Sources of entropy in representative agent models[*]

David Backus,[†] Mikhail Chernov,[‡] and Stanley Zin[§]

Rough draft: March 30, 2011
This revision: July 11, 2011

Abstract

We propose two metrics for asset pricing models and apply them to representative agent models with recursive preferences, habits, and jumps. The metrics describe the pricing kernel's dispersion (the entropy of the title) and dynamics (time dependence, a measure of how entropy varies over different time horizons). We show how each model generates entropy and time dependence and compare their magnitudes to estimates derived from asset returns. This exercise — and transparent loglinear approximations — clarifies the mechanisms underlying these models. It also reveals, in some cases, tension between entropy, which should be large enough to account for observed excess returns, and time dependence, which should be small enough to account for mean yield spreads.

JEL Classification Codes: E44, G12.

Keywords: pricing kernel, asset returns, bond yields, recursive preferences, habits, jumps, disasters.

1 Introduction

Representative agent models are the workhorses of the macro-finance literature, with models based on recursive preferences and habits contributing prominently. The two approaches are often treated as substitutes, two ways to generate realistic asset prices and returns, yet their mechanisms are quite different. Are they similar, then, or different?

We explore the mechanics of asset pricing models, paying special attention to these popular examples of representative agent models. Our first step is to develop metrics that summarize the properties of asset pricing models. We base them on the pricing kernel, because every model has one, and on entropy, for reasons we outline shortly. One metric concerns the pricing kernel's dispersion, which we measure with *entropy*. We show that the entropy of the pricing kernel is an upper bound on mean excess returns. Roughly speaking, this entropy bound is a logarithmic version of the well-known Hansen-Jagannathan (1991) bound. It differs in using entropy as its measure of dispersion of the pricing kernel and in defining excess returns as differences in logarithms of gross returns. The other metric concerns the pricing kernel's dynamics. We summarize its dynamics with what we call *time dependence*, a measure of how entropy varies with the time horizon. As with entropy, we can infer its magnitude from asset returns: negative (positive) time dependence is associated with an upward-sloping (downward-sloping) mean yield curve and positive (negative) mean yield spreads.

These entropy-based metrics are, in our view, natural ones for our purpose. One reason is that most popular asset pricing models are loglinear, or nearly so. Logarithmic measures like entropy and log-differences in returns are easily computed for them. A second reason is that entropy extends more easily to multiple periods than, say, the mean and standard deviation of the pricing kernel used in the Hansen-Jagannathan bound. Finally, entropy incorporates nonnormal components of the pricing kernel and returns in a particularly simple and transparent way. Similar reasoning underlies the treatment of long-run risk in Hansen (2008) and Hansen and Scheinkman (2009). All of this will be clearer once we've developed the appropriate tools.

We think these metrics give us some insight into the behavior of popular asset pricing models. The data suggests that a realistic model should have substantial entropy (to account for observed mean excess returns) and modest time dependence (to account for observed mean yield spreads). In representative agent models, the two features are often linked. Although the specifics differ, dynamics play a fundamental role in models based on recursive preferences and habits. With recursive preferences, dynamics in the consumption growth process are required to distinguish them from additive power utility. In the popular Bansal-Yaron (2004) model this takes the form of persistence in consumption growth and its conditional variance. With habits, dynamics enter preferences directly. The challenge in each case is to use these dynamics to increase entropy relative to power utility without producing unrealistically large time dependence. The models are similar, therefore, in this respect: tension between entropy and time dependence is an inherent feature.

We illustrate the tension between entropy and time dependence and point to a number of ways of resolving it. One of the most useful approaches, in our view, is to introduce jumps: nonnormal innovations in consumption growth. Skewness and kurtosis are evident in asset returns, so it seems natural to include them in asset pricing models. Such jump risk can be added to either class of models. With recursive preferences, persistent jump risk can increase entropy substantially with only a small impact on time dependence.

These and other approaches are described below. We use closed-form loglinear approximations 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. Latin letters are variables or collections of 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 $\log x_t$, but the subtle difference between the two has no bearing on anything important. (iii) The letter 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: $a(B) = a_0 + a_1 B + a_2 B^2 + \cdots$.

2 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 show that mean excess returns on equity, bonds, and other assets correspond to properties of the pricing kernel, specifically its dispersion and time dependence. We base these properties on entropy, a dispersion concept that is particularly convenient in the loglinear environments common in the asset pricing literature.

2.1 Properties of asset returns

We begin with a summary of the salient properties of excess returns. In Table 1 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's helpful to collect it in one place. Excess returns are measured as differences in logs of gross returns over the one-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 over 5%, skewness of -0.4, and excess kurtosis of 7.9.

Nonzero values of skewness and excess kurtosis are a clear 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-French 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 (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 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 five-year US Treasury over the one-month Treasury (0.15% in our sample) is evident in the one-year bond (0.08%). The increase in excess returns with maturity corresponds to a mean yield curve that also increases with maturity over this range. (See, for example, Appendix A.1.) The numbers are similar: the mean spread between yields on one-month and five-year Treasuries over the last three decades has been 1.2% annually or 0.1% monthly. Backus, Foresi, Mozumdar and Wu (2001, Table 2) is one of the many sources of evidence on this point. All of these numbers refer to nominal bonds. Data on inflation-indexed bonds is available for only a short sample and a limited range of maturities, leaving some range of opinion about their properties, including the mean slope of the real yield curve. 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 (2008) suggest that yield spreads are smaller on real than nominal bonds, which would make our estimates upper bounds.

These properties of returns are estimates, but they're suggestive of the facts a theoretical model might try to explain. Our list includes: (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 $0.01 = 1\%$ per month. The exact number isn't critical, but it's 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 that are unlikely to be larger than $0.001 = 0.1\%$ monthly in absolute value. (iii) Excess returns on many assets are decidedly nonnormal.

2.2 Entropy

The next step is to use this evidence to characterize the pricing kernel. Here we introduce entropy as a measure of dispersion and show that the entropy of the pricing kernel bounds mean excess returns from above. This entropy bound is a close relative of the well-known Hansen-Jagannathan (1991) bound and extends related work by Alvarez and Jermann (2005) and Bansal and Lehman (1997).

The foundation for both bounds is the fundamental result in asset pricing theory: in arbitrage-free environments, there is a positive random variable m that satisfies

$$E_t\left(m_{t+1}r_{t+1}\right) \;=\; 1 \tag{1}$$

for gross returns r on all traded assets. We refer to m as the *pricing kernel*. Here E_t denotes the expectation based on the distribution conditioned on the state at date t. In the stationary ergodic environments we study, the unconditional expectation E is computed from the unique equilibrium or invariant distribution.

We define the conditional entropy of the pricing kernel by

$$L_t(m_{t+1}) \;\equiv\; \log E_t m_{t+1} - E_t \log m_{t+1}. \tag{2}$$

This is the entropy of the risk-neutral distribution relative to the true distribution or the Kullback-Leibler divergence; see Backus, Chernov, and Martin (2011, Section II). Note that entropy is a measure of dispersion: it is nonnegative and equals zero only if m_{t+1} is (conditionally) constant. If m_{t+1} is lognormal, with $\log m_{t+1}|t \sim \mathcal{N}(\kappa_{1t}, \kappa_{2t})$, conditional entropy is

$$L_t(m_{t+1}) \;=\; \log E_t m_{t+1} - E_t \log m_{t+1} \;=\; (\kappa_{1t} + \kappa_{2t}/2) - \kappa_{1t} \;=\; \kappa_{2t}/2.$$

It includes additional terms if $\log m_{t+1}$ is not normal.

There's a nice connection between entropy and the cumulants of $\log m_{t+1}$. Cumulants, of course, are close relatives of moments. The (conditional) cumulant-generating function (if it exists) for $\log m_{t+1}$ is the log of its (conditional) moment-generating function:

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

a function of the real variable s. As before, we denote conditioning with a subscript t. With enough regularity, it has the power series expansion

$$k_t(s) \;=\; \log E_t\left(e^{s \log m_{t+1}}\right) \;=\; \sum_{j=1}^{\infty} \kappa_{jt} s^j / j!$$

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

γ_{2t} are scaled versions of the third and fourth cumulants: $\gamma_{1t} = \kappa_{3t}/\kappa_{2t}^{3/2}$ and $\gamma_{2t} = \kappa_{4t}/\kappa_{2t}^2$. Entropy is therefore

$$
\begin{aligned}
L_t(m_{t+1}) &= k_t(1) - \kappa_{1t} \\
&= \underbrace{\kappa_{2t}(\log m_{t+1})/2!}_{\text{normal term}} + \underbrace{\kappa_{3t}(\log m_{t+1})/3! + \kappa_{4t}(\log m_{t+1})/4! + \cdots}_{\text{high-order cumulants}}.
\end{aligned}
\tag{3}
$$

If the conditional distribution of $\log m_{t+1}$ is normal, then only the variance κ_{2t} appears. In other cases, high-order cumulants (those of order $j \geq 3$) also appear. These high-order terms are a defining feature of models with jumps and disasters.

These two ingredients — the pricing relation (1) and the conditional entropy of the pricing kernel (2) — lead to the entropy bound:

$$
EL_t(m_{t+1}) \geq E\left(\log r_{t+1} - \log r_{t+1}^1\right),
\tag{4}
$$

where r^1 is the (gross) return on a one-period riskfree bond. In words: mean excess returns are bounded above by mean conditional entropy of the pricing kernel. Closely related results are reported by Alvarez and Jermann (2005) and Bansal and Lehman (1997). For convenience, we refer to mean conditional entropy simply as entropy.

We derive the bound (4) as follows. Since log is a concave function, the pricing relation (1) and Jensen's inequality imply that for any positive return r_{t+1},

$$
E_t \log m_{t+1} + E_t \log r_{t+1} \leq \log(1) = 0,
\tag{5}
$$

with equality iff $m_{t+1} r_{t+1} = 1$. The log return with the highest mean is, evidently, $r_{t+1} = 1/m_{t+1}$. For the excess return, we need the return on a one-period bond (a claim to "one" next period). Its price is $q_t^1 = E_t m_{t+1}$, so the short rate is $r_{t+1}^1 = 1/q_t^1$. The log of the short rate is therefore

$$
\log r_{t+1}^1 = -\log q_t^1 = -\log E_t m_{t+1} = -L_t(m_{t+1}) - E_t \log m_{t+1}.
$$

If we subtract this from (5), we have

$$
L_t(m_{t+1}) \geq E_t\left(\log r_{t+1} - \log r_{t+1}^1\right).
$$

We take the unconditional expectation of both sides to produce (4).

The entropy bound (4), like the Hansen-Jagannathan (1991) bound, produces an upper bound on excess returns from the dispersion of the pricing kernel. In this 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. We explore the differences further in Appendix A.2.

5

2.3 Time dependence

With the entropy bound, we use information about excess returns to describe the pricing kernel's dispersion. We now show how bond yields can be used to describe its dynamics.

Bond prices follow from the pricing kernel. Let q_t^n be the price at date t of a claim to one at $t + n$. The one-period return on this bond is $r_{t+1}^n = q_{t+1}^{n-1}/q_t^n$. Equation (1) then gives us recursive pricing of bonds:

$$q_t^n \;\; = \;\; E_t\left(m_{t+1}q_{t+1}^{n-1}\right) \tag{6}$$

starting with $q_t^0 = 1$. Repeated substitution leads to

$$q_t^n \;\; = \;\; E_t\left(m_{t+1}m_{t+2}\cdots m_{t+n}\right) \;\; = \;\; E_t m_{t,t+n}, \tag{7}$$

where $m_{t,t+n} \equiv m_{t+1}m_{t+2}\cdots m_{t+n}$ is compact notation for the *multiperiod pricing kernel*. Yields are defined from prices by $q_t^n = \exp(-ny_t^n)$ or $ny_t^n = -\log q_t^n$.

We summarize the pricing kernel's dynamics with the entropy of the multiperiod pricing kernel, which we term *multiperiod entropy*. From the pricing relation (7) for bonds and the definition of entropy, you might guess that multiperiod entropy is closely related to long bond prices. Multiperiod conditional entropy is

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

Mean conditional multiperiod entropy (multiperiod entropy, for short) is therefore

$$EL_t(m_{t,t+n}) \;\; = \;\; E\log q^n - nE\log m \;\; = \;\; -nEy^n - nE\log m. \tag{8}$$

The second equality follows from the definition of the n-period yield y^n.

We use multiperiod entropy to describe the pricing kernel's dynamics. Suppose, for example, that successive pricing kernels are iid (independent and identically distributed). Then mean multiperiod conditional entropy is simply a scaled-up version of one-period entropy:

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

This is a generalization of a well-known result: the variance of a random walk is proportional to the time interval. We refer to this as a case of neutral or zero time dependence. More formally, we define time dependence by the per-period difference between the left and right sides:

$$T(n) \;\; = \;\; n^{-1}EL_t(m_{t,t+n}) - EL_t(m_{t+1}). \tag{9}$$

If $T(n)$ is positive we say there is positive time dependence, if negative, we say there is negative time dependence. It's possible for $T(n)$ to be positive for some values of n and negative for others, but that doesn't occur in the examples we study. In a loglinear model, such as the one in the next section, time dependence is governed by the autocorrelation of the log pricing kernel. If autocorrelations are negative, then time dependence is, too.

More important for our purposes, time dependence is tied to bond yields. The difference between equation (8) for arbitrary n and $n = 1$ implies

$$T(n) \;=\; n^{-1}EL_t(m_{t,t+n}) - EL_t(m_{t+1}) \;=\; -E(y_t^n - y_t^1). \tag{10}$$

In words: time dependence is negative if the mean yield curve slopes upward, positive if it slopes downward, and zero if it's flat.

If we look at time dependence through the lens of the evidence, it's clear that it must be small relative to entropy. Why? Because mean yield spreads (and returns on long bonds) are small relative to returns on other assets.

3 Models of pricing kernels

We apply the concepts of entropy and time dependence to a loglinear example, a modest generalization of the Vasicek (1977) model. We show, in this relatively simple setting, how parameter values based on observed asset returns generate the two properties we've emphasized: entropy must be larger than observed excess returns (say, $0.01 = 1\%$ per month) and time dependence must be small in comparison (no larger in absolute value than $0.001 = 0.1\%$ at 60 months). We then compare the model to one based on a representative agent with additive power utility.

3.1 The Vasicek model

Consider the loglinear statistical model of the pricing kernel,

$$\log m_t \;=\; \log m + \sum_{j=0}^{\infty} a_j w_{t-j} \;=\; \log m + a(B)w_t, \tag{11}$$

where w_t is iid $\mathcal{N}(0,1)$, $\sum_j a_j^2 < \infty$, and B is the lag or backshift operator. The implicitly defined lag polynomial $a(B)$ is described in Appendix A.3 along with some of its uses. The infinite moving average gives us control over the pricing kernel's dynamics, which reappear in bond yields and time dependence. We'll see that a_0 governs entropy and the other coefficients govern time dependence. Here m_t is the real pricing kernel, which provides a direct comparison with marginal rates of substitution in representative agent models.

