# Stripping Coupons with Linear Programming

**DAVID E. ALLEN, LYN C. THOMAS, AND HARRY ZHENG**

**DAVID E. ALLEN** is professor of finance at the School of Finance and Business Economics of Edith Cowan University in Western Australia, Australia

**LYN C. THOMAS** is professor of operational research in the Department of Business Studies of the University of Edinburgh in the U.K.

**HARRY ZHENG** is a lecturer of operational research in the Department of Business Studies at the University of Edinburgh.

The first step in using market prices to fit the parameters of models for the price of bonds is to strip the bonds of their coupons. This is because most bond pricing models really model the current term structures of spot rates of benchmark risk-free and risky securities (Treasury and corporate bonds) — that is, the prices of zero-coupon bonds.

There are few zero-coupon bonds available in the market, however. Although Treasury STRIPS can be used to represent these theoretical risk-free spot rates, there are some problems with this approach. The main one is that the Treasury STRIPS market is less liquid than the Treasury coupon market, which means that the observed rates on STRIPS reflect a premium for liquidity. It is thus necessary to extract spot rates from yields of coupon bonds of different maturities, both in the Treasury and corporate bond markets.

The standard methods of stripping coupons are bootstrapping (Fabozzi [1998]) or linear regression (Carleton and Cooper [1976]). If for each period there is one and only one coupon bond that matures, these techniques generate a unique set of spot rates over the periods. If there are no bonds that mature for some periods, however, or if there are several bonds that mature at the same time, then there are not unique answers, and in some cases the techniques give rise to rates with unacceptable features, particularly in the case of risky bonds.

Jarrow, Lando, and Turnbull [1997], for example, use these methods to strip out the risky zero-coupon bond prices, and point out several mispricings, such as five-year AA zero-coupon bonds priced above five-year AAA zero-coupon bonds, and four-year B zero-coupon bonds priced below five-year B zero-coupon bonds. The authors attribute these mispricings to the noise of the data and the call features of some bonds.

These mispricings are bothersome, because it becomes difficult to estimate the parameters in the credit bond pricing models. There has been a resurgence of interest in such models, as they not only give investors a clear indication of current market perceptions of the riskiness of particular bonds, but are also a stepping stone to pricing many credit-sensitive fixed-income derivatives, such as callable and putable bonds, caps and floors, and mortgage-backed securities.

Jarrow and Turnbull [1998], for example, derive the default probabilities of risky bonds by combining a default process with an interest rate model. They apply the Black-Derman-Toy model to build a recombined binomial short rate tree, then combine it with the default process to form a larger tree for credit-risky bonds, and finally obtain default probabilities by forward and backward induction methods.

To remedy the mispricing caused by bootstrapping, Thomas, Allen, and Morkel-Kingsbury [1998] suggest using linear programming to strip out risky zero-coupon bond prices. This produces the same spot rates as the bootstrapping technique if there is one and only one coupon bond that matures for each period, but is always able to ensure that, for the same maturity the higher-rated zero-coupon bond is priced above the lower-rated zero-coupon bond, and for the same credit rating the shorter-maturity zero-coupon is priced above the longer-maturity zero-coupon bond.

Although the LP formulation avoids the difficulties encountered in Jarrow, Lando, and Turnbull [1997], other problems have crept in. The main one is that the gap between zero-coupon bond prices of different credit ratings widens and then narrows as time goes by, which means the forward rate of higher credit-rated zero-coupon bonds is higher than that of lower credit-rated zero-coupon bonds. Such a result again suggests that there are potential arbitrage opportunities.

We suggest a new linear programming formulation to strip out risky zero-coupon bond prices that resolves these problems. First, we use an extension of the original LP approach to strip Treasury bonds, which works whatever the current date, coupon dates, and sampling dates. Then we introduce a new LP formulation for stripping the coupons from risky corporate bonds that ensures that the spreads increase over time.

We discuss how the LP formulation can be modified to deal with liquidity issues and extend it to arbitrary time intervals between the sampling points at which the zero-coupon bond price is calculated. There is a connection between the zero-coupon bond prices obtained by the LP formulation and the default probabilities that the market is imputing to the risky corporate bonds.

