Review of Finance, 2017, 945–985
doi: 10.1093/rof/rfx010
Advance Access Publication Date: 2 March 2017

OXFORD

Macro-Finance*

John H. Cochrane

Hoover Institution, Stanford University. Also NBER, Stanford GSB and SIEPR, University of Chicago Booth School of Business and Becker-Friedman Institute, and Cato Institute

Abstract

Macro-finance addresses the link between asset prices and economic fluctuations. Many models reflect the same rough idea: the market's ability to bear risk is greater in good times, and less in bad times. Models achieve this similar result by quite different mechanisms. I contrast their strengths and weaknesses. I highlight directions for future research, including additional facts to be matched, and limitations of the models that should prod future theoretical work. I describe how macro-finance models can fundamentally alter macroeconomics, by putting time-varying risk premiums and risk-bearing capacity at the center of recessions rather than variation in the interest rate and intertemporal substitution.

JEL classification: G1, E1.

Keywords: Macro-finance, Equity premium, Volatility

Received December 30, 2016; accepted January 23, 2017 by Editor Alex Edmans.

1. Facts

Macro-finance studies the relationship between asset prices and economic fluctuations. These theories are built on some simple facts.

Asset prices and returns are correlated with business cycles. Stocks rise in good times, and fall in bad times. Real and nominal interest rates rise and fall with the business cycle. Stock returns and bond yields also help to forecast macroeconomic events such as GDP growth and inflation.[1]

Stocks have a substantially higher average return than bonds. Typical estimates put the equity premium between 4% and 8%. Even 4% is puzzling. Why do people not try to hold more stocks, given the power of compound returns to increase wealth dramatically over long horizons?

* This essay is based on a keynote speech at the University of Melbourne 2016 "Finance Down Under" conference. I am grateful to Carole Comerton-Forde, Vincent Gregoire, Bruce Grundy, and Federico Nardari for inviting me. I am grateful to Alex Edmans, Ivo Welch, and an anonymous referee for extensive and thoughtful comments.

1 To save space, I do not provide citations to this extensive literature here. See reviews in Cochrane (2004, 2007, 2011).

The answer is, of course, that stocks are risky. But people accept many risks in life. In lotteries and at casinos they even seek out risks. The answer must be that stocks have a special kind of risk, that stock values fall at particularly inconvenient times or in particularly inconvenient states of nature.

The canonical theory of finance captures this special fear. It starts with the pricing formula

$$0 = E(M_{t+1}R^e_{t+1})$$

or equivalently (as an approximation, and exact in continuous time)

$$E(R^e_{t+1}) = -\mathrm{cov}(M_{t+1}, R^e_{t+1}),$$

where M denotes the stochastic discount factor, or growth of marginal utility, and R^e is an excess return, that is the difference between the returns on two securities.

In this expression, expected returns are high because stocks fall when investors are already hungry—high marginal utility, or high discount factor. Other risks, which investors take more happily, are not correlated with such bad times.

So, just what are the bad times or bad states of nature, in which investors are particularly anxious that their stocks do not fall? Well, something about recessions is an obvious candidate. Losing money in the stock market is especially fearsome if that event tends to happen just as you lose your job, your business is losing money, you may lose your house, and so on.

But what is the feared event exactly? How do we measure that event? And what does this fear that stocks might fall in recessions tell us about the macroeconomics of recessions? These questions are what macro-finance is all about.

The standard power-utility consumption-based model is the simplest macro-finance model:

$$M_{t+1} = e^{-\delta}\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}$$

or

$$E(R^e_{t+1}) = \gamma\,\mathrm{cov}(\Delta c_{t+1}, R^e_{t+1}), \tag{1}$$

where Δc represents consumption growth and γ is the risk aversion coefficient. This model identifies the precise, quantifiable, and measurable feature of recessions that induces fear: consumption falls. (The latter equation is again an approximation in discrete time and exact in the continuous time version of the model.)

But, as crystallized by the equity premium–risk-free rate puzzle (Mehra and Prescott, 1985; Hansen and Jagannathan, 1991), consumption is just not volatile enough to generate the observed equity premium in this model, without very large risk aversion coefficients. From (1),

$$\frac{E(R^e)}{\sigma(R^e)} \leq \gamma\sigma(\Delta c_{t+1}).$$

With market volatility about 16% on an annual basis, and 4%–8% average returns, the Sharpe ratio on the left is 0.25–0.5. Aggregate consumption growth only has a 1%–2% standard deviation on an annual basis, 0.01–0.02. Reconciling these numbers takes a very high degree of risk aversion γ. Therefore, though the sign is right, and consumption is positively correlated with stock returns, this model does not quantitatively answer our motivating question, why are people so afraid of stocks when they do not seem that afraid of other events?

One may accept high risk aversion, at least for the representative agent, but the power-utility model then has trouble with the level of the risk-free rate. This problem is best seen in the continuous time version of the model, where $R_t^f = 1/E_t(M_{t+1})$ becomes

$$r_t^f dt = \delta dt + \gamma E_t\left(\frac{dC}{C}\right) - \frac{1}{2}\gamma(\gamma+1)\sigma_t^2\left(\frac{dC}{C}\right).$$

With 1%–2% mean consumption growth, a high γ such as 25 implies by the second term 25%–50% risk-free rates. Worse, $\gamma = 25$ implies that a one percentage point rise in mean consumption growth must correspond to a 25 percentage point rise in risk-free rates. The third, precautionary savings term can come to the rescue for very high γ, but then we require a knife-edge balance between conditional mean, conditional variance, and risk aversion to produce the observed low and relatively stable risk-free rate.

Risk premiums also vary over time, with a clear business-cycle correlation. You can forecast stock, bond, and currency returns by regressions of the form

$$R_{t+1}^e = a + by_t + \epsilon_{t+1} \tag{2}$$

using as the forecasting variable y_t the price/dividend or price/earnings ratio of stocks, yield spreads of bonds, or interest rate spreads across countries.

In each case the 1-month or 1-year R^2 and t statistics are not overwhelming. But measures of *economic* importance are large. Expected returns vary over time as much as their level: $\sigma[E_t(R_{t+1}^e)] = \sigma(a + by_t)$ is large compared to $E(R^e)$. If the equity premium is 6% on average, it is as likely to be 1% or 11% at any moment in time. (A regression of returns on dividend yields gives a standard error of expected returns $\sigma[E_t(R_{t+1}^e)] = \sigma(by_t)$ of 5.5 percentage points. See Cochrane (2011), Table I.)

Furthermore, expected returns are high, prices are low, and risk premiums are high, in a coordinated way across many asset classes, in the bottoms of recessions. Expected returns are low, prices are high, and risk premiums are low at the tops of booms (Fama and French, 1989).

Price volatility is another measure of the economic significance of expected-return variation. Shiller (1981) (see also Shiller, 2014) famously found that higher or lower stock prices do not signal higher or lower subsequent dividends. This observation is arithmetically equivalent to regressions of the form (2) (Cochrane, 1991). High prices relative to current dividends must imply higher future dividends or lower future returns. If higher prices do not correspond to higher future dividends, then high prices mechanically correspond to lower future returns. The "excess" volatility of prices is exactly the same phenomenon as the predictability of returns and time-variation of the risk premium.

In sum, we face two main questions. First, the equity premium question: What is there about recessions, or some other measure of economic bad times, that makes people particularly afraid that stocks will fall during those bad times—and so people require a large up-front premium to bear that risk? Second, the predictability question: What is there about recessions, or some other measure of economic bad times, that makes that premium rise—that makes people, in bad times, even more afraid of taking the same risk going forward?

These are two separate questions. People could hate the event of a recession, but not become more risk averse during recessions. Power utility has this property—people dislike losses, but losses do not make them more averse to taking risk going forward. Some gamblers have the opposite response, doubling up on risk when they lose. Or people could become more risk averse at times that do not involve painful losses. Recessions seem to combine both effects, current pain and additional risk aversion about future prospects. But the two effects may not be perfectly correlated and different mechanisms or aspects of recessions—job loss versus financial crisis, say—may control each one.

The questions are related, however. A mechanism that makes people more risk averse in recessions will drive them to try to sell stocks. With inelastic supply, they will drive down prices and cause prices to be lower in recessions, so if recessions are also painful, the betas will be higher.

The challenge is not one of telling stories or "explaining" facts or events ex post. The consumption-based model works well at a qualitative level, as does the story that people are afraid of recessions, and become more risk averse during recessions. The challenge is to find concrete, quantitative, and theoretically explicit measures of fearful outcomes and of risk aversion, that quantitatively account for asset pricing facts.

2. Theories

To explain these facts, the macro-finance literature explored a wide range of alternative preferences and market structures. A sampling with a prominent example of each case:[2]

1. Habits (Campbell and Cochrane, 1999a, 1999b).
2. Recursive utility (Epstein and Zin, 1989).
3. Long-run risks (Bansal and Yaron, 2004; Bansal, Kiku, and Yaron, 2012).
4. Idiosyncratic risk (Constantinides and Duffie, 1996).
5. Heterogeneous preferences (Gârleanu and Panageas, 2015).
6. Rare Disasters (Reitz, 1988; Barro, 2006).
7. Utility nonseparable across goods (Piazzesi, Schneider, and Tuzel, 2007).
8. Leverage; balance sheet; institutional finance (Brunnermeier, 2009; Krishnamurthy and He, 2013).
9. Ambiguity aversion, min–max preferences (Hansen and Sargent, 2001).
10. Behavioral finance; probability mistakes (Shiller, 1981, 2014).

These approaches look different, but in the end the ideas are quite similar. Each of them boils down to a generalization of marginal utility or discount factor, most of the same form,

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

The new variable Y_{t+1} does most of the work.

Even the behavioral and probability distortion views are basically of this form. Expressing the expectation as a sum over states s, the basic first-order condition is

$$p_t u'(C_t) = \beta \sum_s \pi_s(Y) u'(C_{t+1,s}) x_{t+1,s},$$

2 The following sections cover more examples of each case, but in the interest of space, and with apologies to authors whose papers are omitted, I do not attempt a comprehensive literature review. I focus on the ideas through these examples.

where x denotes a payoff with price p. Probability and marginal utility always enter together, so distorting marginal utility is the same thing as distorting probabilities. The state variables Y driving probability distortions act then just like state variables driving marginal utility.

The source of additional risk Y and of time-varying risk-bearing ability varies. In the habit model, endogenous time-varying individual risk aversion is at work—people are less willing to take risks in bad times. Nonseparable goods models work in a related way—past decisions such as the size of house you buy affect marginal utility of consumption. In behavioral or ambiguity aversion models, people's probability assessments vary over time. In long-run risks, rare disasters and idiosyncratic risks models, the risk itself is time-varying. In heterogeneous agent models and institutional finance models, the market has a time-varying risk-bearing capacity, though neither risks, individual risk aversion, or individual probability mis-perceptions need vary over time. In heterogeneous agent models, changes in the wealth distribution that favor more or less risk averse agents induce the shift in risk-bearing capacity. In institutional finance models, preferences do not change but the changing fortunes of leveraged intermediaries induce changes in the market's risk-bearing capacity.

The models also differ in their tractability, elegance, and the number and fragility of extra assumptions (or "dark matter" in the colorful analogy of Chen, Dou, and Kogan, 2015) needed to get from theory to central facts.

These features matter. In explaining which models become popular throughout economics, tractability, elegance, and parsimony matter more than probability values of test statistics. Economics needs simple tractable models that help to capture the bewildering number of mechanisms people like to talk about. Elegance matters. Economic models are quantitative parables. Elegant parables are more convincing than black boxes. Dark matter is particularly inelegant. Models that need an extra assumption for every fact are less convincing than are models that tie several facts together with a small number of assumptions. Financial economics is always in danger of being simply an interpretive or poetic discipline: Markets went down, sentiment must have fallen. Markets went down, risk aversion must have risen. Markets went down, there must have been selling pressure. Markets went down, the Gods must be displeased. Models that rejectably tie their central explanations to other data, and cannot "explain" any event are more convincing—even if they are formally rejected as perfect descriptors of the data.

2.1 Habits

Campbell and Cochrane (1999a) address the facts, focusing on predictability and volatility, by introducing a habit, or subsistence point X into the standard power utility function,

$$u(C) = (C - X)^{1-\gamma}/(1-\gamma).$$

We furthermore assume that the habit X is external, generated by observing others' consumption, so the consumer ignores the fact that more current consumption will affect future habits, and risk aversion becomes

$$-\frac{Cu''(C)}{u'(C)} = \gamma\left(\frac{C}{C-X}\right) = \frac{\gamma}{S}.$$

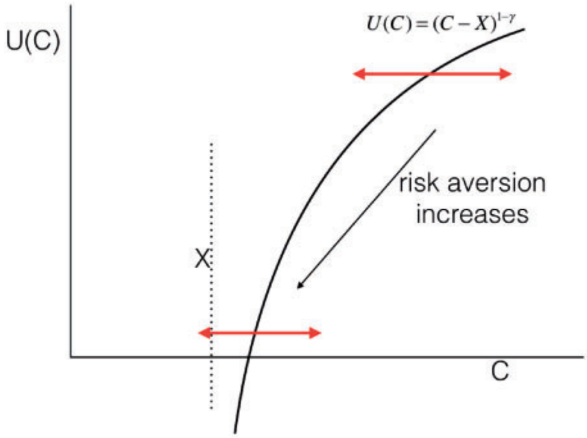

Figure 1. Utility function with habit. The curved line is the utility function. The vertical dashed line denotes the habit or subsistence level X. Horizontal arrows represent the same proportional risk to consumption.

As consumption C or the surplus consumption ratio S decline, risk aversion rises. (Risk aversion is properly the curvature of the value function, not the curvature of the utility function. However, true risk aversion behaves much as this local curvature in the habit model. Also, external habit is a convenience, but not essential.)

Figure 1 illustrates the idea. The same proportional risk to consumption, indicated by the horizontal arrows, is a more fearful event when consumption starts closer to habit, on the left in the graph. In the example, the risk is merely unpleasant at a high level of consumption on the right. However if consumption is low on the left, the same risk can send future consumption below habit, a fate worse than death.

