Consistent Yield Curves via Static Hedging

Yury Rojek
Deutsche Bank[1]
Paul Wilmott
Willmott Associates, paul@wilmott.com

Abstract
Interpolating yield curves between observable points can be done in various ways. In this paper, we show that market making in tradeable instruments with static hedging can be considered as a consistent basis for such interpolation. This approach leads to solutions where parameters match the historical rates while currently quoted prices are also perfectly reconstructed. We also show that the interpolation solution derived from the minimum variance optimization for the static hedge ratios is equivalent to statistical inference with Gaussian Processes Regression (GPR). We offer closed-form solutions for affine interest-rate models for market making as well as for profit-seeking scenarios.

Keywords
yield curve, calibration, interest rates, Gaussian processes

1 Introduction

The task of inferring the non-observable continuous parts of a yield curve from a limited number of point observations, i.e. a finite set of tradeable securities, has attracted considerable interest due to high practical demand. Complete continuous yield curves are needed for valuing almost all derivative securities as well as some primary instruments.

Considering a single curve, it seems desirable to base value calculations on prices that can actually be obtained on the market. There are, however, situations when we can only use a proxy for the price of a target instrument because it is not traded. It is tempting, in such scenarios, to impute missing values using a model that has desirable properties or calibrate yield curves to some implied characteristics of quoted instruments or, simply, interpolate between closest observable points.

A compendium of interpolation methods was presented by Hagan and West [6] with a summary of desirable properties from a practical point of view. The main proposed criteria for an interpolation method were numerical stability, preservation of continuity (for yield curves and bond prices, not forward rates), positivity, monotonicity, locality of changes, and locality of hedges. These properties are as much about mathematical niceness as about financial reality. An apparent winner turned out to be a modification of a cubic spline-based interpolation scheme built on an idea borrowed from non-financial engineering work. While properties such as curve smoothness, absence of spurious oscillations, and numerical stability can potentially be explicitly translated into plausibility in a financial sense, the paper, however, presents formulas that are independent of the underlying financial model. Nevertheless, empirical tests very often favor numerically stable and smooth solutions.

In the present paper, we are trying to retain the most important (but not all) numerical properties advocated by [6] and at the same time take interest-rate evolution dynamics into account. We fully subscribe to the view that tradeable prices used as inputs should be reconstructed precisely on a smooth curve without spurious oscillations. But these and other desirable features can emerge naturally (where possible) as statistical properties of a statically hedged portfolio within a chosen stochastic process framework.

We argue that it is possible to base interpolation of non-observable curve parts on realistic hedging considerations and, as a result, compute values that can be realistically achieved via trading available instruments.

Lastly, when pricing instruments, not only should we take the bid–ask spread of tradeable instruments into account but aim to produce uncertainty bounds that can be translated into a spread given one's own risk preferences. This resonates with earlier work of Epstein and Wilmott on interest modeling, where the "yield envelope" obtained via static hedging is suggested (see [4], [5]). In the current paper we present closed-form formulae of a similar yield envelope as well as the mean yield curve for affine interest-rate models. We also show that obtaining the "yield envelope" in our current settings is based on calculations identical to those used in Gaussian Processes Regression when analytical functional forms of covariances and means of traded bonds are available.

2 Curve interpolation via static hedging

The main idea is fairly simple to explain. Suppose, we are asked (by a client) to quote a zero-coupon bond with non-standard maturity, say, 4 months. In order to do that, we consult the list of available bonds and discover that only bonds with maturities 3, 6, 9, and 12 months are traded. We cannot predict interest-rate behavior or find a perfect hedge, but it should be possible, in principle, to find an optimal static hedge as a linear combination of traded instruments. If our client agrees to buy a 4-month bond from us, we enter into our static hedge position and then keep all bonds until maturity. At the end of a 12-month period we will close our bond positions completely and calculate PnL of the whole transaction. There may be several conflicting notions of static hedge "optimality" for such a scenario, and we will consider some of them later on. One optimality criterion can be minimal variance of the transaction outcome. If that is our purpose then we can find the expected bond yield and its standard deviation, and for all affine interest-rate models there are closed-form

[1] The opinions and ideas expressed in this article are those of the author alone, and do not necessarily reflect the views of Deutsche Bank AG, its subsidiaries or affiliates.

solutions for these. Other optimization tasks, especially those involving inequality constraints, do not necessarily lead to closed-form formulas but can be solved numerically with a quadratic programming solver or with a nonlinear convex optimization software. Our method can be easily generalized to other types of bonds provided that all cashflows are defined ahead of trading.

2.1 The Partial Differential Equation for the Mean

For the purposes of our analysis, we will be dealing mostly with zero-coupon bonds. We denote, in a rather loose but obvious notation, by $V(t,T)$ $(0 \le t \le T)$ the present value of the bond at time t. Then, for simplicity, we also specify $V(T,T) = 1$, i.e. all bonds pay 1 at maturity. In general, $V(t,T)$ is a random variable since the bond present value depends on instantaneous spot rate at all times between t and T. We assume that the real spot rate (*not risk-neutral*) r evolves according to the stochastic differential equation

$$dr = \alpha(r,t)dt + \beta(r,t)dX(t), \tag{1}$$

where $dX(t)$ denotes the Wiener process term. Note, that since present value V will depend on initial state $r(t)$ of the dynamic system (1) we should make it explicit from now on via notation $V(r(t), t, T)$.

