



NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Victor Lapshin, Vadim Kaushanskiy

A NONPARAMETRIC METHOD FOR TERM STRUCTURE FITTING WITH AUTOMATIC SMOOTHING

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: Financial Economics
WP BRP 39/FE/2014

A Nonparametric Method for Term Structure Fitting with Automatic Smoothing

We present a new nonparametric method for fitting the term structure of interest rates from bond prices. Our method is a variant of the smoothing spline approach, but within our framework we are able to determine the smoothing coefficient automatically from the data using generalized cross-validation or maximum likelihood estimates. We present an effective numerical algorithm to simultaneously find the term structure and the optimal smoothing coefficient. Finally, we compare the proposed nonparametric fitting method with other parametric and nonparametric methods to show its superior performance.

JEL Classification: G12, C14.

Key words: regularization, smoothing splines, term structure of interest rates.

¹ National Research University Higher School of Economics. Laboratory for Financial Engineering and Risk Management. Research Fellow; E-mail: vlapshin@hse.ru

² National Research University Higher School of Economics. Laboratory for Financial Engineering and Risk Management. Research Assistant; E-mail: vkaushanskiy@hse.ru

1 Introduction

The term structure of interest rates is fundamental to many areas of financial analysis: pricing and hedging derivatives, risk management, project analysis, accounting, actuarial science. While all these areas use the term structure of interest rates, different notions are (or should be) assumed by these words. For example, before the recent financial crisis, in pricing and hedging derivatives were used swap rates. Nowadays to price or hedge a derivative instrument correctly, a whole range of different interest rates are modelled: Libor (or Euribor), Interest Rate Swap rates, Overnight Indexed Swap rates, repo rates, bond rates, and more. Nevertheless, for some other applications we still need the term structure of bond rates. For example, pension funds are usually legally restricted in their investments, so they may find the term structure of (government) bond rates the interest rate of their choice. Unfortunately, the term structure of bond rates is not readily available on the market. In this paper we deal with the problem of estimating it from the market data.

Different applications require different approaches to interest rate modelling. Pricing and hedging derivatives is usually done via stochastic interest rate models, which may describe the risk-neutral stochastic evolution of the short rate and/or other variables such as its variance, long rate or even the whole yield curve (as in the Heath-Jarrow-Morton framework). However, these models are designed to model the short-term fluctuations of interest rates and either imply a non-realistic term structure or assume that the term structure is directly observable on the market at any moment. Risk management applications also usually employ stochastic models (although, no more risk-neutral) to sample from a distribution of future rates in order to assess risks, perform stress-testing and so on. Once again, the stochastic properties of the model come in front of its term structure fitting possibilities, except when a significant part of the portfolio is comprised of bonds, which may be the case for pension funds, life insurance companies, and fixed-income hedge funds. However, fitting the term structure to market data is most useful when performing fair value pricing for accounting purposes or actuarial assessments. Of course, for routine marking to market and setting trading limits, the fitting procedure has to be modified to ensure temporal stability, as for these applications stability and simplicity is usually preferable over quality and precision. That said, we consider a single-time fitting of the term structure to a snapshot of the bond market. This problem naturally arises during investment analysis, monetary decision making,

periodic and occasional reporting and various assessments, external auditing, and cases when the data sample is limited to a snapshot of the market or when we have reasons to believe that the market has changed completely since the last time we performed this calculation, so historic data is irrelevant. Generalizing the setting to continuous everyday fitting poses many new problems, which we partially address in ongoing research. From now on, we consider only snapshot term structure fitting.

Term structure estimation has a long history, of which we limit ourselves to only the key milestones. For a more detailed exposition we refer the reader to the review article by Schmidt (2011) or to specialized treatises by James and Webber (2000) and by Andersen and Piterbarg (2010). It should be noted again that we do not consider stochastic (dynamic) interest rate models here, these are used mainly for pricing/hedging derivatives and risk management. A recent article by Jarrow (2014) presents an overview of dynamic term structure fitting methods. Before reviewing the literature we need to introduce some notation. We use the words 'term structure' to denote either the discount function $d(t)$, the zero-coupon yield $r(t)$ or the instantaneous forward rate $s(t)$. Under continuous compounding convention, they are linked by a simple relationship:

$$d(t) = \exp[-r(t)t] = \exp\left[\int_0^t s(\tau)d\tau\right].$$

The term structure estimation will mean obtaining an estimate of any of these functions. The existing methods of the term structure estimation may be divided into three main groups.

1. Ad hoc methods. These are techniques designed to obtain a plausible result, but sometimes lacking financial soundness, intuition, explanation, exploiting a unique idea, or estimating something else. In the beginning, people used to draw the dots on the (t, r) -plane: one for each bond, its term to maturity and yield, and connect them with a smooth line by hand. Such a curve was called a 'yield curve' and could give a general notion of the interest rate term structure. The bootstrap method of Fama (1987), kernel smoothing methods by Linton, Mammen, Nielsen, and Tanggaard (2001) are other examples of this family.

Before introducing the other two groups, we note that the term structure estimation problem is ill-posed, because the input data, whatever it is (we come to this later), is finite-dimensional

while the term structure is an infinite-dimensional object (a function). Therefore to estimate it, a priori assumptions are made to regularize the problem in some way. According to the nature of these a priori assumptions or regularizations, we divide the methods into two classes.

2. Parametric methods. One may assume that the unknown function $d(\cdot)$, $r(\cdot)$ or $s(\cdot)$ has a known parametric form (belongs to a finite-dimensional manifold). Once this assumption is made and given the number of parameters is sufficiently low, the problem becomes well-posed and may be solved usually in the least squares sense (although we'll return to this point later). The key examples of parametric families used for term structure fitting are by Nelson and Siegel (1987) and Svensson (1994), which are widely used.
3. Spline methods. Instead of assuming a specific parametric form, one may assume that the unknown function possesses an extreme property of some kind: minimum length or curvature, minimum potential energy, etc. Since this kind of regularization usually results in some kinds of splines, we call these techniques spline methods. However if we just search for the solution in the form of a particular spline this should be considered as a parametric method, since we simply assume a specific parametric form of a piecewise defined function. Splines were used for term structure estimations by McCulloch (1971, 1975); Vasicek and Fong (1982) and many others.

Since our proposed method belongs to the class of spline methods, we continue reviewing the literature regarding spline methods. Adams and van Deventer (1994) used smoothing splines, Fisher, Nychka and Zervos (1995) introduced automatic choice of the smoothing parameter, Waggoner (1997) had the smoothing parameter depend on t – the term to maturity. Smirnov and Zakharov (2003) exposed the methodologically correct routine by starting from an assumption about an extreme property and arriving at an exponential-sinusoidal spline via formal mathematical methods.

