American Economic Review* 91, 149-166.
- Brainard, William C., William R. Nelson, and Matthew D. Shapiro, 1991, "The Consumption Beta Explains Expected Returns at Long Horizons," Manuscript, Economics Department, Yale University.
- 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.

- 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 2005, "Estimation and Test of a Simple Model of Intertemporal Asset Pricing," *Journal of Finance* 59, 1743-1776.
- Buraschi, Andrea and Alexei Jiltsov, 2005, "Inflation Risk Premia and the Expectations Hypothesis," *Journal of Financial Economics* 75, 429-90.
- Campbell, John Y., 1993, "Intertemporal Asset Pricing without Consumption Data," *American Economic Review* 83, 487-512.
- Campbell, John Y., 1995, "Some Lessons from the Yield Curve," *Journal of Economic Perspectives* 9, 129-152.
- Campbell, John Y., 1996, "Understanding Risk and Return," *Journal of Political Economy* 104, 298-345.
- 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.
- Campbell, John Y. and John H. Cochrane, 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, 1999, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy* 107, 205-251.
- 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.
- Campbell; John Y. and Robert J. Shiller, 1991, "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *The Review of Economic Studies* 58 (3), Special Issue: The Econometrics of Financial Markets, 495-514.
- Campbell, John Y., Robert J. Shiller and Kermit L. Schoenholtz, 1983, "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates," *Brookings Papers on Economic Activity* 1983, 173-223.

- Campbell, John Y. and Luis M. Viceira, 1999, "Consumption and Portfolio Decisions When Expected Returns are Time Varying" *Quarterly Journal of Economics* 114, 433-495.
- Campbell, John Y., and Tuomo Vuolteenaho, 2004, "Good Beta, Bad Beta," *American Economic Review* 94, 1249 – 1275.
- Carhart, Mark, 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance* 52, 57-82.
- 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.
- Chetty, Raj and Adam Szeidl, 2004, "Consumption Commitments: Neoclassical Foundations for Habit Formation," Manuscript, University of California at Berkeley.
- 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., 1991b, "Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations," *Journal of Finance* 46, 207-234.
- 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., 1996, "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy* 104, 572-621.
- 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)
- Cochrane, John H., 1999a, "New Facts in Finance," *Economic Perspectives* Federal Reserve Bank of Chicago 23 (3) 36-58.

- Cochrane, John H., 1999b, "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, Revised Edition.
- Cochrane, John H., 2005a, "Liquidity Trading and Asset Prices," *NBER Reporter*, National Bureau of Economic Research, www.nber.org/reporter.
- Cochrane, John H., 2005b, "Financial Markets and the Real Economy," *Foundations and Trends in Finance* 1, 1-101.
- Cochrane, John H., 2006a, "Financial Markets and the Real Economy" in John H. Cochrane, ed., *Financial Markets and the Real Economy* Volume 18 of the International Library of Critical Writings in Financial Economics, London: Edward Elgar, p. xi-lxix.
- Cochrane, John H., 2006b, "The Dog That Did Not Bark: A Defense of Return Predictability," NBER Working paper 12026.
- 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.
- 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," *American Economic Review* 95, 138-160.
- Cogley, Timothy, 2002, "Idiosyncratic Risk and the Equity Premium: Evidence From The Consumer Expenditure Survey," *Journal of Monetary Economics* 49, 309-334.
- Constantinides, George, 1990, "Habit Formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy* 98, 519-543.
- Constantinides, George, and Darrell Duffie, 1996, "Asset Pricing With Heterogeneous Consumers," *Journal of Political Economy* 104, 219-240.
- Cooper, Ilan and Richard Priestley, 2005, "Stock Return Predictability in a Production Economy," Manuscript, Norwegian School of Management.
- Craine, Roger, 1993, "Rational Bubbles: A Test," *Journal of Economic Dynamics and Control* 17, 829-46.
- Daniel, Kent, and David Marshall, 1997, "Equity-Premium and Risk-Free-Rate Puzzles at Long Horizons," *Macroeconomic Dynamics* 1, 452-84.
- Daniel, Kent and Sheridan Titman, 2005, "Testing Factor-Model Explanations of Market Anomalies," Manuscript, Northwestern University and University of Texas, Austin.

