The Long Bond

Ian Martin Steve Ross*

June, 2013

Abstract

We study the behavior of the *long bond*, a zero-coupon bond that pays off in the far-distant future, under the assumptions that (i) the fixed-income market is complete and (ii) the state vector that drives interest rates follows a Markov chain. We show that if the pricing kernel is transition-independent—in particular, if there is an investor with separable utility or Epstein–Zin preferences—then the yield curve must slope up on average, and derive a formula that expresses the expected return on the long bond in terms of the prices of long bond options.

*Martin: Stanford GSB. Ross: MIT Sloan. We are grateful to Dave Backus, Ravi Bansal, John Campbell, Mike Chernov, Darrell Duffie, and to participants at the TIGER Forum for their comments.

In this paper, we present some theoretical results on the properties of the long end of the yield curve. Our results relate to two literatures. The first is the Recovery Theorem of Ross (2013). Ross showed that, given a matrix of Arrow–Debreu prices, it is possible to infer the objective state transition probabilities and implied marginal utilities. Although it is a familiar fact that sufficiently rich asset price data pins down the risk-neutral probabilities, it is initially surprising that we can do the same for the objective, or real-world, probabilities.

It is, however, challenging in practice to construct the matrix of Arrow–Debreu prices. We sidestep the need to do so by showing that recovery can be effected by observing the behavior of the long end of the yield curve. We show that the yield on the (infinitely) long zero-coupon bond reveals the time preference rate of what we call a pseudo-representative agent. (Such an agent exists and is uniquely determined. If there is a representative agent in the conventional sense, the representative and pseudo-representative agents coincide.) We also show that the time-series of returns on the long bond reveals the pseudo-representative agent's "marginal utilities." These "marginal utilities" are a convenient guide to intuition, but our analysis is founded in the logic of no-arbitrage: we do not need to assume that the concept of expected utility is meaningful.

These results connect the Recovery Theorem to earlier results of Backus, Gregory and Zin (1989), Kazemi (1992), Alvarez and Jermann (2005) and Hansen and Scheinkman (2009), though our focus differs in that we focus exclusively on bonds and interest-rate derivatives rather than on the prices of claims to growing cashflows, for reasons discussed further below. A side benefit is that we can avoid some of the technical complication of this literature, and provide simple and short proofs of our results.

We present two applications. First, we show that the yield curve must slope upwards on average. This is an empirical regularity. Second, we provide a formula for the expected excess return on the long bond in terms of the prices of options on the long bond. Consistent with the theme of the paper, the matrix of Arrow–Debreu prices is in the background of both results, but we avoid the need to observe it directly.

1 The pseudo-representative agent and the yield curve

We work in discrete time. Our focus is on fixed-income assets, including fixed- or floating-rate bonds, interest-rate swaps, and derivatives on them, including caps, floors, bond options and swaptions. We assume that fixed-income markets are (i) complete, and (ii) governed by a (potentially multidimensional) state variable that follows a Markov chain on $\{1, 2, \ldots, m\}$ with transition probabilities $\pi(i, j)$. The associated stationary distribution $(\pi(1), \ldots, \pi(m))$ satisfies $\pi(j) = \sum_i \pi(i)\pi(i, j)$. The ith entry, $\pi(i)$, can be interpreted as the long-run proportion of time spent in state i.

Although we have in mind that the state vector might in principle comprise, say, the current short rate, measures of the yield curve slope and curvature, the VIX index, state of the business cycle—even, perhaps, some measure of "animal spirits"—we will not need, in this paper, to be specific about its constituents. But since we assume that the state variable follows a Markov chain, it *is* important that all elements of the vector can reasonably be thought of as stationary.

In a complete market, it is convenient to summarize all available asset-price data in a matrix, $\boldsymbol{A}$, of Arrow–Debreu prices

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}(1,1) & \boldsymbol{A}(1,2) & \cdots & \boldsymbol{A}(1,m) \\ \boldsymbol{A}(2,1) & \boldsymbol{A}(2,2) & & \vdots \\ \vdots & & \ddots & \vdots \\ \boldsymbol{A}(m,1) & \cdots & \cdots & \boldsymbol{A}(m,m) \end{pmatrix},$$

where we write $\boldsymbol{A}(i, j)$ for the price, in state i, of the Arrow–Debreu security that pays off \$1 if state j materializes next period. The absence of arbitrage requires that the entries of $\boldsymbol{A}$ are nonnegative, and that $\boldsymbol{A}(i, j)$ is strictly positive if the transition from state i to state j occurs with positive probability. The row sums of $\boldsymbol{A}$ are the prices of one-period riskless bonds in each state. We assume that not all of these row sums are equal, except where explicitly stated otherwise. This assumption is inessential, but it lets us avoid constantly having to qualify statements to account for the somewhat degenerate case in which interest rates are constant and the yield curve is flat.

We assume that for a sufficiently large time horizon, T, any Arrow–Debreu security that pays off in T periods has a strictly positive price in every state; this assumption ensures that $\boldsymbol{A}$ is a *primitive* matrix. This framework is extremely flexible: it allows the yield curve to take any shape consistent with no arbitrage. For now, we think of the matrix $\boldsymbol{A}$ as directly observable; we will address the issue that it may be hard to observe in practice below.

