A nonparametric zero–coupon yield curve

Mojmír Simerský*

January 2016

Keywords: yield curve estimation, nonparametric regression, optimal smoothness

Abstract

The paper adds to a volume of publications on the construction of a zero–coupon yield curve (YC) from coupon–bearing instruments, primarily from sovereign coupon bonds.

In general, our method belongs to the penalized spline category. The starting point is a piecewise linear model of forward rates on a time partition of some 20 knots, from zero to 50 years. These knots are fixed and cover also commonly quoted maturity benchmarks. Such number of knots is sufficiently high to reflect a structure of the curve. In our current implementation, this number can be increased up to the limit of the matrix inversion capability built in MS Excel. The number of instruments is arbitrary.

The nonlinear optimization problem is solved iteratively by the Newton–Raphson (N–R) method. Given the rank of the Hessian matrix – 19 in our case – concerns about numerical stability are quite important. We hope having found reasonable additional constraints that improve stability under a variety of initial conditions.

Our ambition to find a procedure generating an adaptive smoothing factor that, under different bond portfolios and prices, leads to the same degree of "smoothness" of the spot curve, has been met with only a partial success. A subjective metrics to estimate the "fairness" of the YC is proposed and implemented.

1 Introduction

In our view, there does not exist any substantiated theory answering the question of the yield curve functional form. Therefore, some *ad hoc* assumptions are indispensable. (For instance, the successful parametric Nelson–Siegel model [1] stands on a rather vague claim that the instantaneous forward rate should satisfy a second-order differential equation.)

We start with a simplified problem of estimating the YC from a portfolio of instruments whose future cash flows occur only at a limited number of fixed times, henceforth referred to as knots. We maintain that as we have no market information pertaining to the times between the knots, it is reasonable to model the forward rate by a piecewise linear continuous function with linear segments between the knots. (Let us remark that a similar

*Permanent address: info@zyc.cz

assumption is inherent in the bootstrap method, where the forward rate has the simplest form – the step function.)

The position of the knots is arbitrary. For practical use, nevertheless, the knots should be denser at shorter maturities and should include all conventional benchmarks. In the case of sovereign bonds, maturities may reach up to some 40 or 50 years. Having in mind the current Czech government bond market, which will serve us as a "case study", we have chosen the following knot partition: $0.25, 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 10, 12, 15, 20, 25, 35$, and 50 years. Not surprisingly, real debt instruments with cash flows only at those knots do not exist. Therefore, we introduce a procedure that splits any cash flow into two cash flows at the adjacent knots. In this way, any cash flow scheme can be treated as if all the cash flows occurred only at the knot times.

The penalized spline method seeks optimal coefficients of a spline (here the forward rate) under constraints related to the spline "roughness". The goal is to minimize the sum of squared (and, in general, weighted) pricing errors of constituting bonds and the roughness penalty. The minimization is carried out in the space of spline coefficients, or equivalently, in the space of particular discount functions. The exponential form of the discount function leads to a nonlinear optimization problem with the sums of exponentials, which is known to be ill–conditioned (e.g., [2]). From numerous numerical experiments we have learned that the stability can be substantially improved by constraints at the longest maturities. We have devised and applied the following two.

Some models, the N–S including, give an asymptotic forward rate at infinity. In actuarial parlance, the concept is known as UFR – the ultimate forward rate. Such idea, however, stands in a stark contradiction to another interest rate concept: the reversion to the mean. In our endeavor to fix the forward rate at the longest maturity, we opted for the mean reversion scheme: at the time of, say, twice the longest maturity, the forward rate is expected to attain the same value as at $t = 0$. In other words, the rightest knot on the time axis (50 years) is set as the "mean reversion point". The time partition is therefore augmented with one knot at (for instance) twice the time of the preceding knot (100 years) and the augmented problem is solved with the aforestated constraint. In this manner, we could fix the rate at the longest maturity using the value at $t = 0$. The rate fixed at zero time, which becomes now an input parameter, stabilizes the short end of the curve as well.

The second stabilizing constraint relates to the spot curve extrema. Supposing a continuous forward rate, it is trivial to show that the YC local extrema occur at the points where the YC equals the forward rate. There could be several such local extrema, depending on the price data and allowed degree of roughness. We will require that one of the extrema be placed at the mean reversion point. In other words, the spot curve leaves the right end of maturities with a zero first derivative.

The roughness penalty is usually computed as an integral of the squared second derivative of the spline in question. As we are working with a linear spline, we take, as a reasonable proxy for the second derivative, the difference of the first derivatives (i.e., slopes) at the knot points. The penalty is then the sum of squared slope differences at the knots. These differences are weighted in a way to satisfy the two stabilizing constraints just described.

