TERM STRUCTURES OF ASSET PRICES AND RETURNS

David Backus
Nina Boyarchenko
Mikhail Chernov

Term Structures of Asset Prices and Returns
David Backus, Nina Boyarchenko, and Mikhail Chernov
NBER Working Paper No. 22162
April 2016, Revised September 2017
JEL No. G12,G13

ABSTRACT

We explore the term structures of claims to a variety of cash flows, namely, 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). The average term structures reflect the dynamics of the dollar pricing kernel, cash flow growth, and the interaction between the two. We use an affine model to illustrate how these two components can deliver term structures with a wide range of levels and shapes. Finally, we calibrate a representative agent economy to show that the evidence we document is consistent with the equilibrium models.

David Backus
Stern School of Business
NYU
44 West 4th Street
New York, NY 10012-1126
dbackus@stern.nyu.edu

Nina Boyarchenko
Federal Reserve Bank of New York
33 Liberty Street
New York, NY 10045
Nina.boyarchenko@ny.frb.org

Mikhail Chernov
Anderson School of Management
University of California, Los Angeles
110 Westwood Plaza, Suite C-417
Los Angeles, CA 90095
and NBER
mikhail.chernov@anderson.ucla.edu

1 Introduction

Perhaps the most striking recent challenge to the representative agent models comes from evidence about the term structure of risk premiums. Several papers have argued that the patterns 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, 2017, provide a comprehensive review). In an endowment economy, representative agent models have two basic components: an equilibrium-based pricing kernel that prices all of the assets in the economy and an exogenously specified cash flow process for a given asset. The evidence implicitly suggests which features these two components must possess. In this paper, we develop a methodology that allows researchers to establish these required features.

We start by introducing complementary evidence on the average buy-and-hold log excess returns across different horizons of a diverse set of assets, namely, foreign-currency bonds, inflation-protected bonds, and the dividend yields associated with equity dividend strips. We find log excess returns attractive because, as we show, the change in their averages with the horizon tracks the difference between the term spreads of two yield curves: one associated with U.S. dollar (USD) bonds and the other corresponding to the other individual assets. Because the foreign-currency, inflation-protected, and equity dividend assets are claims to different cash flows, the recovered term structures of their average log excess returns have disparate levels and shapes.

What determines the level and shape of the term structure for a given asset? We follow the approach of Alvarez and Jermann (2005), Backus, Chernov, and Zin (2014), and Bansal and Lehmann (1997) in articulating the key modeling points. These authors connect the average level of excess returns to the entropy of the pricing kernel or, equivalently, to the largest risk premium in a given economy. The entropy of the USD pricing kernel has to be sufficiently large to be consistent with the magnitudes of the observed returns. Next, the changes in the entropy of the pricing kernel associated with changes in an investment horizon, known as horizon dependence, should be moving one for one with how the yield term spreads change with the horizon. These changes are small, and this, in general, imposes a limit on how large the entropy can be.

We can apply the same logic to the average log excess returns by suitably redefining the pricing kernel. We define the transformed pricing kernel as the product of the USD pricing kernel and the growth rate of the cash flows of a given asset. This type of pricing kernel is often referred to as the foreign pricing kernel in the context of currencies and the real pricing kernel in the context of inflation, although it does not have any special moniker in the context of equity dividends. Given the empirical slopes of the term structures, each of these transformed pricing kernels should have a large entropy and small horizon dependency. As with the pricing kernel itself, it is convenient to separate the consideration of the shape of the term structure of the log excess returns from the consideration of the level of the term structure. Because this term structure is driven by the differences between the U.S. curves

and asset-specific yields, we can infer an empirically plausible specification of the cash flow dynamics for each asset by comparing the transformed pricing kernel to the nominal one.

As it turns out, it is difficult to match the cash flow dynamics capable of replicating the term structure pattern in the excess returns with the level of one-period excess returns. We characterize this tension between shapes and levels quantitatively using an affine term structure model. The U.S. nominal term structure allows us to fix an empirically plausible model of the nominal pricing kernel. We follow the logic of term structure models in identifying the transformed pricing kernels that correspond to the yield curves for each of the other assets. We establish that the cross-sectional differences in the shapes of the yield curves for these assets are driven by the cross-sectional differences between the levels of persistence of the expected cash flows and by the difference between their persistence and the persistence of the U.S. nominal pricing kernel. In practice, this means that the expected cash flow growth should be affected by at least two state variables. One is common across all assets, including the U.S. nominal term structure, while the other is asset-specific and has a different level of persistence.

Turning next to the levels of the term structures, we find that the observed one-period excess returns are too high; that is, the entropy is too low in the calibrated model. We argue that the affine term structure models that are used to describe the shape of the yield curve have to be augmented with non-normal innovations, which are frequently modeled via jumps. For non-normal innovations to the cash flows to affect the level of excess returns, the pricing kernel and the cash flow growth process should have coincident jumps.

The shape of the yield curve imposes an important constraint on the dynamics of this additional shock. To maintain the empirically plausible horizon dependence of log excess returns, this joint jump must be iid, that is, neither the probability of a jump occurring nor the conditional distribution of the jump sizes can have persistent components. This separation between horizon dependence and the level of the yield curve allows us to calibrate the jump distribution separately by matching the average and variance of the one-period risk premiums of the corresponding assets.

In log-normal environments, log risk premiums are captured by the covariance of the log pricing kernel and log cash flows. When both the pricing kernel and the cash flow process have non-normal innovations, covariance is no longer a sufficient statistic for the comovement between the two. Building on the entropy research, we introduce the concept of coentropy as a measure of dependence that directly generalizes the computation of log risk premiums to non-normal environments.

Coentropy is equal to an infinite sum of the joint cumulants of the log pricing kernel and log cash flows, with the first joint cumulant being the covariance. This concept is useful for computing risk premiums in our models. Furthermore, the interpretation of coentropy as an infinite sum of joint cumulants allows us to highlight the role of non-normal innovations in generating realistic risk premiums.

Indeed, we show that our proposed extension to the affine term structure model successfully matches the one-period risk premiums. Quantitatively, a modest non-normality, i.e., small cumulants of shocks to the cash flow growth process, translates into large risk premiums. This happens because the cash-flow cumulants interact with large cumulants of the non-normal innovations to the pricing kernel.

We introduce a model of an endowment economy featuring a representative agent with recursive preferences to illustrate how the modeling insights of the affine model manifest themselves in an equilibrium setting. The model shows which features need to be incorporated in the equilibrium models to satisfy the empirical targets presented by the term structure evidence.

We highlight three important modeling components. First, unlike the literature that models the variance of consumption growth as either an AR(1) or ARG(1) (square-root) process, we assume that the *volatility* of the consumption growth is an AR(1) process. This feature allows the generation of upward sloping nominal and real yield curves. Second, the consumption growth features an iid jump similar to the one in Barro (2006). This is important for resolving the highlighted tension between matching the shapes of the yield curves and levels of the risk premiums. Third, the expected cash flow growth depends on two state variables. One of the state variables also affects the expected consumption growth, which is the traditional, albeit less persistent, "long-run risk" component of consumption growth. The other state variable is asset-specific as in our affine model.

Related literature

Our work is primarily motivated by two strands of recent literature. First, there is growing evidence, both non-parametric and model-based, on the risk premium patterns of zero-coupon securities across different horizons. A partial list of the research in this area includes Belo, Collin-Dufresne, and Goldstein (2015), Binsbergen, Brandt, and Koijen (2012), Binsbergen, Hueskes, Koijen, and Vrugt (2012), Dahlquist and Hasseltoft (2013, 2014), Dew-Becker, Giglio, Le, and Rodriguez (2015), Giglio, Maggiori, and Stroebel (2015), Hansen, Heaton and Li (2008), Hasler and Marfe (2015), Lustig, Stathopolous, and Verdelhan (2014), and Zviadadze (2013). We complement this body of research by offering evidence on the log excess returns, which are cousins of risk premiums. This switch allows us to connect evidence across the different horizons in a more transparent way. We also differ from the literature in that we use the evidence to establish features that a successful asset-pricing model should possess instead of estimating and testing specific models.

Second, an important stream of theoretical literature, exemplified by Alvarez and Jermann (2005), Hansen (2012), Hansen, Heaton, and Li (2008), and Hansen and Scheinkman (2009), analyzes the interaction of cash flows and the pricing kernel at the infinite horizon. Our approach has a deep connection to these papers, which we highlight in the main text. The main difference is that we rely on the existing evidence at intermediate horizons to characterize the transition in the risk premiums across these horizons.

3

Our paper is also connected to earlier research seeking to understand the value premium in the cross-section of equities, such as Hansen, Heaton, and Li (2008), Lettau and Wachter (2007), and Santos and Veronesi (2010). The evidence on zero-coupon assets was not available at that time, so these papers confront a different set of facts that do not have an explicit horizon dependence, which forces them to make different modeling choices.

The work of Lettau and Wachter (2011), who reach across different types of assets by modelling the aggregate stock market, cross-section of equities, and the yield curve via an affine pricing kernel, is particularly close to our paper. Because of limited data availability, they focus on different moments than we do. Moreover, cross-sectional differences in cash-flows arise from the additivity constraint on individual firms, whereas we do not explore this mechanism at all in this paper. Because we do not study individual firms, our model of cross-sectional differences in cash flows arises from the different exposures to the common component and from the asset-specific cash-flow components. Finally, we emphasize the tension in the term structure of zero-coupon asset returns and their one-period counterparts. We argue that the only way to resolve this tension is to allow for a non-normal shock to the pricing kernel and cash flows.

In this last respect, our paper is related to the literature on the impact of jumps on asset prices dating back to Merton (1976). More recent work, such as Backus, Chernov, and Martin (2011), Barro (2006), Longstaff and Piazzesi (2004), Rietz (1988), and Wachter (2013), has focused on the ability of jumps to explain the asset risk premiums. Our approach differs from this body of literature because it emphasizes that the jump component is not only helpful but must also be present in our asset-pricing models. Moreover, this component must be iid to resolve the tension between the relatively flat term structures of returns and relatively high one-period risk premiums. Finally, we introduce the concept of coentropy, which is helpful in characterizing the log risk premiums and expected log excess returns in the presence of jumps.

2 Evidence

We focus on the properties of the observed term structures of prices and returns, so it is helpful to begin with the 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 the "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 U.S. dollar, the 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 the special notation $y_t^n \equiv n^{-1}\log r_{t,t+n}^n$ for a yield or, equivalently, an $n-$period holding period return on a U.S. nominal bond); foreign bonds if d_t is an exchange rate; inflation-linked bonds if d_t is the price level; and equities if d_t is a dividend.

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

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

We can thus express the term spread between the average per-period 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 the per-period excess holding return as

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

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

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

This connection between yields and excess returns simplifies the ordinarily difficult task of reliably computing the holding period returns over long horizons. The number of non-overlapping data points available decreases when computing the historical average of realized returns. In contrast, yields are available in every period, so the number of available data points does not change with the horizon n and does not require observations of the cash flows. We only need to compute the average excess return for $n = 1$ and then propagate it across horizons using the yields.

We report the summary statistics for the one-period excess returns for a number of examples in Table 1. We choose assets for which zero-coupon approximations exist, namely the various bonds and dividend strips. This exercise is meant to be illustrative, so we do not exhaustively analyze all possible assets (see Giglio and Kelly, 2015, and Binsbergen and Koijen, 2017, for a more comprehensive list). Based on data availability, we select one quarter as one period. We observe a large cross-sectional dispersion in the returns of around 1.36 percent per quarter or about 5.5 percent per year. Departures of excess returns from normality are evident despite the relatively low frequency, with the skewness of the returns ranging from -0.5 (Australian dollars) to 0.75 (S&P dividends).

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

Assets with cash flows also have term structures, although they typically have less market depth at long maturities than bonds. In general, they differ 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 are flatter, and some have different shapes.

In Figure 1, we plot the term spreads of the U.S. Treasury yield, $y_t^n - y_t^1$, and the differences between the mean term spreads on a number of other assets and U.S. Treasury yields, $E(\widehat{y}_t^n - \widehat{y}_t^1) - E(y_t^n - y_t^1)$. Because the latter is equal to the average difference between one-

and n-period excess returns, excess returns decline with the horizon in all of the examples with the exception of the dividend strips. Moreover, the cross-sectional spread in the excess returns widens as the horizon increases. The additional spread is about 1 percent higher annually than the one-quarter excess returns.

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

3   Term structures of prices and returns

We now model the term structures of asset prices and returns. We do this by showing how the concepts of entropy and horizon dependence can help translate the evidence on excess returns into the language of term structure modeling. In particular, we highlight the tension between a model's ability to fit how risk premiums change with the horizon vs how large the risk premiums are over one period.

3.1   Term structure of zero-coupon bonds

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

$$E_t\big(m_{t,t+1}r_{t,t+1}\big) \;\;=\;\; 1 \tag{3}$$

for all returns r. An asset pricing model is then a stochastic process for m and r. We want to characterize what the asset prices tell us about these stochastic processes. In this section, we start with a process for m.

The equality (3) implies a bound on the expected log excess returns:

$$E(\log r_{t,t+1} - \log r^1_{t,t+1}) \;\;\leq\;\; E[\log E_t m_{t,t+1} - E_t \log m_{t,t+1}] \equiv E[L_t(m_{t,t+1})]. \tag{4}$$

We refer to the inequality as the *entropy bound*, to L_t as conditional entropy, and to its unconditional expectation as entropy (see Alvarez and Jermann (2005, proof of Proposition 2), Backus, Chernov, and Martin (2011, Section I.C), Backus, Chernov, and Zin (2014, Sections I.C and I.D), and Bansal and Lehmann (1997, Section 2.3)). Thus, entropy is the highest possible expected excess return that an asset can generate in an economy that features the pricing kernel $m_{t,t+1}$.

