

Sources of Entropy in Representative Agent Models

DAVID BACKUS, MIKHAIL CHERNOV, and STANLEY ZIN*

ABSTRACT

We propose two data-based performance measures for asset pricing models and apply them to models with recursive utility and habits. Excess returns on risky securities are reflected in the pricing kernel's dispersion and riskless bond yields are reflected in its dynamics. We measure dispersion with entropy and dynamics with horizon dependence, the difference between entropy over several periods and one. We compare their magnitudes to estimates derived from asset returns. This exercise reveals tension between a model's ability to generate one-period entropy, which should be large, and horizon dependence, which should be small.

WE HAVE SEEN SIGNIFICANT PROGRESS in the recent past in research linking asset returns to macroeconomic fundamentals. Existing models provide quantitatively realistic predictions for the mean, variance, and other moments of asset returns from similarly realistic macroeconomic inputs. The most popular models have representative agents, with prominent examples based on recursive utility, including long-run risk, and habits, both internal and external. Recursive utility and habits are different preference orderings, but they share one important feature: dynamics play a central role. With recursive preferences, dynamics in the consumption growth process are required to distinguish them from additive power utility. With habits, dynamics enter preferences directly. The question we address is whether these dynamics, which are essential to explaining average excess returns, are realistic along other dimensions.

What other dimensions, you might ask. We propose two performance measures that summarize the behavior of asset pricing models. We base them on the pricing kernel, because every arbitrage-free model has one. One measure

*Backus and Zin are from New York University and NBER, and Chernov is from UCLA Anderson School and CEPR. We are grateful to many people for help with this project, including: Jarda Borovicka, Nina Boyarchenko, Adam Brandenburger, Wayne Ferson, Lars Hansen, Christian Heyerdahl-Larsen, Hanno Lustig, Ian Martin, Monika Piazzesi, Bryan Routledge, Andrea Tamoni, and Harald Uhlig as well as participants in seminars at, and conferences sponsored by, AHL, CEPR, CERGE, Columbia, CREATES/SoFiE, Duke, ECB, Federal Reserve Board, Federal Reserve Banks of Atlanta, Minneapolis, and San Francisco, Geneva, IE Business School, LSE, LUISS Guido Carli University, Minnesota, NBER, NYU, Penn State, Reading, SED, SIFR, and USC. We also thank Campbell Harvey, an Associate Editor, and two referees for helpful comments on earlier versions.

DOI: 10.1111/jofi.12090

concerns the pricing kernel's dispersion, which we capture with *entropy*. We show that the (one-period) entropy of the pricing kernel is an upper bound on mean excess returns (also over one period). The second measure concerns the pricing kernel's dynamics. We summarize dynamics with what we call *horizon dependence*, a measure of how entropy varies with the investment horizon. As with entropy, we can infer its magnitude from asset prices: negative (positive) horizon dependence is associated with an increasing (decreasing) mean yield curve and positive (negative) mean yield spreads.

The approach is similar in spirit to Hansen and Jagannathan (1991), who compare properties of theoretical models to those implied by observed returns. In their case, the property is the standard deviation of the pricing kernel. In ours, the properties are entropy and horizon dependence. Entropy is a measure of dispersion, a generalization of variance. Horizon dependence has no counterpart in the Hansen–Jagannathan methodology. We think it captures the dynamics essential to representative agent models in a convenient and informative way.

Concepts of entropy have proved useful in a wide range of fields, so it is not surprising they have started to make inroads into economics and finance. We find entropy-based measures to be natural tools for our purpose. One reason is that entropy extends more easily to multiple periods than, say, the standard deviation of the pricing kernel. Similar reasoning underlies the treatment of long-horizon returns in Alvarez and Jermann (2005), Hansen (2012), and Hansen and Scheinkman (2009). A second reason is that many popular asset pricing models are loglinear, or nearly so. Logarithmic measures like entropy and log-differences in returns are easily computed for them. Finally, entropy extends to nonnormal distributions of the pricing kernel and returns in a simple and transparent way. All of this will be clearer once we have developed the appropriate tools.

Our performance measures give us new insight into the behavior of popular asset pricing models. The evidence suggests that a realistic model should have substantial one-period entropy (to match observed mean excess returns) and modest horizon dependence (to match observed differences between mean yields on long and short bonds). In models with recursive preferences or habits, the two features are often linked: dynamic ingredients designed to increase the pricing kernel's entropy often generate excessive horizon dependence.

This tension between entropy and horizon dependence is a common feature: to generate enough of the former we end up with too much of the latter. We illustrate this tension and point to ways of resolving it. One is illustrated by the Campbell–Cochrane (1999) model: offsetting effects of a state variable on the conditional mean and variance of the log pricing kernel. Entropy comes from the conditional variance and horizon dependence comes from both, which allows us to hit both targets. Another approach is to introduce jumps, that is, nonnormal innovations in consumption growth. Asset returns are decidedly nonnormal, so it seems natural to allow the same in asset pricing models. Jumps can be added to either class of models. With recursive utility, jump risk

can increase entropy substantially. Depending on their dynamic structure, they can have either a large or modest impact on horizon dependence.

All of these topics are developed later. We use closed-form loglinear approximations throughout to make all the moving parts visible. We think this brings us some useful intuition even in models that have been explored extensively elsewhere.

We use a number of conventions to keep the notation, if not simple, as simple as possible. (i) For the most part, Greek letters are parameters and Latin letters are variables or coefficients. (ii) We use a t subscript (x_t , for example) to represent a random variable and the same letter without a subscript (x) to represent its mean. In some cases, $\log x$ represents the mean of $\log x_t$ rather than the log of the mean of x_t , but the subtle difference between the two has no bearing on anything important. (iii) The term B is the backshift or lag operator, shifting what follows back one period: $Bx_t = x_{t-1}$, $B^k x_t = x_{t-k}$, and so on. (iv) Lag polynomials are one-sided and possibly infinite: $a(B) = a_0 + a_1 B + a_2 B^2 + \dots$. (v) The expression $a(1)$ is the same polynomial evaluated at $B = 1$, which generates the sum $a(1) = \sum_j a_j$.

I. Properties of Pricing Kernels

In modern asset pricing theory, a pricing kernel accounts for asset returns. The reverse is also true: asset returns contain information about the pricing kernel that gave rise to them. We summarize some well-known properties of asset returns, show what they imply for the entropy of the pricing kernel over different time horizons, and illustrate the entropy consequences of fitting a loglinear model to bond yields.

A. Properties of Asset Returns

We begin with a summary of the salient properties of excess returns. In Table I, we report the sample mean, standard deviation, skewness, and excess kurtosis of monthly excess returns on a diverse collection of assets. None of this evidence is new, but it is helpful to collect it in one place. Excess returns are measured as differences in logs of gross U.S. dollar returns over the 1-month Treasury.

We see, first, the equity premium. The mean excess return on a broad-based equity index is $0.0040 = 0.40\%$ per month or 4.8% a year. This return comes with risk: its sample distribution has a standard deviation of 0.05 , skewness of -0.4 , and excess kurtosis of 7.9 . Nonzero values of skewness and excess kurtosis are an indication that excess returns on the equity index are not normal.

Other equity portfolios exhibit a range of behavior. Some have larger mean excess returns and come with larger standard deviations and excess kurtosis. Consider the popular Fama and French (1992) portfolios, constructed from a five-by-five matrix of stocks sorted by size (small to large) and book-to-market (low to high). Small firms with high book-to-market have mean excess returns more than twice the equity premium (0.90% per month). Option strategies

Table I
Properties of Monthly Excess Returns

Entries are sample moments of monthly observations of (monthly) log excess returns: $\log r - \log r^1$, where r is a (gross) return and r^1 is the return on a 1-month bond. Sample periods: S&P 500, 1927 to 2008 (source: CRSP), Fama–French (1992), 1927 to 2008 (source: Kenneth French’s website); nominal bonds, 1952 to 2008 (source: Fama–Bliss data set, CRSP); currencies, 1985 to 2008 (source: Datastream); options, 1987 to 2005 (source: Broadie, Chernov, and Johannes (2009)). For options, OTM means out-of-the-money and ATM means at-the-money

Asset	Mean	Standard Deviation	Skewness	Excess Kurtosis
<i>Equity</i>				
S&P 500	0.0040	0.0556	−0.40	7.90
Fama–French (small, low)	−0.0030	0.1140	0.28	9.40
Fama–French (small, high)	0.0090	0.0894	1.00	12.80
Fama–French (large, low)	0.0040	0.0548	−0.58	5.37
Fama–French (large, high)	0.0060	0.0775	−0.64	11.57
<i>Equity Options</i>				
S&P 500 6% OTM puts (delta-hedged)	−0.0184	0.0538	2.77	16.64
S&P 500 ATM straddles	−0.6215	1.1940	−1.61	6.52
<i>Currencies</i>				
CAD	0.0013	0.0173	−0.80	4.70
JPY	0.0001	0.0346	0.50	1.90
AUD	−0.0015	0.0332	−0.90	2.50
GBP	0.0035	0.0316	−0.50	1.50
<i>Nominal Bonds</i>				
1 year	0.0008	0.0049	0.98	14.48
2 years	0.0011	0.0086	0.52	9.55
3 years	0.0013	0.0119	−0.01	6.77
4 years	0.0014	0.0155	0.11	4.78
5 years	0.0015	0.0190	0.10	4.87

(buying out-of-the-money puts and at-the-money straddles on the S&P 500 index) have large negative excess returns, suggesting that short positions will have large positive returns, on average. Both exhibit substantial skewness and excess kurtosis.

Currencies have smaller mean excess returns and standard deviations but comparable excess kurtosis, although more sophisticated currency strategies have been found to generate large excess returns. Here we see that buying the British pound generates substantial excess returns in this sample.

Bonds have smaller mean excess returns than the equity index. About half the excess return of the 5-year U.S. Treasury bond over the 1-month Treasury bill (0.15% in our sample) is evident in the 1-year bond (0.08%). The increase in mean excess returns with maturity corresponds to a mean yield curve that also increases with maturity over this range.

The mean spread between yields on 1-month and 10-year Treasuries over the last four decades has been about 1.5% annually or 0.125% monthly. Alvarez and Jermann (2005, Section 4) show that mean excess returns and yield spreads are somewhat smaller if we consider longer samples, longer maturities, or evidence

from the United Kingdom. All of these numbers refer to nominal bonds. Data on inflation-indexed bonds are available for only a short sample and a limited range of maturities, leaving some range of opinion about their properties. However, none of the evidence suggests that the absolute magnitudes, whether positive or negative, are significantly greater than we see for nominal bonds. Chernov and Mueller (2012) suggest instead that yield spreads are about half as large on real bonds, which would make our estimates upper bounds.

These properties of returns are estimates, but they are suggestive of the facts a theoretical model might try to explain. Such facts are as follows. (i) Many assets have positive mean excess returns, and some have returns substantially greater than a broad-based equity index such as the S&P 500. We use a lower bound of $0.0100 = 1\%$ per month. The exact number is not critical, but it is helpful to have a clear numerical benchmark. (ii) Excess returns on long bonds are smaller than excess returns on an equity index and positive for nominal bonds. We are agnostic about the sign of mean yield spreads, but suggest they are unlikely to be larger than $0.0010 = 0.1\%$ monthly in absolute value. (iii) Excess returns on many assets are decidedly nonnormal.

B. Entropy

Our goal is to connect these properties of excess returns to features of pricing kernels. We summarize these features using entropy, a concept that has been applied productively in such disparate fields as physics, information theory, statistics, and (increasingly) economics and finance. Among notable examples of the latter, Hansen and Sargent (2008) use entropy to quantify ambiguity; Sims (2003) and Van Nieuwerburgh and Veldkamp (2010) use it to measure learning capacity; and Ghosh, Julliard, and Taylor (2011) and Stutzer (1996) use it to limit differences between true and risk-adjusted probabilities subject to pricing assets correctly.

The distinction between true and risk-adjusted probabilities is central to asset pricing. Consider a Markovian environment based on a state variable x_t . We denote (true) probability of the state at date $t + 1$ conditional on the state at date t by $p_{t,t+1} = p(x_{t+1} | x_t)$. We use $p_{t,t+n}$ as shorthand notation for $\prod_{j=1}^n p_{t+j-1,t+j} = \prod_{j=1}^n p(x_{t+j} | x_{t+j-1})$. Similarly, $p_{t,t+n}^*$ is the analogous risk-adjusted probability. The *relative entropy* of the risk-adjusted distribution is then

$$L_t(p_{t,t+n}^*/p_{t,t+n}) = -E_t \log(p_{t,t+n}^*/p_{t,t+n}),$$

where E_t is the conditional expectation based on the true distribution. This object, sometimes referred to as the Kullback–Leibler divergence, quantifies the difference between the two probability distributions. In the next subsection, we refer to it as conditional entropy, but the distinction is more than we need here.

Intuitively, we associate large risk premiums with large differences between true and risk-adjusted probabilities. One way to capture this difference is with

a log-likelihood ratio. For instance, we could use the log-likelihood ratio to test the null model p against the alternative p^* . A large statistic is evidence against the null and thus suggests significant prices of risk. Entropy is the population value of this statistic.

Another way to look at the same issue is to associate risk premiums with variability in the ratio $p_{t,t+n}^*/p_{t,t+n}$. Entropy captures this notion as well. Because $E_t(p_{t,t+n}^*/p_{t,t+n}) = 1$, we can rewrite entropy as

$$L_t(p_{t,t+n}^*/p_{t,t+n}) = \log E_t(p_{t,t+n}^*/p_{t,t+n}) - E_t \log(p_{t,t+n}^*/p_{t,t+n}). \quad (1)$$

If the ratio is constant, it must equal one and entropy is zero. The concavity of the log function tells us that entropy is nonnegative and increases with variability, in the sense of a mean-preserving spread to the ratio $p_{t,t+n}^*/p_{t,t+n}$. These properties are consistent with a measure of dispersion.

We think the concept of entropy is useful here because of its properties. It is connected to excess returns on assets and real bond yields in a convenient way. This allows us to link theoretical models to data in a constructive manner. We make these ideas precise in the next section.

C. Entropy over Short and Long Horizons

Entropy, suitably defined, supplies an upper bound on mean excess returns and a measure of the dynamics of the pricing kernel. The foundation for both results is a stationary environment and the familiar no-arbitrage theorem: in environments that are free of arbitrage opportunities, there is a positive random variable $m_{t,t+n}$ that satisfies

$$E_t(m_{t,t+n}r_{t,t+n}) = 1, \quad (2)$$

for any positive time interval n . Here, $m_{t,t+n}$ is the *pricing kernel* over the period t to $t+n$ and $r_{t,t+n}$ is the gross return on a traded asset over the same period.

Both can be decomposed into one-period components, $m_{t,t+n} = \prod_{j=1}^n m_{t+j-1,t+j}$ and $r_{t,t+n} = \prod_{j=1}^n r_{t+j-1,t+j}$.

We approach entropy by a somewhat different route from the previous section. We also scale it by the time horizon n . We define *conditional entropy* by

$$L_t(m_{t,t+n}) = \log E_t m_{t,t+n} - E_t \log m_{t,t+n}. \quad (3)$$

We connect this to our earlier definition using the relation between the pricing kernel and conditional probabilities: $m_{t,t+n} = q_t^n p_{t,t+n}^*/p_{t,t+n}$, where $q_t^n = E_t m_{t,t+n}$ is the price of an n -period bond (a claim to “one” in n periods). Since (3) is invariant to scaling (the multiplicative factor q_t^n), it is equivalent to (1). Mean conditional entropy is

$$EL_t(m_{t,t+n}) = E \log E_t m_{t,t+n} - E \log m_{t,t+n},$$

where E is the expectation based on the stationary distribution. If we scale this by the time horizon n , we have mean conditional entropy per period:

$$I(n) = n^{-1} E L_t(m_{t,t+n}). \quad (4)$$

We refer to this simply as *entropy* from here on. We develop this definition of entropy in two directions, the first focusing on its value over one period, the second on how it varies with time horizon n .

Our first result, which we refer to as the *entropy bound*, connects one-period entropy to one-period excess returns:

$$I(1) = E L_t(m_{t,t+1}) \geq E (\log r_{t,t+1} - \log r_{t,t+1}^1), \quad (5)$$

where $r_{t,t+1}^1 = 1/q_t^1$ is the return on a one-period bond. In words: mean excess log returns are bounded above by the (mean conditional) entropy of the pricing kernel. The bound tells us entropy can be expressed in units of log returns per period.

The entropy bound (5) starts with the pricing relation (2) and the definition of conditional entropy (3). Since log is a concave function, the pricing relation (2) and Jensen's inequality imply that for any positive return $r_{t,t+n}$,

$$E_t \log m_{t,t+n} + E_t \log r_{t,t+n} \leq \log(1) = 0, \quad (6)$$

with equality if and only if $m_{t,t+n} r_{t,t+n} = 1$. This is the conditional version of an inequality reported by Bansal and Lehmann (1997, Section 2.3) and Cochrane (1992, Section 3.2). The log return with the highest mean is, evidently, $\log r_{t,t+n} = -\log m_{t,t+n}$.

The first term in (6) is one component of conditional entropy. The other is $\log E_t m_{t,t+n} = \log q_t^n$. We set $n = 1$ in (3) and note that $r_{t,t+1}^1 = 1/q_t^1$ and $\log E_t m_{t,t+1} = \log q_t^1 = -\log r_{t,t+1}^1$. If we subtract this from (6), we have

$$L_t(m_{t,t+1}) \geq E_t \log r_{t,t+1} - \log r_{t,t+1}^1. \quad (7)$$

We take the expectation of both sides to produce the entropy bound (5).

