CHAPTER 7

Financial Markets and the Real Economy

John H. Cochrane*
University of Chicago

*This is a substantially reworked version of two papers that appeard under the same title, Cochrane (2005b, 2006a). I gratefully acknowledge research support from the NSF in a grant administered by the NBER and from the CRSP. I thank Ron Balvers, Frederico Belo, John Campbell, George Constantinides, Hugo Garduno, François Gourio, Robert Ditmar, Lars Hansen, John Heaton, Hanno Lustig, Rajnish Mehra, Marcus Opp, Dino Palazzo, Monika Piazzesi, Nick Roussanov, Alsdair Scott, Luis Viceira, Mike Wickens, and Motohiro Yogo for comments.

Copyright © 2008 by Elsevier BV. All rights of reproduction in any form reserved.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 238 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 cover the time-series and cross-sectional facts, the equity premium, consumption-based models, general equilibrium models, and labor income/idiosyncratic risk approaches.

JEL Classification: G12, E44

Keywords: asset pricing, cross-sectional tests, empirical work, intangible capital, long horizons, macroeconomics, microdata, proprietary income, risk premia, time series tests, risk aversion, state variables, bad times

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 239 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

1. INTRODUCTION

1.1. Risk Premia

Some assets offer higher average returns than other assets, or, equivalently, they attract lower prices. These "risk premia" 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,

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

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

$$p_t^i = \frac{E_t\left(x_{t+1}^i\right)}{R_t^f} + \mathrm{Cov}_t\left(m_{t+1}, x_{t+1}^i\right). \tag{2}$$

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 becomes[1]

$$E_t\left(R_{t+1}^{ei}\right) = -\mathrm{Cov}_t\left(R_{t+1}^{ei}, m_{t+1}\right). \tag{3}$$

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

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

$$1 = E(mR),$$

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

$$0 = E\left(mR^e\right).$$

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

$$0 = E(m)E\left(R^e\right) + \mathrm{Cov}\left(m, R^e\right),$$
$$E\left(R^e\right) = -R^f\mathrm{Cov}\left(m, R^e\right).$$

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

$$E\left(R^e\right) = -\mathrm{Cov}\left(m, R^e\right).$$

This equation holds exactly in continuous time.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 240 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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

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

This is a simple statement of an investor's first-order conditions. The marginal value of wealth[2] 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.[3] 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 Eqs. (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,

[2]Formally, the value of wealth is the achieved level of utility given the investor has wealth W,

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

subject to an appropriate budget constraint that is limited by initial wealth W_t. It can be a function $V(W_t, z_t)$ of other "state variables" z_t, for example, the expected returns of assets or the amount of outside income the investor expects to recieve, since higher values of these variables allow the investor to generate more utility.

[3]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 $V_W(t+1)$.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 241 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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,

$$E_t\left(R^{ei}_{t+1}\right) = \mathrm{Cov}_t\left(R^{ei}_{t+1}, R^{m}_{t+1}\right) \times \gamma,$$

where R^e denote excess returns, γ is a constant of proportionality equal to the average investor's risk aversion, and R^m is the market portfolio.[4] Multifactor models such as the popular Fama–French (1996) three-factor model use returns on multiple portfolios to measure the marginal value of wealth.

Research connecting financial markets to the real economy—the subject of this 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,

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

which states that assets must offer high returns if they pay off badly in "bad times" as measured by consumption growth.[5] 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

[4]To derive this expression of the CAPM, assume the investor lives one period and has quadratic utility $u(c_{t+1}) = -\frac{1}{2}\left(c^* - c_{t+1}\right)^2$. The investor's problem is

$$\max E\left(-\frac{1}{2}\left(c^* - c_{t+1}\right)^2\right) \quad \text{subject to} \quad c_{t+1} = R^p_{t+1}W_t = \left(R^f + \sum_{j=1}^{N} w_j R^{ej}_{t+1}\right)W_t,$$

where R^e denotes excess returns and R^f is the risk-free rate. Taking the derivative with respect to w_j, we obtain $0 = E\left[\left(c^* - R^p_{t+1}W_t\right)R^{ej}_{t+1}\right]$. Using the definition of covariance,

$$E\left(R^{ej}_{t+1}\right) = -\frac{\mathrm{Cov}\left[\left(c^* - R^p_{t+1}W_t\right), R^{ej}_{t+1}\right]}{E\left(c^* - R^p_{t+1}W_t\right)} = \mathrm{Cov}\left(R^p_{t+1}, R^{ej}_{t+1}\right)\frac{W_t}{\left(c^* - E\left(R^p_{t+1}\right)W_t\right)}.$$

The risk aversion coefficient is $\gamma = -cu''(c)/u'(c) = c/(c^* - c)$. Thus, we can express the term multiplying the covariance as the local risk aversion coefficient γ, at a value of consumption $\hat{c}$ given by $1/c = (1/W_t) - (E(R^p_{t+1}) - 1/b)$. If consumers are enough alike, then the indvidual portfolio is the market portfolio, $R^p = R^m$.

[5]One may derive this expression quickly by a Taylor expansion of the investor's first-order conditions, and using $R^f = 1/E(m) \approx 1$ for short horizons,

$$0 = E\left(mR^{ei}\right) = E\left(\beta\frac{u'(c_{t+1})}{u'(c_t)}R^{ei}_{t+1}\right),$$

$$E\left(R^{ei}_{t+1}\right) = -R^f\mathrm{Cov}\left(R^{ei}_{t+1}, \frac{u'(c_{t+1})}{u'(c_t)}\right)$$

$$\approx \mathrm{Cov}\left(R^{ei}_{t+1}, \frac{-c_t u''(c_t)}{u'(c_t)}\left(\frac{c_{t+1} - c_t}{c_t}\right)\right) = \mathrm{Cov}\left(R^{ei}_{t+1}, \frac{c_{t+1}}{c_t}\right) \times \gamma.$$

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 242 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

in measures of economic conditions such as aggregate consumption, that explain the pattern by which mean returns $E_t(R^{ei}_{t+1})$ vary across assets i and over time t.

1.2. 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?

1.2.1. 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.

1.2.2. 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 243 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 of 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.

1.3. 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 well.[6] And this "mimicking portfolio" will have better-measured and more frequent data, so it will work better in sample and in practice. It will be the right model to recommend for many applications.

[6]Start 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 constant,

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

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

$$1 = E[(b'R)R].$$

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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 244 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 toward 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)?

2.1. 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 245 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

TABLE 1
OLS Regressions of Excess Returns (value-weighted NYSE—Treasury bill) and Real Dividend Growth on the Value-Weighted NYSE Dividend-Price Ratio

Horizon k (years)	$R^e_{t\to t+k} = a + b\frac{D_t}{P_t} + \varepsilon_{t+k}$			$\frac{D_{t+k}}{D_t} = a + b\frac{D_t}{P_t} + \varepsilon_{t+k}$		
	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	0.0010
3	12.6	3.0	0.20	0.16	0.13	0.0001
5	20.6	2.6	0.22	2.42	1.11	0.0200

Sample 1927–2005, annual data. $R^e_{t\to t+k}$ 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.

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.

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.

2.2. 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 246 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 $x_{i,t}$ that give significant coefficients in

$$R^i_{t+1} - R^f_t = a + bx_{i,t} + \varepsilon_{i,t+1}.$$

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* by 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.)

2.3. 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 are 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 247 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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,[7] 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 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 have been extended and stand up well over time. Stambaugh (1988) extended the results for short-term bonds and Cochrane and Piazzesi (2005) 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 Piazzesi 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

[7]Evidence against the expectations hypothesis of bond yields goes back at least to Macaulay (1938). Shiller, Campbell, and Schoenholtz cite Hansen and Sargent (1981), Roll (1970), Sargent (1978, 1972), and Shiller (1979). Fama says, "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 Shoenholtz.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 248 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 percent. (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).

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, univariate forecastability or mean reversion are small (see, for example, Cochrane (2004), pp. 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,

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

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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 249 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

It would be lovely if variation in price-dividend ratios corresponded to dividend-growth forecasts. Investors, knowing future dividends will be higher than they are today, bid up stock prices relative to current dividends; then today's 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 cash flows.

2.3.1. Macro Variables and Forecastability

The forecasting variables in return regressions are so far all based on market prices, 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), document that the price variables that forecast returns also forecast economic activity.

One can, of course, run regressions of returns on macroeconomic variables. A number of 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)), 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 to capital ratio and consumption to 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 250 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

easily by watching the consumption to 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 to 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 to wealth ratio forecasts returns, even in the presence of D/P. In the context of (4), the consumption to wealth ratio sends dividend growth and returns in the same direction, so its effects on the price to dividend ratio are 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, with offsetting effects on prices. So we end up with a new variable, and an opening for additional variables, that forecast *both* returns and cash flows, giving stronger links from macroeconomics to finance.

2.3.2. Statistics

Return forecastability has come with a long statistical controversy. The first round of statistical investigation asked whether the 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 first answer was 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 $E(r_{t+1}x_t) = 0$—noone 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 regressions are implied by one-year regressions. If

$$x_{t+1} = \rho x_t + \upsilon_{t+1},$$
$$r_{t+1} = bx_t + \varepsilon_{t+1},$$

then

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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 251 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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.

More seriously, even the one-period regressions are suspect. 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 toward 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-forecasting 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 cash flow) 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 are constant. Thus, the *lack* of dividend forecastability is important evidence *for* return forecastability, and this is ignored in the statistical studies. In Cochrane (2006b), I formalize this argument. I show 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.) I also show that long horizon return forecasts *can* add important statistical evidence for return forecastability and that long horizon return forecasts are closely related to dividend growth forecasts.

2.4. 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.[8] See the average returns in their Table 1 panel A, reproduced in Figure 1.

Again, this pattern is not by itself a puzzle. High expected returns *should* be revealed by low market values (see Eq. (4)). The puzzle is that the value and small firms do *not*

[8]These 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).

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 252 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

TABLE 1
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

	Book-to-market equity (BE/ME) quintiles									
Size	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.

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 (see Figure 2), 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 that 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 subtract the market return, we would not expect to see a 95 percent R^2 in a regression of the A portfolio on an A–L minus M–Z "factor," because we would expect no common movement among 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,

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 253 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Table I–*Continued*

	Book-to-market equity (BE/ME) quintiles									
Size	Low	2	3	4	High	Low	2	3	4	High
Panel B: Regressions: $R_i - R_f = a_i + b_i\left(R_M - R_f\right) + s_i\text{smb} + h_i\text{hml} + e_t$										
	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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 254 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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, a fact that turns conventional investment advice on its head. They do not have higher market betas, but they do have higher betas on the Fama–French factors. In this sense, the Fama–French three-factor model "explains" this additional pattern in expected returns. In this kind of application, the Fama–French three-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 evaluating 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 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 of both returns (high R^2) and expected returns. One could state the result that three dominant eigenvalues in the covariance matrix of the 25 portfolios 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 nonzero expected returns. Given these factors, the expected returns of the 25 portfolios (and any other, different, portfolios that are explained by the three-factor model) follow automatically. The point of the 25 portfolios is to show "non-parametrically" 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).

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 255 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 that 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 ahead 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 astructural 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 moved much more closely with the aggregate economy, so more recent estimates do show a

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 256 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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* hml 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 they 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 percent 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 percent 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 257 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 or not 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* fall into 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 258 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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) becomes[9]

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

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

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

This model is a natural first place to link asset returns to macroeconomics. It has a great economic and intuitive appeal. Assets should give a high premium if they pay off badly in "bad times." What better measure of "bad times" than consumption? People may complain, or seem to be in bad straits, but if they're going out to fancy dinners you can tell that times aren't so bad after all. More formally, consumption subsumes or reveals all we need to know about wealth, income prospects, etc. in a wide class of models starting with 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

[9]In discrete time, the actual equation is

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

with

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

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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 259 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

mean market return as exogenously given. We are asking what are the economics behind the mean market return.) From (6) write

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

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

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

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 U.S. data, the mean return of stocks over bonds is about 8 percent, with a standard deviation of about 16 percent, 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 non-durable and services consumption volatility is much smaller, about 1.5 percent per year in the postwar U.S. 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 $1 + r^f = 1/E(e^{-\delta}(u'(c_{t+1})/u'(c_t)))$, is

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

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

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

Thus, with $\delta = 0$, the model predicts a 20 percent interest rate. To generate a (say) 5 percent interest rate, we need a *negative* 15 percent discount rate δ. Worse, (9) with $\gamma = 33$ predicts that the interest rate will be extraordinarily sensitive to changes in

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 260 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 non-durable 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 percent 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 percent mean and a 16 percent standard deviation, this calculation predicts 100 percent 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—miscited, 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 percent 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 percent in a typical 1σ year and consumption moves 2 percent in the same 1σ year, either consumption must eventually rise 14 percent or wealth must eventually decline 14 percent, as the consumption to wealth ratio is stable in the long run. This is a powerful motivation for Lettau and Ludvigson's use of consumption and 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 261 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

3.1. 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, Eq. (8), is similar to a relation derived by 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) points 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) named and announced the "puzzle" and launched the literature devoted to "explaining" it. 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 their model).

The history of the equity premium puzzle is an interesting case study for how ideas form, catch on, and evolve in economics and finance. The pattern does not fit well into the familiar stylized models of intellectual evolution such as Kuhn (1962) or McCloskey (1983).

Like many famous papers, this one has precursors. Shiller (1982) derived the first bound on discount-factor volatility. On p. 221, Shiller writes,

> It is also possible to arrive at a lower bound on the standard deviation of the marginal rate of substitution ... by using data on asset returns alone.... One finds that
>
> $$\sigma(S) \geq \frac{E(R^j) - E(R^i)}{\sigma(R^i)E(R^j) - \sigma(R^j)E(R^i)}$$
>
> [Shiller uses S for what I have denoted m]. This inequality puts a lower bound on the standard deviation of S in terms only of the means and standard deviations [of returns].... This inequality asserts that if two assets have very different average returns and their standard deviations are not sufficiently large, then $\sigma(S)$ must be large if the covariance [of returns] with S is to explain the difference in average returns. If one uses the Standard & Poor's portfolio as the jth asset, prime 4–6 month commercial paper as the ith asset and the sample means and sample standard deviations of after-tax real one-year returns for 1891 to 1980 in the right-hand side of the above inequality, then the lower bound on $\sigma(S)$ is 0.20. ... The large standard deviation for S arises because of the large difference between the after-tax average real return on stocks (... 5.7 percent per year for 1891 to 1980), and

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 262 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

> [the] average after-tax real return on commercial paper (... 1.4 percent per year for 1891 to 1980), while the standard deviations of the real after-tax returns are not sufficiently high (0.154 for stocks and 0.059 for commercial paper) to account for the average return spread unless $\sigma(S)$ is very high. A high $\sigma(S)$ suggests a high coefficient of relative risk aversion A (γ) since $\sigma(S) \approx A\sigma(\Delta C/C)$. For 1891 to 1980 $\sigma(\Delta C/C)$ was 0.035 so a lower bound for $\sigma(S)$ of 0.20 suggests A be over five....
>
> ... the conventional notion that stocks have a much higher return than does short term debt, coupled with the notion that pretax stock real returns have a standard deviation in the vicinity of 20 percent per year (commercial paper much less) implies that the standard deviation of S is very high.

There it is, in a nutshell. Interestingly, we have come full circle, as my summary boils the calculation down to much the same sort of inequality Shiller started with. This work appeared in the context of a number of studies in the early 1980s that found very high risk aversion popping up in estimates of consumption-based first-order conditions, and Grossman and Shiller (1981) and Hansen and Singelton (1983) in particular, but the latter do not have as clear a statement of the puzzle.

It's interesting that Mehra and Prescott's more complex approach was so much more influential. (A quick count in the Social Sciences Citation index gives 679 citations to Mehra and Prescott (1985), and only 35 to Shiller (1982).) Mostly, it seems to me that Mehra and Prescott were the first to argue and to persuade others that this puzzle, among so many in fitting the consumption-based model to data, is particularly *important* and that solving it would lead to some fundamental revision of the economics in the consumption-based model. This really is their distinctive, and central, contribution. Columbus "discovered" America, though Leif Ericson and a thousand Basque fishermen had been there before.

Shiller's (1982) result is presented in Section IV of a long survey paper, most of which covers volatility tests. The equity premium is, to Shiller, one of many interesting aspects of fitting the consumption-based model to data, and not the most important. The introduction makes no mention of the calculation. Instead, it advises that "the bulk of this paper will be an exploratory data analysis," and will present "the broadest possible array of evidence relevant to judging the plausibility of the model." It advertises that the paper will focus on ... "three substantive questions," the business-cycle behavior of interest rates, the accuracy of consumption data and the fact that few consumers hold stocks, and whether prices are too volatile—and does not include risk aversion and the equity premium in this list. Section IV first reviews other risk aversion estimates, gives a reminder of a different, volatility-test-based discount factor volatility calculation in Shiller (1991), and only then presents the result quoted above. The conclusion (p. 231) briefly mentions the calculation, among many others, but phrases it as "encouraging for the model" since large $\sigma(S)$ can rationalize volatile prices, not noting that large $\sigma(S)$ and smooth $\sigma(\Delta C)$ imply huge risk aversion. It is not a surprise that readers did not seize on the puzzle and run with it as they did after reading Mehra and Prescott. (Hansen's (1982) comment on Shiller did notice the bounds on the volatility of marginal rates of

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 263 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

substitution, and sharpened and extended Shiller's calculations; one can see the roots of the Hansen–Jagannathan (1991) bounds here very clearly.)

Grossman and Shiller (1981) devote almost their entire paper to tests of price volatility. Only in the very last paragraph, in a section titled "Further research," do they write

> We have some preliminary results on the estimation of A (γ) and β. Estimates of both parameters can be derived using expression (3) ($1 = E(mR^i)$) for two different assets which we took as stocks and short-term bonds. 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 ...

They do not present the actual estimates or document them in any more detail than these sentences, though one may surmise that working paper versions of this paper presented more details. It would have been truly extraordinary if a verbal report of "preliminary" and "implausible" results, attributed to peculiarities of one data sample, at the end of a Papers and Proceedings elaboration of volatility tests, were to launch the equity-premium ship. (Volatility tests are also an important contribution, and with 211 citations this is a highly influential paper. The point here is not to diminish volatility tests but to track down why this paper did not *also* launch the equity premium.)

Grossman, Melino, and Shiller (1987) is the other published work to result from Grossman and Shiller's early 1980s', risk aversion estimates. This paper starts with a simple table (Table 1, p. 318) of risk aversion estimates based on $E(R^e) \approx \gamma \operatorname{Cov}(R^e, \Delta c)$, and reports estimates between 13.8 and 398, depending on data set. "Table 1 shows that the mean excess return on stocks is associated with a relatively small covariance with consumption changes. If we ignore sampling and measurement error, this can be justified only by an implausibly high estimate of the risk-aversion parameter (see also Mehra and Prescott 1985)." This calculation shows that the low correlation between consumption growth and returns is another part of the problem, already extending the puzzle. At this point, though, the paper has become an explain-the-equity-premium paper, devoted to the question of whether a sophisticated treatment of time aggregation in consumption will overturn the result, and coming to the conclusion that it doesn't do so.

Hansen and Singleton (1983) also report a high risk aversion estimate. Hansen and Singleton describe the result in 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 264 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

> 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.

Clearly, the point of this paper is to introduce instruments and to study varying conditioning information and how that conditioning information can be used to sharpen estimates. The bulk of this paper studies intertemporal substitution, how consumption-growth forecasts line up with interest rate forecasts, which involves one asset at a time and many instruments. The introduction (p. 250) summarizes the crucial idea of the paper as "The predictable components of the logarithms of asset returns are proportional to the predictable components of the change in the logarithm of consumption, with the proportionality factor being minus the coefficient of relative risk aversion." Table 5 is the only table in this paper or in Hansen and Singleton (1982, 1984) that does *not* have instruments, or that *does* given a high risk aversion estimate. These are groundbreaking contributions, as I discuss in detail ahead, but again it's clear how readers can easily miss the equity premium, introduced only as "for the sake of comparison" with Grossman and Shiller, buried in Table 5, summarized as an illustration of the sensitivity of the method to serial correlation, and the finding of high risk aversion needed to explain the unconditional equity premium ignored in the introduction or conclusion.

By contrast, Mehra and Prescott (1985) claim that high risk aversion is a robust and unavoidable feature of any method for matching the model to data. They also argue 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:

> 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 sum, while it's clear the central result can be found in Shiller (1982), Grossman and Shiller (1981), and Hansen and Singleton (1983), it is also pretty clear why readers missed it there.

Part of Mehra and Prescott's influence might also be traced to things they left out. Mehra and Prescott completely avoided inference or standard errors. Alas, the equity premium is not that well measured. $\sigma/\sqrt{T}$ with $\sigma \approx 20$ percent means that in 50 years of data the sample mean is estimated with a $20/\sqrt{50} = 2.8$ percent standard error, so a 6 percent equity premium is barely two standard errors above zero. By ignoring standard errors, they focused attention on an economically interesting moment. But standard

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 265 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

errors are not that hard. Shiller (1982, p. 221) already had them, directly below the above paragraph:

> Of course expected returns and standard deviations of returns are not precisely measured, even in a hundred years of data. An asymptotic standard error for the estimate of the right hand side of the inequality ... was 0.078. Thus, the estimated lower bound for $\sigma(S)$ is only two and a half standard deviations from zero.

Hansen and Singleton (1983) also calculate standard errors. In fact, it is exactly the greater precision of estimates based on predictable movements in consumption growth and returns that drives them to pay more attention to moments with one return and many instruments and their indications of low risk aversion (which we now label "intertemporal substitution") rather than the apparently less well measured moment consisting of stock and bond returns and no instruments, which is central to the equity premium.

In fact, even reading Mehra and Prescott as saying "one needs high risk aversion to explain the equity premium" involves some hindsight. The introduction does not mention high risk aversion, it simply says that the equity premium "cannot" be accounted for by frictionless Arrow–Debreu models. The text on p. 155 documents this fact, in their two-state model, for risk aversion "calibrated" to be less than 10. The possibility that the model might work with high risk aversion is only acknowledged in a footnote describing a private communication with Fischer Black, and stated in the context of a different model.

Mehra and Prescott also gave a *structure* that many people found useful for thinking about variations on the puzzle. A very large number of alternative explicitly-calculated two-state endowment economies followed Mehra and Prescott, though we now understand that the equity premium point really only needs first-order conditions as Shiller derived them and as I summarized earlier. Even the latter approach needed Hansen and Jagannathan's (1991) paper to be revived. It took another army of papers calculating Hansen–Jagannathan bounds to come back in the end to the simple sorts of calculations in Shiller's (1982) original article. Leaving a complex structure for others to play with seems to be a crucial piece of generating followers. Answering a question too quickly is dangerous to your influence.

Mehra and Prescott's general equilibrium modeling imposes extra discipline on this kind of research and has a separate and fully justified place of honor as the progenitor of the general equilibrium models described ahead. In a general equilibrium model, the covariance of consumption with returns is generated endogenously. You can't just take $\mathrm{Cov}(R, \Delta c)$ as given and crank up γ (see Eq. (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 generate 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 266 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 that they are a problem if you only look at first-order conditions.

3.2. The Future of the Equity Premium

My view of the literature is that work "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 later, 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. (Frictions, as advocated by Mehra and Prescott (1985), have not emerged as the consensus answer to the puzzle. In part, this is because high Sharpe ratios occur between pairs of stocks as well as between stocks and bonds.)

At the same time, many economists' beliefs about the size of the equity premium are declining from the 8 percent postwar average, past the 6 percent average in longer samples, down to 2 or 3 percent or less. The U.S. 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 percent per year more than bonds, and shy away from buying more stocks in the full knowledge of this mean, because the 16 percent annual standard deviation of stock returns seemed like too much risk? Or was the 8 percent 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 percent and have declined to about 2 percent. Dividend yields are known ahead of time and 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 percent (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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 267 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 the "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-to-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*, by 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 268 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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

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

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

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

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

Hansen and Singleton's (1984) results provide a useful baseline. If we take a single asset and multiply it by instruments (Hansen and Singleton's Table 1), 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,

$$r^f_t = \delta + \gamma E_t(\Delta c_{t+1}) - \frac{1}{2}\gamma(\gamma+1)\sigma^2_t(\Delta c_{t+1}). \tag{12}$$

If risk premia are not well forecast by these instruments (and they aren't) and consumption is homoskedastic (pretty close), then the main thing underlying 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's (1983) study this linearized version of the consumption-based model, and their Table 4 studies this interest rate equation explicitly.)

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 269 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

TABLE 4

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 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's (1983) Table 5, reproduced here, makes the story clear.

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 weigh the evidence provided by different moments according to the statistical significance of those moments. 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 270 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

4.1.1. 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 above 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 percent) 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 for asset pricing. This is the implicit assumption that allowed us to use only non-durable 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 non-separability is *across goods*. Perhaps the marginal utility of non-durable 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 non-durable 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 271 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

break from all that work, or less valuable since you have less time to enjoy it? will you believe an estimate that relies strongly on one or the other effect?

We can also consider non-separability *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 non-separability 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 non-separable, 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

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

or, directly,

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

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

A difficulty in adding multiple goods is that if the non-separability 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 non-separable *across states of nature*. Epstein and Zin (1991) pioneered this idea in the asset pricing literature, following the thoretical development

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 272 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

by Kreps and Porteus (1978). The expected utility function adds over states, just as separable utility adds over goods,

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

Epstein and Zin propose a recursive formulation of utility:

$$U_t = \left((1-\beta) c_t^{1-\rho} + \beta \left(E_t \left(U_{t+1}^{1-\gamma} \right) \right)^{\frac{1-\rho}{1-\gamma}} \right)^{\frac{1}{1-\rho}}, \tag{13}$$

which, among other things, abandons separability across states of nature. The term $\left(E_t \left(U_{t+1}^{1-\gamma} \right) \right)^{\frac{1}{1-\gamma}}$ 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

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

(The Appendix contains a short derivation.) If $\rho \neq \gamma$, we see a second term; expected returns depend on covariances of returns with the utility index, capturing news about the investor's future prospects, as well as on covariances of returns 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 term.

The utility index itself is not directly measurable, so to make this formula operational we need some procedure for its 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

$$m_{t+1} = \left(\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\rho} \right)^{\theta} \left(\frac{1}{R_{t+1}^W} \right)^{1-\theta}, \tag{15}$$

where

$$\theta = \frac{1-\gamma}{1-\rho}.$$

(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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 273 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 non-traded 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, we can think of the discount factor in terms only of current and future consumption. In the discount factor (14), the utility index is a function of the distribution of *future* consumption, so the essence of the discount factor is that news about *future* consumption matters as well as current consumption in the discount factor.

To see this effect more concretely, we can derive the discount factor for the case $\rho = 1$, and log-normal heteroskedastic consumption. I present the algebra in the Appendix. The result is

$$(E_{t+1} - E_t)\ln m_{t+1} = -\gamma(E_{t+1} - E_t)(\Delta c_{t+1}) + (1-\gamma)\left[\sum_{j=1}^{\infty} \beta^j (E_{t+1} - E_t)(\Delta c_{t+1j})\right], \tag{16}$$

where Δc is log consumption growth, $\Delta c_t = \ln c_t - \ln c_{t-1}$. News about *future* long horizon consumption growth enters the *current* period marginal rate of substitution. Shocks to variables that predict future consumption growth will appear as additional risk factors even with (perfectly measured) current consumption growth. (Campbell (1996, p. 306) pursues the mirror-image expression, in which assets are priced by covariance with current and future wealth-portfolio *returns*, substituting out *consumption*. Restoy and Weil (1998, p. 10) derive an approximation similar to (16) and make this point. Hansen, Heaton, and Li (2006) and Hansen, Heaton, Lee, and Roussanov (2006) derive (16) and show how to make similar approximations for $\rho \neq 1$.)