Since the introducing of the STRIPS program into the US bond market, estimating the term structure has become a much more simple task and has gradually faded out of the view of US researchers. However, researchers in other countries, especially in those having less liquid bond markets, have been trying to devise a perfect estimation method. We mention several of these works. Gimeno (2009) employ genetic algorithms to fit Nelson-Siegel and Svensson parametric forms to Spanish bond data,

Sanchez (2004) use fuzzy mathematics to obtain an estimate of the discount function, Lin (2002) uses B-splines to fit Taiwanese interest rates. Andersen (2008) uses tension splines, Hagan and West (2006) compare different spline and interpolation techniques for a model dataset. Laurini and Hotta (2010) introduces Bayesian a extension of a dynamic parametric method of Diebold and Li (2006) and uses it on Brazilian data. There is a general understanding that parametric methods may do well in developed bond markets such as in US or Europe but they lack flexibility and the ability to reflect complex term structures usually found in developing markets, where spline methods are usually used.

Unfortunately, almost all spline methods except the one by Fisher, Nychka and Zervos (1995) (FNZ) require manual (expert) choice for the smoothing parameter. We present a new spline term structure estimation method, which does not require manual tuning and compare it to FNZ and some parametric alternatives.

The rest of the article is organized as follows. Section 2 introduces the rest of the notation and formulate the mathematical problem. Section 3 solves the mathematical problem to obtain a spline. Sections 4 and 5 present a numerical procedure to estimate both the spline coefficients and the regularization parameter. Finally, Section 6 describes our data and compare our method with several widespread parametric families and with FNZ, which is also a fully automatic spline method. Section 7 concludes.

2 Problem Formulation

Now suppose that there are N bonds traded on the market, with prices $P_k, k = 1, \dots, N$, paying $F_{i,k}, i = 0, \dots, n, k = 1, \dots, N$, at times t_i respectively. Note that in a typical situation most of the $F_{i,k}$ will be zeros since payment times usually do not coincide.

We assume that any coupon bond is equivalent to a portfolio of zero-coupon bonds with appropriate notionals and maturities. Therefore we can express the theoretical price of the $k - th$ bond, P_k^* :

$$P_k^* = \sum_{i=0}^n d(t_i) F_{i,k},$$

where $d(t)$ is the (unknown) discount function.

A discount function must satisfy several constraints in order to be economically reasonable.

1. $d(t)$ should be non-strictly decreasing on its domain;
2. $d(t) > 0$;
3. $d(0) = 1$.

Therefore, the problem is to find the discount function $d(t)$ on $[0, t_n]$ satisfying the conditions above and such that.

$$\sum_{i=0}^n d(t_i) F_{i,k} \approx P_k, k = 1, \dots, N,$$

where the latter is usually considered in the least squares sense, although we consider an improved version in this paper. For now we just state that the function $d(t)$ minimizes

$$J_1 = \sum_{k=1}^N \left(P_k - \sum_{i=1}^n F_{i,k} d(t_i) \right)^2.$$

Observe that this problem is ill-posed. Actually, only values $d(t_i)$ can be determined from these conditions.

Principles of solving ill-posed problems are described by Tikhonov and Arsenin (1979). According to these principles, we suppose that the sought term structure is smooth, more precisely, that it minimizes

$$J_2 = \int_0^T \xi(s^{[p]}(\tau)) d\tau,$$

where $T = t_n$, $s(t)$ is the spot forward rate for the time t , p is the derivative order and $\xi(\cdot)$ is some pre-specified transformation function.

To satisfy the conditions stated above, we consider the following change of variables:

$$d(t) = \exp \left(- \int_0^t F(f(\tau)) d\tau \right), \quad (1)$$

where $f = \xi(s)$ is the unknown and $F = \xi^{-1}$ is a specified function.

Plugging this into J_1 and adding J_2 , we get that the unknown function $f(t)$ is the solution of the following minimization problem:

$$J_1(f) + \alpha J_2(f) = \sum_{k=1}^N \left(P_k - \sum_{i=1}^n F_{i,k} e^{-\int_0^{t_i} F(f(\tau)) d\tau} \right)^2 + \alpha \int_0^T (f^{[p]})^2(\tau) d\tau \rightarrow \min_{f(\cdot)}, \quad (2)$$

where α is the regularization parameter governing the interplay between precision (J_1 term) and smoothness (J_2 term). The regularization parameter is usually set exogenously by an human expert. In this article we present an algorithm to determine its value from the data.

3 Solving the minimization problem

In the previous section we arrived at the problem

$$\sum_{k=1}^N \left(\sum_{i=0}^n F_{ik} d(t_i) - P_k \right)^2 + \alpha \int_0^T (f^{[p]})^2(\tau) d\tau \rightarrow \min_{f(\cdot)}, \quad (3)$$

where

$$d(t_i) = \exp \left(- \int_0^{t_i} F(f(\tau)) d\tau \right).$$

For simplicity consider $G(x) = x^2$ in this section. While further we will extend our solution for more general case.

Let

$$c_i = \int_{t_{i-1}}^{t_i} f^2(\tau) d\tau, i = 1, \dots, n. \quad (4)$$

Notice that this substitution allows us to split our problem into 2 subproblems:

1. Given fixed $\{c_i\}$ (this fixes J_1), find the function f_{c_1, \dots, c_n} minimizing J_2 . This is a conditional minimization problem.
2. Find $\{c_i\}$, for which f_{c_1, \dots, c_n} minimizes the functional $J_1 + \alpha J_2$. This is a finite-dimensional

optimization problem.

Note that the set $\{c_i\}$ determines $d(t)$ in partition points t_1, \dots, t_n while the regularization is responsible for the values of $d(t)$ at intermediate points.

First of all, we solve the minimization problem (3) with the condition (4). The Lagrangian is

$$L = \alpha \int_0^T (f^{[p]})^2(\tau) d\tau + \sum_{i=1}^n \lambda_i \left(\int_{t_{i-1}}^{t_i} F(f(\tau)) d\tau - c_i \right). \quad (5)$$

A necessary condition for an extremum is:

$$\sum_{i=1}^n \left(2\alpha \int_{t_{i-1}}^{t_i} f^{[p]}(\tau) h^{[p]}(\tau) d\tau + \lambda_i \int_{t_{i-1}}^{t_i} F'(f(\tau)) h(\tau) d\tau \right) = 0, \quad (6)$$

for any p times continuously differentiable function $h(t)$.