- De Bondt, Werner F. M. and Richard Thaler, 1985, "Does the Stock Market Overreact?" *Journal of Finance* 40, 793-805.
- De Santis, Massimiliano, 2005, "Interpreting Aggregate Stock Market Behavior: How Far Can the Standard Model Go?" Manuscript, University of California Davis.
- Eichenbaum, Martin, and Lars Peter Hansen, 1990, "Estimating Models With Intertemporal Substitution Using Aggregate Time Series Data," *Journal of Business and 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, 1991, "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., 1976, "Forward Rates as Predictors of Future Spot Rates," *Journal of Financial Economics* 3, 361-377.
- Fama, Eugene F., 1984a, "Forward and Spot Exchange Rates," *Journal of Monetary Economics* 14, 319-338.
- Fama, Eugene F. 1984b, "The Information in the Term Structure," *Journal of Financial Economics* 13, 509-28.
- Fama, Eugene F., 1990, "Stock Returns, Expected Returns, and Real Activity," *Journal of Finance* 45, 1089 - 1108.
- Fama, Eugene F., 1991, "Efficient Markets: II," Fiftieth Anniversary Invited Paper, *Journal of Finance* 46, 1575-1617.
- 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, 1989, "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics* 25, 23-49.
- 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, 1996, "Multifactor Explanations of Asset-Pricing Anomalies," *Journal of Finance* 51, 55-84.
- 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," *Journal of Finance* 54, 1325-1360.
- Gala, Vito D., 2006, "Investment and Returns," Manuscript, University of Chicago GSB.
- Goetzmann, William N., and Philippe Jorion, 1993, "Testing the Predictive Power of Dividend Yields," *Journal of Finance* 48, 663-679.
- Gomes, Francisco, and Alexander Michaelides, 2004, "Asset Pricing with Limited Risk Sharing and Heterogeneous Agents," Manuscript, London Business School.
- 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 Cyclicalities and the Cross-Section of Returns," Manuscript, University of Chicago.
- Goyal, Amit and Ivo Welch, 2003, "Predicting the Equity Premium with Dividend Ratios," *Management Science* 49, 639-654.
- Goyal, Amit and Ivo Welch, 2005, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," Manuscript, Brown University, Revision of NBER Working Paper 10483.

- 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 J. and Robert J. Shiller, 1982, "Consumption Correlatedness and Risk Measurement in Economies with Non-Traded Assets and Heterogeneous Information," *Journal of Financial Economics* 10, 195-210.
- Grossman, Sanford, Angelo Melino, and Robert J. Shiller, 1987, "Estimating the Continuous-Time Consumption-Based Asset-Pricing Model," *Journal of Business and 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.
- Hall, Robert E., 1988, "Intertemporal Substitution in Consumption," *Journal of Political Economy* 96, 339-57.
- Hall, Robert E., 2001, "The Stock Market and Capital Accumulation," *American Economic Review* 91, 1185 - 1202.
- Hamburger, Michael J. and EN. Platt, 1975, "The Expectations Hypothesis and the Efficiency of the Treasury Bill Market," *Review of Economics and Statistics* 57, 190-199.
- 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 Peter and Thomas J. Sargent, 1981, "Exact Linear Rational Expectations Models: Specification and Estimation," Federal Reserve Bank of Minneapolis Staff Report 71.
- 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 Kenneth J. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50, 1269-1288.
- 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 and Kenneth J. Singleton 1984, "Errata," *Econometrica* 52, 267-268
- 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 Ravi Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52, 557–590.
- 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* 61, 353–385.
- Heaton, John C., 1995, "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications," *Econometrica* 63, 681–717.
- Heaton John, and Deborah J. 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 and Deborah J. 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 and Deborah J. 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 J. Lucas, 2000, "Stock Prices and Fundamentals," *NBER Macroeconomics Annual 1999*, 213–42.
- Heaton, John and Deborah Lucas, 2000, "Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk," *Journal of Finance* 55, 1163–1198.
- Hodrick, Robert J., 1992, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies* 5, 357–86.
- Jacobs, Kris and Kevin Q. Wang, 2004, "Idiosyncratic Consumption Risk and the Cross-Section of Asset Returns," *Journal of Finance* 59, 2211–2252.
- 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, 1998, "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.

