ON BOOTSTRAPPING HAZARD RATES FROM CDS SPREADS

GIUSEPPE CASTELLACCI, PHD

Abstract. We present a simple procedure to construct credit curves by bootstrapping a hazard rate curve from observed CDS spreads. The hazard rate is assumed constant between subsequent CDS maturities. In order to link survival probabilities to market spreads, we use the JPMorgan model, a common market practice. We also derive approximate closed formulas for "cumulative" or "average" hazard rates and illustrate the procedure with examples from observed credit curves.

Contents

Date: April 18, 2012, 2:48pm; Version 2. First Version: August 14, 2008.

Key words and phrases. Hazard rates, risk-neutral hazard rates, risk-neutral default probabilities, CDS spread.

An early version of this paper was presented at the CITI Quant seminar on November 5, 2008.

Typeset in LaTeX 2_ε.

1. Introduction

One of the key tasks in the valuation of credit derivatives is the estimation of default and/or survival probabilities for individual names. The so called *credit curve*, that is, the term structure of such probabilities, is a fundamental input to the valuation of both single-name and portfolio credit derivatives.[1] If a credit curve is estimated from prices or other observables corresponding to liquid securities for a given name, then we obtain risk-neutral default and survival probabilities for that name. In general, such probabilities will reflect not only the maturity of the security, but also other features such as seniority and restructuring type.

In this note we deal with the problem of estimating default probabilities and the corresponding hazard rates from Credit Default Swap (CDS) spreads. We assume that the latter are liquid and do not deal with the delicate problem of filtering illiquid quotes. The estimation will be dependent on the model we choose. We select the so-called JPMorgan model, which we introduce below. This model is rather crude, but as it is usually the case for liquid securities, the gist of the valuation lies in the quality of the data. Models (such as the celebrated Black-Scholes-Merton for implying the volatility of vanilla options) serve rather as interpolators or quoting tools. The JPMorgan model seems to have acquired such status for CDSs.

1.1. **Reminder on Hazard Rates.** The key driver of the value of a single-name credit derivative is the time of default τ. In the mathematical modeling of these securities, τ is assumed to be a stopping time (in a filtration satisfying the "usual conditions"—for further details, cf., e.g., [Pro90]). The default probability up to time t is defined as the cumulative probability distribution function of τ, namely,

$$F(t) := \mathrm{Prob}\left(\tau \leq t\right). \tag{1.1}$$

The corresponding survival probability, that is, the probability that no default occurs until time t, is

$$S(t) = 1 - F(t) = \mathrm{Prob}\left(\tau > t\right). \tag{1.2}$$

The *hazard rate* corresponding to τ can be defined as the deterministic function[2] h such that

$$S(t) = \exp\left(-\int_0^t h(u)du\right) \tag{1.3}$$

---

[1] Of course, in the case of portfolio credit derivatives, the other key ingredient is the dependence structure, which is often modeled via copulas.

[2] We assume h is integrable on the range of τ, usually $[0, \infty]$.

provided such function exists (i.e., $\ln S(t)$ is absolutely continuous). Conversely, if S is differentiable one can obtain the hazard rate from the survival probability function as[3]

$$h(t) = -\frac{d}{dt} \ln S(t) \tag{1.4}$$

One can equivalently write the hazard rate as a function of the probability of default:

$$h(t) = \frac{F'(t)}{1 - F(t)}. \tag{1.5}$$

Remark 1.1. As a basic example, note that the arbitrage price, $\bar{P}(0,T)$, of the zero coupon bond (ZCB) with zero recovery rate delivering one unit of cash at time T can be expressed as the following risk-neutral expectation:

$$\bar{P}(0,T) = E\left[D(0,T)\mathbb{1}_{\{\tau > T\}}\right] = P(0,T)S(T), \tag{1.6}$$

where $D(0,t) := \exp\left(-\int_0^t r(u)du\right)$ is the reciprocal of the money market numéraire, r is the risk-free short rate process, $P(0,T)$ is the corresponding risk-free ZCB price, and, crucially, we are assuming interest rates are independent of default times. Hence, survival probabilities are analogous to discount factors and can be read off the risk-free and risky discount curves:

$$S(T) = \frac{\bar{P}(0,T)}{P(0,T)} \ . \tag{1.7}$$

Note that absence of arbitrage implies that $\bar{P}(0,T) < P(0,T)$, hence survival probabilities for non-trivial maturities are smaller than one. Note also that $S(T)$ can be interpreted as the forward price of the risky ZCB maturing at time T. (cf., e.g., [MR05, Section 9.6.1]).

Similarly, if survival probabilities are differentiable, hazard rates correspond the the short rate of risk-less interest rate modeling. The price of a risky ZCB with zero recovery can be written in "risk-adjusted" form:[4]

$$\bar{P}(0,T) = E\left[\exp\left(-\int_0^t (r(u) + h(u))du\right)\right] \tag{1.8}$$

[3]Given h, S could be obtained by integrating h along with the initial condition $S(0) = 1$.

[4]Note that these basic arbitrage valuations have conditional generalizations corresponding to unknown future values. For example, the value of the risky ZCB at a future time $t > t$ can be written as

$$\bar{P}(t,T) = E\left[\exp\left(-\int_0^t (r(u) + h(u))du\right) \|\mathcal{F}_t\right].$$

The hazard rate itself has a forward generalization that correspond to conditional survival probabilities, i.e., $h(t,T) = d\ln S(t,T)/dt$ where $S(t,T) = \bar{P}(t,T)/P(t,T)$

More generally, a zero coupon bond with random recovery rate[5] $\tilde{R}$ maturing at T has (pre-default) arbitrage price

$$\bar{P}(0,T) := RE\left[D(0,\tau)\mathbb{1}_{\{\tau\leq T\}}\right] + P(0,T)S(T), \tag{1.9}$$

where $R = E[\tilde{R}]$ and the expectation are taken with respect to the risk-neutral measure (cf., e.g., [JR00]). The complication in this formula derives from the assumption that recovery payments occur at default. If instead one assumes that recovery payments occur at a pre-specified set of times (e.g., coupon payment dates if this were a coupon bond or other security with intermediate payments) $T_1, T_2, \dots, T_n$, the value of such bond can be written as

$$\bar{P}(0,T) = R\sum_{i=1}^{n} P(0,T_i)(S(T_i) - S(T_{i-1})) + P(0,T)S(T), \tag{1.10}$$

where $T_0 = 0$. As we will shortly see, this is analogous to the value of the fee leg of a CDS in the JPMorgan model and it suggests that a CDS can be regarded as the exchange of two suitable risky bonds.

