COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

Multifactor models of interest rates and gain from an additional factor

MASTER'S THESIS

Bratislava 2017 Bc. Matúš KOŠÍK

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

Multifactor models of interest rates and gain from an additional factor

MASTER'S THESIS

Study Programme:	Mathematical Economy, Finance and Modelling
Field of Study:	1114 Applied Mathematics
Department:	Department of Applied Mathematics and Statistics
Supervisor:	doc. RNDr. Beáta Stehlíková, PhD.

Bratislava 2017 **Bc. Matúš Košík**

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Bc. Matúš Košík
Study programme:	Mathematical Economics, Finance and Modelling (Single degree study, master II. deg., full time form)
Field of Study:	Applied Mathematics
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Multifactor models of interest rates and gain from an additional factor.

Aim: More factors in short-rate models allow more flexibility in fitting term structures; however, also here it can happen that "less is more". In the thesis [Sutoris, 2009] it was found that if the short rate was modelled as a sum of n random factors, the objective function measuring difference between market and theoretical yields decreased significantly after adding the second factor, while adding more factors led to a smaller decrease. Also, there is a known phenomenon of overfitting - a lot of parameters or factors allow a very good fit, but these models fail when predicting the test data. They copy the training data and the noise in them instead of modelling the real relations. In this master thesis, these ideas will be used in studying selected models of interest rates an they will be applied to real data.

Supervisor:	doc. RNDr. Beáta Stehlíková, PhD.
Department:	FMFI.KAMŠ - Department of Applied Mathematics and Statistics
Head of department:	prof. RNDr. Daniel Ševčovič, CSc.
Assigned:	21.01.2016
Approved:	25.01.2016

prof. RNDr. Daniel Ševčovič, CSc.
Guarantor of Study Programme

.....
Student

.....
Supervisor

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Matúš Košík
Študijný program: ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Multifactor models of interest rates and gain from an additional factor.
Viacfaktorové modely úrokových mier a prínos dodatočného faktora.

Cieľ: Viac faktorov v short-rate modeloch úrokových mier umožňuje väčšiu flexibilitu pri fitovaní výnosových kriviek, aj tu sa však môže stať, že "menej je viac". V diplomovej práci (Sutóris, 2009) sa zistilo, že ak sa short rate modelovala ako súčet n náhodných faktorov, účelová funkcia merajúca rozdiel medzi trhovými a teoretickými výnosmi výrazne poklesla pri pridaní druhého faktora, kým pridávanie ďalších faktorov už viedlo k málo výraznému poklesu. Takisto je známy fenomén "overfitting" - veľké množstvo parametrov, resp. faktorov umožňuje veľmi dobre fitovať dáta, pri predikciách na testovacích dátach však takéto modely zlyhávajú. Kopírujú totiž tréningové dáta a v nich obsiahnutý šum namiesto modelovania skutočných závislostí. V diplomovej práci sa tieto myšlienky skonkretizujú na skúmanie vybraných modelov úrokových mier a aplikujú sa na reálne dáta.

Vedúci: doc. RNDr. Beáta Stehlíková, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.

Dátum zadania: 21.01.2016

Dátum schválenia: 25.01.2016 prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Acknowledgements I am grateful to my supervisor doc. RNDr. Beáta Stehlíková, PhD. for support and guidance throughout the elaboration of this thesis. Gratitude belongs to my family as well of their patience and support.

Abstrakt v štátnom jazyku

KOŠÍK, Matúš: Viacfaktorové modely úrokových mier a prínos dodatočného faktora [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: doc. RNDr. Beáta Stehlíková, PhD., Bratislava, 2017.

Úrokové miery podliehajú množstvu vplyvov v súčasnej ekonomike. Modely short rate sú populárnym nástrojom modelovania vývoja časovej štruktúry úrokových mier a oceňovania derivátov. V posledných dekádach vzniklo mnoho modelov s rozličným levelom komplexnosti. Komplexnosť však prichádza so stratou prehľadu a jednoduchosti riešenia. Venujeme sa CKLS short rate modelom s jedným, dvomi alebo tromi nezávislými stochastickými faktormi. Analyzujeme výsledky aplikácie týchto modelov na generované a aj reálne dáta. Cieľom je porovnanie týchto modelov na základe ich schopnosti popisovať a predpovedať dáta.

Kľúčové slová: Úrokové miery, Short rate, CKLS model, Viacfaktorové modely, Metóda priamok, Nelder-Mead optimalizácia.

Abstract

KOŠÍK, Matúš: Multifactor models of interest rates and gain from an additional factor [Master's thesis],Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: doc. RNDr. Beáta Stehlíková, PhD., Bratislava, 2017.

There are many influences that determine development of interest rates in nowadays economy. Short rate models are very popular tool of modelling term structure of interest rates and pricing derivatives. Various models were created in last several decades with various level of complexity. Complexity comes with price of losing tractability and the results become harder to explain. We focus on CKLS short rate models with one, two or three independent stochastic factors. We aim to analyse the results of application of these models to simulated and real datasets. The goal is to compare these models based on their ability to fit and predict data.

Keywords: Interest rates, Short rate, CKLS model, Multi-factor models, Method of lines solution, Nelder-Mead optimization.

Obsah

Introduction

Bond, one of the most common asset, is derivative of interest rate, meaning the price of the bond depends on the underlying interest rate. Understanding of interest rate is very desirable and modelling interest rate is important part of quantitative finance. Various models were developed throughout the history, each more complex than the previous one. The more complex models are able to capture wider range of interest rate behaviours and often explain reality better. However the added complexity comes with lower tractability of the model. Often analytical mathematics is no longer sufficient for more complex models and numerical approach is needed.

In this paper we focus on short rate models with various number of independent stochastic factors. These models estimate interest rates based on short rate, which is modelled by stochastic factors. One-factor short rate model is considered as the basic model. Multi-factor model uses more factors to describe short rate and is generally much more complex model. The advantage is that each factor may represent different influence on the interest rates.

In current global economy a single local event like elections can influence foreign countries and global development. Many seemingly uncorrelated areas become connected through globalization. With that in mind, are simple short rate models sufficient to describe the development of interest rates? Multi-factor models are better equipped to simulate various influences at once, but is it worth the added difficulty? How big of an improvement is the addition of another factor to the model? These are the questions we aim to answer in this paper.

1 Term structure modelling

Let us briefly introduce the theory behind modelling of term structures using stochastic differential equations. There are many assumptions that these models require and neglecting them can lead to misleading information or incorrect pricing. Theory gives us much needed insight in this problem, which will be useful for model calibration later on.

1.1 Term structure and short rate

Bonds are one of the most common and most traded securities on the market. Often they represent safe and stable investment, which makes them desirable for many undertakings. Bonds are customizable and various forms of bonds occurred throughout the history. However, this study is focused on the simple bonds without coupon payments - zero-coupon bonds. The issuer writes bonds with an expiration day - maturity, which are then bought by investors by providing money to issuer. The issuer has bound himself to pay the same value back with some additional value upon the expiration day. The additional value is usually predetermined and is called interest rate or yield. The price of the bond depends on its interest rate and the interest rate depends on expiration date and the probability of issuer's inability to pay back - default. The dependency of yields over time to expiration is called term structure of interest rates.

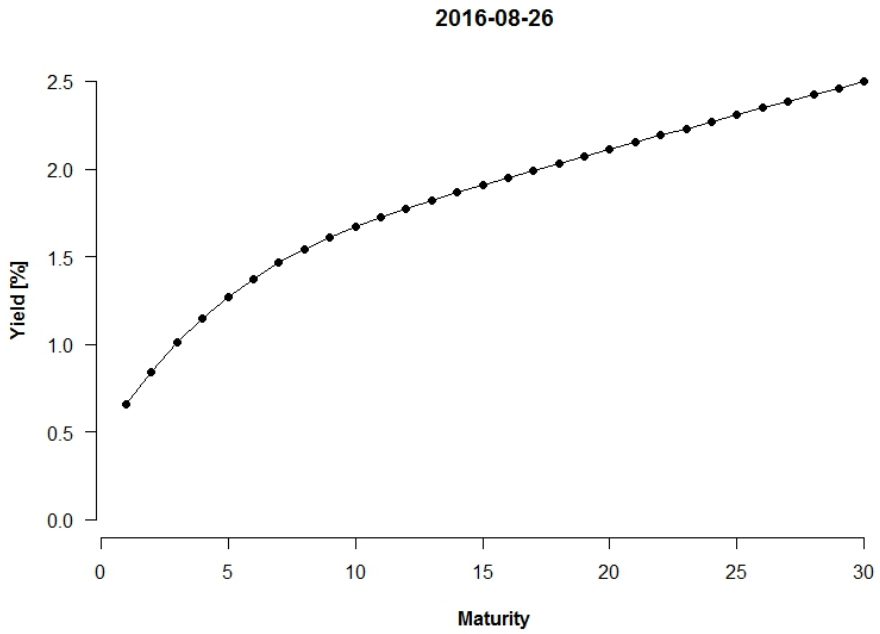
Relation between bond price and bond's yield is described by formula

$$P(t,T) = \exp^{-R(t,T)(T-t)}, \tag{1}$$

where $P(t,T)$ is the bond price, $R(t,T)$ is the interest rate of the bond and $T-t$ represents time to bond's maturity. From now on we will denote time to maturity by $\tau = T - t$. Taking logarithm of equation (1) leads to the formula for yields

$$R(t,\tau) = -\frac{P(t,\tau)}{\tau}. \tag{2}$$

Bonds are emitted with various maturities depending mostly on the current market situation and the needs of the issuer. Many government bonds have maturity measurable by decades, but there are also many short term investments. The shortest maturity of money market investment is one day, for example EURIBOR overnight rate.



Obr. 1: Zero-coupon yield curve of United States government bonds. Source [11].

Short rate is defined as instantaneous interest rate. It is interest rate for infinitesimally short time period. Mathematically it is defined as a limit

$$r(t) = \lim_{\tau \to 0^+} R(t, \tau), \tag{3}$$

where $r(t)$ is short rate. In Figure 1 short rate would be depicted by intersection of yield curve and Y-axis. Short rate is unobservable as the shortest traded maturity is one day. The short rate can be either approximated by a short term yield like the overnight rate or estimated by a model.

1.2 Short rate models - One-factor models

Modelling term structure of interest rate is subject of many studies as it is very useful for predicting future development or even better understanding of the present development. Various models and methods were created throughout the last decades. In this paper we focus on short rate models, where the short rate is assumed to follow stochastic differential equation (SDE)

$$dr = \nu(r, t)dt + \sigma(r, t)dW, \tag{4}$$

where dW is differential of Wiener process W. Wiener process is stochastic process where each increment of the process is independent and normally distributed with zero

Tabuľka 1: Table of basic short rate models and their stochastic differential equations. Third column shows whether the modelled short rate needs to be positive and fourth column whether the explicit solution exists. Table is based from [7].

Model	SDE	$r_t > 0$	Explicit
Vasicek	$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t$	No	Yes
CIR	$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t$	Yes	Yes
Hull & White	$dr_t = \kappa(\theta_t - r_t)dt + \sigma dW_t$	No	Yes
CKLS	$dr_t = \kappa(\theta - r_t)dt + \sigma r_t^\gamma dW_t$	Yes	No

mean value $W_{t+dt} - W_t \sim N(0, dt)$. Variable $\nu(r, t)$ denotes drift of the process, which defines the trend of the process over time. Variable $\sigma(r, t)$ stands for volatility of the process, which defines the variance of the process and range of fluctuations around the trend. The models differ mainly by different choices of drift or volatility variables. See Table 1 for examples of basic short rate models and their corresponding SDEs.

In Table 1 the models are also divided by their two properties. From these four models only Vasicek and Hull & White are able to model negative rates. Difference between these two is the time dependency of model's drift, in this case parameter θ. CKLS model (see [3]) is generalization of Vasicek and CIR model for any γ, but there is no closed form solution for bond price of CKLS model short rate.

There is common drift pattern in these models and it is called *mean-reversion*. Mean-reverting drift ensures that the average short rate converges towards parameter θ in the long term scale, while κ determines the speed of the convergence.

The CIR and CKLS model's volatility terms depend on the value of short rate. The volatility term is decreasing as the short rate is moving towards zero. Even if the short rate would happen to be zero, the positive trend and zero volatility would only lead to increase. Under specific conditions the zero short rate is unobtainable and we will address this topic later in chapter 1.5.

1.3 Bond pricing

In this section we derive partial differential equation for bond price of one-factor short rate model same as in [7]. Bond price P depends on the the time t, time till expiration

τ and short rate r_t. Short rate r_t is defined as stochastic markovian process and if the drift and volatility are $\mathcal{F}_t^W$-measurable, Itô lemma may be applied.

Lemma 1.1. *Let X be stochastic markovian process following*

$$dX = \nu(x,t)dt + \sigma(x,t)dW,$$

where W is Wiener process and $\nu(x,t)$ and $\sigma(x,t)$ are $\mathcal{F}_t^W$-measurable. If $f(x,t)$ is twice-differentiable transformation of process X, then the first differential of transformed process is described

$$df = \frac{\delta f}{\delta x}dx + \left(\frac{\delta f}{\delta t} + \frac{1}{2}\sigma^2(x,t)\frac{\delta^2 f}{\delta x^2}\right)dt.$$

Substituting $x = X$ and $dx = dX$ leads to

$$df = \left(\frac{\delta f}{\delta t} + \mu(x,t)\frac{\delta f}{\delta x} + \frac{1}{2}\sigma^2(x,t)\frac{\delta^2 f}{\delta x^2}\right)dt + \sigma(x,t)\frac{\delta f}{\delta x}dW.$$

Similarly bond price P is function of short rate, time and maturity $P(r,t,\tau)$. Using Itô lemma for $x = r$ and $f = P$ leads to

$$dP = \underbrace{\left(\frac{\delta P}{\delta t} + \mu(r,t)\frac{\delta P}{\delta r} + \frac{1}{2}\sigma^2(r,t)\frac{\delta^2 P}{\delta r^2}\right)}_{\mu_B(r,t)}dt + \underbrace{\sigma(r,t)\frac{\delta P}{\delta r}}_{\sigma_B(r,t)}dW, \tag{5}$$

where $\mu_B(r,t)$ and $\sigma_B(r,t)$ denotes drift and volatility of the process for the bond price.

Consider a portfolio comprised of one bond with maturity τ_1 and Δ bonds with maturity τ_2. The total value of portfolio is

$$\Pi = P(r,t,\tau_1) + \Delta P(r,t,\tau_2).$$

Change in the portfolio value

$$\begin{aligned} d\Pi &= dP(r,t,\tau_1) + \Delta dP(r,t,\tau_2) \\ &= \left(\mu_B(r,t,\tau_1) + \Delta\mu_B(r,t,\tau_2)\right)dt + \left(\sigma_B(r,t,\tau_1) + \Delta\sigma_B(r,t,\tau_2)\right)dW \end{aligned}$$

This equation holds for any given Δ and taking

$$\Delta = -\frac{\sigma_B(r,t,\tau_1)}{\sigma_B(r,t,\tau_2)}$$

eliminates the stochastic term dW and thus change of portfolio value $d\Pi$ becomes deterministic. At the same time due to no-arbitrage principle, yield of a riskless portfolio

Π has to be equal to riskless instantaneous short rate r, i.e. $d\Pi = r\Pi dt$. This leads to equation

$$\mu_B(r,t,\tau_1) - \frac{\sigma_B(r,t,\tau_1)}{\sigma_B(r,t,\tau_2)}\mu_B(r,t,\tau_2) = r\Pi.$$

Substituting Π into equation gives

$$\mu_B(r,t,\tau_1) - \frac{\sigma_B(r,t,\tau_1)}{\sigma_B(r,t,\tau_2)}\mu_B(r,t,\tau_2) = r\left(P(r,t,\tau_1) + -\frac{\sigma_B(r,t,\tau_1)}{\sigma_B(r,t,\tau_2)}P(r,t,\tau_2)\right),$$

which implies

$$\frac{\mu_B(r,t,\tau_1) - rP(r,t,\tau_1)}{\sigma_B(r,t,\tau_1)} = \frac{\mu_B(r,t,\tau_2) - rP(r,t,\tau_2)}{\sigma_B(r,t,\tau_2)}.$$

This equality holds for any τ_1 and τ_2, therefore the ratio on each side of the equation does not depend on the maturity. The ratio is called market price of risk

$$\lambda(r,t) = \frac{\mu_B(r,t,\tau) - rP(r,t,\tau)}{\sigma_B(r,t,\tau)} \tag{6}$$

and it describes the added value for taking risk of the investment. Rational investor would never invest in riskier assets unless they provide higher yield. Substituting μ_B and σ_B from (5) into (6) gives

$$\frac{\delta P}{\delta t} + (\mu(r,t) - \lambda(r,t)\sigma(r,t))\frac{\delta P}{\delta r} + \frac{1}{2}\sigma^2(r,t)\frac{\delta^2 P}{\delta r^2} - rP = 0. \tag{7}$$

This partial differential equation (PDE) for bond price P is core foundation of this paper. The zero-coupon bond's payoff is its face value when it reaches its maturity date. Without consequences we normalize the payoff to a unit 1 and use it as an initial condition for the PDE.

$$P(r,0) = 1 \tag{8}$$

1.4 Risk-neutral probability measure

In real world each asset is subject to risk, even government bonds to some extent. Real probability measure does not disregard these risks, but pricing assets becomes difficult. Each asset has different risks and each investor differently averse to risk. It is much easier to price assets in risk-neutral probability measure and then add the cost of risk. Similarly as we priced the bonds in previous section. In last step we introduced market price of risk to the equation, which basically added the cost of risk.

The drift μ in the PDE (7) is in real probability measure, thus the market price of risk needs to be subtracted. This means that some form of market risk is present. Our goal is the estimation of drift and volatility, while the bond prices are observable. Therefore the unknown market price of risk makes the estimation more difficult.

Defining a risk-neutral drift and volatility would lead to much more convenient PDE. Under risk-neutral probability measure each asset is priced based on discounted future payoffs, while the discount factor is equal to the risk free interest rate. Under this measure the price does not reflect any underlying risk and thus riskier assets may have same price as safer assets. This is equivalent to the existence of no-arbitrage principle, since two differently priced assets with same payoff would lead to arbitrage.

So to decrease the difficulty of estimation it is convenient to formulate PDE with risk-neutral parameters

$$\frac{\delta P}{\delta t} + \widetilde{\mu}(r,t)\frac{\delta P}{\delta r} + \frac{1}{2}\sigma^2(r,t)\frac{\delta^2 P}{\delta r^2} - rP = 0, \tag{9}$$

where $\widetilde{\mu}(r,t)$ is risk-neutral drift. Instead of subtracting the market price of risk from the real drift μ, we introduce risk-neutral drift $\widetilde{\mu}$. The risk-neutral drift will be different as it will assume no risks are present. Consider real drift μ of CKLS model

$$\mu(r,t) = \kappa(\theta - r),$$

where both κ, θ are real parameters. Then risk-neutral drift $\widetilde{\mu}$ is

$$\widetilde{\mu}(r,t) = \widetilde{\kappa}(\widetilde{\theta} - r),$$

where $\widetilde{\kappa}, \widetilde{\theta}$ are risk-neutral parameters. To avoid confusion we will use $\kappa(\theta - r)$ as risk-neutral drift for the rest of this paper and when changing to real probability measure we add the market price of risk to the risk-neutral drift $\kappa(\theta - r) + \lambda(r,t)\sigma(r,t)$.

Using PDE (9) estimation of risk-neutral parameters can be done based on observed bond prices. The advantage, compared to using PDE(7), is that we do not have to estimate market price of risk. The disadvantage is that the resulting short rate model will not produce correct predictions. The observations will be correctly estimated, but the predictions have to be done in real probability measure. Later we show how to estimate market price of risk and use risk-neutral drift to predict future data.

1.5 Fichera theory

Recall Table 1 and the drift of CKLS model. Together with CIR model the are restricted to positive short rate only. Volatility decreases as short rate decreases and the mean-reverting drift term becomes more dominant and push the short further from zero. Even if short rate reaches zero, the SDE for short rate simplifies to only $dr = \kappa\theta dt$, where the right hand side is positive. Nevertheless, under specific conditions there is zero probability that the random process reaches zero.

For example consider CIR model, where the coefficient γ determining the power of short rate is 0.5. If the short rate randomly decreases towards zero, the volatility of next random „jump“ decreases too, since it depends on the value of current short rate. Unless the „jumps“ are too big, the zero remains unreachable. If the condition

$$2\kappa\theta \geqslant \sigma^2 \tag{10}$$

is met, it can be shown, that the probability of zero short rate is zero (Feller condition).

In previous section we derived PDE for bond price (9) with its initial condition. However, we haven't mentioned any boundary conditions for the PDE. Behaviour of bond price at $r = 0$ and $r = \infty$, which are the boundaries of short rate, is not easily described. There is no clear indication how the bond price behaves at the boundaries. It becomes even harder for multi-factor models, where the boundaries are more complicated and the behaviour is even less clear. That is why we need Fichera theory.

Fichera theory focuses on the question of boundary conditions for parabolic PDE, which degenerate at a boundary. The parabolic PDE for bond price degenerates at boundary $r = 0$, since it becomes hyperbolic. Under specific conditions the boundary condition may be obsolete and thus does not need to be described. On the other hand, if the condition is violated, the boundary conditions must be supplied. The core principle of the theory is based on whether there is an inflow or an outflow at the boundary. Naturally inflow boundaries needs to be described by the boundary conditions, since it will directly impact the behaviour in the interior of the boundaries. To determine inflow and outflow boundaries one has to calculate *Fichera function*.

In article [1] authors describe the basics of the theory and propose its application

in short rate models. The Fichera function for one-factor model reads

$$b(r) = \beta(r,\tau) - \frac{\delta\alpha(r,\tau)}{\delta r}, \tag{11}$$

where $\beta(r,\tau)$ and $\alpha(r,\tau)$ are the terms in front of spatial derivatives $\frac{\delta P}{\delta r}$ and $\frac{\delta^2 P}{\delta r^2}$ in PDE for bond price (9). Specially for CKLS model $\beta(r,\tau) = \kappa(\theta - r)$ and $\alpha(r,\tau) = \frac{1}{2}\sigma^2 r^{2\gamma}$. Substituting in (11) the Fichera function for CKLS model reads

$$b(r) = \kappa(\theta - r) - \sigma^2\gamma r^{2\gamma-1}. \tag{12}$$

The Fichera function at $r \to 0^+$ determines whether the boundary condition are needed. Based on the sign two cases are possible:

- if $\lim_{r\to 0^+} b(r) \geqslant 0$ the outflow is present at the boundary and we must not define boundary condition.
- if $\lim_{r\to 0^+} b(r) < 0$ the inflow is present and we need to supply boundary condition.

Calculating the limit of (12) for $r \to 0^+$ leads to three possible cases:

- if $\gamma = 0.5 \Rightarrow$ Boundary conditions are not needed if $\kappa\theta - \frac{1}{2}\sigma^2 \geqslant 0$.
- if $\gamma > 0.5 \Rightarrow$ Boundary conditions are not needed if $\kappa\theta \geqslant 0$.
- if $0 < \gamma < 0.5 \Rightarrow$ Boundary condition are needed always, since the $\lim_{r\to 0^+} b(r) = -\infty$.