- Jegadeesh, Narasimhan, and Sheridan Titman, 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance* 48, 65-91.
- Jermann, Urban, 1998, "Asset Pricing in Production Economies," *Journal of Monetary Economics* 41, 257-275.
- Jermann, Urban, 2005, "The Equity Premium Implied by Production," Manuscript, University of Pennsylvania.
- 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, Narayana R., 1990, "Disentangling the Coefficient of Relative Risk Aversion from the Elasticity of Intertemporal Substitution: An Irrelevance Result" *Journal of Finance*, 45, 175-90.
- Kocherlakota, Narayana, 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," *Journal of Finance*, 49, 1541-1578.
- Lamont, Owen A., 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, 2003, "Inspecting the Mechanism: Closed-Form Solutions for Asset Prices in Real Business Cycle Models," *Economic Journal* 113, 550-75.

- 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, 376-380.
- Lettau, Martin, and Sydney Ludvigson, 2001a, "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance* 56, 815-49.
- Lettau, Martin, and Sydney Ludvigson, 2001b, "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, 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, 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.
- Lewellen, Jonathan, Stefan Nagel and Jay Shanken, 2006, "A Skeptical Appraisal of Asset Pricing Tests," Manuscript, Dartmouth College, Stanford University and Emory University.
- 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.
- 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 an 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.
- Lustig, Hanno, and Stijn Van Nieuwerburgh, 2004a, "Housing Collateral, Consumption Insurance and Risk Premia: An Empirical Perspective," *Journal of Finance* Forthcoming.
- Lustig, Hanno, and Stijn Van Nieuwerburgh 2004b, "A Theory of Housing Collateral, Consumption Insurance and Risk Premia," Manuscript UCLA and NYU.

- Macaulay, Frederick Robertson, 1938, *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields and Stock Prices in the United States Since 1856* Publications of the National Bureau of Economic Research no. 33. Reprinted in Risk Classics Library, Risk Books, 1999.
- Malloy, Christopher, Tobias Moskowitz, and Annette Vissing-Jorgenson, 2005, "Job Risk and Asset Returns," Manuscript, University of Chicago.
- 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.
- Menzly, Lior, 2001, "Influential Observations in Cross-Sectional Asset Pricing Tests," manuscript, University of Chicago.
- Menzly, Lior, Tano Santos and Pietro Veronesi, 2004, "Understanding Predictability," *Journal of Political Economy* 112, 1-47.
- Merton, Robert C., 1973, "An Intertemporal Capital Asset Pricing Model," *Econometrica* 41, 867-887.
- Merz, Monika, and Eran Yashiv, 2005, "Labor and the Market Value of the Firm," Manuscript, University of Bonn.
- Nelson, Charles R., and Myung J. Kim, 1993, "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance* 48, 641-661.
- 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, 185-222.
- Pastor, Lubos, and Pietro Veronesi, 2004, "Rational IPO Waves," *Journal of Finance*, forthcoming.
- Petkova, Ralitsa 2006, Do the Fama-French Factors Proxy for Innovations in Predictive Variables?, *Journal of Finance* 61, 581-612
- Piazzesi, Monika, 2005, "Bond yields and the Federal Reserve," *Journal of Political Economy* 113, 311-344.