We specify a slow-moving habit. Roughly,

$$X_t \approx \phi X_{t-1} + kC_t.$$

This specification allows us to incorporate growth, which a fixed subsistence level would not do. As consumption rises, people slowly get used to the higher level of consumption. Then, as consumption declines relative to the level they have gotten used to, it hurts more than the same level did back when consumption were rising. As I once overheard a hedge-fund manager's spouse say at a cocktail party, "I'd sooner die than fly commercial again." The one-period habits $U = \sum \beta^t (C_t - \theta C_{t-1})^{1-\gamma}$ common in macroeconomics give rise to large quarterly fluctuations in asset prices, not the business-cycle pattern we see in the data.

Figure 2 graphs the basic idea of the slow-moving habit. As consumption declines toward habit in bad times, risk aversion rises. Therefore, expected excess returns rise. Higher expected returns mean lower prices relative to cash flows, consumption, or dividends. Thus a lower price-dividend ratio forecasts a long period of higher returns.

Expected cash flows (consumption or dividend growth) are constant in our model, so if prices reflected expected dividends discounted at a constant rate, then the price–dividend ratio would be constant. The large variation in the model's price–dividend ratio is driven entirely by varying risk premiums. Thus, the model accounts for the "excess volatility" of stock prices relative to expected dividends.

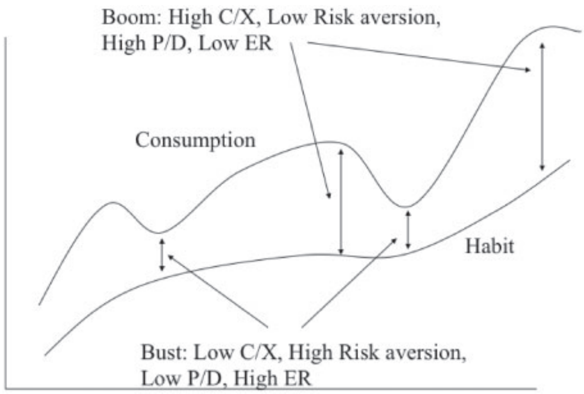

Figure 2. Stylized sample from the habit model. The upper line represents a sample path of consumption. The lower line represents a slow-moving habit induced by movements in consumption.

As Figure 2 illustrates, at the top of an economic boom, prices seem too high or to be in a "bubble," as prospective returns are low. But the representative investor in this model knows that expected returns are low going forward. Still, he or she answers, times are good, he or she can afford to take some risk, and what else is the investor going to do with the money? He or she "reaches for yield," as so many investors are alleged to do in good economic times.

Conversely, in bad times, such as the wake of the financial crisis, prices are indeed temporarily depressed. It is a buying opportunity; expected returns are high. But the average investor looks at this situation and answers "I know it's a good time to buy. But I might lose my job. If things get any worse, I could lose the house too. There is a minimum standard of living I just can't put at risk."

In sum, as Figure 2 illustrates, the habit model naturally delivers a time-varying, recession-driven risk premium. It naturally delivers returns that are forecastable from dividend yields, and more so at longer horizons. It naturally delivers the "excess" volatility of stock prices.

This habit model is proudly reverse-engineered. This graph gives our basic intuition going into the project. A note to Ph.D. students: All good economic models are reverse-engineered! If you pour plausible sounding ingredients in the pot and stir, you will never get anywhere.

We engineer the habit accumulation function to deliver a constant interest rate, or in an easy generalization, a real interest rate that varies slowly and pro-cyclically, as we observe.

With $(C-X)^{-\gamma}$ marginal utility and fixed X, the interest rate is

$$rdt = \delta dt + \gamma\left(\frac{C}{C-X}\right)E\left(\frac{dC}{C}\right) - \frac{1}{2}\gamma(\gamma+1)\left(\frac{C}{C-X}\right)^2\sigma^2\left(\frac{dC}{C}\right). \tag{3}$$

The real interest rate equals the subjective discount factor δ, plus the inverse elasticity of intertemporal substitution times expected consumption growth, plus risk aversion squared times the variance of consumption growth.

Habit models typically have trouble with risk-free rates. As $C-X$ varies, the second term leads to strong movement in risk-free rates r or in expected consumption growth $E(dC/C)$. In a bad time, marginal utility is high, and the consumer expects better (lower marginal utility) times ahead, if not by a rise in consumption, then by a downward adjustment in habit. He or she would like very much to borrow against that brighter future to cushion the blow today. If consumers can borrow, that desire leads to persistent movements in consumption growth. If not, the attempt drives up the interest rate. The data show neither strongly persistent consumption growth nor large time-variation in real interest rates.

But in our model, precautionary savings in the third term are large and vary over time. For example, if $\gamma = 2$ but $S = (C - X)/X = 0.05$ so $\gamma/S = 40$ to accommodate the equity premium puzzle, and with 2% standard deviation of consumption growth, then $1/2 \times 2 \times 3/(0.05)^2 \times 0.02^2 = 0.48$, so precautionary savings subtracts 48 percentage points from the risk-free rate. This term addresses the risk-free rate puzzle, that high risk aversion in the first term otherwise implies a large risk-free rate. More importantly here, movement in precautionary savings in the third term offsets movement in intertemporal substitution in the second term. In the simplest form of the habit model, the two terms offset exactly to produce a constant risk-free rate and i.i.d. consumption growth. In bad times, people want to borrow more against a better future, but they want to save more against a risky future, and in the end they do neither.

Expressed in terms of a discount factor, the habit model adds a recession indicator $S \equiv (C - X)/C$ to consumption growth of the power utility model,

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

Consumers want to avoid stocks that fall when consumption is low, yes. But with $\gamma = 2$ this is a small effect. Consumers really want to avoid stocks that fall when S is low—when the economy is in a recession.

2.2 Evaluation

So, what does the habit model accomplish? And, by example, what is the standard first set of empirical successes that similar macro-finance models aim for?

We compared the habit model to data by comparing interesting statistics of simulated data from the model to those from the data. We picked most parameters directly to match data, such as the mean and standard deviation of consumption growth. We picked the curvature parameter γ to match the sample equity premium and the habit persistence parameter to match the autocorrelation of dividend yields. Additional moments are then somewhat like tests of the model.

Equity premium. The model delivers the equity premium $E(R^e)$ and market Sharpe ratio $E(R^e)/\sigma(R^e)$, with low consumption volatility $\sigma(\Delta c)$, unpredictable consumption growth $E_t(\Delta c_t) =$ constant, and a low and constant (or slowly varying) risk-free rate.

But the model does not have low risk aversion. The coefficient $\gamma = 2$, but utility curvature γ/S and risk aversion are large. In the latter sense, the habit model does not solve the equity premium–risk-free rate puzzle. The puzzle as now distilled includes the equity premium $E(R^e)$, the market Sharpe ratio $E(R^e)/\sigma(R^e)$ and thus market volatility $\sigma(R^e)$, a low and stable risk-free rate R^f, realistic mean, volatility, and predictability (not much) of consumption growth, with a positive subjective discount factor δ and low-risk aversion. The habit model has everything but low-risk aversion. So far no model has achieved a full solution of the equity premium puzzle as stated.

Predictability and volatility. The model delivers the observed return predictability from dividend yields, and price–dividend ratio volatility, despite i.i.d. cash flows—high price/dividend ratios do not forecast cash flow growth at all—and despite a low and constant risk-free rate. One of its functions has been to point out how predictability, volatility and time-varying risk aversion and risk premium are really the same.

The model also delivers conditionally heteroskedastic returns—volatility is higher after a price fall. However, the conditional mean and conditional standard deviation of returns are different functions of the state variable, so the conditional Sharpe ratio varies over time, higher in bad times.

Long-run equity premium. The long-run equity premium was to us the most unexpected result. Look again at the habit discount factor, this time at a k year horizon,

$$M_{t,t+k} = e^{-k\delta}\left(\frac{C_{t+k}}{C_t}\right)^{-\gamma}\left(\frac{S_{t+k}}{S_t}\right)^{-\gamma}.$$

The equity premium, as distilled by Hansen and Jagannathan (1991), is centrally the need for a higher volatility $\sigma(M_{t,t+k})$ than aggregate consumption alone, raised to small powers γ, provides. The S term provides that extra volatility in the habit model, and the similar terms do so in other models. In the short run, S and C are perfectly correlated—a positive shock to C raises $C - X$—so the second S factor just amplifies consumption volatility. But in the long run, S_{t+k}/S_t—whether we are in a recession—and C_{t+k}/C_t—long run growth—become uncorrelated. Risks to the surplus consumption ratio are a separate pricing factor, and the dominant one for driving asset prices and long-run expected returns.

Now, consumption is a random walk, so the standard deviation of the consumption-growth term rises approximately linearly with horizon. But the second term, like the second term of most other models in this class, is stationary. Therefore, the volatility of the recession indicator $\sigma(S_{t+k}/S_t)$ eventually stops growing with horizon k. If you look far enough out, any model with a stationary extra factor Y_t is going to end up with the consumption model and no extra equity premium at long horizons. Intuitively, temporary price movements really do melt away, so a patient investor collects long-run returns and no long-run volatility. In the long run, growth fluctuations drown out business cycle fluctuations.

In the nonlinear habit model, it turns out that though S_{t+k}/S_t is stationary, $(S_{t+k}/S_t)^{-\gamma}$ is not stationary. Its volatility increases linearly with horizon, so the model produces a high long-run equity premium. Marginal utility has a fat tail, a rare event, a min–max, or super-salient state of nature that keeps the equity premium high at all horizons. I deliberately use words to connect to the other literatures here, as one of my points is the commonality of the different kinds of models, and the fact that habit models do incorporate many of the intuitions that motivate related models. And vice versa. However, most of the other explicit models do not capture the long-run equity premium.

Fitting data. In the habit model, the price dividend ratio is a function of the surplus consumption ratio $S = (C - X)/C$. Thus, one can construct a model-implied price/dividend ratio from the history of consumption data and compare that to the actual price–dividend ratio, which we do.

Figure 3 presents a simpler version of this calculation, to highlight the central intuition and robust fact of the model in a more transparent way. Figure 3 plots the NYSE price–dividend ratio $\log(P/D)$ together with $C - X$, log consumption minus a habit that is simply a

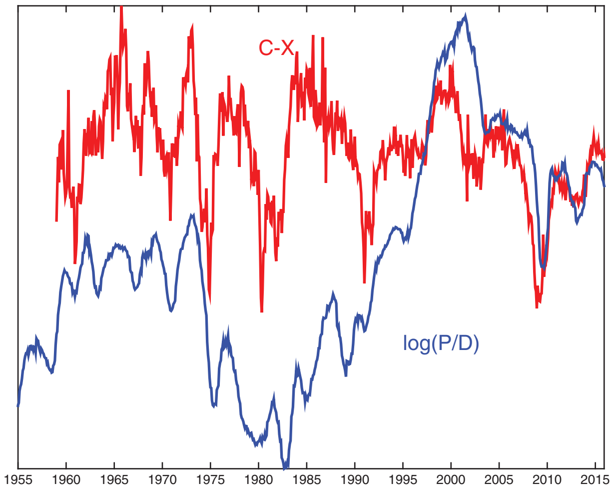

Figure 3. Price–dividend ratio and detrended consumption. P/D is the ratio of price to dividends of the value-weighted NYSE CRSP index. $C - X$ is the difference between log total real per capita consumption C and its moving average $X_t = \phi X_{t-1} + (1 - \phi)C_t$. The vertical scales are shifted so the lines fit on the same graph, by maximizing the correlation between the two lines 1990–2015.

moving average of past log consumption, $X_t = \phi X_{t-1} + (1 - \phi)C_t$, with ϕ chosen arbitrarily at $\phi = 0.9$.

The figure shows a strong correlation between detrended consumption and cyclical movements in stock prices. In fact, the correlation is stronger after 1999 than before. The stock boom of the 1990s corresponds to a consumption boom. Most of all, the stock plunge in 2008, recovery in 2010 and even the variation in the slowdown of 2013–14 mirror those of detrended consumption. The brickbats thrown at modern efficient-market finance for being unable to accommodate the financial crisis are simply false. This model works better in the big shock of the financial crisis than at other times.

Many questions about the habit model remain. It does not fit the data perfectly, and it can and should be generalized to address these facts better and many more asset pricing facts. One may question its micro foundations—do people really behave this way in micro data, and does that matter? I address these questions below after surveying parallel approaches.

2.3 Recursive Utility and Long-Run Risk

The bulk of other work in macro-finance has adopted seemingly much different fundamental specifications of preferences, markets, and technology. Though quite different in their underpinnings, the end result of these models is quite similar. Even the mode of analysis is similar, as models all capture similar lists of moments.

The recursive utility approach uses a nonlinear aggregator to unite present utility and future value,

$$U_t = \left[(1 - \beta)C_t^{1-\rho} + \beta\left[E_t\left(U_{t+1}^{1-\gamma}\right)\right]^{\frac{1-\rho}{1-\gamma}}\right]^{\frac{1}{1-\rho}}. \tag{4}$$

Here γ is the risk aversion coefficient and $1/\rho$ is the elasticity of intertemporal substitution. This function reduces to time-separable power utility for $\rho = \gamma$.

The discount factor, or growth in marginal utility, is

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

The innovation in the utility index takes the role of the new variable Y in my general classification. (Cochrane, 2007 contains a derivation.)

The utility index U_t itself is not observable, so the trick is to substitute for it in terms of observable variables. Epstein and Zin (1989) used the market return, as a proxy for the wealth portfolio return. The most common approach recently, exemplified by Bansal, Kiku, and Yaron (2012), and Hansen, Heaton, and Li (2008), is to substitute out the utility index in terms of the stream of consumptions that generate utility. This substitution delivers the long-run risk model. For $\rho \approx 1$,

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

where $\Delta E_{t+1} \equiv E_{t+1} - E_t$.

In this formulation, news about long-run future consumption growth is the extra state variable Y_t. As usual this extra state variable does the bulk of the work to explain risk premiums. In this model, people are afraid of stocks because stocks go down when there is bad news about long-run future consumption growth, not necessarily when the economy is currently in a recession, when current consumption is low (power utility), when the market is low (CAPM), or at a time when consumption is low relative to its recent past (habits).

The Bansal, Kiku, and Yaron (2012) consumption process is

$$\Delta c_{t+1} = \mu_c + x_t + \sigma_t \eta_{t+1} \tag{5}$$

$$x_{t+1} = \rho x_t + \phi_e \sigma_t e_{t+1} \tag{6}$$

$$\sigma_{t+1}^2 = \bar{\sigma}^2 + \nu(\sigma_t^2 - \bar{\sigma}^2) + \sigma_w w_{t+1} \tag{7}$$

$$\Delta d_{t+1} = \mu_d + \phi x_t + \pi \sigma_t \eta_{t+1} + \phi \sigma_t u_{d,t+1}. \tag{8}$$

The x process generates positive serial correlation in consumption growth. Thus, a small change in current consumption is linked to a big change in long-run consumption, and it is the long-run consumption news that agents fear.