Not supplying boundary conditions when they are needed leads to incorrect solution of PDE and thus incorrect bond pricing. Since we want to avoid that, we need to follow Fichera theory. At the same time, we are unable to provide boundary conditions and thus we imply restrictions of parameters such that no boundary conditions are needed. For example $\gamma < 0.5$ violates these restrictions and therefore we are limited only to CKLS models with $\gamma \geqslant 0.5$. Each time we estimate parameters of CKLS model we need to check if they satisfy all restrictions mentioned above.

1.6 Multi-factor models

In section 1.2 we described one-factor short rate models. One-factor models are quite simple and the parameters are easily explainable. However, one-factor models often over-simplify the problem leading to poor results.

Possible solution are multi-factor short rate models. There are various types of the multi-factor models, but the general idea is to add another random process. Each random process counts as one factor, hence the name. In addition to short rate process, models can have random process to describe volatility, drift or another short rate. For example, multi-factor models are widely used for convergence models, when one interest rates is converging towards another and each one is modelled with a different factor (more information in [10]).

In this paper we focus on models where short rate is described as sum of multiple factors. Each factor is defined as a CKLS short rate with different parameters and Wiener processes.

Consider a two-factor CKLS model

$$dr_1 = \kappa_1(\theta_1 - r_1)dt + \sigma_1 r_1^{\gamma_1} dW_1$$
$$dr_2 = \kappa_2(\theta_2 - r_2)dt + \sigma_2 r_2^{\gamma_2} dW_2,$$

where $cor(W_1, W_2) = \rho$.

The bond price $P(r_1, r_2, t, \tau)$ now depends on each factor r_1, r_2, time t and maturity τ. Itô lemma (1.1) is applicable also when process dX is multivariate, as is in the case of multi-factor models. The PDE of bond price is constructed similarly as for one-factor model - using Itô lemma and eliminating randomness. To eliminate randomness one has to consider a portfolio of three bonds with various maturities and volumes. Under specific volumes of these bonds, the stochastic term gets eliminated. Applying no-arbitrage principle the PDE reads

$$\frac{\delta P}{\delta t} + (\mu_1 - \lambda_1\Sigma_1)\frac{\delta P}{\delta r_1} + (\mu_2 - \lambda_2\Sigma_2)\frac{\delta P}{\delta r_2} + \rho\Sigma_1\Sigma_2\frac{\delta^2 P}{\delta r_1 \delta r_2} + \\ + \frac{1}{2}\Sigma_1^2\frac{\delta^2 P}{\delta r_1^2} + \frac{1}{2}\Sigma_2^2\frac{\delta^2 P}{\delta r_2^2} - r_1 P - r_2 P = 0,$$

where μ_1, μ_2 are drifts of corresponding factors, Σ_1, Σ_2 volatilities and λ_1, λ_2 market prices of risk of each factor.

It can be seen from the PDE, that if the factors are independent $\rho = 0$, the bond price is product of prices calculated for each of the model separately. Similarly after taking logarithm, the total yield is sum of yields calculated separately for each model:

$$P = P_1 P_2 \qquad R = R_1 + R_2,$$

where P_1, P_2 are prices calculated as for one-factor model and P is price of bonds for two-factor model with independent factors. The bond price P_1 satisfies one-factor PDE (7) with parameters $\kappa_1, \theta_1, \gamma_1, \sigma_1$. Similarly P_2 with parameters of second factor. Therefore Same with the yields R_1, R_2. This makes calculating bond price easier, since instead of solving different and more complicated PDE, it is possible to solve bond price for each factor using the easier PDE (7) and summing up results. Problem of solving bond price PDE of model with any number of factors reduces to problem of solving bond price PDE of one-factor CKLS model.

2 Theory in practice

In previous section we briefly introduced basic theory and terminology needed for our work with the short rate models. Purpose of this work is to compare various models and naturally, the obvious comparison is based on how well the models describe the real data. Many factors influence the fit of the model such as model quality or correct model calibration and parameters estimation. As a building stone of our practice, we chose CKLS one-factor short rate model, which is one of the more sophisticated models and basically includes CIR and Vasicek model.

Since we would like to objectively compare various models, we need to design a robust algorithm, which could correctly calibrate not only CKLS model, but also multi-factor models. This way we avoid using various methods for various models, which would lead to subjective comparison.

The final algorithm must be able to solve each of these tasks:

- Calculate the correct bond price for given parameters based on the PDE.
- Evaluate the measure of difference between the real and estimated yield curve for given parameters.
- Effectively optimize parameters with aim to minimise the measure of difference.
- Predict future short rate and yields based on the model with optimized parameters.

In following chapters we describe each of these tasks and propose our solution, while demonstrating each used method on showcase data. The showcase data are theoretically ideal for CKLS model and therefore the base problem, our algorithm should easily solve. The data are simulated based on CKLS process with known parameters.

2.1 Bond pricing

Each short rate model is representing bond prices differently. Models like CKLS define partial differential equation (PDE) for bond price, which is simple transformation of interest rate (2). Closed form solution of the PDE is available only for Vasicek and CIR model. As for CKLS model, other methods must be used, often leading to approximate

solutions. Recall PDE for bond price (7). Later we show that under risk-neutral probability measure the parameter λ is part of the risk-neutral drift. In other words the PDE can be rewritten

$$P_\tau(r,\tau) = \kappa(\theta - r)P_r(r,\tau) + \frac{1}{2}\sigma^2 r^{2\gamma} P_{rr}(r,\tau) - rP(r,\tau) \tag{13}$$

$$P(r,0) = 1, \tag{14}$$

where the parameters of CKLS model $\kappa, \theta, \gamma, \sigma$ are risk-neutral parameters. The bond price $P(r,\tau)$ depends on time left until maturity of the bond τ and on the actual value of short rate r. The initial condition is in reality a terminal condition described by bond nominal value. The important thing to notice is the absence of the boundary conditions. The boundaries would be very useful, but they are very hard to construct. The behaviour of short rate at the boundaries is not clear, even less in case of multi-factor models. We either find a way to construct the boundaries or find solution without them. Already mentioned Fichera theory in section 1.5 proposes conditions, under which the lower boundary is reached with zero probability and therefore makes the boundary condition mathematically redundant. Basically the solution has to satisfy the Fichera conditions.

There are many available numerical tools for solving similar PDE. The basic is the method of finite difference. The core of the method is approximating differentials with finite differences and using discretization of time and spatial variable - short rate variable. The solution is then obtained for defined discrete set of time and space variables. Using both forward and backward differences as differential approximates is known as *Crank-Nicholson* method and has higher degree of precision, but is costly in terms of computing power.

Upgrade of the basic finite difference method is *Method of lines* using *Exponential time integration (ETI)* method described in [8]. It still uses differences for spatial differential approximation and discretization of spatial variable. This part is common with the basic method and leads to ordinary differential equation

$$P_\tau(r,\tau) = AP(r,\tau), \tag{15}$$

where A is tridiagonal matrix and P_τ is time differentiation of bond price. However, instead of approximating time differential the method simply solves ordinary differential

equation

$$P(\tau, r) = e^{A\tau} P(0, r). \tag{16}$$

For the solution the term $e^{A\tau}$ is needed and for the calculation following definition is essential.

Definition 2.1. *Let X be $n \times n$ matrix. Exponential matrix is $n \times n$ matrix defined by infinite series*

$$e^X = \sum_{k=0}^{\infty} \frac{X^k}{k!}, \tag{17}$$

while $X^0 = I$.

Task of solving the PDE breaks down to calculating infinite series defined by (17). Fortunately algorithm for such infinite series is available in R package *expm*, which uses several methods including Higham's algorithm described in his book [5]. Thanks to this package implementing ETI method into code is very easy. Understandably calculation of such series can be computation power demanding especially for large matrices A, but using basic finite difference method like Crank-Nicholson proves to be much more time consuming. In case the expm takes too much time, authors in article [8] propose simplified approximation of infinite series and even less demanding algorithm.

Implementation of ETI method consists of two steps:

- Construct fully-discretized space and matrix A
- Compute $e^{A\tau}$ with expm.

As mentioned before, the implementation of expm is straightforward with the built in functions. For any given matrix A and maturity τ the *expm()* function returns a vector of bond prices, where each price corresponds to short rate from the discretized set.

The first step can be more troublesome. In addition to (13) and (14) we need to know the boundary behaviour of bond price $P(r, \tau)$. Since we are using CKLS model, the short rate is restricted by model to interval $r \in \langle 0, \inf \rangle$. Practically we bound the short rate to interval $r \in \langle 0, r_{max} \rangle$ since we are not interested in cases with too high short rate, which are useless in reality. Nevertheless, the parameter r_{max} should still be large enough to represent infinity and shall be adjusted according to data. Using step

of length h, discrete set for space variable is created with length N, where N and h is chosen to match $r_N = r_{max}$. The choice of parameter h greatly impacts precision and time consumption of expm. Let us denote discrete short rate points and bond prices corresponding to each of those points for any given maturity τ:

$$r = \{r_i | r_i = ih, i = 0, 1, 2, ..., N\} \qquad P(r, \tau) = \{P_i | P_i = P(r_i, \tau), \forall \tau\}.$$

Method of finite difference is using the PDE, its initial and boundary conditions to approximate solution for any point in time and space. Initial and boundary conditions, if we know them, serve as known building blocks from which we can derive solution for points around them using the PDE with differential approximations. Iteratively one can derive solution for whole set of points providing image about the solution. The differential approximations are given by central difference formulas:

$$P_r(r_i, \tau) \approx \frac{P_{i+1} - P_{i-1}}{2h}, \qquad \forall i = 1, 2, ..., N-1 \tag{18}$$

$$P_{rr}(r_i, \tau) \approx \frac{P_{i+1} - 2P_i + P_{i-1}}{h^2}, \qquad \forall i = 1, 2, ..., N-1 \tag{19}$$

If Fichera conditions are satisfied, the boundaries are redundant. However, there still have to be something happening on boundaries $P_0 = P(r_0, \tau)$ and $P_N = P(r_N, \tau)$. Inspired by [4] we describe these bounds based on the values of $P_i = P(r_i, \tau), i = 1, 2, ..., N-1$, which are known from the PDE. In other words, we apply the behaviour of the PDE also to the boundaries. To obtain sufficient approximations of the spatial differentials on boundaries, we use formulas derived from Taylor series expansion. Detailed guide and theory is well studied for example in [6] in the first chapter. Derived formulas needed for boundary behaviour are

$$P_r(r_0, \tau) \approx \frac{-3P_0 + 4P_1 - P_2}{2h} \tag{20}$$

$$P_r(r_N, \tau) \approx \frac{3P_N - 4P_{N-1} + P_{N-2}}{2h} \tag{21}$$

$$P_{rr}(r_N, \tau) \approx \frac{2P_N - 5P_{N-1} + 4P_{N-2} - P_{N-3}}{h^2}. \tag{22}$$

All three approximation formulas have precision of order $O(h^2)$ and for higher accuracy longer Taylor series expansion is needed.

Combining (18) and (19) with the PDE (13) gives

$$P_\tau(r_i, \tau) = \kappa(\theta - r_i)\frac{P_{i+1} - P_{i-1}}{2h} + \frac{1}{2}\sigma^2 r_i^{2\gamma}\frac{P_{i+1} - 2P_i + P_{i-1}}{h^2} - r_i P_i, \quad \forall i = 1, ..., N-1. \tag{23}$$

This system of linear equations forms core rows of matrix A giving tridiagonal matrix

$$A^* = \frac{1}{2h^2}\sigma^2 r^{2\gamma} tridiag(1,-2,1) + \frac{1}{2h}\kappa(\theta - r)tridiag(-1,0,1) - rI, \tag{24}$$

where I is identity matrix and A^* represents matrix A without first and last row, where the boundary behaviour has to be defined by (20), (21) and (22). Next two equations define first and last row of matrix A.

$$P_\tau(0,\tau) = \kappa\theta\frac{-3P_0 + 4P_1 - P_2}{2h} \tag{25}$$

$$\begin{aligned} P_\tau(r_N,\tau) = {} & \frac{1}{2}\sigma^2 r_N^{2\gamma}\frac{2P_N - 5P_{N-1} + 4P_{N-2} - P_{N-3}}{h^2} \\ & + \kappa(\theta - r_N)\frac{3P_N - 4P_{N-1} + P_{N-2}}{2h} - r_N P_N \end{aligned} \tag{26}$$

All approximations of differentials are second order accurate and thus precision of this procedure should be no less than $O(h^2)$. However, accuracy of the whole method depends also on precision of solving ordinary differential equation (15), which depends on spatial step h.

Either we use the Crank-Nicholson finite difference method, where the principle is similar, but with time discretization and the resulting precision will depend also on the length of time step. Or we proceed with ETI method, which is much faster.

The algorithm for ETI method for given parameters κ, θ, γ and σ follows:

1. Perform discretization of interval $< 0, r_{max} >$ into N points with spatial spacing h.

2. Construct the inside elements of matrix A using (24).

3. Add first and last row of matrix A based on (25) and (26).

4. Calculate value of Fichera function. This is optional, but lets us know when the boundaries become incorrect.

5. With matrix A, maturity (e.g. 1 year) and terminal condition (bond's nominal value or simply 1 as a unit nominal value) calculate (16) using exponential matrix *expm()*.

6. For given short rate r find the two closest r_i from the spatial discretization and interpolate between corresponding bond prices.

7. Transform bond price to yield by relation (2).

The input of this algorithm are the parameters of the CKLS model$\kappa, \theta, \gamma, \sigma$, maturity τ, short rate r and numerical parameters of the spatial discretization h, r_{max}. The interpolation is done by R function *approx()*, which performs linear interpolation and is very fast.

To test which of the methods is better, we created simple experiment with simulation. For CIR model exact solution of the price $P(r, \tau)$ is known and therefore CIR model is suitable to test both numerical methods. Accuracy is not the only concern. During optimization this pricing procedure has to run countless times, so the procedure must be as fast as possible. Following Table 2 shows comparison between exact solution, ETI method and Crank-Nicholson method for selected points of short rate and maturity equal to 5 years.

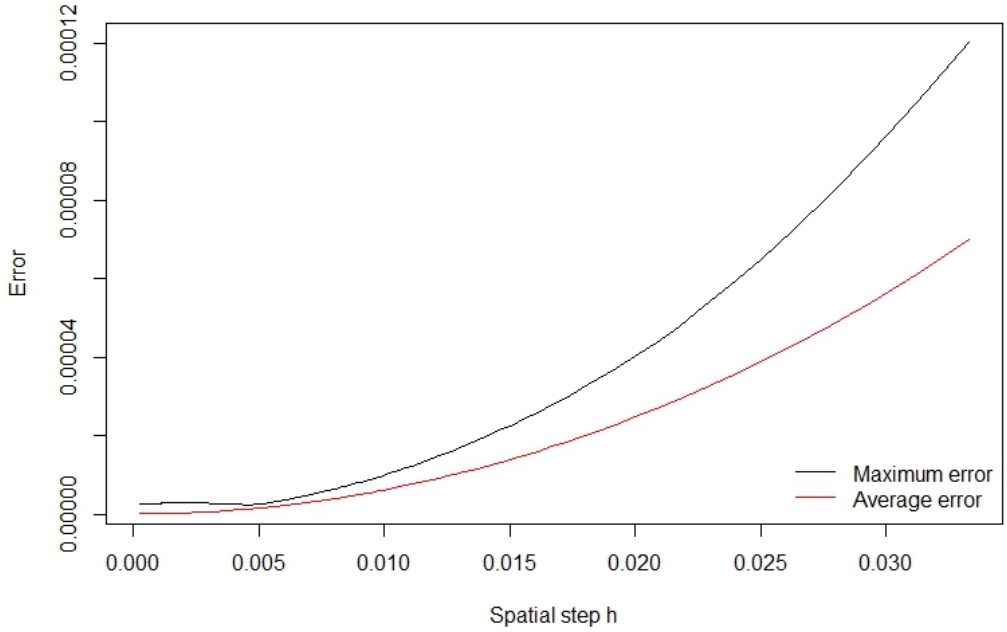
r	Exact P	Error expm	Error CN
0.000	0.8819199	4.216906e-07	3.239927e-05
0.065	0.7838823	2.994203e-07	3.205034e-05
0.195	0.6192903	1.495321e-06	3.026675e-05
0.260	0.5504477	1.888429e-06	2.910039e-05
0.455	0.3865282	2.440998e-06	2.512872e-05

Tabuľka 2: Comparison of exact CIR price and prices obtained from ETI or Crank-Nicholson numerical methods for selected points of short rate.

Of course Crank-Nicholson method in Table 2 uses time step 0.001 and same spatial spacing as ETI method. Taking smaller time step would result to higher accuracy, but also more time consumption, which is already much higher than with ETI method as we can in Table 3.

	ETI	CN_a	CN_b
Time step [years]	-	0.1	0.0001
Avg. error	1.531e-06	3.035e-03	1.618e-06
Max error	2.465e-06	3.227e-03	4.775e-06
Time elapsed [s]	0.0205	0.0234	20.1985

Tabuľka 3: Comparison of ETI and Crank-Nicholson methods. Error is averaged over all observed short rate points and time elapsed is averaged by repeating the procedure multiple times.

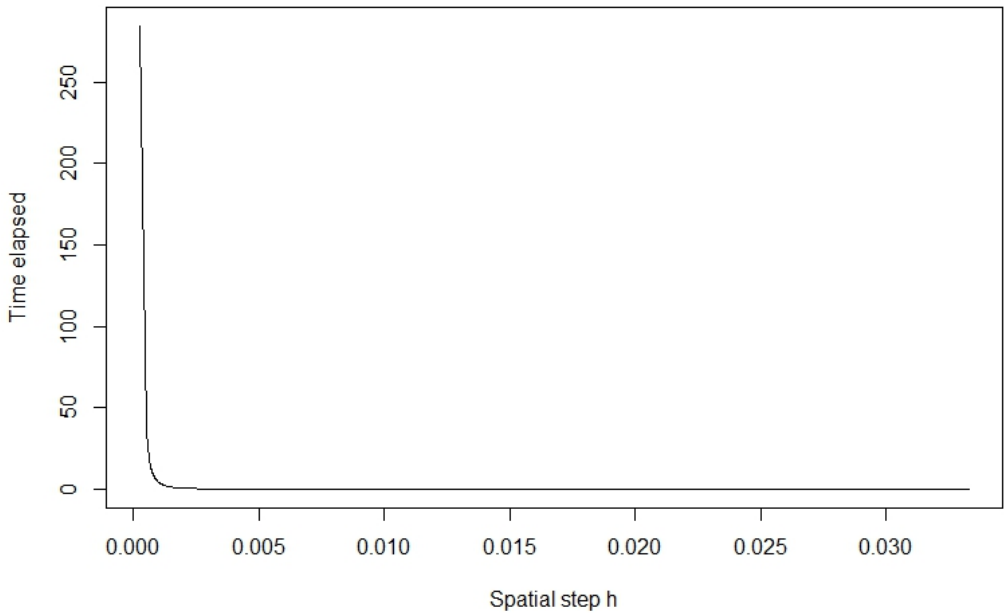


Obr. 2: Dependence of precision on spatial step h.

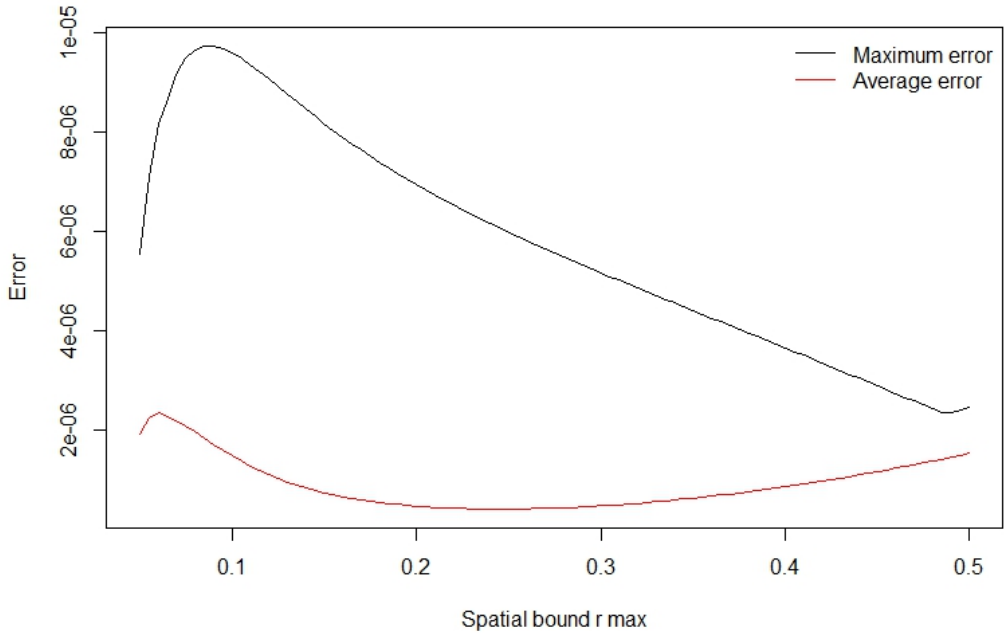
This experiments confirms the superiority of ETI method over basic finite difference method. Under same conditions ETI performs with higher precision and less time consumption. To match ETI's precision, Crank-Nicholson method would need more than 20 seconds. On the other hand to match ETI's speed, Crank-Nicholson would be 1000 times less accurate.

It is smart to know the effect of the numerical parameters like h and $rmax$ on the whole procedure. In Figure 2 is plotted dependence of average precision and maximum error on choice of spatial step h. Naturally smaller h gives more accurate results, but the time consumption rises non-linearly (see Figure 3). We can observe unexpected slight rise of maximal error for smaller h, which is probably caused by the change in source of the maximum error, for example from one boundary to another.

The parameter r_{max} is closely connected with h, so to separate the effect of h, we always take $h = r_{max}/100$ for all values of r_{max} we tested. The h is different, but the number of points in spatial discretization remains the same. In Figure 4 we see that the optimal choice of r_{max} is around 0.2, which could be explained as a compromise between satisfying large enough bound to practically represent infinity and small enough spatial discretization. Choosing too small r_{max} offsets the solution, because doing so inputs



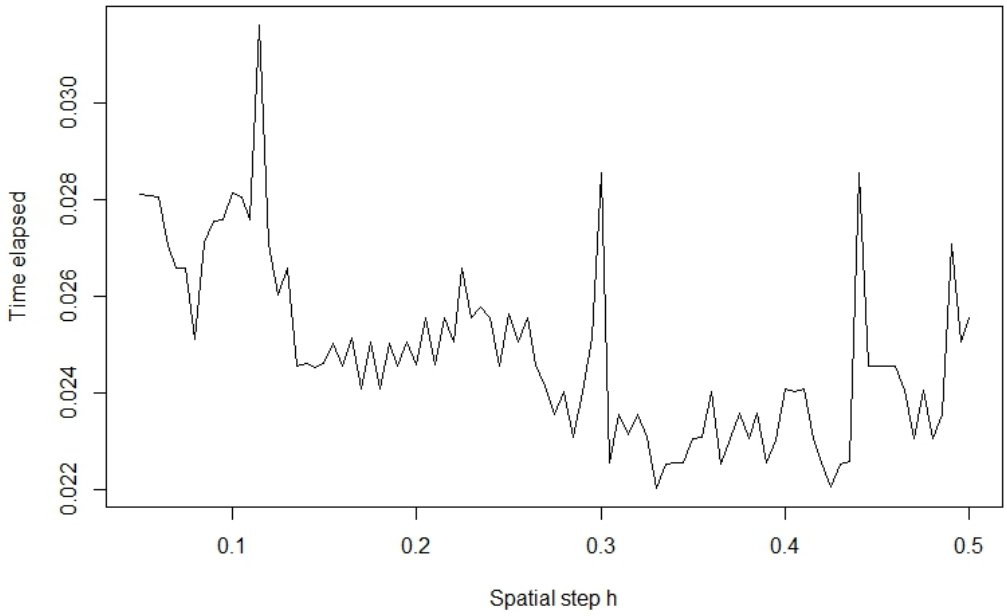
Obr. 3: Dependence of time consumption on spatial step h.



