Term structures of asset prices and returns*

David Backus,† Nina Boyarchenko,‡ and Mikhail Chernov§

Preliminary and incomplete
June 1, 2015

Abstract

We explore the term structures of claims to a variety of cash flows: US government bonds (claims to dollars), foreign government bonds (claims to foreign currency), inflation-adjusted bonds (claims to the price index), and equity (claims to future equity indexes or dividends). Average term structures reflect the dynamics of the dollar pricing kernel, of cash flow growth, and of their interaction. We use simple models to illustrate how relations between the two components can deliver term structures with a wide range of levels and shapes.

JEL Classification Codes: G12, G13.

Keywords: entropy; coentropy; term structure; yields; forward rates; forward contracts.

*Working draft, no guarantees. Comments welcome, including references to related work we inadvertently overlooked. We thank Bryan Routledge, Raghu Sundaram, Bruce Tuckman, Stijn Van Nieuwerburgh, Jonathan Wright, and Liuren Wu for help along the way. Disclaimer: The views expressed here are the authors' and are not representative of the views of the Federal Reserve Bank of New York or the Federal Reserve System.

† Stern School of Business, New York University, and NBER; db3@nyu.edu.

‡ Federal Reserve Bank of New York; nina.boyarchenko@ny.frb.org.

§ Anderson School of Management, UCLA, and CEPR; mikhail.chernov@anderson.ucla.edu.

1 Introduction

[This is half rough draft, half authors' notes to each other, with lots of worked examples to see how things work. It's probably hard to read on its own, but you're welcome to give it a try.]

We study the term structures of a diverse set of assets: dollar bonds, foreign-currency bonds, inflation-protected bonds, and equity indexes and dividends. These assets are claims to different cash flows, which gives their term structures different levels and shapes. The question is where these levels and shapes come from.

Bonds are a useful benchmark. Their cash flows are fixed, so bond prices, yields, forward rates, and returns are functions of the pricing kernel alone. Since the pricing kernel is not directly observed, estimated bond pricing models are essentially reverse engineering exercises, in which properties of the pricing kernel are inferred from bond prices. A central feature of the pricing kernel is its dispersion, which we measure with entropy. We show how the average slopes of yield and forward rate curves are mirrored by the behavior of entropy over different time horizons. In some cases, the long forward rate converges at long maturities to a constant, a feature that has a number of observable consequences for bond prices.

Other assets also have maturity dimensions, which we see in a broad range of forward, futures, and swap contracts. We approach them in a similar way. The term structures in this case are functions of a transformed pricing kernel, the product of the original pricing kernel and the growth rate of the cash flow to which the assets are claims: the spot price of foreign currency, the consumer price index, or an equity dividend. In terms of the original pricing kernel, entropy here is connected to the dispersion of the pricing kernel, the dispersion of cash flow growth, and the relation between the two. We measure dispersion, as before, with entropy, and define a new concept — coentropy — to measure dependence. The cash flows are typically observed, which allows us to estimate their properties, but their coentropy with the pricing kernel is a critical unseen feature that affects their term structures. We show how various elements of models affect the term structures of a number of assets.

Mention debt to Alvarez and Jermann (2005), Hansen and Scheinkman (2008), Martin (2013).

2 Evidence

Our focus is the properties of observed term structures of prices and returns, so it's helpful to begin with data. Consider returns. We measure excess returns with differences in the logarithms of gross monthly returns; you'll see why shortly.

We report summary statistics for some examples in Table 1. A broad-based equity index, for example, has an average excess return of about 0.4 percent monthly or 4.8 percent annually. Some of the Fama-French portfolios have mean excess returns in the neighborhood of 1 percent monthly. We'll use this 1 percent number as an informal lower bound on the maximum mean excess return.

Nominal bonds with maturities greater than a month also have positive average excess returns, but they're significantly smaller. We see in the table mean values of 0.11 percent monthly for 2-year US Treasuries and 0.15 for 5-year Treasuries.

[Currencies? TIPS?]

The US dollar term structure. The US dollar term structure starts low, on average, reflecting low average returns on short-term default-free dollar bonds. Mean yields, forward rates, and returns increase with maturity. The mean spread between one-month and 120-month yields and forward rates have been about 2 percent annually. See Figure 1.

Other term structures. Assets with cash flows also have term structures, although there's not often as much market depth at long maturities as there is with bonds. They differ, in general, in both the starting point (the one-period return on a spot contract) and in how they vary with maturity. The former is covered in Table 1. In Figure 2 we plot the differences between mean yields on a number of other assets and US Treasury yields. We see that some assets have steeper yield curves, some flatter, and some have completely different shapes.

References to related work (more to come): Binsbergen, Brandt, and Koijen (2012), Binsbergen, Hueskes, Koijen, and Vrugt (2012), Binsbergen and Koijen (2015), Boguth, Carlson, Fisher, and Simutin (2013), Boudoukh, Richardson, and Whitelaw (2015), Dahlquist and Hasseltoft (2013, 2014), Dai and Singleton (2003), Hasler and Marfe (2015), Lettau and Wachter (2007), many more ...

3 Entropy, coentropy, and returns

We define entropy and coentropy and connect them to expected excess returns. We'll see in the next section that these concepts generalize easily to time horizons of any length.

3.1 Entropy and coentropy

We start with definitions of entropy, a measure of dispersion, and coentropy, a measure of dependence. The *entropy* of a positive random variable x is

$$L(x) \quad = \quad \log E(x) - E(\log x). \tag{1}$$

Entropy $L(x)$ is nonnegative and positive unless x is constant (Jensen's inequality applied to the log function). It's also invariant to scale: $L(ax) = L(x)$ for any positive constant a.

If we choose $a = 1/E(x)$, then ax is a ratio of probability measures (or Radon-Nikodym derivative) and $L(ax) = L(x)$ is its relative entropy. See Alvarez and Jermann (2005, Section 3), Backus, Chernov, and Martin (2011, Section I.C), Backus, Chernov and Zin (2014, Section I.C), and Cover and Thomas (2006, Chapter 2).

We find it instructive to express entropy in terms of the cumulants and cumulant generating function (cgf) $\log x$. The cgf of $\log x$, if it exists, is the log of its moment generating function,

$$k(s) \;\;=\;\; \log E(e^{s \log x}).$$

The function k is convex in s; see, for example, Figure 3. Given sufficient regularity, it has the Taylor series expansion

$$k(s) \;\;=\;\; \sum_{j=1}^{\infty} \kappa_j s^j / j!,$$

where the jth *cumulant* κ_j is the jth derivative of $k(s)$ at $s = 0$. More concretely, κ_1 is the mean, κ_2 is the variance, $\kappa_3/(\kappa_2)^{3/2}$ is skewness, $\kappa_4/(\kappa_2)^2$ is excess kurtosis, and so on. Entropy is therefore

$$L(x) \;\;=\;\; k(1) - E(\log x) \;\;=\;\; \kappa_2/2! + \kappa_3/3! + \kappa_4/4! + \cdots \;\;=\;\; \sum_{j=2}^{\infty} \kappa_j / j!. \qquad (2)$$

If $E(\log x) = 0$, entropy is simply $k(1)$. See Backus, Chernov, and Martin (2011, Section I.C) and Martin (2013, Sections 1 and 3).

Two examples show how this might work:

Example 1 (normal). Let $\log x \sim \mathcal{N}(\mu, \sigma^2)$. The cgf is $k(s) = \mu s + (\sigma s)^2/2$ and entropy is $L(x) = (\mu + \sigma^2/2) - \mu = \sigma^2/2$. If we compare this to the cumulant expansion (2), we see that normality gives us the variance term $\kappa_2/2$, but all the higher-order terms are zero (κ_j for $j \geq 3$).

Example 2 (Poisson). Let $\log x = j\theta$ where j is Poisson with intensity parameter $\omega > 0$: j takes on nonnegative integer values with probabilities $e^{-\omega}\omega^j/j!$. The cgf of $\log x$ is $k(s) = \omega(e^{\theta s} - 1)$. The mean is $\omega\theta$, the variance is $\omega\theta^2$, and entropy is $\omega(e^{\theta} - 1) - \omega\theta$. Expanding the exponential, we can express entropy in terms of the cumulants of $\log x$:

$$L(x) \;\;=\;\; \omega(\theta^2/2! + \theta^3/3! + \theta^4/4! + \cdots).$$

The first term is half the variance — what we might think of as the normal term. The other terms represent higher-order cumulants. Numerical examples suggest that we can make their overall impact as large or as small as we like. For example, entropy can be smaller than half the variance (try $\theta = -1$) or greater ($\theta = 1$). Or it can be much greater: If $\omega = 1.5$ and $\theta = 5$, half the variance is 18.75 and entropy is 213.62.

We plot both cgf's in Figure 3. The random variables $\log x$ have been standardized, so that they have mean zero and variance one, but they are otherwise the examples described above. In the normal case, the cgf is the parabola $k(s) = s^2/2$ and is symmetric around zero. In the Poisson case, the cgf's asymmetry reflects the positive skewness of a Poisson random variable with positive scale parameter θ. The positive contribution of high-order cumulants in this case drives entropy — the valaue of the cgf k at $s = 1$ — above its normal value of half the variance.

We turn next to the relation between two random variables — what is commonly referred to as *dependence*. If entropy is an analog of variance, then coentropy is an analog of covariance. We define the *coentropy* of two positive random variables x_1 and x_2 as the difference between the entropy of their product and the sum of their entropies:

$$C(x_1, x_2) \quad = \quad L(x_1 x_2) - L(x_1) - L(x_2). \tag{3}$$

This definition is new. If x_1 and x_2 are independent, then $L(x_1 x_2) = L(x_1) + L(x_2)$ and $C(x_1, x_2) = 0$. If $x_1 = a x_2$ for $a > 0$, then coentropy is positive. If $x_1 = a/x_2$, then $L(x_1 x_2) = L(a) = 0$ and coentropy is negative. Coentropy is also invariant to noise. Consider a positive random variable y, independent of x_1 and x_2 — noise, in other words. Then $C(x_1 y, x_2) = C(x_1, x_2 y) = C(x_1, x_2)$.

As with entropy, we can express coentropy in terms of cgf's. The cgf of $\log x = (\log x_1, \log x_2)$ is $k(s_1, s_2) = \log E(e^{s_1 \log x_1 + s_2 \log x_2})$. The cgf's of the components are $k(s_1, 0)$ and $k(0, s_2)$. Coentropy is therefore

$$C(x_1, x_2) \quad = \quad k(1, 1) - k(1, 0) - k(0, 1). \tag{4}$$

The cgf has the Taylor series representation

$$k(s_1, s_2) \quad = \quad \sum_{i,j=0}^{\infty} \kappa_{ij} s_1^i s_2^j / i! j!,$$

where κ_{ij} is the (i, j)th joint cumulant, the (i, j)th cross derivative of k at $s = 0$. Here κ_{i0} is the ith cumulant of $\log x_1$, κ_{0j} is the jth cumulant of $\log x_2$, and κ_{ij} is a joint cumulant — κ_{11}, for example, is the covariance. The details are mind-numbing, but the idea is that coentropy includes contributions from both the covariance and high-order cumulants.