4.3. Empirics with New Utility Functions

4.3.1. Non-separabilities Across Goods

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

$$U = \sum \beta^t \frac{\left(c_t^{*\theta} l_t^{*1-\theta}\right)^{1-\gamma} - 1}{1-\gamma};$$
$$c_t^* = c_t + \alpha c_{t-1},$$
$$l_t^* = l_t + b l_{t-1} \quad \text{or} \quad l_t^* = l_t + b\sum_{j=0}^{\infty} \eta^j l_{t-j},$$

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 274 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 of whether this extension of the consumption-based model can address what we now understand are the interesting questions.[10]

Eichenbaum and Hansen (1990) investigate a similar model with non-separability between durables and non-durables. This is harder because one needs also to model the relation between observed durable purchases and the service flow that enters the utility function. Also, any model with multiple goods gives rise to an *intratemporal* first-order condition, marginal utility of non-durables/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 cross-sectional risk premia.

In the consumption-based revival, Yogo (2004) reconsiders non-separability 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-to-market portfolios, along with the 3-month T-bill rate, and including the intratemporal first-order condition for durables vs. non-durables, 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 non-durables. As in the discussion of Piazzesi, Schneider, and Tuzel ahead, the difference between this modest elasticity and the much smaller σ and $1/\gamma$ means that the non-separabilities 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:

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

[10]Lettau's (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 are 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 275 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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. non-durables, using the non-linear specification, dealing with the intratemporal first-order condition (durable vs. non-durable 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 non-durable consumption in order to make quantitatively important differences to asset pricing. To keep the durable vs. non-durable first-order condition happy, given the downward trend in the ratio of durables to non-durables, he adds an income elasticity (non-homothetic preferences).

4.3.2. Habits

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

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

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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 276 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 to capture this observation would be an AR(1),

$$X_t=\rho X_{t-1}+\lambda C_t. \tag{18}$$

(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 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 non-linear version of (18). This non-linear version allows us to avoid an Achilles heel of many habit models, a 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 anxiety 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 277 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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,

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

where $S_t = (C_t - X_t)/C_t$ and X_t is habit. A large number of models amount to something like Eq. (19), in which the discount factor generalizes the power utility case by adding another state variable. 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).

There is a danger in models of the form (19) 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 a large Sharpe ratio, we need a large volatility of the discount factor $\sigma(m)$, and to generate a large Sharpe ratio 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.

The Campbell–Cochrane model turns out not to suffer from this problem: while S_t is stationary, the conditional variance of $S_t^{-\gamma}$ 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 non-linear habit accumulation process. It may not be there in many superficially attractive simplifications or linearizations of the habit model.

However, though the maximum Sharpe ratio, driven by $\sigma(m_{t,t+k})$ remains high at long horizons, this fact does not necessarily mean that the average returns of all assets remain high at long horizons. For example, a consumption claim gets a high premium at a one-year horizon, since C_{t+1} and S_{t+1} are correlated, so the consumption claim payoff

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 278 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

covaries a great deal with the discount factor. However, at long horizons, consumption and $S_{t+k}^{-\gamma}$ become uncorrelated, so a long-term consumption claim will not attain the still-high Sharpe ratio bound.

Simulation is a prequel to empirical work, not a substitute for it, 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-to-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-to-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 that the resulting consumption-based model performs quite well, even better than the Fama–French three-factor model. Within this structure, they find that 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 later.

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 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 at all with U.S. 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). Voilá, the interest rate spread forecasts a time-varying exchange rate risk premium. More generally, these papers pave the way to go beyond equity, value, size, and momentum premiums to start thinking about bond risk premia and foreign exchange risk premia.

4.3.3. 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 279 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Constantinides and Duffie (1996), discussed aheads, 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 non-separable between non-housing consumption and consumption of housing services; you need a roof to enjoy the new TV. Thus, the marginal rate of substitution or stochastic discount factor is

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

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 non-durable and services aggregate, of course, and the fact that utility is non-separable across two components of the index does not invalidate the theory behind the use of aggregate consumption. Therefore, 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 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, 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 ratios. They verify the latter prediction, adding to the list of macro variables that forecast returns. (See Tables 4 and 5.) Finally, the model generates slow-moving variation in price-dividend ratios and stock return volatility, all coming from risk premia rather than

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 280 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 or any high long-run Sharpe ratios.

Lustig and Van Niewerburgh (2004a, 2004b) explore a similar model. Here, variations in housing collateral play the role of the "habit." Consumer-investors (-homeowners) whose housing collateral declines become effectively more risk-averse. Lustig and Van Niewerburgh show that variations in housing collateral predict stock returns in the data, as the surplus consumption ratio predicts stock returns in the Cambpell–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, 2001b) *cay* state variable.

Chetty and 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. They also show that people who have recently moved for exogenous reasons hold a smaller proportion of stocks, acting in more risk-averse manner.

4.3.4. 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* is 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 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 a high frequency, too. Clothing is "non-durable."

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 to 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 281 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 non-durable and services consumption, the simple consumption-based model can account for the Fama–French 25 size and book-to-market portfolios. Figure 3 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.

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

$$1 = E_t\left(\beta^k\left(\frac{C_{t+k}}{C_t}\right)^{-\gamma} R_{t+1}R^f_{t+1}R^f_{t+2}\ldots R^f_{t+k-1}\right). \tag{21}$$

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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 282 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

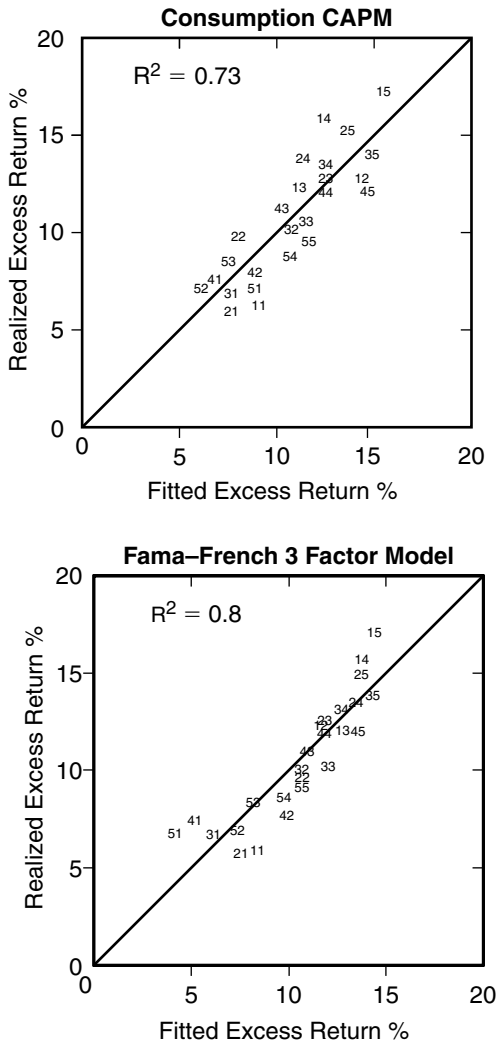

FIGURE 3 Top panel: Average returns of Fama–French 25 portfolios vs. predictions of the linearized consumption-based model (essentially, consumption betas) and vs. predictions of the Fama–French 3 factor model. Fourth-quarter to fourth-quarter data, 1954–2003.
(*Source*: Jagannathan and Wang (2005),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

4.3.5. 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 283 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

more power to forecast returns. In essence, their empirical work, using the discount factor

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

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,

$$m_{t+1} = a - bR^W_{t+1}.$$

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 model ties the discount factor to news about future consumption as well as to current consumption; in the $\rho = 1$ log-normal homoskedastic case,

$$\begin{aligned}(E_{t+1} - E_t)\ln m_{t+1} = &-\gamma(E_{t+1} - E_t)(\Delta c_{t+1}) \\ &+ (1-\gamma)(E_{t+1} - E_t)\left[\sum_{j=1}^{\infty}\beta^j(\Delta c_{t+1+j})\right].\end{aligned} \tag{22}$$

Hansen, Heaton, and Li (2006) 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^f_{t+j} 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 (22) in a simulation economy context. Concentrating 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 (22) can then act as an "extra factor" to generate a high equity premium, a 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 cash flows 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 (2006) show that Bansal, Dittmar, and Lundblad's evidence that value stocks have much different long-run consumption betas than do

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 284 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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.

4.3.6. 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 earlier, 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)?

Consumption 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 (22), 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 $\sum_{j=2}^{k} \Delta c_{t+k}$, 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 percent at a one-year horizon. That R^2 is a good deal lower at longer horizons we are interested in, 1.23 percent at 3 years and 0.15 percent at nearly 4 years, and some of that predictability comes from the 1.78 percent R^2 from explaining Δc_{t+1} from returns at time $t + 1$.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 285 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Long-run properties of anything are hard to measure, as made clear in this context by the Hansen, Heaton, and Li (2006) 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 (22)) act like unobservable shifts in marginal utility, or shifts in "sentiment," which are always suspicious explanations for anything. At a minimum, explanation-based, difficult-to-observe shifts in long-run consumption growth should parsimoniously tie together many asset pricing phenomena.

Epstein–Zin utility has another unfortunate implication, that we really have to consider all components of consumption. We usually focus on non-durable and services consumption, ignoring durables. This is justified if the utility function is separable across goods, $u(c_{nds}) + v(c_d)$, where c_{nds} is consumption of nondurables and services, and c_d is the flow of services from durables. Alas, even if the *period* utility function is separable in this way, the resulting Epstein–Zin utility *index* responds to news about future durables consumption. In this way, the non-separability across *states* induces a non-separability across *goods*, which really cannot be avoided (see Uhlig (2006)).

4.3.7. 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) start 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 $E_t(R^e_{t+1})/\sigma_t(R^e_{t+1}) \approx \gamma\sigma_t(\Delta c_{t+1})$, 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 Epstien–Zin literature is to some extent going back to this framework. Bansal and Yaron (2004), for example, add conditional heteroskedasiticty 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 286 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

4.4. Consumption and Factor Models

A second tradition also has re-emerged 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 simply ask whether average returns line up with betas computed relative to consumption growth, after correcting for a number of problems with consumption data and using 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 instruments variables such as the dividend price ratio that forecast returns much better than lagged returns.

Lettau and Ludvigson (2001b) is the first modern re-examination 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,

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

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 to 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:

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

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,

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

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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 287 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Lettau and Ludvigson's Figure 1, reproduced here as Figure 4, 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

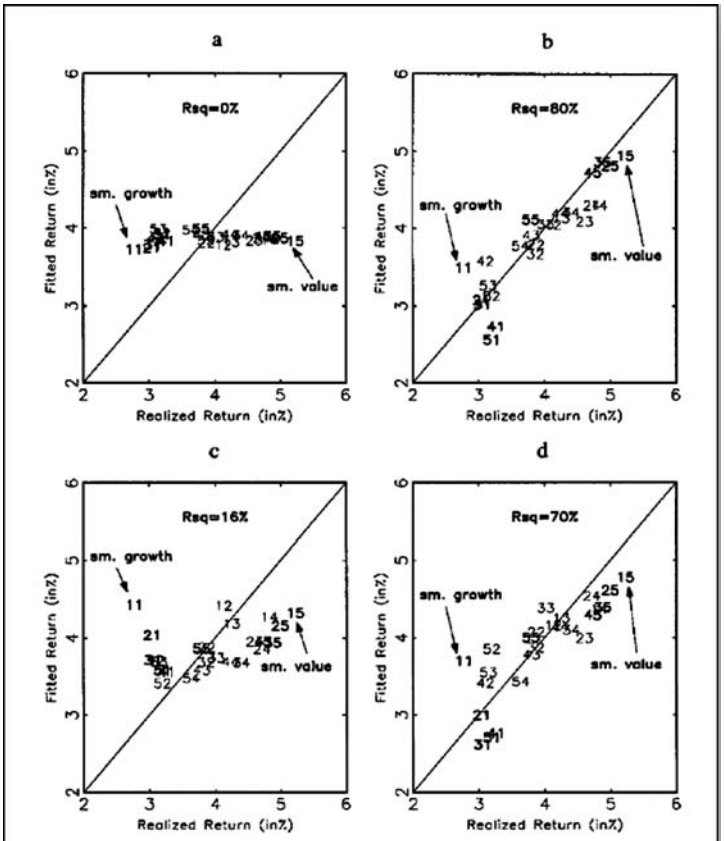

FIG. 1.—Realized vs. fitted returns: 25 Fama-French portfolios: *a*, CAPM; *b*, Fama-French; *c*, consumption CAPM; *d*, consumpion CAPM scaled. The figure shows the pricing errors for each of the 25 Fama-French portfolios for the four models. Each two-digit number represents one portfolio. The first digit refers to the size quintiles (1 indicating the smallest firms, 5 the largest), and the second digit refers to book-to-market quintiles (1 indicating the portfolio with the lowest book-to-market ratio, 5 with the highest). The pricing errors are generated using the Fama-MacBeth regressions in table 3 below. The scaling variable is $\widehat{cay}$.

FIGURE 4 Lettau and Ludvigson's Figure 1.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 288 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 percent pricing errors is a lot more interesting than a "non-rejected" model with 10 percent 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 encourage 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 of computing R^2 can give wildly different results.[11]

These criticisms are, of course, solved by statistical measures; test statistics based on $\alpha' \operatorname{Cov}(\alpha, \alpha')^{-1}\alpha$, 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 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

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

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

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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 289 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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. Correctly-done standard errors will 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 that forecast GDP growth can price the Fama–French 25 portfolios. Let's not repeat the mistake of the CAPM that hid the implied 16 percent volatility of consumption growth or extraordinary risk aversion for so many years.

4.4.1. 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 30 years of effort, the consumption-based model and its variants have 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-to-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 the prediction of perfect correlations) that are easily rejected by formal statistical tests.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 290 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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), and 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 non-separability across time, goods, and states behave 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 291 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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.

5.1.1. 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 empirical result in Cochrane (1991b) 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 to 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 conceive 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 to 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 292 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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). Merz and Yashiv's Figure 2 summarizes this central finding well.

5.1.2. 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 of whether the investment returns from these technologies span stock returns, i.e., whether a discount factor of the form

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

satisfies

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

for a cross section of asset returns R_{t+1}. Here $R^{(i)}_{t+1}$ denote the investment returns, functions of investment, and capital only, i.e., $R^{(i)}_{t+1} = f(I^i_{t+1}/K^i_{t+1}, I^i_t/K^i_t)$. The paper also explores scaled factors and returns to incorporate conditioning information, (though Cochrane (2004) does a better job of summarizing this technique) and 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-to-market portfolios, extending the list of macro models that can account for the value effect.

5.1.3. 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 293 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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,

$$\Theta: \left[E\left(\frac{\varepsilon_{t+1}}{\theta_{t+1}}\right)^{\alpha}\right]^{\frac{1}{\alpha}} = \left[\sum_s \pi_s \left(\frac{\varepsilon_s}{\theta_s}\right)^{\alpha}\right]^{\frac{1}{\alpha}} = 1, \tag{23}$$

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

$$m_{t+1} = \lambda \frac{y_{t+1}^{\alpha-1}}{\theta_{t+1}^{\alpha} f(k_t)^{\alpha}}. \tag{24}$$

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 non-stochastic models and then tacked on shocks. They did not come from a detailed microeconomic investigation that 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 294 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Belo (2005) takes a crucial step to making this approach work, by proposing a solution to the problem of identifying θ_{t+1} in (24). 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 $c(I, Y, \ldots)$ 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.

5.2.1. 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 in 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 295 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 non-linear 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 (23) 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 296 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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.

5.2.2. 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 didn't anybody notice all the missing pieces before? Why did a generation of macroeconomists trying to match quantity dynamics not find themselves forced to adopt 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.

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

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

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

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

where X is stochastic productivity and N is labor. The Epstein–Zin preferences allow him to raise risk aversion while keeping intertemporal substitution constant. As he does so, he is better 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 297 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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. Or perhaps we should chide them further for continuing to ignore the asset-market prediction of their models.

5.2.3. 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 multiplied by a large change in the relative *price* of (installed) capital. Yes, there was a surge of measured investment, but the resulting increase in the quantity of capital did not come close to accounting for the large increase in stock market valuations.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 298 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 "intangible output?" 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. Among others, see Hansen, Heaton, and Li (2005).) 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 into 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.

5.2.4. The Cross Section of Returns

Obviously, the range of asset pricing phenomena addressed by this sort of general-equilibrium 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 cash flows of the multiple technologies, but not the investment and labor decisions that go behind these cash flows. They specify a clever model for the *shares* of each cash flow 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 cash flow is temporarily low, the duration of that cash flow is longer since more of the expected cash flows are pushed out to the future. This makes the cash flow more exposed to the aggregate risk premium, giving it a higher expected return and a lower price.

The obvious next step is to amplify the model's 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) and Gomes, Kogan, and Zhang (2003) derive size and book-to-market effects in general equilibrium models with a bit more explicit, but also fairly stylized, technologies. For example, Gomes, Kogan, and Zhang envision

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 299 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

"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-to-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 are 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.

5.2.5. 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 wants to address the cross section of returns. Since the extra technologies represent non-zero 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 cash-flow 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 cash flows and some firms have low-beta cash flows 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 300 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 non-stochastic 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 percent is much larger than the interest rate of 1 percent. Thinking of risk as a "second-order" effect, expanding around a 1 percent 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 non-linear 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 is 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 $\mathrm{Cov}_t(R_{t+1}, \Delta c_{t+1})$ or $\mathrm{Cov}_t(R_{t+1}, R^m_{t+1})$? 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 301 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 $E_t(R^e_{t+1}) \approx \gamma \operatorname{Cov}_t(R^e_{t+1}, \Delta c_{t+1})$, 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 about 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.[12] 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.

[12]The 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 302 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 303 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Still, the intuition for the importance of labor income risk is strong. The paragraph from Fama and French (1996, p. 77) quoted earlier 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.

6.1.1. 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,

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

where the betas are defined as usual from time-series regressions,

$$R_t^i = a + \beta_i^{VW} VW_t + \beta_i^{\text{prem}}\text{prem}_t + \beta_i^{\text{labor}}\text{labor}_t + \varepsilon_t^i;$$

where 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 × VW and prem × labor as factors, as in Lettau and Ludvigson's (2001b) conditional CAPM.)

With VW and prem alone, Jagannathan and Wang report only 30 percent 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 percent cross-sectional[13] 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 three-factor model does no better (55 percent cross-sectional R^2) on their portfolios, but we don't know from the paper if labor income prices the book-to-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 $\text{labor}_t = (L_{t-1} + L_{t-2})/(L_{t-2} - L_{t-3})$ 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.

Much of Jagannathan and Wang's empirical point can be seen in Table 1 of Lettau and Ludvigson (2001b), reproduced here as Figure 5. Δy is labor income growth,

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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 304 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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-to-market portfolios, which allows us better to compare across models in this standard playground. They actually find reasonable performance (58 percent 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.,

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

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

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.

6.1.2. 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, and (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 de-trended 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) and 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-to-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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 305 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

TABLE 1
Fama–MacBeth Regressions Using 25 Fama–French Portfolios: λ_j Coefficient Estimates on Betas in Cross–Sectional Regression

			Factors$_{t+1}$				$\widehat{cay}_t \cdot$ Factors$_{t+1}$		R^2
Row	Constant	$\widehat{cay}_t$	R_{uw}	Δ_y	smb	hml	R_{uw}	Δ_y	$\left(\overline{R}^2\right)$
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: $E\left[R_{i,t+1}\right] = E\left[R_{0,t}\right] + \beta'\lambda$. 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_{uw} 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 $\widehat{cay}$. 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^2denotes the unadjusted cross-sectional R^2 statistic, and R^2 adjusts for the degrees of freedom.

FIGURE 5 Lettau and Ludvigson's Table 1.

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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 306 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Campbell and Vuolteenaho (2004) follow on the ICAPM component of Campbell (1996). They break the standard CAPM beta into two components, a "bad" cash-flow beta that measures how much an asset return declines if expected future market cash flows 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.

6.1.3. 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.

6.1.4. 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-to-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 of human capital that should matter in theory. They find good performance of the model in book-to-market sorted portfolios, suggesting that labor income risk (or associated macroeconomic risk) really is behind the "value effect."

6.1.5. 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 307 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 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-to-market portfolios (Santos and Veronesi's 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 percent 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 Micro Data

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 of whether cross-sectional distributions matter. Two particular strains of this investigation are important for us. First, 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.

6.2.1. 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 308 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 to 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:

$$E_t\left\{\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\exp\left[\frac{\gamma(\gamma+1)}{2}y_{t+1}^2\right]R_{t+1}\right\}=1. \tag{25}$$

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 (25) 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}^2varies 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 non-linear function of consumption. Examining $E(m_{t+1}^i R_{t+1})=E\left[E(m_{t+1}^i|R_{t+1})R_{t+1}\right]$, we see that a non-linear m will lead to a Jensen's inequality $1/2\sigma^2$-term, which is exactly the exponential term in (25). Thus, if the *cross-sectional* variance of idiosyncratic shocks is higher when the returns R_{t+1} are higher, we will see a premium that does not make sense from aggregate consumption. The derivation of (25) follows exactly this logic and doesn't take much extra algebra.[14]

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*, 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*.

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

$$\ln\left(\frac{C_{t+1}^i}{C_t^i}\right)=\ln\left(\frac{C_{t+1}}{C_t}\right)+\eta_{i,t+1}y_{t+1}-\frac{1}{2}y_{t+1}^2. \tag{26}$$

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 $-1/2y_{t+1}^2$ term is there exactly for this reason:

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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 309 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 percent or so. ($E(R^e) \approx 8$ percent, $\sigma(R^e) \approx 16$ percent, $R^f = 1/E(m) \approx 1.01$.) We can make some back-of-the-envelope calculations with the approximation

$$\sigma\left\{\exp\left[\frac{\gamma(\gamma+1)}{2} y_{t+1}^2\right]\right\} \approx \frac{\gamma(\gamma+1)}{2}\sigma\left(y_{t+1}^2\right). \tag{27}$$

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 percent, the typical person you meet either has +73 percent or −69 percent 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 an 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.

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

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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 310 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 $\sigma_t(y^2_{t+h})$ 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^2_{t+h} governs the *variance* of individual consumption growth, not its *standard deviation*, and variances usually decline linearly with horizon. If $\sigma_t(y^2_{t+h})$ declines only with the square root of horizon, then typical values of the *level* of y^2_{t+h} must also decline only with the square root of horizon, since y^2_{t+h} 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^2_t 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,

$$E_t(dR_t) - r^f_t\,dt = \gamma E_t\left(dR_t \frac{dC^i_t}{C^i_t}\right), \tag{28}$$

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

$$E_t(dR_t) - r^f_t\,dt = \gamma E_t\left(dR_t \frac{dC_t}{C_t}\right).$$

Aggregation holds even with incomplete markets and non-linear 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 diffusion.[15]

Conversely, we may anticipate the same generic problem that many models have at long horizons. Like many models (see the Cambpbell–Cochrane discussion earlier), the Constantinides–Duffie model (25) 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 of

[15]There is another logical possibility. $E_t(dR_t) = r^f_t\,dt$ does not imply $E_t(R_{t+1}) = R^f_t$ 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.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 311 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

idiosyncratic shocks, usually do not grow with horizon, leaving us back to the power utility model at long horizons.

6.2.2. 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 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:

$$1 = E\left(\beta \frac{u'(C^i_{t+1})}{u'(C^i_t)} R_{t+1}\right). \tag{29}$$

We therefore can always "aggregate" by averaging *marginal utilities*:

$$1 = E\left(\left[\frac{1}{N}\sum_i \beta \frac{u'(C^i_{t+1})}{u'(C^i_t)}\right] R_{t+1}\right). \tag{30}$$

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 312 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

We cannot in general aggregate by averaging *consumption*:

$$1 \neq E\left(\beta \frac{u'\left(\frac{1}{N}\sum_i C^i_{t+1}\right)}{u'\left(\frac{1}{N}\sum_i C^i_t\right)} R_{t+1}\right). \tag{31}$$

Brav, Constantinides, and Geczy contrast calculations of (30) with those of (31). This analysis also shows again how important non-linearities 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 by 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 do 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.

6.2.3. Micro Data

Of course, *individuals* still price assets exactly as before. Equation (29) still holds for each individual's consumption in all these models. So, once we have opened the CES or PSID database, we could simply test whether asset returns are correctly related to household level consumption with (29) and forget about aggregation either of consumption (31) or of marginal utility (30). 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 313 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 percent 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 non-durable consumption and the flow of durables services, not durables purchases. Buying a house once in 10 years or a car once in 3 does not count toward 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 microdata. 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 (29) in microdate, 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 micro data? While initially appealing, it's not clear that the stockholder/non-stockholder 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 suggest. 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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 314 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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.

References

Abel, A. B. Asset prices under habit formation and catching up with the Joneses. *American Economic Review* 80 (1990): 38–42.

Ait-Sahalia, Y., J. Parker, and M. Yogo. Luxury goods and the equity premium. *Journal of Finance* 59 (2004): 2959–3004.

Alvarez, F., and U. J. Jermann. Using asset prices to measure the cost of business cycles. *Journal of Political Economy* 112 (2004): 1223–1256.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 315 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Ang, A., M. Piazzesi, and M. Wei. What does the yield curve tell us about GDP growth? *Journal of Econometrics* (2004).

Attanasio, O. P., J. Banks, and S. Tanner. Asset holding and consumption volatility. *Journal of Political Economy* 110 (2002): 771–792.

Ball, R. Anomalies in relationships between securities' yields and yield—surrogates. *Journal of Financial Economics* 6 (1978): 103–126.

Bansal, R., R. F. Dittmar, and C. Lundblad. Consumption, dividends, and the cross-section of equity returns. *Journal of Finance* 60 (2005): 1639–1672.

Bansal, R., and A. Yaron. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance* 59(4) (2004): 1481–1509.

Banz, R. W. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9 (1981): 3–18.

Basu, S. The relationship between earnings yield, market value, and return for NYSE common stocks: Further evidence. *Journal of Financial Economics* 12 (1983): 129–156.

Belo, F. A pure production-based asset pricing model. Manuscript, University of Chicago (2005).

Berk, J. B., R. C. Green, and V. Naik. Optimal investment, growth options and security returns. *Journal of Finance* 54 (1999): 1153–1607.

Boldrin, M., L. J. Christiano, and J. Fisher. Habit persistence, asset returns, and the business cycle. *American Economic Review* 91 (2001): 149–166.

Brainard, W. C., W. R. Nelson, and M. D. Shapiro. The consumption beta explains expected returns at long horizons. Manuscript, Economics Department. Yale University (1991).

Brav, A., G. Constantinides, and C. Geczy. Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. *Journal of Political Economy* 110 (2002): 793–824.

Breeden, D. T. An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics* 7 (1979): 265–296.

Breeden, D., M. Gibbons, and R. Litzenberger. Empirical tests of the consumption-oriented CAPM. *Journal of Finance* 44 (1989): 231–262.

Brennan, M. J., Y. Xia, and A. Wang. Estimation and test of a simple model of intertemporal asset pricing. *Journal of Finance* 59 (2005): 1743–1776.

Buraschi, A., and A. Jiltsov. Inflation risk premia and the expectations hypothesis. *Journal of Financial Economics* 75 (2005): 429–490.

Campbell, J. Y. Intertemporal asset pricing without consumption data. *American Economic Review* 83 (1993): 487–512.

Campbell, J. Y. Some lessons from the yield curve. *Journal of Economic Perspectives* 9 (1995): 129–152.

Campbell, J. Y. Understanding risk and return. *Journal of Political Economy* 104 (1996): 298–345.

Campbell, J. Y. Asset pricing at the millennium. *Journal of Finance* 55 (2000): 1515–1567.

Campbell, J. Y. Consumption-based asset pricing. Chapter 13 in G. Constantinides, M. Harris, and R. Stulz, eds., *Handbook of the Economics of Finance* Vol. IB. North-Holland, Amsterdam, (2003): pages 803–887. .

Campbell, J. Y., and J. H. Cochrane. By force of habit: A consumption-based explanation of aggregate stock market behavior. NBER Working paper 4995 (1995).

Campbell, J. Y., and J. H. Cochrane. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107 (1999): 205–251.

Campbell, J. Y., and J. H. Cochrane. Explaining the poor performance of consumption based asset pricing models. *Journal of Finance* 55 (2000): 2863–2878.

Campbell, J. Y., R. J. Shiller, and K. L. Schoenholtz. Forward rates and future policy: Interpreting the term structure of interest rates. *Brookings Papers on Economic Activity* (1983): 173–223.

Campbell, J. Y., and R. J. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1 (1988): 195–228.

Campbell, J. Y., and R. J. Shiller. 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 (1991): 495–514.