The long-run risk model, like the habit model, produces the equity premium with a low and stable risk-free rate and realistic (low) one-period consumption volatility. It can use high-risk aversion, as in the habit model. It can also produce the equity premium with relatively low risk aversion, by imagining a lot of positive serial correlation in consumption growth—a lot of long-run news. In this case, though, long-run consumption growth volatility is high, so it is in the class of theories that abandon the low consumption volatility ingredient of the equity premium puzzle statement. ($E(R^e)/\sigma(R^e) = \gamma\sigma(\Delta c)$ can be achieved with high $\sigma(\Delta c)$.) Therefore, the recursive utility model also does not solve the classic statement of the equity premium puzzle. No model yet does so.

Return predictability and time-varying volatility are the more interesting and challenging phenomena, and the ones more tied to macroeconomics. The long-run risk model does not endogenously produce time-varying risk premia. These are added by assuming an exogenous pattern of consumption volatility. In Equation (7) σ_t gives the time-varying long-run consumption risk which drives time-varying expected returns. This explanation of predictability goes back to Kandel and Stambaugh (1990) with power utility: To get $E_t(R^e)/\sigma_t(R^e) \approx \gamma\sigma_t(\Delta c_{t+1})$ to vary over time with constant γ, you need to imagine that $\sigma_t(\Delta c_{t+1})$ varies over time.

This model is very popular. Still, it carries some long-standing difficulties.

First, the model crucially needs there to *be* news about long-run consumption growth—variation in $\Delta E_{t+1}(\Delta c_{t+j})$, $j>1$—to get anywhere. If consumption is a random walk; if each day consumers' expectations of consumption growth in 2030 are the same, say 1%, then there is no long-run consumption news and the model reduces to time-separable power utility.

Current conditions Δc_t are essentially irrelevant to investor's fear. Investors only seem to fear stocks that go down when current consumption goes down (fall 2008, say) because, by coincidence, current consumption declines are correlated with the bad news about far-off long-run future consumption growth that investors really care about.

So is there a lot of news about long-run consumption growth? And is it at all believable that this is really what investors care about? The former is hard to find in the data. Apart from a first-order autocorrelation due to the Working effect (a time-averaged random walk follows an MA(1) with an 0.25 coefficient) and the effects of seasonal adjustment (our data are passed through a 7 year, two-sided bandpass filter), nondurable and services consumption looks awfully close to a random walk. (Beeler and Campbell (2012) elaborate this point.) The evidence is largely about short-run correlations, and Inferring long-run predictability from a few short-run correlations is a dubious business in the first place. Maximum likelihood and related econometric techniques value short-run forecasts, and are happy to get long-run forecasts wrong, or to miss many high-order autocorrelations, in order to better fit one-step ahead predictions (Cochrane, 1988).

Similarly, there needs to be substantial variation over time in the uncertainty about future long-run growth rates for the model to generate a time-varying risk premium. If consumers' uncertainty about consumption growth in 2025 is the same, each day, say also 1 percentage, point, then there are no time-varying risk premiums.

One might retort, well, the standard errors are big, so you cannot prove there is not a lot of long-run positive autocorrelation in consumption growth and its volatility. But demoting the central ingredient of the model from a robust feature of the data to an assumption that is hard to falsify clearly weakens the whole business.

I often advise students to write the op-ed or teaching note version of their paper. If you cannot explain the central idea to a lay audience in 900 words, then maybe it is not such a good idea after all.

In this case, that oped would go something like this: Why were people so unhappy in fall 2008? What was there about fall 2008 that made the fall in stock prices so much more painful than a similar fall in good times—and contemplating such events ahead of time is why people in good times did not buy even more stocks? (That is the equity premium question.) It was not, really, because the economy was in a recession, that investors had lost their jobs and houses and they were cutting back on consumption. Those facts, per se, were irrelevant. Instead, it was because 2008 came with bad news about the long-run future. Investors figured out what no professional forecaster did, that we would enter the current decade or more of low growth. If that bad news about long-run growth happened to be correlated with a boom rather than bust in 2008, people would have paid dearly ex ante to avoid stocks that did particularly badly in the boom. People did not fundamentally care at all about what was happening in 2008—it is only the long-run news that mattered to them.

Similarly, why were people in 2008 unwilling to take advantage of a buying opportunity, a higher than usual expected returns and buy more stock? (This is the predictability, volatility and time-varying risk premium question.) Why were university endowments, despite websites declaring themselves to be "long-run" investors who ride out "temporary" market drops, trying to sell in a panic? It was not because consumption fell toward habitual levels, or a reduced cash flow from endowment might force universities to fire tenured faculty, or people fear becoming unable to pay debts. It was because the conditional variance of such long-run growth expectations rose. They were less sure about conditions in 2028 than they had been before, and this, and only this, drove them to panic.

This strikes me as a difficult essay to write, and a difficult proposition to explain honestly to an MBA class on any day but the first of April.

To understand the long-run risk model, ask this (a good exam question): How is the long-run risk model different from Merton's ICAPM (Merton, 1973)? After all, the ICAPM also includes additional pricing factors, that are "state variables for investment opportunities." News about long-run consumption growth would certainly qualify as an ICAPM state variable. Yet the ICAPM has power utility. Why did we need recursive utility to get long-run consumption growth expectations to matter for asset prices?

The answer is that the ICAPM is a subset of the power-utility consumption-based model. Its multiple factors are the market return and state variables, not consumption growth and state variables. In response to bad news about future consumption, ICAPM consumers reduce consumption today. That reduction in today's consumption reveals all we need to know about how much the bad news hurts.

In contrast, the long-run risks model weights news about future consumption that is not reflected in consumption today. Somehow, you get news that you will be poor in the future. You rue the decision to buy stocks, yet still choose to consume a lot today. This is the kind of bad news about which you are really afraid. If you did react by lowering consumption today then today's consumption would be a sufficient statistic for the bad long-run news, and that news would have no extra explanatory power.

In the habit model as other models, people really are worried about stocks falling in 2008—because of events going on in 2008.

Fear of news about the far off future, unrelated except by coincidence and correlation to macroeconomic events today, is closely related to the central theoretical advertisement for recursive utility. It is a feature, not a bug. Recursive utility captures—and requires—a preference for early resolution of uncertainty. Psychology lab experiments seemed to find such a preference, motivating the development of the theory. This is a tricky concept. In almost all of your experience you prefer to resolve uncertainty early because you can do something with that knowledge. If you know what your salary will be next year, you can start looking for a better house, or a different job. If you learn what the stock market will do next year, you can buy or sell today. The preference for early resolution of uncertainty that these preferences capture is a pure pleasure of knowing the future, even when you cannot do anything in response to the news.

I find lab experiments documenting such preference unpersuasive, because there is essentially no circumstance in daily life in which one gets news that one can do absolutely nothing about. People respond to surveys and experiments with rules of thumb adapted to the circumstances of their lives.

Epstein, Farhi, and Strzalecki (2014) address the question this way: How much would the consumer in the Bansal–Yaron economy pay, by accepting a lower overall level of consumption, to know in advance what that consumption will be, even though they could not do anything about it; just for the psychic pleasure of knowing what it will be in advance? The answer is around 20%–30%. That seems like a lot.

So, capturing a strong preference for early resolution of uncertainty starts to me to look more like a bug than a feature.

The other apparent theoretical advantage is that recursive utility separates risk aversion from intertemporal substitution, allowing high-risk aversion for the equity premium and a low and steady risk-free rate.

But so do habits. The habit model delicately offsets time-varying intertemporal substitution demands with a time-varying precautionary saving and thereby generates the same result.

Recursive utility may achieve the result more elegantly. Elegance and tractability are important in economic theories. Elegance is a plausible argument for the popularity of the recursive-utility approach. But elegance and tractability can also lead us astray. If in fact time-varying precautionary saving is important—if, say, fall 2008 had a large fall in consumption because people were scared to death—then the recursive-utility model is missing the crucial feature of reality. Furthermore, though the square root habit adjustment process in the habit model may seem inelegant, in fact it requires much less algebra than one must surmount to solve recursive utility models.

There is also little direct evidence for the proposition that the conditional variance of long-run consumption growth varies significantly over time and is tightly correlated to price–dividend ratios in the manner of Figure 3. Moreover, the presence of time-varying conditional long-run consumption growth volatility and its correlation with time-varying long-run news are additional exogenous assumptions.

To avoid vacuousness, all extra state-variable models must propose some independent way to measure the extra (Y) variable. In the habit model, the extra state variable—surplus consumption ratio—is directly and independently measurable from the history of consumption.

The long-run risk model ties its extra state variables—volatility and news about long-run consumption growth—to observables by the assumption of a time-series process in which short-run consumption growth is correlated with volatility and long-run news. That assumption makes long-run news (almost) independently measurable. But the crucial link is driven by the exogenous driving process, not the economic structure of the model. (I say "almost" because the state variables x_t and σ_t in (6)–(8), though observed by agents, cannot be directly recovered from the history of consumption and dividends.)

Finally, substituting the market return as in Epstein and Zin (1989), or long-run consumption growth for the utility index in (4), requires that we use the entire wealth portfolio (claim to total consumption stream) or total consumption. The usual trick in separable utility, that the asset pricing implications of $u(c_{nd}) + v(c_d)$ are the same as those of $u(c_{nd})$ alone, where c_{nd} and c_d represent consumption of nondurables and durables respectively, does not work for nonseparable utility. As with the CAPM, one ignores this fact.

However, the habit and recursive utility models have a lot in common, and that commonality is my greater theme. Both models capture a quite similar idea. There is an extra state variable, which explains why people are afraid of holding stocks in ways not described by consumption growth alone. That extra state variable has something to do with recessions, bad macroeconomic times. Both models capture an equity premium and time-varying predictability, one with time-varying risk, the other with time-varying risk aversion. No model has gotten significantly ahead of the others in terms of the number of phenomena it captures. All models have inconvenient truths that we ignore, as the original CAPM required no investor to hold a job, and predicted that consumption volatility is the same as market volatility. That did not stop it from being a useful model for many years. The habit model carefully reverse-engineers preferences to deliver the equity premium and predictability. The long-run risks model carefully reverse-engineers the exogenous consumption process to deliver the same phenomena. One observer's fragile assumption is another observer's well-identified parameter. Though I have argued that model-specification assumptions are prettier than driving-process assumptions, that is an esthetic judgment.

2.4 Idiosyncratic Risk

Idiosyncratic risk, such as in Constantinides and Duffie (1996), is another fundamentally different microeconomic story that generates similar results.

The bottom line is again a discount factor that adds a state variable beyond consumption growth,

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} e^{\frac{\gamma(\gamma+1)}{2} y_{t+1}^2}.$$

Here y_{t+1} denotes the cross-sectional variance of individual consumption growth. The log of each individual's consumption follows

$$\Delta c_{t+1}^i = \Delta c_{t+1} + \eta_{i,t+1} y_{t+1} - \frac{1}{2} y_{t+1}^2; \; \sigma^2 \left(\eta_{i,t+1} \right) = 1.$$

Therefore, y_{t+1} plays the role of the second, recession-related state variable in place of the surplus consumption ratio or long-run risk.

The story: People are afraid of idiosyncratic consumption risk. Some people might get great consumption gains, some might face great consumption losses. With risk aversion, which implies nonlinear marginal utility, fear of the losses outweighs pleasure at the gains, so overall people (the representative consumer) fear times of large idiosyncratic consumption risk and fear assets that do badly at times of great idiosyncratic consumption risk.

The Constantinides and Duffie paper is brilliant because it is so simple, and it provides directions by which you can reverse-engineer any asset pricing results you want. Just assume the desired cross-sectional variance y_{t+1} process. This reverse engineering also circumvents many problems with the previous idiosyncratic risk literature.

As with the long-run risks model, however, the level and any time-variation and business cycle correlation of the equity premium all are baked in by the exogenous variation in the moments of the consumption process, rather than the endogenous response of risk aversion to bad times. Cross-sectional consumption volatility must be large, must vary a good deal over time, and at just the right times.

One can check the facts, and so far the empirical work has been a bit disappointing. What matters for risk premiums is not the level of cross-sectional risks, but unexpected increases in cross-sectional risks. y_{t+1} must vary over time to generate volatility in the discount factor $\sigma(M_{t+1})$. Cross-sectional risks do rise in recessions, and when asset prices are low, but that rise does not seem large enough to generate the risk premiums we see, at least with low levels of risk aversion. Consumption risks are much smaller than transitory income or employment risks, because people tend to smooth consumption.

However, this is still an active area of empirical research. For example, Schmidt (2015) investigates whether the nonnormality of idiosyncratic risks can help—whether a time-varying probability of an idiosyncratic rare disaster dominates the cross-sectional risks to marginal utility. Such events are intuitively plausible.

These models and empirical investigation have not seen much extension to generate return predictability. The theoretical path is straightforward. To generate $\sigma_t(M_{t+1})$ that varies over time, we need $\sigma_t(y_{t+1})$ to vary over time—time variation in the conditional variance of the conditional variance of cross-sectional risks (a mouthful indeed). Constantinides and Ghosh (2017) is the state of the art, both in theory and in empirical work to demonstrate the appropriate time-varying moments in micro data.

Again, you can see the essential unity of the ideas. A second state variable, associated with recessions, drives marginal utility. People are afraid that stocks might fall in recessions, and being in a recession and a time of low price–dividend ratios raises that fear. Here recessions are measured by an increase in idiosyncratic risk, and an increase in the chance of further shocks to idiosyncratic risk, rather than by a fall of average consumption relative to its recent past or a rise in the conditional variance of long-run aggregate consumption. But those events are likely to be highly correlated. The state variable is exogenous and requires an extra set of assumptions or measurements. But that is an esthetic difference. The moments of cross-sectional risk are at least more tightly tied to data and measurable than the inference about long-run risk from its correlation with short-run risks, and more theoretically restricted and measurable than the extra state variables in psychological models to come.