Two examples highlight the differences between covariance and coentropy:

Example 3 (bivariate lognormal). Let $\log x = (\log x_1, \log x_2) \sim \mathcal{N}(\mu, \Sigma)$, where μ is a 2-vector and Σ is a 2 by 2 matrix. The cgf is $k(s) = s^\top \mu + s^\top \Sigma s/2$ where $s^\top = (s_1, s_2)$. Entropies are $L(x_i) = \sigma_{ii}/2$ for $i = 1, 2$ and $L(x_1 x_2) = (\sigma_{11} + \sigma_{22} + 2\sigma_{12})/2$. Coentropy is the covariance: $C(x_1, x_2) = \sigma_{12} = \text{Cov}(\log x_1, \log x_2)$.

Example 4 (bivariate Poisson mixture). Jumps j are Poisson with intensity ω. Conditional on j jumps, $\log x \sim \mathcal{N}(j\theta, j\Delta)$ where the matrix Δ has elements δ_{ij}. The cgf is $k(s) =$

$\omega(e^{s^\top \theta + s^\top \Delta s/2} - 1)$. Entropies are

$$\begin{aligned} L(x_i) &= \omega(e^{\theta_i + \delta_{ii}/2} - 1) - \omega\theta_i \\ L(x_1 x_2) &= \omega\Big(e^{(\theta_1+\theta_2)+(\delta_{11}+\delta_{22}+2\delta_{12})/2} - 1\Big) - \omega(\theta_1 + \theta_2). \end{aligned}$$

Coentropy is therefore

$$C(x_1, x_2) = \omega\Big(e^{(\theta_1+\theta_2)+(\delta_{11}+\delta_{22}+2\delta_{12})/2} - e^{\theta_1+\delta_{11}/2} - e^{\theta_2+\delta_{22}/2} + 1\Big).$$

The covariance is $\text{Cov}(\log x_1, \log x_2) = \omega(\theta_1\theta_2 + \delta_{12})$, so coentropy is clearly different. A numerical example makes the point. Let $\omega = \theta_1 = 1$ and $\Delta = 0$ (a 2 by 2 matrix of zeros). If $\theta_2 = 1$, $C(x_1, x_1) > \text{Cov}(x_1, x_2)$, but if $\theta_2 = -1$, the inequality goes the other way as the odd high-order cumulants flip sign. For similar reasons, it's not hard to construct examples in which the covariance and coentropy have opposite signs.

Another numerical example shows how different they can be. Let $\theta_1 = \theta_2 = -0.5$ and

$$\Delta = \delta \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

We set $\rho = 0$ and $\delta = 1/\omega$. We then vary ω to see what happens to the covariance and coentropy. We see in Figure 4 that the two can be very different.

3.2 Returns and risk premiums

Our interest in these concepts lies in their application to asset pricing, specifically the returns documented in Table 1. Consider an ergodic Markovian environment with state variable x. In such an environment we distinguish between the probability distribution conditional on the state at a specific date and the unconditional or stationary distribution. Entropy and coentropy can be computed with either one. We define conditional entropy and coentropy in terms of the conditional distribution. Entropy and coentropy are their (unconditional) means.

We denote by $r_{t,t+1}$ the (gross) return on an arbitrary asset between dates t and $t+1$. The subscripts are shorthand for dependence on the state at dates t and $t+1$ — that is, $r(x_t, x_{t+1})$. We define the *risk premium* as the expected excess return in logs: $E_t(\log r_{t,t+1} - \log r^1_{t,t+1})$ where E_t is the expectation conditional on the state at date t and $r^1_{t,t+1}$ is the one-period riskfree rate. If E is the expectation computed from the unconditional or equilibrium distribution, the mean risk premium is $E(\log r_{t,t+1} - \log r^1_{t,t+1})$.

Returns and risk premiums follow from the no-arbitrage theorem: There exists a positive pricing kernel m that satisfies

$$E_t(m_{t,t+1} r_{t,t+1}) = 1 \qquad (5)$$

for all returns r. An asset pricing model is then a stochastic process for m. We'll come back later to what asset prices tell us about this stochastic process.

Risk premiums reflect the coentropy of the pricing kernel m with the return r. Jensen's inequality applied to the log of (5) implies

$$E_t(\log r_{t,t+1}) \quad \leq \quad -E_t(\log m_{t,t+1}).$$

Given a pricing kernel m, the price of a one-period riskfree bond is $q_t^1 = E_t(m_{t,t+1})$ and the riskfree rate is $r_{t,t+1}^1 = 1/q_t^1 = 1/E_t(m_{t,t+1})$. The risk premium is therefore bounded above by the entropy of m computed from its conditional distribution:

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

The inequality characterizes the maximum risk premium that can be generated by this pricing kernel. The high-return asset — the one that attains the bound — has return $\log r_{t,t+1} = -\log m_{t,t+1}$. Taking expectations of both sides gives us

$$E(\log r_{t,t+1} - \log r_{t,t+1}^1) \quad \leq \quad E[L_t(m_{t,t+1})]. \tag{6}$$

We refer to the right side as entropy and (6) as the *entropy bound*. See Alvarez and Jermann (2005, Proposition 2), Backus, Chernov, and Martin (2011, Section I.C), and Backus, Chernov, and Zin (2014, Sections I.C and I.D).

The entropy bound gives us the risk premium on an asset whose return has a perfect loglinear relation to the pricing kernel. More generally, risk premiums are governed by the dependence of the return and the pricing kernel, which we measure with coentropy. The pricing relation (5) implies $\log E_t(m_{t,t+1} r_{t,t+1}) = 0$. If we substitute the definition of coentropy and rearrange terms, we have

$$\begin{aligned} E_t(\log r_{t,t+1} - \log r_{t,t+1}^1) &= L_t(m_{t,t+1}) - L_t(m_{t,t+1} r_{t,t+1}) \\ &= -L_t(r_{t,t+1}) - C_t(m_{t,t+1}, r_{t,t+1}). \end{aligned} \tag{7}$$

In general, conditional entropy L_t and coentropy C_t depend on the current state. Unconditionally we have

$$\begin{aligned} E(\log r_{t,t+1} - \log r_{t,t+1}^1) &= E[L_t(m_{t,t+1})] - E[L_t(m_{t,t+1} r_{t,t+1})] \\ &= -E[L_t(r_{t,t+1})] - E[C_t(m_{t,t+1}, r_{t,t+1})]. \end{aligned} \tag{8}$$

We refer to the two terms on the right as the entropy of the return and the coentropy of the return and the pricing kernel. If we defined the risk premium as $\log E_t(r_{t,t+1}) - \log r_{t,t+1}^1$, as some suggest, then the entropy term drops out and we are left with coentropy.

Equation (8) gives us a framework for thinking about the excess returns summarized in Table 1. The table gives us estimates of the left side of equation (8); the right side gives us an interpretation of it. We estimate that the upper bound is at least 1 percent monthly.

Whether expected excess returns on other assets are close to the bound or well below it depends on their entropy and their coentropy. The maximum risk premium comes, as we've seen, when $r_{t,t+1} = 1/m_{t,t+1}$. Then coentropy is

$$E[C_t(m_{t,t+1}, r_{t,t+1})] \;=\; -E[L_t(m_{t,t+1})] - E[L_t(r_{t,t+1} = 1/m_{t,t+1})] \;<\; 0.$$

Equation (8) then reproduces the entropy bound (6). What about the minimum? We can make the risk premium as small as we like by adding random noise to the return, independent of the pricing kernel. That increases the entropy of the return and drives down the risk premium. We can also drive down the coentropy term. If the return is independent of the pricing kernel, coentropy is zero and the risk premium is $-E[L_t(r_{t,t+1})]$, as we just saw. And if we hold the entropy of the return constant, we can make coentropy positive and reduce the risk premium further.

The role of coentropy mirrors that of the covariance in traditional approaches to asset pricing in which risk premiums are defined in terms of levels of returns: $E_t(r_{t,t+1} - r^1_{t,t+1})$. A risk premium defined this way is connected, via (5), to the covariance of the pricing kernel and the return:

$$\begin{aligned} E_t(r_{t,t+1} - r^1_{t,t+1}) &= -\mathrm{Cov}_t(m_{t,t+1}, r_{t,t+1} - r^1_{t,t+1})/E_t(m_{t,t+1}) \\ &= -\mathrm{Cov}_t(m_{t,t+1}, r_{t,t+1})/E_t(m_{t,t+1}). \end{aligned}$$

The high return asset is then defined as the one with the highest Sharpe ratio. Given a pricing kernel, the maximum Sharpe ratio is given by the Hansen-Jagannathan (1991) bound:

$$E_t(r_{t,t+1} - r^1_{t,t+1})/\mathrm{Var}_t(r_{t,t+1} - r^1_{t,t+1})^{1/2} \;\leq\; \mathrm{Var}_t(m_{t,t+1})^{1/2}/E_t(m_{t,t+1}).$$

The expression on the right can be expressed compactly with the cumulant generating function $k_t(s) = \log E_t(e^{s \log m_{t,t+1}})$:

$$\mathrm{Var}_t(m_{t,t+1})^{1/2}/E_t(m_{t,t+1}) \;=\; \left(e^{k_t(2) - 2k_t(1)} - 1\right)^{1/2}. \tag{9}$$

The return that attains the bound is linear, rather than loglinear, in the pricing kernel:

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

We can do the same with unconditional moments, but there's no simple relation between the conditional and unconditional versions of the bound.

Example 5 (Markov pricing kernels). Let

$$\log m_{t,t+1} \;=\; \log \beta + a^\top x_t + b^\top x_{t+1} \tag{10}$$

$$x_{t+1} \;=\; Ax_t + Bw_{t+1}, \tag{11}$$

where $\{w_t\}$ is a sequence of independent random vectors with mean zero, variance one, and (multivariate) cgf $k(s)$. The pricing kernel for this model is often written

$$\log m_{t,t+1} \;=\; \log\beta + (a^\top + b^\top A)x_t + b^\top B w_{t+1} \;=\; \log\beta + \delta^\top x_t + \lambda^\top w_{t+1}. \quad (12)$$

Entropy is $E[L_t(m_{t,t+1})] = L_t(m_{t,t+1}) = k(B^\top b) = k(\lambda)$. If the innovations are multivariate normal, then $k(s) = s^\top s/2$ and entropy is $E[L_t(m_{t,t+1})] = L_t(m_{t,t+1}) = b^\top BB^\top b/2 = \lambda^\top\lambda/2$.