We are going to be working with real random walks, and the present values of the real moments of random variables. We use "real" to distinguish from "risk neutral." Indeed, in this paper there is no mention of risk neutrality, other than to say we aren't mentioning it. For example, the real first moment, the expectation, of the real present value at time t of one dollar received at time T is

$$m_1(x,t,T) = E[V(r(t),t,T)|r(t)=x] = E\left[e^{-\int_t^T r(\tau)d\tau}\middle| r(t)=x\right].$$

Note that the expectation is conditioned on current spot rate being equal to x. Consider for some twice differentiable function $\phi(x,s)$ a quantity

$$z(x,s) = \phi(x,s)\exp\left\{-\int_t^s r(\tau)d\tau\right\}.$$

Using multi-dimensional Ito's lemma we can write

$$dz(x,s) = \phi d\left(e^{-\int_t^s r(\tau)d\tau}\right) + e^{-\int_t^s r(\tau)d\tau}d\phi + d\phi d\left(e^{-\int_t^s r(\tau)d\tau}\right).$$

The third term is zero. The remaining ones are

$$dz(x,s) = e^{-\int_t^s r(\tau)d\tau}\left(-r(s)\phi ds + \phi_s ds + \phi_x dr + \frac{(\beta(r(s),s))^2}{2}\phi_{xx}ds\right)$$

$$= e^{-\int_t^s r(\tau)d\tau}\left[\left(\phi_s + \alpha\phi_x + \frac{\beta^2}{2}\phi_{xx} - r(s)\phi\right)ds + \beta dX(s)\right],$$

where $\alpha = \alpha(r(s),s)$, $\beta = \beta(r(s),s)$, $\phi = \phi(r(x),s)$ for some $x > 0$, and similarly for the partial derivatives of ϕ.

Now we define the function $\phi(x,s)$ to be a solution of PDE

$$\phi_s + \alpha(x,s)\phi_x + \frac{(\beta(x,s))^2}{2}\phi_{xx} - x\phi = 0$$

with final condition $\phi(x,T) = 1$. Then

$$E[z(x,T)] = z(x,t) + E\left[\int_t^T dz(x,s)\right] = z(x,t)$$

$$+ E\left[\int_t^T\left(\phi_s + \alpha\phi_x + \frac{\beta^2}{2}\phi_{xx} - x\phi\right)e^{-\int_t^s r(\tau)d\tau}ds\right] + E\left[\int_t^T \beta(x,s)e^{-\int_t^s r(\tau)d\tau}dX(s)\right].$$

The last integral is zero according to Fubini's theorem, subject to technical conditions on $\beta(x,s)$ (see [13]), while the former is zero by construction of $\phi(x,s)$. It is easy to see also that $m_1(x,t,T) = E[z(x,T)]$ using definition of $z(x,s)$ and the final condition on $\phi(x,s)$. Then it is true that $m_1(x,t,T) = \phi(x,t)$ which concludes the derivation of the PDE for mean of V.

To state it explicitly, we are looking for a solution $m_1(r,t)$ (with slight abuse of notation, in order to avoid clutter) of

$$\frac{\partial m_1}{\partial t} + \alpha\frac{\partial m_1}{\partial r} + \frac{\beta^2}{2}\frac{\partial^2 m_1}{\partial r^2} - rm_1 = 0$$

subject to final condition

$$m_1(r,T) = 1,$$

where r is the current spot rate. From now on we are going to use simple r and x interchangeably for the currently observed spot rate value and $r(t)$ for the stochastically evolving spot rate.

2.2 General mixed moments PDE

It will be convenient to derive PDEs for a more general case of the mixed moments of two bonds' discounted values

$$m_{ij}(x,t,T_1,T_2) = E\left[V(x,t,T_1)^i V(x,t,T_2)^j \middle| r(t)=x\right]$$

for some nonnegative integers i and j such that $i + j > 0$. Unless explicitly mentioned, we always assume that $T_1 \le T_2$ if both i, j are positive. To avoid clutter, we will also use single subscript notation $m_i(x,t,T)$:

$$m_i(x,t,T) = m_{i,0}(x,t,T,\cdot) = m_{0,i}(x,t,\cdot,T).$$

We have already introduced this notation for $m_1(x,t,T)$ implicitly in the previous section.

It is easy to see that all the quantities we are interested in can be expressed via mixed moments of two bonds. Namely, mean, variance and co-variance are the examples:

$$E[V(r,t,T)] = m_1(r,t,T), \tag{2}$$

$$Var[V(r,t,T)] = m_2(r,t,T) - (m_1(r,t,T))^2, \tag{3}$$

$$Cov[V(r,t,T_1),V(r,t,T_2)] = m_{1,1}(r,t,T_1,T_2) - m_1(r,t,T_1)m_1(r,t,T_2). \tag{4}$$

We are going to show that $m_{ij}(r,t,T_1,T_2)$ can be found as a solution of PDE

$$\frac{\partial m_{ij}}{\partial t} + \alpha\frac{\partial m_{ij}}{\partial r} + \frac{\beta^2}{2}\frac{\partial^2 m_{ij}}{\partial r^2} - (i+j)rm_{ij} = 0 \tag{5}$$

for $t \in [0, T_1)$ subject to final condition

$$m_{ij}(r,T_1,T_1,T_2) = m_j(r,0,T_2 - T_1) \tag{6}$$

if both $i, j > 0$, or either

$$m_j(r,T,T) = 1, \quad m_i(r,T,T) = 1 \tag{7}$$

otherwise.

In order to derive this, we introduce quantities

$$R_{ij}(t) = \begin{cases} (i+j)r(t) & 0 \le t < T_1, \\ jr(t) & T_1 \le t < T_2 \end{cases}$$

and

$$z(x,s) = \phi(x,s)e^{-\int_t^s R_{ij}(\tau)d\tau}$$

for some twice piecewise-differentiable function $\phi(x, s)$. Then the change dz equals

$$dz = e^{-\int_t^s R_{ij}(\tau)d\tau} \left[\left(\phi_s + \alpha\phi_x + \frac{\beta^2}{2}\phi_{xx} - R_{ij}(s)\phi \right) ds + \beta dX(s) \right].$$