Remark 1.2. Note that, by definition, survival probabilities must be non-increasing. Indeed, if $T < T'$, the event $\{\tau > T'\}$ is contained in the event $\{\tau > T\}$, hence $S(T') \leq S(T)$. This implies that the hazard rate function must be non-negative (another analogy with risk-free short rates). If one assumes differentiability of S, this follows from (1.4). In the full generality of (1.3) this holds only almost surely.

One can equivalently prescribe a credit curve in terms of the hazard rate function. This approach can be also extended to include stochastic hazard rates, which can be thought of as the intensity of a Cox process.[6] In this case the link between hazard rates and survival probabilities is only in the mean:

$$S(t) = E\left[\exp\left(-\int_0^t h(u)du\right)\right], \tag{1.11}$$

where the expectation is taken with respect to the martingale measure corresponding to the chosen numéraire (usually the money market one, so that this is the risk-neutral measure). Such generality is needed when valuing CD Swaptions or

[5]The recovery rate, short rate and default time are assumed pairwise independent.

[6]The default process, or more generally, a point process, admits an intensity when the predictable compensator is absolutely continuous. Roughly, this boils down to the condition that $\ln S(t)$ can be written as a definite integral (with respect to the Lebesgue measure on the Borel sigma-algebra of $[0,\infty]$. When an intensity exists, it coincides with the hazard rate. For a precise statement of the conditions under which intensities exist and coincide with hazard rates, cf. [Sch03, Chap. 4]. Note also that our definition of hazard rate is unconditional. More generally, one can define it conditionally on the information available up to a certain time.

other derivatives where the dynamics (and in particular the volatility) of spreads plays a key role. However, it can be dispensed with when dealing with CDSs.

2. Bootstrapping a Hazard Rate Curve

Because the fundamental driver of single credit derivative valuation is the curve of survival probabilities $S(T_1), S(T_2), \ldots, S(T_n)$, one immediately realizes the importance of constructing such a curve. The task is formally identical to the construction of any other term structure, and the same or similar algorithms may be applied. In particular, two stages can be isolated:

- Estimate as many as possible $S(T_i)$'s from observed market data
- Determine the remaining $S(T_i)$'s by some form of interpolation and/or extrapolation that is consistent with the previously estimated ones.

The first stage requires a valuation model, one that inputs the $S(T_i)$'s and outputs the securities values, which is the de facto bridge between the theoretical curve (1.3) and an observed term structure of spreads or prices. Then one "inverts" the model to "imply" the $S(T_i)$'s from the observed market prices (or spreads, or other trading variables). Because typically there are many more $S(T_i)$'s than traded contracts, some kind of parametrization is needed. The most parsimonious and natural is arguably the assumption that the hazard rate is constant between maturities. Accordingly, we can overlay the following recursive structure onto the previously mentioned curve construction stages:

- Fist estimate the $S(T_i)$ for the earliest maturing security.
- Assuming the curve has been estimated up to a give maturity, extend it to the next maturity consistently with what has been built and the next available maturity's market datum.

This is what is broadly known as *term structure bootstrapping*.[7] We stress that this approach is generally superior to a the fitting of a parametric functional form of $S(T)$ (or $h(T)$) to observed market data, what is dubbed as *curve fitting*. The latter tends to impose an artificial shape (such as quadratic, Bzier, or cubic spline curves) to the curve, especially for unobserved data. Moreover, depending on the number of parameters, the calibration may well be only approximate, so that the valuation

[7] For a practical survey of bootstrapping and fitting algorithms we refer to [**?**]. The procedure is by no means inherently one-dimensional. Indeed, two- and three-dimensional term structure have crucial application in the modeling of the volatility implied from vanilla option and swaptions, respectively.

model would not match observed data. Increasing the number of parameters may afford exact calibration at the expense of overfitting. All of these drawbacks may create arbitrage opportunities.

In this section we first formalize the structure of the $S(T_i)$'s when hazard rate are piecewise constant and then the general equations in such constants to which the bootstrapping procedure reduces.

2.1. **Piecewise Constant Hazard Rates.** The assumption that the hazard rate function is piecewise constant boils down to assuming a partition of the time axis, $0 = T_0 \leq T_1 \leq \cdots \leq T_n$, such that

$$h(t) \equiv h_i \quad \text{for all } t \in (T_{i-1}, T_i], \tag{2.1}$$

for some fixed real constants $h_1, h_2, \ldots, h_n$. Under this assumption, the survival probability can be written as

$$S(t) = \exp\left(-\sum_{i=1}^{n(t)} h_i \Delta T_i + h_{n(t)+1}(t - T_{n(t)}) \right), \tag{2.2}$$

where $n(t) := \max\{i \leq n : T_i \leq t\}$ and $\Delta T_i := T_i - T_{i-1}$.