Example 6 (Vasicek model). A popular univariate model is

$$\log m_{t,t+1} \;=\; \log\beta + x_t + \lambda w_{t+1} \quad (13)$$

$$x_{t+1} \;=\; \varphi x_t + \sigma w_t, \quad (14)$$

where φ is between zero and one and $\{w_t\}$ is an iid sequence with mean zero, variance one, and cgf $k(s)$. Entropy is

$$E[L_t(m_{t,t+1})] \;=\; L_t(m_{t,t+1}) \;=\; \log E_t(m_{t,t+1}) - E_t(\log m_{t,t+1}) \;=\; k(\lambda),$$

an upper bound on expected excess log returns. This maximum risk premium is determined by the coefficient λ on the risk w_{t+1} and by the distribution of the risk, represented by the cgf k.

Example 7 (moving average pricing kernel). Let the log pricing kernel have moving average form:

$$\log m_{t,t+1} \;=\; \log\beta + \sum_{j=0}^{\infty} a_j w_{t+1-j} \quad (15)$$

with $\sum_j a_j^2 < \infty$. The Vasicek model is a special case with $a_0 = \lambda$, $a_1 = \sigma$, and $a_{j+1} = \varphi a_j$ for $j \geq 1$. Consider, in the general model, the risk premium on an asset with arbitrary loglinear return

$$\log r_{t,t+1} \;=\; \log\gamma + \sum_{j=0}^{\infty} c_j w_{t+1-j}.$$

The pricing relation (5) implies the restrictions $\log\beta + \log\gamma + k(a_0 + c_0) = 0$ and $a_j + c_j = 0$ for $j \geq 1$. The expected excess return is therefore

$$E_t(\log r_{t,t+1} - \log r^1_{t,t+1}) \;=\; -k(a_0 + c_0) + k(a_0).$$

Since k is convex and $k'(0) = 0$ (w has mean zero), the asset with the largest risk premium has $c_0 = -a_0$ and attains the entropy bound. With the substitutions $E[C_t(m_{t,t+1}, r_{t,t+1})] = C_t(m_{t,t+1}, r_{t,t+1}) = k(a_0 + c_0) - k(a_0) - k(c_0)$ and $E[L_t(r_{t,t+1})] = L_t(r_{t,t+1}) = k(c_0)$, we reproduce equation (8). The dependence of m and r is captured by $k(a_0 + c_0)$, which depends on the distribution of w through its cgf k.

Example 8 (state-dependent price of risk). The examples so far have had constant conditional entropy. Duffee (2002) developed an alternative that's been widely used in studies of bond prices. The univariate version is

$$\log m_{t,t+1} \;\; = \;\; \log\beta - (\lambda_0 + \lambda_1 x_t)^2/2 + x_t + (\lambda_0 + \lambda_1 x_t) w_{t+1} \qquad (16)$$

with transition equation (14) for x and $\{w_t\}$ iid standard normal. The critical ingredient is the coefficient $\lambda_0 + \lambda_1 x_t$ of w_t, a linear function of the state. Conditional entropy,

$$L_t(m_{t,t+1}) \;\; = \;\; (\lambda_0 + \lambda_1 x_t)^2/2,$$

is the maximum risk premium in state x_t. Entropy is its mean: $E[L_t(m_{t,t+1})] = [\lambda_0^2 + \lambda_1^2/(1-\varphi)^2]/2$.

4 Term structures of prices and returns

We're now ready to attack term structures of asset prices and returns. We do this by highlighting the connection to entropy over different time horizons. If that seems like a detour, we would argue it gives us a useful framework for interpreting the evidence we reviewed in Section 2.

4.1 The term structure of zero-coupon bonds

We start with definitions. Let p_t^n be the price at date t of an n-period zero-coupon bond, a claim to a cash flow of one at $t+n$. By convention $p_t^0 = 1$ (the price of one now is one). Continuously-compounded *yields* are $y_t^n = -n^{-1}\log p_t^n$. One-period *forward rates* are $f_t^n = -\log(p_t^{n+1}/p_t^n)$, which implies $-\log p_t^n = \sum_{j=1}^n f_t^{j-1}$ and $y_t^n = n^{-1}\sum_{j=1}^n f_t^{j-1}$. Yield and forward rate curves, such as those described in Section 2, are the sequences $(y_t^1, y_t^2, \ldots)$ and $(f_t^0, f_t^1, f_t^2, \ldots)$.

Returns depend on changes in bond prices from one period to another, but we think it's helpful to express these changes in terms of forward rates. The *one-period return* on an n-period bond is $r_{t,t+1}^n = p_{t+1}^{n-1}/p_t^n$. Since $p_{t+n}^0 = 1$, we have

$$p_t^n \prod_{j=0}^{n-1} r_{t+j,t+j+1}^{n+1-j} \;\; = \;\; 1.$$

The average log return over the life of the bond is therefore the yield or average forward rate:

$$n^{-1}\sum_{j=0}^{n-1} \log r_{t+j,t+j+1}^{n-j} \;\; = \;\; y_t^n \;\; = \;\; n^{-1}\sum_{j=0}^{n-1} f_t^{n-1-j}.$$

The one-period return on an n-period bond is

$$\log r^n_{t,t+1} \;=\; -\sum_{j=1}^{n-1} f^{j-1}_{t+1} + \sum_{j=1}^{n} f^{j-1}_t \;=\; \sum_{j=1}^{n-1} (f^{j-1}_t - f^{j-1}_{t+1}) + f^{n-1}_t. \qquad (17)$$

In an ergodic environment, the mean return equals the analogous mean forward rate: $E(\log r^n) = E(f^{n-1})$. Similarly, mean excess returns $E(\log r^n - \log r^1)$ equal mean forward spreads $E(f^{n-1} - f^0)$. Thus the evidence in Section 2 on bond returns and forward rates represent two approaches to the same thing.

In an arbitrage-free setting, bond prices inherit their properties from the pricing kernel. Pricing has a simple recursive structure. Applying the pricing relation (5) to bond returns gives us

$$p^n_t \;=\; E_t(m_{t,t+1} p^{n-1}_{t+1}) \;=\; E_t(m_{t,t+n}), \qquad (18)$$

where $m_{t,t+n} = m_{t,t+1} m_{t+1,t+2} \cdots m_{t+n-1,t+n}$.

The right side of (18) suggests a link between the n-period bond price and the conditional entropy of the n-period pricing kernel:

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

Taking expectations as before, we define entropy for horizon n by

$$\mathcal{L}_m(n) \;\equiv\; E[L_t(m_{t,t+n})] \;=\; E[\log E_t(m_{t,t+n})] - E(\log m_{t,t+n}).$$

The first term on the right is the mean log bond price, which is easily expressed in terms of mean yields or forward rates:

$$E[\log E_t(m_{t,t+n})] \;=\; -nE(y_n) \;=\; -\sum_{j=1}^{n} E(f^{j-1}).$$

By convention, $m_{t,t} = 1$, so $\mathcal{L}_m(0) = 0$. If $n = 1$, we're back where we were in Section 3.1.

The dynamics of the pricing kernel are reflected in what we call *horizon dependence*, the relation between entropy and the time horizon represented by the function $\mathcal{L}_m(n)$. The value of this function in a term structure context is its relation to mean yields and forward rates. If one-period pricing kernels $\{m_{t,t+1}\}$ are iid, entropy is proportional to n. Bond yields and forward rates are then the same at all maturities and constant over time. Differences from this proportional benchmark reflect dynamics in the pricing kernel. One version of this difference is captured in what we call *average horizon dependence*:

$$H(n) \;=\; n^{-1}\mathcal{L}_m(n) - \mathcal{L}_m(1).$$

The connection with bond yields then gives us $H(n) = -E(y^n - y^1)$. In the iid case, $H(n) = 0$ and the yield curve is flat. If the mean yield curve slopes upwards, then $H(n)$ is negative and slopes downward. Another version is *marginal horizon dependence*:

$$F(n) \;=\; [\mathcal{L}_m(n+1) - \mathcal{L}_m(n)] - \mathcal{L}_m(1).$$

This is connected to forward rates by $F(n) = -E(f^n - f^0)$. Average and marginal horizon dependence are connected in the same way yields and forwards rates are connected: $H(n) = n^{-1}\sum_{j=1}^{n} F(j-1)$. The two concepts capture the same information.

Horizon dependence has a coentropy concept hidden inside it. This is clearest in the two-period case:

$$\mathcal{L}_m(2) \;=\; E[L_t(m_{t,t+2})] \;=\; 2E[L_t(m_{t,t+1})] - E[C_t(m_{t,t+1}, m_{t+1,t+2})].$$

If the coentropy of successive one-period pricing kernels is zero, then average and marginal horizon dependence are zero as well.

Two of our earlier examples illustrates how the dynamics of the pricing kernel reappear in forward rates and horizon dependence:

Example 5 (Markov pricing kernel, continued). Bond prices follow from the pricing kernel (12), the transition equation (11), and the pricing relation (5). They imply bond prices of the form $\log q^n(x) = a_n + b_n^\top x$ with coefficients (a_n, b_n) satisfying

$$\begin{aligned} a_{n+1} &= a_n + \log\beta + k(\lambda + B^\top b_n) \\ b_{n+1} &= \delta^\top + b_n^\top A \;=\; \delta^\top (I + A + \cdots + A^n) \end{aligned}$$

starting with $a_n = b_n = 0$. Entropy is therefore

$$\mathcal{L}_m(n) \;=\; E(\log q^n - n\log m) \;=\; a_n - n\log\beta \;=\; \sum_{j=1}^{n} k(\lambda + B^\top b_{j-1}).$$

Example 7 (moving average pricing kernel, continued). With the moving average pricing kernel (15), forward rates are

$$-f_t^n \;=\; \log\beta + k(A_n) + \sum_{j=0}^{\infty} a_{n+1+j} w_{t-j}, \qquad (19)$$

where $A_n = \sum_{j=0}^{n} a_j$. The mean forward spread is therefore $E(f^n - f^0) = k(A_0) - k(A_n)$. Here we see the dynamics of the pricing kernel reflected in the partial sums A_n and the distribution of innovations reflected in the cgf k. Similarly, entropy is

$$\mathcal{L}_m(n) \;=\; E[L_t(m_{t,t+n})] \;=\; \sum_{j=1}^{n} k(A_{j-1}).$$

Therefore

$$H(n) \;=\; n^{-1}\sum_{j=1}^{n} k(A_{j-1}) - k(A_0), \qquad F(n) \;=\; k(A_n) - k(A_0),$$

and we have the suggested connection between horizon dependence and mean forward rate spreads.