The relation between one-period entropy and the conditional distribution of $\log m_{t,t+1}$ is captured in a convenient way by its cumulant generating function and cumulants. The conditional cumulant generating function of $\log m_{t,t+1}$ is

$$k_t(s) = \log E_t (e^{s \log m_{t,t+1}}),$$

the log of the moment generating function. Conditioning is indicated by the subscript t . With the appropriate regularity conditions, it has the power series expansion

$$k_t(s) = \sum_{j=1}^{\infty} \kappa_{jt} s^j / j!$$

over some suitable range of s . The conditional cumulant κ_{jt} is the j th derivative of $k_t(s)$ at $s = 0$; κ_{1t} is the mean, κ_{2t} is the variance, and so on. The third

and fourth cumulants capture skewness and excess kurtosis, respectively. If the conditional distribution of $\log m_{t,t+1}$ is normal, then high-order cumulants (those of order $j \geq 3$) are zero. In general we have

$$L_t(m_{t,t+1}) = k_t(1) - \kappa_{1t} \\ = \underbrace{\kappa_{2t}(\log m_{t,t+1})/2!}_{\text{normal term}} + \underbrace{\kappa_{3t}(\log m_{t,t+1})/3! + \kappa_{4t}(\log m_{t,t+1})/4! + \dots}_{\text{nonnormal terms}}, \quad (8)$$

a convenient representation of the potential role played by departures from normality. We take the expectation with respect to the stationary distribution to convert this to one-period entropy.

Our second result, which we refer to as *horizon dependence*, uses the behavior of entropy over different time horizons to characterize the dynamics of the pricing kernel. We define horizon dependence as the difference in entropy over horizons of n and one, respectively:

$$H(n) = I(n) - I(1) = n^{-1}EL_t(m_{t,t+n}) - EL_t(m_{t,t+1}). \quad (9)$$

To see how this works, consider a benchmark in which successive one-period pricing kernels $m_{t,t+1}$ are iid (independent and identically distributed). Then mean conditional entropy over n periods is simply a scaled-up version of one-period entropy,

$$EL_t(m_{t,t+n}) = nEL_t(m_{t,t+1}).$$

This is a generalization of a well-known property of random walks: the variance is proportional to the time interval. As a result, entropy $I(n)$ is the same for all n and horizon dependence is zero. In other cases, horizon dependence reflects departures from the iid case, and in this sense is a measure of the pricing kernel's dynamics. It captures not only the autocorrelation of the log pricing kernel, but variations in all aspects of the conditional distribution. This will become apparent when we study models with stochastic variance and jumps, in Sections II.C and II.D, respectively.

Perhaps the most useful feature of horizon dependence is that it is observable, in principle, through its connection to bond yields. In a stationary environment, conditional entropy over n periods is

$$L_t(m_{t,t+n}) = \log E_t m_{t,t+n} - E_t \log m_{t,t+n} = \log q_t^n - E_t \sum_{j=1}^n \log m_{t+j-1,t+j}.$$

Entropy (4) is therefore

$$I(n) = n^{-1}E \log q_t^n - E \log m_{t,t+1}.$$

Bond yields are related to prices by $y_t^n = -n^{-1} \log q_t^n$; see Appendix A. Therefore, horizon dependence is related to mean yield spreads by

$$H(n) = -E(y_t^n - y_t^1).$$

In words: horizon dependence is negative if the mean yield curve is increasing, positive if it is decreasing, and zero if it is flat. Since mean forward rates and returns are closely related to mean yields, we can express horizon dependence with them too. See Appendix A.

Entropy and horizon dependence give us two properties of the pricing kernel that we can quantify with asset prices. Observed excess returns tell us that one-period entropy is probably greater than 1% monthly. Observed bond yields tell us that horizon dependence is smaller, probably less than 0.1% at observable time horizons. We use these bounds as diagnostics for candidate pricing kernels. The exercise has the same motivation as Hansen and Jagannathan (1991), but extends their work in looking at pricing kernels' dynamics as well as dispersion.

D. Related Approaches

Our entropy bound and horizon dependence touch on issues and approaches addressed in other work. A summary follows.

The entropy bound (5), like the Hansen–Jagannathan (1991) bound, produces an upper bound on excess returns from the dispersion of the pricing kernel. In this broad sense the ideas are similar, but the bounds use different measures of dispersion and excess returns. They are not equivalent and neither is a special case of the other. One issue is extending these results to different time intervals. The relationship between entropy at two different horizons is easily computed, a byproduct of judicious use of the log function. The Hansen–Jagannathan bound, on the other hand, is not. Another issue is the role of departures from lognormality, which are easily accommodated with entropy. These and related issues are explored further in Appendix B.

Closer to our work is a bound derived by Alvarez and Jermann (2005). Ours differs from theirs in using conditioning information. The conditional entropy bound (7) characterizes the maximum excess return as a function of the state at date t . Our definition of entropy is the mean across such states. Alvarez and Jermann (2005, Section 3) derive a similar bound based on unconditional entropy,

$$L(m_{t,t+1}) = \log E m_{t,t+1} - E \log m_{t,t+1}.$$

The two are related by

$$L(m_{t,t+1}) = E L_t(m_{t,t+1}) + L(E_t m_{t,t+1}).$$

There is a close analog for the variance: the unconditional variance of a random variable is the mean of its conditional variance plus the variance of its conditional mean. This relation converts (5) into an “Alvarez–Jermann bound,”

$$L(m_{t,t+1}) \geq E (\log r_{t,t+1} - \log r_{t,t+1}^1) + L(E_t m_{t,t+1}),$$

a component of their Proposition 2. Our bound is tighter, but since the last term is usually small, it is not a critical issue in practice. More important to us

is that our use of mean conditional entropy provides a link to bond prices and yields.

Also related is an influential body of work on long-horizon dynamics that includes notable contributions from Alvarez and Jermann (2005), Hansen and Scheinkman (2009), and Hansen (2012). Hansen and Scheinkman (2009, Section 6) show that, since pricing is a linear operation, logic following the Perron and Frobenius theorem tells us there exist a positive eigenvalue λ and associated positive eigenfunction e that solve

$$E_t(m_{t,t+1}e_{t+1}) = \lambda e_t. \quad (10)$$

As before, subscript t denotes dependence on the state at date t ; e_t , for example, stands for $e(x_t)$.

One consequence is Alvarez and Jermann's (2005) multiplicative decomposition of the pricing kernel into $m_{t,t+1} = m_{t,t+1}^1 m_{t,t+1}^2$, where

$$\begin{aligned} m_{t,t+1}^1 &= m_{t,t+1} e_{t+1} / (\lambda e_t), \\ m_{t,t+1}^2 &= \lambda e_t / e_{t+1}. \end{aligned}$$

They refer to the components as permanent and transitory, respectively. By construction, $E_t m_{t,t+1}^1 = 1$. They also show that $1/m_{t,t+1}^2 = r_{t,t+1}^\infty$, the one-period return on a bond of infinite maturity. The mean log return is therefore $E \log r_{t,t+1}^\infty = -\log \lambda$. Long bond yields and forward rates converge to the same value. Hansen and Scheinkman (2009) suggest a three-way decomposition of the pricing kernel into a long-run discount factor λ , a multiplicative martingale component $m_{t,t+1}^1$, and a ratio of positive functionals e_t/e_{t+1} . Hansen (2012) introduces an additive decomposition of $\log m_{t,t+1}$ and identifies permanent shocks with the additive counterpart to $m_{t,t+1}^1$.

Alvarez and Jermann summarize the dynamics of pricing kernels by constructing a lower bound for $L(m_{t,t+1}^1)/L(m_{t,t+1})$. Bakshi and Chabi-Yo (2012) refine this bound. More closely related to what we do is an exact relation between the entropy of the pricing kernel and its first component:

$$EL_t(m_{t,t+1}) = EL_t(m_{t,t+1}^1) + E(\log r_{t,t+1}^\infty - \log r_{t,t+1}^1);$$

see Alvarez and Jermann (2005, proof of Proposition 2). Since the term on the left is big (at least 1% monthly by our calculations) and the one on the far right is small (say, 0.1% or smaller), most entropy must come from their first component. The term structure shows up here in the infinite-maturity return, but Alvarez and Jermann do not develop the connection between entropy and bond yields further.

Another consequence is an alternative route to long-horizon entropy: entropy for an infinite time horizon. This line of work implies, in our terms,

$$I(\infty) = \log \lambda - E \log m_{t,t+1}. \quad (11)$$

We now have the two ends of the entropy spectrum. The short end $I(1)$ is the essential ingredient of our entropy bound (5). The long end $I(\infty)$ is given

by equation (11). Horizon dependence $H(n) = I(n) - I(1)$ describes how we get from one to the other as we vary the time horizon n .

E. An Example: The Vasicek Model

We illustrate entropy and horizon dependence in a loglinear example, a modest generalization of the Vasicek (1977) model. The pricing kernel is

$$\log m_{t,t+1} = \log m + \sum_{j=0}^{\infty} a_j w_{t+1-j} = \log m + a(B)w_{t+1}, \quad (12)$$

where $a_0 > 0$ (a convention), $\sum_j a_j^2 < \infty$ (“square summable”), and B is the lag or backshift operator. The lag polynomial $a(B)$ is described in Appendix C along with some of its uses. The innovations w_t are iid with mean zero, variance one, and (arbitrary) cumulant generating function $k(s) = \log E(e^{sw_t})$. The infinite moving average gives us control over the pricing kernel’s dynamics. The cumulant generating function gives us similar control over the distribution.

The pricing kernel dictates bond prices and related objects; see Appendix A. The solution is most easily expressed in terms of forward rates, which are connected to bond prices by $f_t^n = \log(q_t^n/q_t^{n+1})$ and yields by $y_t^n = n^{-1} \sum_{j=1}^n f_t^{j-1}$. Forward rates in this model are

$$-f_t^n = \log m + k(A_n) + [a(B)/B^n]_+ w_t \quad (13)$$

for $n \geq 0$ and $A_n = \sum_{j=0}^n a_j$; see Appendix D. The subscript “+” means ignore negative powers of B . Mean forward rates are therefore $-E(f_t^n) = \log m + k(A_n)$. Mean yields follow as averages of forward rates: $-E(y_t^n) = \log m + n^{-1} \sum_{j=1}^n k(A_{j-1})$.

In this setting, the initial coefficient (a_0) governs one-period entropy and the others (a_j for $j \geq 1$) combine with it to govern horizon dependence. Entropy is

$$I(n) = n^{-1} E L_t(m_{t,t+n}) = n^{-1} \sum_{j=1}^n k(A_{j-1})$$

for any positive time horizon n . Horizon dependence is therefore

$$H(n) = I(n) - I(1) = n^{-1} \sum_{j=1}^n [k(A_{j-1}) - k(A_0)].$$

Here we see the role of dynamics. In the iid case ($a_j = 0$ for $j \geq 1$), $A_j = A_0 = a_0$ for all j and horizon dependence is zero at all horizons. Otherwise horizon dependence depends on the relative magnitudes of $k(A_{j-1})$ and $k(A_0)$. We also see the role of the distribution of w_t . Our benchmarks suggest $k(A_0)$ is big (at least $0.0100 = 1\%$ monthly) and $k(A_{j-1}) - k(A_0)$ is small (no larger than $0.0010 = 0.1\%$ on average). The latter requires, in practice, small differences between A_0 and A_{j-1} , and hence small values of a_j .

We see more clearly how this works if we add some structure and choose parameter values to approximate the salient features of interest rates. We make $\log m_{t,t+1}$ an ARMA(1,1) process. Its three parameters are (a_0, a_1, φ) , with $a_0 > 0$ and $|\varphi| < 1$ (to ensure square summability). They imply moving average coefficients $a_{j+1} = \varphi a_j$ for $j \geq 1$; see Appendix C. This leads to an AR(1) for the short rate, which turns the model into a legitimate discrete-time version of Vasicek. We choose φ and a_1 to match the autocorrelation and variance of the short rate and a_0 to match the mean spread between 1-month and 10-year bonds. The result is a statistical model of the pricing kernel that captures some of its central features.

The short rate is $\log r_{t,t+1}^1 = f^0 = y_t^1$. Equation (13) tells us that the short rate is AR(1) with autocorrelation φ . We set $\varphi = 0.85$, an estimate of the monthly autocorrelation of the real short rate reported by Chernov and Mueller (2012). The variance of the short rate is

$$\text{Var}(\log r_{t+1}^1) = \sum_{j=1}^{\infty} a_j^2 = a_1^2 / (1 - \varphi^2).$$

Chernov and Mueller report a standard deviation of (0.02/12; 2% annually), which implies $|a_1| = 0.878 \times 10^{-3}$. Neither of these numbers depends on the distribution of w_t .

We choose a_0 to match the mean yield spread on the 10-year bond. This calculation depends on the distribution of w_t through the cumulant generating function $k(s)$. We do this here for the normal case, where $k(s) = s^2/2$, but the calculation is easily repeated for other distributions. If the yield spread is $E(y^{120} - y^1) = 0.0100$, this implies $a_0 = 0.1837$ and $a_1 < 0$. We can reproduce a negative yield spread of similar magnitude by making a_1 positive.

We see the impact of these numbers on the moving average coefficients in Figure 1. The first bar in each pair corresponds to a negative value of a_1 and a positive yield spread, the second bar the reverse. In both cases we see that the initial coefficient a_0 is larger than the others—by two orders of magnitude. It continues well beyond the figure, which we truncate to make the others visible. The only difference is the sign: an upward sloping mean yield curve requires a_0 and a_1 to have opposite signs, a downward sloping curve the reverse.

The configuration of moving average coefficients, with a_0 much larger than the others, means that the pricing kernel is only modestly different from white noise. Stated in our terms: one-period entropy is large relative to horizon dependence. We see that in Figure 2. The dotted line in the middle is our estimated 0.0100 lower bound for one-period entropy. The two thick lines at the top are entropy for the two versions of the model. The dashed one is associated with negative mean yield spreads. We see that entropy rises (slightly) with the horizon. The solid line below it is associated with positive mean yield spreads, which result in a modest decline in entropy with maturity. The dotted lines around them are the horizon dependence bounds: one-period entropy plus and minus 0.0010. The models hit the bounds by construction.

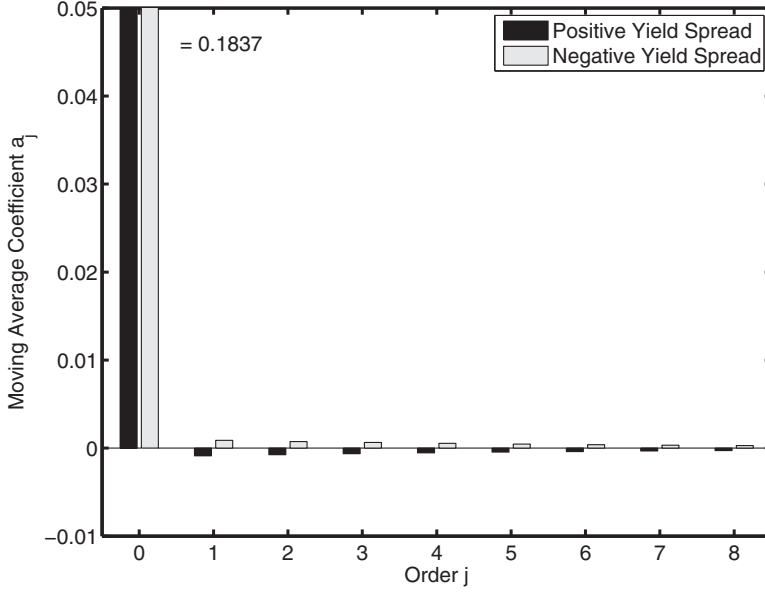


Figure 1. The Vasicek model: moving average coefficients. The bars depict moving average coefficients a_j of the pricing kernel for two versions of the Vasicek model of Section I.E. For each j , the first bar corresponds to parameters chosen to produce a positive mean yield spread, the second to parameters that produce a negative yield spread of comparable size. The initial coefficient a_0 is 0.1837 in both cases, as labeled in the figure. It has been truncated to make the others visible.

The model also provides a clear illustration of long-horizon analysis. The state here is the infinite history of innovations: $x_t = (w_t, w_{t-1}, w_{t-2}, \dots)$. Suppose

$$A_\infty = a(1) = \lim_{n \rightarrow \infty} \sum_{j=0}^n A_n$$

exists. Then the principal eigenvalue λ and eigenfunction e_t are

$$\begin{aligned} \log \lambda &= \log m + k(A_\infty), \\ \log e_t &= \sum_{j=0}^{\infty} (A_\infty - A_j) w_{t-j}. \end{aligned}$$

Long-horizon entropy is $I(\infty) = k(A_\infty)$.

II. Properties of Representative Agent Models

In representative agent models, pricing kernels are marginal rates of substitution. A pricing kernel follows from computing the marginal rate of substitution for a given consumption growth process. We show how this works with

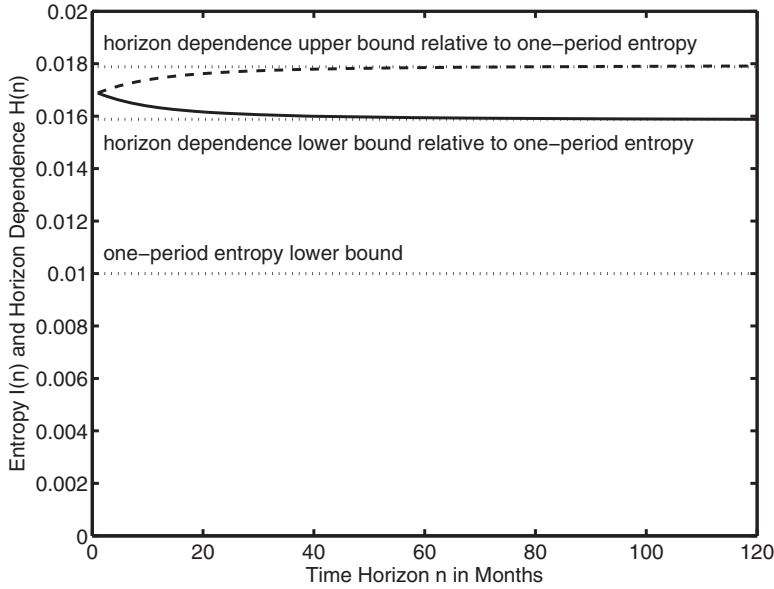


Figure 2. The Vasicek model: entropy and horizon dependence. The lines represent entropy $I(n)$ and horizon dependence $H(n) = I(n) - I(1)$ for two versions of the Vasicek model based on positive and negative mean yield spreads. The dashed line near the top corresponds to a negative mean yield spread and indicates positive horizon dependence. The solid line below it corresponds to a positive mean yield spread and indicates negative horizon dependence. The dotted lines represent bounds on entropy and horizon dependence. The dotted line in the middle is the one-period entropy lower bound (0.0100). The dotted lines near the top are horizon dependence bounds around one-period entropy (plus and minus 0.0010).

several versions of models with recursive utility and habits, the two workhorses of macrofinance. We examine models with dynamics in consumption growth, habits, the conditional variance of consumption growth, and jumps. We report entropy and horizon dependence for each one and compare them to the benchmarks we established earlier.