And if $\phi(x, s)$ satisfies

$$\phi_x + \alpha\phi_s + \frac{\beta^2}{2}\phi_{ss} - R_{ij}(s)\phi = 0 \tag{8}$$

with final condition $\phi(x, T_2)$ then $m_{ij}(r, t, T_1, T_2) = \varphi(r, t)$. Now (8) doesn't look exactly like (5) with (6). But it is quite straightforward to notice that one can solve a simplified PDE

$$\tilde{\phi}_x + \alpha(x, s+T_1)\tilde{\phi}_s + \frac{(\beta(x, s+T_1))^2}{2}\tilde{\phi}_{ss} - jx\tilde{\phi} = 0$$

on $s \in [0, T_2 - T_1)$ with final condition $\tilde{\phi}(x, T_2 - T_1) = 1$ first, then another PDE

$$\overline{\phi}_x + \alpha(x, s)\overline{\phi}_s + \frac{(\beta(x, s))^2}{2}\overline{\phi}_{ss} - (i+j)x\overline{\phi} = 0$$

on $s \in [0, T_1)$ with final condition $\overline{\phi}(x, T_1) = \tilde{\phi}(x, 0)$ and obtain thus a continuous solution of the original (8) on the whole time interval $[0, T_2]$ as a combination of the two

$$\phi(x, s) = \begin{cases} \overline{\phi}(x, s) & s \in [0, T_1], \\ \tilde{\phi}(x, s+T_1) & s \in (T_1, T_2). \end{cases}$$

3 Analytic solutions for affine models

Problem (5),(6),(7) can be solved analytically for some popular spot-rate models, namely, models of the affine class. The spot rate process for this class is modeled by the following SDE with constants $k > 0, \theta > 0, \sigma \geq 0$ and $\gamma \geq 0$.

$$dr = k(\theta - r)dt + \sqrt{\sigma^2 r + \gamma}\, dX(t). \tag{9}$$

The affine model family can also be characterized (see [3], [14]) by the bond present value taking the functional form

$$m_1(r, t, T) = e^{A_1(T-t) + B_1(T-t)r}, \tag{10}$$

where $A_1(\tau)$ and $B_1(\tau)$ are functions to be defined. As it turns out, the representation (10) can be generalized to all other moments of the distribution:

$$m_j(r, t, T) = e^{A_j(T-t) + B_j(T-t)r}, \tag{11}$$

$$m_{i,j}(r, t, T_1, T_2) = \begin{cases} e^{A_{i,j}(T_1 - t, T_2 - T_1) + B_{i,j}(T_1 - t, T_2 - T_1)r} & t \leq T_1, \\ e^{A_j(T_2 - T_1 - t) + B_j(T_2 - T_1 - t)r} & T_1 < t \leq T_2. \end{cases} \tag{12}$$

We present a simplified solution for the Vasicek model first, then derive a general formulae for the affine class. Solutions for the mean and variance of a zero-coupon bond for affine models have been already known for a while (see, for example, [14]). Souza et al. in [11] have also obtained similar formulae for Vasicek model. We choose to present here derivations for all the moments for completeness.

3.1 Vasicek model

Setting $\alpha(r, t) = k(\theta - r)$ and $\beta(r, t) = \sigma$ we obtain a traditionally written SDE of Vasicek model [12].

$$dr = k(\theta - r)dt + \sigma dX(t).$$

For illustrative purposes, we will only pursue first and second moments here, although a general mixed-moment formula can also be easily obtained.

3.1.1 Bond present value

We will be looking for a solution in the form (10) satisfying initial conditions $A_1(0) = 0, B_1(0) = 0$.

Taking partial derivatives, we have

$$\frac{\partial m_1}{\partial t} = -(A_1' + B_1' r)m_1,$$

$$\frac{\partial m_1}{\partial r} = B_1 m_1, \quad \frac{\partial^2 m_1}{\partial r^2} = (B_1)^2 m_1.$$

Substituting these into the PDE (5) we obtain

$$-\left(A_1' + B_1' r\right) + k(\theta - r)B_1 + \frac{\sigma^2}{2}(B_1)^2 - r = 0.$$

Since functions $A_1(\tau)$ and $B_1(\tau)$ do not depend on r we collect terms with and without r separately and obtain 2 ODEs in the variable τ:

$$B_1' + kB_1 = -1,$$

$$A_1' = k\theta B_1 + \frac{\sigma^2}{2}(B_1)^2$$

both with initial conditions $A_1(0) = 0, B_1(0) = 0$. It is easy to see that

$$B_1(\tau) = \frac{1}{k}\left(e^{-k\tau} - 1\right)$$

and

$$A_1(\tau) = \frac{1}{k}\left(\theta - \frac{3\sigma^2}{4k^2}\right) + \frac{1}{k}\left(\frac{\sigma^2}{k^2} - \theta\right)e^{-k\tau} - \frac{\sigma^2}{4k^3}e^{-2k\tau} + \left(\frac{\sigma^2}{2k^2} - \theta\right)\tau$$

with $\tau = T - t$.

3.1.2 Bond variance

As it was suggested in the derivations above, we do not directly solve PDEs for the variance but calculate second moments instead and then combine it with the already obtained solution for the bond mean using (3). The derivations are very similar to the bond mean and we are going to jump straight to the solution in closed form. We are looking again for an affine solution in the form

$$m_2(r, t, T) = e^{A_2(\tau) + B_2(\tau)r} \tag{13}$$

and it turns out that

$$B_2(\tau) = \frac{2}{k}\left(e^{-k\tau} - 1\right)$$

and

$$A_2(\tau) = \frac{1}{k}\left(2\theta - \frac{3\sigma^2}{k^2}\right) + \frac{1}{k}\left(\frac{4\sigma^2}{k^2} - 2\theta\right)e^{-k\tau} - \frac{\sigma^2}{k^3}e^{-2k\tau} + \left(\frac{2\sigma^2}{k^2} - 2\theta\right)\tau.$$

With this the variance can be expressed as

$$v(r, t, T) = e^{2A_1(\tau) + 2B_1(\tau)r}\left(e^{A_2(\tau) - 2A_1(\tau)} - 1\right),$$

where, as before, $\tau = T - t$.

