Yield Curve Modelling: A Comparison of Principal Components Analysis and the Discrete-Time Vasicek Model

Irene Sekyere Asare

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Mathematics) at

Concordia University

Montreal, Quebec, Canada

August 2019

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Irene Sekyere Asare

Entitled: Yield Curve Modelling: A Comparison of Principal Components Analysis and the Discrete-Time Vasicek Model

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Examiner

Dr. Patrice Gaillardetz

_____ Examiner

Dr. Cody Hyndman

_____ Thesis Supervisor

Prof. Frédéric Godin

Approved by _____

Chair of Department or Graduate Program Director

_____

Dean of Faculty

_____

Date

Abstract

The term structure of interest rates is relevant to economists as it reflects the information available to the market about the time value of money in the future. Affine term structure models such as short rate models have been used in interest rate modelling over the past years to determine the mechanisms driving the term structure. Machine learning approaches are explored in this thesis and compared to the traditional econometric approach, specifically the Vasicek model. Multifactor Vasicek models are considered as the one factor model is found not adequate to characterize the term structure of interest rates. Since the short rates are not observable the Kalman filter approach is used in estimating the parameters of the Vasicek model. This thesis utilizes the Canadian zero-coupon bond price data in the implementation of both methods and it is observed from both methods that increasing the number of factors to three increases the ability to capture the curvature of the yield curve. The first factor is identified to be responsible for the level of the yield curve, the second factor the slope and third factor the curvature of the yield curve. This is consistent with results obtained from previous work on term structure models. The results from this work indicates that the machine learning technique, specifically the first three principal components of the Principal Component Analysis (PCA), outperforms the Vasicek model in fitting the yield curve.

Acknowledgments

I would like to thank my supervisor Prof. Frédéric Godin for his guidance, excellent advice and continuous support for my thesis. I could not have imagined having a better supervisor for my research. I would like to also thank the rest of my thesis committee: Dr. Patrice Gaillardetz and Dr. Cody Hyndman for their comments and contributions.

A special thanks goes to my family and dearest friends who supported me through good and bad times and made it possible for me to be the person I am today. A special thank you also goes to my deceased mother, Paulina Amobea Ofosu, who was my inspiration to pursue my master's degree and taught me hard work, but did not live to see me graduate.

Finally, I thank my dear husband who has been my source of encouragement. I am very thankful for having you in my life.

Contents

List of Figures

List of Tables

Introduction

A zero-coupon bond is a financial debt instrument which promises to repay its face value at a predetermined future time of maturity. The future value, and thus price of a bond at maturity is determined among other things by the daily interest rate, which is not observable. Interest rates are used in practice to forecast the yields of financial instruments and to determine their price. They are very important to investors as they influence the total return of many investments. Interest rates are not directly traded on the financial markets. They are obtained from the market prices of the related tradeable products such as bonds.

Interest rates are generally estimated using mathematical and statistical models. These models determine the interest rates, and consequently yields for the zero-coupon bonds, using the daily time series of spot rates i.e. the yield on each individual zero-coupon bond on a particular day. Traditionally, the modelling of interest rates has been done using affine term structure models such as short-rate models but in recent times, machine learning approaches have become popular.

One of the aims of interest rate modelling for zero-coupon bonds is to determine the mechanisms driving the term structure or yield curve so that we can make predictions of future interest rates. The yield curve is the functional relationship between the spot rates and the time to maturity. The traditional econometric approach to fitting the yield curve is to estimate the parameters of a mathematical model relating the spot rates and the time to maturity. Modern machine learning approaches on the other hand, fit the yield curve model without necessarily elucidating an underlying parametric structure.

One of the challenges of interest rate modelling and prediction is how to capture the considerable daily variation in the interest rates. Models using single factors or components often fail to adequately fit the yield curve. One approach to improving the fit of the yield curve model is to use multiple factors. Using multiple factors increases the flexibility of the model and thus its fit to the past interest rate trends as well as its ability to predict future movements.

In this thesis the discrete time Vasicek model was chosen as the short rate model to be used in fitting the yield curve. The machine learning method we explore in this paper is principal components analysis (PCA), which is an unsupervised learning technique that extracts a number of unobservable factors which drive the interest rates. This thesis complements the work performed by Bolder et al. [2004] by applying the PCA method to recent data for the Canadian yield curve and comparing its ability to characterize the yield curve to that of the discrete time Vasicek model.

The first generation econometrics models for fitting the yield curve such as the Vasicek model are known to poorly fit interest rate data. In particular, the single factor Vasicek model struggles to capture the curvature of the yield curve adequately. Increasing the number of factors should theoretically improve the fit of the model. In this thesis, we examine the potential improvement of the fit of the Vasicek model with increasing number of factors. We use the Vasicek model because it was the first model introduced to describe the short rate $r(t)$ and also because of its simplicity. The Vasicek model as described by Vasicek [1977] is a one factor short rate model which describes the evolution of interest rates. Following the approach used in the one factor Vasicek model we implement a method for the multifactor Vasicek model. Multifactor models for short rate models have been discussed by Brigo and Mercurio [2007] and the discrete time version of these models were considered by Wüthrich and Merz [2013]. We will concentrate on the first three factors of the model developed in performing our analysis.

The objective of this thesis is to compare the multifactor Vasicek and PCA approaches to

deriving the yield curve. We will perform a descriptive analysis of the Canadian zero-coupon bond data for the period 2004 to 2018 thus extending similar work done for 1986 to 2003 by Bolder et al. [2004]. We will then fit the single factor, two-factor and three-factor Vasicek models to the data using a Kalman filter approach. We will also fit a principal components analytic model to the data and compare the PCA and multifactor Vasicek models.

This thesis consists of three main chapters, all related to its central theme. Chapter 1 reviews the literature on interest rate modeling using econometric and machine learning techniques. We discuss various affine term structure models and develop the formula for the discrete time Vasicek model. Various other methods of estimation are discussed including methods based on Bayesian statistics such as the Kalman filter which we use in this paper. We also discuss machine learning methods such as PCA in detail. In Chapter 2, we present a descriptive analysis of the data to be used in our yield curve modeling. In Chapter 3, we model the yield curve using first the Kalman filter approach and then the PCA method, and finally compare the two.

Chapter 1

Interest Rate Models

In this chapter we discuss the modelling approaches used in this thesis—particularly affine term structure modelling and machine learning yield curve modelling—as well as the data to be used. It is organized in 3 sections.

The first section provides an overview of the affine term structure models of interest rates and the various ways of modelling interest rates including short rate models, for both discrete-time and continuous-time models although only discrete-time model is used later on in the thesis. We derive equations for the discrete time Vasicek model, our chosen short rate model, and extend this model from the one-factor case to the multi-factor situation in this section. We also discuss various estimation procedures for the Vasicek model with emphasis on the method to be used in the thesis. In the second section, we discuss various machine learning techniques for fitting the yield curve and provide a detailed explanation of the PCA method. We finally describe the data source we are to use in this thesis in the third section.

1.1 Affine Term Structure modeling

A zero-coupon bond with maturity T (also called T-bond), as defined by Björk [2009], is a contract that guarantees its holder a payment of one unit of currency at time T. The price at time t of a zero-coupon bond is denoted as $P(t,T)$ and the price at maturity $P(T,T) = 1$

for all $t \leq T$.

The following assumptions are made concerning zero-coupon bonds:

1) That there exists a market for the zero coupon bond for every $T > 0$.

2) That for each fixed t, the bond price is differentiable with respect to time to maturity.

We define forward rates and spot rates which are related to zero-coupon bonds for continuous time models.

A spot rate is the constant rate at which we would recover the price at time t of the zero-coupon bond maturing at a future date T. This is also referred to as the yield on a zero-coupon bond. Brigo and Mercurio [2007] define the continuously compounded spot interest rate as follows.

Definition 1.1.1. Continuously-compounded spot interest rate *The continuously-compounded spot interest rate prevailing at time t for the maturity T is denoted by $s(t,T)$ and is the constant rate at which an investment of $P(t,T)$ units of currency at time t accrues continuously to yield a par value at maturity T.*

$$s(t,T) = \frac{-\log\left(P(t,T)\right)}{T-t}. \tag{1.1}$$

The forward rate is the rate allowing discounting between two future dates, time T_1 and T_2. It is the constant interest rate over (T_1,T_2) which would allow discounting to time T_1, a cashflow occuring at time T_2. The forward rate $F(t,T_1,T_2)$ is given as,

$$F(t,T_1,T_2) = -\log\left(\frac{P(t,T_2)}{P(t,T_1)}\right)/T_2 - T_1. \tag{1.2}$$

The instantaneous forward rate $f(t,T)$ in continuous time is given as

$$f(t,T) = \frac{-\partial \log\left(P(t,T)\right)}{\partial T}. \tag{1.3}$$

with

$$P(t,T) = e^{-\int_t^T f(t,T)du}. \tag{1.4}$$

The instantaneous short rate at time t is defined as

$$r(t) = f(t,t). \tag{1.5}$$

The yield to maturity (YTM) is the total return anticipated on a bond if the bond is held until it matures. The yield to maturity is relevant to investors as it helps them assess the return they could make on a bond.

The term structure of interest rates refers to the relationships between bond prices of different maturity. At time t, the term structure, also referred to as the spot curve or yield curve, is the function $s(t,\cdot)$ which associates to each maturity, T, a spot rate $s(t,T)$. The term of a debt instrument with a fixed maturity date is the time until the maturity date. A term structure study gives indications on the evolution of the yield curve. The shape of the yield curve can sometimes be downward sloping, upward sloping, humped or flattened. Investors usually demand higher interest rates for long-term investments as compensation for the risk taken in investing their money in long-term bonds. Due to this the shape of the yield curve is usually upward sloping. However, the long-term bond yields may sometimes fall below the short-term bond yields resulting in an inverted shape, or the intermediate terms may have higher interest rates causing the curve to have a humped shape. Inverted shapes are usually observed when short term expectations for interest rates are higher than for the long term.

Term structure models can be put into three categories, short rate models, forward rate models and market dynamics (LIBOR Market Model (LMM)).

Short Rate Models

Short rate models have been used in pricing of interest rate derivatives over the years in various ways. The Vasicek model, Vasicek [1977], is one of the oldest methods used. One feature of the Vasicek model is that the short rate follows a Gaussian distribution, hence, there is the possibility of interest rates being negative.

The Vasicek model follows an Ornstein-Uhlenbeck process with constant coefficients under the risk-neutral measure as follows:

$$dr(t) = \kappa(\theta - r(t))dt + \sigma dW(t), \quad r(0) = r_0 \tag{1.6}$$

where r_0, κ, θ and σ are positive constants. κ represents the mean reversion speed, θ, the long term mean of the process $r(t)$ and σ the volatility of the interest rate. Intergrating (1.6) gives,

$$r(t) = r(s)e^{-k(t-s)} + \theta(1 - e^{-k(t-s)}) + \sigma \int_s^t e^{-k(t-u)dW(u)}, \tag{1.7}$$

with $s < t$. The major drawback of the Vasicek model is the possiblity of having negative interest rates.

The Dothan Model, Brigo and Mercurio [2007], is a continuous time version of the Rendleman and Bartter model. Rendleman and Bartter [1980] is an improvement of the Vasicek model to make the short rate always positive by assuming a lognormal distribution. The short rate under the risk-neutral dynamics for the Dothan model is given as

$$dr(t) = ar(t)dt + \sigma r(t)dW(t), \quad r(0) = r_0 \tag{1.8}$$

where a is a real constant. Integrating the dynamics in the equation above gives

$$r(t) = r(s)\exp\left((a - \frac{1}{2}\sigma^2)(t-s) + \sigma(W(t) - W(s))\right) \quad \text{for} \quad s < t. \tag{1.9}$$

One main drawback of the Dothan model is the potential explosion of the bank account value.