Remark 2.1. There may be a simple way to determine the h_i's recursively from the survival probability curve and average hazard rates. Indeed, let us define $\bar{h}_i$ as the constant hazard rate such that

$$\exp\left(-\sum_{j=1}^{i} h_j \Delta T_j \right) = S(T_i) = e^{-\bar{h}_i T_i}. \tag{2.3}$$

Assuming that we have already determined $h_1, \ldots h_{i-1}$ as well as $S(T_i)$, one can first calculate

$$\bar{h}_i = -\frac{1}{T_i} \ln S(T_i) \tag{2.4}$$

and then solve for h_i explicitly:

$$h_i = \frac{\bar{h}_i T_i - h_1 \Delta T_1 - \cdots - h_{i-1} \Delta T_{i-1}}{\Delta T_i}. \tag{2.5}$$

Alternately, we can derive the h_i directly from successive survival probabilities:

$$h_i = -\frac{1}{\Delta T_i} \ln\left(\frac{S(T_i)}{S(T_{i-1})} \right). \tag{2.6}$$

This approach in itself can be regarded as a form of bootstrapping. It may well appear as the final stage of a realistic bootstrapping process, where the initial stages determine the survival probabilities from market data via a specific model (cf. below).

2.2. **The Bootstrapping Equations.** We proceed to set forth the equations for the bootstrapping procedure that was informally presented at the beginning of the section. For simplicity and practicality, we will assume that the hazard rate function is piecewise constant and that the partition corresponds to the observed term structure of market data. In other words, the market is providing liquid data[8] $s_1, s_2, \ldots, s_m$ for securities with maturity $T_{n_1}, T_{n_2}, \ldots, T_{n_m}$, respectively, where $m \leq n$ and $\{n_k\}_{k=1}^{m}$ is a (usually proper) subsequence of the original indices $1, 2, \ldots, n$. Let $V_k(s_1, s_2, \ldots, s_k; h_1, \ldots, h_k)$ denote the model value[9] of the security maturing at T_{n_k}[10] and let $\hat{V}_k$ be the corresponding market price.[11] Then the bootstrapping process can be describe as the following recursive solution of equations in the h_k's. First solve for

$$\hat{V}_1 = V_1(s_1; h_1), \tag{2.7}$$

that is, express the model value of the earliest maturing security as a function of the first constant hazard rate and impose equality with the market price after inputting the market datum s_1. Then, assuming we have estimated $h_1, h_2, \ldots, h_{k-1}$ solve for h_k in the following equation

$$\hat{V}_k = V_i(s_1, \ldots, s_k; h_1, \ldots, h_k). \tag{2.8}$$

Note that in general these are not explicit equations, however, for most commonly used valuation models, they are implicitly solvable. Incidentally, one of the criteria that should be kept in mind in selecting a model is the ease and speed with which these equations can be solved, for this is the analogue of the calibration process for more exotic models.

Remark 2.2. We stress that the h_k as well as other possible parameters are not "parameters" in the sense that they define the model's credit curve in the usual parametric sense, that is, assuming a given functional form, such as quadratic, cubic spline, etc... Instead these come directly from the observed term structure without imposing additional restrictions. Earlier, we hinted at the advantages of

[8]If there is a bid-ask spread, one can take the mid-value as representative of these data.

[9]Although we speak of "model value," the s_i's and corresponding model outputs may well refer to quantities that are not prices. Such will be the case for CDS, in which the market quotes spreads, i.e., rates, rather than prices.

[10]We emphasize only the dependence of the V_k on the hazard rate and the market data s_k. They may well depend on other parameters as long as there are sufficient market data to estimate them through the bootstrapping process.

[11]In the case of CDSs, we will assume $\hat{V}_k \equiv 0$ for all k's, which corresponds to the assumption that the s_k's are the fair spreads of spot CDSs. Note that in a sense the dependence on the s_k's is redundant for it is equivalent to giving the $\hat{V}_k$'s. The market quotes the the s_k's with the understanding that these make the corresponding contract fair, i.e., worthless at inception.

this approach as opposed to the classical parametric one, the curve fitting. For further discussion in this context, cf., e.g., [Sch03, Sec. 3.5.2, p. 72-74]

3. Bootstrapping Hazard Rates from CDS Spreads via the JPMorgan Model

To concretely illustrate the bootstrapping method proposed above we consider the problem of estimating the hazard rate curve from CDS spreads. As a valuation model we select the so called the JPMorgan model, one of the earliest model to appear for the valuation of these securities and still a popular one among practitioners. As we will see the model leads to reasonably accurate explicit approximations.

Recall that a CDS gives the right and the obligation to be compensated for a loss given the default of a given reference security. The insurance "premium" is paid in the form of a spread, which is a per annum rate and must be multiplied by the notional (usually the face value of the reference security) to obtain the actual payments. We will denote with $PV_{\text{FLOAT}}(T)$ the present (arbitrage) value of the default leg, that is, the value of the payment that a buyer of default protection up to time T, the maturity of the CDS, would receive in default. $PV_{\text{FIX}}(T)$, on the other hand, denotes the value of the payment, or fee leg, that is the present value of the cash stream that must be paid in exchange for default protection until T.[12] By contractual stipulation, at inception a CDS must be worthless, that is,

$$PV_{\text{FIX}}(T) = PV_{\text{FLOAT}}(T). \tag{3.1}$$

Let $s = s(T)$ denote the fair spread at inception for such CDS. Since s can be factored out of the payment leg as $PV_{\text{FIX}}(T) = sPV_{\text{FIX},0}(T)$, where $PV_{\text{FIX},0}(T)$ is the value of protection per rate unit, one has the fundamental relation

$$s = \frac{PV_{\text{FLOAT}}(T)}{PV_{\text{FIX},0}(T)}. \tag{3.2}$$

3.1. The JPMorgan Model.

The so called JPMorgan model assumes that the (risk-less) interest rate process is independent of the default process and the default leg pays at the end of each accrual period, so that the present value of this leg can

[12] The choice of the "fix for floating" terminology comes from the analogy with interest rate swaps.