If there were a utility-maximizing investor with time discount factor ϕ, marginal utility $u'(i)$ in state i, and subjective transition probabilities $\pi(i,j)$, then the Arrow–Debreu prices would satisfy

$$\boldsymbol{A}(i,j) = \phi\pi(i,j)\frac{u'(j)}{u'(i)}.$$

In terms of inverse marginal utility, $v(i) \equiv 1/u'(i)$, we would have $\boldsymbol{A}(i,j) = \phi v(i)\pi(i,j)/v(j)$ for each i and j or, written more concisely as a matrix equation,

$$\boldsymbol{A} = \phi \boldsymbol{D}\boldsymbol{\Pi}\boldsymbol{D}^{-1},$$

where $\boldsymbol{D}$ is a diagonal matrix with positive diagonal elements $\{v(i)\}$, and $\boldsymbol{\Pi} = \{\pi(i,j)\}$ is a *stochastic* matrix, i.e. a matrix whose row sums all equal 1 (because they sum over the probabilities of moving from the current state to any other state).

The next result is a variant of the Recovery Theorem of Ross (2013) that emphasizes that the existence of a decomposition of asset price data, $\boldsymbol{A}$, into a diagonal matrix $\boldsymbol{D}$ with positive diagonal entries and a stochastic matrix $\boldsymbol{\Pi}$ is guaranteed whether or not there exists such a utility-maximizing investor. We think of this result as establishing the existence of a *pseudo-representative agent*.

Result 1. *For arbitrary asset price data $\boldsymbol{A}$, there exists a unique decomposition*

$$\boldsymbol{A} = \phi \boldsymbol{D}\boldsymbol{\Pi}\boldsymbol{D}^{-1}$$

where $\boldsymbol{D}$ is a diagonal matrix and $\boldsymbol{\Pi}$ is a transition matrix. (More precisely, the matrix $\boldsymbol{D}$ in the above decomposition is determined only up to positive scalar multiples, but this corresponds to economically irrelevant changes of utility units.)

This can be interpreted as establishing existence and uniqueness of a pseudo-representative agent (ϕ and $\boldsymbol{D}$) and probability measure ($\boldsymbol{\Pi}$) that rationalize the asset prices ($\boldsymbol{A}$).

Proof. The Perron–Frobenius theorem guarantees that $\boldsymbol{A}$ has an eigenvector $\boldsymbol{v}$ and associated eigenvalue ϕ—so that $\boldsymbol{Av} = \phi\boldsymbol{v}$—with the properties that (i) the entries of $\boldsymbol{v}$ are positive; and (ii) ϕ is real and positive (and the largest, in absolute value, of all the eigenvalues of $\boldsymbol{A}$). Let $\boldsymbol{D}$ be the diagonal matrix with $\boldsymbol{v}$ along its diagonal, so that $\boldsymbol{v} = \boldsymbol{De}$, where $\boldsymbol{e}$ is the vector of ones. Then, $\frac{1}{\phi}\boldsymbol{D}^{-1}\boldsymbol{ADe} = \frac{1}{\phi}\boldsymbol{D}^{-1}\boldsymbol{Av} = \boldsymbol{D}^{-1}\boldsymbol{v} = \boldsymbol{e}$. This shows that $\frac{1}{\phi}\boldsymbol{D}^{-1}\boldsymbol{AD}$ is a stochastic matrix, which we will call $\boldsymbol{\Pi}$, so that $\boldsymbol{A} = \phi\boldsymbol{D\Pi D}^{-1}$.

Now suppose that we have some potentially different representation $\boldsymbol{A} = \widetilde{\phi}\widetilde{\boldsymbol{D}}\widetilde{\boldsymbol{\Pi}}\widetilde{\boldsymbol{D}}^{-1}$, where $\widetilde{\boldsymbol{D}}$ is a diagonal matrix with positive entries along its diagonal and $\widetilde{\boldsymbol{\Pi}}$ is a stochastic matrix. The goal is to show that $\widetilde{\phi} = \phi$ and that $\widetilde{\boldsymbol{D}}$ is proportional to $\boldsymbol{D}$; having done so, it will immediately follow that $\widetilde{\boldsymbol{\Pi}} = \boldsymbol{\Pi}$, and hence that the representation $\boldsymbol{A} = \phi\boldsymbol{D\Pi D}^{-1}$ is unique (up to economically irrelevant scalings of $\boldsymbol{D}$). Since $\widetilde{\phi}\widetilde{\boldsymbol{D}}\widetilde{\boldsymbol{\Pi}}\widetilde{\boldsymbol{D}}^{-1}\boldsymbol{v} = \boldsymbol{Av} = \phi\boldsymbol{v}$, we have $\widetilde{\boldsymbol{\Pi}}\widetilde{\boldsymbol{D}}^{-1}\boldsymbol{v} = (\phi/\widetilde{\phi})\widetilde{\boldsymbol{D}}^{-1}\boldsymbol{v}$, so $\widetilde{\boldsymbol{D}}^{-1}\boldsymbol{v}$ is a positive eigenvector of $\widetilde{\boldsymbol{\Pi}}$ with eigenvalue $\phi/\widetilde{\phi}$. But $\widetilde{\boldsymbol{\Pi}}$ is a stochastic matrix, so $\phi/\widetilde{\phi} = 1$; and the Perron–Frobenius theorem implies that $\boldsymbol{e}$ is the *unique* positive eigenvector of $\widetilde{\boldsymbol{\Pi}}$, up to positive scalar multiples. So $\widetilde{\boldsymbol{D}}^{-1}\boldsymbol{v} = \boldsymbol{e}$ and hence $\widetilde{\boldsymbol{D}} = \boldsymbol{D}$, up to positive scalar multiples. □