Cox, Ingersoll and Ross (CIR) model, Brigo and Mercurio [2007], introduced a square-root term in the diffusion coefficient of the instantaneous short rate dynamics proposed by Vasicek. This makes the short rate always positive and in this model the short rate follows a non-central chi-squared distribution. The model under the risk-neutral dynamics is

$$dr(t) = \kappa(\theta - r(t))dt + \sigma\sqrt{r(t)}dW(t), \quad r(0) = r_0 \tag{1.10}$$

where r_0, κ, θ and σ are positive constants. The Feller condition $2\kappa\theta > \sigma^2$ must always be imposed for the process $r(t)$ to be always strictly positive for the CIR model.

The above models are endogenous term structure models. That is the initial term structure does not necessarily match the observed yield curve no matter the model parameter chosen. This problem was solved by introducing the exogenous models, which allow the inclusion of a time-varying parameter and lead to perfectly reproducing the term structure at a given point in time. Some exogenous models are the Hull and White (1990), extended Vasicek model, possible extensions of the Cox, Ingersoll and Ross (1985) model, the Black and Karasinski (1991) model.

Forward Rate Models (Heath-Jarrow Morton Models)

In the short rate models described above we observe that $r(t)$ being a solution of the stochastic differential equation allows us to work within a PDE framework and also explicitly obtain formulas for bond prices and derivatives. The main drawback of these models is that an exact calibration of the observed term structure is difficult to achieve. Ho and Lee (1986) provided a solution to this problem by modelling the evolution of the yield curve in a binomial tree setting. Heath, Jarrow and Morton developed a general framework for modelling interest rates by extending the work of Ho and Lee to continuous time. Any (exogenous

term-structure) interest-rate model can be derived within the HJM [1] framework.

In the risk neutral HJM framework the instantaneous forward rate process $f(t,T)$ representing the time T instantaneous rate of return as seen at time t is assumed to satisfy the stochastic equations

$$\begin{aligned} f(t,T) &= f(0,T) + \int_t^T \sigma(u,T) \int_u^T \sigma^{\intercal}(u,s) ds du + \int_0^T \sigma(s,T) dW(s) \\ &= f(0,T) + \sum_{i=1}^{N} \int_0^t \sigma_i(u,T) \int_u^T \sigma^{\intercal}{}_i(u,s) ds du + \sum_{i=1}^{N} \int_0^t \sigma_i(s,T) dW_i(s). \end{aligned} \tag{1.11}$$

where, T is the time to maturity, $W = (W_1, ..., W_N)$ is an N-dimensional Brownian motion, $\sigma(t,T) = (\sigma_1(t,T), ..., \sigma_N(t,T))$ is a vector of adapted processes. The spot rate process $r(t)$ representing the instantaneous rate of return at time is defined by the equation
$r(t) = f(t,t)$, that is

$$r(t) = f(0,t) + \sum_{i=1}^{N} \int_0^t \sigma_i(u,t) \int_u^t \sigma^{\intercal}{}_i(u,s) ds du + \sum_{i=1}^{N} \int_0^t \sigma_i(s,t) dW_i(s). \tag{1.12}$$

The time t price of a zero coupon bond maturing at time T is defined as in (1.4).

Market Dynamics (LIBOR Market Model (LMM))

The Libor Market Model [2] is the market standard model used for pricing interest rate derivatives. Each forward rate is modelled by a log-normal process under its forward measure. From (1.2) the LIBOR forward rate according to Brigo and Mercurio [2007] is defined as

$$L(t,T_1,T_2) = \frac{P(t,T_2) - P(t,T_1)}{(T_2 - T_1)P(t,T_2)}. \tag{1.13}$$

and LIBOR spot rate defined as

$$L(T_1,T_2) = \frac{P(T_1,T_2) - 1}{(T_2 - T_1)P(T_1,T_2)}. \tag{1.14}$$

[1] For details of the model kindly refer to Björk [2009] and Brigo and Mercurio [2007]
[2] For details of the model kindly refer to Björk [2009] and Brigo and Mercurio [2007]

Some recent works performed with the LIBOR market model are Miltersen et al. [1997] and Brace et al. [1997]. Most implementations of this model use Monte Carlo simulations to price European-style and Bermudian-style swaptions.

Affine Term Structure Models

An affine term structure is a model for the short rate where spot rates have the form:

$$s(t,T) = \alpha(t,T) + \beta(t,T)r(t). \tag{1.15}$$

For the affine term structure model above the zero-coupon price is given as

$$P(t,T) = A(t,T)e^{B(t,T)r(t)}. \tag{1.16}$$

$$A(t,T) = e^{-(T-t)\alpha(t,T)}. \tag{1.17}$$

$$B(t,T) = (T-t)\beta(t,T). \tag{1.18}$$

In the aforementioned short rate models, the short term interest rate is assumed to follow a one factor continuous time process. The term structure and bond prices of most of these models can be obtained explicitly. However, the one-factor model is not able to fully characterize the term structure of interest rates and its changing shape over time. Some work has been performed on increasing the number of parameters and risk factors to assess the term structure of interest rates. Chen and Scott [1993] developed a method for estimating the parameters of a continuous time model of the term structure. In their work they used the CIR, developed a multifactor equilibrium model of the term structure and determined the number of factors necessary to characterize the term structure adequately over time. Brigo and Mercurio [2007] also discussed two-factor and three-factor versions of some interest rate models. A multi factor version of the Vasicek model was discussed and presented by Babbs and Nowman [1998]. They applied the one-factor and two-factor models to assess the term

structure of interest rates for the UK gilt-edged market.

In this thesis we use the two-factor and three-factor discrete time Vasicek models for our multifactor modeling. Following Sanders and Unal [1988] and others, we estimate the parameters of the Vasicek model using discrete time. Our interest rate however assumes a continuous range of values and not a discrete set of values. The Vasicek model is an affine model and its term structure and associated bond prices can be obtained explicitly. The analysis of most multifactor short rate models is computationally intensive. The Vasicek model which was one of the first to be developed is less complex. We choose this model mainly because of its simplicity and the existence of a closed-form solution. In the next section we discuss the dynamics under the discrete time Vasicek model.

1.1.1 Dynamics Under the Discrete Time Vasicek Model

The short rate is the growth rate of a risk-free asset between time t and $t+1$. The instantaneous short rate process under the three factor Vasicek model is defined as:

$$r(t) = \sum_{p=1}^{3} x_t^{(p)}. \tag{1.19}$$

where

$$x_{t+1}^{(p)} - x_t^{(p)} = \kappa_p(\theta_p - x_t^{(p)}) + \sigma_p \epsilon_{t+1,p}^{\mathbb{P}}. \tag{1.20}$$

κ_p, θ_p and σ_p, $p = 1, 2, 3$ are model parameters under $\mathbb{P}$. $\epsilon^{\mathbb{P}} = \{(\epsilon_{t,1}^{\mathbb{P}}, \epsilon_{t,2}^{\mathbb{P}}, \epsilon_{t,3}^{\mathbb{P}})\}_{t=1}^{T}$ is a strong standardized Gaussian white noise under $\mathbb{P}$, that is they are not only uncorrelated but independent of each other. We assume that $\epsilon_{t+1,1}^{\mathbb{P}}$, $\epsilon_{t+1,2}^{\mathbb{P}}$ and $\epsilon_{t+1,3}^{\mathbb{P}}$ are independent of each other.

Using the Girsanov transformation, (1.21), we perform a change of measure from the physical measure to the risk neutral measure:

$$\epsilon_{t+1,p}^{\mathbb{Q}} = \epsilon_{t+1,p}^{\mathbb{P}} + \lambda_t^{\mathbb{P}} \tag{1.21}$$

where $\epsilon^{\mathbb{Q}}$ is a strong standardized Gaussian white noise under $\mathbb{Q}$.

It is very important to work under the risk neutral measure because bonds are priced under the risk neutral measure. Substituting (1.21) in (1.20) we obtain

$$x_{t+1}^{(p)} = x_t^{(p)} + \kappa_p^{\mathbb{P}}((\theta_p^{\mathbb{P}} - \frac{\lambda_t^{\mathbb{P}}\sigma_p^{\mathbb{P}}}{\kappa_p}) - x_t^{(p)}) + \sigma_p^{\mathbb{P}}\epsilon_{t+1,p}^{\mathbb{Q}}. \tag{1.22}$$

With the following substitutions

$$\begin{aligned}
\sigma_p^{\mathbb{Q}} &= \sigma_p^{\mathbb{P}}, \\
\kappa_p^{\mathbb{P}} &= \kappa_p^{\mathbb{Q}}, \\
\theta_p^{\mathbb{Q}} &= \theta_p^{\mathbb{P}} - \frac{\sigma_p^{\mathbb{P}}\lambda_t^{\mathbb{P}}}{\kappa_p},
\end{aligned}$$

we obtain the dynamics of $x_t^{(p)}$ under the risk neutral measure as

$$x_{t+1}^{(p)} = x_t^{(p)} + \kappa_p^{\mathbb{Q}}(\theta_p^{\mathbb{Q}} - x_t^{(p)}) + \sigma_p^{\mathbb{Q}}\epsilon_{t+1,p}^{\mathbb{Q}}. \tag{1.23}$$

In the next section we drop the superscripts $\mathbb{Q}$ for κ, θ and σ. The bond price at time t of a zero coupon bond maturing in n periods is given as:

$$P(t, t+n) = E^{\mathbb{Q}}[e^{-\sum_{i=0}^{n-1} r(t+i)} \big| \mathcal{F}_t] \tag{1.24}$$

where $r(t) = x_t^{(1)} + x_t^{(2)} + x_t^{(3)}$. Given $\mathcal{F}_t$, the information at time t, generated by the process $(x_1, \ldots, x_t)$, $\sum_{i=0}^{n-1} r(t+i)$ is Gaussian. This is a discrete-time formula which differs from the continuous-time formulas used in models described in the literature above. The resulting affine term structure under the discrete time model as proven below is:

$$s(t, t+n) = \alpha(t, t+n) + \sum_{p=1}^{3} \beta_p(t, t+n) x_t^{(p)}. \tag{1.25}$$

The zero coupon price is therefore

$$P(t,t+n) = A(t,t+n)e^{-\sum_{p=1}^{3} B_p(t,t+n)x_t^{(p)}}. \tag{1.26}$$

with

$$A(t,t+n) = e^{-n\alpha(t,t+n)} \quad \text{and} \quad B_p(t,t+n) = n\beta_p(t,t+n).$$

We also obtain the expression:

$$\ln P(t,t+n) = \ln A(t,t+n) - \sum_{p=1}^{3} B_p(t,t+n)x_t^{(p)}.$$

$$P(t,t+n) = E^{\mathbb{Q}}[\exp(-\sum_{i=0}^{n-1}(\sum_{p=1}^{3} x_{t+i}^{(p)}))\big|\mathcal{F}_t].$$

We obtain the formulas for A and B_p by first introducing the following lemma.

Lemma 1.1.1. $x_{t+i} = x_t(1-\kappa)^i + \theta\kappa\sum_{j=1}^{i}(1-\kappa)^{i-j} + \sigma\sum_{j=1}^{i}(1-\kappa)^{i-j}\epsilon_{t+j}^{\mathbb{Q}}$, *for* $i = 0,...,n-1$.