be written as

$$
\begin{aligned}
PV_{\text{FLOAT}}(T) &= (1-R)\sum_{i=1}^{n} P(0,T_i)\,\text{Prob}\,(T_{i-1} < \tau \le T_i) = \\
&= (1-R)\sum_{i=1}^{n} P(0,T_i)\,(S(T_{i-1}) - S(T_i))\,.
\end{aligned}
\tag{3.3}
$$

Concerning the fee leg, regular fee payments occur at the end of each period. However, when default occurs, one last fee payment corresponding to the accrued fee for that period is contractually required. One way to model the value of such accrual payment is by making assumption on the occurrence of default as well as their actual payment. In this regard, the JPMorgan model assumes that defaults occur midway during each payment period, but the accrual payment is made at the end of the periods. These assumptions yield the following value for the fee leg:

$$
\begin{aligned}
&PV_{\text{FIX}}(T) = \\
&= s\sum_{i=1}^{n}\alpha_i P(0,T_i)\,\text{Prob}\,(T_i < \tau) + \frac{s}{2}\sum_{i=1}^{n}\alpha_i P(0,T_i)\,\text{Prob}\,(T_{i-1} < \tau \le T_i) = \\
&= s\sum_{i=1}^{n}\alpha_i P(0,T_i)S(T_i) + \frac{s}{2}\sum_{i=1}^{n}\alpha_i P(0,T_i)\,(S(T_{i-1}) - S(T_i)) = \\
&= s\sum_{i=1}^{n}\alpha_i P(0,T_i)\frac{S(T_{i-1}) + S(T_i)}{2},
\end{aligned}
\tag{3.4}
$$

where the α_i are year fractions corresponding to the period $[T_{i-1}, T_i]$.

3.2. **Bootstrapping in JPMorgan Model.** Given the above expression (3.3) and (3.4) for the value of the default and fee leg, respectively, in the JPMorgan model, one writes the value of the CDS with maturity $T = T_n$ as (from the viewpoint of the buyer of protection):

$$
C(T) = PV_{\text{FLOAT}}(T) - sPV_{\text{FIX},0}(T) \tag{3.5}
$$

Now, let us assume that we have liquid spreads $s_1, s_2, \ldots, s_m$ for the m maturities $T_{n_1}, T_{n_2}, \ldots, T_{n_m}$, respectively, as well as a complete liquid curve of risk-free ZCB prices. We assume that the hazard rate is piecewise constant on the intervals that correspond to the maturities of the liquid CDS contracts, namely

$$h(t) \equiv h_k \quad \text{for } t \in (T_{n_{k-1}}, T_{n_k}] \tag{3.6}$$

and $k = 1, 2, \ldots, m$. Then, we start by solving for h_1 using the value of the first contract spread, which, being fair, makes the contract itself worthless:

$$\begin{aligned}
0 = C(T_{n_1}) = \\
&= (1-R)\sum_{i=1}^{n_1} P(0,T_i)\left(S(T_{i-1}) - S(T_i)\right) - \\
&- s_1 \sum_{i=1}^{n_1} \alpha_i P(0,T_i) \frac{S(T_{i-1}) + S(T_i)}{2} = \\
&= (1-R)\sum_{i=1}^{n_1} P(0,T_i) e^{-h_1 T_{i-1}} \left(1 - e^{-h_1 \Delta T_i}\right) - \\
&- s_1 \sum_{i=1}^{n_1} \alpha_i P(0,T_i) e^{-h_1 T_{i-1}} \frac{1 + e^{-h_1 \Delta T_i}}{2} = \\
&= \sum_{i=1}^{n_1} P(0,T_i) e^{-h_1 T_{i-1}} \left[1 - R - \frac{s_1 \alpha_i}{2} - e^{-h_1 \Delta T_i}(1 - R + \frac{s_1 \alpha_i}{2})\right],
\end{aligned} \tag{3.7}$$

where $\Delta T_i = T_i - T_{i-1}$. Because all parameters except the hazard rate are known, this is an implicit equation in h_1, one that can be easily solved using mainstream numerical solvers. The assumption that the hazard rate was constant over $(T_0, T_{n_1}]$ was crucial. In particular, we would not have been able to solve for the survival probabilities $S(T_i)$ (for $1 \leq i \leq T_{n_1}$) simply because there are, in general,[13] many more such unknowns than (one) equation.

[13] The trivial case $n_1 = 1$, has an explicit solution, namely

$$h_1 = -\frac{1}{T_1} \ln\left(\frac{1 - R - s_1\alpha_1/2}{1 - R + s_1\alpha_1/2}\right).$$

Assuming that the maturity of the first liquid contract coincides with the first payment date is, however, not realistic for, in general, payments occur at a much higher frequency (quarterly, or semi-annually) than maturities (measured in years). On the other hand, as we will see shortly, the formula just given can be assumed as a good approximation of the "average" hazard rate for the entire life of the contract.

Remark 3.1. Note that Equation (3.7) must admit a solution as the right-hand side is continuous function in h_1 attaining both negative and positive values. This is more clearly seen by rewriting the right-hand side as

$$\phi(h_1) = \sum_{i=1}^{n_1} P(0,T_i) e^{-h_1 T_{i-1}} \left(1 - R + \frac{s_1 \alpha_i}{2}\right) \left[A_i - e^{-h_1 \Delta T_i}\right]. \tag{3.8}$$

where we have put

$$A_i := \frac{1 - R - \frac{s_1 \alpha_i}{2}}{1 - R + \frac{s_1 \alpha_i}{2}}.$$