Bond prices, returns, and yields follow from their definitions, the pricing relation (6), and the pricing kernel (11). See Appendix A.4. The short rate is

$$\log r_{t+1}^1 = -(\log m + A_0^2/2) - \sum_{j=0}^{\infty} a_{j+1} w_{t-j} = -(\log m + A_0^2/2) + [a(B)/B]_+ w_t. \quad (12)$$

The subscript "+" means ignore negative powers of B; see Appendix A.3. The partial sums $A_n = \sum_{j=0}^{n} a_j$ include $A_0 = a_0$. Mean excess returns and yield spreads are

$$E(\log r_{t+1}^n - \log r_{t+1}^1) = \left(A_0^2 - A_{n-1}^2\right)/2$$

$$E(y_t^n - y_t^1) = n^{-1} \sum_{j=1}^{n} (A_0^2 - A_{j-1}^2)/2.$$

We see here how the dynamics of the pricing kernel, represented by the moving average coefficients a_j and their partial sums A_j, affect yields and returns. If $a_j = 0$ for $j \geq 1$ (the iid case), then $A_j = A_0$ for all j, interest rates are constant, the mean yield curve is flat, and mean excess returns on long bonds are zero. Otherwise, excess returns and yield spreads are governed by terms like $A_0^2 - A_{j-1}^2$.

The same components drive entropy and time dependence. Conditional entropy is

$$L_t(m_{t+1}) = \log E_t m_{t+1} - E_t \log m_{t+1} = a_0^2/2 = A_0^2/2. \quad (13)$$

Entropy is the same. Multiperiod entropy is

$$L_t(m_{t,t+n}) = \sum_{j=1}^{n} A_{j-1}^2/2.$$

Time dependence is the difference, averaged over the number of periods:

$$T(n) = n^{-1} \sum_{j=1}^{n} \left(A_{j-1}^2 - A_0^2\right)/2, \quad (14)$$

the mirror image of the mean yield spread. Here, too, you can see how the pricing kernel's dynamics interact with its properties. A typical term in the sum can be expressed

$$A_j^2 - A_0^2 = [(A_j - A_0) + A_0]^2 - A_0^2$$
$$= (A_j - A_0)^2 + 2(A_j - A_0)A_0.$$

A realistic model must have large enough entropy to account for observed excess returns and small enough time dependence to account for the size of term premiums relative to excess returns on other assets. The former requires large (in absolute value) $a_0 = A_0$. The latter requires small $A_j - A_0$, which in turn requires the individual moving average coefficients a_j for $j \geq 1$ to be small relative to the initial term a_0. A useful indicator in this regard is the ratio, $\theta = a_1/a_0$. For $|a_1| \ll |a_0|$ we need $|\theta| \ll 1$. Departures from the iid case ($a_j = 0$ for $j \geq 1$) can't be too big.

3.2 Entropy and time dependence in the Vasicek model

We can see more clearly how this works if we add some structure and choose parameter values that approximate the salient features of interest rates. We make $\log m_t$ an ARMA(1,1) process; see Appendix A.3. This leads to an AR(1) for the short rate, which turns the model into a legitimate discrete-time version of Vasicek. We set $a_{j+1} = \varphi a_j$ for $j \geq 1$ for $0 \leq \varphi < 1$. That gives us three free parameters: the initial value a_0 (positive by convention), the second value a_1, and the rate of decay φ. All of these parameters can be inferred from bond yields. 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 one-month and five-year bonds. The result is a statistical model of the pricing kernel that captures some of its central features.

With this structure, we can connect parameters to properties of bond returns. The (log) short rate, equation (12), is AR(1) with autocorrelation parameter φ. We set $\varphi = 0.85$, an estimate of the monthly autocorrelation of the real short rate reported by Chernov and Mueller (2008). The variance of the short rate is

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

Chernov and Mueller report a standard deviation of $(0.02/12)$ (2% annually), which implies $|a_1| = 8.78 \times 10^{-4}$. Finally, we choose a_0 to match the mean yield spread on the five-year bond. If the yield spread is $E(y^{60} - y^1) = 0.001 = 0.1\%$ a month, this implies $a_0 = 0.1953$ and $a_1 < 0$. The ratio $\theta = a_1/a_0 = -0.0045$ is, indeed, much smaller than one in absolute value. If we make a_1 positive, the yield spread is negative.

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 to the reverse. We see in both cases that the initial coefficient a_0 is larger by a wide margin. In fact, we have truncated it to make the others visible. The primary difference is in sign: an upward sloping mean yield curve requires a_0 and a_1 to have opposite signs, which we accomplish here by making a_1 and its successors negative. Negative mean yield spreads lead us to make a_j positive for $j \geq 1$.

We see the same thing from multiple perspectives in Figure 2. In each quadrant, the solid line corresponds to parameter values chosen to generate a positive mean yield spread and the dashed line corresponds to a negative mean yield spread. The upper left quadrant contains sequences of moving average coefficients: $a_0, a_1, a_2, \ldots$. From this perspective, the coefficients are essentially zero after the first one and the log pricing kernel is approximately white noise. Departures from white noise are evident in the lower left quadrant, where we zoom in on the sequence starting with a_1. Both sequences approach zero at rate φ starting with a_1. The upper right quadrant shows the resulting partial sums A_j, which appear in expressions for yield spreads and time dependence. The lower right quadrant shows the mean yield spreads.

Entropy and time dependence are pictured in Figure 3. The dotted line in the middle is our estimated entropy lower bound. The top three lines report average multiperiod entropy: multiperiod entropy divided by the number of periods. The top dashed line corresponds to positive time dependence: per period entropy rises (slightly) with the number of periods. The center line is (one-period) entropy, which serves here as a benchmark. The downward-sloping line below it corresponds to negative time dependence. The three lines at the bottom are time dependence, which we see is small relative to entropy. The dotted lines surrounding them are our estimated upper and lower bounds. Time dependence hits the bounds by construction: we chose parameters precisely to match these quantities at a time horizon of 60 months.

3.3   A representative agent with additive power utility

In representative agent models, pricing kernels are marginal rates of substitution. Typically a pricing kernel follows from applying a preference ordering to a given consumption growth process. The form is particularly simple if the representative agent has additive power utility. To fix notation, let period utility be c^ρ/ρ with $\rho < 1$. Then the pricing kernel is

$$m_{t+1} \;\;=\;\; \beta(c_{t+1}/c_t)^{\rho-1} \;\;=\;\; \beta g_{t+1}^{\rho-1}, \tag{15}$$

where $g_{t+1} = c_{t+1}/c_t$ is consumption growth. Here ρ governs both risk aversion and intertemporal substitution.

If the consumption growth process is loglinear, the pricing kernel has the same form as the Vasicek model. Let

$$\log g_t \;\;=\;\; \log g + \gamma(B)v^{1/2}w_t, \tag{16}$$

where $\{w_t\} \sim \mathrm{NID}(0,1)$, $\gamma_0 = 1$ (a normalization), and v is the conditional variance. The pricing kernel is then

$$\log m_t \;\;=\;\; \log \beta + (\rho-1)\log g + (\rho-1)\gamma(B)v^{1/2}w_t.$$

This is equivalent to the Vasicek model (11), with all the moving average coefficients scaled by $(\rho-1)$: $a_j = (\rho-1)v^{1/2}\gamma_j$. The ratio $\theta = a_1/a_0 = \gamma_1/\gamma_0$ is controlled entirely by the consumption process.

With this structure, entropy and time dependence are both scaled by $(\rho-1)^2$: if we increase $1-\rho$ to magnify entropy, we also raise the absolute value of time dependence. Consider a given consumption growth process with partial sums $\Gamma_n = \sum_{j=0}^n \gamma_j$. Entropy is

$$EL_t(m_{t+1}) \;\;=\;\; a_0^2/2 \;\;=\;\; (\rho-1)^2\Gamma_0^2 v/2 \;\;=\;\; (\rho-1)^2 v/2,$$

which puts a lower bound on $1 - \rho$. If the conditional variance of monthly consumption growth is 0.0072^2 (see Table 2), then to generate entropy of (say) 0.01, we need $1 - \rho \geq (2 \times 0.01)^{1/2}/0.0072 = 19.6$. Time dependence is also scaled by $(\rho - 1)^2$:

$$T(n) \;\; = \;\; n^{-1}(\rho - 1)^2 \sum_{j=1}^{n} (\Gamma_{j-1}^2 - \Gamma_0^2) v/2.$$

Our upper bound on the absolute value of time dependence thus places an upper bound on $1 - \rho$. Whether we can satisfy both bounds at the same time depends on the consumption growth process. This tension between entropy and time dependence is a feature of many of the models we study.

We study three classes of models in the following sections: models based on recursive preferences, habits, and jumps. Recursive preferences and habits each add an extra term to the pricing kernel relative to the additive case. With recursive preferences, the extra term involves future utility, which in turn depends on the consumption growth process. With habits, the extra term involves a state variable (the habit) whose dynamics depend on past consumption. Jumps are a device for introducing nonnormal innovations into models. We show how each class of models works, focusing on their pricing kernels and the ways in which they generate entropy and time dependence.

4 Pricing kernels with recursive preferences

Our first class of representative agent models is based on Bansal and Yaron (2004), who show that a combination of consumption dynamics and recursive preferences can generate risk premiums similar to those we observe. They build on related work by Campbell (1993), Epstein and Zin (1989), and Weil (1989).

4.1 Recursive preferences

We define "utility from date t on" U_t recursively with the time aggregator,

$$U_t \;\; = \;\; [(1 - \beta)c_t^\rho + \beta \mu_t(U_{t+1})^\rho]^{1/\rho}, \tag{17}$$

and certainty equivalent function,

$$\mu_t(U_{t+1}) \;\; = \;\; \left[E_t(U_{t+1}^\alpha)\right]^{1/\alpha}. \tag{18}$$

Additive power utility is a special case with $\alpha - \rho = 0$. 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 terminology is a

useful shortcut, but it's somewhat misleading: α describes risk aversion over future utility, which depends on (among other things) ρ. As in other multigood environments, there is no clear separation between preference across goods and preference across states.

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 (17) becomes

$$u_t = [(1-\beta) + \beta\mu_t(g_{t+1}u_{t+1})^\rho]^{1/\rho}, \tag{19}$$

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

We use a loglinear approximation of (19) to give us transparent closed-form expressions for pricing kernels. The approach is similar to that of Hansen, Heaton, and Li (2008, Section III). The loglinear approximation is

$$\begin{aligned} \log u_t &= \rho^{-1} \log\left[(1-\beta) + \beta\mu_t(g_{t+1}u_{t+1})^\rho\right] \\ &= \rho^{-1} \log\left[(1-\beta) + \beta e^{\rho \log \mu_t(g_{t+1}u_{t+1})}\right] \\ &\approx b_0 + b_1 \log \mu_t(g_{t+1}u_{t+1}). \end{aligned} \tag{20}$$

The last line is a first-order approximation of $\log u_t$ in $\log \mu_t$ around the point $\log \mu_t = \log \mu$, with

$$\begin{aligned} b_1 &= \beta e^{\rho \log \mu}/[(1-\beta) + \beta e^{\rho \log \mu}] \\ b_0 &= \rho^{-1} \log[(1-\beta) + \beta e^{\rho \log \mu}] - b_1 \log \mu. \end{aligned} \tag{21}$$

The equation is exact when $\rho = 0$, in which case $b_0 = 0$ and $b_1 = \beta$. Otherwise, it is the only source of approximation in what follows.

The pricing kernel (marginal rate of substitution) is

$$\begin{aligned} m_{t+1} &= \beta(c_{t+1}/c_t)^{\rho-1} \left[U_{t+1}/\mu_t(U_{t+1})\right]^{\alpha-\rho} \\ &= \beta g_{t+1}^{\rho-1} \left[g_{t+1}u_{t+1}/\mu_t(g_{t+1}u_{t+1})\right]^{\alpha-\rho}. \end{aligned} \tag{22}$$

See Appendix A.5. This has a convenient loglinear structure as long as g and u do. The pricing kernel reduces to additive power utility (15) in two cases: when $\alpha - \rho = 0$ 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 can't 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.

12

4.2 Entropy and time dependence with consumption dynamics

Our first example generates additional entropy relative to additive power utility through a combination of recursive preferences and a persistent component in consumption growth. The model is Bansal and Yaron's (2004) Case I with one change: we replace their bivariate process for log consumption growth with a univariate process that has the same autocovariance function. We think this captures their idea in a (slightly) simpler way.

Consider the univariate loglinear consumption growth process (16). Persistence is reflected in nonzero values of γ_j for $j \geq 1$. The loglinear approximation to the pricing kernel is

$$
\begin{aligned}
\log m_{t+1} &= \log \beta + (\rho - 1) \log g - (\alpha - \rho)(\alpha/2)\gamma(b_1)^2 v \\
&\quad + [(\rho - 1)\gamma_0 + (\alpha - \rho)\gamma(b_1)]v^{1/2}w_{t+1} + (\rho - 1)[\gamma(B)/B]_+ v^{1/2}w_t. \quad (23)
\end{aligned}
$$

See Appendix A.6. This has the same form as the Vasicek model (11) with

$$
a_j = \left\{ \begin{array}{ll} [(\rho - 1)\gamma_0 + (\alpha - \rho)\gamma(b_1)]v^{1/2} & j = 0 \\ (\rho - 1)\gamma_j v^{1/2} & j \geq 1. \end{array} \right.
$$

Entropy then follows from equation (13):

$$
EL_t(m_{t+1}) = a_0^2/2 = [(\rho - 1)\gamma_0 + (\alpha - \rho)\gamma(b_1)]^2 v/2.
$$

Time dependence follows from equation (14).

The only departure from additive power utility is in the initial moving average coefficient, a_0. The key ingredient is the term

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

The logic for this sum is that the impact of shocks to consumption growth on future utility depends on their persistence (represented by γ) and discounting (represented by b_1). If $\alpha - \rho = 0$, none of this matters: future utility does not appear in the pricing kernel and we're back to the additive case. Similarly, if consumption growth is iid, so that $\gamma_j = 0$ for $j \geq 1$ and $\gamma(b_1) = \gamma_0$, the coefficient of w_{t+1} becomes $(\alpha - 1)\gamma_0 v^{1/2}$, which is indistinguishable from additive power utility. The impact of recursive preferences depends, then, on having dynamics in consumption growth. With $\gamma(b_1) > 0$ and $\alpha - \rho < 0$, the ratio $\theta_g = a_1/a_0$ is smaller than with additive power utility.

We get a quantitative sense of the importance of consumption growth dynamics for entropy from a numerical example. We impose the ARMA(1,1) structure we used in the Vasicek model: $\gamma_{j+1} = \varphi_g \gamma_j$ for $j \geq 1$ with $0 \leq \varphi_g < 1$. With this structure, we have

$$
\gamma(b_1) = 1 + b_1 \gamma_1/(1 - b_1 \varphi_g),
$$

which increases as b_1 and φ_g approach one. We use parameter values adapted from Bansal, Kiku, and Yaron (2009); see Appendix A.9.

We consider two thought experiments that highlight the difference between recursive preferences and additive power utility. Bansal, Kiku, and Yaron (2009) use, in our notation, $\alpha = -9$ and $\rho = 1/3$. One thought experiment is to move toward power utility by setting $\alpha = \rho = 1/3$. This has the advantage of keeping intertemporal preferences the same in the two models. The two sets of moving average coefficients are pictured in Figure 4. The impact of recursive preferences shows up entirely in the initial moving average coefficient, specifically the term $(\alpha - \rho)\gamma(b_1)$ in (23). Subsequent coefficients are identical.