Proof. Trivially, this is true for $i = 0$. Furthermore we assume it holds for a given i. Then

$$\begin{aligned}
x_{t+i+1} &= x_{t+i} + \kappa(\theta - x_{t+i}) + \sigma\epsilon_{t+i+1}^{\mathbb{Q}}. \\
&= x_{t+i}(1-\kappa) + \kappa\theta + \sigma\epsilon_{t+i+1}^{\mathbb{Q}} \\
&= (1-k)\left(x_t(1-\kappa)^i + \kappa\theta\sum_{j=1}^{i}(1-\kappa)^{i-j} + \sigma\sum_{j=1}^{i}(1-\kappa)^{i-j}\epsilon_{t+j}^{\mathbb{Q}}\right) + \kappa\theta + \sigma\epsilon_{t+i+1}^{\mathbb{Q}}. \\
&= x_t(1-\kappa)^{i+1} + \kappa\theta\sum_{j=1}^{i}(1-\kappa)^{i+1-j} + \sigma\sum_{j=1}^{i}(1-\kappa)^{i+1-j}\epsilon_{t+j}^{\mathbb{Q}} + \kappa\theta + \sigma\epsilon_{t+i+1}^{\mathbb{Q}}. \\
&= x_t(1-\kappa)^{i+1} + \kappa\theta\sum_{j=1}^{i+1}(1-\kappa)^{i+1-j} + \sigma\sum_{j=1}^{i+1}(1-\kappa)^{i+1-j}\epsilon_{t+j}^{\mathbb{Q}}.
\end{aligned}$$

which completes the induction. □

The formulas for A and B_p are therefore obtained as follows;

Theorem 1.1.2.

$$\ln(A(t,t+n)) = \sum_{p=1}^{3} \theta_p \left[\frac{1-(1-k_p)^n}{k_p} - n \right] + \sum_{p=1}^{3} \frac{{\sigma^2}_p}{2{k_p}^2} \left[\left(\frac{1-((1-k_p)^{2n}}{k_p(2-k_p)} \right) - 2\left(\frac{1-(1-k_p)^n}{k_p} \right) + n \right],$$

$B_p(t,t+n) = \frac{1-(1-k_p)^n}{k_p}$.

Proof.

$$\begin{aligned}
\sum_{i=0}^{n-1}\sum_{p=1}^{3} x_{t+i}^{(p)} &= \sum_{p=1}^{3}\sum_{i=0}^{n-1} \left(x_t^{(p)}(1-k_p)^i + \theta_p k_p \sum_{j=1}^{i}(1-k_p)^{i-j} + \sigma_p \sum_{j=1}^{i}(1-k_p)^{i-j}\epsilon_{t+j,p} \right) \\
&= \sum_{p=1}^{3} \left(x_t^{(p)} \sum_{i=0}^{n-1}(1-k_p)^i + \theta_p k_p \sum_{i=0}^{n-1}\sum_{j=1}^{i}(1-k_p)^{i-j} + \sigma_p \sum_{i=0}^{n-1}\sum_{j=1}^{i}(1-k_p)^{i-j}\epsilon_{t+j,p} \right) \\
&= \sum_{p=1}^{3} \left(x_t^{(p)} \frac{1-(1-k_p)^n}{1-(1-k_p)} + \theta_p k_p \sum_{i=0}^{n-1}\sum_{j=0}^{i-1}(1-k_p)^{j} + \sigma_p \sum_{i=0}^{n-1}\sum_{j=1}^{n-1}\mathbb{I}_{j\le i}(1-k_p)^{i-j}\epsilon_{t+j,p} \right) \\
&= \sum_{p=1}^{3} \left(x_t^{(p)} \frac{1-(1-k_p)^n}{k_p} + \theta_p k_p \sum_{i=0}^{n-1} \frac{1-(1-k_p)^i}{k_p} + \sigma_p \sum_{j=1}^{n-1} \left[\sum_{i=1}^{n-1} \mathbb{I}_{j\le i}(1-k_p)^{i-j} \right] \epsilon_{t+j,p} \right) \\
&= \sum_{p=1}^{3} \left(x_t^{(p)} \frac{1-(1-k_p)^n}{k_p} + \theta_p n - \theta_p \sum_{i=0}^{n-1}(1-k_p)^i + \sigma_p \sum_{j=1}^{n-1}\sum_{i=j}^{n-1}(1-k_p)^{i-j}{\epsilon_{t+j}}^p \right) \\
&= \sum_{p=1}^{3} \left((x_t^{(p)} - \theta_p) \frac{1-(1-k_p)^n}{k_p} + \theta_p n + \sigma_p \sum_{j=1}^{n-1} \frac{1-(1-k_p)^{n-j}}{k_p} \epsilon_{t+j,p} \right).
\end{aligned}$$

Finding the expectation and variance

$$\mathbb{E}\left[\sum_{i=0}^{n-1}\sum_{p=1}^{3} x_{t+i}^{(p)} \middle| \mathcal{F}_t \right] = \sum_{p=1}^{3} (x_t^{(p)} - \theta_p) \left[\frac{1-(1-k_p)^n}{k_p} \right] + \theta_p n.$$

$$
\begin{aligned}
Var\left[\sum_{i=0}^{n-1}\sum_{p=1}^{3} x_{t+i}^{(p)}\middle|\mathcal{F}_t\right] &= Var\left(\sum_{p=1}^{3}\left[\sigma_p\sum_{j=1}^{n-1}\frac{1-(1-k_p)^{n-j}}{k_p}\epsilon_{t+j,p}\right]\middle|\mathcal{F}_t\right)\\
&= \sum_{p=1}^{3}\sum_{j=1}^{n-1}\sigma_p^2\left[\frac{1-(1-k_p)^{n-j}}{k_p}\right]^2\\
&= \sum_{p=1}^{3}\sum_{j=1}^{n-1}\left[\frac{\sigma_p^2}{k_p^2}((1-k_p)^2)^{n-j}-2(1-k_p)^{n-j}+1\right]\\
&= \sum_{p=1}^{3}\sum_{j=1}^{n-1}\left[\frac{\sigma_p^2}{k_p^2}((1-k_p)^2)^{j}-2(1-k_p)^{j}+1\right]\\
&= \sum_{p=1}^{3}\frac{\sigma_p^2}{k_p^2}\left[\left(\frac{1-((1-k_p)^2)^n}{k_p(2-k_p)}-1\right)-2\left(\frac{1-(1-k_p)^n}{k_p}-1\right)+n-1\right]\\
&= \sum_{p=1}^{3}\frac{\sigma_p^2}{k_p^2}\left[\left(\frac{1-(1-k_p)^{2n}}{k_p(2-k_p)}\right)-2\left(\frac{1-(1-k_p)^n}{k_p}\right)+n\right].
\end{aligned}
$$

This implies,

$$P(t,t+n)=\exp(-\text{mean}+\frac{1}{2}\text{Variance})$$

Therefore

$P(t,t+n)=\exp\Bigg(-\sum_{p=1}^{3}(x_t^{(p)}-\theta_p)\left[\frac{1-(1-k_p)^n}{k_p}\right]-\theta_p n+\sum_{p=1}^{3}\frac{\sigma_p^2}{2k_p^2}\left[\left(\frac{1-((1-k_p)^{2n}}{k_p(2-k_p)}\right)$
$-2\left(\frac{1-(1-k_p)^n}{k_p}\right)+n\right]\Bigg).$

Thus,

$$
\begin{aligned}
\ln P(t,t+n) &= \log(A(t,t+n))-\sum_{p=1}^{3}B_p(t,t+n)x_t^{(p)}\\
&= -\sum_{p=1}^{3}(x_t^{(p)}-\theta_p)\left[\frac{1-(1-k_p)^n}{k_p}\right]-\theta_p n+\sum_{p=1}^{3}\frac{\sigma_p^2}{2k_p^2}\left[\left(\frac{1-(1-k_p)^{2n}}{k_p(2-k_p)}\right)-\right.\\
&\left.2\left(\frac{1-(1-k_p)^n}{k_p}\right)+n\right]\\
&= -\sum_{p=1}^{3}x_t^{(p)}\left[\frac{1-(1-k_p)^n}{k_p}\right]+\sum_{p=1}^{3}\theta_p\left[\frac{1-(1-k_p)^n}{k_p}\right]-\sum_{p=1}^{3}\theta_p n\\
&+\sum_{p=1}^{3}\frac{\sigma_p^2}{2k_p^2}\left[\left(\frac{1-(1-k_p)^{2n}}{k_p(2-k_p)}\right)-2\left(\frac{1-(1-k_p)^n}{k_p}\right)+n\right]. \qquad (1.27)
\end{aligned}
$$

Comparing the bond price formula to (1.27) we obtain

$$\ln(A(t,t+n)) = \sum_{p=1}^{3} \theta_p \left[\frac{1-(1-k_p)^n}{k_p} - n \right] + \sum_{p=1}^{3} \frac{{\sigma^2}_p}{2{k_p}^2} \left[\left(\frac{1-(1-k_p)^{2n}}{k_p(2-k_p)} \right) - 2\left(\frac{1-(1-k_p)^n}{k_p} \right) + n \right],$$

$$B_p(t,t+n) = \frac{1-(1-k_p)^n}{k_p}.$$

□

1.1.2 Types of Estimation Methods

In the previous section, the bond price was developed as an affine function of the functions $A(t,t+n)$ and $B_p(t,t+n) \quad p=1,2,3$. This relationship contains the parameter set (κ_p, θ_p, and σ_p) for each factor $x_t^{(p)}$. We seek to estimate these parameters.

A lot of work has been done in the estimation and calibration of interest rate models using various techniques. We will discuss some of these calibration methods. For the purposes of discussion, we have grouped the calibration methods under these three main headings: methods based on moments, maximum likelihood-based methods, and methods based on Bayesian statistics.

Methods based on moments

Some estimation methods based on moments are the method of moments (MOM), generalised method of moments (GMM) and efficient method of moment estimation (EMM). Methods based on moments involve equating sample moments with theoretical moments. Theoretical moments are defined as $\mathbb{E}[X^k]$, the kth moment of the distribution about the origin and the $\mathbb{E}[(X-\mu)^k]$, the kth moment of the distribution about the mean. The general idea behind this method is to equate the first few sample moments around the origin to respectively the first few sample moments of the distribution. We continue equating sample moments around

the origin to the moments of the distribution until we have as many equations as the number of parameters. We then solve for the parameters. The method of moments is based on law of large numbers. The law of large numbers states that given an independent random variable $X_1, X_2, ...$ with the common mean of X_i being μ_x then

$$\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu_x \quad \text{as} \quad n \to \infty.$$

Let $f_X(x|\lambda)$ be the probability density function of X with parameter λ and μ_x is a function $q(\lambda)$ of λ. Then it means that if $\mu = q(\lambda)$ then $\bar{X} \simeq q(\lambda)$ for large n. Another method based on moment is the generalised method of moments estimator. In the MOM we generate k equations in k unknowns to find the estimator but in the GMM method as described by Hansen [1982] we have more moments than parameters. In the generalised method of moments to obtain the optimal parameters we minimise the sum of square errors of the differences between the sample moments and the population moments. These methods are often used as starting points for MLE and least squares estimation. Chan et al. [1992] estimates the parameters of short term interest rate models in discrete time using the generalised method of moments. In their work they tested how various short rate models fit the term structure of short term treasury bill yields.