If the observed Arrow–Debreu prices were generated by a utility-maximizing investor as described above, then this analysis recovers $\boldsymbol{\Pi}$ and $v(i) = 1/u'(i)$, the investor's subjective probability distribution and marginal rates of substitution. But the above decomposition is well-defined whether or not such an investor exists. Quite generally, this procedure identifies the unique pricing kernel that is *transition-independent*, in the terminology of Ross (2013). From now on we will assume that the recovered $\boldsymbol{\Pi}$ is the true probability distribution. However, if prices are not actually generated by a transition-independent kernel, then $\boldsymbol{\Pi}$ generally will not be the true probability distribution. This is not an issue to be determined on *a priori* theoretical grounds. It is an empirical question that can only be resolved by comparing the derived $\boldsymbol{\Pi}$ with observed realizations in the market.

The above result identifies the unique positive eigenvector, $\boldsymbol{v}$, of $\boldsymbol{A}$ with the (inverse) marginal utilities of the pseudo-representative agent; and the associated time discount factor with the eigenvalue, ϕ. On the face of it, though, we are still confronted with the difficult task of having to compute the matrix $\boldsymbol{A}$ in order to find its eigenvalues and eigenvectors. It turns out that we can avoid this difficulty by connecting $\boldsymbol{v}$ and ϕ to the long end of the yield curve (which we *will* take to be directly observable).

We will use the notation $B_t(i)$ for the price of a t-period zero-coupon bond if the current (time zero) state is i, and define the t-period continuously-compounded yield $y_t(i)$ and simple yield $Y_t(i)$ by

$$e^{-y_t(i)t} = \frac{1}{[1+Y_t(i)]^t} = B_t(i),$$

so that $y_t(i) = \log(1+Y_t(i))$. We have

$$B_t(i) = \sum_j \boldsymbol{A}^t(i,j)\,, \tag{1}$$

where $\boldsymbol{A}^t$ is the matrix $\underbrace{A\cdots A}_{t\text{ times}}$ and $\boldsymbol{A}^t(i,j)$ is the entry that appears in its ith row and jth column. In other words, to calculate the price in state i of a t-period zero-coupon bond, we sum the elements of the ith row of $\boldsymbol{A}^t$.

The t-period zero-coupon yield is therefore $y_t(i) = -\frac{1}{t}\log\sum_j \boldsymbol{A}^t(i,j)$. We will be most interested in the two ends of the yield curve. The *short rate* in state i is

$$y_1(i) = r_f(i) = -\log\sum_j \boldsymbol{A}(i,j),$$

and the *long rate* in state i is

$$y_\infty(i) = -\lim_{t\to\infty}\frac{1}{t}\log\sum_j \boldsymbol{A}^t(i,j)\,.$$

The realized return on a t-period bond following a transition from state i to state j is

$$R_t(i,j) \equiv \frac{B_{t-1}(j)}{B_t(i)},$$

and the log return is $r_t(i,j) \equiv \log R_t(i,j)$. The corresponding long bond returns are defined in the obvious way: $R_\infty(i,j) \equiv \lim_{t\to\infty} R_t(i,j)$ and $r_\infty(i,j) \equiv \lim_{t\to\infty} r_t(i,j)$.

The return on a t-period bond depends on its yield, on its maturity, and on its realized yield change,

$$r_t(i,j) = \underbrace{y_t(i)}_{\text{yield}} - \underbrace{(t-1)}_{\text{duration}} \times \underbrace{(y_{t-1}(j) - y_t(i))}_{\text{realized yield change}}.$$

For long-dated bonds (large t), we will see that the realized yield change is very small (indeed constant in the large-t limit). Pulling in the opposite direction, the duration $t-1$ is large for long bonds. When the two effects are combined, long-dated bonds will have volatile returns.

We can calculate expected bond returns either conditionally or unconditionally, and we can work with either log returns or simple returns. For example, the conditionally expected simple return on the long bond in state i is

$$\overline{R}_\infty(i) \equiv \sum_j \pi(i,j) R_\infty(i,j),$$

and the (unconditional) expected log return on the long bond is

$$\overline{r}_\infty \equiv \sum_{i,j} \pi(i)\pi(i,j) r_\infty(i,j).$$

Result 1 provided one economic interpretation of the vector $\boldsymbol{v}$. But $\boldsymbol{v}$ can also be viewed as representing a one-period asset that pays $v(i) > 0$ in state i. Our next result shows that this $\boldsymbol{v}$-*asset* has another interesting economic interpretation.

Result 2. *Returns on the $\boldsymbol{v}$-asset replicate the returns on the long bond.*

Proof. The return on the long bond, if we move from state i to state j, is

$$R_\infty(i,j) = \lim_{t\to\infty} \frac{\sum_k \boldsymbol{A}^{t-1}(j,k)}{\sum_k \boldsymbol{A}^t(i,k)} = \frac{1}{\phi} \cdot \lim_{t\to\infty} \frac{\sum_k \frac{\boldsymbol{A}^{t-1}(j,k)}{\phi^{t-1}}}{\sum_k \frac{\boldsymbol{A}^t(i,k)}{\phi^t}} = \frac{1}{\phi} \cdot \frac{v(j)}{v(i)}. \tag{2}$$

The first equality follows from (1). The last equality follows from the well-known result that $(\boldsymbol{A}/\phi)^t$ tends to a matrix whose row sums are proportional to the entries of $\boldsymbol{v}$ (see, for example, Theorem 8.5.1 of Horn and Johnson (1990)).