We would like to stress that this weighting does not function as a "variable roughness penalty" sometimes used in the YC splining. The variable penalty, in short, allows for more details at the short maturities and suppresses oscillations at the long end of the YC. It is argued that the estimate of the yield curve should be more "accurate" for the near future than for, say, 30 years. Here we see two different problems: the sparse market data at long maturities can and do affect excessively the shape of the YC, and, a continuation of the YC for maturities with no available cash flows cannot be but speculative. The first effect can be mitigated by minimizing the sum of squared yield errors instead of price errors, as often used in the zero–coupon YC estimation ([3]).

For us, the most subjective point seems to be the optimal degree of smoothing, i.e., the relative contributions of the yield errors and the roughness penalty. A sophisticated way, developed, however, for a different type of problems, is the GCV – the general cross validation (applied in, e.g., [4], [5]). The method, roughly speaking, seeks a degree of smoothing that gives the best estimate of the price or yield of one left–out instrument, based on the estimation using the remaining ones; such estimation errors are summed, of course, over all sequentially omitted instruments. After days and days of rumination, we had come to a conclusion that this was not the goal we wanted to achieve. For us, the shape of the YC is associated with the words like "reasonable", "plausible", "likely" rather than with some predictive power related to left–out instruments. In a word, the YC should be "fair". Hardly can we find a fairer reflection on this elusive quality than this one, quoted from [6]:

Fairness, while clearly associated with smoothly varying curvature, is nonetheless a poorly defined concept. Even though most people think they recognize fairness (or the lack thereof) when they see it, there is no agreed–upon mathematical formalism that defines it. Fairness does not simply increase with the degree n of continuity (G^n). It also is not properly captured with the behavior of an ideal elastic strip, as in the minimum–energy spline (MEC), nor by minimizing the variation of curvature, as in the minimum–variation curve (MVC). There is no known functional that completely captures the notion of fairness, nor any agreed–upon way to measure it.

We tried, partly exonerated by the opinion above, to find some measure of fairness, carrying out a number of numerical experiments. The result leaves much to be desired but it is our view that fixed knots and aptly fixed curve fairness can eliminate, at least partially, data–dependent estimation artifacts that may occur due to, e.g., dynamic knot insertion and/or cross–validation. Time series of yields and spreads computed with fixed fairness may be corrupted with some systematic errors, but those errors are likely to be persistent in time.

2 Estimation

2.1 Spline representation

Let's consider a set of M debt instruments with known future cash flows that occur only at discrete times $t_1 < t_2 < \ldots < t_{N-1}$, where $t_1 \geq 0$. The first cash flow of each instrument is negative and represents the payment equal to the gross market price at the time of the cash settlement t_1. Other cash flows, if present, are positive. To include the rate reversion, we augment these knots with the last one: t_N, where $t_N = m\, t_{N-1}\ , \quad m > 1\ .$

The estimated yield curve generates a set of discount factors at the knots. We denote by $f(t)$ the instantaneous forward rate; in continuous compounding, the discount factor at a time t is

$$d(t) = \exp\Big(- \int_0^t f(\tau)\,\mathrm{d}\tau \Big)\ , \tag{1}$$

and the spot curve

$$s(t) = \frac{\displaystyle\int_0^t f(\tau)\,\mathrm{d}\tau}{t}\ , \qquad \text{for } t > 0 \quad \text{and } s(0) = f(0)\,. \tag{2}$$

The discount factors $d(t_i)$ at the knots $t_1, \ldots, t_N$ form a vector, denoted by $\mathbf{d}$. In the following, vectors and matrices will be set in bold. So, we arrange the payments and the nominal cash flows into an $(M \times N)$ matrix $\mathbf{T}$. The $(M \times N)$ matrix $\mathbf{F}$ of the discounted cash flows is a transformation of $\mathbf{T}$, where each column has been multiplied by the corresponding discount factor,

$$\mathbf{F} = \mathbf{T}\ \mathrm{diag}(\mathbf{d})\,.$$

The price errors $\boldsymbol{\epsilon}$ of the instruments with respect to the YC are simply $\boldsymbol{\epsilon} = \mathbf{F}\ \mathbf{1}$, where $\mathbf{1}$ denotes the vector of ones. The piecewise linear forward rate with knots $\mathbf{t} = (t_1, \ldots, t_N)$ can be expressed in terms of its final value $f_N = f(t_N)$ and a set of $N-1$ slopes

$$k_i = \frac{f_{i+1} - f_i}{t_{i+1} - t_i}, \quad i = 1, 2, \ldots, N-1\,. \tag{3}$$