A second thought experiment is to move toward power utility by setting $\rho = \alpha = -9$. The results are reported in column (1) of Table 2. Entropy is 0.0026, well below our estimated lower bound of 0.01, and time dependence is 0.0305, well above our estimated upper bound of 0.001. We can increase entropy by making $\alpha = \rho$ larger in absolute value or decrease time dependence by making them smaller, but we can't do both at once. Recursive preferences make progress along both fronts — see column (2). When we set $\rho = 1/3$, keeping $\alpha = -9$, we increase entropy by increasing $(\alpha - \rho)$ in the initial moving average coefficient. And with a smaller value of $1 - \rho$, we also reduce time dependence.

Time dependence, however, remains large: 0.0042 at a maturity of 60 months. This corresponds to an annualized mean yield spread of $-5\% (= -0.0042 \times 1200)$, which is well outside our bounds. Beeler and Campbell (2009, Section 7) note the same thing. There is some uncertainty about the slope of the real yield curve, but the magnitude here is well beyond existing estimates.

4.3 Entropy and time dependence with volatility dynamics

The most popular version of the Bansal-Yaron model has two persistent components: consumption growth and volatility. Both affect future utility and for that reason interact with recursive preferences to generate additional entropy in the pricing kernel.

The model is based on the bivariate consumption growth process

$$
\begin{align}
\log g_t &= \log g + \gamma(B)v_{t-1}^{1/2}w_{gt} \tag{24}\\
v_t &= v + \nu(B)w_{vt}, \tag{25}
\end{align}
$$

where w_{gt} and w_{vt} are $\mathcal{N}(0,1)$ series that are independent of each other and over time. The volatility process allows v_t to be negative in some states. We can keep it positive by using a square-root process instead, but with some loss in the transparency of the solution. See Appendix A.7.

14

We take the same approach to the solution as in the previous example. The pricing kernel is

$$\begin{aligned}
\log m_{t+1} &= \log\beta + (\rho-1)\log g - (\alpha-\rho)(\alpha/2)\gamma(b_1)^2\left[v + (\alpha/2)^2\gamma(b_1)^2 b_1^2\nu(b_1)^2\right] \\
&\quad + [(\rho-1)\gamma_0 + (\alpha-\rho)\gamma(b_1)]v_t^{1/2}w_{gt+1} + (\rho-1)[\gamma(B)/B]_+ v_{t-1}^{1/2}w_{gt} \\
&\quad + (\alpha-\rho)(\alpha/2)\gamma(b_1)^2[b_1\nu(b_1)w_{vt+1} - \nu(B)w_{vt}]. \tag{26}
\end{aligned}$$

See Appendix A.6. The new features concern volatility. Although innovations in consumption growth and volatility are independent, their roles interact in the pricing kernel. The coefficient of the volatility innovation w_{vt+1} depends on the dynamics of volatility [represented by $\nu(b_1)$], the dynamics of consumption growth [$\gamma(b_1)$], and recursive preferences [$(\alpha-\rho)$]. This interaction between the two components shows up in many of its properties.

Stochastic volatility gives the pricing kernel a different form from the Vasicek model. We might express it as

$$\log m_t = \log m + a_g(B)(v_{t-1}/v)^{1/2}w_{gt} + a_v(B)w_{vt}$$

with

$$a_{gj} = \begin{cases} [(\rho-1)\gamma_0 + (\alpha-\rho)\gamma(b_1)]v^{1/2} & j=0 \\ (\rho-1)\gamma_j v^{1/2} & j\geq 1 \end{cases}$$

$$a_{vj} = \begin{cases} (\alpha-\rho)(\alpha/2)\gamma(b_1)^2 b_1\nu(b_1) & j=0 \\ -(\alpha-\rho)(\alpha/2)\gamma(b_1)^2\nu_{j-1} & j\geq 1. \end{cases}$$

The first is the same as the previous example, but second is new. That gives us an additional term in entropy:

$$EL_t(m_{t+1}) = [(\rho-1)\gamma_0 + (\alpha-\rho)\gamma(b_1)]^2/2 + (\alpha-\rho)^2(\alpha/2)^2\gamma(b_1)^4[b_1\nu(b_1)]^2/2.$$

The second term is the contribution of stochastic volatility. Multiperiod entropy and time dependence are reported in Appendix A.8.

We see the results in Table 2. As in the previous example, the structure and parameter values are adapted from Bansal, Kiku, and Yaron (2009). The structure is the ARMA(1,1) process for $\log g_t$ ($\gamma_{j+1} = \varphi_g\gamma_j$ for $j \geq 1$) and an AR(1) process for v_t ($\nu_{j+1} = \varphi_v\nu_j$ for $j \geq 0$). The result, reported in column (3), is a substantial increase in entropy. Most of the increase comes from the volatility term, but its magnitude depends (as we've seen) on consumption dynamics. There is also an increase in (positive) time dependence.

It's not evident in these numbers, but the two components generate qualitatively different time dependence. In the calculations they're intertwined, but you can get a sense from the moving average coefficients pictured in Figure 6. The consumption coefficients in the top panel all have the same sign and therefore generate positive time dependence.

The volatility coefficients in the bottom panel switch signs and therefore generate negative time dependence. Note, too, that the volatility coefficients are zero in the additive case ($\alpha - \rho = 0$). Both components have small ratios: $\theta_g = a_{g1}/a_{g0} = 0.0018$ and $\theta_v = a_{v1}/a_{v0} = \nu_0/[b_1\nu(b_1)] = -0.0077$. The positive time dependence reported for this model indicates that the consumption component is more important to time dependence with these parameter values.

4.4 Discussion

The combination of recursive preferences and consumption dynamics is capable of generating a huge increase in entropy relative to additive power utility. In this respect the approach is a clear success. However, the persistence in consumption growth that underlies this success also generates strong positive time dependence with standard parameter values. Even with some uncertainty about the slope of the real yield curve, time dependence is too large.

Since recursive preferences rely on dynamics to affect the pricing kernel, the tension between entropy and time dependence is an inherent feature of the approach. However, the magnitudes are not. It's not hard to imagine alternative parameter values that could reduce time dependence without reducing entropy unduly. One approach is to reduce the size of the persistent component of consumption growth. Column (4) of Table 2 shows how this might work: with a smaller persistent component in consumption growth (smaller γ_1), time dependence falls below our estimated upper bound. The associated decline in entropy can be reversed, if desired, by increasing risk aversion $1 - \alpha$. See equation (26). Another approach is to place more weight on the volatility term, which (on its own) generates negative time dependence. Gallmeyer, Hollifield, Palomino, and Zin (2007) provide an example. Yet another approach is to specify direct interaction between consumption and volatility dynamics, as in Backus, Routledge, and Zin (2008, Section 3) and Hansen and Scheinkman (2009, Section 3.3). The model is sufficiently complex that it's hard to know in advance which of these alternatives will work best, but we think they show enough promise to merit further exploration.

5 Pricing kernels with habit formation

Our second class of models introduces dynamics into the pricing kernel directly through preferences. This mechanism has a long history, with applications ranging from microeconomic studies of consumption behavior (Deaton, 1993) to business cycles (Lettau and Uhlig, 2000, and Smets and Wouters, 2003). The asset pricing literature includes notable contributions by Abel (1992), Campbell and Cochrane (1999), Chan and Kogan (2002), Constantinides (1990), Heaton (1993, 1995), and Sundaresan (1989). We consider three examples here based on different functional forms.

All of our examples start with utility functions that include a state variable x_t, which we refer to as the "habit." A recursive formulation is

$$U_t = (1 - \beta)f(c_t, x_t) + \beta E_t U_{t+1}. \tag{27}$$

Typically x_t is tied to past consumption in some way. The examples we study have "external" habits: the agent ignores any impact of her consumption choices on future values of x_t. They differ in the functional form of $f(c_t, x_t)$ and in the law of motion for x_t.

5.1 Entropy and time dependence with a ratio habit

With so-called ratio habits, preferences follow (27) with $f(c_t, x_t) = (c_t/x_t)^\rho/\rho$ and $\rho \leq 1$. Examples include Abel (1992) and Chan and Kogan (2002). The pricing kernel is

$$m_{t+1} = \beta \left(c_{t+1}/c_t\right)^{\rho-1} \left(x_{t+1}/x_t\right)^{-\rho}. \tag{28}$$

As with recursive models, we add an extra term to the additive power utility pricing kernel.

Our first result follows from the habit being predetermined: x_{t+1} is known at date t. A direct consequence is that the habit has no impact on conditional entropy:

$$L_t\left(m_{t+1}\right) = L_t\left(g_{t+1}^{\rho-1}\right).$$

There is no entropy contribution beyond additive power utility.

We can be more specific if we impose additional structure on the processes for consumption growth and the habit. We use our reliable loglinear moving averages, the consumption growth process (16) and the habit

$$\log x_{t+1} = \log x + \chi(B) \log c_t. \tag{29}$$

Note the timing in the second equation: the habit at date $t+1$ depends on consumption at dates t and before. The log pricing kernel is then

$$
\begin{aligned}
\log m_{t+1} &= \log \beta + (\rho - 1)\log g_{t+1} - \rho\chi(B)\log g_t \\
&= \log \beta + (\rho - 1)\log g + [(\rho - 1) - \rho\chi(B)B]\gamma(B)v^{1/2}w_{t+1}.
\end{aligned}
$$

Its dynamics combine those of consumption growth $[\gamma(B)]$ and the habit $[\chi(B)]$. Chan and Kogan (2002) use an AR(1) habit, with $\chi_{j+1} = \varphi_x\chi_j$ for $j \geq 0$ and $0 \leq \varphi_x < 1$. Abel's (1990) one-period habit corresponds to $\varphi_x = 0$. They differ in one other respect. Chan and Kogan set $\chi_0 = 1 - \varphi_x$ so that the coefficients sum to one. Abel allows χ_0 to vary, allowing attenuation or exaggeration of the habit — or even durability if we allow $\chi_0 < 0$.

The impact of the habit lies in the pricing kernel's dynamics. We can see this most simply when consumption growth is iid: $\gamma(B) = \gamma_0 = 1$ (that is, $\gamma_j = 0$ for $j \geq 1$). Then all the dynamics come from the habit and the pricing kernel is

$$\log m_{t+1} = \log \beta + (\rho - 1)\log g + (\rho - 1)v^{1/2}w_{t+1} - \rho\chi(B)v^{1/2}w_t.$$

17

This corresponds to the Vasicek model with

$$a_j = \begin{cases} (\rho - 1)v^{1/2} & j = 0 \\ -\rho\chi_{j-1}v^{1/2} & j \geq 1. \end{cases}$$

Then $\theta = a_1/a_0 = \rho\chi_0/(\rho-1)$. Apparently $1 - \rho$ must be large to generate enough entropy, just as with additive power utility, and χ_0 must be small to keep time dependence modest.

We see the result in Figure 7. The initial term in the moving average is unchanged relative to additive power utility, but the others change from zero to negative. Entropy is, of course, unchanged, but time dependence is negative; see Table 3. We set $\varphi_x = 0.75$ here, which is close to Chan and Kogan's (2002, Table 1) choice converted to a monthly time interval, but it makes little difference to any of these properties. Time dependence is governed, in large part, by the choice of χ_0. If $\chi_0 = 1 - \varphi_x$, so that the χ_js sum to one, time dependence is -0.0025 at $n = 60$, corresponding to a mean yield spread of 3.1% per year at a maturity of 60 months. See column (2). Time dependence declines if we choose smaller values, and if we use $\chi_0 < 0$ ("durability") it changes sign. To summarize: a ratio habit has no impact on entropy, but it introduces a second source of time dependence beyond the dynamics in the consumption growth process.

5.2   Entropy and time dependence with a difference habit

A second functional form has significantly different properties. So-called difference habits are based on $f(c_t, x_t) = (c_t - x_t)^\rho/\rho$ and (again) a law of motion for the habit x_t that ties it to past consumption. Examples include Campbell and Cochrane (1999), Constantinides (1990), Heaton (1993, 1995), and Sundaresan (1989).

Campbell and Cochrane (1999) define the "surplus consumption ratio" $s_t = (c_t - x_t)/c_t = 1 - x_t/c_t$, which takes on values between zero and one. The pricing kernel becomes

$$m_{t+1} = \beta\left(\frac{c_{t+1} - x_{t+1}}{c_t - x_t}\right)^{\rho-1} = \beta g_{t+1}^{\rho-1}(s_{t+1}/s_t)^{\rho-1}.$$

As in our other examples, we gain an extra term relative to additive power utility.

The challenge lies in transforming this into something tractable. One approach is to use a loglinear approximation. Define $z_t = \log x_t - \log c_t$ so that $s_t = 1 - e^{z_t}$. If z_t is stationary with mean $z = \log x - \log c$, then a linear approximation of $\log s_t$ around z is

$$\log s_t = \text{constant} - [(1-s)/s]z_t = \text{constant} - [(1-s)/s]\log(x_t/c_t).$$

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

$$\log m_{t+1} = \log \beta + (\rho - 1)(1/s)\log g_{t+1} - (\rho - 1)[(1-s)/s]\log(x_{t+1}/x_t).$$

Campbell (1999, Section 5.1) has a similar analysis. This pricing kernel differs in a couple ways from the ratio habit case, equation (28). The first difference is the coefficient of log consumption growth, which now includes a magnification factor $(1/s)$. If the habit is zero, then $s = 1$ and there's no change. But if $s < 1$, the habit increases the sensitivity of marginal utility to changes in consumption. In this respect the habit works like an increase in $|\rho - 1|$. The second difference is the impact of the habit. The coefficient of $\log(x_{t+1}/x_t)$ changes from $-\rho$ in the ratio case to $(1 - \rho)(1 - s)/s$.

Using the same moving average representations for consumption growth (16) and the habit (29), the pricing kernel becomes

$$\log m_{t+1} = \log \beta + (\rho - 1) \log g + (\rho - 1)(1/s)[1 - (1 - s)\chi(B)B]\gamma(B)v^{1/2}w_{t+1}.$$

As with ratio habits, the pricing kernel reflects a combination of consumption and habit dynamics. We get a sense of the impact from a numerical example. Column (3) of Table 3 uses $s = 1/2$, but is otherwise the same as the ratio habit example reported in column (2), including iid consumption growth. Entropy rises from 0.0026 to 0.0104 through the magnification effect noted above. Time dependence also increases sharply. Looking at the two components, we see that entropy and time dependence are governed by different features of the habit: entropy is driven by the average habit (through $s = 1 - x/c$), while time dependence is governed by the volatility (the magnitude of the χ_js).

The loglinear approximation highlights a familiar tension between entropy and time dependence: to generate enough of the former we end up with too much of the latter. To resolve this tension, Campbell and Cochrane (1999) suggest the nonlinear surplus process

$$\log s_{t+1} - \log s_t = (\varphi_s - 1)(\log s_t - \log s) + \lambda(\log s_t)v^{1/2}w_{t+1}$$

$$1 + \lambda(\log s_t) = v^{-1/2}\left(\frac{(1 - \rho)(1 - \varphi_s) - b}{(1 - \rho)^2}\right)^{1/2}(1 - 2[\log s_t - \log s])^{1/2}.$$

The pricing kernel is

$$\begin{aligned}
\log m_{t+1} &= \log \beta + (\rho - 1)\log g_{t+1} + (\rho - 1)\log(s_{t+1}/s_t) \\
&= \log \beta - (\rho - 1)\log g + (\rho - 1)(\varphi_s - 1)(\log s_t - \log s) \\
&\quad + (\rho - 1)\left[1 + \lambda(\log s_t)\right]v^{1/2}w_{t+1}.
\end{aligned}$$

As in the previous model, dynamics in the pricing kernel follow from those of the habit, in this case through the surplus. The difference here is that the conditional variance also depends on the habit.