Maximum likelihood

Maximum likelihood estimation and methods based on maximum likelihood have been employed in the estimation of interest rate models. See for instance Vojtek [2004] for its application in the context of a term structure model driven by a GARCH process. Maximum likelihood estimation (MLE) is a method which is used to estimate the parameters of a statistical model given observations. It attempts to find the parameter values that maximize the likelihood function given the data. Under the assumption that the probability density function of the data has a particular parametric form, the method ascertains which parame-

ter value would yield the greatest likelihood of obtaining the observed data. In other words, the method chooses the probability density function under which the observed data would have the highest likelihood of occurring.

The MLE method requires knowledge of the short rate as the likelihood function is derived from the equation of the state variable. The closed form of the likelihood function is known only in some cases for the various short rate models. For example the CIR model implies a non-central χ^2 and the geometric Brownian motion implies a log-normal density, for which closed forms of the likelihood functions are available. Also for the Vasicek model, as $r(t)$ follows a normal distribution the distribution of each of the $r(t+1)$ is given as

$$f\big(r(t+1)\big|r(t)\big) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[r(t+1)-(r(t)+k(\theta-r(t)))]^2}{2\sigma^2}}$$

The likelihood function is simply the joint density function of the sample data and the log likelihood function is defined as

$$\mathcal{L} = \ln \prod_{i-1}^{n} f\big(r(i+1)\big|r(i)\big) = \sum_{i=1}^{n} \ln f\big(r(i+1)\big|r(i)\big).$$

To estimate the parameters, the likelihood is maximized with respect to the parameters κ, θ and σ. Chen and Scott [1993] used maximum likelihood estimation for estimating the parameters of the multifactor CIR model. In their work they used the rates of both short-term treasury bills and long term bonds to represent values for the unobservable state variables. Pearson and Sun [1989] also used maximum likelihood to estimate a two factor version of the CIR model. However, in their model they used only short-term treasury bills to infer values for the unobservable state variables.

Methods based on Bayesian Statistics

One of the methods based on Bayesian statistics is the Kalman filter method. It is an iterative mathematical process that uses a set of equations and consecutive data inputs to

quickly estimate an unobservable process when the measured values contain random errors. Rudolph E. Kalman published a paper in 1960, Kalman [1960], which introduced the linear filtering technique named after him. A good explanation of the Kalman filter method can be found in the book Harvey [1990].

It is very useful in situations where the underlying state variables are not observable. This implies that Kalman filter can be used to estimate parameters for both CIR and Vasicek models. It has been used to estimate the parameters of the continuous time Vasicek and CIR models using Canadian zero-coupon bond yields by Bolder [2001]. To perform the Kalman filter a state space formulation which consists of a measurement system and a transition system is required.

When used to fit the yield curve, the measurement system is an observed system which represents the affine relationship between the market zero coupon bonds and the state variables. The transition system is then an unobserved system of equations which describes the dynamics of the state system. A recursive approach is used to infer values about the unobserved state variables. These variables are used to construct and maximize a log likelihood function to find the optimal values of the unknown parameters for the underlying model (for example κ, θ, σ and λ for the Vasicek model).

The main calculations done in the performance of the Kalman filter are the computations of the Kalman gain, the current estimate and the estimation error. The Kalman gain is a weight between the measurement and the previous estimate. If the Kalman gain is close to one it implies the measurements are very accurate and such measurements contribute more to the underlying estimates. If the Kalman gain is very small, that is if the error in the measurement is large relative to the error in the estimate, the Kalman gain will ensure we put less reliance on the measured value.

One advantage the Kalman filter has over other statistical methods such as OLS and LSM is that the Kalman filter uses all the present and past information to determine the current estimate and the yield curve, and also does not require that the state variable is

observable. Also, the Kalman filter is used for MLE in situations where the state variable is not observable. The Kalman filter method will be used in the estimation of parameters in this thesis.

1.1.3 The Kalman Filter In Detail

The Kalman filter begins by estimating the state transition space $X \in \mathbb{R}^n$ of a discrete time process

$$X_t = FX_{t-1} + G\mu_{t-1} + W_t,$$

with a measured space $S_t \in \mathbb{R}^m$ that is observed, thus $S_t = \alpha + \beta X_t + \varepsilon_t$. The variable μ is an optional control variable matrix, and it can be assumed to be an identity matrix. The dimensions of F, G, W, α, β and ε and what they represent are to be explained below. The random variables W_t and ε_t represent the process and measurement errors respectively. They are assumed to be independent and follow a normal distribution. The state space formulation is necessary for the Kalman filter procedure. Using our discrete time Vasicek model developed in Section 1.1.1 we perform a state space formulation for a three factor discrete time Vasicek model. The framework developed in this section has some features in common with Bolder [2001].

Let m be the number of bond maturities considered within the dataset, $s(t_i, T_j)$ for $j = 1, ..., m$ being the spot rate of the bond at time t_i with maturity T_j. T_j are a set of fixed maturities considered for bonds within the estimation period. We denote each time step as $\Delta t = t_i - t_{i-1}$. The measurement system equation is then described as

$$\begin{bmatrix} S(t_i,T_1) \\ S(t_i,T_2) \\ \vdots \\ S(t_i,T_m) \end{bmatrix} = \begin{bmatrix} \alpha(t_i,T_1) \\ \alpha(t_i,T_2) \\ \vdots \\ \alpha(t_i,T_m) \end{bmatrix} + \begin{bmatrix} \beta_1(t_i,T_1) & \beta_2(t_i,T_1) & \beta_3(t_i,T_1) \\ \beta_1(t_i,T_2) & \beta_2(t_i,T_2) & \beta_3(t_i,T_2) \\ \vdots & & \\ \beta_1(t_i,T_m) & \beta_2(t_i,T_m) & \beta_3(t_i,T_m) \end{bmatrix} \begin{bmatrix} x_{t_i}^{(1)} \\ x_{t_i}^{(2)} \\ x_{t_i}^{(3)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t_i}^{(1)} \\ \varepsilon_{t_i}^{(2)} \\ \vdots \\ \varepsilon_{t_i}^{(m)} \end{bmatrix} \tag{1.28}$$

or

$$S_{t_i} = \alpha + \beta X_{t_i} + \varepsilon_{t_i} \tag{1.29}$$

where

- $S_{t_i} : m \times 1$ matrix of the spot rates for all bonds at time t_i,
- $\alpha : m \times 1$ matrix for all bonds characterized by (2.29) ,
- $\beta : m \times 3$ matrix for all bonds and all factors for each time characterized by (2.30)
- $\varepsilon_{t_i} \sim N(0, P)$ with ε_{t_i} being the vector containing $\varepsilon_{(1,t_i)} \dots \varepsilon_{(m,t_i)}$

$$P = \begin{bmatrix} {r_1}^2 & 0 & 0 & \dots & 0 \\ 0 & {r_2}^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & {r_m}^2 \end{bmatrix}. \tag{1.30}$$

The transition system is also defined as

$$\begin{bmatrix} x_{t_i}^{(1)} \\ x_{t_i}^{(2)} \\ x_{t_i}^{(3)} \end{bmatrix} = \begin{bmatrix} \kappa_1\theta_1 \\ \kappa_2\theta_2 \\ \kappa_3\theta_3 \end{bmatrix} + \begin{bmatrix} (1-\kappa_1) & 0 & 0 \\ 0 & (1-\kappa_2) & 0 \\ 0 & 0 & (1-\kappa_3) \end{bmatrix} \begin{bmatrix} x_{t_{i-1}}^{(1)} \\ x_{t_{i-1}}^{(2)} \\ x_{t_{i-1}}^{(3)} \end{bmatrix} + \begin{bmatrix} w_{t_i}^{(1)} \\ w_{t_i}^{(2)} \\ w_{t_i}^{(3)} \end{bmatrix} \tag{1.31}$$

or

$$X_{t_i} = G + FX_{t_{i-1}} + W_{t_i} \tag{1.32}$$

where

$X = [x_1, x_2, x_3]$

$W_{t_i} \sim N(0, R)$

$$R = \begin{bmatrix} {\sigma_1}^2 & 0 & 0 \\ 0 & {\sigma_2}^2 & 0 \\ 0 & 0 & {\sigma_3}^2 \end{bmatrix}. \tag{1.33}$$

With the state space formulation obtained we set our initial values to start the recursion. We use the unconditional mean and variance of our transition system as our starting values. The discrete time Vasicek model follows an AR(1) process. From (1.20)

$$\begin{aligned} x_{t+1} &= \kappa\theta + (1-\kappa)x_t + \sigma\epsilon_{t+1} \\ E[x_{t+1}] &= \kappa\theta + (1-\kappa)E[x_t] + E[\sigma\epsilon_{t+1}] \\ \mu &= \kappa\theta + (1-\kappa)\mu + 0 \\ \mu &= \theta \\ E[x_t] &= \theta, \end{aligned}$$

therefore for the three-factor model, $E[x_1] = [\theta_1 \quad \theta_2 \quad \theta_3]^T$.
The unconditional variance for the transition Equation (1.20) is given by

$$Var[x_t] = E[x_t^2] - \mu^2 = \frac{\sigma^2}{\kappa(2-\kappa)}.$$

Therefore for the three factor,

$$Var[X_1] = \begin{bmatrix} \frac{{\sigma_1}^2}{\kappa_1(2-\kappa_1)} & 0 & 0 \\ 0 & \frac{{\sigma_2}^2}{\kappa_2(2-\kappa_2)} & 0 \\ 0 & 0 & \frac{{\sigma_3}^2}{\kappa_3(2-\kappa_3)} \end{bmatrix}.$$

Using the initial guess we begin the recursive approach by finding the estimated values of the measurement equation. We let $\hat{X}_{t_i|t_{i-1}}$ be the estimated value of the conditional

expectation of X_{t_i} at time t_i given information at time t_{i-1}. Therefore

$$\hat{X}_{t_i|t_{i-1}} = G + F\hat{X}_{t_{i-1}}. \tag{1.34}$$

The estimated values of the measurement equation are based on the initial values. The conditional expectation of the measurement equation is:

$$E[S_{t_i|t_{i-1}}] = \alpha + \beta E[X_{t_i|t_{i-1}}], \tag{1.35}$$

and its corresponding conditional variance is

$$Var[S_{t_i|t_{i-1}}] = \beta \mathrm{Var}[X_{t_i|t_{i-1}}]\beta^T + P, \tag{1.36}$$

where $\mathrm{Var}[X_{t_i|t_{i-1}}]$ is the conditional variance of the transition matrix, and P is from (1.30). Having the expected measurement equation which is our estimated spot rates we define $\zeta_{t_i|t_{i-1}}$ to be the measurement errors, the difference between the observed data and the expected spot rates.

$$\zeta_{t_i|t_{i-1}} = S_{t_i} - E[S_{t_i|t_{i-1}}]. \tag{1.37}$$

Next we let K, represent the Kalman gain. The Kalman gain is explained in Section 1.1.2. This is a correction we will add to the estimate. The estimated state variable then becomes

$$\hat{X}_{t_i|t_i} = \hat{X}_{t_i|t_{i-1}} + K_{t_i}\zeta_{t_i|t_{i-1}}, \tag{1.38}$$

with

$$K_{t_i} = \frac{\mathrm{Var}[X_{t_i|t_{i-1}}]\beta^T}{\beta \mathrm{Var}[X_{t_i|t_{i-1}}]\beta^T + P}. \tag{1.39}$$