The iid case is a useful benchmark: $a_j = 0$ for $j \geq 1$, $A_n = a_0$ for all n, the mean yield and forward rate curves are flat, $\mathcal{L}_m(n) = nk(a_0)$, and $H(n) = F(n) = 0$. Any departure from proportionality in entropy $\mathcal{L}_m(n)$ is evidence against this case. The n-period Hansen-Jagannathan upper bound (9) is then

$$\text{Var}_t(m_{t,t+n})^{1/2}/E_t(m_{t,t+n}) \quad = \quad \left(e^{n[k(2a_0)-2k(a_0)]} - 1\right)^{1/2}.$$

The term in brackets is a positive constant. That gives us, even in this case, a nonlinear relation between the maximum Sharpe ratio and maturity n.

We use the term *long horizon* to refer to the behavior of asset prices and entropy as the time horizon approaches infinity. Hansen and Scheinkman (2008) echo the Perron-Frobenius theorem and consider the problem of finding a positive dominant eigenvalue ν and associated positive eigenfunction v_t satisfying

$$E_t(m_{t,t+1}v_{t+1}) \quad = \quad \nu v_t. \tag{20}$$

If such a pair exists, we can construct the Alvarez-Jermann (2005) decomposition $m_{t,t+1} = m^1_{t,t+1}m^2_{t,t+1}$ with

$$\begin{aligned} m^1_{t,t+1} &= m_{t,t+1}v_{t+1}/(\nu v_t) \\ m^2_{t,t+1} &= \nu v_t/v_{t+1}. \end{aligned}$$

By construction $E_t(m^1_{t,t+1}) = 1$.

Given such an eigenvalue-eigenfunction pair, the long forward rate converges to a constant: $f^n_t \to f^\infty = -\log \nu$. Since yields are averages of forward rates, the long yield converges to the same value. The long bond return is not constant, but its expected value also converges: $r^\infty_{t,t+1} = \lim_{n\to\infty} r^n_{t,t+1} = 1/m^2_{t,t+1} = v_{t+1}/(\nu v_t)$, so that $E(\log r^\infty) = -\log \nu$. See Alvarez and Jermann (2005, Section 3).

The special case $m^1_{t,t+1} = 1$ has gotten a lot of recent attention; see, for example, the review in Borovicka, Hansen, and Scheinkman (2014). The pricing kernel becomes $m_{t,t+1} = m^2_{t,t+1}$. Since the long bond return is its inverse, the long bond is the high return asset. As Alvarez and Jermann (2005) and Borovicka, Hansen, and Scheinkman (2014) note, this isn't remotely realistic, but it's an interesting special case. In logs, the pricing kernel becomes

$$\log m_{t,t+1} \quad = \quad \log \nu + \log v_t - \log v_{t+1}.$$

The log pricing kernel is the first difference of a stationary object, namely v, plus a constant. In a sense, it's been over differenced.

Example 5 (Markov pricing kernel, continued). We guess an eigenvector of the form $\log v_t = c^\top x_t$. If we substitute into (20) we find:

$$c^\top \;=\; (a^\top + b^\top A)(I-A)^{-1}, \;\; \log\nu \;=\; \log\beta + k\Big(B^\top c\Big).$$

If $b=-a$, then $c=a$ and $m^1_{t,t+1}=1$.

Example 7 (moving average pricing kernels, continued). We consider the existence of the Hansen-Scheinkman eigenvalue and eigenfunction in models with moving average pricing kernels.

- Moving average. In the general moving average case (15) with partial sums $A_n = \sum_{j=0}^n a_j$, suppose $\lim_{n\to\infty} A_n = A_\infty$ exists. Then the Hansen-Scheinkman eigenvalue ν and eigenfunction v are

  $$\log\nu \;=\; \log\beta + k(A_\infty), \;\; \log v_t \;=\; \sum_{j=0}^{\infty}(A_\infty - A_j)w_{t-j}, \;\; \log m^1_{t,t+1} \;=\; -k(A_\infty) + A_\infty w_{t+1}.$$

  The long forward rate converges to $-\log\nu$. If, in addition, $A_\infty = 0$, then $m^1_{t,t+1}=1$.

  A special case is the Vasicek model with log pricing kernel (13) and AR(1) state variable x following (14). The moving average coefficients are then $a_0=\lambda$, $a_1=\sigma$, and $a_{j+1} = \varphi a_j = \varphi^j a_1$ for $j\ge 1$. The partial sums are $A_n = \lambda+\sigma(1+\varphi+\cdots+\varphi^{n-1})$, which converge to $A_\infty = \lambda + \sigma/(1-\varphi)$. Forward rates are given by (19). As we increase maturity n, the mean converges to $-\log\beta - k(A_\infty)$. The variance,

  $$\mathrm{Var}(f^n_t) \;=\; \sum_{j=0}^{\infty} a^2_{n+1-j}, \;=\; \sigma^2\varphi^{2n}/(1-\varphi^2),$$

  converges to zero.

- Long memory. Suppose $(1-B)^\delta x_t = \sigma w_t$ with $0<\delta<1/2$, where B is the lag or backshift operator. This has a moving average representation with $a_0=\lambda$, $a_1=\sigma$, and $a_{j+1} = [(j+\delta)/(j+1)]a_j$ for $j\ge 1$. It's said to have long memory because the coefficients approach zero slowly; specifically, the partial sum A_n doesn't converge. See Granger and Joyeux (1980). As a result, the pair (ν, v) doesn't exist. Here the mean of the long forward rate doesn't converge, but its variance goes to zero.

Example 8 (state-dependent price of risk, continued). Recall the model consisting of pricing kernel (16) and transition equation (14). (The Vasicek model of Example 6 is a special case with $\lambda_1 = 0$.) Bond prices satisfy $\log p^n(x) = a_n + b_n x$ with

$$\begin{aligned} a_{n+1} &= a_n + \log\beta + (b_n\sigma)^2/2 + \lambda_0 b_n\sigma \\ b_{n+1} &= 1 + b_n(\varphi+\lambda_1\sigma) \;=\; 1+\varphi^* + \varphi^{*2} + \cdots + \varphi^{*(n-1)}, \end{aligned}$$

where $a_0 = b_0 = 0$ and $\varphi^* = \varphi + \lambda_1\sigma$. Forward rates are therefore

$$\begin{aligned} -f^n(x) &= \log\left[p^{n+1}(x)/p^n(x)\right] \\ &= (a_{n+1}-a_n) + (b_{n+1}-b_n)x \;=\; \log\beta + (b_n\sigma)^2/2 + \lambda_0 b_n\sigma + \varphi^{*n}x. \end{aligned}$$

If $|\varphi^*|<1$, this converges to a constant, independent of x.

4.2 Term structures of other assets

Bonds are simple assets in the sense that their cash flows are known. All the action in valuation comes from the pricing kernel. When we introduce uncertain cash flows, pricing reflects the interaction of the pricing kernel and the cash flows. Nevertheless, we can think about the term structures of these other assets in a similar way. The approach mirrors Hansen and Scheinkman (2009, Sections 3.5 and 4.4).

Consider claims to an arbitrary positive cash flow d with one-period growth rate $g_{t,t+1} = d_{t+1}/d_t$. Examples include the price of a foreign currency, the consumer price index, the dividend on a stock market index, or the spot price of a commodity. If the price at date t of a claim to the growth rate $g_{t,t+n} = d_{t+n}/d_t$ is $\widehat{p}^n_t$, we can define yields and forward rates as before: $\widehat{y}^n_t = -\log \widehat{p}^n_t$ and $\widehat{f}^n_t = -\log(\widehat{p}^{n+1}_t/\widehat{p}^n_t)$. Every cash flow d has its own term structure. In an international context, d might be the dollar price of one euro and g the depreciation rate of the dollar with respect to the euro. The euro-denominated yield and forward rates curves are $(\widehat{y}^1_t, \widehat{y}^2_t, \widehat{y}^3_t, \ldots)$ and $(\widehat{f}^0_t, \widehat{f}^1_t, \widehat{f}^2_t, \ldots)$.

Term structures are connected by forward contracts. If q^n_t is the price of an n-period forward contract, specifying at date t the exchange of q^n_t for $g_{t,t+1}$ at date $t+n$, then arbitrage implies $p^n_t q^n_t = \widehat{p}^n_t$. That leads to the so-called (covered) parity relations

$$
\begin{aligned}
n^{-1} \log q^n_t &= y^n_t - \widehat{y}^n_t \\
\log(q^{n+1}_t/q^n_t) &= f^n_t - \widehat{f}^n_t .
\end{aligned}
$$

If we have forward prices, we can back out the interest differentials directly.

One-period returns are again connected to forward rates. The one-period return on a claim to $g_{t,t+n}$ is $\widehat{r}^n_{t,t+1} = g_{t,t+1}\widehat{p}^{n-1}_{t+1}/\widehat{p}^n_t$. Expressed in terms of forward rates, we have

$$
\log \widehat{r}^n_{t,t+1} \;=\; \log g_{t,t+1} - \sum_{j=1}^{n-1} \widehat{f}^{j-1}_{t+1} + \sum_{j=1}^{n} \widehat{f}^{j-1}_t \;=\; \log g_{t,t+1} + \sum_{j=1}^{n-1} (\widehat{f}^{j-1}_t - \widehat{f}^{j-1}_{t+1}) + \widehat{f}^{n-1}_t .
$$

The mean return is therefore $E(\log \widehat{r}^n) = E(\log g + \widehat{f}^{n-1})$. If $g_{t,t+1} = 1$ at all dates, we're back in the world of the Section 4.1.

We value these assets in the usual way. The pricing relation (5) gives us

$$
\widehat{p}^n_t \;=\; E_t(m_{t,t+1} g_{t,t+1} \widehat{p}^{n-1}_{t+1}) \;=\; E_t(\widehat{m}_{t,t+1} \widehat{p}^{n-1}_{t+1}) \;=\; E_t(\widehat{m}_{t,t+n}), \tag{21}
$$

with $\widehat{m}_{t,t+1} = m_{t,t+1} g_{t,t+1}$, $\widehat{m}_{t,t+n} = \widehat{m}_{t,t+1}\widehat{m}_{t+1,t+2}\cdots\widehat{m}_{t+n-1,t+n}$, and $\widehat{p}^0_t = 1$. This has the same form as the bond pricing equation (21), with $\widehat{m}$ replacing m.

Our focus is on the differences between the two term structures, specifically the differences documented in Section 2 in mean excess returns and in slopes and shapes of mean yield and forward rate curves. The spot market return,

$$
\widehat{r}^1_{t,t+1} \;=\; g_{t,t+1}/E_t\left(m_{t,t+1} g_{t,t+1}\right),
$$

has mean excess return

$$E(\log \widehat{r}^1_{t,t+1} - \log r^1_{t,t+1}) \;\; = \;\; E(\log g + \widehat{f}^0 - f^0). \tag{22}$$

As in equation (8), this reflects both the entropy of the return and its coentropy with the pricing kernel.