3.1.3 Bonds' co-variance

Here we will also rely on a mixed moment's solution using (4) instead of modeling co-variance directly.

As before, we assume that the two maturities of the bonds are $0 < T_1 \le T_2$. We denote $\Delta T = T_2 - T_1$ for convenience.

As demonstrated in Section 2.2, mixed moment $m_{1,1}(r, t, T_1, T_2)$ equals to mean value $m_1(r, t, -T_1, \Delta T)$ for $T_1 \le t \le T_2$. However, for $t < T_1$ our mixed moment satisfies

$$\frac{\partial m_{1,1}}{\partial r} + k(\theta - r)\frac{\partial m_{1,1}}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 m_{1,1}}{\partial t^2} - 2rm_{1,1} = 0$$

with final condition

$$m_{1,1}(r, T_1, T_1, T_2) = m_1(r, 0, \Delta T).$$

Again we are looking for a solution in the form (12). Let's introduce the quantity $\phi_{1,1}(\Delta T) = e^{-k\Delta T} + 1$. Then, skipping similar derivations, we obtain

$$B_{1,1}(s, \Delta T) = \frac{1}{k}\left(\phi_{1,1}(\Delta T)e^{-ks} - 2\right),$$

$$A_{1,1}(s, \Delta T) = 2s\left(\frac{\sigma^2}{k^2} - \theta\right) + \frac{\phi_{1,1}(\Delta T)}{k}\left(\theta - \frac{2\sigma^2}{k^2}\right)\left(1 - e^{-ks}\right)$$
$$+ \left(\phi_{1,1}(\Delta T)\right)^2 \frac{\sigma^2}{4k^3}\left(1 - e^{-2ks}\right) + A_1(\Delta T),$$

where the function A_1 is the function introduced earlier and $s = T_1 - t$.

3.2 CIR model

In this section we present solutions for the spot rate model of Cox, Ingersoll and Ross (CIR) [2].

$$dr = k(\theta - r)dt + \sigma\sqrt{r}dX. \tag{14}$$

The mean value is still found to be in the form (10) but the system of individual ODEs for functions A_1 and B_1 is nonlinear now:

$$B_1' = \frac{\sigma^2}{2}B_1^2 - kB_1 - 1, \tag{15}$$

$$A_1' = k\theta B_1. \tag{16}$$

Equation (15) can be written as

$$\frac{dB_1}{d\tau} = \frac{\sigma^2}{2}\left(B_1 - \psi_1^+\right)\left(B_1 - \psi_1^-\right), \tag{17}$$

where $\psi_1^+ = (k + \kappa_1)/\sigma^2$, $\psi_1^- = (k - \kappa_1)/\sigma^2$, $\kappa_1 = \sqrt{k^2 + 2\sigma^2}$.

Solving

$$\int \frac{dB_1}{\left(B_1 - \psi_1^+\right)\left(B_1 - \psi_1^-\right)} = \frac{\sigma^2}{2}\int d\tau$$

with initial condition $B_1(0) = 0$, we obtain

$$B_1(\tau) = \frac{2}{\sigma^2}\frac{1 - e^{\kappa_1\tau}}{\psi_1^+ e^{\kappa_1\tau} - \psi_1^-}.$$

If we divide (16) by (17) then another ODE is obtained

$$\frac{dA_1}{dB_1} = \frac{2k\theta}{\sigma^2}\frac{B_1}{\left(B_1 - \psi_1^+\right)\left(B_1 - \psi_1^-\right)}$$

for A_1 being a function of B_1, with initial condition $A_1(0) = 0$. The solution is

$$A_1(\tau) = \frac{k\theta}{\kappa_1}\left[\psi_1^+ \ln\left(1 - \frac{B_1(\tau)}{\psi_1^+}\right) - \psi_1^- \ln\left(1 - \frac{B_1(\tau)}{\psi_1^-}\right)\right],$$

where $\tau = T - t$.

Analogously, the second moment is of the form (13) with

$$B_2(\tau) = \frac{4}{\sigma^2}\frac{1 - e^{\kappa_2\tau}}{\psi_2^+ e^{\kappa_2\tau} - \psi_2^-},$$

$$A_2(\tau) = \frac{k\theta}{\kappa_2}\left[\psi_2^+ \ln\left(1 - \frac{B_2(\tau)}{\psi_2^+}\right) - \psi_2^- \ln\left(1 - \frac{B_2(\tau)}{\psi_2^-}\right)\right],$$

with $\psi_2^+ = (k + \kappa_2)/\sigma^2$, $\psi_2^- = (k - \kappa_2)/\sigma^2$, $\kappa_2 = \sqrt{k^2 + 4\sigma^2}$.

Derivation of the mixed moment for the bonds' co-variance is not dramatically different but looks a bit more complex due to the non-trivial final condition at $t = T_1$. We omit the derivations and present the final result here for $t \in [0, T_1]$.

We define an additional quantity (where, as before, $\Delta T = T_2 - T_1$)

$$\phi_{1,1}(\Delta T) = \frac{B_1(\Delta T) - \psi_2^+}{B_1(\Delta T) - \psi_2^-}$$

and express through it the solution (where, as before, $s = T_1 - t$)

$$B_{1,1}(s, \Delta T) = \frac{\psi_2^+ - \phi_{1,1}(\Delta T)\psi_2^- e^{\kappa_2 s}}{1 - \phi_{1,1}(\Delta T)e^{\kappa_2 s}},$$

$$A_{1,1}(s, \Delta T) = \frac{k\theta}{\kappa_2}\left[\psi_2^+ \ln\left(\frac{B_{1,1}(s) - \psi_2^+}{B_1(\Delta T) - \psi_2^+}\right) - \psi_2^- \ln\left(\frac{B_{1,1}(s) - \psi_2^-}{B_1(\Delta T) - \psi_2^-}\right)\right] + A_1(\Delta T).$$

3.3 General formula

Now we are in a position to give components $A_{i,j}$ and $B_{i,j}$ of the most general formula (11),(12) for mixed moments in the affine models (9).