To facilitate computation, we express entropy in terms of the cumulant generating function (cgf) of $\log x$. The cgf of $\log x$, if it exists, is the log of its moment-generating function,

$$k_t(s; \log x_{t+1}) \;\;=\;\; \log E_t\big(e^{s \log x_{t+1}}\big). \tag{5}$$

Conditional entropy is therefore

$$L_t(x_{t+1}) \equiv \log E_t x_{t+1} - E_t \log x_{t+1} = k_t(1; \log x_{t+1}) - E_t \log x_{t+1}.$$

Example 1 ("two-horizon" price of risk). Consider the following model of the pricing kernel:

$$\log m_{t,t+1} = \log \beta + a_0 w_{t+1} + a_1 w_t. \tag{6}$$

with $\{w_t\}$ iid standard normal. Although this model has valuation implications for horizons beyond two, the information about one- and two-horizon assets is sufficient to identify its properties. The price of risk a_0 is constant in this model.

Conditional entropy,

$$L_t(m_{t,t+1}) = a_0^2/2,$$

is the maximum risk premium regardless of the state. So, the entropy, $E[L_t(m_{t,t+1})]$, is the same. The model can generate high risk premiums via large values of a_0. Although one could be tempted to choose a high a value as desired, the issue is whether discipline can be imposed on the choice of this value.

One source of such discipline is the yield curve. In an arbitrage-free setting, the bond prices inherit their properties from the pricing kernel. Pricing has a simple recursive structure. Applying the pricing relation (3) to the bond prices gives us

$$p_t^n = E_t\big(m_{t,t+1}p_{t+1}^{n-1}\big) = E_t m_{t,t+n}, \tag{7}$$

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

The dynamics of the pricing kernel are reflected in what Backus, Chernov, and Zin (2014) call horizon dependence, which is the relation between the $n-$period entropy $\mathcal{L}_m(n)$,

$$\mathcal{L}_m(n) \equiv n^{-1}E[L_t(m_{t,t+n})] = n^{-1}E[\log E_t m_{t,t+n}] - E \log m_{t,t+1}. \tag{8}$$

and the time horizon represented by the function

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

Backus, Chernov, and Zin (2014) show that horizon dependence is connected to bond yields via

$$\mathcal{H}_m(n) = -E(y_t^n - y_t^1). \tag{9}$$

In the iid case, $\mathcal{H}_m(n) = 0$, and the yield curve is flat or, equivalently, the entropy does not change with n. Bond yields are then the same at all maturities and are constant over time. If the mean yield curve slopes upwards, then $\mathcal{H}_m(n)$ is negative and slopes downward,

7

reflecting the dynamics in the pricing kernel. One important implication of this result is that the iid components of m will affect the level of the yield curve but not its shape.

These concepts relate to the work of Alvarez and Jermann (2005), Hansen (2012), Hansen, Heaton, and Li (2008), and Hansen and Scheinkman (2009) when the horizon n is pushed to infinity. The connection is detailed in Appendix A. The major difference is that by considering the intermediate n, we are able to connect the theory to data.

Example 1 ("two-horizon" price of risk, continued). Using the cgf definition (5), the guess-and-verify approach, and the law of iterated expectations, we obtain the cgf of the $n-$period pricing kernel:

$$k_t(s; \log m_{t,t+n}) = C_n(s) + sa_1 w_t,$$

where the constant is $C_n(s) = ns\log\beta + (n-1)s^2(a_0 + a_1)^2/2 + s^2a_0^2/2$. Thus, the (log) bond prices are

$$\log p_t^n = k_t(1; \log m_{t,t+n}) = n\log\beta + (n-1)(a_0 + a_1)^2/2 + a_0^2/2 + a_1 w_t. \tag{10}$$

The one-period yield is

$$y_t^1 = -\log p_t^1 = -\log\beta - a_0^2/2 - a_1 w_t. \tag{11}$$

The horizon dependence is

$$\mathcal{H}_m(n) = (1 - 1/n)[(a_0 + a_1)^2 - a_0^2]/2. \tag{12}$$

The pricing kernel becomes iid when $a_1 = 0$. Consistent with the earlier observation, the horizon dependence is constant across horizons in this case. That is, a_1 affects the slope of the yield curve in this model. Also, a_1 is pinned down by the volatility of the one-period interest rate (11). Thus, the quantity that is helpful in generating the large one-period risk premium, a_0, is constrained by the value of a_1 and the need to fit the term spreads, $-\mathcal{H}_m(n)$, or, equivalently, by how the largest n-period risk premium differs from the one-period premium.

To demonstrate this, we put some numbers on the parameters. We use the properties of the U.S. nominal Treasury data described in Tables 1 and 2 for calibration. We focus on one- and two-period bonds only because of the two-horizon structure of the pricing kernel. At a quarterly frequency, the short rate y_t^1 in equation (11) has a standard deviation of 0.0084. Thus, we set the absolute value of a_1 to this value. The mean of the two-quarter yield spread $y^2 - y^1$ is 0.0004; equivalently, the two-period horizon dependence in equation (12) is -0.0004. We reproduce this value by setting $a_0 = -0.0994$. This value of a_0 corresponds to the maximum risk premium of 2 percent per year $(a_0^2/2 \cdot 400)$.

This low magnitude of the maximum risk premium reflects the tension between fitting the yield curve and generating large one-period risk premiums within the same pricing kernel. We tackle the question of how to resolve this tension in section 5.

To conclude this section, we contrast the result (9) with Hansen and Jagannathan's (1991) characterization of the pricing kernel via the bound

$$E_t(r_{t,t+1} - r^1_{t,t+1})/\text{Var}_t(r_{t,t+1} - r^1_{t,t+1})^{1/2} \leq \text{Var}_t(m_{t,t+1})^{1/2}/E_t(m_{t,t+1}). \tag{13}$$

We extend this to an $n-$period case by characterizing the mean and variance of the pricing kernel via the cgf:

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

The n-period bound is then

$$
\begin{aligned}
\text{Var}_t(m_{t,t+n})^{1/2}/E_t(m_{t,t+n}) &= \left(e^{k_t(2;\log m_{t,t+n})-2k_t(1;\log m_{t,t+n})} - 1\right)^{1/2} \\
&= \left(e^{(n-1)(a_0+a_1)^2+a_0^2} - 1\right)^{1/2},
\end{aligned}
$$

where the last line corresponds to the simple model from our example. The term in the exponent is a positive constant that gives us a nonlinear relation between the maximum Sharpe ratio and maturity n, even in the iid case.

Thus, the entropy conveys the term structure effects in a more intuitive fashion. Figure 2 compares the Sharpe ratios with the entropies for the iid and non-iid cases at different horizons. The dashed lines show the departures from iid for the "two-horizon" model, which are evident in the case of entropy. This is why the evidence in section 2 is presented in terms of log excess returns rather than Sharpe ratios.

3.2 Term structures of other assets

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

The pricing relation (3) gives us

$$\widehat{p}^n_t = E_t(m_{t,t+1}g_{t,t+1}\widehat{p}^{n-1}_{t+1}) = E_t(\widehat{m}_{t,t+1}\widehat{p}^{n-1}_{t+1}) = E_t\widehat{m}_{t,t+n}, \tag{14}$$

where $\widehat{m}_{t,t+1} = m_{t,t+1}g_{t,t+1}$ is the transformed pricing kernel, $\widehat{m}_{t,t+n} = \widehat{m}_{t,t+1}\widehat{m}_{t+1,t+2}\cdots\widehat{m}_{t+n-1,t+n}$, and $\widehat{p}^0_t = 1$. This has the same form as the bond pricing equation (7), with $\widehat{m}$ replacing m.

Example 2 (cash flows). We complement the pricing kernel of example 1 by adding a process for cash flow growth,

$$\log g_{t,t+1} = \log \gamma + b_0 w_{t+1} + b_1 w_t. \tag{15}$$

The transformed pricing kernel is then

$$\log \widehat{m}_{t,t+1} = \log \beta + \log \gamma + (a_0 + b_0)w_{t+1} + (a_1 + b_1)w_t \equiv \log \widehat{\beta} + \widehat{a}_0 w_{t+1} + \widehat{a}_1 w_t.$$

Our focus is on the differences between the two term structures, specifically those documented in section 2 in the mean excess returns and in the slopes and shapes of the mean yield curves. Combining equation (2) with the definition of horizon dependence, we see that the term difference in the log excess return on an asset is equal to

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

The term spread of the U.S. nominal yield curve tells us about the properties of the nominal pricing kernel m via its horizon dependence $\mathcal{H}_m(n)$, while the term difference in the log excess returns tells us about the properties of the cash flow process g because the differences in g are the only source of the differences in $\mathcal{H}_m(n) - \mathcal{H}_{\widehat{m}}(n)$.

Example 2 (cash flows, continued). We need bond prices to compute horizon dependence. When the pricing kernel is $\widehat{m}$, the expression is the same as (10), but with "hats" over the appropriate parameters. In particular, the one-period yield is

$$\widehat{y}_t^1 = -\log \widehat{\beta} - \widehat{a}_0^2/2 - \widehat{a}_1 w_t.$$

Therefore, the horizon dependence is

$$\mathcal{H}_{\widehat{m}}(n) = (1 - 1/n)[(\widehat{a}_0 + \widehat{a}_1)^2 - \widehat{a}_0^2]/2. \tag{16}$$

Thus, the average term spreads in the log excess returns are determined by the properties of the cash flow growth process. Cross-sectional differences in the term spreads are driven by cross-sectional differences in the cash flows. In our simple model, these are manifested by b_0 and b_1.

As an illustration, we use the evidence on S&P 500 dividend futures (S&P) and the British Pound (GBP) described in Tables 1 and 2. In the former case, the cash flow growth process (15) represents the equity index dividend growth, and in the latter case, it is the currency depreciation rate. The calibration proceeds similarly to the nominal U.S. yield curve. The short interest rate $\widehat{y}_t^1$ corresponds to the dividend yield on a one-period S&P strip and to the yield on a one-period U.K. bond. Their respective volatilities are 0.0402 and 0.0105. These values determine $\widehat{a}_1$ for each of the assets. The horizon dependence $\mathcal{H}_{\widehat{m}}(2)$ is -0.0016 and -0.0005, implying values of -0.0997 and -0.1005 for $\widehat{a}_0$,, respectively. These coefficients imply $b_0 = -0.0003$ and $b_1 = 0.0318$ for S&P, and $b_0 = -0.0011$ and $b_1 = 0.0021$ for GBP. This simple example indicates the potential cross-sectional differences between the cash flows that are implied by the respective term structures of their respective asset prices.

10

3.3 One-period returns

We have already shown that there is tension between the fit of the model to the term structure of nominal yields and the implied largest risk premium. We can highlight this tension further by computing the one-period expected log excess returns for different assets.

From equation (1), the one-period log excess return is

$$\log rx_{t,t+1} = \log g_{t,t+1} + \widehat{y}_t^1 - y_t^1. \tag{17}$$

In our example, its expectation is

$$E \log rx_{t,t+1} = \log \gamma - \log \widehat{\beta} - \widehat{a}_0^2/2 + \log \beta + a_0^2/2 = (a_0^2 - \widehat{a}_0^2)/2 = -b_0^2/2 - a_0 b_0.$$

In the case of our two asset classes, S&P and GBP, this expression implies $-0.29 \cdot 10^{-4}$ and $-1.09 \cdot 10^{-4}$, respectively. Consistent with our observation about the relatively small maximum risk premium in this model, the absolute values of these quantities are much smaller than those of their sample counterparts reported in Table 1.

It could be argued that the observed tension between the term spreads and one-period premiums stems from an overly simple model of the pricing kernel and cash flows. It may very well be that specific values are due to model misspecification. However, consider the general implications. The loading on the shock w_{t+1} in the pricing kernel, a_0 in this model, controls the one-period risk premium in any model. The loading has to be large because some of the one-period risk premiums are large. In contrast, the term spreads, which are controlled by the same loading on w_{t+1}, are an order of magnitude smaller than these one-period risk premiums. We confirm these observations and explore a resolution of this tension in a more realistic model in the next section.

4 Interpreting term structure evidence using an affine model

One of the implications of our discussion is that the term spreads in the log excess returns can be viewed as the term spreads in yields of bonds corresponding to the suitably transformed pricing kernel, $\widehat{m}$. Thus, the tools developed in the term structure literature can be used to model the behavior of the risk premiums on a cross-section of assets. As is the case with the U.S. nominal pricing kernel m, we want our models to deliver a large entropy of $\widehat{m}$ and changes in entropy that are consistent with the yield curve corresponding to $\widehat{m}$. Because cross-sectional differences in $\widehat{m}$ are solely due to cross-sectional differences in g, considering different transformed pricing kernels together with the U.S. nominal pricing kernel will help us to identify the properties of the cash flows. We present an affine term structure model that captures these features.

4.1 The model

Consider a simplified version of the model in Koijen, Lustig, and Van Nieuwerburgh (2015, Appendix), which we refer to as the KLV model. Specifically, we complement the (essentially) affine model of the pricing kernel in Duffee (2002) by adding a process for cash flow growth:

$$
\begin{aligned}
\log m_{t,t+1} &= \log \beta + \theta_m^\top x_t - \lambda_t^2/2 + \lambda_t w_{t+1}, \\
\log g_{t,t+1} &= \log \gamma + \theta_g^\top x_t + \eta_0 w_{t+1}, \\
x_{t+1} &= \Phi x_t + I w_{t+1},
\end{aligned}
$$

where $x_t = (x_{1t}, x_{2t})^\top$, $\lambda_t = \lambda_0 + \lambda_1 x_{1t}$, $\theta_m = (\theta_{m1}, 0)^\top$, $\Phi = \mathrm{diag}(\varphi_1, \varphi_2)$, I is the identity matrix, and $\{w_t\}$ is iid standard normal. This model is intentionally restricted compared to the most general identifiable two-factor affine model (e.g., the risk premium depends on one state only, and only one disturbance drives the dynamics of the state). Our goal is to present the simplest model that highlights the features necessary to capture the evidence.

Note that if $\lambda_1 = 0$, the pricing kernel can be re-written as

$$
\log m_{t,t+1} = \log \beta - \lambda_0^2/2 + a_0 w_{t+1} + a_1 w_t + a_2 w_{t-1} + \dots,
$$

where $a_0 = \lambda_0$, $a_1 = \theta_{m1}$, $a_j = a_{j-1}\varphi_1$, and $j \geq 2$. We recover the model of example 1 with $\varphi_1 = 0$. Thus, our simple model misses two important features: the time-variation in risk premiums and persistent state variables.