2.5 Heterogeneous Preferences

Gârleanu and Panageas (2015) offer a related but diametrically opposed model. For Constantinides and Duffie, people have the same preferences, risks are not insured across people, and exposure to this time-varying cross-sectional risk drives asset prices. For Gârleanu and Panageas, people have different preferences—some are more risk averse, and some are less risk averse—risks are perfectly insured across people, and time-varying wealth across more or less risk averse people drives asset prices. Less risk averse people hold more stock. But when the market goes down, these big stockholders lose more money, and so they become a smaller part of the overall market. and the market as a whole becomes more risk averse after a fall in value.

More precisely, in a complete market the unique discount factor Λ_t and consumer A, B consumptions follow

$$\Lambda_t = \lambda_A e^{-\delta t} C_{A,t}^{-\gamma_A} = \lambda_B e^{-\delta t} C_{B,t}^{-\gamma \lambda_B}. \tag{9}$$

(Here, λ_i are time-invariant Pareto weights, the weight of each consumer in the associated planning problem, or reflecting initial wealth in equilibrium. $M_{t+1} = e^{-\delta}\Lambda_{t+1}/\Lambda_t$.)

In bad times, with high Λ_t, the less risk averse consumer accepts greater consumption losses, while in good times, that consumer enjoys greater gains. Mechanically, this sensitivity is implemented via greater investment in the market.

Differentiating these relationships, we can express the discount factor in terms of aggregate consumption $C_t = C_{A,t} + C_{B,t}$ raised to an aggregate risk aversion, which is the consumption-weighted average of individual's inverse risk aversion,[3]

$$\frac{1}{\gamma_{mt}} = \frac{1}{\gamma_B}\frac{C_{B,t}}{C_t} + \frac{1}{\gamma_A}\frac{C_{At}}{C_t}. \tag{12}$$

You see here exactly the sort of mechanism of a habit model—the representative agent becomes more risk averse after a fall in consumption. But here, that rise does not come because each individual becomes more risk averse. It comes because the mechanism of aggregation puts more weight on the risk averse people in bad times.

This is a beautiful model, which emphasizes just how many micro stories are consistent with the same macro phenomenon. The representative consumer has time-varying risk aversion though individuals do not. Markets display less risk bearing capacity in bad times, though people do not. Time-varying risk bearing capacity of the market can be driven by market structures—a point the institutional finance and leveraged-intermediary literature below makes with a different mechanism—as well as by individual preferences.

This model faces challenges and opportunities in the micro data just as the idiosyncratic risk model does. Do the "high-beta rich" really lose so much in bad times? Can the model quantitatively account for return predictability? But that investigation has not really started.

3 Differentiating Equation (9),

$$\frac{d\Lambda_t}{\Lambda_t} = -\delta dt - \gamma_A \frac{dC_{A,t}}{C_{A,t}} + \frac{1}{2}\gamma_A(1+\gamma_A)\frac{dC_{A,t}^2}{C_{A,t}^2} \tag{10}$$

and likewise for B. Therefore,

$$\frac{d\Lambda_t^2}{\Lambda_t^2} = \gamma_A^2 \frac{dC_{A,t}^2}{C_{A,t}^2},$$

and we can solve

$$\frac{dC_{A,t}}{C_{A,t}} = -\frac{1}{\gamma_A}\delta dt - \frac{1}{\gamma_A}\frac{d\Lambda_t}{\Lambda_t} + \frac{1}{2}\frac{1+\gamma_A}{\gamma_A}\frac{d\Lambda_t^2}{\Lambda_t^2}. \tag{11}$$

Now,

$$\frac{dC_t}{C_t} = \frac{C_{A,t}}{C_t}\frac{dC_{A,t}}{C_{A,t}} + \frac{C_{B,t}}{C_t}\frac{dC_{B,t}}{C_{B,t}}.$$

Substituting from (11), and with (12), and its corollary

$$\frac{1+\gamma_m}{\gamma_m} = \frac{1+\gamma_A}{\gamma_A}\frac{C_{A,t}}{C_t} + \frac{1+\gamma_B}{\gamma_B}\frac{C_{B,t}}{C_t},$$

we have

$$\frac{dC_t}{C_t} = -\frac{1}{\gamma_m}\delta dt - \frac{1}{\gamma_m}\frac{d\Lambda_t}{\Lambda_t} + \frac{1}{2}\frac{1+\gamma_m}{\gamma_m}\frac{d\Lambda_t^2}{\Lambda_t^2}.$$

So we have (11) and (10) with aggregate consumption and market risk aversion.

2.6 Debt, Balance Sheets, and Institutional Finance

A different category of model has become much more popular since the 2008 financial crisis: models involving debt, balance sheets, mortgage overhang; institutional or intermediated finance.

The basic story works much like habit persistence. Imagine that an investor has taken on a level of debt X, which he or she must repay. Now, as income declines toward X, the investor takes on less and less risk, to make sure that even in bad subsequent states of the world he or she can repay the debt. The intuition of Figure 1 applies exactly.

Moreover, as consumption rises in good times, such people slowly take on more debt. As consumption falls in bad times, people “delever,” “repair balance sheets” and so forth. Debt moves slowly, following consumption, much like slow-moving habit.

Though the mechanism is broadly similar, however, debt-based finance models are deeply different from all the others in this survey. In all the other models, even psychological ones, markets equate margins between borrowers and lenders. Asset price variations result from preferences or perceptions of each individual, and each individual is “marginal” at all times. Aggregate risks are shared. (Behavioral models have some frictions on occasion to keep arbitrageurs from removing pricing errors, but the source of pricing errors remains misperceptions by individual final investors, who are able to buy and sell.) In intermediated-finance models, in contrast, the absence of most investors from the market, is central to the story. In this story, for example, the vast bulk of people did not change risk preferences or probability mis-perceptions in 2008. They would have loved to have bought at fire-sale prices. But they were not “marginal,” unable or unwilling to buy cheaply priced stocks directly. Only the leveraged intermediaries were active in markets, and they, and only they, were suddenly more risk averse because of recent losses. Similarly, households and businesses would have loved to borrow more to finance purchases or investment, but leverage and capital constraints at banks stopped money from flowing from willing lenders to these willing borrowers.

These models also face theoretical and empirical difficulties.

First, why do people get more risk averse as they approach bankruptcy, not less? Bankruptcy is the point at which you do not have to pay your debts any more. It is usually modeled as a call option. Failure to pay debt in our economy does not result in debtors’ prison, destitution, or worse. The usual concern is therefore that people and businesses near bankruptcy have incentives to take too much risk, not too little.

The costs, benefits, reputational concerns, and so forth surrounding bankruptcy are subtle, of course, and I do not mean to argue that we know exactly one way or another in all circumstances. I do point out that it is not at all obvious that debt should induce more risk aversion rather than less, and it takes modeling effort and special assumptions to produce the “more” answer.

Second, not everyone is in debt. My debt is your asset—net debt is zero. For this reason, most institutional finance models center on segmented markets, so that the problems of borrowers weigh more heavily on markets than the problems of their creditors.

The typical institutional finance story told of the financial crisis goes like this: Fundamental investors—you and me—give our money to intermediaries. The intermediaries take on leverage, so we split our funding of the intermediaries into debt and equity tranches. When the intermediaries start losing money, they get more risk averse, and start selling assets. Assumptions are layered on to keep them from raising more equity, giving us securities directly, betting the farm on riskier trades, or borrowing more. You and I do not trade in the underlying assets, so there is nobody around to sell to. Only other highly levered intermediaries are "marginal." Hence, when they try to sell, prices go down. That puts them closer to bankruptcy, so they sell more, with colorful names like "liquidity spiral," or "fire sale."

The objections to this sort of model are straightforward. OK, for obscure collateralized debt obligations, credit default swaps, or other hard to trade instruments, and this may explain why small arbitrages opened up between more obscure derivatives and more commonly traded fundamentals. But how does this story explain widespread, coordinated, long-lasting movements in stock and bond markets around the world? After all, these assets are part of everybody's opportunity set via Vanguard and E-Trade, and most people have them in 401(k) accounts. We are all "marginal," at least at the month to years horizons over which business cycles evolve.

Moreover, large, sophisticated, unconstrained, debt-free wealthy investors and institutions such as university endowments, family offices, sovereign wealth funds, and pension funds all trade stock indices and corporate bonds every day. If leveraged intermediaries push such prices down nothing stops these investors from buying. Where were they in the crisis? Answer: they were selling in a panic like everyone else. That surely smacks of time-varying risk aversion, induced by recent losses, not a segmented market in which every investor, fundamental and intermediary alike, wants to buy but leverage and agency problems cause the only active agents—leveraged intermediaries—to sell.

Furthermore, if there is such an extreme agency problem, that delegated managers were selling during the buying opportunity of a generation, why do fundamental investors put up with it? Why not invest directly, or find a better contract?

To be clear, I think the evidence is compelling that "small" arbitrage opportunities in hard-to-trade markets during the fall of 2008 were linked to intermediary problems. I put "small" in quotes, because an economically small arbitrage opportunity—say, a 1% deviation from covered interest parity—while not enough to attract long-only interest on one side or the other, represents a potentially enormous profit for a highly leveraged arbitrageur. Still, a 1% price deviation is still small from the perspective of the overall economy.

But the presence of those frictions and arbitrages does not mean that leveraged intermediaries are responsible for the bulk of the large movements in stocks, corporate and government bonds, and foreign exchange that we saw during the crisis. Their presence means even less that perpetually constrained, leveraged intermediaries and absent fundamental investors are always the story for financial market movements, continuing to this day. Inequality constraints do not bind when they are slack, and people who run in to them take care not to have them bind forever.

Business and consumer debt, "leveraging" and "deleveraging" or "balance sheets" are an attractive related mechanism for inducing time-varying risk aversion. The models also can look a lot like a habit model. But I have similar doubts about the view that business and consumer debt is the major driver of asset prices and macroeconomics, rather than contributing relatively minor, if important, epicycles. If bad times mean that some consumers will be close to the default limit, then why borrow so much in the first place? Buffer stock models require very high discount rates to eliminate this natural tendency to save up enough assets to avoid the bankruptcy constraint Though the average person may be constrained, the average dollar driving the risk-bearing capacity of the market is held by an unconstrained consumer. Bill Gates and the Harvard endowment have a lot more money than you and I do.

The institutional finance view also does not easily explain why asset prices are so related to macroeconomic events. Losing money on intermediated and obscure securities does not always lead to a recession.

One might imagine reverse causality, a new model of macroeconomics by which financial events spread to the real economy not vice versa. That is an exciting possibility, actually, and the core of the bustling frictions-based macro-finance research agenda. (Mian and Sufi, 2014 is a prominent example.) But at this stage it is really no more than a vision—models adduce frictions far beyond reality, such as that no agent can buy stocks directly, and data analysis is still limited to one event.

Also, institutional frictions are fleeting, but the sense that economic downturns correspond to less risk-bearing capacity in markets goes back centuries. If our financial crisis occurred because only leveraged hedge funds and dealer banks could buy and sell collateralized debt obligations, and therefore a great buying opportunity opened up, well, next time mortgage backed securities will be held in long-only exchange-traded funds, and the entire phenomenon will disappear.

So, in my view, institutional finance and small arbitrages are surely important frosting on the macro-finance cake, needed to get a complete description of financial markets in times of crisis. When a recession happens, they are likely amplifying mechanisms for financial markets and potentially real activity. But are they also the cake? And are they the meat and vegetables of normal times, and the bulk of movements in broad market indices, and the full explanation for macroeconomic events? Or can we understand the big picture of macro-finance without widespread frictions, and leave the frictions to understand the smaller puzzles, much as we conventionally leave the last 10 basis points to market microstructure, but do not feel that microstructure issues drive the large business cycle movements in broad indices?

Again, though, my main point is to point out the many commonalities, and only slightly to highlight differences. Theories based on debt deliver the same central idea, that the market fears recessions and fears assets whose values fall in recessions, and that the risk-bearing capacity of the market declines in bad times.

The theories outlined so far differ mainly in the exact state variable for expected returns—consumption relative to recent values, news about long-run future consumption, cross-sectional risk, or leverage, that is, balance sheets of individual consumers or those of leveraged intermediaries. But all of these state variables are highly correlated, and all capture a similar idea.

2.7 Rare Disasters

Barro (2006) has recently taken up an idea of Reitz (1988), that the equity premium and other asset pricing phenomena can be understood by the fear of rare disasters. With Barro's inspiration, this idea has expanded substantially.

Look back at the basic asset pricing equation,

$$E_t(R^e_{t+1}) = \mathrm{cov}_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}, R^e_{t+1}\right] \le \sigma_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\right]\sigma_t(R^e_{t+1}).$$

If people worry about rare events with very low consumption growth, then the variance of marginal utility in investors' heads will be larger than the variance we measure in a sample that does not include any rare events.

The basic idea is reasonable, that people worry about rare and severe events when buying securities. People in California still worry about large earthquakes, though we have not seen one since 1906, and rare events are priced in to earthquake insurance.

One objection to this view is that we should have seen more disasters if they are large or frequent enough to account quantitatively for the equity premium with low risk aversion (Constantinides, 2008). This observation has led much work quantifying just how many disasters we have seen, in the USA and abroad, over long spans of time, how to define a disaster, and what it constitutes. (Events in which both stocks and bonds become worthless do not justify an equity premium.) Calibration of time-varying rare disaster models to account for predictability and volatility is still in its infancy. (Welch, 2016 also finds that the probabilities of rare disasters implied by put option prices are too low to account for much of the equity premium.)

To get rare disasters to account for the more interesting business-cycle related return predictability and stock price volatility, one could specify that the risk of a rare consumption or return disaster changes over time—that $\sigma_t[(c_{t+1}/c_t)^{-\gamma}]$ or $\sigma_t(R_{t+1})$ vary over time, due to changing tail probabilities. Wachter (2013) and Gabaix (2012) follow this approach.

Alternatively, a rare disasters perspective could posit that expected returns really do not change over time. People see time-varying probabilities of a rare disaster in dividend payouts. Prices really are lower because expected dividends are lower, not because expected returns are higher. But in a sample that has no rare disasters we suffer Peso-problem regressions that falsely indicate return predictability rather than dividend predictability, and consequently falsely indicate "excess" volatility.

Dark matter is a deeper objection. Rare events, unobserved in the postwar sample, are already to some extent a dark matter assumption. Time-varying probabilities of rare disasters seems like dark energy (i.e., even more obscure)—unless one proposes some way of independently tying the time-varying probability of rare disasters to some data.