Campbell, J. Y., and L. M. Viceira. Consumption and portfolio decisions when expected returns are time varying. *Quarterly Journal of Economics* 114 (1999): 433–495.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 316 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Campbell, J. Y., and T. Vuolteenaho. Good beta, bad beta. *American Economic Review* 94 (2004): 1249–1275.

Carhart, M. On persistence in mutual fund performance. *Journal of Finance* 52 (1997): 57–82.

Chan, L., and L. Kogan. Catching up with the Jones: Heterogeneous preferences and the dynamics of asset prices. *Journal of Political Economy* 110 (2001): 1255–1285.

Chen, N. F. Financial investment opportunities and the macroeconomy. *Journal of Finance* 46 (1991): 529–554.

Chen, N. F., R. Roll, and S. A. Ross. Economic forces and the stock market. *Journal of Business* 59 (1986): 383–403.

Chen, X., and S. Ludvigson. Land of addicts? An empirical investigation of habit-based asset pricing models. Manuscript, New York University (2004).

Chetty, R., and A. Szeidl. Consumption commitments: Neoclassical foundations for habit formation. Manuscript, University of California at Berkeley (2004).

Cochrane, J. H. The sensitivity of tests of the intertemporal allocation of consumption to near-rational alternatives. *American Economic Review* 79 (1989): 319–337.

Cochrane, J. H. Explaining the variance of price-dividend ratios. *Review of Financial Studies* 5 (1991a): 243–280.

Cochrane, J. H. Production-based asset pricing and the link between stock returns and economic fluctuations. *Journal of Finance* 46 (1991b): 207–234.

Cochrane, J. H. Rethinking production under uncertainty. Manuscript, University of Chicago (1993).

Cochrane, J. H. Permanent and transitory components of GNP and stock prices. *Quarterly Journal of Economics* 109 (1994): 241–266.

Cochrane, J. H. A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy* 104 (1996): 572–621.

Cochrane, J. H. Where is the market going? Uncertain facts and novel theories. *Economic Perspectives*, Federal Reserve Bank of Chicago 21(6) (November/December) (1997).

Cochrane, J. H. New facts in finance. *Economic Perspectives*, Federal Reserve Bank of Chicago 23(3) (1999a): 36–58.

Cochrane, J. H. Portfolio advice for a multifactor world. *Economic Perspectives*, Federal Reserve Bank of Chicago 23(3) (1999b): 59–78.

Cochrane, J. H. *Asset Pricing*. Princeton University Press, Revised Edition, Princeton, NJ (2004).

Cochrane, J. H. Liquidity trading and asset prices. *NBER Reporter*, National Bureau of Economic Research, www.nber.org/reporter (2005a).

Cochrane, J. H. Financial markets and the real economy. *Foundations and Trends in Finance* 1 (2005b): 1–101.

Cochrane, J. H. Financial markets and the real economy, in J. H. Cochrane, ed., *Financial Markets and the Real Economy*, Volume 18 of the International Library of Critical Writings in Financial Economics, Edward Elgar, London, (2006a): pages xi–lxix.

Cochrane, J. H. The dog that did not bark: A defense of return predictability. *Review of Financial Studies* (2006b).

Cochrane, J. H., and L. P. Hansen. Asset pricing explorations for macroeconomics. In O. Blanchard, and S. Fisher, eds., *NBER Macroeconomics Annual* (1992): pages 115–165.

Cochrane, J. H., and M. Piazzesi. The Fed and interest rates: A high-frequency identification. *American Economic Review* 92 (2002): 90–95.

Cochrane, J. H., and M. Piazzesi. Bond risk premia. *American Economic Review* 95 (2005): 138–160.

Cogley, T. Idiosyncratic risk and the equity premium: Evidence from the consumer expenditure survey. *Journal of Monetary Economics* 49 (2002): 309–334.

Constantinides, G. Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy* 98 (1990): 519–543.

Constantinides, G., and D. Duffie. Asset pricing with heterogeneous consumers. *Journal of Political Economy* 104 (1996): 219–240.

Cooper, I., and R. Priestley. Stock return predictability in a production economy. Manuscript, Norwegian School of Management (2005).

Craine, R. Rational bubbles: A test. *Journal of Economic Dynamics and Control* 17 (1993): 829–846.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 317 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Daniel, K., and D. Marshall. Equity-premium and risk-free-rate puzzles at long horizons. *Macroeconomic Dynamics* 1 (1997): 452–484.

Daniel, K., and S. Titman. Testing factor-model explanations of market anomalies. Manuscript, Northwestern University and University of Texas, Austin (2005).

De Bondt, W. F. M., and R. Thaler. Does the stock market overreact? *Journal of Finance* 40 (1985): 793–805.

De Santis, M. Interpreting aggregate stock market behavior: How far can the standard model go? Manuscript, University of California, Davis (2005).

Eichenbaum, M., and L. P. Hansen. Estimating models with intertemporal substitution using aggregate time series data. *Journal of Business and Economic Statistics* 8 (1990): 53–69.

Eichenbaum, M., L. P. Hansen, and K. Singleton. A time-series analysis of representative agent models of consumption and leisure choice under uncertainty. *Quarterly Journal of Economics* 103 (1988): 51–78.

Engel, C. The forward discount anomaly and the risk premium: A survey of recent evidence. *Journal of Empirical Finance* 3 (1996): 123–192.

Epstein, L. G., and S. E. Zin. Substitution, risk aversion and the temporal behavior of asset returns. *Journal of Political Economy* 99 (1991): 263–286.

Estrella, A., and G. Hardouvelis. The term structure as a predictor of real economic activity. *Journal of Finance* 46 (1991): 555–576.

Fama, E. F. Short-term interest rates as predictors of inflation. *American Economic Review* 65 (1975): 269–282.

Fama, E. F. Forward rates as predictors of future spot rates. *Journal of Financial Economics* 3 (1976): 361–377.

Fama, E. F. Forward and spot exchange rates. *Journal of Monetary Economics* 14 (1984a): 319–338.

Fama, E. F. The information in the term structure. *Journal of Financial Economics* 13 (1984b): 509–528.

Fama, E. F. Stock returns, expected returns, and real activity. *Journal of Finance* 45 (1990): 1089–1108.

Fama, E. F. Efficient markets: II, Fiftieth Anniversary Invited Paper. *Journal of Finance* 46 (1991): 1575–1617.

Fama, E. F., and R. R. Bliss. The information in long-maturity forward rates. *American Economic Review* 77 (1987): 680–692.

Fama, E. F., and K. R. French. Permanent and temporary components of stock prices. *Journal of Political Economy* 96 (1988a): 246–273.

Fama, E. F., and K. R. French. Dividend yields and expected stock returns. *Journal of Financial Economics* 22 (1988b): 3–27.

Fama, E. F., and K. R. French. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25 (1989): 23–49.

Fama, E. F., and K. R. French. The cross-section of expected stock returns. *Journal of Finance* 47 (1992): 427–465.

Fama, E. F., and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1993): 3–56.

Fama, E. F., and K. R. French. Multifactor explanations of asset-pricing anomalies. *Journal of Finance* 51 (1996): 55–84.

Fama, E. F., and K. R. French. Size and book-to-market factors in earnings and returns. *Journal of Finance* 50 (1997a): 131–155.

Fama, E. F., and K. R. French. Industry costs of equity. *Journal of Financial Economics* 43 (1997b): 153–193.

Fama, E. F., and M. R. Gibbons. Inflation, real returns and capital investment. *Journal of Monetary Economics* 9 (1982): 297–323.

Fama, E. F., and G. W. Schwert. Asset returns and inflation. *Journal of Financial Economics* 5 (1977): 115–146.

Ferson, W. E., and G. Constantinides. Habit persistence and durability in aggregate consumption: Empirical tests. *Journal of Financial Economics* 29 (1991): 199–240.

Ferson, W. E., and C. R. Harvey. Conditioning variables and the cross section of stock returns. *Journal of Finance* 54 (1999): 1325–1360.

Gala, V. D. Investment and returns. Manuscript, University of Chicago (2006).

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 318 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Goetzmann, W. N., and P. Jorion. Testing the predictive power of dividend yields. *Journal of Finance* 48 (1993): 663–679.

Gomes, F., and A. Michaelides. Asset pricing with limited risk sharing and heterogeneous agents. Manuscript, London Business School (2004).

Gomes, J. F., L. Kogan, and L. Zhang. Equilibrium cross-section of returns. *Journal of Political Economy* 111 (2003): 693–732.

Gourio, F. Operating leverage, stock market cyclicality and the cross-section of returns. Manuscript, University of Chicago (2004).

Goyal, A., and I. Welch. Predicting the equity premium with dividend ratios. *Management Science* 49 (2003): 639–654.

Goyal, A., and I. Welch. A comprehensive look at the empirical performance of equity premium prediction. Manuscript, Brown University, Revision of NBER Working Paper 10483 (2005).

Grossman, S., A. Melino, and R. J. Shiller. Estimating the continuous-time consumption-based asset-pricing model. *Journal of Business and Economic Statistics* 5 (1987): 315–328.

Grossman, S. J., and R. J. Shiller. The determinants of the variability of stock market prices. *American Economic Review* 71 (1981): 222–227.

Grossman, S. J., and R. J. Shiller. Consumption correlatedness and risk measurement in economies with non-traded assets and heterogeneous information. *Journal of Financial Economics* 10 (1982): 195–210.

Hall, R. E. Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. *Journal of Political Economy* 86 (1978): 971–987.

Hall, R. E. Intertemporal substitution in consumption. *Journal of Political Economy* 96 (1988): 339–357.

Hall, R. E. The stock market and capital accumulation. *American Economic Review* 91 (2001): 1185–1202.

Hamburger, M. J., and E. N. Platt. The expectations hypothesis and the efficiency of the Treasury bill market. *Review of Economics and Statistics* 57 (1975): 190–199.

Hansen, L. P. Consumption, asset markets and macroeconomic fluctuations: A comment. *Carnegie-Rochester Conference Series on Public Policy* 17 (1982): 239–250.

Hansen, L. P. Calculating asset prices in three example economies, in T. F. Bewley, ed., *Advances in Econometrics*, Fifth World Congress. Cambridge University Press, Cambridge, UK (1987).

Hansen, L. P., and R. J. Hodrick. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy* 88 (1980): 829–853.

Hansen, L. P., and T. J. Sargent. Exact linear rational expectations models: Specification and estimation, Federal Reserve Bank of Minneapolis Staff Report 71 (1981).

Hansen, L. P., T. J. Sargent, and T. Tallarini. Robust permanent income and pricing. *Review of Economic Studies* 66 (1998): 873–907.

Hansen, L. P., and K. J. Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50 (1982): 1269–1288.

Hansen, L. P., and K. J. Singleton. Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy* 91 (1983): 249–268.

Hansen, L. P., and K. J. Singleton. Errata. *Econometrica* 52 (1984): 267–268.

Hansen, L. P., and R. Jagannathan. Implications of security market data for models of dynamic economies. *Journal of Political Economy* 99 (1991): 225–262.

Hansen, L. P., and R. Jagannathan. Assessing specification errors in stochastic discount factor models. *Journal of Finance* 52 (1997): 557–590.

Hansen, L. P., J. C. Heaton, J. Lee, and N. Roussanov. Intertemporal substitution and risk aversion. Manuscript, University of Chicago, Forthcoming in J. Heckman, ed., *Handbook of Econometrics*. North-Holland, Amsterdam (2006).

Hansen, L. P., J. C. Heaton, and N. Li. Intangible risk? in C. Corrado, J. Haltiwanger, and D. Sichel, eds., *Measuring Capital in the New Economy*. University of Chicago Press, Chicago (2005): pages 111–152.

Hansen, L. P., J. C. Heaton, and N. Li. Consumption strikes back? Measuring long-run risk. Manuscript. University of Chicago (2006).

Heaton, J. C. The interaction between time-nonseparable preferences and time aggregation. *Econometrica* 61 (1993): 353–385.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 319 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Heaton, J. C. An empirical investigation of asset pricing with temporally dependent preference. specifications. *Econometrica* 63 (1995): 681–717.

Heaton, J. C., and D. J. Lucas. The effects of incomplete insurance markets and trading costs in a consumption-based asset pricing model. *Journal of Economic Dynamics and Control* 16 (1992): 601–620.

Heaton, J. C., and D. J. Lucas. The importance of investor heterogeneity and financial market imperfections for the behavior of asset prices. *Carnegie-Rochester Conference Series on Public Policy* 42 (1995): 1–32.

Heaton, J. C., and D. J. Lucas. Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy* 104 (1996): 443–487.

Heaton, J. C., and D. J. Lucas. Stock prices and fundamentals. *NBER Macroeconomics Annual* 1999, (2000a): 213–242.

Heaton, J. C., and D. J. Lucas. Portfolio choice and asset prices: The importance of entrepreneurial risk. *Journal of Finance* 55 (2000b): 1163–1198.

Hodrick, R. J. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies* 5 (1992): 357–386.

Jacobs, K., and K. Q. Wang. Idiosyncratic consumption risk and the cross-section of asset returns. *Journal of Finance* 59 (2004): 2211–2252.

Jagannathan, R., and Z. Wang. The conditional CAPM and the cross-section of expected returns. *Journal of Finance* 51 (1996): 3–53.

Jagannathan, R., and Y. Wang. Consumption risk and the cost of equity capital. *NBER Working Paper* 11026 (2005).

Jagannathan, R., K. Kubota, and H. Takehara. Relationship between labor-income risk and average return: Empirical evidence from the Japanese stock market. *Journal of Business* 71 (1998): 319–347.

Jegadeesh, N., and S. Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48 (1993): 65–91.

Jermann, U. Asset pricing in production economies. *Journal of Monetary Economics* 41 (1998): 257–275.

Jermann, U. The equity premium implied by production. Manuscript, University of Pennsylvania (2005).

Kandel, S., and R. F. Stambaugh. Expectations and volatility of consumption and asset returns. *Review of Financial Studies* 3 (1990): 207–232.

Kandel, S., and R. F. Stambaugh. Asset returns and intertemporal preferences. *Journal of Monetary Economics* 27 (1991): 39–71.

Kandel, S., and R. F. Stambaugh. Portfolio inefficiency and the cross-section of expected returns. *Journal of Finance* 50 (1995): 157–184.

Kocherlakota, N. R. Disentangling the coefficient of relative risk aversion from the elasticity of intertemporal substitution: An irrelevance result. *Journal of Finance* 45 (1990): 175–190.

Kocherlakota, N. R. The equity premium: It's still a puzzle. *Journal of Economic Literature* 34 (1996): 42–71.

Kogan, L. Asset prices and real investment. *Journal of Financial Economics* 73 (2004): 411–431.

Kreps, D. M., and E. L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46 (1978): 185–200.

Kuhn, T. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago (third edition, 1996) (1962).

Krusell, P., and A. A. Smith, Jr. Income and wealth heterogeneity, portfolio choice, and equilibrium asset returns. *Macroeconomic Dynamics* 1 (1997): 387–422.

Lakonishok, J., A. Shleifer, and R. W. Vishny. Contrarian investment, extrapolation, and risk. *Journal of Finance* 49 (1994): 1541–1578.

Lamont, O. A. Earnings and expected returns. *Journal of Finance* 53 (1998): 1563–1587.

Lamont, O. A. Investment plans and stock returns. *Journal of Finance* 55 (2000): 2719–2745.

LeRoy, S. F. Risk aversion and the martingale property of stock prices. *International Economic Review* 14 (1973): 436–446.

LeRoy, S. F., and R. D. Porter. The present-value relation: Tests based on implied variance bounds. *Econometrica* 49 (1981): 555–574.

Lettau, M. Idiosyncratic risk and volatility bounds, or, can models with idiosyncratic risk solve the equity premium puzzle? *Review of Economics and Statistics* 84 (2002): 376–380.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 320 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Lettau, M. Inspecting the mechanism: Closed-form solutions for asset prices in real business cycle models. *Economic Journal* 113 (2003): 550–575.

Lettau, M., and S. Ludvigson. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 56 (2001a): 815–849.

Lettau, M., and S. Ludvigson. Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109 (2001b): 1238–1287.

Lettau, M., and S. Ludvigson. Time-varying risk premia and the cost of capital: An alternative implication of the Q theory of investment. *Journal of Monetary Economics* 49 (2002): 31–66.

Lettau, M., and S. Ludvigson. Expected returns and expected dividend growth. *Journal of Financial Economics* (2004).

Lewellen, J., and S. Nagel. The conditional CAPM does not explain asset-pricing anomalies. Manuscript, MIT (2004).

Lewellen, J., S. Nagel, and J. Shanken. A skeptical appraisal of asset pricing tests. Manuscript, Dartmouth College, Stanford University, and Emory University (2006).

Li, Q., M. Vassalou, and Y. Xing. Investment growth rates and the cross-section of equity returns. Manuscript, Columbia University (2003).

Liew, J., and M. Vassalou. Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics* 57 (2000): 221–245.

Lucas, D. J. Asset pricing with undiversifiable risk and short sales constraints: Deepening the equity premium puzzle. *Journal of Monetary Economics* 34 (2001): 325–341.

Lucas, R. E., Jr. Asset prices in an exchange economy. *Econometrica* 46 (1978): 1429–1446.

Lustig, H., and S. Van Nieuwerburgh. Housing collateral, consumption insurance and risk premia: An empirical perspective. *Journal of Finance* (2004a).

Lustig, H., and S. Van Nieuwerburgh. A theory of housing collateral, consumption insurance and risk premia. Manuscript, UCLA and NYU (2004b).

Lustig, H., and A. Verdelhan. The cross-section of foreign currency risk premia and U. S. consumption growth risk. Manuscript, University of Chicago and UCLA (2004).

Macaulay, F. R. *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 (1938). Reprinted in Risk Classics Library, Risk Books (1999).

Malloy, C., T. Moskowitz, and A. Vissing-Jorgenson. Job risk and asset returns. Manuscript, University of Chicago (2005).

Mankiw, N. G. The equity premium and the concentration of aggregate shocks. *Journal of Financial Economics* 17 (1986): 211–219.

Mankiw, N. G., and S. Zeldes. The consumption of stockholders and non-stockholders. *Journal of Financial Economics* 29 (1991): 97–112.

McCloskey, D. N. The rhetoric of economics. *Journal of Economic Literature* 21 (1983): 481–517.

Mehra, R., and E. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics* 15 (1985): 145–161.

Menzly, L. Influential observations in cross-sectional asset pricing tests. Manuscript, University of Chicago (2001).

Menzly, L., T. Santos, and P. Veronesi. Understanding predictability. *Journal of Political Economy* 112 (2004): 1–47.

Merton, R. C. An intertemporal capital asset pricing model. *Econometrica* 41 (1973): 867–887.

Merz, M., and E. Yashiv. Labor and the market value of the firm. Manuscript, University of Bonn (2005).

Nelson, C. R., and M. J. Kim. Predictable stock returns: The role of small sample bias. *Journal of Finance* 48 (1993): 641–661.

Ogaki, M., and C. M. Reinhart. Measuring intertemporal substitution: The role of durable goods. *Journal of Political Economy* 106 (1998): 1078–1098.

Pakos, M. Asset pricing with durable goods and non-homothetic preferences. Manuscript, University of Chicago (2004).

Parker, J., and C. Julliard. Consumption risk and the cross-section of expected returns. *Journal of Political Economy* 113 (2005): 185–222.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 321 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Pastor, L., and P. Veronesi. Rational IPO waves. *Journal of Finance* (2004).

Petkova, R. Do the Fama–French factors proxy for innovations in predictive variables? *Journal of Finance* 61 (2006): 581–612.

Piazzesi, M. Bond yields and the Federal Reserve. *Journal of Political Economy* 113 (2005): 311–344.

Piazzesi, M., and M. Schneider. Equilibrium yield curves. Manuscript, University of Chicago and NYU, prepared for the *2006 Macroeconomics Annual* (2006).

Piazzesi, M., M. Schneider, and S. Tuzel. Housing, consumption, and asset pricing. Manuscript, University of Chicago, NYU, and UCLA (2004).

Poterba, J., and L. H. Summers. Mean reversion in stock returns: Evidence and implications. *Journal of Financial Economics* 22 (1988): 27–60.

Restoy, F., and P. Weil. Approximate equilibrium asset prices. *NBER Working Paper* 6611 (1998).

Roll, R. *The Behavior of Interest Rates*, Basic Books, New York (1970).

Roll, R. A critique of the asset pricing theory's tests part I: On past and potential testability of the theory. *Journal of Financial Economics* 4 (1977): 129–176.

Roll, R., and S. A. Ross. On the cross-sectional relation between expected returns and betas. *Journal of Finance* 49 (1994): 101–121.

Rozeff, M. S. Dividend yields are equity risk premiums. *Journal of Portfolio Management* 11 (1984): 68–75.

Santos, T., and P. Veronesi. Labor income and predictable stock returns. *Review of Fiancial Studies* (2005).

Sargent, T. J. Rational expectations and the term structure of interest rates. *Journal of Money Credit and Banking* 4 (1972): 74–97.

Sargent, T. J. A note on maximum likelihood estimation of the rational expectations model of the term structure. *Journal of Monetary Economics* 5 (1978): 133–143.

Schwert, G. W. Anomalies and market efficiency. Chapter 15 of G. Constantinides, M. Harris, and S. Stulz, eds., *Handbook of the Economics of Finance*. North-Holland, Amsterdam, (2003): 937–972.

Shiller, R. J. The volatility of long-term interest rates and expectations models of the term structure. *Journal of Political Economy* 87 (1979): 1190–1219.

Shiller, R. J. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71 (1981): 421–436.

Shiller, R. J. Consumption, asset markets, and economic fluctuations. *Carnegie-Rochester Conference on Public Policy* 17 (1982): 203–238.

Shiller, R. J. Stock prices and social dynamics. *Brookings Papers on Economic Activity* (1984): pages 457–510.

Shiller, R. J., J. Y. Campbell, and K. L. Schoenholz. Forward rates and future policy: Interpreting the term structure of interest rates. *Brookings Papers on Economic Activity* (1983): pages 173–217.

Stambaugh, R. F. The information in forward rates: Implications for models of the term structure. *Journal of Financial Economics* 21 (1988): 41–70.

Stambaugh, R. F. Predictive regressions. *Journal of Financial Economics* 54 (1999): 375–421.

Storesletten, K., C. I. Telmer, and A. Yaron. Asset pricing with idiosyncratic risk and overlapping generations. Manuscript. Carnegie Mellon University (2000).

Storesletten, K., C. I. Telmer, and A. Yaron. Cyclical dynamics of idiosyncratic labor market risk. *Journal of Political Economy* (2005).

Sundaresan, S. M. Intertemporally dependent preferences and the volatility of consumption and wealth. *Review of Financial Studies* 2 (1989): 73–88.

Tallarini, T. D., Jr. Risk-sensitive real business cycles. *Journal of Monetary Economics* 45 (2000): 507–532.

Telmer, C. I. Asset pricing puzzles and incomplete markets. *Journal of Finance* 48 (1993): 1803–1832.

Uhlig, H. Asset pricing with Epstein-Zin preferences. Manuscript, Humboldt University (2006).

Vassalou, M. News related to future GDP growth as a risk factor in equity returns. *Journal of Financial Economics* 68 (2003): 47–73.

Verdelhan, A. A habit-based explanation of the exchange rate risk premium. Manuscript, University of Chicago (2004).

Vissing-Jorgensen, A. Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy* 110 (2002): 825–853.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 322 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Wachter, J. A consumption-based model of the term structure of interest rates. Manuscript, University of Pennsylvania (2004).

Weil, P. The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics* 24 (1989): 401–421.

Yogo, M. A consumption-based explanation of expected stock returns. *Journal of Finance* (2006).

Zhang, L. Anomalies. Manuscript, University of Rochester (2004).

Zhang, L. The value premium. *Journal of Finance* 60 (2005): 67–104.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 323 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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

$$U_t = \left((1-\beta)c_t^{1-\rho} + \beta\left[E_t\left(U_{t+1}^{1-\gamma}\right)\right]^{\frac{1-\rho}{1-\gamma}}\right)^{\frac{1}{1-\rho}}, \tag{A1}$$

we have

$$\frac{\partial U_t}{\partial c_t} = U_t^{\rho}(1-\beta)c_t^{-\rho}. \tag{A2}$$

Then, the first order condition is

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

Substituting from (A2) and canceling gives

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

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

$$m_{t+1} = \beta\left(\frac{U_{t+1}}{\left[E_t\left(U_{t+1}^{1-\gamma}\right)\right]^{\frac{1}{1-\gamma}}}\right)^{\rho-\gamma}\left(\frac{c_{t+1}}{c_t}\right)^{-\rho}. \tag{A3}$$

Market return

The utility function (A1) is linearly homogeneous. Thus,

$$U_t = \sum_{j=0}^{\infty}\frac{\partial U_t}{\partial c_{t+j}}c_{t+j} = E_t\sum_{j=0}^{\infty}\frac{\partial U_t}{\partial c_{t+j}}c_{t+j}$$

$$\frac{U_t}{\partial U_t/\partial c_t} = E_t\sum_{j=0}^{\infty}m_{t,t+j}c_{t+j} = W_t \tag{A4}$$

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 324 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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 (A3).

We want an expression with the market *return*, not wealth itself, so we proceed as follows. Use the utility function (A1) to express the denominator of (A3) in terms of time t observables:

$$\left(E_t\left(U_{t+1}^{1-\gamma}\right)\right)^{\frac{1}{1-\gamma}} = \left(\frac{1}{\beta}\right)^{\frac{1}{1-\rho}} \left(U_t^{1-\rho} - (1-\beta)c_t^{1-\rho}\right)^{\frac{1}{1-\rho}}. \tag{A5}$$

Now, substitute for U_t and U_{t+1} from (A4), with (A2):

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

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

$$U_t = \left(W_t(1-\beta)c_t^{-\rho}\right)^{\frac{1}{1-\rho}}. \tag{A6}$$

Now, use (A5) and (A6) in (A3):

$$m_{t+1} = \beta\left(\frac{U_{t+1}}{\left[E_t\left(U_{t+1}^{1-\gamma}\right)\right]^{\frac{1}{1-\gamma}}}\right)^{\rho-\gamma}\left(\frac{c_{t+1}}{c_t}\right)^{-\rho}. \tag{A7}$$

Substituting into (A3) gives

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

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 325 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

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

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

so we have in the end

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

If we define

$$\theta = \frac{1-\gamma}{1-\rho}, \quad 1-\theta = \frac{\gamma-\rho}{1-\rho},$$

then we can express the result as a combination of the standard consumption-based discount factor and the inverse of the market return:

$$m_{t+1} = \left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\rho}\right]^{\theta} \left(\frac{1}{R^W_{t+1}}\right)^{1-\theta}.$$

Discount factor in the $\rho = 1$ case

From (A1), let $v = \ln U$ and let c now denote log consumption. Then we can write (A1) as

$$v_t = \frac{1}{1-\rho} \ln((1-\beta)e^{(1-\rho)c_t} + \beta e^{(1-\rho)Q_t}),$$
$$Q_t = \frac{1}{1-\gamma} \ln E_t(e^{(1-\gamma)v_{t+1}}).$$

In the limit $\rho = 1$ (differentiating numerator and denominator),

$$v_t(1) = (1-\beta)c_t + \beta Q_t(1),$$

where I use the notation $v_t(1), Q_t(1)$ to remind ourselves that v_t is a function of the preference parameter ρ, and results that only hold when $\rho = 1$.