The corresponding return on the $\boldsymbol{v}$-asset is

$$\frac{\text{payoff in state } j}{\text{price in state } i} = \frac{v(j)}{\sum_k A(i,k)v(k)} = \frac{v(j)}{\phi v(i)}. \qquad \square$$

This result means that the kernel can be observed over time empirically from the returns on the long bond.[1] Having linked the eigenvector $\boldsymbol{v}$ to the long end of the yield curve, we can now do the same for the eigenvalue ϕ. The next result[2] verifies that the long rate is independent of the current state i—no surprise, given Dybvig, Ingersoll and Ross (1986)—and equals the unconditional expected log return on the long bond; and shows that both are determined by the eigenvalue, ϕ.

Result 3. *The long rate and unconditional expected log return on the long bond are determined by ϕ, the largest eigenvalue of $\boldsymbol{A}$: $y_\infty = \overline{r}_\infty = -\log\phi$.*

Proof. To show that $y_\infty = -\log\phi$, we must show that $\lim_{t\to\infty}\frac{1}{t}\log\sum_j \boldsymbol{A}^t(i,j) = \log\phi$ for arbitrary i, or equivalently that $\lim_{t\to\infty}\frac{1}{t}\log\sum_j(\boldsymbol{A}/\phi)^t(i,j) = 0$. But this holds because (as noted in the proof above) $(\boldsymbol{A}/\phi)^t$ tends to a fixed matrix.

For the unconditional expected log return, note from (2) that $r_\infty(i,j) = -\log\phi + \log[v(j)/v(i)]$. The expected return on the long bond is therefore

$$\overline{r}_\infty = -\log\phi + \sum_{i,j}\pi(i)\pi(i,j)\log\frac{v(j)}{v(i)}\,.$$

The result follows, since

$$\sum_{i,j}\pi(i)\pi(i,j)\log\frac{v(j)}{v(i)} = \sum_j \log v(j)\underbrace{\sum_i \pi(i)\pi(i,j)}_{\pi(j)} - \sum_i \pi(i)\log v(i)\underbrace{\sum_j \pi(i,j)}_{1} = 0. \quad \square$$

To summarize, returns on the long bond reveal the eigenvector $\boldsymbol{v}$, and hence also marginal utilities; and the long yield reveals the eigenvalue ϕ, and hence the time discount factor of the pseudo-representative agent.

While the long bond is a theoretically appealing construct, the question remains of how closely it can be approximated by bonds of long but finite maturity. To address this

[1] Kazemi (1992) made this observation in a continuous-time diffusion model. The eigenvector $\boldsymbol{v}$ can be given other, related, interpretations: Appendix A.1 links it to the return on *any* sufficiently long-dated asset, and Appendix A.2 links it to the unique infinitely-lived asset that has a constant dividend yield.

[2] Backus, Gregory and Zin (1989) essentially have the first part of this result: they solve for the long rate in a power utility model, but their approach—which, as far as we are aware, represents the earliest use of the Perron–Frobenius theorem in asset pricing—applies more generally.

issue, we introduce the quantity

$$Q \equiv \log \frac{\max_k v(k)}{\min_k v(k)}.$$

We can think of Q as an index of the extent to which risk considerations matter for pricing. It is nonnegative, and equals zero if and only if all the entries of $\boldsymbol{v}$ are equal. Put differently, Q is zero if and only if $\boldsymbol{e}$ is an eigenvector of $\boldsymbol{A}$; then, riskless rates are constant and the yield curve is flat, so pricing (of fixed-income securities) is effectively risk-neutral and the risk-neutral and objective probabilities coincide. It also measures the dispersion of possible returns on the long bond: from equation (2), the maximum and minimum possible log returns on the long bond are given by

$$\min_{i,j} r_\infty(i,j) = y_\infty - Q \qquad \text{and} \qquad \max_{i,j} r_\infty(i,j) = y_\infty + Q.$$

Our next result shows that Q controls the rate at which finite-maturity yields approach the long yield.

Result 4. *The T-period yield cannot diverge too far from the long yield: for arbitrary i,*

$$|y_T(i) - y_\infty| \leq \frac{Q}{T}.$$

It follows that for finite T_1 and T_2 (which may be equal) the T_1-period and T_2-period yields cannot diverge too far from each other in any two states i and j:

$$|y_{T_1}(i) - y_{T_2}(j)| \leq Q\left(\frac{1}{T_1} + \frac{1}{T_2}\right).$$

Proof. From Corollary 8.1.33 of Horn and Johnson (1990), we have $\phi^T e^{-Q} \leq B_T(i) \leq \phi^T e^Q$ for arbitrary i. It follows that

$$y_\infty - \frac{Q}{T} \leq y_T(i) \leq y_\infty + \frac{Q}{T},$$

from which both statements follow. □

Result 4 implies that if Q is small, then the long end of the yield curve must be fairly flat and yield volatility low. Conversely, if there is substantial variation in long-dated yields either across *maturities* or across *states of the world*—if the yield curve has significant slope at the long end, or if long-dated yields are volatile—then Q is large and risk considerations are important for fixed-income pricing.