It is safe to assume that

$$0 < A_i < 1,$$

for all $i \leq n_1$. For, a negative value of the numerator would imply

$$s_1 \geq 2 \frac{1-R}{\alpha_i}.$$

Since, typically, $1 - R \geq 0.2$ and $\alpha_i \approx 1/4$, this would imply spreads of greater than 160% or 1.6×10^6 bps. The function ϕ can be regarded as sum of terms $A_i - e^{-h_1 \Delta T_i}$ with positive coefficients $P(0,T_i) e^{-h_1 T_{i-1}} \left(1 - R + \frac{s_1 \alpha_i}{2}\right)$. Next, notice that

$$\lim_{h_1 \to 0^+} \phi(h_1) = \sum_{i=1}^{n_1} P(0,T_i) \left(1 - R + \frac{s_1 \alpha_i}{2}\right) [A_i - 1] < 0.$$

Because the coefficients are positive, ϕ will attain positive values if for some h_1 all $A_i - e^{-h_1 \Delta T_i} > 0$. But this must occur as soon as[14]

$$h_1 \geq \max_{1 \leq i \leq n_1} \frac{1}{\Delta T_i} \ln A_i$$

We have thereby shown that ϕ attains negative as well as positive values, which along with its continuity implies it must vanish for some h_1.[15]

Let us now formulate the inductive step of the bootstrapping procedure. Assuming that we have already estimated $h_1, h_2, \ldots, h_{k-1}$, as well as that the CDS spreads

[14] The fact that $\lim_{h_1 \to \infty} \phi(h_1) = 0$ is irrelevant. The function asymptotically vanished from positive values.

[15] Notice that the explicit solution to the equation $\phi(h_1) = 0$ which was noticed earlier in the case of $n_1 = 1$ holds also when the accrual factors α_i and the calendar time durations ΔT_i are constant. In our new notation this can be written as

$$h_1 = \frac{1}{\Delta T} \ln A.$$

where $\Delta T_i = \Delta T$ and $A_i = A$ for all i. This formula will reappear later as a useful approximation for "average" hazard rates.

s_k for the CDS maturing at T_{n_k} is available, the worthlessness of the corresponding contract translates into the following equation in h_k:

$$\begin{aligned} 0 = C(T_{n_k}) = \\ &= (1-R)\sum_{i=1}^{n_k} P(0,T_i)\left(S(T_{i-1}) - S(T_i)\right) - \\ &- s_k \sum_{i=1}^{n_k} \alpha_i P(0,T_i) \frac{S(T_{i-1}) + S(T_i)}{2}, \end{aligned} \tag{3.9}$$

where the dependence on h_k appears through the unfolding of the survival probabilities as

$$S(T_i) = \exp\left(- \sum_{\substack{1 \le j \le i \\ T_{n_{a-1}} < T_j \le T_{n_a}}} h_a T_j \right). \tag{3.10}$$

To make this dependence more explicit as well as emphasize that the h_a's are assumed known for $a < k$, we can write (3.9) in this equivalent form:

$$0 = C(T_{n_k}) =$$

$$= (1-R) \sum_{i=1}^{n_{k-1}} P(0,T_i) \left(S(T_{i-1}) - S(T_i) \right) -$$

$$- s_k \sum_{i=1}^{n_{k-1}} \alpha_i P(0,T_i) \frac{S(T_{i-1}) + S(T_i)}{2} +$$

$$+ (1-R) \sum_{i=n_{k-1}+1}^{n_k} P(0,T_i) S(T_{n_{k-1}}) e^{-h_k T_{n_{k-1}}} \left(e^{-h_k T_{i-1}} - e^{-h_k T_i} \right) -$$

$$- s_k \sum_{i=n_{k-1}+1}^{n_k} \alpha_i P(0,T_i) S(T_{n_{k-1}}) e^{-h_k T_{n_{k-1}}} \frac{e^{-h_k T_{i-1}} + e^{-h_k T_i}}{2} =$$

$$= \sum_{i=1}^{n_{k-1}} P(0,T_i) \left(1 - R + \frac{s_k \alpha_i}{2} \right) \left(A_{k,i} S(T_{i-1}) - S(T_i) \right) +$$

$$+ \sum_{i=n_{k-1}+1}^{n_k} P(0,T_i) \left(1 - R + \frac{s_k \alpha_i}{2} \right) S(T_{n_{k-1}}) e^{-h_k T_{n_{k-1}}} \left(A_{k,i} e^{-h_k T_{i-1}} - e^{-h_k T_i} \right),$$

where

$$A_{k,i} := \frac{1 - R - \frac{s_k \alpha_i}{2}}{1 - R + \frac{s_k \alpha_i}{2}}.$$

To emphasize that the first sum is known, we can write the above equation in non-homogeneous form as follows:

$$\sum_{i=n_{k-1}+1}^{n_k} P(0,T_i) \left(1 - R + \frac{s_k \alpha_i}{2} \right) e^{-h_k (T_{i-1} - T_{n_{k-1}})} \left(A_{k,i} - e^{-h_k \Delta T_i} \right) =$$

$$= \frac{1}{S(T_{n_{k-1}})} \sum_{i=1}^{n_{k-1}} P(0,T_i) \left(R - 1 - \frac{s_k \alpha_i}{2} \right) \left(A_{k,i} S(T_{i-1}) - S(T_i) \right). \tag{3.11}$$

Further, if one assumes that the accrual factors are identical, $\alpha_i \equiv \alpha$ for $i = 1, 2, \ldots n_k$, so that one can set $A_{k,i} = A_k$, this last equation can be nicely simplified as follows:

$$\sum_{i=n_{k-1}+1}^{n_k} P(0,T_i)e^{-h_k(T_{i-1}-T_{n_{k-1}})}\left(e^{-h_k\Delta T_i}-A_k\right)=$$
$$=\frac{1}{S(T_{n_{k-1}})}\sum_{i=1}^{n_{k-1}} P(0,T_i)\left(A_kS(T_{i-1})-S(T_i)\right). \tag{3.12}$$

4. Average Bootstrapping and a Useful Approximation

We give an alternative bootstrapping procedure that solves for hazard rates that are constant throughout the life of the contract. This is essentially the first step of the previous approach applied to all maturities. As a byproduct we derive a useful approximation that is apparently widely used among practitioners.