A. Preferences and Pricing Kernels

Our first class of representative agent models is based on what has come to be known as recursive preferences or utility. The theoretical foundations were laid by Koopmans (1960) and Kreps and Porteus (1978). Notable applications to asset pricing include Bansal and Yaron (2004), Campbell (1993), Epstein and Zin (1989), Garcia, Luger, and Renault (2003), Hansen, Heaton, and Li (2008), Koijen et al. (2009), and Weil (1989).

We define utility recursively with the time aggregator,

$$U_t = \left[(1 - \beta)c_t^\rho + \beta\mu_t(U_{t+1})^\rho \right]^{1/\rho}, \quad (14)$$

and certainty equivalent function,

$$\mu_t(U_{t+1}) = [E_t(U_{t+1}^\alpha)]^{1/\alpha}. \quad (15)$$

Here U_t is “utility from date t on,” or continuation utility. Additive power utility is a special case with $\alpha = \rho$. In standard terminology, $\rho < 1$ captures time preference (with intertemporal elasticity of substitution $1/(1 - \rho)$) and $\alpha < 1$ captures risk aversion (with coefficient of relative risk aversion $1 - \alpha$). The time aggregator and certainty equivalent functions are both homogeneous of degree one, which allows us to scale everything by current consumption. If we define scaled utility $u_t = U_t/c_t$, equation (14) becomes

$$u_t = [(1 - \beta) + \beta \mu_t(g_{t+1} u_{t+1})^\rho]^{1/\rho}, \quad (16)$$

where $g_{t+1} = c_{t+1}/c_t$ is consumption growth. This relation serves, essentially, as a Bellman equation.

With this utility function, the pricing kernel is

$$m_{t,t+1} = \beta g_{t+1}^{\rho-1} [g_{t+1} u_{t+1} / \mu_t(g_{t+1} u_{t+1})]^\alpha. \quad (17)$$

By comparison, the pricing kernel with additive power utility is

$$m_{t,t+1} = \beta g_{t+1}^{\rho-1}. \quad (18)$$

Recursive utility adds another term. It reduces to power utility in two cases: when $\alpha = \rho$ and when g_{t+1} is iid. The latter illustrates the central role of dynamics. If g_{t+1} is iid, u_{t+1} is constant and the pricing kernel is proportional to $g_{t+1}^{\alpha-1}$. This is arguably different from power utility, where the exponent is $\rho - 1$, but with no intertemporal variation in consumption growth we cannot tell the two apart. Beyond the iid case, dynamics in consumption growth introduce an extra term to the pricing kernel—in logs, the innovation in future utility plus a risk adjustment.

Our second class of models introduces dynamics to the pricing kernel directly through preferences. This mechanism has a long history, with applications ranging from microeconomic studies of consumption behavior (Deaton (1992)) to business cycles (Lettau and Uhlig (2000) and Smets and Wouters (2003)). The asset pricing literature includes notable contributions from Abel (1990), Bansal and Lehmann (1997), Campbell and Cochrane (1999), Chan and Kogan (2002), Chapman (2002), Constantinides (1990), Heaton (1995), Otrok, Ravikumar, and Whiteman (2002), and Sundaresan (1989).

All of our habit models start with utility functions that include a state variable h_t that we refer to as the “habit.” A recursive formulation is

$$U_t = (1 - \beta) f(c_t, h_t) + \beta E_t U_{t+1}. \quad (19)$$

Typically h_t is predetermined (known at $t - 1$) and tied to past consumption in some way. Approaches vary, but they all assume h_t/c_t is stationary. The examples we study have “external” habits: the agent ignores any impact of her

consumption choices on future values of h_t . They differ in the functional form of $f(c_t, h_t)$ and in the law of motion for h_t .

Two common functional forms are ratio and difference habits. With ratio habits, $f(c_t, h_t) = (c_t/h_t)^\rho/\rho$ and $\rho \leq 1$. The pricing kernel is

$$m_{t,t+1} = \beta g_{t+1}^{\rho-1} (h_{t+1}/h_t)^{-\rho}. \quad (20)$$

Because the habit is predetermined, it has no impact on one-period entropy. With difference habits, $f(c_t, h_t) = (c_t - h_t)^\rho/\rho$. The pricing kernel becomes

$$m_{t,t+1} = \beta \left(\frac{c_{t+1} - h_{t+1}}{c_t - h_t} \right)^{\rho-1} = \beta g_{t+1}^{\rho-1} (s_{t+1}/s_t)^{\rho-1}, \quad (21)$$

where $s_t = (c_t - h_t)/c_t = 1 - h_t/c_t$ is the surplus consumption ratio. In both cases, we gain an extra term relative to additive power utility.

These models have different properties, but their long-horizon entropies are similar to some version of power utility. Consider models that can be expressed in the form

$$m_{t,t+1} = \beta g_{t+1}^\varepsilon d_{t+1}/d_t, \quad (22)$$

where d_t is stationary and ε is an exponent to be determined. Then long-horizon entropy $I(\infty)$ is the same as for a power utility agent (18) with $\rho - 1 = \varepsilon$. Elements of this proposition are reported by Bansal and Lehmann (1997) and Hansen (2012, Sections 7 and 8).

The proposition follows from the decomposition of the pricing kernel (equation (22)), the definition of the principal eigenvalue and eigenfunction (equation (10)), and the connection between the principal eigenvalue and long-horizon entropy (equation (11)). Suppose an arbitrary pricing kernel $m_{t,t+1}$ has principal eigenvalue λ and associated eigenfunction e_t . Long-horizon entropy is $I(\infty) = \log \lambda - E \log m_{t,t+1}$. Now consider a second pricing kernel $m'_{t,t+1} = m_{t,t+1} d_{t+1}/d_t$, with d_t stationary. The same eigenvalue λ now satisfies (10) with pricing kernel $m'_{t,t+1}$ and eigenfunction $e'_t = e_t/d_t$. Since d_t is stationary, the logs of the two pricing kernels have the same mean: $E \log(m_{t,t+1} d_{t+1}/d_t) = E \log m_{t,t+1}$. Thus, they have the same long-horizon entropy. Power utility is a special case with $m_{t,t+1} = \beta g_{t+1}^\varepsilon$.

We illustrate the impact of this result on our examples, which we review in reverse order. With difference habits, the pricing kernel (21) is already in the form of equation (22) with $\varepsilon = \rho - 1$ and $d_t = s_t^{\rho-1}$. With ratio habits, the pricing kernel (20) does not have the right form, because h_t is not stationary in a growing economy. An alternative is

$$m_{t,t+1} = \beta g_{t+1}^{-1} [(h_{t+1}/c_{t+1})/(h_t/c_t)]^{-\rho},$$

which has the form of (22) with $\varepsilon = -1$ (corresponding to $\rho = 0$, log utility) and $d_t = (h_t/c_t)^{-\rho}$. Bansal and Lehmann (1997, Section 3.4) report a similar decomposition for a model with an internal habit.

Recursive utility can be expressed in approximately the same form. The pricing kernel (17) can be written as

$$m_{t,t+1} = \beta g_{t+1}^{\alpha-1} [u_{t+1}/\mu_t(g_{t+1}u_{t+1})]^{\alpha-\rho}.$$

If μ_t is approximately proportional to u_t , as suggested by Hansen (2012, Section 8.2), then

$$m_{t,t+1} \cong \beta' g_{t+1}^{\alpha-1} (u_{t+1}/u_t)^{\alpha-\rho},$$

where β' includes the constant of proportionality. The change from β to β' is irrelevant here, because entropy is invariant to such changes in scale. Thus, the model has (approximately) the form of (22) with $\varepsilon = \alpha - 1$ and $d_t = u_t^{\alpha-\rho}$.

All of these models are similar to some form of power utility at long horizons. We will see shortly that they can be considerably different at short horizons.

B. Models with Constant Variance

We derive specific pricing kernels for each of these preferences based on loglinear processes for consumption growth and, for habits, the relation between the habit and consumption. When the pricing kernels are not already loglinear, we use loglinear approximations. The resulting pricing kernels have the same form as the Vasicek model. We use normal innovations in our numerical examples to focus attention on the models' dynamics, but consider other distributions at some length in Section II.D. Parameters are representative numbers from the literature chosen to illustrate the impact of preferences on entropy and horizon dependence.

The primary input to the pricing kernels of these models is a consumption growth process. We use the loglinear process

$$\log g_t = \log g + \gamma(B)v^{1/2}w_t, \quad (23)$$

where $\gamma_0 = 1$, $\sum_j \gamma_j^2 < \infty$, and innovations w_t are iid with mean zero, variance one, and cumulant generating function $k(s)$. With normal innovations, $k(s) = s^2/2$.

With power utility (18) and the loglinear consumption growth process (23), the pricing kernel takes the form

$$\log m_{t,t+1} = \text{constant} + (\rho - 1)\gamma(B)v^{1/2}w_{t+1}.$$

Here the moving average coefficients (a_j in Vasicek notation) are proportional to those of the consumption growth process: $a(B) = (\rho - 1)\gamma(B)v^{1/2}$, so $a_j = (\rho - 1)\gamma_j v^{1/2}$ for all $j \geq 0$. The infinite sum is $A_\infty = a(1) = (\rho - 1)\gamma(1)v^{1/2}$.

With recursive utility, we derive the pricing kernel from a loglinear approximation of (16),

$$\log u_t \approx b_0 + b_1 \log \mu_t(g_{t+1}u_{t+1}), \quad (24)$$

a linear approximation of $\log u_t$ in $\log \mu_t$ around the point $\log \mu_t = \log \mu$. See Hansen, Heaton, and Li (2008, Section III). This is exact when $\rho = 0$, in which case $b_0 = 0$ and $b_1 = \beta$, the approximation used to derive long-horizon entropy. With the loglinear approximation (24), the pricing kernel becomes

$$\log m_{t,t+1} = \text{constant} + [(\rho - 1)\gamma(B) + (\alpha - \rho)\gamma(b_1)]v^{1/2}w_{t+1},$$

see Appendix E. The key term is

$$\gamma(b_1) = \sum_{j=0}^{\infty} b_1^j \gamma_j,$$

the impact of an innovation to consumption growth on current utility. The action is in the moving average coefficients. For $j \geq 1$ we reproduce power utility: $a_j = (\rho - 1)\gamma_j v^{1/2}$. The initial term, however, is affected by $\gamma(b_1)$: $a_0 = [(\rho - 1)\gamma_0 + (\alpha - \rho)\gamma(b_1)]v^{1/2}$. If $\gamma(b_1) \neq \gamma_0$, we can make a_0 large and a_j small for $j \geq 1$, as needed, by choosing α and ρ judiciously. The infinite sum is $A_{\infty} = a(1) = \{(\alpha - 1)\gamma(1) + (\alpha - \rho)[\gamma(b_1) - \gamma(1)]\}v^{1/2}$, which is close to the power utility result if $\gamma(b_1) - \gamma(1)$ is small.

With habits, we add the law of motion

$$\log h_{t+1} = \log h + \eta(B)\log c_t.$$

We set $\eta(1) = 1$ to guarantee that h_t/c_t is stationary. For the ratio habit model (20), the log pricing kernel is

$$\log m_{t,t+1} = \text{constant} + [(\rho - 1) - \rho\eta(B)B]\gamma(B)v^{1/2}w_{t+1}.$$

Here $a_0 = (\rho - 1)\gamma_0 v^{1/2}$ and $A_{\infty} = -\gamma(1)v^{1/2}$. The first is the same as power utility with curvature $1 - \rho$, the second is the same as log utility ($\rho = 0$). The other terms combine the dynamics of consumption growth and the habit.

For the difference habit model (21), the challenge lies in transforming the pricing kernel into something tractable. We use a loglinear approximation. Define $z_t = \log(h_t/c_t)$ so that $s_t = 1 - e^{z_t}$. If z_t is stationary with mean $z = \log h - \log c$, then a linear approximation of $\log s_t$ around z is

$$\log s_t \cong \text{constant} - [(1 - s)/s]z_t = \text{constant} - [(1 - s)/s]\log(h_t/c_t),$$

where $s = 1 - h/c = 1 - e^z$ is the surplus ratio corresponding to z . The pricing kernel becomes

$$\log m_{t,t+1} = \text{constant} + (\rho - 1)(1/s)[1 - (1 - s)\eta(B)B]\gamma(B)v^{1/2}w_{t+1}.$$

Campbell (1999, Section 5.1) and Lettau and Uhlig (2000) have similar analyses. Here $a_0 = (\rho - 1)(1/s)\gamma_0 v^{1/2}$, which differs from power utility in the $(1/s)$ term, and $A_{\infty} = (\rho - 1)\gamma(1)v^{1/2}$, which is the same as power utility.

We illustrate the properties of these models with numerical examples based on parameter values used in earlier work. We use the same consumption growth process in all four models, which helps to align their long-horizon properties. We

Table II
Representative Agent Models with Constant Variance

The columns summarize the properties of representative agent pricing kernels when the variance of consumption growth is constant. See Section II.B. The consumption growth process is the same for each one, an ARMA(1,1) version of equation (23) in which $\gamma_{j+1} = \varphi_g \gamma_j$ for $j \geq 1$. Parameter values are $\gamma_0 = 1$, $\gamma_1 = 0.0271$, $\varphi_g = 0.9790$, and $v^{1/2} = 0.0099$.

Parameter or Property	Power Utility (1)	Recursive Utility (2)	Ratio Habit (3)	Difference Habit (4)
<i>Preference Parameters</i>				
ρ	−9	1/3	−9	−9
α	−9	−9		
β	0.9980	0.9980	0.9980	0.9980
φ_h			0.9000	0.9000
s				1/2
<i>Derived Quantities</i>				
b_1		0.9978		
$\gamma(b_1)$		2.165		
$\gamma(1)$		2.290		
$A_0 = a_0$	−0.0991	−0.2069	−0.0991	−0.1983
$A_\infty = a(1)$	−0.2270	−0.2154	−0.0227	−0.2270
<i>Entropy and Horizon Dependence</i>				
$I(1) = EL_t(m_{t,t+1})$	0.0049	0.0214	0.0049	0.0197
$I(\infty)$	0.0258	0.0232	0.0003	0.0258
$H(120) = I(120) - I(1)$	0.0119	0.0011	−0.0042	0.0001
$H(\infty) = I(\infty) - I(1)$	0.0208	0.0018	−0.0047	0.0061

use an ARMA(1,1) that reproduces the mean, variance, and autocorrelations of Bansal and Yaron (2004, Case I); see Appendix I. The moving average coefficients are $\gamma_0 = 1$, $\gamma_1 = 0.0271$, and $\gamma_{j+1} = \varphi_g \gamma_j$ for $j \geq 1$ with $\varphi_g = 0.9790$. This introduces a small but highly persistent component to consumption growth. The mean is $\log g = 0.0015$, the conditional variance is $v = 0.0099^2$, and the (unconditional) variance is 0.01^2 . In the habit models, we use Chan and Kogan’s (2002) AR(1) habit: $\eta_0 = 1 - \varphi_h$ and $\eta_{j+1} = \varphi_h \eta_j$ for $j \geq 0$ and $0 \leq \varphi_h < 1$. We set $\varphi_h = 0.9$, which is between the Chan–Kogan choice of 0.7 and the Campbell–Cochrane (1999) choice of 0.9885. Finally, we set the mean surplus s for the difference habit model equal to one half.

We summarize the properties of these models in Table II (parameters and selected calculations), Figure 3 (moving average coefficients), and Figure 4 (entropy versus time horizon). In each panel of Figure 3, we compare a representative agent model to the Vasicek model of Section I.E. We use absolute values of coefficients in the figure to focus attention on magnitudes.