We will need the following quantities to be defined

$$\kappa_i = \sqrt{k^2 + 2i\sigma^2},\quad \psi_i^+ = \frac{k + \kappa_i}{\sigma^2},\quad \psi_i^- = \frac{k - \kappa_i}{\sigma^2},$$

$$\phi_i^+ = \frac{k(\gamma + \sigma^2\theta)\psi_i^+ + i\gamma}{\kappa_i\sigma^2},\quad \phi_i^- = \frac{k(\gamma + \sigma^2\theta)\psi_i^- + i\gamma}{\kappa_i\sigma^2},$$

$$\phi_{i,j}(\Delta T) = \frac{B_j(\Delta T) - \psi_{i+j}^+}{B_j(\Delta T) - \psi_{i+j}^-}.$$

Then

$$B_i(\tau) = \frac{2i}{\sigma^2}\frac{1 - e^{\kappa_i\tau}}{\psi_i^+ e^{\kappa_i\tau} - \psi_i^-},$$

$$A_i(\tau) = \phi_i^+ \ln\left(1 - \frac{B_i(\tau)}{\psi_i^+}\right) - \phi_i^- \ln\left(1 - \frac{B_i(\tau)}{\psi_i^-}\right) + \frac{\gamma}{\sigma^2}B_i(\tau),$$

$$B_{i,j}(s, \Delta T) = \frac{\psi_{i+j}^+ - \phi_{i,j}(\Delta T)\psi_{i+j}^- e^{\kappa_{i+j}s}}{1 - \phi_{i,j}(\Delta T)e^{\kappa_{i+j}s}},$$

$$A_{i,j}(s,\Delta T)=\phi^+_{i+j}\ln\left(\frac{B_{i,j}(s,\Delta T)-\psi^+_{i+j}}{B_j(\Delta T)-\psi^+_{i+j}}\right)-\phi^-_{i+j}\ln\left(\frac{B_{i,j}(s,\Delta T)-\psi^-_{i+j}}{B_j(\Delta T)-\psi^-_{i+j}}\right)$$
$$+\frac{\gamma}{\sigma^2}\left(B_{i,j}(s,\Delta T)-B_j(\Delta T)\right)+A_j(\Delta T),$$

where $\tau = T - t$, $s = T_1 - t$, $\Delta T = T_2 - T_1$.

4 Construction of hedging portfolios

4.1 Minimum variance portfolio

Getting back to the interpolation task, we will try to price a zero-coupon bond with a given maturity then extract the effective yield for that maturity from the obtained price.

There are three key principles:

- we hedge the bond with other tradeable bonds and plan to hold the assets until maturity;
- hedge ratios in our portfolio are chosen so that variance of the discounted value of the portfolio is minimal;
- price of the original bond should depend on present values of all portfolio constituents.

4.1.1 Simple case: Single-bond hedge

There follows a step-by-step example using only one other zero-coupon bond as a static hedge. Suppose someone is willing to buy a bond with maturity T_0 from us. We quote a price p_0 for a bond with present value V_0. Note that the present value is a random variable and depends on the realized yield. We sell one bond but at the same time we buy some quantity h of another bond with present value V for the quoted price p. This bond has maturity T different from our T_0. Both V_0 and V are random variables while p is not. One way of setting the price p_0 is to equate it with the expectation of our total portfolio present value:

$$p_0=\mu=E\left[V_0+h(V-p)\right]=E\left[V_0\right]+h\left(E[V]-p\right)$$

and the expected yield curve value for that maturity will be

$$y=-\frac{\ln p_0}{T_0-t}=-\frac{\ln\left[m_1(r,t,T_0)+h\left(m_1(r,t,T)-p\right)\right]}{T_0-t}.$$

Minimizing the variance of the portfolio

$$Var\left[V_0\right]+2hCov\left[V_0,V\right]+h^2Var[V]$$

yields the hedge ratio

$$h=-\frac{Cov\left[V_0,V\right]}{Var[V]}=-\frac{m_{1,1}(r,t,T_0,T)-m_1(r,t,T_0)m_1(r,t,T)}{m_2(r,t,T)-\left(m_1(r,t,T)\right)^2}.$$

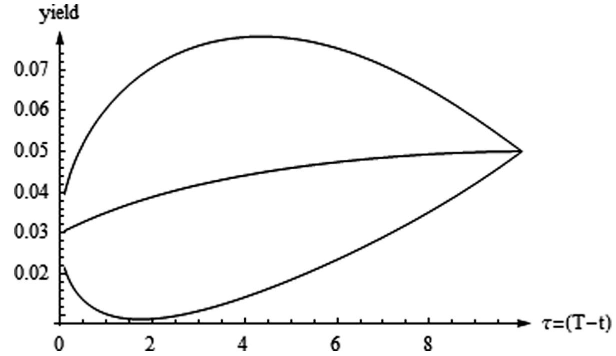
The portfolio variance then becomes

$$v^2=Var\left[V_0\right]-\frac{\left(Cov\left[V_0,V\right]\right)^2}{Var[V]}.$$

Note that it is not necessary that $T_0 < T$ but if this is not the case we will need to swap T_0 and T when computing the moments above.

Finally, we often want to incorporate the obtained quantities into utility-function considerations. A very simple way of augmenting the quoted price with one's risk

Figure 1: Yield curve for Vasicek model with one standard deviation upper/lower bounds obtained $\tau = T_0 - t \in [0.01, 10]$ with single bond hedge with fixed maturity T = 10.0 and quoted yield 0.05. Used parameters: $r = 0.03$, $\kappa = 0.2$, $\theta = 0.06$, $\sigma = 0.05$.



preferences is to add standard deviation bands $\pm\xi\sqrt{v^2}$ with some selected level ξ ([4], [5]). If we now vary maturity T_0 we obtain a mean yield curve as well as a yield curve envelope as illustrated in Figure 1. There, for convenience, mean value and standard deviation bands are plotted as functions of time to maturity $\tau = T_0 - t$ while T is kept fixed.