Result 8 in Appendix A.1 presents a corresponding result for returns: the return on any sufficiently long-dated fixed-income asset must be very close to the return on the long bond. Interestingly, returns converge faster (at exponential rate) than yields (at rate $1/T$).

We now return to the question of making inferences about the objective probability distribution from observable data. As a crude example of the kind of thing we are interested in, note that Result 3 implies (by Jensen's inequality) that the expected return on the long bond is at least as great as its simple yield, $\overline{R}_\infty > 1+Y_\infty$. We will shortly be able to provide more refined information on the expected return on the long bond: in fact we will derive a formula that expresses its conditionally expected return in terms of option prices. The next result is a stepping stone towards this goal.

Result 5. *Within the class of fixed-income assets, the long bond is growth-optimal; that is, in every state the long bond has the highest expected log return of all fixed-income assets.*

Proof. The state-i price of an asset with payoffs $\boldsymbol{x}$ is

$$\begin{aligned}\sum_j \boldsymbol{A}(i,j)x(j) &\overset{(a)}{=} \phi \sum_j \pi(i,j)\frac{v(i)}{v(j)}x(j) \\ &\overset{(b)}{=} \sum_j \pi(i,j)\frac{x(j)}{R_\infty(i,j)} \qquad (3)\end{aligned}$$

Equality (a) follows from Result 1, and equality (b) follows from Result 2. It follows that the reciprocal of the return on the long bond is a stochastic discount factor.

But then the long bond's return must be growth-optimal, since (i) the reciprocal of the growth-optimal return is also a stochastic discount factor, and (ii) the growth-optimal return is the *unique* fixed-income return with this property. For, suppose that for two returns R_1 and R_2, $1/R_1$ and $1/R_2$ are both stochastic discount factors. It follows that $\mathbb{E}(R_1/R_2) = \mathbb{E}(R_2/R_1) = 1$. But then we must have $R_1 = R_2$; otherwise, we would have a violation of Jensen's inequality. □

A related result can be found in Alvarez and Jermann (2005), where it is used to argue that the stochastic discount factor that prices all assets cannot be stationary, since the long bond has a lower expected log return than, for example, certain equity portfolios. Setting aside the difficulties in constructing ex ante growth-optimal portfolios, we simply emphasize that the long bond is growth-optimal in a *relative* sense: it maximizes the expected log return among fixed-income assets. There may be other assets, not spanned by the Arrow–Debreu securities whose prices appear in $\boldsymbol{A}$, that have higher expected log returns.

Equation (3) shows that a fixed-income payoff can be priced either via state prices—equivalently, by discounting at the short riskless rate and using risk-neutral probabilities—as on the left-hand side, or by discounting at the long bond's return and using the true probabilities, as on the right-hand side. Loosely speaking, Result 5 can be thought of as saying that the cheapest way of generating a payoff in the far-distant future is to buy the growth-optimal portfolio, or equivalently the long bond.

2 Applications

Having set the scene, we now turn to two implications of our framework.

Result 6. *The yield curve cannot always slope down. Nor can it always slope up. But it must slope up on average.*

Proof. The first two claims follow because the eigenvalue ϕ must lie between the minimum and maximum row sums of $\boldsymbol{A}$; that is, $\min_k P_1(k) \le \phi \le \max_k P_1(k)$ (see, for example, Theorem 1.1 of Minc (1988, p. 24)), and hence $\min_i y_1(i) \le y_\infty \le \max_i y_1(i)$. Other than in the i.i.d. case—in which the rows of $\boldsymbol{A}$ are all the same and the yield curve is flat, and unchanging—both inequalities are strict.

By Result 5, $\overline{r}_\infty(i) \ge r_f(i)$ in every state, with strict inequality in some state, so $\overline{r}_\infty > \overline{r}_f$. Trivially, $\overline{r}_f = \overline{y}_1$. And by Result 3, $\overline{r}_\infty = y_\infty$. Together, these imply that $y_\infty > \overline{y}_1$, establishing the claim that the yield curve slopes up on average. $\square$

This result is borne out in the data; see, for example, Alvarez and Jermann (2005) and Backus, Chernov and Zin (2013), who report positive average nominal yield spreads, and Chernov and Mueller (2012), who report positive average real yield spreads.

In combination with our earlier results, Result 6 implies that the time discount factor $\phi < e^{-\overline{y}_1}$: average short rates provide an upper bound on the subjective time discount factor. This strengthens the finding in Ross (2013) that the *maximal* short rate provides an upper bound on the time discount factor.

What do market prices imply about expected future yields? We have already seen that for the long bond, this is the wrong question, since its *yield* is constant over time. But long-dated bonds have volatile returns due to their large durations, which scale up the influence of tiny fluctuations in the yield curve, so the more natural question is: what do market prices imply about the expected future *return* on the long bond?

We already know that the expected excess return on the long bond is positive in every state, for Result 5 implies that $\overline{r}_\infty(i) \geq r_f(i)$ and hence, by Jensen's inequality, that $\overline{R}_\infty(i) > R_f(i)$. (Here our assumption that the short rate is not constant has bite. If it *is* constant, then the strict inequality becomes a weak inequality—in fact, $\overline{R}_\infty(i) = R_f(i)$, the yield curve is flat and all bonds earn the riskless rate.) But we can do better: the next result, which is closely related to results in Martin (2013), shows that the prices of options on the long bond reveal its conditional expected excess return, and indeed the higher moments of its return.