We seek to minimise the variance in the error term ε_{t_i} in (1.29) to get reliable estimated parameters. If $\mathrm{Var}[X_{t_i|t_{i-1}}] \to 0$ then the previous measurement updates are mostly ignored. We want to make sure that $\mathrm{Var}[X_{t_i|t_{i-1}}]$ does not go to zero as this will make the measurement

space unreliable. Using the Kalman gain we update the conditional variance of the state system

$$\text{Var}[X_{t_i|t_i}] = (1 - K_{t_i}\beta)\text{Var}[X_{t_i|t_{i-1}}]. \tag{1.40}$$

Using

$$\hat{X}_{t_{i+1}|t_i} = G + F\hat{X}_{t_i|t_i} \tag{1.41}$$

and

$$\text{Var}[X_{t_{i+1}|t_i}] = \beta\text{Var}[X_{t_i|t_i}]\beta^T + R \tag{1.42}$$

as the new predicted state and error in estimate we forecast the unknown values of our next time period conditional on the values of the previous time period. We repeat the steps for all the time steps in our data. The new estimate $\hat{X}_t$ is called the filtered state. In doing this we obtain the measurement error ζ_{t_i} and the estimate error $\text{Var}[X_{t_i|t_{i-1}}]$ at every time step. With the assumption that the error terms are normally distributed we construct the log likelihood function as

$$\mathcal{L} = -\frac{nN}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{N}\ln[\text{detVar}(X_{t_i|t_{i-1}})] - \frac{1}{2}\sum_{t=1}^{N}\zeta_{t_i}\text{Var}[X_{t_i|t_{i-1}}]^{-1}{\zeta_{t_i}}^T \tag{1.43}$$

with n being the dimension space of X_t and N being the number of collected data. To obtain our optimal parameters optimisation techniques are applied to maximize the log likelihood. Putting all the steps together we have the Kalman algorithm as

1. Choose initial values X_0 the unconditional mean and P_0 the unconditional variance, with previous state denoted as $X_{t_{i-1}}$ and $P_{t_{i-1}}$.

2. $\hat{X}_{t_i|t_{i-1}} = G + F\hat{X}_{t_{i-1}}$.

3. $E[S_{t_i|t_{i-1}}] = \alpha + \beta E[X_{t_i|t_{i-1}}]$ and $Var[S_{t_i|t_{i-1}}] = \beta\text{Var}[X_{t_i|t_{i-1}}]\beta^T + P$.

4. $\zeta_{t_i|t_{i-1}} = S_{t_i} - E[S_{t_i|t_{i-1}}]$.

5. $K_{t_i} = \frac{\text{Var}[X_{t_i|t_{i-1}}]\beta^T}{\beta\text{Var}[X_{t_i|t_{i-1}}]\beta^T + P}$.

6. $\hat{X}_{t_i|t_i} = \hat{X}_{t_i|t_{i-1}} + K_{t_i}\zeta_{t_i|t_{i-1}}$.

7. $\text{Var}[X_{t_i|t_{i-1}}] = (1 - K_{t_i}\beta)\text{Var}[X_{t_i|t_{i-1}}]$.

8. $\hat{X}_{t_{i+1}|t_i} = G + F\hat{X}_{t_i}$ and $\text{Var}[X_{t_{i+1}|t_i}] = \beta\text{Var}[X_{t_i}]\beta^T + R$.

1.2 Machine Learning Yield Curve modeling

Machine learning often involves the scientific study or use of statistical methods to perform a specific task. Computers can perform tasks without being explicitly programmed, but by learning from data, identifying patterns in them and using these to make decisions with little or no human intervention. Machine learning techniques can be used to make predictions or statistical inference. There are many different types of machine learning algorithms and these are grouped according to the type of learning involved—that is either supervised learning or unsupervised learning. In supervised learning, models are trained by comparing their output against known true answers while with unsupervised learning, no correct answer is provided during the training phase. Supervised learning uses classification and regression techniques to develop predictive models. Some supervised learning techniques are the K-nearest neighbor, logistic regression and neural networks. Unsupervised learning techniques seek to find patterns or ways of categorizing or subgrouping the data based on variables and observations. Some unsupervised learning techniques are principal component analysis and various clustering methods. Unsupervised learning will be explored in this thesis—specifically principal components analysis.

1.2.1 Principal Component Analysis (PCA)

Principal component analysis is an unsupervised learning technique which is used to reduce the dimensions of a large data set by deriving a low-dimensional set of features from the large

dataset. It is a procedure that uses linear transformation to transform correlated variables into a set of linearly uncorrelated features.

The goal of the principal component analysis in this thesis is to determine the number of principal components required to sufficiently explain the yield curve for the Canadian zero coupon data. This information has been shown to be important for investors. Litterman and Scheinkman [1991] demonstrated that the number of factors from the principal component analysis that affect zero coupon data is very useful for market participants who hold fixed income securities–particularly for investors looking to hedge against future volatility by helping them categorize instruments by their potential future performance. This helps them manage their return risk. In their paper, they identified that three factors—level, steepness and curvature—affect the return on U.S. government bonds and related securities.

Bolder et al. [2004] also used principal component analysis on the Canadian zero coupon bond yields over the period 1986 to 2003, to determine that the variation in the yields can be explained by similar factors. In their work they observed that the first three principal components explains 99.5% of the variation in the data. This is consistent with the observations made by Litterman and Scheinkman [1991]. We replicate this analysis for the period 2004 to 2018 to determine if the pattern of results for the period before the financial crisis also holds true for the period after the financial crisis. We will further compare the PCA as a machine learning yield curve modeling technique with econometrics methods, specifically, the multi-factor Vasicek model.

The methodology for performing PCA to fit the yield curve is explained as follows. Let us consider an X^{obs} matrix with dimensions $n \times m$ where n is the number of dates of the sample, and m being the different maturities for which the yield curve is being analyzed. PCA is applied on a zero mean set of predictors as we are interested in the variance, therefore we let X be a set of centered predictors,

$$X = X_j^{obs} - \bar{X}_j^{obs}. \tag{1.44}$$

Let $a_{1i}, ..., a_{mi}$ be elements of the vector a_i for $i = 1, \ldots, m$. We seek to obtain vectors of an orthonormal base where the first basis vector maximizes the sample average squared projection length of each data points on the basis vector, as defined by James et al. [2013], that is

$$\underset{a_{11},...a_{p1}}{\text{maximize}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_{j1} X_{ij} \right)^2 \right\} \quad \text{subject to} \sum_{j=1}^{m} {a_{j1}}^2 = 1. \tag{1.45}$$

In general we seek to solve sequentially for $j = 1, \ldots, m$ an optimisation procedure to identify a vector a_j solving

$$\max_{a_j \in \mathbb{R}^m} \sum_{i=1}^{n} \langle a_j, X_i \rangle^2 \tag{1.46}$$

under the constraints $\|a_j\| = \sqrt{\langle a_j, a_j \rangle} = 1$ and $\langle a_j, a_k \rangle = 0$ for $k = 1, \ldots, j-1$.

The $(a_{1i}, ..., a_{mi})$, $i = 1, \ldots, m$ are the orthornormal basis belonging to the PCA optimisation problem and these are called the principal components. The first principal component is one that has the largest eigenvalue. The row i of matrix X can be expressed as

$$X_i = [X_{i,1} \quad \cdots \quad X_{i,m}] = \sum_{j=1}^{m} Z_{i,j} a_j, \tag{1.47}$$

for some coordinates $Z_{i,j}$, $j = 1, \ldots, m$, $i = 1, \ldots, n$. This leads to the equation $Z = XA^{-1}$ where X is the centered data, $A = [a_1, ..., a_m]$ and Z is the $n \times m$ matrix containing the $Z_{i,j}$. Since A is orthogonal it is invertible with $A^{-1} = A^T$. The co-ordinates $Z_{1,j}....Z_{n,j}$ for $j = 1, ..m$ are called the scores of the j^{th} principal component.

Next we consider the number of principal components required to capture the behavior of the yield curve. We want to retain sufficient principal components such that the retained data explains a high percentage of the total variance of the initial data.

The percentage of variance explained for the j^{th} principal component is defined as

$$PVE_j = \frac{S^2_{z(j)}}{\sum_{j=1}^{m} S^2_{z(j)}} \quad \text{or} \quad PVE_j = \frac{\lambda_j}{\sum_{j=1}^{m} \lambda_j} \tag{1.48}$$

$S^2_{z(j)} = \sum_{i=1}^{n} Z^2_{i,j}$ where $j = 1, .., m$ is the variance such that $S^2_{z(j)}$ is large when j is small.

λ_j are the eigenvalues of the matrix $X^T X$. The cumulative percentage of variance explained (CPVE) is given as

$$CPVE_m = \sum_{k=1}^{m} PVE_k.$$

We want to ensure that the retained observation explains at least $100\alpha\%$ of the total variance of observation.

1.3 Data

The historical data used for this thesis are the Canadian zero-coupon bond pseudo rates which were obtained from Bank of Canada [2018]. The data were recorded on a daily basis for the period 1986 to 2018. The term structure was backed-out from bond price data by Bolder et al. [2004] using the Merrill Lynch Exponential Spline (MLES) model as described by Li et al. [2001]. The MLES was assessed and evaluated by Bolder and Gusba [2002] against other estimation algorithms and he concluded that it was the most desirable to use to derive the zero-coupon bond rates. Bolder and Gusba [2002] assessed a model as good if it provided zero-coupon rates that were a good approximation to the corresponding set of coupon bonds in the economy. Coupon bonds constitute a portfolio of pure-discount bonds and the price of these coupon bonds is merely the sum of these pure discount bonds.

The Merrill Lynch Exponential spline model is a function based model, which uses single-piece functions defined over the entire term to maturity domain rather than piecewise cubic polynomials or splines. The basis function in a spline model can take different forms. In Li et al. [2001] the discount function is modelled as a linear combination of exponentials. That is

$$d(t) = \sum_{k=1}^{K} \eta_k e^{-k\alpha t} \tag{1.49}$$

where η_k are unknown parameters and α is the long term instantaneous forward rate.

The basis function was chosen as $e^{-k\alpha t}$ for two reasons. Firstly exponentials are strongly related to the discount function, and secondly the parameter α represents a long term instantaneous forward rate such that, for large values of t, the discount function is approximately $e^{-k\alpha t}$ (see Bolder and Gusba [2002]).

The discount function was used in computing the theoretical bond price which was then used in generating the yield curve. A summary of how this was performed by Bolder et al. [2004] is as follows.

Let c_{ij} be the jth payment of the ith bond, τ_{ij} be the time the jth payment of the ith bond occurs and m_i be the remaining number of payments for the ith bond. The theoretical bond price was defined as

$$\hat{P}_i = \sum_{j=1}^{m_i} c_{ij} d(\tau_{ij}) \tag{1.50}$$

Substituting our basis function in (1.50) we have our theoretical price now defined as $\hat{P} = QZ$ where $\hat{P}$ is a vector which contains all $\hat{P}_i$, $Z = (\eta_1, ...\eta_K)^\intercal$, K is the number of basis function used and Q is a matrix of dimension $N \times K$ where N is number of bonds. We seek to minimize the pricing error across the sample of bonds as we are generating the term structure from the theoretical bond prices. Given the higher sensitivity of price to one unit of yield for long term bonds it is necessary to assign weights to each bond to ensure that the price error on the various bonds are treated differently (see Bolder et al. [2004]). The maximum likelihood estimate of $\eta_1, ...\eta_K$ was found using generalised least squares since the theoretical prices are a linear function of the unknown parameters. It was assumed that the pricing errors are normally distributed with zero mean and variance that is proportional to $1/w_j$ where w_j is weight assigned to bond j. After obtaining the parameters $\eta_1, ...\eta_K$ the unknown parameter α was estimated by using a numerical optimization technique to solve for the value of α which minimizes the residual pricing errors. This gave us the yield curve.