The shapes of mean forward rate curves also reflect entropy and coentropy. Mean forward spreads are tied to the entropy of $\widehat{m}$,

$$\mathcal{L}_{\widehat{m}}(n) \;\; \equiv \;\; E[L_t(\widehat{m}_{t,t+n})].$$

Forward horizon dependence

$$\widehat{F}(n) \;\; = \;\; \mathcal{L}_{\widehat{m}}(n+1) - \mathcal{L}_{\widehat{m}}(n) - \mathcal{L}_{\widehat{m}}(1)$$

As before, this is connected to the mean forward spread: $\widehat{F}(n) = -E(\widehat{f}^n - \widehat{f}^0)$. In short, the mean forward spread is tied to the dynamics of the transformed pricing kernel $\widehat{m}$.

The entropy of $\widehat{m}$ over a time horizon of n is connected to the dependence of the dollar pricing kernel m and the growth rate of cash flows g. More concretely, its entropy from the definition of coentropy, equation (3):

$$L_t(\widehat{m}_{t,t+n}) \;\; = \;\; L_t(m_{t,t+n} g_{t,t+n}) \;\; = \;\; C_t(m_{t,t+n}, g_{t,t+n}) + L_t(m_{t,t+n}) + L_t(g_{t,t+n}).$$

Taking expectations of both sides gives us

$$\begin{aligned} \mathcal{L}_{\widehat{m}}(n) \;\; &= \;\; \mathcal{L}_m(n) + E[C_t(m_{t,t+n}, g_{t,t+n})] + E[L_t(g_{t,t+n})] \\ &= \;\; \mathcal{L}_m(n) + \mathcal{C}_{mg}(n) + \mathcal{L}_g(n), \end{aligned} \tag{23}$$

where the second line is simply new notation for the first. The difference between $\mathcal{L}_{\widehat{m}}(n)$ and $\mathcal{L}(n)$, and therefore between mean forward spreads, thus stems from two things: the entropy of the growth rate and the coentropy of the growth rate and the pricing kernel. This is a natural extension to arbitrary time horizons of our earlier claim: that mean excess returns reflect the entropy of the return and the coentropy of the return and the pricing kernel. We'll need more structure to make sense of this, but it's a good place to start.

Example 5 (Markov pricing kernel, continued). We add a process for cash flow growth,

$$\log g_{t,t+1} \;\; = \;\; \log \gamma + \delta_g^\top x_t + \lambda_g^\top w_{t+1}.$$

The transformed pricing kernel is then

$$\begin{aligned} \log \widehat{m}_{t,t+1} \;\; &= \;\; \log m_{t,t+1} + \log g_{t,t+1} \\ &= \;\; (\log \beta + \log \gamma) + (\delta + \delta_g)^\top x_t + (\lambda + \lambda_g)^\top w_{t+1} \\ &= \;\; \log \widehat{\beta} + \widehat{\delta}^\top x_t + \widehat{\lambda}^\top w_{t+1}. \end{aligned}$$

The expressions for bond prices and entropy are the same as before, but with hats.

Long horizon properties have the same structure as before. Suppose m has the eigenvalue-eigenfunction pair (ν, v_t) described by (20). The analogous equation for $\widehat{m}$ has solution $(\widehat{\nu}, \widehat{v}_t)$:

$$E_t(\widehat{m}_{t,t+1}\widehat{v}_{t+1}) \;=\; \widehat{\nu}\widehat{v}_t. \tag{24}$$

There's not, in general, a close relation between them, but there is in some special cases. One special case is a stationary cash flow d. If (ν, v) is a solution to (20), then $\widehat{\nu} = \nu$ and $\widehat{v}_t = v_t/d_t$ is a solution to the "hat" equation (24). There's a similar result in Lustig, Stathopoulos, and Verdelhan (2014) for the real exchange rate. Another special case is one in which the "price-dividend" ratio $\widehat{p}$ is constant. See the unpublished version of Hansen, Heaton, and Li (2008).

Example 5 (Markov pricing kernel, continued). We revert to the original Markov pricing kernel, equation (10), and posit cash flow growth of

$$\log g_{t,t+1} \;=\; \log\gamma + a_g^\top x_t + b_g^\top x_{t+1}.$$

The transformed pricing kernel is therefore

$$\begin{aligned}\log \widehat{m}_{t,t+1} &= (\log\beta + \log\gamma) + (a+a_g)^\top x_t + (b+b_g)^\top x_{t+1} \\ &= \log\widehat{\beta} + \widehat{a}^\top x_t + \widehat{b}^\top x_{t+1},\end{aligned}$$

which has the same form as (10).

Example 7 (moving average pricing kernel, continued). Suppose cash flow growth is

$$\log g_{t,t+1} \;=\; \log\gamma + \sum_{j=0}^{\infty} c_j w_{t+1-j}.$$

The transformed pricing kernel is

$$\log\widehat{m}_{t,t+1} \;=\; (\log\beta + \log\gamma) + \sum_{j=0}^{\infty}(a_j + c_j)w_{t+1-j} \;=\; \log\widehat{\beta} + \sum_{j=0}^{\infty}\widehat{a}_j w_{t+1-j}.$$

Entropy is

$$\mathcal{L}_{\widehat{m}}(n) \;=\; \sum_{j=1}^{n} k(\widehat{A}_{j-1}) \;=\; \sum_{j=1}^{n} k(A_{j-1} + C_{j-1}),$$

where $C_n = \sum_{j=0}^{n} c_j$.

5 Interpreting term structure evidence

We breathe some life into our theoretical framework and examples by linking them to data. There is, of course, a long history of doing just that for bonds and a growing body of work on other assets. We illustrate some basic features with examples and show how simple term structure models might be extended to account for term structures of other assets.

5.1 US dollar bonds

Consider the Vasicek model: Example 6 with normal innovations. We use properties of the Gurkaynak, Sack, and Wright (2007) Treasury data over the period 1970-2015. Expressed as an annual percentage (that is, multiplied by 1200), the short rate f^0 has a standard deviation of 3.476 and an autocorrelation of 0.980. The mean of the 120-month forward spread $f^{120} - f^0$ is 2.349. We reproduce each of these features by choosing the parameter values $a_0 = \lambda = 0.088$, $a_1 = -\sigma = -0.00070$, $\varphi = 0.980$, and $a_j = \varphi^{j-1} a_1$ for $j \geq 1$. The level of the term structure can then be set however we want by adjusting $\log \beta$.

It's important to be clear about the roles of the various parameters. Here $a_1 = -\sigma$ and φ control the variance and autocorrelation of the short rate and $a_0 = \lambda$ controls the slope of the mean forward rate curve. The different signs of a_0 and a_1 produce the upward slope in the mean yield curve. You can see the result in the top panel of Figure 5. The solid line represents the model, the dots are sample means. The bottom panel illustrates the consequences for entropy in the same example. It increases with the time horizon, but less than proportionally. This departure from the iid benchmark is a mirror image of mean forward rate spreads.

The moving average coefficients are plotted in Figure 6. The point is that the initial coefficient $a_0 = \lambda$ is much larger than the later coefficients — roughly two orders of magnitude greater. It implies (in the normal case) entropy of $\mathcal{L}_m(1) = k(a_0) = (a_0)^2/2 \approx 0.004$, which is less than the upper bound we suggested earlier (0.010). Presumably other risks would appear in the pricing kernel in a more complete model.

5.2 Other term structures

The Vasicek model gives us a rough approximation to bond prices and returns, but it does less well with other assets. Excess returns on equity, for example, have only a small correlation (roughly 0.1) with bond returns, which we can't replicate in a one-innovation model.

Consider then a streamlined version of Koijen, Lustig, and Van Nieuwerburgh (2015, Appendix), which we refer to as the KLV model:

$$
\begin{aligned}
\log m_{t,t+1} &= \log \beta + x_t + \lambda_1 w_{1t+1} + \lambda_2 w_{2t+1} \\
x_{t+1} &= \varphi x_t + \sigma w_{1t+1} \\
\log g_{t,t+1} &= \log \gamma + \theta x_t + \eta_1 w_{1t+1} + \eta_2 w_{2t+1},
\end{aligned}
$$

with $(w_{1t}, w_{2t}) \sim \text{NID}(0, I)$. The added disturbance w_2 is white noise, so it has no impact on bond prices, but potentially plays a role in the pricing of claims to cash flow growth g. By varying the weights (η_1, η_2) we can alter the correlation of stock and bond returns.

The transformed pricing kernel has a similar structure:

$$\begin{aligned} \log \widehat{m}_{t,t+1} &= \log m_{t,t+1} + \log g_{t,t+1} \\ &= (\log \beta + \log \gamma) + (1+\theta) x_t + (\lambda_1 + \eta_2) w_{1t+1} + (\lambda_2 + \eta_2) w_{2t+1}. \end{aligned} \quad (25)$$

Asset prices are easily computed by the same approach we used with Vasicek.

This model has a triangular structure, in which $(\sigma, \varphi, \lambda_1)$ control bonds prices, and $(\eta_1, \eta_2, \lambda_2)$ control the return on the cash flow g and its relation to bond returns. That allows us to keep the parameter values we used earlier for bonds and choose the others to mimic the behavior of the cash flow of interest. We consider several in turn.

5.3 Foreign currency bonds

There is an extensive set of markets for bonds denominated in foreign currencies, and a similarly extensive set of currency markets linking them. As we saw in Section 4.2, the term structure in a foreign currency depends on the interaction of the dollar pricing kernel and the growth rate of the cash flow, which here is the depreciation rate of the dollar relative to a specific foreign currency.

We can approximate the depreciation rate g reasonably well in the KLV model with $\theta = 0$, which implies that (log) currency prices are random walks. For major currencies, the standard deviation of monthly depreciation rates is about 3 percent, so we have

$$\eta_1^2 + \eta_2^2 = 0.03^2.$$

The question, then, is how to divide the variance between the two components.

We see the result in Figures 7 and 8. In the first figure, we compare mean forward rate curves in dollars and "yen" (foreign-currency). In the former, we see that the yen curve can be significantly different, even without changing the growth rate $\log \gamma$ of the currency price. In the latter, we see that these changes also show up in the slope, with The differences here reflect differences in the coentropy term. As we increase η_1, we change the price of risk $\lambda_1 + \eta_1$ on the innovation w_1 that drives dollar bond prices. With $\eta_1 = 0$, there is no such effect, and the impact operates solely through the entropy of cash flow growth.

5.4 Equity

Forward contracts on an equity index give us another example of a term structure. Here there's a range of evidence suggesting a downward-sloping relative term structure. We explore this possibility in the KLV model, this time focusing on the coefficient θ of long-run risk x.