One might surmount the dark-matter criticism if one assumption about time-varying disaster probability could reconcile multiple asset prices, but as Gabaix (2012) finds, to make sense of the different asset classes, one needs to assume an asset-specific time-varying loading on the disaster risk.

Finally, the correlation of asset prices with business cycles relies on a correlation of business cycles with a time-varying disaster probability. A correlation between business cycles and time-varying disasters is not implausible, as a correlation between a recession and lower long-run growth is not implausible. Each recession could turn in to a great depression or worse, they just have not done so yet. But it is one more exogenous assumption, and one that is one step harder to measure directly than the unconditional frequency of rare disasters.

I admit a final esthetic prejudice—but these aesthetic considerations are important in which theories survive. Though it is possible that time-varying fears of rare events drive all asset price movements, if this is so our discipline will never really be able to measure anything. Day to day betas do not matter. All that matters is time-varying subjective correlation of asset prices with each other and with macroeconomic events in once per century Armageddons. It could be true. But if that is true, we will never really be able to move past interpretive, ex-post storytelling. We will never be able to distinguish rare disaster probability from "sentiment," we will never tie our state variables to some independent measurement, and workaday applied finance that relates average returns to measurable betas is doomed. I prefer to go back and look for the car keys under the light.

2.8 Probability Assessments

Another class of models generalizes rational expectations. Suppose people's probability assessments are wrong. I include behavioral finance here, which uses survey, psychology, and lab experiments to motivate wrong probability assessments, as well as modifications of preferences under the labels "Knightian uncertainty," "ambiguity aversion," and "robust control."

The basic asset pricing equation, with the expectation written as a sum over states s, is

$$p_t u'(C_t) = \beta \sum_s \pi_s u'(C_{t+1,s}) x_{t+1,s},$$

where p_t is price at time t, s indexes states of nature at time $t+1$, and $x_{t+1,s}$ is a payoff. (Typically $x_{t+1,s} = d_{t+1,s} + p_{t+1,s}$ includes a dividend and tomorrow's price.)

As this equation emphasizes, probability and marginal utility always enter together. There is no way to tell risk aversion—marginal utility—from a probability distortion, using price p and payoff x data alone. That is, there is no way to do it without some restriction—some model that ties either probability distortions or marginal utility to observables. This statement is just the modern form of Fama's joint hypothesis theorem that you cannot test efficiency (π) without specifying a model of market equilibrium ($u'(C)$). Likewise, absent arbitrage opportunities, there is always a "rational" model, a specification of $u'(C)$ that can rationalize any data.

Given these facts, one would have thought that arguments over rational versus irrational pricing, using only price and payoff data, would have ended the minute Fama (1970) and its joint hypothesis theorem were published. That financial economists have spent so much effort fighting over theories all sides know were proved observationally equivalent a half century ago will hopefully be a useful observation for future historians and sociologists of science.

The solution, of course, is to tie either probabilities or marginal utility to observable data, in some rejectable way. In our general formula, if subjective (wrong) probability assessments are $\pi_s(Y)$, where Y is measurable, then it becomes a testable theory, just as we have so far added potentially measurable state variables Y to investor's marginal utilities or objective probability assessments. A model of sentiment can be a model of market equilibrium for Fama's theorem.

Barberis *et al.* (2015) is a good example of a paper that takes on these problems. They build a model based on extrapolative expectations (a probability distortion), in which people believe good growth will keep going more than it really does. As a result prices are "too high" in good times, and forecast lower returns, but probability distortion takes the place of marginal utility. "Sentiment" is tied to the history of consumption so becomes observable and testable. And the model generates the standard suite of predictability results.

But without such a specification, "sentiment" is all too often just a dark-matter, ex-post, interpretive explanation. Time-varying rare-disaster probabilities, not separately measured, or time-varying news about far-future incomes, not separately measured, or time-varying risk aversion, not separately measured, are equally dark-matter, interpretive stories.

Behavioral economists point to surveys, in which people report amazing possibilities as their "expectation." (Greenwood and Shleifer, 2014 is a good recent example.) But it is a big leap from "What do you expect?" in a survey to "What is your true-measure conditional mean?" in a model. Surveys never ask "By the way, did you report your risk-neutral or true-measure mean?" They do not ask that question for the obvious reason that people would have no idea what the question means. But the question is crucial. The risk-neutral probability is the actual probability times marginal utility,

$$\pi_s^* = \pi_s \beta \frac{u'(C_s)}{u'(C_0)} R^f.$$

With risk-neutral probabilities, price is the expected payoff, discounted at the risk-free rate.

$$p_0 = \frac{1}{R^f} \sum_s \pi_s^* x_s = \frac{1}{R^f} E^*(x).$$

Now, imagine that prices are absurdly high, true expected returns are extremely low, you ask in a survey what investors "expect," and they answer that they "expect" good returns (good expected x), justifying the price. Irrationality confirmed! But without the follow-up question, if respondents reported the risk-neutral probabilities, they are not being irrational at all. The price *is* the risk-neutral expectation of payoff! So the question "are those true-measure or risk neutral probabilities?" is not a technicality, it is the whole question.

And it would be entirely sensible for people to think about and report risk-neutral probabilities, not true probabilities. Since probability and marginal utility always enter together, risk-neutral probabilities are a good sufficient statistic to make decisions. Risk neutral probabilities mix "How likely is the event?" with "How much will it hurt if it happens?" That combination is really what matters. Avoid stubbing your toe on the door jamb, yes. But put more effort into avoiding getting run over by a truck. Though it is much less probable, it hurts a lot more.

More generally, the colloquial word "expect" is centuries older than the mathematical concept of true-measure conditional mean. Statisticians borrowed a colloquial word to describe their new concept, as they borrowed the colloquial words "efficient," "unbiased," "consistent" and so forth, and endowed them with new technical meanings. But unless trained in statistics or economics (and, as teachers will ruefully note, actually remembering anything from that training) there is no reason to believe that a surveyed person has the statistical definition in mind rather than the colloquial definition.

The online Oxford English Dictionary[4] defines "expect" as to "regard (something) as likely to happen." Its five definitions do not even mention means, let alone true versus risk neutral measure, or the distinction between mean, mode, and median and conditional versus unconditional. So even a literate layperson cannot know that the survey questioner is asking for the true-measure conditional mean. The OED says the word comes from the mid-16th century (in the sense "defer action, wait"), and derives from Latin exspectare, to look out for. The word goes back a long before anyone dreamed up the concept of conditional mean. The distinction between risk-neutral and real probabilities was only formalized in Harrison and Kreps (1979).

4 https://en.oxforddictionaries.com/definition/expect

So why do we expect nontrained survey respondents to use the word "expect" as we defined it about 30 years ago, and not in the sense reflected in centuries of common usage, and the dictionary definition?

More deeply, of course, economics has long argued whether it matters what people think and say, rather than how they act. If they act "as if" rational, if rational maximizing describes their behavior, who cares how they fill out surveys? Much animal behavior is well described by optimization—how bees search flowers for pollen, for example—yet they do not answer surveys coherently either.

I do not mean to disparage survey information. This is a fascinating and very useful source of data. Even with "as if" skepticism, it is comforting when people report something like the mechanisms of our models and troubling when they do not. People do report an unwillingness to take on risk in recessions, or when facing a lot of debt. People do not, as far as I know, report any concern about stocks falling at a time of poor long-run consumption growth expectations independent of such events. These are important observations. That people report what seem to us as incredible "expectations" is a nut to be swallowed, if only as a report on the subjective perception of low-risk aversion.

My complaint is only with interpreting what people say they "expect" directly and unquestioningly as true-measure full-information conditional means, and looking at people somewhat condescendingly as "irrational" if those answers do not make sense. Other uses and interpretations of the data—for example, running regressions $R_{t+1} = a + b(\text{survey}_t) + cy_t + \epsilon_{t+1}$ on survey expectations, and looking at everything but whether $b = 1$ and $c = 0$—are potentially very revealing.

The ambiguity-aversion and robust-control literature also distorts probabilities. (For an excellent overview see Hansen, 2014.) An heuristic equation describes this approach,

$$p_0 u'(C_0) = \beta \sum_s \pi_s u'(C_s) x_s$$
$$\{\pi_s\} = \arg\min_{\{\pi \in \Theta\}} \max_{\{c\}} \sum_s \pi_s u(C_s)$$

(I call this an heuristic equation because the real equations are much harder, but this conveys the idea.) The probabilities π are chosen, in a restricted set Θ, as those that minimize the maximum attainable utility. The investor focuses on the worst-case scenario in a set, and devotes all his or her attention to that case.

Obviously, hard questions remain. Most of all, just what is the restricted set Θ? If you worry about meteorites falling from the sky, maybe you should worry about anvils and pianos too? Again, also, tying the distorted probabilities to measurable data remains the key to understanding variation in prices over time.

3. Research Agenda

A rich smorgasbord of research possibilities appears before us: Distinguish the models, by their microeconomic or macroeconomic foundations, and by their ability to fit current or new asset pricing facts. Extend the range of asset pricing phenomena the models can address. Unite the models where the answer appears to be that two or more mechanisms coexist and interact in an important way. Unite the models with macroeconomics—place endowments in a general equilibrium context, think about where cash flows come from, and change macroeconomics in turn.

3.1 Additional Facts

The habit model has been extended to capture additional macro-finance facts. Verdelhan (2010) shows how two habit economies with varying interest rates produce the forecastability of currency returns. Roughly, the low interest rate country has higher risk premiums, so the high interest rate country's bonds, with exchange rate risk, must deliver higher expected returns. Lopez, Lopez-Salido, and Vazquez-Grande (2015) use slow-moving habits, extended to the utility of leisure, production using capital and labor, investment with adjustment costs, and Calvo-style price rigidities, to address the term structure of risk premiums. Wachter (2006) produces a time-varying term premium that is not perfectly correlated with the time-varying equity premium. She adds a new state variable to fit that fact. The currency risk premium and term premium at least should now also be part of the standard set of facts we ask models to replicate.

There are also straightforward and valuable opportunities to bring up some of the other macro-finance models to describe the standard set of predictability facts, as I have indicated above. The plausibility of the extensions needed to do so may be an important distinguishing lesson.

3.2 New Facts

One of the most active areas of macro-finance research addresses a new fact, the term structure of risk premiums. This literature is notable because it has been used as an explicit testing ground to see how macro-finance models work on new facts, how different models compare on new facts, and it pursues sensible extensions and generalizations of the models toward that goal.

Roughly speaking, this work distinguishes conditional and unconditional expected returns $E_t R^e_{t+k}$ across different horizons k, generalizing the above long run equity premium discussion. It also studies the expected one-period returns $E_t R^{e,k}_{t+1}$ of dividend strips, claims to one dividend D_{t+k}, as a function of k. The concept of a risk premium that varies by horizon as well as time is familiar in fixed income (e.g., Cochrane and Piazzesi, 2008), but this rapidly-expanding literature brings the same concepts to the much harder world of equities with risky payoffs. Notable examples include Hansen (2013), Borovička *et al.* (2011), and Croce, Lettau, and Ludvigson (2015), the latter notable for a bounded rationality model of long-run risk to generate a downward sloping equity term structure. van Binsbergen and Koijen (2017) summarize the facts and evaluate the ability or failure of several different classes of model to explain them, including habits, long-run risks, and variable rare disasters, finding all current models wanting, even including some sensible generalizations.

However, the curse of $\sigma/\sqrt{T}$ means that all asset pricing facts are somewhat contentious, especially those based on short samples. There is an especially large contention whether the term structure of equity risk premium "facts" are facts.

For example, van Binsbergen, Brandt, and Koijen (2012), one of the first and most celebrated papers documenting facts around which models are built, claims that average returns to short-dated dividend strips (a claim to the first year of dividends only, say) are higher than average returns on the market and the average returns on long-dated strips. However, while point estimates of mean returns are in their Table 1, the standard errors for the key difference $E(R_{1,t} - R_{\mathrm{S\&P500},t})$ do not show up until row 5 and 6 of their Table 6, where the authors acknowledge that this central fact of the paper is not statistically significant:

> We also formally test whether the risk premium on the dividend strategies is higher than the risk premium on the index. We cannot reject this null hypothesis at conventional significance levels. (p. 1609)

The paper does not even present a similar statistic for the difference between dividend strips of different maturity $E(R_{1,t} - R_{2,t})$. In our era of rampant t-statistic fishing, the AER and the authors are perhaps to be congratulated for publishing interesting results with t-statistics of 0.75 and 0.97. But it remains a tenuous fact for decisive model discrimination.

Also, their Table 2 also shows negative serial correlation of −0.27 and −0.37, classic signs of measurement error in prices. Their Figure 4 makes it clear that any return comes from one data point in 2001. Additional critiques are in Boguth *et al.* (2012) and Schulz (2016), with a response in van Binsbergen and Koijen (2016), focusing on the data construction and tax treatment.

This is not the place to exhaustively summarize a data controversy. The point here: this is just the kind of new fact one should use to evaluate and contrast macro-finance models; This literature includes just the kind of sensible generalizations one must pursue to thoughtfully match models to data. That in this particular case, the fact itself is more disputed than, say, the equity premium, volatility, predictability and so forth—themselves still disputed—just echoes what should be a constant warning: Theorists, beware. Read empirical papers carefully before jumping to explain new "facts."

3.3 More Facts

There are substantial additional discrepancies in macro-finance models that seem to have gone unnoticed, and inviting directions for improvement. I illustrate with habits, but the same points hold more generally.

More shocks; match VARs. The consumption-claim version of the habit model has one shock, the shock to consumption growth. This shock is simultaneously a cash flow shock and a discount rate shock, so the cash flow and discount rate shocks are perfectly negatively correlated. When consumption declines (cash flow shock), the discount rate rises.

The standard VAR representation of returns and dividend yields has at least two distinct and economically important shocks. In the simplest VAR, cash flow shocks and discount rate shocks are uncorrelated.

