

Backus, David; Boyarchenko, Nina; Chernov, Mikhail

**Working Paper**

## Term structures of asset prices and returns

Staff Report, No. 774

**Provided in Cooperation with:**

Federal Reserve Bank of New York

*Suggested Citation:* Backus, David; Boyarchenko, Nina; Chernov, Mikhail (2016) : Term structures of asset prices and returns, Staff Report, No. 774, Federal Reserve Bank of New York, New York, NY

This Version is available at:

<http://hdl.handle.net/10419/146673>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Federal Reserve Bank of New York  
Staff Reports

# Term Structures of Asset Prices and Returns

David Backus  
Nina Boyarchenko  
Mikhail Chernov

Staff Report No. 774  
April 2016



This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

## **Term Structures of Asset Prices and Returns**

David Backus, Nina Boyarchenko, and Mikhail Chernov

*Federal Reserve Bank of New York Staff Reports*, no. 774

April 2016

JEL classification: G12, G13

### **Abstract**

We explore the term structures of claims to a variety of cash flows: U.S. 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 relationships between the two components can deliver term structures with a wide range of levels and shapes.

Key words: entropy, coentropy, term structure, yields, excess returns

---

Backus: New York University and NBER (e-mail: db3@nyu.edu). Boyarchenko: Federal Reserve Bank of New York (e-mail: nina.boyarchenko@ny.frb.org). Chernov: UCLA and CEPR (e-mail: mikhail.chernov@anderson.ucla.edu). Comments are welcome, including references to related work the authors may have inadvertently overlooked. The authors thank Jaroslav Borovička, Lars Hansen, Christian Heyerdahl-Larsen, Mahyar Kargar, Lars Lochstoer, Bryan Routledge, Raghu Sundaram, Fabio Trojani, Bruce Tuckman, Stijn Van Nieuwerburgh, Jonathan Wright, Liuren Wu, and Irina Zviadadze for comments on earlier drafts, as well as participants in seminars at and conferences sponsored by the 2015 BI-SHoF conference in Oslo, the 2014 Brazilian Finance Meeting in Recife, Carnegie Mellon University, City University of Hong Kong, the Board of Governors of the Federal Reserve System, Goethe University, ITAM, the Sixth Macro-Finance Workshop, McGill University, the 2015 NBER meeting at Stanford University, New York University, the 2014 SoFiE conference in Toronto, the Swedish House of Finance, UCLA, and the Vienna Graduate School of Finance. The latest version of this paper is available at [https://sites.google.com/site/mbchernov/BBC\\_coentropy\\_latest.pdf](https://sites.google.com/site/mbchernov/BBC_coentropy_latest.pdf).

# 1 Introduction

Perhaps the most striking recent challenge to representative agent models comes from the evidence about the term structure of risk premiums. Several papers make a forceful argument that the pattern of Sharpe ratios computed for “zero-coupon” assets across different investment horizons cannot be replicated using workhorse models, such as long-run risk, habits, or disasters (Binsbergen and Koijen, 2015 provide a comprehensive review). Usually, representative agent models offer an equilibrium-based pricing kernel and exogenously specified cash flow process for a given asset. The question is then whether the documented failure of the models comes from an equilibrium pricing kernel, cash flow specification, or both.

In this paper, we develop a methodology that allows us to consider these issues in a unified fashion accounting for term structure and cross-sectional effects at the same time. We use an illustrative affine model with regular shocks and disasters to characterize, using our methodology and basic summary statistics, the desired features of both the pricing kernel and cash flows. We subsequently develop a model with recursive preferences that, by and large, satisfies these desired properties.

Our approach is motivated by the work of Hansen, Heaton, and Li (2008), Hansen and Scheinkman (2009), and Hansen (2012) who seek to analyze the interaction of cash flows and the pricing kernel, and by Backus, Chernov, and Zin (2014) who characterize the properties of the pricing kernel alone at multiple intermediate horizons. We extend the entropy-based approach of the latter paper to the cross-section by introducing the concept of coentropy. Coentropy is a new measure of co-dependence between random variables. It serves as a natural generalization of covariance to non-normal cases and, as we show, has a useful application in asset pricing because of its connection to yield curves.

Our evidence is based on the term structures of a diverse set of assets: US dollar bonds, foreign-currency bonds, inflation-protected bonds, and equity dividend strips. These assets are claims to different cash flows, which gives their term structures different levels and shapes. The question is where do these levels and shapes come from.

Bonds provide a useful benchmark. Their cash flows are fixed, so bond prices, yields, 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 curves are mirrored by the behavior of entropy over different time horizons.

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 use 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 that the average difference between log excess buy-and-hold return on a given asset over multiple horizons and that over one period is equal to the average difference between two term spreads implied by the term structure of a given asset and by the term structure of US dollar yields. Thus, we do not need to use the data on underlying cash flows over multiple horizons, which makes computation of multi-horizon returns feasible. We report evidence on one one-period excess returns and, separately, on how they change with horizon.

We know a lot about the cross-section of one-period asset returns from the gigantic asset-pricing literature. In our limited sample, we continue to observe large cross-sectional difference in average returns and evidence of non-normality in realized returns. As investment horizon increases, the cross-sectional spread widens out, and average returns on all assets in our sample decline with horizon. Because we are working with log returns, the latter is similar to the pattern documented for Sharpe ratios of several asset classes. Finally, excess returns decline with horizon at different rates, depending on the asset. Because we are subtracting US nominal term spreads, this difference must be coming from differences in cash flows. Specifically, this indicates cross-sectional differences in the persistence of expected cash flow rates.

We use a series of affine models to show how their various elements affect the term structures of multiple assets, both in theory and in the data. We rely on our separation result and focus on modeling short-term returns and intermediate-term returns in two steps. We can do so because iid elements of a model will not affect term spreads, so we focus on getting the right magnitude of cross-sectional differences in one-period returns without worrying about persistence of expected cash flows.

We uncover three critical components that are helpful in capturing the cross-sectional and horizon dimensions of asset prices. First, in order to reflect non-normalities and to capture large magnitudes of one period excess returns, a model should feature a jump, or a disaster, component. Second, once this component is featured in a model, there is less pressure on the persistent component of a model to be high in order to match one period returns. As a result, the persistence of this component could be selected to match the shape of the yield curve. Thus, the presence of an iid jump component alleviates the tension between matching short-term returns and term structure of yields. Third, the cross-sectional differences in these term spreads are driven by the cross-sectional differences between the persistence of expected cash flows and by the difference of these persistences from the persistence of the US nominal pricing kernel.

These observations allow us to reverse engineer an example of a model featuring the representative agent with recursive preferences. Such a model delivers an equilibrium real pricing kernel. Following a big part of the literature, we assume an exogenous specification of cash flows. The key features of cash flows follow what we have uncovered in the reduced form case: iid jumps in consumption growth which allows for smaller persistence of expected consumption growth and persistence of expected cash flows that is different from that of expected consumption growth.

The models that we explore in our examples can be made more realistic, and some of the data sources could be improved, albeit with a passage of time. Thus our discussion should not be viewed as our claim to offer the definitive explanation of existing evidence. Our empirical examples are intended to be illustrative. We hope that our research offers a sufficiently clear path for further study and improvements.

## 2 Evidence

Our focus is the properties of observed term structures of prices and returns, so it is helpful to begin with data. Consider a cash-flow process  $d_t$  with growth rate  $g_{t,t+n} = d_{t+n}/d_t$  over  $n$  periods. We are interested in “zero-coupon” claims to  $g_{t,t+n}$  with a price denoted by  $\widehat{p}_t^n$ . In the special case of a claim to the cash flow of one US dollar, its price is denoted by  $p_t^n$ . We define a yield on such an asset as:  $\widehat{y}_t^n = -n^{-1} \log \widehat{p}_t^n$ . Examples include nominal risk-free bonds with  $g_{t,t+n} = 1$  (we reserve a special notation  $y_t^n \equiv n^{-1} \log r_{t,t+n}^n$  for a yield, or equivalently  $n$ -period holding period return, on a US nominal bond); foreign bonds if  $d_t$  is an exchange rate; inflation-linked bonds if  $d_t$  is price level; and equities if  $d_t$  is a dividend.

Returns are connected to yields. Consider a hold-to-maturity  $n$ -period log return:

$$\log r_{t,t+n} = \log(g_{t,t+n}/\widehat{p}_t^n) = \log g_{t,t+n} + n\widehat{y}_t^n.$$

Therefore, we can express the term spread between average returns as:

$$n^{-1}E \log r_{t,t+n} - E \log r_{t,t+1} = E(\widehat{y}_t^n - \widehat{y}_t^1).$$

Define excess holding return per period as

$$\log rx_{t,t+n} = n^{-1}(\log r_{t,t+n} - \log r_{t,t+n}^n). \tag{1}$$

Therefore, the average difference between one- and  $n$ -period excess returns is equal to difference between average term spreads:

$$E(\log rx_{t,t+n} - \log rx_{t,t+1}) = E(\widehat{y}_t^n - \widehat{y}_t^1) - E(y_t^n - y_t^1). \tag{2}$$

This connection between yields and excess returns simplifies an ordinarily difficult task: reliably computing holding period returns over long horizons. One faces declining number of

non-overlapping data points available when computing historical average of realized returns. In contrast, yields are available every period, so the number of available data points does not change with the horizon  $n$  and does not require observations of cash flows. All we need to compute is the average excess return for  $n = 1$  and then propagate it across horizons using yields.

We report summary statistics for one-period excess returns for some examples in Table 1. We choose assets for which zero-coupon approximations exist: various bonds and dividend strips. This exercise is meant to be illustrative, so we do not perform exhaustive analysis of all possible assets (see Giglio and Kelly, 2015 and Binsbergen and Koijen, 2015 for a more exhaustive list). Based on data availability, we select one quarter to be one period. We observe quite large cross-sectional dispersion in returns, on the order of 0.0136 per quarter or about 5.5 percent per year. Departures of excess returns from normality are evident despite the relatively low frequency.

Table 2 reports the yield curves and departures of term spreads from that of the US term structure. The US dollar term structure starts low, on average, reflecting low average returns on short-term default-free dollar bonds. Mean yields increase with maturity. The mean spread between one-quarter and 40-quarter yields have been about 2 percent annually.

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. Some assets have steeper yield curves, some flatter, and some have completely different shapes.

In Figure 1 we plot term spreads of US Treasury yield,  $y_t^n - y_t^1$ , and the differences between mean term spreads on a number of other assets and US Treasury yields,  $E(\hat{y}_t^n - \hat{y}_t^1) - E(y_t^n - y_t^1)$ . Because, the latter object is equal to the average difference between one- and  $n$ -period excess returns, excess returns decline with horizon in all examples with the exception of dividend strips. Moreover, there is a widening cross-sectional spread in excess returns as the horizon increases. As compared to one-quarter excess returns, the additional spread is about 1 percent extra, annually.

To summarize, the evidence points to large cross-sectional differences in excess returns. Because short-term excess returns are non-normal, part of the returns may be coming from the compensation for tail risk. The differences in returns increase with horizon, suggesting that persistence of asset yields is different from the persistence of interest rates.

All of this evidence is related to the recent literature on term structure of asset returns, such as Belo, Collin-Dufresne, and Goldstein (2015), Binsbergen, Brandt, and Koijen (2012), Binsbergen, Hueskes, Koijen, and Vrugt (2012), Boguth, Carlson, Fisher, and Simutin (2013), Boudoukh, Richardson, and Whitelaw (2015), Dahlquist and Hasseltoft (2013, 2014), Dew-Becker, Giglio, Le, and Rodriguez (2015), Doskov, Pekkala, and Ribeiro (2013), Giglio, Maggiori, and Stroebel (2015), Hasler and Marfe (2015), Lettau and Wachter (2007), Lustig, Stathopoulous, and Verdelhan (2014), and Zviadadze (2013).

### 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

$$L(x) = \log E(x) - E(\log x). \quad (3)$$

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 2. Given sufficient regularity, it has the Taylor series expansion

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

where the  $j$ th *cumulant*  $\kappa_j$  is the  $j$ th derivative of  $k(s)$  at  $s = 0$ . More concretely,  $\kappa_1$  is the mean,  $\kappa_2$  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

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

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 (4), we see



that normality gives us the variance term  $\kappa_2/2$ , but all the higher-order terms are zero ( $\kappa_j$  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  $\log x$ :

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

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 2. 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  $\theta$ . The positive contribution of high-order cumulants in this case drives entropy — the value 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:

$$C(x_1, x_2) = L(x_1x_2) - L(x_1) - L(x_2). \quad (5)$$

Appendix A shows how it is different from earlier concepts of dependence introduced in the literature. If  $x_1$  and  $x_2$  are independent, then  $L(x_1x_2) = L(x_1) + L(x_2)$  and  $C(x_1, x_2) = 0$ . If  $x_1 = ax_2$  for  $a > 0$ , then coentropy is positive. If  $x_1 = a/x_2$ , then  $L(x_1x_2) = L(a) = 0$  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  $C(x_1y, x_2) = C(x_1, x_2y) = C(x_1, x_2)$ .