## I. TREASURY STRIPS PRICES

To derive pure discount bond prices $v_0(t)$ of risk-free zero-coupon bonds paying 1 at a set of prechosen times t = 0, 1, ..., T, we use the observed market prices of $N_0$ bonds to solve the linear programming problem:

$$\text{LP1:} \quad \text{Minimize} \quad \sum_{i=1}^{N_0} (a_i + b_i) \tag{1}$$

$$\text{subject to} \quad P_i + a_i = \sum_{t=1}^{T} c_i(t) v_0(t) + b_i$$

$$v_0(t) \geq [1 + m(t)] v_0(t+1)$$

$$a_i, b_i \geq 0$$

for i = 1, ..., $N_0$ and t = 0, 1, ..., T – 1, where $P_i$ is the present value of the bond i; $c_i(t)$ is its cash flow at time t; $v_0(0) = 1$; and m(t) is the minimum expected forward rate from t to t + 1.

The first constraints seek to match the present value $P_i$ to the discounted cash flows $c_i(t)$, and $a_i$ and $b_i$ are the mispricing errors. $a_i$ is positive and $b_i = 0$ if the price is "too low"; $b_i$ is positive and $a_i = 0$ if the price is "too high." The second constraint ensures that there is no mispricing with respect to maturity. (When m(t) = 0, the constraint corresponds to saying bonds of longer maturity should be priced lower than those of shorter maturity.)

The cash flow $c_i(t)$ is decided by the coupon payment, coupon date, sampling date, and current date. As an example, assume $t_1, t_2, ..., t_T$ are fixed semiannual sampling dates. A bond pays a coupon c every six months with a principal F and a maturity date before or at time $t_T$; there is one cash flow in each sampling period. Let $v_i$ be the price of the risk-free zero-coupon bond paying 1 at time $t_i$. Assume $\alpha$ is the proportion of the time between a coupon date and the next sampling date compared with the time between two sampling dates (therefore $\alpha$ is a number between 0 and 1). Let $\beta$ be the proportion of a sampling interval between the current date (when the market price of the bond is observed) and the next sampling date.

There are two cases to consider:

1. There is no coupon payment between the current date and the next sampling date.
2. There is one coupon payment.

In the first case, we have $\alpha \geq \beta$, and the relation between the market price and the future cash flows is approximated by:

$$\begin{aligned} P &= P_C + (\alpha - \beta)c \\ &= \alpha c v_1 + c v_2 + \ldots + \\ &\quad c v_{T-2} + (c + \alpha F) v_{T-1} + (1 - \alpha)(c + F) v_T \end{aligned} \tag{2}$$

Here P is the present value, $P_C$ is the "clean" market price, and $(\alpha - \beta)c$ is the accrued interest. We have split each coupon payment and the principal into two parts. One $\alpha c$ is paid at the previous sampling date, and one $(1 - \alpha)c$ is paid at the subsequent sampling date. In the second case, we have $\alpha < \beta$ and the resulting equation is

$$
\begin{aligned}
P &= P_C + (1 + \alpha - \beta)c \\
&= (\alpha/\beta)c + [\alpha + (\beta - \alpha)/\beta]cv_1 + cv_2 + \ldots + \\
&\quad cv_{T-2} + (c + \alpha F)v_{T-1} + (1 - \alpha)(c + F)v_T \qquad (3)
\end{aligned}
$$

Here $(1 + \alpha - \beta)c$ is the accrued interest.

Note that the two equations are basically the same except for the cash value at the present date and the cash paid out at $t_1$. In the case of $\alpha = 0$ and $\beta = 1$ (the current date is the sampling date, and the next coupon payment is on the next sampling date), we have a very simple equation:

$$P_C = cv_1 + \ldots + cv_{T-1} + (c + F)v_T \qquad (4)$$

This set of timings leads to the special case of LP1 where, if bond i has coupon $c_i$ and principal $F_i$,

$$c_i(t) = c_i \qquad \text{for } t = 1, 2, \ldots, T - 1$$

and

$$c_i(T) = c_i + F_i \qquad (5)$$

For more general current dates, we have that if $\alpha \geq \beta$, then the cash flows are

$$
\begin{aligned}
&c_i(1) = \alpha c_i \\
&c_i(t) = c_i \qquad \text{for } t = 2, \ldots, T - 2 \qquad (6) \\
&c_i(T - 1) = c_i + \alpha F_i
\end{aligned}
$$

and

$$c_i(T) = (1 - \alpha)(c_i + F_i)$$

For $\alpha < \beta$ with $t_0$ the current date, the cash flows are

$$
\begin{aligned}
&c_i(0) = (\alpha/\beta)c_i \\
&c_i(1) = [\alpha + (\beta - \alpha)/\beta]c_i \\
&c_i(t) = c_i \qquad \text{for } t = 2, \ldots, T - 2 \qquad (7) \\
&c_i(T - 1) = c_i + \alpha F_i
\end{aligned}
$$

and

$$c_i(T) = (1 - \alpha)(c_i + F_i)$$

The proof of Equations (2) and (3) is given in the appendix, with some other cases.