4.1. **Bootstrapping through average hazard rates.** We can pursue the approach outlined in (2.4) and (2.5), that is, let $\bar{h}_k$ be the "average hazard rate" that matches the survival probabilities implied by s_k:

$$S(T_i)=e^{-\bar{h}_kT_i}\quad\text{for } 1\le i\le n_k. \tag{4.1}$$

The first step of the bootstrapping process is identical to the one presented above (leading to Equation (3.7)) and $\bar{h}_1=h_1$. However, the inductive step yields a considerably simpler equation.

$$
\begin{aligned}
0 = C(T_{n_k}) = \\
&= (1-R)\sum_{i=1}^{n_k} P(0,T_i)\left(S(T_{i-1}) - S(T_i)\right) - \\
&- s_k \sum_{i=1}^{n_k} \alpha_i P(0,T_i)\frac{S(T_{i-1}) + S(T_i)}{2} = \\
&= (1-R)\sum_{i=1}^{n_k} P(0,T_i)e^{-\bar{h}_k T_{i-1}}\left(1 - e^{-\bar{h}_k \Delta T_i}\right) - \\
&- s_k \sum_{i=1}^{n_k} \alpha_i P(0,T_i)e^{-\bar{h}_k T_{i-1}}\frac{1 + e^{-\bar{h}_k \Delta T_i}}{2} = \\
&= \sum_{i=1}^{n_k} P(0,T_i)\left(1 - R + \frac{s_k \alpha_i}{2}\right)e^{-\bar{h}_k T_{i-1}}\left(A_{k,i} - e^{-\bar{h}_k \Delta T_i}\right).
\end{aligned}
\tag{4.2}
$$

As remarked above for the A_i's one can safely assume that $0 < A_{k,i} < 1$ so that the equation is solvable.

4.2. **A Useful Approximation.** If in (4.2) we further assume that all the accrual factors are identical, say $\alpha_i \equiv \alpha$ for $i = 1, 2, \ldots n_k$, so that we can set $A_{k,i} = A_k$, then the equation simplifies to

$$
\sum_{i=1}^{n_k} P(0,T_i)e^{-\bar{h}_k T_{i-1}}\left(A_k - e^{-\bar{h}_k \Delta T_i}\right) = 0. \tag{4.3}
$$

This equation in $\bar{h}_k$ is not solvable unless we further assume that the calendar times between payments are identical, that is, $\Delta T_i = \Delta T$ for $i = 1, 2, \ldots, n_k$. Note that this assumption is quite compatible, and nearly[16] implied by the identity of accrual factors. Because the coefficients $P(0,T_i)e^{-\bar{h}_k T_{i-1}}$ in the sum are positive the solution must be

$$
e^{-\bar{h}_k \Delta T} = A_k, \tag{4.4}
$$

[16] Up to holidays, business days, and other conventions.

or, equivalently,

$$\bar{h}_k = -\frac{1}{\Delta T}\ln A_k = -\frac{1}{\Delta T}\ln\left(\frac{1-R-\frac{s_k\alpha}{2}}{1-R+\frac{s_k\alpha}{2}}\right). \tag{4.5}$$

Furthermore, noticing that

$$\ln\left(\frac{1-R-\frac{s_k\alpha}{2}}{1-R+\frac{s_k\alpha}{2}}\right) = \ln\left(1-\frac{s_k\alpha}{1-R+\frac{s_k\alpha}{2}}\right) \approx -\frac{s_k\alpha}{1-R+\frac{s_k\alpha}{2}}$$

we recover an approximation common amongst practitioners:[17]

$$\bar{h}_k \approx \frac{\alpha}{\Delta T}\frac{s_k}{1-R+\frac{s_k\alpha}{2}}. \tag{4.6}$$

Note that and that $\alpha/\Delta T$ can be thought of as a day count basis adjustment factor. For example, if the day count basis is act/360, then one can assume $\alpha/\Delta T = 365/360$.

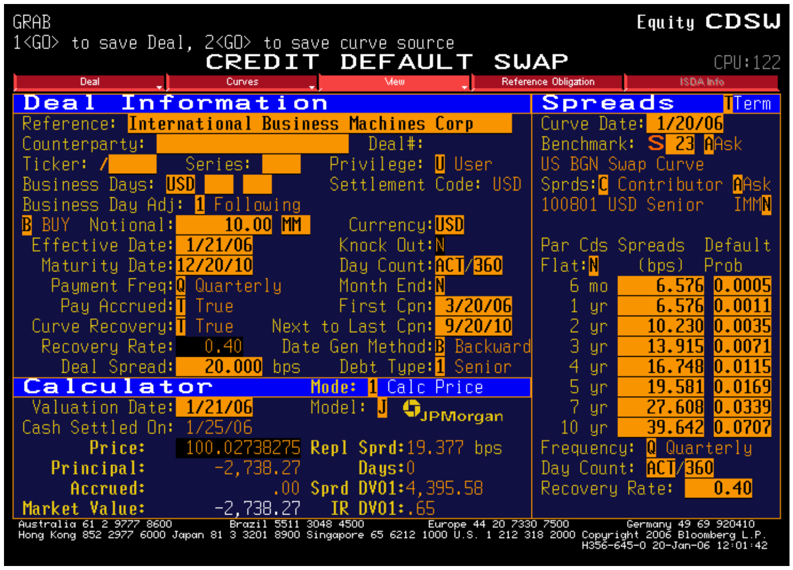
5. Testing with Actual Data

We illustrate the application of the above approximation (4.5) using spread data from Bloomberg (BLP) (cf. Figure 1). In Table 1 we compare the default probabilities estimated via (4.5) with Bloomberg's. The error is relatively small for near maturities except for the first one. We suspect that Bloomberg obtained the six month maturity spread by extrapolating the one year spread (indeed the spreads are identical). Moreover, there may be some numerical rounding in the default probabilities, which makes the error particularly sensitive especially for the six month maturity. Indeed, if we round our model's estimate for this maturity down (or truncate) to the fourth decimal place we obtain zero error. Similar rounding to the fourth decimal place for other maturities generally gives smaller relative errors and in particular no error for $T = 1, 2$. However, the error increase with maturity, which is to be expected as the assumption of a constant hazard rate over the entire life of the contract becomes less and less tenable.