Further, for convenience, let $f_i(t)$ be $f(t)$ restricted to the segment $[t_{i-1}, t_i]$. Integrating by parts, we have:

$$\begin{aligned} \alpha \sum_{i=1}^n \left(\sum_{s=1}^p (-1)^s \left(f_i^{[p+s-1]}(t_i - 0) h^{[p-s]}(t_i) - f_i^{[p+s-1]}(t_{i-1} + 0) h^{[p-s]}(t_{i-1}) \right) + \right. \\ \left. + \int_{t_{i-1}}^{t_i} \left((-1)^p f_i^{[2p]}(\tau) + \frac{\lambda_i}{2\alpha} f_i(\tau) \right) h(\tau) d\tau \right) = 0. \quad (7) \end{aligned}$$

Assuming $f_i \in C^{2p}[t_{i-1}, t_i]$,

$$f_i^{[s]}(t_i - 0) = f_{i+1}^{[s]}(t_i + 0), i = 2, \dots, n-1, s = 1, \dots, p \quad (8)$$

$$f_1^{[s]}(0 + 0) = f_n^{[s]}(T - 0) = 0, s = 1, \dots, p \quad (9)$$

$$f_i^{[2p]}(t) = (-1)^{p+1} \frac{\lambda_i}{2\alpha} F'(f_i(t)), t \in (t_{i-1}, t_i), i = 1, \dots, n. \quad (10)$$

In general case, the analytic solution stops here. However, we present a simple solution for $p = 1, F(x) = x^2$ for illustration purposes.

This minimization problem could be addressed via the Pontryagin maximum principle along the lines of Lapshin (2009). However, since an analytic solution is presented only for illustration, we would like to keep it simple.

Let $F(x) = x^2, p = 1$.

$$f'_i(t_i - 0) = f'_{i+1}(t_i + 0), i = 2, \dots, n - 1, \quad (11)$$

$$f'_1(0 + 0) = f'_n(T - 0) = 0, \quad (12)$$

$$f''_i(t) = \frac{\lambda_i}{\alpha} f_i(t), t \in (t_{i-1}, t_i), i = 1, \dots, n. \quad (13)$$

Solving this system, we get

$$f_i(t) = \begin{cases} C_i^1 \sin \sqrt{-\gamma_i t} + C_i^2 \cos \sqrt{-\gamma_i t}, \gamma_i < 0, \\ C_i^1 \exp(\sqrt{\gamma_i t}) + C_i^2 \exp(-\sqrt{\gamma_i t}), \gamma_i > 0, \\ C_i^1 + C_i^2 t, \gamma_i = 0, \end{cases} \quad (14)$$

where $\gamma_i = \frac{\lambda_i}{\alpha}$, and C_i^1, C_i^2 are constants.

Notice that C_i^1, C_i^2, λ_i depend on c_1, \dots, c_n . Consequently, we have that $f(t)$ has the form of an sinusoidal-exponential spline (14). Spline coefficients C_i^1, C_i^2, λ_i can be determined from $\{c_i\}$ by solving $3n$ equations.

$$\int_{t_{i-1}}^{t_i} f^2(\tau) d\tau = c_i, i = 1, \dots, n \quad (15)$$

$$f_{i+1}(t_i + 0) = f_i(t_i - 0), i = 1, \dots, n - 1 \quad (16)$$

$$f'_{i+1}(t_i + 0) = f'_i(t_i - 0), i = 1, \dots, n - 1 \quad (17)$$

$$f'_1(0 + 0) = f'_n(T - 0) = 0. \quad (18)$$

The rest is a finite-dimensional optimization problem. This is not the most computationally efficient way to address this particular problem, however it illustrates the spline nature of the method.

4 Numerical method

For arbitrary p and F an analytic solution cannot be found, therefore we consider a numerical method.

Let

$$\begin{cases} J_1 = \sum_{k=1}^N \left(\sum_{i=1}^n F_{ik} e^{-\int_0^{t_i} F(f(\tau)) d\tau} - P_k \right)^2, \\ J_2 = \int_0^T (f^{(p)})^2(\tau) d\tau, \end{cases}$$

in the following minimization problem:

$$J_1 + \alpha J_2 \rightarrow \min_{f(\cdot)}. \quad (19)$$

For convenience we introduce the following notation:

$$\mathcal{N}_k f = \sum_{i=1}^n F_{ik} e^{-\int_0^{t_i} F(f(\tau)) d\tau}, \quad (20)$$

\mathcal{N}_k is an operator mapping the curve space W to R .

We construct an iterative minimization process as follows. Let the previous approximation of $f(t)$ be denoted as $f_0(t)$ and let $\mathcal{D}_k = \left. \frac{\partial \mathcal{N}_k}{\partial f} \right|_{f=f_0}$ be the Fréchet derivative of the operator \mathcal{N}_k at the point f_0 . \mathcal{D}_k is a linear operator mapping $W \rightarrow R$, such that for any function $f(\cdot) \in W$

$$\mathcal{N}_k f - \mathcal{N}_k f_0 = \mathcal{D}_k(f - f_0) + r_0(f, f_0),$$

where the reminder $r_0(f, f_0)$ satisfies

$$\frac{||r_0(f, f_0)||}{||f - f_0||} \rightarrow 0, \text{ when } ||f - f_0|| \rightarrow 0.$$

Let

$$\mathcal{H}_i f = \int_0^{t_i} F(f(t)) dt$$

By the chain rule,

$$\mathcal{D}_k f = - \sum_i F_{ik} e^{-\mathcal{H}_i f_0} \frac{\partial \mathcal{H}_i}{\partial f} \Big|_{f=f_0}. \quad (21)$$

The derivative of \mathcal{H}_i may be calculated via

$$\begin{aligned} \mathcal{H}_i f - \mathcal{H}_i f_0 &= \int_0^{t_i} (F(f(\tau)) - F(f_0(\tau))) d\tau = \int_0^{t_i} (F'(f_0(\tau))(f(\tau) - f_0(\tau)) + o(f(\tau) - f_0(\tau))) d\tau = \\ &= \int_0^{t_i} F'(f_0(\tau))(f(\tau) - f_0(\tau)) d\tau + o(\|f - f_0\|). \end{aligned}$$

Consequently,

$$\frac{\partial \mathcal{H}_i}{\partial f} \Big|_{f=f_0} = \int_0^{t_i} F'(f_0(\tau)) f(\tau) d\tau.$$

Plugging this into (21), we obtain the expression for the operator \mathcal{D}_k :

$$\mathcal{D}_k f = - \sum_{i=1}^n e^{-\int_0^{t_i} F(f_0(\tau)) d\tau} F_{ik} \int_0^{t_i} F'(f_0(\tau)) f(\tau) d\tau. \quad (22)$$

By the Riesz representation theorem, for a bounded linear operator \mathcal{D}_k there is a unique $\eta_k \in W$, s.t. for any $f \in W$ $\mathcal{D}_k f = \langle \eta_k, f \rangle$ and also $\|\mathcal{D}_k\| = \|\eta_k\|$.