Let us remind that the final value f_N is not estimated; its value is fixed and supplied as an input parameter. Our roughness penalty is a function of slope differences, thus reflecting the "bends" of the forward rate,

$$b_i = k_{i+1} - k_i,\ i = 1, 2, \ldots, N-1\,.$$

We denote by $\mathbf{b}$ the vector of the bends, with the last, so far undefined, component b_N set at $b_N = -k_N$. It is evident that the relation between the bends and the values of the forward rate at the knots is a linear transformation, and it holds

$$f_i = f_N + \sum_{j=i+1}^{N} b_j (t_j - t_i)\,. \tag{4}$$

This formula is easily extended to any t in the interval $\langle 0, t_N \rangle$. The forward rate is a linear spline, so it can be expressed as a sum of basis functions

$$f(t) = f_N + \sum_{k=1}^{N} b_k (t_k - t)_+\ ,$$

i.e., b_k are the spline coefficients in the basis of truncated power functions of the order 1, here with the opposite direction of the time axis.

2.2 The N–R method

We need the relation between the discount vector $\mathbf{d}$ and the bend vector $\mathbf{b}$. The integral of the basis function is

$$A_{ij} = \int_0^{t_i} (t_j - t)_+ \,\mathrm{d}t = \begin{cases} t_i \left(t_j - \dfrac{t_i}{2} \right) & \text{if } i \le j\,, \\ \dfrac{t_j^2}{2} & \text{if } i > j\,, \end{cases} \tag{5}$$

and the integral in the exponent of (1) can be expressed as

$$\int_0^{t_i} f(t)\,\mathrm{d}t = f_N t_i + (\mathbf{A}\,\mathbf{b})_i, \tag{6}$$

where the elements A_{ij} of the $(N \times N)$ matrix $\mathbf{A}$ are given by the relations (5). The discount vector $\mathbf{d}$ as a function of $\mathbf{b}$ and f_N is now

$$\mathbf{d} = \exp\left(-(\mathbf{A}\,\mathbf{b} + f_N \mathbf{t})\right);$$

this shorthand notation should be understood component–wise.

We are now ready to write a standard form of the penalty function (loss function) to be minimized

$$L(\mathbf{b}) = \frac{1}{2}\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} + \frac{1}{2}\beta\,\mathbf{b}^\top \mathbf{B}\,\mathbf{b}\,, \tag{7}$$

where we have introduced a smoothing parameter β and a bend weighting matrix $\mathbf{B}$. The matrix $\mathbf{B}$ is supposed to be symmetric and non–negative, with elements independent of $\mathbf{b}$; the superscript $\top$ stands for the matrix transpose.

The gradient $\mathbf{g}$ of L with respect to $\mathbf{b}$ is a row vector of length N:

$$\mathbf{g} = \frac{\partial L}{\partial \mathbf{b}} = \boldsymbol{\epsilon}^\top \left(\frac{\partial \boldsymbol{\epsilon}}{\partial \mathbf{b}} \right) + \beta \mathbf{b}^\top \mathbf{B}$$

The derivative $\left(\dfrac{\partial \boldsymbol{\epsilon}}{\partial \mathbf{b}} \right)$ is an $(M \times N)$ matrix

$$\frac{\partial \boldsymbol{\epsilon}}{\partial \mathbf{b}} = \frac{\partial (\mathbf{T}\,\mathrm{diag}(\mathbf{d})\,\mathbf{1})}{\partial \mathbf{b}} = \mathbf{T}\,\frac{\partial \mathbf{d}}{\partial \mathbf{b}} = -\mathbf{T}\,\mathrm{diag}(\mathbf{d})\,\mathbf{A} = -\mathbf{F}\,\mathbf{A}\,,$$

so that the transposed gradient has the form

$$\mathbf{g}^\top = \beta\,\mathbf{B}\,\mathbf{b} - \mathbf{A}^\top\,\mathbf{F}^\top\,\boldsymbol{\epsilon}\,. \tag{8}$$

To minimize L, this gradient should approach the zero vector by iterating $\mathbf{b}$. In line with the N–R method, we write

$$\mathbf{g}^{(i+1)} = \mathbf{g}^{(i)} + \mathbf{H}\,(\mathbf{b}^{(i+1)} - \mathbf{b}^{(i)})\,.$$

The Hessian $\mathbf{H}$ can be obtained, after some algebra, as

$$\mathbf{H} = \frac{\partial \mathbf{g}^\top}{\partial \mathbf{b}} = \beta\,\mathbf{B} + \mathbf{A}^\top \mathbf{F}^\top \mathbf{F}\,\mathbf{A} + \mathbf{A}^\top \mathrm{diag}(\mathbf{F}^\top \boldsymbol{\epsilon})\,\mathbf{A}\,.$$

In the last term, the vector $\mathbf{F}^\top \boldsymbol{\epsilon}$ should be small for good estimates of the yield curve and we will omit it in the N–R implementation. Empirically, it was found that this omission improves the convergence of the numerical procedure.