We use 113 Treasury bonds in the market on February 7, 2000, with maturity dates up to the second half of 2008. Data come from Datastream. Since coupons are paid semiannually, we choose a six-month time interval for a period, and set the sampling dates on May 15 and November 15.

Exhibit 1 is the stripped Treasury zero-coupon bond prices v(t) on February 7, 2000, given by LP1. We apply pricing Equations (2) and (3) to model the cash flows and market prices. The last column is the observed U.S. STRIPS prices on the same day from Datastream.

The total error between the market prices and the estimated prices is 2.13, and the total market value of these bonds is 7163. So the relative error is less than 0.03%, a very good fit. The results are exactly the same for several different minimum forward rates m(t) from 0

## EXHIBIT 1
**U.S. Treasury Zero-Coupon Bond Prices**

| Date | LP1 Price | Yield (%) | U.S. STRIPS |
|---|---|---|---|
| 05/15/00 | 0.9849 | 5.60 | 0.9854 |
| 11/15/00 | 0.9539 | 6.12 | 0.9548 |
| 05/15/01 | 0.9213 | 6.44 | 0.9215 |
| 11/15/01 | 0.8898 | 6.59 | 0.8904 |
| 05/15/02 | 0.8604 | 6.62 | 0.8629 |
| 11/15/02 | 0.8320 | 6.63 | 0.8363 |
| 05/15/03 | 0.8040 | 6.67 | 0.8060 |
| 11/15/03 | 0.7791 | 6.62 | 0.7795 |
| 05/15/04 | 0.7528 | 6.65 | 0.7526 |
| 11/15/04 | 0.7256 | 6.72 | 0.7239 |
| 05/15/05 | 0.7017 | 6.72 | 0.7018 |
| 11/15/05 | 0.6784 | 6.72 | 0.6789 |
| 05/15/06 | 0.6556 | 6.73 | 0.6564 |
| 11/15/06 | 0.6350 | 6.71 | 0.6359 |
| 05/15/07 | 0.6158 | 6.67 | 0.6136 |
| 11/15/07 | 0.5957 | 6.66 | 0.5974 |
| 05/15/08 | 0.5785 | 6.62 | 0.5742 |
| 11/15/08 | 0.5608 | 6.59 | 0.5560 |

to 0.03, which implies that the choice of m(t) is fairly robust. Comparing the result with the observed U.S. STRIPS prices, we see they are very close.

## II. RISKY ZERO-COUPON BOND PRICES

Suppose bonds are classified according to their riskiness into ratings from 1 to M. The bond rated 1 has the highest quality and the lowest default risk, and the bond rated M has the lowest quality and the highest default risk. Suppose there are N bonds observable in the market. Bond i has present value $P_i$, maturity date $T_i$, cash flows $c_i(t)$ for t = 1, 2, ..., $T_i$, and credit rating d(i). Define $c_i(t) = 0$ for i = $T_1$ + 1, ..., T where T is the longest-maturity date among N bonds. Suppose for the class of bonds with credit rating j the price of a bond stripped of its coupons paying 1 at date t is $v_j(t)$ for t = 1, ..., T.

To construct these term structures of spot rate curves of credit-risky bonds, assuming we have already calculated the zero-coupon Treasury bond prices $v_0(t)$, t = 1, 2, ..., T, we can formulate and solve the linear programming problem:

$$\text{LP2: Minimize} \quad \sum_{i=1}^{N_0}(a_i + b_i) \qquad (8)$$

$$\text{subject to} \quad P_i + a_i = \sum_{t=1}^{T} c_i(t) v_{d(i)}(t) + b_i$$

$$v_j(t+1) - v_{j+1}(t+1) \geq v_j(t) - v_{j+1}(t)$$

$$a_i, b_i \geq 0$$

for i = 1, ..., N, j = 0, ..., M – 1, and t = 0, 1, ..., T – 1, where $v_j(0) = 1$.