Result 7. *Options on the long bond reveal its conditional expected excess return:*

$$\overline{R}_\infty(i) - R_f(i) = 2\left\{\int_0^{R_f(i)} \text{put}(K;i)\,dK + \int_{R_f(i)}^\infty \text{call}(K;i)\,dK\right\}, \tag{4}$$

where call$(K;i)$ *is the price, in state i, of a call option with strike K, maturing next period, on the long bond return, and* put$(K;i)$ *is the corresponding put price.*

More generally, option prices reveal all the conditional moments of the return on the long bond. The nth conditional moment of the long bond return, $\overline{R^n_\infty}(i)$, satisfies

$$\overline{R^n_\infty}(i) - R_f(i)^n = n(n+1)\left\{\int_0^{R_f(i)} K^{n-1}\,\text{put}(K;i)\,dK + \int_{R_f(i)}^\infty K^{n-1}\,\text{call}(K;i)\,dK\right\}.$$

Proof. The first statement is a special case of the second, which we now prove. Suppose we are in state i. Substitute $x(j) = R_\infty(i,j)^{n+1}$ in equation (3):

$$\sum_j \boldsymbol{A}(i,j) R_\infty(i,j)^{n+1} = \sum_j \pi(i,j) R_\infty(i,j)^n. \tag{5}$$

The right-hand side is the desired conditional moment, $\overline{R_\infty^n}(i)$. The left-hand side is the price of a claim to the $(n+1)$-th power of the long bond return, settled next period.

If options on the long bond return are traded, this payoff can be priced by a static no-arbitrage argument. To do so, note that $x^{n+1} = n(n+1)\int_0^\infty K^{n-1} \max\{0, x-K\}\, dK$ for arbitrary $x \geq 0$. Setting $x = R_\infty(i,j)$, multiplying on both sides by $\boldsymbol{A}(i,j)$, summing over j, and interchanging sum and integral, this implies that

$$\sum_j \boldsymbol{A}(i,j) R_\infty(i,j)^{n+1} = n(n+1)\int_0^\infty K^{n-1} \underbrace{\sum_j \boldsymbol{A}(i,j) \max\{0, R_\infty(i,j) - K\}}_{\text{call}(K;i)}\, dK.$$

Now we are essentially done. Splitting the range of integration and using the put–call parity relation $\text{call}(K;i) - \text{put}(K;i) = 1 - K/R_f(i)$ we can rewrite the above equation as

$$\sum_j \boldsymbol{A}(i,j) R_\infty(i,j)^{n+1} = n(n+1)\left\{\int_0^{R_f(i)} K^{n-1}\, \text{put}(K;i)\, dK + \int_{R_f(i)}^\infty K^{n-1}\, \text{call}(K;i)\, dK\right\} + R_f(i)^n.$$

Together with equation (5), this gives the result. □

Result 7 connects the forward-looking expected return on the long bond to its volatility surface. The right-hand side of (4) can be thought of as a bond variance index analogous to (the square of) the VIX index, with two modifications: first, the options that enter the index are on the long bond (rather than on an equity index) and, second, the options are equally weighted by strike (rather than by $1/K^2$). Note, in particular, that the result makes no use of historical time-series data. Equation (4) suggests a novel direct test of the weak form of the expectations hypothesis, one version of which implies that the expected excess return on the long bond should be constant. For this to be true, the index on the right-hand side of (4) would also have to be constant.

3 Discussion

We have focussed on fixed-income markets for two reasons. First, our assumption that the state variable follows a Markov chain implies that it is stationary. This accords well with familiar properties of the yield curve but is less appropriate for pricing assets with growing cashflows; see Hansen and Scheinkman (2009) for more on the latter. Second, in support of our complete-market assumption, the fixed-income derivatives market is the most well-developed of all derivatives markets.

Our starting point was a matrix, $\boldsymbol{A}$, of Arrow–Debreu prices. Associated with this matrix are a pure time discount factor, ϕ; a diagonal matrix, $\boldsymbol{D}$, with the inverse marginal utility vector $\boldsymbol{v}$ along its diagonal; and a matrix, $\boldsymbol{\Pi}$, of state transition probabilities:

$$\boldsymbol{A} = \phi \boldsymbol{D} \boldsymbol{\Pi} \boldsymbol{D}^{-1}. \tag{6}$$

Given $\boldsymbol{A}$, the quantities ϕ, $\boldsymbol{D}$, and $\boldsymbol{\Pi}$, are determined uniquely. This is surprising: we do not normally expect to back out real-world probabilities, $\boldsymbol{\Pi}$, from asset price data alone. The issue of determining the matrix $\boldsymbol{A}$ remains, however, and is nontrivial.

Consider the problem of computing Arrow–Debreu prices in today's state, i. (It is only slightly more difficult to compute prices in a state other than the current one; the essential problem is the same.) Assuming options on the state variable are traded, we can back out the prices of Arrow–Debreu securities by constructing butterfly spreads (Breeden and Litzenberger (1978)). If, for example, the state variable is the short rate, Arrow–Debreu prices can be determined by observing caps and floors across a range of strikes.

Things are more complicated if the state variable is multi-dimensional. For the sake of argument, suppose that the state variable can be expressed as a pair (r, s) where r is a one-period yield and s is a T-period yield, and that options are traded on the one-period and on the T-period yields. Then we can create, via butterfly spreads, securities that pay off only if the one-period yield equals some pre-specified value, say 3%; and we can create securities that pay off only if the T-period yield equals, say, 5%. But these are not what we need, which is to synthesize Arrow–Debreu securities that pay off only if the one-period yield equals 3% *and* the T-period yield equals 5%.