Consider power utility with curvature $1 - \alpha = 1 - \rho = 10$. The comparison with the Vasicek model suggests that the initial coefficient is too small (note the labels next to the bars) and the subsequent coefficients are too large. As a result, the model has too little one-period entropy and too much horizon dependence. We see exactly that in Figure 4. The solid line at the center of the figure

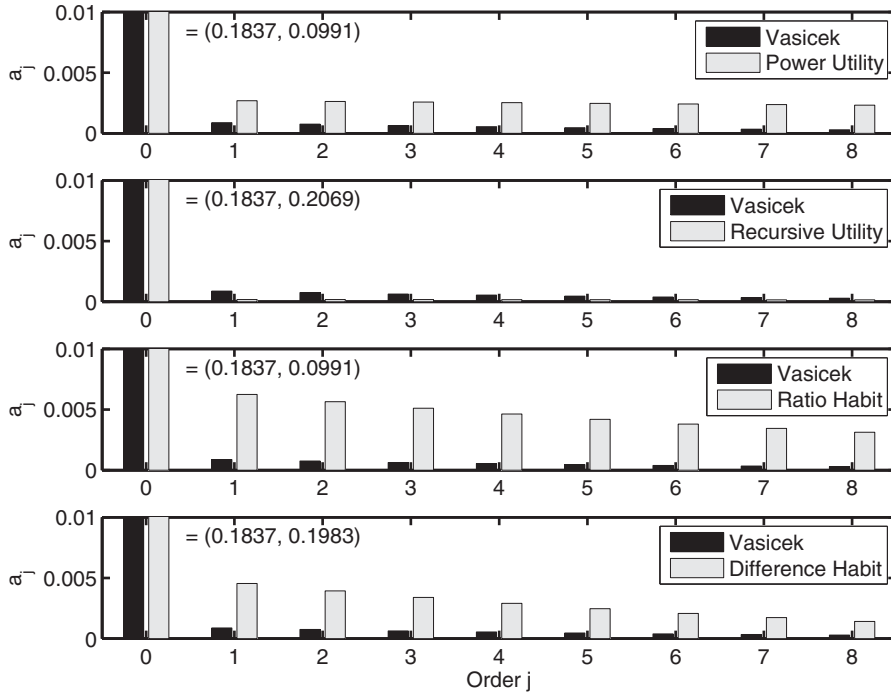


Figure 3. Representative agent models with constant variance: absolute values of moving average coefficients. The bars compare absolute values of moving average coefficients for the Vasicek model of Section I.E and the four representative agent models of Section II.B.

represents entropy for the power utility case with curvature $1 - \alpha = 1 - \rho = 10$. One-period entropy (0.0049) is well below our estimated lower bound (0.0100), the dotted horizontal line near the middle of the figure. Entropy rises quickly as we increase the time horizon, which violates our horizon dependence bounds (plus and minus 0.0010). The bounds are represented by the two dotted lines near the bottom of the figure, centered at power utility's one-period entropy. The model exceeds the bound almost immediately. The increase in entropy with time horizon is, in this case, entirely the result of the positive autocorrelation of the consumption growth process.

The recursive utility model, in contrast, has more entropy at short horizons and less horizon dependence. Here we set $1 - \alpha = 10$ and $1 - \rho = 2/3$, the values used by Bansal and Yaron (2004). Recursive and power utility have similar long-horizon properties, in particular, similar values for $A_\infty = a(1)$, the infinite sum of moving average coefficients. Recursive utility takes some of this total away from later coefficients (a_j for $j \geq 1$) by reducing $1 - \rho$ from 10 to $2/3$, and adds it to the initial coefficient a_0 . As a result, horizon dependence at 120 months falls from 0.0119 with power utility to 0.0011. This is a clear improvement over power utility, but it is still slightly above our bound (0.0010). Further, $H(\infty)$ of 0.0018 hints that entropy at longer horizons is inconsistent

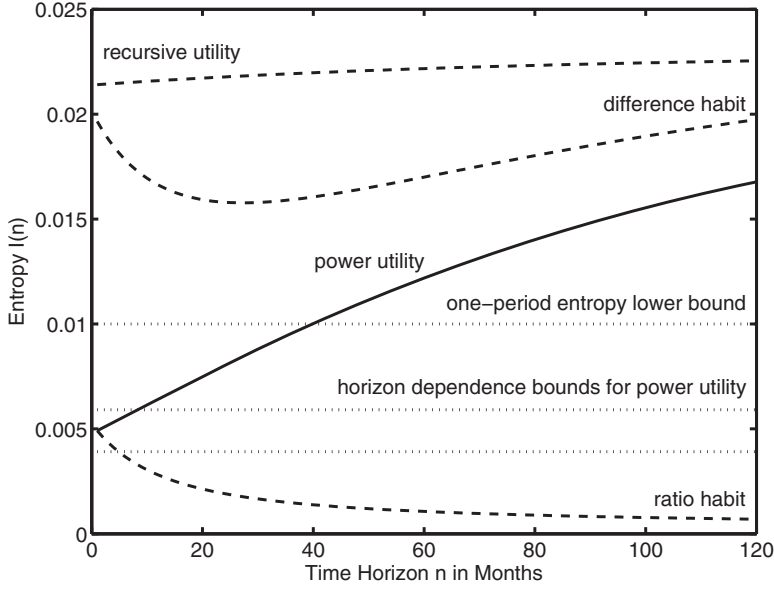


Figure 4. Representative agent models with constant variance: entropy and horizon dependence. The lines plot entropy $I(n)$ against the time horizon n for the representative agent models of Section II.B. The consumption growth process is the same for each one, an ARMA(1,1) version of equation (23) with positive autocorrelations.

with the tendency of long bond yields to level off or decline between 10 and 30 years. See, for example, Alvarez and Jermann (2005, Figure 1).

The difference habit model has greater one-period entropy than power utility (the effect of $1/s$) but the same long-horizon entropy. In between, it has negative horizon dependence, the result of the negative autocorrelation in the pricing kernel induced by the habit. Horizon dependence satisfies our bound at a horizon of 120 months, but violates it for horizons between 4 and 93 months. Relative to power utility, this model reallocates some of the infinite sum A_∞ to the initial term, but it affects subsequent terms in different ways. In our example, the early terms are negative, but later terms turn positive. The result is nonmonotonic behavior of entropy, which is mimicked, of course, by the mean yield spread.

The ratio habit model has, as we noted earlier, the same one-period entropy as power utility with $1 - \rho = 10$. Like the difference habit, it has excessive negative horizon dependence at short horizons, but, unlike that model, the same is true at long horizons, too, as it approaches log utility ($1 - \rho = 1$).

Overall, these models differ in both their one-period entropy and their horizon dependence. They are clearly different from each other. With the parameter values we use, some of them have too little one-period entropy and all of them have too much horizon dependence. The challenge is to clear both hurdles.

C. Models with Stochastic Variance

In the models of the previous section, all of the variability in the distribution of the log pricing kernel is in its conditional mean. Here we consider examples proposed by Bansal and Yaron (2004, Case II) and Campbell and Cochrane (1999) that have variability in the conditional variance as well. They illustrate in different ways how variation in the conditional mean and variance can interact in generating entropy and horizon dependence.

One perspective on the conditional variance comes from recursive utility. The Bansal–Yaron (2004, Case II) model is based on the bivariate consumption growth process:

$$\begin{aligned}\log g_t &= \log g + \gamma(B)v_{t-1}^{1/2}w_{gt}, \\ v_t &= v + v(B)w_{vt},\end{aligned}\tag{25}$$

where w_{gt} and w_{vt} are independent iid standard normal random variables. The first equation governs movements in the conditional mean of log consumption growth, the second movements in the conditional variance.

This linear volatility process is analytically convenient, but it implies that v_t is normal and therefore negative in some states. We think of it as an approximation to a censored process $v'_t = \max\{0, v_t\}$. We show in Appendix G that, if the true conditional variance process is v'_t , then an approximation based on (25) is reasonably accurate for the numerical examples reported later, where the stationary probability that v_t is negative is small.

With this process for consumption growth and the loglinear approximation (24), the Bansal–Yaron pricing kernel is

$$\begin{aligned}\log m_{t,t+1} &= \text{constant} + [(\rho - 1)\gamma(B) + (\alpha - \rho)\gamma(b_1)]v_t^{1/2}w_{gt+1} \\ &\quad + (\alpha - \rho)(\alpha/2)\gamma(b_1)^2[b_1v(b_1) - v(B)B]w_{vt+1},\end{aligned}$$

see Appendix E. The coefficients on the consumption growth innovation w_{gt} now vary with v_t , but they are otherwise the same as before. The volatility innovation w_{vt} is new. Its coefficients depend on the dynamics of volatility (represented by $v(b_1)$), the dynamics of consumption growth ($\gamma(b_1)$), and recursive preferences ($(\alpha - \rho)$). One-period conditional entropy is

$$L_t(m_{t+1}) = [(\rho - 1)\gamma_0 + (\alpha - \rho)\gamma(b_1)]^2 v_t/2 + (\alpha - \rho)^2 (\alpha/2)^2 \gamma(b_1)^4 [b_1 v(b_1)]^2/2,$$

which now varies with v_t . One-period entropy is the same with v_t replaced by its mean v , because the log pricing kernel is linear in v_t .

The pricing kernel looks like a two-shock Vasicek model, but the interaction between the conditional variance and consumption growth innovations gives it a different form. The pricing kernel can be expressed as

$$\log m_{t,t+1} = \log m + a_g(B)(v_t/v)^{1/2}w_{gt+1} + a_v(B)w_{vt+1},$$

Table III
Representative Agent Models with Stochastic Variance

The columns summarize the properties of representative agent pricing kernels with stochastic variance. See Section II.C. Model (1) is recursive utility with a stochastic variance process. Model (2) is the same with more persistent conditional variance. Model (3) is the Campbell and Cochrane (1999) model with their parameter values. Its entropy and horizon dependence do not depend on the discount factor β or variance v

Parameter or Property	Recursive Utility 1 (1)	Recursive Utility 2 (2)	Campbell- Cochrane (3)
<i>Preference Parameters</i>			
ρ	1/3	1/3	−1
α	−9	−9	
β	0.9980	0.9980	
φ_s			0.9885
b			0
<i>Consumption Growth Parameters</i>			
γ_0	1	1	1
γ_1	0.0271	0.0271	
φ_g	0.9790	0.9790	
$v^{1/2}$	0.0099	0.0099	
ν_0	0.23×10^{-5}	0.23×10^{-5}	
φ_v	0.9870	0.9970	
<i>Derived Quantities</i>			
b_1	0.9977	0.9977	
$\gamma(b_1)$	2.164	2.1603	
$\nu(b_1)$	0.0002	0.0004	
<i>Entropy and Horizon Dependence</i>			
$I(1) = EL_t(m_{t,t+1})$	0.0218	0.0249	0.0230
$I(\infty)$	0.0238	0.0293	0.0230
$H(120) = I(120) - I(1)$	0.0012	0.0014	0
$H(\infty) = I(\infty) - I(1)$	0.0020	0.0044	0

with

$$a_g(B) = (\rho - 1)\gamma(B) + (\alpha - \rho)\gamma(b_1),$$

$$a_v(B) = (\alpha - \rho)(\alpha/2)\gamma(b_1)^2[b_1\nu(b_1) - \nu(B)B].$$

In our examples, consumption growth innovations lead to positive horizon dependence, just as in the previous section. Variance innovations lead to negative horizon dependence, the result of the different signs of the initial and subsequent moving average coefficients in $a_v(B)$. The overall impact on horizon dependence depends on the relative magnitudes of the two effects and the non-linear interaction between the consumption growth and conditional variance processes; see Appendix F.

We see the result in the first two columns of Table III. We follow Bansal and Yaron (2004) in using an AR(1) volatility process, so that $\nu_{j+1} = \varphi_v \nu_j$ for $j \geq 1$. With their parameter values (column (1)), the stationary distribution

of v_t is normal with mean $v = 0.0099^2 = 9.8 \times 10^{-5}$ and standard deviation $v_0/(1 - \varphi_v^2)^{1/2} = 1.4 \times 10^{-5}$. The zero bound is therefore almost seven standard deviations away from the mean. The impact of the stochastic variance on entropy and horizon dependence is small. Relative to the constant variance case (column (2) of Table II), one-period entropy rises from 0.0214 to 0.0218 and 120-month horizon dependence from 0.0011 to 0.0012. This suggests that horizon dependence is dominated, with these parameter values, by the dynamics of consumption growth. The increase in horizon dependence over the constant variance case indicates that nonlinear interactions between the two processes are quantitatively significant.

We increase the impact if we make the “variance of the variance” larger, as in Bansal, Kiku, and Yaron (2009). We do so in column (2) of Table III, where we increase φ_v from 0.987 to 0.997. With this value, the unconditional standard deviation roughly doubles and zero is a little more than three standard deviations from the mean. We see that one-period entropy and horizon dependence both rise. The latter increases slowly with maturity and exceeds our bound for maturities above 100 months. Bansal, Kiku, and Yaron (2009) increase φ_v further to 0.999. This increases substantially the probability of violating the zero bound and makes our approximation of the variance process less reliable. Further exploration of this channel of influence likely calls for some modification of the volatility process, such as the continuous-time square-root process used by Hansen (2012, Section 8.3) or the discrete-time ARG process discussed in Appendix H.

A second perspective comes from the Campbell–Cochrane (1999) habit model. They suggest the nonlinear surplus process

$$\begin{aligned} \log s_{t+1} - \log s_t &= (\varphi_s - 1)(\log s_t - \log s) + \lambda(\log s_t) v^{1/2} w_{t+1}, \\ 1 + \lambda(\log s_t) &= v^{-1/2} \left[\frac{(1 - \rho)(1 - \varphi_s) - b}{(1 - \rho)^2} \right]^{1/2} (1 - 2[\log s_t - \log s])^{1/2}, \end{aligned}$$

where w_t is iid standard normal. The pricing kernel is then

$$\begin{aligned} \log m_{t,t+1} &= \text{constant} + (\rho - 1)(\varphi_s - 1)(\log s_t - \log s) \\ &\quad + (\rho - 1)[1 + \lambda(\log s_t)] v^{1/2} w_{t+1}. \end{aligned}$$

The essential change from our earlier approximation of the difference habit model is that the conditional variance now depends on the habit as well as the conditional mean. This functional form implies one-period conditional entropy of

$$\begin{aligned} L_t(m_{t,t+1}) &= (\rho - 1)^2 [1 + \lambda(\log s_t)]^2 \\ &= [(1 - \rho)(1 - \varphi_s) - b/2] + b(\log s_t - \log s). \end{aligned}$$

One-period entropy is therefore $I(1) = EL_t(m_{t+1}) = [(1 - \rho)(1 - \varphi_s) - b/2]$.

Campbell and Cochrane (1999) set $b = 0$. In this case, conditional entropy is constant and horizon dependence is zero at all horizons. Entropy is governed by

curvature $1 - \rho$ and the autoregressive parameter φ_s of the surplus. With their suggested values of $1 - \rho = 2$ and $\varphi_s = 0.9885 = 0.87^{1/12}$, entropy is 0.0231, far more than we get with additive power utility when $1 - \rho = 10$ and comparable to Bansal and Yaron's version of recursive utility.

The mechanism is novel. The Campbell–Cochrane model keeps horizon dependence low by giving the state variable $\log s_t$ offsetting effects on the conditional mean and variance of the log pricing kernel. In its original form with $b = 0$, horizon dependence is zero by construction. In later work, Verdelhan (2010) and Wachter (2006) study versions of the model with nonzero values of b . The interaction between the mean and variance is a useful device that we think is worth examining in other models, including those with recursive preferences, where the tradition has been to make them independent.

These two models also illustrate how conditioning information could be used more intensively. The conditional entropy bound (7) shows how the maximum excess return varies with the state. With recursive preferences the relevant component of the state is the conditional variance v_t . With habits, the relevant state is the surplus s_t , but it affects conditional entropy only when b is nonzero. We do not explore conditioning further here, but it strikes us as a promising avenue for future research.

D. Models with Jumps

An influential body of research has developed the idea that departures from normality, including so-called disasters in consumption growth, can play a significant role in asset returns. There is, moreover, strong evidence of non-normality in both macroeconomic data and asset returns. Prominent examples of this line of work include Barro (2006), Barro et al. (2009), Bekaert and Engstrom (2010), Benzoni, Collin-Dufresne, and Goldstein (2011), Branger, Rodrigues, and Schlag (2011), Drechsler and Yaron (2011), Eraker and Shaliastovich (2008), Gabaix (2012), Garcia, Luger, and Renault (2003), Longstaff and Piazzesi (2004), Martin (2013), and Wachter (2013). Although nonnormal innovations can be added to any model, we follow a number of these papers in adding them to models with recursive preferences.

We generate departures from normality by decomposing the innovation in log consumption growth into normal and “jump” components. Consider the process

$$\begin{aligned}\log g_t &= \log g + \gamma(B)v^{1/2}w_{gt} + \psi(B)z_{gt} - \psi(1)h\theta, \\ h_t &= h + \eta(B)w_{ht},\end{aligned}$$

where $\{w_{gt}, z_{gt}, w_{ht}\}$ are standard normal random variables, independent of each other and over time. (Note that we are repurposing h and η here; we have run out of letters.) The last term is constant: it adjusts the mean so that $\log g$ is, in fact, the mean of $\log g_t$. The jump component z_{gt} is a Poisson mixture of normals, a specification that has been widely used in the options literature. Its central ingredient is a Poisson random variable j . At date t ,

j (the number of jumps, so to speak) takes on nonnegative integer values with probabilities $p(j) = e^{-h_{t-1}} h_{t-1}^j / j!$. The “jump intensity” h_{t-1} is the mean of j . Each jump triggers a draw from a normal distribution with mean θ and variance δ^2 . Conditional on the number of jumps, the jump component is normal with mean $j\theta$ and variance $j\delta^2$. That makes z_{gt} a Poisson mixture of normals, which is clearly not normal.

We use a linear process for h_t with standard normal innovations w_{ht} . As with volatility, we think of this as an approximation to a censored process that keeps h_t nonnegative. We show in Appendix G that the approximation is reasonably accurate here, too, in the examples we study.

With this consumption growth process and recursive utility, the pricing kernel is

$$\begin{aligned} \log m_{t,t+1} = & \text{constant} + [(\rho - 1)\gamma(B) + (\alpha - \rho)\gamma(b_1)]v^{1/2}w_{gt+1} \\ & + [(\rho - 1)\psi(B) + (\alpha - \rho)\psi(b_1)]z_{gt+1} \\ & + (\alpha - \rho)[(e^{\alpha\psi(b_1)\theta + (\alpha\psi(b_1)\delta)^2/2} - 1)/\alpha][b_1\eta(b_1) - \eta(B)B]w_{ht+1}, \end{aligned}$$

see Appendix E. The pricing kernel falls into the generalized Vasicek example of Section I.E when persistence of the normal and jump components is the same, $\gamma(B) = \psi(B)$, and the jump intensity is constant, $h_t = h$.