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

$$C(x_1, x_2) = k(1, 1) - k(1, 0) - k(0, 1). \quad (6)$$

The cgf has the Taylor series representation

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

where  $\kappa_{ij}$  is the  $(i, j)$ th joint cumulant, the  $(i, j)$ th cross derivative of  $k$  at  $s = 0$ . Here  $\kappa_{i0}$  is the  $i$ th cumulant of  $\log x_1$ ,  $\kappa_{0j}$  is the  $j$ th cumulant of  $\log x_2$ , and  $\kappa_{ij}$  is a joint cumulant —  $\kappa_{11}$ , for example, is the covariance.

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  $\mu$  is a 2-vector and  $\Sigma$  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  $\omega$ . Conditional on  $j$  jumps,  $\log x \sim \mathcal{N}(j\theta, j\Delta)$  where the matrix  $\Delta$  has elements  $\delta_{ij}$ . The cgf is  $k(s) = \omega(e^{s^\top \theta + s^\top \Delta s/2} - 1)$ . Entropies are

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

Coentropy is therefore

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

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  $\omega$  to see what happens to the covariance and coentropy. We see in Figure 3 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 (log) *risk premium* as:  $\log E_t(r_{t,t+1}/r_{t,t+1}^1)$  where  $E_t$  is the expectation conditional on the state at date  $t$  and  $r_{t,t+1}^1$  is the one-period riskfree rate. Risk premium is closely related to expected excess returns,  $E_t \log r_{t,t+1}$ , which we've discussed earlier.

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

$$E_t(m_{t,t+1}r_{t,t+1}) = 1 \tag{7}$$

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 (7) implies

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

See Bansal and Lehmann (1997, Section 2.3) and Cochrane (1992, Section 3.2). 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 excess return is therefore bounded above by the entropy of  $m$  computed from its conditional distribution:

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

The inequality characterizes the maximum excess return 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

$$E(\log r_{t,t+1} - \log r_{t,t+1}^1) \leq E[L_t(m_{t,t+1})]. \tag{8}$$

We refer to the right side as entropy and (8) 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 (7) 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 for the (log) risk premium

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

Hansen (2012) observes that the log risk premiums can be represented as the difference between the sum of individual entropies of  $m$  and  $r$  and the entropy of their product — the first time risk premiums are linked to an idea of an entropy-based measure of co-dependence.

Average log excess returns are much easier to measure, so (7) can also be manipulated to yield

$$\begin{aligned} E_t(\log r_{t,t+1} - \log r_{t,t+1}^1) &= L_t(m_{t,t+1}) - L_t(m_{t,t+1}r_{t,t+1}) \\ &= -L_t(r_{t,t+1}) - C_t(m_{t,t+1}, r_{t,t+1}). \end{aligned} \quad (9)$$

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

$$\begin{aligned} E(\log r_{t,t+1} - \log r_{t,t+1}^1) &= E[L_t(m_{t,t+1})] - E[L_t(m_{t,t+1}r_{t,t+1})] \\ &= -E[L_t(r_{t,t+1})] - E[C_t(m_{t,t+1}, r_{t,t+1})]. \end{aligned} \quad (10)$$

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. The “extra” term  $E[L_t(r_{t,t+1})]$  reflects a generalization of the usual convexity adjustment that appears in the log-normal case. As a result, idiosyncratic dynamics may be helpful in matching observed log excess returns. One has to be mindful of this when interpreting a model’s ability to explain the evidence.

Equation (10) 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 (10); the right side gives us an interpretation of it. Backus, Chernov, and Zin (2014) estimate that the upper bound on expected excess returns is at least 3 percent quarterly. 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 (10) then reproduces the entropy bound (8). 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 excess return 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 excess return 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:  $E_t(r_{t,t+1} - r_{t,t+1}^1)$ . A risk premium defined this way is connected, via (7), to the covariance of the pricing kernel and the return:

$$\begin{aligned} E_t(r_{t,t+1} - r_{t,t+1}^1) &= -\text{Cov}_t(m_{t,t+1}, r_{t,t+1} - r_{t,t+1}^1)/E_t(m_{t,t+1}) \\ &= -\text{Cov}_t(m_{t,t+1}, r_{t,t+1})/E_t(m_{t,t+1}). \end{aligned} \quad (11)$$

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:

$$E_t(r_{t,t+1} - r_{t,t+1}^1)/\text{Var}_t(r_{t,t+1} - r_{t,t+1}^1)^{1/2} \leq \text{Var}_t(m_{t,t+1})^{1/2}/E_t(m_{t,t+1}). \quad (12)$$

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

$$\text{Var}_t(m_{t,t+1})^{1/2}/E_t(m_{t,t+1}) = \left( e^{k_t(2)} - 2k_t(1) - 1 \right)^{1/2}. \quad (13)$$

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

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

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

$$\log m_{t,t+1} = \log \beta + a^\top x_t + b^\top x_{t+1} \quad (14)$$

$$x_{t+1} = Ax_t + Bw_{t+1}, \quad (15)$$

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

$$\log m_{t,t+1} = \log \beta + (a^\top + b^\top A)x_t + b^\top Bw_{t+1} = \log \beta + \theta_m^\top x_t + \lambda^\top w_{t+1}. \quad (16)$$

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  $E[L_t(m_{t,t+1})] = L_t(m_{t,t+1}) = b^\top B B^\top b/2 = \lambda^\top \lambda/2$ . The Vasicek model is special case when  $x$  and  $w$  are one-dimensional.

*Example 6 (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

$$\log m_{t,t+1} = \log \beta - (\lambda_0 + \lambda_1 x_t)^2/2 + \theta_m x_t + (\lambda_0 + \lambda_1 x_t)w_{t+1} \quad (17)$$

$$x_{t+1} = \varphi x_t + w_t, \quad (18)$$

with  $\{w_t\}$  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. We 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

In an arbitrage-free setting, bond prices inherit their properties from the pricing kernel. Pricing has a simple recursive structure. Applying the pricing relation (7) to bond returns gives us

$$p_t^n = E_t(m_{t,t+1}p_{t+1}^{n-1}) = E_t(m_{t,t+n}), \quad (19)$$

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 (19) 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:

$$E[\log E_t(m_{t,t+n})] = -nE(y^n).$$

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 Backus, Chernov, and Zin (2014) call *horizon dependence*, the relation between entropy and the time horizon represented by the function  $\mathcal{L}_m(n)$ . In the term structure context, this function maps directly to mean yields. If one-period pricing kernels  $\{m_{t,t+1}\}$  are iid, entropy is proportional to  $n$ . Bond yields are then the same at all maturities and constant over time. Differences from this proportional benchmark reflect dynamics in the pricing kernel. Horizon dependence is defined as:

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

The connection with bond yields then gives us  $\mathcal{H}_m(n) = -E(y^n - y^1)$ .

In the iid case,  $\mathcal{H}_m(n) = 0$  and the yield curve is flat. If the mean yield curve slopes upwards, then  $\mathcal{H}_m(n)$  is negative and slopes downward. One implication of this result is that iid components of  $m$  will affect only the level of the yield curve, but not its shape.

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

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

If the coentropy of successive one-period pricing kernels is zero, then horizon dependence is zero as well. Borovicka and Hansen (2014, section 3) characterize this intertemporal dependence via an entropy counterpart to an impulse response.

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

*Example 5 (Markov pricing kernel, continued).* Bond prices follow from the pricing kernel (16), the transition equation (15), and the pricing relation (7). They imply bond prices of the form  $\log q^n(x) = a_n + b_n^\top x$  with coefficients  $(a_n, b_n)$  satisfying

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

starting with  $a_0 = b_0 = 0$ . Entropy is therefore

$$\mathcal{L}_m(n) = E(\log q^n - n \log m) = a_n - n \log \beta = \sum_{j=0}^{n-1} k(\lambda + B^\top b_j).$$

The iid case is a useful benchmark:  $\theta_m = 0$ , the mean yield curve is flat,  $\mathcal{L}_m(n) = nk(\lambda)$ , and  $\mathcal{H}_m(n) = 0$ . Any departure from proportionality in entropy  $\mathcal{L}_m(n)$  is evidence against this case. The  $n$ -period Hansen-Jagannathan upper bound (13) is then

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

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  $n$ .

Thus, entropy conveys term structure effects in a more intuitive fashion. Figure 4 compares Sharpe ratios with entropies for the iid and non-iid cases at different horizons. The dashed lines show departures from iid for the Vasicek model. Departures from iid are evident in the case of entropy.

*Example 6 (state-dependent price of risk, continued).* Recall the model consisting of pricing kernel (17) and transition equation (18). (The Vasicek model is a special case with  $\lambda_1 = 0$ .) Bond prices satisfy  $\log p^n(x) = a_n + b_n x$  with

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

where  $a_0 = b_0 = 0$  and  $\varphi^* = \varphi + \lambda_1$ . In particular, one-period yield is

$$y_t^1 = -\log p^1(x_t) = -\log \beta - \theta_m x_t. \quad (20)$$

Horizon dependence is

$$\mathcal{H}_m(n) = n^{-1}a_n - a_1 = n^{-1} \left[ \lambda_0 \sum_{j=0}^{n-1} b_j + 1/2 \sum_{j=0}^{n-1} b_j^2 \right]. \quad (21)$$

## 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.

We value these assets in the usual way. The pricing relation (7) gives us

$$\widehat{p}_t^n = E_t(m_{t,t+1}g_{t,t+1}\widehat{p}_{t+1}^{n-1}) = E_t(\widehat{m}_{t,t+1}\widehat{p}_{t+1}^{n-1}) = E_t(\widehat{m}_{t,t+n}), \quad (22)$$

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  $\widehat{p}_t^0 = 1$ . This has the same form as the bond pricing equation (22), 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 curves. By analogy with equation (10), we can show, using equation (1), that

$$\begin{aligned} nE \log r x_{t,t+n} &= E(\log r_{t,t+n} - \log r_{t,t+n}^n) \\ &= \mathcal{L}_m(n) - \mathcal{L}_{\widehat{m}}(n) = -\mathcal{L}_g(n) - \mathcal{C}_{mg}(n), \end{aligned} \quad (23)$$

where  $\mathcal{C}_{mg}(n)$  is a notation for  $E[C_t(m_{t,t+n}, g_{t,t+n})]$ . This expression shows how 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  $g$ . The difference between  $\mathcal{L}_{\widehat{m}}(n)$  and  $\mathcal{L}_m(n)$ , and therefore average excess returns, 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 multi-period extension of our earlier claim: that mean excess returns reflect the entropy of the return and the coentropy of the return and the pricing kernel.

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

$$\log g_{t,t+1} = \log \gamma + \theta_g^\top x_t + \eta^\top w_{t+1}.$$

The transformed pricing kernel is then

$$\begin{aligned} \log \widehat{m}_{t,t+1} &= \log m_{t,t+1} + \log g_{t,t+1} \\ &= (\log \beta + \log \gamma) + (\theta_m + \theta_g)^\top x_t + (\lambda + \eta)^\top w_{t+1} \\ &= \log \widehat{\beta} + \widehat{\theta}_m^\top x_t + \widehat{\lambda}^\top w_{t+1}. \end{aligned}$$



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

Combining equation (2) with the definition of horizon dependence, we see that the term difference in log excess return on an asset is equal to:

$$E(\log rx_{t,t+n} - \log rx_{t,t+1}) = \mathcal{H}_m(n) - \mathcal{H}_{\hat{m}}(n).$$

Combining this with equation (23), we can characterize how coentropy changes with horizon:

$$n^{-1}\mathcal{C}_{mg}(n) - \mathcal{C}_{mg}(1) = \mathcal{H}_{\hat{m}}(n) - \mathcal{H}_m(n) - \mathcal{H}_g(n) \quad (24)$$

This expression can also be obtained from the definitions of coentropy and horizon dependence. In words, the difference between the  $n$ -period and one-period coentropies is equal to the differences between the horizon dependence of the transformed pricing kernel and those of its two constituents: the pricing kernel and cash flows.