The conditional entropy of the pricing kernel,

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

is the maximum risk premium in state x_{1t}. The entropy is its mean:

$$
\mathcal{L}_m(1) = [\lambda_0^2 + \lambda_1^2/(1 - \varphi_1)^2]/2.
$$

Thus, the model has two avenues for generating high risk premiums. The first is the large values of the coefficients λ_0 and λ_1, which control the exposure to the state affecting the volatility of m. The second is the high persistence φ_1 of the state. Horizon dependence imposes discipline on the choice of their values, as we demonstrate below.

The bond prices satisfy $\log p_t^n = A_n + B_n x_t$ with

$$
\begin{aligned}
B_n &= \theta_m^\top (I + \Phi^* + \Phi^{*2} + \cdots + \Phi^{*n-1}) = \theta_m^\top (I - \Phi^*)^{-1}(I - \Phi^{*n}), \\
A_n &= n \log \beta + \lambda_0 \sum_{j=0}^{n-1} B_j^\top e + 1/2 \sum_{j=0}^{n-1} (B_j^\top e)^2, \quad e^\top = (1,1),
\end{aligned}
$$

where

$$
\Phi^* = \left(\begin{array}{cc} \varphi_1 + \lambda_1 & 0 \\ \lambda_1 & \varphi_2 \end{array} \right)
$$

is the matrix of persistence coefficients under the risk-neutral measure. These expressions are obtained by using the guess for the log bond price and applying the law of iterated expectations to (7). The result is standard in the affine term structure literature.

In particular, the one-period yield is

$$y_t^1 = -\log p_t^1 = -\log \beta - \theta_{m1} x_{1t}. \tag{18}$$

The horizon dependence is

$$\mathcal{H}_m(n) = n^{-1} A_n - A_1 = n^{-1} \left[\lambda_0 \sum_{j=0}^{n-1} B_j^\top e + 1/2 \sum_{j=0}^{n-1} (B_j^\top e)^2 \right]. \tag{19}$$

The quantities that are helpful in generating the large one-period risk premiums (λ_0, λ_1) and φ_1 are constrained by the need to fit the n-period term spreads or, equivalently, by how the largest n-period risk premium differs from the one-period premium (recall that B_n depends on λ_1).

The transformed pricing kernel has the same form

$$\log \widehat{m}_{t,t+1} = \log m_{t,t+1} + \log g_{t,t+1} \equiv \log \widehat{\beta} + \widehat{\theta}_m^\top x_t - \widehat{\lambda}_t^2/2 + \widehat{\lambda}_t w_{t+1} \tag{20}$$

with suitably redefined parameters: $\log \widehat{\beta} = \log \beta + \log \gamma + \lambda_0 \eta_0 + \eta_0^2/2$, $\widehat{\theta}_m^\top = (\theta_{m1} + \theta_{g1} + \eta_0 \lambda_1, \theta_{g2})$, $\widehat{\lambda}_t = \widehat{\lambda}_0 + \lambda_1 x_{1t}$, and $\widehat{\lambda}_0 = \lambda_0 + \eta_0$. The expression for horizon dependence when the pricing kernel is $\widehat{m}$ is the same but with parameters with a "hat":

$$
\begin{aligned}
\mathcal{H}_{\widehat{m}}(n) &= n^{-1} \left[\widehat{\lambda}_0 \sum_{j=0}^{n-1} \widehat{B}_j^\top e + 1/2 \sum_{j=0}^{n-1} (\widehat{B}_j^\top e)^2 \right], \\
\widehat{B}_n &= \widehat{\theta}_m^\top (I - \Phi^*)^{-1} (I - \Phi^{*n}).
\end{aligned}
\tag{21}
$$

Thus, the horizon dependence of the log excess returns is determined by the differences between the impacts of the state variables x_t on the nominal pricing kernel and the transformed pricing kernel (θ_m and $\widehat{\theta}_m$, respectively) and by the differences between the loadings on the shocks common to the nominal and transformed pricing kernels (λ_0 and $\widehat{\lambda}_0$). These differences, $-(\theta_{g1} + \eta_0 \lambda_1, \theta_{g2})^\top$, and η_0, respectively, are determined by the properties of the cash flow growth process, that is, by the exposure of its conditional mean to the state variables and by its exposure to the shock.

As a next step, we relate this model to data.

4.2 U.S. dollar bonds

We use the properties of the U.S. nominal Treasury data described in Tables 1 and 2 to calibrate the pricing kernel m. At a quarterly frequency, the short rate y_t^1 in equation (18) has a standard deviation of 0.0084 and an autocorrelation of 0.9487. The mean of the 40-quarter (10-year) yield spread $y^{40} - y^1$ is 0.0045; equivalently, the horizon dependence in equation (19) is -0.0045. We reproduce each of these features by choosing the parameter values $\theta_{m1} = 0.0026$, $\varphi_1 = 0.9487$, and $\lambda_0 = -0.1225$. The parameter controlling the time variation in the risk premium is set to match the curvature of the yield curve. Typically, this results in $\Phi_{11}^* \equiv \varphi_1^*$ being very close to 1. We set it to 0.9999, implying that $\lambda_1 = 0.0512$. All of these values are summarized in Panel A of Table 3. The level of the term structure can then be set by adjusting $\log \beta$.

It is important to clarify the roles of the various parameters. Here, θ_{m1} and φ_1 control the variance and autocorrelation of the short rate and λ_0 controls the slope of the mean yield curve. The different signs of θ_{m1} and λ_0 produce the upward slope in the mean yield curve. The difference in the absolute values of λ_0 and θ_{m1} (the former is roughly two orders of magnitude greater) implies a large entropy and small horizon dependence.

4.3 Other term structures

We complement the analysis in section 4.2 by characterizing the empirical properties of cash flow growth g. We keep the parameter values we used earlier for U.S. bonds, $(\theta_m, \varphi_1, \lambda_0, \lambda_1)$, and choose others, $(\theta_g, \varphi_2, \eta_0)$, to mimic the behavior of the cash flow of interest. Thus, the first set of parameters is common to all assets, while the second is asset-specific. We suppress an asset-specific notation for simplicity.

4.3.1 Foreign currency bonds

There is an extensive set of markets for bonds denominated in foreign currencies, which are linked by a similarly extensive set of currency markets. The term structure of a foreign sovereign yield curve depends on the interaction of the dollar pricing kernel and the depreciation rate of the dollar relative to a specific foreign currency, with the depreciation rate corresponding to the growth rate of the cash flow in this setting.

For symmetry between the interest rates in the U.S. and other countries and for simplicity of calibration, we assume that $\theta_{g1} = -\theta_{m1} - \lambda_1 \eta_0$ (so that $\widehat{\theta}_{m1} = 0$ in (20)). Then the one-period yield is

$$\widehat{y}_t^1 = -\log \widehat{\beta} - \theta_{g2} x_{2t}.$$

Thus, the asset-specific parameters φ_2, and θ_{g2} are calibrated by analogy with U.S. nominal bonds using serial correlation and the variance of the one-period yields. The term spread

of the foreign curve can then be used to back out $\widehat{\lambda}_0 = \lambda_0 + \eta_0$ from equation (21). Because we already know λ_0 from the U.S. curve, we can determine η_0. Panel B of Table 3 lists the calibrated values.

We observe a dramatic difference in the persistence of the cash-flow specific shock, φ_2, across the different countries. The volatility θ_{g2} and the risk premium contribution η_0 retain the same qualitative features as their U.S. counterparts in that they have different signs and the former is much smaller than the latter. Quantitatively, we observe cross-sectional variations in both parameters.

The literature views foreign exchange rates as being close to a random walk. In our model, this means $\theta_{g2} = 0$ and $\theta_{g1} = 0$. These values imply that $\widehat{B}_{n1} = (1 + \eta_0\lambda_1/\theta_{m1})B_{n1}$ and $\widehat{B}_{n2} = B_{n2} = 0$. The foreign term spread is (approximately) a scaled version of the U.S. term spread, contradicting the term structure evidence. Thus, the information captured in the term structure of the sovereign bonds provides additional information that may be useful in modeling the one-period dynamics of exchange rates.

Another implication of the calibrated model is that in contrast to many theoretical models of exchange rates, the "domestic" and "foreign" pricing kernels are asymmetric. We use quotation marks because we simply express the same projection of the pricing kernel in different units. Thus, depending on the setup of the general equilibrium model, the marginal rates of substitution of domestic and foreign economic agents could still be symmetric.

4.3.2   Inflation-linked bonds

Conceptually, the analysis of inflation-linked bonds is similar to that of foreign bonds. Exchange rates and foreign bonds tell us about the transitions between domestic and foreign economies, while the price level (CPI) and TIPS tell us about transition between the real and nominal economy. Therefore, we use the same model and the same calibration strategy in this case. We maintain the same U.S. nominal pricing kernel, so the calibration of cash flow growth, or inflation in this case, is the only novel part relative to the previous section.

Assuming that $\widehat{\theta}_{m1} = 0$ would be too restrictive in this case because of the extremely low volatility of returns associated with trading TIPS at quarterly frequency. Table 1 shows that the volatility of returns to holding TIPS is two orders of magnitude smaller than those of foreign bonds, and Table 2 shows that the difference in the term spreads of TIPS and U.S. nominal bonds is in the middle of the range for the spreads of foreign bonds. These quantities create tension between the dual role of η_0, which controls both the cross section (term spreads of bonds) and the time series (conditional volatility of cash flow growth and, therefore, returns).

Thus, we reconsider the calibration strategy of cash flow growth in the case of inflation by relaxing the zero constraint on $\widehat{\theta}_{m1}$. In this case, the real short interest rate is equal to a

linear combination of two AR(1) processes; that is, it is an ARMA(2,1):

$$
\begin{aligned}
\widehat{y}_t^1 &= \text{constant} - (\widehat{\theta}_{m1} + \widehat{\theta}_{m2})s_t, \\
s_t &= (\varphi_1 + \varphi_2)s_{t-1} - \varphi_1\varphi_2 s_{t-2} + w_t - (\widehat{\theta}_{m1}\varphi_2 + \widehat{\theta}_{m2}\varphi_1)(\widehat{\theta}_{m1} + \widehat{\theta}_{m2})^{-1}w_{t-1}.
\end{aligned}
\tag{22}
$$

See Appendix B.

First, we can calibrate $\widehat{\theta}_m$ and φ_2 by matching the variance and first- and second-order autocorrelations of $\widehat{y}_t^1$. As a second step, we can calibrate η_0 using $\mathcal{H}_{\widehat{m}}(40)$. Finally, we can calibrate θ_{g1} because $\widehat{\theta}_{m1} = \theta_{m1} + \theta_{g1} + \eta_0\lambda_1$. All of the required expressions are provided in Appendix B.

The results are reported in the first line of Table 3B. We see that the persistence of x_{2t} is much lower than in the currency examples. This is natural because we rely on two factors to model the real short interest rate.

4.3.3   Equity

Dividend strips have recently attracted interest in the literature because the term structure of the associated Sharpe ratios seems to offer *prima facie* evidence against the major asset pricing models. Although we study excess log returns instead of Sharpe ratios, a comparison of equations (4) and (13) clearly demonstrates that these objects are related.

We make best use of the available data by mixing 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 making a number of bold assumptions (see the description in Table 2 and Appendix C). All of this evidence is worth revisiting as more data become available.

Our calibrated model shares the qualitative traits of those matched to bond prices in the preceding sections. Quantitatively, we observe a dramatic drop in persistence φ_2. We note the cross-sectional variation in φ_2 appears earlier, but the equity model is the lowest (excluding the CPI model that features a two-factor structure of the relevant short interest rate). Most of the representative agent models that have been 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 that different levels of persistence of cash flows and the pricing kernel must be explored before the final opinion on the equilibrium component of these models can be expressed.

5   One-period risk premiums

The discussion in the previous section shows that evidence on the behavior of average excess returns can be translated across different horizons into the language of term structure

modeling. Given the remarkable success of the no-arbitrage modeling of zero coupon yield curves, it is not surprising that we can model the yield curves for other assets by suitably redefining the pricing kernel.

This approach circumvents two issues. First, the term structure approach focuses on the *differences* in excess returns along the maturity curve ("slope" in the term structure literature) and does not address the question of the *level* of excess returns. If the term structure model successfully captures the horizon dependence of returns, we can recast the question of the level of excess returns in terms of the one-period excess return, that is, in terms of $\log rx_{t,t+1}$. The second concern with the term structure approach is whether the same dynamic behavior can be generated in an equilibrium model. In this section, we focus on the question of the level of the term structure.

5.1 Term structure implications for one-period returns

From equation (1), the one-period log excess return is

$$\log rx_{t,t+1} = \log g_{t,t+1} + \widehat{y}_t^1 - y_t^1.$$

In the log-normal environment of the models that we have discussed, the expected log excess return is given by

$$
\begin{aligned}
E_t \log rx_{t,t+1} &= -\mathrm{var}_t(\log rx_{t,t+1})/2 - \mathrm{cov}_t(\log m_{t,t+1}, \log rx_{t,t+1}) \\
&= -\mathrm{var}_t(\log g_{t,t+1})/2 - \mathrm{cov}_t(\log m_{t,t+1}, \log g_{t,t+1}).
\end{aligned}
\tag{23}
$$

Equation (23) implies for the KLV model:

$$E_t \log rx_{t,t+1} = -\eta_0^2/2 - (\lambda_0 + \lambda_1 x_{1t})\eta_0.$$

Conditional expectations are not observable, so the theoretical counterpart to average excess returns is

$$E \log rx_{t,t+1} = -\eta_0^2/2 - \lambda_0\eta_0.$$

Here, we consider what the model that was calibrated to match the term structure evidence implies for one-period excess returns. The first column of Table 3C reports the calculated values, which depart dramatically from their data counterparts displayed in Table 1. Thus, the KLV model does a good job in matching the term structure of excess returns but not their level.

In the language of entropy, the presented model can match the horizon dependences associated with the various assets but not the one-period entropies of the respective pricing kernels. In fact, the message is more refined because one-period entropy is related to the unobserved maximal risk premium. Here, we show that the observed risk premiums on specific assets that cannot be matched exceed this model-based maximal risk premium.

Thus, we reach the same conclusion as in the simple two-horizon example. In the remainder of this section, we discuss the possible extensions of the model to rectify this shortcoming.