The properties of the pricing kernel follow immediately. Conditional entropy is

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

19

Entropy is therefore

$$EL_t(m_{t+1}) \;\;=\;\; [(1-\rho)(1-\varphi_s) - b/2].$$

Campbell and Cochrane (1999) use $b = 0$, so entropy is governed by the curvature parameter ρ and the autoregressive parameter φ_s of the surplus. With their suggested values of $\rho = -1$ and $\varphi_s = 0.9885 = 0.87^{1/12}$, entropy is 0.0231, far more than we get with additive power utility and $\rho = -9$. See columns (1) and (4) of Table 3.

5.3 Discussion

The impact of habits depends, clearly, on their form. With ratio habits, the sole impact is on time dependence: there is no additional contribution to entropy. With difference habits, there is an impact on both entropy and time dependence. In both cases, time dependence is larger in absolute value than the bound we estimated earlier.

The Campbell-Cochrane model reduces time dependence by giving the state variable offsetting effects on the conditional mean and variance of the pricing kernel. In its original form, time dependence is zero by construction. In later applications by Verdelhan (2010) and Wachter (2006), the model is parameterized to generate realistic bond yields, hence time dependence. The interaction between the mean and variance is a useful device that we think is worth exploring in other models, including those with recursive preferences.

6 Pricing kernels with jumps and disasters

The examples so far are based on normal innovations, yet the evidence is overwhelming that monthly log returns on common assets are not normal. See Table 1, for example. If that's true of the high-return asset, then the log pricing kernel must be nonnormal as well. We therefore consider a third class of models, in which departures from normality contribute to entropy and time dependence.

The standard mechanism in finance for generating nonnormal innovations is to introduce a "jump" component. The word jump is used here to mean an innovation with a distribution other than normal, but the connection to jump processes should be clear shortly. Prominent examples of jumps in the macro-finance literature include Barro (2006), Barro, Nakamura, Steinsson, and Ursua (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 (2010), Garcia, Luger, and Renault (2003), Longstaff and Piazzesi (2004), and Wachter (2008).

We illustrate the potential impact of jumps on entropy and time dependence with a series of examples. We start by adding jumps to iid environments with additive power utility, then go on to explore the interaction of persistent jump risk and recursive preferences.

6.1 Entropy and time dependence with iid jump risk

Our starting point is a jump component in log consumption growth. Let log consumption growth be iid, the sum of a normal component w_{gt} and a jump component z_{gt}:

$$\log g_t \quad = \quad w_{gt} + z_{gt}. \tag{30}$$

The normal component has arbitrary mean and variance: $w_{gt} \sim \mathcal{N}(\mu, v)$. The jump component is a Poisson mixture of normals, a specification that has been widely used in the option pricing literature. Its central ingredient is a Poisson random variable j (the number of jumps) that takes on nonnegative integer values with probabilities $p(j) = e^{-h}h^j/j!$. The parameter $h \geq 0$ (the "jump intensity") 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 j, the component is normal:

$$z_{gt}|j \quad \sim \quad \mathcal{N}(j\theta, j\delta^2) \text{ for } j = 0, 1, 2, \ldots. \tag{31}$$

If h is small, the jump component is well approximated by a Bernoulli mixture of normals: there is at most one jump per period and it occurs with probability h. If h is large there can be a significant probability of multiple jumps.

With additive power utility, the pricing kernel is

$$\log m_t \quad = \quad \log \beta + (\rho - 1)(w_{gt} + z_t).$$

Its entropy is

$$L_t(m_{t+1}) \quad = \quad (\rho - 1)^2 v/2 + \left\{ [e^{(\rho-1)\theta + (\rho-1)^2\delta^2/2} - 1] - (\rho - 1)\theta \right\} h.$$

The first term represents the normal component, the second the jump component. Time dependence is, of course, zero, because the pricing kernel is iid.

Even with additive power utility, the impact of jumps rises sharply with $1 - \rho$. The expansion (3) shows how high-order cumulants contribute to entropy. With power utility, the cumulants of $\log m_{t+1}$ are closely tied to those of log consumption growth:

$$\kappa_j(\log m_{t+1}) \quad = \quad (\rho - 1)^j \kappa_j(\log g_{t+1}).$$

In words, the jth cumulant of the pricing kernel is the same cumulant of log consumption growth multiplied by $(\rho - 1)^j$, which can be large for $\rho < 0$ and $j > 1$. As a result, a small jump component in consumption growth can make a large contribution to the pricing kernel.

We report numerical examples in Table 4. In each one, we hold the mean and variance of log consumption growth constant and equal to the same values we've used in other examples. With this structure, they are

$$E(\log g_t) \quad = \quad \mu + h\theta$$
$$Var(\log g_t) \quad = \quad v + h(\theta^2 + \delta^2).$$

In column (1), there is no jump component: we're simply reproducing column (1) of Table 3 with lower risk aversion. We introduce a jump component in column (2). The parameter values are taken from Backus, Chernov, and Martin (2011, Section III) and are similar to those suggested by Barro, Nakamura, Steinsson, and Ursua (2009). We see that entropy rises sharply. The point is a general one: increasing the probability of adverse outcomes relative to the normal distribution increases entropy, even when we hold constant the mean and variance of log consumption growth.

6.2 Entropy and time dependence with persistent jump risk

We can get further magnification of entropy if we make jumps persistent and apply recursive preferences. The following example is similar to Wachter's (2008).

There are a number of ways to introduce dynamics into this model. We use one here: we make the jump intensity h_t stochastic. Let consumption growth be

$$
\begin{aligned}
\log g_t &= w_{gt} + z_{gt} \\
h_t &= h + \eta(B)w_{ht},
\end{aligned}
$$

where (w_{gt}, z_{gt}) are shocks of the same form as the previous section and w_{ht} is an independent iid $\mathcal{N}(0,1)$ series. The only change is in the dynamics of jump risk h_t, which are controlled by $\eta(B)$. If $\eta(B) = 0$, h_t is constant and we are back to the model of the previous subsection. Otherwise $\eta(B)$ governs the persistence of jump risk.

With this process and recursive preferences, the pricing kernel is

$$
\begin{aligned}
\log m_{t+1} &= \log \beta + (\rho - 1)\log g - (\alpha - \rho)(\alpha/2)(v + \eta_0^2) - (\alpha - \rho)[(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha]h \\
&\quad + (\alpha - 1)(v^{1/2}w_{gt+1} + z_{gt+1}) \\
&\quad + (\alpha - \rho)[(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha][b_1\eta(b_1)w_{ht+1} - \eta(B)w_{ht}].
\end{aligned}
$$

See Appendix A.6. Note the similarity of the last term to that of equation (26). Persistent jump risk and stochastic volatility are different mechanisms, but they have a similar impact on the pricing kernel with recursive preferences. Neither appears in the pricing kernel with additive power utility ($\alpha - \rho = 0$).

From the pricing kernel we see that conditional entropy is

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

Multiperiod entropy and time dependence are reported in Appendix A.8. For entropy (mean conditional entropy), we simply replace h_t with its mean h. Relative to the previous

example, we have one additional term — the last one — which reflects a combination of recursive preferences $[(\alpha - \rho)]$ and dynamics in the intensity of jump risk $[\eta(b_1)]$.

We report the features of a numerical example in column (3) of Table 4. Here the jump intensity is AR(1): $\eta_{j+1} = \varphi_h \eta_j$ for $j \geq 0$. We adapt parameter values for (φ_h, η_0) from Wachter (2008); see Appendix A.9. The results are striking. Entropy rises above 0.5 even with risk aversion of $1 - \alpha = 5$. We think these jump parameters are somewhat extreme, but they make the point that jumps and disasters can increase entropy enormously. Moreover, they do it without adding much to time dependence. As we saw with stochastic volatility, and for the same reason, there is negative time dependence, but the magnitude is close to zero.

6.3 Discussion

Jumps and disasters can generate enormous amounts of entropy without generating much in the way of time dependence. It's not hard to imagine that jumps could play a useful role, then, in accounting not only for entropy and time dependence, but for the skewness and excess kurtosis evident in asset returns.

These examples barely scratch the surface, as our earlier list of related work makes clear. In the class of models with recursive preferences, we can add dynamics to both components of consumption growth, stochastic volatility, and jump intensity, and jumps can be added to volatility as well as consumption growth. Jumps can also be added to models with habits, as Bekaert and Engstrom (2010) show. With such a wide range of possibilities, it will take some time to sort out which ones are most useful.

7 Final thoughts

We've 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 must be at least as large as the largest mean excess return, which is likely well above the equity premium. The second property is time dependence, a measure of the pricing kernel's dynamics. Time 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: to generate enough entropy but not too much time dependence. Representative agent models with recursive preferences and habits use dynamics to increase entropy, but as a result they often increase time dependence as well. Figure 8 is a summary of how a number of representative agent models do along these two dimensions. The top panel reports entropy, which should be above the estimated lower bound marked by the dotted line. The bottom panel reports time dependence, which

23

should lie between the bounds also noted by dotted lines. All of these numbers depend on parameter values and are therefore subject to change, but they provide targets for the future evolution of these models. We've suggested a number of directions this evolution might take.

A Appendix

A.1 Premiums in forward rates and bond returns

We connect term premiums reflected in bond yields, returns, and forward rates. We start with the latter; the others follow from their definitions.

Let q_t^n be the price at date t of an n-period zero-coupon bond. Yields y and forward rates f are defined from prices by

$$-\log q_t^n \;=\; ny_t^n \;=\; f_t^0 + f_t^1 + \cdots + f_t^{n-1}.$$

The definition of forward rates implies $f_t^n = \log(q_t^n/q_t^{n+1})$. One-period returns are $r_{t+1}^n = q_{t+1}^{n-1}/q_t^n$. The *short rate* is $\log r_{t+1}^1 = y_t^1 = f_t^0$.

We define (forward) *term premiums tp* by

$$f_t^n \;=\; E_t f_{t+n}^0 + tp_t^n.$$

(Aficionados will see the expectations hypothesis lurking here.) We define *bond premiums* from (log) excess returns:

$$bp_t^n \;=\; E_t(\log r_{t+1}^n - \log r_{t+1}^1).$$

The question is how the two are related.

The answer follows from the definitions. The excess return on an $(n+1)$-period bond is

$$\log r_{t+1}^n - \log r_{t+1}^1 \;=\; (f_t^1 - f_{t+1}^0) + ... + (f_t^n - f_{t+1}^{n-1}).$$

The typical term is

$$f_t^j - f_{t+1}^{j-1} \;=\; (E_t f_{t+j}^0 - E_{t+1} f_{t+j}^0) + tp_t^j - tp_{t+1}^{j-1}.$$

That is, you get innovations in forward rates and changes in term premiums. Since the former have (conditional) mean zero by construction, only the latter show up in bond risk premiums:

$$
\begin{aligned}
bp_t^{n+1} \;=\;& E_t(tp_{t+1}^1 - tp_t^0) + E_t(tp_{t+1}^2 - tp_t^1) + ... + E_t(tp_{t+1}^n - tp_t^{n-1}) \\
\;=\;& E_t(tp_{t+1}^1 - tp_t^1) + E_t(tp_{t+1}^2 - tp_t^2) + ... + E_t(tp_{t+1}^{n-1} - tp_t^{n-1}) + tp_t^n.
\end{aligned}
$$

In short, bond premiums are functions of term premiums. In stationary ergodic settings, the unconditional means have a more simple form: $E(bp^{n+1}) = E(tp^n)$.

Yield premiums also follow from term premiums. Note that yields are averages of forward rates: $y_t^n = n^{-1} \sum_{j=1}^n f_t^{j-1}$. The yield spread is therefore

$$y_t^n - y_t^1 \;=\; n^{-1} \sum_{j=0}^{n-1} \left[E_t(f_{t+j}^0 - f_t^0) + tp_t^j \right].$$

A natural definition of the yield premium is the last term,

$$yp_t^n = n^{-1} \sum_{j=0}^{n-1} tp_t^j,$$

a simple average of term premiums. In stationary settings, the mean has the same form: $E(yp^n) = n^{-1} \sum_{j=0}^{n-1} E(tp^j)$.

Many models have the property that long terms premiums are constant:

$$\lim_{n \to \infty} tp_t^n = tp^\infty.$$

If the limit exists, bond premiums and yield premiums converge to the same value.

A.2 Entropy and Hansen-Jagannathan bounds

The entropy and Hansen-Jagannathan 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+1} maximizes the Sharpe ratio: given a pricing kernel m_{t+1}, its excess return $x_{t+1} = r_{t+1} - r_{t+1}^1$ maximizes $SR_t = E_t(x_{t+1})/Var_t(x_{t+1})^{1/2}$ subject to the pricing relation (1). The maximization leads to the bound,

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

and the return that hits the bound,

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

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

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

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

which connects it directly to the pricing kernel. The first term is a constant: the mean excess return. The second shows that the return is a decreasing linear function of the pricing kernel.

Bryan Routledge suggested a similar approach to the entropy bound. The bound defines a high-return asset as one whose return r_{t+1} maximizes $E_t(\log r_{t+1} - \log r_{t+1}^1)$ subject to the pricing relation (1). The maximization leads to the return

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

Its mean log excess return $E_t(\log r_{t+1} - \log r_{t+1}^1)$ hits the entropy bound (6).

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

Entropy and maximum Sharpe ratios over different time horizons. 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 was suggested by Ian Martin and is summarized in Backus, Chernov, and Martin (2011, Appendix A.2). Suppose $\log m_{t+1}$ has conditional cumulant-generating function and associated power series expansion

$$k_t(s) = \log E_t\left(e^{s \log m_{t+1}}\right) = \sum_{j=1}^{\infty} \kappa_{jt} s^j / j!.$$

Entropy is then

$$L_t(m_{t+1}) = k_t(1) - \kappa_{1t} = \sum_{j=2}^{\infty} \kappa_{jt}/j!.$$

In the lognormal case, only the variance is nonzero: $L_t(m_{t+1}) = \kappa_{2t}/2$. The maximum Sharpe ratio follows from the mean and variance of m_{t+1}, which depends on the cumulants of $\log m_{t+1}$ of all order:

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

The maximum squared Sharpe ratio is therefore

$$Var_t(m_{t+1})/E_t(m_{t+1})^2 = e^{k_t(2) - 2k_t(1)} - 1.$$

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 and $k_t(2) - 2k_t(1) = \kappa_{2t}$.

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

27

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:

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

This is just the zero time 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:

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

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.

Excess returns and Sharpe ratios in lognormal settings. It's common to report the mean, standard deviation, and Sharpe ratio for data and models. We show how this works if returns are lognormal.

Suppose asset j's return is conditionally lognormal: $\log r^j_{t+1}|t \sim \mathcal{N}(\log r^1_{t+1} + \kappa^j_{1t}, \kappa^j_{2t})$. Our analysis focuses on the mean *log* excess return:

$$E_t(\log r^j_{t+1} - \log r^1_{t+1}) = \kappa^j_{1t}.$$

That's it. Most asset pricing research focuses on the simple excess return: $x_{t+1} = r^j_{t+1} - r^1_{t+1}$. Here we see that the Sharpe ratio reflects both moments of the log return. The mean and variance of the excess return are

$$E_t(x_{t+1}) = r^1_{t+1}\left(e^{\kappa^j_{1t}+\kappa^j_{2t}/2} - 1\right)$$

$$Var_t(x_{t+1}) = \left(r^1_{t+1}e^{\kappa^j_{1t}+\kappa^j_{2t}/2}\right)^2\left(e^{\kappa^j_{2t}} - 1\right).$$

The conditional Sharpe ratio is therefore

$$SR_t = \frac{E_t(x_{t+1})}{Var_t(x_{t+1})^{1/2}} = \frac{e^{\kappa^j_{1t}+\kappa^j_{2t}/2} - 1}{e^{\kappa^j_{1t}+\kappa^j_{2t}/2}\left(e^{\kappa^j_{2t}} - 1\right)^{1/2}}.$$

Evidently there are two ways to generate a large Sharpe ratio. The first is to have a large mean log return: 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. If we go beyond lognormal, other cumulants enter, too.