Let's find η_k :

$$- \sum_{i=1}^n e^{-\int_0^{t_i} F(f_0(\tau)) d\tau} F_{ik} \int_0^{t_i} F'(f_0(\tau)) h(\tau) d\tau = \langle \eta_k, h \rangle, \forall h \in W. \quad (23)$$

Let W be equipped with the following scalar product:

$$\langle \eta_k, h \rangle = \sum_{s=0}^{p-1} \eta_k^{(s)}(0) h^{(s)}(0) + \int_0^T \eta_k^{(p)}(\tau) h^{(p)}(\tau) d\tau.$$

Statement 1. $\eta_k(t)$ has the form:

$$\begin{aligned} \eta_k(t) = & - \sum_{i=1}^n \exp \left(- \int_0^{t_i} F(f_-(\tau)) d\tau \right) F_{ik} \times \\ & \times \int_0^{t_i} F'(f_-(\tau)) \left(\sum_{s=0}^{p-1} \frac{t^s \tau^s}{(s!)^2} + \int_0^T \frac{(t-u)_+^{p-1} (\tau-u)_+^{p-1}}{(p-1)!^2} du \right) d\tau \end{aligned} \quad (24)$$

Proof. Write the Taylor expansion for $f(t) \in W$ with a reminder in the integral form:

$$f(t) = \sum_{s=0}^{p-1} \frac{t^s}{s!} f^{(s)}(0) + \int_0^T \frac{(t-u)_+^{p-1}}{(p-1)!} f^{(p)}(u) du. \quad (25)$$

The proof is done via direct verification. Compute $\eta_k^{(s)}$:

$$\begin{aligned} \eta_k^{(s)}(0) &= \sum_{i=1}^n \exp \left(- \int_0^{t_i} F(f_0(\tau)) d\tau \right) F_{ik} \int_0^{t_i} F'(f_0(\tau)) \frac{\tau^s}{s!} d\tau, \quad s = 0, \dots, p-1, \\ \eta_k^{(p)}(\tau) &= \sum_{i=1}^n \exp \left(- \int_0^{t_i} F(f_0(\tau)) d\tau \right) F_{ik} \int_0^{t_i} F'(f_0(\tau)) \frac{(\tau-u)_+^{p-1}}{(p-1)!} du. \end{aligned}$$

Consequently,

$$\begin{aligned} \langle \eta_k, h \rangle &= \sum_{s=0}^{p-1} \sum_{i=1}^n \exp \left(- \int_0^{t_i} F(f_0(\tau)) d\tau \right) F_{ik} \int_0^{t_i} F'(f_0(\tau)) \frac{\tau^s}{s!} d\tau h^{(s)}(0) + \\ &+ \int_0^T \sum_{i=1}^n \exp \left(- \int_0^{t_i} F(f_0(\tau)) d\tau \right) F_{ik} \int_0^{t_i} F'(f_0(\tau)) \frac{(\tau-u)_+^{p-1}}{(p-1)!} du h^{(p)}(\tau) d\tau = \\ &= \sum_{i=1}^n \exp \left(- \int_0^{t_i} F(f_0(\tau)) d\tau \right) F_{ik} \left(\sum_{s=0}^{p-1} \int_0^{t_i} F'(f_0(\tau)) \frac{\tau^s}{s!} d\tau h^{(s)}(0) + \right. \\ &\quad \left. + \int_0^T \int_0^{t_i} F'(f_0(\tau)) \frac{(\tau-u)_+^{p-1}}{(p-1)!} du h^{(p)}(\tau) d\tau \right) \end{aligned} \quad (26)$$

Function $F'(f_0(\tau)) \frac{(\tau-u)_+^{p-1}}{(p-1)!} h^{(p)}(\tau)$ is integrable on $[0, T] \times [0, t_i]$, therefore by the Fubini theorem the

order of integration in the second term can be changed.

$$\begin{aligned}
& \sum_{i=1}^n \exp \left(- \int_0^{t_i} F(f_0(\tau)) \right) F_{ik} \left(\sum_{s=0}^{p-1} \int_0^{t_i} F'(f_0(\tau)) \frac{\tau^s}{s!} d\tau h^{(s)}(0) + \right. \\
& \quad \left. + \int_0^T \int_0^{t_i} F'(f_0(\tau)) \frac{(\tau-u)_+^{p-1}}{(p-1)!} du h^{(p)}(\tau) d\tau \right) = \\
& = \sum_{i=1}^n \exp \left(- \int_0^{t_i} F(f_0(\tau)) \right) F_{ik} \int_0^{t_i} F'(f(\tau)) \left(\sum_{s=0}^{p-1} \frac{\tau^s}{s!} h^{(s)}(0) + \int_0^T \frac{(\tau-u)_+^{p-1}}{(p-1)!} h^{(p)}(u) du \right) d\tau \quad (27)
\end{aligned}$$

Plugging (25) in, we arrive at the proof. \square

Now that we have found the operators $\mathcal{D}_k : \mathcal{D}_k f = \langle \eta_k, f \rangle$, we may describe the iterative minimization process.

4.1 Iteration description

For convenience we denote $\lambda = \frac{\alpha}{N}$. Then rewrite minimization problem:

$$\sum_{k=1}^N (\mathcal{N}_k f - P_k)^2 + N\lambda \int_0^T (f^{(p)})^2(\tau) d\tau \rightarrow \min_{f(\cdot)}. \quad (28)$$

Let us consider the linear approximation of operators \mathcal{N}_k :

$$\mathcal{N}_k f \approx \mathcal{N}_k f_- + \mathcal{D}_k(f - f_-),$$

where f_- is still the current approximation to the unknown function f .