*Example 5 (Markov pricing kernel, continued).* With cash flow growth of

$$\log g_{t,t+1} = \log \gamma + \theta_g^\top x_t + \eta^\top w_{t+1}$$

we can compute its horizon dependence similarly to bond prices:  $\log E_t(g_{t,t+n})(x) = a_{gn} + b_{gn}^\top x$  with coefficients  $(a_{gn}, b_{gn})$  satisfying

$$\begin{aligned} a_{gn+1} &= a_{gn} + \log \gamma + k(\eta + B^\top b_{gn}) \\ b_{gn+1} &= \theta_g^\top + b_{gn}^\top A = \theta_g^\top (I + A + \dots + A^n) \end{aligned}$$

starting with  $a_{g0} = b_{g0} = 0$ . Entropy is therefore

$$\mathcal{L}_g(n) = E(\log E_t(g_{t,t+n}) - n \log g_{t,t+1}) = a_{gn} - n \log \gamma = \sum_{j=0}^{n-1} k(\eta + B^\top b_{gj}),$$

horizon dependence of cash flows is

$$\mathcal{H}_g(n) = n^{-1} \sum_{j=0}^{n-1} [k(\eta + B^\top b_{gj}) - k(\eta)],$$

and coentropy changes with horizon according to

$$\begin{aligned} \mathcal{C}_{mg}(n) - n\mathcal{C}_{mg}(1) &= \sum_{j=0}^{n-1} [k(\lambda + \eta + B^\top (b_j + b_{gj})) - k(\lambda + B^\top b_j) - k(\eta + B^\top b_{gj})] \\ &\quad - n[k(\lambda + \eta) - k(\lambda) - k(\eta)]. \end{aligned}$$

### 4.3 Long horizons

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  $\nu$  and associated positive eigenfunction  $v_t$  satisfying

$$E_t(m_{t,t+1}v_{t+1}) = \nu v_t. \quad (25)$$

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

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

By construction  $E_t(m_{t,t+1}^1) = 1$ , hence Hansen and Scheinkman (2009) refer to it as a martingale component of the pricing kernel. Qin and Linetsky (2015) demonstrate how this decomposition works in non Markovian environments.

Given such an eigenvalue-eigenfunction pair, the long yield converges to  $-\log \nu$ . The long bond one-period return is not constant, but its expected value also converges:  $r_{t,t+1}^\infty = \lim_{n \rightarrow \infty} r_{t,t+1}^n = 1/m_{t,t+1}^2 = v_{t+1}/(\nu v_t)$ , so that  $E(\log r^\infty) = -\log \nu$ . See Alvarez and Jermann (2005, Section 3).

The special case  $m_{t,t+1}^1 = 1$  has gotten a lot of recent attention; see, for example, the review in Borovicka, Hansen, and Scheinkman (2014). The pricing kernel becomes  $m_{t,t+1} = m_{t,t+1}^2$ . Since the long bond return is its inverse, the long bond is the high return asset. Realistic or not, it's an interesting special case. In logs, the pricing kernel becomes

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

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 (25) we find:

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

If  $b = -a$ , then  $c = a$ ,  $\log \nu = \log \beta$ , and  $m_{t,t+1}^1 = 1$ .

Moving on to other assets, we introduce two equation analogous to (25). One is for cashflow growth:

$$E_t(g_{t,t+1}u_{t+1}) = \xi u_t$$

leading to a decomposition  $g_{t,t+1} = \xi g_{t,t+1}^1 u_t / u_{t+1}$ . The other is for transformed pricing kernel:

$$E_t(\widehat{m}_{t,t+1} \widehat{v}_{t+1}) = \widehat{v}_t. \quad (26)$$

leading to a decomposition  $\widehat{m}_{t,t+1} = \widehat{v} \widehat{m}_{t,t+1}^1 \widehat{v}_t / \widehat{v}_{t+1}$ . These decompositions allow us to characterize behavior of coentropy at long horizons. Using the definition of coentropy and exploiting stationarity of  $v_t$ ,  $\widehat{v}_t$ , and  $e_t$  we obtain

$$n^{-1} \mathcal{C}_{mg}(n) \rightarrow \log \widehat{v} - \log \nu - \log \xi, \quad \text{as } n \rightarrow \infty. \quad (27)$$

The decompositions are related to each other via:

$$\widehat{v} \widehat{m}_{t,t+1}^1 \widehat{v}_t / \widehat{v}_{t+1} = \widehat{m}_{t,t+1} \equiv m_{t,t+1} g_{t,t+1} = \nu \xi m_{t,t+1}^1 g_{t,t+1}^1 (v_t u_t) / (v_{t+1} u_{t+1}). \quad (28)$$

There's not, in general, a close relation between  $\widehat{v}$ ,  $\nu$ , and  $\xi$ , but there is in some special cases. One special case is a stationary cash flow  $d$ , which leads to the martingale component  $g_{t,t+1}^1 = 1$  as in the example above. In this case, the simplified equation (28) implies that the value  $\nu \xi$  and function  $v_t u_t$  solve equation (26). Therefore,  $\widehat{v} = \nu \xi$ , the martingale components coincide,  $\widehat{m}_{t,t+1}^1 = m_{t,t+1}^1$ , long-horizon coentropy is equal to zero, and so are long-horizon excess returns:

$$E \log r x_{t,t+n} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (29)$$

The reverse is also true: if  $\widehat{m}_{t,t+1}^1 = m_{t,t+1}^1$ , it must be the case that  $g_{t,t+1}^1 = 1$ . Indeed, in this case equation (28) implies that the level of  $g_{t,t+1}^1 (v_t u_t) / (v_{t+1} u_{t+1})$  must be stationary because  $\widehat{v}_t$  is. Because  $v_t$  and  $u_t$  are stationary as well, the martingale  $g_{t,t+1}^1$  must be a constant (we can normalize it to one w.l.o.g.).

*Example 5 (Markov pricing kernel, continued).* We revert to the original Markov pricing kernel, equation (14), and posit cash flow growth of

$$\log g_{t,t+1} = \log \gamma + a_g^\top x_t + b_g^\top x_{t+1}. \quad (30)$$

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 (14). The Perron-Frobenius theory implies  $\log u_t = c_g^\top x_t$  with

$$c_g^\top = (a_g^\top + b_g^\top A)(I - A)^{-1}, \quad \log \xi = \log \gamma + k \left( B^\top (b_g + c_g) \right).$$

and  $\log \widehat{v}_t = \widehat{c}^\top x_t$  with

$$\widehat{c}^\top = (\widehat{a}^\top + \widehat{b}^\top A)(I - A)^{-1}, \quad \log \widehat{v} = \log \widehat{\beta} + k \left( B^\top (\widehat{b} + \widehat{c}) \right).$$

If  $b_g = -a_g$ , then  $d_t$  is stationary,  $\log \xi = \log \gamma$ , and  $\log \hat{\nu} = \log \beta + \log \gamma + k \left( B^\top (b + c) \right) = \log \nu + \log \xi$ .

Another special case is one in which the “price-dividend” ratio  $\hat{p}$  is constant, see the October 2005 version of Hansen, Heaton, and Li (2008), section 3.2. Consider a factorization of the dividend into a growth component  $d_t^*$  and a stationary component  $s_t$ , so that  $d_t = d_t^* \cdot s_t$ , and  $g_{t,t+1}^* \equiv d_{t+1}^*/d_t^*$  (if  $g_{t,t+1}^*$  is a constant, then  $g_{t,t+1}^1 = 1$ .) Because  $s_t$  is stationary, the two transformed pricing kernels  $\hat{m}_{t,t+1}$  and  $m_{t,t+1}^* \equiv m_{t,t+1} g_{t,t+1}^*$  will have the same eigenvalue  $\hat{\nu}$ . The eigenfunctions will be  $\hat{v}_t$  and  $\hat{v}_t \cdot s_t$ , respectively. Thus, if a dividend is such that its  $\hat{v}_t = 1$ , or, equivalently,  $s_t$  equals the eigenfunction associated with  $m_{t,t+1}^*$ , then  $\hat{p}$  is constant. Long-horizon entropy is still going to be as in (27) because long-run properties are affected by eigenvalues, not eigenfunctions.

## 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 with time-varying risk premium: example 6 with normal innovations. We use properties of the US nominal Treasury data described in Tables 1 and 2. At a quarterly frequency the short rate  $y_t^1$  in equation (20) has a standard deviation of 0.0084 and an autocorrelation of 0.9487. The mean of the 40-quarter yield spread  $y^{40} - y^1$  is 0.0045, or, equivalently, horizon dependence in equation (21) is  $-0.0045$ . We reproduce each of these features by choosing the parameter values  $\theta_m = 0.0026$ ,  $\varphi = 0.9487$ , and  $\lambda_0 = -0.1225$ . The parameter controlling time variation in risk premium is set to match the curvature of the yield curve. Typically, this results in  $\varphi^*$  being very close to 1. We set it to 0.9999 implying the value of  $\lambda_1 = 0.0512$ . All of these values are summarized in Panel A of Table 3. 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  $\theta_m$  and  $\varphi$  control the variance and autocorrelation of the short rate and  $\lambda_0$  controls the slope of the mean yield curve. The different signs of  $\theta_m$  and  $\lambda_0$  produce the upward slope in the mean yield curve. The difference in absolute values of  $\lambda_0$  and  $\theta_m$  — the former is roughly two orders of magnitude greater — implies a large entropy and small horizon dependence. This allows us to generate large one-period excess returns and small departures from them as the horizon changes.

## 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. Further, departures from normality documented in Table 1 cannot be captured with a normal innovation.

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

$$\begin{aligned}
 \log m_{t,t+1} &= \log \beta + \theta_m x_{1t} - (\lambda_0 + \lambda_1 x_{1t})^2/2 + (\lambda_0 + \lambda_1 x_{1t})w_{t+1} + \lambda_2 z_{t+1}^m, \\
 \log g_{t,t+1} &= \log \gamma + \theta x_{1t} + \theta_g x_{2t} + \eta_0 w_{t+1} + \eta_2 z_{t+1}^g, \\
 x_{1t+1} &= \varphi_1 x_{1t} + w_{t+1}, \\
 x_{2t+1} &= \varphi_2 x_{2t} + w_{t+1}.
 \end{aligned} \tag{31}$$

with  $w_t \sim \mathcal{N}(0, 1)$  and  $z_t^m$  and  $z_t^g$  are compound Poisson process with the same arrival rate of  $\omega$  and jump size distributions of  $\mathcal{N}(\mu_m, \delta_m^2)$  and  $\mathcal{N}(\mu_g, \delta_g^2)$ , respectively.

The added disturbance  $z^m$  is designed to capture pricing of the disaster risk. It is iid, so it has no impact on US nominal bond prices, but potentially plays a role in the pricing of claims to cash flow growth  $g$ . By varying the weights  $(\eta_0, \eta_2)$  we can alter the dependence of stock and bond returns. Setting  $\varphi_1 = \varphi_2 = \varphi$  recovers the Vasicek model with time-varying risk premiums. Figure 1 suggests differences between  $\varphi_1$  and  $\varphi_2$ , and between  $\varphi_2$ 's of different assets.

Afficionados of careful bond curve modeling would prefer to see separate shocks driving  $x_{1t}$  and  $x_{2t}$  but we intentionally limit ourselves to one normal and one Poisson shocks in order to highlight the most critical features a model needs to capture the key facts.

As far as the US pricing kernel is concerned, this is the same model as in example 6 with an added iid jump component. Thus, this addition does not affect horizon dependence in equation (21). What's affected is entropy of the pricing kernel:

$$\mathcal{L}_m(1) = \lambda_0^2/2 + \lambda_1^2(1 - \varphi_1^2)^{-1}/2 - \omega \lambda_2 \mu_m + \omega \left( e^{\lambda_2 \mu_m + \lambda_2^2 \delta_m^2/2} - 1 \right).$$

Given that, it is easy to compute  $n$ -period entropy via  $\mathcal{L}_m(n) = n(\mathcal{L}_m(1) + \mathcal{H}_m(n))$ .

The transformed pricing kernel has a similar structure:

$$\begin{aligned}
 \log \widehat{m}_{t,t+1} &= \log m_{t,t+1} + \log g_{t,t+1} \\
 &= (\log \beta + \log \gamma) + (\theta_m + \theta)x_{1t} + \theta_g x_{2t} - (\lambda_0 + \lambda_1 x_{1t})^2/2 \\
 &\quad + (\lambda_0 + \eta_0 + \lambda_1 x_{1t})w_{t+1} + \lambda_2 z_{t+1}^m + \eta_2 z_{t+1}^g.
 \end{aligned} \tag{32}$$

Asset prices are easily computed by the same approach we used with Vasicek. In particular, we guess the (log) bond price to be a linear function of  $x_t$  :

$$\log \widehat{p}_t^n = \widehat{a}_n + \widehat{b}_n x_{1t} + \widehat{c}_n x_{2t}.$$

Then, following the same steps as before, we get

$$\begin{aligned} \widehat{c}_n &= \theta_g \frac{1 - \varphi_2^n}{1 - \varphi_2} \\ \widehat{b}_n &= \left( \theta^* + \frac{\theta_g \lambda_1}{1 - \varphi_2} \right) \frac{1 - \varphi_1^{*n}}{1 - \varphi_1^*} - \frac{\theta_g \lambda_1}{1 - \varphi_2} \frac{1 - (\varphi_2/\varphi_1^*)^n}{1 - \varphi_2/\varphi_1^*} \varphi_1^{*n-1} \\ \widehat{a}_n &= \log \beta + \log \gamma + \eta_0 \lambda_0 + \eta_0^2/2 + k_z(\lambda_2, \eta_2) + \widehat{a}_{n-1} \\ &\quad + (\widehat{b}_{n-1} + \widehat{c}_{n-1})^2/2 + (\widehat{b}_{n-1} + \widehat{c}_{n-1})(\lambda_0 + \eta_0) \end{aligned}$$

with  $\theta^* = \theta + \theta_m + \eta_0 \lambda_1$ ,  $\varphi_1^* = \varphi_1 + \lambda_1$ , and  $k_z(s_1, s_2) = \omega(e^{s_1 \mu_m + s_2 \mu_g + (s_1 \delta_m + s_2 \delta_g)^2/2} - 1)$ . Horizon dependence is

$$\mathcal{H}_{\widehat{m}}(n) = n^{-1} \widehat{a}_n - \widehat{a}_1 = n^{-1} \left[ (\lambda_0 + \eta_0) \sum_{j=0}^{n-1} (\widehat{b}_j + \widehat{c}_j) + 1/2 \sum_{j=0}^{n-1} (\widehat{b}_j + \widehat{c}_j)^2 \right]. \quad (33)$$

Horizon dependence of cash flows is computed similarly (see example 5):

$$\mathcal{H}_g(n) = n^{-1} \left[ \eta_0 \sum_{j=0}^{n-1} (b_{gj} + c_{gj}) + 1/2 \sum_{j=0}^{n-1} (b_{gj} + c_{gj})^2 \right],$$

where

$$b_{gn} = \theta \frac{1 - \varphi_1^n}{1 - \varphi_1}, \quad c_{gn} = \theta_g \frac{1 - \varphi_2^n}{1 - \varphi_2}.$$

One-period coentropy is

$$\mathcal{C}_{mg}(1) = \lambda_0 \eta_0 + k_z(\lambda_2, \eta_2) - k_z(\lambda_2, 0) - k_z(0, \eta_2).$$

Equation (24) implies the  $n$ -period one.