17

5.2 Proposed extension I: normal shocks

As noted earlier, an iid component of the pricing kernel has identical implications for horizon dependence regardless of the horizon. Thus, adding an iid component to the pricing kernel is the only avenue that will allow the one-period risk premium to be changed without affecting the implications for the term spreads.

We start by adding a normal iid shock to the KLV model:

$$\log m_{t,t+1} = \log \beta + \theta_m^\top x_t - \lambda_t^2/2 + \lambda_t w_{t+1} + \lambda_2 \varepsilon_{t+1},$$
$$\log g_{t,t+1} = \log \gamma + \theta_g^\top x_t + \eta_0 w_{t+1} + \eta_2 \varepsilon_{t+1}.$$

We calibrate (λ_2, η_2) to match the expected excess returns

$$E \log rx_{t,t+1} = -\eta_0^2/2 - \eta_2^2/2 - \lambda_0 \eta_0 - \lambda_2 \eta_2.$$

The expected excess return gives us one target for two parameters. To calibrate both parameters, we also use the unconditional variance of excess returns

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

Thus, the variance of the excess returns implies η_2; then, given η_2, the expected excess returns imply λ_2.

Table 3C reports the calculated values in the second and third columns. Both parameters have indeterminate signs, so we report their absolute values. The inferred values of λ_2 are dramatically different across the different assets. In fact, the values have to be the same because λ_2 reflects the exposure of the U.S. nominal pricing kernel to the shock ε. Thus, a normal shock is not capable of capturing the levels of the risk premiums. The statistics in Table 1 also indicate that the observed one period excess returns are non-normal, further suggesting that a non-normal shock is needed.

5.3 Coentropy

Before we proceed with a non-normal extension of the KLV model, we introduce the concept of coentropy and its properties. This will be helpful for developing a non-normal counterpart to the risk premium formula in (23) and for understanding the role of non-normality in generating realistic risk premiums.

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:

$$C(x_1, x_2) = L(x_1 x_2) - [L(x_1) + L(x_2)]. \tag{24}$$

Coentropy captures a notion of dependence of two variables. If x_1 and x_2 are independent, then $L(x_1 x_2) = L(x_1) + L(x_2)$ and $C(x_1, x_2) = 0$. If $x_1 = ax_2$ for $a > 0$, then the coentropy

is positive. If $x_1 = a/x_2$, then $L(x_1x_2) = L(a) = 0$ and the coentropy is negative. Coentropy is also invariant to noise. Consider a positive random variable y independent of x_1 and x_2 or, in other words, noise. 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 cgfs. 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 cgfs of the components are $k(s_1, 0)$ and $k(0, s_2)$. The coentropy is therefore

$$C(x_1, x_2) = k(1,1) - k(1,0) - k(0,1). \tag{25}$$

The connection to the joint cgf can be made more explicit by representing it as a sum of joint cumulants, $\kappa^{i,j}$

$$k(s_1, s_2) = \sum_{j=1}^{\infty} \sum_{p=0}^{j} \frac{\kappa^{j-p,p}}{j!} \frac{j!}{p!(j-p)!} s_1^{j-p} s_2^p,$$

This representation defines joint cumulants as the respective partial derivatives of the cgf evaluated at zero:

$$\kappa^{i,j} = \frac{\partial^{i+j} k(0,0)}{\partial s_1^i \partial s_2^j}.$$

Joint cumulants are close relatives of co-moments:

$$
\begin{aligned}
\text{mean}_1 &= \kappa^{1,0} \\
\text{mean}_2 &= \kappa^{0,1} \\
\text{variance}_1 &= \kappa^{2,0} \\
\text{variance}_2 &= \kappa^{0,2} \\
\text{covariance} &= \kappa^{1,1}
\end{aligned}
$$

Coentropy can be expressed in terms of joint cumulants:

$$
\begin{aligned}
C(x_1, x_2) &= \sum_{j=2}^{\infty} \sum_{p=1}^{j-1} \frac{\kappa^{j-p,p}}{p!(j-p)!} \\
&= \underbrace{\kappa^{1,1}}_{\text{(log)normal term}} + \underbrace{\kappa^{2,1}/2! + \kappa^{1,2}/2! + \cdots}_{\text{high-order joint cumulants}} \tag{26}
\end{aligned}
$$

Intuitively, coentropy focuses on the joint distribution of the two variables by removing all of the terms pertaining to the respective marginal distributions.

Expression (26) suggests that coentropy could deviate from covariance in a non-normal case. Consider a Poisson mixture of normals. Jumps j are Poisson with intensity ω. Conditional

19

on j jumps, $\log x \sim \mathcal{N}(j\mu, j\Delta)$, where the matrix Δ has the elements δ_{ij}. The cgf is $k(s) = \omega(e^{s^\top \mu + s^\top \Delta s/2} - 1)$. The entropies are

$$
\begin{aligned}
L(x_i) &= \omega(e^{\mu_i + \delta_{ii}/2} - 1) - \omega\mu_i \\
L(x_1 x_2) &= \omega(e^{(\mu_1 + \mu_2) + (\delta_{11} + \delta_{22} + 2\delta_{12})/2} - 1) - \omega(\mu_1 + \mu_2).
\end{aligned}
$$

The coentropy is therefore

$$
C(x_1, x_2) = \omega(e^{(\mu_1 + \mu_2) + (\delta_{11} + \delta_{22} + 2\delta_{12})/2} - e^{\mu_1 + \delta_{11}/2} - e^{\mu_2 + \delta_{22}/2} + 1).
$$

In contrast, the covariance is $\operatorname{cov}(\log x_1, \log x_2) = \omega(\mu_1\mu_2 + \delta_{12})$. We show in Appendix D that coentropy also differs from the other concepts of dependence introduced in the literature.

A numerical example illustrates this point. Let $\omega = \mu_1 = 1$ and $\Delta = 0$ (a 2-by-2 matrix of zeros). If $\mu_2 = 1$, $C(x_1, x_1) > \operatorname{cov}(x_1, x_2)$, but if $\mu_2 = -1$, the inequality goes the other way because the odd high-order cumulants change signs. Similarly, it is not difficult to construct examples in which the covariance and coentropy have opposite signs.

Another numerical example shows how different they can be. Let $\mu_1 = \mu_2 = -0.5$ and

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

We set $\rho = 0$ and $\delta = 1/\omega$. We then vary ω to identify what happens to the covariance and coentropy. We see in Figure 3 that the two can be very different. When the jump intensity ω is low, the coentropy exceeds the covariance, but as the jump intensity increases, the coentropy becomes smaller than the covariance between the two processes.

5.4 Proposed extension II: Poisson shocks

We introduce jumps into the KLV model so that the new pricing kernel and cash flow growth specification is

$$
\begin{aligned}
\log m_{t,t+1} &= \log \beta + \theta_m^\top x_t - \lambda_t^2/2 + \lambda_t w_{t+1} + \lambda_2 z_{t+1}^m, &(27) \\
\log g_{t,t+1} &= \log \gamma + \theta_g^\top x_t + \eta_0 w_{t+1} + \eta_2 z_{t+1}^g, &(28)
\end{aligned}
$$

where z_t^m and z_t^g are compound Poisson processes with the same arrival rate of ω and jump size distributions of $\mathcal{N}(\mu_m, \delta_m^2)$ and $\mathcal{N}(\mu_g, \delta_g^2)$, respectively. For the jumps to cash flow growth to be priced, the jump processes z_t^m and z_t^g must have coincident jumps. Thus, the only difference between the non-normal innovations to the pricing kernel and the cash flow is in the jump size.

In this case, the expression for expected excess returns (23) no longer applies. However, the main asset valuation equation (3) still applies, and we use it to derive a generalization of (23) to non-normal settings. Taking the logs of (3), we obtain

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

This equation implies that the (log) risk premium is

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

and the expected (log) excess return is

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

Here, $C_t(x_{1t+1}, x_{2t+1})$ is the conditional version of the definition in (24).

The last equation implies that in the case of zero-coupon claims,

$$
E_t \log rx_{t,t+1} = -L_t g_{t,t+1} - C_t(m_{t,t+1}, g_{t,t+1}).
$$

As in the normal examples, we have unconditional expectations, so it is helpful to introduce the additional notation $\mathcal{C}_{mg}(n) = n^{-1} E C_t(m_{t,t+n}, g_{t,t+n})$. The definition of coentropy implies

$$
\mathcal{C}_{mg}(n) - \mathcal{C}_{mg}(1) = \mathcal{H}_{\widehat{m}}(n) - \mathcal{H}_m(n) - \mathcal{H}_g(n). \tag{29}
$$

Although the horizon dependence is not affected by the addition of the Poisson iid shock, the entropy of the pricing kernel is affected:

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

Thus, it is easy to compute the $n-$period entropy via $\mathcal{L}_m(n) = \mathcal{L}_m(1) + \mathcal{H}_m(n)$, where $\mathcal{H}_m(n)$ is exactly the same as in (19). The one-period coentropy is

$$
\mathcal{C}_{mg}(1) = \lambda_0 \eta_0 + k_z(\lambda_2, \eta_2) - k_z(\lambda_2, 0) - k_z(0, \eta_2) \tag{30}
$$

with

$$
k_z(s_1, s_2) = \omega(e^{s_1 \mu_m + s_2 \mu_g + (s_1^2 \delta_m^2 + s_2^2 \delta_g^2)/2} - 1). \tag{31}
$$

Equation (29) implies the $n-$period coentropy.

Although $\mathcal{C}_{mg}(n)$ is affected by the jump components for any n, the change in coentropy across horizons is not – a result of the jump components being iid. Thus, $\mathcal{C}_{mg}(n) - \mathcal{C}_{mg}(1)$

is affected only by the differences in the covariances between $\log m$ and $\log g$ at different horizons.

To calibrate the jump parameters, we normalize the jump loadings λ_2 and η_2 to 1 because they are not identified separately from the jump volatilities δ_m and δ_g, respectively. We borrow the parameters that control the jumps in the pricing kernel from the CI2 model of Backus, Chernov, and Zin (2014): $\omega = 0.01/4$, $\mu_m = -10 \cdot (-0.15) = 1.5$, $\delta_m^2 = (-10 \cdot 0.15)^2 = 1.5^2$. These parameter values represent a milder version of the Barro (2006) disaster calibration, as we discuss in the context of a consumption-based model in the next section.

We can use information about cash flows or, equivalently, about one-period excess returns, to infer the asset-specific η_2, and μ_g. The one-period excess (log) returns are

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

Thus,

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

and

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

Table 3B reports the results of the calibration procedure. As discussed in the example of a bivariate Poisson process, the non-normality of the pricing kernel and cash flow growth manifests itself in large differences between coentropy and covariance, as reported in Table 3C. These substantial deviations of coentropy from covariance highlight the ability of models with non-normal innovations to generate large expected returns and large cross-sectional differences between them.

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

A second reality check is to use the calibrated model of equity to see what it implies for the equity premium. The value of an equity claim is an infinite sum of zero-coupon claims, so the model-implied premium should be consistent with the one in the data. In Appendix E,

22

we explain how we solve this for the equity premium in our model and demonstrate that the premium is indeed matched.

We want to highlight how the non-normalities featured in our model affect the coentropy and, therefore, the risk premiums via the joint cumulants of the pricing kernel and cash flows. Because the (log) normal component, $\lambda_0\eta_0$, of coentropy in (30) is standard and well-explored in the literature, we focus on the effect the jumps have on coentropy using the decomposition in equation (26).

The joint cgf associated with jumps in our model (27)-(28) is provided in equation (31). The cumulants corresponding to the marginal distributions have the following well-known form:

$$
\begin{aligned}
\kappa^{1,0} &= \omega\mu_m, & \kappa^{0,1} &= \omega\mu_g, \\
\kappa^{2,0} &= \omega(\mu_m^2 + \delta_m^2), & \kappa^{0,2} &= \omega(\mu_g^2 + \delta_g^2), \\
\kappa^{3,0} &= \omega\mu_m(\mu_m^2 + 3\delta_m^2), & \kappa^{0,3} &= \omega\mu_g(\mu_g^2 + 3\delta_g^2), \\
\kappa^{4,0} &= \omega(\mu_m^4 + 6\mu_m^2\delta_m^2 + 3\delta_m^4), & \kappa^{0,4} &= \omega(\mu_g^4 + 6\mu_g^2\delta_g^2 + 3\delta_g^4),
\end{aligned}
$$

and so on. Because the jump size distributions of the two variables are independent of each other, the joint cumulants have a particularly transparent representation:

$$
\kappa^{i,j} = \omega^{-1}\kappa^{i,0}\kappa^{0,j}. \tag{32}
$$

This expression clarifies how joint cumulants contribute to generating risk premiums. The non-normalities present in both variables and reflected in the magnitudes of the respective cumulants reinforce each other multiplicatively in the joint cumulants. Further, because the first term, $\kappa^{i,0}$, reflects the properties of the pricing kernel, the cross-sectional differences in the risk premiums are determined by the marginal cumulants of the cash flow process.

The question is which of the joint cumulants matter quantitatively in the model presented here. The property (32) allows us to obtain a simple interpretation of coentropy. The high-order cumulants of the log pricing kernel are very large—larger than 10 for $i > 2$. Regardless of the asset, the cumulants of the cash flows are relatively small. Thus, it is helpful to use the log scale to compare the quantitative effects of the two. On the log scale, each joint cumulant is simply a sum of the two log marginal cumulants, up to a constant.

To illustrate the quantitative effect of coentropy, consider the case of GBP. The first row of Figure 4 displays the log marginal cumulants. Each log joint cumulant with an index (i, j) is equal to the sum of the two marginals with indexes i and j, as displayed in panel C of the figure. We see that the log joint cumulants corresponding to a large i are dominated by the properties of the pricing kernel, and vice versa for large j.

Panel C ignores the fact that the contributions of joint cumulants to coentropy are divided by $i!j!$, so the effects of the higher-order terms on risk premiums and coentropy should be diminished. Figure 4D demonstrates how fast this occurs for GBP by showing the joint

23

cumulants (not logs) scaled by the factorials. We see that the quantitative effect of high-order joint cumulants decreases quickly. The cash flow matters for $j = 1, 2$, while the effect of the pricing kernel starts to decline at $i = 4$ but is still visible at $i = 10$.