Finally, the N–R iteration is, at least in principle, straightforward; from the input parameters f_N, β, and the input matrix $\mathbf{T}$ the initial gradient is

$$\mathbf{g}^{(0)} = -\mathbf{1}^\top \mathbf{F}_0^\top \mathbf{F}_0 \, \mathbf{A}, \text{ (see (8)), } \text{ where } \mathbf{F}_0 = \mathbf{T} \, \mathrm{diag}(\exp(-f_N \mathbf{t})) \; .$$

The starting bends are set to zero, $\mathbf{b}^{(0)} = \mathbf{0}$, and the starting Hessian computed from $\mathbf{F}_0$ and a constant matrix $\mathbf{B}$.

Now we iterate the values $\mathbf{b}^{(i+1)}$, $\mathbf{g}^{(i+1)}$ in the steps as follows:

$$\mathbf{b}^{(i+1)} = \mathbf{b}^{(i)} + \mathbf{H}^{-1} \left(\mathbf{g}^{(i)} \right)^\top$$

$$\mathbf{d} = \exp \left(-(\mathbf{A} \, \mathbf{b}^{(i+1)} + f_N \mathbf{t}) \right)$$

$$\mathbf{F} = \mathbf{T} \, \mathrm{diag}(\mathbf{d})$$

$$\mathbf{H} = \beta \, \mathbf{B} + \mathbf{A}^\top \mathbf{F}^\top \mathbf{F} \, \mathbf{A}$$

$$\boldsymbol{\epsilon} = \mathbf{F} \, \mathbf{1}$$

$$\left(\mathbf{g}^{(i+1)} \right)^\top = \beta \, \mathbf{B} \, \mathbf{b}^{(i+1)} - \mathbf{A}^\top \, \mathbf{F}^\top \, \boldsymbol{\epsilon}$$

while the variables without indexation, i.e., $\boldsymbol{\epsilon}, \mathbf{d}, \mathbf{F}, \mathbf{H}$, are computed from scratch in each iteration.

The stopping rule is not critical; we adopted a condition demanding that the new iteration not change the spot curve at any knot by more than a given small value.

2.3 The weighting matrices

The weighting matrix $\mathbf{B}$ plays an important role in fulfilling the stability conditions described above. The smoothing part of $\mathbf{B}$ is simply the identity matrix $\mathbf{I}$; now we add two more matrices, each controlling one condition.

From (4), it follows that

$$f(0) = f_N + \sum_{j=1}^{N} b_j t_j \, .$$

For the condition $f(0) = f_N$ to hold, it is necessary that the product $\mathbf{t}^\top \mathbf{b}$ be zero or attain some small value negligible with respect to the yield curve precision. In the minimization procedure this can be achieved by requiring that $\mathbf{b}^\top \, \mathbf{t} \, \mathbf{t}^\top \mathbf{b}$ tend to zero, or equivalently, that the weighting matrix is of the form

$$\mathbf{B} = \mathbf{I} + c_t \, \mathbf{t} \, \mathbf{t}^\top$$

where c_t is a coefficient sufficiently large to ensure that the difference between $f(0)$ and f_N is relatively small (in practice, tenths of the basis point).

The second condition, the zero derivative at t_{N-1}, can be treated similarly. This condition is equivalent to the equality of the spot and forward curves at t_{N-1}, $s(t_{N-1}) = f(t_{N-1})$. From (2) and (6) it follows that

$$s(t_{N-1}) - f(t_{N-1}) = \frac{f_N t_{N-1} + (\mathbf{A}\,\mathbf{b})_{N-1}}{t_{N-1}} - f_{N-1}$$

The last row of $\mathbf{A}$ is simply $(\mathbf{t}^2)^\top/2$, i.e., the vector of squared knot times divided by 2; we notice that the last row but one reads

$$\mathbf{e}_{N-1}^\top \mathbf{A} = \frac{1}{2}\left((\mathbf{t}^2)^\top - (t_N - t_{N-1})^2 \mathbf{e}_N^\top\right),$$

where $\mathbf{e}_k$ is the unit basis vector. Employing this and (4)

$$s(t_{N-1}) - f(t_{N-1}) = \frac{1}{2}\left(\frac{(\mathbf{t}^2)^\top \mathbf{b} - (t_N - t_{N-1})^2\, b_N}{t_{N-1}}\right) - (t_N - t_{N-1}) b_N =$$

$$= \frac{1}{2}\left(\frac{(\mathbf{t}^2)^\top \mathbf{b} - (t_N^2 - t_{N-1}^2)\, b_N}{t_{N-1}}\right)$$