This model has a triangular structure, in which  $(\theta_m, \varphi_1, \lambda_0, \lambda_1)$  control bond prices, and  $(\theta_g, \eta_0, \eta_2, \lambda_2)$  control the return on the cash flow  $g$  and its relation to bond returns. This 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.

For symmetry between the US and other economies, and for simplicity of calibration we assume that  $\theta = -\theta_m - \lambda_1\eta_0$  (so that  $\theta^* = 0$ ). As a result, one-period yield is

$$\widehat{y}_t^1 = -\log \beta - \log \gamma - \lambda_0\eta_0 - \eta_0^2/2 - k_z(\lambda_2, \eta_2) - \theta_g x_{2t}.$$

Thus, asset-specific parameters  $\varphi_2$ , and  $\theta_g$  are calibrated by analogy with US nominal bonds using serial correlation and variance of the one-period yields. Then, one can use the term spread of the foreign curve to back out  $\lambda_0 + \eta_0$  from equation (33). Because we already know  $\lambda_0$  from the US curve, we can determine  $\eta_0$ . Panel B of Table 3 lists the calibrated values.

We observe quite dramatic difference in  $\varphi_2$ 's across the different countries. The volatility  $\theta_g$  and risk premium  $\lambda_0 + \eta_0$  retain the same qualitative features as their US counterparts: they have different signs, and the former is much smaller than the latter. Quantitatively, we observe cross-sectional variation in both parameters.

The literature views foreign exchange rates as being close to random walk. In our model this would mean  $\theta_g = 0$ , and  $\theta = 0$ . Such a value would imply  $\widehat{c}_n = 0$  and  $\widehat{b}_n = (\theta_m + \eta_0\lambda_1)(1 - \varphi_1^{*n})/(1 - \varphi_1^*)$ . Thus, the foreign term spread will be (approximately) a scaled version of the US term spread, which contradicts the evidence.

We were able to characterize the properties of the US and foreign yield curves without discussing the Poisson parameters. This is because disasters have iid distribution in the model.

To calibrate the jump parameters, we normalize jump loadings  $\lambda_2$  and  $\eta_2$  to 1 because they are not identified separately from jump volatilities  $\delta_m$  and  $\delta_g$ , respectively. We borrow parameters controlling jumps in the pricing kernel from Backus, Chernov, and Zin (2014), the CI2 model:  $\omega = 0.01/4$ ,  $\mu_m = -10 \cdot (-0.15) = 1.5$ ,  $\delta_m^2 = (-10 \cdot 0.15)^2 = 1.5^2$ . We can use information about cash flows, or, equivalently, about one-period excess returns to infer asset-specific  $\eta_2$ , and  $\mu_g$ . One-period excess (log) returns are:

$$\begin{aligned} \log r x_{t,t+1} &= \log g_{t,t+1} + \widehat{y}_t^1 - y_t^1 \\ &= -\lambda_0\eta_0 - \eta_0^2/2 - k_z(\lambda_2, \eta_2) + k_z(\lambda_2, 0) - \lambda_1\eta_0 x_{1t} + \eta_0 w_{t+1} + \eta_2 z_{t+1}^g. \end{aligned}$$

Thus,

$$E \log r x_{t,t+1} = -\lambda_0\eta_0 - \eta_0^2/2 - k_z(\lambda_2, \eta_2) + k_z(\lambda_2, 0) + \omega\eta_2\mu_g$$

and

$$\text{var}(\log rx_{t,t+1}) = [\lambda_1^2(1 - \varphi_1^2)^{-1} + 1]\eta_0^2 + \omega\eta_2^2(\mu_g^2 + \delta_g^2).$$

The variance of the normal component, that is, the first element of the sum, must evidently be no greater than the observed variance. Variants of our model with  $\varphi_1^* = \varphi_1$ , or  $\varphi_1 = \varphi_2$  does not have this property if parameters are calibrated to the yield curves. In other words, the persistence structure implied by these restrictions is so rigid that the values of  $\eta_0$  inferred from the yield curves are much larger than those implied by the time-series of excess returns even assuming no disaster component. The combination of German yields and the Euro is an exception in that a model with  $\varphi_1^* = \varphi_1$  does not feature this tension.

Table 3B reports the results. The non-normality manifests itself in the differences between coentropy and covariance that we discussed in example 4. The differences are substantive highlighting an ability of non-normal models to generate large expected returns and large cross-sectional difference between them.

As a reality check, we verify if the calibrated process for exchange rates,  $\log g$  resembles the data. We focus on two basic summary statistics: variance and serial correlation (mean can be mechanically matched by adjusting  $\log \gamma$ ). We use the model to compute the population values of these two statistics at calibrated parameters. Further, we simulate 100,000 artificial histories of the respective exchange rates which allows us to compute finite-sample distribution of the same two statistics. Table 4 compares these theoretical results with empirical values. We see that theoretical values are sufficiently close to the data.

Figure 5 displays the term structure of coentropies, a difference between the  $n$ -period and one-period ones. Given that a negative of coentropy reflects risk premium, this figure tells us about cross-sectional differences of how risk premiums change with horizon. For Australia and the UK, risk premiums continue to increase. The increase has a similar magnitude. In the case of Germany, they increase out to 11 quarters but not much compared to the other two countries, and then they start to decline. These differences reflect the differences in the persistence coefficient  $\varphi_2$ .

Lustig, Stathopolous, and Verdelhan (2014) study log excess returns on a strategy that borrows via an  $n$ -period US bond, converts into foreign currency, invests in an  $n$ -period foreign bond, and then unwinds in one period. In our notation, this would be:

$$\begin{aligned} E \log rx_{t,t+1}^n &= E[\log g_{t,t+1} + (\log \hat{p}_{t+1}^{n-1} - \log \hat{p}_t^n) - (\log p_{t+1}^{n-1} - \log p_t^n)] \\ &= E \log rx_{t,t+1} - (n\mathcal{H}_{\hat{m}}(n) - (n-1)\mathcal{H}_{\hat{m}}(n-1)) \\ &\quad + (n\mathcal{H}_m(n) - (n-1)\mathcal{H}_m(n-1)). \end{aligned}$$

So, their object of interest contains elements of both one-period and  $n$ -period holding returns.



They find that at long maturities  $n$  the average excess return is negative, but not significantly different from zero. The long horizon results from section 4.3 and the first line of the equation imply that

$$E \log rx_{t,t+1}^\infty = \log \xi - \log \hat{\nu} + \log \nu + E \log g_{t,t+1}^1.$$

If exchange rate is stationary, then  $g_{t,t+1}^1 = 1$ , and  $E \log rx_{t,t+1}^\infty = 0$ . As we noted in section 4.3, this is equivalent to  $\hat{m}_{t,t+1}^1 = m_{t,t+1}^1$  – a condition highlighted in Proposition 3 of Lustig, Stathopoulos, and Verdelhan (2014). Thus, the modern language of the pricing kernel decomposition translates into the old question of stationarity of nominal exchange rates.

## 5.4 Inflation-linked bonds

Analysis of inflation-linked bonds is very similar to the foreign ones. Exchange rates and foreign bonds tell us about transitions between domestic and foreign economies. The price level (CPI) and TIPS tell us about transition between the real and nominal economy. For this reason, we use exactly the same model and the same calibration strategy in this case. We maintain the same US nominal pricing kernel, so calibration of the cash flow growth, or inflation in this case, is the only novel part relative to the previous section. The results are reported in the first line of Table 3B. Figure 5 shows term structure of coentropy – it is similar to that of Germany.

The key difference from the foreign-bond case is the highlighted tension in calibrating  $\eta_0$  that we were not able to resolve. The reason is extremely low volatility of returns associated with trading TIPS at quarterly frequency. Table 1 shows that it is two orders of magnitude smaller than those of foreign bonds. Table 2 shows that the difference in the term spreads of TIPS and US nominal bonds is right in the middle of those for foreign bonds. Hence, the time-series and term structure information about  $\eta_0$  are in conflict under the null of our model. The figures in Table 4 reflect the model’s difficulty in capturing variance and serial correlation of inflation – potentially a manifestation of the same issue.

Perhaps, one could suggest a more elaborate model that would be able to reconcile these facts. We were hesitant to do so because these numbers could be an outcome of poor quality of data, especially at the short end of the curve. As is well known, the TIPS data are considered reliable after 2003. The data prior to 2003 are extrapolated by Chernov and Mueller (2012) using their preferred model. TIPS experienced distorted prices during the credit crisis, so the yields of maturities of up to eight quarters had to be discarded during the last three quarters of 2008. Thus, we leave more refined analysis of inflation-linked bonds for future research.

## 5.5 Equity

Dividend strips have attracted recent interest in the literature, as the term structure of associated Sharpe ratios seems to offer prima facie evidence against major asset-pricing

models. We study excess log returns instead of Sharpe ratios, but it is clear that these are related objects by comparing equations (8), (9) and (11), (12).

We try to make the best from the available data and mix two-quarter strip prices from Binsbergen, Brandt, and Koijen (2012) with summary statistics for  $\widehat{y}_t^n - y_t^n$ ,  $n \geq 4$  quarters from Binsbergen, Hueskes, Koijen, and Vrugt (2013) and pepper them with admittedly heroic assumptions. See the description in Table 2 and Appendix B. All of this evidence is worth revisiting as more data become available in the future.

Our calibrated model shares qualitative traits of those matched to bond prices in the preceding sections. Quantitatively, we observe a dramatic drop in persistence  $\varphi_2$ . We've noted cross-sectional variation in  $\varphi_2$  earlier, but the equity one is the lowest. Most representative-agent models that were confronted with the Sharpe ratio evidence feature exogenously specified cash flows with persistence connected to that of expected consumption growth and, therefore, the real pricing kernel. Our results suggest exploring different persistence of cash flows and the pricing kernel before the final opinion on the equilibrium component of these models can be expressed.

Further, in the context of recursive preferences, high persistence of expected consumption growth is needed to generate high one-period risk premiums. This high persistence leads to unrealistically steep yield curves. Our model illustrates that an iid disaster component is helpful in separating the modelling of one-period high returns and relatively low term spreads in yields.

## 6 The representative agent with recursive preferences

In this section we offer an example of a representative-agent model that captures the basic features that we've highlighted in the previous sections. We hope this illustration would be useful for further development and improvement of existing models.

We use a model that is based on recursive preferences developed by Kreps and Porteus (1978), Epstein and Zin (1989), and Weill (1989), among many others. We define utility with the time aggregator,

$$U_t = [(1 - \beta)c_t^\rho + \beta\mu_t(U_{t+1})^\rho]^{1/\rho}, \quad (34)$$

and certainty equivalent function,

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

where  $c_t$  is the aggregate consumption. Additive power utility is a special case with  $\alpha = \rho$ . In standard terminology,  $\rho < 1$  captures time preference (with intertemporal elasticity of substitution  $1/(1 - \rho)$ ) and  $\alpha < 1$  captures risk aversion (with coefficient of relative risk aversion  $1 - \alpha$ ).

The time aggregator and certainty equivalent functions are both homogeneous of degree one, which allows us to scale everything by current consumption. If we define scaled utility  $u_t = U_t/c_t$ , equation (34) becomes

$$u_t = [(1 - \beta) + \beta\mu_t(g_{t+1}^c u_{t+1})^\rho]^{1/\rho}, \quad (35)$$

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

## 6.1 Real pricing kernel

With this utility function, the real pricing kernel is

$$\hat{m}_{t,t+1} = \beta(g_{t,t+1}^c)^{\rho-1} [g_{t,t+1}^c u_{t+1} / \mu_t(g_{t,t+1}^c u_{t+1})]^{1-\rho}.$$

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

$$\log g_{t,t+1}^c = g^c + \theta_c x_{2t} + \sigma w_{t+1} + z_{t+1}^c, \quad (36)$$

where jumps arrive at the rate  $\omega$  and jump sizes are distributed  $\mathcal{N}(\mu_c, \delta_c^2)$ . The factor  $x_{2t}$  is as above.