Figure 5 displays the contribution of the joint cumulants to coentropy for all of the other assets. Qualitatively, the implications are the same in that the non-normalities in the pricing kernel matter much more than those of the cash flows. Obviously, if the cash flows have no non-normalities, the joint cumulants will be equal to zero. Thus, despite being small, the non-normal shocks in the cash flows play an important role for the risk premiums.

6   The representative agent with recursive preferences

We have demonstrated that the evidence on risk premiums can be represented with affine term structure models that feature different levels of persistence of the pricing kernel and cash flows to capture the term differences in the risk premiums and an iid jump component to capture the level of premiums. These insights are important because they highlight the features that a pricing kernel and cash flows should possess to match the empirical evidence. However, this analysis does not address the question of whether the evidence is consistent with an equilibrium model.

We would like to emphasize the reasons for considering an equilibrium model. It is not our intent to offer an improvement of the existing models. We merely want to demonstrate how certain features of the data translate into required features in a model. We have already highlighted such features in the affine framework. The issue is whether the restrictions imposed by the economic theory affect the model's ability to produce quantitatively realistic results. Although we argue that the resulting model is reasonable according to basic metrics, we leave it to our readers to pursue this formulation or its extensions in their research.

We focus on an endowment economy with a representative agent. Because such an economy delivers implications for the real pricing kernel, we focus on whether we can generate something similar to the real affine pricing kernel implied by the combination of the nominal pricing kernel and the growth process for CPI (inflation). To recap, we have

$$
\begin{aligned}
\log \widehat{m}_{t,t+1} &= (\log \beta + \log \gamma + \lambda_0 \eta_0 + \eta_0^2/2) + (\theta_{m1} + \theta_{g1} + \eta_0 \lambda_1) x_{1t} + \theta_{g2} x_{2t} \\
&\quad - (\lambda_0 + \eta_0 + \lambda_1 x_{1t})^2/2 + (\lambda_0 + \eta_0 + \lambda_1 x_{1t}) w_{t+1} + \lambda_2 z_{t+1}^m + \eta_2 z_{t+1}^g \\
&\equiv \log \widehat{\beta} + \widehat{\theta}_m^\top x_t - \widehat{\lambda}_t^2/2 + \widehat{\lambda}_t w_{t+1} + \widehat{z}_{t+1}^m,
\end{aligned}
\tag{33}
$$

where jumps $\widehat{z}_{t+1}^m$ arrive at rate ω with jump sizes $\mathcal{N}(\widehat{\mu}_m, \widehat{\delta}_m^2)$ with $\widehat{\mu}_m = \lambda_2 \mu_m + \eta_2 \mu_g$ and $\widehat{\delta}_m^2 = \lambda_2^2 \delta_m^2 + \eta_2^2 \delta_g^2$. We conduct a reverse-engineering exercise in which the specification of consumption growth is motivated by the objective of having a similar functional form of the consumption-based real pricing kernel.

6.1 Equilibrum real pricing kernel

We assume that there is a representative agent with recursive preferences, as developed by Kreps and Porteus (1978), Epstein and Zin (1989), and Weill (1989), among many others. The benefit of this specification is that a log-linear approximation of the agent's value function leads to a (restricted) affine pricing kernel. Thus, the insights that we have gained from the no-arbitrage case have the best chance of being successfully translated into an equilibrium setting.

We define utility with the time aggregator,

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

and certainty equivalent function,

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

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

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

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

where $g_{t,t+1}^c = c_{t+1}/c_t$ is consumption growth. This relation serves as the recursive utility analog of a classical Bellman equation.

With this utility function, the real pricing kernel is

$$\widehat{m}_{t,t+1} = \beta(g_{t,t+1}^c)^{\rho-1}[g_{t,t+1}^c u_{t+1}/\mu_t(g_{t,t+1}^c u_{t+1})]^{\alpha-\rho}. \tag{36}$$

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

$$\log g_{t,t+1}^c = g^c + \theta_c^\top x_t + (\sigma_0 + \sigma_1 x_{1t})w_{t+1} + z_{t+1}^c, \tag{37}$$

where jumps arrive at the rate of ω and the jump sizes are distributed as $\mathcal{N}(\mu_c, \delta_c^2)$. The factors x_t are as the same as in the KLV model.

This specification has two novel features that arise directly from the functional form of the affine pricing kernel. First, the conditional volatility of the consumption growth is captured by a normally distributed variable. Because the sign of the volatility is not determined, it is appropriate to use such a specification. Commonly used specifications use either a normal variable for the variance of consumption growth or model the variance via the square-root

or ARG processes. We propose our specification to match the essentially affine form of the affine $\widehat{m}_{t,t+1}$. It could be a useful alternative in future applications. Second, the expected consumption growth follows an ARMA(2,1) process instead of the commonly used AR(1) process.

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

$$\log u_t \approx u_0 + u_1 \log \mu_t(g^c_{t,t+1} u_{t+1}) \tag{38}$$

around the point $\log \mu_t = E(\log \mu_t)$. This approximation is exact when $\rho = 0$, in which case $u_0 = 0$ and $u_1 = \beta$.

We guess a value function of the form

$$\log u_{t+1} = u + a_1 x_{1t+1} + a_2 x^2_{1t+1} + b x_{2t+1}.$$

We verify this guess and derive the values of a_1, a_2, and b in Appendix F.1. Then equation (36) implies the real pricing kernel

$$
\begin{aligned}
\log \widehat{m}_{t+1} &= \widehat{m} + [(\rho-1)\theta_{c1} - (\alpha-\rho)\alpha(\sigma_0 + a_1 + b)(\sigma_1 + 2a_2\varphi_1)(1-2\alpha a_2)^{-1}]x_{1t} \\
&+ (\rho-1)\theta_{c2}x_{2t} - (\alpha-\rho)\alpha(\sigma_1 + 2a_2\varphi_1)^2(1-2\alpha a_2)^{-1}x^2_{1t}/2 \\
&+ [(\alpha-1)\sigma_0 + (\alpha-\rho)(a_1 + b) + ((\alpha-1)\sigma_1 + (\alpha-\rho)2a_2\varphi_1)x_{1t}]w_{t+1} \\
&+ (\alpha-\rho)a_2 w^2_{t+1} + (\alpha-1)z^c_{t+1}, \tag{39}
\end{aligned}
$$

where $\widehat{m}$ is a constant whose explicit expression as a function of the model parameters is omitted.

We calibrate the model to match the properties of the affine pricing kernel (33), see Appendix F.2. The calibrated preference parameters and parameters controlling the dynamics of consumption are listed in Panel D of Table 3. Table 4 shows that the model-implied positive serial correlation of consumption growth is similar to the empirical one. Yet in contrast to the traditional implementations of the long-run risk paradigm, the real yield curve is upward-sloping. This property arises from the specification of the consumption volatility. Intuitively, x_{1t}, which affects the conditional volatility of the pricing kernel in line 3 of (39), acts similarly to a habit or a preference shock (e.g., Creal and Wu, 2016; Wachter, 2006). The technical details of how this works are provided in Appendix F.3. Thus, our model of consumption growth transcends the specific objective of matching the affine pricing kernel and could be useful in other equilibrium setups.

The serial correlation of the expected consumption growth can be computed using the formulas in Appendix B and is equal to 0.6786. This number is much lower than is typically used in the long-run risk literature. Implicitly, this number is determined by the shape of the real yield curve. As a result, the normal component of the model will not be able to generate risk premiums of realistic magnitudes. Here, jumps in consumption help in matching the levels of the risk premiums. The calibrated jump parameters are consistent with a modest version of the Barro (2006) disaster model in that jumps take place once in a hundred years, and the average jump in consumption is -15%, with a volatility of 15%.

6.2 Nominal pricing kernel and other assets

To obtain the nominal pricing kernel in our endowment economy, we assume an exogenous process of inflation as in Bansal and Shaliastovich (2013), Piazzesi and Schneider (2006), and Wachter (2006). Specifically, we use the process specified and calibrated in sections 4.3.2 and 5.4. Given that our consumption-based real pricing kernel is quite close to the affine one, the nominal pricing kernel will be (nearly) matched by construction. Following the same logic, we can match the transformed pricing kernels associated with currencies and equities using the calibrated cash flows (28).

Given that the state variables x_{1t} and x_{2t} control the expected consumption growth in the recursive model, we can reinterpret the cash flow model in the context of what is typically used in endowment economies. The Bansal and Yaron (2004) specification expects consumption growth to be determined by a single AR(1) factor, and an asset's expected cash flow growth has a different exposure to this factor. Our model has two AR(1) factors, and the exposure to both changes as we move from expected consumption growth to expected cash flow growth. Put differently, we can write expected consumption growth as an ARMA(2,1) process, while expected cash flow growth cannot be written as a different exposure to the same ARMA(2,1) process, because both the exposure to common shocks and the process change. These two departures from the traditional specifications are helpful in matching the observed patterns of average multi-horizon returns.

7 Concluding remarks

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

Theoretically, we introduce the concept of coentropy, which serves as a generalized measure of covariance in the non-normal, multi-period world. Coentropy of the pricing kernel and cash flows is closely related to the documented cross-sectional differences in yields. Thus, these differences in yields must reflect the differences in cash flows. We show that 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 and consumption that features disasters and persistent variation in its expected value is capable of capturing the evidence.

A 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 ν and associated positive eigenfunction v_t satisfying

$$E_t\big(m_{t,t+1}v_{t+1}\big) \;\;=\;\; \nu v_t. \tag{40}$$

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

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

By construction $E_t(m^1_{t,t+1}) = 1$, 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: $r^\infty_{t,t+1} = \lim_{n\to\infty} r^n_{t,t+1} = 1/m^2_{t,t+1} = v_{t+1}/(\nu v_t)$, so that $E(\log r^\infty) = -\log \nu$. See Alvarez and Jermann (2005, Section 3).

The special case $m^1_{t,t+1} = 1$ has gotten a lot of recent attention; see, for example, the review in Borovicka, Hansen, and Scheinkman (2016). The pricing kernel becomes $m_{t,t+1} = m^2_{t,t+1}$. Since the long bond return is its inverse, the long bond is the high return asset. 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 1 ("two-horizon" price of risk, continued). We guess an eigenvector of the form $\log v_t = c_0 w_t + c_1 w_{t-1}$. If we substitute into (40) we find:

$$c_0 = a_1, \quad c_1 = 0, \quad \log \nu = \log \beta + (a_0 + a_1)^2/2.$$

Horizon dependence is

$$\mathcal{H}_m(\infty) = \log \nu - E \log m_{t,t+1} = \log \nu - \log \beta = (a_0 + a_1)^2/2.$$

If $a_1 = -a_0$, then $m^1_{t,t+1} = 1$, and $\mathcal{H}_m(\infty) = 0$.

Moving on to other assets, we introduce two equation analogous to (40). One is for cashflow growth:

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

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

$$E_t(\widehat{m}_{t,t+1}\widehat{v}_{t+1}) = \widehat{\nu}\widehat{v}_t. \tag{41}$$

leading to a decomposition $\widehat{m}_{t,t+1} = \widehat{\nu}\widehat{m}^1_{t,t+1}\widehat{v}_t/\widehat{v}_{t+1}$.

The decompositions are related to each other via:

$$\widehat{\nu}\widehat{m}^1_{t,t+1}\widehat{v}_t/\widehat{v}_{t+1} = \widehat{m}_{t,t+1} \equiv m_{t,t+1}g_{t,t+1} = \nu\xi m^1_{t,t+1}g^1_{t,t+1}(v_tu_t)/(v_{t+1}u_{t+1}). \tag{42}$$

There's not, in general, a close relation between $\widehat{\nu}$, ν, and ξ, but there is in some special cases. One special case is a stationary cash flow, which leads to the martingale component $g^1_{t,t+1} = 1$ as in the example above. In this case, the simplified equation (42) implies that the value $\nu\xi$ and function v_tu_t solve equation (41). Therefore, $\widehat{\nu} = \nu\xi$, the martingale components coincide, $\widehat{m}^1_{t,t+1} = m^1_{t,t+1}$, and long-horizon excess returns are equal to zero:

$$E\log rx_{t,t+n} \to 0, \quad \text{as } n \to \infty. \tag{43}$$

The reverse is also true: if $\widehat{m}^1_{t,t+1} = m^1_{t,t+1}$, it must be the case that $g^1_{t,t+1} = 1$. Indeed, in this case equation (42) implies that the level of $g^1_{t,t+1}(v_tu_t)/(v_{t+1}u_{t+1})$ must be stationary because $\widehat{v}_t$ is. Because v_t and u_t are stationary as well, the martingale $g^1_{t,t+1}$ must be a constant (we can normalize it to one w.l.o.g.).

Example 2 (cash flows, continued). The Perron-Frobenius theory implies $\log u_t = d_0w_t + d_1w_{t-1}$ with

$$d_0 = b_1, \quad d_1 = 0, \quad \log\xi = \log\gamma + (b_0 + b_1)^2/2.$$

and $\log\widehat{v}_t = \widehat{c}_0w_t + \widehat{c}_1w_{t-1}$ with

$$\widehat{c}_0 = \widehat{a}_1, \quad \widehat{c}_1 = 0, \quad \log\widehat{\nu} = \log\widehat{\beta} + (\widehat{a}_0 + \widehat{a}_1)^2/2.$$

If If $b_1 = -b_0$, then cash flow is stationary, $\log\xi = \log\gamma$, and $\log\widehat{\nu} = \log\beta + \log\gamma + (a_0 + a_1)^2/2 = \log\nu + \log\xi$.

Another special case is one in which the "price-dividend" ratio $\widehat{p}$ is constant, see the October 2005 version of Hansen, Heaton, and Li (2008), section 3.2. Consider a factorization of the dividend d_t 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 $g^1_{t,t+1} = 1$.) Because s_t is stationary, the two transformed pricing kernels $\widehat{m}_{t,t+1}$ and $m^*_{t,t+1} \equiv m_{t,t+1}g^*_{t,t+1}$ will have the same eigenvalue $\widehat{\nu}$. The eigenfunctions will be $\widehat{v}_t$ and $\widehat{v}_t \cdot s_t$, respectively. Thus, if a dividend is such that its $\widehat{v}_t = 1$, or, equivalently, s_t equals the eigenfunction associated with $m^*_{t,t+1}$, then $\widehat{p}$ is constant.