The inequalities $v_j(t+1) - v_{j+1}(t+1) \geq v_j(t) - v_{j+1}(t)$ are used to characterize these bond properties: The price of a longer-maturity bond is cheaper than that of a shorter-maturity bond, and the price of a higher-rated bond is higher than that of a lower-rated bond. The first condition is satisfied by rewriting the constraint as $v_{j+1}(t) - v_{j+1}(t+1) \geq v_j(t) - v_j(t+1)$ and repeatedly applying it from rating j to 0 using the fact that $v_0(t) - v_0(t+1) \geq 0$. The second condition is satisfied by repeatedly applying the constraint from time 0 to t since $v_j(0) - v_{j+1}(0) = 1 - 1 = 0$.

The constraint actually conveys more information. It says that the forward rates of higher-rated bonds are lower than those of lower-rated bonds. This will become clear when we study default probabilities of credit-risky bonds later.

We downloaded the list of U.S. industry corporate bonds on February 7, 2000, from Datastream, which provides information on S&P rating, amount issued, amount outstanding, next call date, and last date price changed, as well as all standard bond information. We use twenty-six AA bonds, thirty-two A bonds, and thirty-two BBB bonds with maturity up to November 15, 2005 (six years), after excluding bonds that are unrated, or have call options embedded, or have different issuing amount and outstanding amount, or have not been traded for at least two months, or are of market value less than 100,000. The last two criteria try to remove bonds whose prices may be irrelevant because of illiquidity. We do not include AAA, BB, or B bonds either, as there are relatively few such bonds available.

Exhibit 2 shows the corporate pure discount bond prices derived using the linear programming model with the Treasury pure discount bond prices as reference $v_0(t)$. The last three columns are the yield spreads between the Treasury bonds and the corporate bonds (in basis points).

The increasing gap between risk-free and risky bond prices indicates the increasing default risks over longer terms. The relative errors of LP prices and observed market prices are 0.25% for AA bonds, 0.41% for A bonds, and 1.16% for BBB bonds. The increased errors may be partly due to the ripple effects of higher-rated bond pricing errors. We do not need to calculate the Treasury bonds and the corporate bond prices separately, but can calculate their zero-coupon prices in the same LP problem. This incorporates the constraints of LP1 and LP2 into

$$\text{LP3: Minimize} \quad \sum_{i=1}^{N_0+N}(a_i + b_i) \qquad (9)$$

$$\text{subject to} \quad P_i + a_i = \sum_{t=1}^{T} c_i(t) v_{d(i)}(t) + b_i$$

$$v_0(t) \geq [1 + m(t)] v_0(t+1)$$

$$v_j(t+1) - v_{j+1}(t+1) \geq v_j(t) - v_{j+1}(t)$$

$$a_i, b_i \geq 0$$

## EXHIBIT 2
**Risky Zero-Coupon Bond Prices and Yield Spreads**

| | | Bond Prices | | | Yield Spreads (bp) | | |
|---|---|---|---|---|---|---|---|
| **Date** | **Treasury** | **AA** | **A** | **BBB** | **AA** | **A** | **BBB** |
| 05/15/00 | 0.9849 | 0.9831 | 0.9831 | 0.9828 | 67 | 67 | 76 |
| 11/15/00 | 0.9538 | 0.9478 | 0.9478 | 0.9476 | 82 | 82 | 85 |
| 05/15/01 | 0.9214 | 0.9153 | 0.9134 | 0.9109 | 52 | 68 | 90 |
| 11/15/01 | 0.8898 | 0.8837 | 0.8818 | 0.8748 | 38 | 51 | 95 |
| 05/15/02 | 0.8604 | 0.8503 | 0.8483 | 0.8414 | 52 | 62 | 98 |
| 11/15/02 | 0.8320 | 0.8122 | 0.8103 | 0.8034 | 87 | 96 | 127 |
| 05/15/03 | 0.8040 | 0.7842 | 0.7823 | 0.7753 | 76 | 84 | 111 |
| 11/15/03 | 0.7791 | 0.7551 | 0.7531 | 0.7462 | 83 | 90 | 115 |
| 05/15/04 | 0.7528 | 0.7252 | 0.7212 | 0.7143 | 87 | 100 | 123 |
| 11/15/04 | 0.7256 | 0.6980 | 0.6940 | 0.6871 | 81 | 93 | 114 |
| 05/15/05 | 0.7017 | 0.6737 | 0.6697 | 0.6545 | 77 | 88 | 132 |
| 11/15/05 | 0.6784 | 0.6386 | 0.6346 | 0.6194 | 105 | 115 | 158 |

for $i = 1, ..., N_0 + N$, $j = 0, 1, ..., M - 1$, and $t = 0, 1, ..., T - 1$.