Define $\alpha^* - 1 = (\rho - 1)\psi_0 + (\alpha - \rho)\psi(b_1)$. Then one-period conditional entropy is

$$\begin{aligned} L_t(m_{t,t+1}) = & [(\rho - 1)\gamma_0 + (\alpha - \rho)\gamma(b_1)]^2 v/2 \\ & + \left\{ \left(e^{(\alpha^* - 1)\theta + [(\alpha^* - 1)\delta]^2/2} - 1 \right) - (\alpha^* - 1)\theta \right\} h_t \\ & + \left\{ (\alpha - \rho) \left[(e^{\alpha\psi(b_1)\theta + (\alpha\psi(b_1)\delta)^2/2} - 1)/\alpha \right] b_1\eta(b_1) \right\}^2 /2. \quad (26) \end{aligned}$$

New features include the dynamics of intensity h_t , $\eta(b_1)$, and jumps, $\psi(b_1)$. Horizon dependence includes nonlinear interactions between these features and consumption growth analogous to those we saw with stochastic variance; see Appendix F.

We report properties of several versions in Table IV. The initial parameters of the jump component z_{gt} are taken from Backus, Chernov, and Martin (2011, Section III) and are designed to mimic those estimated by Barro et al. (2009) from international macroeconomic data. The mean and variance of the normal component are then chosen to keep the stationary mean and variance of log consumption growth the same as in our earlier examples.

In our first example (column (1) of Table IV), both components of consumption growth are iid. This eliminates the familiar Bansal–Yaron mechanism in which persistence magnifies the impact of shocks on the pricing kernel. Nevertheless, the jumps increase one-period entropy by a factor of 10 relative to the normal case (column (1) of Table II). The key ingredient in this example is the exponential term $\exp\{(\alpha^* - 1)\theta + [(\alpha^* - 1)\delta]^2/2\}$ in (26). We know from

Table IV
Representative Agent Models with Jumps

The columns summarize the properties of representative agent models with jumps. See Section II.D. The mean and variance of the normal component w_{gt} are adjusted to have the same stationary mean and variance of log consumption growth in each case. Model (1) has iid jumps. Model (2) has stochastic jump intensity. Model (3) has constant jump intensity but a persistent component in consumption growth. Model (4) is the same with a smaller persistent component and less extreme jumps.

Parameter or Property	iid w/Jumps (1)	Stochastic Intensity (2)	Constant Intensity 1 (3)	Constant Intensity 2 (4)
<i>Preference Parameters</i>				
ρ	1/3	1/3	1/3	1/3
α	-9	-9	-9	-9
β	0.9980	0.9980	0.9980	0.9980
<i>Consumption Growth Process</i>				
$v^{1/2}$	0.0025	0.0025	0.0021	0.0079
h	0.0008	0.0008	0.0008	0.0008
θ	-0.3000	-0.3000	-0.3000	-0.1500
δ	0.1500	0.1500	0.1500	0.1500
η_0	0	0.0001	0	0
φ_h		0.9500		
γ_0	1	1	1	1
γ_1			0.0271	0.0281
φ_g			0.9790	0.9690
ψ_0	1	1	1	1
ψ_1			0.0271	
φ_z			0.9790	
<i>Derived Quantities</i>				
b_1	0.9974	0.9973	0.9750	0.9979
$\gamma(b_1)$	1	1	1.5806	1.8481
$\psi(b_1)$	1	1	1.5806	1
$\eta(b_1)$	0	0.0016	0	0
<i>Entropy and Horizon Dependence</i>				
$I(1) = EL_t(m_{t,t+1})$	0.0485	0.0502	1.2299	0.0193
$I(\infty)$	0.0485	0.0532	15.7300	0.0200
$H(120) = I(120) - I(1)$	0	0.0025	9.0900	0.0005
$H(\infty) = I(\infty) - I(1)$	0	0.0030	14.5000	0.0007

earlier work that this function increases sharply with $1 - \alpha^*$, as the nonnormal terms in (8) increase in importance; see, for example, Backus, Chernov, and Martin (2011, figure 2). Evidently setting $1 - \alpha^* = 1 - \alpha = 10$, as we do here, is enough to have a large impact on entropy. The example shows clearly that departures from normality are a significant potential source of entropy. And since consumption growth is iid, horizon dependence is zero at all time horizons.

The next two columns show that, when we introduce dynamics to this model, either through intensity h_t (column (2)) or by making consumption growth persistent (column (3)), both one-period entropy and horizon dependence rise substantially. In column (2), we use an AR(1) intensity process: $\eta_{j+1} = \varphi_h \eta_j$ for

$j \geq 0$. We choose parameters to keep h_t far enough from zero for our approximation to be accurate. This requirement leads to a tiny value of the volatility of jump intensity, η_0 . One-period entropy increases by a small amount, but horizon dependence is now two-and-a-half times our upper bound. Evidently, even this modest amount of volatility in h_t is enough to drive horizon dependence outside the range we established earlier.

In column (3), we reintroduce persistence in consumption growth. Intensity is constant, but the normal and jump components of log consumption growth have the same ARMA(1,1) structure we used in Section II.B. With intensity constant, the model is an example of a Vasicek model with nonnormal innovations. The impact is dramatic. One-period entropy and horizon dependence increase by orders of magnitude. The issue is the dynamics of the jump component, represented by the lag polynomial $\psi(B)$. Here $\psi(b_1) = 1.58$, which raises $1 - \alpha^*$ from 10 in column (1) to 15.4 and drives entropy two orders of magnitude beyond our lower bound. It has a similar impact on horizon dependence, which is now almost three orders of magnitude beyond our bound.

These two models illustrate the pros and cons of mixing jumps with dynamics. We know from earlier work that jumps give us enormous power to generate large expected excess returns. Here we see that, when they come with dynamics, they can also generate unreasonably large horizon dependence, which is inconsistent with the evidence on bond yields.

The last example (column (4)) illustrates what we might do to reconcile the two, that is, to use jumps to increase one-period entropy without also increasing horizon dependence to unrealistic levels. We cut the mean jump size θ in half, eliminate dynamics in the jump ($\psi_1 = 0$), and reduce the persistence of the normal component (by reducing φ_g and increasing γ_1). In this case, we exceed our lower bound on one-period entropy by a factor of two and are well within our bounds for horizon dependence.

We do not claim any particular realism for this example, but it illustrates what we think could be a useful approach to modeling jumps. Since jumps have such a powerful effect on entropy, we can rely less on the persistent component of consumption growth that has played such a central role in work with recursive preferences since Bansal and Yaron (2004).

III. Final Thoughts

We have shown that an asset pricing model, represented here by its pricing kernel, must have two properties to be consistent with the evidence on asset returns. The first is entropy, a measure of the pricing kernel's dispersion. Entropy over a given time interval must be at least as large as the largest mean log excess return over the same time interval. The second property is horizon dependence, a measure of the pricing kernel's dynamics derived from entropy over different time horizons. Horizon dependence must be small enough to account for the relatively small premiums we observe on long bonds.

The challenge is to accomplish both at once, that is, to generate enough entropy without too much horizon dependence. Representative agent models with

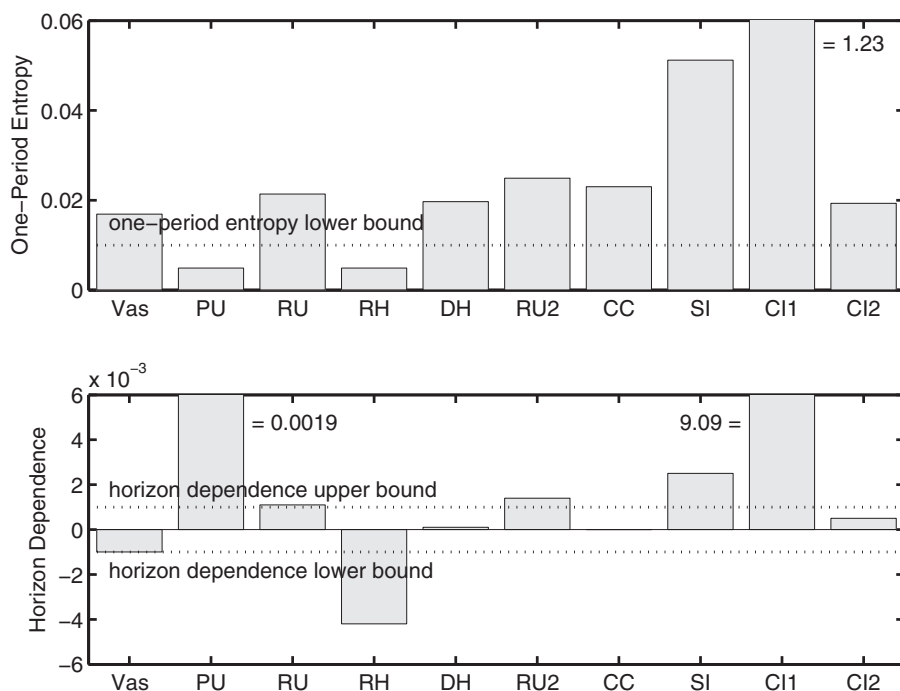


Figure 5. Model summary: one-period entropy and horizon dependence. The figure summarizes one-period entropy $I(1)$ and horizon dependence $H(120)$ for a number of models. They include: Vas (Vasicek); PU (power utility, column (1) of Table II); RU (recursive utility, column (2) of Table II); RH (ratio habit, column (3) of Table II); DH (difference habit, column (4) of Table II); RU2 (recursive utility 2 with stochastic variance, column (2) of Table III); CC (Campbell–Cochrane (1999), column (3) of Table III); SI (stochastic intensity, column (2) of Table IV); CI1 (constant intensity 1, column (3) of Table IV); and CI2 (constant intensity 2, column (4) of Table IV). Some of the bars have been truncated; their values are noted in the figure. The idea is that a good model should have more entropy than the lower bound in the upper panel, but no more horizon dependence than the bounds in the lower panel. The difference habit model here looks relatively good, but we noted earlier that horizon dependence violates the bounds at most horizons between 1 and 120 months.

recursive preferences and habits use dynamics to increase entropy, but as a result they often increase horizon dependence as well. Figure 5 summarizes how a number of representative agent models do along these two dimensions. In the top panel we report entropy, which should be above the estimated lower bound marked by the dotted line. In the bottom panel we report horizon dependence, which should lie between the bounds also noted by dotted lines.

We identify two approaches that we think hold some promise. One is to specify the interaction between the conditional mean and variance designed, as in the Campbell–Cochrane model, to reduce their impact on horizon dependence. See the bars labeled CC. The other is to introduce jumps with little in the way of additional dynamics. An example of this kind is labeled CI2 in the figure.

All of these numbers depend on parameter values and are therefore subject to change, but they suggest directions for the future evolution of these models.

Initial submission: August 9, 2011; Final version received: June 21, 2013
Editor: Campbell Harvey

Appendix A: Bond Prices, Yields, and Forward Rates

We refer to prices, yields, and forward rates on discount bonds throughout the paper. Given a term structure of one of these objects, we can construct the other two. Let q_t^n be the price at date t of an n -period zero-coupon bond, a claim to one at date $t + n$. Yields y and forward rates f are defined from prices by

$$-\log q_t^n = ny_t^n = \sum_{j=1}^n f_t^{j-1}.$$

Equivalently, yields are averages of forward rates: $y_t^n = n^{-1} \sum_{j=1}^n f_t^{j-1}$. Forward rates can be constructed directly from bond prices by $f_t^n = \log(q_t^n/q_t^{n+1})$.

A related concept is the holding period return. The one-period (gross) return on an n -period bond is $r_{t,t+1}^n = q_{t+1}^{n-1}/q_t^n$. The *short rate* is $\log r_{t+1}^1 = y_t^1 = f_t^0$.

Bond pricing follows directly from bond returns and the pricing relation (2). The direct approach follows from the n -period return $r_{t,t+n} = 1/q_t^n$. It implies

$$q_t^n = E_t m_{t,t+n}.$$

The recursive approach follows from the one-period return, which implies

$$q_t^{n+1} = E_t(m_{t,t+1} q_{t+1}^n).$$

In words: an $n + 1$ -period bond is a claim to an n -period bond in one period.

There is also a connection between bond prices and returns. An n -period bond price is connected to its n -period return by

$$\log q_t^n = - \sum_{j=1}^n \log r_{t+j-1,t+j}^j.$$

This allows us to express yields as functions of returns and relate horizon dependence to mean returns.

These relations are exact. There are analogous relations for means in stationary environments. Mean yields are averages of mean forward rates:

$$Ey_t^n = n^{-1} \sum_{j=1}^n Ef_t^{j-1}.$$

Mean log returns are also connected to mean forward rates:

$$E \log r_{t,t+1}^{n+1} = E \log q_{t+1}^n - E \log q_t^{n+1} = Ef_t^n,$$

where the t subscript in the last term simply marks the forward rate as a random variable rather than its mean.

Appendix B: Entropy and Hansen–Jagannathan Bounds

The entropy and Hansen–Jagannathan (HJ hereafter) bounds play similar roles, but the bounds and the maximum returns they imply are different. We describe them both, show how they differ, and illustrate their differences further with an extension to multiple periods and an application to lognormal returns.

Bounds and Returns: The HJ bound defines a high-return asset as one whose return $r_{t,t+1}$ maximizes the Sharpe ratio: given a pricing kernel $m_{t,t+1}$, its excess return $x_{t,t+1} = r_{t,t+1} - r_{t,t+1}^1$ maximizes $SR_t = E_t(x_{t,t+1})/\text{Var}_t(x_{t,t+1})^{1/2}$ subject to the pricing relation (2) for $n = 1$. The maximization leads to the bound

$$SR_t = E_t(x_{t,t+1})/\text{Var}_t(x_{t,t+1})^{1/2} \leq \text{Var}_t(m_{t,t+1})^{1/2}/E_t m_{t,t+1},$$

and the return that hits the bound

$$x_{t,t+1} = E_t(x_{t,t+1}) + [E_t(m_{t,t+1}) - m_{t,t+1}] \cdot \frac{\text{Var}_t(x_{t,t+1})^{1/2}}{\text{Var}_t(m_{t,t+1})^{1/2}},$$

$$r_{t,t+1} = x_{t,t+1} + r_{t,t+1}^1.$$

There is one degree of indeterminacy in $x_{t,t+1}$: if $x_{t,t+1}$ is a solution, then so is $\lambda x_{t,t+1}$ for $\lambda > 0$ (the Sharpe ratio is invariant to leverage). If we use the normalization $\text{Var}_t(x_{t,t+1}) = 1$, the return becomes

$$r_{t,t+1} = \frac{1 + \text{Var}_t(m_{t,t+1})^{1/2}}{E_t(m_{t,t+1})} + \frac{E_t(m_{t,t+1}) - m_{t,t+1}}{\text{Var}_t(m_{t,t+1})^{1/2}},$$

which connects it directly to the pricing kernel.

We can take a similar approach to the entropy bound. The bound defines a high-return asset as one whose return $r_{t,t+1}$ maximizes $E_t(\log r_{t,t+1} - \log r_{t,t+1}^1)$ subject (again) to the pricing relation (2) for $n = 1$. The maximization leads to the return

$$r_{t,t+1} = -1/m_{t,t+1} \Leftrightarrow \log r_{t,t+1} = -\log m_{t,t+1}.$$

Its mean log excess return $E_t(\log r_{t,t+1} - \log r_{t,t+1}^1)$ hits the entropy bound (7).

It is clear, then, that the returns that attain the HJ and entropy bounds are different: the former is linear in the pricing kernel, the latter loglinear. They are solutions to two different problems.

Entropy and Maximum Sharpe Ratios: We find it helpful in comparing the two bounds to express each in terms of the (conditional) cumulant generating function of the log pricing kernel. The approach is summarized in Backus, Chernov, and Martin (2011, Appendix A.2) and Martin (2013, Section III.A). Suppose $\log m_{t,t+1}$ has conditional cumulant generating function $k_t(s)$. The

maximum Sharpe ratio follows from the mean and variance of $m_{t,t+1}$:

$$\begin{aligned} E_t m_{t,t+1} &= e^{k_t(1)}, \\ \text{Var}_t(m_{t,t+1}) &= E_t(m_{t,t+1}^2) - (E_t m_{t,t+1})^2 = e^{k_t(2)} - e^{2k_t(1)}. \end{aligned}$$

The maximum squared Sharpe ratio is therefore

$$\text{Var}_t(m_{t,t+1})/E_t(m_{t,t+1})^2 = e^{k_t(2)-2k_t(1)} - 1.$$

The exponent has the expansion

$$k_t(2) - 2k_t(1) = \sum_{j=1}^{\infty} \kappa_{jt}(2^j - 2)/j!,$$

a complicated combination of cumulants. In the lognormal case, cumulants above order two are zero, $k_t(2) - 2k_t(1) = \kappa_{2t}$, and the squared Sharpe ratio is $e^{\kappa_{2t}} - 1$. For small κ_2 it is approximately κ_{2t} and entropy is exactly $\kappa_{2t}/2$, so the two reflect the same information. Otherwise they do not.

Lognormal Settings: Suppose asset j 's return is conditionally lognormal: $\log r_{t,t+1}^j$ is normal with mean $\log r_{t,t+1}^1 + \kappa_{1t}^j$ and variance κ_{2t}^j . Our entropy bound focuses on the mean log excess return

$$E_t(\log r_{t,t+1}^j - \log r_{t,t+1}^1) = \kappa_{1t}^j.$$

That is it.

The Sharpe ratio focuses on the simple excess return, $x_{t,t+1} = r_{t,t+1}^j - r_{t,t+1}^1$, which we will see reflects both moments of the log return. The mean and variance of the excess return are

$$\begin{aligned} E_t(x_{t,t+1}) &= r_{t,t+1}^1 \left(e^{\kappa_{1t}^j + \kappa_{2t}^j/2} - 1 \right), \\ \text{Var}_t(x_{t,t+1}) &= \left(r_{t,t+1}^1 e^{\kappa_{1t}^j + \kappa_{2t}^j/2} \right)^2 \left(e^{\kappa_{2t}^j} - 1 \right). \end{aligned}$$

The conditional Sharpe ratio is therefore

$$SR_t = \frac{E_t(x_{t,t+1})}{\text{Var}_t(x_{t,t+1})^{1/2}} = \frac{e^{\kappa_{1t}^j + \kappa_{2t}^j/2} - 1}{e^{\kappa_{1t}^j + \kappa_{2t}^j/2} \left(e^{\kappa_{2t}^j} - 1 \right)^{1/2}}.$$

Evidently there are two ways to generate a large Sharpe ratio. The first is to have a large mean log return, that is, a large value of κ_{1t}^j . The second is to have a small variance: as κ_{2t}^j approaches zero, so does the denominator.