B Details of the inflation model

First we show that a sum of two AR(1) processes is an ARMA(2,1). Consider

$$s_t = \frac{\widehat{\theta}_{m1}}{\widehat{\theta}_{m1} + \widehat{\theta}_{m2}} x_{1t} + \frac{\widehat{\theta}_{m2}}{\widehat{\theta}_{m1} + \widehat{\theta}_{m2}} x_{2t}.$$

Then

$$
\begin{aligned}
(1 - \varphi_1 L)(1 - \varphi_2 L)s_t &= \frac{\widehat{\theta}_{m1}}{\widehat{\theta}_{m1} + \widehat{\theta}_{m2}}(1 - \varphi_2 L)w_t + \frac{\widehat{\theta}_{m2}}{\widehat{\theta}_{m1} + \widehat{\theta}_{m2}}(1 - \varphi_1 L)w_t \\
&= \left(1 - \left(\frac{\widehat{\theta}_{m1}}{\widehat{\theta}_{m1} + \widehat{\theta}_{m2}}\varphi_2 + \frac{\widehat{\theta}_{m2}}{\widehat{\theta}_{m1} + \widehat{\theta}_{m2}}\varphi_1\right) L\right) w_t.
\end{aligned}
$$

So

$$s_t = \phi_1 s_{t-1} + \phi_2 s_{t-2} + w_t + \theta_1 w_{t-1} \tag{44}$$

with

$$\phi_1 = \varphi_1 + \varphi_2, \quad \phi_2 = -\varphi_1\varphi_2, \quad \theta_1 = -\left(\frac{\widehat{\theta}_{m1}}{\widehat{\theta}_{m1} + \widehat{\theta}_{m2}}\varphi_2 + \frac{\widehat{\theta}_{m2}}{\widehat{\theta}_{m1} + \widehat{\theta}_{m2}}\varphi_1\right).$$

The first step of calibration requires the knowledge of the unconditional moments of ARMA(2,1). Denote variance by v_0 and autocovariances by v_j. Then multiply s_{t-j} on both side of (44) and take expectation:

$$v_j = \phi_1 v_{j-1} + \phi_2 v_{j-2} + Ew_t s_{t-j} + \theta_1 Ew_{t-1} s_{t-j}.$$

We are interested in $j = 0, 1, 2$. We use $v_{-i} = v_i$; $Ew_{t-i}s_{t-j} = 0$ if $j > i$, and $Ew_{t-i}s_{t-j} = \psi_{i-j}$, if $j \leq i$ where ψ_j is a coefficient in the MA representation of s_t. As a result we get the following linear system of equations in the variables of interest:

$$
\begin{aligned}
v_0 &= \phi_1 v_1 + \phi_2 v_2 + 1 + \theta_1(\phi_1 + \theta_1), \\
v_1 &= \phi_1 v_0 + \phi_2 v_1 + \theta_1, \\
v_2 &= \phi_1 v_1 + \phi_2 v_0.
\end{aligned}
$$

The solution is:

$$
\begin{aligned}
v_0 &= D^{-1}(1 + 2\theta_1\phi_1 - \theta_1^2(\phi_2 - 1) - \phi_2), \\
v_1 &= D^{-1}(\phi_1 + \theta_1^2\phi_1 + \theta_1(1 + \phi_1^2 - \phi_2^2)), \\
v_2 &= D^{-1}(\phi_2 + \phi_1^2 - \phi_2^2 + \theta_1^2(\phi_2 + \phi_1^2 - \phi_2^2) + \theta_1\phi_1(1 + \phi_1^2 + 2\phi_2 - \phi_2^2)), \\
D &= ((\phi_2 - 1)^2 - \phi_1^2)(1 + \phi_2).
\end{aligned}
$$

Having obtained these moments, we can construct serial autocorrelations of $\widehat{y}_t^1 = \text{constant} - (\widehat{\theta}_{m1} + \widehat{\theta}_{m2})s_t$: v_1/v_0, and v_2/v_0, and its variance: $v_0(\widehat{\theta}_{m1} + \widehat{\theta}_{m2})^2$.

C 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 U.S. nominal zero-coupon bond that pays \$1 at time $t+n$.

Then,

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

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

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

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 $\widehat{y}_t^1$:

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

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.

D   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:

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

To relate coentropy to mutual information and copula, define new probabilities: $\tilde{p}(x_1, x_2) = p(x_1, x_2) x_1 x_2 / E(x_1 x_2)$, and $-j$ denotes "not j". We have the following marginals

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

Therefore,

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

In words, coentropy is the difference between mutual informations corresponding to two different joint probabilities. 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:

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

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.

E   The equity premium

The main objective of this section is to establish equity premium in the KLV model that was calibrated to match returns on dividend strips. We exploit the Campbell-Shiller loglinearaization

$$\log r_{t,t+1} = \log g_{t,t+1} - \log pd_t + \kappa \log pd_{t+1},$$

where $pd_t = \widehat{p}_t/d_t$ is the price-to-dividend ratio, and $\kappa = E(pd_t)/(1 + E(pd_t))$.

Then (3) implies

$$1 = E_t e^{\log m_{t,t+1} + \log g_{t,t+1} - \log pd_t + \kappa \log pd_{t+1}},$$

or, equivalently,

$$\log pd_t = \log E_t e^{\log \widehat{m}_{t,t+1} + \kappa \log pd_{t+1}}.$$

Guess $\log pd_t = A + Bx_t$, and solve for A and B by plugging the guess into the valuation equation. We obtain:

$$
\begin{aligned}
A &= (1-\kappa)^{-1}[\log \widehat{\beta} + \widehat{\lambda}_0 \kappa B^\top e + (\kappa B^\top e)^2/2 + k_z(\lambda_2, \eta_2)], \\
B &= \widehat{\theta}^\top (I - \kappa \Phi^*)^{-1}.
\end{aligned}
$$

The risk free rate in the presence of jumps is

$$y_t^1 = -\log E_t e^{\log m_{t,t+1}} = -\log \beta - k_z(\lambda_2, 0) - \theta_{m1} x_{1t}.$$

Therefore, excess log returns are

$$\log rx_{t,t+1} = \log g_{t,t+1} - A - Bx_t + \kappa A + \kappa B x_{t+1} + \log \beta + k_z(\lambda_2, 0) + \theta_{m1} x_{1t}$$

and expected excess returns are

$$
\begin{aligned}
E \log rx_{t,t+1} &= \log \gamma + \omega \eta_2 \mu_g - [\log \widehat{\beta} + \widehat{\lambda}_0 \kappa B^\top e + (\kappa B^\top e)^2/2 + k_z(\lambda_2, \eta_2)] + \log \beta + k_z(\lambda_2, 0) \\
&= -\lambda_0 \eta_0 - \eta_0^2/2 - \widehat{\lambda}_0 \kappa B^\top e - (\kappa B^\top e)^2/2 + k_z(\lambda_2, 0) + \omega \eta_2 \mu_g - k_z(\lambda_2, \eta_2).
\end{aligned}
$$

Given the calibrated parameters for zero-coupon equity claim and a commonly used value of $\kappa = 0.963$, the implied equity premium is 0.0033, or 1.33% per year. While this appears to be on the low end of the customary equity premium estimates (the Shiller dataset implies 4% with a standard error of 1.65%), the number is close to average realized log excess returns on S&P 500 of 1.34% during 1996-2011. This is the period that was effectively used for calibration of equity cash flow growth as it is the sample corresponding to data from Binsbergen, Brandt, and Koijen (2012) and Binsbergen, Hueskes, Koijen, and Vrugt (2013).

F   The recursive utility pricing kernel

F.1   Derivation

We derive the pricing kernel for a representative agent model with recursive utility, loglinear consumption growth dynamics, stochastic volatility, and jumps with constant intensity.

We guess a value function of the form

$$
\begin{aligned}
\log u_{t+1} &= u + a_1 x_{1t+1} + a_2 x_{1t+1}^2 + b x_{2t+1} \\
&= u + a_1 \varphi_1 x_{1t} + a_2 \varphi_1^2 x_{1t}^2 + b \varphi_2 x_{2t} + (a_1 + b + 2 a_2 \varphi_1 x_{1t}) w_{t+1} + a_2 w_{t+1}^2.
\end{aligned}
$$

Then

$$
\begin{aligned}
\log(g_{t,t+1} u_{t+1}) &= g + u + (\theta_{c1} + a_1 \varphi_1) x_{1t} + (\theta_{c2} + b \varphi_2) x_{2t} + a_2 \varphi_1^2 x_{1t}^2 \\
&+ [\sigma_0 + a_1 + b + (\sigma_1 + 2 a_2 \varphi_1) x_{1t}] w_{t+1} + a_2 w_{t+1}^2 + z_{t+1}^c.
\end{aligned}
$$

Therefore, using the identity

$$
\log E_t e^{\alpha_1 w_{t+1} + \alpha_2 w_{t+1}^2} = -\frac{1}{2}[\log(1 - 2\alpha_2) - \alpha_1^2 (1 - 2\alpha_2)^{-1}],
$$

we obtain

$$
\begin{aligned}
\log \mu_t(g_{t,t+1} u_{t+1}) &= g + u - \frac{1}{2\alpha} \log(1 - 2\alpha a_2) + \frac{\alpha}{2}(\sigma_0 + a_1 + b)^2 (1 - 2\alpha a_2)^{-1} + \alpha^{-1} \omega (e^{\alpha \mu_c + \alpha^2 \delta_c^2/2} - 1) \\
&+ [\theta_{c1} + a_1 \varphi_1 + \alpha(\sigma_0 + a_1 + b)(\sigma_1 + 2 a_2 \varphi_1)(1 - 2\alpha a_2)^{-1}] x_{1t} + [\theta_{c2} + b \varphi_2] x_{2t} \\
&+ [a_2 \varphi_1^2 + 0.5\alpha(\sigma_1 + 2 a_2 \varphi_1)^2 (1 - 2\alpha a_2)^{-1}] x_{1t}^2.
\end{aligned}
$$

Lining up terms in expression (38), we get $b = u_1 \theta_{c2}(1 - u_1 \varphi_2)^{-1}$, and

$$
a_1 = u_1(\theta_{c1} + \alpha(\sigma_0 + b)(\sigma_1 + 2 a_2 \varphi_1)(1 - 2\alpha a_2)^{-1})(1 - u_1(\varphi_1 + \alpha(\sigma_1 + 2 a_2 \varphi_1)(1 - 2\alpha a_2)^{-1}))^{-1},
$$

while a_2 is the negative root of

$$
a_2 = u_1(a_2 \varphi_1^2 + 0.5\alpha(\sigma_1 + 2 a_2 \varphi_1)^2 (1 - 2\alpha a_2)^{-1}).
$$

We select the negative root because in this case $a_2 \to 0$ when $\sigma_1 \to 0$.

Then equation (36) implies the real pricing kernel in (39).

F.2 Matching the affine real pricing kernel

We introduce the following shorthand notation for the various elements of the real pricing kernel (39):

$$
\begin{aligned}
\log \widehat{m}_{t+1} &= \widehat{m} + [(\rho - 1)\theta_{c1} - (\alpha - \rho)\alpha(\sigma_0 + a_1 + b)(\sigma_1 + 2a_2\varphi_1)(1 - 2\alpha a_2)^{-1}]x_{1t} \\
&+ (\rho - 1)\theta_{c2}x_{2t} - (\alpha - \rho)\alpha(\sigma_1 + 2a_2\varphi_1)^2(1 - 2\alpha a_2)^{-1}x_{1t}^2/2 \\
&+ [(\alpha - 1)\sigma_0 + (\alpha - \rho)(a_1 + b) + ((\alpha - 1)\sigma_1 + (\alpha - \rho)2a_2\varphi_1)x_{1t}]w_{t+1} \\
&+ (\alpha - \rho)a_2 w_{t+1}^2 + (\alpha - 1)z_{t+1}^c, \\
&\equiv \widehat{m} + \theta_1^{rec}x_{1t} + \theta_2^{rec}x_{2t} - \vartheta^{rec}x_{1t}^2/2 + (\lambda_0^{rec} + \lambda_1^{rec}x_{1t})w_{t+1} \\
&+ \lambda_2^{rec}w_{t+1}^2 + z_{t+1}^{rec}. \qquad\qquad (45)
\end{aligned}
$$

Observe that both the affine, (33), and the recursive, (45), pricing kernels can be written as

$$
\log \widehat{m}_{t+1} = -\widehat{y}_t^1 - \text{convexity} + v_t^{1/2}w_{t+1} + v_2 w_{t+1}^2 + \widehat{z}_{t+1}^m,
$$

where $\widehat{y}_t^1$ is given in (22) in the affine case and

$$
\widehat{y}_t^1 = \text{constant} - (\theta_1^{rec} + \lambda_0^{rec}\lambda_1^{rec})x_{1t} - \theta_2^{rec}x_{2t} + (\vartheta^{rec} - [\lambda_1^{rec}]^2)x_{1t}^2/2
$$

in the recursive case; the convexity, ignoring constants, is $\widehat{\lambda}_0\widehat{\lambda}_1 x_{1t} + \widehat{\lambda}_1^2 x_{1t}^2/2$ and $\lambda_0^{rec}\lambda_1^{rec}x_{1t} + [\lambda_1^{rec}]^2 x_{1t}^2/2$. The volatility of the pricing kernel, $v_t^{1/2}$, is $\widehat{\lambda}_0 + \widehat{\lambda}_0 x_{1t}$ and $\lambda_0^{rec} + \lambda_1^{rec}x_{1t}$. Finally, v_2 is zero in the affine case and λ_2^{rec} in the recursive case.

The convexity term in (45) does not offset the quadratic term in the interest rate completely, as it does in (33). Moreover, the structural model features the squared shock w_{t+1}^2 in the pricing kernel, and such a term is absent from the affine pricing kernel. Thus, we will not be able to construct a perfect equilibrium counterpart to our affine model. However, this is not our goal, as we simply want to obtain a model that is quantitatively consistent with the highlighted features of the data, and we use an affine model to guide this search.