Comparisons of Sharpe ratios thus reflect both the mean and variance of the log return — and possibly higher-order cumulants as well. Binsbergen, Brandt, and Koijen (2010) 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, an increase in the variance.

A.3   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 building block is the infinite one-sided 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 result is a stationary process if $\sum_j a_j^2 < \infty$ (we say the sequence of a_js is square summable).

In this form, prediction is simple. If the information set at date t includes current and past values of w, forecasts of future x 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 $\varphi(B) = 1 - \varphi_1 B$ and $\theta(B) = 1 - \theta_1 B$. Special cases include the AR(1) (set $\theta_1 = 0$) and the MA(1) (set $\varphi_1 = 0$). The infinite moving average representation is $x_t = [\varphi(B)/\theta(B)] w_t = a(B) w_t$, whose coefficients include $a_0 = 1$, $a_1 = \varphi - \theta$, and $a_j = (\varphi - \theta)\varphi^{j-1}$. We typically choose φ and a_1, leaving θ implicit.

A.4   Bond prices, yields, and returns in the Vasicek model

We report bond prices, entropy, and related objects for the Vasicek model of Section 3. If the pricing kernel is (11), then (6) implies

$$\log q_t^n = n \log m + \sum_{j=1}^{n} A_{j-1}^2/2 + \sum_{j=0}^{\infty}(A_{n+j} - A_j)w_{t-j},$$

where $A_n = \sum_{j=0}^{n} a_j$. See Backus and Zin (1994, Section 3). Forward rates, yields, and returns follow from prices (see Appendix A.1):

$$f_t^n = -\log m - A_n^2/2 - [a(B)/B^n]_+ w_t$$

$$y_t^n = -\log m - n^{-1}\sum_{j=1}^{n} A_{j-1}^2/2 - n^{-1}\sum_{j=0}^{\infty}(A_{n+j} - A_j)w_{t-j}$$

$$\log r_{t+1}^n = -\log m - A_{n-1}^2/2 + (A_n - A_0)w_{t+1} - [a(B)/B]_+ w_t.$$

The short rate $r_{t+1}^1 = 1/q_t^1$ is

$$\log r_{t+1}^1 = y_t^1 = f_t^0 = -(\log m + A_0^2/2) - [a(B)/B]_+ w_t.$$

Mean excess returns and yield spreads are

$$E(\log r_{t+1}^n - \log r_{t+1}^1) = \left(A_0^2 - A_{n-1}^2\right)/2$$

$$E(y_t^n - y_t^1) = n^{-1}\sum_{j=1}^{n}(A_0^2 - A_{j-1}^2)/2.$$

Multiperiod entropy is

$$L_t(m_{t,t+n}) = \log E_t m_{t,t+n} - E_t \log m_{t,t+n} = \sum_{j=1}^{n} A_{j-1}^2/2.$$

Time dependence is therefore

$$T(n) = n^{-1}EL_t(m_{t,t+n}) - EL_t(m_{t+1}) = n^{-1}\sum_{j=1}^{n}(A_{j-1}^2 - A_0^2)/2.$$

The (unconditional) expectation is unnecessary in this case: conditional entropy is constant.

A.5   The marginal rate of substitution with recursive preferences

Consider a stationary markovian environment based on a state variable s_t at date t. Denote transition probabilities by $p(s_{t+1}|s_t)$. In this setting, the pricing kernel in a representative

agent model is the marginal rate of substitution between (say) consumption in state s_t [$c(s_t)$, or c_t for short] and consumption in state s_{t+1} [$c(s_{t+1})$, or c_{t+1} for short].

Here's how that works with the recursive preferences described in Section 4. With this notation and a finite set of states, the certainty equivalent (18) might be expressed less compactly as

$$\mu_t(U_{t+1}) = \left[\sum_{s_{t+1}} p(s_{t+1}|s_t)U(s_{t+1})^\alpha \right]^{1/\alpha} ,$$

where $U(s_{t+1})$ is continuation utility. With this in hand, we differentiate (17) and (18):

$$\partial U_t/\partial c_t = U_t^{1-\rho}(1-\beta)c_t^{\rho-1}$$
$$\partial U_t/\partial \mu_t(U_{t+1}) = U_t^{1-\rho}\beta\mu_t(U_{t+1})^{\rho-1}$$
$$\partial \mu_t(U_{t+1})/\partial U(s_{t+1}) = \mu_t(U_{t+1})^{1-\alpha}p(s_{t+1}|s_t)U(s_{t+1})^{\alpha-1}.$$

The marginal rate of substitution between consumption at date t and consumption in state s_{t+1} is therefore

$$\frac{\partial U_t/\partial c(s_{t+1})}{\partial U_t/\partial c_t} = \frac{[\partial U_t/\partial \mu_t(U_{t+1})][\partial \mu_t(U_{t+1})/\partial U(s_{t+1})][\partial U(s_{t+1})/\partial c(s_{t+1})]}{\partial U_t/\partial c_t}$$
$$= p(s_{t+1}|s_t)\, \beta \left(\frac{c(s_{t+1})}{c_t} \right)^{\rho-1} \left(\frac{U(s_{t+1})}{\mu_t(U_{t+1})} \right)^{\alpha-\rho}.$$

Equation (22) is the same with the probability left out and the state left implicit.

A.6 The pricing kernel with recursive preferences

We derive the pricing kernel for a representative agent model with recursive preferences, consumption growth dynamics, stochastic volatility, and jumps with time-varying intensity. The models in Sections 4 and 6 are special cases.

We posit a consumption growth process with

$$\log g_t = \log g + \gamma(B)v_{t-1}^{1/2}w_{gt} + z_{gt}$$
$$v_t = v + \nu(B)w_{vt}$$
$$h_t = h + \eta(B)w_{ht}.$$

Here $\{w_{gt}, w_{vt}, w_{ht}\}$ are $\mathcal{N}(0,1)$ and independent of each other and over time. z_{gt} is a jump component. Conditional on the number of jumps j, z_{gt} is normal: $z_{gt}|j \sim \mathcal{N}(j\theta, j\delta^2)$. The probability of $j \geq 0$ jumps at date t is $e^{-h_{t-1}}h_{t-1}^j/j!$. This covers most of the models in the literature, but we can generalize further by (for example) adding dynamics to the jump component, jumps to volatility and intensity, or interdependence among the three processes.

Given a value of b_1, we solve the model by using the "Bellman equation" (20) to characterize the value function and substituting the result into the marginal rate of substitution cum pricing kernel (22). 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, Section XI.19).

We use two results about certainty equivalents, evaluations of the power certainty equivalent function (18). The first one applies to normal risks. If $w_{t+1}|t \sim \mathcal{N}(\kappa_{1t}, \kappa_{2t})$, then the log of the certainty equivalent of $e^{aw_{t+1}}$ is $\log \mu_t(e^{aw_{t+1}}) = \kappa_{1t} + (\alpha/2)a^2\kappa_{2t}$. The second applies to Poisson-normal mixtures like the jump component z_{gt+1}. The analogous certainty equivalent result is $\log \mu_t(e^{az_{t+1}}) = h_t(e^{a\alpha\theta + (a\alpha\delta)^2/2} - 1)/\alpha$.

We find the value function by guess and verify:

- Guess. We guess a value function of the form

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

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

- Compute. We need the certainty equivalent $\mu_t(g_{t+1}u_{t+1})$ for (20). Given our guess, $\log(g_{t+1}u_{t+1})$ is

$$
\begin{aligned}
\log(g_{t+1}u_{t+1}) &= \log g + \log u + [\gamma(B) + p_g(B)]v_t^{1/2}w_{gt+1} + z_{gt+1} \\
&\quad + p_v(B)w_{vt+1} + p_h(B)w_{ht+1} \\
&= \log(gu) + [\gamma(B) + p_g(B) - (\gamma_0 + p_{g0})]v_t^{1/2}w_{gt+1} \\
&\quad + [p_v(B) - p_{v0}]w_{vt+1} + [p_h(B) - p_{h0}]w_{ht+1} \\
&\quad + (\gamma_0 + p_{g0})v_t^{1/2}w_{gt+1} + z_{gt+1} + p_{v0}w_{vt+1} + p_{h0}w_{ht+1}.
\end{aligned}
$$

Clever subtraction here of the initial coefficients in the lag polynomials eliminates the $t+1$ terms and leaves the resulting expressions in a convenient form. The last line consists of innovations, the others of (conditional) constants. The certainty equivalent treats them differently:

$$
\begin{aligned}
\log \mu_t(g_{t+1}u_{t+1}) &= \log(gu) + [\gamma(B) + p_g(B) - (\gamma_0 + p_{g0})]v_t^{1/2}w_{gt+1} \\
&\quad + [p_v(B) - p_{v0}]w_{vt+1} + [p_h(B) - p_{h0}]w_{ht+1} \\
&\quad + (\alpha/2)(\gamma_0 + p_{g0})^2 v_t + (\alpha/2)(p_{v0}^2 + p_{h0}^2) + [(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha]h_t \\
&= \log(gu) + [\gamma(B) + p_g(B) - (\gamma_0 + p_{g0})]v_t^{1/2}w_{gt+1} \\
&\quad + [p_v(B) - p_{v0}]w_{vt+1} + [p_h(B) - p_{h0}]w_{ht+1} \\
&\quad + (\alpha/2)(\gamma_0 + p_{g0})^2[v + \nu(B)w_{vt}] + (\alpha/2)(p_{v0}^2 + p_{h0}^2) \\
&\quad + [(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha][h + \eta(B)w_{ht}].
\end{aligned}
$$

- Verify. We substitute the certainty equivalent into (20) and solve for the parameters. Matching like terms, we have

$$\text{constant}: \quad \log u = b_0 + b_1[\log(gu) + (\alpha/2)(p_{v0}^2 + p_{h0}^2)]$$

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

$$w_{vt}: \quad p_v(B) = b_1 \left[\frac{p_v(B) - p_{v0}}{B} + (\alpha/2)(\gamma_0 + p_{g0})^2 \nu(B)\right]$$

$$w_{ht}: \quad p_h(B) = b_1 \left[\frac{p_h(B) - p_{h0}}{B} + [(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha]\eta(B)\right].$$

We skipped the constants v and h in the first equation, because we're going to include them in v_t and h_t below.

We solve the equations one at a time. 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 $L = b_1$ to get $\gamma_0 + p_{g0} = \gamma(b_1)$. The other coefficients of $p_g(B)$ are of no concern to us: they don't show up in the pricing kernel. The third equation has a similar structure. Setting $B = b_1$ gives us $p_{v0} = (\alpha/2)\gamma(b_1)^2 b_1 \nu(b_1)$. In the last equation, setting $B = b_1$ gives us $p_{h0} = [(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha]b_1\eta(b_1)$. Finally, the first equation implies

$$\log u = (1 - b_1)^{-1}\left\{b_0 + b_1\left[\log g + (\alpha/2)(p_{v0}^2 + p_{h0}^2)\right]\right\}.$$

We can substitute for p_{v0} and p_{h0} with some loss of simplicity.

Now that we know the value function, we construct the pricing kernel from (22). One component is

$$\log(g_{t+1}u_{t+1}) - \log\mu_t(g_{t+1}u_{t+1}) = \gamma(b_1)v_t^{1/2}w_{gt+1} + z_{gt+1} + p_{v0}w_{vt+1} + p_{h0}w_{ht+1}$$
$$- (\alpha/2)[\gamma(b_1)^2 v_t + (p_{v0}^2 + p_{h0}^2)] - [(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha]h_t,$$

a combination of innovations to future utility (the first line) and adjustments for risk (the second). The pricing kernel then becomes

$$\log m_{t+1} = \log\beta + (\rho - 1)\log g - (\alpha - \rho)(\alpha/2)(p_{v0}^2 + p_{h0}^2) + (\alpha - 1)z_{gt+1}$$
$$+ [(\rho - 1)\gamma_0 + (\alpha - \rho)\gamma(b_1)]v_t^{1/2}w_{gt+1} + (\rho - 1)[\gamma(B)/B]_+ v_{t-1}^{1/2}w_{gt}$$
$$+ (\alpha - \rho)[p_{v0}w_{vt+1} - (\alpha/2)\gamma(b_1)^2 v_t]$$
$$+ (\alpha - \rho)\{p_{h0}w_{ht+1} - [(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha]h_t\}.$$

The examples in Sections 4 and 6 follow from setting various combinations of $\gamma(B)$, $\nu(B)$, and $\eta(B)$ equal to zero and rearranging terms.

A.7 Recursive models based on square root processes

We like the simplicity and transparency of linear processes; expressions like $\nu(b_1)$ 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. Here we describe and solve an analogous model based on square-root processes, discrete-time analogs of well-known continuous-time processes. The analysis parallels Appendix A.6.

Consider the consumption process

$$
\begin{aligned}
\log g_t &= \log g + \gamma(B)v_{t-1}^{1/2}w_{gt} + z_{gt} \\
v_t &= (1 - \varphi_v)v + \varphi_v v_{t-1} + \sigma_v v_{t-1}^{1/2}w_{vt} \\
h_t &= (1 - \varphi_h)h + \varphi_h h_{t-1} + \sigma_h h_{t-1}^{1/2}w_{ht}.
\end{aligned}
$$

The innovations $(w_{gt}, z_{gt}, w_{vt}, w_{ht})$ are described in Appendix A.6. With the exception of the processes for v_t and h_t, the model is the same one we saw earlier.

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

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

  its certainty equivalent is

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

34

- Verify. We substitute the certainty equivalent into (20) and collect similar terms:

$$\text{constant}: \quad \log u \;=\; b_0 + b_1[\log(gu) + p_v(1-\varphi_v)v + p_h(1-\varphi_h)h]$$

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

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

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

The second equation is the same one we saw in Appendix A.6 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've 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.

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

$$b_1(\alpha/2)(p_v\sigma_v)^2 - (1 - b_1\varphi_v)p_v + b_1(\alpha/2)\gamma(b_1)^2 \;=\; 0.$$

It has two real roots if

$$(1 - b_1\varphi_v)^2 - (b_1\alpha\sigma_v)^2\gamma(b_1)^2 \;\geq\; 0.$$

(If not, there's no solution of the specified form.) This requires that α and σ_v not be too large in absolute value. Descartes' rule of signs tells us that both roots have the same sign as α. The quadratic formula gives us their values:

$$p_v \;=\; \frac{(1 - b_1\varphi_v) \pm \left[(1 - b_1\varphi_v)^2 - (b_1\alpha\sigma_v)^2\gamma(b_1)^2\right]^{1/2}}{b_1\alpha\sigma_v^2}.$$

Now consider the stability of our value function iterations,

$$p_v \;\leftarrow\; b_1[p_v\varphi_v + (\alpha/2)\gamma(b_1)^2 + (\alpha/2)(p_v\sigma_v)^2].$$

The slope of the right-hand side at p_v is

$$b_1\varphi_v + b_1\alpha\sigma_v^2 p_v.$$

This is positive since $\alpha p_v > 0$ and less than one (locally stable) if

$$p_v \;<\; (1 - b_1\varphi_v)/(b_1\alpha\sigma_v^2).$$

The smaller root is therefore the only one that is locally stable. Hansen and Scheinkman (2009, Appendix A) make the same argument.