- Piazzesi, Monika, Martin Schneider and Selale Tuzel, 2004, "Housing, Consumption, and Asset Pricing," Manuscript, University of Chicago, NYU and UCLA.
- Piazzesi, Monika, and Martin Schneider, 2006, "Equilibrium Yield Curves," Manuscript, University of Chicago and NYU prepared for the *2006 Macroeconomics Annual*.
- Poterba, James, and Lawrence H. Summers, 1988, "Mean Reversion in Stock Returns: Evidence and Implications," *Journal of Financial Economics* 22, 27-60.
- Roll, Richard, 1970, *The Behavior of Interest Rates*, Basic Books.
- 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, 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.
- Rozeff, Michael S., 1984, "Dividend Yields are Equity Risk Premiums," *Journal of Portfolio Management* 11, 68-75.
- Santos, Tano and Pietro Veronesi (2005) "Labor Income and Predictable Stock Returns," *Review of Financial Studies* forthcoming.
- Sargent, Thomas J. 1972, "Rational Expectations and the Term Structure of Interest Rates," *Journal of Money Credit and Banking* 4, 74-97.
- Sargent, Thomas J., 1978, "A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure," *Journal of Monetary Economics* 5, 133-143.
- Schwert, G. William, 2003, "Anomalies and Market Efficiency," Chapter 15 of George Constantinides, Milton Harris, and René Stulz, eds., *Handbook of the Economics of Finance*, North-Holland, pp. 937-972.
- Shiller, Robert J., 1979, "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," *Journal of Political Economy* 87, 1190-1219.
- Shiller, Robert J., 1981, "Do Stock Prices Move too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review* 71, 421-436.
- Shiller, Robert J., 1982, "Consumption, Asset Markets, and Economic Fluctuations," *Carnegie-Rochester Conference on Public Policy* 17, 203-238.
- Shiller, Robert J. Shiller, Robert J., 1984, "Stock Prices and Social Dynamics," *Brookings Papers on Economic Activity*, 1984, 457-510.
- Shiller, Robert J., John Y. Campbell and Kermit L. Schoenholz, 1983, "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates," *Brookings Papers on Economic Activity* 1983, 173-217.

- Stambaugh, Robert F., 1988, "The Information in Forward Rates: Implications for Models of the Term Structure," *Journal of Financial Economics* 21, 41-70.
- Stambaugh, Robert F., 1999, "Predictive Regressions," *Journal of Financial Economics* 54, 375-421.
- 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," 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.
- Tallarini, Thomas D. Jr., 2000, "Risk-Sensitive Real Business Cycles," *Journal of Monetary Economics* 45, 507-532.
- 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," *Review of Financial Studies* forthcoming.
- 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, 2006, "A Consumption-Based Explanation of Expected Stock Returns," *Journal of Finance*, forthcoming.
- Zhang, Lu, 2005, "The Value Premium," *Journal of Finance* 60, 67-104.
- Zhang, Lu, 2004, "Anomalies," Manuscript, University of Rochester.

9 Appendix

This appendix gives a self-contained derivation of the discount factor under Epstein-Zin (1991) preferences.

Utility index

The consumer contemplates the purchase of ξ shares at price p_t with payoff x_{t+1} . The maximum is achieved where $\frac{\partial}{\partial \xi} U_t(c_t - p_t \xi, c_{t+1} + x_{t+1} \xi) \Big|_{\xi=0} = 0$. From the utility function

$$U_t = \left((1 - \beta) c_t^{1-\rho} + \beta [E_t (U_{t+1}^{1-\gamma})]^{\frac{1-\rho}{1-\gamma}} \right)^{\frac{1}{1-\rho}} \quad (33)$$

we have

$$\frac{\partial U_t}{\partial c_t} = U_t^\rho (1 - \beta) c_t^{-\rho}. \quad (34)$$

Then, the first order condition is

$$\frac{\partial U_t}{\partial c_t} p_t = \frac{1}{1-\rho} U_t^\rho \beta \frac{1-\rho}{1-\gamma} [E_t (U_{t+1}^{1-\gamma})]^{\frac{\gamma-\rho}{1-\gamma}} \left[E_t \left((1-\gamma) U_{t+1}^{-\gamma} \frac{\partial U_{t+1}}{\partial c_{t+1}} x_{t+1} \right) \right]$$

Substituting from (34) and canceling,

$$c_t^{-\rho} p_t = \beta [E_t (U_{t+1}^{1-\gamma})]^{\frac{\gamma-\rho}{1-\gamma}} [E_t (U_{t+1}^{\rho-\gamma} c_{t+1}^{-\rho} x_{t+1})]$$

Thus, defining the discount factor from $p_t = E(m_{t+1} x_{t+1})$,

$$M_{t+1} = \beta \left(\frac{U_{t+1}}{[E_t (U_{t+1}^{1-\gamma})]^{\frac{1}{1-\gamma}}} \right)^{\rho-\gamma} \left(\frac{c_{t+1}}{c_t} \right)^{-\rho} \quad (35)$$