So, instead of (28) we consider the following partially linearized problem

$$\sum_{k=1}^N (\mathcal{D}_k f - y_k)^2 + N\lambda \int_0^T (f^{(p)})^2(\tau) d\tau \rightarrow \min_{f(\cdot)}, \quad (29)$$

where $y_k = P_k - \mathcal{N}_k f_- + \mathcal{D}_k f_-$.

Now we find the minimizer of (29).

For applying the theory of Wahba (1990), we break the space W into the direct sum of subspaces $W = W^0 + W^1$, where

- W^0 is the subspace of polynomials, whose degree is less or equal than $p - 1$. Functions $\varphi_j(t) = \frac{t^j}{j!}$, $j = 0, \dots, p - 1$ form a basis of W^0 .
- W^1 is the subspace of functions satisfying $f(0) = f'(0) = \dots f^{(p-1)}(0) = 0$.

As \mathcal{D}_k is a bounded linear operator, the following theorem, proved by Wahba (1990) (Theorem 1.3.1, p. 11) is applicable:

Theorem 4.1. *The solution of (29) $f_\lambda(t)$ has the form:*

$$f_\lambda(t) = \sum_{j=1}^p d_j \varphi_j(t) + \sum_{k=1}^N c_k \xi_k(t), \quad (30)$$

where

$$\mathbf{d} = (d_1, \dots, d_p)' = (T' M^{-1} T)^{-1} T' M^{-1} y, \quad (31)$$

$$\mathbf{c} = (c_1, \dots, c_N)' = M^{-1} (I - T(T' M^{-1} T)^{-1} T' M^{-1}) y, \quad (32)$$

$$T = \{\mathcal{D}_k \varphi_j\}, k = 1, \dots, N, j = 1, \dots, p, \quad (33)$$

$$M = \Sigma + N \lambda I, \quad (34)$$

$$\Sigma = \{\langle \xi_i, \xi_j \rangle\}, \quad (35)$$

$$\varphi_j(t) = \frac{t^j}{j!}, \quad (36)$$

$$\xi_k = P_1 \eta_k, \quad (37)$$

and P_1 is the orthogonal projector from W onto W^1 .

Using (31) and (32), we can find the solution $f_\lambda(t)$ to the problem (29).

After this we consider a new linearization in the vicinity of a new current approximation and repeat the iterative process for finding the minimizer of the problem (19) using new linearizations of operators \mathcal{N}_k .

This procedure is an analog of the Gauss-Newton method for nonlinear models Bates and Watts (1988), but we do not linearize the full functional, only a part of it, while the second part remains nonlinear.

4.2 Different bond weights

Our procedure can easily accommodate different weights for different bonds. It might be necessary, for example, to reflect the fact that the prices of different bonds can be observed with varying precision, which is usually measured by the bid-ask spread, and therefore require different fitting accuracy. For this, our minimization problem (3) should be modified as follows:

$$\sum_{k=1}^N w_k \left(\sum_{i=1}^n F_{ik} d(t_i) - P_k \right)^2 + \alpha \int_0^T (f^{[p]})^2(\tau) d\tau \rightarrow \min_{f(\cdot)}, \quad (38)$$

The numerical algorithm remains virtually unchanged. Indeed, the operator \mathcal{N}_k can be redefined as

$$\mathcal{N}_k f = \sqrt{w_k} \sum_{i=1}^n F_{ik} \exp \left(- \int_0^{t_i} F(f(\tau)) \right),$$

and prices P_k can be replaced by $P'_k = \sqrt{w_k} P_k$. Thus, after the change of variables we are back at the same minimization problem.

5 Estimation of the smoothing parameter

The selection of the smoothing parameter is an important task, because the wrong choice might yield economically unsound results. We suggest using two statistical methods: a generalized cross-validation method and a generalized maximum likelihood method to select the smoothing parameter (Wahba, 1990; Ke and Wang, 2002). In what follows, we briefly describe each of the methods, and write the final expressions for the choice of the smoothing parameter for both of them.

5.1 Cross-validation method

The idea of the ordinary leave-out-one cross-validation method is that the parameter λ should minimize the OCV score:

$$OCV(\lambda) = \frac{1}{N} \sum_{k=1}^N (y_k - \mathcal{N}_k f^{[k]})^2, \quad (39)$$

where $f^{[k]}$ is the minimizer of

$$\sum_{i \neq k} (y_i - \mathcal{N}_i f)^2 + N\lambda \int_0^T (f^{(p)})^2(\tau) d\tau. \quad (40)$$

The OCV score is the residual of fitting any single bond price via the term structure estimated from all data, except for this very single bond, summarized over all bonds.

Ke and Wang (2002) show that $OCV(\lambda)$ can be approximated as

$$OCV(\lambda) \approx \frac{1}{N} \sum_{k=1}^N (y_k - \mathcal{N}_k f)^2 / (1 - a_{kk})^2, \quad (41)$$

where

$$a_{ij} = \frac{\partial \mathcal{N}_i f}{\partial f} \frac{\partial f}{\partial y_j}. \quad (42)$$

Elements a_{ij} forms the matrix A . From we can get

$$\begin{pmatrix} \mathcal{D}_1 f_\lambda \\ \vdots \\ \mathcal{D}_N f_\lambda \end{pmatrix} = A(\lambda) y.$$

Wahba (1990) obtains the following formula:

$$I - A(\lambda) = n\lambda Q_2 (Q_2' M Q_2)^{-1} Q_2', \quad (43)$$

where Q_1, Q_2 are the orthogonal matrices forming the QR-decomposition of the matrix T :

$$T = (Q_1 : Q_2) \begin{pmatrix} R \\ 0 \end{pmatrix}. \quad (44)$$

Now let us consider a generalized version of this method. Instead of a_{kk} we use the value $\mu_1 = N^{-1} \sum_{i=1}^N a_{ii} = \text{tr}(A)/N$ in (41). Then the parameter λ is the minimizer of

$$V(\lambda) = \frac{1}{N} \sum_{k=1}^N (\mathcal{N}_k f - P_k)^2 / \left[\frac{1}{N} \text{tr}(I - A) \right]^2. \quad (45)$$

The generalized method allows to the obtaining of certain invariant properties as opposed to the ordinary cross-validation method Wahba (1990).