The left–hand side is zero, if the product $\mathbf{q}^\top \mathbf{b}$ with a vector $\mathbf{q}$ is zero, where

$$\mathbf{q}^\top = \frac{1}{2}(t_1^2, t_2^2, t_3^2, \ldots, t_{N-2}^2, t_{N-1}^2, t_{N-1}^2),$$

i.e., almost identical to the last row of $\mathbf{A}$, but with the last element changed so that it is equal to the last but one. In the same way, adding a second matrix to $\mathbf{B}$

$$\mathbf{B} = \mathbf{I} + c_t \mathbf{t}\,\mathbf{t}^\top + c_q \mathbf{q}\,\mathbf{q}^\top$$

would ensure the optimization with both constraints, to a sufficient precision that depends on c_t and c_q.

The weighting of the price errors, which is not shown in (7), should transform the price errors to the yield errors with respect to the estimated spot curve. Let a spot curve price an instrument with a price error ε, i.e., $\varepsilon = \sum_1^M F_i$. We want to find a (small) parallel shift of this curve so that the shifted curve render the price error equal to zero.

The discount factor at a knot t_k is $d_k = \exp(-s(t_k)t_k)$; a shift by σ will change it to

$$d_k^\sigma = \exp(-s_k + \sigma)t_k) \approx d_k(1 - \sigma t_k), \quad \text{so that } \sigma \approx \frac{\varepsilon}{\sum_1^M F_i\, t_i}\,.$$

The transformation from the price to the yield errors is carried out, as shown, by multiplying each ϵ_i by $1/\delta_i$, where $\boldsymbol{\delta} = \mathbf{F}\mathbf{t}$. The first term in (7) is now

$$\frac{1}{2}\sum_1^M \frac{\epsilon_i^2}{\delta_i^2}\,, \tag{9}$$

but it should be clear that the gradient and the Hessian are computed as if $\boldsymbol{\delta}$ were a constant vector. The weights $\boldsymbol{\delta}$ are updated after each iteration of the matrix $\mathbf{F}$.

The minimization of the yield errors instead of the price ones removes the dependence of the smoothing parameter β on the units in which the cash flows are quoted. The number M of processed instruments clearly affects β; the sum (9) should be normalized somehow. The dependence on M seems to be quadratic, but due to random nature of errors, we may suppose that the sum of the squared errors is linear in M. The penalty function, used in the minimization procedure, is finally

$$L(\mathbf{b}) = \sum_{1}^{M} \frac{\epsilon_i^2}{\delta_i^2} + \beta\, M\, \mathbf{b}^\top (\mathbf{I} + c_t \mathbf{t}\, \mathbf{t}^\top + c_q \mathbf{q}\, \mathbf{q}^\top) \mathbf{b}\,.$$

2.4 The splitting procedure

All this elementary calculus exposed above would be a useless exercise if we were not capable to accommodate real debt instruments. The procedure we propose is as follows:

Let there be a nominal cash flow T occurring at a time t, where

$$t_i \leq t \leq t_{i+1}\,.$$

The discount factor at t, as given by the estimated yield curve, is $d(t)$. We would like to split $F = d(t)T$ into two flows F_i and F_{i+1} in a way that the first two moments, the sum and the duration, remain unchanged

$$T\,d = F_i + F_{i+1}, \qquad T\,d\,t = F_i t_i + F_{i+1} t_{i+1}.$$

Consequently, the splitting of all cash flows conserves the pricing errors and durations of all instruments. The discount factor $d(t)$, $t_i \leq t \leq t_{i+1}$, is easily computed from the value at the knot d_i and the forward rates at t_i, t_{i+1}, by adding to the exponent the area of the trapezoid delimited by the forward rate from t_i to t

$$d(t) = d_i \exp\left(-(t - t_i) \left[f_i + \frac{k_i}{2}\,(t - t_i) \right] \right)\,,$$

where k_i stands for a slope as in (3).

Anyway, even if $\boldsymbol{\epsilon}$ remains unchanged, the vector $\mathbf{F}^\top \boldsymbol{\epsilon}$ does not, so the gradient and the Hessian change. To treat this complication analytically could be perhaps feasible, but it would be impractical for computational purposes. We have to rely on a two–stage iteration process: to update, after each N–R iteration, the splitting of all cash flows. In practice, this comes down to the recomputation of $\mathbf{F}$ after each iteration, which would be necessary even if the cash flows coincided with the knots. Intuitively, the splitting may be detrimental to the convergence and/or stability of the iterations. We, however, have not encountered numerical problems of this kind so far.