In round numbers, the standard VAR representation for log returns r, log dividend growth Δd, and log dividend yield dp is

$$\begin{aligned} r_{t+1} &\approx 0.1 \times dp_t + \epsilon^r_{t+1} \\ \Delta d_{t+1} &\approx 0 \times dp_t + \epsilon^d_{t+1} \\ dp_{t+1} &\approx 0.94 \times dp_t + \epsilon^{dp}_{t+1} \end{aligned}$$

and the covariance matrix of the shocks is

$$\mathrm{cov}(\epsilon\epsilon') = \begin{array}{c|ccc} & r & \Delta d & dp \\ \hline r & \sigma \approx 20\% & +\text{big} & -\text{big} \\ \Delta d & & \sigma \approx 14\% & \mathbf{0\ not\ -1} \\ dp & & & \sigma \approx 15\% \end{array}.$$

The definition of return means that only two of the three equations are needed, and the other one follows. If prices rise or dividends rise, returns must rise. In equations, the Campbell and Shiller (1987) log return approximation is

$$r_{t+1} \approx dp_t - \rho dp_{t+1} + \Delta d_{t+1}, \tag{13}$$

where dp is log dividend yield, Δd is long dividend growth, and $\rho \approx 0.96$ is a constant of approximation. This equation is just a log-linearization of the definition of a return, $R_{t+1} = (P_{t+1} + D_{t+1})/P_t$. As a result of this identity, the VAR regression coefficients b and shocks ϵ are linked by identities

$$b_r = 1 - \rho b_{dp} + b_d \tag{14}$$

$$\epsilon^r_{t+1} = -\rho\epsilon^{dp}_{t+1} + \epsilon^d_{t+1}. \tag{15}$$

With any two coefficients (14), shocks (15), or data series (13), you can find the third.

It is common to write the VAR with dividend yields and returns, $\{dp_t, r_t\}$ and let dividend growth be the implied variable. I like to think of it instead in terms of dividend growth and dividend yields $\{dp_t, \Delta d_t\}$ with returns the implied variable. (Think of it, yes, but do not run it that way. Never run a return forecasting regression with less than a pure return. Small approximation errors can make returns look much more forecastable than they really are.) The reason for this preference is that, while dp and r shocks are negatively correlated—when prices go up, dividend yields go down and returns go up—dp and Δd shocks are essentially uncorrelated.

Thus, the easy-to-remember summary of the canonical three-variable VAR is this: There are two shocks in the data: a cash flow shock ϵ^d, and a discount rate shock ϵ^{dp}, and these two shocks are uncorrelated. The negative correlation of return and dividend yield shocks ϵ^r, ϵ^{dp}, and the positive correlation of return and dividend growth shocks ϵ^d, ϵ^r then just follows from the last identity.

Clearly, this little VAR paints a different picture than the consumption-claim habit model in which the cash flow and discount rate shocks are perfectly correlated. We need to think of and model a world with separate cash-flow and discount-rate shocks.

Campbell and Cochrane (1999a) includes a model with a claim to dividends poorly correlated with consumption, which makes progress toward a two-shock model. However, that model does not replicate the VAR. It suffers a worse problem too as follows.

Cointegration. Consumption, stock market value, and dividends are cointegrated. Consumption, dividends, and wealth are steady fractions of GDP in the long run. Campbell and Cochrane (1999a) just specify imperfectly correlated growth rates of consumption and dividends Δc and Δd. But the levels of consumption and dividends wander away from each other.

Many models have imperfectly correlated Δc and Δd. I have not seen one yet that properly delivers the long-run stability of the ratios of stock market value, consumption, and dividends.

Cointegration is tricky, however. Total dividends and total consumption in the economy are cointegrated, because if dividend payments go up, people start to spend them. The dividend growth one calculates in the usual CRSP data, and uses in identities such as (13), are dividends accruing to an initial dollar investment. These are not the same as total dividends in the economy, and they are not cointegrated with consumption, because they do not (and should not, for their purpose) account for the effect of new issues or repurchases on total dividends paid in the economy.

But cointegrating relations are powerful long-run forecasters. Cointegration tells us that the ratio of prices to dividends must forecast long-run price changes or long-run dividend changes. So it is a good guess that a model that imposes the correct long-run cointegration between total consumption, total dividends, and stock market wealth will enlighten at least the long-run equity premium and long-run consumption risk. The cay variable inaugurated by Lettau and Ludvigson (2001) and studied more in their following work is an important start, integrating consumption-wealth and price–dividend cointegration. But the larger program, to fully specify macro-finance models in a way that incorporates the single common trend that we seem to see in the data, remains for the future.

More state variables? The habit model has one state variable, the surplus consumption ratio $S_t = (C_t - X_t)/C_t$. The dividend yield perfectly reveals this state variable, so no other variable should help to forecast stock returns, bond returns, volatility, or anything else. And all variables that are a function of this state variable should be perfectly correlated, as the surplus consumption ratio and price–dividend ratio should be perfectly correlated. Conditional variances move over time, and the conditional Sharpe ratio moves over time as well, because $E(R^e_{t+1}|S_t)$ and $\sigma(R^e_{t+1}|S_t)$ are different functions of the state variable S_t. The version the habit model that allows for time-varying interest rates, Campbell and Cochrane (1999b), also has time-varying bond risk premiums forecast by yield spreads. But all of these variables are functions of the same state variable, so perfectly correlated with dividend yields and with each other. (Note, the number of state variables and the number of shocks, discussed above, are different issues.)

In the empirical literature, many variables beyond the dividend yield seem to forecast both stock returns and dividend growth. The Lettau and Ludvigson (2001) consumption to wealth ratio cay is a good example. Bond returns are forecastable by bond forward-spot spreads, and foreign exchange returns by international interest spreads. In the cross-section of returns, size, book-market, momentum, earnings quality and now literally hundreds of other variables are said to forecast returns.

Now, a big empirical question remains: Just how many of these state variables do we really need, in a multiple regression sense? The forecasting variables are correlated with each other. Are they all proxies for a single underlying state variable? Or maybe two or three state variables, not hundreds?

The question is, what is the factor structure of expected returns? If we run regressions

$$R^{ei}_{t+1} = a_i + b_i x_t + c^i y_t + \ldots + \epsilon^i_{t+1};\ E_t(R^{ei}_{t+1}) = a_i + b_i x_t + c^i y_t,$$

how many state variables—orthogonal linear combinations of x, y, z—do we really need? What is the factor structure of $\text{cov}[E_t(R^{ei}_{t+1})]$? Look at that question closely—this is not the factor structure of returns, $\text{cov}(R^{ei}_{t+1})$, time $t+1$ random variables. It is the factor structure of expected returns, time t random variables. This covariance and its factor structure may have nothing to do with the factor structure of ex-post returns. But what is that factor structure? Across stocks, bonds, foreign exchange etc.? As a small first step, Cochrane and Piazzesi (2005, 2008) find that the covariance of bond expected returns across maturities has one dominant factor, though the covariance of yields or returns has three factors. Does that observation extend to bonds and stocks together? Probably not. But the bond-forecasting factor forecasts stocks, and dividend yields forecast bonds, so there is some commonality. How much of a second factor do we really need?

Conditional variances $\sigma_t(R_{t+1})$ vary over time as well. The habit model has such variation, but like everything else conditional variance is a function of the surplus consumption ratio and thus of the divided yield only. The empirical literature seems to focus on realized volatility—lagged squared returns—and volatilities implied by options prices as the state variables for conditional variance. These variables decay much more quickly than typical expected return forecasters like dividend yield. Realized volatility also forecasts mean returns, though, and dividend yields forecast volatility. How many state variables are there really driving means and variances?

If we then add state variables for returns at different horizons, $E_t(R_{t+k})$, we see a huge project—and the huge simplification and integration if a small factor structure emerges.

The answer is unlikely to be one factor, as specified in the habit model. Hence, the natural generalization of theory must be to include more state variables, to match the more state variables in the data. Since fishing around among highly correlated factors is tricky, and since time-varying mean returns are hard to measure, theory and empirical work may have an important interaction to sorting out the factor structure of expected returns. Wachter (2006) has taken a step in this direction, but there is a long way to go.

3.4 Finance Facts

Meanwhile, empirical finance has moved on, and now presents us with a zoo of factors in the cross-section of equity returns, including the market, size, value, momentum, earnings, accounting factors, carry trade and others.

(Newcomers beware: Finance uses the same word "factor" to describe a portfolio or other variable at time $t + 1$ that helps to capture return $t + 1$ variance across securities, in the classic statistical sense of the word; to describe a portfolio or other variable at time $t + 1$ against which one runs regressions to obtain betas that then explain cross-sectional variation in average E_t returns; and to describe a time t variable that helps to forecast returns at time $t + 1$, and equivalently a variable on which one can sort assets into portfolios. I use "characteristic" for the latter to avoid some confusion. These are all different concepts. Here I mean "factors" such as the market return and value-growth portfolio HML that fulfill primarily the second function.)

Empirical asset pricing summarizes the cross-section of returns in terms of a few factor portfolios, following Fama and French (1993). This summary provides a great simplification for macro-finance. Our job is to explain the premiums of the factor portfolios. We do not have to test macro-finance models in the full cross-section of returns.

Indeed, though it is tempting to do such tests, and to see if a macro-finance model drives out ad hoc factors such as Fama and French's in a cross-section of test assets, that kind of horse race is a mistake. Even if one's macro model is perfect, ad-hoc factors will always do better in any sample.

Conversely it is the sole job of macro-finance to understand why the asset-pricing factors earn a premium. Finance alone can never explain the premiums of the factor portfolios. The CAPM explains stock average returns given the average return on the market, but leaves the equity premium or average return on the market as a free parameter, whose explanation requires macro-finance. The same point holds for multifactor models.

Macro-finance has had some limited success with the value premium. Jagannathan and Wang (2007), find that even the simple linearized consumption-based model can explain the value premium at an annual horizon. Lettau and Wachter (2007) and Santos and Veronesi (2010) also present models of the value premium with habit preferences. But even this success is limited, compared to the stirring prose of Fama and French's papers suggesting a premium for firms prone to financial distress in recessions, just when broad categories of investors suffer losses to their nonmarketed businesses and human capital. (Fama and French, 1996 section VI. A. is a personal favorite.) That prose needs a model. And the rest of the factor zoo is wide open.

Again, though, theorists beware empiricists bearing facts. Finance has yet to settle down on just how many new factors there are. Harvey, Liu, and Zhu (2016) list 316 variables in the published literature, and an exponential growth rate. Linnainmaa and Roberts (2016), Harvey (2017), and Mclean and Pontiff (2016) offer sobering caution that many factors may be spurious. My above call for a factor structure of expected returns may reveal that even among those that are not spurious, many return forecasters are be highly correlated, so we only need models with a few factors. Many factors that explain common variation, such as industry portfolios, are not needed for average returns.

In addition, theory and empirical work interplay. We no longer live in the stifling world before about 1990, in which every empirical paper must pose as a "test" of a "theory" made in advance of looking at the data. But the current world in which every fact, no matter how crazy, is established on its own may go too far in the other directions. Sensible theories may help us to fish through the claims of factors.

Still, the basic value, momentum, earnings premiums are well established, and macro-finance can get going.

It is curious that macro-finance has spent quite so much effort on a tenuous new fact, the term structure of equity premiums, and so little on the much more extensively documented finance factors. That may be a selection bias that nobody has gotten a positive result so far. Great research consists of solid answers to little questions, not tenuous answers to big questions.

But it is also possible that most of the above macro-finance approaches will not be useful to understand the zoo of cross-sectional premiums, and they will be the province of institutional or frictions finance.

The central assumption of most of the macro-finance models is that all risks are perfectly shared. Most investors—or at least most dollars—are "marginal" at all times, meaning that even if they choose not to trade, they could, and prices are not far from those investors' marginal rates of substitution. This is true even of the behavioral finance and probability distortion models.

Cross-sectional factors such as momentum require frequent trading, and low-transactions trading expertise. Even large hedge funds struggle to trade it profitably. Premiums available in more obscure securities require expertise. When risks are narrowly held, markets are segmented, and premiums unrelated to aggregate conditions can emerge—at least until large mutual funds or hedge funds make those premiums available to the average investor.

The small arbitrages that appeared during the financial crisis, and some like covered interest parity that persist to this day are a particularly clear example of this phenomenon. There is obviously no reason at all to expect a macro-finance model like habits to explain the premiums of such strategies.

So we are likely to end up with an economic picture of asset markets that, in the end, unites two or more of these fundamental approaches. A representative-consumer model such as habits may well describe large movements in widely-available securities and funds, leaving near-arbitrages and premiums of high-frequency trading strategies to the economics of institutional finance and mechanics of information trading.

But where the boundary will be is not so obvious and an area ripe for research opportunity. Institutional finance is not happy being just the frosting on the cake. It would like to be the paradigm even for broad movements of the level of, say, stocks, not just for small price differences among similar specialized securities.

Conversely, limits to arbitrage and institutional finance must get past showing that price anomalies can occur, to describing with elegance and parsimony why, when, and in which direction they occur. There are also indications of the pervasiveness and hence macro foundations of many anomalies. Why do momentum stocks all comove ex-post? Why is there a momentum factor? As emphasized by Asness, Moskowitz, and Pedersen (2013) and Koijen *et al.* (2015), why are momentum and value so pervasive and so correlated with each other? Why do anomalies appear in the depth of recessions? (At a minimum, there is a macro-finance reason for shifts in the supply of capital to intermediaries.) Why is the volume of trading so suggestively correlated with the level of asset prices?

If we do end up merging models, though, that will be more interesting if the models interact: if habit persistence kicks off a price decline that hurts the leveraged intermediaries, for example.

3.5 Foundations

There is an obvious opportunity to examine more the economic foundations of various models, either to refine and extend the models or to compare them. Does the conditional distribution of aggregate consumption vary as much and in the way that long-run risk or rare disaster models specify? Do cross-sectional income and wealth distributions change as idiosyncratic risk and heterogeneous agent models specify? Do individuals display behavior in recessions—other than sitting on the market portfolio despite higher expected returns—indicative of higher aversion to aggregate risks? My comments on each model asked many such questions.

To be fair, similar questions and doubts remain for the habit model too. To generate the size of the observed risk premium, it requires habit quite close to consumption. Is habit persistence really that strong? Do these microfoundations make sense, and do they hold up in micro data? Is habit persistence really habit persistence, or does it pick up something else, such as irreversible debt-financed expenditures on durable goods? These are subtle questions. In particular, we provided only a very simple aggregation theory, while aggregation in a realistic market setting is much harder.

A small sampling of this extensive literature: Ravinia (2011) finds evidence for external habit, but notes the difficulties of aggregation. Brunnermeier and Nagel (2008) find that people do not rebalance their portfolios when they suffer wealth shocks, and behave inertially instead. But the representative investor must seem to behave inertially, and hold the market portfolio, raising the interesting aggregation question of aggregate versus idiosyncratic wealth shocks, as well as the distinction between wealth and consumption measures. Carroll, Overland, and Weil (2000) find a habit specification is important to understand cross-country correlations between savings and growth rates.