## III. LIQUIDITY ISSUES

The objective of a linear programming model is to minimize the sum of all under/over errors. Such a formulation indicates all bonds are treated equally. Yet the issue amount of each bond may be quite different, from hundreds of thousands of dollars for a corporate bond to tens of millions of dollars for a Treasury bond. This has a significant impact on the liquidity of individual bonds. If the amount outstanding of a bond is small (the bid-ask spread tends to widen (to compensate for possible illiquidity), which may result in higher/lower bond prices than for other more liquid bonds. Therefore, we should treat each bond differently depending on its liquidity.

One way to do this is to use the amount outstanding information of all bonds in the market, which is readily available from financial information services such as Datastream. If some bonds have much lower amounts outstanding than other bonds, we may treat them as illiquid and remove them from the data set. This approach is easily implemented by setting a threshold value and removing any bonds whose amount outstanding is below that value.

This is the method we have used so far. For the Treasury bonds, the cutoff point is set to be $10 million, which is less than the amount outstanding of most Treasury bonds. For corporate bonds the cutoff point is $100,000. This approach retains most liquid bonds while removes some possibly illiquid bonds.

The disadvantage of this approach is how to choose a threshold value. This problem can be easily solved by reformulating the LP model. Instead of the simple sum of under/over errors of the objective function, we can use the weighted sum of under/over errors. The weight of a bond is the proportion of its amount outstanding to the total amount outstanding of all bonds in the market.

To write out this idea mathematically, suppose there are N bonds to be used to derive pure discount bond prices, and bond i has amount outstanding $M_i$. Then the objective function is defined as

$$\text{Minimize} \quad \sum_i w_i(a_i + b_i) \tag{10}$$

where weights $w_i = M_i/M$ and $M = M_1 + ... + M_N$.

The obvious advantage of this approach is that we do not need to set a threshold value to remove possible illiquid bonds. If a bond has a smaller amount outstanding, its weight is also small compared to other bonds. Since weights act as penalty costs in the objective function, the LP model will try to minimize errors of the bonds with greater weights and pay less attention to those with smaller weights. This in turn removes the effect of bonds with small amounts outstanding.

Exhibit 3 shows the Treasury pure discount bond prices using the weighted LP model and all relevant Trea-

## EXHIBIT 3
**Bond Prices Using Weighted LP Models**

| Date | LP Price | Weighted LP | U.S. STRIPS |
|---|---|---|---|
| 05/15/00 | 0.9849 | 0.9849 | 0.9854 |
| 11/15/00 | 0.9539 | 0.9539 | 0.9548 |
| 05/15/01 | 0.9213 | 0.9214 | 0.9215 |
| 11/15/01 | 0.8898 | 0.8898 | 0.8904 |
| 05/15/02 | 0.8604 | 0.8604 | 0.8629 |
| 11/15/02 | 0.8320 | 0.8322 | 0.8363 |
| 05/15/03 | 0.8040 | 0.8038 | 0.8060 |
| 11/15/03 | 0.7791 | 0.7791 | 0.7795 |
| 05/15/04 | 0.7528 | 0.7528 | 0.7526 |
| 11/15/04 | 0.7256 | 0.7256 | 0.7239 |
| 05/15/05 | 0.7017 | 0.7022 | 0.7018 |
| 11/15/05 | 0.6784 | 0.6783 | 0.6789 |
| 05/15/06 | 0.6556 | 0.6555 | 0.6564 |
| 11/15/06 | 0.6350 | 0.6354 | 0.6359 |
| 05/15/07 | 0.6158 | 0.6158 | 0.6136 |
| 11/15/07 | 0.5957 | 0.5957 | 0.5974 |
| 05/15/08 | 0.5785 | 0.5785 | 0.5742 |
| 11/15/08 | 0.5608 | 0.5604 | 0.5560 |

sury bonds. We note that 90% of pricing errors are caused by 30% of the most illiquid bonds. The prices derived from the two LP models are remarkably close, which may be because a threshold value is used in the original LP model.