2.5 The fairness metrics

After many error–and–trial experiments, we have chosen as our measure of the YC fairness a ratio of the norms of the second and the first forward rate differences:

$$\varphi = \frac{\sum_2^k (b_i - b_{i-1})^2}{\sum_1^k b_i^2} + \frac{\sum_{k+1}^{N-1} (b_i - b_{i-1})^2}{\sum_k^{N-1} b_i^2} \tag{10}$$

It is a dimensionless quantity and can be seen as a sum of two Rayleigh quotients $\mathbf{c}^\top \mathbf{M} \mathbf{c} \,/\, \mathbf{c}^\top \mathbf{c}$, where $\mathbf{c}$ represents part of $\mathbf{b}$. The difference matrix $\mathbf{M}$ is tridiagonal, with the upper and lower diagonals filled with -1, and the main diagonal $(1, 2, 2, \ldots, 2, 1)$. The largest value of $\mathbf{c}^\top \mathbf{M} \mathbf{c} \,/\, \mathbf{c}^\top \mathbf{c}$ is then the value of the largest eigenvalue of $\mathbf{M}$, λ_{max}. From the Gershgorin circle theorem it follows that $\lambda_{max} \leq 4$; a numerical solution [7] gives values approaching this limit with an increasing rank of $\mathbf{M}$. For $\text{rank}(\mathbf{M}) = 6$ is $\lambda_{max} \approx 3.73$, for $\text{rank}(\mathbf{M}) = 12$ we get $\lambda_{max} \approx 3.93$. The partition into two quotients separates measures of roughness in the short and the long parts of the curve; we set $k = 6$, i.e., the short part is delimited by three years.

The value of φ lies therefore within the interval $\langle 0, 7.7 \rangle$. The greater the value of φ, the less fair is the curve (perhaps φ should be named as "unfairness"); the difficult question remains, however, how to fix its value for the curve to look sufficiently fair. In the Czech bond market, we have found that values of φ not greater than 3 lead to a subjectively acceptable fairness; for $\varphi > 1$ the YC gives price errors comparable with the bid/ask spreads quoted by market makers.

From a technical point, a requested value of φ is achieved by iterating the value of β, while repeating the estimation of the YC. This is already the third level of iteration: the N–R algorithm, the update of cash flow splitting and error weights, the update of β. The programming becomes a bit tricky, especially if we do not want to sacrifice the computing speed. Anyway, the good message is that, for all ends and purposes, the YC iteration can be stopped if a change in subsequent values does not exceed some tenths of basis point.

3 Case study

As a practical example of the proposed YC estimation, we have chosen the Czech government bond market, with which we are technically acquainted. The number of bonds and the range of maturities are rather sufficient; a worse situation prevails in the market liquidity and the number of quoting agents. These shortcomings result in large bid/ask spreads and price stagnation, as can be seen from the available fixing source [8].

Anyway, the story of the Czech government yield curves in the two recent years is interesting in itself. In November 2013, the Czech central bank made an idiosyncratic move to artificially weaken the CZK exchange rate by long–term currency interventions, in a good intention to revive the Czech economy (and to embellish their inflation target, too). Leaving aside the adage that "the road to hell is paved with good intentions", we note that the effect on the Czech YC has been dramatic. Some two years later, a squeezing due to a speculative capital invested into the short–term Czech bonds led to negative interest rates up to five years. The 10y benchmark spread against Germany has lowered to units of bp, a fact hardly explainable from the fundamental point of view.

A graph of three estimated YCs in this period is plotted in Fig.1. The bond portfolios on the particular dates contain all fixed–coupon government bonds quoted at the MTS fixing [8]. The YCs in Fig.1 are computed from mid prices, with added accrued interest as of the settlement date T+2. The yield at $t = 0$ is taken from the money market and equals the one–day PRIBOR mid–rate.