In principle, Theorem 4 of Ross (1976) offers a way around this problem: there exists some number λ such that if options on $r + \lambda s$ are traded, we can back out Arrow–Debreu prices. (Choose λ so that $r + \lambda s$ takes a different value in every state. Then appropriate butterfly spreads on $r + \lambda s$ pick out individual states.) Unfortunately we do not observe liquid markets in options on arbitrary linear combinations of r and s, and the task is even harder if the state variable has more dimensions: with a three-dimensional state vector (ρ, σ, τ), we need to observe options on $\rho + \lambda\sigma + \mu\tau$ for some λ and μ.

We think that the problem of computing $\boldsymbol{A}$ from observable data is interesting and important. But in this paper we have taken an alternative route. The matrix $\boldsymbol{A}$ remains in the background: our results sidestep the need to observe its elements directly. The fundamental decomposition (6) breaks $\boldsymbol{A}$ into a scalar, ϕ, a vector, $\boldsymbol{v}$, and a matrix, $\boldsymbol{\Pi}$. One might perhaps expect that the scalar should be easiest, and the matrix hardest, to infer from observables. This turns out to be true, in the sense that ϕ can be observed directly from the long end of the yield curve, and $\boldsymbol{v}$ can be inferred from the time series behavior of the long bond.

We use this approach to establish two results. We show that the yield curve must slope upwards on average; this is a well-known empirical regularity. And we bring $\boldsymbol{\Pi}$ into the picture by showing how to compute the expected return, and all higher moments, of the long bond in terms of the prices of options on the long bond. Aside from the intrinsic interest of forecasting long bond returns, this suggests a novel approach to testing the expectations hypothesis.

4 References

Alvarez, F., and U. J. Jermann (2005), "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73:6:1977–2016.

Backus, D. K., M. Chernov, and S. E. Zin (2013), "Sources of Entropy in Representative Agent Models," *Journal of Finance*, forthcoming.

Backus, D. K., A. W. Gregory, and S. E. Zin (1989), "Risk Premiums in the Term Structure: Evidence from Artificial Economies," *Journal of Monetary Economics*, 24:371–399.

Breeden, D. T., and R. H. Litzenberger (1978), "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51:4:621–651.

Chernov, M., and P. Mueller (2012), "The Term Structure of Inflation Expectations," *Journal of Financial Economics*, 106:367–394.

Dybvig, P. H., J. E. Ingersoll, and S. A. Ross (1996), "Long Forward and Zero-Coupon Rates Can Never Fall," *Journal of Business*, 69:1:1–25.

Hansen, L. P., and J. A. Scheinkman (2009), "Long-Term Risk: An Operator Approach," *Econometrica*, 77:1:177–234.

Horn, R. A., and C. R. Johnson (1990), *Matrix Analysis*, Cambridge University Press, Cambridge, UK.

Kazemi, H. B. (1992), "An Intemporal Model of Asset Prices in a Markov Economy with a Limiting Stationary Distribution," *Review of Financial Studies*, 5:1:85–104.

Martin, I. W. R. (2013), "Simple Variance Swaps," working paper.

Minc, H. (1988), *Nonnegative Matrices*, John Wiley & Sons, New York, NY.

Ross, S. A. (1976), "Options and Efficiency," *Quarterly Journal of Economics*, 90:1:75–89.

Ross, S. A. (2013), "The Recovery Theorem," *Journal of Finance*, forthcoming.

A Appendix

A.1 Long-dated assets and the kernel

Consider pricing a long-dated asset with a single time-T payoff of $x(i)$ if the economy is in state i at time T. In state i, the price of this asset is

$$\sum_{j=1}^{m} \boldsymbol{A}^T(i,j)x(j).$$

The following result provides a way to view pricing of long-dated assets, that is, when T is large. Define the matrix $\boldsymbol{L}$ to have (i,j)-th element $v(i)w(j)$, where $v(i)$ are the elements of the right-eigenvector $\boldsymbol{v}$ and $w(j)$ are the elements of the left-eigenvector $\boldsymbol{w}$ (the positive eigenvector that satisfies $\boldsymbol{w}'\boldsymbol{A} = \phi\boldsymbol{w}'$), normalized so that $\boldsymbol{w}'\boldsymbol{v} = 1$; and let ψ be the second-largest of the absolute values of the eigenvalues of $\boldsymbol{A}$, so $0 \leq \psi < \phi < 1$.

Fact 1 (Horn and Johnson, Theorem 8.5.1)**.** *The largest difference between the elements of $\boldsymbol{A}^T$ and $\phi^T\boldsymbol{L}$ decays exponentially in T: it is at most $C\delta^T$, where $\delta > \psi$ can be chosen arbitrarily close to ψ—and in particular, to be less than ϕ—and C is a constant.*

This fact formalizes the sense in which $\boldsymbol{A}^T \approx \phi^T \boldsymbol{L}$. The price in state i of the asset above is therefore $\sum_{j=1}^m \phi^T \boldsymbol{L}(i,j)x(j) + O(\delta^T) = \sum_{j=1}^m \phi^T v(i)w(j)x(j) + O(\delta^T)$, where $O(\cdot)$ is the order symbol. The term of order δ^T decays exponentially in T, and at a faster rate than ϕ^T, so is negligible for large T.