Similar logic applies to p_h. The two roots are

$$p_h = \frac{(1 - b_1\varphi_h) \pm \left[(1 - b_1\varphi_h)^2 - 2(b_1\sigma_h)^2(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)\right]^{1/2}}{b_1\alpha\sigma_h^2}.$$

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

Finally, the first equation implies

$$\log u = (1 - b_1)^{-1}\{b_0 + b_1[\log g + p_v(1 - \varphi_v)v + p_h(1 - \varphi_h)h], \}.$$

where p_v and p_h are understood to be functions of primitive parameters.

Given these value function coefficients, the pricing kernel is

$$
\begin{aligned}
\log m_{t+1} = {} & \log\beta + (\rho - 1)\log g + (\alpha - 1)z_{gt+1} \\
& + [(\rho - 1)\gamma_0 + (\alpha - \rho)\gamma(b_1)]v_t^{1/2}w_{gt+1} + (\rho - 1)[\gamma(B)/B]_+ v_{t-1}^{1/2}w_{gt} \\
& + (\alpha - \rho)\left\{p_v\sigma_v v_t^{1/2}w_{vt+1} - (\alpha/2)[\gamma(b_1)^2 + (p_v\sigma_v)^2]^2 v_t\right\} \\
& + (\alpha - \rho)\left\{p_h\sigma_h h_t^{1/2}w_{ht+1} - [(e^{\alpha\theta + (\alpha\delta)^2/2} - 1)/\alpha + (\alpha/2)(p_h\sigma_h)^2]h_t\right\}.
\end{aligned}
$$

There's a marked resemblance to the pricing kernel in Appendix A.6, but they're not, of course, identical.

A.8 Time dependence with recursive models

We derive time dependence for the model described in Appendix A.6. The results are used on Sections 4 and 6. The pricing kernel has the form

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

with $\{w_{gt}, w_{vt}, z_{gt}, w_{ht}\}$ defined above. 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 $x = g, v, h$ and $H_n = \sum_{j=0}^{n} \eta_{xj}$,

We derive multiperiod entropy from bond prices as described in (8). Suppose bond prices have the form

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

36

(If we had dynamics in z_{gt}, that would be here, too. Be thankful for small mercies.) Then the pricing relation (6) tells us

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

The log of the term in brackets is

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

Evaluating the expectation and lining up terms gives us

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

The second equation mirrors the Vasicek model:

$$
\gamma_{gj}^n = \sum_{i=1}^{n} a_{gj+i} = A_{gn+j} - A_{gj}.
$$

The third equation implies

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

The fourth equation implies

$$
\begin{aligned}
\gamma_{hj}^n &= \sum_{i=1}^{n} [a_{hj+i} + (e^{a_z\theta + (a_z\delta)^2/2} - 1)\eta_{j+i-1}] \\
&= A_{hn+j} - A_{hj} + (e^{a_z\theta + (a_z\delta)^2/2} - 1) \cdot (H_{n+j-1} - H_{j-1}).
\end{aligned}
$$

And at long last, the first equation implies

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

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

$$
\begin{aligned}
T(n) = \;& (2n)^{-1} \sum_{j=1}^{n} (A_{gj-1}^2 - A_{g0}^2) + (2n)^{-1} \sum_{j=2}^{n} \left[\left(A_{vj-1} + (2v)^{-1} \sum_{i=0}^{j-2} \nu_i A_{gi}^2\right)^2 - A_{v0}^2\right] \\
& + (2n)^{-1} \sum_{j=2}^{n} \left[(A_{hj-1} + (e^{a_z\theta + a_z^2\delta^2/2} - 1)H_{j-2})^2 - A_{h0}^2\right].
\end{aligned}
$$

The first term is a Vasicek-like expression for w_{gt}. The second reflects the interaction between w_{gt} and volatility. The last term is the interaction between the jump component z_{gt} and the intensity process.

A.9 Parameter values for recursive models

Bansal-Yaron models. Bansal, Kiku, and Yaron (2009, Table I) report parameters for a monthly model, so we can take them as is. The only change is to approximate their two-component model of log consumption growth with a univariate model. We use the univariate model with the same autocovariance function, which is an ARMA(1,1) in this case.

Their model consists of two equations,

$$
\begin{aligned}
\log g_t &= \log g + x_{t-1} + \sigma_g w_{gt} \\
x_t &= \varphi x_{t-1} + \sigma_x w_{xt},
\end{aligned}
$$

with $|\varphi| < 1$ and (w_g, w_x) independent iid series with zero mean and unit variance. The autocovariance function is

$$
\begin{aligned}
Var(\log g_t) &= \sigma_g^2 + \sigma_x^2/(1-\varphi^2) \\
Cov(\log g_t, \log g_{t-k}) &= \varphi^{k-1}\sigma_x^2/(1-\varphi^2), \quad k \geq 1,
\end{aligned}
$$

which has the form of an ARMA(1,1).

Our model can be expressed as an infinite moving average with coefficients γ_j satisfying $\gamma_{j+1} = \varphi\gamma_j = \varphi^j\gamma_1$ for $j \geq 1$. The kth autocovariance is

$$
Cov(\log g_t, \log g_{t-k}) = \sum_{j=0}^{\infty} \gamma_j \gamma_{j+k}
$$

for any $k \geq 0$. Therefore

$$
\begin{aligned}
Var(\log g_t) &= \gamma_0^2 + \gamma_1^2/(1-\varphi^2) \\
Cov(\log g_t, \log g_{t-k}) &= \varphi^{k-1}\left[\gamma_0\gamma_1 + \varphi\gamma_1^2/(1-\varphi^2)\right] \quad k \geq 1.
\end{aligned}
$$

Evidently φ plays the same role in each model. We then choose (γ_0, γ_1) to match the variance and first autocovariance in the first model. This involves a quadratic equation in the ratio γ_1/γ_0. We choose the root associated with an invertible moving average coefficient for reasons outlined in Sargent (1987, Section XI.15).

Jump models. We adapt the parameters governing the dynamics of the intensity process h_t from Wachter (2008, Table 1). 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 we used in our iid example: $h = 0.01\tau$. The remaining parameters are

connected to hers as follows (our parameters on the left, her parameters and numbers on the right):

$$
\begin{aligned}
\varphi_h &= e^{-\kappa\tau} = e^{-0.08/12} = 0.9934 \\
\eta_0 &= 0.0114 \cdot \tau^{1/2} = 0.0033.
\end{aligned}
$$

This gives our process the same autocorrelation and variance as hers.

Finding b_1. We've described approximate solutions to recursive models given a value of the approximating constant b_1. See equations (20, 21) and the surrounding discussion. We compute b_1 this way:

0. Initialize. Set $b_1 = \beta$.

1. Solve. Compute the solution to the model, including $\log\mu$, the mean value of $\log\mu_t$.

2. Verify. Substitute $\log\mu$ into equation (21) and compute a new approximation for b_1.

3. Check. If the new value of b_1 is close to the previous one, stop. Otherwise, return to step 1.

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*, in press.

Backus, David, Silverio Foresi, Abon Mozumdar, Liuren Wu, 2001, "Predictable changes in yields and forward rates," *Journal of Financial Economics* 59, 281-311.

Backus, David, Bryan Routledge, and Stanley Zin, 2008, "The cyclical component of US asset returns," manuscript, November.

Backus, David K., and Stanley E. Zin, 1994, "Reverse engineering the yield curve," NBER Working Paper No. 4676.

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.

Bansal, Ravi, Dana Kiku, and Amir Yaron, 2009, "An empirical evaluation of the long-run risks model for asset prices," manuscript.

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," manuscript, June.

Beeler, Jason, and John Y. Campbell, 2009, "The long-run risks model and aggregate asset prices: an empirical assessment," NBER Working Paper No. 14788.

Bekaert, Geert, and Eric Engstrom, 2010, "Asset return dynamics under bad environment-good environment fundamentals," manuscript, June.

Benzoni, Luca, Pierre Collin-Dufresne, and Robert S. Goldstein, 2011, "Explaining asset pricing puzzles associated with the 1987 market crash," *Journal of Financial Economics*, in press.

Binsbergen, Jules van, Michael Brandt, and Ralph Koijen, 2010, "On the timing and pricing of dividends," manuscript, October.

Branger, Nicole, Paulo Rodrigues, and Christian Schlag, 2011, "The role of volatility shocks and rare events in long-run risk models," manuscript, March.

Broadie, Mark, Mikhail Chernov, and Michael Johannes, 2009, "Understanding index option

returns," *Review of Financial Studies*, in press.

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 *Handbook of Macroeconomics, Volume 1*, J.B. Taylor and M. Woodford, eds., New York: Elsevier.

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.

Chernov, Mikhail, and Philippe Mueller, 2008, "The term structure of inflation expectations," CEPR Discussion Paper No. 6809.

Constantinides, George, 1990, "Habit formation: a resolution of the equity premium puzzle," *Journal of Political Economy* 98, 519-543.

Deaton, Angus, 1993, *Understanding Consumption*, New York: Oxford University Press.

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," manuscript, Johns Hopkins.

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, 519543.

Gabaix, Xavier, 2010, "Variable rare disasters: an exactly solved framework for ten puzzles in macro-finance," manuscript, March.

Gallmeyer, Michael, Burton Hollifield, Francisco Palomino, and Stanley Zin, 2007, "Arbitrage-free bond pricing with dynamic macroeconomic models," *Federal Reserve Bank of St Louis Review*, 205-326.

Garcia, Rene, Richard Luger, and Eric Renault, 2003, "Empirical assessment of an intertemporal option pricing model with latent variables," *Journal of Econometrics* 116, 49-83.

Hansen, Lars Peter, 2008, "Modeling the Long Run: Valuation in Dynamic Stochastic Economies," manuscript, August.

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 Jose Scheinkman, 2009, "Long term risk: an operator approach," *Econometrica* 77, 177-234.

Heaton, John, 1993, "The interaction between time-nonseparable preferences and time aggregation," *Econometrica* 61, 353-385.

Heaton, John, 1995, "An empirical investigation of asset pricing with temporally dependent preference specifications," *Econometrica* 63, 681-717.

Koijen, Ralph S.J., Hanno Lustig, and Stijn Van Nieuwerburgh, 2010, "The cross-section and time-series of stock and bond returns, manuscript, December.

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.

Sargent, Thomas J., 1987, *Macroeconomic Theory (Second Edition)*, Academic Press: San Diego.

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.

Sundaresan, Suresh, 1989, "Intertemporally dependent preferences and the volatility of consumption and wealth," *Review of Financial Studies* 2, 73-89.

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, 2008, "Can time-varying risk of rare disasters explain aggregate stock market volatility?," NBER Working Paper No. 14386.

Weil, Philippe, 1989, "The equity premium puzzle and the risk-free rate puzzle," *Journal of Monetary Economics* 24, 401-421.

Table 1
Properties of monthly excess returns

Asset	Mean	Standard Deviation	Skewness	Excess Kurtosis
Equity				
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
Equity options				
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
Currencies				
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
Nominal bonds				
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

Notes. 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 one-month bond. Sample periods: S&P 500, 1927-2008 (source: CRSP), Fama-French, 1927-2008 (source: Kenneth French's website); nominal bonds, 1952-2008 (source: Fama-Bliss dataset, CRSP); currencies, 1985-2008 (source: Datastream); options, 1987-2005 (source: Broadie, Chernov and Johannes, 2009). For options, OTM means out-of-the-money and ATM means at-the-money.

Table 2
Parameters and properties of Bansal-Yaron models

Parameter/ Property	Power Utility (1)	Bansal-Yaron Version 1 (2)	Bansal-Yaron Version 2 (3)	Bansal-Yaron Version 3 (4)
Preferences				
α	-9	-9	-9	-9
ρ	-9	$1/3$	$1/3$	$1/3$
β	0.9970	0.9970	0.9970	0.9970
Consumption growth				
$\log g$	0.0015	0.0015	0.0015	0.0015
γ_0	1	1	1	1
γ_1	0.1246	0.1246	0.1246	0.0250
φ_g	0.9750	0.9750	0.9750	0.9750
$v^{1/2}$	0.0072	0.0072	0.0072	0.0110
ν_0	0	0	0.280×10^{-5}	0.280×10^{-5}
φ_v			0.9990	0.9990
Derived				
CRRA	10	10	10	10
IES	0.1	1.5	1.5	1.5
$E(\log g_t)$	0.0015	0.0015	0.0015	0.0015
$Var(\log g_t)^{1/2}$	0.0135	0.0135	0.0135	0.0135
b_1		0.9954	0.9933	0.9967
$\gamma(b_1)$		5.2054	4.9247	1.8829
$\nu(b_1)$			0.364×10^{-3}	0.651×10^{-3}
θ_g	0.1246	0.0017	0.0018	0.0009
θ_v			-0.0077	-0.0043
Entropy and time dependence				
$EL_t(m_{t+1})$	0.0026	0.0631	0.1241	0.0248
$T(60)$	0.0305	0.0042	0.0077	0.0009

Notes. Entries summarize the parameters and properties of Bansal-Yaron models (Section 4). All models have persistent consumption growth. Column (1) is additive power utility ($\alpha - \rho = 0$). Column (2) is recursive preferences. Column (3) adds stochastic volatility. Column (4) has less persistence in consumption growth (smaller γ_1).

Table 3
Parameters and properties of habit models

Parameter/ Property	Power Utility (1)	Ratio Habit (2)	Difference Habit (3)	Campbell-Cochrane (4)
Preferences				
ρ	-9	-9	-9	-1
Consumption growth				
γ_0	1	1	1	1
γ_1	0	0	0	0
$v^{1/2} = Var(\log g_t)^{1/2}$	0.0135	0.0135	0.0135	0.0135
Habit				
χ_0		0.25	0.25	
φ_x or φ_s		0.75	0.75	0.9885
s			0.5	
b				0
Derived				
$\theta = a_1/a_0$	0	-0.2250	-0.1250	
Entropy and time dependence				
$EL_t(m_{t+1})$	0.0091	0.0091	0.0365	0.0231
$T(60)$	0	-0.0086	-0.0258	0

Notes. Entries summarize the parameters and properties of habit models (Section 5). Consumption growth is iid in all cases. Column (1) is additive power utility (no habit). Column (2) is a ratio habit. Column (3) is a loglinear approximation of a difference habit. Column (4) is the Campbell-Cochrane model.