Comparisons of Sharpe ratios thus reflect both the mean and the variance of the log return, and possibly higher-order cumulants as well. Binsbergen, Brandt, and Kojen (2012) and Duffee (2010) are interesting examples. They show that Sharpe ratios for dividends and bonds, respectively, decline with maturity. In the former, this reflects a decline in the mean; in the latter, this reflects an increase in the variance.

Varying the Time Horizon: We can get a sense of how entropy and the Sharpe ratio vary with the time horizon by looking at the iid case. We drop the subscript t from k (there's no conditioning) and add a superscript n denoting the time horizon. In the iid case, the n -period cumulant generating function is n times the one-period function:

$$k^n(s) = nk^1(s).$$

The same is true of cumulants. As a result, entropy is proportional to n :

$$L(m_{t,t+n}) = n[k^1(1) - \kappa_1].$$

This is the zero-horizon dependence result we saw earlier for the iid case. The time horizon n is an integer in our environment, but if the distribution is infinitely divisible we can extend it to any positive real number.

The maximum Sharpe ratio also varies with the time horizon. We can adapt our earlier result:

$$\text{Var}(m_{t,t+n})/E(m_{t,t+n})^2 = e^{k^n(2)-2k^n(1)} - 1 = e^{n[k^1(2)-2k^1(1)]} - 1.$$

For small time intervals n , this is approximately

$$e^{n[k^1(2)-2k^1(1)]} - 1 \approx n[k^1(2) - 2k^1(1)],$$

which is also proportional to n . In general, however, the squared Sharpe ratio increases exponentially with n .

Another perspective on dynamics comes from Chretien (2012), who notes that one- and two-period bond prices are related to the first autocovariance of the pricing kernel by

$$E(q_t^2) - E(q_t^1)^2 = \text{Cov}(m_{t,t+1}, m_{t+1,t+2}).$$

The left side is negative in U.S. data, the price analog of an increasing mean yield curve. The first autocorrelation is therefore

$$\text{Corr}(m_{t,t+1}, m_{t+1,t+2}) = \frac{\text{Cov}(m_{t,t+1}, m_{t+1,t+2})}{\text{Var}(m_{t,t+1})} = \frac{E(q_t^2) - E(q_t^1)^2}{\text{Var}(m_{t,t+1})}.$$

The unconditional HJ bound gives us an upper bound on the variance,

$$\text{Var}(m_{t,t+1}) \geq SR^2 E(q_t^1)^2,$$

which gives us bounds on the autocorrelation,

$$\text{Corr}(m_{t,t+1}, m_{t+1,t+2}) \leq \frac{E(q_t^2) - E(q_t^1)^2}{SR^2 E(q_t^1)^2} \leq 0.$$

This is an interesting result, but it is more complicated than horizon dependence and does not extend in any obvious way to horizons greater than two periods.

Appendix C: Lag Polynomials

We use notation and results from Hansen and Sargent (1980, Section 2) and Sargent (1987, Chapter XI), who supply references to the related mathematical literature. Our primary tool is the one-sided infinite moving average,

$$x_t = \sum_{j=0}^{\infty} a_j w_{t-j} = a(B)w_t,$$

where $\{w_t\}$ is an iid sequence with zero mean and unit variance. This defines implicitly the lag polynomial

$$a(B) = \sum_{j=0}^{\infty} a_j B^j.$$

The lag or backshift operator B shifts what follows back one period in time: $Bw_t = w_{t-1}$, $B^2 w_t = w_{t-2}$, and so on. The result is a stationary process if $\sum_j a_j^2 < \infty$; we say the sequence of a_j s is square summable.

In this form, prediction is simple. If the information set at date t includes current and past values of w_t , forecasts of future values of x_t are

$$E_t x_{t+k} = E_t \sum_{j=0}^{\infty} a_j w_{t+k-j} = \sum_{j=k}^{\infty} a_j w_{t+k-j} = [a(B)/B^k]_+ w_t$$

for $k \geq 0$. We simply chop off the terms that involve future values of w . The subscript “+” applied to the final expression is compact notation for the same thing—it means ignore negative powers of B .

We use the ARMA(1,1) repeatedly:

$$\varphi(B)x_t = \theta(B)v^{1/2}w_t,$$

with $\varphi(B) = 1 - \varphi B$ and $\theta(B) = 1 - \theta B$. Special cases include the AR(1) (set $\theta = 0$) and the MA(1) (set $\varphi = 0$). The infinite moving average representation is $x_t = [\varphi(B)/\theta(B)]v^{1/2}w_t = a(B)v^{1/2}w_t$, with $a_0 = 1$, $a_1 = \varphi - \theta$, and $a_{j+1} = \varphi^j(\varphi - \theta)$ for $j \geq 1$. We typically choose φ and a_1 , leaving θ implicit. Then $a_{j+1} = \varphi^j a_1 = \varphi a_j$ for $j \geq 1$. An AR(1) has $a_{j+1} = \varphi a_j$ for $j \geq 0$.

Appendix D: Bond Prices, Yields, and Returns in the Vasicek Model

Consider the pricing kernel (12) for the Vasicek model of Section II.E. We show that the proposed forward rates (13) satisfy the pricing relation $q_t^{n+1} = E_t(m_{t,t+1}q_{t+1}^n)$.

The proposed forward rates imply bond prices of

$$\log q_t^n = \sum_{j=1}^n f_t^{j-1} = n \log m + \sum_{j=1}^n k(A_{j-1}) + \sum_{j=0}^{\infty} (A_{n+j} - A_j)w_{t-j}.$$

Therefore,

$$\log(m_{t,t+1}q_{t+1}^n) = (n+1)\log m + \sum_{j=1}^n k(A_{j-1}) + A_n w_{t+1} + \sum_{j=0}^{\infty} (A_{n+1+j} - A_j) w_{t-j}.$$

The next step is to evaluate $\log E_t(m_{t,t+1}q_{t+1}^n)$. The only stochastic term is $\log E_t(e^{A_n w_{t+1}})$, which is the cumulant generating function $k(s)$ evaluated at $s = A_n$. Therefore, we have

$$\log E_t(m_{t,t+1}q_{t+1}^n) = (n+1)\log m + \sum_{j=1}^{n+1} k(A_{j-1}) + \sum_{j=0}^{\infty} (A_{n+1+j} - A_j) w_{t-j},$$

which is $\log q_t^{n+1}$. Thus, the proposed forward rates and associated bond prices satisfy the pricing relation as stated.

Appendix E: The Recursive Utility Pricing Kernel

We derive the pricing kernel for a representative agent model with recursive utility, loglinear consumption growth dynamics, stochastic volatility, and jumps with time-varying intensity. The recursive utility models in Sections II.B, II.C, and II.D are all special cases.

The consumption growth process is

$$\begin{aligned} \log g_t &= \log g' + \gamma(B)v_{t-1}^{1/2}w_{gt} + \psi(B)z_{gt}, \\ v_t &= v + v(B)w_{vt}, \\ h_t &= h + \eta(B)w_{ht}, \end{aligned}$$

where $\{w_{gt}, w_{vt}, w_{ht}\}$ are independent standard normals and $\log g' = \log g - \psi(1)h\theta$. The jump component z_{gt} is a Poisson mixture of normals: conditional on the number of jumps j , z_{gt} is normal with mean $j\theta$ and variance $j\delta^2$. The probability of $j \geq 0$ jumps at date $t+1$ is $e^{-h_t}h_t^j/j!$.

Given a value of b_1 , we use equation 24 to characterize the value function and substitute the result into the pricing kernel (17). Our use of value functions mirrors Hansen, Heaton, and Li (2008) and Hansen and Scheinkman (2009). Our use of lag polynomials mirrors Hansen and Sargent (1980) and Sargent (1987).

The certainty equivalents needed for the recursion (24) are closely related to the cumulant generating functions of the relevant random variables. Consider an arbitrary random variable y_{t+1} whose conditional cumulant generating function is $k_t(s; y) = \log E_t(e^{sy_{t+1}})$. Then the log of the certainty equivalent (15) of $e^{a_t + b_t y_{t+1}}$ is

$$\log \mu_t(e^{a_t + b_t y_{t+1}}) = a_t + k_t(\alpha b_t)/\alpha.$$

We use two kinds of cumulant generating functions below: for the standard normals we have $k_t(s; w_{t+1}) = s^2/2$, and for the jump component we have $k_t(s; z_{t+1}) = (e^{s\theta + (s\delta)^2/2} - 1)h_t$. Both functions occur repeatedly in what follows.

We find the value function by guess-and-verify:

- Guess. We guess a value function of the form

$$\log u_t = \log u + p_g(B)v_{t-1}^{1/2}w_{gt} + p_z(B)z_{gt} + p_v(B)w_{vt} + p_h(B)w_{ht}$$

with parameters (u, p_g, p_z, p_v, p_h) to be determined.

- Compute certainty equivalent. Given our guess, $\log(g_{t+1}u_{t+1})$ is

$$\begin{aligned} \log(g_{t+1}u_{t+1}) &= \log g' + \log u + [\gamma(B) + p_g(B)]v_t^{1/2}w_{gt+1} + [\psi(B) + p_z(B)]z_{gt+1} \\ &\quad + p_v(B)w_{vt+1} + p_h(B)w_{ht+1} \\ &= \log(g'u) + [\gamma(B) + p_g(B) - (\gamma_0 + p_{g0})]v_t^{1/2}w_{gt+1} \\ &\quad + [\psi(B) + p_z(B) - (\psi_0 + p_{z0})]z_{gt+1} + [p_v(B) - p_{v0}]w_{vt+1} \\ &\quad + [p_h(B) - p_{h0}]w_{ht+1} + (\gamma_0 + p_{g0})v_t^{1/2}w_{gt+1} \\ &\quad + p_{v0}w_{vt+1} + p_{h0}w_{ht+1} + (\psi_0 + p_{z0})z_{gt+1}. \end{aligned}$$

We use a clever trick here from Sargent (1987, Section XI.19): we rewrite, for example, $p_v(B)w_{vt+1} = (p_v(B) - p_{v0})w_{vt+1} + p_{v0}w_{vt+1}$. As of date t , the first term is constant (despite appearances, it does not depend on w_{vt+1}) but the second is not. The other terms are treated the same way. As a result, the last line consists of innovations, and the others of (conditional) constants. The certainty equivalent treats them differently:

$$\begin{aligned} \log \mu_t(g_{t+1}u_{t+1}) &= \log(g'u) + [\gamma(B) + p_g(B) - (\gamma_0 + p_{g0})]v_t^{1/2}w_{gt+1} \\ &\quad + [\psi(B) + p_z(B) - (\psi_0 + p_{z0})]z_{gt+1} \\ &\quad + [p_v(B) - p_{v0}]w_{vt+1} + [p_h(B) - p_{h0}]w_{ht+1} \\ &\quad + (\alpha/2)(\gamma_0 + p_{g0})^2v_t + (\alpha/2)(p_{v0}^2 + p_{h0}^2) \\ &\quad + [(e^{\alpha(\psi_0 + p_{z0})\theta + (\alpha(\psi_0 + p_{z0})\delta)^2/2} - 1)/\alpha]h_t \\ &= \log(g'u) + [\gamma(B) + p_g(B) - (\gamma_0 + p_{g0})]v_t^{1/2}w_{gt+1} \\ &\quad + [\psi(B) + p_z(B) - (\psi_0 + p_{z0})]z_{gt+1} \\ &\quad + [p_v(B) - p_{v0}]w_{vt+1} + [p_h(B) - p_{h0}]w_{ht+1} \\ &\quad + (\alpha/2)(\gamma_0 + p_{g0})^2[v + v(B)w_{vt}] + (\alpha/2)(p_{v0}^2 + p_{h0}^2) \\ &\quad + [(e^{\alpha(\psi_0 + p_{z0})\theta + (\alpha(\psi_0 + p_{z0})\delta)^2/2} - 1)/\alpha][h + \eta(B)w_{ht}]. \end{aligned}$$

- Verify. We substitute the certainty equivalent into (24) and solve for the parameters. Matching like terms, we have

$$\begin{aligned}
\text{constant : } \log u &= b_0 + b_1 [\log(g'u) + (\alpha/2)(p_{v0}^2 + p_{h0}^2) + (\alpha/2)(\gamma_0 + p_{g0})^2 v] \\
&\quad + b_1 [(e^{\alpha(\psi_0 + p_{z0})\theta + (\alpha(\psi_0 + p_{z0})\delta)^2/2} - 1)/\alpha] h; \\
v_{t-1}^{1/2} w_{gt+1} : p_g(B)B &= b_1 [\gamma(B) + p_g(B) - (\gamma_0 + p_{g0})] \\
z_{gt+1} : p_z(B)B &= b_1 [\psi(B) + p_z(B) - (\psi_0 + p_{z0})] \\
w_{vt+1} : p_v(B)B &= b_1 [p_v(B) - p_{v0} + (\alpha/2)(\gamma_0 + p_{g0})^2 v(B)B] \\
w_{ht+1} : p_h(B)B &= b_1 \\
&\quad \times \left[p_h(B) - p_{h0} + [(e^{\alpha(\psi_0 + p_{z0})\theta + (\alpha(\psi_0 + p_{z0})\delta)^2/2} - 1)/\alpha] \eta(B)B \right].
\end{aligned}$$

The second equation leads to forward-looking geometric sums like those in Hansen and Sargent (1980, Section 2) and Sargent (1987, Section XI.19). Following their lead, we set $B = b_1$ to get $\gamma_0 + p_{g0} = \gamma(b_1)$. The other coefficients of $p_g(B)$ are of no concern to us: they do not show up in the pricing kernel. The third equation is similar and implies $\psi_0 + p_{z0} = \psi(b_1)$. In the fourth equation, setting $B = b_1$ gives us $p_{v0} = (\alpha/2)\gamma(b_1)^2 b_1 v(b_1)$. Proceeding the same way with the fifth equation gives us $p_{h0} = [(e^{\alpha\psi(b_1)\theta + (\alpha\psi(b_1)\delta)^2/2} - 1)/\alpha] b_1 \eta(b_1)$. For future reference, define $D = (\alpha/2)\gamma(b_1)^2$ and $J = [(e^{\alpha\psi(b_1)\theta + (\alpha\psi(b_1)\delta)^2/2} - 1)/\alpha]$.

Now that we know the value function, we construct the pricing kernel from (17). One component is

$$\begin{aligned}
\log(g_{t+1}u_{t+1}) - \log \mu_t(g_{t+1}u_{t+1}) &= -Dv - Jh - (\alpha/2) \{ [Db_1 v(b_1)]^2 + [Jb_1 \eta(b_1)]^2 \} \\
&\quad + \gamma(b_1)v_t^{1/2}w_{gt+1} + \psi(b_1)z_{gt+1} \\
&\quad + D[b_1 v(b_1) - v(B)B]w_{vt+1} \\
&\quad + J[b_1 \eta(b_1) - \eta(B)B]w_{ht+1},
\end{aligned}$$

a combination of innovations to future utility and adjustments for risk. The pricing kernel is

$$\begin{aligned}
\log m_{t,t+1} &= \log \beta + (\rho - 1) \log g \\
&\quad - (\alpha - \rho)(Dv - Jh) - (\alpha - \rho)(\alpha/2) \{ [Db_1 v(b_1)]^2 + [Jb_1 \eta(b_1)]^2 \} \\
&\quad + [(\rho - 1)\gamma(B) + (\alpha - \rho)\gamma(b_1)]v_t^{1/2}w_{gt+1} \\
&\quad + [(\rho - 1)\psi(B) + (\alpha - \rho)\psi(b_1)]z_{gt+1} \\
&\quad + (\alpha - \rho)D[b_1 v(b_1) - v(B)B]w_{vt+1} \\
&\quad + (\alpha - \rho)J[b_1 \eta(b_1) - \eta(B)B]w_{ht+1}.
\end{aligned}$$

The special cases used in the paper come from setting some terms equal to zero.

Appendix F: Horizon Dependence with Recursive Models

We derive horizon dependence for the model described in Appendix I.E. The pricing kernel has the form

$$\begin{aligned}\log m_{t,t+1} &= \log m + a_g(B)(v_t/v)^{1/2}w_{gt+1} + a_z(B)z_{gt+1} + a_v(B)w_{vt+1} + a_h(B)w_{ht+1} \\ v_t &= v + \nu(B)w_{vt} \\ h_t &= h + \eta(B)w_{ht}\end{aligned}$$

with $\{w_{gt}, w_{vt}, z_{gt}, w_{ht}\}$ defined earlier. This differs from the Vasicek model in the roles of v_t in scaling w_{gt} and of the intensity h_t in the jump component z_{gt} . For future reference, we define the partial sums $A_{xn} = \sum_{j=0}^n a_{xj}$ for $x = g, v, h, z$.