### IV. SAMPLING INTERVALS

So far we have dealt with six monthly intervals between the sampling dates, but we might want to have finer sampling dates for the near future and sparser ones for the distant future. The general pricing equations can be extended to allow for this as follows. For each cash flow $c_k$ at time $s_k$, we can find two adjacent sampling dates $t_n$ and $t_{n+1}$ such that $s_k$ lies in between. Define

$$\alpha_k = \frac{t_{n+1} - s_k}{t_{n+1} - t_n} \tag{11}$$

Then the discount factor $\tilde{v}_k$ at time $s_k$ can be approximated as

$$\tilde{v}_k = \alpha_k v_n + (1 - \alpha_k) v_{n+1} \tag{12}$$

The present value of all cash flows is the sum of $c_k \tilde{v}_k$, which then leads to a pricing equation.

Suppose the sampling periods are six months for the first five years, and then one year for the next ten years. If a bond has three years to maturity, no change is required. If a bond has ten years to maturity, then in the first five years, the contributions of each cash flow to its adjacent sampling dates are $1 - \alpha$ and $\alpha$, respectively. From year six, there are two cash flows in each interval. The contributions of the first cash flow to its adjacent sampling dates are $(1 + \alpha)/2$ and $(1 - \alpha)/2$, respectively; those of the second cash flow are $\alpha/2$ and $(2 - \alpha)/2$, respectively. This approach can simplify derivation of discount factors for bonds covering very long periods.

For the same bond data as above, we use semiannual intervals for years 2000 to 2003, and annual intervals for years 2004 to 2008. The results, given in Exhibit 4, are very similar to the equal sampling period results.

### V. DEFAULT PROBABILITIES

We have described a way of constructing theoretical Treasury and corporate pure discount bond prices from the observed coupon bond prices. The yield spread between Treasury STRIPS and corporate strips represents the premium of several risk factors, e.g., default risk, liquidity risk, sector risk. To simplify matters, we assume the yield spread is due purely to default risk. This assumption obviously exaggerates the default risk, but it makes calculation of default probabilities easier, and at least it gives

## EXHIBIT 4
**Bond Prices Using Varying Sampling Intervals**

| Date | Price | Yield (%) | Semiannual Price |
|---|---|---|---|
| 05/15/00 | 0.9849 | 5.60 | 0.9849 |
| 11/15/00 | 0.9539 | 6.12 | 0.9539 |
| 05/15/01 | 0.9213 | 6.44 | 0.9213 |
| 11/15/01 | 0.8898 | 6.59 | 0.8898 |
| 05/15/02 | 0.8604 | 6.62 | 0.8604 |
| 11/15/02 | 0.8320 | 6.63 | 0.8320 |
| 05/15/03 | 0.8040 | 6.67 | 0.8040 |
| 11/15/03 | 0.7791 | 6.62 | 0.7791 |
| 05/15/04 | 0.7256 | 6.72 | 0.7256 |
| 11/15/05 | 0.6784 | 6.72 | 0.6784 |
| 05/15/06 | 0.6348 | 6.71 | 0.6350 |
| 11/15/07 | 0.5967 | 6.64 | 0.5957 |
| 05/15/08 | 0.5608 | 6.59 | 0.5608 |

the upper bound of the risk perceived by the market.

Suppose the Treasury STRIPS prices $v_0(t)$ and the corporate zero-coupon bond prices $v_i(t)$ are given, where i is the credit rating. If a company defaults before its bond matures, then a proportion $\delta$ (the recovery rate) of the face value, discounted by the remaining years to maturity, is given to bondholders. This is the assumption that most authors make (Jarrow and Turnbull [1998], for example). (We assume the recovery rate is the same for all bonds, whether AAA bonds or C bonds, for simplicity. This assumption can be relaxed to make $\delta$ credit-rating dependent.)