5.2 Generalized maximum likelihood method

For the unknown parameter λ , we a construct maximum likelihood estimate in a certain stochastic model corresponding to the minimization problem (30). Consider the following stochastic model linking observations Y_k to a known functional \mathcal{D}_k of an unknown random function F :

$$Y_k = \mathcal{D}_k F + \varepsilon_k, \quad k = 1, \dots, N, \quad (46)$$

where

$$\varepsilon \sim \mathcal{N}(0, \sigma^2 I),$$

$$F(t) = \sum_{i=1}^p \theta_i \varphi_i(t) + \sqrt{b} X(t),$$

$$\theta \sim \mathcal{N}(0, aI),$$

b is a positive constant,

$$\varphi_i(t) = \frac{t^i}{i!},$$

$X(t)$ is a Gaussian stochastic process:

$$\mathbb{E}X(t) = 0, \quad \mathbb{E}X(t)X(s) = R^1(t, s).$$

We repeat some of the notion used:

$$T = \{\mathcal{D}_k \varphi_j\}, \quad k = 1, \dots, N, j = 1, \dots, p,$$

$$\Sigma = \{\langle \xi_i, \xi_j \rangle\},$$

$$\varphi_j(t) = \frac{t^j}{j!},$$

$$\xi_k = P_1 \eta_k,$$

where P_1 is the orthogonal projector onto W^1 .

We will also use the QR-decomposition of the matrix T :

$$T = (Q_1 : Q_2) \begin{pmatrix} R \\ 0 \end{pmatrix}.$$

Using (46) and the fact that the sum of independent normal random variables is normal, we find that the observations are normally distributed with known parameters:

$$Y_k \sim \mathcal{N}(0, b(\eta T T' + \Sigma + N \lambda I)), \quad (47)$$

where $\eta = a/b$.

Now we formulate a theorem by Wahba (1990) stating the relationship of the original problem and the stochastic model:

Theorem 5.1. *Let*

$$\tilde{F}_a(t) = \mathbb{E}(F(t) \mid Y_k = y_k, k = 1, \dots, N),$$

and also $f_\lambda(t)$ is the minimizer of

$$\sum_{k=1}^N (y_k - \mathcal{D}_k f)^2 + N\lambda \int_0^T (f^{(k)})^2(\tau) d\tau,$$

where $\lambda = \sigma^2/Nb$. Then $\forall t$ the following equality holds

$$\lim_{a \rightarrow +\infty} \tilde{F}_a(t) = f_\lambda(t).$$

Thus, if we get the maximum likelihood estimate in (47) with $\eta \rightarrow +\infty$, the corresponding function $f_\lambda(t)$ will be the solution of the original minimization problem. To obtain the maximum likelihood estimate, we make the change of variables

$$\begin{pmatrix} z \\ \dots \\ w \end{pmatrix} = \begin{pmatrix} Q'_2 \\ \dots \\ \frac{1}{\sqrt{\eta}} T' \end{pmatrix} y. \quad (48)$$

This substitution allows us to split the vector y into two parts, where w in the limit does not depend on the parameter λ :

$$z \sim \mathcal{N}(0, b(Q'_2 \Sigma Q_2 + N\lambda I)), \quad \lim_{\eta \rightarrow +\infty} \mathbb{E} z w' = 0,$$

$$\lim_{\eta \rightarrow +\infty} \mathbb{E} w w' = b(T' T)(T' T).$$

Therefore, to obtain the negative log-likelihood function it is sufficient to use only z :

$$GML(\lambda) = \frac{z'(Q_2'\Sigma Q_2 + N\lambda I)^{-1}z}{[\det(Q_2'\Sigma Q_2 + N\lambda I)^{-1}]^{1/(n-p)}}, \quad (49)$$

Then, returning to y , we find that λ is the minimizer of

$$GML(\lambda) = \frac{y'(I - A(\lambda))y}{[\det^+(I - A(\lambda))]^{1/(n-p)}}, \quad (50)$$

where \det^+ is a product of nonzero eigenvalues.

The minimizer will be the maximum likelihood estimator of our regularization parameter λ .

5.3 Applying results

We described two methods of choosing the regularization parameter: generalized cross-validation and generalized maximum likelihood. As a result, we have that λ is determined as the minimum of (45) for GCV or of (50) for GML.

We can now write the final algorithm for yield curve fitting:

1. Select the initial approximation. Practice shows that a constant interest rate works quite effectively.
2. For a fixed λ compute the new approximation for f using (30)–(35).
3. Estimate λ using (45) or (50).
4. The iterative process 2–3 then continues until the difference between the old and the new approximations becomes less than a pre-specified value.

Although the convergence of the algorithm has not been studied because of the complexity of the problem, in practice it converges in few iterations. Table 1 presents the statistics of the number of trials until convergence for our dataset.

Table 1: Iterations for convergence

Number of iterations	Number of trials
3	127
4	163
5	47
> 5	3

Table 2: Market liquidity statistics

Number of traded bonds	10	11	12	13	14	15	16	17	18	19	20
Number of days	1	3	5	16	10	23	40	94	80	61	9

6 Testing

We compare our method with others using data on Russian bonds provided by Cbonds.info. The data includes daily closing price and bid & ask closing quotes for Russian domestic government bonds from 10 Jan 2012 until 14 May 2013. For each bond we also have the schedule of its promised cash flows. Typically bonds are coupon bearing with interest semi-annually or quarterly. This particular market was selected for two reasons. Russia is one of the few countries where bonds are exchange-traded, and the Russian bond market is rather illiquid, which makes term structure estimation a worthy problem. Table 2 presents the liquidity statistics of the market. The first row is the number of bonds traded in the market. The second row is the number of days with this number of bonds traded on this day. One can see from the table 2 that the liquidity of the market is relatively low.

We compare our method with several alternatives. Parametric methods by Nelson and Siegel (1987) and Svensson (1994) are the most used parametric term structure fitting methods. Svensson's generalization of Nelson-Siegel parametrization models the spot forward rate $s(t)$ as

$$s(t) = \beta_0 + \beta_1 e^{-\frac{t}{\tau_1}} + \beta_2 \frac{t}{\tau_1} e^{-\frac{t}{\tau_1}} + \beta_3 \frac{t}{\tau_2} e^{-\frac{t}{\tau_2}},$$

where $\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2$ are the model parameters.

The model by Cox, Ingersoll and Ross (1985) is a general equilibrium model, which describes the stochastic dynamics of the short rate. One can infer the term structure equation from these dynamics.