4.1.2 Multi-bond portfolios

Extending it to the case where we have access to bonds of more than one maturity for hedging is relatively easy. The bond we need to price has present value which we denote $m_0 = E[V_0]$ and variance $v_0^2 = Var[V_0]$. We denote the present values of available bonds as $\mathbf{m} = (m_1, m_2, \dots, m_n)^T$, their co-variance matrix as K, and the vector of their co-variances with the bond being priced as $\mathbf{k}$. We also have a vector of available bond prices $\mathbf{p}$. What we need to find is the optimal hedge ratios vector $\mathbf{h}$ and corresponding price bands.

Solving an unconstrained optimization task

$$\mathbf{h}^*=\arg\min_{\mathbf{h}}\left(v_0^2+2\mathbf{k}^T\mathbf{h}+\mathbf{h}^TK\mathbf{h}\right)$$

we obtain $\mathbf{h}^* = -K^{-1}\mathbf{k}$ that gives us mean and variance

$$\mu=m_0-\mathbf{k}^TK^{-1}(\mathbf{m}-\mathbf{p}),\ v^2=v_0^2-\mathbf{k}^TK^{-1}\mathbf{k}. \quad (18)$$

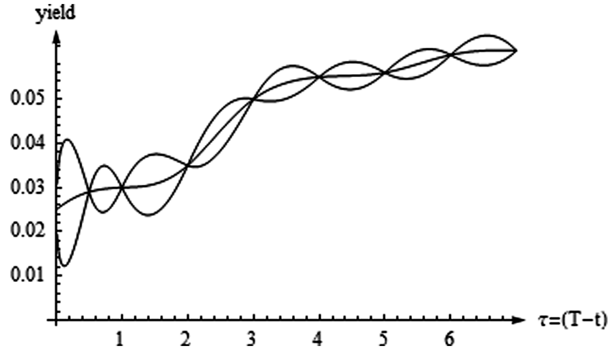
Figure 2 shows what the price bands would look like if we use multiple instruments with which to hedge.

Note that without additional constraints the hedge ratios are not necessarily one sided (all negative or all positive). For example, for the inputs used in Figure 2 a long bond position with maturity $\tau = 4.5$ will have to be hedged optimally with ratios:

$$\mathbf{h}^*=(-0.0076, 0.0143, -0.0382, 0.137, -0.6178, -0.5799, 0.1072, -0.0168)^T.$$

The good news is that hedge ratios are independent of quoted prices (i.e., yields). As is usually the case, there is non-zero bid–ask spread in tradeable instruments. But we are able to plug in the right price of a particular bond into (18) depending on its hedge ratio sign afterwards. Note that this is because our optimization is based solely

Figure 2: Yield curve for general affine model with one standard deviation upper/lower bounds for $\tau = T_0 - t \in [0,8]$ obtained with multiple bond hedges with maturities (0.5,1,2,3,4,5,6,7) and corresponding yields (0.029,0.03, 0.035,0.05,0.055,0.056,0.06,0.061). Used parameters: $r = 0.025$, $\kappa = 0.2$, $\theta = 0.03$, $\sigma = 0.04$, $\gamma = 0.01$.



on variance. Had we chosen a static-hedge optimization procedure that involved price, then this would no longer be the case.

4.1.3 Pricing multiple bonds at once

A slightly more elaborate solution is obtained for the case where we are pricing a portfolio of bonds $\mathbf{V}_0$ with a pre-specified vector of weights $\mathbf{w}$. Although this task may seem to be a bit contrived at first, it allows us to establish a connection to GPR as will be described later. Another reason is that it allows us to price coupon-based bonds (consider coupons as separate bonds with pre-defined weights).

We denote $\mathbf{m}_0 = E[\mathbf{V}_0]$, and $\mathbf{V}$ as the vector of bonds used for hedging. The co-variance matrix of composite bond vector $(\mathbf{V}_0, \mathbf{V})$ can be written in a block form as

$$K = \begin{pmatrix} K_{00} & K_{01} \\ K_{01}^T & K_{11} \end{pmatrix}.$$

When trying to minimize $Var\left[\mathbf{V}_0^T\mathbf{w} + \mathbf{V}^T\mathbf{h}\right]$ we obtain mean portfolio price

$$\mu = \left(\mathbf{m}_0 - K_{01}^T K_{11}^{-1}(\mathbf{m} - \mathbf{p})\right)^T \mathbf{w} \quad (19)$$

and variance

$$v^2 = \mathbf{w}^T\left(K_{00} - K_{01}^T K_{11}^{-1} K_{01}\right)\mathbf{w}. \quad (20)$$

4.2 Fixed variance portfolio

What would be the worst/best price for a bond when hedged portfolio variance is constrained? Suppose, the portfolio variance should be equal to $\tilde{v}^2$:

$$\tilde{v}^2 = v_0^2 + 2\mathbf{h}^T\mathbf{k} + \mathbf{h}^T K\mathbf{h}.$$

We solve the constrained optimization problem

$$m_0 + \mathbf{h}^T(\mathbf{m} - \mathbf{p}) \rightarrow \min_{\mathbf{h}} / \max_{\mathbf{h}}$$

with the Lagrange multipliers method and obtain

$$\mu = m_0 - \mathbf{k}^T K^{-1}(\mathbf{m} - \mathbf{p}) \pm \sqrt{(\mathbf{k}^T K^{-1}\mathbf{k} + \tilde{v}^2 - v_0^2)(\mathbf{m} - \mathbf{p})^T K^{-1}(\mathbf{m} - \mathbf{p})}.$$

4.3 Pricing and using coupon-paying bonds for hedging

As noted above, pricing a single coupon-paying bond is equivalent to pricing a portfolio of bonds with pre-specified weights.