In conducting our descriptive analysis, we considered the daily observations for the period 2004 to 2018, as the data for previous years has been descriptively analysed by Bolder et al. [2004]. We subsampled the data into 2 periods—before and after the 2008 financial crises. i.e.

2004 to 2007, and 2008 to 2018 respectively. For our analysis in chapter four, we considered only data for the third day of each month (substituting the subsequent day if data for the third day were not available). We limited our analysis to the third day because it generally had less missing data.

For our principal component analysis we used data from the period 2004 to 2018 as Bolder et al. [2004] has done similar work for previous years. For the Vasicek model however, we considered the period 1991 to 2018. We used a total of 30 bonds with the shortest maturity being 3-months and the longest being 30 years. The bonds were chosen to capture different points of the yield curve. This was done by using equally maturity intervals from the 1year bond to the 30yr bond. We however omitted the 13^{th}(156-month bond) and 23^{rd}(276-month bond) bonds and rather include two short term bonds—the 3-month bond and the 9-month bond. This was done in consideration of the fact that long term bonds are sensitive to mean reversion and the risk premium (see Chen and Scott [2003]).

Chapter 2

Descriptive Statistics

In this chapter we perform a descriptive analysis of the data that is used in Chapter 3 for the yield curve modelling. In Section 1, we discuss the average yield curve levels and yield statistics, and in Section 2, we discuss the descriptive statistics of the first differences as it was done by Bolder et al. [2004] in the Bank of Canada working paper.

We will focus on the 3-month and 30-year yields as representative of the short and long term yields respectively. We will also discuss the slope of the yield curve (the difference between the 3-month and 30-year rates) and the degree of curvature. The data and period to be used for this work are as explained in Chapter 1, i.e. with the full period being from 2004 to 2018, and the sub-sample periods being from 2004 to 2007 (before the financial crisis), and 2008 to 2018 (after the financial crisis). Following Bolder et al. [2004], the curvature is defined as

$$C = 16y - 0.5(2y + 30y). \tag{2.1}$$

That is, the curvature is equal to the difference between the yield on a 16-year bond and a linear interpolation between the 2-year and 30-year yields. The choice of these bonds for estimating the curvature allows us to compare equivalent durations [1]. That is because the duration of a zero-coupon bond is equivalent to its maturity.

[1] if the distance between A and B, $\overline{AB}$, is the same as the distance B and C, $\overline{BC}$, we can compute the curvature by (2.1)

2.1 Descriptive Statistics of Yield-Curve Levels

In this section we discuss the yield curve levels and present the statistics for the various yield curve measures. We consider the statistics for the full period as well as for the sub-sample periods.

Figure 2.1 **Average Zero Coupon Yield Curve-Full Period**

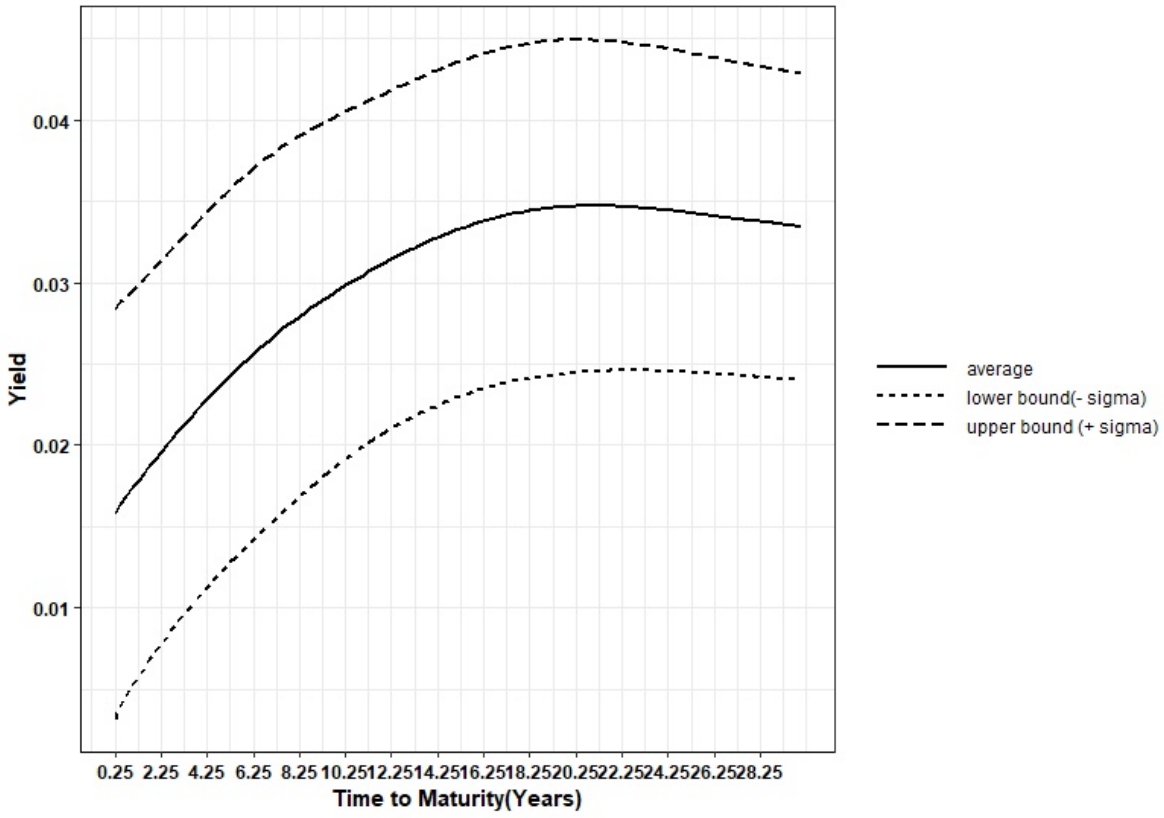
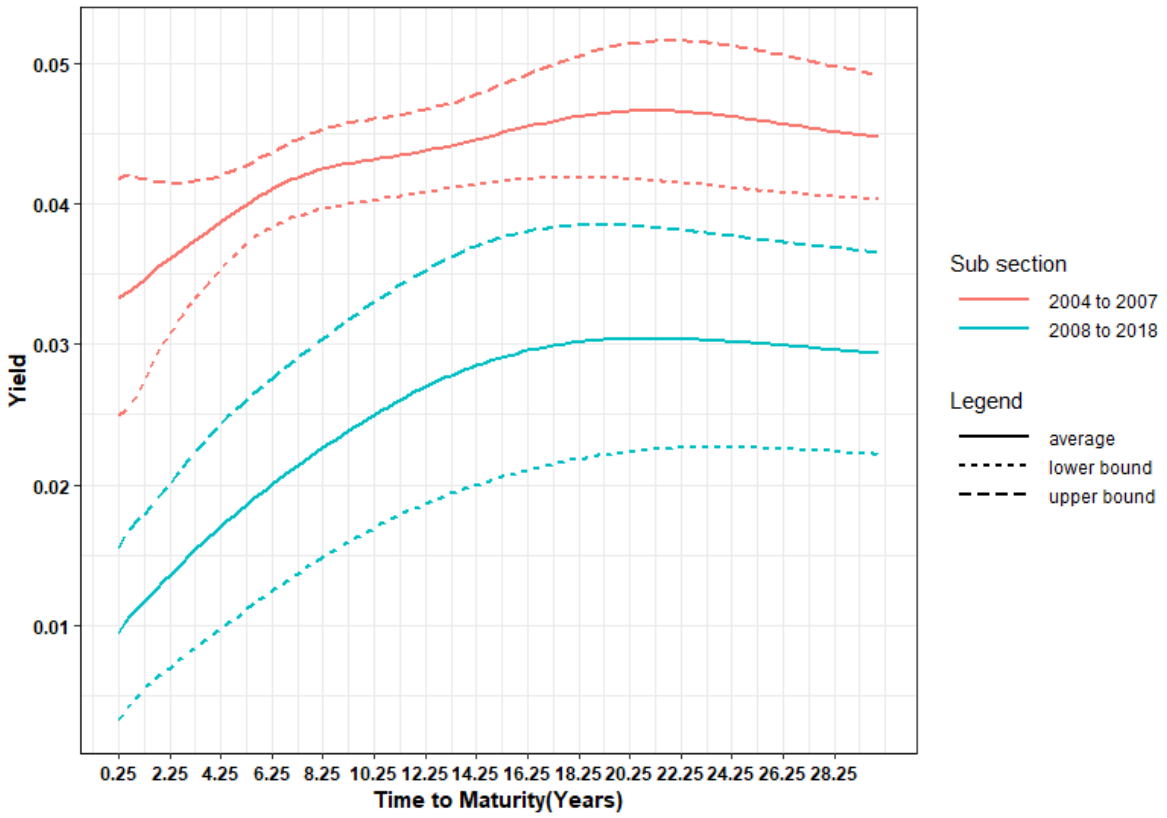


Figure 2.1 shows the average yield curve and the one-sample standard deviation bands (+/- σ) for the period 2004 to 2018. The average yield is an increasing curve starting from approximately 1.6 percent until it reaches it's maximum level of approximately 3.5 percent. From this maximum, it starts declining steadily to 3.34 percent. The maximum level occurs between the 19.5-year bond and 23-year bond.

The bands also follow the same increasing trend with the short term bond having a lower band of 0.32 percent and an upper band of 2.8 percent. The upper band increases gradually

and peaks at 4.5 percent and starts declining steadily afterwards. The lower band also increases at the same rate and becomes steady between the 20.5 year bond and the 22.5 year bond. The long term bond has a lower band of 2.4 percent and an upper band of 4.2 percent. The variation for the short term yield was higher than the long term yield. The scattering of the yields around the average remains approximately stable.

Figure 2.2 **Average Zero Coupon Yield Curve-Sub-Sample**



In Figure 2.2 we consider the average yield curve and the one standard deviation bands for the sub-sample period 2004 to 2007 (before the financial crisis) and 2008 to 2018 (after the financial crisis).

For the period before the financial crisis the trend of the average yield curve is non monotonous. It has slight fluctuations for the short term bonds, and starts declining steadily with the long term bonds.

The lowest yielding bond in the pre-financial crisis period was the 3-month bond with an

average yield of 3.3% (standard deviation (SD) = 0.8%)). In the post-crisis period, the same bond had an average yield of 0.9% (SD = 0.6%). It was again the lowest yielding bond for the period. The highest yielding bond was the 21-year bond in both pre- and post-financial crisis periods. It had an average yield of 4.7% (SD = 0.5%) in the pre-crisis period and this declined to 3.0% (SD = 0.8%) in the post-crisis period. The longest-maturing bond (30-year bond) had an average yield in the pre-crisis period of 4.5% (SD = 0.4%) and this went down to an average of 2.9% (SD = 0.7%) in the post-crisis period.

The average yield curve for the period before the financial crisis is higher than after the financial crisis. This is consistent with our understanding of the market situation following the Great Recession when the U.S. Federal Reserve dropped its Fed Funds Rate to a range of 0.00 percent to 0.25 percent Kenton [2018].