Market return

The utility function 33 is linearly homogeneous. Thus,

$$\begin{aligned} U_t &= \sum_{j=0}^{\infty} \frac{\partial U_t}{\partial c_{t+j}} c_{t+j} = E_t \sum_{j=0}^{\infty} \frac{\partial U_t}{\partial c_{t+j}} c_{t+j} \\ \frac{U_t}{\partial U_t / \partial c_t} &= E_t \sum_{j=0}^{\infty} M_{t,t+j} c_{t+j} = W_t \end{aligned} \quad (36)$$

The final equality is the definition of total wealth – the value of the claim to consumption (including time t consumption). This is the heart of the idea – wealth reveals the utility index in (35).

We want an expression with the market *return*, not wealth itself, so we proceed as follows. Use the utility function (33) to express the denominator of (35) in terms of time-t observables:

$$\frac{1}{\beta} [U_t^{1-\rho} - (1-\beta)c_t^{1-\rho}]^{\frac{1}{1-\rho}} = [E_t (U_{t+1}^{1-\gamma})]^{\frac{1}{1-\gamma}} \quad (37)$$

Now, substitute for U_t and U_{t+1} from (36), with (34),

$$W_t = \frac{U_t}{\partial U_t / \partial c_t} = \frac{U_t}{U_t^\rho (1-\beta)c_t^{-\rho}} = \frac{1}{1-\beta} U_t^{1-\rho} c_t^\rho$$

(Note with $\rho = 1$ the wealth-consumption ratio is constant $\frac{W_t}{c_t} = \frac{1}{1-\beta} \left(\frac{U_t}{c_t}\right)^{1-\rho}$.) Solving for U_t

$$U_t = [W_t(1-\beta)c_t^{-\rho}]^{\frac{1}{1-\rho}} \quad (38)$$

Now, use (37) and (38) in (35)

$$M_{t+1} = \beta \left(\frac{U_{t+1}}{[E_t (U_{t+1}^{1-\gamma})]^{\frac{1}{1-\gamma}}} \right)^{\rho-\gamma} \left(\frac{c_{t+1}}{c_t} \right)^{-\rho}$$

Substituting into (35),

$$\begin{aligned} M_{t+1} &= \beta \left(\frac{[W_{t+1}(1-\beta)c_{t+1}^{-\rho}]^{\frac{1}{1-\rho}}}{\frac{1}{\beta} [W_t(1-\beta)c_t^{-\rho} - (1-\beta)c_t^{1-\rho}]^{\frac{1}{1-\rho}}} \right)^{\rho-\gamma} \left(\frac{c_{t+1}}{c_t} \right)^{-\rho} \\ &= \beta^{1+\gamma-\rho} \left(\frac{W_{t+1}c_{t+1}^{-\rho}}{W_t c_t^{-\rho} - c_t^{1-\rho}} \right)^{\frac{\rho-\gamma}{1-\rho}} \left(\frac{c_{t+1}}{c_t} \right)^{-\rho} \\ &= \beta^{1+\gamma-\rho} \left(\frac{W_{t+1}}{W_t - c_t} \right)^{\frac{\rho-\gamma}{1-\rho}} \left(\frac{c_{t+1}}{c_t} \right)^{-\rho(\frac{\rho-\gamma}{1-\rho}+1)} \\ &= \beta^{1+\gamma-\rho} \left(\frac{W_{t+1}}{W_t - c_t} \right)^{\frac{\rho-\gamma}{1-\rho}} \left(\frac{c_{t+1}}{c_t} \right)^{-\rho(\frac{1-\gamma}{1-\rho})} \end{aligned}$$

Since this definition of wealth includes current consumption (dividend), the return on the wealth portfolio is

$$R_{t+1}^W = \frac{W_{t+1}}{W_t - c_t}$$

so we have in the end

$$M_{t+1} = \beta^{1+\gamma-\rho} (R_{t+1}^W)^{\frac{\rho-\gamma}{1-\rho}} \left(\frac{c_{t+1}}{c_t} \right)^{-\rho(\frac{1-\gamma}{1-\rho})}$$