Hedging with coupon-paying bonds is also possible to consider within the proposed framework. If all the cash-flows are specified in advance, we can easily incorporate it into the optimization task in the form of equality constraints on hedge ratios. Although a closed-form solution is possible in this situation too (using the Lagrange multipliers method again), we do not pursue this here to avoid clutter—most realistic scenarios are likely to incorporate inequality constraints as well and that would often render a closed-form solution impossible. See below the examples of inequality constraints.

4.4 Limited hedging with inequality constraints

Various inequality constraints can be easily incorporated into the portfolio construction task although closed-form solutions are rarely possible. Nevertheless, one can utilize widely available convex optimization software to find numerical solutions.

As an example, we can impose a maximum position size limit on every constituent. That will be equivalent to having a constraint

$$||\mathbf{h}||_\infty \leq \tilde{h}.$$

Alternatively, we may demand that hedging notional does not exceed certain value, i.e.

$$|\mathbf{h}^T\mathbf{p}| \leq \tilde{h}.$$

The goal function can be portfolio variance as before or min/max price (given an additional, perhaps, inequality constraint on variance).

As mentioned previously, hedge ratios might need to be one-sided in practical situations, which gives us yet another type of optimization task.

An even more practical task may be to incorporate lot-sizes for hedging instruments. This will turn the task into mixed-integer programming problem.

We can anticipate that in all these scenarios the price bands will remain a smooth "sausage" shape, although the widths will vary.

5 Gaussian Process Regression

One cannot help noticing that the expression inside the brackets in (20) is called *Schur complement* of matrix Σ, while the expression in (19) is the mean of conditional multivariate Gaussian distribution. Going one step further we can also point at connections with Gaussian Processes (GPs) (see [8] for a comprehensive reference). Equations (18, 19, 20) coincide with those for Gaussian Processes Regression (GPR) with kernel (co-variance) function defined by (4).

This coincidence may give us additional insights into the interpolation task and, so, we will give here a very short and targeted introduction into GPR theory. We will follow [8] in our introduction and consider only the "noise-free" case.

Formal definition states that a Gaussian Process is a collection of random variables any finite number of which have a joint Gaussian distribution. Such a collection can be thought of as a particular realization of GP and be represented as a function $f(x)$ where x can be, for example, a multidimensional continuous variable ($x \in R^D$). Thus, for any finite set $\{x_1, x_2, \ldots, x_n\}$ corresponding function values are drawn from a multivariate Gaussian distribution which can be completely described by its first two moments. This fact is usually denoted as

$$f(x) \sim \mathcal{GP}\left(m(x), k\left(x, x'\right)\right). \quad (21)$$

meaning that

$$\left(f(x_1), f(x_2), \dots, f(x_n)\right)^T \sim \mathcal{N}(\mathbf{m}, K),$$

where $\mathbf{m} = (m(x_1), m(x_2), \dots, m(x_n))^T$, $K = \left(k(x_i, x_j)\right)_{ij}^n$, the co-variance matrix with individual entries expressed through some function of two variables $k(x, x')$ which is sometimes called the *kernel function*.

It remains to define functions $m(x)$ and $k(x, x')$ in order to fully specify the Gaussian distribution in (21).

So far so good, but this might seem not especially useful yet. GPs acquire practical value when we consider inference problems. In particular, given some observations of a particular realization of GP, called *training set* and denoted as $\{(x_i, f_i)|i = 1, \dots, n\}$ we want to pinpoint locations of other (unobserved) points of the same realization. In particular, we are given *test inputs* $X_* = \{x_i^*|i = 1, 2, \dots, p\}$. We do not normally know exactly where the unobserved *test outputs* are. The question is then asked: What can we say about the distribution of test outputs?

Using the definition of GP, we conclude that vectors of training outputs $\mathbf{f}$ and testing outputs $\mathbf{f}_*$ are jointly distributed as

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m} \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} K(X, X) & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*). \end{bmatrix}\right).$$

Conditional on observable training outputs, the distribution of test outputs only can be expressed using Schur's complement:

$$\mathbf{f}_* \sim \mathcal{N}(\mu, \nu),$$

where

$$\mu = \mathbf{m}_* - K(X_*, X)K(X, X)^{-1}(\mathbf{m} - \mathbf{f})$$

and

$$\nu = K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*).$$

The last two formulas constitute Gaussian Process Regression algorithm in the noise-free case. In practice, the kernel function $k(x, x')$ is usually unknown and needs to be learned from data. Some parametric form of the function is usually assumed and the parameters of the kernel function (called *hyperparameters*) are fitted/calibrated to available training data.

As noted previously, coincidentally, GPR produces formulae for the minimum variance portfolio case.

Are we stating here that bond prices are Gaussian Processes? Obviously not. One can easily see this if we remember that a single bond value cannot be negative while any Gaussian random variable is able to take values in the range $(-\infty, \infty)$. What seems apparent is that equivalence of our curve interpolation task and GPR algorithm is merely the consequence of us using only the first two moments to summarize a distribution of non-traded bond prices as random variables.

It is interesting to note that with a particular choice of kernel function GPR algorithm is equivalent to spline smoothing (see references within [8]).

6 Model calibration/fitting

Can GPs be used in order to fit parameters of interest-rate models? Indeed, in [11] Sousa et al. suggest using GPs for fitting parameters in Vasicek model. We will consider the calibration process in detail.

Here we come to an interesting junction. There are two disjoint sets of data: historical quotes and current prices. Practitioners have long faced the dilemma of which dataset to use for calibration. On one hand, historical prices are abundant and fitting with Maximum Likelihood approach is a well-posed problem. Parameter estimations usually have tight confidence intervals. On the other hand, when these parameters are plugged into the model a poor match to the current prices is often observed. This might lead to development of a more complex model, the one that gives a better current price match, although still not ideal. More complex models are usually harder to fit and more computationally demanding. Another approach would be to discard the historical data completely and use only the current prices. This gives wild parameter estimates, estimates that inconsistently change from day to day, but the price match is perfect!