Next, consider an example of a structurally different type of debt, namely, sovereign credit instruments. In Figure 2 we report the credit spreads for the Russian Government International Bond for April 18, 2006. This corresponds to the Russian Federation debt denominated in US Dollars. Besides the slightly different year

[17] We allude to the stylized equation that relates the "vertices" of the credit triangle, namely, spreads, default probabilities, and recovery rates. Such equation reads (cf., e.g., [?])

$$h = \frac{s}{1-R}.$$



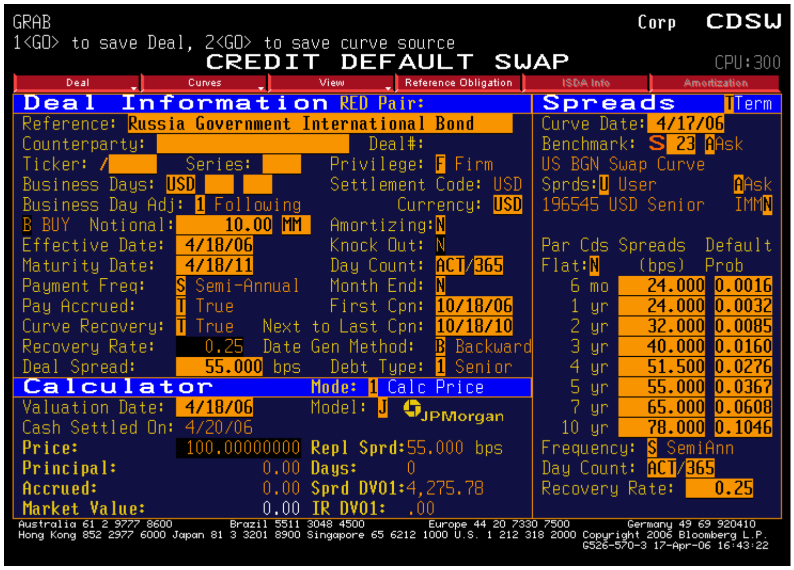
(a) IBM CDS Credit Curve

Figure 1. IBM's CDS Credit Curve (source: Bloomberg LLP)

Years	**BLP's** s_k	**BLP's** $1-S(T_{n_k})$	$\bar{h}_k$	$S(T_{n_k})$	$1-S(T_{n_k})$	**% Error**
0.5	0.06576%	0.0005	0.001111222	0.999444543	0.000555457	9.98 %
1	0.06576%	0.0011	0.001111222	0.998889395	0.001110605	0.95 %
2	0.10230%	0.0035	0.001728681	0.996548609	0.003451391	-1.41 %
3	0.13915%	0.0071	0.002351377	0.99297069	0.00702931	-1.01 %
4	0.16748%	0.0115	0.002830102	0.988743427	0.011256573	-2.16 %
5	0.19581%	0.0169	0.003308827	0.98359197	0.01640803	-3.00 %
7	0.27608%	0.0339	0.004665241	0.967870783	0.032129217	-5.51 %
10	0.39642%	0.0707	0.006698765	0.935206747	0.064793253	-9.12 %

Table 1. Comparison of Bloomberg and Model Approximate Default Probabilities from IBM CDS spreads

fractions of ACT/365 (so that $\alpha/\Delta T = 1$), we notice the recovery rate of $R = 0.25$, significantly lower that the typical recovery rate ($R = 0.4$) for North American corporate debt.



(A) RFSF CDS Credit Curve

FIGURE 2. Russian Federation CDS Credit Curve (source: Bloomberg LLP)

Our cumulative hazard rate approximation (4.5) gives the default probabilities tabulated in Table 2. Overall the approximation is better than that for the corporate credit curve we considered above (cf. Table 1), with a similarly decreasing accuracy as a function of maturity. Furthermore, we notice that the approximation formula (4.5) consistently underestimates the default probabilities (hence, overestimates the survival probabilities).

5.1. **Some Ad Hoc Improvement.** In the latter example, we can almost perfectly match Bloomberg's estimates if instead of the formally more rigorous approximation for default probabilities deriving from (4.5), we use the further approximation[18]

$$1 - S(T_{n_k}) = 1 - e^{-\bar{h}_k T} \approx \bar{h}_k T. \tag{5.1}$$

Further, we round the results to the fourth decimal place. We report the results in Table 3.

In the former example, a slightly more complex ad hoc treatment, where we take into account the nontrivial year fraction ACT/360 ($\frac{\alpha}{\Delta T} \approx 1.013888889$) and then truncate to the fourth decimal place also dramatically improve our matching Bloomberg's estimates. We tabulate default probabilities obtained by

$$1 - S(T_{n_k}) = 1 - e^{-\bar{h}_k T} \approx \mathrm{Trunc}_4\left(\bar{h}_k T \frac{\alpha}{\Delta T}\right), \tag{5.2}$$

where Trunc_4 denotes truncation to the fourth decimal place, report the results in Table 4.