Obr. 4: Dependence of precision on spatial bound r_{max}.

initial error into the boundary, which by nature of the PDE diffuses to whole solution. The deciding factor for choosing correct r_{max} is to fit the data correctly. Parameter r_{max} needs to be few times bigger than parameter θ, because otherwise there is high chance

the short rate will exceed these bounds. Same with parameter σ, which represents the volatility of short rate. Too big volatility can lead to significant jumps in short rate, which can easily jump above the bound r_{max}. Therefore in case of incorrect pricing it is important to take larger r_{max} to prevent it. In Figure 5 the time elapsed fluctuates around the 0.025 seconds and does not appear to be dependent on r_{max}.



Obr. 5: Dependence of time consumption on spatial bound r_{max}.

Thanks to ETI method we are able to effectively solve PDE for any parameters that satisfy condition from Fichera theory. It is important to correctly choose numerical parameters r_{max} and h to prevent incorrect solutions. The output of the method is vector of prices P_i for each short rate r_i from spatial discretization. From these we can simply interpolate for any value of short rate, which is not directly present in the discretization.

2.2 Parameter optimization

In previous section we described method to calculate bond price for given parameters κ, θ, γ and σ. In reality these parameters are unknown, but they can be estimated from data about term structures.

There is no general or standard approach to calibration of short rate models. For

simpler models, such as Vasicek or CIR, it is possible to derive the probability distribution of the short rate processes and use method of maximum likelihood to estimate parameters. Other estimation techniques may be based on methods of moments, where the estimation uses the information about the statistical moments of random variable, in this case short rate. For example since it is mean-reverting process, the first moment - mean value can be used. Even more complex statistical tools are used for estimation like Kalman filter, which has the advantage of modelling both current term structure and the dynamics of the rates. One of the simpler methods is basic regression, in other words optimizing the parameters so estimated yield curves fit real dataset as much as possible. More information regarding these methods and their use for term structure modelling can be read in thesis [9].

Our goal is to compare various models and for that we need a versatile calibration method, which is able to calibrate more than just specific model. This motivation led us to use regression method, since it does not require knowledge of probability distribution and is able to estimate even multi-factor models.

To find the optimum one must first be able to evaluate the optimality, the goodness of the fit. This evaluation must capture differences between estimated and observed yield curve and its dynamics over time. Standard L_2 regression uses sum of square residuals, which corresponds to searching for mean value within the data. We use same principle, except we have to capture both residuals in yield curves and residuals in development over measured time period. Let us define the function to evaluate the goodness of fit as a loss function F equal

$$F = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} (R^{est}(\tau_j, r_i) - R_{ij}^{real})^2, \tag{27}$$

where m is the number of observed maturities, n number of observed time points and w_{ij} constants representing weights of each residuals. For example, if we want to fit long-term rates, we can use weights $w_{ij} = \tau_j$.

The objective is to minimize the loss function by changing R^{est}, which is done by changing parameters κ, θ, γ and σ. These parameters are technically restricted only by condition of non-negativity, but in real world some values are unrealistic. Using common sense it is easy to restrict these parameters, which can sometimes prevent cases when the optimization diverges. Implementing these restrictions for parameters is quite simple

using penalization function. Penalization function checks each restriction and returns value, which is then added to the objective function. For example restriction $\gamma \geqslant 0.5$ could be evaluated in form $\Lambda min(0, \gamma - 0.5)^2$, where Λ is coefficient of strictness of the penalization. In some cases wrong Λ may impede convergence, so in general, it is better to run optimization multiple times with different Λ. The mentioned restriction for γ rises from condition of Fichera theory (section 1.5). Other restrictions, such as non-negativity and reasonable upper bounds for parameters, are formulated in same fashion and in the end all penalizations are added together. One way to ensure not violating Fichera conditions is to implement whole condition into penalization function, but due to complexity of this condition that could have negative effect on convergence. We keep only the obvious restriction for $\gamma > 0.5$.

The inputs for the calculation of objective function consist of parameters κ, θ, γ and σ in form of vector x, vector of maturities, input short rate time series, dataset of bond prices for each observation and maturity, numerical parameters for ETI method and penalization function. First we perform the ETI algorithm to obtain estimated yields for each observation and each maturity. The resulting matrix of estimations is compared with observed yields and the loss function is calculated as in equation (27). Afterwards penalizations are added if any bounds are breached. The result is the objective value.

There is no indication whether the loss function is convex, which means there might be multiple local minima along with global minimum. Therefore using traditional convex programming algorithms might not converge correctly. Let us introduce few possible algorithms which are widely used for non-convex problems.

Multi-start BFGS In convex programming one of the top performance algorithm is BFGS, which is based on numerically estimating gradient and using it to find the way towards minimum. In non-convex problems the convergence is very limited with the choice of starting point. As a workaround one can run the algorithm from various starting points and hope one converges correctly. Advantage of BFGS method are its known convergence properties.

Nelder-Mead Widely used non-convex problem method is Nelder-Mead heuristics. Its convergence properties are unknown, but in many cases it outperforms other methods like BFGS. Especially for functions that are very flat and the approximated

gradient would basically be zero. The core of the algorithm is based on systematic and reasonable simplex manipulations in the space of feasible solutions. Advantage of this method is the simplicity and instead of calculating complicated multivariate gradients, Nelder-Mead only needs to evaluate objective function few times in each iteration, depending on the number of parameters to optimize.

Genetic algorithm Genetic algorithm (also known as evolution algorithm) belongs to class of stochastic optimization methods, more specifically a population heuristics. Since the method works with multiple points, it has bigger chance of finding global minimum instead of local minima. There are many variations of genetic algorithms, but the core remains same. Create population of points from set of feasible solutions and evaluate loss function for each of them. Order them by the loss function and combine the points between them, while the better points combine more frequently. Last step is to mutate whole population, which means slight adjustment to the points. Mutation can lead to better points as well as to worse points, but the better points will combine more. In the end we keep only the best points and repeat whole procedure. The advantage of genetic algorithm is the ability to explore completely new points, which enables the solution to jump from local minima to another potential global minimum.

Each of these methods looks for ideal combination of parameters differently and each has different advantages. To pick one we tested each of these methods on simulated data. The objective is to find most precise and fastest method, even though the speed of method mainly depends on the speed of ETI method for bond pricing.

Short rate is simulated using CKLS process with know parameters $\kappa = 0.5, \theta = 0.04, \gamma = 0.7$ and $\sigma = 0.1$ and the interest rates are estimated for maturities $\tau = 1, 2, 3, 4, 5, 10$ years. In practice the short rate time series is always unknown, but at the moment we are interested in comparison of these methods. In case of Nelder-Mead and BFGS we use the R built-in function *optim*, which covers both of these methods. Genetic algorithm is designed by us and uses basic recombination based on the best points and mutation using normal distribution. We test these methods on same data, same starting point (except genetic algorithm, which does not require starting point) and same termination criteria (except genetic algorithm, which is terminated after N iterations). The results are in Table 4.

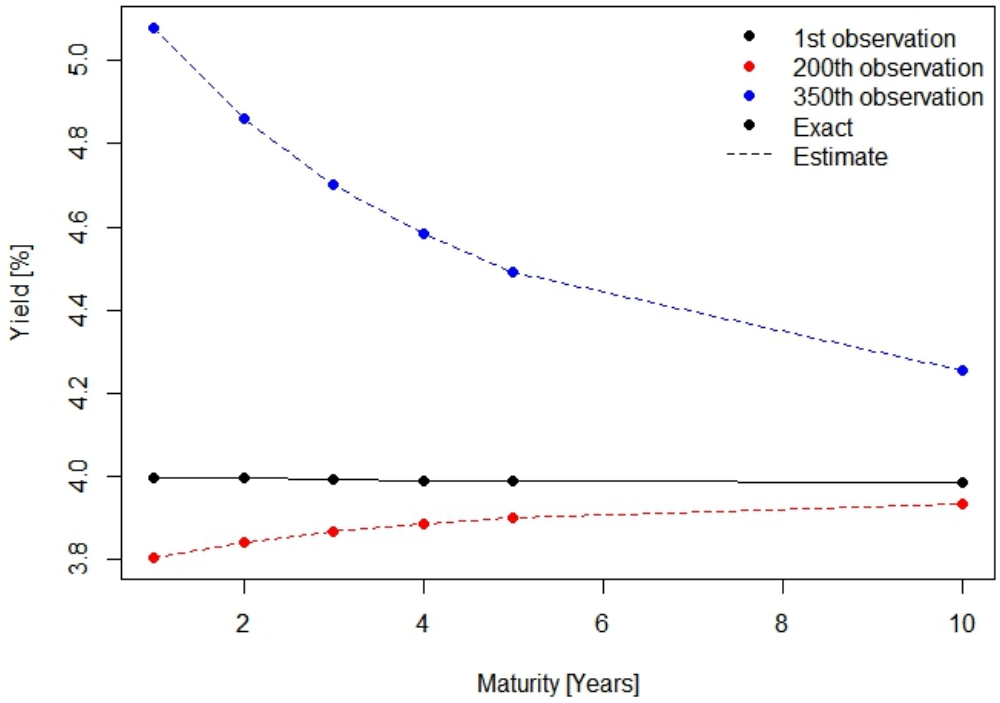
	Exact	Nelder-Mead	BFGS	Genetic algorithm
F	0	1.7e-17	4.8e-13	1.3e-09
κ	0.50	0.5000016	0.5007356	0.0388817
θ	0.04	0.0399999	0.0400058	0.0719658
γ	0.70	0.6995516	0.5040209	1.6749183
σ	0.10	0.0998374	0.05461255	0.0950745
time	-	216s	400s	15min

Tabuľka 4: Comparison of results with different parameters optimization approaches.

Of course the time elapsed for each algorithm depends on the computing power of machine on which it is running, but this is irrelevant for the purpose of comparison. From Table 4 its obvious that our genetic algorithm is the worst choice for this task. Performance of genetic algorithms hugely depends on choice of recombination and mutation methods, which basically fill out the core of the algorithm. Poor choice of these can lead to under-performance or even failure. Therefore we can not deny possibility where a different genetic algorithm would perform much better. The stochastic mutation can be very effective in some cases, but in continuous space with four dimensions (one for each parameter) is rather ineffective and requires huge amounts of iterations. Reasonable combination of genetic algorithm and convex methods is proposed in [9] with good results. BFGS cost less iterations to find the optimal κ and θ, but converged to suboptimal result. Most of the time elapsed BFGS algorithm moved in tiny steps and even then struggled to find the optimal γ and σ. Keep in mind, that BFGS and Nelder-Mead have same starting point and same termination criteria, meaning that the algorithm stops, if the relative change in objective function is less than 1×10^{-18}. Nelder-Mead algorithm proves to be the best option out of these three and so all optimization in following sections will be using Nelder-Mead algorithm.

At this point we only need to provide the bond yields for various maturities and corresponding short-rate as inputs for the optimization algorithm. The algorithm finds the best estimates $\widehat{\kappa}, \widehat{\theta}, \widehat{\gamma}, \widehat{\sigma}$ for which the objective function (27) is minimized. In case the objective value equals zero, the estimated bond yields would fit perfectly the real values. Since we are dealing with numerical approach and numerics, absolute zero

is impossible to accomplish. Next figures show simulated data points together with estimates based on the optimized parameters and short-rate time series. The data were simulated based on CKLS process with parameters same as the exact parameters specified in Table 4.

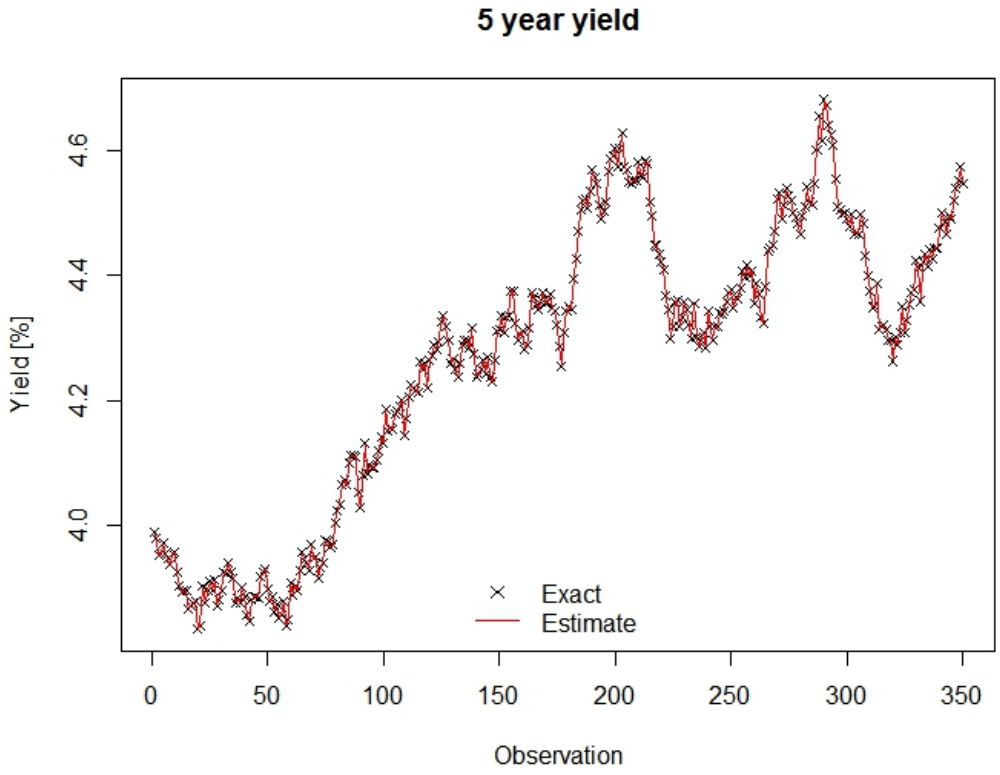


Obr. 6: Exact yield curve and estimated yield curve for 3 different observations.

Figure 6 shows how well the estimation fits the data for all observed maturities and Figure 7 for all observed points. The estimation captures all shifts in the yield curve and looks like its a perfect fit. Although there are differences on some level as we can see in Table 5. These differences are on average 10^{-7} of percentage point and that is negligible. Such small differences rise up from numerical approach in both bond pricing equation and parameter optimization.

2.3 Short rate optimization

Short rate is an abstract time series and unknown in real world. Not knowing the short rate makes all previous paragraphs about optimization worthless since the algorithm's key input is the short rate. We already mentioned choosing the yield with shortest



Obr. 7: Exact time series of 5 year yield and its estimate.

maturity as proxy for short rate could avoid this issue. This workaround can work for some cases, but often it does not. Therefore we propose alternate solution using, once again, computing power.

Pricing the bonds numerically provides whole set of prices each valid for different short rate value based on short rate discretization. Instead of providing short rate value as outside variable, we let the algorithm choose the bond price, that leads to best fit possible over all maturities. In other words, we let the algorithm choose the unknown short rate value for each observation point. With this upgrade the algorithm requires only yield data. However, this adds to level of freedom during the optimization and could impair the ability to find the correct parameters. Meaning that even incorrect parameters can become relevant if the short rate is incorrect, which can happen easily when we choose bad starting point for the optimization. So there is possibility this upgrade correctly optimizes short rate and then converges to correct parameters, but also possibility that it adjusts the short rate for incorrect parameters and justifies them. Therefore it is important to try both short maturity proxy and short rate optimization

Y1	Y2	Y3
Min. :6.868e-10	Min. :1.853e-09	Min. :5.958e-09
1st Qu.:4.029e-07	1st Qu.:5.170e-07	1st Qu.:4.886e-07
Median :7.197e-07	Median :8.620e-07	Median :7.849e-07
Mean :7.155e-07	Mean :9.036e-07	Mean :8.984e-07
3rd Qu.:9.618e-07	3rd Qu.:1.168e-06	3rd Qu.:1.214e-06
Max. :1.890e-06	Max. :2.610e-06	Max. :2.811e-06
Y4	**Y5**	**Y10**
Min. :1.314e-10	Min. :4.827e-09	Min. :9.122e-10
1st Qu.:4.111e-07	1st Qu.:3.402e-07	1st Qu.:2.583e-07
Median :6.854e-07	Median :5.878e-07	Median :4.452e-07
Mean :8.319e-07	Mean :7.517e-07	Mean :4.795e-07
3rd Qu.:1.143e-06	3rd Qu.:1.052e-06	3rd Qu.:6.698e-07
Max. :2.713e-06	Max. :2.496e-06	Max. :1.333e-06

Tabuľka 5: Summary of absolute residuals between exact and estimated yields in [%].

approaches and compare them. For example, if we see that the optimized short rate looks completely out of place by comparing it with the yield data, we then incline towards rejecting the short rate optimization approach. On the other hand if we see that the optimized short rate looks similar as the short maturity proxy, except some parts of the curve, we then incline towards using optimized short rate, rather than proxy.

We use following algorithm to optimize short rate and parameters in R:

1. For initial parameters calculate estimated yields without interpolating for the short rate (the ETI algorithm omits the 6th task).
2. For each observation separately:
   - For each maturity subtract the observed yield from all estimated yields.
   - Square the results and aggregate them into a vector of same length as is the number of discretization points.
   - Find which two discretization points lead to the smallest square difference and interpolate between them based on the squared difference.
   - The interpolated point is the optimized short rate for given observation.
3. Calculate objective function with given parameters and optimized short rate.
4. Find new parameters and repeat until minimum is obtained. This is achieved by Nelder-Mead optimization.

To showcase the situation we use same simulated data as before, but this time we know only 6 months yield as approximate short rate instead of the correct short rate. As expected the original optimization does not find the optimum and stops at very bad parameters. Sure we can argue that it can be coincidence and maybe a different choice of starting point or density of space grid in ETI method could perform much better. To test this we set the starting point to correct parameters and run the optimization again. But even under such perfect conditions the algorithm fails and we have no other choice than using short rate optimization.

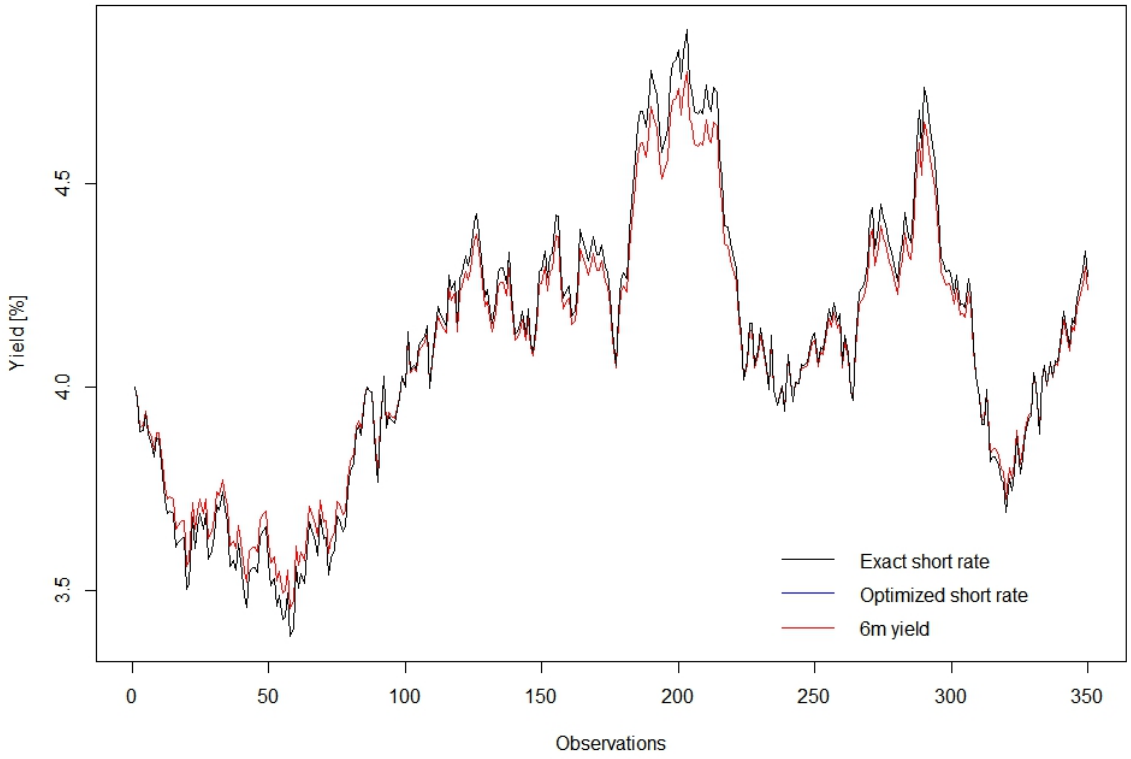
The disadvantage is that the upgraded algorithm takes little bit longer, which multiplied by hundreds of iterations can prolong the process by few minutes. But for those few minutes we get the correct parameters and optimized short rate, which is important for our following work. In Table 6 we show the result of these two approaches, while the proxy approach did not even converged and stopped far away from solution, the optimal short rate approach found the correct solution with precision as good as in case of known short rate.

	Exact	6m proxy	Opt. short rate
F	0	8.5e-09	2.2e-18
κ	0.5	0.390684	0.5000008
θ	0.04	0.040142	0.0399999
γ	0.7	1.351749	0.6996681
σ	0.1	0.883663	0.0998904

Tabuľka 6: Comparison of optimization using 6 month yield as short rate and optimizing short rate each iteration. Data are simulated from CKLS process.

As we mentioned in the beginning of this section it is important to compare the optimized short rate with the proxy short rate, in this case 6 month yields. From this view we can atleast guess if the optimization of the short rate is meaningful and does not deviate from the actual yields too much. As we see in Figure 8 the optimial short rate mimics 6 month yields most of the time and since this dataset is simulated we can easily compare them to the exact short rate. The difference between exact and optimized short rate is so small, that it is not visible in Figure 8. In numbers the

difference is on average 8.5×10^{-9}. Still we need to emphasize the significance of the starting point and the fact that these data are theoretically ideal for CKLS process. We have yet to see if this methods will be of any use in practice.



Obr. 8: Difference between 6 month yields, exact and optimized short rate. Optimized is so close to exact, that its not visible on the graph.

2.4 Market price of risk

All previous calculations and optimizations are done in risk-neutral probability measure, meaning that we disregard the parameter market price of risk $\lambda(r,t)$ in bond pricing Equation (7). Optimized parameters are risk-neutral and even though they fit the data perfectly, they will not produce correct predictions for future. The reason is that the mean value of short rate defined by CKLS process equals the drift of the process. Risk-neutral drift will produce different predictions than the real drift. Therefore we want to estimate the market price of risk based on what we have available, which are the yield data, optimized parameters and optimized short rate or short rate proxy.