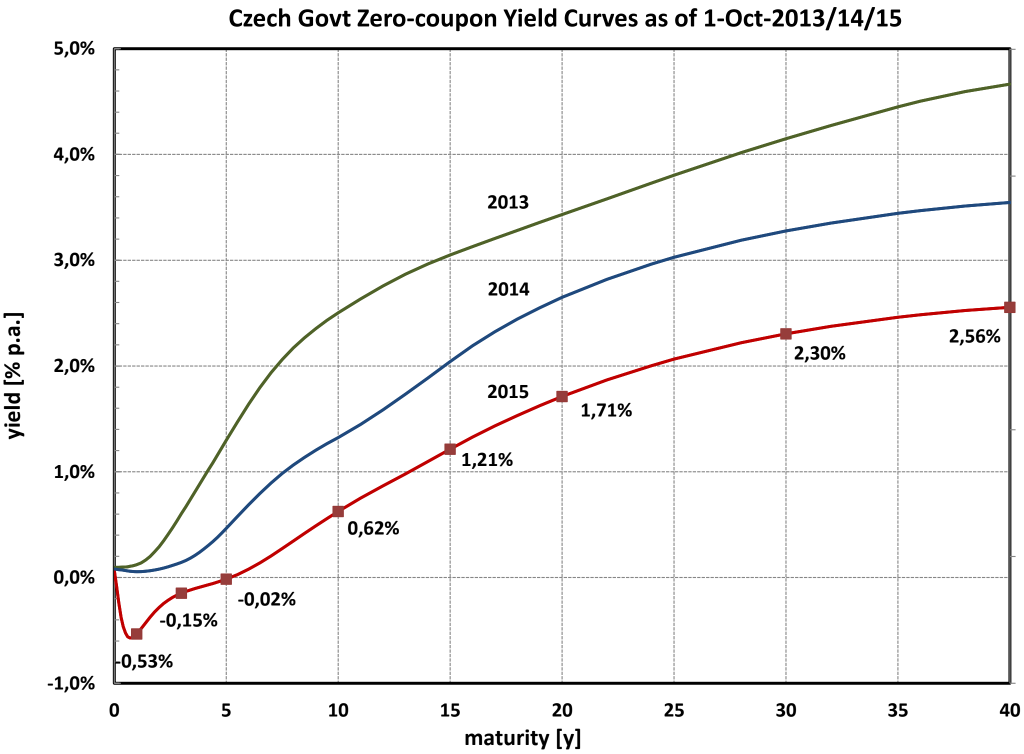

Figure 1: Estimated yield curves at the indicated dates. Fairness parameter $\varphi = 2$.

Negative interest rates, a recent phenomenon, lead to an uncommon shape of the last YC at the start. It can be argued that some artificial zero rate would change that for the better. We maintain, vindicated by our experiments with a modified estimation procedure where not only $\mathbf{b}$, but also the zero rate was estimated, that the cash paid for the bonds is provided by the money market and that predicting or estimating something clearly within our good knowledge is not a good idea. The level of fairness deemed subjectively acceptable lies somewhere between 1 and 2. An example of recent YCs and the effect of our fairness metrics (10) are seen in Fig.2.

As for the estimation errors, we have to distinguish between the price and the yield errors. The former are the differences between the gross price and the price generated by

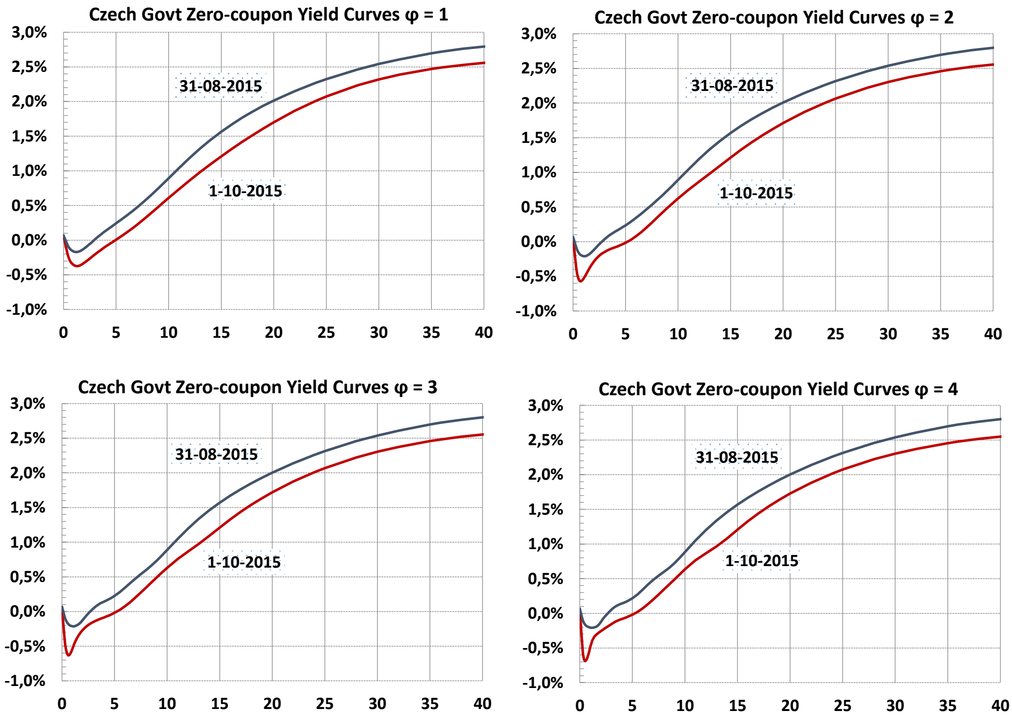

Figure 2: The effect of different fairness levels on estimated curves.

the YC, the latter the parallel shifts of the YC to zero the price error out. The former are in units of price quotation, the latter in yields p.a. We remind that the optimization procedure strives to minimize the sum of squared yield errors.