Next, assuming consumption and hence $v_{t+1}(1)$ are log-normal and conditionally homoskedastic, we have

$$v_t(1) = (1-\beta)c_t + \beta\frac{1}{1-\gamma} \ln E_t(e^{(1-\gamma)v_{t+1}(1)})$$
$$= (1-\beta)c_t + \beta E_t[v_{t+1}(1)] + \frac{1}{2}\beta(1-\gamma)\sigma^2[v_{t+1}(1)],$$
$$v_t(1) = (1-\beta)\sum_{j=0}^{\infty} \beta^j E_t(c_{t+j}) + \frac{1}{2}\beta\frac{(1-\gamma)}{(1-\beta)}\sigma^2[v_{t+1}(1)].$$

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 326 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

The discount factor is, from (A7),

$$\ln m_{t+1} = \ln(\beta) - \rho\,(c_{t+1} - c_t) + (\rho - \gamma)\,(v_{t+1} - Q_t)$$
$$(E_{t+1} - E_t)\ln m_{t+1} = -\rho\,(E_{t+1} - E_t)\,c_{t+1} + (\rho - \gamma)(E_{t+1} - E_t)v_{t+1}$$

In the case $\rho = 1$, with normal and homoskedastic consumption, we then have

$$(E_{t+1} - E_t)\ln m_{t+1} = -(E_{t+1} - E_t)c_{t+1} + (1-\gamma)\,(1-\beta)(E_{t+1} - E_t)\sum_{j=0}^{\infty}\beta^j(c_{t+1+j}).$$

It's convenient to rewrite the discount factor in terms of consumption growth, as follows:

$$W = \sum_{j=0}^{\infty}\beta^j c_{t+1+j} = (c_{t+1} - c_t) + \beta(c_{t+2} - c_{t+1}) + \beta^2(c_{t+3} - c_{t+2}) + \cdots + c_t + \beta c_{t+1} + \beta^2 c_{t+2} + \cdots,$$
$$W = \sum_{j=0}^{\infty}\beta^j \Delta c_{t+1+j} + c_t + \beta W,$$
$$W = \frac{1}{1-\beta}\sum_{j=0}^{\infty}\beta^j \Delta c_{t+1+j} + \frac{1}{1-\beta}c_t.$$

Then, since $(E_{t+1} - E_t)\,c_t = 0$,

$$(E_{t+1} - E_t)\ln m_{t+1} = -(E_{t+1} - E_t)\,\Delta c_{t+1} + (1-\gamma)(E_{t+1} - E_t)\sum_{j=0}^{\infty}\beta^j \Delta c_{t+1+j}$$

or

$$(E_{t+1} - E_t)\ln m_{t+1} = -\gamma\,(E_{t+1} - E_t)(\Delta c_{t+1}) + (1-\gamma)\,(E_{t+1} - E_t)\left[\sum_{j=1}^{\infty}\beta^j\left(\Delta c_{t+1+j}\right)\right].$$

Here we see the familiar consumption growth raised to the power γ, plus a newterm-reflecting innovations in long-run consumption growth.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 327 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

Financial Markets and the Real Economy: A Comment[1]

Lars Peter Hansen
Dept of Economics, University of Chicago

John Cochrane has done an admirable job of summarizing a rather extensive empirical literature. The work is so exhaustive that I will not even attempt to comment on it in a systematic way. There are many very nice aspects to his discussion, and what follows merely provides some minor amendments.

To his credit, Cochrane considers again some of the early literature on consumption-based asset pricing and compares quotes across papers in an attempt at intellectual history. This is interesting reading, but I would urge others to read the whole papers, not just quotes. Some important breakthroughs occurred prior to Mehra and Prescott (1985) and Hansen and Singleton (1983). While the Shiller (1982) paper that Cochrane features is a nice paper, I am personally a big fan of Grossman and Shiller (1980, 1981). These two joint papers really got researchers like Singleton and myself and others thinking of empirical implications of the consumption-based capital asset pricing model along with the earlier theoretical work of Rubinstein (1976), Lucas (1978), and Breeden (1979). It is unfortunate that only an abbreviated version, Grossman and Shiller (1981), was published because Grossman and Shiller (1980) was familiar to many people at the time. In this sense, the analogy to Columbus versus Erikson in the discovery of America is a bit misleading, although the important influence of Mehra and Prescott (1985) in subsequent research is undeniable. Given my Nordic origins, I have always been a bit partial to Erikson.

It is interesting that the Shiller inequality that Cochrane refers to differs from the ones Cochrane uses to frame most of his discussion. Shiller deduced his inequality using information about the marginal distribution for consumption or more generally a stochastic discount factor along with the marginal distributions for separate returns. Hansen and Singleton (1982) used information from the joint distribution of stochastic discount factors and returns following in part Grossman and Shiller (1980). Hansen

[1]I thank John Heaton and Ken Singleton for helpful comments. This material is based upon work supported by the National Science Foundation under Award Number SES0519372.

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 328 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

and Jagannathan (1991) and the Hansen comment on Shiller used marginal information on stochastic discount factors in conjunction with information on the joint distribution of returns. This latter approach was motivated by an aim to produce a common set of diagnostics for a rich family of stochastic discount factor models. All three approaches are interesting and arguably serve different purposes. It is certainly true that the Shiller paper was a natural precursor to my work with Jagannathan.

In Cochrane's discussion of the Hansen and Singleton (1983) paper in his section of the equity-premium puzzle, it is not clear why Ken and I are even mentioned as part of the same discovery game. We focused on monthly postwar data and used a sample with a shorter span (but observed more frequently) for estimation and inference in contrast to both Grossman and Shiller (1980, 1981) and Mehra and Prescott (1985). For the postwar data sample we used, the mean returns could not be estimated with enough accuracy for reliable inference. A narrowly framed equity premium puzzle based on postwar data would have been much less dramatic and much easier to debunk. Perhaps we erred in focusing on such a short time period, but this choice is non-trival and has important consequences. It revolves in part around the following question: Did postwar investors presume the prewar volatility was germane when making investments?

The whole point of Hansen and Singleton (1983) is to show that by exploiting conditioning information one can make non-trivial inferences with postwar data. Unfortunately, this led us to a related problem. While conditioning information could be helpful in identifying the intertemporal elasticity of substitution from asset market data and consumption, actual use of this information put us in a bind. You cannot simultaneously explain the conditional distribution of consumption as well as multiple returns. This bind was reflected in the conclusion of our paper, but certainly our prose did not match the elegance of Mehra and Prescott (1985).

While our paper in the *Journal of Political Economy* exploited log-normality, our companion paper Hansen and Singleton (1982) (and awkwardly the errata in Hansen and Singleton (1984)) published in *Econometrica* found comparable results with multiple returns and conditioning information constructed as scaling factors using an estimation method that avoided log-normality. It is evident from our work that the heterogeneity in the risk exposure of returns including those we constructed through scaling posed a serious challenge to the power utility, representative consumer model. On the other hand, we were not as clever as Mehra and Prescott in describing and framing this as a puzzle. In contrast to Mehra and Prescott, statistical inference was front and center in our analysis and formally shaped how we looked at evidence, but this is only part of the difference in approaches.

Although vast in its coverage, there is a missing link in Cochrane's essay that is worth further consideration. Cochrane has separate discussions of the Fama and French (1995) empirical evidence based on portfolio constructed using ratios of book equity to market equity and Hall (2001)'s analysis of intangible capital. While I share Cochrane's interest in Hall's work, in Hansen et al. (2005), we view the Fama–French work as suggesting possibly important differences in the risk exposure of technologies that feature different mixes of tangible and intangible capital. If intangible capital is a primary source of

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 329 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

divergence in measures of book equity and market equity, then the Fama and French (1995) analysis suggests that the macroeconomic risk exposure of intangible capital may be fundamentally different from that of measured capital. This has potentially important modeling implications that are worth exploring further.

Restoy and Weil (1998), Hansen et al. (1998), Tallarini (2000), and others feature the use of continuation values computed from consumption dynamics in conjunction with recursive utility. While Restoy and Weil (1998) focus on the role of consumption, they exploit its link to wealth and the return on the wealth portfolio. The link between continuation values and wealth becomes degenerate when the intertemporal elasticity of substitution is unity. This leads Restoy and Weil (1998) to exclude this case. Even with a unitary elasticity of substitution, however, continuation values still can be inferred from consumption dynamics by solving the utility recursion exactly or at least approximately. In fact, a unitary elasticity of substitution simplifies the calculation, as is evident from Cochrane's discussion.

By working with continuation values, Hansen et al. (2006) show that an approximation around $\rho = 1$, where ρ is the reciprocal of the intertemporal elasticity of substitution, is straightforward to compute for some alternative models of consumption dynamics. From Eq. (14) in Cochrane's essay, the logarithm of the marginal utility of consumption is

$$\log m_{t+1} = -\rho(\log c_{t+1} - \log c_t) + (\rho - \gamma)\log U_{t+1} + \pi_t,$$

where π_t is in the date t information set and U_{t+1} is the continuation value for consumption at date $t + 1$. The term π_t is inconsequential when characterizing the innovation to the logarithm of the marginal rate of substitution. Differentiating $\log m_{t+1} - \pi_t$ with respect to ρ gives

$$-\log c_{t+1} + \log c_t + \log U_{t+1} + (\rho - \gamma)\frac{d\log U_{t+1}}{d\rho}.$$

To localize around unity, we evaluate both $\log m_{t+1} - \pi_t$ and its derivative at $\rho = 1$, scale the latter by $\rho - 1$, and add the terms:

$$\begin{aligned}\log m_{t+1} - \pi_t \approx & -(\log c_{t+1} - \log c_t) + (1 - \gamma)\log U_{t+1}|_{\rho=1} \\ & + (\rho - 1)\left[-\log c_{t+1} + \log c_t + \log U_{t+1}|_{\rho=1} + (\rho - \gamma)\frac{d\log U_{t+1}}{d\rho}|_{\rho=1}\right].\end{aligned}$$

Hansen et al. (2006) compute continuation values and derivatives for log-normal consumption dynamics and for consumption dynamics that include some forms of stochastic volatility. These are analogous approximation formulas that characterize how asset values and local risk prices change as a function of the intertemporal substitution elasticity of investors.

My final thought is a reflection about how explorations into alternative preferences have been or will be useful in macroeconomic analyses. Recently, Backus et al. (2004) wrote a useful summary on so-called exotic preferences and why they might or should

Els UK ch07-n50899.pdf 2007/9/5 5:33 pm Page: 330 Trim: 7.5in × 9.25in Floats: Top/Bot TS: diacriTech, India

be of interest to macroeconomists. In asset pricing, are exotic preferences merely a device to account for asset pricing facts, or do we aim for this evidence to be formally integrated into, say, macroeconomics models to be used in policy analysis? Similarly, what role will the asset pricing-based models with market imperfections have to play in constructing heterogenous agent models for use in addressing macroeconomic policy questions? It will be interesting to see how this empirically ambitious literature summarized by Cochrane will influence the construction of dynamic general equilibrium models. Will there be an analogous ambition that will pervade dynamic economic modelling more generally, or will asset pricing evidence be viewed in isolation? The jury is still out on such questions.

References

Backus, D. K., B. R. Routledge, and S. E. Zin. Exotic preferences for macroeconomics. In M. Gertler, and K. Rogoff, eds., *NBER Macroeconomics Annual* (2004).

Breeden, D. An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics* 7 (1979): 265–296.

Fama, E. F., and K. R. French. Size and book-to-market factors in earnings and returns. *Journal of Finance* 50 (1995): 131–155.

Grossman, S. J., and R. J. Shiller. Preliminary results on the determinants of the variability of stock market prices. University of Pennslyvania Press Philadelphia (1980).

Grossman, S. J., and R. J. Shiller. The determinants of the variability of stock market prices. *American Economic Review Papers and Proceedings* 71 (1981): 222–227.

Hall, R. E. The stock market and capital accumulation. *American Economic Review* 96 (2001): 222–227.

Hansen, L. P., and R. Jagannathan. Implications of security market data for models of dynamic economies. *Journal of Political Economy* 99 (1991): 225–262.

Hansen, L. P., and K. J. Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50 (1982): 1269–1288.

Hansen, L. P., and K. J. Singleton. Stochastic consumption, risk aversion, and the intertemporal behavior of asset returns. *Journal of Political Economy* 91 (1983): 249–268.

Hansen, L. P., and K. J. Singleton. Errata. *Econometrica* 52 (1984): 267–268.

Hansen, L. P., T. J. Sargent, and T. D. Tallarini. Robust permanent income and pricing. *Review of Economic Studies* 66 (1998): 873–907.

Hansen, L. P., J. Heaton, and N. Li. Intangible risk? in C. Corrado, J. Haltiwanger, and D. Sichel, eds., *Measuring Capital in the New Economy*. University of Chicago Press, Chicago (2005), pages 111–152.

Hansen, L. P., J. C. Heaton, J. Lee, and N. Rousanov. Intertemporal substitution and risk aversion. Forthcoming in the *Handbook of Econometrics* (2006).

Lucas, R. E. Asset prices in an exchange economy. *Econometrica* 46 (1978): 1429–1445.

Mehra, R. and E. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics* 15 (1985): 145–161.

Restoy, F., and P. Weil. Approximate equilibrium asset prices. NBER Working paper 6611 (1998).

Rubinstein, M. The valuation of uncertain income streams and the pricing of options. *Bell Journal of Economics* 7 (1976): 407–425.

Shiller, R. J. Consumption, asset prices and economic fluctuations. *Carnegie-Rochester Conference on Public Policy* 17 (1982): 203–238.

Tallarini, T. D. Risk-sensitive real business cycles. *Journal of Monetary Economics* 45 (2000): 507–532.

Asset Pricing under Quantile Utility Maximization*

Bruno C. Giovannetti†

July 26, 2011

"Focus on the downside, and the upside will take care of itself" is a famous quote among professional investors. By considering an agent who follows this advice, we reproduce the first and second moments of stock returns, risk-free rate and consumption growth. The agent's behavior towards risk is analogous to a relative risk aversion of about 3 under expected utility, the elasticity of intertemporal substitution is about 0.5 and the time discount factor is below 1. In particular, the proposed model separates time and risk preferences in an innovative way.

*I would like to thank Dennis Kristensen for invaluable support on this work. I also thank Andrew Ang, Pierre-André Chiappori and Marcelo Moreira for constant and crucial advice, and Ricardo Brito, John Donaldson, Guilherme Martins, Marcos Nakaguma, Walker Hanlon, Ricardo Reis, Bernard Salanié, Samer Shousha and the participants at the Columbia Econometrics Colloquium, Applied Micro Colloquium and Finance PhD Student Seminar, and the Economics Seminars at Fundação Getúlio Vargas (EESP), Insper and Universidade de São Paulo for important comments and discussions. All errors left are only mine.

†Economics Department, University of Sao Paulo. E-mail: bgiovannetti@fipe.org.br, bruno.giovannetti@gmail.com

A famous quote among professional investors is "Focus on the downside, and the upside will take care of itself".[1] In this paper, we consider a consumer-investor who follows this advice. Surprisingly, the consumption-based asset pricing model that emerges from this idea explains the main existing puzzles found within the asset pricing literature. These include the equity premium and the risk-free rate puzzles, the countercyclicality of the equity premium and the procyclicality of the risk-free rate.

In the proposed model, the consumer-investor is concerned with the so-called downside risk. This is done by replacing the standard setting of expected utility optimizing agents with the concept of quantile utility. Under this framework, the agent summarizes a risky situation using a worst-case scenario which is a function of his downside risk aversion. The more downside risk averse the agent, the worse the worst-case scenario he considers. The τ quantile of a continuous random variable can be interpreted as the worst possible outcome that can occur with probability $1 - \tau$. Hence, instead of maximizing the expected value of his utility function, the agent maximizes a given τ quantile of it. As we will see, τ defines his downside risk aversion: the lower τ, the higher the downside risk aversion.[2]

This is a novel extension of the static decision-theoretical framework developed by Manski (1988) and Rostek (2010) for a dynamic asset pricing setting. In a standard economy with one risky and one risk-free asset, we can derive an arbitrage-free asset pricing model,

1. A search of this sentece on the internet returns many results.

2. One could say that the agent's objective function is given by the value at risk (VaR) of his utility. However, since τ here is a free parameter defining preference towards risk, it is not restricted to being close to zero (as in standard VaR applications).

where both main characteristics of the canonical expected utility consumption-based approach (Hansen and Singleton (1982), Mehra and Prescott (1985), hereinafter, the canonical model) are modified. The equity premium is no longer based on the covariance between the risky return and the consumption growth. Instead, it is a linear function of the risky return standard deviation. In addition, risk aversion and elasticity of intertemporal substitution (EIS), which are linked throughout a single parameter in the canonical model, are automatically disentangled in a simple way.

These two endogenous changes are the main drivers of the good empirical results. Since stock returns historically have a high standard deviation, the price of such a risk, i.e., the level of downside risk aversion, will not have to be high to match the empirical excess returns. Moreover, the attitude towards intertemporal substitution is not polluted by risk preferences.

To reproduce (i) the first and second moments of the risk-free return, the equity premium, and the consumption growth, (ii) the low covariance between risky return and consumption growth, (iii) the countercyclical risk premium, and (iv) the procyclical risk-free rate that we see in data, our model requires only three parameters related to preferences: a downside risk aversion (τ) of about 0.43, an EIS (ψ) of about 0.5 and a time discount factor (β) of less than 1. A downside risk aversion of such a magnitude is reasonable in that it produces reasonable certainty equivalents for bets on continuously distributed random variables (stock indexes, for example). By comparing certainty equivalents under quantile and expected utility maximization, an agent with this level of downside risk aversion is analogous to an

expected utility agent with a relative risk aversion coefficient of 3. According to Mehra and Prescott (1985) reasonable values for such a parameter would be between 1 and 10. An EIS of about 0.5 is also an acceptable value. In a recent work using microdata, Engelhardt and Humar (2009) estimate the EIS to be 0.74, with a 95% confidence interval that ranges from 0.37 to 1.21. Using macrodata and separating stockholders from nonstockholders, Vissing-Jorgensen (2002) estimates the EIS around 0.4 and 0.9 for these respective groups.

To illustrate the main differences between the predictions of our framework and the predictions of the canonical model, we first derive equations in closed-form for the risky return, the risk-free rate, and the equity premium. These equations come from combining the Euler equations of the quantile agent with the standard assumption of joint lognormality of returns and consumption growth. In order to replicate the well-evidenced existence of predictability in future excess returns, we then allow for time-varying economic uncertainty in the aggregate economy dynamics. From this, a countercyclical risk premium and a procyclical risk-free rate are produced.

Taking the model to data, we first perform simulation exercises, matching the first and second moments of consumption growth, risk-free rate and excess returns. Then, to evaluate the model free of distributional assumptions, we propose a GMM-based estimation method for its parameters.

The derived Euler equations impose restrictions on the functional forms of the conditional τ quantiles of consumption growth and excess return. They are well-defined functions of the

period-by-period risk-free rate and of the other parameters related to preferences. However, as τ in this framework is not given (it is the downside risk aversion to be estimated), the standard asymptotic results for quantile regressions as a GMM problem do not apply. Hence, we derive sufficient conditions for the parameters to be globally identified and for the proposed estimator to be consistent.

The fact that the model separates risk and time preferences allows us to estimate the EIS. This is a useful result of this paper. Under the standard technology for disentangling EIS and risk aversion (Epstein and Zin's (1989) preferences), one has to use instrumental variables to estimate the EIS. This is what Hall (1988) and Campbell (2003) do for example. Such estimations were recently found to suffer from weak-instruments related issues[3] and therefore are not reliable (see Neely, Roy, and Whiteman (2001) and Yogo (2004), for instance). However, the EIS estimation under our model does not require the use of any instrument.

We conclude the introduction by positioning this study in the related literature. The research in asset pricing can be separated according to the modifications proposed with respect to the canonical model. Such modifications are about (i) preferences, (ii) market and asset structure, and (iii) the endowment process. Group (i) could be further divided into two branches: (i.i) preferences inside and (i.ii) preferences outside the expected utility framework. Kocherlakota (1996), Cochrane (1997, 2006), Campbell (1999, 2003), and Donaldson and Mehra (2008a, 2008b) provide good surveys of this literature.

3. To estimate the EIS under Epstein and Zin's preferences one has to use instruments for consumption growth or returns. Since both of these variables are only weakly predictable, the instruments are weak.

The current study belongs to branch (i.ii), which was initiated by Epstein and Zin (1989) and Weil (1989). These authors use the recursive preferences of Kreps and Porteus (1978) as a way of separating time and risk preferences, something that is not possible under the canonical model. By disentangling risk aversion and EIS, they end up with a three-parameter model which is able to generate a reasonable level for the risk-free rate. However, since no innovation in the risk dimension is made, a high level of risk aversion is still necessary to fit the equity premium.

Epstein and Zin (1990, 2001) and Bekaert, Hodrick and Marshall (1997) investigate the use of Gul's (1991) disappointment aversion preferences to explain the equity premium puzzle.[4] According to these preferences, outcomes below the certainty equivalent are overweighted relative to outcomes above it. Although such preferences are a one-parameter extension of the expected utility framework, these papers extend the canonical model in two parameters, since they also use the model of Epstein and Zin (1989) to disentangle risk aversion and EIS. However, they are able to fit the equity premium with only a slightly lower, still unreasonable, risk aversion level.[5]

Going further, Routledge and Zin (2010) extend the disappointment aversion model in one additional dimension. They generalize Gul's preferences by defining an outcome as

4. Single-period portfolio allocation is studied under disappointment aversion by Ang, Bekart and Liu (2005). Basset, Koenker and Kordas (2004) also study sinlge-period allocation using preferences that accentuate the likelihood of the least favorable outcomes.

5. Bonomo and Garcia (1993) show that it is crucial to combine Gul's preferences with a joint process for consumption and dividends that follows a Markov switching model in order to match the first and second moments of risk-free and excess returns under reasonable parameter values. However, a model such as that would be in both groups (i.ii) and (iii) defined above.

disappointing only when it is sufficiently far (defined by the new parameter) from the certainty equivalent. Since their model also separates risk aversion and EIS using Epstein and Zin (1989) preferences, they are a three-parameter extension of the expected utility model, resulting in a total of five preference-related parameters. Under this richer structure, the disappointment aversion-based framework is finally able to address the financial puzzles successfully.

An alternative way of considering the fact that people care asymmetrically about good and bad outcomes is provided by the prospect theory of Kahneman and Tversky (1979). Applying prospect theory to asset pricing, Barberis, Huang and Santos (2001) are also able to reproduce the financial data patterns under reasonable parameter values.[6] In their model, the representative agent derives direct utility not only from consumption, but also from changes in the value of his financial wealth. Moreover, he is more sensitive to negative movements in his financial wealth than to positive movements. Besides that, such a sensitivity also is a function of the agent's past portfolio experience: if he had losses in the past relative to a time-varying benchmark, he now is more sensitive to further losses. A functional form reflecting this mechanism is imposed by the researchers.

Barberis, Huang and Santos's (2001) model also employs a large number of preference-related parameters; six, to be exact. The first two are the time discount factor and the relative risk aversion related to consumption. The third is the agent's extra sensitivity

6. Benartzi and Thaler (1995) investigate single-period portfolio allocation under prospect theory.

to losses in his portfolio wealth. The fourth defines how previous losses impact the third parameter. The fifth determines how the benchmark used by the agent to define gains and losses evolves over time. The sixth controls the overall importance of utility from gains and losses in financial wealth relative to utility from consumption.

The good empirical results from Barberis, Huang and Santos (2001) and Routledge and Zin (2010) indicate that consideration of asymmetric preferences over good and bad outcomes is a promising path for theories on choices and, in particular, for a well-accepted resolution of the asset pricing puzzles. Nevertheless, the large number of preference-related parameters in these models, which is crucial for their success, is a delicate issue. First, it is not easy to translate the models into a comprehensive view of the whole process. Second, it is hard to assign precisely the corresponding importance of each parameter to the obtained results. Finally, and perhaps most problematic, matching data by augmenting the parametric dimension is subject to the standard over-fitting critique. According to this critique, the larger number of parameters may simply describe better the noise in the data, rather than the underlying economic relationships. In other words, these models could be providing spurious data-fitting.[7]

The present paper addresses these issues. The quantile utility criterion comes from a

7. This tense relationship between the augmentation of the expected utility framework with additional parameters and the over-fitting critique is raised, for instance, by Zin (2002). Based on that article, Watcher (2002) claims that "behavioral models leave room for multiple degrees of freedom in the utility function. Taken to an extreme, this approach could reduce structural modeling to a tautological, data-fitting exercise" and "I believe that parsimony lies at the root of what Zin refers to as reasonableness. A parsimonious model is a model in which the number of phenomena to be explained is much greater than the number of free parameters."

loss-function that asymmetrically weighs good and bad outcomes, the well-known check loss-function. Hence, the derived model under this framework belongs to the class of models related to asymmetric preferences. Moreover, the model is quite parsimonious, requiring only three preference-related parameters: the time discount factor; the EIS; and the downside risk aversion. Finally, it solves the main asset pricing puzzles addressed by Barberis, Huang and Santos (2001) and Routledge and Zin (2010).

The rest of this work is organized as follows. Section I presents the quantile utility agent in its general form and derives some basic results of asset pricing under quantile maximization. Section II solves the model under lognormality and simulates from it. Section III discusses how to estimate the model free of distributional assumptions and presents the results. Section IV concludes.

I. Quantile Utility Maximization and Asset Pricing

In this section, we first present the elements of the quantile utility model, following Manski (1988) and Rostek (2010). Then, we apply this theoretical-decision framework to asset pricing.

A Quantile utility maximization elements

A general choice theory for quantile maximizing agents was developed recently. Rostek (2010) is the first study to axiomatize the quantile utility agent. Notwithstanding, the quantile maximization model for decision making under uncertainty was first proposed 23 years ago by Manski (1988).

The main idea is simple. An agent, when facing a situation where he has to choose among uncertain alternatives, picks the one that maximizes some given quantile of the utility distribution instead of its mean, as in the expected utility model. In this framework, the agent cares about the worst outcome that can happen with a given probability. For instance, the given quantile can be the median of the utility distribution, or the 0.25 quantile. In the case of the 0.25 quantile for example, when evaluating an uncertain situation, he looks at the worst outcome that can occur with 75 percent probability (i.e., the chance of the realized scenario being better than the scenario he considers is 75 percent).

The quantile of concern is an intuitive measure of pessimism. If agent A looks at the worst that may happen in 90 percent of the situations, i.e., quantile 0.10, and agent B looks at the worst that may happen in 60 percent of the situations, i.e., quantile 0.40, we would naturally classify agent B as more optimistic than agent A : agent A picks a more conservative scenario to summarize the lottery. Figure 1 illustrates this for a lottery that follows a normal distribution.

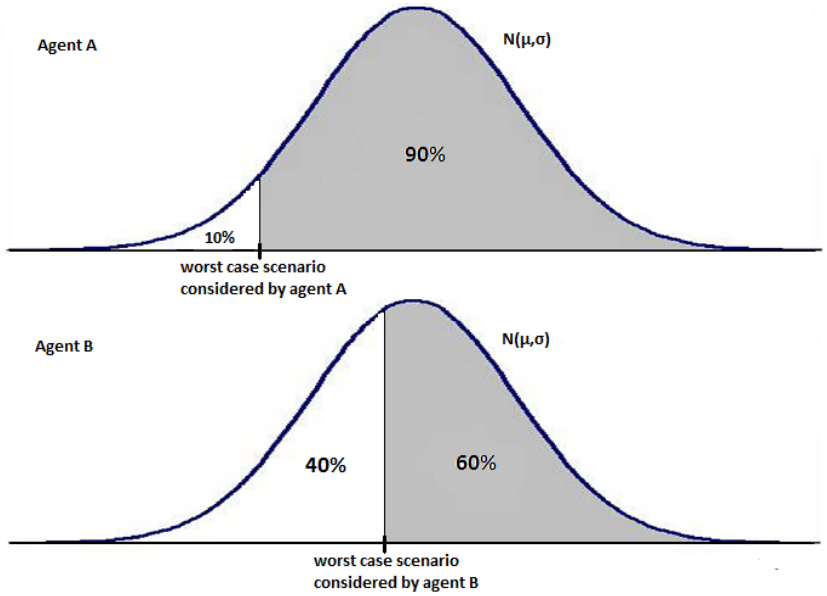

Figure 1. The quantile utility agent's reasoning.

As we shall see below, the quantile of concern defines also the agent's downside risk preference. Hence, downside risk preference is closely related to our standard notion of optimism-pessimism.

(i) Asymmetric preference

Because of the characteristics of his loss-function, we can say that the quantile agent cares asymmetrically about good and bad outcomes. This intuition comes from Manski (1988), based on the work of Wald (1937).

Assume that an agent has to evaluate an uncertain situation where U is his utility level which can have different values in different states of the world. This uncertain situation is represented by the cumulative distribution function of U, denoted by F_U. According to

the standard framework in decision theory introduced by Wald (1937), this agent should summarize (evaluate) F_U using the criterion ω^* that minimizes the expected value of his loss-function, i.e., his risk-function.