There are many ways to define $\lambda(r,t)$ and how to estimate it. We tried few methods and chose a rather simple approach, but easy to estimate. We experimented with the form $\lambda(r,t) = \lambda r^{\delta}$, but the estimation method failed to give reasonable results for other than $\delta = 0$. We define market price of risk:

$$\lambda(r,t) = \lambda, \tag{28}$$

where λ is an unknown constant.

To estimate the constant we use the fact that CKLS short rate is normally distributed. With known distribution we can perform maximum likelihood estimation (MLE), where we maximize the likelihood function. MLE is looking for most probable parameter λ for given data points and distribution. The MLE estimator is biased, so the estimation is not perfect.

The MLE estimation is based on the known distribution for the next short rate realization, which means the probability density function for the random variable needs to be known. The distribution for short rate is normal with these parameters:

$$r_{t+\Delta t} \sim N(r_t + (\kappa(\theta - r_t) + \lambda\sigma r_t^{\gamma})\Delta t, \sigma^2 r_t^{2\gamma}\Delta t) \tag{29}$$

The Δt stands for the time step between each observation, for example day ($\Delta t = 1/250$). MLE estimator maximizes the likelihood function:

$$\begin{aligned} L(r,\lambda) &= \prod_{t=1}^{n} f_t(r,\lambda) \\ L(r,\lambda) &= \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2 r^{2\gamma}\Delta t}} \exp\left(\frac{-(r_{t+\Delta t} - (r_t + (\kappa(\theta - r) + \lambda\sigma r_t^{\gamma})\Delta t))^2}{2\sigma^2 r^{2\gamma}\Delta t}\right) \end{aligned}$$

We then take logarithm of the likelihood function since the optimum remains the same for both, but log-likelihood function is easier to manipulate.

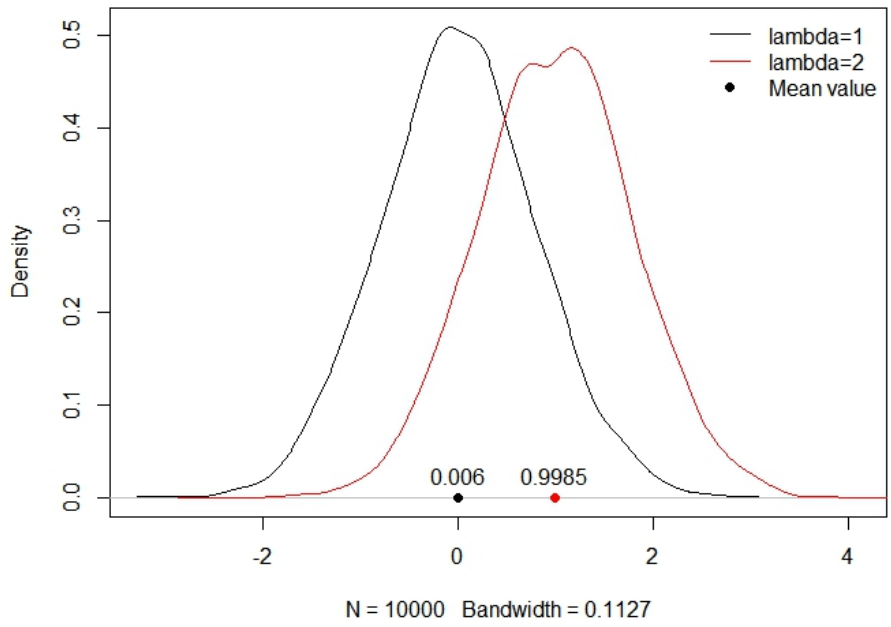
$$log(L(r,\lambda)) = \sum_{t=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2 r^{2\gamma}\Delta t}}\right) - \frac{-(r_{t+\Delta t} - (r_t + (\kappa(\theta - r) + \lambda\sigma r_t^{\gamma})\Delta t))^2}{2\sigma^2 r^{2\gamma}\Delta t}$$

We now search λ that satisfies optimality condition.

$$\frac{\delta L(x,\lambda)}{\delta\lambda} = 0 \tag{30}$$

$$\widehat{\lambda} = \frac{1}{n\Delta t}\sum_{t=1}^{n} \frac{r_{t+\Delta t} - (r_t + \kappa(\theta - r_t)\Delta t)}{\sigma r_t^{\gamma}} \tag{31}$$

The optimum is maximum, because the second order derivative is negative, which means the log-likelihood function is concave. Equation (31) can be easily solved if the parameters and short rate series are known. Following steps from previous sections we have both these requirements and we simply aggregate differences between expected short rate and actual short rate divided by the volatility and then divide it by the total number of days multiplied by the time step.



Obr. 9: Density of MLE estimated λ for simulated 10000 short rates with $\lambda = 1$ and as reference another 10000 short rates with $\lambda = 0$. The mean of 10000 estimations is close to correct values.

To test this we simulated data once again, but this time we added known market price of risk to the drift function. Since we already know the estimator will be biased with limited number of observations we simulate the data 10000 times. On average the estimator should hit the correct value of λ. In addition we add second dataset with zero market price of risk as a reference.

In Figure 9 we see that the estimated market prices of risk are on average very close to correct values, but the variance of the estimation is high and in unlucky cases the estimation gets very biased. Adding more observations points would narrow the interval.

2.5 Predictions

At this point we have all we need to predict future short rates and with that also future yield curves. Equation (29) describes the distribution of future short rate and the mean value is basically the prediction. With the variance it is possible to construct 95% confidence interval for the average of the future short rates.

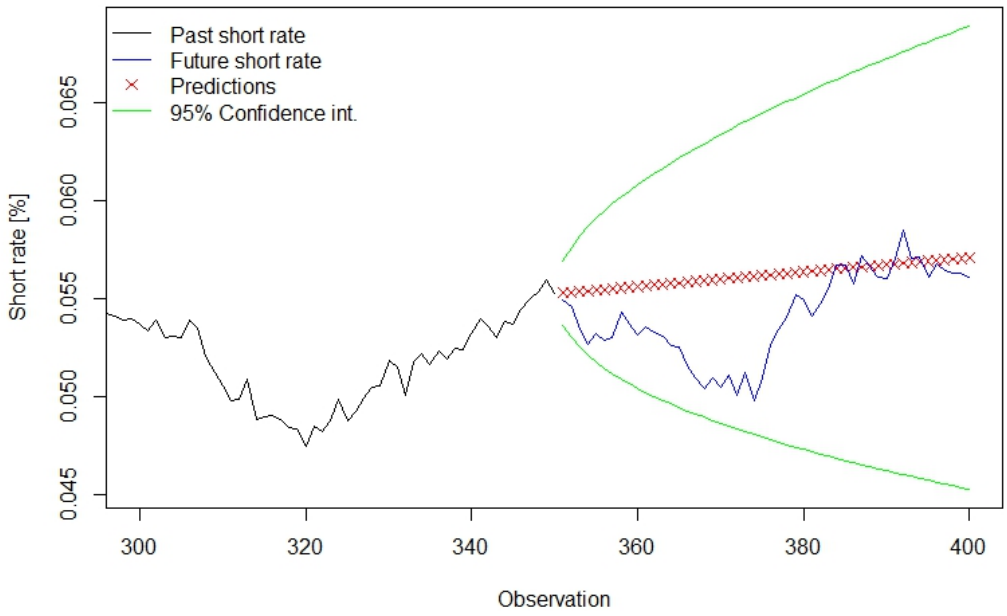
Naturally the accuracy of prediction decreases for longer predictions. The mean value basically captures only the trend of the most recent development.

To test our algorithm we leave last 50 days from simulated dataset and construct predictions for these days, based on 350 previous observations. The predictions are fairly easy to construct using the mean value from (29). Iteratively we calculate the mean value, while all other parameters remain fixed. The confidence interval is calculated using the standard formula $E(r) \pm qnorm(97.5\%)\sqrt{\frac{var(r)}{n}}$, where $qnorm(97.5\%)$ is 97.5% quantile of normal distribution. In this case the $n = 1$ since the prediction is based on only the last short rate value. We plot the real short rate, the prediction and the confidence interval in Figure 10. The predicted interest rates, which are based on predicted short rate and optimized parameters, are on average 0.1% different than the exact values (see Table 7).

1Y	2Y	3Y
Min. :0.003485	Min. :0.002797	Min. :0.002289
1st Qu.:0.056825	1st Qu.:0.045591	1st Qu.:0.037305
Median :0.149360	Median :0.119834	Median :0.098053
Mean :0.174393	Mean :0.139918	Mean :0.114487
3rd Qu.:0.244603	3rd Qu.:0.196249	3rd Qu.:0.160579
Max. :0.501343	Max. :0.402238	Max. :0.329131
4Y	**5Y**	**10Y**
Min. :0.001909	Min. :0.001619	Min. :0.0008741
1st Qu.:0.031102	1st Qu.:0.026387	1st Qu.:0.0142415
Median :0.081750	Median :0.069356	Median :0.0374333
Mean :0.095452	Mean :0.080981	Mean :0.0437077
3rd Qu.:0.133881	3rd Qu.:0.113584	3rd Qu.:0.0613041
Max. :0.274410	Max. :0.232809	Max. :0.1256544

Tabuľka 7: Summary of absolute residuals between exact and predicted interest rates for 50 days in [%].

It is obvious the predictions of short rate are not too good, where the third prediction is already far from the reality. The confidence interval is getting wide fast due to

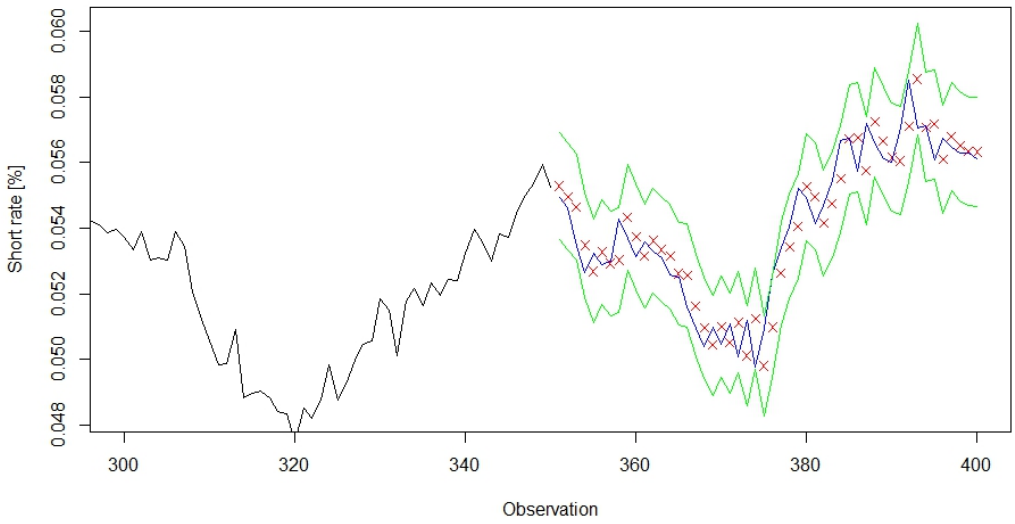


Obr. 10: Prediction of short rate for 50 days with 95% confidence interval in comparison with exact short rate.

increasing variance. To fix this, we try different approach.

Instead of predicting all 50 days at once, we predict only 1 day ahead, wait one day to obtain real data for that day and then again predict 1 day ahead. This way we always use the correct information from previous point and prevent accumulating of errors. The parameters κ, θ, γ and σ remain fixed, the market price of risk is iteratively estimated based on all days previous of the prediction, so even the λ is estimated only 1 day ahead. This prediction should be much more accurate than the 50 days predictions above and Figure 11 and Table 8 confirm it. The confidence interval is much more narrow since the errors are not accumulating and predictions are much more precise. The interest rates are on average 0.03% different than the exact values.

Being able to predict interest rates with precision around few basis points is good enough for us. However, we have to keep in mind that this dataset was simulated by CKLS process and therefore ideal for CKLS one-factor model. Later we use different datasets, which may not be suited for this model and the predictions may serve as an indication whether the model is appropriate for the dataset. The objective is to compare predictions of various models and finding out whether adding more stochastic



Obr. 11: Prediction of short rate 1 day ahead 50 times with 95% confidence intervals in comparison with exact short rate.

1Y		2Y		3Y	
Min.	:0.001141	Min.	:0.0009149	Min.	:0.0007482
1st Qu.	:0.027792	1st Qu.	:0.0222975	1st Qu.	:0.0182444
Median	:0.047485	Median	:0.0380982	Median	:0.0311737
Mean	:0.054328	Mean	:0.0435882	Mean	:0.0356658
3rd Qu.	:0.082232	3rd Qu.	:0.0659757	3rd Qu.	:0.0539841
Max.	:0.126187	Max.	:0.1012427	Max.	:0.0828422
4Y		5Y		10Y	
Min.	:0.0006236	Min.	:0.000529	Min.	:0.0002854
1st Qu.	:0.0152108	1st Qu.	:0.012905	1st Qu.	:0.0069649
Median	:0.0259907	Median	:0.022050	Median	:0.0119012
Mean	:0.0297360	Mean	:0.025228	Mean	:0.0136162
3rd Qu.	:0.0450084	3rd Qu.	:0.038185	3rd Qu.	:0.0206094
Max.	:0.0690693	Max.	:0.058598	Max.	:0.0316275

Tabuľka 8: Summary of absolute residuals between exact and predicted interest rates 1 day ahead 50 times in [%].

factors is worth to achieve better predictions.

2.6 Two factor model

Previous sections described core methods for calculating bond price, optimizing parameters and predicting future observations. These core methods are easily applicable to one-factor CKLS model with great results. For the ideal dataset the algorithm finds correct parameters and short rate and the predictions are relatively precise. However,

for other datasets, not simulated exactly for CKLS one-factor model, the model might produce worse results.

That is what multi-factor models may improve. Introducing another random factor into the equation the process becomes more complex, but can describe more complex behaviour of interest rates. For example, two-factor models are widely used for convergence models, when one interest rates is converging towards another and each one is modelled with a different factor (more information in [10]). We are interested in the difference between the one-factor and multi-factor models, while using the same dataset. How many factors is the optimal number to get the best fit or predictions? What are the costs and benefits of adding additional factors? To answer these questions, we must first be able to calibrate such models. Calibrating one-factor model in previous sections is core of how we want to approach calibration of multi-factor models. Lets begin with two-factor CKLS model.

In section 1.6 we reduced the problem of solving multi-factor bond price PDE to one-factor bond price PDE if the factors are independent. Thanks to that we are able to use the ETI method from section 2.1 again. We perform discretization of each factor separately and compute the bond prices for each. Product of these prices equals prices under two-factor model with independent factors. Basic logarithm transformation leads to yields for the two factor model.

Same as before comparing estimated yields with the observed is done by calculating objective function F, which is the same as in Equation(27). The difference is that, the estimations come from a different model.

Optimizing parameters $\kappa_1, \theta_1, \gamma_1, \sigma_1, \kappa_2, \theta_2, \gamma_2, \sigma_2$ is done by minimizing the objective function using again Nelder-Mead algorithm. The convergence is highly influenced by the choice of the starting point and from our empirical results the objective function does not appear to be convex at all. Often times the optimization converges only to local minima depending on the initial point, but even such initial points that are close to optimum does not ensure convergence to optimum.

There are many options how to deal with incorrect convergence. As mentioned before, other algorithms, like genetic heuristics, are capable to explore the objective function and find new local minima until it reaches global minima. These methods are

usually worse at exploiting the current minima, meaning they are much less effective when it comes to precision of the optima. A solution could be combination of both exploring and exploiting methods.

Simpler way to reach global minima instead of local minima using Nelder-Mead algorithm is resetting the algorithm. Nelder-Mead algorithm works with a simplex in the space of parameters which are to be optimized. Using clever manipulations the simplex test new points and if they lead towards better objective value, the simplex moves to the new points. If none manipulation leads to better objective value, the simplex shrinks its volume until termination criteria. Theoretically and with a bit of luck the simplex manipulations are able to explore new local optima instead of stopping at first local optimum. However, if the simplex shrinks too much, the reach of the simplex manipulations decreases and chances to explore new optimum with it. It is convenient to reset the volume of the simplex after sufficient amount of iterations to improve the reach of the simplex manipulations. This volume resetting can be done simply by optimizing again, while taking the result of previous optimization as a new starting point. However, it is unknown how many local optimums the function has and thus how many times the optimization has to be reset.

We use data simulated by a two factor CKLS model as an example. These data are ideal for the model and optimal parameters should not be hard to find. The data are generated by independent Wiener processes, added to two CKLS factors and then summed up. From short rate of each factor yields are estimated using ETI method. The exact parameters used for simulation are shown in Table 9. To demonstrate the effect of restarting Nelder-Mead optimization again from previous optimum, we show results both after first optimization and after second optimization using the first optimum. The starting point we used for results shown in the Table 9 is $\kappa_1^{(0)} = 0.60, \theta_1^{(0)} = 0.07, \gamma_1^{(0)} = 0.6, \sigma_1^{(0)} = 0.15, \kappa_2^{(0)} = 0.50, \theta_2^{(0)} = 0.01, \gamma_2^{(0)} = 0.8, \sigma_2^{(0)} = 0.13$.

Table 9 shows the achieved optimum for two-factor model together with exact values and optimum achieved by one-factor model. The objective value for exact parameters is 8.3×10^{-16}, while theoretically it should be zero. This difference originates from numerics and possibly is bottlenecked by the precision of the ETI method for solving PDE. It is almost negligible, but creates an issue, when another slightly different set of

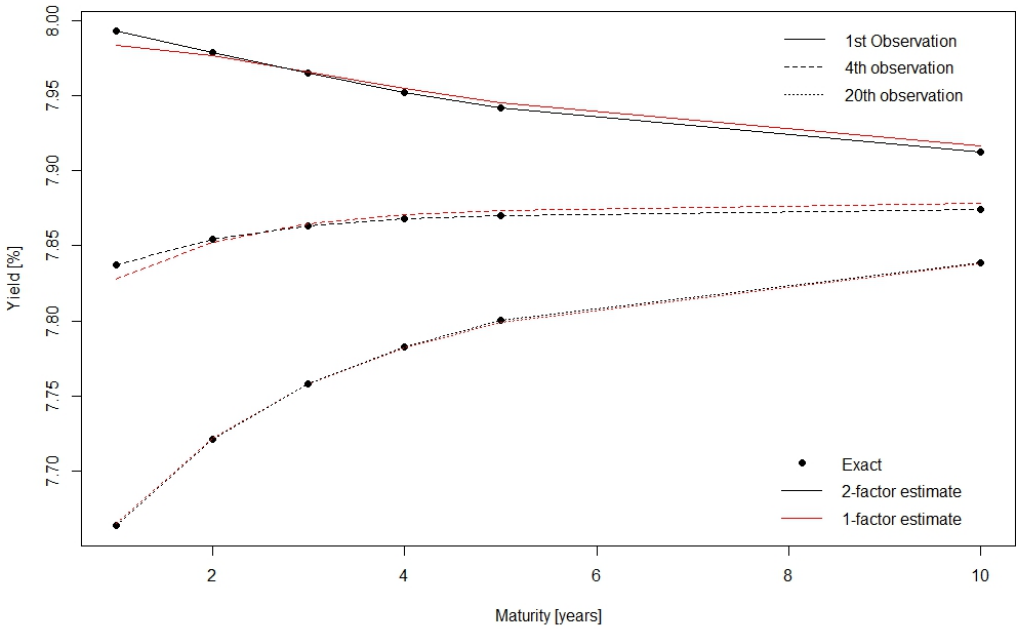
	κ_1	θ_1	γ_1	σ_1	κ_2	θ_2	γ_2	σ_2	F
Exact	0.50	0.06	0.50	0.10	0.40	0.02	0.70	0.08	8.3e-16
2F 1st opt.	0.500	0.060	0.583	0.121	0.398	0.021	0.568	0.085	2.0e-13
2F 2nd opt.	0.500	0.060	0.500	0.100	0.400	0.020	0.918	0.169	1.7e-15
1F	0.498	0.081	0.684	0.172	-	-	-	-	7.3e-10

Tabuľka 9: Results of parameters optimization for simulated data for two-factor CKLS model. First row contains true parameters and value of objective function with exact parameters. Second and third row contain results of first and second optimizations of two-factor CKLS model parameters. Last row serves as a comparison between two-factor and one-factor CKLS model, for which the optimization approach from previous sections is used.

parameters might achieve slightly better objective value F. For example if γ_2 equal to 0.9 instead of 0.7 leads to lower values of F because of the numerics error and precision of ETI method. Looking at the second optimization run for two-factor CKLS model it appears that the situation from previous example is very much possible. Indeed after resetting optimization few more times, new optimum with $\gamma_2 = 0.9$ and $\sigma_2 = 0.175$ occurs with objective value slightly lower than the 8.3×10^{-16}. Therefore the parameters are slightly different, but the objective value difference is very small and both sets of parameters fit data almost equally.

It is interesting to see the difference between using two-factor and one-factor model for the same dataset. While certainly the fit of one-factor model is not bad with objective value 7.3×10^{-10} it is worse than two-factor model. This is the expected effect of two-factor models, since the more parameters model has, the better precision it will achieve. Similar logic as in basic linear regression. This simulated dataset is ideal for two-factor model and so the one-factor model can not fit the data perfectly. Following figure compare these two models and how well they fit the data.

Figure 12 proves that two-factor model is better suited for this dataset, but for example in the 20th observation the difference between the two models is barely visible. Of course the fact that one-factor model is relatively good is influenced by the parameters we chose for the data simulation. Different choice of parameters to simulate could lead to more distinct short rate processes which would be harder to estimate for one-factor model. The mean value of error between estimated yields in % and observa-



Obr. 12: Exact yields with two-factor and one-factor CKLS model estimates for 3 separate observations.

tions over all maturities is 4.2×10^{-6} for two-factor model and 1.9×10^{-3} for one-factor model. The difference is significant, even though it is not as visible in Figure 12. The estimation from one-factor model seems to have troubles with shortest and longest maturities, while the middle maturities seem to have better precision. This phenomenon is common when using one-factor models on datasets, which are not ideal for it. Using different weights in objective function (27) could lead to different estimations, where for example long term yields are more precise. Following Table 10 shows detailed summary of absolute differences between observed yields and estimations of the two-factor CKLS model.