We choose parameters to match features of the data, specifically the variance of excess returns on the asset of interest, the correlation of the excess return with that on bonds, and the mean excess return. In the model, these features are

$$\begin{aligned}
\mathrm{Var}(\log r_{t,t+1} - \log r^1_{t,t+1}) &= 0.05^2 &= \eta_1^2 + \eta_2^2 \\
\mathrm{Corr}(\log r_{t,t+1} - \log r^1_{t,t+1}, \log r^{60}_{t,t+1} - \log r^1_{t,t+1}) &= 0.10 &= \mathrm{sgn}(\sigma)\eta_1/(\eta_1^2 + \eta_2^2) \\
E(\log r_{t,t+1} - \log r^1_{t,t+1}) &= 0.004 &= -(\eta_1^2 + \eta_2^2)/2 - (\lambda_1\eta_1 + \lambda_2\eta_2)
\end{aligned}$$

(We've approximated moments here with conditional moments, will fix up later.) The solution is $\eta_1 = -0.005$, $\eta_2 = -0.050$, and $\lambda_2 = 0.097$.

We see the results in Figures 9 and 10. In the former, we see the US dollar benchmark and three other curves. The dashed blue line reflects the impact of entropy in the transformed pricing kernel from the second innovation w_2. Since it's iid, this is constant across maturities and simply shifts the curve down. The other two lines reflect changes in θ from its benchmark value of zero. In one, we set $\theta = -0.25$, which reduces the amount of long-run risk in the transformed pricing kernel and flattens the mean forward rate curve. In the other, we set $\theta = 0.25$ and steepen the curve.

5.5 Inflation-protected bonds

Inflation is very persistent, we probably need another state variable to do it justice.

6 Last thoughts

We wish we had some.

References

Alvarez, Fernando, and Urban Jermann, 2005, "Using asset prices to measure the persistence of the marginal utility of wealth," *Econometrica* 73, 1977-2016.

Backus, David, Mikhail Chernov, and Ian Martin, 2011, "Disasters implied by equity index options," *Journal of Finance* 66, 1969-2012.

Backus, David, Mikhail Chernov, and Stanley Zin, 2014, "Sources of entropy in representative agent models," *Journal of Finance* 69, 51-99.

Bansal, Ravi, and Ivan Shaliastovich, 2013, "A long-run risks explanation of predictability puzzles in bond and currency markets," *Review of Financial Studies* 26, 1-33.

Binsbergen, Jules van, Michael Brandt, and Ralph Koijen, 2012, "On the timing and pricing of dividends," *American Economic Review* 102, 1596-1618.

Binsbergen, Jules van, and Ralph Koijen, 2015, "The term structure of returns: facts and theory," manuscript, March.

Binsbergen, Jules van, Wouter Hueskes, Ralph Koijen, and Evert Vrugt, 2012, "Equity yields," manuscript, March.

Boguth, Oliver, Murray Carlson, Adlai Fisher, Mikhail Simutin, 2013, "Heterogeneous information diffusion and horizon effects in average returns," manuscript, May.

Borovicka, Jaroslav, Lars Peter Hansen, and Jose Scheinkman, 2014, "Misspecified recovery," manuscript, May.

Boudoukh, Jacob, Matthew Richardson, and Robert Whitelaw, 2015, "New evidence on the forward premium puzzle," *Journal of Financial and Quantitative Analysis*, forthcoming.

Chernov, Mikhail, and Philippe Mueller, 2012, "The term structure of inflation expectations," *Journal of Financial Economics* 106, 367-394.

Cover, Thomas, and Joy Thomas, 2006, *Elements of Information Theory (Second Edition)*, New York: John Wiley & Sons.

Dahlquist, Magnus, and Henrik Hasseltoft, 2013, "International bond risk premia," *Journal of International Economics* 90, 17-32.

Dai, Qiang, and Kenneth Singleton, 2003, "Term structure dynamics in theory and reality," *Review of Financial Studies* 16, 631-678.

Dahlquist, Magnus, and Henrik Hasseltoft, 2014, "Empirical evidence on international bond risk premia," manuscript, June.

Doskov, Nikolay, Tapio Pekkala, and Ruy M. Ribeiro, 2013, "Tradable aggregate risk factors and the cross-section of stock returns," manuscript, November.

Duffee, Gregory, 2002, "Term premia and interest rate forecasts in affine models," *Journal of Finance* 57, 405-443.

Gurkaynak, Refet, Brian Sack, Jonathan Wright, 2007, "The US Treasury yield curve: 1961 to the present," *Journal of Monetary Economics* 54, 2291-2304.

Gurkaynak, Refet, Brian Sack, Jonathan Wright, 2010, "The TIPS yield curve and inflation compensation," *American Economic Journal: Macroeconomics* 54, 70-92.

Granger, C.W.J., and Roselyne Joyeux, 1980, "An introduction to long-memory time series models and fractional differencing," *Journal of Time Series Analysis* 1, 15-29.

Hansen, Lars Peter, 2012, "Dynamic value decomposition in stochastic economies," *Econometrica* 80, 911-967.

Hansen, Lars Peter, John C. Heaton, and Nan Li, 2008, "Consumption strikes back? Measuring long-run risk," *Journal of Political Economy* 116, 260-302.

Hansen, Lars Peter, and Ravi Jagannathan, 1991, "Implications of security market data for models of dynamic economies," *Journal of Political Economy* 99, 225-262.

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

Hasler, Michael, and Roberto Marfe, 2015, "Disaster recovery and the term structure of dividend strips," manuscript.

Koijen, Ralph, Hanno Lustig, and Stijn Van Nieuwerburgh, 2015, "The bond risk premium and the cross-section of equity returns," manuscript, April.

Lettau, Martin, and Jessica Wachter, 2007, "Why is long-horizon equity less risky? A duration-based explanation of the value premium," *Journal of Finance* 62, 55-92.

Lustig, Hanno, Andreas Stathopoulos, and Adrien Verdelhan, 2014, "The term structure of currency carry trade risk premia," manuscript, May.

Martin, Ian, 2013, "Consumption-based asset pricing with higher cumulants," *Review of Economic Studies* 80, 745-773.

Wright, Jonathan, 2011, "Term premia and inflation uncertainty: empirical evidence from an international panel dataset," *American Economic Review* 101, 1514-1534.

Table 1. Properties of excess dollar returns. Entries are sample moments of monthly observations of (monthly) log excess returns: $\log r - \log r^1$, where r is a (gross) return and r^1 is the (gross) return on a one-month bond. All of these returns are measured in dollars. 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).

Asset	Mean	Standard Deviation	Skewness	Excess Kurtosis	First Autocorr
Equity					
S&P 500	0.0040	0.0556	−0.40	7.90	
Fama-French (small, high)	0.0090	0.0894	1.00	12.80	
Currencies					
AUD	0.0087	0.0567	−1.25	6.50	
JPY	0.0001	0.0346	0.50	1.90	
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	
Inflation-protected bonds (TIPS)					
2 years					
5 years					
10 years					

Table 2. Properties of cash flow growth. Entries are sample moments of monthly observations of log growth rates of various cash flows: foreign currencies, inflation-protected bonds, and equity dividends. The sample periods are 1985-2015 for currencies, with the exception of the euro which starts in 1999; 1985-2015 for the consumer price index.

Variable	Mean	Standard Deviation	Skewness	Excess Kurtosis	First Autocorr
Foreign currencies					
USD/CAD	0.0002	0.0212	–0.78	6.22	–0.07
USD/EUR	–0.0002	0.0299	–0.20	0.94	0.05
USD/JPY	0.0021	0.0325	0.37	1.58	0.05
USD/GBP	0.0008	0.0294	–0.27	2.71	0.06
USD/CHF	0.0028	0.0339	0.02	0.77	–0.01
Other cash flows					
Consumer price index (sa)	0.0022	0.0026	–1.49	11.34	0.44
Consumer price index (nsa)	0.0022	0.0033	–1.11	6.09	0.48
Equity					
S&P 500 dividends	0.0029	0.0113	–1.40	13.00	

Figure 1. Average yield curve for US Treasuries. The lines represent mean zero-coupon yields on US Treasury securities over the period 1970 to 2015. Yields are continuously compounded and expressed as annual percentages. The data are an updated version of Gurkaynak, Sack, and Wright (2007).

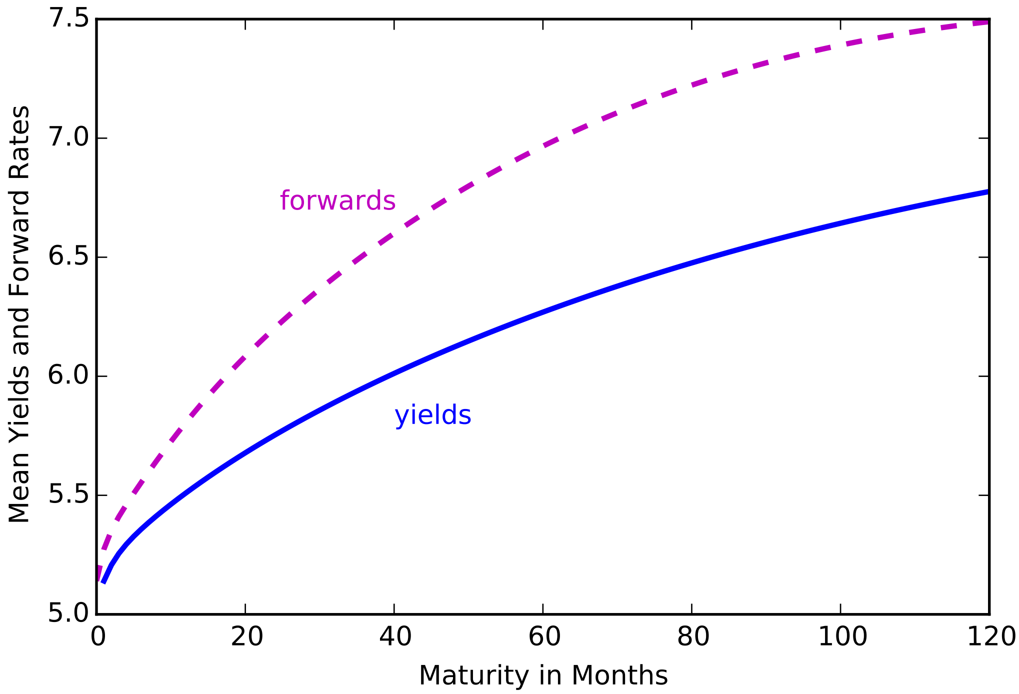

Figure 2. Average forward rate curves. The lines represent differences of average yield curves on several assets relative to US Treasuries.