A possible loss-function could be the square loss. In this case, he would summarize F_U using

$$\begin{aligned} \omega^* &= \arg\min_{\omega \in \mathbb{R}^1} \int_{\mathbb{R}} (z-\omega)^2 \, dF_U(z) \\ &= \int_{\mathbb{R}} z \, dF_U(z) . \end{aligned}$$

Hence, he would use the expected utility criterion of von Neumann and Morgenstern (1944) and Savage (1954). This allows us to interpret the expected utility agent as someone who is evenly worried with underpredictions and overpredictions of his utility level in a risky situation and uses squares (L^2 norm) to compute the distances between the utility level predictions and realizations.

What if the decision maker was asymmetrically worried about under and overpredictions of his future utility level? We could describe a situation like that by the check loss-function of Koenker and Basset (1978). In this case, he would evaluate F_U using

$$\begin{aligned}\omega^* &= \arg\min_{\omega\in\mathbb{R}^1}\int_{\mathbb{R}}(1-\tau)\,|z-\omega|\times 1\,[z<\omega]+\tau\,|z-\omega|\times 1\,[z\geq\omega]\,dF_U(z)\\ &= Q^\tau(U),\end{aligned}$$

where $Q^\tau(U)$ is the τ_{th} quantile of the random variable U (if F_U is continuous, $Q^\tau(U) = F_U^{-1}(\tau)$).

Therefore, a quantile maximizer can be described as someone who asymmetrically weighs underpredictions and overpredictions of his future utility level, in the ratio $(1-\tau)/\tau$, and uses absolute values (L_1 norm) to compute the distances.[8] In this case, the agent's evaluation criterion is the τ_{th} quantile of his utility, that is, the worst possible utility level that may happen with probability $(1-\tau)$. This is the optimal criterion to summarize F_U given his asymmetric concern with the upper tails of utility distributions relative to their lower tails.

(ii) Quantile agent definition

We now define the quantile agent in a more formal way. Let $\mathcal{S}$ be a set of states of the world $s\in\mathcal{S}$, and $\mathcal{X}$ be an arbitrary set of payoffs $x, y\in\mathcal{X}$. Then, the agent has to choose among simple acts $h:\mathcal{S}\to\mathcal{X}$, which map from states to payoffs. Let $\mathcal{A}$ be the set of all such acts, and $E = 2^{\mathcal{S}}$ be the set of all events. Define π to be a probability measure on

8. Such an agent could also compute distances under the L^2 norm. In this case, his criterion to evaluate F_U would be the expectiles of Newey and Powell (1987)

$$\omega^*(\tau) = E(U) + \left(\frac{2\tau-1}{1-\tau}\right)E\left[(U-\omega^*(\tau))\times 1\left[U<\omega^*(\tau)\right]\right].$$

E, and u a utility function over payoffs $u : \mathcal{X} \to \mathbb{R}$. For each act, π induces a probability distribution over payoffs, referred to as a lottery. Given that, let G, H denote the random variables (payoffs) induced by the acts $g, h \in \mathcal{A}$, respectively. Finally, define F_G and F_H as the lotteries induced by the acts g and h, i.e., the cumulative distribution functions of G and H, respectively.

A decision maker is defined as a τ-quantile maximizer if there exists a unique $\tau \in [0, 1]$, a probability measure π on E, and a utility function u, such that for all $g, h \in \mathcal{A}$,

$$g \succ h \Leftrightarrow Q^{\tau}(u(G)) > Q^{\tau}(u(H)).$$

As always, we can think in terms of the lotteries:

$$F_G \succeq F_H \Leftrightarrow Q^{\tau}(u(G)) \geq Q^{\tau}(u(H)).$$

(iii) Downside risk aversion

For the standard expected utility agent, we may understand risk preferences using the following logic.

First we define riskiness. We say that the lottery F_H is riskier than the lottery F_G if F_G second-order stochastic dominates[9] (SSD) F_H (see Rothschild and Stiglitz (1970)). Then, we define Υ to be the class of all pairs of lotteries that SSD one another, i.e., $\Upsilon = \{(F_G, F_H) : F_G$

9. F_G SSD F_H if and only if

SSD $F_H\}$. It is natural to classify agent A as more risk averse than agent B if for all pairs of distributions in Υ, whenever B prefers a distribution which SSD the other, so does A. Finally, we show that this will be the case if and only if the utility function of agent A is "more concave" than the utility function of agent B, i.e., $u_A(x) = h(u_B(x))$, where $h(\cdot)$ is an increasing concave function. Given that, we conclude that risk-aversion is described by the concavity of the utility function.

Manski (1988) and Rostek (2010) follow the same logic to attach the quantile maximizer's attitude toward risk to the quantile he maximizes. The central point is that riskiness is characterized in a different way, the so-called downside risk: F_H involves more downside risk than F_G if F_G crosses F_H from below. We say that lottery F_G crosses lottery F_H from below if there exists $x, y \in \mathcal{X}$, such that $F_G(y) \leq F_H(y)$ for all $y < x$ and $F_G(y) \geq F_H(y)$ for all $y > x$. That is, downside risk is related to the probability of bad outcomes.[10]

Just as above, considering the class of all pairs of lotteries with the single-crossing property, $\Phi = \{(F_G, F_H) : F_G \text{ crosses } F_H \text{ from below}\}$, we say that individual A is more downside risk averse than individual B if, for all pairs of distributions in Φ, whenever B prefers a distribution which crosses the other from below, so does A. Given that, we can show that agent A is more downside risk averse than agent B if and only if $\tau_A < \tau_B$, and then τ can

$$\int_{-\infty}^{x} [F_H(t) - F_G(t)]\, dt \geq 0, \text{ for any } x \in \mathcal{X}.$$

10. If F_G and F_H have the same mean, and F_H has more downside risk than F_G, then F_H has also more (second-order stochastic dominance) risk than F_G. However, under different means, this is not true.

be defined as the downside risk aversion parameter in the decision model: the lower τ, the more downside risk averse the agent.

But what role does the concavity of the utility function play under this framework? Because of the property of equivariance of quantiles to monotonic transformations, the answer to this question is "none", at least for static decision problems.

(iv) Equivariance of quantiles to monotonic transformations and its implications

A key aspect of the quantile utility model is that static decisions are invariant to any strictly increasing transformation of the utility function. This is described in Proposition 1 in Manski (1988).

If $m : \mathbb{R} \to \mathbb{R}$ is a strictly increasing function, and X is a random variable, then[11]

$$Q^{\tau}\left(m\left(X\right)\right) = m\left(Q^{\tau}\left(X\right)\right). \tag{1}$$

Hence, for lotteries F_G and F_H,

$$\begin{aligned} F_G \succeq F_H \quad &\Leftrightarrow \quad Q^{\tau}\left(u\left(G\right)\right) \geq Q^{\tau}\left(u\left(H\right)\right) \\ &\Leftrightarrow \quad u^{-1}\left(Q^{\tau}\left(u\left(G\right)\right)\right) \geq u^{-1}\left(Q^{\tau}\left(u\left(H\right)\right)\right) \\ &\Leftrightarrow \quad Q^{\tau}\left(G\right) \geq Q^{\tau}\left(H\right), \end{aligned}$$

11. The intution under this result is that a strictly increasing transformation of the random variables doesn't change the order of the values of their support.

where the second line follows from the fact that u is a strictly increasing function.

Therefore, for static problems, the agent's decision does not depend on u. Manski (1988) and Rostek (2010) refer to this as a robustness property: the choice is unaffected by misspecification of the utility function.

However, the utility function is relevant in intertemporal choices. When the utility function has more than one argument, it is not possible to use the equivariance property to get rid of u. In particular, under time-separability, the concavity of the utility function defines the preference towards intertemporal substitution as usual. This is going to play an important role in the asset pricing theory, allowing the downside risk aversion and the EIS to be disentangled. This idea is not in Manski (1988) or in Rostek (2010) and, to the best of our knowledge, is explored for the first time in the present study.

B Asset pricing

We now apply the quantile maximization decision theory to the standard intertemporal problem of a consumer-investor agent. First, we define the consumption-investment problem and solve for the Euler equations that the agent must respect in equilibrium. Then we discuss the Law of One Price and the no-arbitrage condition under this framework.

The model to be considered has two periods. This has two reasons. First, as Karni and Schmeidler (1991) show, once we depart from expected utility, one of the following three assumptions has to be relaxed: (i) time consistency; (ii) consequentialism; or, (iii)

reduction of compounded lotteries. Assumptions (i) and (ii) are in the heart of the Principle of Optimality of dynamic programming (see Rust (2006), section 3.6). Therefore, to be able to solve a multiple-period problem outside of the expected utility framework by standard dynamic programming, one must relax assumption (iii). However, by relaxing (iii), one would be including preferences about the time of resolution of the uncertainty in the model, just as in the recursive preferences of Kreps and Porteus (1978) and Epstein and Zin (1989).[12] Since the central goal of this study is to develop a simple, parsimonious and stylized model to address the over-fitting critique within the asymmetric preferences literature, we restrict the model to a two-period framework. The second reason refers to the separability of time and risk preferences and we will return to it very soon (after Proposition 1 below).

The economy has two assets, one risky and one risk-free. Define the value of the risky asset at $t+1$ to be $X_{t+1} = P_{t+1} + D_{t+1}$, where P_{t+1} is the price of the asset at $t+1$ and D_{t+1} is the value of some cash flow the investor received between t and $t+1$ (in the case of a stock, D is the dividend). Define X^f_{t+1} to be the value of the risk-free asset at $t+1$ and P^f_t its price at t. Let C_t be the agent's consumption at t, ξ and ξ^f be the quantity of the risky and risk-free assets he buys at t respectively, and W_t be his initial wealth. Then, under time-separability, he solves:

12. Indeed, according to Rust (2006), recursive preference is the only class of non-expected utility preferences that allows the use of standard dynammic programming (backward induction) to solve multi-period problems.

$$\underset{\xi,\xi^f\in\mathbb{R}^2}{Max}\ Q_t^{\tau}\left(u\left(C_t\right)+\beta u\left(C_{t+1}\right)\right) \tag{2}$$

$$s.t.\quad C_t = W_t - P_t\xi - P_t^f\xi^f$$

$$C_{t+1} = X_{t+1}\xi + X_{t+1}^f\xi^f$$

where β is the time discount factor, u is the utility function, $Q_t^{\tau}\left(x\right)$ is the τ^{th} quantile of the conditional distribution of the random variable x (conditional on the information set available at time t).

This agent derives utility only from consumption, as usual, and cares about the worst outcome (in terms of the utility for both periods) that may occur with probability $\left(1-\tau\right)$. In other words, this agent follows the famous advice "Focus on the downside, and the upside will take care of itself". As discussed in sub-section I.A, the higher his level of downside risk aversion, the lower τ.

A key feature of problem (2) is that downside risk aversion and elasticity of intertemporal substitution (EIS) are automatically disentangled. This is a direct consequence of the quantile's equivariance for monotonic transformations. Note that, according to equation (1), we have

$$Q_t^{\tau}\left(u\left(C_t\right)+\beta u\left(C_{t+1}\right)\right) \tag{3}$$

$$= \quad u\left(C_t\right)+\beta u\left(Q_t^{\tau}\left(C_{t+1}\right)\right),$$

since u is a strictly increasing function.

Hence, all uncertainty in problem (2) is resolved by parameter τ, since $Q_t^{\tau}\left(C_{t+1}\right)$ is deterministic at t. The only role played by u is to discount consumption across time: depending on the concavity of u, the agent will combine present consumption, C_t, and the certainty equivalent of future consumption (which, for the quantile maximizer, is equal to $Q_t^{\tau}\left(C_{t+1}\right)$). In other words, the concavity of u will only define the EIS, denoted by ψ. Specializing $u\left(c\right)=\frac{c^{1-\gamma}-1}{1-\gamma}$, we have $\psi=\frac{1}{\gamma}$.[13] Note that such an assumption for the functional form of u imposes no restriction on risk preference: it simply restricts the EIS to being constant.

The EIS parameter, $\psi=\frac{1}{\gamma}$, defines the degree of substitutability-complementarity between consumption today, C_t, and the certainty equivalent of consumption tomorrow, $Q_t^{\tau}\left(C_{t+1}\right)$. For $\psi\rightarrow 0$, C_t and $Q_t^{\tau}\left(C_{t+1}\right)$ become perfect complements, and we have the agent's objective function given by

13. Defining $U\left(C_t,Q_t^{\tau}\left(C_{t+1}\right)\right)=\frac{C_t^{1-\gamma}-1}{1-\gamma}+\beta\frac{\left(Q_t^{\tau}\left(C_{t+1}\right)\right)^{1-\gamma}-1}{1-\gamma}$

we have that

$$\psi\equiv-\frac{\frac{\partial U}{\partial Q_t^{\tau}\left(C_{t+1}\right)}/\frac{\partial U}{\partial C_t}}{Q_t^{\tau}\left(C_{t+1}\right)/C_t}\frac{d\left(Q_t^{\tau}\left(C_{t+1}\right)/C_t\right)}{d\left(\frac{\partial U}{\partial Q_t^{\tau}\left(C_{t+1}\right)}/\frac{\partial U}{\partial C_t}\right)}=\frac{1}{\gamma}.$$

$$U\left(C_t, Q_t^{\tau}\left(C_{t+1}\right)\right) = \min\left\{C_t, Q_t^{\tau}\left(C_{t+1}\right)\right\}.$$

At the other extreme, for $\psi \to \infty$, C_t and $Q_t^{\tau}\left(C_{t+1}\right)$ become perfect substitutes, i.e., the agent maximizes

$$U\left(C_t, Q_t^{\tau}\left(C_{t+1}\right)\right) = C_t + \beta Q_t^{\tau}\left(C_{t+1}\right).$$

For the intermediate case of $\psi = 1$, we end up with the Cobb-Douglas

$$U\left(C_t, Q_t^{\tau}\left(C_{t+1}\right)\right) = C_t\left(Q_t^{\tau}\left(C_{t+1}\right)\right)^{\beta}.$$

With respect to the time discount factor β, its role is to determine the marginal rate of substitution between C_t and $Q_t^{\tau}\left(C_{t+1}\right)$. Therefore, ψ defines the degree of substitutability-complementarity between C_t and $Q_t^{\tau}\left(C_{t+1}\right)$, and β parameterizes such a relation.[14]

What are the implications of the quantile maximization asset pricing model? With the following proposition, proved in the appendix, we initiate this analysis.

PROPOSITION 1. Suppose a consumer-investor solves problem (2) and $u\left(c\right) = \frac{c^{1-\gamma}-1}{1-\gamma}$. Then, the Euler equations are given by

$$P_t = \beta\left(Q_t^{\tau}\left(\frac{C_{t+1}}{C_t}\right)\right)^{-\gamma} Q_t^{\tau}\left(X_{t+1}\right) \tag{4}$$

14. On the empirical side, we will see that both parameters are also separately identified by our estimation method.

$$P_t^f = \beta \left(Q_t^\tau \left(\frac{C_{t+1}}{C_t} \right) \right)^{-\gamma} X_{t+1}^f \tag{5}$$

We should now study the asset pricing implications of equations (4) and (5). Before proceeding, however, we discuss the second reason, as mentioned above, for the employment of a two-period model. As one may have already noticed, the equivalence stated in equation (3) only holds for a two-period setting, where there is only one random variable, namely, the consumption level in period $t+1$. In a three-period model, for instance, one would not be able to interchange the quantile and the utility functions as in (3), since the quantile function is not a linear operator when applied to more than one random variable. This would have at least two important consequences: first, risk and intertemporal preferences would not be automatically disentangled for the quantile agent in a multi-period setting; second, the quantile agent's Euler equations would not be computed as in Proposition 1, since it would be necessary to differentiate inside the quantile function.

Note, however, that the results presented above would continue to hold in a multi-period framework if we modeled the resolution of the uncertainty through a scenario-based reasoning. Accordingly, in this case, the agent would already present

$$u\left(C_t\right) + \beta u\left(Q_t^\tau\left(C_{t+1}\right)\right) + \beta^2 u\left(Q_t^\tau\left(C_{t+2}\right)\right) + ...$$

as his initial objective function and nothing would change in a setting with more than two periods. Such a variant of the model is worth exploring and is object of future research.

We turn now to the analysis of equations (4) and (5). The first step is to understand whether they respect the Law of One Price and the no-arbitrage condition. Then, we solve the model under the standard assumption of joint lognormality for returns and consumption growth, deriving closed-forms for the risky return, the risk-free rate and the equity premium in equilibrium.

Since we ignore transaction costs, any candidate for an equilibrium pricing system has to respect the Law of One Price: prices should be linear. That is, denoting $\Xi_t = \left(\xi_t, \xi_t^f\right)$ to be a portfolio formed at t, with price given by P_t^{Ξ}, the pricing system has to imply $P_t^{\Xi} = \xi_t P_t + \xi_t^f P_t^f$. Otherwise, P_t and P_t^f cannot be equilibrium prices because of arbitrage opportunities among the individual assets and the portfolio. Equations (4) and (5) respect this condition. Defining $\eta_t = \beta \left(Q_t^{\tau}\left(\frac{C_{t+1}}{C_t}\right)\right)^{-\gamma}$, we have

$$\begin{aligned} P_t^{\Xi} &= \eta_t Q_t^{\tau}\left(\xi_t X_{t+1} + \xi_t^f X_{t+1}^f\right) \\ &= \eta_t\left(Q_t^{\tau}\left(\xi_t X_{t+1}\right) + \xi_t^f X_{t+1}^f\right) \\ &= \eta_t Q_t^{\tau}\left(\xi_t X_{t+1}\right) + \eta_t \xi_t^f X_{t+1}^f \\ &= \eta_t Q_t^{\tau}\left(\xi_t X_{t+1}\right) + \eta_t Q_t^{\tau}\left(\xi_t^f X_{t+1}^f\right) \\ &= \xi_t P_t + \xi_t^f P_t^f, \end{aligned}$$

where the second line follows from the quantile equivariance. Note that for a degenerate

random variable x, $Q^{\tau}(x) = x$ for any $\tau \in [0,1]$, and this implies $Q_t^{\tau}\left(X_{t+1}^f\right) = X_{t+1}^f$.

As is well-known, a linear pricing system does not completely rule out arbitrage opportunities. Hence, we need to impose two mild conditions to end up with an arbitrage-free model.

PROPOSITION 2. Suppose that (i) the risky asset payoff X_{t+1} is a continuous random variable and (ii) $\tau \in (0,1)$. Then, the pricing model given by equation (4) rules out arbitrage opportunities.

Both conditions of Proposition 2 (proved in the appendix) are reasonable. The continuity of the risky asset payoff comes for free for stock prices. The second condition, more subtle, rules out two well known agents in decision theory, the so-called MaxMin and MaxMax. The MaxMin agent ($\tau = 0$) summarizes a lottery by looking at the very worst case scenario that may take place (that is, the worst case scenario that may occur with probability 1). On the other hand, the MaxMax ($\tau = 1$) summarizes a lottery by looking at the very best case scenario that may take place (or, in other words, the worst case scenario that may occur with probability 0). Since both agents represent extreme behaviors (the extremely pessimistic and the extremely optimistic), excluding them is not a restrictive assumption.

In the next section, we solve the model under the standard assumption of joint lognormality for returns and consumption growth, deriving closed-forms for the risky return, the risk-free rate and the equity premium in equilibrium.

II. Dynamics, Model Solution, and Simulation

We solve the model with both constant and fluctuating economic uncertainty. Although the solution under constant economic uncertainty is enough to match both the risk-free rate and the risk premium under reasonable levels for the preference-related parameters, it does not generate a time-varying risk premium. To improve the model in this direction, we allow stochastic volatility in the economy dynamics. The model is then simulated under this richer environment.

A Dynamics 1: constant economic uncertainty

Assume

$$\begin{aligned} g_{t+1} &= \mu_c + \eta_{t+1}, \quad \eta_{t+1} \sim iid\ N(0, \sigma_c) \\ r_{t+1} &= \mu_r + u_{t+1}, \quad u_{t+1} \sim iid\ N\left(0, \sigma_r^2\right) \end{aligned} \tag{6}$$

where $g_{t+1} = \log\left(C_{t+1}/C_t\right)$, $r_{t+1} = \log\left(X_{t+1}/P_t\right)$ and $Cov\left(\eta_{t+1}, u_{t+1}\right) = \sigma_{cr}$.

Under this framework, the closed-forms for the risky return, the risk-free rate and the equity premium are given by the following proposition.

Proposition 3. If returns and consumption growth are jointly lognormally distributed,

following (6), and the pricing system is given by equations (4) and (5), then

$$\begin{aligned} r_{t+1} &= -\log(\beta) + \gamma\mu_c + \Phi^{-1}(\tau)(\gamma\sigma_c - \sigma_r) + u_{t+1} && (7) \\ r^f_{t+1} &= -\log(\beta) + \gamma\mu_c + \gamma\sigma_c\Phi^{-1}(\tau) && (8) \\ E_t\left(r_{t+1} - r^f_{t+1}\right) &= -\sigma_r\Phi^{-1}(\tau) && (9) \end{aligned}$$

where r^f_{t+1} refers to the risk-free asset return and Φ^{-1} is the inverse of the cumulative distribution function of a standard normal random variable.

To gain intuition on equations (8) and (9), it is useful to compare them to the analogous equations from the canonical expected utility model. As first derived by Hansen and Singleton (1983), it is well-known that under expected utility maximization and lognormality of returns and consumption growth we have

$$r^f_{t+1} = -\log(\beta) + \gamma\mu_c - \frac{1}{2}\gamma^2\sigma_c^2, \tag{10}$$

and

$$E_t\left(r_{t+1} - r^f_{t+1}\right) = -\frac{1}{2}\sigma_r^2 + \gamma\sigma_{cr}. \tag{11}$$

We first focus on the predictions for the risk-free return. First, in both models, the risk-free rate is linear in expected consumption growth with the slope equal to the inverse of the

elasticity of intertemporal substitution. The lower the EIS (i.e., the higher the desire for consumption smoothing across time), the higher the risk-free rate. This effect is increasing in the expected consumption growth, meaning that the agent will be less willing to save if he expects tomorrow's consumption to be higher.

Second, also common to both models, the higher the rate at which the agent discounts future utility (the lower β), the higher the risk-free rate he requires in order to save.

Third, and this is a first novelty of the quantile approach, a higher variability of consumption growth may have either positive or negative effects on the level of the risk-free rate under the quantile model. If $\tau > 0.5$, a high standard deviation of consumption growth generates a high risk-free rate. If $\tau < 0.5$, a high standard deviation of consumption growth generates a low risk-free rate. The intuition for this is clear: if the agent is optimistic ($\tau > 0.5$), a higher variability is interpreted by him as a higher chance of getting a high level of consumption tomorrow and hence, he becomes less willing to save (higher risk-free rate). In the case of pessimism ($\tau < 0.5$), a higher variability is interpreted as a higher chance of getting a low level of consumption tomorrow, which leads the agent to save more (lower risk-free rate). The strength of this effect, as expected, is increasing in the desire of smoothing consumption across time (γ).

The separation of intertemporal and risk preferences under the quantile model becomes evident when we compare the third terms of equations (8) and (10). In equation (10), we have γ^2, where one γ stands for the risk aversion and the other γ is the inverse of the EIS.

In equation (8), we have the product between the inverse of the EIS and a function of the downside risk aversion.

We now turn to the equity premium equation (9). The risk premium does not depend on the covariance between consumption and stock returns as in the canonical model but, instead, on the standard deviation of the stock return.[15] A higher standard deviation may require either a higher or a lower expected return, depending again on whether τ is greater or less than 0.5. The intuition is the same as above: under optimism ($\tau > 0.5$), a high variability is interpreted as a high chance of getting good returns which, therefore, increases prices (decreasing expected returns). Under pessimism ($\tau < 0.5$) a high variability means a high chance of getting bad returns which causes prices to decrease (increasing expected returns).

These differences imply a better performance of the quantile model when taken to data. Because risk and time preferences are now disentangled we have degrees of freedom to fit both the risk-free rate and the equity premium (just as in Epstein and Zin (1989)). Moreover, the source of risk has now changed. Under expected utility, the covariance between consumption and risky return is the source of risk. This is empirically low, generating the necessity of a high risk aversion to match the equity premium. However, under quantile utility, risk is determined by the standard deviation of the risky return. This value is high in data and,

15. The variance term that shows up in equation ((11)) is simply a Jensen's inequality adjustment (since the expression is about log returns). All that matters for the difference between the risky and the risk-free returns is the covariance term.

therefore, we attenuate the role of the downside risk aversion.

Yearly US data on consumption and returns ranging from 1889 to 2009 can be found on Professor Robert Shiller's website.[16] The risky and risk-free returns are from the S&P 500 and 1-year treasury bill, respectively. The series for per capita consumption are based on the NIPA and NBER series of consumption.

According to this data set, the average real stock log return has exceeded the average treasury bills log return in about 5 percent per year in the post-war period. Stock log return has had a standard deviation about 17 percent per year, and the covariance between stock log return and per capita log consumption growth has been about 0.2 percent. Inserting these values into equation (11) and solving for γ, we have $\gamma = 32$. Hence, in order to fit these patterns of the data, the canonical model requires a risk aversion coefficient that is too high (equity premium puzzle).

But let us suppose one is willing to accept $\gamma = 32$. Then we run into the risk-free rate puzzle. The per capita log consumption growth series has presented annual mean and standard deviation of about 2.1 and 2.2 percent, respectively. The risk-free log return has been about 1.4 percent. Calibrating equation (10) with these values and solving for the time discount factor (β), we have an absurd $\beta = 1.59$ (it is unreasonable to assume that people prefer later utility).

Doing the same exercise using the quantile model equations, we first impose the left hand

16. http://www.econ.yale.edu/~shiller/data.htm, as in November 2010.

side of (9) to be 5 percent and the standard deviation of the risky log return to be 17 percent. Solving for τ, we have $\tau = 0.38$. So, in order to fit the equity premium, the agent has to care about the worst that may happen with probability 62 percent. At a first glance, this does not seem to be a high degree of pessimism. We soon will return to this point.

To compute the time discount factor (β) necessary to fit the observable risk-free rate we should calibrate equation (8) with empirically acceptable values for the EIS. In a recent work using microdata, Engelhardt and Humar (2009) estimate the EIS to be 0.74, with a 95% confidence interval ranging from 0.37 to 1.21. By differentiating between stockholders and nonstockholders and using macrodata, Vissing-Jorgensen (2002) estimates the EIS to be around 0.4 and 0.9, respectively. Given that, we use $\gamma = 1.5$ (i.e., EIS equal to 0.67).[17]

Calibrating equation (8) with $r^f_{t+1} = 1\%$, $\mu_c = 1.9\%$, $\tau = 0.36$, $\sigma_c = 0.021$ and $\gamma = 1.5$, and solving for β, we have $\beta = 1.007$, which is much better than 1.46. By increasing r^f_{t+1} to 2%, we have $\beta = 0.997$, a qualitatively acceptable value (2% is reasonable number for the average risk-free rate as well).

17. All of these estimates are obtained under the expected utility framework. Even though the EIS has nothing to do with risk, one could conjecture that if the true model is related to quantile maximization, such estimates might be biased, which would complicate the calibration of γ under the quantile model. However, the forthcoming estimates for the EIS that I obtain under the quantile model are around these values as well.

B Dynamics 2: stochastic economic uncertainty

A limitation of the quantile model presented so far is that it does not generate a time-varying equity premium (or a time-varying risk-free rate). Because of that, the model cannot theoretically explain two well documented empirical facts: the existence of excess returns predictability and countercyclical risk premia.[18] Since a significant part of the current literature on consumption-based asset pricing addresses matching time variation in expected returns, it is important to improve the quantile model in this direction.

One possible way of doing that is to incorporate fluctuating economic uncertainty into the model. Bansal and Yaron (2004) provide empirical evidence that justifies such a modification. Bansal, Khatchatrian and Yaron (2002) extensively document that a time-varying consumption volatility holds up quite well across different samples and economies. Therefore, we now assume the following dynamics for the real economy:

$$g_{t+1} = \mu_c + \sigma_t \eta_{t+1} \tag{12}$$

$$r_{t+1} = \mu_{r,t} + \varphi \sigma_t u_{t+1} \tag{13}$$

$$\sigma^2_{t+1} = \alpha + \rho\left(\sigma^2_t - \alpha\right) + \sigma_v v_{t+1} \tag{14}$$

where η_{t+1}, v_{t+1} and u_{t+1} are now standard gaussian random variables and $Cov\left(\eta_{t+1}, u_{t+1}\right) =$

18. See Fama and French (1989), Ludvigson and Ng (2007) and Cooper and Priestley (2009), for instance, on the countercyclicality of the risk premium.