Table 2.1 shows the summary of the yield curve statistics for the full data and Figure 2.3 shows the time series plot and the histogram for the various yield curve measures.

Table 2.1 **Summary of Yield Curve Statistics-Full period**

Yield Curve Measure	**Mean**	**Max**	**Min**	**Std Dev**	**Kurt.**	**Skew.**
3-Month Yield	1.58%	4.61%	0.077%	1.26%	2.72	1.015
30-Year Yield	3.35%	5.41%	1.54%	0.95%	1.88	0.184
Slope	1.77%	4.04%	-0.39%	1.00%	2.77	0.032
Curvature	73.96bps	183.74bps	0.84bps	42.00bps	3.09	0.609

From Figure 2.3 we observe that the 3-month rate shows an increasing trend during the initial part of the curve which peaks between mid 2006 and 2007, and then starts decreasing afterwards. The 30 year rate also shows a decreasing trend. The curve drops considerably

from about 2006 to 2007 and then increases between the period from approximately 2008 to 2011. It then decreases gradually afterwards for the remaining years with a few fluctuations.

Table 2.2 **Summary of Yield Curve Statistics-2004 to 2007**

Yield Curve Measure	**Mean**	**Max**	**Min**	**Std Dev**	**Kurtosis**	**Skewness**
3-Month Yield	3.33%	4.61%	2.01%	0.84%	1.33	-0.12
30-Year Yield	4.47%	5.41%	3.86%	0.44%	2.03	0.68
Slope	1.14%	3.34%	-0.39%	1.21%	1.56	0.39
Curvature	51.15bps	125.26bps	0.84bps	37.09bps	1.83	0.52

Table 2.3 **Summary of Yield Curve Statistics-2008 to 2018**

Yield Curve Measure	**Mean**	**Max**	**Min**	**Std Dev**	**Kurtosis**	**Skewness**
3-Month Yield	0.94%	3.83%	0.077%	0.611%	7.81	1.93
30-Year Yield	2.93%	4.29%	1.59%	0.719%	1.74	0.36
Slope	1.99%	4.045%	-0.39%	0.800 %	3.16	0.73
Curvature	82.25bps	183.74bps	0.84bps	40.59bps	3.219	0.855

Table 2.2 and 2.3 shows the summary of the statistics for the sub-sample periods.

For the two sub-sample periods we observe that the short term yield shows an increasing trend from 2004 to 2007 and then declines sharply for the early part of 2008 with a few fluctuations. It then decreases steadily from 2011 to 2015 and finally increases again from later 2017 to 2018.

The long term maturity shows a general decreasing trend for both the period 2004-2007 and 2008-2018. The slope and curvature both indicate a decreasing trend for the period 2004 to 2007. There were significant spikes between the period 2009-2011 for both the slope and curvature with the trend generally decreasing afterwards.

The means for the 3-month yields and 30-year yields are both lower for the period 2008-2018 compared to 2004-2007. The variation for the 30 year rates is higher in the period 2008-2018 than in 2004-2007, however that of 3-month bonds is the opposite. The slope and curvature for the latter period is higher than the former. The distribution for both periods are non-normal.

2.2 Descriptive Statistics of First Differences

In this section we look at the daily changes in the yield curve measures. This is important to consider as daily changes in the level and shape of the yield curve affect the short-term risk and return behavior for government bonds. Figure 2.4 and Table 2.4 give the behaviorial and statistical details for the full period while Table 2.5 and 2.6 present the same results for the sub-samples.

Table 2.4 **Summary of Yield Curve Measure First Differences-Full period**

Yield Curve Measure	**Mean (bps)**	**Max (bps)**	**Min (bps)**	**Std Dev (bps)**	**Kurtosis (bps)**	**Skewness (bps)**
3-Month Yield	-0.03	69.28	-50.99	3.58	100.98	2.11
30-Year Yield	-0.08	19.50	-14.41	3.48	3.83	0.05
Slope	-0.06	58.55	-67.89	4.68	32.51	-0.36
Curvature	-0.03	11.62	-11.95	1.71	7.58	0.19

Table 2.5 **Summary of Yield Curve Measure First Differences-2004 to 2007**

Yield Curve Measure	Mean (bps)	Max (bps)	Min (bps)	Std Dev (bps)	Kurtosis (bps)	Skewness (bps)
3-Month Yield	0.12	33.01	27.37	3.18	22.57	0.38
30-Year Yield	-0.13	10.67	-10.09	3.10	3.38	0.13
Slope	-0.25	35.31	-33.76	4.13	12.96	0.25
Curvature	-0.09	8.41	-6.67	1.45	6.93	0.41

Table 2.6 **Summary of Yield Curve Measure First Differences-2008 to 2018**

Yield Curve Measure	Mean (bps)	Max (bps)	Min (bps)	Std Dev (bps)	Kurtosis (bps)	Skewness (bps)
3-Month Yield	-0.08	69.28	-50.99	3.71	115.01	2.51
30-Year Yield	-0.07	19.50	-14.41	3.61	3.86	0.03
Slope	0.01	58.55	-67.89	4.87	35.56	-0.51
Curvature	-0.01	11.62	-11.95	1.79	7.47	0.13

The average change for all the yield curve measures in both the full period and sub-sample periods is very small. This is reasonable as we do not expect the average daily changes of the yield to be large. There is a significant temporary spike in the spot rates right around the time of the financial crisis in 2007-2008. The standard deviations for the various measures do not vary much with respect for either period. The kurtosis are very high with large portions of the observations close to the mean. This is not usual with observations which are normally distributed. The results skewness and kurtosis suggest that the daily changes in the yield curves are not normally distributed.

The returns on a zero coupon bond are driven by price changes as there are no interest payments. Daily price changes are driven by changes in the yield. Short term returns for the zero coupon bond are not normally distributed as the daily changes in the yield curve are not normally distributed. As discussed in Bolder et al. [2004], this characteristic of the short term returns is of considerable importance as many derivative pricing algorithms and risk management models assume that returns are normally distributed. Since this assumption does not appear to hold in our results, any model using our data that makes the assumption of normality could be generating inaccurate prices or risk measures as indicated by Bolder et al. [2004].

Figure 2.3 **Yield Curve Measures-Full Sample**

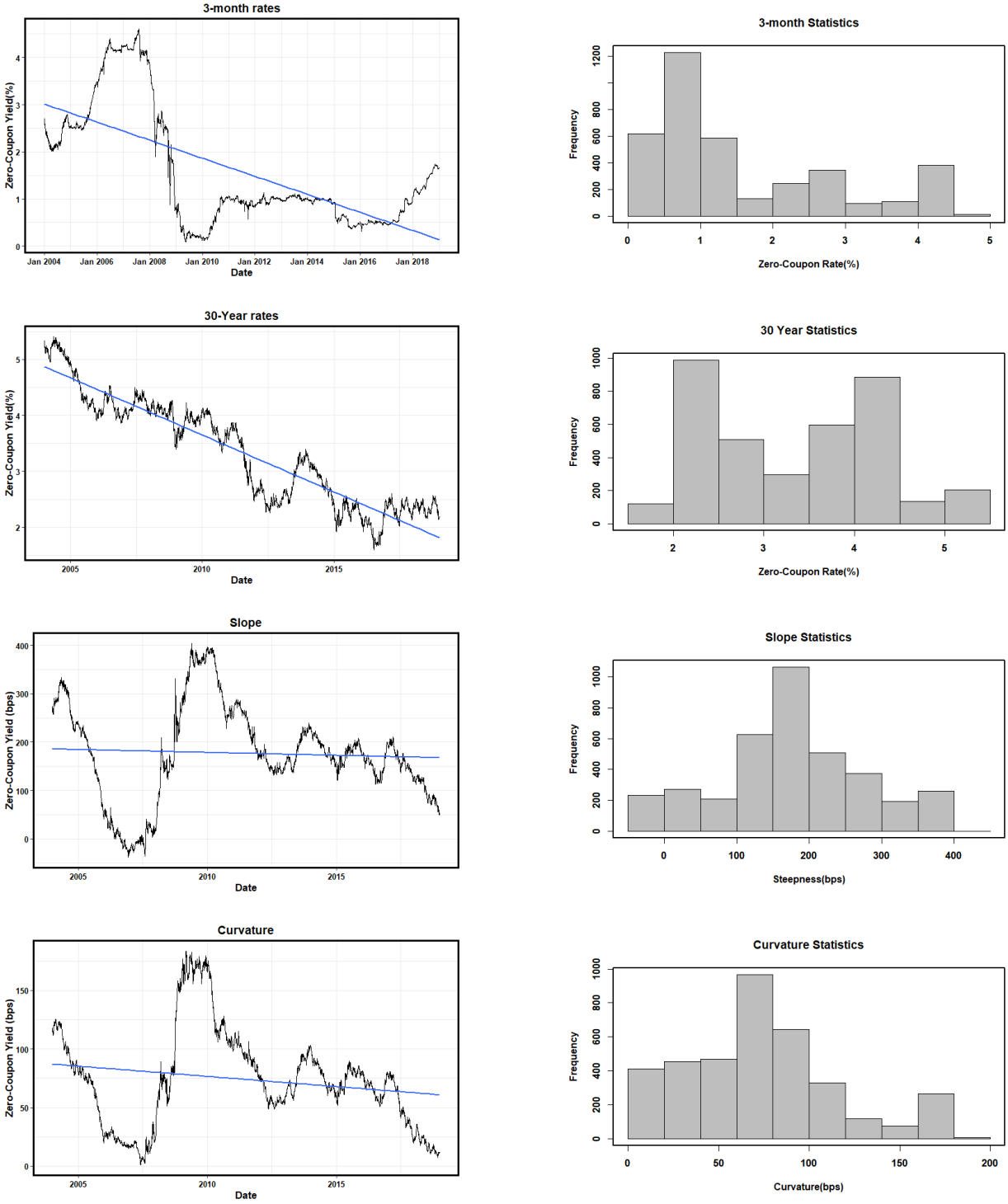
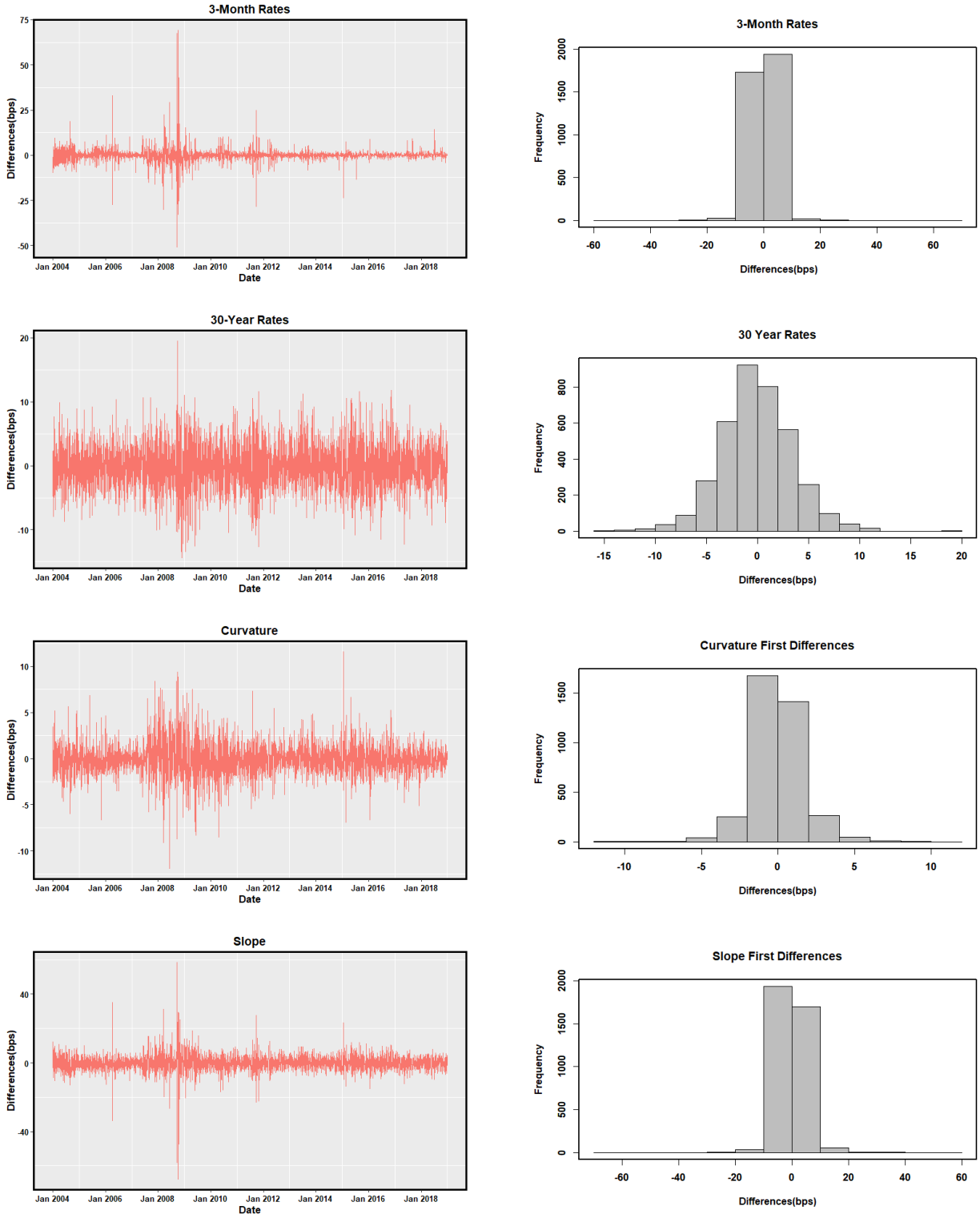