We focus on calibrating the parameters controlling consumption growth θ_{1c}, θ_{2c}, σ_0, and σ_1 to match the linear components of the interest rates and conditional volatilities of the pricing kernels. The jump parameters have an intuitive one-to-one mapping. In addition, $\theta_{c2} = \widehat{\theta}_{m2}(\rho - 1)^{-1}$, $\mu_c = \widehat{\mu}_m(\alpha - 1)^{-1}$, $\delta_c = \widehat{\delta}_m(\alpha - 1)^{-1}$, while θ_{c1}, σ_0, and σ_1 solve a nonlinear system of equations:

$$
\begin{aligned}
\widehat{\theta}_{m1} &= \theta_1^{rec} + \lambda_0^{rec}\lambda_1^{rec}, \\
\widehat{\lambda}_0 &= \lambda_0^{rec}, \\
\widehat{\lambda}_1 &= \lambda_1^{rec}.
\end{aligned}
$$

We select $\alpha = -9$ and $\rho = 1/3$, which are ubiquitous in the literature.

F.3 Understanding implications for the real yield curve

In this appendix, we follow the approach of Backus, Chernov, and Zin (2014) and focus on the serial covariance of the real pricing kernel, which governs the slope of the real yield curve. Note that we can represent the consumption growth process in (37) as:

$$\log g_{t,t+1}^c = g^c + \gamma(B)w_{t+1} + \bar{\sigma}_t w_{t+1},$$

where $\bar{\sigma}_t = \sigma_1 x_{1t}$ is the demeaned conditional volatility of consumption growth. We omit the iid jumps here because they have no implications for the shape of the yield curve.

The term $\gamma(B) = \sum_{j=0}^{\infty} B^j \gamma_j$ is a lag polynomial that captures the serial dependence of consumption growth. In the specific case of our model, it arises from the ARMA(2,1) structure of the expected consumption growth process; the first three elements of the lag polynomial in our model are $\gamma_0 = \sigma_0$, $\gamma_1 = \theta_{c1} + \theta_{c2}$, $\gamma_2 = (\theta_{c1} + \theta_{c2})(\phi_1 + \theta_1)$, with $\phi_1 = \varphi_1 + \varphi_2$ and $\theta_1 = -(\theta_{c1}\varphi_2 + \theta_{c2}\varphi_1)(\theta_{c1} + \theta_{c2})^{-1}$ as in appendix B. The lag polynomial notation $\gamma(B)$ is more general than our particular model. We use this notation to emphasize that the specifics of expected consumption growth are not important for our discussion here.

Finally, notice that, in our model, the time-varying component of volatility is additive. Instead, in the standard ARG(1) and AR(1) specifications of variance, volatility is multiplicative, and the consumption growth process is represented as:

$$\log g_{t,t+1}^c = g^c + \gamma(B)\sigma_t w_{t+1},$$

where σ_t is the volatility.

Following the same notation, the real pricing kernel (39) can be re-written as:

$$
\begin{aligned}
\log \widehat{m}_{t+1} &= \text{constant} + [(\rho - 1)\gamma(B) + (\alpha - \rho)\gamma(u_1)]w_{t+1} \\
&\quad - (\alpha - \rho)\alpha\gamma(u_1)(\sigma_1 + 2a_2\varphi_1)(1 - 2\alpha a_2)^{-1}\nu(B)Bw_{t+1} \\
&\quad + [(\alpha - 1)\sigma_1 + (\alpha - \rho)2a_2\varphi_1]\sigma_t w_{t+1} + \dots,
\end{aligned}
$$

where $\gamma(u_1) = \sum_{j=0}^{\infty} u_1^j \gamma_j$, lower dots represent the omitted quadratic terms, and $\nu(B)$ is a lag polynomial that arises from the AR(1) structure of x_{1t} and whose first three elements equal to $\nu_0 = 1$, $\nu_1 = \varphi_1$, $\nu_2 = \varphi_1^2$. The last two lines of this expression represent components of the pricing kernel associated with time-varying volatility of consumption growth: if $\sigma_1 = 0$, then $a_2 = 0$, and these terms vanish.

The first line in this expression represents the pricing kernel corresponding to Bansal and Yaron (2004), Model I – time-varying expected consumption growth with constant volatility. As pointed out by Backus, Chernov, and Zin (2014), extensions to Model I using ARG(1) or AR(1) processes to capture the volatility dynamics do not help in changing the sign of the term spread. In contrast, the linear specification of volatility, reflected in the second line of the recursive pricing kernel expression, offers additional flexibility in generating

serial covariance of the linear part of the pricing kernel. For instance, the leading term $(\rho - 1)\gamma_0 + (\alpha - \rho)\gamma(u_1)$ is unaltered, but the second one changes from $(\rho - 1)\gamma_1$ to $(\rho - 1)\gamma_1 - (\alpha - \rho)\alpha\gamma(u_1)(\sigma_1 + 2a_2\varphi_1)(1 - 2\alpha a_2)^{-1}\nu_0$. Using the calibrated parameters in Table 3D, we see that indeed the extra term changes the sign of the serial covariance of the pricing kernel as compared to the traditional specifications of variance. Thus, serial covariance of the pricing kernel becomes negative and the slope of the real curve becomes positive.

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.

Barro, Robert, 2006, "Rare disasters and asset markets in the twentieth century," *Quarterly Journal of Economics* 121, 823-867.

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, 2017, "The term structure of returns: facts and theory," *Journal of Financial Economics* 124, 1-21.

Binsbergen, Jules van, Wouter Hueskes, Ralph Koijen, and Evert Vrugt, 2013, "Equity yields," *Journal of Financial Economics* 110, 503-519.

Borovicka, Jaroslav, Lars Peter Hansen, and Jose Scheinkman, 2016, "Misspecified recovery," *Journal of Finance* 71, 2493-2544.

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 Financial Studies* 5, 243-280.

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

Creal, Drew, and Jing Cynthia Wu, 2016, "Bond Risk Premia in Consumption-Based Models," manuscript, April.

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.

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

Epstein, Larry G., and Stanley E. Zin, 1989, "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework," *Econometrica* 57, 937-969.

Garlappi, Lorenzo, Georgios Skoulakis, and Jinming Xue, 2015, " Do industry portfolios predict the aggregate market?," manuscript.

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 U.S. 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.

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.

Lettau, Martin, and Jessica Wachter, 2011, "The term structures of equity and interest

rates," *Journal of Financial Economics* 101, 90-113.

Longstaff, Francis A., and Monika Piazzesi, 2004, "Corporate earnings and the equity premium," *Journal of Financial Economics* 74, 401-421.

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.

Merton, Robert, 1976, "Option pricing when underlying stock returns are discontinuous," *Journal of Financial Economics* 3, 125-144.

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).

Rietz, Thomas A., 1988, "The equity risk premium: A solution," *Journal of Monetary Economics* 22, 117-131.

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

Santos, Tano, and Pietro Veronesi, 2010, "Habit formation, the cross section of stock returns and the cash-flow risk puzzle," *Journal of Financial Economics* 98, 385-413.

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," *Journal of Finance*, forthcoming

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: U.S. TIPS, 1971-2014 (source: Gurkaynak, Sack, and Wright, 2010; Chernov and Mueller, 2012); U.S. 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 C.

Asset	Mean	Standard Deviation	Skewness	Excess Kurtosis	Entropy, $L(rx)$
Inflation-protected bonds (TIPS)					
CPI	0.0022	0.0078	0.1785	0.8223	0.00002
Currencies					
AUD	0.0108	0.0677	−0.5134	0.7206	0.0016
EUR (Germany)	−0.0015	0.0614	0.2748	0.6517	0.0018
GBP	−0.0008	0.0607	−0.0816	1.4681	0.0015
Equity					
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 U.S. nominal curve. A term spread is defined as the difference between an $n-$quarter yield and a one-quarter yield. Sample periods: U.S. nominal bonds, 1971-2014 (source: Gurkaynak, Sack, and Wright, 2007; FRED); U.S. 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 C.

Asset or Country	Maturity, quarters								
	1	2	4	8	12	20	24	28	40
U.S.	0.0124	0.0128	0.0138	0.0144	0.0149	0.0157	0.0160	0.0163	0.0169
U.S. TIPS	0.0042	0.0043	0.0043	0.0045	0.0047	0.0052		0.0056	0.0063
		−0.0003	−0.0013	−0.0017	−0.0020	−0.0023		−0.0025	−0.0024
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. With the exception of CPI, θ_{g1} in Panel B is a derived parameter. The first column of Panel C displays expected log excess returns implied by the model in Example 2. The subsequent two columns display the loadings of cash flows and the pricing kernel, respectively, on a normal shock in the extension of this model in section 5.2. The last two columns highlight differences in covariance and entropy implied by the final model featuring jumps.

Panel A. Common parameters
(U.S. nominal economy)

φ_1	θ_{m1}	λ_0	λ_1	ω	μ_m	δ_m
0.9487	0.0026	-0.1225	0.0512	0.0025	1.5000	1.5000

Panel B. Asset-specific parameters

Asset	φ_2	θ_{g1}	θ_{g2}	η_0	μ_g	δ_g
Inflation-protected bonds (TIPS)						
CPI	0.2240	-0.0016	0.0016	0.0077	-0.0438	0.0200
Currencies						
AUD	0.9404	-0.0058	0.0029	0.0642	-0.1728	0.0173
EUR	0.8356	-0.0039	0.0045	0.0254	-0.4365	1.0268
GBP	0.9664	-0.0056	0.0027	0.0587	0.1765	0.0589
Equity						
S&P 500	0.6846	-0.0014	0.0292	-0.0225	-0.0799	0.3736

Panel C. Some derived quantities

Asset	$E \log rx_{t,t+1}$	$\|\eta_2\|$	$\|\lambda_2\|$	cov_{mg}	$\mathcal{C}_{mg}(1)$
Inflation-protected bonds (TIPS)					
CPI	0.0013	0.0004	3.3067	-0.0001	-0.0023
Currencies					
AUD	0.0058	0.0188	0.2749	-0.0079	-0.0129
EUR	0.0028	0.0557	0.0491	0.0007	-0.0001
GBP	0.0055	0.0123	0.5014	-0.0069	-0.0009
Equity					
S&P 500	-0.0030	0.0145	0.8834	0.0042	0.0024

Panel D. Parameters from the representative agent model

Consumption						*Preferences*	
θ_{c1}	θ_{c2}	σ_0	σ_1	μ_c	δ_c	α	ρ
0.0009	-0.0025	-0.0085	-0.0045	-0.1456	0.1500	-9	$1/3$

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 highlighted in Garlappi, Skoulakis, and Xue (2015). 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.

	Data		Model	
Asset	Var$\times 10^2$	AR(1)	Var$\times 10^2$	AR(1)
Inflation-protected bonds (TIPS)				
CPI	0.0078	0.5889	0.0084	0.2202
	(0.0008)	(0.0625)	(0.0062, 0.0123)	(0.0520, 0.3848)
Currencies				
AUD	0.3132	0.0598	0.44309	−0.0191
	(0.0422)	(0.0950)	(0.3224, 0.5577)	(−0.2018, 0.1625)
EUR	0.3748	0.0015	0.3763	0.0655
	(0.0410)	(0.0788)	(0.0562, 2.9027)	(−0.0726, 0.2490)
GBP	0.3126	0.1271	0.3604	−0.0278
	(0.0370)	(0.0828)	(0.2784, 0.4586)	(−0.1886, 0.1328)
Equity				
S&P 500	0.3829	0.2599	0.2947	0.1097
	(0.0459)	(0.0812)	(0.1981, 0.4033)	(−0.0819, 0.2961)
Macro				
Cons. growth	0.0004	0.0877	0.0004	0.0487
	−	(0.0600)	−	(−0.0941, 0.2081)