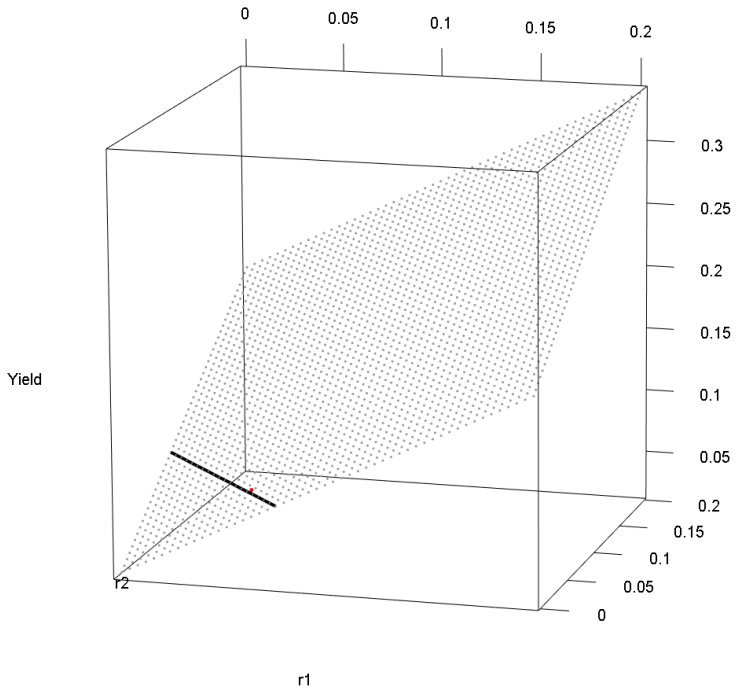
In previous paragraphs we omitted the fact that short rate is unknown. In case of two-factor models we need knowledge of both short rates of each factor to be able to estimate bond yields. In reality we do not know these short rates and we need to estimate them. For one-factor model we simply picked the best possible estimated yield and corresponding short rate value for each observation. However, similar approach for two-factor models becomes more or less impossible. Instead of vector of yields as in one-factor model we now have matrix of yields, where each element corresponds to different combination of two short rates. Therefore the computability becomes costly. Adapting

1Y	2Y	3Y
Min. :1.866e-07	Min. :5.770e-08	Min. :4.895e-09
1st Qu.:2.501e-06	1st Qu.:1.854e-06	1st Qu.:2.352e-06
Median :4.834e-06	Median :3.900e-06	Median :3.326e-06
Mean :4.819e-06	Mean :3.869e-06	Mean :3.419e-06
3rd Qu.:6.772e-06	3rd Qu.:5.177e-06	3rd Qu.:4.611e-06
Max. :1.581e-05	Max. :1.157e-05	Max. :7.290e-06
4Y	5Y	Y10
Min. :1.494e-07	Min. :8.917e-08	Min. :1.569e-07
1st Qu.:2.974e-06	1st Qu.:1.780e-06	1st Qu.:2.029e-06
Median :4.166e-06	Median :4.392e-06	Median :3.554e-06
Mean :4.313e-06	Mean :4.766e-06	Mean :4.160e-06
3rd Qu.:5.807e-06	3rd Qu.:7.086e-06	3rd Qu.:5.938e-06
Max. :9.289e-06	Max. :1.203e-05	Max. :1.313e-05

Tabuľka 10: Summary of absolute differences between observed and estimated yields in [%]. Estimation with known short rates.

this approach for two-factor model would require searching huge matrix for closest match with the data point for each maturity and each observation. Sure the matrix consist of points, which are impossible or unlikely to even be relevant, but even striping matrices of these points does not help enough. The method is still time consuming and the precision is very bad, even if we implement interpolations between 2 or more closest matches. In Figure 13 we visualize the problem in three dimensions - first short rate, second short rate and the corresponding yield. The gray points represent matrix of bond prices resulting form all possible combinations of short rates from discrete set r_i coming from ETI method. The red point represents correct observation of given maturity. The black line segment highlights all estimates with yield closest to the observed yield. Luckily with correct parameters of model the black line is not constant and it is possible to find out which point on the line is closest to the correct red point just by looking at the yields. However, the precision of this estimation highly depends on the discretization step in each factor and desired precision would require too much time.

The solution seems to be another optimization algorithm such as BFGS, since the problem appears to be convex. Using BFGS we are able to find short rates of each factor under reasonable time. The time obviously depends on the number of observations in the dataset, since each observation creates two short rate variables to optimize. With correct parameters of the model we are able to estimate short rates with average

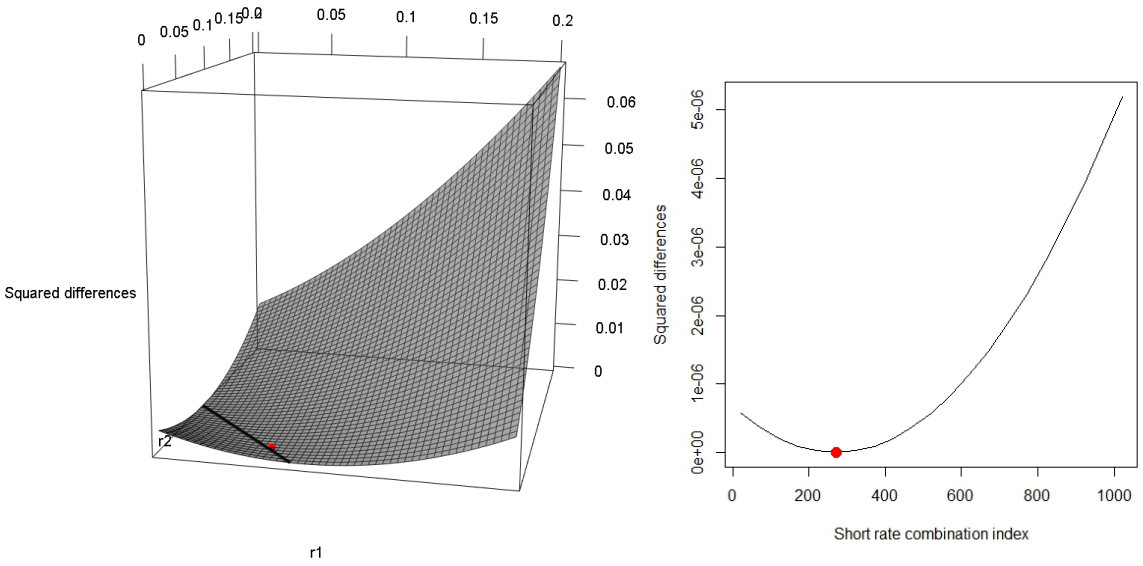


Obr. 13: Visualization of the task to find the optimal short rate based on the observed yield (Red). Gray points represent possible estimated yields based on all possible combinations of short rates. The black line highlights area of points with estimated yield close to observed yield.

precision 9.6×10^{-6} for the simulated dataset. The precision should not be underestimated since even from Figure 13 we see how difficult it is to derive both short rates just using the observed yield. With correct model parameters simply subtracting the observed yield and taking squared differences will lead towards the correct red point in the image. In Figure 14a we visualize the squared differences and in Figure 14b we zoom into the squared differences on the black line from previous graphs.

Figures 14a and 14b suggests that the short rate optimization problem is convex. Therefore using algorithm like BFGS might be very efficient and outperform the simple search through the matrix element by element. Finding both short rates is possible, even though we work with just one observed yield for one given maturity.

Summing up both short rates results into total short rate for both factors. It is clear that even the the optimized short rates should result into same total short rate. Indeed looking at the Figure 14a the best combinations of short rates, which are denoted by black line, appear to follow this rule $r2 = r_{total} - r1$. Total short rate is also unknown, but this information can pave the way for new more sophisticated and more efficient



(a) Squared differences of observed and possible estimations.

(b) Detailed view of differences in area of closest match. Red dot marks optimum.

Obr. 14: Visualization of short rate optimization problem in 3D and then detailed 2D view. The black line denotes combinations of short rates leading to closest match with observed yield. The problem appears to be convex under correct model parameters.

optimization methods.

So far we have worked either with known short rates or correct model parameters. We are able to optimize parameters with known short rates and optimize short rates with know parameters. The challenge is doing both simultaneously. At start best we can do is roughly guess some parameters. We can even use short maturity yields as a proxy for short rate, because we have data only about the total short rate, which is aggregate of the two short rates, that we need. This problem is not very well documented and we had to improvise and design at least somewhat reliable heuristics.

One option is select starting parameters for which we optimize both short rates and then use them to estimate new parameters. Repeating this procedure iteratively would be way to move and explore new sets of parameters, but we had very little success with this method. The estimated short rates are very sensitive to the parameters and the starting point would need to be close to perfect for this method to work reliably. However, choosing perfect starting point is impossible without luck.

Another option is to just optimize short rates within parameters optimization, meaning in each iteration of Nelder-Mead algorithm we estimate short rates using BFGS

for given parameters and then evaluate objective function. This may sound very time demanding, but at this point all methods will require much more power and time than the algorithm for one-factor model. Again the convergence is highly influenced by the choice of the starting point, but the range of starting points is certainly wider than in the first approach. This method quite reliably finds local minima with very low objective value, which is what we need. Optimizing again could lead towards the global minimum. The method is very time consuming, so each optimization reset is very costly.

Third option is using simpler two-factor models with analytical solutions, which may help with the short rate optimization. In article [2] they use clever linear regression to estimate two-factor Vasicek model without correlation. The input for the regression are only parameters κ_1, κ_2 and data about yields. The output of the model are approximate short rates. In the article the authors prove, that the estimated short rates are biased by a constant equal to θ_2. The disadvantage of using Vasicek model is possibility of negative short rates, which are unacceptable for a CKLS model. Using this approach the estimation of total short rate is fairly easy and precise with average accuracy 1.8×10^{-5}. We can use the information about total short rate as a better staring point for previous methods or to optimize only one of the short rates and calculating the other one.

The total short rate can be also approximated by the optimized short rate using one-factor CKLS model. Even though the model is not suitable for the data, the optimized short rate is very close to the total short rate of the two-factor model. The difference between them is on average 9.5×10^{-5}, so the precision is little bit lower than with the approach using two-factor Vasicek model.

We designed various algorithms each using different option of short rate optimization. With them we are able to find suboptimal points with low objective value, but none of them succeeded in finding the true optimum. The problem is that with so many parameters uniqueness of the solution is in question. Parameters θ_1, θ_2 are closely tied with the short rates of the factors. Bias in parameters θ_1, θ_2 will gen compensated with adjusted short rates, therefore we can never obtain true optimum without knowing either true θ_1, θ_2 or true short rates. Similar relation is also mentioned in [2] for the case of two-factor Vasicek model, where different choices of parameters α_1, α_2 would

lead to same objective value.

Without external information it is impossible to reach the true optimum except trying various starting θ_1, θ_2 parameters. However, even these suboptimal parameters and short rates are able to fit the data sufficiently well. After some testing we designed two algorithms that lead to best results. Both use BFGS short rate optimization within each iteration of Nelder-Mead parameter optimization and both use the two-factor Vasicek model to guess parameters κ_1 and κ_2 for better starting point. The difference is whether we optimize both short rates or only one short rate, while the second short rate is calculated with estimated total short rate from two-factor Vasicek model. The second methods relies on the precision of total short rate, which is often not sufficient. However, the first method requires significantly more time, because it optimizes both short rates. The final algorithm consists of:

1. Using linear regression estimate **two-factor Vasicek model** for various combinations of κ_1, κ_2. The combination leading to the best fit is used for the initial point. Using the best κ_1, κ_2 estimate total short rate as sum of both short rates, which are estimated during linear regression.

2. Choose parameters for the **initial point**.

3. **Evaluate estimated** bond prices and yields with ETI method for the given parameters separately for each factor. The estimated yields depend on short rates.

4. **Find short rates** that lead to best match between estimated and observed yields:

   A Use BFGS to optimize both short rates to the observed yields.

   B Use BFGS to optimize only one short rate, while the second short rate is calculated as $r_2 = r_{total} - r_1$, where r_1 is the optimized first short rate.

5. Calculate the **objective function** F with the penalization for the boundaries similar as in one-factor algorithm.

6. **Find new parameters** leading to lower objective value F and **repeat** from point 3. This is done using Nelder-Mead algorithm for all eight parameters.

The variant B restricts the optimized short rates, while the variant A is unrestricted. Therefore with variant A often ends up at the same point as the initial point, because the short rates are optimized for those parameters. That may prove to be a problem for correct predictions, which are based also on the parameters. The variant B slightly restricts the short rates, but not enough to prevent similar phenomenon. Instead of searching for more suitable parameters, the short rates are optimized to compensate the error. For larger datasets the variant A becomes unusable, since the optimization will require very long time.

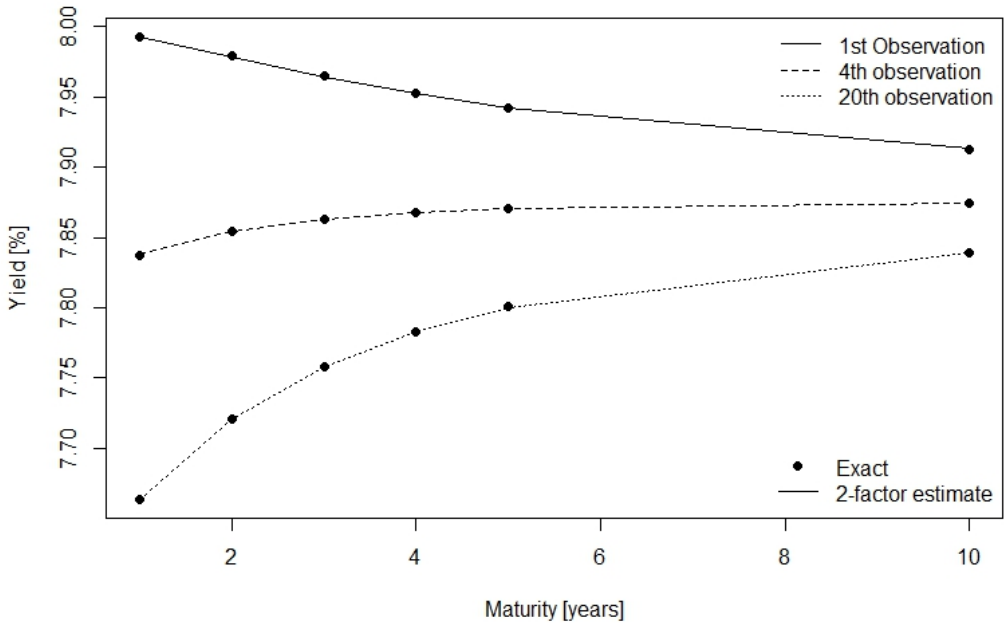
	κ_1	θ_1	γ_1	σ_1	κ_2	θ_2	γ_2	σ_2	F
Exact	0.50	0.06	0.50	0.10	0.40	0.02	0.70	0.08	8.3e-16
initial parameters	0.60	0.04	0.60	0.15	0.50	0.04	0.80	0.13	5.1e-9
2F optimum B	0.52	0.041	0.50	0.10	0.36	0.04	0.78	0.20	4.7e-12

Tabuľka 11: Results of parameter optimization for simulated data for two-factor CKLS model. First row contains true parameters and value of objective function using exact short rates. Second row consists of the initial parameters entering the optimization and the objective value using optimized short rates with method B for the initial parameters. Third row contains optimum after parameter and short rate optimization using algorithm B.

The optimized parameters in Table 11 are obviously distant from the exact parameters. However, note the objective values. All objective values are relatively small due to the short rate optimization, which optimizes short rates even for the incorrect parameters. The objective value using algorithm with variant B uses total short rate as input. If we use all parameters and short rates exact, the result is the objective value for the first row in the table. But the total short rate is approximated from the two-factor Vasicek model with precision around 1×10^{-5}. With such approximated total short rate the objective value for the exact parameters decreases to 2.8×10^{-11}. The optimum with incorrect parameters then appears better than the true optimum, which is unfortunate. This means that the estimation of total short rate with two-factor Vasicek model is not accurate enough, but is still the most robust and reliable approximation we have when the exact parameters are unknown.

Even though the optimized parameters are incorrect and short rates with them, the objective value is very low. Let us compare the observed yields with estimations using

algorithm with variant B.



Obr. 15: Exact yields for 3 separate observations compared to estimated yield using two-factor CKLS model without known short rates.

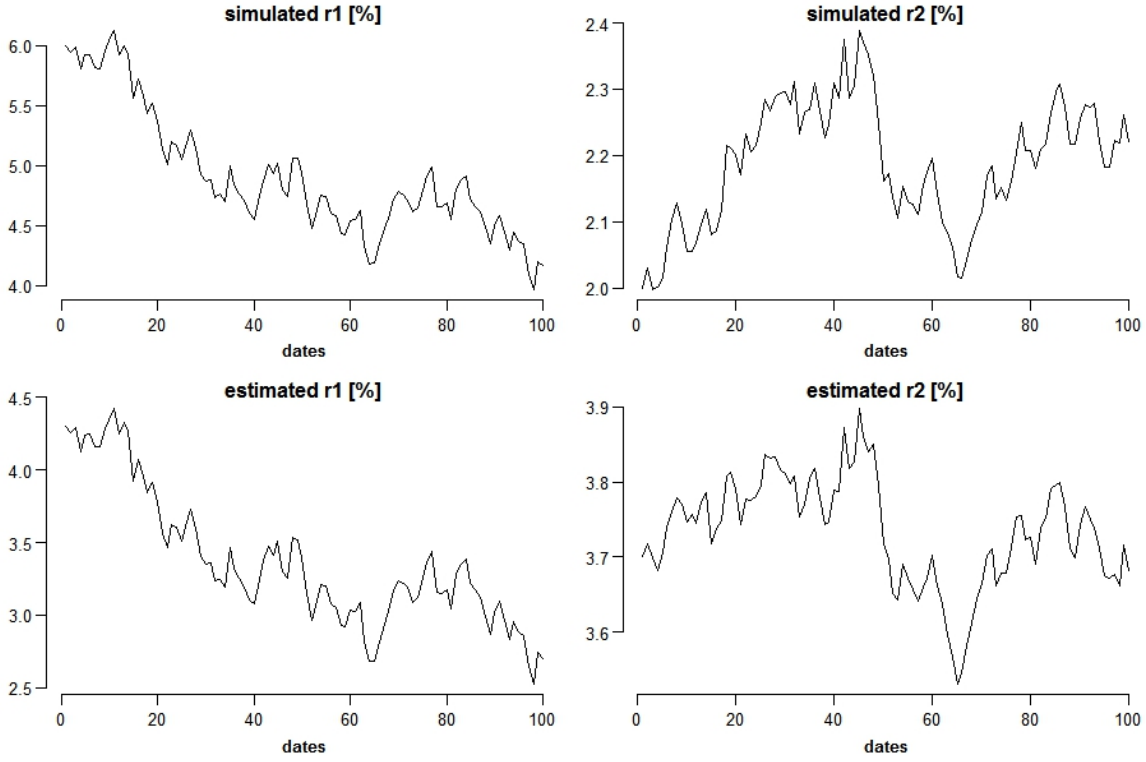
In Figure 15 we demonstrate how well the estimations fit the data for the same three observations as in Figure 12, where we used known short rates for optimization. Even though the optimized parameters and short rates are heavily biased the differences between estimations and observations are almost invisible. To see full picture of the precision we show summary of absolute differences in % for each maturity separately in the following Table 12.

Keep in mind that all differences in Table 12 are in percentage points, so for example for the average difference for 4 year yields is 0.00006%. The precision is worse than with known short rates summarized in Table 10, but still the precision is sufficient for our needs. The average absolute difference over all maturities is around 1.3×10^{-4}, which is still better than the one-factor model applied to the same data.

In Figure 16 we visualize estimated short rates with the exact short rates, that were used for the simulation of the dataset. The graphs of estimated and exact short rates are very similar, almost identical except the range on Y axis - yield in %. This means the short rates are almost optimal, but each observation is shifted by an unknown constant.

1Y	2Y	3Y
Min. :1.180e-06	Min. :4.593e-07	Min. :8.079e-06
1st Qu.:8.649e-05	1st Qu.:6.752e-05	1st Qu.:8.698e-05
Median :1.350e-04	Median :1.128e-04	Median :1.393e-04
Mean :1.551e-04	Mean :1.300e-04	Mean :1.507e-04
3rd Qu.:2.198e-04	3rd Qu.:1.803e-04	3rd Qu.:2.187e-04
Max. :4.397e-04	Max. :3.160e-04	Max. :3.549e-04
4Y	**5Y**	**Y10**
Min. :3.367e-07	Min. :2.893e-06	Min. :2.373e-06
1st Qu.:2.731e-05	1st Qu.:3.810e-05	1st Qu.:1.111e-04
Median :5.088e-05	Median :5.803e-05	Median :1.742e-04
Mean :6.156e-05	Mean :5.811e-05	Mean :2.363e-04
3rd Qu.:8.670e-05	3rd Qu.:7.396e-05	3rd Qu.:3.411e-04
Max. :2.137e-04	Max. :1.248e-04	Max. :6.838e-04

Tabuľka 12: Summary of absolute differences between observed and estimated yields in [%]. Estimation without knowing correct short rates.



Obr. 16: Comparison of estimated short rates factors and the exact short rates used in simulation.

In article [2] authors observe similar phenomenon, but for two-factor Vasicek model with independent factors. Due to the analytical tractability of Vasicek model, authors were possible to identify the off-set for all observations as a constant equal to $\pm\theta_2$. This fact comes from the direct relation between each short rate and parameters $\alpha_{1,2}, \beta_{1,2}$

(in our notation the risk-neutral parameters $\theta_{1,2}, \kappa_{1,2}$). Therefore in optimization of parameters, authors optimize only „total“ α and can not determine $\alpha_{1,2}$ without any additional input. Any choice of α_1 and calculated corresponding α_2 would lead to the same optimum.

In our case the off-set appears to be constant over all observations as well. The offset for the same simulation is between $\pm < 0.01343, 0.01527 >$ and on average ± 0.01412. Unlike in the case of two-factor Vasicek model, the constant offset is not equal to the parameter θ_2.

So the fit appears to be fine even if the optimization did not find the correct optimum. What about predictions? Will the offset of parameters and short rates become recognizable when predicting new yields? To test this we use similar procedure as with one-factor model in chapters 2.5.

The data are simulated with market price of risk for each factor, specifically $\lambda_1 = 0$ and $\lambda_2 = 0.5$. In chapter 2.4 we use MLE estimator for the approximate market price of risk for one-factor CKLS model. In case of two-factor CKLS model with independent factors, which is important, the factors are simply summed up. Therefore we estimate market price for each factor separately based on parameters and short rate of the corresponding factor. Of course the MLE is based on the assumption that the short rate is normally distributed with the correct parameters (29). On the other hand the short rates are biased too and in such way, that the incorrect parameters are correct for the biased short rates.

The estimated market prices are $\widehat{\lambda}_1 = -2.262$ and $\widehat{\lambda}_2 = -0.054$. The MLE estimator is biased, so the difference compared to the simulated market prices of risk is reasonable. As a test we estimated the market prices of risk using the the exact parameters and short rates used for simulating data. These exact market prices are $\lambda_1^* = -2.195$ and $\lambda_2^* = 1.271$. There is apparent difference and again the estimates appear to be shifted by a similar constant. This might be caused by the shift in the short rates, which are used to estimate market price of risk. Since we find no reason to reject these estimations of market price of risk, we continue to work with them.

All is ready for constructing predictions of the model for the future yields. Using the same simulation we generate 50 future observations, which will be used to test our

model. Since we are unsure whether the incorrect parameters impair the prediction ability of the model, we also construct predictions with a model with exact parameters and short rates that were used to simulate the dataset. Therefore the prediction of the second model are certainly correct and serve as a control group for the first model.