Table 3: Fitting accuracy for different methods

Method	MAE	RMSE	MAE Normalized	RMSE Normalized
N-S	0.2911	0.4787	3.9201	10.4407
Svensson	0.2122	0.2643	3.2797	8.8879
CIR	0.2170	0.2861	5.5097	17.8905
FNZ	0.2717	0.3004	2.9625	7.4074
G-curve	0.2452	0.3263	3.5723	9.7635
Our Method	0.1132	0.1512	0.7878	1.1857

Here we treat this term structure equation as just one more parametric fitting method, setting aside all dynamic interpretations.

$$\begin{cases} Q(t) = \exp[-A(t) - B(t)r_0], \\ A(t) = \frac{2\alpha}{\sigma^2} \left\{ \ln \left[\frac{(\gamma+\beta)(1-e^{-\gamma t})}{2\gamma} + e^{-\gamma t} \right] + \gamma t \right\} - \frac{\alpha(\gamma+\beta)t}{\sigma^2}, \\ B(t) = \frac{2(1-e^{-\gamma t})}{(\gamma+\beta)(1-e^{-\gamma t}) + 2\gamma e^{-\gamma t}}, \\ \gamma = \sqrt{\beta^2 + 2\sigma^2}. \end{cases}$$

This is not the most common parametrization, but our experience shows that this parametrization is the easiest to estimate. When considering stochastic dynamics, the restriction $2\beta > \sigma^2$ is usually imposed to ensure the positivity of the short rate. We relax this restriction since we only consider snapshot fitting. Fisher, Nychka and Zervos (1995) model is one of the most popular spline term structure fitting models, and almost the only one which does not require an exogenous smoothing parameter. We also add to our comparison the G-curve by Balabushkin (2004), a dynamic parametric approach based on an extension of Svensson parametrization and Extended Kalman Filter, because it is the official term structure model used by the exchange in the Russian market.

Fitting results are reported in Table 3. For each method we report Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the normalized MAE and RMSE. Normalized versions are obtained by normalizing the fitting error by the quoted bid-ask spread of the bond being fitted. All compared methods have been modified to accommodate these weights.

One can see that our method significantly outperforms others. One might think that this may be due to the low values of α (that is, due to the lack of smoothing). However, a visual inspection (see Figure 1) reveals that the constructed term structures are all quite similar, and the one constructed by

Table 4: Measures of smoothness

Method	Length of the yield curve	Length of the forward curve	$\int (r'')^2 dx$ of the yield curve	$\int (f'')^2 dx$ of the forward curve
Our method	16.1136	16.7672	0.1486	0.1511
N-S	16.4865	18.4099	0.5712	0.5654
Svensson	18.4607	18.6050	0.6201	0.6148
CIR	16.1384	17.7763	0.2134	0.3452
FNZ	17.2605	24.0861	0.3810	0.4433
G-curve	16.1809	18.2778	0.5512	0.6028

our method is even more smooth than the one constructed by FNZ. To further back this statement, we also report several measures that are often used for evaluating the smoothness of a zero-coupon yield curve (van Deventer and Uyemura (1992)). In particular, the smoothness of the zero-coupon yield curve can be measured as its length (or the length of the forward curve), another approach is to calculate the integral of the square of the second derivative (of spot or forward rates). In Table 4 we present different smoothness measures for all methods. Our method uses the smoothness of the forward curve in the problem formulation (although our objective does not coincide with the reported smoothness measure), so we would expect the values in the last column to be slightly biased towards the new method. The smoothness is not the key fitting performance measure. The fitted curve only has to be smooth enough, although if one method produces fitted curves which exhibit at the same time more accuracy and more smoothness, then this method should be considered superior, because accuracy and smoothness represent the two conflicting goals. The same holds if one estimate is better in accuracy and has the same smoothness or vice versa.

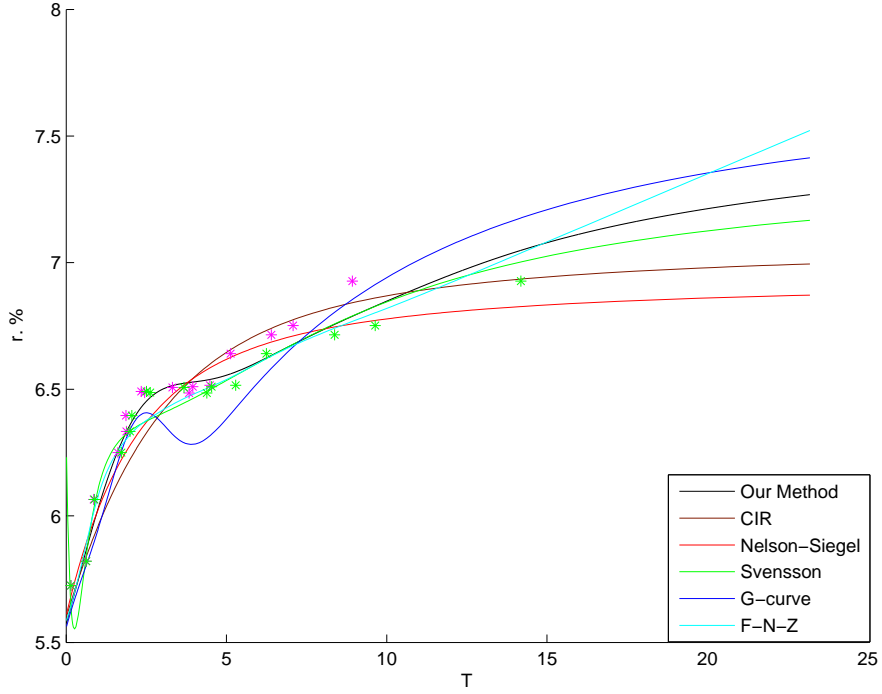


Figure 1: Term structures fitted by different methods.

7 Conclusion

We have described the mathematical problem of the nonparametric estimation of interest rate term structure, proposed its regularization and presented an iterative numerical algorithm that automatically estimates the regularization parameter. The proposed algorithm has several desirable properties unlike most alternative methods, it guarantees the non-negativeness of interest rates, it takes into account the liquidity of the market and the uncertainties of observable quoted prices (via different weights for different bonds). However, the main advantage of the proposed method is that it does not need an exogenously defined smoothing (regularization) parameter, as it is automatically determined within the algorithm, which is essential when dealing with noisy data.