Table 4
Parameters and properties of jump models

Parameter/ Property	Power Utility No Jump (1)	Power Utility With Jump (2)	Persistent Jump Risk (3)
Preferences			
α	-4	-4	-4
ρ	-4	-4	0
β	0.9990	0.9990	0.9990
Consumption growth			
$v^{1/2}$	0.0135	0.0094	0.0094
$h \times 12$		0.0100	0.0100
θ		-0.3000	-0.3000
δ		0.1500	0.1500
η_0			0.0033
φ_h			0.9934
Derived			
CRRA	5	5	5
IES	0.2	0.2	1
$E(\log g_t)$	0.0015	0.0015	0.0015
$Var(\log g_t)^{1/2}$	0.0135	0.0135	0.0135
$\eta(b_1)$			0.3419
θ_h			
Entropy and time dependence			
$EL_t(m_{t+1})$	0.0023	0.0051	0.5193
$T(60)$	0	0	-0.0002

Notes. Entries summarize the parameters and properties of jump models (Section 6). Column (1) is additive power utility with no jump. Column (2) adds a jump. Column (3) has a jump with persistent intensity h_t and recursive preferences.
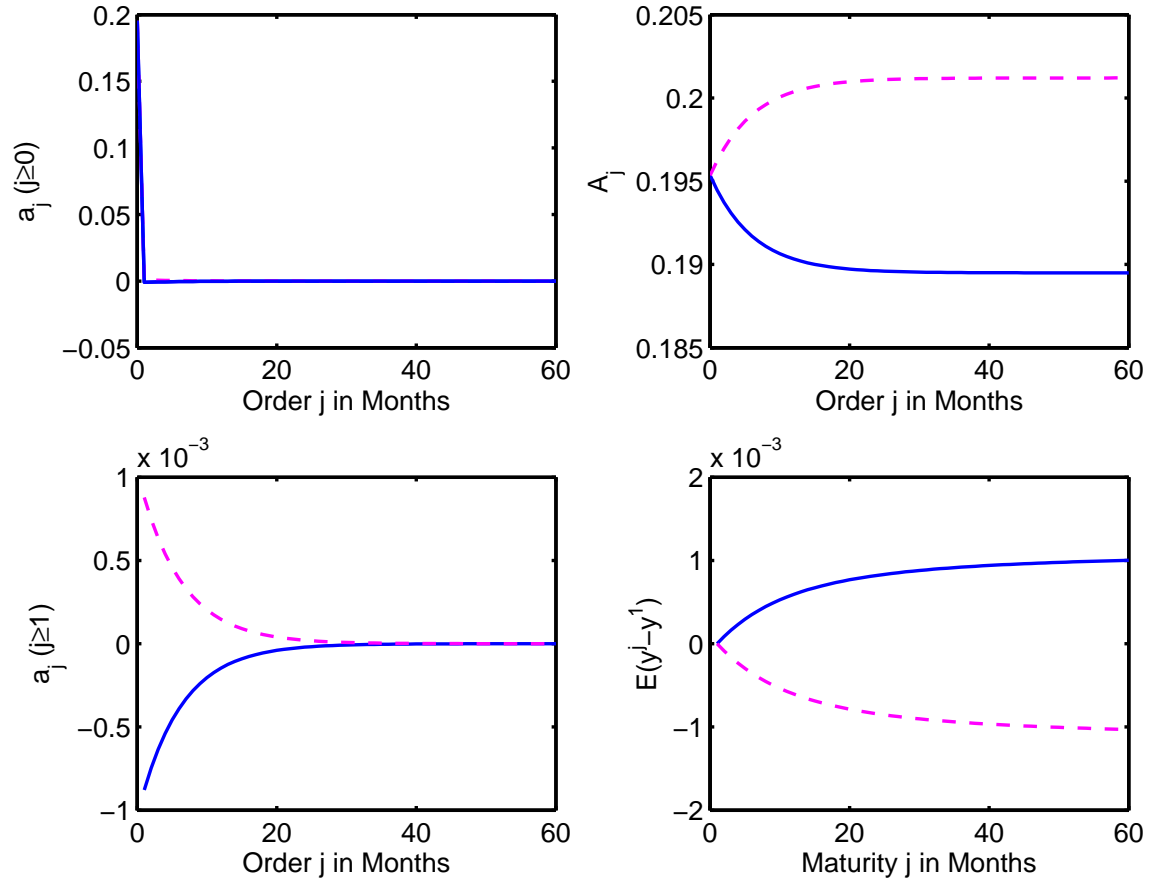
Figure 1
The Vasicek model: moving average coefficients
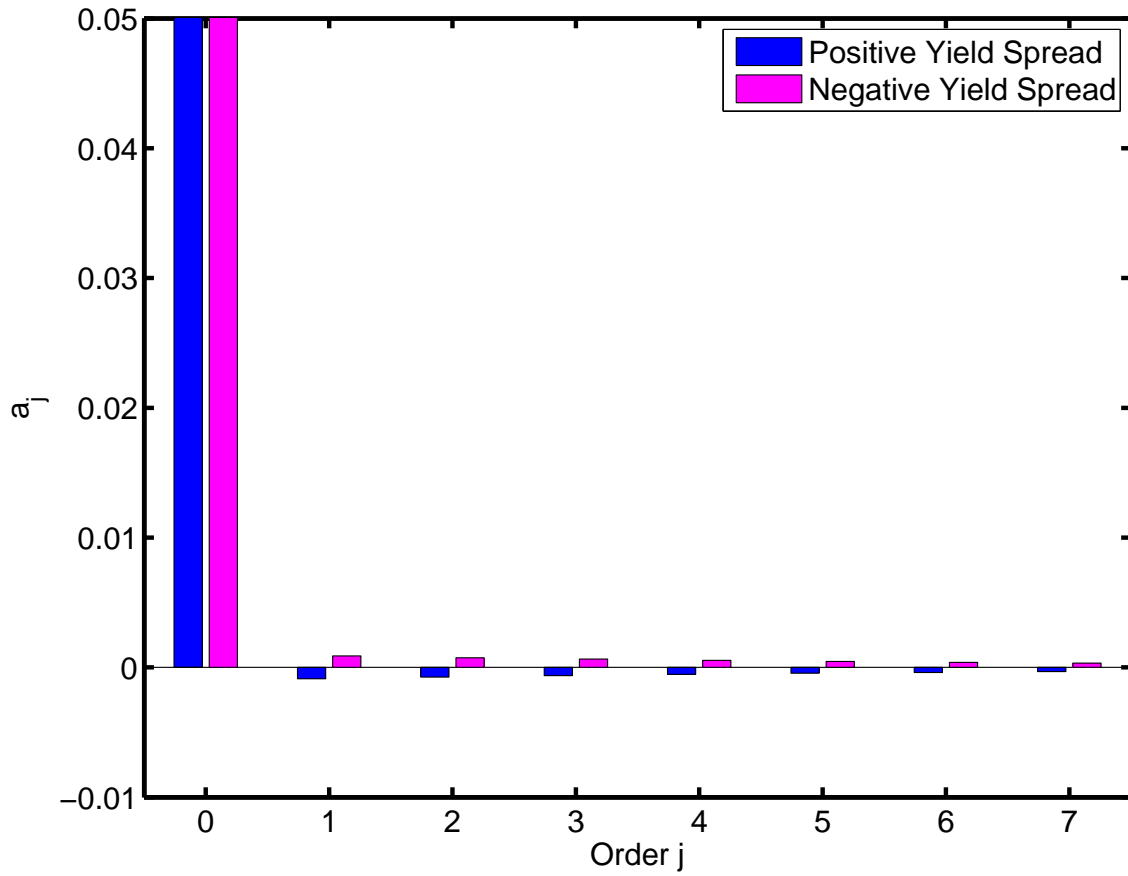
Notes. The bars depict moving average coefficients a_j of the pricing kernel for two versions of the Vasicek model (Section 3). For each j, the first bar corresponds to the model with parameters chosen to produce a positive mean yield spread, the second with parameters that produce a negative yield spread of comparable size. The initial term has been truncated in both cases to make the others visible.

Figure 2
The Vasicek model: moving average coefficients and mean yields

Notes. Like Figure 1, the four panels describe two versions of the Vasicek model (Section 3). In each panel, the solid line corresponds to the model with a positive mean yield spread and the dashed line to one with a negative yield spread. The upper left panel reports the moving average coefficients we saw earlier. The lower left panel gives us a closer look at the same coefficients starting with a_1. The upper right panel reports the partial sums $A_j = \sum_{i=0}^{j} a_i$ that show up in expressions for bond yields, entropy, and time dependence. The lower right panel reports the corresponding mean yield spreads.

Figure 3
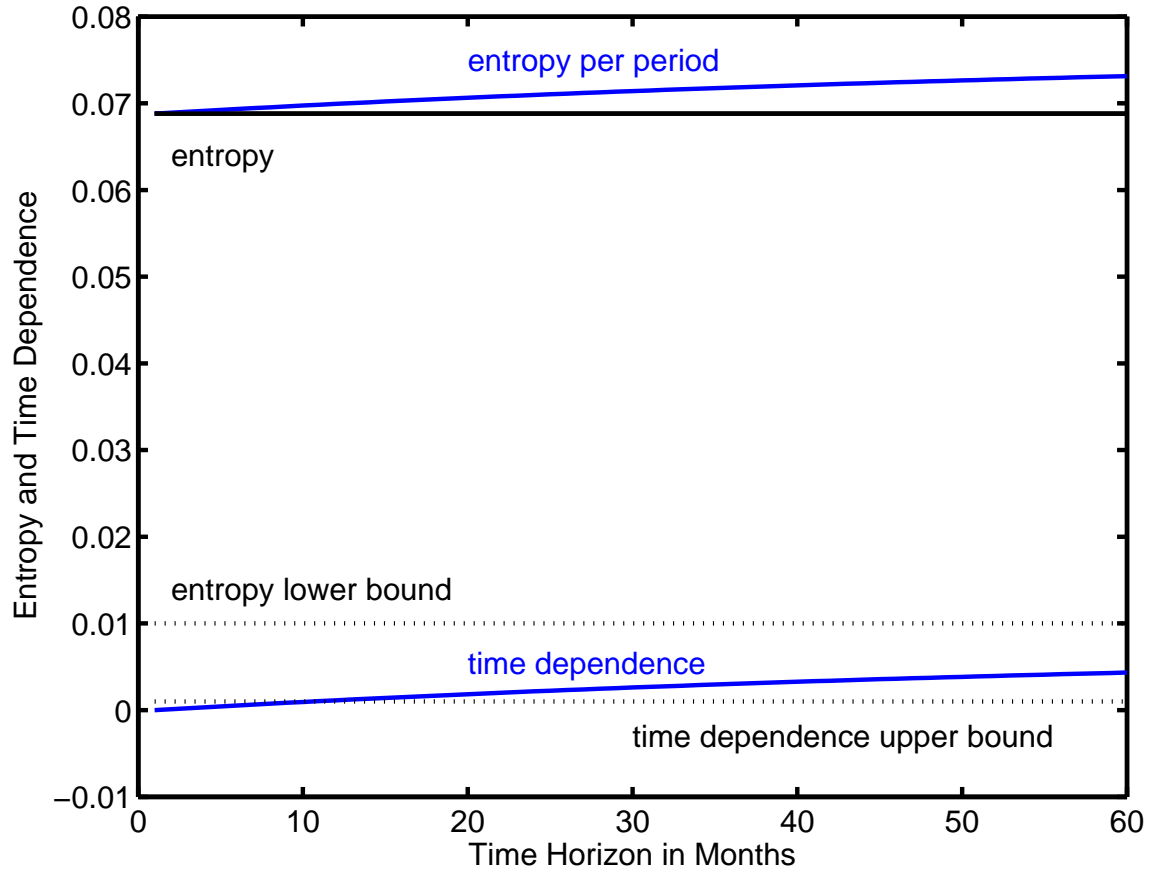**The Vasicek model: entropy and time dependence**
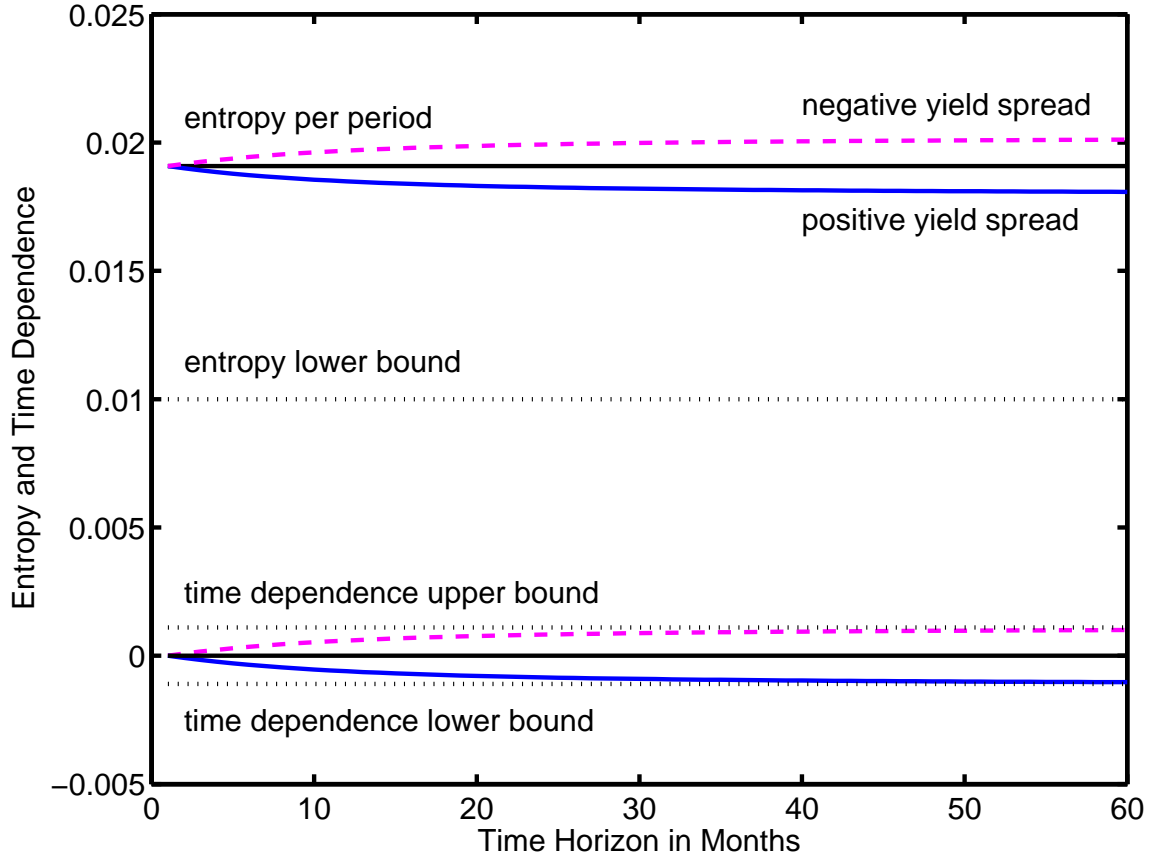
Notes. The lines represent entropy per period and time dependence for three versions of the Vasicek model based, respectively, on positive, zero, and negative mean yield spreads. The trio of lines at the top depict entropy per period over different time horizons. The top (dashed) line is derived from the Vasicek model with a negative yield spread, which corresponds to increasing entropy per period and positive time dependence. The bottom line is derived from the same model with a negative yield spread, which corresponds to decreasing entropy per period and negative time dependence. The center line is horizontal; it implies zero yield spreads and time dependence. The dotted line in the center is an estimated lower bound for entropy. The lines at the bottom of the figure are the same three from the top of the figure, shifted down to represent time dependence, the difference between entropy per period and one-period entropy. The dotted lines around them are estimates of maximum and minimum time dependence.