Figure 1. Average U.S. curve and excess returns. The black solid line shows average U.S. 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 U.S. Treasuries, $E(\widehat{y}_t^n - \widehat{y}_t^1) - E(y_t^n - y_t^1)$. Data sources are the same as in Table 2.
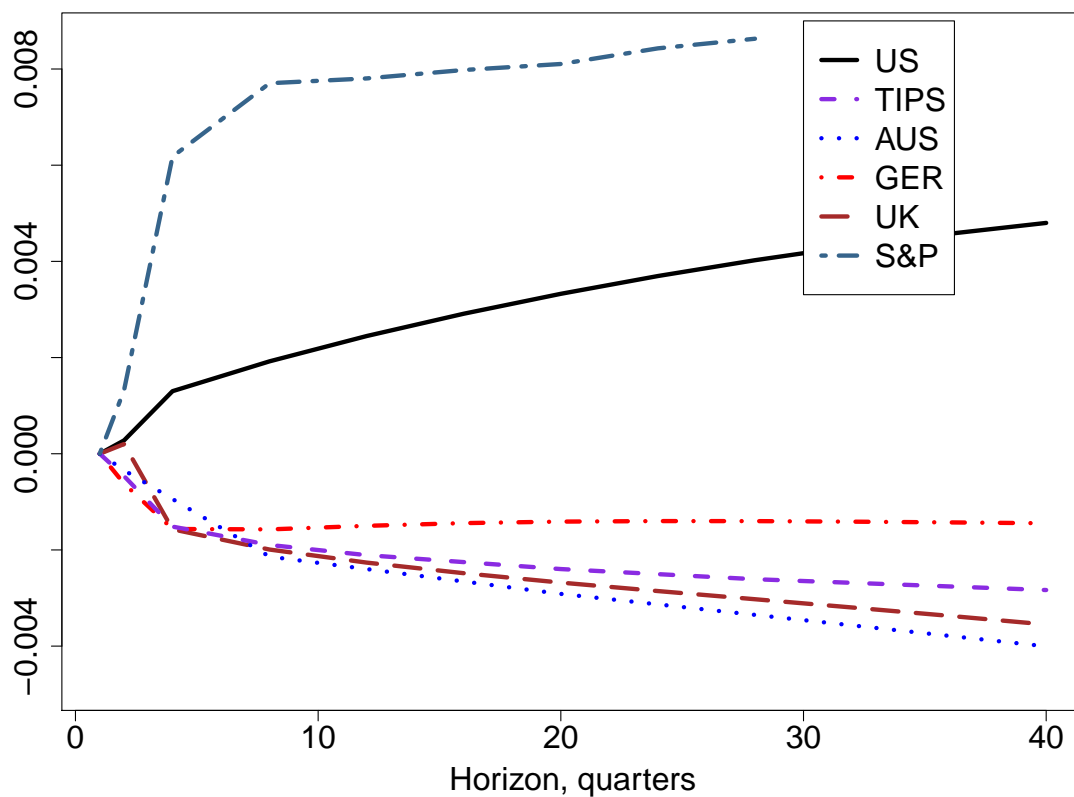
Figure 2. HJ bound and entopy in the two-horizon model. The figure compares how the HJ bound (purple lines) and entropy, $n\mathcal{L}_m(n)$, (blue lines) change with horizon in the benchmark iid case (solid lines) and in the two-horizon model (dashed lines).
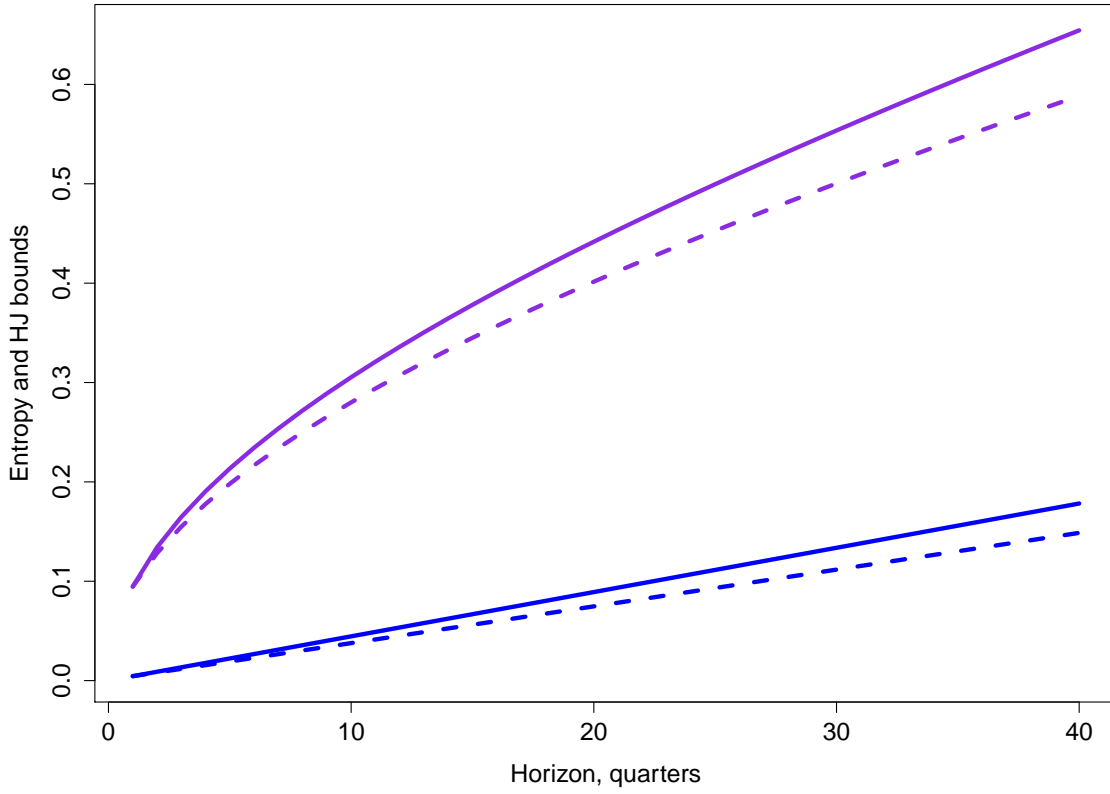
Figure 3. Coentropy and covariance. The figure compares coentropy and covariance for the Poisson mixture of bivariate normals described at the end of section 5.3. As we vary ω, we adjust δ to hold the variance constant.
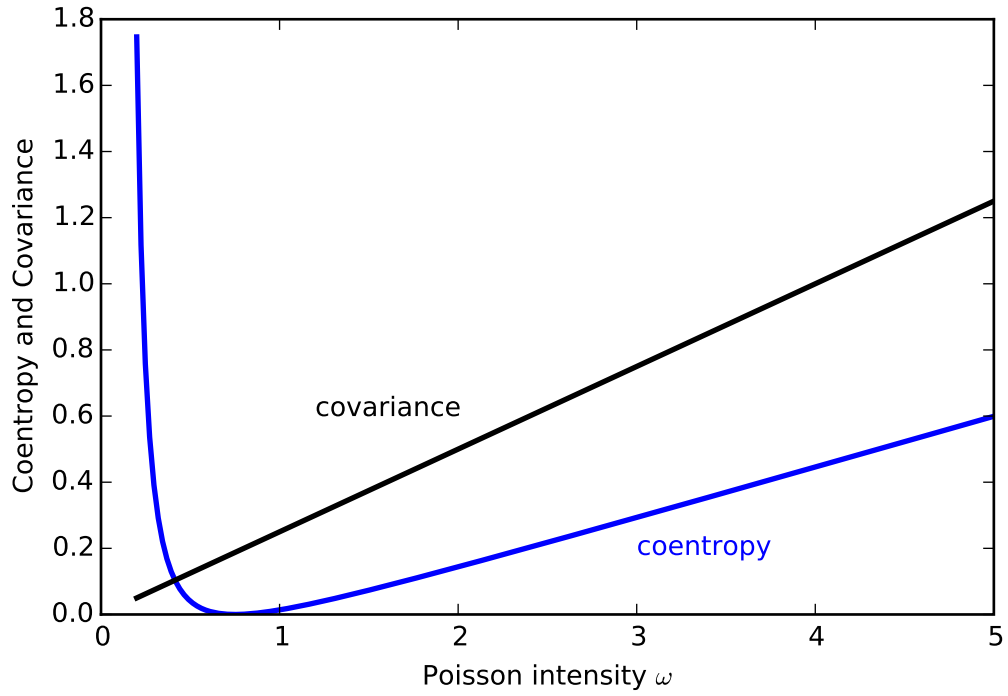
Figure 4. Ingridients of coentropy, a case of GBP. The figure displays marginal cumulants of the jump component of the pricing kernel and the GBP depreciation rate in panels (A) and (B), respectively. Panels (C) and (D) show how they combine to contribute to coentropy. In contrast to plots in other panels that are depicted on the log scale, the plot in panel (D) is in levels. Further, in contrast to panel (C), it accounts for factorials that strongly discount contribution of high order joint cumulants to coentropy.
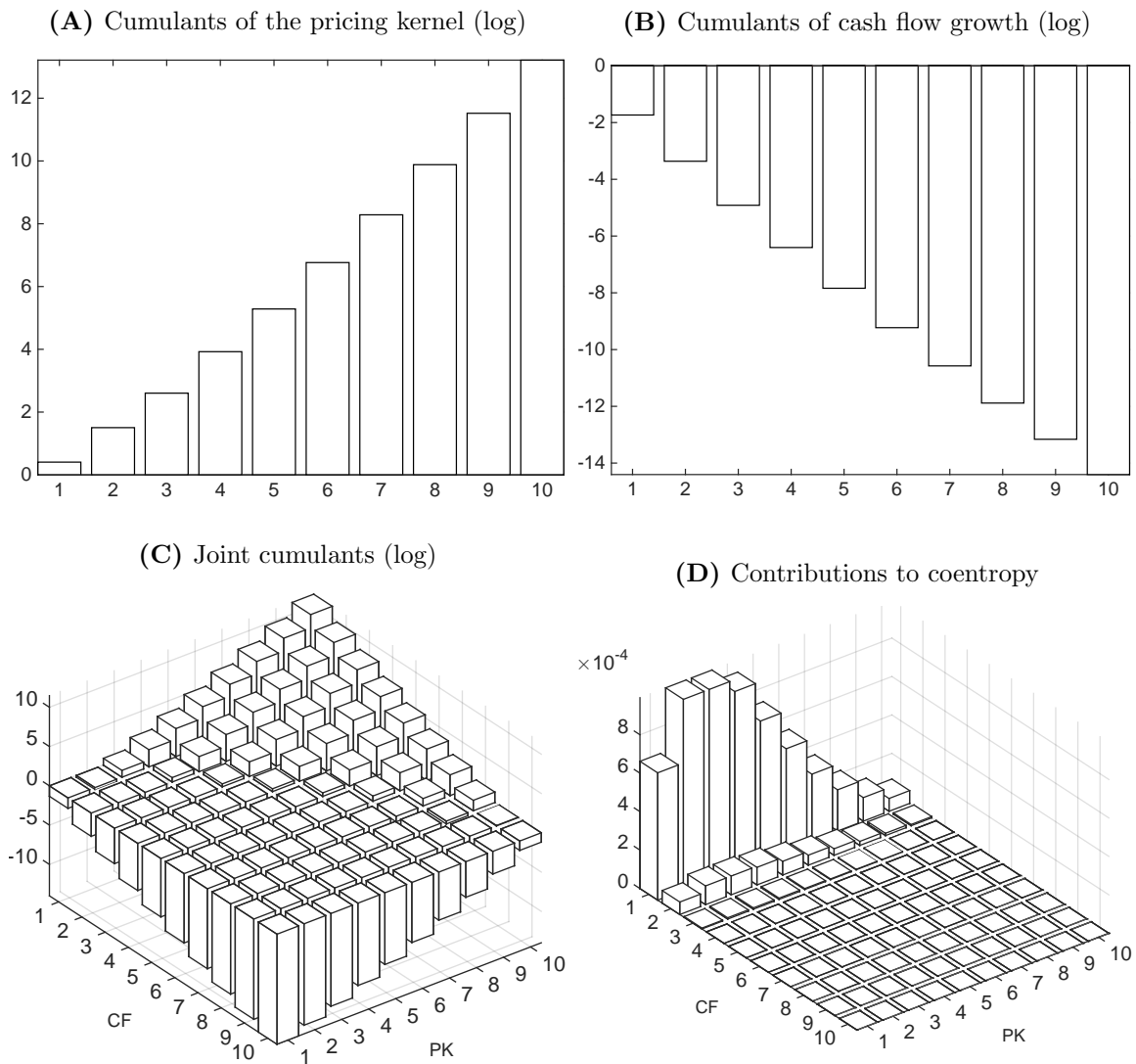

(A) Cumulants of the pricing kernel (log)

(B) Cumulants of cash flow growth (log)

(C) Joint cumulants (log)

(D) Contributions to coentropy

Figure 5. Coentropy. The figure compares contribution of joint cumulants to coentropies of different assets.

(A) AUD

(B) EUR

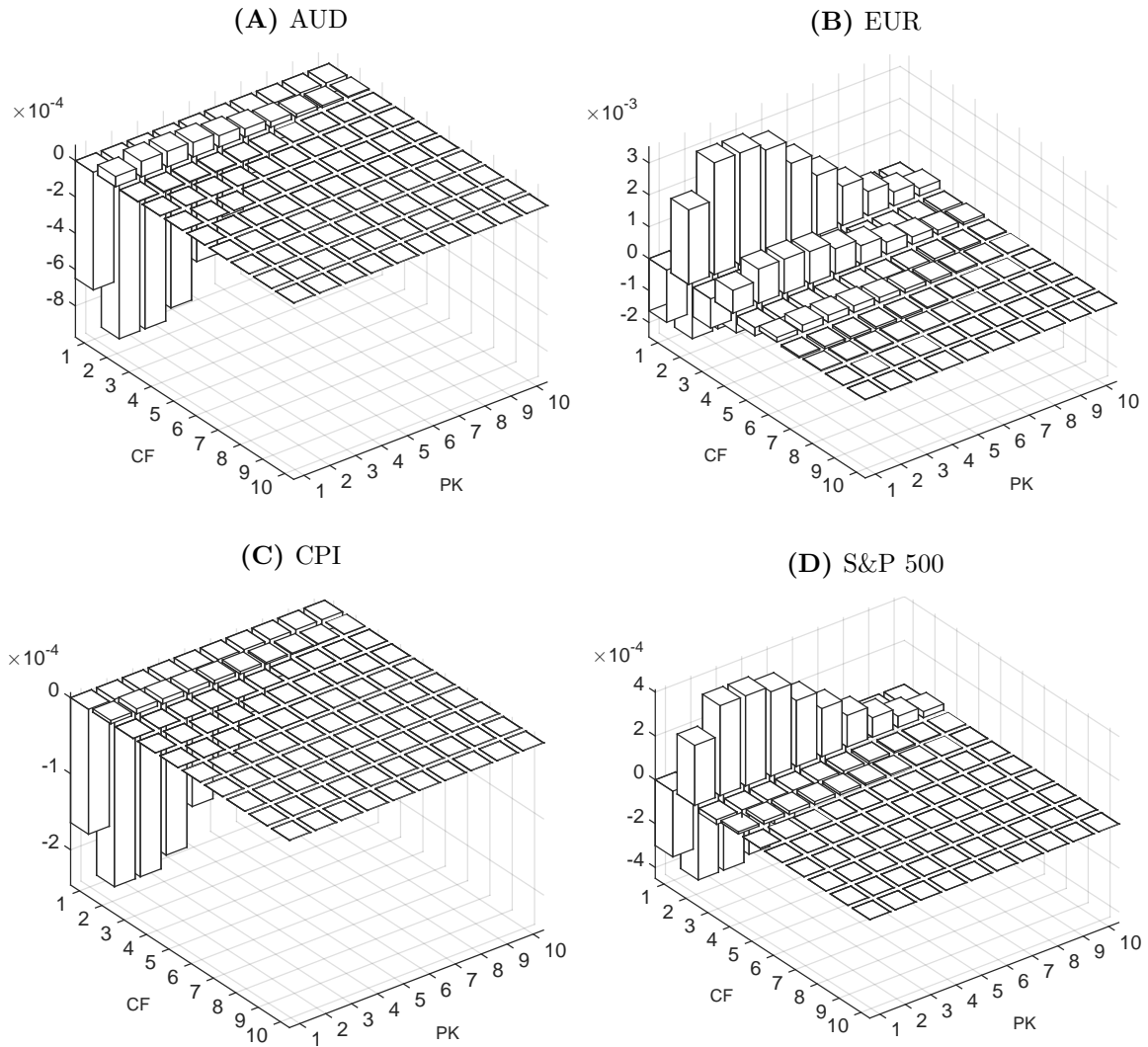**(C)** CPI

(D) S&P 500

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 Koijen (2012).