References

- Adams K. and van Deventer D. Fitting yield curves and forward rate curves with maximum smoothness // *The Journal of Fixed Income*. — 1994. — Vol. 4, no. 1. — P. 52–62.
- Andersen L. Discount curve construction with tension splines // *Review of Derivatives Research*. — 2008. — Vol. 10, no. 3. — P. 227–267.
URL: <http://link.springer.com/article/10.1007%2Fs11147-008-9021-2>.
- Andersen L. and Piterbarg V. *Interest Rate Modeling. Term Structure Models*. — 2010.
- Balabushkin A., Gambarov G. and Shevtchuk I. Fitting the Term Structure of Interest Rates [In Russian] // *Rynok Tsennykh Bumag*. — 2004. — Vol. 11,13. —
- Bates D. M., Watts D. G. *Nonlinear Regression Analysis and Its Applications*. Wiley Series in Probability and Statistics. — Hoboken, NJ, USA : John Wiley and Sons, Inc., 1988. — ISBN: 9780470316757.
- Cox J. C, Ingersoll Jr J. E, Ross S. A. A Theory of the Term Structure of Interest Rates // *Econometrica*. — 1985. — Vol. 53, no. 2. — P. 385–408. — URL: <http://www.jstor.org/stable/10.2307/1911242>.
- Francis X. Diebold and Canlin Li. Forecasting the term structure of government bond yields. // *Journal of Econometrics*. — 2006. —
URL: <http://www.sciencedirect.com/science/article/pii/S0304407605000795>.
- Fama E. and Bliss R.R. The information in long-maturity forward rates // *The American Economic Review*. — 1987. — Vol. 77 no. 4. — P. 680–692.
- Fisher M., Nychka D., Zervos D.. Fitting the term structure of interest rates with smoothing splines. — 1995. — URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=6260.
- Gimeno R. and Nave J.M. A genetic algorithm estimation of the term structure of interest rates // *Computational Statistics and Data Analysis*. — 2009. — Vol. 53 no. 6. — P. 2236–2250.
URL: <http://www.sciencedirect.com/science/article/pii/S0167947308004994>.

- Hagan P.S. and West G. Interpolation Methods for Curve Construction // *Applied Mathematical Finance*. — 2006. — Vol. 13, no. 2. — P. 89–129.
- URL: <http://www.tandfonline.com/doi/abs/10.1080/13504860500396032>.
- James J. and Webber N. *Interest Rate Modelling*. — 2000.
- Jarrow R. Forward Rate Curve Smoothing. // *Annual Review of Financial Economics*. — 2014.
- Ke C, Wang Y. Nonlinear Nonparametric Regression Models. — 2002. — URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.392.8123&rep=rep1&type=pdf>.
- Lapshin V. Determining the Term Structure of Interest Rates // *Moscow University Computational Mathematics and Cybernetics*. — 2009. — Vol. 33, no. 4. — P. 206–213.
- Laurini M.P. and Hotta L.K. Bayesian extensions to Diebold-Li term structure model // *International Review of Financial Analysis*. — 2010. — Vol. 19, no. 5. — P. 342–350. URL: <http://www.sciencedirect.com/science/article/pii/S1057521910000578>.
- Lin B.H. Fitting term structure of interest rates using B-splines: the case of Taiwanese Government bonds // *Applied Financial Economics*. — 2002. — Vol. 12. — P. 57–54.
- Linton O., Mammen E., Nielsen J.P., Tanggaard C. Yield curve estimation by kernel smoothing methods // *Journal of Econometrics*. — 2001. — Vol. 105 no. 1. — P. 185–223.
- URL: <http://www.sciencedirect.com/science/article/pii/S0304407601000756>.
- McCulloch J. Measuring the term structure of interest rates // *The Journal of Business*. — 1971. — Vol. 44, no. 1. — P. 19–31. — URL: <http://www.jstor.org/stable/10.2307/2351832>.
- McCulloch J. The tax-adjusted yield curve // *The Journal of Finance*. — 1975. — Vol. 30, no. 3. — P. 811–830. — URL: <http://www.jstor.org/stable/10.2307/2326860>.
- Nelson C. R and Siegel A. F. Parsimonious Modeling of Yield Curves // *Journal of Business*. — 1987. — Vol. 60, no. 4. — P. 473–489.

- Sanchez J. and Gomez A.T. Estimating a fuzzy term structure of interest rates using fuzzy regression techniques // *European Journal of Operational Research*. — 2004. — Vol. 154 no. 3. — P. 803–818.
URL: <http://www.sciencedirect.com/science/article/pii/S0377221702008548>.
- Schmidt W.M. Interest rate term structure modelling // *European Journal of Operational Research*. — 2011. — Vol. 214, no. 1. — P. 1–14.
URL: <http://www.sciencedirect.com/science/article/pii/S0377221711000877>.
- Shea G. S. Pitfalls in Smoothing Interest Rate Term Structure Data: Equilibrium Models and Spline Approximations // *Journal of Financial and Quantitative Analysis*. — 1984. — Vol. 19, no. 3. — P. 253–269.
- Smirnov S and Zakharov A. A Liquidity-Based Robust Spline Fitting of Spot Yield Curve Providing Positive Forward Rates: EFFAS-EBC Working Paper. — 2003.
- Steeley J. Estimating the Gilt-Edged Term Structure: Basis Splines and Confidence Intervals // *Journal of Business Finance and Accounting*. — 1991. — Vol. 18, no. 4. — P. 513–529. — URL: <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-5957.1991.tb01119.x/abstract>.
- Svensson L. E. Estimating and Interpreting Forward Interest Rates: Sweden 1992 - 1994 : Working Paper : 4871 / National Bureau of Economic Research .— 1994. — URL: <http://www.nber.org/papers/w4871>.
- Tikhonov A.N and Arsenin V. Ya. Solution of Ill-Posed Problems. — Nauka, Moscow, 1979.
- van Deventer D., Uyemura D. Financial Risk Management in Banking: The Theory and Application of Asset and Liability Management. — 1992.
- Vasicek O. A, Fong H. G. Term Structure Modeling Using Exponential Splines // *The Journal of Finance*. — 1982. — Vol. 37, no. 2. — P. 339–348.
- Waggoner D. F. Spline methods for extracting interest rate curves from coupon bond prices / Federal Reserve Bank of Atlanta Working Paper .— 1997. — URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=86789.

Wahba G. Spline models for observational data. — SIAM, 1990. — P. 169. — ISBN: 0898712440.

Authors:

1. Victor Lapshin, Higher School of Economics (Moscow, Russia), Financial Engineering and Risk Management Laboratory. Research Fellow;

E-mail: vlapshin@hse.ru

2. Victor Lapshin, Higher School of Economics (Moscow, Russia), Financial Engineering and Risk Management Laboratory. Research Assistant;

E-mail: vkaushanskiy@hse.ru

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of National Research University Higher School of Economics.

©Lapshin, Kaushanskiy, 2014