To assess price errors, consideration should be given to quoted bid-ask spreads, because mid prices, used in the estimation, cannot be but less reliable if the spreads are larger. We have made use of a mean error-to-spread (MES) ratio:

$$MES = \frac{2}{M} \sum_{1}^{M} \frac{|\epsilon_k|}{q_k^{ask} - q_k^{bid}}$$

where q^{bid} , q^{ask} are bid, ask quotations. The value of MES is dimensionless; $MES = 1$ means that the YC price equals, on average, either the quoted bid or ask price. The lower the value of MES, the closer the prices generated by the YC lie, on average, to the center of the bid–ask interval.

Numerical values of errors related to two YCs plotted in Fig.1 can be found in Tab.1. The differences between the curves are substantial: a "normal" shape of the YC as of 2014 contrasts with a more complicated and "unnatural" curve of 2015. In terms of price errors, as measured by MES, the precision of the former is almost twice as good; the mean absolute error (MAE) of the yields of all bonds with remaining maturity longer than one year is more than three times smaller. (This can be tracked down to a large yield error at the shortest maturities that dominates the sum of squared yield errors; the curve has to be a bit oversmoothed to accommodate a sharp dip close to the zero time.)

Date	1–Oct–2014				1–Oct–2015			
φ	1	2	3	4	1	2	3	4
MES	0.81	0.35	0.36	0.36	0.78	0.59	0.59	0.59
MAE $0-50$ y (bp)	1.3	0.6	0.6	0.6	6.9	5.2	4.9	4.4
MAE $1-50$ y (bp)	1.4	0.6	0.6	0.6	3.3	2.2	2.2	2.4

Table 1: Price and yield errors of two yield curves at indicated dates

At least in these two cases, we see that a level of φ greater than 2 does not bring substantial improvements in the precision of the estimate. Also, in the case of a normal YC shape, the mean yield error of 0.6 bp is quite satisfactory, while 2 bp in the second case, given the current secondary Czech bond market, tolerable.

4 Conclusion

We have presented a viable approach to the estimation of the zero-coupon yield curve from coupon instruments. The design and debugging of the underlying program have taken a long time, since our endeavor has been extensive rather than intensive, and is continuing in the same manner. Even if we have carried out a considerable amount of testing and simulations, the algorithm, and in particular, its implementation, may turn out to suffer from errors we have not been able to detect.

This is the reason we intend to offer the computer program, which can be run on any standard PC or NTB with installed MS Excel, as a freeware, expecting some feedback in case of user problems. Run times in the examples shown above were about 200 ms on a standard Lenovo Yoga NTB; when only intraday price updates are processed, the times are much shorter. So, such estimation can be run as a real-time application.

For further development, we see currently two issues. The first is purely technical - the integrated spline matrix **A** allows a simple LU decomposition and we could take advantage of it to improve the N–R convergence speed. The second is methodological: the part of the **B** matrix responsible for smoothing, now the identity matrix, may play a role in defining the shape of the curve at long maturities when there are few or no cashflows. Clearly, this approach could not be considered purely nonparametric, as the shape of the curve at long maturities would be determined solely by the method itself rather than by the (missing) data.

References

[1] Nelson, C.R., Siegel, A.F., *Parsimonious modeling of yield curves*, Journal of Business, 60(4), (1987), pp. 473–489.

[2] Acton F. S., *Numerical Methods that Work*, Mathematical Association of America, Washington D.C., (1990)

[3] BIS Papers, *Zero–coupon yield curves: technical documentation*, Monetary and Economic Department, No 25. (2005)

[4] Waggoner D.F., *Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices*, "Working Paper", No. 97–10, Federal Reserve Bank of Atlanta, Atlanta, (1997)

[5] Tanggaard, C., *Nonparametric Smoothing of Yield Curves*, Review of Quantitative Finance and Accounting, No. 9 (1997), pp. 251–267.

[6] Levien R., Séquin C.H., *Interpolating Splines: Which is the fairest of them all?*, Computer–Aided Design and Applications, 6(1) (2009), pp. 91–102.

[7] http://www.bluebit.gr/matrix-calculator/calculate.aspx

[8] https://www.mtsdata.com/content/data/public/cze/fixing/index.php?