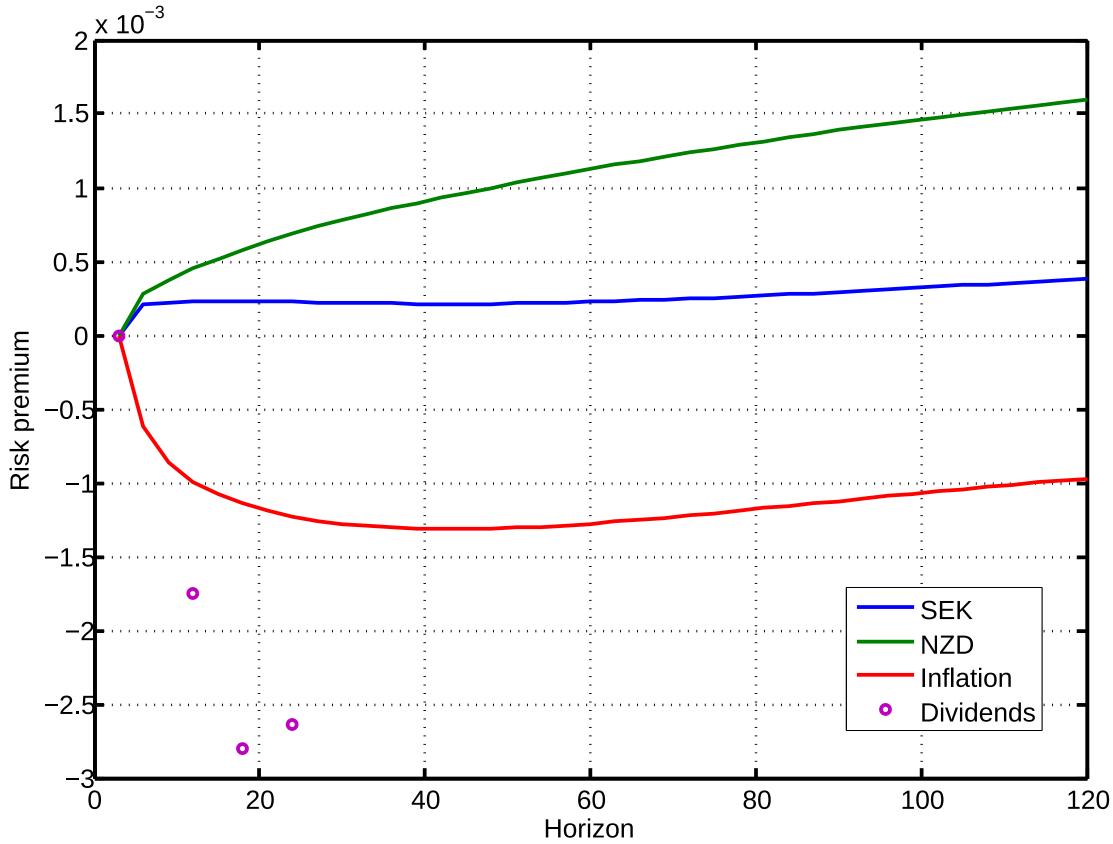

Figure 3. Two cumulant generating functions. The functions $k(s)$ are properties of the distributions of $\log x$. In one, $\log x$ is normal, in the other Poisson. Both are standardized: they have mean zero and variance one. The Poisson has intensity parameter $\omega = 1$ and scale parameter $\theta > 0$. Since the mean is zero, the entropy of x is the value of the cgf at $s = 1$, noted by the dotted line. In the normal example entropy is 0.5 (half the variance). In the Poisson example, entropy is 0.72.

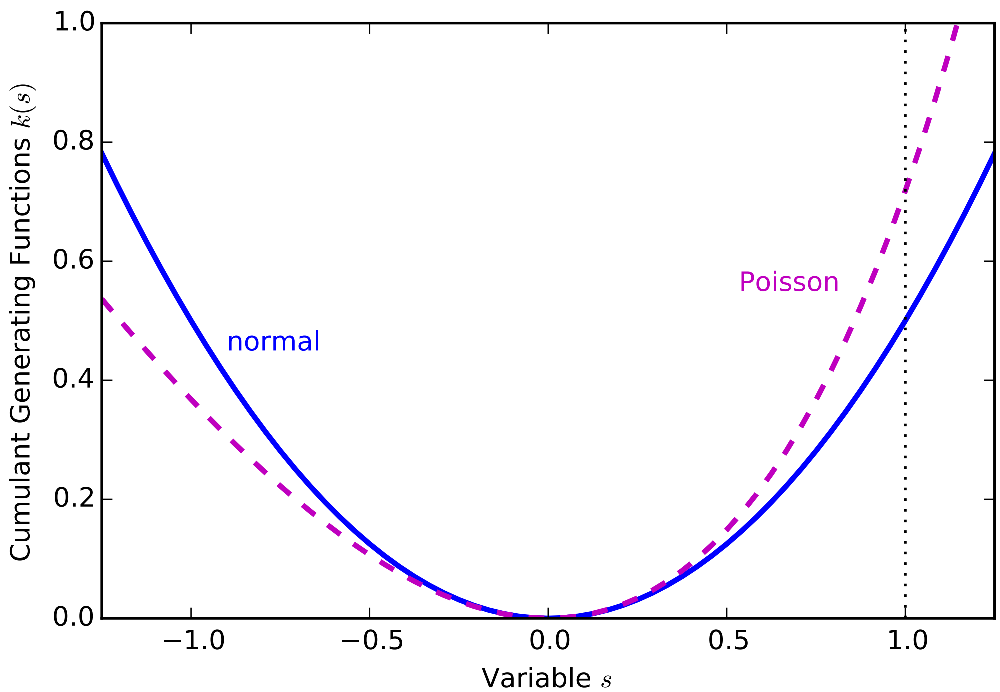

Figure 4. Coentropy and covariance. The figure compares coentropy and covariance for the Poisson mixture of bivariate normals described in Example 4. As we vary ω, we adjust δ to hold the variance constant.

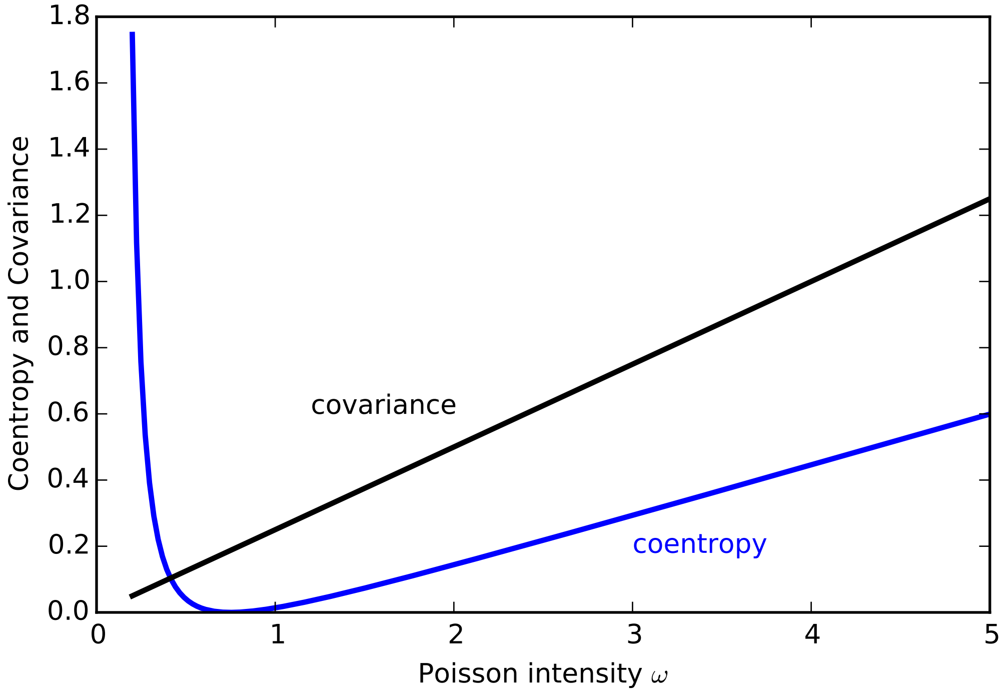

Figure 5. Mean forward spreads and entropy in the Vasicek model. Both panels refer to the numerical example examples of the Vasicek model described in Section 5.1. The top panel compares mean forward rate spreads $(f^n - f^0)$ in the model (the line) to those in US data over the period 1970-2015 (the dots). The bottom panel compares entropy $\mathcal{L}_m(n)$ over different horizons n with a benchmark in which the pricing kernel $m_{t,t+1}$ is iid and entropy is therefore proportional to the time horizon.

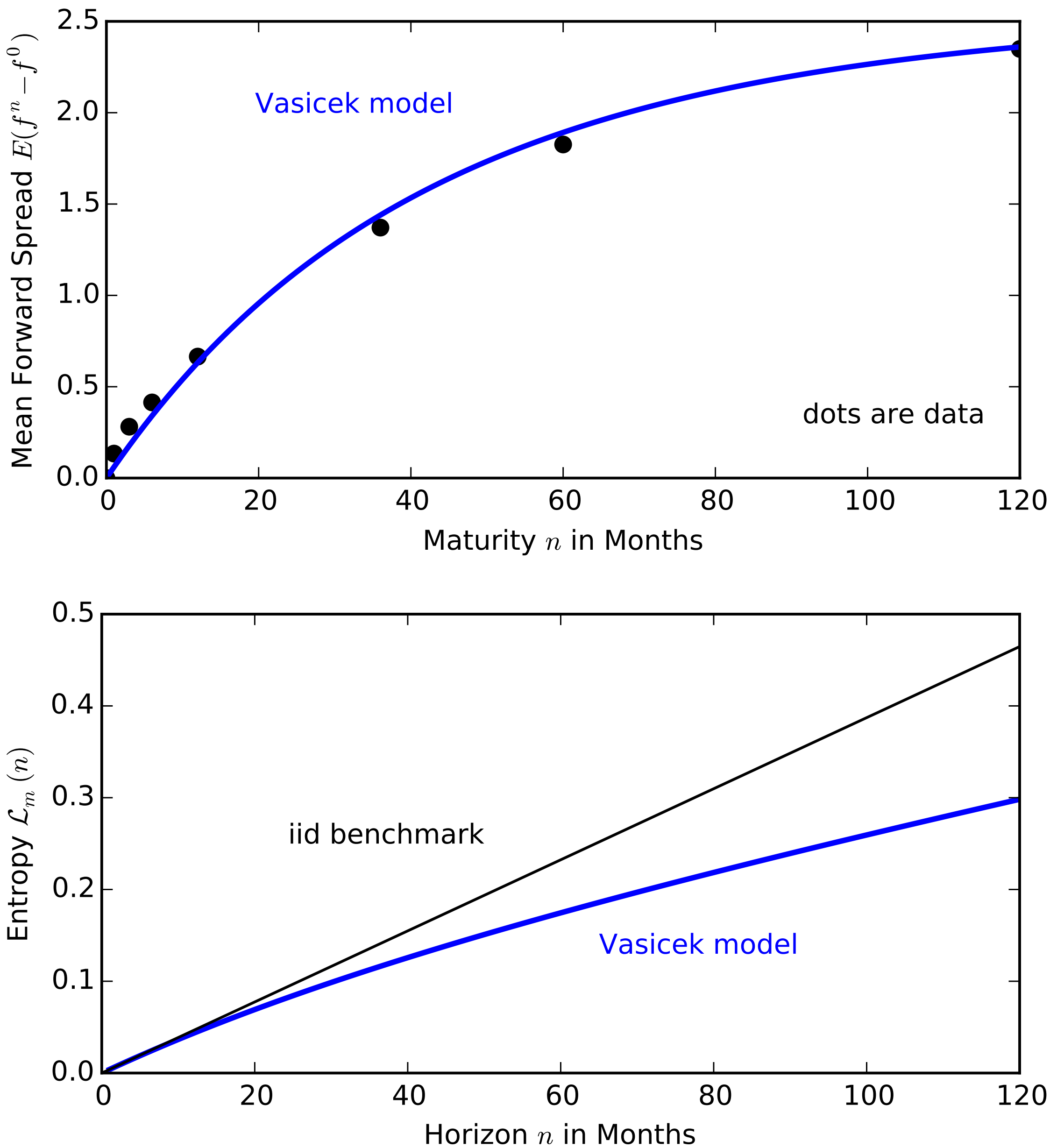

Figure 6. Moving average coefficients in the Vasicek model.