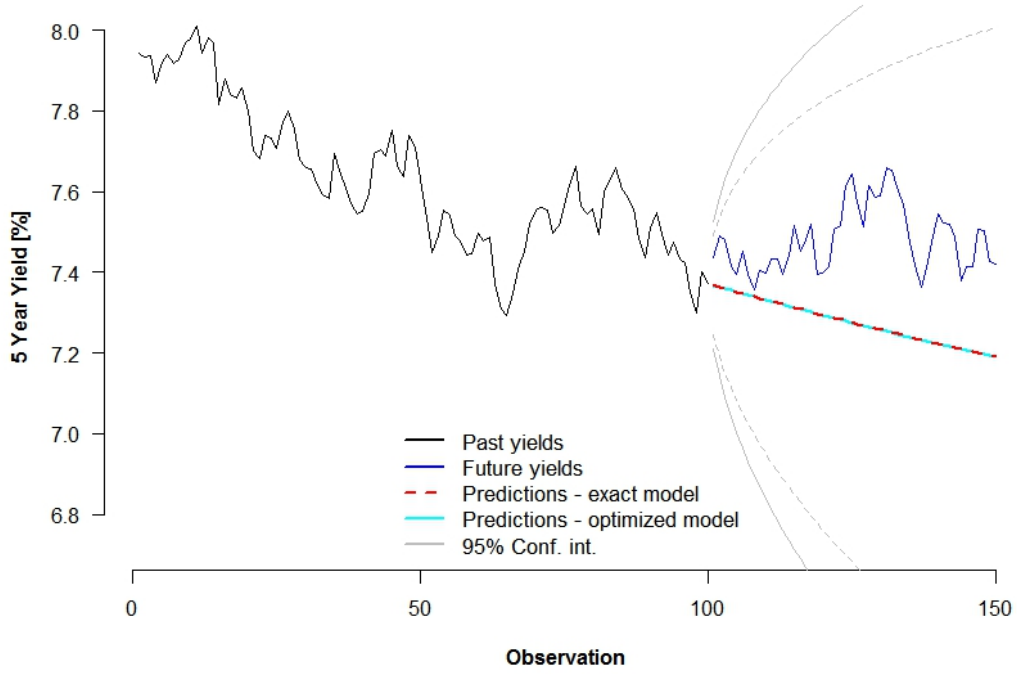
First we will use both models to predict 50 days ahead at once. The model parameters are used to construct matrices A_1 and A_2, which are the the result of finite differences method of solving PDE for each of the factors separately. Each of the short rates follow normal distribution described in (29) except now with the corresponding parameters,

$$r_{1,t+\Delta t} \sim N(r_{1,t} + (\kappa_1(\theta_1 - r_{1,t}) + \lambda_1\sigma_1 r_{1,t}^{\gamma_1})\Delta t, \sigma_1^2 r_{1,t}^{2\gamma_1}\Delta t)$$
$$r_{2,t+\Delta t} \sim N(r_{2,t} + (\kappa_2(\theta_2 - r_{2,t}) + \lambda_2\sigma_2 r_{2,t}^{\gamma_2})\Delta t, \sigma_2^2 r_{2,t}^{2\gamma_2}\Delta t).$$

The prediction of short rates are the mean values of their distribution and the variance is useful for calculation of confidence intervals for each of the factors. Both parameters and short rates decide the future predictions, which means the predictions of the two models should be different. Since the data are simulated, the exact short rate is available for exact predictions. In Figure 17 we portrait predictions of 5 year bond yields in % for the simulated dataset.

Note how the predictions of optimized model align with predictions of exact model in Figure 17. The difference between them is negligible, except the confidence interval. The confidence interval for optimized model is visible wider, which can be caused by the small difference in the models. These small differences then accumulate over 50 days widening the gap between the confidence intervals. Same phenomenon occurs for other maturities as well. The predictions appear to follow significantly different trend than the future observations. The cause of such imperfect predictions can not be the bias in optimized parameters and short rates, since both optimized and exact model lead to same predictions.

The problem lies with the parameter λ, it appears that the MLE estimates for both optimized and exact model are greatly biased. The variance of MLE estimator is influenced by the number of observations, so for smaller samples the bias may be a problem. In addition to that in two-factor model we estimate two market prices of risk and both are subject to the bias. In this case it appears the bias was too great, which



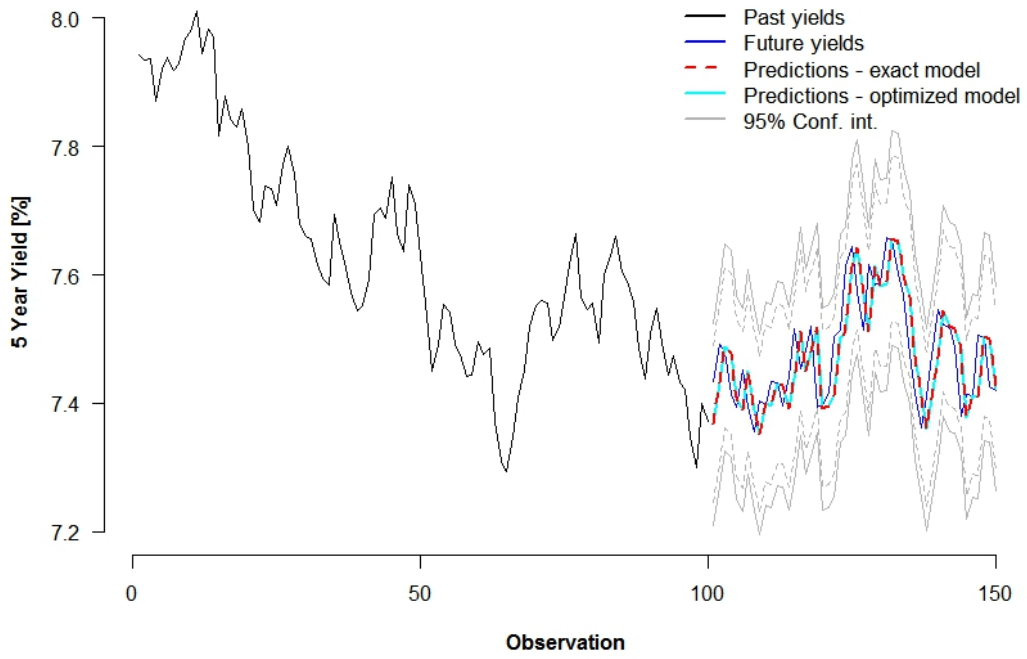
Obr. 17: Predicted yields for 50 days ahead using model with optimized parameters and short rates and model with exact parameters and short rates. The gray lines represent 95% confidence intervals one for each of the models (full line for the optimized model, dashed for the exact model).

led to poorer predictions.

Since using bigger dataset may sometimes be unavailable or significantly prolong the procedure, we deal with poor predictions differently. Instead of predicting 50-days ahead, we settle down for 1-day prediction. Same as we done with one-factor model predictions. Today we predict for tomorrow using today's observation, tomorrow we predict again for next day using tomorrow's observation and so on for 50 days. Figure 18 plots such predictions again comparing both exact and optimized model.

These predictions appear to be better, as it should be when predicting for shorter periods. The predictions now basically trace the future values with 1 day lag. It is because the predictions are unable to anticipate sudden random jumps in short rate and thus predicting only most probable value.

We already compared estimates of one-factor and two-factor model in Table 9 and Figure 12. The two-factor model visibly beat the one-factor model in the precision, but as we already mentioned the fit of the one-factor model is relatively good. Are the predictions visibly worse too? Or is one-factor model enough for predicting this

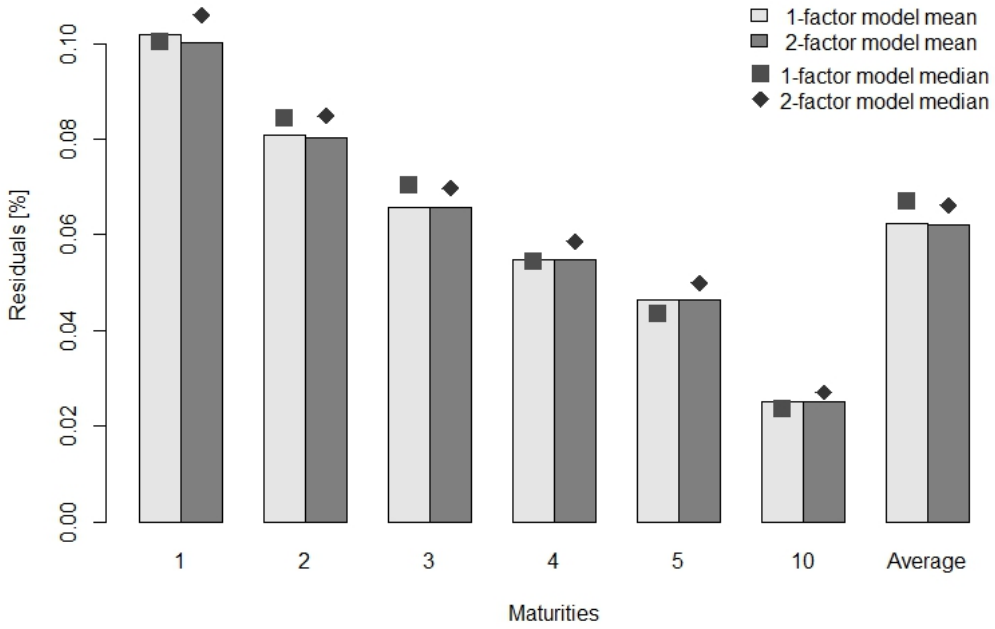


Obr. 18: 50 predictions of yields each for 1 day ahead using model with optimized parameters and short rates and model with exact parameters and short rates. The gray lines represent 95% confidence intervals one for each of the models (full line for the optimized model, dashed for the exact model).

dataset.

In Figure 19 are shown mean and median errors in 1-day ahead predictions of one-factor and two-factor model. For example average error of both models for yield with 4 year maturity is around 0.055% and the median error for one-factor model is 0.054% and for two-factor model 0.059%. It appears that the two-factor model based predictions bring little improvement over one-factor model. The average error of two-factor model is slightly lower for short maturities and almost equal in other maturities, but the median error is in favour of one-factor model. Both models produce almost same predictions. Figure 20 shows the comparison of predictions 50 days ahead at once.

Figure 20 suggest that two-factor model indeed produces slightly better predictions for longer periods. Both average and median residuals are slightly lower for predictions based on two-factor optimized model. The improvement in predictions is still very small. However, this dataset is simulated by two-factor model, which means the dataset may still be usable for one-factor model with good precision. Dealing with more complicated data may further impair predictions of one-factor model, while the two-factor model



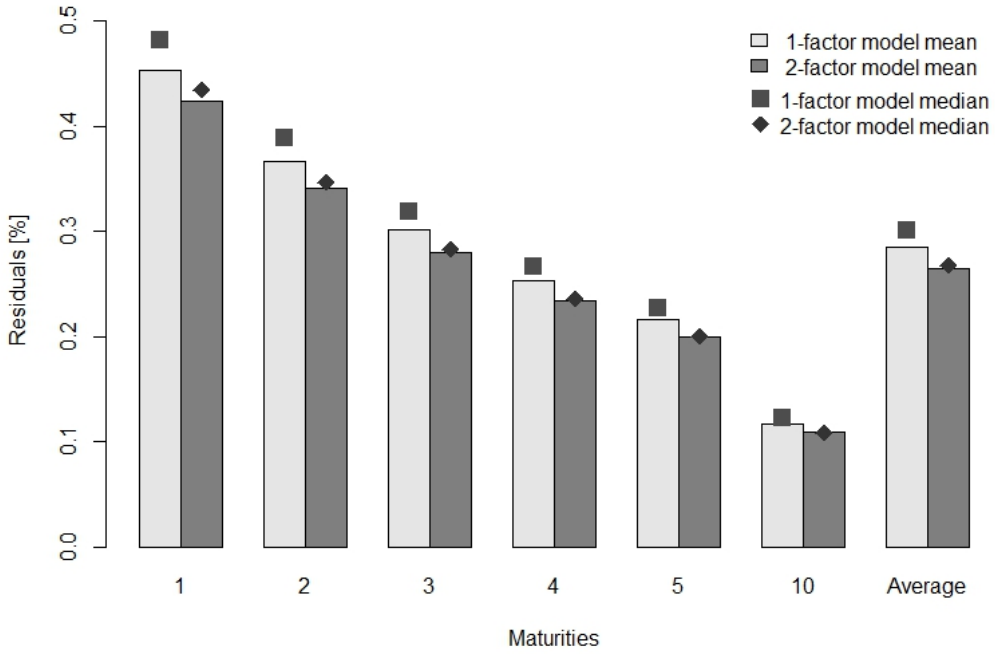
Obr. 19: Comparison of residuals in [%] between one-factor and two-factor 1 day ahead predictions.

may perform significantly better. If one-factor model is enough to fit the data then the added value of multi-factor models is minimal. Later we compare predictions of these models on real datasets.

2.7 Three-factor model

The algorithm used to calibrate two-factor model is theoretically usable for any number of independent factors. Unfortunately each added factor requires much more time. We had some difficulties even with two-factor model, when the algorithm often converges to local minima and needs resetting. The local optimum highly depends on the initial point. Having another factor means having 4 more parameters and thus guessing good initial point becomes very difficult. Based on this we are sceptical about using our algorithm on too many factors. In this section we demonstrate the possible us of the algorithm on three-factor model and analyse the results of its application to simulated data.

To calibrate model with three independent factors we use same algorithm as for

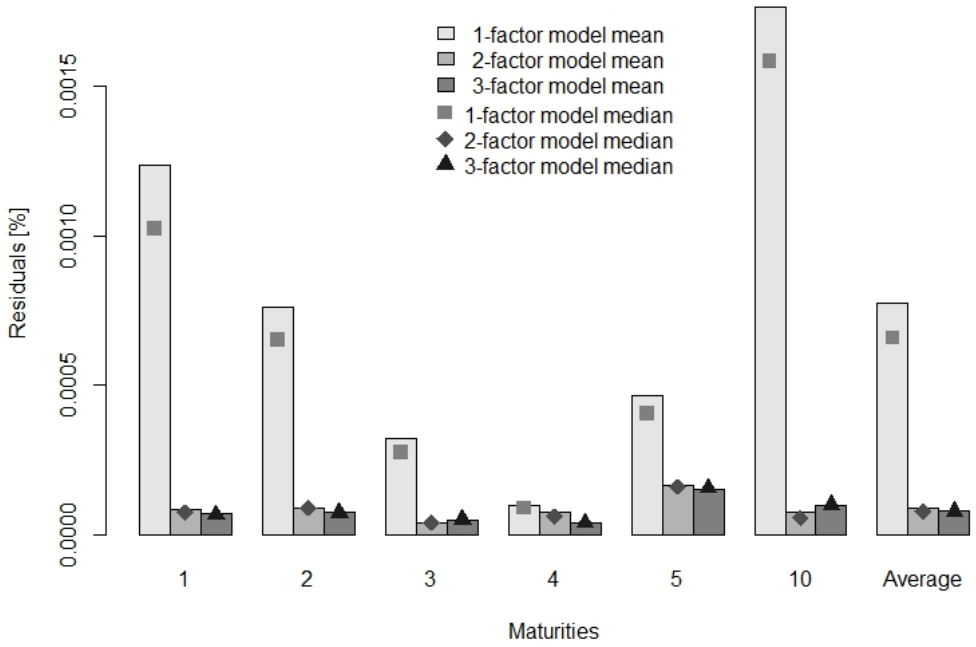


Obr. 20: Comparison of residuals in [%] between one-factor and two-factor predictions 50 days ahead.

two-factor model. First we estimate total short rate, which is the sum of all factors. Total short rate can be estimated by one-factor model or two-factor Vasicek model. Second we guess few initial points and use parallel optimization. In each iteration the short rates of factor one and two are optimized and third short rate is calculated from the total short rate. This doubles the number of optimization variables in comparison with two-factor model. Therefore the time consumption is much higher.

The data are simulated with a three-factor CKLS model and should be ideal for three-factor model. Due to the time requirements of the algorithm we work with smaller dataset. Since the data are simulated, we know exact parameters and short rates.

With the simulated dataset, which should be ideal for three-factor model, we calibrated one-factor, two-factor and three-factor CKLS models with independent factors. Listing all optimized parameters becomes pointless, since we do not know what each factor represents and thus parameters lose their explanations. What matters is the value of objective function, which should be minimal for three-factor model. Optimal objective value for three-factor model is 9.6×10^{-13}, for two-factor model 1.2×10^{-12} and for one-factor model just 1.4×10^{-10}. As expected three-factor model fits the data best,

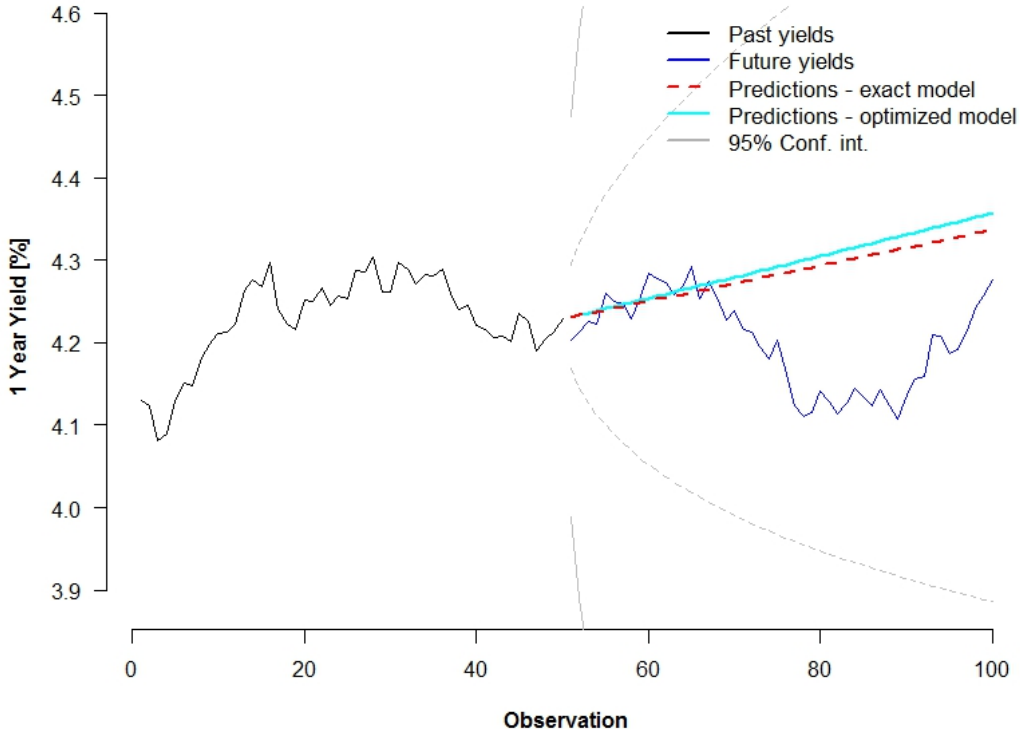


Obr. 21: Comparison of residuals in [%] between one-factor, two-factor and three-factor model estimates.

but two-factor model is not far behind. Figure 21 shows mean and median residuals of the estimated yields for each maturity with each of these models. Three-factor model consistently outperforms both one-factor and two-factor model, but the difference from two-factor model is very small.

The optimized parameters and short rates are again highly biased in comparison with exact parameters and short rates. Same phenomenon occurred while calibrating two-factor model, but the fit and predictions of exact and optimized model were alike. Also for three-factor optimized model the objective value of 9.6×10^{-13} proves that the fit is very good. Figure 22 compares predictions of exact and optimized three-factor model, while the predictions are done for 50 days ahead at once. The predictions are different, but only slightly.

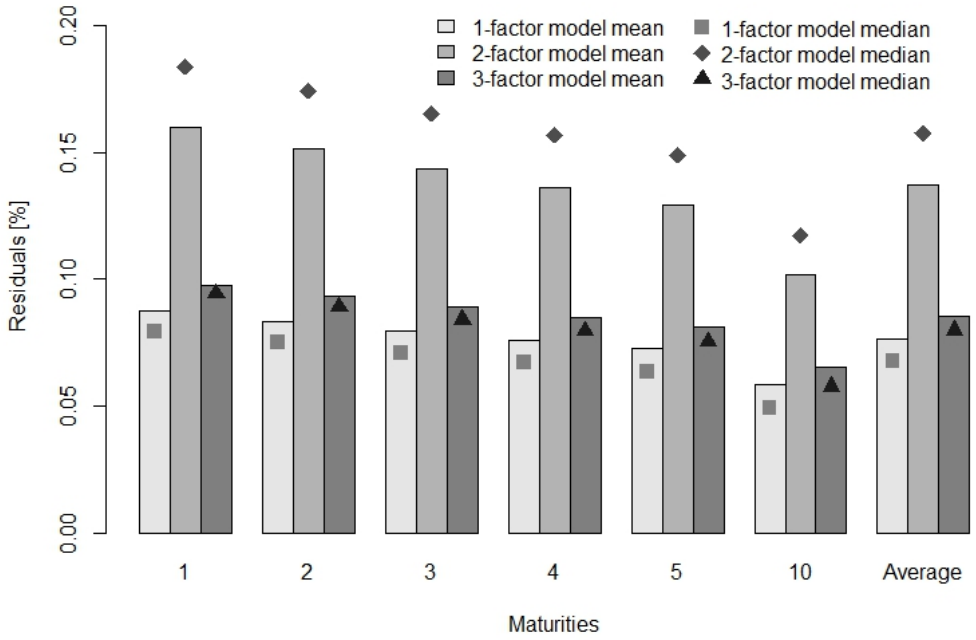
At last we compare predictions of one-factor, two-factor and three-factor model similarly as we did in previous section. Figure 23 shows the average and median errors in 50 day predictions. Two-factor appears to provide the worst predictions for this dataset compared to the other two models. Even though two-factor and three-factor models fitted the data comparably well, the predictions are significantly different in



Obr. 22: Predicted yields for 50 days ahead using three-factor model with optimized parameters and short rates and three-factor model with exact parameters and short rates. The gray lines represent 95% confidence intervals one for each of the models (full line for the optimized model, dashed for the exact model).

favour of three-factor model. However, the predictions of one-factor model appear as the best for this dataset, while performing slightly better than three-factor model. This is unexpected and very unintuitive result. Possible explanation is that the one-factor model wins for this dataset, but may perform worse for other dataset, while three-factor model would overall perform better. However, this is just a hypothesis and requires additional simulations, proofs and tests.

Overall the general idea is that more complicated models tend to fit data much better. When it comes to predictions, the differences in the fits of observed yields become inferior to the randomness of the future yields. The market price of risk has great influence on the predictions and the MLE estimator is biased. In the end it comes down to how badly biased are the market prices of risk in the models, which is partly random.



Obr. 23: Comparison of 50 days ahead predictions error in [%] between one-factor, two-factor and three-factor model.

3 Real data application

The purpose of the term structure models is to be able to describe real world problems and interest rate development. In this section we use available theory and empirical results from simulated datasets and apply it on real data. However, the theory is often simplified and real data might introduce many more unanticipated issues.

Since we are limited to CKLS models, where $\gamma \geqslant 0.5$ due to the Fichera theory, all yields have to be positive. Negative yield would lead to negative short rate, which is unobtainable with CKLS model. At times of deflation and recession interest rates tend to reach zero or even negative values. Such times were recently observable in eurozone, when EURIBOR overnight interbank rate reached -0.4% last year. Massive quantitative easing for past several months suppressed deflation and interest rates show positive trend lately. Therefore modelling negative yields may be desirable. Technically, substituting one of the CKLS factor in multi-factor models with Vasicek factor could be helpful. Vasicek implies constant volatility and does not restrict itself to positive yields only. Correctly calibrated CKLS-Vasicek two-factor model could be able to model

negative yields should it be needed.

We chose to work with yields that are positive. Data leading to short rate equal zero can prove to be problematic as well. Especially when estimating market price of risk calculated by (31). Numerical evaluation of the fraction may easily end in infinite numbers, which should be avoided.