A nice feature of our approach is a perfect match to current prices regardless of the fitting process. This property is borrowed from GPR, as already noted. The parameter estimates, however, matter. And the question of a suitable fitting procedure remains.

The authors of [11] note that using only one GP realization path as a training set for parameters fitting is unsatisfactory. They experiment with currently quoted prices and conclude that the confidence intervals resulting from parameter uncertainties are large.

We would add that the GP framework is imposed on bond pricing rather superficially in our view. There is no guarantee that even a sample from a simulated bond price path has multivariate Gaussian distribution with the desired kernel function. Therefore, having more data points might not necessarily give better estimates. This is especially true for models other than Vasicek (CIR and more general affine models). One might end up imposing additional constraints in the form of inequalities or Bayesian prior probabilities on hyperparameters in order to achieve numerically convergent optimization.

Another problem with GPR-based parameter estimation is the necessity of introducing noise into the model even for Vasicek model. Even a brief encounter with some available implementations ([7], [9]) of GPR reveals that it is very often difficult to switch off fitting of the noise variance parameter. The reason is, again, reliable optimization convergence to finite values. But adding noise into our framework is equivalent to destroying the "sausage" shape and the perfect price match for traded instruments.

While parameter fitting in GPR models may be a fruitful area of further research, we can also suggest alternative approaches for parameter fitting. Approaches that derive the probability distributions from the actual model dynamics (as described by SDEs) and then use these distributions to fit parameters. Examples include Kalman filter ([1]) for the Vasicek model, while for general affine models particle filtering (PF) might be a more natural algorithm (see [10]).

7 Conclusion

We have presented a framework for constructing individual yield curves with uncertainty bands for spot rate models that can be described by a single-factor SDE. The only assumption made is ability to statically hedge with available instruments. Closed-form formulae are derived for affine models and portfolios with minimal variance. Using the framework and regardless of parameter fitting procedure, one can always obtain a perfect price match with tradeable instruments without a need for statistically inconsistent calibration to current prices.

8 Acknowledgments

We are grateful to Thijs van den Berg for useful discussions and for pointing out the connection with Gaussian Processes.

Yury Rojek obtained his PhD in Applied Mathematics (Differential Equations) from Voronezh State University (Russia). He is currently employed by Deutsche Bank AG as an Algorithmic Trading Quantitative Analyst with responsibilities covering research, design,

and implementation of liquidity seeking algorithms for European equities. His research interests include derivatives pricing, machine/reinforcement learning, signal processing, Bayesian statistics, optimisation methods, optimal control, and empirical data analysis.

Paul Wilmott is a financial consultant, specializing in derivatives, risk management, and quantitative finance. He has worked with many leading US and European financial institutions. He studied mathematics at St Catherine's College, Oxford, where he also received his D.Phil. He founded the Diploma in Mathematical Finance at Oxford University and the journal *Applied Mathematical Finance*. He is the author of *Paul Wilmott Introduces Quantitative Finance* (Wiley 2007), *Paul Wilmott On Quantitative Finance* (Wiley 2006), *Frequently Asked Questions in Quantitative Finance* (Wiley 2009), and other financial textbooks. He has written over 100 research articles on finance and mathematics. He was a founding partner of the volatility arbitrage hedge fund Caissa Capital which managed $170 million. His responsibilities included forecasting, derivatives pricing, and risk management. He is the proprietor of the popular quantitative finance community website www.wilmott.com and the quant magazine *Wilmott*, and the creator of the Certificate in Quantitative Finance.

REFERENCES

[1] D. J. Bolder. 2001. Affine term-structure models: Theory and implementation. Bank of Canada Working Paper, 2001-15, October. Available at SSRN: http://ssrn.com/abstract=1082826 or http://dx.doi.org/10.2139/ssrn.1082826.

[2] J. C. Cox, J. E. Ingersoll, S. A. Ross. 1985. A theory of the term structure of interest rates. *Economet.* **53**(2): 385–407.

[3] D. Duffie, D. Filipovic, W. Schachermayer. 2003. Affine processes and applications in finance. *Ann. Appl. Prob.* **13**(3): 984–1053.

[4] D. Epstein, P. Wilmott. 1999. The Yield Envelope: Price Ranges for Fixed Income Products. OFRC Working Papers Series. Oxford Financial Research Centre: Oxford. Available at: http://www.finance.ox.ac.uk/file_links/finecon_papers/1999mf10.pdf.

[5] D. Epstein, P. Wilmott. 1998. A new model for interest rates. *Int. J. Theor. Appl. Fin.* **1**(2): 195–226.

[6] P. S. Hagan, G. West. 2006. Interpolation methods for curve construction. *Appl. Math. Fin.* **13**(2): 89–129.

[7] F. Pedregosa *et al.* 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**: 2825–2830.

[8] C. E. Rasmussen, C. K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. MIT Press: Boston, MA.

[9] C. E. Rasmussen, H. Nickisch. 2010. Gaussian Process Regression and Classification. Toolbox version 3.1, 2010-09-27. Available at: http://gaussianprocess.org/gpml/code.

[10] G. de Rossi. 2010. Maximum likelihood estimation of the Cox–Ingersoll–Ross model using particle filters. *Comput. Econ.* **36**(1): 1–16.

[11] J. B. Sousa, M. L. Esquvel, R. M. Gaspar. 2012. Machine learning Vasicek model calibration with Gaussian processes. *Commun. Stat. Simul. Comput.* **41**(6): 776–786.

[12] O. Vasicek. 1977. An equilibrium characterization of the term structure. *J. Fin. Econ.* **5**: 177–188.

[13] M. Veraar. 2012. The stochastic Fubini theorem revisited. *Int. J. Prob. Stochast. Process.* **84**(4): 543–551.

[14] P. Wilmott. 2006. *Paul Wilmott on Quantitative Finance*, Vol. II, 2nd edn. Wiley: New York.