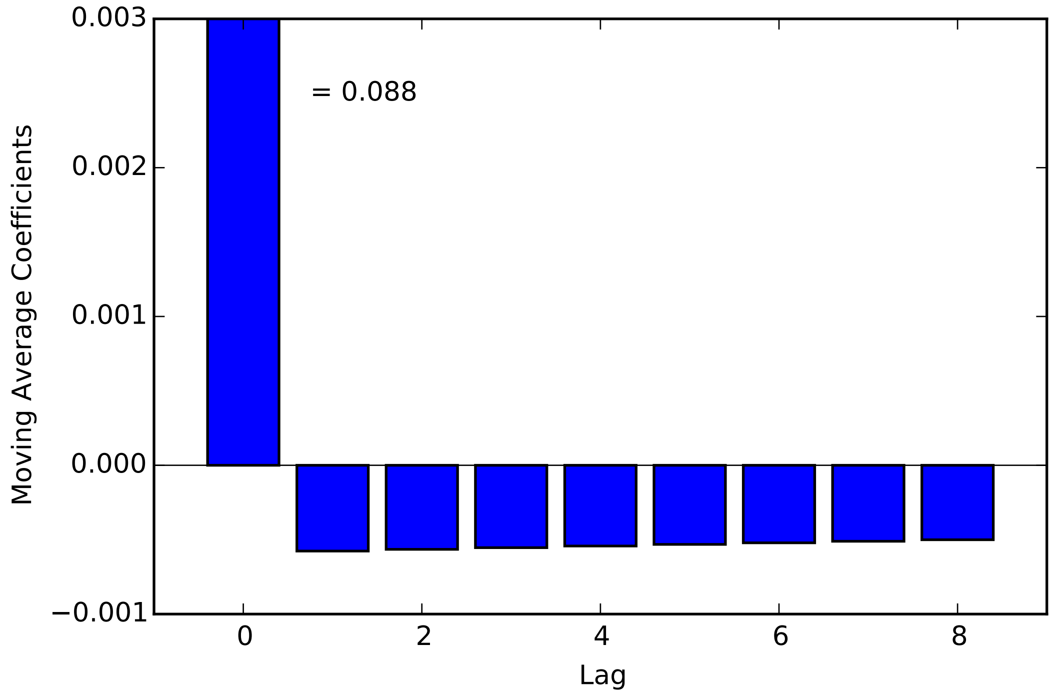

Figure 7. Dollar and "yen" forward rates. The figure shows numerical examples of forward rate curves. The solid line, labeled USD, is the Vasicek model with parameter values chosen to approximate the properties of US Treasury yields. The dashed lines are the possible JPY curves for iid depreciation with positive, zero, and negative coentropy with the dollar pricing kernel.

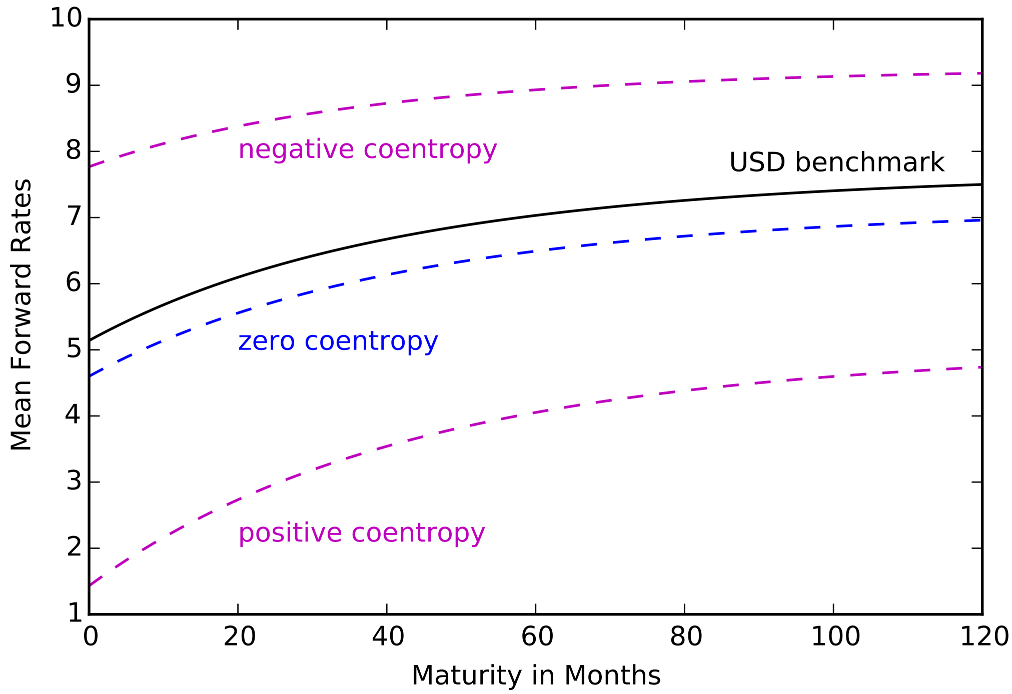

Figure 8. Dollar and "yen" forward spreads. The same forward rates as the previous figure but expressed as spreads.

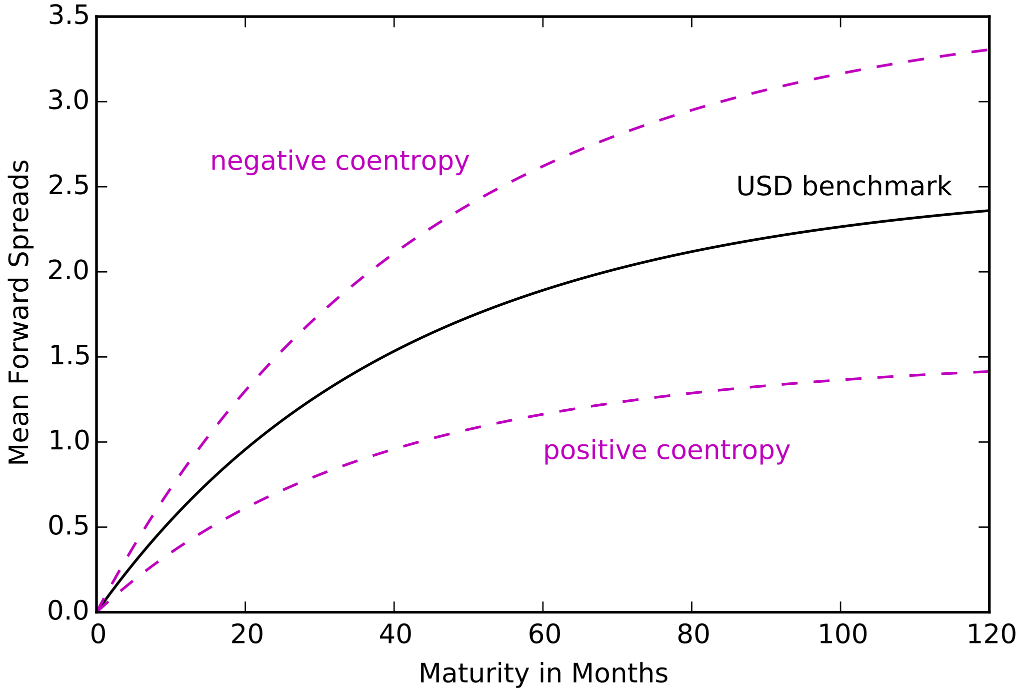

Figure 9. Equity forward spreads.

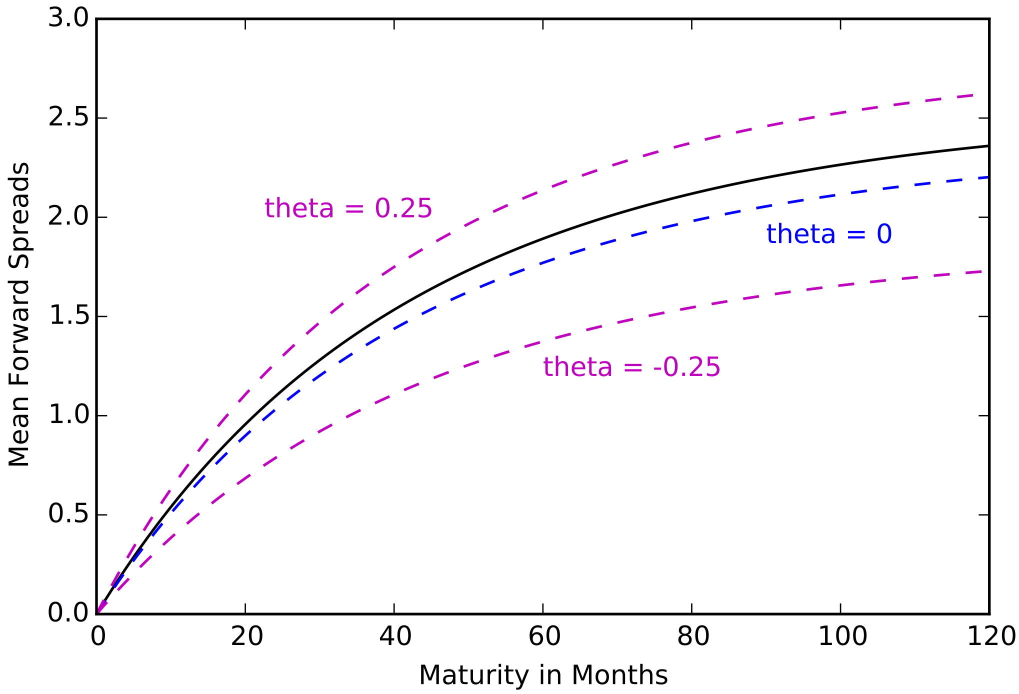

Figure 10. Equity forward differentials.