We derive the pricing implications from a loglinear approximation of (35):

$$\log u_t \approx b_0 + b_1 \log \mu_t(g_{t,t+1}^c u_{t+1}).$$

around the point  $\log \mu_t = E(\log \mu_t)$ . This is exact when  $\rho = 0$ , in which case  $b_0 = 0$  and  $b_1 = \beta$ .

We guess a value function of the form

$$\log u_{t+1} = u + u_x x_{2t+1} = u + u_x \varphi_2 x_{2t} + u_x w_{t+1}.$$

Then

$$\log(g_{t,t+1}^c u_{t+1}) = g^c + u + (\theta_c + u_x \varphi_2) x_{2t} + (\sigma + u_x) w_{t+1} + z_{t+1}^c.$$

Therefore,

$$\log \mu_t(g_{t,t+1}^c u_{t+1}) = g^c + u + \alpha(\sigma + u_x)^2/2 + \alpha^{-1}\omega(e^{\alpha\mu_c + \alpha^2\delta_c^2/2} - 1) + (\theta_c + u_x \varphi_2) x_{2t}.$$

Lining up terms, we get  $u_x = b_1\theta_c(1 - b_1\varphi_2)^{-1}$ . As a result, the real pricing kernel is:

$$\log \hat{m}_{t+1} = \hat{m} + (\rho - 1)\theta_c x_{2t} + [(\alpha - 1)\sigma + (\alpha - \rho)u_x] w_{t+1} + (\alpha - 1)z_{t+1}^c.$$

## 6.2 Nominal pricing kernel

In order to obtain the nominal pricing kernel, we can assume the process for inflation as is done in Bansal and Shaliastovich (2013), Piazzesi and Schneider (2006), and Wachter (2006). For example,

$$\log g_{t,t+1}^\pi = g^\pi - \theta_m x_{1t} + \theta_g^\pi x_{2t} + \eta_0^\pi w_{t+1} + \eta_2^\pi z_{t+1}^g. \quad (37)$$

Then the nominal pricing kernel is:

$$\begin{aligned} \log m_{t+1} &= \log \widehat{m}_{t+1} - \log g_{t,t+1}^\pi \\ &= m + \theta_m x_{1t} + [(\rho - 1)\theta_c - \theta_g^\pi]x_{2t} + [(\alpha - 1)\sigma + (\alpha - \rho)u_x - \eta_0^\pi]w_{t+1} \\ &\quad + (\alpha - 1)z_{t+1}^c - \eta_2^\pi z_{t+1}^g. \end{aligned}$$

## 6.3 Calibration and implications

The calibrated preference parameters and parameters controlling dynamics of consumption are listed in Panel C of Table 3. Our starting point is calibration of the inflation process (37). Because it is specified exogenously, we take it to be identical to CPI in Table 3B. Next, the preference parameters are selected to match the standard choice in the literature.

In calibrating consumption we start with the CI2 model presented in Backus, Chernov, and Zin (2014). By focussing on the US economy, they showed that introduction of iid disasters into the homoscedastic version of the Bansal and Yaron (2004) model and decrease of persistence of expected consumption growth leads to a realistic yield curve without giving up much of one-period entropy (largest risk premium). The difference in our and their calibration is in persistence and in conditional volatility of expected consumption growth. The former is lower in our case and matches our earlier calibration of  $\varphi_2$  for CPI. The latter,  $|\theta_c|$ , is larger in our case. Our calibrated values for the persistence and the conditional volatility of expected consumption growth are very close to the ones estimated by Zviadadze (2013) as a part of a comprehensive analysis of the US consumption dynamics, 0.81 and 0.0016, respectively.

We calibrate  $\theta_c$  with three objectives in mind: (i) to have  $x_{2t}$  affecting the nominal pricing kernel as little as possible to be close to our affine model of section 5; (ii) to ensure upward sloping real yield curve (negative serial covariance of the real pricing kernel); and (iii) to match  $\lambda_0 = (\alpha - 1)\sigma + (\alpha - \rho)u_x - \eta_0$  ( $u_x$  depends on  $\theta_c$ ). Objectives (i) and (ii) are conflicting:  $|\theta_c|$  would have to be larger to satisfy (i) perfectly. Condition (i) is arbitrary – it is chosen for esthetic reasons – so it is not essential that it is satisfied perfectly. It is essential for  $\theta_c$  to be negative to satisfy (ii) in our homoscedastic model. Finally,  $\sigma$  is selected to match the variance of consumption growth:  $0.018^2 = \theta_c^2(1 - \varphi_2^2)^{-1} + \sigma^2 + \omega(\mu_c^2 + \delta_c^2)$ . The cash flow processes are specified exogenously and taken directly from the model of section 5. Table 4 shows that the calibrated consumption process serves as a sensible representation of actual consumption data.

The jump in the nominal pricing kernel can be re-written as:

$$(\alpha - 1)z_{t+1}^c - \eta_2^\pi z_{t+1}^g = z_{t+1}^m,$$

where jumps arrive at the rate  $\omega$  with jump sizes  $\mathcal{N}(\mu_m, \delta_m^2)$  with  $\mu_m = [(\alpha - 1)\mu_c - \eta_2^\pi \mu_g]$  and  $\delta_m^2 = (\alpha - 1)^2 \delta_c^2 + \eta_2^{\pi^2} \delta_g^2$ . In our calibration  $\mu_m$  and  $\delta_m$  match those in the affine model. As a result, we have reverse-engineered the nominal pricing kernel that closely resembles the homoscedastic version of the one we've obtained in the reduced-form model.

We can specify exactly the same nominal cash flows as in equation (31). As a result, cash flows are calibrated exactly the same way as in the affine model of the previous section. Appendix C offers a motivation for this specification that is based on the change from real to nominal units.

## 6.4 Persistent jump component

Throughout the paper we have insisted on featuring an iid jump component and a persistent normal component in our models. Is there a scope for a persistent jump component? The issue is that additional persistence in the model would affect term spreads. Perhaps, it would be possible to setup a reduced-form model in such a way that the “right” amount of persistence is shared between the normal and jump components. The more pertinent question is whether this is feasible in an equilibrium model, such as the one introduced in this section, when there are additional cross-equation restrictions on parameters that are implied by the model.

To illustrate the issues involved we augment the model of consumption growth with a persistent jump component. We follow Wachter (2013) by introducing persistence through time-varying jump arrival rate  $\omega_t$ . We follow Backus, Chernov, and Zin (2014) by specifying it as:

$$\omega_{t+1} = \omega(1 - \varphi_\omega) + \varphi_\omega \omega_t + \sigma_\omega e_{t+1}.$$

Like these authors, we treat the specification as an approximation to a true process that truncates  $\omega_t$  at zero.

Repeating the same steps as above, one can show that the real pricing kernel is, in this case:

$$\begin{aligned} \log \hat{m}_{t+1} &= \hat{m} + (\rho - 1)\theta_c x_{2t} + [(\alpha - 1)\sigma + (\alpha - \rho)u_x]w_{t+1} + (\alpha - 1)z_{t+1}^c \\ &+ (\alpha - \rho)\alpha^{-1}(e^{\alpha\mu_c + \alpha^2\delta_c^2/2} - 1)\sigma_\omega(b_1(1 - b_1\varphi_\omega)^{-1}e_{t+1} - \omega_t/\sigma_\omega). \end{aligned}$$

We have factored out  $\sigma_\omega$  so that the persistent component associated with jumps is standardized similarly to the normal component  $x_{2t}$ . The jump component is multiplied by a more complicated expression featuring an exponential which could lead to large values of the loading on  $\omega_t$ .

If  $x_{2t}$  and  $\omega_t$  have similar persistence, then they will contribute equally to the shape of the yield curve if they have similar loadings in the pricing kernel. The normal persistent component is multiplied by  $(\rho - 1)\theta_c = 0.0009$  at calibrated parameter values. The jump arrival rate is multiplied by  $(\alpha - \rho)\alpha^{-1}(e^{\alpha\mu_c + \alpha^2\delta_c^2/2} - 1)\sigma_\omega = 9\sigma_\omega$ . As a benchmark, the value of  $\sigma_\omega$  should be around 0.0001 for the two components to have a similar impact on the pricing kernel. Wachter (2013) entertains a value of  $0.0355^{1/2} \cdot 0.067 \cdot (1/4)^{1/2} = 0.0063$  and Backus, Chernov, and Zin (2013) use  $0.0001 \cdot 3^{1/2} = 0.0002$ , so there is a range of opinion of where this value could be. The point is that if  $\sigma_\omega = 0.0001$ , then one is introducing double the persistence of what we've seen to be realistic.

So something has to adjust. One can set persistence of  $x_{2t}$  to zero as is done in Wachter (2013). Then, as Backus, Chernov, and Zin (2014) demonstrate in model SI, the issue is that  $\sigma_\omega$  and  $\varphi_\omega$  should have modest values of 0.0002 and  $0.95^3 = 0.8573$ , respectively to get anywhere close to the shape of the US yield curve. But at these modest values, the largest one-period risk premium captured by entropy is not much different from the iid jump case. If one needs modest values of  $\sigma_\omega$  and  $\varphi_\omega$  when there is no persistence in  $x_2$ , it is clear that once  $x_{2t}$  is persistent, the role for persistence in the jump component would have to be even smaller.

To summarize, quantitatively, there is no scope for having persistence in both normal and jump components of the pricing kernel. Given the differences in mathematical structure of the two components, persistence in jumps has a much larger impact on the term structure of asset prices. So, at least as a first order effect, the jump component is the one that should be iid.

## 7 Last thoughts

We focus on how risk is priced in the cross-section of assets and across investment horizons. Empirically, we link average log holding period returns on a given asset in excess of US interest rates to the difference between the yield curve corresponding to this asset (dividend yield, foreign yield, real yield) and the US yield curve. The cross-sectional dispersion of one-period excess returns is very large and continues to increase with horizon. For a given asset, excess log returns decline with horizon, but the rate of decline is different in the cross-section.

Theoretically, we introduce a concept of coentropy that serves as a generalized measure of covariance in the non-normal and multi-period world. Coentropy of the pricing kernel and cash flows is closely related to the aforementioned cross-sectional differences in yields. Thus, these differences in yields must reflect the differences in cash flows. We show that in order to capture the documented patterns in excess log returns an asset pricing model has to feature iid extreme outcomes, a persistent component, and cross-sectional variation in the persistence of cash flows. A model of the representative agent with recursive preferences whose consumption features disasters and persistent variation in its expected value is capable of capturing the evidence.

## A Copula, mutual information, and coentropy

Consider two random variables  $x_1$  and  $x_2$  with a joint pdf  $p(x_1, x_2)$  and marginals  $p_1(x_1)$  and  $p_2(x_2)$ . The corresponding marginal cdf's are  $P_1(x_1)$  and  $P_2(x_2)$ . Sklar's theorem enables one to decompose  $p$  using copula "density"  $c$ :

$$p(x_1, x_2) = c(P_1(x_1), P_2(x_2)) \cdot p_1(x_1) \cdot p_2(x_2).$$

(The general result is  $P(x_1, x_2) = Cop(P_1(x_1), P_2(x_2))$ , where  $Cop$  is copula.) Mutual information is

$$I(x_1, x_2) \equiv E \log \frac{p(x_1, x_2)}{p_1(x_1) \cdot p_2(x_2)} = E \log c(P_1(x_1), P_2(x_2)).$$

Coentropy:

$$\begin{aligned} C(x_1, x_2) &\equiv L(x_1 x_2) - L(x_1) - L(x_2) \\ &= \log E(x_1 x_2) - E \log(x_1 x_2) - (\log E x_1 - E \log x_1 + \log E x_2 - E \log x_2) \\ &= -E \log \frac{x_1 x_2}{E(x_1 x_2)} + E \log \frac{x_1}{E(x_1)} + E \log \frac{x_2}{E(x_2)}. \end{aligned}$$

Define new probabilities:  $\tilde{p}(x_1, x_2) = p(x_1, x_2)x_1x_2/E(x_1x_2)$ , and  $-j$  denotes "not  $j$ ". We have the following marginals

$$\begin{aligned} \tilde{p}_j(x_j) &= \int \tilde{p}(x_1, x_2) dx_{-j} = \int p(x_1, x_2)x_1x_2/E(x_1x_2) dx_{-j} \\ &= x_j p_j(x_j)/E(x_1x_2) \int p(x_{-j}|x_j)x_{-j} dx_{-j} = x_j p_j(x_j)E(x_{-j}|x_j)/E(x_1x_2) \\ &= p_j(x_j)x_j/E(x_j). \end{aligned}$$

Therefore,

$$\begin{aligned} C(x_1, x_2) &= -E \log \tilde{p}/p + E \log \tilde{p}_1/p_1 + E \log \tilde{p}_2/p_2 = -E \log \frac{\tilde{p}/p}{\tilde{p}_1/p_1 \cdot \tilde{p}_2/p_2} \\ &= -E \log \frac{\tilde{p}}{\tilde{p}_1 \cdot \tilde{p}_2} + E \log \frac{p}{p_1 \cdot p_2} \\ &= -E \log \tilde{c}(\tilde{P}_1(x_1), \tilde{P}_2(x_2)) + E \log c(P_1(x_1), P_2(x_2)) \\ &= -E \log \tilde{c}/c. \end{aligned}$$

Consider a specific example when the new probability is defined by  $\tilde{p}(m, g) = p(m, g)mg/E(mg)$ . Then the first marginal,  $\tilde{p}_1$  is the risk-adjusted probability.