Up to the negligibly small error term we have, therefore, a pricing decomposition: the price in state i reflects the effect of horizon (through ϕ^T), asset-specific risk (through $x(j)$, with risk-weights determined by the left-eigenvector $w(j)$), and an economy-wide state-specific effect (through $v(i)$):

$$\text{price} \approx \underbrace{\phi^T}_{\text{time}} \times \underbrace{\sum_{j=1}^m w(j)x(j)}_{\text{asset-specific risk}} \times \underbrace{v(i)}_{\text{state}}.$$

This approximate equality implies that all sufficiently long-dated assets have highly correlated returns, because—up to the negligible error term—they all share the same dependence on the current state via $v(i)$, so all have the return $v(j)/[\phi v(i)]$ if the state moves from i to j.

We summarize the above discussion in the following result, which is related to results of Hansen and Scheinkman (2009).

Result 8. *The return on an arbitrary fixed-income asset maturing at time T is very close to the return on the long bond, if T is sufficiently large. More precisely, the realized return following a transition from state i to state j is*

$$R_\infty(i,j) + O\left(\varepsilon^T\right), \tag{7}$$

where $O(\cdot)$ indicates order notation and ε, which satisfies $\psi/\phi < \varepsilon < 1$, can be chosen arbitrarily close to ψ/ϕ; recall that ϕ and ψ are the largest and second-largest, respectively, of the absolute values of the eigenvalues of $\boldsymbol{A}$. Thus the eigenvalue gap ψ/ϕ *determines how rapidly long-dated assets' returns converge to the return on the long bond.*

Proof. Given the discussion preceding the result, this is almost immediate. Writing $K = \sum_j w(j)x(j)$ for the asset-specific risk term, the given realized return equals

$$\frac{K\phi^{T-1}v(j) + O(\delta^{T-1})}{K\phi^T v(i) + O(\delta^T)}.$$

Since $R_\infty(i,j) = v(j)/[\phi v(i)]$ by Result 2, this simplifies to (7) after defining $\varepsilon = \delta/\phi$. $\square$

Thus, in principle, the realized return on *any* sufficiently long-dated asset is a satisfactory proxy for the return on the long bond.

A.2 The constant-dividend-yield asset

Much of our analysis above focussed on one-period assets defined by their payoff vector $\boldsymbol{x}$, the price of such an asset, in vector form (i.e. listing the asset's price by state), being $\boldsymbol{Ax}$. We can also consider an infinitely-lived asset that pays $\boldsymbol{x}$ in *every* period: its price is $\boldsymbol{Ax} + \boldsymbol{A}^2\boldsymbol{x} + \boldsymbol{A}^3\boldsymbol{x} + \cdots$. Since we have assumed that $\phi < 1$, we can define

$$\boldsymbol{A}^* = \boldsymbol{A}(\boldsymbol{I} - \boldsymbol{A})^{-1} = \boldsymbol{A} + \boldsymbol{A}^2 + \boldsymbol{A}^3 + \cdots ,$$

so that the price of an infinitely-lived asset that pays $\boldsymbol{x}$ every period is $\boldsymbol{A}^*\boldsymbol{x}$. The matrix $\boldsymbol{A}^*$ has the same unique positive eigenvector $\boldsymbol{v}$; the associated eigenvalue is now $\frac{\phi}{1-\phi}$.

We now provide an alternative interpretation for the eigenvector $\boldsymbol{v}$.

Result 9. *There is a unique infinitely-lived, limited liability asset with a constant dividend yield, namely the perpetual $\boldsymbol{v}$-asset, which pays $\boldsymbol{v}$ every period. Its dividend yield is $D/P_\infty = Y_\infty = 1/\phi - 1$. This asset's returns perfectly replicate the returns on the long bond and on the one-period $\boldsymbol{v}$-asset.*

No asset's dividend yield can be always higher, or always lower, than D/P_∞.

Proof. The price of the perpetual $\boldsymbol{v}$-asset is $\boldsymbol{A}^*\boldsymbol{v} = \frac{\phi}{1-\phi}\boldsymbol{v}$, so it is immediate that its dividend yield is constant at $1/\phi - 1 = Y_\infty$. Uniqueness follows because any asset with constant dividend yield is an eigenvector of $\boldsymbol{A}^*$; and we have seen that up to multiples there is a unique eigenvector that is positive (as required for limited liability). The return on this asset, on transitioning from state i to state j, is $[v(j) + \frac{\phi}{1-\phi}v(j)]/[\frac{\phi}{1-\phi}v(i)] = v(j)/[\phi v(i)]$. The final claim follows by applying Theorem 8.1.26 of Horn and Johnson (1990) to $\boldsymbol{A}^*$. $\square$

The last part of Result 9 can be thought of as a generalization of the first part of Result 6, which can be rephrased as saying that $\min_i R_f(i) \le 1 + Y_\infty \le \max_i R_f(i)$. It implies that if one can find bounds on the dividend yield of any asset, then these bounds also apply to the dividend yield of the perpetual $\boldsymbol{v}$-asset, i.e. to $1/\phi - 1$. Suppose, for example, that for some asset i, $D/P_i > \alpha$ in every state. Then we must have $D/P_\infty > \alpha$. For if not, then we would have $D/P_\infty \leq \alpha < D/P_i$ in every state, a contradiction.