Figure 4
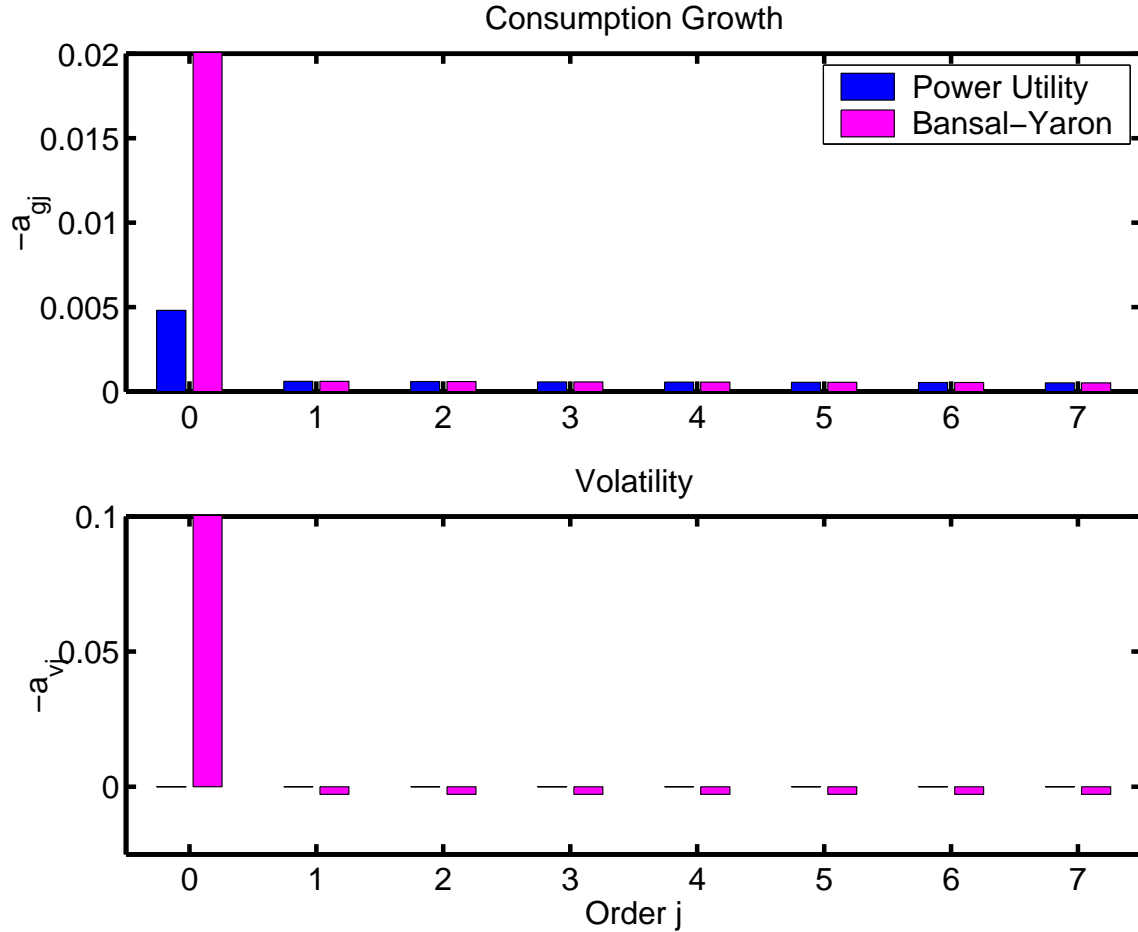**The Bansal-Yaron model with consumption dynamics: moving average coefficients**
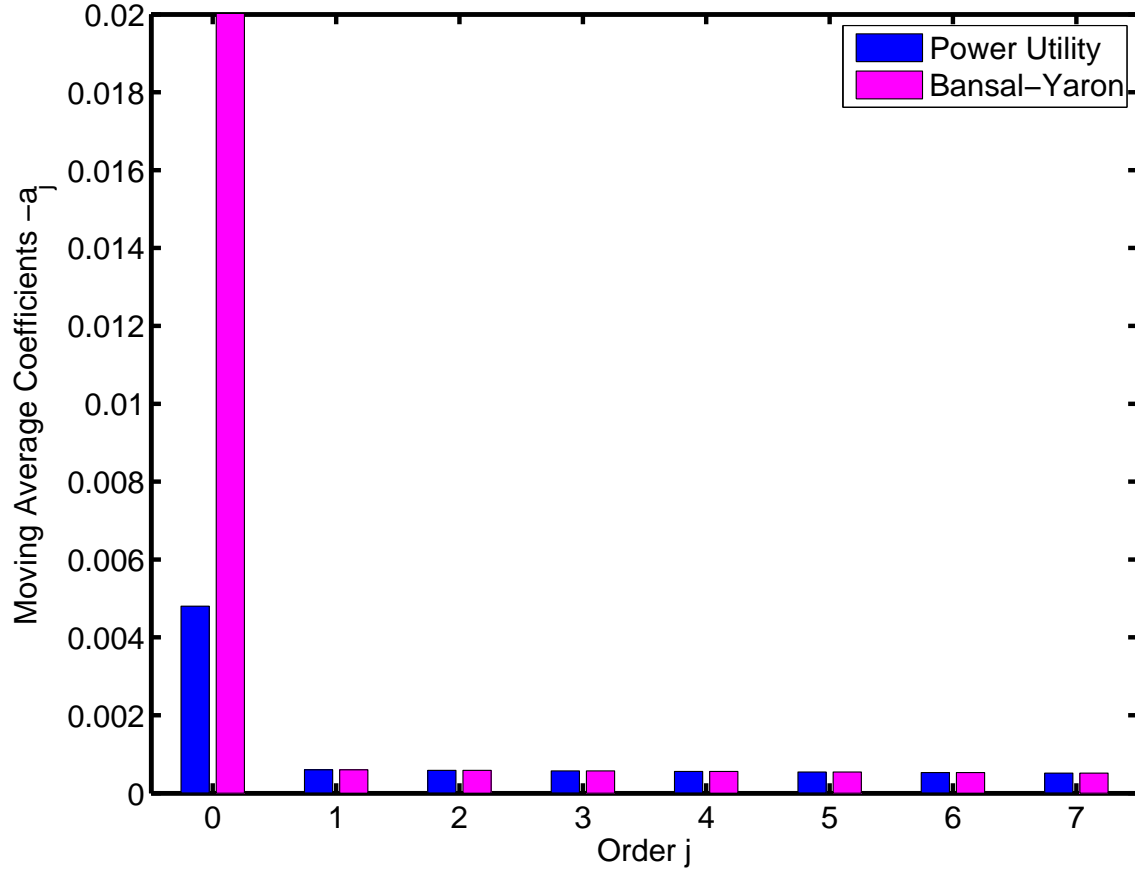
Notes. The figure depicts the moving average coefficients of the pricing kernel for the Bansal-Yaron model with consumption dynamics and constant volatility (Section 4.2). For each j, the first bar corresponds to moving average coefficients with power utility, the second with recursive utility (the Bansal-Yaron model). Power utility, in this case, is based on the same value of ρ as recursive preferences with α set equal to ρ to eliminate the recursive terms in the pricing kernel. We have reversed the sign to correspond to the convention used for the Vasicek model (positive initial coefficient). The initial term has been truncated to make the others visible.

Figure 5
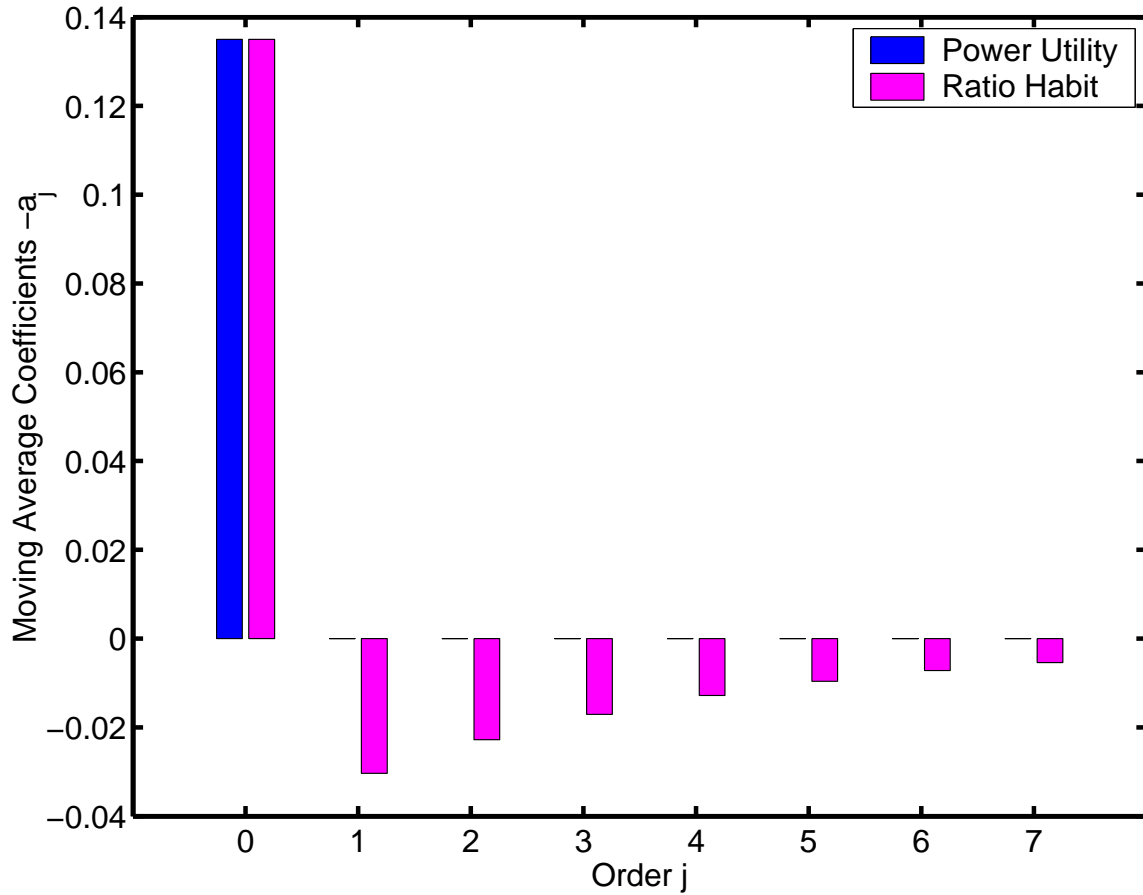**The Bansal-Yaron model with consumption dynamics: entropy and time dependence**

Notes. The lines represent entropy and time dependence for the Bansal-Yaron model with consumption dynamics and constant volatility (Section 4.2). The solid line at the top is entropy. The dotted line below it is an estimated lower bound. The upward-sloping line below that is the model's time dependence as a function of the time horizon. The last dotted line is an estimated upper bound on time dependence.

Figure 6
The Bansal-Yaron model with consumption and volatility dynamics: moving average coefficients


Notes. The figure depicts the moving average coefficients of the pricing kernel for the Bansal-Yaron model with consumption dynamics and stochastic volatility (Section 4.3). For each j, the first bar corresponds to the moving average coefficient with power utility, the second with recursive utility (the Bansal-Yaron model). As in Figure 4, power utility is based on the same value of ρ as recursive preferences, with α set equal to ρ. We have reversed the signs to correspond to the convention used for the Vasicek model (positive initial coefficient). The top panel reports the consumption growth coefficients a_{gj}, the bottom panel the volatility coefficients a_{vj}. In the bottom panel, the coefficients are all zero in the power utility case. In both panels, the initial terms for recursive preferences have been truncated to make the others visible.

Figure 7
The ratio habit model: moving average coefficients

Notes. The figure depicts the moving average coefficients of the pricing kernel for a model with a ratio habit (Section 5.1). For each j, the first bar corresponds to the moving average coefficient with power utility, the second with a ratio habit (the Bansal-Yaron model). We have reversed their signs to correspond to the convention used for the Vasicek model (positive initial coefficient).

Figure 8
Model summary: entropy and time dependence
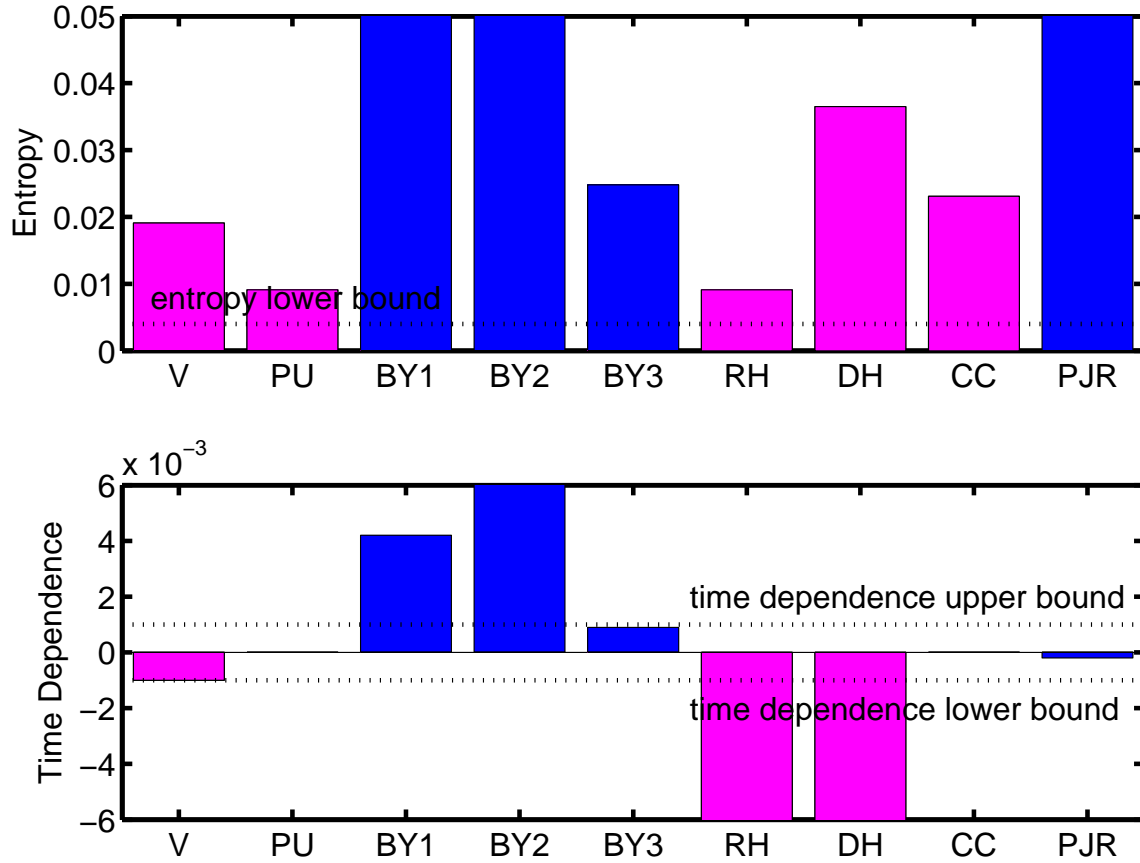
Notes. The figure summarizes entropy and time dependence (at maturity 60 months) for a range of models. They include: V (Vasicek); PU (power utility, column (1) of Table 3); BY1 (Bansal-Yaron with consumption dynamics, column (2) of Table 2); BY2 (Bansal-Yaron with consumption and volatility dynamics, column (3) of Table 2); BY3 (alternative parameterization of BY2, smaller persistent component of consumption growth, column (4) of Table 2); RH (ratio habit, column (2) of Table 3); DH (difference habit, column (3) of Table 3); CC (Campbell-Cochrane, column (4) of Table 3); and PJR (persistent jump risk, column (3) of Table 4). Some of the bars have been truncated to make the others visible.