Chabi-Yo and Colacito (2013) introduce a concept of coentropy. It is different from coentropy in this paper despite the same name. Expanding on their definition, we obtain:

$$K(x_1, x_2) \equiv 1 - \frac{L(x_2)}{L(x_1x_2) + L(x_1)} = \frac{L(x_1x_2) + L(x_1) - L(x_2)}{L(x_1x_2) + L(x_1)} = \frac{C(x_1, x_2) + 2L(x_1)}{C(x_1, x_2) + 2L(x_1) + L(x_2)}.$$

In their notation,  $x_1 = x$  and  $x_2 = y/x$ . We have relabeled the variables to match our use with theirs:  $x_1$  ultimately becomes  $m$ , and  $x_2$  is  $g$ .

## B Details of the dividend strips

Dividend strips are forward contracts on *annual* dividends paid out  $n$  years from now. So, assuming a time step of one quarter, these are not zero-coupon claims. The issue is how to summarize the data and to value these contracts in a setup where one quarter is the shortest time step.

Suppose  $d_{t+1}$  is a one-quarter dividend that is paid out at time  $t + 1$ . The corresponding (log) growth rate is  $\log g_{t,t+1} = \log(d_{t+1}/d_t)$ . One-year dividend is

$$d_t^{(m)} = \sum_{i=1}^m d_{t-m+i}, \quad m = 4.$$

A  $k$ -year forward contract specifies at date  $t$  the exchange of its price, or strike, for  $d_{t+km}^{(m)}$  at date  $t + n$ ,  $n = km$ . Denote its price by  $Q_t^n$ . Binsbergen, Hueskes, Koijen, and Vrugt (2013) report summary statistics for  $k^{-1}[\log d_t^{(m)} - \log Q_t^n]$ . Specifically, they report averages that are estimates of  $k^{-1}[E \log d_t^{(m)} - E \log Q_t^n]$ . This section establishes how is this object related to  $E \log rx_{t,t+n}$  in our paper.

Consider a claim to  $g_{t,t+n}^{(m)} \equiv d_{t+km}^{(m)}/d_t^{(m)}$  with a price denoted by  $\widehat{p}_t^n$ . The corresponding yield, as before, is  $\widehat{y}_t^n = -n^{-1} \log \widehat{p}_t^n$ . By no-arbitrage,  $p_t^n q_t^n = \widehat{p}_t^n$ , with  $q_t^n = Q_t^n/d_t^{(m)}$ , and  $p_t^n$  is a price of a US nominal zero-coupon bond that pays \$1 at time  $t + n$ .

Then,

$$k^{-1}E[\log d_t^{(m)} - \log Q_t^n] = k^{-1}E[-\log \widehat{p}_t^n + \log p_t^n] = mE[\widehat{y}_t^n - y_t^n].$$

Now consider return on the claim to  $g_{t,t+n}^{(m)}$ :

$$\begin{aligned} \log rx_{t,t+n} &= n^{-1}[\log g_{t,t+n}^{(m)} - \log \widehat{p}_t^n - \log r_{t,t+n}^n] \\ &= n^{-1}\left[\sum_{j=1}^n \log g_{t+j-1,t+j}^{(m)}\right] + \widehat{y}_t^n - y_t^n. \end{aligned}$$



Therefore,

$$\begin{aligned}
E(\log rx_{t,t+n} - \log rx_{t,t+1}) &= E(\widehat{y}_t^n - y_t^n) - E(\widehat{y}_t^1 - y_t^1) \\
&+ n^{-1} \sum_{j=1}^n E[\log g_{t+j-1,t+j}^{(m)} - \log g_{t,t+1}^{(m)}] \\
&= E(\widehat{y}_t^n - \widehat{y}_t^1) - E(y_t^n - y_t^1).
\end{aligned}$$

So, the Binsbergen, Hueskes, Koijen, and Vrugt (2013) statistic allows computing average term spread in excess returns.

We need to clarify what  $\widehat{y}_t^1$  is because the smallest  $n = 4$  in Binsbergen, Hueskes, Koijen, and Vrugt (2013). We will use the results from Binsbergen, Brandt, and Koijen (2012) to approximate this quantity. One-period asset yield corresponds to a claim to  $g_{t,t+1}^{(m)} \equiv d_{t+1}^{(m)}/d_t^{(m)}$ . Its price is

$$\begin{aligned}
\widehat{p}_t^1 &= E_t(m_{t,t+1}g_{t,t+1}^{(m)}) = (d_t^{(m)})^{-1} E_t(m_{t,t+1}d_{t+1}^{(m)}) \\
&= (d_t^{(m)})^{-1} \sum_{i=1}^{m-1} d_{t-m+1+i} + (d_t^{(m)})^{-1} E_t(m_{t,t+1}d_{t+1}).
\end{aligned}$$

Prices of six-month contracts, that is, claims to  $g_{t,t+2}^{(m)}$  are:

$$\begin{aligned}
\widehat{p}_t^2 &= E_t[m_{t,t+2}g_{t,t+2}^{(m)}] = (d_t^{(m)})^{-1} E_t[m_{t,t+2} \sum_{i=0}^m d_{t-m+2+i}] \\
&= (d_t^{(m)})^{-1} [\sum_{i=1}^{m-2} d_{t-m+2+i} + E_t(m_{t,t+1}d_{t+1}) + E_t(m_{t,t+2}d_{t+2})].
\end{aligned}$$

Binsbergen, Brandt, and Koijen (2012) report  $P_t^2 = E_t(m_{t,t+1}d_{t+1}) + E_t(m_{t,t+2}d_{t+2})$ . If we assume that  $E_t(m_{t,t+1}d_{t+1}) \approx 1.02E_t(m_{t,t+2}d_{t+2})$  (the one-period price is just a bit higher than the two-period price) then we can obtain an estimate of  $\widehat{y}_t^1$ :

$$\widehat{y}_t^1 = -\log \widehat{p}_t^1 \approx \log d_t^{(m)} - \log(d_{t-2} + d_{t-1} + d_t + P_t^2 * 0.495).$$

The reported shape of the corresponding curve does not materially depend on reasonable variations in the approximating assumption.

The issue with theoretical valuation of these securities is that they are not literally zero-coupon. Therefore, computation of yields would involve taking logs of sums of variables, which is not convenient. For this reason, we will exploit the persistence of dividends. That is, annual dividend divided by 4 (quarterly average) should not be too much different from the quarterly dividend. Figure 6 confirms this intuition. As a result, our theoretical model will treat dividend strips as if they were claims on quarterly dividends.

## C Real and nominal cashflows

Because we consider an endowment economy we can specify the cash flow process directly as in equation (31). To motivate why cash flows would have a two-factor structure, consider a more traditional specification of real cash flows in an endowment economy (e.g., Bansal and Yaron, 2004):

$$\log \hat{g}_{t,t+1} = \log \hat{\gamma} + \hat{\theta}_g \hat{x}_{2t} + \hat{\eta}_0 w_{t+1} + \hat{\eta}_2 z_{t+1}^g,$$

where  $\hat{x}_{2t}$  is an AR(1) process with persistence  $\hat{\varphi}_2$  and unit variance. Then, given the inflation process (37), the nominal cash flows follow:

$$\begin{aligned} \log g_{t,t+1} &= \log \hat{g}_{t,t+1} + \log g_{t,t+1}^\pi \\ &= (\log \hat{\gamma} + g^\pi) - \theta_m x_{1t} + (\hat{\theta}_g \hat{x}_{2t} + \theta_g^\pi x_{2t}) + (\hat{\eta}_0 + \eta_0^\pi) w_{t+1} + (\hat{\eta}_2 + \eta_2^\pi) z_{t+1}^g. \end{aligned}$$

We show in this appendix that the term  $x_t \equiv (\hat{\theta}_g \hat{x}_{2t} + \theta_g^\pi x_{2t}) / (\hat{\theta}_g + \theta_g^\pi)$  can be approximated by an AR(1) process. Thus, two persistent components in the nominal cash flows can be justified by the adjustment of the usual real cash flow process for inflation.

The specification of  $x_{2t}$  and  $\hat{x}_{2t}$  implies:

$$\begin{aligned} (1 - \varphi_2 L) x_{2t} &= w_t, \\ (1 - \hat{\varphi}_2 L) \hat{x}_{2t} &= w_t, \end{aligned}$$

where  $L$  is the lag operator. Therefore,

$$(1 - \varphi_2 L)(1 - \hat{\varphi}_2 L) x_t = \left( 1 - \left( \frac{\hat{\theta}_g}{\hat{\theta}_g + \theta_g^\pi} \varphi_2 + \frac{\theta_g^\pi}{\hat{\theta}_g + \theta_g^\pi} \hat{\varphi}_2 \right) L \right) w_t.$$

Therefore, the term  $x_t$  is an ARMA(2,1) process, which can be re-written as an AR( $\infty$ ) process:

$$\begin{aligned} w_t &= \frac{1 - (\varphi_2 + \hat{\varphi}_2)L + \varphi_2 \hat{\varphi}_2 L^2}{1 - \left( \frac{\hat{\theta}_g}{\hat{\theta}_g + \theta_g^\pi} \varphi_2 + \frac{\theta_g^\pi}{\hat{\theta}_g + \theta_g^\pi} \hat{\varphi}_2 \right) L} x_t \\ &= x_t - \left( \frac{\theta_g^\pi}{\hat{\theta}_g + \theta_g^\pi} \varphi_2 + \frac{\hat{\theta}_g}{\hat{\theta}_g + \theta_g^\pi} \hat{\varphi}_2 \right) x_{t-1} + \dots \end{aligned}$$

Thus, as a first-order approximation, the term  $x_t$  is an AR(1) with persistence coefficient of  $\frac{\theta_g^\pi}{\hat{\theta}_g + \theta_g^\pi} \varphi_2 + \frac{\hat{\theta}_g}{\hat{\theta}_g + \theta_g^\pi} \hat{\varphi}_2$ .

## 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 Bruce N. Lehmann, 1997, "Growth-optimal portfolio restrictions on asset pricing models," *Macroeconomic Dynamics* 1, 333-354.
- 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.
- Bansal, Ravi, and Amir Yaron, 2004, "Risks for the long run: A potential resolution of asset pricing puzzles," *Journal of Finance* 59, 1481-1509.
- Belo, Frederico, Pierre Collin-Dufresne, and Robert Goldstein, 2015, "Dividend Dynamics and the Term Structure of Dividend Strips," *Journal of Finance* 70, 1115-1160.
- 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, 2013, "Equity yields," *Journal of Financial Economics* 110, 503-519.
- Boguth, Oliver, Murray Carlson, Adlai Fisher, Mikhail Simutin, 2013, "Heterogeneous information diffusion and horizon effects in average returns," manuscript, May.
- Borovicka, Jaroslav, and Lars Peter Hansen, 2014, "Misspecified recovery," manuscript, May.
- Borovicka, Jaroslav, Lars Peter Hansen, and Jose Scheinkman, 2014, "Examining macroeconomic models through the lens of asset pricing," *Journal of Econometrics* 183, 67-90.
- Boudoukh, Jacob, Matthew Richardson, and Robert Whitelaw, 2015, "New evidence on the forward premium puzzle," *Journal of Financial and Quantitative Analysis*, forthcoming.
- Chabi-Yo, Fousseini and Riccardo Colacito, 2013, "The Term Structures of Co-Entropy in International Financial Markets," manuscript, November.
- Chernov, Mikhail, and Philippe Mueller, 2012, "The term structure of inflation expectations," *Journal of Financial Economics* 106, 367-394.
- Cochrane, John, 1992, "Explaining the variance of price-dividend ratios," *Review of Finan-*

*cial Studies* 5, 243-280.