We derive entropy and horizon dependence using (3) and its connection to bond prices: $q_t^n = E_t m_{t,t+n}$. Recursive pricing of bonds gives us

$$\log q_t^{n+1} = \log E_t(m_{t,t+1}q_{t+1}^n).$$

Suppose bond prices have the form

$$\log q_{t+1}^n = \gamma_0^n + \gamma_g^n(B)(v_t/v)^{1/2}w_{gt+1} + \gamma_v^n(B)w_{vt+1} + \gamma_h^n(B)w_{ht+1} + \gamma_z^n(B)z_{t+1}. \quad (\text{F1})$$

Then we have

$$\begin{aligned}\log(m_{t,t+1}q_{t+1}^n) &= \log m + \gamma_0^n + [a_g(B) + \gamma_g^n(B)](v_t/v)^{1/2}w_{gt+1} \\ &\quad + [a_v(B) + \gamma_v^n(B)]w_{vt+1} \\ &\quad + [a_z(B) + \gamma_z^n(B)]z_{gt+1} + [a_h(B) + \gamma_h^n(B)]w_{ht+1}.\end{aligned}$$

Evaluating the expectation and lining up terms gives us

$$\begin{aligned}\gamma_0^{n+1} &= \log m + \gamma_0^n + \left[(a_{g0} + \gamma_{g0}^n)^2 + (a_{v0} + \gamma_{v0}^n)^2 + (a_{h0} + \gamma_{h0}^n)^2 \right] / 2 \\ &\quad + h(e^{(a_{z0} + \gamma_{z0}^n)\theta + ((a_{z0} + \gamma_{z0}^n)\delta)^2/2} - 1); \\ \gamma_{gj}^{n+1} &= \gamma_{gj+1}^n + a_{gj+1}; \\ \gamma_{vj}^{n+1} &= \gamma_{vj+1}^n + a_{vj+1} + (a_{g0} + \gamma_{g0}^n)^2 v_j / (2v); \\ \gamma_{hj}^{n+1} &= \gamma_{hj+1}^n + a_{hj+1} + (e^{(a_{z0} + \gamma_{z0}^n)\theta + ((a_{z0} + \gamma_{z0}^n)\delta)^2/2} - 1)\eta_j; \\ \gamma_{zj}^{n+1} &= \gamma_{zj+1}^n + a_{zj+1}.\end{aligned}$$

The second and fourth equations mirror the Vasicek model:

$$\begin{aligned}\gamma_{gj}^n &= \sum_{i=1}^n a_{gj+i} = A_{gn+j} - A_{gj}; \\ \gamma_{zj}^n &= \sum_{i=1}^n a_{zj+i} = A_{zn+j} - A_{zj}.\end{aligned}$$

The third equation implies

$$\gamma_{vj}^n = A_{vn+j} - A_{vj} + (2v)^{-1} \sum_{i=0}^{n-1} v_{j+n-1-i} A_{gi}^2.$$

The fourth equation implies

$$\gamma_{hj}^n = A_{hn+j} - A_{hj} + \sum_{i=0}^{n-1} \eta_{j+n-1-i} (e^{A_{zi}\theta + (A_{zi}\delta)^2/2} - 1).$$

The first equation implies

$$\begin{aligned} \gamma_0^n = & n \log m + \frac{1}{2} \sum_{j=1}^n A_{gj-1}^2 + \frac{1}{2} \sum_{j=1}^n A_{zj-1}^2 + h \sum_{j=1}^n (e^{A_{zj-1}\theta + (A_{zj-1}\delta)^2/2} - 1) \\ & + \frac{1}{2} \sum_{j=1}^n \left[A_{vj-1} + (2v)^{-1} \sum_{i=0}^{j-2} v_{j-2-i} A_{gi}^2 \right]^2 \\ & + \frac{1}{2} \sum_{j=1}^n \left[A_{hj-1} + \sum_{i=0}^{j-2} \eta_{j-2-i} (e^{A_{zi}\theta + (A_{zi}\delta)^2/2} - 1) \right]^2. \end{aligned}$$

If subscripts are beyond their bounds, the expression is zero.

Horizon dependence is determined by unconditional expectations of yields. The z_g component in the log-price (F1) is nonzero, so we have to take this into account:

$$E(\gamma_z^n(B)z_{t+1}) = \theta h \gamma_z^n(1) = \theta h \sum_{j=0}^{\infty} (A_{zn+j} - A_{zj}).$$

Horizon dependence is therefore

$$\begin{aligned} H(n) = & (2n)^{-1} \sum_{j=1}^n (A_{gj-1}^2 - A_{g0}^2) + (2n)^{-1} \sum_{j=1}^n (A_{zj-1}^2 - A_{z0}^2) \\ & + hn^{-1} \sum_{j=1}^n \left(e^{A_{zj-1}\theta + (A_{zj-1}\delta)^2/2} - e^{A_{z0}\theta + (A_{z0}\delta)^2/2} \right) \\ & + (2n)^{-1} \sum_{j=1}^n \left[\left(A_{vj-1} + (2v)^{-1} \sum_{i=0}^{j-2} v_{j-2-i} A_{gi}^2 \right)^2 - A_{v0}^2 \right] \\ & + (2n)^{-1} \sum_{j=1}^n \left[\left(A_{hj-1} + \sum_{i=0}^{j-2} \eta_{j-2-i} (e^{A_{zi}\theta + (A_{zi}\delta)^2/2} - 1) \right)^2 - A_{h0}^2 \right] \\ & + n^{-1} \theta h \gamma_z^n(1) - \theta h \gamma_z^1(1). \end{aligned}$$

Appendix G: Assessing the Loglinear Approximation

We employ the discrete-grid algorithm of Tauchen (1986) to compute approximate numerical solutions of recursive utility models and compare them to the loglinear approximations used in the paper. This approach generates an arbitrarily good approximation of the value function and related objects if we use a sufficiently fine grid. We compute such approximations for two models, one with stochastic variance and another with stochastic jump intensity. In each case, there are two sources of nonlinearity, namely, the time aggregator (16) and the censored distributions of the variance and intensity.

Stochastic Variance: We use an equivalent state–space representation of consumption growth dynamics:

$$\log g_t = \log g + x_{t-1} + v'_{t-1}/2 w_{gt},$$

$$x_t = \phi_g x_{t-1} + \gamma_1 v'_{t-1}/2 w_{gt},$$

$$v_t = (1 - \phi_v)v + \phi_v v_{t-1} + v_0 w_{vt},$$

$$v'_t = \max\{0, v_t\}.$$

The goal is to compute a numerical approximation of the scaled value function u_t as a function of the state (x_t, v_t) . In our calculations, we use the parameter values reported in column (2) of Table III.

We approximate the law of motion of the state with finite-state Markov chains. We construct a discrete version of v_t that assumes values given by a grid of 100 equally spaced points. We label the distance between points ϵ_v . The points are centered at the mean v and extend five standard deviations in each direction. In the notation of the model, v_t covers the interval $[v - 5v_0/(1 - \phi_v^2)^{1/2}, v + 5v_0/(1 - \phi_v^2)^{1/2}]$. Since the mean is more than five standard deviations from zero in this case, there is no censoring in the discrete approximation: $v'_t = \max\{0, v_t\} = v_t$. The only nonlinearity in this model is in the time aggregator.

Probabilities are assigned as Tauchen suggests. Since the conditional distribution of v_t is normal, we define probabilities using $\Phi(\cdot; a, b)$, the distribution function for a normal random variable with mean a and standard deviation b . The transition probabilities are

$$\begin{aligned} \Pi_{ij}^v &\equiv \text{Prob}(v_t = v_i \mid v_{t-1} = v_j) \\ &= \Phi\left[v_i + \frac{\epsilon_v}{2}; (1 - \phi_v)v + \phi_v v_j, v_0\right] - \Phi\left[v_i - \frac{\epsilon_v}{2}; (1 - \phi_v)v + \phi_v v_j, v_0\right]. \end{aligned}$$

When $v = v_1$ (the first grid point), we set the second term equal to zero, and when $v = v_{100}$ (the last grid point), we set the first term equal to one.

The state variable x_t has a one-step-ahead distribution that is conditional on both x_{t-1} and v_{t-1} . We choose a fixed grid for x_t that takes 200 equally spaced values on an interval five standard deviations either side of its mean. Since we want this grid to remain fixed for all values of the conditional variance, we use the largest value on the grid for v_t to set this interval. Transition probabilities

are then

$$\begin{aligned}\Pi_{ijk}^x &\equiv \text{Prob}(x_t = x_i \mid x_{t-1} = x_j, v_{t-1} = v_k) \\ &= \Phi\left[x_i + \frac{\epsilon_x}{2}; \varphi_x x_j, \gamma_1 v_k^{1/2}\right] - \Phi\left[x_i - \frac{\epsilon_x}{2}; \varphi_x x_j, \gamma_1 v_k^{1/2}\right].\end{aligned}$$

Again, we set the second term equal to zero for the first point and the first term equal to one for the last one.

With these inputs, we can compute a discrete approximation to the value function: scaled utility u_t defined over the grid of states (x_i, v_j) . The Markov chain for x_t implies an approximation for the shock w_{gt} of

$$w_{ijk} = \left(x_i - \sum_l \Pi_{ljk}^x x_l\right) / v_k^{1/2},$$

which implies a consumption growth process with states

$$g_{ijk} = \exp\left(\log g + x_j + v_k^{1/2} w_{ijk}\right).$$

The scaled value function is a function of the states x_t and v_t and solves the system of equations

$$u_{ij} = \left\{ (1 - \beta) + \beta \left[\sum_k \sum_l \Pi_{kij}^x \Pi_{lj}^v (u_{kl} g_{kij})^\alpha \right]^{\rho/\alpha} \right\}^{1/\rho}.$$

We compute a solution by value function iteration: we substitute an initial guess $\{u_{ij}(0)\}$ on the right-hand side, which generates a new value $\{u_{ij}(1)\}$. We repeat this process until the largest percentage change is smaller than 10^{-5} .

The approximation is highly accurate. In the top panel of Figure G1, we plot the discrete-grid and loglinear approximations of the value function against the state variable v_t with $x_t = 0$. The two solutions are literally indistinguishable in the figure. We superimpose the ergodic distribution of the conditional variance to provide some guidance on the relative importance of different regions of the state-space. We find similar agreement with other values of x_{t-1} , with plots of the value function versus x_t , and for calculations of entropy and horizon dependence. These conclusions are not affected by refining the grid or tightening the convergence criterion.

The discrete-grid approximation yields $I(1) = 0.0253$ and $H(120) = 0.0014$. If we use the loglinear approximation but keep the same state-space as in the discrete-grid approximation, we obtain $I(1) = 0.0254$ and $H(120) = 0.0014$. Therefore, the loglinear approximation has almost no effect on the entropy computations. In the case of the analytical loglinear approximation where the state-space allows for negative values of variance, $I(1) = 0.0249$ and $H(120) = 0.0014$ (column (2) of Table III). This small discrepancy in $I(1)$ arises from approximating the true variance with a process that allows for negative values. Neither approximation affects the horizon dependence.

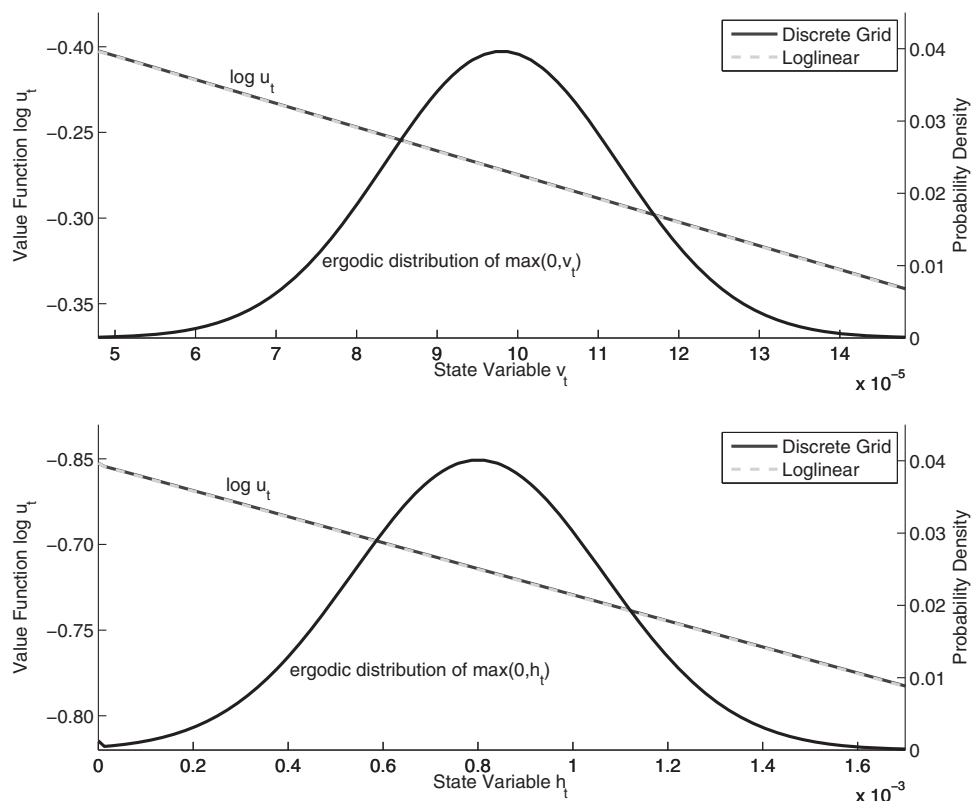


Figure G1. Numerical approximation of value functions with recursive utility. We compare value functions for recursive utility models computed by, respectively, discrete-grid and log-linear approximations. See Appendix I.G. The grid is fine enough to provide a close approximation to the true solution. The top panel refers to the stochastic variance model reported in column (1) of Table III. We plot the log value function $\log u_t$ against the state variable v_t holding x_t constant at zero. The discrete-grid approximation is the solid line, the loglinear approximation is the dashed line. The bell-shaped curve is the ergodic density function for the state, a discrete approximation of a normal density function. The bottom panel refers to the stochastic jump intensity model reported in column (2) of Table IV. Here we plot the log value function against intensity h_t . The curve is the ergodic density for $h'_t = \max(0, h_t)$, which results in a small blip near zero.

Stochastic Jump Intensity: The state-space representation of consumption growth dynamics in this case is

$$\log g_t = \log g' + v^{1/2} w_{gt} + z_{gt},$$

$$z_{gt} | j \sim N(j\theta, j\delta^2),$$

$$\text{Prob}(j) = \exp(-h'_{t-1}) h'^j_{t-1} / j!,$$

$$h_t = (1 - \varphi_h)h + \varphi_h h_{t-1} + \eta_0 w_{ht},$$

$$h'_t = \max\{0, h_t\}.$$

This model has a single state variable, h_t . We use parameter values from column (2) of Table IV.

We discretize the Poisson intensity h_t on a grid of 100 equally spaced points covering the interval $[h - 5\eta_0/(1 - \varphi_h^2)^{1/2}, h + 5\eta_0/(1 - \varphi_h^2)^{1/2}]$. We calculate transition probabilities using the same procedure as for the conditional variance process mentioned above. The true intensity is calculated from its normal counterpart by $h'_t = \max\{0, h_t\}$. For the jump z_{gt} , we use 10 Gauss–Hermite quadrature values, appropriately recentered and rescaled, as the discrete values, along with their associated probabilities. We truncate j at five. The scaled value function solves an equation analogous to the previous case and we use the same method to solve it.

We plot the results in the second panel of Figure G1. Here we see some impact from censoring. The ergodic distribution of intensity h_t has a small blip at the left end, reflecting censoring at zero. The effect is small, because zero is three standard deviations from the mean. This results in curvature of the value function as we approach zero, but it is too small to see in the figure.

The discrete-grid approximation yields $I(1) = 0.0490$ and $H(120) = 0.0025$. The loglinear approximation with the same state–space produces the same values. In the case of the analytical loglinear approximation where the state–space allows for negative values of jump intensity, $I(1) = 0.0502$ and $H(120) = 0.0025$ (column (2) of Table IV). Therefore, as in the case with stochastic variance, the loglinear approximation has almost no effect on entropy. The small discrepancy in $I(1)$ arises from approximating the true jump intensity with a process that allows for negative values. Neither approximation affects the horizon dependence.

Appendix H: Models Based on ARG Processes

We like the simplicity and transparency of linear processes; expressions like $v(b_1)$ summarize clearly and cleanly the impact of volatility dynamics. A less appealing feature is that they allow the conditional variance v_t and intensity h_t to be negative, as we have noted. Here we describe and solve an analogous model based on ARG(1) processes, discrete-time analogs of continuous-time square-root processes. See, for example, Gourieroux and Jasiak (2006) and Le, Singleton, and Dai (2010). The analysis parallels Appendix I.E.

Consider the consumption process

$$\begin{aligned}\log g_t &= \log g + \gamma(B)v_{t-1}^{1/2}w_{gt} + z_{gt}, \\ v_t &\sim \text{ARG}(c_v, \varphi_v, \delta_v), \\ h_t &\sim \text{ARG}(c_h, \varphi_h, \delta_h).\end{aligned}$$

The first-order autoregressive gamma for v_t and h_t implies

$$\begin{aligned}v_t &= \delta_v c_v + \varphi_v v_{t-1} + w_{vt}, \\ h_t &= \delta_h c_h + \varphi_h h_{t-1} + w_{ht},\end{aligned}$$

where w_{vt} and w_{ht} are martingale difference sequences with conditional variances equal to $\delta_v c_v^2 + 2\varphi_v c_v v_{t-1}$ and $\delta_h c_h^2 + 2\varphi_h c_h h_{t-1}$. The cumulant generating functions for v_t and h_t are:

$$\begin{aligned} k_t(s; v_{t+1}) &= \varphi_v s(1 - sc_v)^{-1} v_t - \delta_v \log(1 - sc_v), \\ k_t(s; h_{t+1}) &= \varphi_h s(1 - sc_h)^{-1} h_t - \delta_h \log(1 - sc_h). \end{aligned}$$

If one selects the ARG inputs

$$\begin{aligned} v_t &\sim \text{ARG}(\sigma_v^2/2, \varphi_v, (1 - \varphi_v)v/(\sigma_v^2/2)), \\ h_t &\sim \text{ARG}(\sigma_h^2/2, \varphi_h, (1 - \varphi_h)h/(\sigma_h^2/2)), \end{aligned}$$

then

$$\begin{aligned} v_t &= (1 - \varphi_v)v + \varphi_v v_{t-1} + w_{vt}, \\ h_t &= (1 - \varphi_h)h + \varphi_h h_{t-1} + w_{ht}, \end{aligned}$$

with variances of shocks equal to $\sigma_v^2[(1 - \varphi_v)v/2 + \varphi_v v_{t-1}]$ and $\sigma_h^2[(1 - \varphi_h)h/2 + \varphi_h h_{t-1}]$ and cumulant generating functions:

$$\begin{aligned} k_t(s; v_{t+1}) &= \varphi_v s(1 - s\sigma_v^2/2)^{-1} v_t - (1 - \varphi_v)v \log(1 - s\sigma_v^2/2)/(\sigma_v^2/2), \\ k_t(s; h_{t+1}) &= \varphi_h s(1 - s\sigma_h^2/2)^{-1} h_t - (1 - \varphi_h)h \log(1 - s\sigma_h^2/2)/(\sigma_h^2/2). \end{aligned}$$

We start with the value function:

- Guess. We guess a value function of the form

$$\log u_t = \log u + p_g(B)v_{t-1}^{1/2}w_{gt} + p_v v_t + p_h h_t,$$

with parameters to be determined.