We can and should ask these questions—but we should also beware of the limitations of this inquiry. Macroeconomics has spent half a century looking for micro foundations for aggregate relationships. The impact of such microeconomic observation on macroeconomic modeling has been limited at best. In part, microeconomic evidence often does not aggregate up as obviously as it seems. Stimulus payments to one state may raise its GDP, but if they do so by transferring resources from another state, the aggregate multiplier is zero. In part, the complexity and detailed modeling one needs to understand micro data does not give rise to tractable or elegant macro relationships. Most recently, for example, huge literatures on micro price formation have limited impact on aggregate Phillips curves, and the huge literature on irreversible firm-level investment has seemed to have little effect on aggregate representations.

Perhaps to understand economic fluctuations and their link to asset prices, it is enough to study representative consumer preferences, without worrying too much about aggregation theory and microfoundations, or at least studying the latter separately.

In part, that caution follows from the central theme of this essay: Many micro stories can produce the same or quite similar representative-agent representations. For example, the representative agent becomes more risk averse in recessions. Is that because individuals become more averse to aggregate risk, as their consumption approaches habit levels? Or is it because idiosyncratic risk becomes larger? Constantinides and Duffie (1996) provide a formula by which one can reverse-engineer a cross-sectional risk assumption that exactly mimics the habit model! These two forms of the model are then observationally equivalent, using macro and asset price data.

Different microeconomic stories for the same aggregate outcomes have different policy implications, and different implications for how structural changes to the economy will affect macro-finance. Better insurance for cross-sectional risks will, in one case, and will not, in the other, dampen asset-price fluctuations. Internal versus external habits (habits formed from one's own experience versus a neighbor's experience) have virtually the same asset pricing implications. But they have quite different welfare implications, since external habits have an externality. Balance sheets and consumer debt models look like habits, but with obviously different implications for the effect of structural improvements to debt markets. Behavioral misperceptions lead to policy implications that changing risk aversion does not—at least on the questionable technocratic assumption that federal bureaucracies are less prone to probability misperceptions than investors are, and a deeper paternalistic assumption that benevolent governments should respect agents' crazy preferences but not their crazy probability assessments. So microfoundations matter much more seriously for welfare and policy analysis.

One might take a pure reduced-form attitude, and distinguish models by which data for Y turn out to work best to price assets. But the same similarity among the models makes this approach difficult as well. Most of the candidates are highly correlated with each other—most models end up adding a recession state variable, and it is practically a defining feature of recessions that many variables move together—so telling models apart will be hard this way.

3.6 Tests?

One natural avenue of research strikes me as unproductive: formal testing. All of the models are extreme and simplified. Any nonvacuous (more predictions than free parameters) macro-finance model can easily be formally rejected.

Figure 3 illustrates this statement for the habit model. The correlation is, I hope, impressive. But the model predicts that a nearly linear function of surplus consumption ratio should match the log price-dividend ratio, perfectly, down to the last decimal point. So the model can be formally rejected with zero probability value.

This kind of failure is a general feature of many economic models. Any general equilibrium model with one shock will predict that, up to nonlinearities, all its outcome variables—shocks to asset returns, consumption growth, etc.—are perfectly correlated. Any general equilibrium model with one state variable will predict that all its forward-looking variables such as prices are perfectly correlated.

Clearly in some sense the failure of perfect correlation or $R^2 = 1$ predictions is not interesting. They are not robust tests of the model's basic intuition. But maximum likelihood does not know which moments are interesting or not.

"Testing" a model asks the question, does this mathematical structure produce data that are literally and exactly identical to real-world data? Our models are not candidates for such a test. They are consciously simplified to illustrate specific mechanisms and to roughly reproduce specific phenomena.

Adding measurement error, multiple shocks and state variables, or other inessential features so that models do not fail in such obvious ways would only complicate the models to no real end. And then the test is a test of the additional ingredients. Arbitrarily removing uninteresting moments and testing the others is a bit better, but since it is not testing, why pretend that it is?

Since Kydland and Prescott (1982), we take simple models directly to data, acknowledging they can be formally rejected, but that they produce moments that are economically close to similar interesting moments in the data. In finance, the Fama and French (1993) three-factor model is the most important and practically useful asset pricing model of the last quarter century. And it is blown away by formal Gibbons, Ross, and Shanken (1989) test statistics. Both models have great successes, they explain many features of the data. If a glass is 95% full, that's an interesting fact even if you can prove it is not 100% full. (I do not defend the practice of not reporting standard errors for moments in the data, however. Understanding which moments are close by statistical measures is interesting, if not definitive.)

For this reason, formal testing of economic models has pretty much disappeared, and rightly so.

However, when we rightly abandon the starchy formalism of testing and the pretense that our models are potentially perfect replicators of the data-generating mechanism, we must recognize that model evaluation and comparison is more of an art.

When taking a model critically to data, one must be ready to adapt the model. To reject the habit model, for example, exactly as written and parameterized, because consumption is not exactly iid, because the real interest varies, is just as silly as rejecting it because simulated and actual dividend yields do not exactly match. It is easy to see how one could quickly generalize those simplifications, explicitly made for rhetorical convenience to show that those ingredients are inessential to the basic point. Similarly, if the habit model, exactly as written, fails on some other dimension, or in confronting some new asset price, then similarly it makes little sense to reject it before seeing if reasonable extensions of the basic idea will work. The same point, of course, holds for all of the models.

Furthermore, fitting any of these models to data requires all sorts of data-handling assumptions. In the habit context, both the original Campbell and Cochrane (1999b) and Campbell and Cochrane (2000) show that time-aggregation, usually ignored, is important. Jagannathan and Wang (2007), by just using fourth quarter to fourth quarter data, find unexpected success for the consumption based model, long the subject of rejections. Seeing just how well the models can do by treating carefully durability, seasonality, time aggregation, price indices, regional variation, nontraded assets, and so forth remains basically unexplored territory.

As a concrete example, consider again the fit of Figure 3. The model seems to capture business cycle movements in the dividend yield reasonably well. But it misses lower-frequency movements. Is there something different about cycles and long run? Or does our consumption data under-report the size of long-run booms and busts? Would better measures of risk premiums, considering the fact that other variables forecast both dividend growth and returns, isolate something more like the predicted line? The actual model does worse than Figure 3, largely because it does not adapt to the post-2008 growth slowdown. But would including time-varying real rates, or a time-varying mean consumption growth alter that fact? Do habits adjust other than the simple AR(1) form of the model, so that habits have adjusted to the lower growth path? There is a long list of reasonable generalizations—or, to a critic, a long list of potential excuses for failure.

In sum, getting the models to fit better by looking hard at model simplifications and data-handling assumptions is a great research opportunity. Rejections need to consider reasonable extensions, and can only show that some range of data-handling assumptions fails. For this reason, I highlighted that each of the above model extensions and comparisons did explore a set of extensions and variations along the way.

So extending and comparing models remains an art. Recognizing that fact and doing it well is no sin. But what are interesting successes and failures is a bit subjective. The ratio of ad-hoc assumptions to successful predictions is a bit subjective. How much is reasonable extension and how much is ex-post fishing is a bit subjective. Elegance matters. Economics lives in the world of McCloskey (1983), not Thomas Bayes or R. A. Fisher.

4. A Macroeconomic Agenda: Risk-Averse Recessions

Though called “macro-finance” this literature still stands quite apart from macroeconomics. Macroeconomics by and large does not use, for understanding recession-related quantity and goods-price dynamics, the preferences or market structures that macro-finance uses to understand recession-related asset pricing dynamics. Macroeconomics by and large ignores first-order effects of uncertainty, focusing on “the” short term interest rate and the consequent allocation of consumption over time.

The central lesson of macro-finance denies this approximation: Business cycle-related asset price fluctuations are all about variation in risk premiums. It follows, I think, that recessions are driven by varying risk premiums and risk aversion, by precautionary saving and by allocation of investment funds to riskier versus less risky projects, and not about varying risk-free (government overnight) interest rates and intertemporal substitution of present for future consumption.

In recessions, both consumption and investment fall, and so output and the labor to produce it fall. Most theories of business cycles therefore start with two questions: First, why does consumption fall? Second, why does a rise in desired saving not produce a rise in investment? These questions have been the heart of macroeconomics since Keynes.

In the equilibrium models that dominate current macroeconomics, intertemporal substitution provides the answer to the first question. The key equation is

$$c_t = E_t c_{t+1} - \sigma r_t + \epsilon_t^d,$$

or, expressed as an interest rate equation in asset pricing form,

$$r_t = \gamma E_t(c_{t+1} - c_t) + v_t^d, \tag{16}$$

which is a log-linearization of our standard first-order condition. High real interest rates or preference shocks (ϵ_t^d and v_t^d are shocks to β in $E\sum \beta^t u(c_t)$) drive people to consume less today and more in the future.

But macro-finance suggests that recessions, such as fall 2008, are not times at which people became thrifty, saving more to provide a better tomorrow, and they are certainly not times of high real interest rates. Macro-finance suggests that people consumed and invested less because they were scared to death—because of risk, risk aversion, high risk premiums, precautionary savings, not because of sudden thriftiness and a wrong level of the overnight federal funds rate.

The continuous-time interest rate equation is a good place to see this difference. With a habit X, we have Equation (3),

$$rdt = \delta dt + \gamma\left(\frac{C-X}{C}\right)E\left(\frac{dC}{C}\right) - \frac{1}{2}\gamma(\gamma+1)\left(\frac{C-X}{C}\right)^2\sigma^2\left(\frac{dC}{C}\right) \tag{17}$$

in place of (16). As consumption C starts to fall, risk aversion starts to rise, and the last precautionary savings term rises. For given level of the interest rate, expected consumption growth $E(dC/C)$ has to rise. For expected consumption growth to rise, the level of consumption has to fall. This is the standard new Keynesian aggregate demand mechanism. But falling consumption raises risk aversion even more. In this way, a small shock can propagate through endogenous risk aversion and precautionary savings to deliver a large decline in consumption.

In standard macroeconomic models, the habit term is small or absent, and γ is small. Since σ is of the same order as $E(dC/C)$, about 0.02, σ^2 is much smaller, about 0.02^2, so the last precautionary savings term on the right is unimportant. But with habits or high risk aversion, the last term is all important. Squaring large risk aversion overcomes squaring small standard deviation.

Many macro modelers have approached the 2008 period following the financial crisis by supposing a preference shock, a sudden increase in patience, a shock to v_t^d in (16) or δ in (17). They acknowledge this is a short hand for some other feature of a more fully-fleshed out model. A rise in precautionary savings, in the last term of (17), is exactly such a feature, and gives a useful foundation for the apparent preference shock.

Given that people want to save more, why does investment fall? Already, the macro-finance models predict that when consumption falls, risk aversion rises, and stock prices fall. On the investment side, corporate investment follows stock prices (Q) well as Figure 4 emphasizes. As with consumption and stock prices in Figure 3, the Q theory captures well cyclical movements while missing the long-run trends before 1992, but then captures the internet boom and bust, the financial crisis, and recovery much better than it is commonly given credit for. (Q theory also predicts an R^2 of one, so is easy to formally reject.)

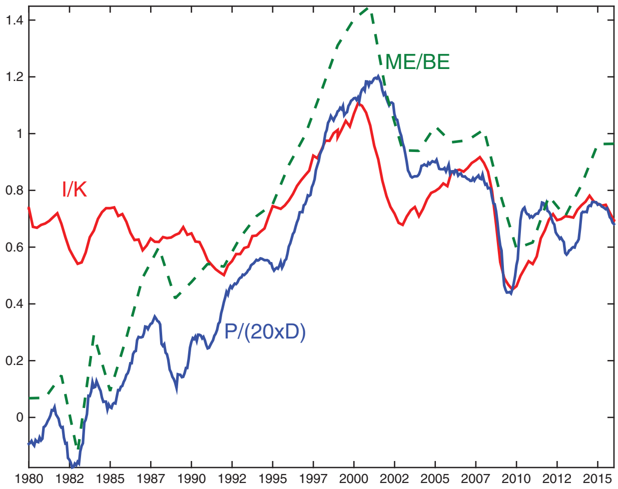

Figure 4. Investment to capital ratio, market-to-book ratio, and price–dividend ratio. Investment is real private nonresidential fixed investment. The investment/capital ratio is formed by $IK_t = I_t/I_{t-1} * IK_{t-1}/[(1-\delta) + IK_{t-1}]$, starting at $IK_0 = E(I_t/I_{t-1}) - 1 + \delta$, with $\delta = 0.1/4$. P/D is the log CRSP value weighted NYSE price–dividend ratio. ME/BE is the log market to book ratio, from Ken French's website. IK is scaled to fit on the graph, by maximizing its correlation with log P/D in the last 25 years of the sample.

Corporate investment has very little relationship with real interest rates, despite the prevalence of this channel in macroeconomic models.

So you can sniff a macroeconomic model emerging out of habits and Q theory. But still, general equilibrium poses the central puzzle of macroeconomics since Keynes: If people want to save more, why do prices not adjust somehow so that investment is larger?

The answer, I think is that investment falls because when risk aversion and precautionary saving rise, because people want not just to save more, they also want to shift their savings into relatively risk-free investments such as government bonds, and away from risky investments funded by stocks and corporate bonds. That is why we see corporate interest rates rise while government rates decline, why we see stock prices fall—expected returns rise—while people are trying to save more, and why investment falls though demand for saving rises. Additional demand for government bonds, at the expense of real assets or goods and services, also accounts for disinflation during recessions. The working paper version of this article, Cochrane (2016), explores this mechanism through a sequence of simple examples.

The key to falling investment, then, is a mismatch between the riskiness of real corporate investment projects, and the higher risk aversion of savers.

This is not the only path to greater unity between macroeconomics and macro-finance, of course. It allows us to merge the relatively frictionless preference- or market-structure based models (habits, recursive utility, idiosyncratic risk, rare disasters) that generate time-varying risk aversion with the standard general-equilibrium aggregative models that pervade macroeconomics. But the behavioral view, as outlined above, might suggest instead a reverse causality by which behavioral misperceptions in stock markets spill over to macroeconomics, or it might suggest a pervasive behavioral misperception behind both macro and finance. And merging macroeconomics with asset pricing is the rallying cry of the institutional finance/frictions research agenda, which aims to put pervasive credit constraints, balance sheet imbalances, agency frictions, and so forth at the heart of macroeconomics as well as of asset pricing. (Brunnermeier and Sannikov, 2017 is a recent, ambitious and comprehensive example.)