σ_{cr}.

The stochastic volatility fluctuates around α, and ρ represents how quickly it gets pulled toward its mean. The evidence in Bansal and Yaron (2004) and Bansal, Khatchatrian and Yaron (2002) are of slow-moving fluctuations in economic uncertainty, implying a ρ close to one. The conditional variances of consumption growth and return are now given by σ_t^2 and $\varphi^2\sigma_t^2$, respectively, and the conditional covariance between consumption growth and return is now $\varphi\sigma_t^2\sigma_{cr}$.

Solving for $\mu_{r,t}$, the next proposition shows that returns and risk premium are now time-variant.

PROPOSITION 4. Under the dynamics defined in equations (12), (13) and (14) and the Euler equations (4) and (5) we have:

$$r_{t+1} = -\ln\beta + \gamma\mu_c + (\gamma - \varphi)\,\sigma_t\Phi^{-1}(\tau) + \varphi\sigma_t u_{t+1} \tag{15}$$

$$r_{t+1}^f = -\ln\beta + \gamma\mu_c + \gamma\sigma_t\Phi^{-1}(\tau) \tag{16}$$

$$E_t\left(r_{t+1} - r_{t+1}^f\right) = -\varphi\sigma_t\Phi^{-1}(\tau) \tag{17}$$

If $\tau < 0.5$ (the pessimistic agent, as discussed in the previous subsection), periods with higher economic uncertainty are periods with higher demand for saving, and hence, lower risk-free rate. This effect is increasing in the desire for consumption smoothing γ, the inverse of the EIS. Moreover, more economic uncertainty raises the risk premium, and this effect

is increasing in φ - the parameter that links economic uncertainty to return uncertainty. Therefore, the time-variation goes in the (theoretically-) intuitive direction.

As Bansal and Yaron (2004) claim, consumption and market volatilities are high during recessions. Given that, the risk premium in equation (17) is countercyclical.[19] In addition, equation (16) implies a procyclical risk-free rate, in line with data as well.

Simulation

We now simulate from this model to better visualize its asset pricing implications. We simulate first the economic uncertainty from equation (14) and then feed equations (12), (15) and (16) with this series. As in Campbell and Cochrane (1999), Barberis, Huang and Santos (2001), Bansal and Yaron (2004), Bansal, Kiku and Yaron (2009) and many others, we assume that the decision interval of the agent is monthly but the targeted data to match are annual. Therefore, we simulate at the monthly frequency and aggregate to annual data.

The stochastic volatility structure added to the model is identical to the one considered in Bansal and Yaron (2004) and Bansal, Kiku and Yaron (2009), and we calibrate parameters (α, ρ, σ_v) with the same values of this last paper.[20] With respect to (μ_c, σ_{cr}), they are set in accordance the sample mean of the consumption growth and the sample covariance between

19. The counter-cyclical feature of the risk premium in the long-run risk model of Bansal and Yaron (2004) also comes from the presence of the stochastic volatility in the risk-premium equation.

20. Equation ((14)) produces a small number (about 5%) of negative values for σ_t^2, as in Bansal and Yaron (2004) and Bansal, Kiku and Yaron (2009). Following them, I replace these negative values with the smallest positive value generated for σ_t^2. Obviously, one could model $\log(\sigma_t^2)$ to get rid of this technical problem (but, in this case, it wouldn't be possible to follow their calibration).

consumption growth and risky return, respectively.

Given such values, we choose the free parameters $(\varphi, \beta, \tau, \gamma)$ seeking to match the first and second moments of the risk-free rate and excess return, and the second moment of consumption growth. Table 1 summarizes the parameters' optimal choices.

parameters for monthly simulation	value
α (mean of economic uncertainty)	0.0072^2
σ_v (standard deviation of log economic uncertainty)	0.28×10^{-5}
ρ (log economic uncertainty persistence)	0.999
μ_c (mean consumption log growth)	0.0018
σ_{cr} (covariance between η and u)	0.5
φ (adjustment of the log return standard deviation)	5.5
β (discount factor)	0.9998
EIS (inverse of γ)	0.6
τ (downside risk aversion)	0.45

Table 1. Configuration of the model parameters.

The preference-related parameters (β, τ, γ) are close to those from the previous subsection. The time discount factor (β) is slightly below one, the EIS of 0.6 implies $\gamma = 1.66$, and the downside risk aversion is now even smaller with $\tau = 0.45$.[21]

Table 2 presents the impacts on the simulated moments of varying both the risk aversion and EIS. The other parameters are kept fixed in accordance with Table 1.

21. Importantly, the quantile model does not need an EIS greater than one to produce good empirical results. This is relevant when compared to Bansal and Yaron (2004). For them, it is crucial for the good results to employ an EIS greater than one, more precisely, equal to 1.5 (and this value is not empirically reasonable, as discussed before.)

τ	EIS	E(r-rf)	σ(r)	E(rf)	σ(rf)	E(g)	σ(g)	cov(g,r)
0.41	0.1	10.0	15.8	3.5	11.2	2.1	2.7	0.2
0.41	0.6	10.0	15.6	0.8	1.9	2.1	2.7	0.2
0.41	1.1	10.0	15.9	0.5	1.0	2.1	2.7	0.2
0.45	0.1	5.5	15.3	11.7	6.2	2.1	2.7	0.2
0.45	**0.6**	5.5	15.3	2.1	1.1	2.1	2.7	0.2
0.45	1.1	5.5	15.3	1.2	0.6	2.1	2.7	0.2
0.49	0.1	1.0	15.0	19.8	1.2	2.1	2.7	0.2
0.49	0.6	1.0	15.0	3.5	0.2	2.1	2.7	0.2
0.49	1.1	1.0	15.0	2.0	0.1	2.1	2.7	0.2
data		4.8	16.8	1.4	1.7	2.1	2.2	0.2
s.e.		(1.5)	(1.8)	(0.5)	(0.3)	(0.3)	(0.5)	(0.0)

other parameters values: following Table 1

Table 2. Varying EIS and downside risk aversion (in %).

From Table 2 we see three effects: (i) higher values of downside risk aversion (i.e., lower values of τ) increase the mean excess return; (ii) lower values for EIS increase the mean risk-risk free return and its volatility; and, (iii) decreasing τ also impacts the mean and standard deviation of the risk-free rate, decreasing the former and increasing the latter.

The theoretical reasons for the effects related to the first moments are the same as those under constant economic uncertainty. A higher downside risk aversion implies a higher price for the risk, and therefore, a higher risk premium, justifying effect (i). A higher complementarity between consumption at t and the certainty equivalent of consumption at $t+1$ implies a higher desire to smooth consumption in time, and therefore, a higher risk-free rate to justify savings from t to $t+1$, which explains effect (ii). Finally, a higher downside risk aversion leads to more savings from period t to period $t+1$ for a given level of economic uncertainty at t, lowering the risk-free rate and justifying (iii).

With respect to the effects related to the second moment of the risk-free rate, the theoretical explanations are the following. The effect in (ii) comes from the natural fact that the volatility of the risk-free rate is a function of the volatility of the economic uncertainty which is decreasing in the EIS (see equation (16)). This makes theoretical sense, since savings should respond more to economic uncertainty, the more the agent cares about smoothing consumption. The reasoning supporting the effect in (iii) follows the same line: the more downside risk averse the agent, the more savings should respond to economic uncertainty.

We therefore conclude that the quantile asset pricing model's predictions are theoretically solid. In addition, when calibrated with empirically reasonable parameters and $\tau = 0.45$, the model is able to reproduce important patterns of financial and macroeconomic data. At this point, a natural question is: how reasonable is $\tau = 0.45$?

C What is a reasonable value for τ ?

Is $\tau = 0.45$ more reasonable than $\gamma = 35$ (the value obtained in sub-section II.A for the risk aversion under expected utility and lognormality) in terms of the implied attitude towards risk? Or, what is a reasonable range for τ?

One way to evaluate τ is to compare the certainty equivalent implicit in a quantile model to the one implicit in a power utility model for risky situations with payoffs following continuous distributions, in accordance with Proposition 2.

Using certainty equivalents of simple bets to relate parameters from different models of

behavior towards risk is a standard procedure in this literature. For instance, Epstein and Zin (1990) use such a strategy to compare the risk aversion levels in Yaari preferences with the risk aversion levels in the expected utility preferences (see their Tables 1 and 2). Bonomo and Garcia (1993), Epstein and Zin (2001), Routledge and Zin (2010), among others, do the same.

A simple and natural risky situation to use is the following. Suppose the agent wants to invest \$1000 and the investment return follows the same distribution considered in (6). Therefore,

$$\ln\left(X_{t+1}\right) \sim N\left(\mu_r + \ln\left(1000\right), \sigma_r^2\right),$$

where, as usual, X_{t+1} is the value of the investment at $t+1$.

For a one-year investment, the sample estimates for μ_r and σ_r^2 are about 0.08 and 0.03 respectively. The initial investment value is immaterial for the forthcoming conclusions.

We can first ask: what are the certainty equivalents for a quantile agent with $\tau = 0.45$ and for an expected power utility agent with $\gamma = 35$ for this uncertain outcome X_{t+1}?

For an expected utility agent with power utility, the certainty equivalent of a lottery with payoff x is given by

$$CE_{EU} = \left[E\left(x^{1-\gamma}\right)\right]^{\frac{1}{1-\gamma}}.$$

For a τ-quantile utility agent, the value of a lottery with payoff x is equal to $Q^{\tau}\left[u\left(x\right)\right]$. So, the certainty equivalent of such a lottery is the solution of $u\left(CE_{QU}\right) = Q^{\tau}\left[u\left(x\right)\right]$. By quantile equivariance,

$$CE_{QU} = Q^{\tau}\left(x\right).$$

Figure 2 presents the histogram of the uncertain investment value at $t+1$, which has mean and standard deviation around \$1103 and \$212, respectively. The vertical dashed lines are the certainty equivalents for the power utility agent with $\gamma = 35$ and for the quantile agent with $\tau = 0.45$ (they are around \$643 and \$1057, respectively).

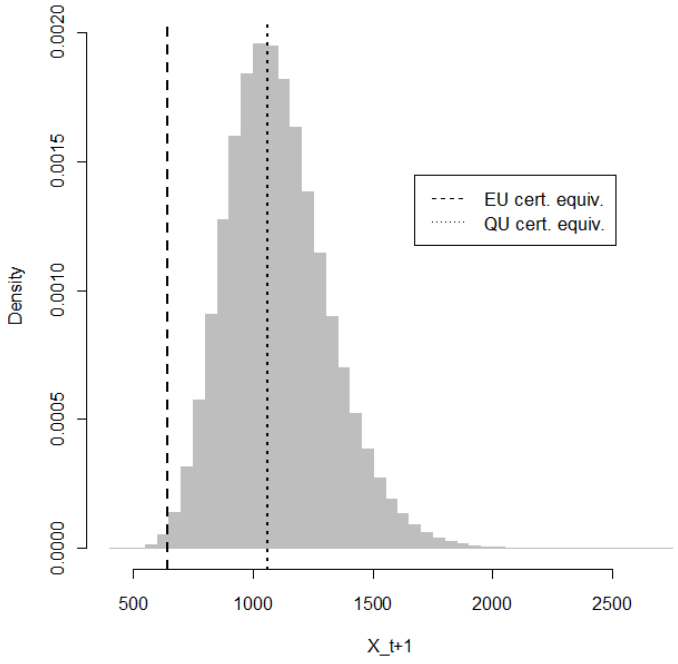

Figure 2. Histogram of uncertain payoff and certainty equivalents for $\gamma = 35$ and $\tau = 0.45$.

A casual review of this figure suggests that the certainty equivalent of a power utility

agent with $\gamma = 35$ is too small compared to what one would expect as reasonable. On the other hand, for a quantile agent with $\tau = 0.45$, his certainty equivalent looks much better. However, it is already well-known in the literature that $\gamma = 35$ generates extreme outcomes in an expected utility setting. So, one can argue that basically any alternative utility specification is going to behave more reasonably. Considering that, perhaps a clearer, more illustrative way to proceed would be to ask: which value of γ would give the certainty equivalent obtained with $\tau = 0.45$? The answer is $\gamma = 2.5$. In other words, in terms of certainty equivalents, a quantile utility agent with $\tau = 0.45$ would be analogous to an expected utility agent with $\gamma = 2.5$, a value which is commonly referred to as reasonable in the literature.

Pursuing this idea further, we can relate many values of τ to many values of γ in terms of producing the same certainty equivalent for the bet defined above. Figure 3 presents this relationship.

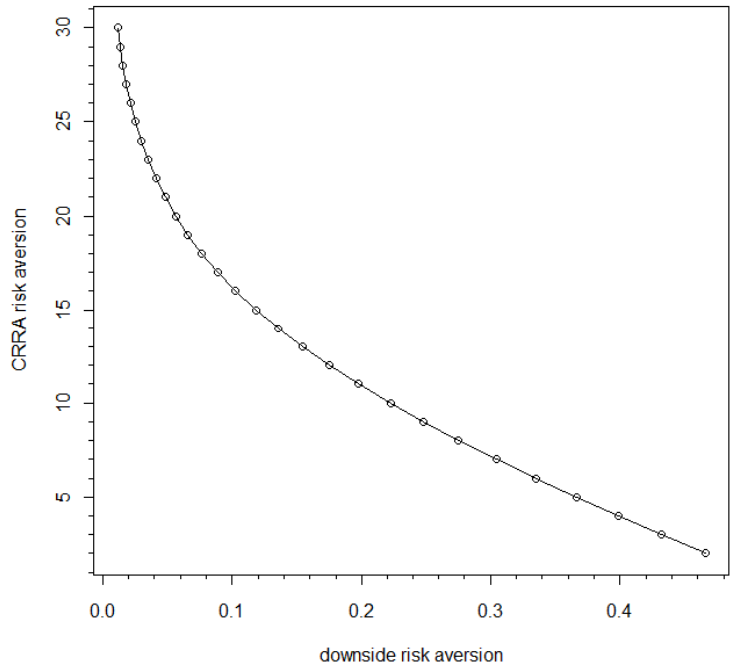

Figure 3. Values of τ and γ that produce the same certainty equivalent in the bet defined above.

Mehra and Prescott (1985) argue that acceptable values for γ would be between 1 and 10. Hence, for the risky situation considered, the analogous interval for τ would be $[0.22, 0.48]$.

D Comparing results

So far we have compared our results only to those from the canonical model. This was done to illustrate the new features of the present approach with respect to the predictions for the risk-free rate and the equity premium.

In this sub-section we briefly compare the results obtained to those of Epstein and Zin (1989) and Weil (1989) (three parameters), Bonomo and Garcia (1993) (four parameters)

and Routledge and Zin (2010) (five parameters), and Barberis, Huang and Santos (2001) (six parameters).

By using recursive preferences, Epstein and Zin (1989) and Weil (1989) disentangle risk aversion and EIS and still have the time discount rate - the same parameters we have here. By doing so, they are able to fit both the equity premium and the risk-free rate. However, the extremely high risk aversion remains crucial. As Table 1 in Weil (1989) shows, in order to match the average of risk-free and excess returns, risk aversion and EIS have to be set at 45 and 0.1, respectively. If risk aversion is decreased to 1, the premium is as low as 0.45 percent, while the mean risk-free rate reaches 25 percent. Furthermore, nothing is said about second moments.

With one extra parameter compared to our model (the one that regulates the disappointment aversion), the model in Bonomo and Garcia (1993) under a joint random walk for consumption and dividend growth rates[22] produces an average equity premium on the order of 2.5 percent with standard deviation about 12.8 percent. The risk-free rate averages about 4.5 percent. This is the best they are able to get using what they consider reasonable values for their parameters.

By adding one more parameter to the disappointment aversion model, Routledge and Zin (2010) are able to generate good results with this framework. By means of a countercyclical risk aversion (produced by an endogenous variation in the probability of disappointment),

22. Comparable to the dynamics I use here.

they produce a large equity premium (about 6 percent) and a risk-free rate with low volatility and mean. However, they still have difficulty with fitting the risky return volatility and maintaining the 6 percent equity premium at the same time.

Barberis, Huang and Santos (2001) assume a functional form for preferences based on prospect theory, which has 6 parameters. Their model succeeds in explaining the first and second moments of the risk-free rate, the equity premium and the consumption growth, and produces a time-varying risk premium (that comes from the impact of the agent's past portfolio result on his sensitivity for future losses).

III. Model Estimation

The previous section presented the quantile utility asset pricing model under the assumption of joint conditional lognormality of asset returns and consumption growth. This was useful for building intuition with respect to the model. However, it is well-known that the lognormality assumption is not consistent with all the properties of historical stock returns. For example, stock log returns show weak evidence of skewness and strong evidence of excess kurtosis, at least for short horizons. Hence, it is important to understand how the model performs if we relax the lognormality assumption.

In this section, we discuss how to estimate the model free of distributional assumptions.

A GMM-based estimator is proposed, the identification of the parameters is analyzed, and sufficient conditions for consistency are established. Moreover, since the proposed estimator is defined over non-differentiable moments, its asymptotic distribution is derived.

In the appendix, we also estimate the model under the lognormality assumption. This complements the simulation exercise performed in section II by providing confidence bands to the parameters.

A A general estimation method

The estimation of β, γ and τ free of any distributional assumption will be performed by combining GMM and quantile regression's elements. However, since τ, the respective conditional quantile, also has to be estimated, the present problem is distinct from the standard quantile regression, where τ is taken as given.

In the case of the canonical expected utility model, the standard way of estimating the model free of distributional assumptions is by applying the GMM of Hansen (1982), as was first proposed by Hansen and Singleton (1982). This is straightforward since it is just a matter of transforming conditional into unconditional expectations. However, this is not the case if we want to estimate the quantile Euler equations (4) and (5). There is nothing analogous to the law of iterated expectations for quantiles. Moreover, equations (4) and (5) are not even moment conditions. But, as we see now, it is possible to overcome such difficulties in a simple fashion.

Let the vector $\theta_0 = (\tau_0, \beta_0, \psi_0)$ represent the populational values for the downside risk aversion, the time discount factor and the EIS, respectively. Define $Y_{t+1} = \left(\frac{C_{t+1}}{C_t}, R_{t+1}, R^f_{t+1}\right)$ and let $Y \equiv \{Y_t : \Omega \longrightarrow \mathbb{R}_+ \times \mathbb{R},\ t = 1, ..., T\}$ be a stochastic process defined on a complete probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F} \equiv \{\mathcal{F}_t : t = 1, ..., T\}$ and $\mathcal{F}_t \equiv \sigma\{Y_s : s \leq t\}$. Define also $\varepsilon_{c,t+1}$ and $\varepsilon_{r,t+1}$ to be the random variables such that

$$\frac{C_{t+1}}{C_t} = Q^{\tau_0}\left(\frac{C_{t+1}}{C_t}|\mathcal{F}_t\right) + \varepsilon_{c,t+1} \tag{18}$$

and

$$R_{t+1} = Q^{\tau_0}\left(R_{t+1}|\mathcal{F}_t\right) + \varepsilon_{r,t+1}. \tag{19}$$

Given this structure, we first note that the asset pricing theory imposes functional forms on the conditional quantiles defined above. From Proposition 1, the risky and risk-free returns in equilibrium should respect the following two equations

$$\beta_0 \left(Q^{\tau_0}\left(\frac{C_{t+1}}{C_t}|\mathcal{F}_t\right)\right)^{-1/\psi_0} Q^{\tau_0}\left(R_{t+1}|\mathcal{F}_t\right) = 1 \tag{20}$$

$$\beta_0 \left(Q^{\tau_0}\left(\frac{C_{t+1}}{C_t}|\mathcal{F}_t\right)\right)^{-1/\psi_0} R^f_{t+1} = 1. \tag{21}$$

where we now use the EIS parameter ψ_0 instead of its inverse γ_0.

By dividing equation (20) with equation (21) we get

$$Q^{\tau_0}\left(R_{t+1}|\mathcal{F}_t\right) = R^f_{t+1}. \tag{22}$$

Rearranging equation (21), we have

$$Q^{\tau_0}\left(\frac{C_{t+1}}{C_t}|\mathcal{F}_t\right) = \left(\beta_0 R^f_{t+1}\right)^{\psi_0}. \tag{23}$$

Hence, the theoretical model imposes that, in equilibrium, all the information that matters for the conditional quantiles of R_{t+1} and C_{t+1}/C_t is R^f_{t+1} (which is already known at t, i.e., $R^f_{t+1} \in \mathcal{F}_t$). More than that, the model defines the whole functional form of such conditional quantiles.

Given (22) and (23), we can state the following proposition.

PROPOSITION 5. Let Z_t be an $m \times 1$ vector such that $Z_t \in \mathcal{F}_t$. Define

$$g\left(Y_{t+1}, Z_t, \theta_0\right) = \begin{pmatrix} \left(\tau - 1\left[\frac{C_{t+1}}{C_t} < \left(\beta_0 R^f_{t+1}\right)^{\psi_0}\right]\right) Z_t \\ \left(\tau - 1\left[R_{t+1} < R^f_{t+1}\right]\right) Z_t \end{pmatrix}$$

where $1\left[\cdot\right]$ is the logical indicator function.

Then,

$$E\left[g\left(Y_{t+1}, Z_t, \theta_0\right)|\mathcal{F}_t\right] = 0. \tag{24}$$

Therefore, we have $2m$ moment conditions and 3 parameters to be estimated. For $m \geq 2$ we may use Hansen's (1982) GMM approach,

$$\widehat{\theta} = \underset{\theta \in \Theta \subseteq \mathbb{R}^3}{\arg\min} \left(\frac{1}{T} \sum_{t=1}^{T} g_t \left(Y_{t+1}, Z_t, \theta \right) \right)' W_T \left(\frac{1}{T} \sum_{t=1}^{T} g \left(Y_{t+1}, Z_t, \theta \right) \right) \tag{25}$$

where W_T is a general weighting matrix.

Even though the interpretation of a quantile regression as a GMM problem is standard, we cannot directly use the established asymptotic results (from Koenker and Basset (1978) and Powell (1984, 1896), for example). In quantile regressions, τ_0 is a given number and not a parameter to be estimated. Hence, the fact that our central task is the estimation of τ_0 places this econometric problem in a new environment.

We have to understand whether the GMM estimation of θ_0 is indeed feasible. In other words, we have to understand whether θ_0 is identified and derive the consistency and asymptotic distribution of $\widehat{\theta}$. Fortunately, as we see now, we can conclude under mild conditions that θ_0 is globally identified and $\widehat{\theta}$ is consistent and asymptotically normal.

The following proposition presents sufficient conditions for consistency.

PROPOSITION 6. Assume that (i) $V_{t+1} \equiv \left(R_{t+1}, C_{t+1}/C_t, R^f_{t+1}, Z_t \right)$ is strictly stationary and α-mixing of size $-r/(r-1)$, with $r > 1$, (ii) $E \left\| Z_t \right\| < \infty$, where $\left\| \cdot \right\|$ denotes the L_∞-norm, (iii) $\Theta \subseteq \mathbb{R}^3$ is a compact set (iv) $W_T \xrightarrow{p} W_0$, where W_0 is a positive definite matrix, (v) C_{t+1}/C_t is a continuous random variable, (vi) $\left(1, R^f_{t+1} \right)' \in Z_t$ and (vii)

$Var\left(R_{t+1}^{f}\right) > 0.$

Then, $\widehat{\theta} \xrightarrow{p} \theta_0$ for $\widehat{\theta}$ defined in equation (25).

Assumptions (i), (ii) and (iii) are technical and often present. Assumption (iv) is satisfied by a special choice for W_T, as Proposition 7 will show. Assumption (v) is standard in quantile regressions and natural for aggregate consumption growth. Assumption (vi) simply says that the instrument set should include a constant and the risk-free rate. Assumption (vii) is a standard rank condition which requires the explanatory variable to be non-degenerate. Assumptions (v), (vi) and (vii) are the crucial ones for global identification, as can be seen in the proof (in the appendix).

The proof of Proposition 6 shows that $E\left[g\left(V_{t+1},\theta\right)\right] = 0$ if and only if $\theta = \theta_0$. By combining this with the fact that W_0 is positive definite, we conclude that the populational object-function of our GMM estimator has a unique optimum at $\theta = \theta_0$, that is, θ_0 is globally identified.[23]

The global identification of the parameters can be seen as a fortunate achievement of the present model. In fact, according to Newey and McFadden (1994), "If $E\left[g\left(z,\theta\right)\right]$ is nonlinear in θ, then specifying primitive conditions for identification becomes quite difficult ... A practical solution to the problem of global GMM identification, that has often been adopted,

23. Lemma 2.3 in Newey and McFadden (1994) shows that if W_0 is positive semi-definite and $W_0 E\left(g\left(V_{t+1},\theta\right)\right) = 0 \Leftrightarrow \theta = \theta_0$, then the populational GMM object-function is uniquely minimized at $\theta = \theta_0$. However, as it is trivial to show, if W_0 is positive definite, one only needs $E\left(g\left(V_{t+1},\theta\right)\right) = 0 \Leftrightarrow \theta = \theta_0$ to get the same result.

is to simply assume identification. This practice is reasonable, given the difficulty of formulating primitive conditions, but it is important to check that it is not a vacuous assumption whenever possible, by showing identification in some special cases." For instance, as Newey and McFadden (1994) points out, in the canonical model of Hansen and Singleton (1982) it is possible to derive global identification only under a particular form of the conditional distribution.

Proposition 7 now proposes a specific choice for W_T.

PROPOSITION 7. Suppose that assumption (i) holds, assumption (ii) is strengthened to (ii') there exists some $\delta > 0$ such that $E\left\|Z_t\right\|^{2r+2\delta}$ and additionally assume (viii) $\tau_0 \in (0,1)$, (ix) $P\left(\varepsilon_{c,t+1} < 0, \varepsilon_{r,t+1} < 0 | Z_t\right) < \tau_0$ and (x) $E\left(Z_t Z_t'\right)$ is nonsingular. Specialize W_T as

$$W_T = \left(\frac{1}{T}\sum_{t=1}^{T} g\left(V_{t+1}, \widetilde{\theta}\right) g\left(V_{t+1}, \widetilde{\theta}\right)'\right)^{-1}, \tag{26}$$

where $\widetilde{\theta}$ is any estimator such that $\widetilde{\theta} \xrightarrow{p} \theta_0$.

Then

$$W_T \xrightarrow{p} \Sigma_0^{-1},$$

where

$$\Sigma_0 \equiv E\left[g\left(V_{t+1}, \theta_0\right) g\left(V_{t+1}, \theta_0\right)'\right]$$

is positive definite.

As usual, estimator $\widetilde{\theta}$ may be computed in a first step by $\widehat{\theta}$, with W_T as the identity matrix (according to Proposition 6). Assumption (viii) rules out the MaxMin and MaxMax agents from the analysis, which had already been done to ensure no-arbitrage in the model. Hence, such agents are not only incompatible with no-arbitrage, but also may jeopardize the identification of the model. Assumption (ix) is also a mild one. First, note that under independence of $\varepsilon_{c,t+1}$ and $\varepsilon_{r,t+1}$, defined in equations (18) and (19), $P\left(\varepsilon_{c,t+1} \leq 0, \varepsilon_{r,t+1} \leq 0 | Z_t\right) = \tau_0^2$, and this is satisfied. Hence, this assumption is about $\varepsilon_{c,t+1}$ and $\varepsilon_{r,t+1}$ not being too positively correlated. But, note that in the extreme case of positive correlation, where $\varepsilon_{c,t+1} = \varepsilon_{r,t+1}$, we have $P\left(\varepsilon_{c,t+1} \leq 0, \varepsilon_{r,t+1} \leq 0 | Z_t\right) = \tau_0$. Therefore, imposing $P\left(\varepsilon_{c,t+1} \leq 0, \varepsilon_{r,t+1} \leq 0 | Z_t\right) < \tau_0$ is not restrictive at all. Assumption (x) is the usual rank condition on the instruments.