- Compute. Since $\log(g_{t+1}u_{t+1})$ is

$$\begin{aligned} \log(g_{t+1}u_{t+1}) &= \log(gu) + [\gamma(B) + p_g(B)]v_t^{1/2}w_{gt+1} + z_{gt+1} + p_v v_{t+1} + p_h h_{t+1} \\ &= \log(gu) + [\gamma(B) + p_g(B) - (\gamma_0 + p_{g0})]v_t^{1/2}w_{gt+1} \\ &\quad + (\gamma_0 + p_{g0})v_t^{1/2}w_{gt+1} + z_{gt+1} + p_v v_{t+1} + p_h h_{t+1}, \end{aligned}$$

its certainty equivalent is

$$\begin{aligned} \log \mu_t(g_{t+1}u_{t+1}) &= \log(gu) + [\gamma(B) + p_g(B) - (\gamma_0 + p_{g0})]v_t^{1/2}w_{gt+1} \\ &\quad + (\alpha/2)(\gamma_0 + p_{g0})^2 v_t + [(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha]h_t \\ &\quad - \delta_v/\alpha \log(1 - \alpha p_v c_v) + \varphi_v p_v (1 - \alpha p_v c_v)^{-1} v_t \\ &\quad - \delta_h/\alpha \log(1 - \alpha p_h c_h) + \varphi_h p_h (1 - \alpha p_h c_h)^{-1} h_t. \end{aligned}$$

- Verify. We substitute the certainty equivalent into (24) and collect similar terms:

$$\text{constant: } \log u = b_0 + b_1[\log(gu) - \delta_v/\alpha \log(1 - \alpha p_v c_v) - \delta_h/\alpha \log(1 - \alpha p_h c_h)];$$

$$v_{t-1}^{1/2} w_{gt}: p_g(B) = b_1 \left[\frac{\gamma(B) + p_g(B) - (\gamma_0 + p_{g0})}{B} \right];$$

$$v_t: p_v = b_1[(\alpha/2)(\gamma_0 + p_{g0})^2 + \varphi_v p_v (1 - \alpha p_v c_v)^{-1}];$$

$$h_t: p_h = b_1[(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha + \varphi_h p_h (1 - \alpha p_h c_h)^{-1}].$$

The second equation is the same one we saw in Appendix I and has the same solution: $\gamma_0 + p_{g0} = \gamma(b_1)$.

The third and fourth equations are new. Their quadratic structure is different from anything we have seen so far, but familiar to anyone who has worked with square-root processes. The quadratic terms arise because risk to future utility depends on h_t and v_t through their innovations. We solve them using value function iterations: starting with zero, we substitute a value into the right side and generate a new value on the left. If this converges, we have the solution as the limit of a finite-horizon problem.

Another approach is to solve the quadratic equations directly and select the appropriate root. The third equation implies

$$\begin{aligned} 0 &= \alpha c_v p_v^2 + b_{pv} p_v + b_1 \alpha (\gamma_0 + p_{g0})^2 / 2, \\ b_{pv} &= b_1 \varphi_v - b_1 c_v \alpha^2 (\gamma_0 + p_{g0})^2 / 2 - 1. \end{aligned}$$

It has two real roots :

$$p_v = \frac{-b_{pv} \pm [b_{pv}^2 - 2b_1 c_v \alpha^2 (\gamma_0 + p_{g0})^2]^{1/2}}{2\alpha c_v}.$$

If the variance of $\log g_t$ is equal to zero, $p_v = 0$ only if we select the smaller root. Similar logic applies to p_h . The fourth equation implies

$$\begin{aligned} 0 &= \alpha c_h p_h^2 + b_{ph} p_h + b_1 (e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha, \\ b_{ph} &= b_1 \varphi_h - b_1 c_h (e^{\alpha\theta + (\alpha\delta)^2/2} - 1) - 1. \end{aligned}$$

The two roots are

$$p_h = \frac{-b_{ph} \pm [b_{ph}^2 - 4b_1 c_h (e^{\alpha\theta + (\alpha\delta)^2/2} - 1)]^{1/2}}{2\alpha c_h}.$$

Again, the discriminant must be positive. If it is, stability leads us to choose the smaller root.

Given these value function coefficients, the pricing kernel is

$$\begin{aligned} \log m_{t,t+1} = & \log \beta + (\rho - 1) \log g + (\alpha - \rho)(\delta_v \log(1 - \alpha p_v c_v)/\alpha + \delta_h \log(1 - \alpha p_h c_h)/\alpha) \\ & + (\alpha - 1)z_{gt+1} + [(\rho - 1)\gamma_0 + (\alpha - \rho)\gamma(b_1)]v_t^{1/2}w_{gt+1} \\ & + (\rho - 1)[\gamma(B)/B]_+ v_{t-1}^{1/2}w_{gt} \\ & + (\alpha - \rho) \{ p_v v_{t+1} - [\alpha(\gamma_0 + p_{g0})^2/2 + \phi_v p_v(1 - \alpha c_v p_v)^{-1}]v_t \} \\ & + (\alpha - \rho) \{ p_h h_{t+1} - [(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha + \phi_h p_h(1 - \alpha c_h p_h)^{-1}]h_t \}. \end{aligned}$$

Appendix I: Parameter Values for Models with Recursive Utility

Bansal–Yaron Models: The Bansal–Yaron (2004) growth rate process is the sum of an AR(1) and white noise. It implies, using their notation,

$$\begin{aligned} \text{Var}(\log g) &= \sigma^2 + (\varphi_e \sigma)^2/(1 - \rho^2), \\ \text{Cov}(\log g_t, \log g_{t-1}) &= \rho(\varphi_e \sigma)^2/(1 - \rho^2), \\ \text{Corr}(\log g_t, \log g_{t-1}) &= \text{Cov}(\log g_t, \log g_{t-1})/\text{Var}(\log g) \equiv \rho(1). \end{aligned}$$

With input from their Table I ($\rho = 0.979$, $\sigma = 0.0078$, $\varphi_e = 0.044$), the unconditional standard deviation is 0.0080 and the first autocorrelation is $\rho(1) = 0.0436$.

We construct an ARMA(1,1) with the same autocovariances. The essential parameters are $(\gamma_0, \gamma_1, \varphi_g)$, with the rest of the MA coefficients defined by $\gamma_{j+1} = \varphi_g \gamma_j = \varphi_g^j \gamma_1$ for $j \geq 1$. Set $\gamma_0 = 1$. This implies

$$\begin{aligned} \text{Var}(\log g) &= v[1 + \gamma_1^2/(1 - \varphi_g^2)]; \\ \text{Cov}(\log g_t, \log g_{t-1}) &= v[\gamma_1 + \varphi_g \gamma_1^2/(1 - \varphi_g^2)]; \\ \text{Corr}(\log g_t, \log g_{t-1}) &= \frac{\gamma_1 + \varphi_g \gamma_1^2/(1 - \varphi_g^2)}{1 + \gamma_1^2/(1 - \varphi_g^2)}. \end{aligned}$$

We set $\varphi_g = 0.979$ (BY's ρ). We choose γ_1 to match the autocorrelation $\rho(1)$, which gives us a quadratic in γ_1 :

$$[\varphi_g - \rho(1)]\gamma_1^2 + (1 - \varphi_g^2)\gamma_1 - \rho(1)(1 - \varphi_g^2) = 0.$$

We choose the root associated with an invertible moving average coefficient for reasons outlined in Sargent (1987, Section XI.15), which implies

$$\gamma_1 = \frac{-(1 - \varphi_g^2)^2 + \{(1 - \varphi_g^2) + 4[\varphi_g - \rho(1)](1 - \varphi_g^2)\rho(1)\}^{1/2}}{2[\varphi_g - \rho(1)]} = 0.0271.$$

Jump Models: Our starting point is calibration of the intensity process h_t used by Wachter (2013, Table I). Most of that consists of converting continuous-time objects to discrete time with a monthly time interval that we represent

by $\tau = 1/12$. We use the same mean value h that we used in our iid example: $h = 0.01\tau$. Monthly analogs to her parameters follow (analogs on the left, hers on the right):

$$\begin{aligned}\varphi_h &= e^{-\kappa\tau} = e^{-0.08/12} = 0.9934; \\ \eta_0 &= \bar{\lambda}^{1/2} \sigma_\lambda \tau^{1/2} = 0.0355^{1/2} \cdot 0.067 \cdot (1/12)^{1/2} = 0.0036.\end{aligned}$$

The process gives us a significant probability of negative intensity, which Wachter avoids by using a square-root process. We scale φ_h and η_0 back significantly, to 0.95 and 0.0001, respectively. Nevertheless, Table IV shows a significant contribution to horizon dependence from stochastic jump intensity.

Finding b_1 : We have described approximate solutions to recursive models given the values of the approximating constants b_0 and b_1 . We construct a fine grid over both and choose the values that come closest to satisfying equation (24).

REFERENCES

- Abel, Andrew, 1990, Asset prices under habit formation and catching up with the Joneses, *American Economic Review* 80, 38–42.
- Alvarez, Fernando, and Urban Jermann, 2005, Using asset prices to measure the persistence of the marginal utility of wealth, *Econometrica* 73, 1977–2016.
- Backus, David, Mikhail Chernov, and Ian Martin, 2011, Disasters implied by equity index options, *Journal of Finance* 66, 1969–2012.
- Bakshi, Gurdip, and Fousseni Chabi-Yo, 2012, Variance bounds on the permanent and transitory components of stochastic discount factors, *Journal of Financial Economics* 105, 191–208.
- Bansal, Ravi, Dana Kiku, and Amir Yaron, 2009, An empirical evaluation of the long-run risks model for asset prices, Working paper, Duke University.
- Bansal, Ravi, and Bruce N. Lehmann, 1997, Growth-optimal portfolio restrictions on asset pricing models, *Macroeconomic Dynamics* 1, 333–354.
- Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.
- Barro, Robert J., 2006, Rare disasters and asset markets in the twentieth century, *Quarterly Journal of Economics* 121, 823–867.
- Barro, Robert J., Emi Nakamura, Jon Steinsson, and Jose F. Ursua, 2009, Crises and recoveries in an empirical model of consumption disasters, Working paper, Harvard University.
- Bekaert, Geert, and Eric Engstrom, 2010, Asset return dynamics under bad environment-good environment fundamentals, Working paper, Columbia University.
- Benzoni, Luca, Pierre Collin-Dufresne, and Robert S. Goldstein, 2011, Explaining asset pricing puzzles associated with the 1987 market crash, *Journal of Financial Economics* 101, 552–573.
- Binsbergen, Jules van, Michael Brandt, and Ralph Koijen, 2012, On the timing and pricing of dividends, *American Economic Review* 102, 1596–1618.
- Branger, Nicole, Paulo Rodrigues, and Christian Schlag, 2011, The role of volatility shocks and rare events in long-run risk models, Working paper, Muenster University.
- Broadie, Mark, Mikhail Chernov, and Michael Johannes, 2009, Understanding index option returns, *Review of Financial Studies* 22, 4493–4529.
- Campbell, John Y., 1993, Intertemporal asset pricing without consumption data, *American Economic Review* 83, 487–512.
- Campbell, John Y., 1999, Asset prices, consumption, and the business cycle, in J.B. Taylor and M. Woodford, eds.: *Handbook of Macroeconomics*, Volume 1 (Elsevier, New York).

- Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.
- Chan, Yeung Lewis, and Leonid Kogan, 2002, Catching up with the Joneses: Heterogeneous preferences and the dynamics of asset prices, *Journal of Political Economy* 110, 1255–1285.
- Chapman, David, 2002, Does intrinsic habit formation actually resolve the equity premium puzzle? *Review of Economic Dynamics* 5, 618–645.
- Chernov, Mikhail, and Philippe Mueller, 2012, The term structure of inflation expectations, *Journal of Financial Economics* 106, 367–394.
- Chretien, Stephane, 2012, Bounds on the autocorrelation of admissible stochastic discount factors, *Journal of Banking and Finance* 36, 1943–1962.
- Cochrane, John, 1992, Explaining the variance of price-dividend ratios, *Review of Financial Studies* 5, 243–280.
- Constantinides, George, 1990, Habit formation: A resolution of the equity premium puzzle, *Journal of Political Economy* 98, 519–543.
- Deaton, Angus, 1993, *Understanding Consumption* (Oxford University Press, New York).
- Drechsler, Itamar, and Amir Yaron, 2011, What's vol got to do with it? *Review of Financial Studies* 24, 1–45.
- Duffee, Gregory R., 2010, Sharpe ratios in term structure models, Working paper, Johns Hopkins University.
- Epstein, Larry G., and Stanley E. Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937–969.
- Eraker, Bjorn, and Ivan Shaliastovich, 2008, An equilibrium guide to designing affine pricing models, *Mathematical Finance* 18, 519–543.
- Fama, Eugene and Kenneth French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.
- Gabaix, Xavier, 2012, Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance, *Quarterly Journal of Economics* 127, 645–700.
- Garcia, René, Richard Luger, and Eric Renault, 2003, Empirical assessment of an intertemporal option pricing model with latent variables, *Journal of Econometrics* 116, 49–83.
- Ghosh, Anisha, Christian Julliard, and Alex Taylor, 2011, What is the consumption-CAPM missing? An information-theoretic framework for the analysis of asset pricing models, Working paper, Carnegie Mellon University.
- Gourieroux, Christian, and Joann Jasiak, 2006, Autoregressive gamma processes, *Journal of Forecasting* 25, 129–152.
- Hansen, Lars Peter, 2012, Dynamic value decomposition in stochastic economies, *Econometrica* 80, 911–967.
- Hansen, Lars Peter, John C. Heaton, and Nan Li, 2008, Consumption strikes back? Measuring long-run risk, *Journal of Political Economy* 116, 260–302.
- Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.
- Hansen, Lars Peter, and Thomas J. Sargent, 1980, Formulating and estimating dynamic linear rational expectations models, *Journal of Economic Dynamics and Control* 2, 7–46.
- Hansen, Lars Peter, and Thomas J. Sargent, 2008, *Robustness* (Princeton University Press, Princeton, NJ).
- Hansen, Lars Peter, and Jose Scheinkman, 2009, Long term risk: An operator approach, *Econometrica* 77, 177–234.
- Heaton, John, 1995, An empirical investigation of asset pricing with temporally dependent preference specifications, *Econometrica* 63, 681–717.
- Koijen, Ralph, Hanno Lustig, Stijn Van Nieuwerburgh, and Adrien Verdelhan, 2009, The wealth-consumption ratio in the long-run risk model, *American Economic Review Papers and Proceedings* 100, 552–556.
- Koopmans, Tjalling C., 1960, Stationary ordinal utility and impatience, *Econometrica* 28, 287–309.
- Kreps, David M., and Evan L. Porteus, 1978, Temporal resolution of uncertainty and dynamic choice theory, *Econometrica* 46, 185–200.

- Le, Ahn, Kenneth Singleton, and Qiang Dai, 2010, Discrete-time affine Q term structure models with generalized market prices of risk, *Review of Financial Studies* 23, 2184–2227.
- Lettau, Martin, and Harald Uhlig, 2000, Can habit formation be reconciled with business cycle facts? *Review of Economic Dynamics* 3, 79–99.
- Longstaff, Francis A., and Monika Piazzesi, 2004, Corporate earnings and the equity premium, *Journal of Financial Economics* 74, 401–421.
- Martin, Ian, 2013, Consumption-based asset pricing with higher cumulants, *Review of Economic Studies* 80, 745–773.
- Otrok, Christopher, B. Ravikumar, and Charles H. Whiteman, 2002, Habit formation: A resolution of the equity premium puzzle? *Journal of Monetary Economics* 49, 1261–1288.
- Sargent, Thomas J., 1987, *Macroeconomic Theory* (Second Edition) (Academic Press, San Diego).
- Sims, Chris, 2003, Implications of rational inattention, *Journal of Monetary Economics* 50, 665–690.
- Smets, Frank, and Raf Wouters, 2003, An estimated dynamic stochastic general equilibrium model of the Euro area, *Journal of the European Economic Association* 1, 1123–1175.
- Stutzer, Michael, 1996, A simple nonparametric approach to derivative security valuation, *Journal of Finance* 51, 1633–1652.
- Sundaresan, Suresh, 1989, Intertemporally dependent preferences and the volatility of consumption and wealth, *Review of Financial Studies* 2, 73–89.
- Tauchen, George, 1986, Finite state Markov-chain approximations to univariate and vector autoregressions, *Economics Letters* 20, 177–181.
- Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2010, Information acquisition and portfolio underdiversification, *Review of Economic Studies* 77, 779–805.
- Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177–188.
- Verdelhan, Adrien, 2010, A habit-based explanation of the exchange rate risk premium, *Journal of Finance* 65, 123–145.
- Wachter, Jessica, 2006, A consumption-based model of the term structure of interest rates, *Journal of Financial Economics* 79, 365–399.
- Wachter, Jessica, 2013, Can time-varying risk of rare disasters explain aggregate stock market volatility? *Journal of Finance* 68, 987–1035.
- Weil, Philippe, 1989, The equity premium puzzle and the risk-free rate puzzle, *Journal of Monetary Economics* 24, 401–421.