Denote $Q_k^i$ and $P_k^i$ as the cumulative default and survival probabilities of a bond currently rated i at the end of period k, respectively, and let $q_k^i$ and $p_k^i$ be the marginal default and survival probabilities in period k. Then, because if a t maturity zero-coupon bond does not default at all it is worth $v_0(t)$, while under the assumption above if it does default it is worth $\delta v_0(t)$, one has

$$v_i(k) = (1 - Q_k^i)v_0(k) + Q_k^i \delta v_0(k)$$

or

$$Q_k^i = \frac{1}{1-\delta}\left(1 - \frac{v_i(k)}{v_0(k)}\right) \tag{13}$$

for k = 1, 2, ... . The other probabilities can be easily computed using the relations

$$P_k^i = 1 - Q_k^i$$

$$p_k^i = P_k^i / P_{k-1}^i$$

$$q_k^i = 1 - p_k^i \tag{14}$$

for k = 1, 2, ..., where $P_0^i = 1$, i.e., a risky bond is not in default at time 0.

Exhibit 5 lists default probabilities derived with Equations (13) and (14) for an example in Jarrow and Turnbull (1998). The recovery rate $\delta$ is assumed to be 0.4.

The result is the same as that Jarrow and Turnbull [1998] derive by building an interest rate tree as well as a default tree. The significance of these recursive formulas is twofold. They provide a quick way to compute default probabilities, and they illustrate the independence between default probabilities and interest rate models. The two issues are decoupled.

## EXHIBIT 5
**An Example by Jarrow and Turnbull [1998]**

| k | $v_0(k)$ | $v_i(k)$ | $q_k^i$ | $Q_k^i$ |
|---|---|---|---|---|
| 1 | 0.953921 | 0.950486 | 0.0060 | 0.0060 |
| 2 | 0.906264 | 0.897056 | 0.0110 | 0.0169 |
| 3 | 0.857820 | 0.841008 | 0.0160 | 0.0327 |

## EXHIBIT 6
**Default Probabilities of Risky Bonds**

| | Marginal Default Probabilities | | | Cumulative Default Probabilities | | |
|---|---|---|---|---|---|---|
| Date | AA | A | BBB | AA | A | BBB |
| 15/05/00 | 0.0030 | 0.0030 | 0.0035 | 0.0030 | 0.0030 | 0.0035 |
| 15/11/00 | 0.0075 | 0.0075 | 0.0075 | 0.0105 | 0.0105 | 0.0109 |
| 15/05/01 | 0.0004 | 0.0040 | 0.0081 | 0.0109 | 0.0144 | 0.0189 |
| 15/11/01 | 0.0004 | 0.0005 | 0.0091 | 0.0113 | 0.0149 | 0.0278 |
| 15/05/02 | 0.0083 | 0.0084 | 0.0089 | 0.0195 | 0.0232 | 0.0365 |
| 15/11/02 | 0.0202 | 0.0204 | 0.0208 | 0.0393 | 0.0431 | 0.0566 |
| 15/05/03 | 0.0014 | 0.0016 | 0.0020 | 0.0407 | 0.0446 | 0.0585 |
| 15/11/03 | 0.0105 | 0.0106 | 0.0111 | 0.0508 | 0.0548 | 0.0690 |
| 15/05/04 | 0.0097 | 0.0143 | 0.0149 | 0.0600 | 0.0683 | 0.0828 |
| 15/11/04 | 0.0023 | 0.0027 | 0.0032 | 0.0622 | 0.0708 | 0.0858 |
| 15/05/05 | 0.0031 | 0.0034 | 0.0238 | 0.0650 | 0.0740 | 0.1075 |
| 15/11/05 | 0.0313 | 0.0316 | 0.0336 | 0.0943 | 0.1033 | 0.1375 |

For the Treasury and corporate zero-coupon bond prices derived in Exhibits 1 and 2, we can quickly compute the cumulative and marginal default probabilities. The results are shown in Exhibit 6.

Marginal or cumulative default probabilities that are negative or greater than 1 clearly indicate that there are mispricings of zero-coupon bonds. The LP formulation ensures that this will not happen.

## VI. CONCLUSION

We have shown that linear programming can be used to strip coupons for both Treasury and corporate bonds. The advantages of the LP approach are that there is no mispricing and the spread structure is built into the model. Real data can be easily analyzed since the LP formulation works whatever the current date, coupon dates, and sampling dates. The weighted LP model can be used to deal with data that may include some less liquid bonds. Finally, default probabilities of risky bonds perceived by the market can be easily calculated without relying on any interest rate models.