5.2. **A Realistic Application to Bootstrapping.** We conclude with a realistic application of the heretofore illustrated techniques. To wit, in practice one does

Years	BLP's Spreads	BLP's Default Prob.s	Average Hazard Rates	Survival Prob.s	Default Prob.s	% Error
0.5	0.24000%	0.0016	0.003200000	0.998401279	0.001598721	-0.08 %
1	0.24000%	0.0032	0.003200000	0.996805114	0.003194886	-0.16 %
2	0.32000%	0.0085	0.004266667	0.991502971	0.008497029	-0.03 %
3	0.40000%	0.016	0.005333334	0.984127318	0.015872682	-0.80 %
4	0.51000%	0.0276	0.006800002	0.973166582	0.026833418	-2.86 %
5	0.55000%	0.0367	0.007333335	0.963997404	0.036002596	-1.94 %
7	0.65000%	0.0608	0.008666670	0.941136877	0.058863123	-3.29 %
10	0.78000%	0.1046	0.010400006	0.901225245	0.098774755	-5.90 %

Table 2. Comparison of Bloomberg and Model Approximate Default Probabilities from Russian Sovereign CDS spreads

Years	BLP's Spreads	BLP's Default Prob.s	Average Hazard Rates	Ad Hoc Def. Prob.s	% Error
0.5	0.24000%	0.0016	0.003200000	0.0016	0.00 %
1	0.24000%	0.0032	0.003200000	0.0032	0.00 %
2	0.32000%	0.0085	0.004266667	0.0085	0.00 %
3	0.40000%	0.016	0.005333334	0.0160	0.00 %
4	0.51000%	0.0276	0.006800002	0.0272	-1.47 %
5	0.55000%	0.0367	0.007333335	0.0367	0.00 %
7	0.65000%	0.0608	0.008666670	0.0607	-0.16 %
10	0.78000%	0.1046	0.010400006	0.1040	-0.58 %

Table 3. An ad hoc approximation compared to Bloomberg's Default Probabilities from Russian Sovereign CDS spreads

hazard rate $\bar{h}_3 = 0.002351377$. Then, assuming we did not observe a spread for the two-year maturity, the hazard rate for the second year (i.e., for the interval $[1, 2)$) in the life of the credit can be determined consistently with this credit curve as

Years	BLP's Spreads	BLP's Default Prob.s	Average Hazard Rates	Ad Hoc Def. Prob.s	% Error
0.5	0.06576%	0.0005	0.001111222	0.0005	0.00 %
1	0.06576%	0.0011	0.001111222	0.0011	0.00 %
2	0.10230%	0.0035	0.001728681	0.0035	0.00 %
3	0.13915%	0.0071	0.002351377	0.0071	0.00 %
4	0.16748%	0.0115	0.002830102	0.0114	-0.88 %
5	0.19581%	0.0169	0.003308827	0.0167	-1.20 %
7	0.27608%	0.0339	0.004665241	0.0331	-2.42 %
10	0.39642%	0.0707	0.006698765	0.0679	-4.12 %

TABLE 4. Ad Hoc Improved Approximation for Default Probabilities from IBM CDS spreads

illustrated in (2.5). In this case, we obtain the equation

$$h_3 = \frac{\bar{h}_3 T_3 - h_1 \Delta T_1 - h_2 \Delta T_2}{\Delta T_3} = 0.002351377 \times 3 - 0.001111222 - h_2, \qquad (5.3)$$

where we have used the fact that $T_3 = 3$ and $\Delta T_i = 1$ for this particular interval. Further, assuming as usual that the hazard rates are constant between liquid maturities so that $h_2 = h_3$, the equation becomes

$$h_2 = h_3 = \frac{0.002351377 \times 3 - 0.001111222}{2} \approx 0.002971455. \qquad (5.4)$$

Notice that this is significantly higher than the cumulative hazard rate $\bar{h}_3$. This is the same behavior as forward rates for a term structure in contango. As it should be

$$0.99297069 = S(T_3) = S(T_1)e^{-h_3(T_3-T_1)} = 0.998889 \times 0.994075. \qquad (5.5)$$

5.3. **Construction of a Full Hazard Rate Curve.** We conclude by constructing a full hazard rate curve from the survival probabilities (or, equivalently, the cumulative hazard rates) we have estimated in the two case studies presented so far using (2.6) (or (2.5), respectively). In Table 5 we present the full hazard rate curve inferred from Bloomberg's spread data. Notice that the hazard rates corresponding to the periods $[0, 0.5)$ and $[0.5, 1)$ are identical because so the corresponding spreads are. In Table 6 we collect the full hazard rate curve for the sec-

Years	BLP's Spreads	Survival Prob.s	Hazard Rates
0.5	0.06576%	0.999444543	0.0011
1	0.06576%	0.998889395	0.0011
2	0.10230%	0.996548609	0.0023
3	0.13915%	0.99297069	0.0036

are monotonic increasing), hazard rates are not necessarily monotonic as maturity increases—$h_4 = 0.0075 < 0.0112 = h_5$ and $h_5 > 0.0095 = h_6$.

Years	**BLP's Spreads**	**Survival Prob.s**	**Hazard Rates**
0.5	0.24000%	0.998401279	0.0032
1	0.24000%	0.996805114	0.0032
2	0.32000%	0.991502971	0.0053
3	0.40000%	0.984127318	0.0075
4	0.51000%	0.973166582	0.0112
5	0.55000%	0.963997404	0.0095
7	0.65000%	0.941136877	0.0120
10	0.78000%	0.901225245	0.0144

Table 6. Full Hazard Rate Curve for Russian Sovereign Debt

References

[JR00] Monique Jeanblanc and Marek Rutkowski. Default Risk and Hazard Process. In *Mathematical Finance Bachelier Congress 2000*, pages 281–312. Springer, 2000.

[MR05] Marek Musiela and Marek Rutkowski. *Martingale Methods in Financial Modelling.* Number 36 in Stochastic Modelling and Applied Probability. Springer-Verlag, 2005.

[Pro90] Philip E. Protter. *Stochastic integration and differential equations : a new approach.* Springer Verlag, 1990.

[Sch03] Philip J. Schönbucher. *Credit Derivatives Pricing Models. Models, Pricing and Implementation.* Wiley Finance Series. John Wiley & Sons, 2003.

E-mail address: castel@alum.mit.edu