In this context, my suggestion is actually quite conservative. The integration of macroeconomics and finance does not have to introduce pervasive financial frictions or irrationality into macroeconomics. It is likely that the relatively frictionless approaches such as habits can be merged with standard representations of technology, pricing frictions and market structure to produce a relatively conventional macroeconomic model with time-varying risk aversion or risk-bearing capacity at its heart. Given the introductions of financial-frictions papers and books, this mere possibility seems to be news. Indeed, "macro-finance" has been appropriated as the label for the view that pervasive frictions are necessary to understand both asset pricing and macroeconomics. I use the term in the title of this essay to try to reclaim it.

The beginnings of this program are evident, though there is further to go than may be easily recognized. Habits are in fact common in macroeconomics. However, they are usually in a one-period form, $(C_t - \theta C_{t-1})$ with a small value of θ such as 0.4. These preferences help to give hump-shaped impulse-response functions. But the low value of θ and log-linearization of the model mean that time-varying risk aversion and precautionary savings channels are largely absent. Similarly, recursive utility is used in macroeconomics, but typically not with large risk aversion, or specifications that lead to long-run risk or the time-varying volatility needed to generate time-varying risk premiums. Heterogeneous-agent macroeconomics is on the rise—for example see the Kaplan, Moll, and Violante (2016) "heterogenous agent new-Keynesian model." But it does not feature the large and time-varying cross-sectional risk that generates a large and time-varying risk premium. Macro has arrived at the banquet, but only sampled the appetizer.

Not to appear imperialistic, finance needs to import macroeconomics as well. All the macro-finance models I reviewed specify endowment processes for consumption. This is a fine shortcut to get going, but eventually we need to know where consumption comes from. Similarly, finance takes the properties of cash flows as given, but we need to specify where they come from as well—betas are not exogenous. General equilibrium holds many surprises. It is easy to specify an endowment economy in which state prices vary enormously, but supply also responds to prices and can quash that variation. For example, a linear production technology gives us a constant interest rate; consumption adjusts, even if we start from an endowment process that would produce a highly varying risk free rate on its own.

Here too, the surface has been scratched, with efforts such as Jermann (1988), who united one-period habits with Q theory, and Gomes, Kogan, and Zhang (2003), Gala (2007), and Gourio (2011). The latter are revealing, as to generate a value premium they have to innovate deeply on the technology, that is beyond $y = \theta f(K, L)$ to deliver that premium. Tallarini (2000) points to the tendency of macroeconomics and finance to separate with $y = \theta f(K, L)$ plus adjustment cost technology, as investment responds to interest rates not risk premia. This thought also drives my view that the choice between multiple risky and less risky technologies is crucial to integrating macroeconomics and finance.

5. Summary

We have learned that asset prices correspond to a large, time-varying, business-cycle correlated risk premium. This risk premium means that price ratios forecast returns, and thus risk premiums—"rational" or not—account entirely for the volatility of price ratios.

Many of the apparently diverse ideas of macro-finance that account for these facts have about the same form. There is an extra, recession-related state variable, Y, so the discount factor is modified to

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

The tendency for assets to fall when Y_t falls drives risk premiums, and changes in the conditional density of Y drive time-varying risk premiums. In other models, an additional recession-related state variable drives variation in conditional moments of consumption, again driving a time-varying risk premium.

Many of the models also capture the same intuitions. I have emphasized that habit models behave much like rare-disaster, probability-distorted, or ambiguity averse models that focus on bad states; the converse interpretations work as well.

No model stands decisively above the others in its ability to describe equity premium/risk-free rate puzzles, and more importantly time-varying, business-cycle related risk premia; return predictability; "excess" volatility; and the long-run equity premium. Many of the models including rare disasters and idiosyncratic risks, have not been explicitly extended to handle predictability and volatility. My favorite, habits, is at least not yet superseded.

Each of the models suggests different candidates for the state variable Y_t. But these candidates are are highly correlated with each other, and each sensibly indicative of fear or bad economic times. Telling them apart empirically is not easy, and possibly not that productive.

The models differ a bit more on aesthetic grounds including the number of assumptions relative to predictions and analytical tractability. The time-varying risk aversion at the center of the habit model is endogenous, and a simple measurable function of consumption relative to its recent past. Many other models require carefully calibrated and complex exogenous driving processes, which in some cases (long-run risks, rare disasters) are nearly invisible in the data, or to date approach ex-post storytelling, such as labeling a market rise a rise in "sentiment" or "selling pressure," without independent measurement. But these are challenges which the other approaches are actively working to surmount. All the models, including habits, have dubious but difficult to verify micro foundations. The more subjective analytical convenience each has in capturing the common ideas may be the most important feature for modeling developments.

As I look to the future, it also seems time for this body of empirical and theoretical knowledge to invade macroeconomics, and for the general equilibrium insights of macroeconomics to invade macro-finance. Recessions are phenomena of risk premiums, risk aversion, risk-bearing capacity, desires to shift the composition of a portfolio from risky to risk free assets, a "flight to quality," not a phenomenon of risk-free interest rates, intertemporal substitution, a desire to consume more tomorrow versus today.

References

Asness, C., Moskowitz, T., and Pedersen, L. (2013) Value and momentum everywhere, *Journal of Finance* **68**, 929–985.

Bansal, R., Kiku, D., and Yaron, A. (2012) An empirical evaluation of the long-run risks model for asset prices, *Critical Finance Review* **1**, 183–221.

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

Barberis, N. et al. (2015) X-CAPM: an extrapolative capital asset pricing model, *Journal of Financial Economics* **115**, 1–24.

Barro, R. J. (2006) Rare disasters and asset markets in the twentieth century, *The Quarterly Journal of Economics* **121**, 823–866.

Beeler, J., and Campbell, J. Y. (2012) The long-run risks model and aggregate asset prices: an empirical assessment, *Critical Finance Review* **1**, 141–182.

van Binsbergen, J., Brandt, M. W., and Koijen, R. (2012) On the timing and pricing of dividends, *American Economic Review* **102**, 1596–1618.

van Binsbergen, J. H., and Koijen, R. (2016) On the timing and pricing of dividends: reply, *American Economic Review* **106**, 3224–3237.

van Binsbergen, J. H., and Koijen, R. (2017) The term structure of returns: facts and theory, *Journal of Financial Economics* Forthcoming.

Boguth, O. et al. (2012) Leverage and the limits of arbitrage pricing: implications for dividend strips and the term structure of equity risk premia. Unpublished Working paper.

Borovička, J. et al. (2011) Risk-price dynamics, *Journal of Financial Econometrics* **9**, 3–65.

Brunnermeier, M. K. (2009) Deciphering the liquidity and credit crunch 2007-08, *Journal of Economic Perspectives* **23**, 77–100.

Brunnermeier, M. K., and Nagel, S. (2008) Do wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals, *The American Economic Review* **98**, 713–736.

Brunnermeier, M. K., and Sannikov, Y. (2017) Macro, money and finance: a continuous-time approach, in: John B. Taylor, and Harald Uhlig (eds.), *Handbook of Macroeconomics*, Vol. **2B**, pp. 1497–1546, North-Holland, Amsterdam.

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

Campbell, J. Y., and Cochrane, J. H. (1999b) By force of habit, Manuscript with unpublished appendices. Unpublished Working paper.

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

Campbell, J. Y., and Shiller, R. (1987) Cointegration and tests of present value models, *Journal of Political Economy* **95**, 1062–1088.

Carroll, C. D., Overland, J., and Weil, D. N. (2000) Saving and growth with habit formation, *The American Economic Review* **90**, 341–355.

Chen, H., Dou, W., and Kogan, L. (2015) Measuring the 'dark matter' in asset pricing models. Unpublished Working paper.

Cochrane, J. H. (1988) How big is the random walk in GNP?, *Journal of Political Economy* **96**, 893–920.

Cochrane, J. H. (1991) Volatility tests and efficient markets: a review essay, *Journal of Monetary Economics* **27**, 463–485.

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

Cochrane, J. H. (2007) Financial markets and the real economy, in: Rajnish Mehra (ed.), *Handbook of the Equity Premium*, pp. 237–325, Elsevier, Amsterdam.

Cochrane, J. H. (2011) Discount rates: American finance association presidential address, *Journal of Finance* **66**, 1047–1108.

Cochrane, J. H. (2016) Macro-finance. Unpublished Working paper.

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

Cochrane, J. H., and Piazzesi, M. (2008) Decomposing the yield curve. Unpublished Working paper.

Constantinides, G. M. (2008) Comment on Barro and Ursùa, in: D. W., Elmendorf Sr., D. W., Elmendorf, N. G., Mankiw, and L. H., Summers (eds.), *Brookings Papers on Economic Activity (Spring 2008)*, pp. 341–350, Brookings Institution, Washington, D.C.

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

Constantinides, G. M., and Ghosh, A. (2017) Asset pricing with countercyclical household consumption risk, *The Journal of Finance* Forthcoming.

Croce, M., Lettau, M., and Ludvigson, S. C. (2015) Investor information, long-run risk, and the duration of risky cash-flows, *The Review of Financial Studies* **28**, 706–742.

Epstein, L. G., Farhi, E., and Strzalecki, T. (2014) How much would you pay to resolve long-run risk?, *American Economic Review* **104**, 2680–2697.

Epstein, L. G., and Zin, S. E. (1989) Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework, *Econometrica* **57**, 937–969.

Fama, E. F. (1970) Efficient capital markets: a review of theory and empirical work, *The Journal of Finance* **25**, 383–417.

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

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

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

Gabaix, X. (2012) Variable rare disasters: an exactly solved framework for ten puzzles in macro-finance, *Quarterly Journal of Economics* **127**, 645–700.

Gala, V. (2007) Irreversible investment and the cross-section of stock returns in general equilibrium. Unpublished Working paper.

Gârleanu, N., and Panageas, S. (2015) Young, old, conservative, and bold: the implications of heterogeneity and finite lives for asset pricing, *Journal of Political Economy* **123**, 670–685.

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989) A test of the efficiency of a given portfolio, *Econometrica* **57**, 1121–1152.

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

Gourio, F. (2011) Putty-clay technology and stock market volatility, *Journal of Monetary Economics* **58**, 117–131.

Greenwood, R., and Shleifer, A. (2014) Expectations of returns and expected returns, *Review of Financial Studies* **27**, 714–746.

Hansen, L. P. (2013) Risk pricing over alternative investment horizons, in: *Handbook of the Economics of Finance, Vol. 2, Part B*, pp. 1571–1611, Elsevier B.V, Amsterdam.

Hansen, L. P. (2014) Uncertainty outside and inside economic models (Nobel lecture), *Journal of Political Economy* **122**, 945–987.

Hansen, L. P., Heaton, J., and Li, N. (2008) Consumption strikes back?: Measuring long run risk, *Journal of Political Economy* **116**, 260–302.

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

Hansen, L. P., and Sargent, T. J. (2001) Robust control and model uncertainty, *American Economic Review* **91**, 60–66.

Harrison, J. M., and Kreps, D. M. (1979) Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* **20**, 381–408.

Harvey, C. R. (2017) Presidential address: the scientific outlook in financial economics, *Journal of Finance* Forthcoming.

Harvey, C. R., Liu, Y., Zhu, H. (2016) . . . and the cross-section of expected returns, *Review of Financial Studies* **29**, 5–68.

Jagannathan, R., and Wang, Y. (2007) Lazy investors, discretionary consumption, and the cross-section of stock returns, *The Journal of Finance* **62**, 1623–1661.

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

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

Kaplan, G., Moll, B., and Violante, G. L. (2016) Monetary policy according to HANK. Unpublished Working paper.

Koijen, R. et al. (2015) Carry. Unpublished Working paper.

Krishnamurthy, A., and He, Z. (2013) Intermediary asset pricing, *American Economic Review* **103**, 732–770.

Kydland, F. E., and Prescott, E. C. (1982) Time to build and aggregate fluctuations, *Econometrica* **50**, 1345–1370.

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

Lettau, M., and Wachter, J. A. (2007) Why is long-horizon equity less risky? A duration-based explanation of the value premium, *Journal of Finance* **62**, 55–92.

Linnainmaa, J. T., and Roberts, M. (2016) The history of the cross section of stock returns? Unpublished Working paper.

Lopez, P., Lopez-Salido, D., and Vazquez-Grande, F. (2015) Nominal rigidities and the term structures of equity and bond returns. Unpublished Working paper.

McCloskey, D. (1983) The rhetoric of economics, *Journal of Economic Literature* **31**, 482–517.

Mclean, R. D., and Pontiff, J. (2016) Does academic research destroy stock return predictability?, *The Journal of Finance* **71**, 5–32.

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

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

Mian, A., and Sufi, A. (2014) *House of Debt: How They (and You) Caused the Great Recession, and How We Can Prevent It from Happening Again*, University of Chicago Press, Chicago, Illinois.

Piazzesi, M., Schneider, M., and Tuzel, S. (2007) Housing, consumption, and asset pricing, *Journal of Financial Economics* **83**, 531–569.

Ravinia, E. (2011) Habit formation and keeping up with the Joneses: evidence from micro data. Unpublished Working paper.

Reitz, T. A. (1988) The equity risk premium: a solution, *Journal of Monetary Economics* **22**, 117–131.

Santos, T., and Veronesi, P. (2010) Habit formation, the cross section of stock returns and the cash flow risk puzzle, *Journal of Financial Economics* **98**, 385–413.

Schmidt, L. (2015) Climbing and falling off the ladder: asset pricing implications of labor market event risk. Unpublished working paper.

Schulz, F. (2016) On the timing and pricing of dividends: comment, *American Economic Review* **106**, 3185–3223.

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

Shiller, R. J. (2014) Speculative asset prices, *American Economic Review* **104**, 1486–1517.

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

Verdelhan, A. (2010) A habit-based explanation of the exchange rate risk premium, *Journal of Finance* **65**, 123–145.

Wachter, J. A. (2006) A consumption-based model of the term structure of interest rates, *Journal of Financial Economics* **79**, 365–399.

Wachter, J. A. (2013) Can time-varying risk of rare disasters explain aggregate stock market volatility?, *Journal of Finance* **68**, 987–1035.

Welch, I. (2016) The (time-varying) importance of disaster risk, *Financial Analysts Journal* **72**, 14–30.