- 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.
- Dahlquist, Magnus, and Henrik Hasseltoft, 2014, "Empirical evidence on international bond risk premia," manuscript, June.
- Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez, 2015, "The price of variance risk," manuscript, April.
- 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.
- Giglio, Stefano, and Bryan Kelly, 2015, " Excess Volatility: Beyond Discount Rates," manuscript.
- Giglio, Stefano, Matteo Maggiori, and Johannes Stroebel, 2015, " Very long-run discount rates," *Quarterly Journal of Economics* 130, 1-53.
- Gurkaynak, Refet, Brian Sack, and Jonathan Wright, 2007, "The US Treasury yield curve: 1961 to the present," *Journal of Monetary Economics* 54, 2291-2304.
- Gurkaynak, Refet, Brian Sack, and 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.
- Piazzesi, Monika, and Martin Schneider, 2006, “Equilibrium yield curves,” in Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, ed.: *NBER Macroeconomics Annual* (MIT Press: Cambridge MA).
- Qin, Likuan, and Vadim Linetsky, 2015, “Long Term Risk: A Martingale Approach,” manuscript
- Shiller, Robert, 1989, “Long term stock, bond, interest rate and consumption data,” spreadsheet posted on <http://www.econ.yale.edu/~shiller/data.htm>
- Wachter, Jessica, 2006, “A consumption-based model of the term structure of interest rates,” *Journal of Financial Economics* 79, 365-399.
- Wachter, Jessica, 2013, “Can time-varying risk of rare disasters explain aggregate stock market volatility?” *Journal of Finance* 68, 987-1035.
- Wright, Jonathan, 2011, “Term premia and inflation uncertainty: empirical evidence from an international panel dataset,” *American Economic Review* 101, 1514-1534.
- Zviadadze, Irina, 2013, “Term structure of consumption risk premia in the cross section of currency returns,” manuscript

**Table 1.** Properties of excess dollar returns. Entries are sample moments of quarterly observations of (quarterly) log excess returns:  $\log r - \log r^1$ , where  $r$  is a (gross) return and  $r^1$  is the (gross) return on a three-month bond. All of these returns are measured in dollars. Sample periods: US TIPS, 1971-2014 (source: Gurkaynak, Sack, and Wright, 2010; Chernov and Mueller, 2012); US nominal bonds, 1971-2014 (source: Gurkaynak, Sack, and Wright, 2007; FRED); Australian nominal bonds, 1987-2014 (source: Reserve Bank of Australia; Wright, 2011); UK nominal bonds, 1979-2014 (source: Bank of England); German nominal bonds, 1973-2014 (source: Bundesbank; Wright, 2011); exchange rate to the USD (source: FRED; EUR was complemented by DM, which was converted using the official EUR/DM rate); S&P 500 dividend strips, 1996-2009 (source: Binsbergen, Brandt, and Koijen, 2012). The shortest maturity available for dividend strips is two quarters, so we extrapolate to one quarter as described in Appendix B.

| Asset                                   | Mean    | Standard Deviation | Skewness | Excess Kurtosis | Entropy, $L(rx)$ |
|---|---------|--------------------|----------|-----------------|------------------|
| <i>Inflation-protected bonds (TIPS)</i> |         |                    |          |                 |                  |
| CPI                                     | 0.0022  | 0.0060             | 0.1785   | 0.8223          | 0.00002          |
| <i>Currencies</i>                       |         |                    |          |                 |                  |
| AUD                                     | 0.0108  | 0.0576             | -0.5134  | 0.7206          | 0.0016           |
| EUR (Germany)                           | -0.0015 | 0.0614             | 0.2748   | 0.6517          | 0.0018           |
| GBP                                     | -0.0008 | 0.0543             | -0.0816  | 1.4681          | 0.0015           |
| <i>Equity</i>                           |         |                    |          |                 |                  |
| S&P 500 div fut                         | -0.0159 | 0.0270             | 0.7491   | 0.6273          | 0.0004           |

**Table 2.** Average curves. Entries are means of yields on various assets of various maturities. All of these yields are expressed in decimals, on a quarterly basis. The second line shows the difference in term spreads relative to the US nominal curve. A term spread is defined as the difference between an  $n$ -quarter yield and a one-quarter yield. Sample periods: US nominal bonds, 1971-2014 (source: Gurkaynak, Sack, and Wright, 2007; FRED); US TIPS, 1971-2014 (source: Gurkaynak, Sack, and Wright, 2010; Chernov and Mueller, 2012); Australian nominal bonds, 1987-2014 (source: Reserve Bank of Australia; Wright, 2011); UK nominal bonds, 1979-2014 (source: Bank of England); German nominal bonds, 1973-2014 (source: Bundesbank; Wright, 2011); 2-quarter S&P 500 dividend strips, 1996-2009 (source: Binsbergen, Brandt, and Koijen, 2012); annual S&P 500 dividend futures, 2002-2011 (source: Binsbergen, Hueskes, Koijen, and Vrugt, 2013). Dividend strip/futures prices are not available at the one-quarter horizon, so we extrapolate to one quarter as described in Appendix B.

| Asset or<br>Country | Maturity, quarters |         |         |         |         |         |         |         |         |
|---------------------|--------------------|---------|---------|---------|---------|---------|---------|---------|---------|
|                     | 1                  | 2       | 4       | 8       | 12      | 20      | 24      | 28      | 40      |
| US                  | 0.0124             | 0.0128  | 0.0138  | 0.0144  | 0.0149  | 0.0157  | 0.0160  | 0.0163  | 0.0169  |
| US TIPS             | 0.0044             | 0.0043  | 0.0043  | 0.0045  | 0.0047  | 0.0052  |         | 0.0056  | 0.0061  |
|                     |                    | -0.0004 | -0.0015 | -0.0018 | -0.0021 | -0.0023 |         | -0.0026 | -0.0028 |
| Australia           | 0.0165             | 0.0164  |         | 0.0161  | 0.0164  | 0.0170  |         |         | 0.0177  |
|                     |                    | -0.0003 |         | -0.0021 | -0.0024 | -0.0029 |         |         | -0.0040 |
| Germany             | 0.0120             | 0.0118  | 0.0118  | 0.0124  | 0.0130  | 0.0139  | 0.0142  | 0.0145  | 0.0151  |
|                     |                    | -0.0006 | -0.0015 | -0.0015 | -0.0014 | -0.0014 | -0.0014 | -0.0014 | -0.0014 |
| UK                  | 0.0168             | 0.0173  | 0.0166  | 0.0168  | 0.0170  | 0.0175  | 0.0177  | 0.0178  | 0.0181  |
|                     |                    | 0.0002  | -0.0016 | -0.0020 | -0.0022 | -0.0026 | -0.0028 | -0.0032 | -0.0035 |
| S&P 500             | -0.0072            | -0.0056 | -0.0001 | 0.0018  | 0.0024  | 0.0035  | 0.0043  | 0.0048  |         |
|                     |                    | 0.0013  | 0.0061  | 0.0077  | 0.0078  | 0.0081  | 0.0084  | 0.0086  |         |

**Table 3.** Calibrated parameters. Entries are the model parameters expressed in quarterly terms. Note that fitted volatility of “CPI returns” is 8 times larger than that in the data.

| <i>Panel A. Common parameters<br/>(US nominal economy)</i> |             |            |             |             |          |         |            |
|--|-------------|------------|-------------|-------------|----------|---------|------------|
|  | $\varphi_1$ | $\theta_m$ | $\lambda_0$ | $\lambda_1$ | $\omega$ | $\mu_m$ | $\delta_m$ |
|  | 0.9487      | 0.0026     | -0.1225     | 0.0512      | 0.0025   | 1.5000  | 1.5000     |

| <i>Panel B. Asset-specific parameters and derived quantities</i> |             |            |          |         |            |            |                       |
|--|-------------|------------|----------|---------|------------|------------|-----------------------|
| Asset  | $\varphi_2$ | $\theta_g$ | $\eta_0$ | $\mu_g$ | $\delta_g$ | $cov_{mg}$ | $\mathcal{C}_{mg}(1)$ |
| <i>Inflation-protected bonds (TIPS)</i>                          |             |            |          |         |            |            |                       |
| CPI  | 0.8023      | 0.0035     | 0.0478   | 0.0198  | 0.0327     | -0.0057    | -0.0034               |
| <i>Currencies</i>  |             |            |          |         |            |            |                       |
| AUD  | 0.9404      | 0.0029     | 0.0642   | -0.0498 | 0.0774     | -0.0081    | -0.0129               |
| EUR  | 0.8356      | 0.0045     | 0.0254   | -0.9606 | 0.5688     | -0.0009    | 0.0002                |
| GBP  | 0.9664      | 0.0027     | 0.0587   | 0.0475  | 0.0800     | -0.0068    | -0.0009               |
| <i>Equity</i>  |             |            |          |         |            |            |                       |
| S&P 500  | 0.6846      | 0.0292     | -0.0225  | -0.1464 | 0.2416     | 0.0036     | 0.0126                |

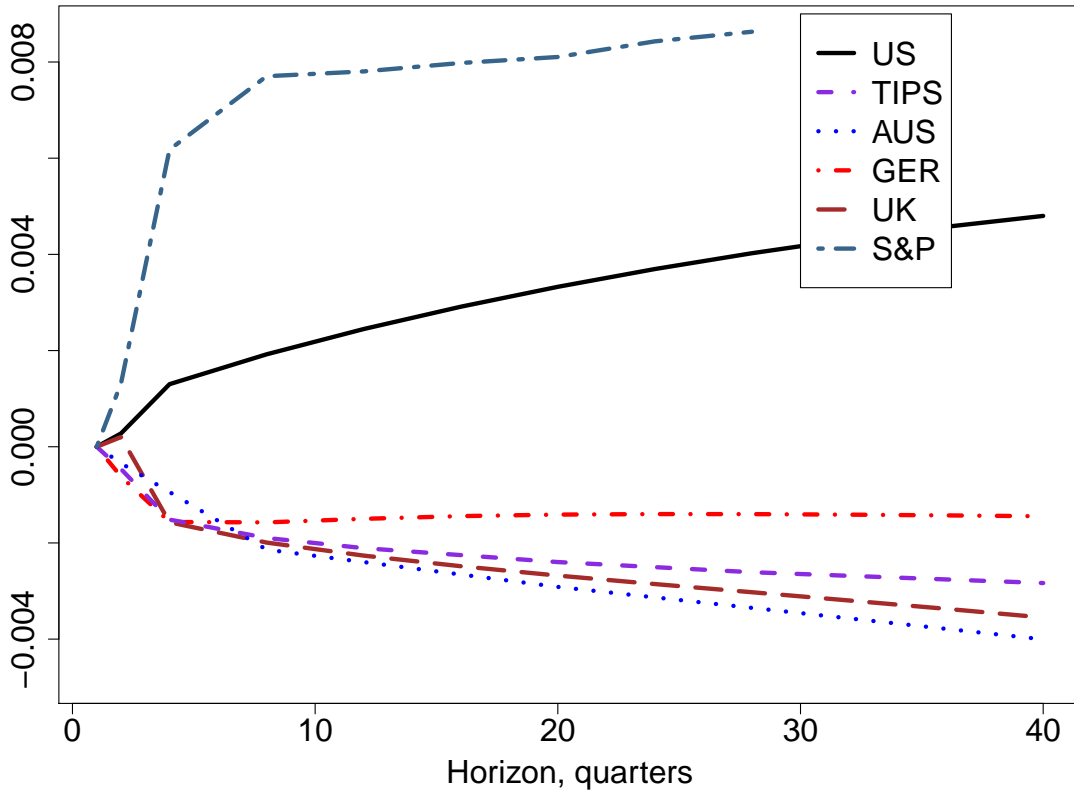
| <i>Panel C. Parameters from the representative agent model</i> |            |          |         |            |                    |          |        |        |
|--|------------|----------|---------|------------|--------------------|----------|--------|--------|
| <i>Consumption</i>   |            |          |         |            | <i>Preferences</i> |          |        |        |
| $\varphi_2$  | $\theta_c$ | $\sigma$ | $\mu_c$ | $\delta_c$ | $\beta$            | $\alpha$ | $\rho$ | $b_1$  |
| 0.8023   | -0.0014    | 0.0143   | -0.1520 | 0.1500     | 0.9980             | -9       | 1/3    | 0.9986 |