The method for estimating yields was described in previous section on simulated datasets, where we had the luxury of known correct answers. In case of real data there is no such thing as correct answer. However, the method was carefully tested with the simulated datasets, so we chose to trust the method.

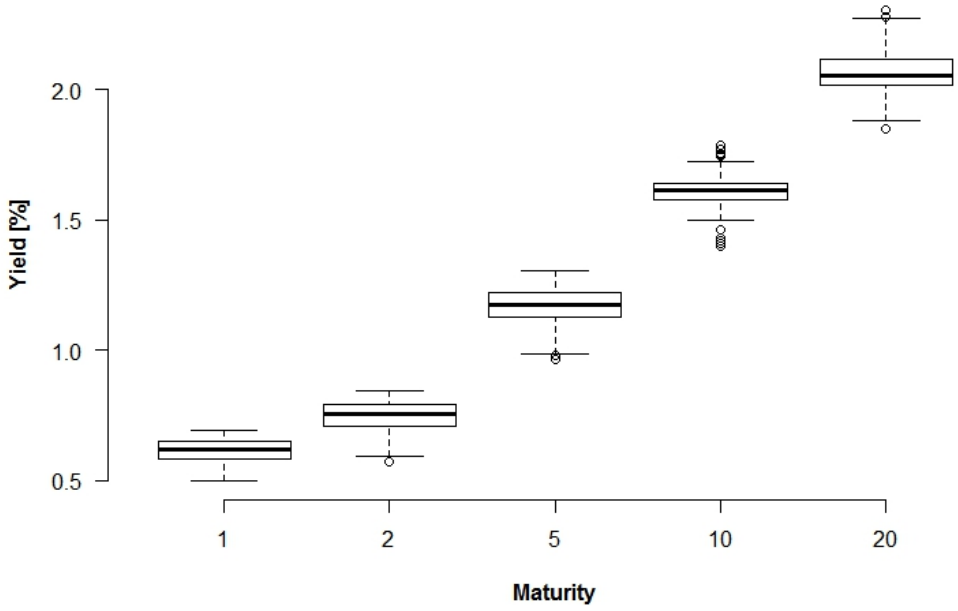
3.1 US treasury bonds

The dataset we chose are US treasury bonds and their zero-coupon yield curve available through QUANDL application or website [11]. The zero-coupon yield curve is calculated based on the real bond market prices. The method of calculation zero coupon curve is described on web page [11]. According to the abstract authors use smoothing method, but the estimates fit the data very well.

The dataset is too vast for our calibration, therefore we chose to use only 1,2,5,10 and 20 year bond yields from 2016-07-01 until 2016-11-01, while leaving out last month for predictions. The yields are quoted in percentage points and need to be converted to decimals. In Figure 24 is boxplot of all data used for calibration of the model. Boxplots are very informative as they capture medians, quantiles and outliers at the same time.

It seems that the yield curve is mostly increasing and that some outliers are present. The outliers are often causing troubles for regression. Optimization will try to fit the outliers, but doing so will worsen the fit of all other observations. In this case the outliers are not too distinct and we leave them in the dataset.

Since the optimum is unknown we advise using various starting points for optimization of each model. Especially for multi-factor models, where the optimum is determined strongly by the choice of initial parameters. This can lead to huge time consumption. Therefore we advise using R package *parallel*, which enables working with multiple calculations at once. The parallel package enables using any number of available logical CPU cores. For short procedures it is ineffective, since each additional core requires



Obr. 24: Boxplot of the dataset. The points represent yields that are too distant from the average - outliers. The lines visualize the range of the yields for each maturity (excluding the outliers). The boxes represent areas where are half of the data points - quantiles. The thick black lines inside the boxes represent median value.

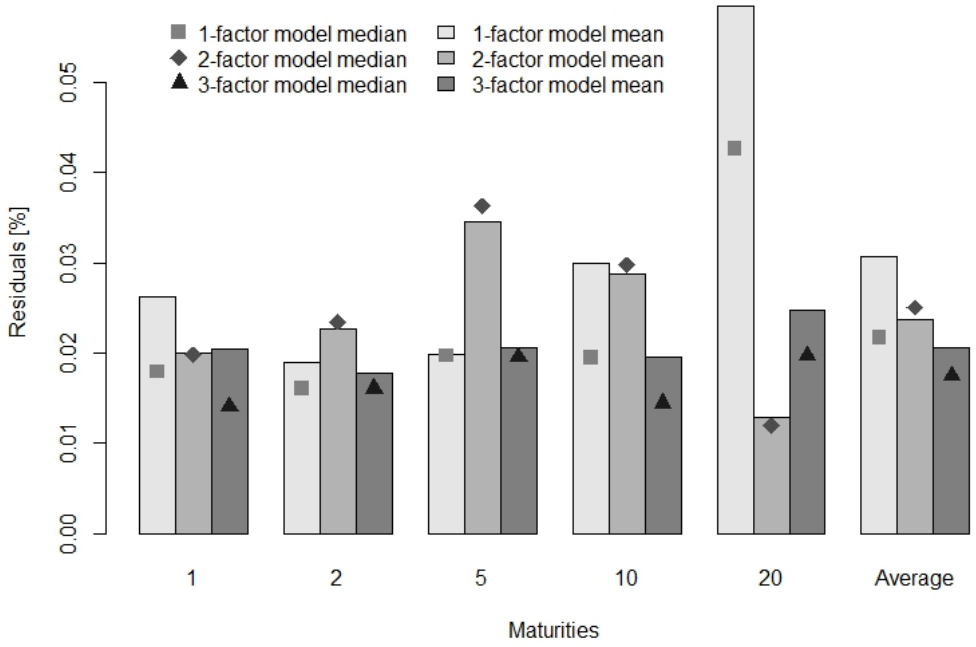
another initialization of R that requires constant amount of time. For costlier methods like multiple independent optimizations it is highly effective. For example calibration of two-factor model from previous chapter took between 40-60 minutes. The parallel computation did it under almost same time, but for seven various starting points. Without parallel computation we would have to optimize one by one and time would add up to few hours.

model		κ	θ	γ	σ	F
1-factor		0.11145	0.03559	0.67696	0.18080	2.1×10^{-7}
2-factor	1st factor	0.00013	0.00265	0.72379	0.01851	8.2×10^{-8}
	2nd factor	0.06304	0.04917	0.63118	0.18594	
3-factor	1st factor	0.00262	0.00117	0.66635	0.05738	1.7×10^{-8}
	2nd factor	0.83267	0.00199	0.86756	0.17024	
	3rd factor	0.13278	0.02752	0.73459	0.15196	

Tabuľka 13: Optimized parameters of the one-factor, two-factor and three-factor CKLS models with independent factors for US yields and the value of objective function F.

For each model we chose seven initial points. Choose initial parameters θ_i reasonably, so that in total for all factors the θ won't exceed θ optimized from one-factor model. It may help prevent factors that are constantly close to zero. Other initial parameters may be random in reasonable intervals. Table 13 shows results of our optimization for US dataset. Judging from one-factor model it appears the data are quite volatile and the short rate slowly converges towards much higher yield. The parameters for multi-factor models are hardly explainable, but can describe the meaning of the factors. For example in three-factor model the second factor has $\kappa^2 = 0.83$ meaning its mean reversion is very fast. Each factor adds new optimization variables and thus making fitting easier. More factors should lead to lower objective value. Objective values in the table indeed confirm this intuition, when the objective value of three-factor model is 1.7×10^{-8} and is the lowest of these three. The average residual in [%] over all data points is 0.0307% for one-factor, 0.0238% for two-factor and 0.0206% for three-factor CKLS models. As a side note, average residual for two-factor Vasicek model, which is a side product, is 0.0331% making it worse than one-factor model. In Figure 25 the

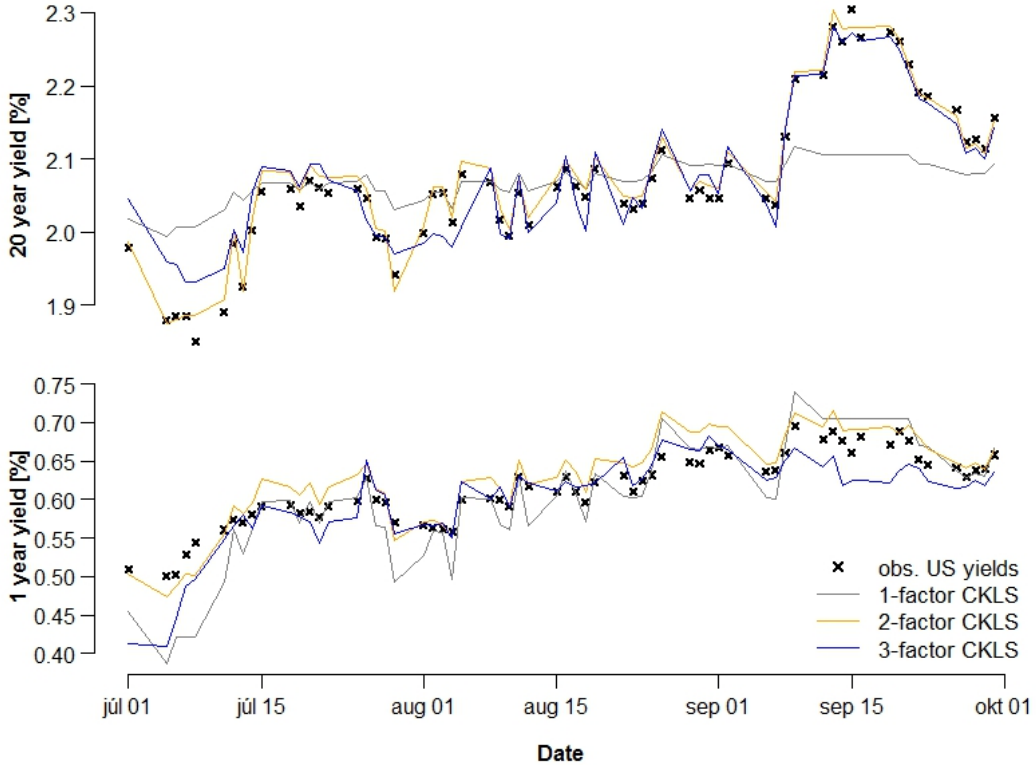
summarize these residuals for each maturity separately.



Obr. 25: Comparison of residuals in [%] between one-factor, two-factor and three-factor model estimates of US yields.

For short maturities all three models fit data similarly on average. The medians are slightly in favour of three-factor model. One-factor model has higher residuals for longer maturities, but fits well the short maturities. Two-factor model is inaccurate for 5 and 10 year maturities. Three-factor model appears to have consistent precision over all maturities. The medians are lower than corresponding means for most cases, which indicates the range of residuals is wide.

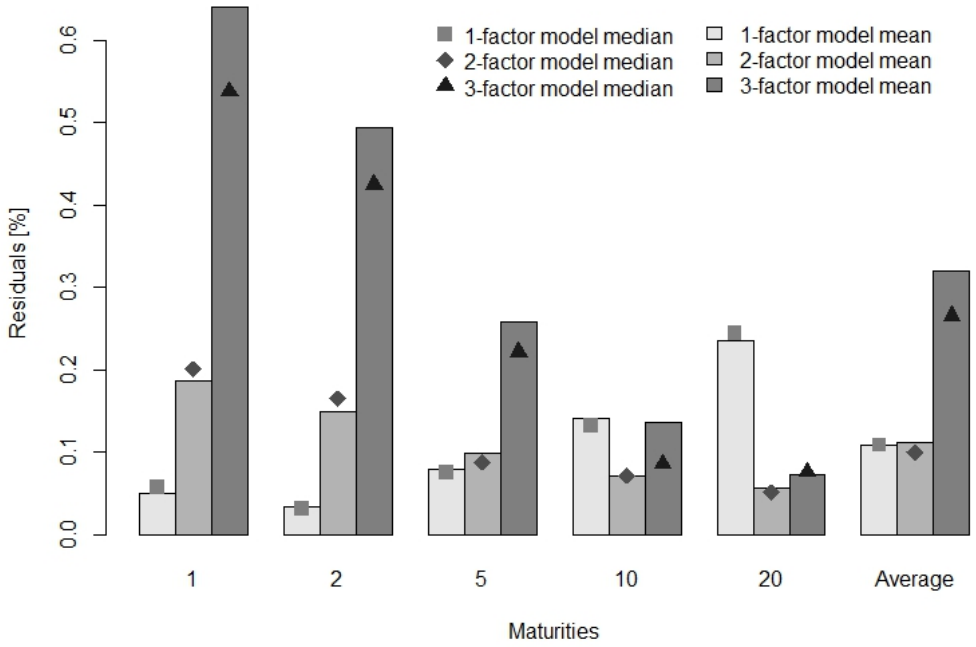
Figure 26 shows development of 1 and 20 year yields and model estimates. This image further confirms that the one-factor model does not fit the long maturities well. The multi-factor models are able to roughly capture development of both 1 and 20 year yields. It appears the two-factor and three-factor estimates are similar for most parts of the time series. But around 1. August the two-factor model fist data visibly better than three-factor model. On the other hand three-factor model performs better overall. The optimization may be altered to specific needs. If there is need to estimate short maturities, specific choice of weights in (27) may achieve much better performance for short maturities. The weights may also depend on time of the observation and by doing



Obr. 26: Development of 1 and 20 year US observed yields with estimates using optimized one-factor, two-factor and three-factor CKLS model.

so one can give the latest observations higher significance than the older observations. Since we don't have such preference, we weighted all data points equally.

Since we optimized parameters and short rates for each model, we can construct predictions. Previously we either predicted whole period at once or only 1 day ahead repeatedly. The 1 day ahead predictions are less influenced by the market price of risk estimation, which proved to be problematic. Figure 27 compares whole period at once predictions for each of the model. Three-factor model predictions are very bad in most cases and strictly worse than two-factor predictions for all maturities. Possible cause may be the incorrect market price of risk. As we already mentioned, short rates close to zero lead to very high market prices of risk estimations. Two optimized future short rates of three-factor model are very close to zero, which may be the reason of the poor predictions. On the other hand 1 day predictions are mostly in favour of multi-factor models. The effect of market price of risk is lower when predicting just 1 day ahead since the error is not accumulating. Figure 28 compares 1 day ahead predictions for each model and maturity. Keep in mind that the 1 day predictions are repeated each day forming predictions for whole period of 21 days.

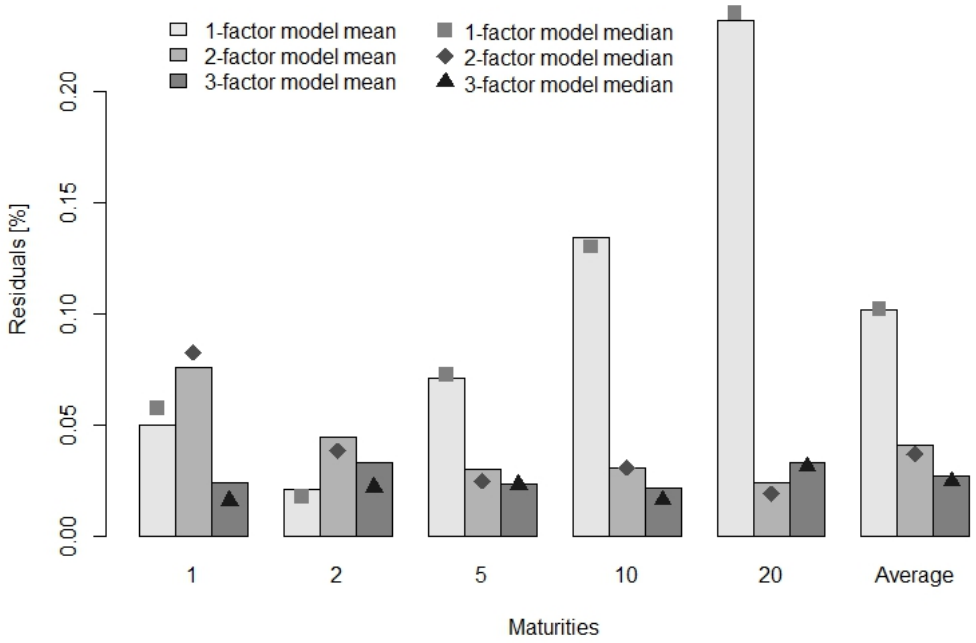


Obr. 27: Predictions of US yields for 21 days ahead for each model and maturity.

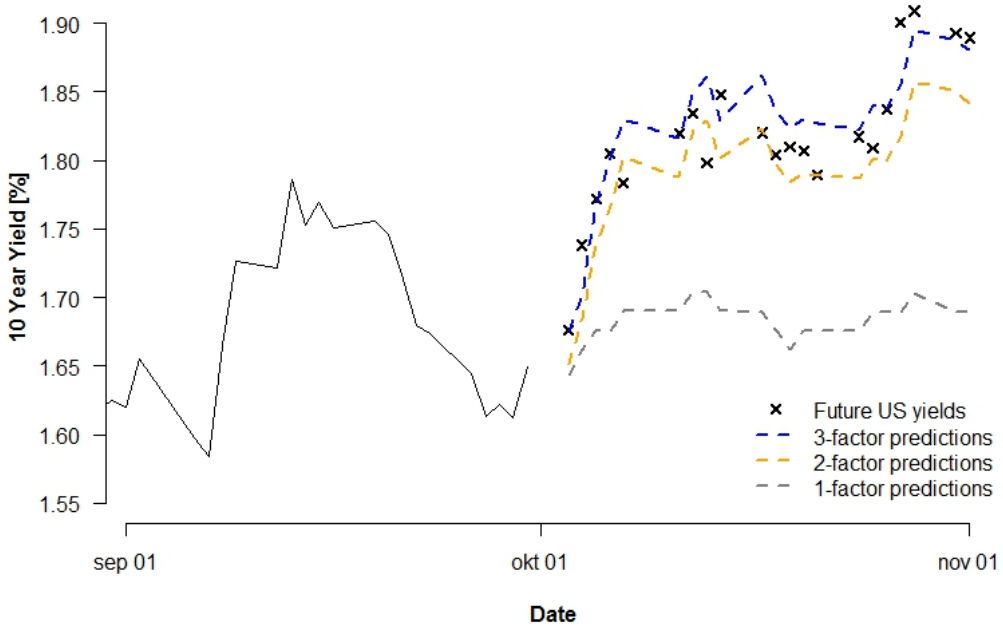
Figure 28 looks similar to Figure 25, which suggests that if model is unable to fit particular data well, it will be unable to predict that particular data too. For example the one-factor model fitted mid term maturities better and the predictions follow this pattern. Also three-factor model produced more or less consistent estimates in terms of residuals. Same can be seen for predictions, where all five average residuals are on roughly the same level.

Figure 29 plots future development of US yields and 1 day ahead predictions of each model for 10 year maturity. For this particular maturity and dataset three-factor model achieves the highest accuracy, closely followed by two-factor model. One-factor model predictions are visibly off.

The future yields follow steep positive trend at first few days, which also may be a cause of poor predictions, especially whole period at once predictions (Figure 27). Predictions for 1 day ahead are able to capture this trend since they always use the data from day before.



Obr. 28: Predictions of US yields repeatedly 1 day ahead for each model and maturity.

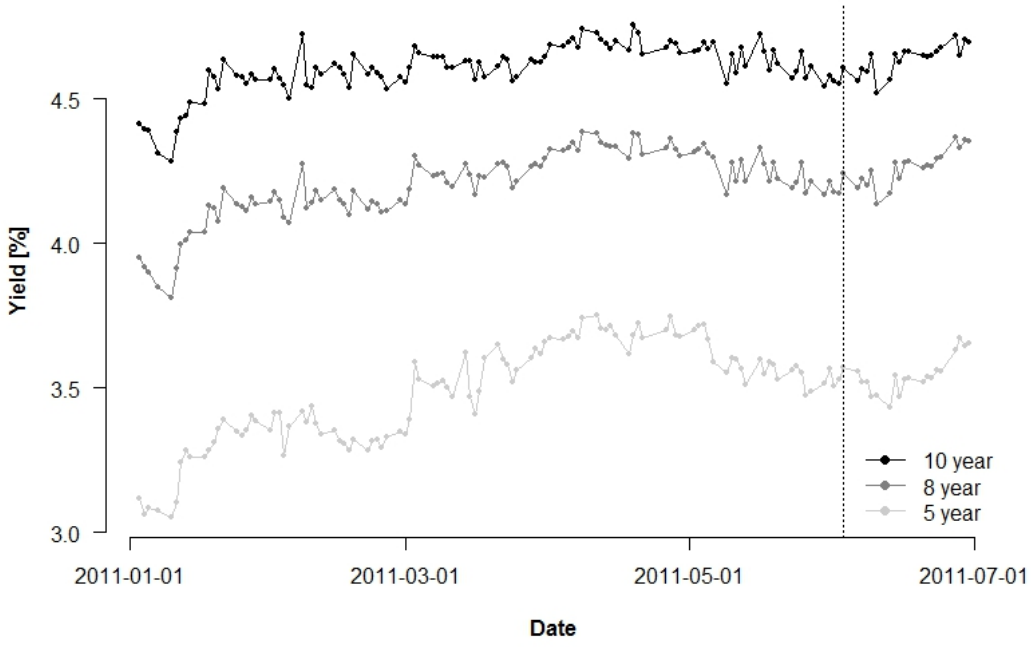


Obr. 29: Time development of future US yields and 1 day ahead predictions of each model for 10 year maturity.

3.2 Slovak government bonds

The US yields are from last year, when the interest rates were globally very low. Therefore we chose another dataset from times with higher yields. As second real dataset we chose Slovak zero-coupon yield curve data available at [12]. The zero coupon yields are converted from market prices using Nelson and Siegel methodology.

The data are from year 2011 and we chose to work with 1, 2, 5, 8, 10 and 15 year yields. Figure 30 plots time development of the mid term maturities. We leave last observations for testing predicting ability of the models.



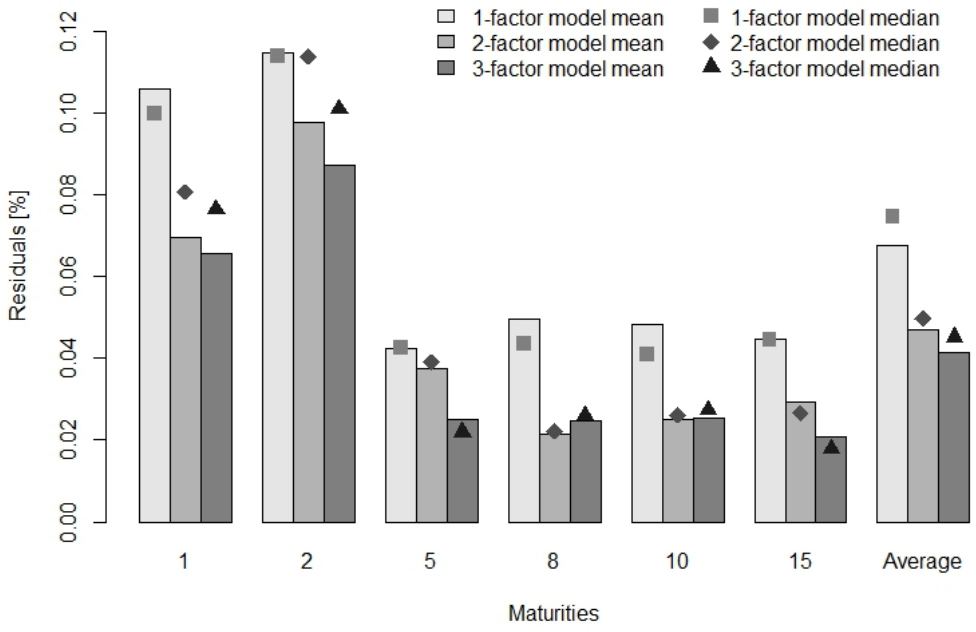
Obr. 30: Time development of 5, 8 and 10 year Slovak zero-coupon government yields. The dotted line divides data for calibration and prediction.