We now turn to the asymptotic distribution of $\widehat{\theta}$. To address the nondifferentiability of $g\left(\cdot\right)$, we use the empirical processes theory approach presented in Andrews (1994) which, under some regularity conditions, replaces the differentiability of $g\left(\cdot\right)$ by the differentiability of $E\left[g\left(\cdot\right)\right]$. The next proposition derives the asymptotic distribution of the estimator.

PROPOSITION 8. Suppose all assumptions of Proposition 6 hold, where assumption (ii) is strengthened to (ii') of Proposition 7. Furthermore, assume that (xi) $f_{\varepsilon_{c,t+1}}\left(0 | Z_t\right)$ is bounded away from zero, and (xii) the matrix $G_0' W_0 G_0$ is nonsingular, where $G_0 \equiv$

$\nabla_\theta E\left(g\left(V_{t+1},\theta_0\right)\right)$ is a $2m\times 3$ matrix with entries

$$\begin{aligned}
G_{i1} &= E\left(Z_{it}\right)\\
G_{i2} &= -\psi_0\beta_0^{(\psi_0-1)}E\left(f_{\varepsilon_{c,t+1}}\left(0|Z_t\right)\left(R_{t+1}^f\right)^{\psi_0}Z_t\right)\\
G_{i3} &= -\beta_0^{\psi_0}E\left(f_{\varepsilon_{c,t+1}}\left(0|Z_t\right)\left(R_{t+1}^f\right)^{\psi_0}\log\left(\beta_0 R_{t+1}^f\right)Z_t\right)\\
G_{j1} &= E\left(Z_{jt}\right)\\
G_{j2} &= 0\\
G_{j3} &= 0
\end{aligned}$$

for $i=1,...,m$ and $j=m+1,...,2m$.

Then

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right)\xrightarrow{d}N\left(0,\left(G_0'W_0G_0\right)^{-1}G_0'W_0\Sigma_0W_0G_0\left(G_0'W_0G_0\right)^{-1}\right).$$

Assumption (xi) is standard in quantile regressions, and rules out having zero in the denominator. Assumption (xii) implies the existence of the term $\left(G_0'W_0G_0\right)^{-1}$ in the asymptotic variance. Proposition 8 tells us that the usual GMM asymptotic distribution for differentiable moments conditions is valid for our nondifferentiable specific case as well. This implies that the optimal choice for W_T is the one that converges in probability to Σ_0^{-1}, which is the weighting matrix defined in Proposition 7. The optimal weighting matrix simplifies

the estimator's asymptotic variance to $\left(G_0'\Sigma_0^{-1}G_0\right)^{-1}$.

B A simple two-step estimation procedure

Functions such as (25) are difficult to optimize by the standard packages algorithms (*fminsearch*, in MATLAB, or *nlm* and *optim* in R, for instance): they are nonsmooth and highly nonconvex, with numerous local optima. However, as we have only 3 parameters with well defined theoretical bounds (such as $\tau_0 \in [01, .99]$, $\beta_0 \in [0.9, 1.1]$ and $\psi_0 \in [0, 5]$), the optimization is feasible using a grid search in our case.

Nevertheless, it is useful to note that θ_0 can be consistently estimated in an even simpler manner, using a two-step procedure. Such an estimator is not going to be efficient, but this discussion builds intuition into the model and provides a rapid and simple technology for estimating, for instance, the EIS (the estimation of the EIS under Epstein and Zin (1989) preferences, the alternative technology of disentangling risk and time preferences, is much more involving).

In a first step, we estimate τ_0. Equation (22) implies

$$E\left[\tau_0 - 1\left[R_{t+1} < R_{t+1}^f\right]\right] = 0.$$

Hence, a consistent estimator of τ_0 is

$$\widetilde{\tau} = \frac{1}{T}\sum_{t=1}^{T} 1\left[R_{t+1} < R_{t+1}^{f}\right], \tag{27}$$

which is the relative number of observations in the sample such that $R_{t+1} < R_{t+1}^{f}$. From standard arguments, its asymptotic distribution is given by

$$\sqrt{T}\left(\widetilde{\tau} - \tau_0\right) \xrightarrow{d} N\left(0, \tau_0\left(1 - \tau_0\right)\right).$$

Given $\widetilde{\tau}$, we can now estimate (β_0, ψ_0) by a standard linear quantile regression. This is the case since, by the equivariance property of quantiles, equation (23) implies

$$Q^{\tau_0}\left(g_{t+1}|\mathcal{F}_t\right) = \lambda_0 + \psi_0 r_{t+1}^{f}, \tag{28}$$

where $g_{t+1} = \log\left(C_{t+1}/C_t\right)$, $r_{t+1}^{f} = \log\left(R_{t+1}^{f}\right)$ and $\lambda_0 = \psi_0 \log\left(\beta_0\right)$.

The only drawback of using $\widetilde{\tau}$ instead of τ_0 in equation (28) is the usual problem with standard errors of the second step. As is well-known, they have to be corrected because of the noise produced in the first-step estimation. However, in practice, this implies no additional computational cost for our two-step procedure. In standard quantile regressions, the coefficients' asymptotic variance contains the unknown conditional distribution of the error term. Because of that it is common to compute standard errors by bootstrap. Hence, to address the two-step estimation issue, it is natural to incorporate the first step in the

bootstrap procedure.[24]

From $\left(\widehat{\lambda},\widehat{\psi}\right)$ one consistently computes $\widehat{\beta}=\exp\left(\widehat{\lambda}/\widehat{\psi}\right)$. The standard error of $\widehat{\beta}$ should be computed from the bootstrapped covariance matrix of $\left(\widehat{\lambda},\widehat{\psi}\right)$ by the delta method. Accordingly,

$$\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\xrightarrow{d}N\left(0,\exp\left(2\frac{\lambda_0}{\psi_0}\right)\left(\frac{1}{\psi_0^2}\sigma_\lambda^2+\frac{\lambda_0^2}{\psi_0^4}\sigma_\psi^2-2\frac{\lambda_0}{\psi_0^3}\sigma_{\lambda\psi}\right)\right),$$

where σ_λ^2 is the asymptotic variance of $\widehat{\lambda}$, σ_ψ^2 is the asymptotic variance of $\widehat{\psi}$, and $\sigma_{\lambda\psi}$ is the asymptotic covariance between both estimators.

C Empirical results

We now apply the estimation procedures discussed above to a monthly data set. Such data frequency is used to maintain the assumption that the decision interval of the agent is monthly, as in the simulation exercise. Per capita consumption is the sum of personal consumption expenditures on services (PCES, St. Louis Fed) and personal consumption expenditures on nondurable goods (PCEND, St. Louis Fed), divided by the total population (POP, St. Louis Fed). The risky return is the S&P 500 return including dividend payments, and the risk-free return is the 1-month risk-free rate series from Professor Fama located in the CRSP data base. All series are deflated by the consumer price index for all urban

24. That is, from S bootstrapped samples one estimates S pairs $\left(\widehat{\lambda},\widehat{\psi}\right)$ and computes their empirical variance matrix.

consumers (CPIAUCSL, St. Louis Fed). Since both consumption series start in January 1959 in the St. Louis Fed data base, the data set ranges from January 1959 to December 2009.

We define three distinct instrument vectors, $Z_t^{(1)} = \left(1, R_{t+1}^f\right)$, $Z_t^{(2)} = \left(1, R_{t+1}^f, R_t^f\right)$, and $Z_t^{(3)} = \left(1, R_{t+1}^f, R_t^f, R_{t-1}^f\right)$, all three satisfying assumption (vi) in Proposition 6. We do not include lags of consumption growth and risky returns since they have very weak forecasting power over their future realizations (see Cochrane (2006), pp 268).

Columns 2, 3 and 4 of Table 3 present the estimates of θ_0 under the general (one-step) estimation method. Standard errors are analytically computed using the asymptotic distribution derived in Proposition 8.[25] The fifth column of Table 3 shows the result from the two-step procedure presented in the last sub-section. Standard errors are calculated by bootstrap according to the previous sub-section, addressing both issues of τ_0 estimated in a previous step and of the unknown distribution in the asymptotic variance.

Table 4 reproduces Table 3, but allows for the presence of auto-correlation in the empirical moments. In columns 2, 3 and 4, W_T is computed by Newey and West's (1987) estimator. In column 5, we employ overlapping block-bootstap to compute the variance matrix of $\left(\widehat{\lambda}, \widehat{\psi}\right)$.

25. We estimate $f_{\varepsilon_{c,t+1}}(0|Z_t)$ nonparametrically, following Powell (1986), using $\widehat{\varepsilon}_{c,t+1} = \frac{C_{t+1}}{C_t} - \left(\widehat{\beta}_0 R_{t+1}^f\right)^{\widehat{\psi}}$.

	1-step procedure			2-step procedure
block 1	$Z^{(1)}$	$Z^{(2)}$	$Z^{(3)}$	
β	1.002	1.002	1.002	1.002
(se)	(0.001)	(0.001)	(0.002)	(0.0001)
EIS	0.37	0.37	0.36	0.39
(se)	(0.07)	(0.07)	(0.07)	(0.05)
τ	0.44	0.44	0.44	0.43
(se)	(0.02)	(0.02)	(0.02)	(0.02)
J stat.	5.1	7.2	10.3	
(p-value)	(0.02)	(0.07)	(0.07)	

Columns 2, 3 and 4 present the 1-step estimates. $Z^{(j)}$ contains up to the j-th lag of the risk-free rate. Column 5 presents the 2-step estimates. For all columns, no serial-correlation is assumed, justified by the fact that moments are martingale difference sequences according to proposition 5.

Table 3: estimates under no serial-correlation

	1-step procedure			2-step procedure
block 1	$Z^{(1)}$	$Z^{(2)}$	$Z^{(3)}$	
β	1.002	1.002	1.002	1.002
(se)	(0.001)	(0.001)	(0.001)	(0.0001)
EIS	0.35	0.35	0.35	0.39
(se)	(0.08)	(0.08)	(0.08)	(0.06)
τ	0.45	0.45	0.44	0.43
(se)	(0.02)	(0.02)	(0.02)	(0.02)
J stat.	3.8	4.9	9.7	
(p-value)	(0.05)	(0.18)	(0.08)	

Columns 2, 3 and 4 present the 1-step estimates. $Z^{(j)}$ contains up to the j-th lag of the risk-free rate. Column 5 presents the 2-step estimates. Serial-correlation is allowed for all columns and asymptotic variance is estimated by Newey-West with 6 lags (columns 2, 3 and 4) and by overlapping block-bootstrap with 6 lags (column 5).

Table 4: estimates under serial-correlation

The estimates from Tables 3 and 4 are very similar. This should be a consequence of

the very low empirical serial-correlation of consumption growth and returns. The estimates across columns in both tables are also very similar, which is evidence of the robustness of the estimation methods. In particular, the results from the one-step and the two-step procedures are very close to each other. This was expected since both procedures are consistent.

Although the time discount factor estimates are slightly above one, it is in general not possible to reject the hypothesis $\beta_0 < 1$. The estimates of the elasticity of intertemporal substitution go from 0.35 to 0.39 and are all significantly different from zero. The downside risk aversion is estimated ranging from 0.43 to 0.45 and are all significantly different from 0.5.

The EIS estimation under an alternative framework is a contribution of the present paper. As discussed in Guvenen (2006), most of the estimated Euler equations deliver extremely low values for such a parameter, often not significantly different from zero. However, macroeconomists calibrate their models using positive values for the EIS, generally between 0.5 and 1. Hence, the present results diminish this contradiction between the dynamic macroeconomics literature and the Euler-equations-based estimates for the EIS.

With respect to the model specification, the overidentifying restrictions test rejects the model at 5% only in the first column of Table 3. This is a remarkable result given the usual rejection of asset prices models by the J-test.

Since these results from estimation are qualitatively the same as those obtained under simulation (the time discount factor used in the simulation exercise was 0.9998, the EIS was

0.6 and the downside risk aversion was 0.45), we conclude that such values are robust.

IV. Conclusion

We considered a framework where a single agent makes his decision about consumption-investment looking at worst-case scenarios, which depend on his degree of pessimism. This agent can be motivated by a well-known quote among professional investors: "Focus on the downside, and the upside will take care of itself".

Using the quantile utility maximizer agent of Manski (1988) and Rostek (2010), we attached the agent's degree of pessimism to a well defined parameter. As a consequence, we disentangled attitude towards risk and attitude towards intertemporal substitution in a novel way.

Two important results emerged. First, with only 3 preference-related parameters, the model was able to reproduce the historical averages and volatilities of the excess return, risk-free rare and consumption growth, the low covariance between stock return and consumption growth, the countercyclicality of the risk premium, and the procyclicality of the risk-free rate. Second, it was possible to estimate the EIS from an Euler equation in which such a parameter was separably identified. Related to the second result, a novel and simple two-step estimation procedure for the EIS was proposed.

The developed model was restricted to a single risky asset and a risk-free security. This was enough to address the proposed questions. From the present discussion, it is not clear

how one could extend the model to allow for $n > 1$ risky assets in order to study the cross-section of the returns. This is an interesting topic for future research.

A pure quantile maximizer agent is probably not a good representation for general behavior towards risk. Given that, the present model should be understood as a stylized and parsimonious study within the class of models that use asymmetric preferences over good and bad outcomes (as in prospect theory and disappointment aversion). As such, this study makes an important contribution to the literature. Given its ability to explain the financial puzzles parsimoniously, it (i) offers a simpler view regarding the relationship between asymmetric preferences and financial data, and (ii) provides evidence that the good empirical results obtained by the studies employing asymmetric preferences are not due to over-fitting.

Appendix

Proof of Proposition 1:

Substituting the restrictions into the object function, the problem is given by

$$\underset{\xi\in\mathbb{R}}{Max}\ Q_t^\tau\left(u\left(W_t - P_t\xi - P_t^f\xi^f\right) + \beta u\left(X_{t+1}\xi + X_{t+1}^f\xi^f\right)\right)$$

By the quantile equivariance, this is equivalent to

$$\underset{\xi\in\mathbb{R}}{Max}\ u\left(W_t - P_t\xi - P_t^f\xi^f\right) + \beta u\left(\xi Q_t^\tau\left(X_{t+1}\right) + X_{t+1}^f\xi^f\right)$$

and the first order conditions are

$$\begin{aligned}
\xi &: u'\left(C_t\right)P_t = \beta u'\left(Q_t^\tau\left(C_{t+1}\right)\right)Q_t^\tau\left(X_{t+1}\right)\\
\xi^f &: u'\left(C_t\right)P_t^f = \beta u'\left(Q_t^\tau\left(C_{t+1}\right)\right)X_{t+1}^f
\end{aligned}$$

which implies

$$\begin{aligned}
P_t &= \beta\frac{u'\left(Q_t^\tau\left(C_{t+1}\right)\right)}{u'\left(C_t\right)}Q_t^\tau\left(X_{t+1}\right)\\
P_t^f &= \beta\frac{u'\left(Q_t^\tau\left(C_{t+1}\right)\right)}{u'\left(C_t\right)}X_{t+1}^f
\end{aligned}$$

Specializing $u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$,

$$\begin{aligned} P_t &= \beta \left(Q_t^\tau \left(\frac{C_{t+1}}{C_t} \right) \right)^{-\gamma} Q_t^\tau (X_{t+1}) \\ P_t^f &= \beta \left(Q_t^\tau \left(\frac{C_{t+1}}{C_t} \right) \right)^{-\gamma} X_{t+1}^f \end{aligned}$$

CQFD. ■

Proof of Proposition 2:

The risky asset and risk-free asset prices are given, respectively, by

$$\begin{aligned} P_t &= \eta_t Q_t^\tau (X_{t+1}) \\ P_t^f &= \eta_t X_{t+1}^f \end{aligned} \tag{29}$$

where $\eta_t \equiv \beta \left(Q_t^\tau \left(\frac{C_{t+1}}{C_t} \right) \right)^{-\gamma}$.

An arbitrage opportunity occurs if and only if it is possible to construct $\Xi_t = \left(\xi_t, \xi_t^f \right)$ such that

$$\begin{aligned} \xi_t P_t + \xi_t^f P_t^f &= 0 \\ \xi_t X_{t+1} + \xi_t^f X_{t+1}^f &\geq 0 \end{aligned} \tag{30}$$

with the second equation holding as an inequality for at least one point in the support of

X_{t+1}.

Substituting (29) into the first equation of (30),

$$\begin{aligned} & \xi_t \eta_t Q_t^\tau \left(X_{t+1}\right) + \xi_t^f \eta_t X_{t+1}^f = 0 \\ \Rightarrow \quad & \xi_t^f X_{t+1}^f = -\xi_t Q_t^\tau \left(X_{t+1}\right) \end{aligned}$$

which, into the second equation of (30) gives the necessary and sufficient condition for arbitrage,

$$\xi_t \left(X_{t+1} - Q_t^\tau \left(X_{t+1}\right)\right) \geq 0$$

with inequality for at least one point in the support of X_{t+1}.

Therefore, all we need to rule out arbitrage is to impose

$$Q_t^\tau \left(X_{t+1}\right) \in \left(\min\left\{supp\left(X_{t+1}\right)\right\}, \max\left\{supp\left(X_{t+1}\right)\right\}\right)$$

If X_{t+1} is a continuous random variable, this is implied by imposing $\tau \in (0,1)$, CQFD.

■

Proof of Proposition 3:

First, note that if $\ln(x) \sim N(\mu, \sigma^2)$ then $Q^\tau(x) = \exp\left(\mu + \sigma\Phi^{-1}(\tau)\right)$. This holds since

$$
\begin{aligned}
F_X(x) &= \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \\
\Rightarrow \quad F_X^{-1}(\tau) &= \exp\left(\mu + \sigma\Phi^{-1}(\tau)\right)
\end{aligned}
$$

According to (6),

$$
\begin{aligned}
\log\left(C_{t+1}/C_t\right)|t &\sim N\left(\mu_c, \sigma_c^2\right) \\
\log\left(R_{t+1}\right)|t &\sim N\left(\mu_r, \sigma_r^2\right)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
Q_t^\tau\left(C_{t+1}/C_t\right) &= \exp\left(\mu_c + \sigma_c\Phi^{-1}(\tau)\right) \qquad (31) \\
Q_t^\tau\left(R_{t+1}\right) &= \exp\left(\mu_r + \sigma_r\Phi^{-1}(\tau)\right)
\end{aligned}
$$

Dividing both sides of (4) and (5) by P_t, and using the quantile equivariance property,

$$
\begin{aligned}
1 &= \beta\left(Q_t^\tau\left(\frac{C_{t+1}}{C_t}\right)\right)^{-\gamma} Q_t^\tau\left(R_{t+1}\right) \qquad (32) \\
1 &= \beta\left(Q_t^\tau\left(\frac{C_{t+1}}{C_t}\right)\right)^{-\gamma} R_{t+1}^f \qquad (33)
\end{aligned}
$$

where $R_{t+1} = \frac{X_{t+1}}{P_t}$.

Substituting (31) into (32) and taking logs from both sides,

$$\log(\beta) - \gamma\mu_c - \gamma\sigma_c\Phi^{-1}(\tau) + \mu_r + \sigma_r\Phi^{-1}(\tau) = 0$$

Hence, since $E_t(r_{t+1}) = \mu_r$,

$$E_t(r_{t+1}) = -\log(\beta) + \gamma\mu_c + \Phi^{-1}(\tau)(\gamma\sigma_c - \sigma_r)$$

For the risk-free rate, using (33) and (31) in the same way,

$$r^f_{t+1} = -\log(\beta) + \gamma\mu_c + \Phi^{-1}(\tau)\gamma\sigma_c$$

Therefore,

$$E_t\left(r_{t+1} - r^f_{t+1}\right) = -\sigma_r\Phi^{-1}(\tau)$$

CQFD. ■

Proof of Proposition 4:

As in the proof of Proposition 3, we use the fact that if $\ln(x) \sim N(\mu, \sigma^2)$ then $Q^\tau(x) =$

$\exp(\mu + \sigma\Phi^{-1}(\tau))$. Given that,

$$\begin{aligned} Q_t^\tau(C_{t+1}/C_t) &= \exp\left(\mu_c + \sigma_t\Phi^{-1}(\tau)\right) \\ Q_t^\tau(R_{t+1}) &= \exp\left(\mu_r + \varphi\sigma_t\Phi^{-1}(\tau)\right) \end{aligned} \tag{34}$$

Hence, using (32),

$$\ln\beta - \gamma\mu_c - \gamma\sigma_t\Phi^{-1}(\tau) + \mu_r + \varphi\sigma_t\Phi^{-1}(\tau) = 0$$

and, since $E_t(r_{t+1}) = \mu_r$, we have

$$E_t(r_{t+1}) = -\ln\beta + \gamma\mu_c + (\gamma - \varphi)\sigma_t\Phi^{-1}(\tau)$$

For the risk-free rate, using (33) and the conditional quantile for consumption growth,

$$r_{t+1}^f = -\ln\beta + \gamma\mu_c + \gamma\sigma_t\Phi^{-1}(\tau)$$

Therefore,

$$E_t\left(r_{t+1} - r_{t+1}^f\right) = -\varphi\sigma_t\Phi^{-1}(\tau)$$

CQFD. ■

Proof of Proposition 5:

$$
\begin{aligned}
& E\left[\left(\tau_0 - 1\left[C_{t+1}/C_t < \left(\beta_0 R^f_{t+1}\right)^{\psi_0}\right]\right) Z_t | \mathcal{F}_t\right] \\
= & \left(\tau_0 - E\left[1\left[C_{t+1}/C_t < \left(\beta_0 R^f_{t+1}\right)^{\psi_0}\right] | \mathcal{F}_t\right]\right) Z_t \\
= & \left(\tau_0 - \Pr\left(\varepsilon_{c,t+1} < 0 | \mathcal{F}_t\right)\right) Z_t \\
= & \ 0, \text{ since } Q^{\tau_0}\left(\varepsilon_{c,t+1} | \mathcal{F}_t\right) = 0.
\end{aligned}
$$

Using the same steps, we also get

$$
E\left[\left(\tau_0 - 1\left[R_{t+1} < R^f_{t+1}\right]\right) Z_t | \mathcal{F}_t\right] = 0.
$$

CQFD. ■

Proof of Proposition 6:

We verify the conditions of Theorem 2.6 of Newey and McFadden (1994) - NM below. First, note that the theorem requires $V_{t+1} \equiv \left(R_{t+1},\ C_{t+1}/C_t,\ R^f_{t+1},\ Z_t\right)$ to be iid. However, as the authors point out on page 2133, the iid assumption may be replaced by strictly stationarity and ergodicity. According to Proposition 3.44 in White (2001), strictly stationarity and α-mixing implies ergodicity, so assumption (i) ensures $V_{t+1} \equiv \left(Y_{t+1}, Z_t, R^f_{t+1}\right)$ to be strictly stationary and ergodic.

(NM 2.6.i) This is the condition that ensures global identification (see lemma 2.3 in NM).

However, if instead of W_0 being positive semi-definite one imposes W_0 to be positive definite, NF2.6.i can be trivially exchanged for $E\left[g\left(V_{t+1},\theta\right)\right]=0$ if and only if $\theta=\theta_0$. By assumption (iv) $W_0>0$ (a choice for W_T that satisfies this will be provided). So, we have to show that $E\left[g\left(V_{t+1},\theta\right)\right]=0$ if and only if $\theta=\theta_0$.

The fact that $E\left[g\left(V_{t+1},\theta_0\right)\right]=0$ was already derived in the body of the text. We are left to show that $E\left[g\left(V_{t+1},\theta\right)\right]=0\Rightarrow\theta=\theta_0$.

First, considering the second set of moment conditions,

$$\begin{aligned}
& E\left[\left(\tau-1\left[R_{t+1}\leq R_{t+1}^{f}\right]\right)Z_t\right] \\
= \; & E\left[\left(\tau-E\left[1\left[R_{t+1}\leq R_{t+1}^{f}\right]|Z_t\right]\right)Z_t\right] \\
= \; & E\left[\left(\tau-F_{R_{t+1}|Z_t}\left(R_{t+1}^{f}|Z_t\right)\right)Z_t\right] \\
= \; & E\left[\left(\tau-\tau_0\right)Z_t\right],\ \text{since } R_{t+1}^{f}\in Z_t \text{ by assumption (vi)} \\
= \; & 0\Rightarrow\tau=\tau_0,\ \text{since } 1\in Z_t \text{ by assumption (vi)}
\end{aligned}$$

Hence, τ_0 is identified. We now consider the first set of moment conditions (with τ_0 already identified):

$$
\begin{aligned}
& E\left[\left(\tau_0 - 1\left[C_{t+1}/C_t < \left(\beta R^f_{t+1}\right)^{\psi}\right]\right) Z_t\right] \\
= \;& E\left[\left(\tau_0 - E\left[1\left[C_{t+1}/C_t < \left(\beta R^f_{t+1}\right)^{\psi}\right] | Z_t\right]\right) Z_t\right] \\
= \;& E\left[\left(\tau_0 - F_{(C_{t+1}/C_t)|Z_t}\left(\left(\beta R^f_{t+1}\right)^{\psi} | Z_t\right)\right) Z_t\right] \\
= \;& 0 \Rightarrow F_{(C_{t+1}/C_t)|Z_t}\left(\left(\beta R^f_{t+1}\right)^{\psi} | Z_t\right) = \tau_0, \text{ since } 1 \in Z_t \text{ by assumption (vi)}
\end{aligned}
$$

By assumption (v), $F_{(C_{t+1}/C_t)|Z_t}$ is a continuous strictly increasing function within its support. By assumption (vi), $R^f_{t+1} \in Z_t$ and hence

$$F_{(C_{t+1}/C_t)|Z_t}\left(\left(\beta_0 R^f_{t+1}\right)^{\psi_0} | Z_t\right) = \tau_0.$$

Therefore, we must have

$$\left(\beta_0 R^f_{t+1}\right)^{\psi_0} = \left(\beta R^f_{t+1}\right)^{\psi},$$

which holds if either

$$(\beta, \psi) = (\beta_0, \psi_0)$$

or

$$R_{t+1}^{f} = \frac{\psi_0 \log(\beta_0) - \psi \log(\beta)}{\psi - \psi_0} \text{ at every } t.$$

By assumption (vii), R_{t+1}^{f} is a non-degenerate random variable. Hence,

$$(\beta, \psi) = (\beta_0, \psi_0) \text{ a.s.}$$

Therefore, we conclude

$$E[g(V_{t+1}, \theta)] = 0 \Rightarrow \theta = \theta_0 \text{ a.s.}$$

(NM 2.6.ii) Assumption (iii) ensures θ_0 as an interior point of Θ.

(NM 2.6.iii) This is satisfied because $g(V_{t+1}, \theta)$ is discontinuous only when $\frac{C_{t+1}}{C_t} = \left(\beta R_{t+1}^{f}\right)^{1/\gamma}$ and $R_{t+1} = R_{t+1}^{f}$. By assumption (v), these two cases have probability zero.

(NM 2.6.iv) Note that since for any value of θ we have

$$\begin{aligned} \left\| \left(\tau - 1\left[\frac{C_{t+1}}{C_t} < \left(\beta R_{t+1}^{f}\right)^{\psi} \right] \right) Z_t \right\| &\leq \|Z_t\| \\ \left\| \left(\tau - 1\left[R_{t+1} < R_{t+1}^{f} \right] \right) Z_t \right\| &\leq \|Z_t\|, \end{aligned}$$

we ensure $E\left(\sup_{\theta \in \Theta} \|g(V_{t+1}, \theta)\|\right) < \infty$ by assumption (ii).