## APPENDIX

**Proof of Equations (2) and (3)**

There are only two cases need to be considered: 1) there is no coupon payment between now and the next sampling date, and 2) there is a coupon payment between now and the next sampling date. The sampling dates (from the next one) are labeled as t = 1, 2, ..., T.

The first case corresponds to $\alpha \geq \beta$. The present value of all cash flows of a bond is equal to the sum of $\tilde{c}(t)\tilde{v}(t)$ over all t from 1 to T – 1, where $\tilde{c}(t)$ is the cash flow, and $\tilde{v}(t)$ is the discount factor at time t. Using linear interpolation, we can write the value of $\tilde{v}(t)$ as a combination of repayments at t and t + 1

$$\tilde{v}(t) = \alpha v_t + (1 - \alpha)v_{t+1}$$

for t = 1, ..., T – 1. Therefore the principal value PV satisfies

$$\begin{aligned} PV &= \tilde{c}(1)\tilde{v}(1) + \dots + \tilde{c}(T-1)\tilde{v}(T-1) \\ &= c(\alpha v_1 + (1-\alpha)v_2) + \dots + (c + F)(\alpha v_{T-1} + (1-\alpha)v_T) \\ &= \alpha c v_1 + c v_2 + \dots + c v_{T-2} + (c + \alpha F)v_{T-1} + (1-\alpha)(c + F)v_T \end{aligned}$$

The accrued interest is taken in the market to be $(\alpha - \beta)c$, and the general equation for bond prices is given by

$$MP + AI = PV$$

where MP is the market (clean) price of the bond. Substituting the accrued interest and the present value into the equation, we obtain the first pricing equation.

The second case corresponds to $\alpha < \beta$. The present value of all cash flows of a bond is equal to the sum of $\tilde{c}(t)\tilde{v}(t)$ over all t from 0 to T – 1. The discount factors $\tilde{v}(t)$ can be computed the same as above for t = 1, ..., T – 1. In computing $\tilde{v}(0)$, however, we must remember that the time interval from now to the first cash flow is $\beta - \alpha$, and the time interval from now to the next sampling date is $\beta$, so linear interpolation gives

$$\tilde{v}_0 = \frac{\alpha}{\beta} + \frac{\beta - \alpha}{\beta} v_1$$

If there is only one cash flow in the future, then

$$PV = \frac{\alpha}{\beta}(c + F) + \frac{\beta - \alpha}{\beta}(c + F)v_1$$

If there are two cash flows in the future, then

$$PV = \frac{\alpha}{\beta}c + \left[\frac{\beta - \alpha}{\beta}c + \alpha(c + F)\right]v_1 + (1-\alpha)(c + F)v_2$$

If there are three or more cash flows in the future, then

$$PV = \frac{\alpha}{\beta}c + \left(\frac{\beta - \alpha}{\beta}c + \alpha\right)cv_1 + cv_2 + \dots + cv_{T-2} + (c + \alpha F)v_{T-1} + (1-\alpha)(c + F)v_T$$

The accrued interest is equal to $(1 - \beta + \alpha)c$. Substituting everything into the general equation for bond price gives the second pricing equation.

## ENDNOTE

The authors are grateful to Datastream for the data and to Nigel Morkel-Kingsbury for advice on extracting the data.

## REFERENCES

Carleton, W., and I. Cooper. “Estimation and Uses of the Term Structure of Interest Rates.” *Journal of Finance*, 31, No. 4 (1976), pp. 1067-1083.

Fabozzi, F. *Valuation of Fixed Income Securities and Derivatives*. New Hope, PA: Frank J. Fabozzi Associates, 1998.

Jarrow, R., D. Lando, and S. Turnbull. “A Markov Model for the Term Structure of Credit Risk Spreads.” *Review of Financial Studies*, 10, No. 2 (1997), pp. 481-523.

Jarrow, R., and S. Turnbull. “Credit Risk.” In C. Alexander, ed., *Risk Management and Analysis*. New York: John Wiley & Sons, 1998, pp. 237-254.

Thomas, L., D. Allen, and N. Morkel-Kingsbury. “A Hidden Markov Chain Model for the Term Structure of Bond Credit Risk Spreads.” Working paper, Centre for Financial Markets Research, University of Edinburgh, 1998.