Results in Table 14 show the objective values slightly higher than for US yields. This suggests the SK yields were harder to fit than the previous dataset. Same as before each added factor lowers objective value. The value F contains information about average squared residual over all data points. The average residual in [%] is 0.0414% for three-factor, 0.0468% for two-factor and 0.0676% for one-factor model. Two-factor Vasicek model performed poorly for this dataset. The three-factor model parameters are interesting since they describe three different processes. High parameter κ implies

model		κ	θ	γ	σ	F
1-factor		0.20218	0.07058	2.53224	0.00002	7.4×10^{-7}
2-factor	1st factor	0.34457	0.03247	0.51512	0.23959	3.6×10^{-7}
	2nd factor	0.02665	0.11302	0.74723	0.00169	
3-factor	1st factor	0.50333	0.02455	0.84033	0.19683	8.7×10^{-8}
	2nd factor	0.06098	0.06378	0.89184	0.00335	
	3rd factor	0.14291	0.00780	0.88248	0.07133	

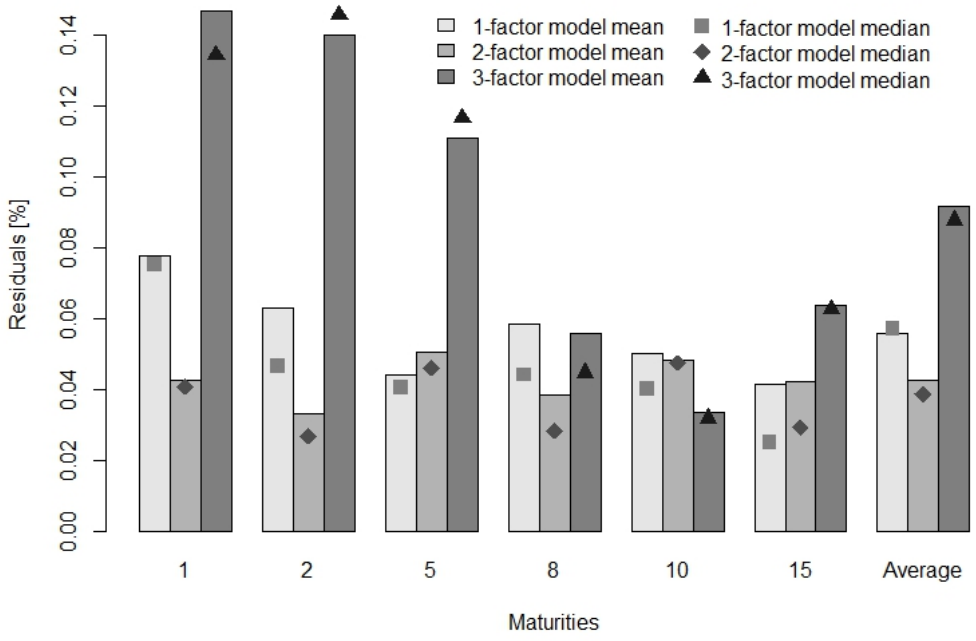
Tabuľka 14: Optimized parameters of the one-factor, two-factor and three-factor CKLS models with independent factors for SK yields and the value of objective function F.

fast mean-reversion of the process. Parameter θ defines the level towards which the short rate converges. First factor converges fast, but the θ is somewhere in middle of other factors. Second factor has small κ and high θ, which means it mostly represents increasing long term trend. And finally third factor with the smallest θ most likely induces decreasing trend. The volatility depends on the optimized short rate, but the first factor appears to be the most volatile.



Obr. 31: Comparison of residuals in [%] between one-factor, two-factor and three-factor model estimates of SK yields.

Judging from Figure 31 the short maturities proved most problematic for all models. The fit of two-factor and three-factor model are very similar and the precision is only slightly in favour of three-factor model. The residuals are higher for SK yields than the residuals for previous dataset.



Obr. 32: Precision of 1 day predictions of SK yields for each model.

The summary of residuals in Figure 32 shows the average precision of 1 day predictions. Predictions of two-factor model are the best for multiple maturities. Three-factor model predictions are good only for 10 year yields, otherwise they are very inaccurate. The reason is not clear. The market price of risk estimation error should be minimal for 1 day predictions, unless the error is too big. Another possible cause is „overfitting“, this effect occurs in regression problems. Adding another variables to a regression model improves the fit, but may impair the predictions. The model with too many variables fits the training dataset so well, it copies the random noise in the training dataset. The model then implies same random noise for future values, which is not correct. Same effect can be happing in our case. Three-factors may introduce too many calibration variables and even though the fit is better, the predictions fail. Either of these causes is possible for the bad precisions of three-factor model. Each dataset is unique and each may suit a different model.

3.3 Remarks

In previous sections we calibrated various models on simulated and also real datasets. We compared these models by the precision of fit and predictions. All results prove that adding another factor to the model generally improves the ability to fit the data, meaning the residuals of the fit are lower. However, predictions for simulated datasets show that adding factor does not improve predicting ability of the model. Both short and long period predictions are almost equal for one-factor, two-factor and three-factor models. In case of real datasets, none of the models are able to fit data perfectly. The residuals are much higher than for simulated datasets. It appears multi-factor models achieve consistently better precision over all data points, while one-factor model focuses on particular data. For example for US yields dataset (Figure 25), the one-factor model focused mostly on short maturities, while three-factor model's residuals remain at roughly the same level for all maturities. This focusing on particular data seems to transfer to predictions too (Figure 28).

The calibration of the models is not flawless. In previous sections, we often avoid topics, which might influence result and we neglect them in our approach. In this section, we describe these issues and propose possible improvements, which may be helpful for further research.

The **estimation of market price of risk** is very simplified in our algorithm. The form of $\lambda(r,t) = \lambda$ leads to $\lambda(r,t)\sigma(r,t) = \lambda\sigma r^{\gamma}$. Models like CIR assume $\lambda(r,t) = \lambda\sqrt{r}$, thus CKLS model should assume even more advanced form of market price of risk. As we mentioned in chapter 2.4, we were unable to estimate market price of risk as $\lambda = \lambda r^{\delta}$. But even then, it may not be a good estimate of true market price of risk.

Market price of risk causes another issue, which strongly determines the long period predictions. The estimation is biased and the range of the bias is sensitive to parameters γ, σ and short rates r. Short rates close to zero and $\gamma > 1$ lead to cases when the denominator in (31) goes faster towards zero than the numerator. Thus even small error in the short rate has great impact on the market price of risk. We have already discussed the precision of short rate optimization and precision is very hard to achieve. In multi-factor models parts of each factors are often close to zero. Which means that even though the fit of the multi-factor model is better, the bias in market price of risk

estimation impairs the predictions for long periods. That is the reason why we provided also 1 day predictions, where the effect of market price of risk is smaller.

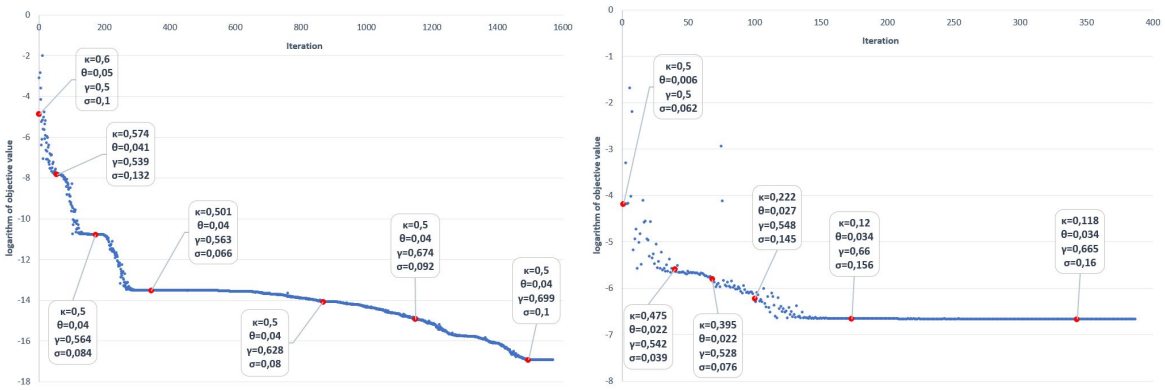
Even though we satisfy Fichera theory the optimization leads to **small non-positive short rates**. Fichera theory ensures that under specific conditions the probability of zero short rate is zero. Negative short rate is restricted completely by the CKLS model definition. Despite that, our algorithm gives small negative short rates. The reason is the way we optimize the short rates. If N is number of factors in model, we optimize only $N-1$ short rates and calculate the remaining factor with total short rate, which is sum of all factors. Under perfect conditions this would be fine, but the total short rate is also estimated from external method (two-factor Vasicek model or one-factor model estimation) with given precision. Thus if the total short rate is not perfect, the factors may be optimized to negative values. Perfect total short rate is impossible to get, since it is unobservable for real datasets. To prevent failures of solving PDE for negative short rates, we basically round the negative short rates to zero and use it to calculate bond price. Numerically this is done instantly and does not cause any problems. But if the short rate is zero, it has to be omitted when calculating market price of risk. Another reason for small negative short rates is that the all is done numerically and some form of error will always be present. If the negative yields are the objective, one could always substitute one CKLS model factor as Vasicek model factor, which enables estimation of negative yields.

Our results of the optimization algorithm may be fine, but there is no guarantee that for different datasets it won't fail. Our algorithm is basically only a **heuristic**. The algorithm may not provide consistently good results and further testing needs to be done. The choice of numerical parameters depends on the dataset and one should always experiment with different numerical parameters like spatial spacing, Nelder-Mead coefficients,...

The choice of starting point in our algorithm is described very vaguely. That is because there is no straight forward guide for picking good starting point. In most optimization problems, the choice of starting point does not matter. However for our algorithm it does, especially for multi factor models. The only guide line we observed is that the initial choices of θ_i parameters for each factors should follow $\sum_i \theta_i = \theta_{1f}$, where

θ_{1f} is parameter of one-factor model. This might help to achieve better performance and prevent having one of the factors basically constant zero. On the other hand, when the data are following a one-factor model perfectly, the multi-factor model optimization fails to converge to one-factor model without using starting point with only one non-zero factor. Possible workaround is using various starting points and resetting the Nelder-Mead optimization few times, but this is hardly a reliable solution of the issue.

Different optimization method might improve the convergence or at least save some time. Nelder-Mead optimization often quickly finds solution and then spends large amounts of time moving around in small steps and shrinking its volume. This can be partly controlled by the maximum allowed iterations, but still the Nelder-Mead algorithm may not be the most suitable for this problem. Interesting union of genetic algorithm and effective convex algorithms is proposed in [9], where they test this algorithm on one-factor models. However, it is unclear how to implement short rate optimization into such genetic algorithms without significantly increasing time needed. Figures 33a and 33b plot iterations of Nelder-Mead optimization for simulated and real datasets and logarithm of objective values. In real dataset the algorithm spends most of the iterations shrinking in optimum. Limiting maximal iterations is possible controlling mechanism to prevent spending time uselessly. On the other hand it may prevent the algorithm to reach better optimum as is seen in Figure 33a. The optimization needs all iterations to reach correct optimum for simulated dataset.

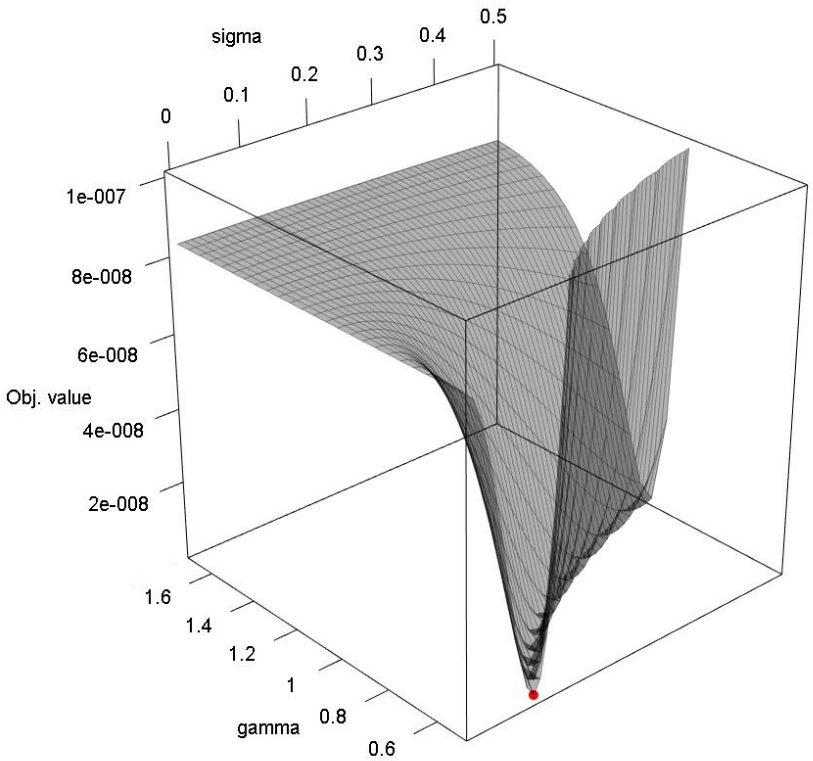


(a) Simulated dataset. **(b)** Real dataset.

Obr. 33: Example developments of objective value during Nelder-Mead optimization of one-factor model.

There are observable steps or stages in the process of optimization in Figure 33a

for simulated dataset. These steps are created when Nelder-Mead algorithm explores new minima. For example the first step comes with finding the correct parameter θ and second step comes with correct κ. Afterwards the Nelder-Mead exploits the found optimum and shrinks to achieve better precision. After each shrink the reach of Nelder-Mead simplex manipulations decreases and thus true optimum may be invisible for the algorithm unless it shrinks enough. When the simplex shrinks enough the third step occurs and is already close to solution. However, to find correct γ and σ, much more shrinking is required. The sensitivity of objective value to γ and σ is complex. For some regions the sensitivity is high, for other low. Usually the optimization ends with two outcomes - $\gamma > 1$ and σ very low or $\gamma < 1$ and σ high. The other combinations are quickly declined, which means the sensitivity is high in those regions.



Obr. 34: Visualization of objective value in relation to γ and σ, while κ and θ are correct. Red dot denotes correct γ and σ.

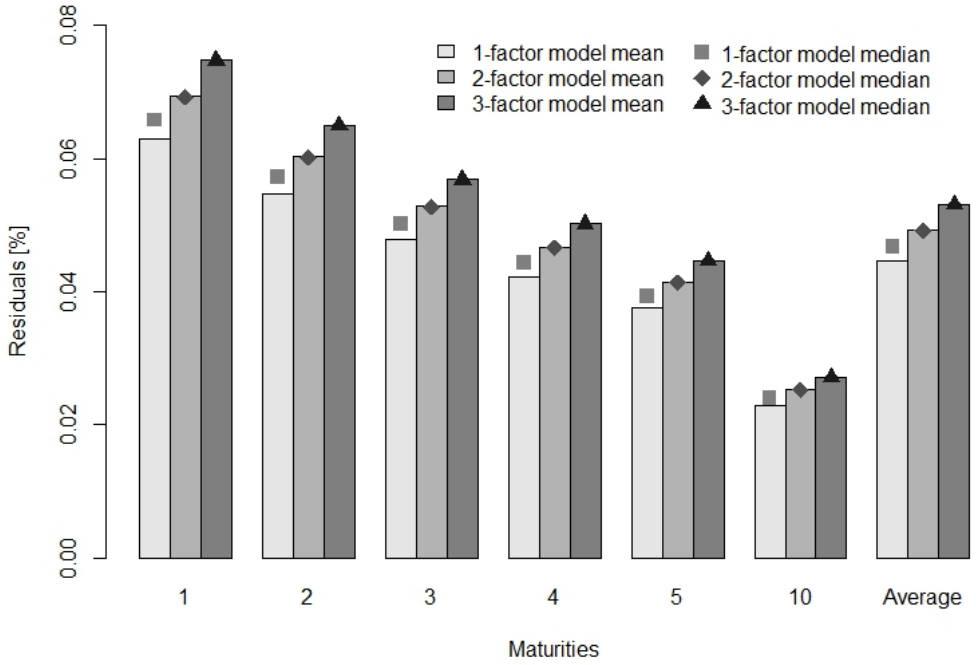
The volatility parameters γ and σ are complicated to optimize. Due to the form σr^{γ}, where r is usually very small, similar volatility can be described by different parameters. Depending on starting point the optimization may end with $\gamma < 1$ and

big σ or $\gamma > 1$ and small σ. Both lead to very similar volatilities. Figure 34 shows the objective value in relation to volatility parameters. Different combinations of γ and σ may result to similar volatility, which can be seen in the graph as „valley“ of low objective values. All points in the valley have similarly low objective values, but the lowest is in correct γ and σ. However, it is easy for optimization algorithm to end in a different point in the valley and thus giving different optimal γ and σ.

The mulit-factor problem is **not convex** and it appears to not be uniquely determined either. In case of two-factor Vasicek model authors in [2] propose relation between drift parameters and short rates. Thus the solution of PDE may not be uniquely determined without implementing similar dependency relation into CKLS multi-factor model. Using convex programming methods like BFGS for parameters optimization leads to bad results, which suggests that the problem has many local minima. These properties make it very difficult to find correct optimum if not impossible.

Phenomenon **overfitting** occurs, when a model uses too many variables or regressors. Model with too many variables may fit the data too well and copy the unwanted random noise contained in the training dataset. Afterwards the copied noise impairs predictions for a new dataset, which contains different random noise. We already mentioned overfitting when we described predictions for Slovak yields dataset in Figure 32. In that case we were unable to determine whether the incorrect predictions of three-factor model are due to overfitting or just bad estimation. To successfully observe overfitting we devise experiment with one-factor CKLS model simulated dataset. We calibrate one-factor, two-factor and three-factor CKLS model for this dataset. Usually the fit is better for more factors in the model, but in this case it is same or worse depending on the starting point for multi-factor models. If we choose starting point with knowing that this data are simulated from one-factor model, we end up with all factors zero except one. However, if we choose starting point otherwise, the optimization finds different optimum and the fit is slightly worse, but still very good. So either we end up with multi-factor models reducing to one-factor model or multi-factor models trying to fit the data otherwise. Only the latter produces different predictions of each model.

Figure 35 summarizes 50 day predictions of each calibrated model. The one-factor model predictions are better than multi-factor model predictions, which may indicate



Obr. 35: Comparison of 50 day predictions of each model for dataset simulated by one-factor model. Overfitting is visible as multi-factor models perform worse.

the presence of overfitting. Although, the effect is not strong and the residuals are very close in this case. The effect may be weak, because the data do not contain much noise to which multi-factor models could overfit.

A better experiment could be done, where the data are simulated in same way, but then exposed to a random noise. Such noise could then transfer to multi-factor models, while one-factor model should remain almost unchanged. However, it is not clear what kind of noise should be used to avoid creating completely different dataset.

Conclusion

Modelling interest rates is widely popular and various approaches emerged throughout the history. Our goal was to compare them and find the best one. The important aspects to compare are the ability to fit data and ability to predict future data. We focused on short rate models, more precisely CKLS short rate models with one, two or three independent factors.

First we described the mathematical background of the relation between modelled short rate and interest rates. The short rate is modelled as stochastic process with parameters κ, θ, γ and σ. Interest rates are calculated as a solution to partial differential equation, which rises up from no-arbitrage principle of derivative pricing. The partial differential equation is solved numerically using method of lines proposed in [8].

In order to calculate interest rates, the model needs parameters κ, θ, γ and σ, which are unknown for real datasets. We use regression to find the correct parameters. The optimization is done using Nelder-Mead algorithm and the result are optimized parameters. However to estimate interest rates, the model needs short rate and maturity of the yield. Maturity is known, but short rate is unobservable. Therefore we designed algorithm which optimizes CKLS model parameters and short rates together. With small changes the algorithm is applicable also for multi-factor models.

Afterwards we calibrated one-factor, two-factor and three-factor CKLS model for simulated and real datasets. We measured the ability to fit the data by average differences between estimated and observed yields. To measure predicting ability we compared predictions with exact future observations, which were left out of the calibration process. Generally the models with more factors are fitting the data better. One-factor model often focuses only on particular maturity or section of dataset, while multi-factor models are able to capture all data points consistently (Fig. 25). Often this focusing effect transfers to predictions as it was the case for US zero-coupon yields dataset (Fig. 28). On the other hand overfitting effect may manifest and principle „less is more“ takes place. This was probably the case for Slovak zero-coupon yields dataset (Fig. 32). In last chapter we discussed many issues that accompanied the calibration process and possible improvements.

List of literature

[1] Bučková Z., Ehrhardt M., Günther M.: *Fichera Theory and its Application in Finance*, Proceedings ECMI, 2014.

[2] Bučkova Z., Halgašová J., Stehlíková B.: *Short rate as a sum of two CKLS-type processes*, Springer Verlag in LNCS, 2016.

[3] Chan K. C., Karolyi G. A., Longstaff F. A., Sanders A. B.: *An Empirical Comparison of Alternative Models of the Short-Term Interest Rate*, The Journal of Finance 47(3), pp. 1209-1227, 1992.

[4] Ekström E., Tysk J.: *Boundary conditions for the single-factor term structure equation*,Ann. Appl. Probab. 21 (2011), no. 1, 332–350. doi:10.1214/10-AAP698. `http://projecteuclid.org/euclid.aoap/1292598037`.

[5] Higham N.J.: *Functions of Matrices: Theory and Computation*, Society for Industrial and Applied Mathematics, 2008

[6] LeVeque R. J.: *Finite Difference Methods for Differential Equations*, University of Washington, pp. 3-11, 2004. doi:10.1137/1.9780898717839.ch1.

[7] Ševčovič D., Stehlíková B., Mikula K.: *Analytické a numerické metódy oceňovania finančných derivátov*, Nakladateľstvo STU, 2009. ISBN 978-80-227-3014-3.

[8] Tangman D.Y., Thakoor N., Dookhitram K., Bhuruth M.:*Fast approximations of bond option prices under CKLS models*, Finance Research Letters 8, 206–212, 2011.

[9] Urbánova A.: *Calibration of term structure models*, Dissertation Thesis, Comenius University, 2007. `http://www.iam.fmph.uniba.sk/studium/efm/phd/urbanova/urbanova-thesis.pdf`

[10] Zíková Z., Stehlíková B.: *Convergence model of interest rates of CKLS type*, Kybernetika 48 (3), 567-586, 2012. `http://www.iam.fmph.uniba.sk/institute/stehlikova/publications.html`

[11] US treasury zero-coupon yields dataset, `https://www.quandl.com/data/FED/SVENY`.

[12] Slovak government zero-coupon yields dataset, `http://www.finance.gov.sk/en/Default.aspx?CatID=713`.