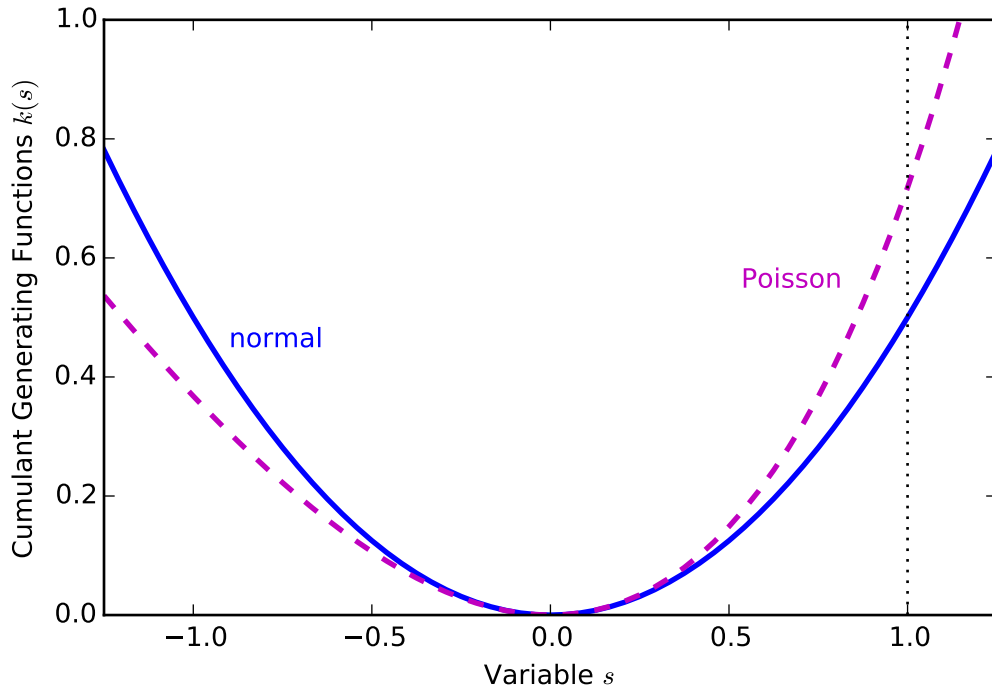
**Table 4.** Variance and serial correlation of cash flows. The Data module reports summary statistics and the corresponding standard errors in parentheses in the second line. The Model module reports population values at the calibrated parameters in the first line. The second line report in parentheses the 2.5th and 97.5th percentiles of the distribution of the respective statistics computed from 100,000 artificial histories of  $\log g$  simulated from the model at calibrated parameters. We use annual data on dividends expressed in quarterly units. The reason is that dividends are highly seasonal and lumpy. As a result, the Shiller (1989) annual data are an accurate representation of annual dividends, but it is oversmoothing at higher frequencies. In order to match the annual data with the quarterly model, we simulate annual dividends. Consumption data are from quarterly NIPA tables from 1947 to 2014. Variance of consumption growth is matched by construction, so we do not report its sampling characteristics to emphasize this.

| Asset                                   | <i>Data</i>        |                    | <i>Model</i>               |                              |
|---|--------------------|--------------------|----------------------------|------------------------------|
|   | Var $\times 10^2$  | AR(1)              | Var $\times 10^2$          | AR(1)                        |
| <i>Inflation-protected bonds (TIPS)</i> |                    |                    |                            |                              |
| CPI                                     | 0.0078<br>(0.0008) | 0.5889<br>(0.0625) | 0.2439<br>(0.1934, 0.2966) | 0.0286<br>(-0.1271, 0.1748)  |
| <i>Currencies</i>                       |                    |                    |                            |                              |
| AUD                                     | 0.3132<br>(0.0422) | 0.0598<br>(0.0950) | 0.4250<br>(0.3189, 0.5470) | -0.0218<br>(-0.2021, 0.1641) |
| EUR                                     | 0.3748<br>(0.0410) | 0.0015<br>(0.0788) | 0.3808<br>(0.0561, 2.3987) | 0.0175<br>(-0.0716, 0.2398)  |
| GBP                                     | 0.3126<br>(0.0370) | 0.1271<br>(0.0828) | 0.3529<br>(0.2757, 0.4427) | -0.0323<br>(-0.1879, 0.1322) |
| <i>Equity</i>                           |                    |                    |                            |                              |
| S&P 500                                 | 0.3829<br>(0.0459) | 0.2599<br>(0.0812) | 0.2781<br>(0.1966, 0.4033) | 0.1138<br>(-0.0775, 0.2974)  |
| <i>Macro</i>                            |                    |                    |                            |                              |
| Cons. growth                            | 0.0324<br>-        | 0.0877<br>(0.0600) | 0.0324<br>-                | -0.0609<br>(-0.1779, 0.0480) |

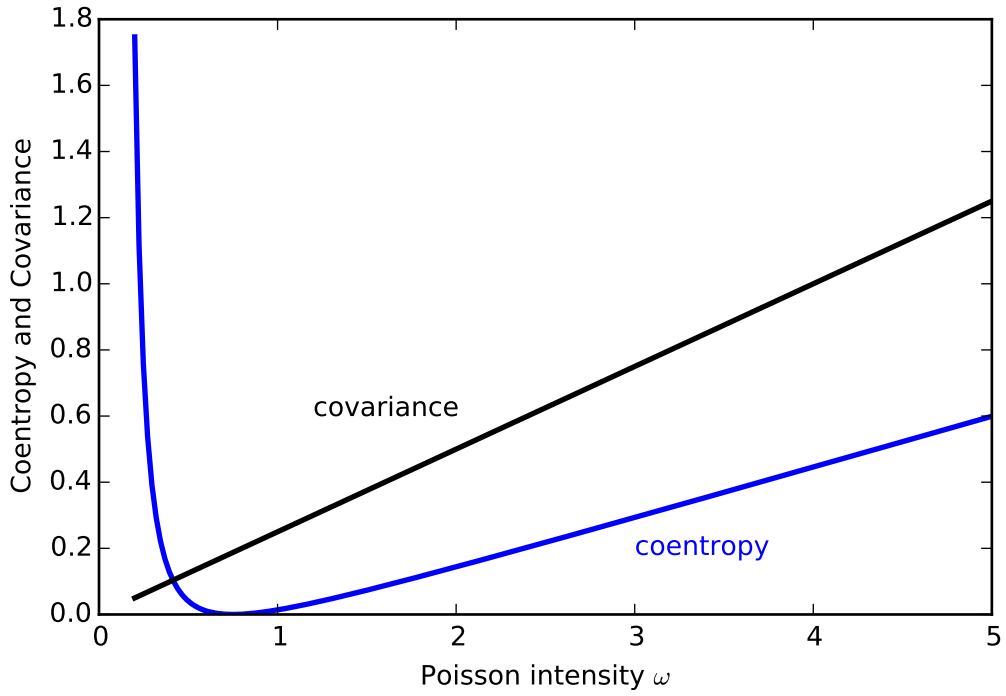
**Figure 1.** Average US curve and excess returns. The black solid line shows average US nominal term spreads,  $E(y_t^n - y_t^1)$  at different maturities  $n$ . The remaining lines represent the term spread in average excess returns,  $E(\log rx_{t,t+n} - \log rx_{t,t+1})$ , measured by differences of average term spreads on several assets relative to US Treasuries,  $E(\hat{y}_t^n - \hat{y}_t^1) - E(y_t^n - y_t^1)$ . Data sources are the same as in Table 2.



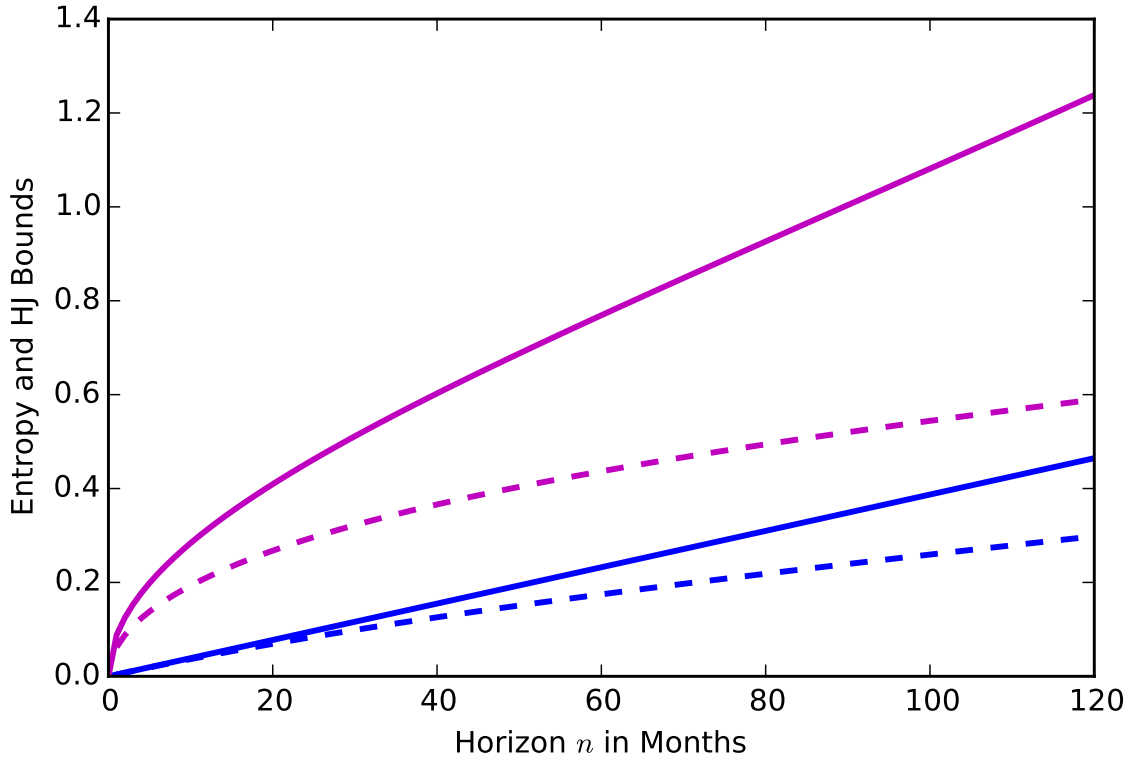
**Figure 2.** Two cumulant generating functions. The functions  $k(s)$  are properties of the distributions of  $\log x$ . In one,  $\log x$  is normal, in the other Poisson. Both are standardized: they have mean zero and variance one. The Poisson has intensity parameter  $\omega = 1$  and scale parameter  $\theta > 0$ . 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.



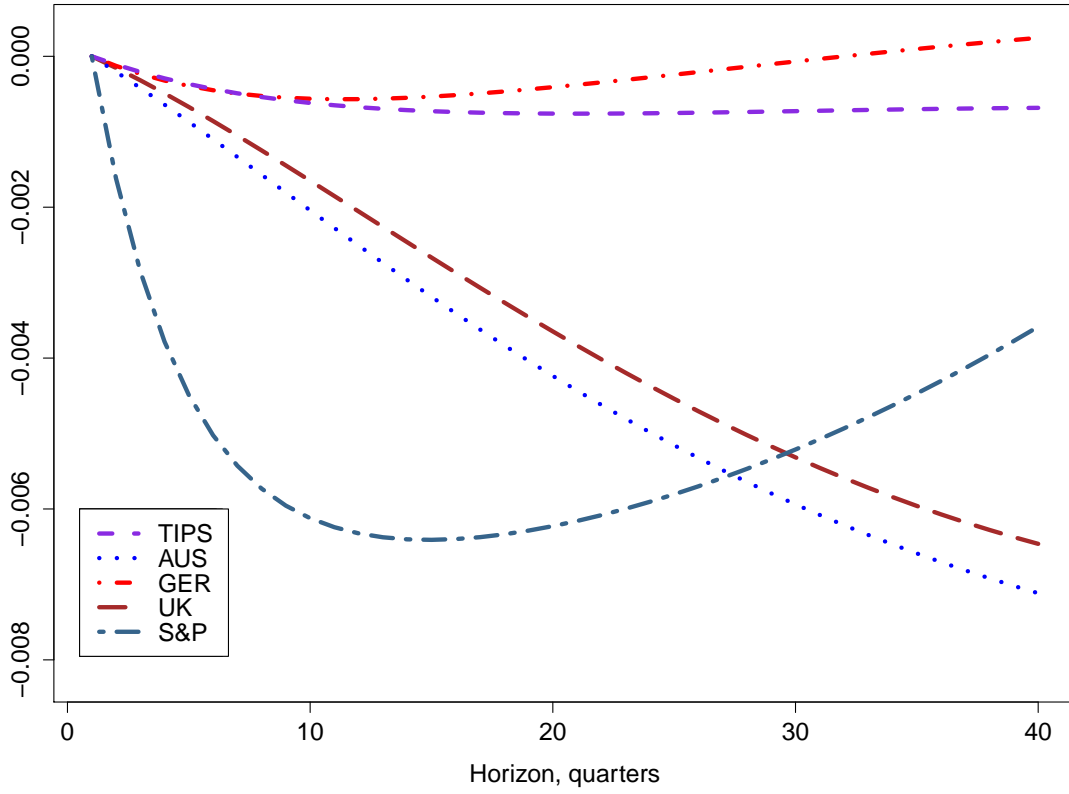
**Figure 3.** Coentropy and covariance. The figure compares coentropy and covariance for the Poisson mixture of bivariate normals described in Example 4. As we vary  $\omega$ , we adjust  $\delta$  to hold the variance constant.



**Figure 4.** HJ bound and entropy in the Vasicek model. The figure compares how the HJ bound (purple lines) and entropy (blue lines) change with horizon in the benchmark iid case (solid lines) and in the Vasicek model (dashed lines).



**Figure 5.** Term structures of coentropies. The figure compares term structures of coentropies,  $n^{-1}\mathcal{C}_{mg}(n) - \mathcal{C}_{mg}(1)$  for the different assets considered in this paper. Coentropies are derived from the illustrative affine model.



**Figure 6.** S&P 500 dividends. The figure displays quarterly dividends (black solid line) and quarterly average of annual dividends (red dashed line). The sample corresponds to the availability of short-term dividend prices in Binsbergen, Brandt, and Kojen (2012).