Therefore, we conclude that $\widehat{\theta} \xrightarrow{p} \theta_0$, by Theorem 2.6 of Newey and McFadden (1994),

CQFD. ■

Proof of Proposition 7: By Lemma 1 below, specializing $\theta = \widetilde{\theta}$, we have $W_T \xrightarrow{p} \Sigma_0^{-1}$ where

$$\Sigma_0 \equiv E\left[g\left(V_{t+1}, \theta_0\right) g\left(V_{t+1}, \theta_0\right)'\right].$$

Now, we prove that Σ_0 is a positive definite matrix (since every positive definite matrix is invertible and its inverse is also positive definite, we then are done: Σ_0^{-1} exists and is positive definite.) First, note that

$$\Sigma_0 = E\left[E\left(A_{t+1}|Z_t\right) \otimes Z_t Z_t'\right]$$

where A_{t+1} is a 2×2 matrix with entries

$$\begin{aligned}
A_{11} &= \left(\tau_0 - 1\left[\frac{C_{t+1}}{C_t} < Q_t^{\tau_0}\left(\frac{C_{t+1}}{C_t}\right)\right]\right)^2 \\
A_{12} &= A_{21} = \left(\tau_0 - 1\left[\frac{C_{t+1}}{C_t} < Q_t^{\tau_0}\left(\frac{C_{t+1}}{C_t}\right)\right]\right)\left(\tau_0 - 1\left[R_{t+1} < Q_t^{\tau_0}\left(R_{t+1}\right)\right]\right) \\
A_{22} &= \left(\tau_0 - 1\left[R_{t+1} < Q_t^{\tau_0}\left(R_{t+1}\right)\right]\right)^2,
\end{aligned}$$

under the theoretical model.

We now compute $E\left(A_{11}|Z_t\right)$, $E\left(A_{12}|Z_t\right)$, $E\left(A_{21}|Z_t\right)$ and $E\left(A_{22}|Z_t\right)$.

$$E\left(A_{11}|Z_t\right) = E\left(A_{22}|Z_t\right) = \tau_0^2\left(1-\tau_0\right) + \left(\tau_0 - 1\right)^2\tau_0 = \tau_0\left(1-\tau_0\right)$$

and

$$\begin{aligned}
E\left(A_{12}|Z_t\right) &= E\left(A_{21}|Z_t\right) \\
&= E\left[\left(\tau_0 - 1\left[\frac{C_{t+1}}{C_t} < Q_t^{\tau_0}\left(\frac{C_{t+1}}{C_t}\right)\right]\right)|Z_t\right] E\left[\left(\tau_0 - 1\left[R_{t+1} < Q_t^{\tau_0}\left(R_{t+1}\right)\right]\right)|Z_t\right] \\
&\quad + Cov\left[\tau_0 - 1\left[\frac{C_{t+1}}{C_t} < Q_t^{\tau_0}\left(\frac{C_{t+1}}{C_t}\right)\right], \tau_0 - 1\left[R_{t+1} < Q_t^{\tau_0}\left(R_{t+1}\right)\right]|Z_t\right] \\
&= \left(\tau_0\left(1-\tau_0\right) + \left(\tau_0 - 1\right)\tau_0\right)\left(\tau_0\left(1-\tau_0\right) + \left(\tau_0 - 1\right)\tau_0\right) \\
&\quad + Cov\left[\tau_0 - 1\left[\frac{C_{t+1}}{C_t} < Q_t^{\tau_0}\left(\frac{C_{t+1}}{C_t}\right)\right], \tau_0 - 1\left[R_{t+1} < Q_t^{\tau_0}\left(R_{t+1}\right)\right]|Z_t\right] \\
&= Cov\left[\tau_0 - 1\left[\frac{C_{t+1}}{C_t} < Q_t^{\tau_0}\left(\frac{C_{t+1}}{C_t}\right)\right], \tau_0 - 1\left[R_{t+1} < Q_t^{\tau_0}\left(R_{t+1}\right)\right]|Z_t\right] \\
&= Cov\left[1\left[\frac{C_{t+1}}{C_t} < Q_t^{\tau_0}\left(\frac{C_{t+1}}{C_t}\right)\right], 1\left[R_{t+1} < Q_t^{\tau_0}\left(R_{t+1}\right)\right]|Z_t\right] \\
&= P\left(\varepsilon_{c,t+1} < 0, \varepsilon_{r,t+1} < 0|Z_t\right) - P\left(\varepsilon_{c,t+1} < 0|Z_t\right)P\left(\varepsilon_{r,t+1} < 0|Z_t\right) \\
&= \varphi_t - \tau_0^2, \text{ for } \varphi_t \equiv P\left(\varepsilon_{c,t+1} < 0, \varepsilon_{r,t+1} < 0|Z_t\right).
\end{aligned}$$

Therefore, $E\left(A_{t+1}|Z_t\right)$ is positive definite if both the following conditions hold,

$$\begin{aligned} \tau_0 (1 - \tau_0) &> 0 \\ \tau_0^2 (1 - \tau_0)^2 - \left(\varphi_t - \tau_0^2\right)^2 &> 0. \end{aligned}$$

The first condition is ensured by assumption (viii). The second condition can be simplified further,

$$\begin{aligned} \tau_0^2 (1 - \tau_0)^2 &> \left(\varphi_t - \tau_0^2\right)^2 \\ \left[\tau_0 (1 - \tau_0)\right]^2 &> \left(\varphi_t - \tau_0^2\right)^2 \\ \left(\tau_0 - \tau_0^2\right)^2 &> \left(\varphi_t - \tau_0^2\right)^2 \\ \tau_0 - \tau_0^2 &> \varphi_t - \tau_0^2 \\ \varphi_t &< \tau_0, \end{aligned}$$

which is assumption (ix).

With respect to $Z_t Z_t'$ we can also show that it is positive definite. In fact, for any $\lambda \in \mathbb{R}^m$,

$$\lambda' Z_t Z_t' \lambda = \left(Z_t' \lambda\right)^2 \geq 0,$$

holding with inequality only if $Z_t' \lambda = 0$. But, given assumption (x), $Z_t' \lambda = 0$ only if $\lambda = 0$.

Therefore, since both $E\left[A_{t+1}|Z_t\right]$ and $Z_t Z_t'$ are positive definite, $E\left[A_{t+1}|Z_t\right] \otimes Z_t Z_t'$ is positive definite and Σ_0 is also positive definite, CQFD. ■

Proof of Proposition 8:

First, an observation:

Even though $g\left(V_{t+1}, \theta\right)$ is not differentiable in θ, $E\left[g\left(V_{t+1}, \theta\right)\right]$ is. In fact, for

$$g_1\left(V_{t+1}, \theta\right) \equiv \left(\tau - 1\left[C_{t+1}/C_t < \left(\beta R_{t+1}^f\right)^{\psi}\right]\right) Z_t$$

and

$$g_2\left(V_{t+1}, \theta\right) \equiv \left(\tau - 1\left[R_{t+1} < R_{t+1}^f\right]\right) Z_t$$

we have:

$$\partial_\tau E\left[g_1\left(V_{t+1}, \theta\right)\right] = E\left(Z_t\right)$$

$$\begin{aligned}
\partial_\beta E\left[g_1\left(V_{t+1},\theta\right)\right] &= -\partial_\beta E\left[1\left[C_{t+1}/C_t < \left(\beta R_{t+1}^f\right)^\psi\right] Z_t\right] \\
&= -\partial_\beta E\left[E\left[1\left[C_{t+1}/C_t < \left(\beta R_{t+1}^f\right)^\psi\right]|Z_t\right] Z_t\right] \\
&= -\partial_\beta E\left[F_{(C_{t+1}/C_t)|Z_t}\left(\left(\beta R_{t+1}^f\right)^\psi|Z_t\right) Z_t\right] \\
&= -E\left[\partial_\beta F_{(C_{t+1}/C_t)|Z_t}\left(\left(\beta R_{t+1}^f\right)^\psi|Z_t\right) Z_t\right] \\
&= -E\left[f_{(C_{t+1}/C_t)|Z_t}\left(\left(\beta R_{t+1}^f\right)^\psi|Z_t\right)\psi\beta^{(\psi-1)}\left(R_{t+1}^f\right)^\psi Z_t\right] \\
&= -E\left[f_{\varepsilon_{c,t+1}}\left(0|Z_t\right)\psi_0\beta_0^{(\psi_0-1)}\left(R_{t+1}^f\right)^{\psi_0} Z_t\right], \text{ for } \theta=\theta_0.
\end{aligned}$$

$$\begin{aligned}
\partial_\psi E\left[g_1\left(V_{t+1},\theta\right)\right] &= -E\left[\partial_\psi F_{C_{t+1}/C_t}\left(\left(\beta R_{t+1}^f\right)^\psi|Z_t\right) Z_t\right] \\
&= -E\left[f_{\varepsilon_{c,t+1}}\left(0|Z_t\right)\left(\beta_0 R_{t+1}^f\right)^{\psi_0}\log\left(\beta_0 R_{t+1}^f\right) Z_t\right], \text{ for } \theta=\theta_0.
\end{aligned}$$

$$\partial_\tau E\left[g_2\left(V_{t+1},\theta\right)\right] = E\left(Z_t\right)$$

$$\partial_\beta E\left[g_2\left(V_{t+1},\theta\right)\right] = 0$$

$$\partial_{\psi} E\left[g_2\left(V_{t+1}, \theta\right)\right] = 0$$

Given that, define $G_0 = \nabla_{\theta} E\left[g\left(V_{t+1}, \theta_0\right)\right]$, where $\nabla_{\theta} E\left[g\left(V_{t+1}, \theta_0\right)\right]$ is the $2m \times 3$ matrix derived above.

(end of observation)

We now check conditions (i) to (v) from Theorem 7.2 of Newey and McFadden (1994) to establish the asymptotic normality of our estimator.

(NF.7.2.i) $E\left[g\left(V_{t+1}, \theta_0\right)\right] = 0$ is shown in the body of the text.

(NF.7.2.ii) The fact that $E\left[g\left(V_{t+1}, \theta\right)\right]$ is differentiable at θ_0 was shown in the observation in the beginning of the proof. $G_0' W_0 G_0$ is nonsingular by assumption (xii).

(NF.7.2.iii) Assumption (iii) ensures θ_0 as an interior point of Θ.

(NF.7.2.iv) We know that $\left\{g\left(V_{t+1}, \theta_0\right), \mathcal{F}_t\right\}$ is a martingale difference sequence. Given that, we check the conditions of Corollary 5.26 in White's (2001). We have

$$\begin{aligned} E\left\|g\left(V_{t+1}, \theta_0\right)\right\|^{2+2\delta} &\leq E\left\|Z_t\right\|^{2+2\delta} \\ &\leq \max\left\{1, E\left\|Z_t\right\|^{2r+2\delta}\right\}, \text{ where } r > 2 \\ &\leq \infty \text{ by assumption (ii')}. \end{aligned}$$

Moreover, applying Lemma 1 below for $\theta = \theta_0$ we have

$$\frac{1}{T}\sum_{t=1}^{T} g\left(V_{t+1},\theta_0\right) g\left(V_{t+1},\theta_0\right)' \xrightarrow{p} \Sigma_0,$$

where $\Sigma_0 \equiv E\left[g\left(V_{t+1},\theta_0\right) g\left(V_{t+1},\theta_0\right)'\right]$.

Therefore, according to Corollary 5.26 in White(2001),

$$\sqrt{T}\left(\frac{1}{T} g\left(V_{t+1},\theta_0\right)\right) \xrightarrow{d} N\left(0,\Sigma_0\right).$$

(NF.7.2.v) Andrews (1994) shows that empirical processes defined from moment conditions as $g\left(V_{t+1},\theta_0\right)$ are stochastically equicontinuous ($g\left(V_{t+1},\theta_0\right)$ fits in what he calls *type I class* of real functions - note that even though $g_1\left(V_{t+1},\theta_0\right)$ has a nonlinear function of the parameters inside the indicator function,

$$g_1\left(V_{t+1},\theta\right) = \left(\tau - 1\left[C_{t+1}/C_t < \left(\beta R^f_{t+1}\right)^{\psi}\right]\right) Z_t$$

this can be written as,

$$g_1\left(V_{t+1},\theta\right) = \left(\tau - 1\left[\log C_{t+1}/C_t < \psi \log\beta + \psi \log R^f_{t+1}\right]\right) Z_t$$

given that the log is a strictly increasing function and $C_{t+1}/C_t, \beta, R^f_{t+1} > 0$).

Therefore, by Theorem 7.2 of Newey and McFadden (1994), we conclude that

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right) \xrightarrow{d} N\left(0,\left(G_0'W_0G_0\right)^{-1}G_0'W_0\Sigma_0W_0G_0'\left(G_0'W_0G_0\right)^{-1}\right),$$

CQFD. ■

Lemma 1: Define $\Sigma(\theta) = E\left[g(V_{t+1},\theta)\,g(V_{t+1},\theta)'\right]$. Then,

$$\frac{1}{T}\sum_{t=1}^{T} g(V_{t+1},\theta)\,g(V_{t+1},\theta)' \xrightarrow{p} \Sigma(\theta).$$

Proof of Lemma 1: First, note that $g(V_{t+1},\theta)$ is an $\mathcal{F}_{t+1}$ measurable function which is strictly stationary and α-mixing what implies that $g(V_{t+1},\theta)\,g(V_{t+1},\theta)'$ is also strictly stationary and α-mixing of the same size (Theorem 3.49 of White (2001)).

Now, all we need is to apply a Law of Large Numbers for α-mixing sequences (Corollary 3.48 of White (2001)). The conditions of White's corollary are (a) $\left\{g(V_{t+1},\theta)\,g(V_{t+1},\theta)'\right\}$ has to be an α-mixing sequence of size $-r/(r-1)$, $r>1$ and (b) $E\left\|g(V_{t+1},\theta)\,g(V_{t+1},\theta)'\right\|^{r+\delta} < \infty$ for some $\delta>0$, where $\|\cdot\|$ denotes the L_∞-norm. Condition (a) is directly satisfied by assumption (i). For condition (b), note that

$$
\begin{aligned}
& \left\| g\left(V_{t+1},\theta\right) g\left(V_{t+1},\theta\right)' \right\| \\
\equiv\ & \left| g_{i_0}\left(V_{t+1},\theta\right) g_{j_0}\left(V_{t+1},\theta\right) \right|, \text{ where } \left(i_0,j_0\right) = \underset{i\geq 1, j\leq \dim(g)}{\arg\max} \left| g_i\left(V_{t+1},\theta\right) g_j\left(V_{t+1},\theta\right) \right| \\
=\ & \left| g_{i0}\left(V_{t+1},\theta\right) \right| \left| g_{j0}\left(V_{t+1},\theta\right) \right| \\
\leq\ & C^2 \left\| g\left(V_{t+1},\theta\right) \right\|^2, \quad \text{by norm equivalence,for some positive constant } C,
\end{aligned}
$$

and hence

$$
E \left\| g\left(V_{t+1},\theta\right) g\left(V_{t+1},\theta\right)' \right\|^{r+\delta} \leq C^2 \max\left\{1, E \left\| g\left(V_{t+1},\theta\right) \right\|^{2r+2\delta}\right\}
$$

by Cauchy-Schwarz. So, we would need some assumption such as "there exist some $\delta > 0$ such that $E \left\| g\left(V_{t+1},\theta\right) \right\|^{2r+2\delta} < \infty$". However, note that

$$
\begin{aligned}
\left\| \left(\tau - 1\left[\frac{C_{t+1}}{C_t} < \left(\beta R^f_{t+1}\right)^{\psi} \right] \right) Z_t \right\| &\leq \left\| Z_t \right\| \\
\left\| \left(\tau - 1\left[R_{t+1} < R^f_{t+1} \right] \right) Z_t \right\| &\leq \left\| Z_t \right\|
\end{aligned}
$$

and, therefore, it is enough to assume that there exist some $\delta > 0$ such that $E \left\| Z_t \right\|^{2r+2\delta} < \infty$, which is our assumption (ii') CQFD. ■

Model estimation under lognormality

The solved model under lognormality and stochastic economic uncertainty is given by

$$r_{t+1} = -\ln\beta_0 + \frac{1}{\psi_0}\mu_c + \left(\frac{1}{\psi_0} - \varphi_0\right)\sigma_t\Phi^{-1}(\tau_0) + \varphi_0\sigma_t u_{t+1} \tag{35}$$

$$r^f_{t+1} = -\ln\beta_0 + \frac{1}{\psi_0}\mu_c + \frac{1}{\psi_0}\sigma_t\Phi^{-1}(\tau_0) \tag{36}$$

$$g_{t+1} = \mu_c + \sigma_t\eta_{t+1} \tag{37}$$

$$\sigma^2_{t+1} = \alpha_0 + \rho_0\left(\sigma^2_t - \alpha_0\right) + \sigma_v v_{t+1} \tag{38}$$

in accordance to section II.B.

A possible estimator for the parameters is the simulated method of moments (SMM) of McFadden (1986), Pakes and Pollard (1987), and Duffie and Singleton (1993), the last one in the context of time-series as we have here.

Analogous to sub-section II.B, we focus only on the estimation of $\theta_0 = (\beta_0, \psi_0, \tau_0)$, fixing the dynamics parameters using the values in Table 1.

Define m_t to be a $p \times 1$ vector of empirical observations on variables whose moments are of interest: the risk-free rate, the excess return, and the consumption growth. Such a vector should contain the moments to be matched by the estimator. In our case, $p = 6$ and

$$m_t = \left(r_t - r^f_t, \left(r_t - r^f_t\right)^2, r^f_t, \left(r^f_t\right)^2, g_t, g^2_t\right).$$

Define $m_t(\theta)$ to be a $p \times 1$ vector with the synthetic counterpart of m_t, whose elements are computed on the basis of artificial data generated by the model using parameter values θ. The number of observations in the artificial time series is given by κT, where T is the sample size and κ is a positive integer.

The SMM estimator of θ_0 is defined as

$$\widehat{\theta}_{SMM} = \underset{\theta \in \Theta \subseteq \mathbb{R}^3}{\arg\min} \left(\frac{1}{T} \sum_{t=1}^{T} m_t - \frac{1}{\kappa T} \sum_{t=1}^{\kappa T} m_t(\theta) \right)' \left(\frac{1}{T} \sum_{t=1}^{T} m_t - \frac{1}{\kappa T} \sum_{t=1}^{\kappa T} m_t(\theta) \right)$$

where, to allow for a direct comparison with the simulation results from section II.B, each moment is equally weighted.

Under the regularity conditions of Duffie and Singleton (1993),

$$\sqrt{T}\left(\widehat{\theta}_{SMM} - \theta_0\right) \xrightarrow{d} N\left(0, (1 + 1/\kappa)\left(D_0' D_0\right)^{-1} D_0' \Omega_0 D_0 \left(D_0' D_0\right)^{-1}\right),$$

where,

$$D_0 = E\left(\partial_\theta m_t(\theta)\,|_{\theta=\theta_0}\right)$$

and

$$\Omega_0 = \sum_{j=-\infty}^{\infty} E\left(\left(m_t - E\left[m_t\right]\right)\left(m_{t-j} - E\left[m_{t-j}\right]\right)'\right).$$

As usual, Ω_T can be obtained by the Newey-West estimator. With respect to D_0, since there is no analytical solution for the differentiation, the derivatives are numerically computed, and the expectation approximated by the average over the κT simulated points.

Under this framework, by drawing monthly observations, aggregating them to yearly, and constructing m_t from the same data used in section III.C (also yearly-aggregated), we end up with the following estimates:

# of draws:	12×10^3	12×10^4	12×10^5
β	1.001	1.001	1.001
(se)	(0.001)	(0.001)	(0.001)
EIS	0.61	0.59	0.61
(se)	(0.25)	(0.17)	(0.19)
τ	0.47	0.46	0.46
(se)	(0.02)	(0.02)	(0.02)

As in the simulation exercise we assume that the decision interval of the agent is monthly but the targeted data to match are annual. Therefore, we simulate at the monthly frequency but match the yearly moments.

These results are in line with the calibrated values in section II.B.

References

Ang, A., Bekaert, G., and Liu, J. (2005). "Why stocks may disappoint." *Journal of Financial Economics*, 76 (3): 471-508.

Bansal, R., and Yaron, A. (2004). "Risks for the long run: a potential resolution of asset pricing puzzles." Journal of Finance 59 (4): 1481–1509.

Bansal, R., Khatchatrian, V., and Yaron, A. (2005). "Interpretable asset markets?" *European Economic Review* 49 (3): 531-560.

Bansal, R., Kiku, D., and Yaron, A. (2009). "An empirical evaluation of the long-run risks model for asset prices.", working paper

Barberis, N., Huang, M., and Santos, T. (2001). "Prospect theory and asset prices." *The Quarterly Journal of Economics,* 116 (1): 1-53.

Basset, G., Koenker, R., and Kordas, G. (2004). "Pessimistic portfolio allocation and choquet expected utility." *Journal of Financial Econometrics*, 2(4): 477-492.

Bekaert, G., Hodrick, R., and Marshall, D. (1997). "The implications of first-order risk aversion for asset market risk premiums", *Journal of Monetary Economics* 40: 3-39.

Benartzi, S., and Thaler, R.H. (1995). "Myopic loss aversion and the equity premium puzzle". *The Quarterly Journal of Economics,* 110 (1): 73-92

Bonomo, M., and Garcia, R. (1993). "Disappointment aversion as a solution to the equity premium and the risk-free rate puzzles." Working paper, Université de Montréal.

Breeden, D., Gibbons, M., and Litzenberger, R. (1989). "Empirical tests of the consumption-oriented CAPM." *Journal of Finance,* 44: 231-262.

Campbell, J. (1996). "Understanding risk and return." *Journal of Political Economy,* 104: 298-345.

——— (1999). "Asset prices, consumption, and the business cycle." in John Taylor and Michael Woodford (eds.), *Handbook of Macroeconomics*, North-Holland, Amsterdam.

——— (2003). "Consumption-based asset pricing." Chapter 13 in G. Constantinides, M. Harris, and R. Stulz (eds.), *Handbook of the Economics of Finance* Vol. IB., North-Holland, Amsterdam.

Campbell, J., Lo, A., and Mackinlay, A.C. (1997). "The econometrics of financial markets." Princeton University Press.

Chernozhukov, V., and Hong, H. (2003). "An MCMC approach to classical estimation." *Journal of Econometrics,* 115: 293–346.

Cochrane, J. (1996). "A cross-sectional test of an investment-based asset pricing model." *Journal of Political Economy*, 104: 572-621.

——— (1997). "Where is the market going? Uncertain facts and novel theories." *Economic Perspectives Federal Reserve Bank of Chicago,* 21, 6.

——— (2006). "Financial markets and real economies." *International Library of Critical Writings in Financial Economics,* volume 18, London: Edward Elgar.

Constantinides, G. (1990). "Habit formation: a resolution of the equity premium puzzle." *Journal of Political Economy*, 98 (3): 519-43.

Cooper, I., and Priestley, R. (2009). "Time-varying risk premiums and the output gap." *Review of Financial Studies*, 22(7): 2801-2833.

Donaldson, J., and Mehra, R. (2008a). "Risk-based explanations of the equity premium". In Mehra, R. (editor), *Handbook of the Equity Risk Premium*: 37-94.

Donaldson, J., and Mehra, R. (2008a). "Non-risk-based explanations of the equity premium". In Mehra, R. (editor), *Handbook of the Equity Risk Premium*: 101-114.

Duffie, D., and Singleton, K. (1993). "Simulated moments estimation of markov models of asset prices". *Econometrica,* 61: 929-952.

Engelhardt, G., and Kumar, A. (2009). "The elasticity of intertemporal substitution: new evidence from 401(k) participation". *Economics Letters,* 103(1): 15-17.

Epstein, L., and Zin, S. (1989). "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework". *Econometrica,* 57: 937-969.

——— (1990). "First-order risk aversion and the equity premium puzzle", *Journal of Monetary Economics*, 26: 387-407.

——— (1991). "Substitution, risk aversion, and temporal behavior of consumption and asset returns II: an empirical analysis", *Journal of Political Economy*, 99: 263-286.

——— (2001). "The independence axiom and asset returns", *Journal of Empirical Finance* 8: 537-572.

Fama, F., and French, K. (1989). "Business conditions and expected returns on stocks and bonds." *Journal of Financial Economics,* 25, 23–49.

Gul, F. (1991). "A theory of disappointment aversion." *Econometrica,* 59 (3): 667-686.

Guvenen, F. (2006). "Reconciling conflicting evidence on the elasticity of intertemporal substitution: a macroeconomic perspective." *Journal of Monetary Economics*, 53 (7): 1451-1472.

Hall, R. (1988). "Intertemporal substitution in consumption." *Journal of Political Economy,* 96, 339-357.

Hansen, L. (1982). "Large sample properties of generalized method of moments estimators." *Econometrica*, 50: 1029-54.

Hansen, L., and Singleton, K.(1982). "Generalized instrumental variables estimation of nonlinear rational expectations models." *Econometrica*, 50, 1269-1286.

——— (1983). "Stochastic consumption, risk aversion and the temporal behavior of asset returns." *Journal of Political Economy*, 91, 249-268.

Heaton, J. (1995). "An empirical investigation of asset pricing with temporally dependent preference specifications." *Econometrica*, 63(3): 681-717.

Jewitt, I. (1987). "Risk aversion and the choice between risky prospects: the preservation of comparative statics results". *Review of Economic Studies*, 54(1): 73-85.

Kahneman, D., and Tversky, A. (1979). "Prospect theory: an analysis of decision under risk." *Econometrica.* 47, 263–291.

Karni, E., and Schmeidler, D. (1991). "Atemporal dynamic consistency and expected utility theory." *Journal of Economic Theory.* 54, 401–408.

Kocherlakota, N. (1996). "The equity premium: it's still a puzzle." *Journal of Economic Literature*, 34: 42-71.

Koenker, R., and Bassett, G. (1978). "Regression quantiles." *Econometrica*, 46(1), 33-50.

Ludvigson, S., and Ng, S. (2009). "Macro factors in bond risk premia." *Review of Financial Studies*, 22(12): 5027-5067.

Mankiw, G., and Shapiro, M. (1986). "Risk and return: consumption beta versus market beta." *Review of Economics and Statistics*, 68: 452-459.

Manski, C. (1988). "Ordinal utility models of decision making under uncertainty." *Theory and Decision*, 25: 79-104.

McFadden, D. (1986). "A method of simulated moments for estimation of discrete response models without numerical integration". *Econometrica*, 57: 995-1026.

Mehra, R., and Prescott,E. C. (1985). "The equity premium: a puzzle." *Journal of Monetary Economics*, 15: 145-61.

Neely, C., Roy, A. and Whiteman, C. "Risk aversion versus intertemporal substitution: a case study of identification failure in the intertemporal consumption capital asset pricing model." *Journal of Business and Economic Statistics*, 19: 395–403.

Newey, W., and McFadden, D. (1994). "Large sample estimation and hypothesis testing." In Engle, R. F., McFadden, D.L. (eds.), *Handbook of Econometrics*, 4: 2113-2247.

Newey, W., and Powell, J. (1987). "Asymmetric least squares estimation and testing." *Econometrica,* 55: 819–847.

Pakes, A., and Pollard, D. (1989). "The asymptotic distribution of simulation experiments." *Econometrica*, 57: 1027-1057.

Powell, J. (1984). "Least absolute deviations estimation for the censored regression model." *Journal of Econometrics*, 25: 303–325.

Powell, J. (1986). "Censored regression quantiles." *Journal of Econometrics*, 32: 143–155.

Rostek, M. (2010). "Quantile maximization in decision theory." *Review of Economic Studies*, 77(1): 339-371.

Routledge, B., and Zin, S. (2010). "Generalized disappointment aversion and asset prices." *The Journal of Finance*, forthcoming.

Rust, J. (2006). "Dynamic programming." *New Palgrave Dictionary of Economics.*

Vissing-Jorgensen, A. (2002). "Limited asset market participation and the elasticity of intertemporal substitution." *Journal of Political Economy*, 110(4): 825-853.

Wald, A. (1939). "Contributions to the theory of statistical estimation and testing hypotheses." *Annals of Mathematical Statistics*, 10(4): 299-326.

Watcher, J. (2002). "Comment on: are behavioral asset-pricing models structural?" *Journal of Monetary Economics,* 49(1): 229-233.

Weil, P. (1989). "The equity premium puzzle and the risk-free rate puzzle." *Journal of Monetary Economics*, 24: 401-421.

White, H. (2001). "Asymptotic theory for econometricians." Academic Press, San Diego.

Yogo, M. (2004). "Estimating the elasticity of intertemporal substitution when instruments are weak." *The Review of Economics and Statistics*, 86(3): 797–810.

Zin, S. (2002). "Are behavioral asset-pricing models structural?" *Journal of Monetary Economics,* 49(1): 215-228.