Figure 2.4 **Yield Curve Measure First Differences -Full Sample**



Chapter 3

Numerical Results

This chapter presents the main results obtained in this work. The first section presents the results obtained from the principal component analysis, and the second does same for the results from the discrete time Vasicek model using a Kalman filter. The last section compares the two methods.

3.1 Modelling Data With The PCA Method

In this section we perform a principal component analysis of the zero coupon Government of Canada bonds using the period 2004 to 2018. Only data from the third day of each month for the period indicated was used as explained in Chapter 2.

The principal components, which are the eigenvectors of the sample covariance matrix from the centered data, as explained in (1.46) were computed. Table 3.1 shows the percentage of variance (PVE) explained from the first three components which was computed using the formula in (1.48).

We find that the first factor is by far the most important factor as determined by Litterman and Scheinkman [1991]. This is true for both sample periods, as well as the total period. The first factor explains 89.8% of the total variability for the full data. Together, the first three (uncorrelated) factors explain almost all the correlation between the zero coupon

Table 3.1 **Percentage of Variance Explained (PVE)**

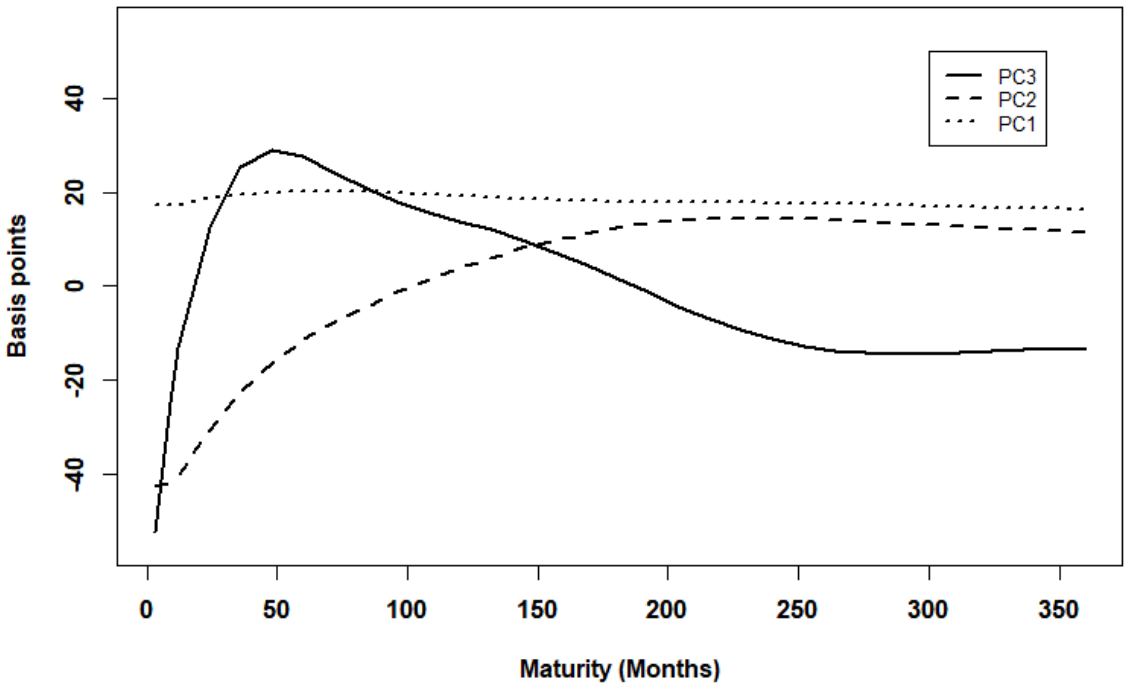
Period	PVE for Comp. 1	PVE for Comp. 2	PVE for Comp. 3	Total
Sub Sample 1	0.830	0.163	0.007	1.000
Sub Sample 2	0.895	0.094	0.011	1.000
Full Sample	0.898	0.097	0.005	1.000

rates. There were 30 bonds used in the sample and therefore 30 predictors for the principal components analysis and consequently, 30 principal components. From the results above we see that the remaining 27 components explain little or none of the variation in the zero coupon bonds. The three factors explain almost all the variation in the data. These observations are consistent with what was obtained by Bolder et al. [2004], which our work follows closely.

We proceed to assess how these factors influence the shape of the yield curve. Figure 3.1 shows the plot of the eigenvectors corresponding to the first three factors. The plot represents the common factors affecting the shape of the yield curve.

The yield for the first factor is constant across all maturities and it drives the overall level of the yield curve. Therefore, a variation in the first principal component score essentially represents a parallel shift of the yield curve. Litterman and Scheinkman [1991] named it the level factor. The second factor, which has been called the steepness factor causes the short end of the yield curve to fall and the long end of the yield curve to rise, which creates the steepness in the yield curve. The third factor also causes the short end of the yield curve to fall but has the opposite effect on the long end. The third factor causes a significant curvature between the 3 months bond and 100 months bond. It increases the curvature of the yield curve across maturities. It is called the curvature factor. These three factors are

Figure 3.1 **Plot of Zero Coupon Bonds as Against the First Three Components**



sufficient to explain the variation of the yield curve.

The figures below represents the estimated yield curve from the factors as compared to the actual yield curve for the selected days. We perform this for the first factor, the first three factors and the first five factors.

Figure 3.2 **Actual Versus PCA First Principal Component**

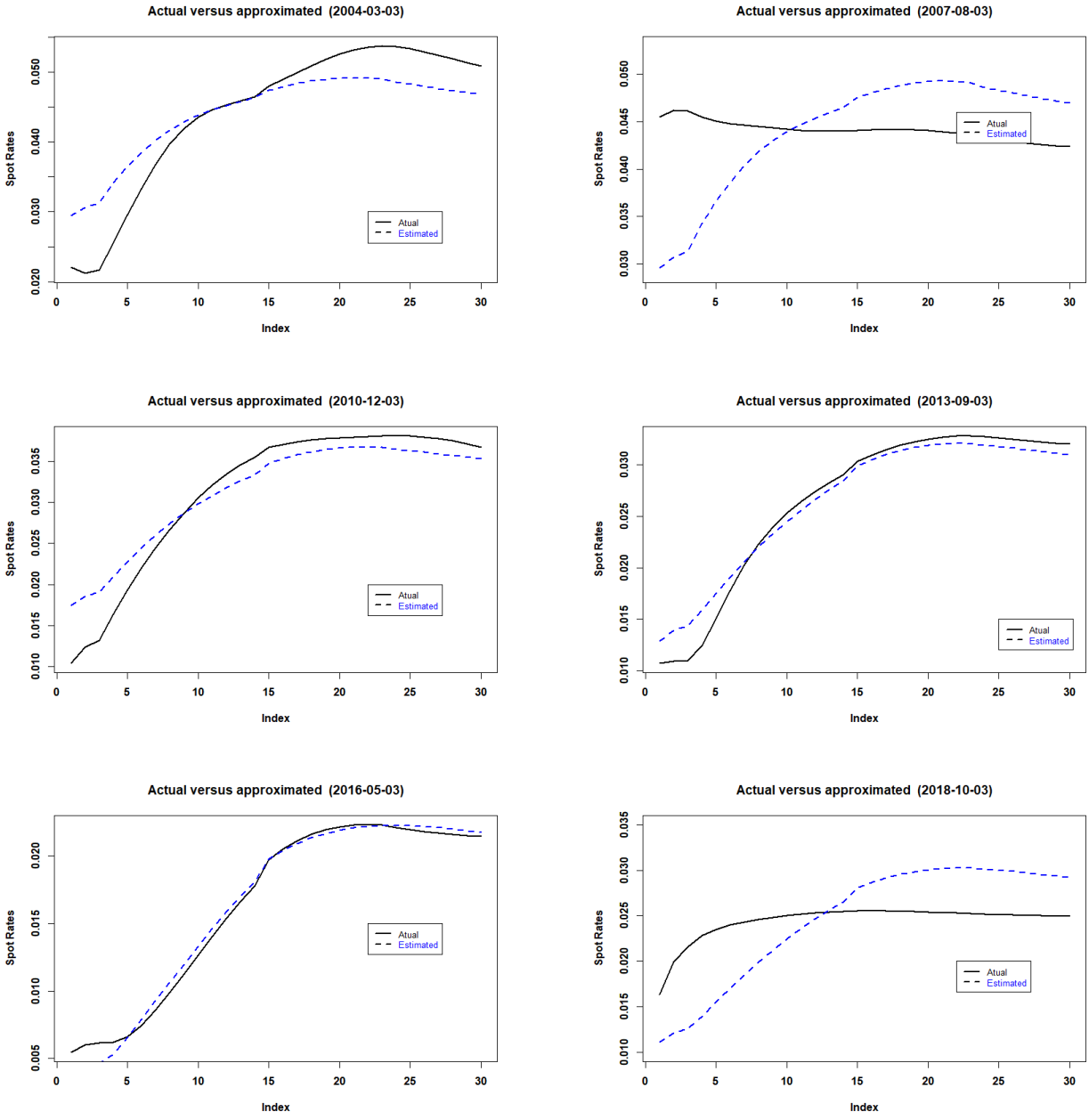


Figure 3.3 **Actual Versus PCA First Three Components**

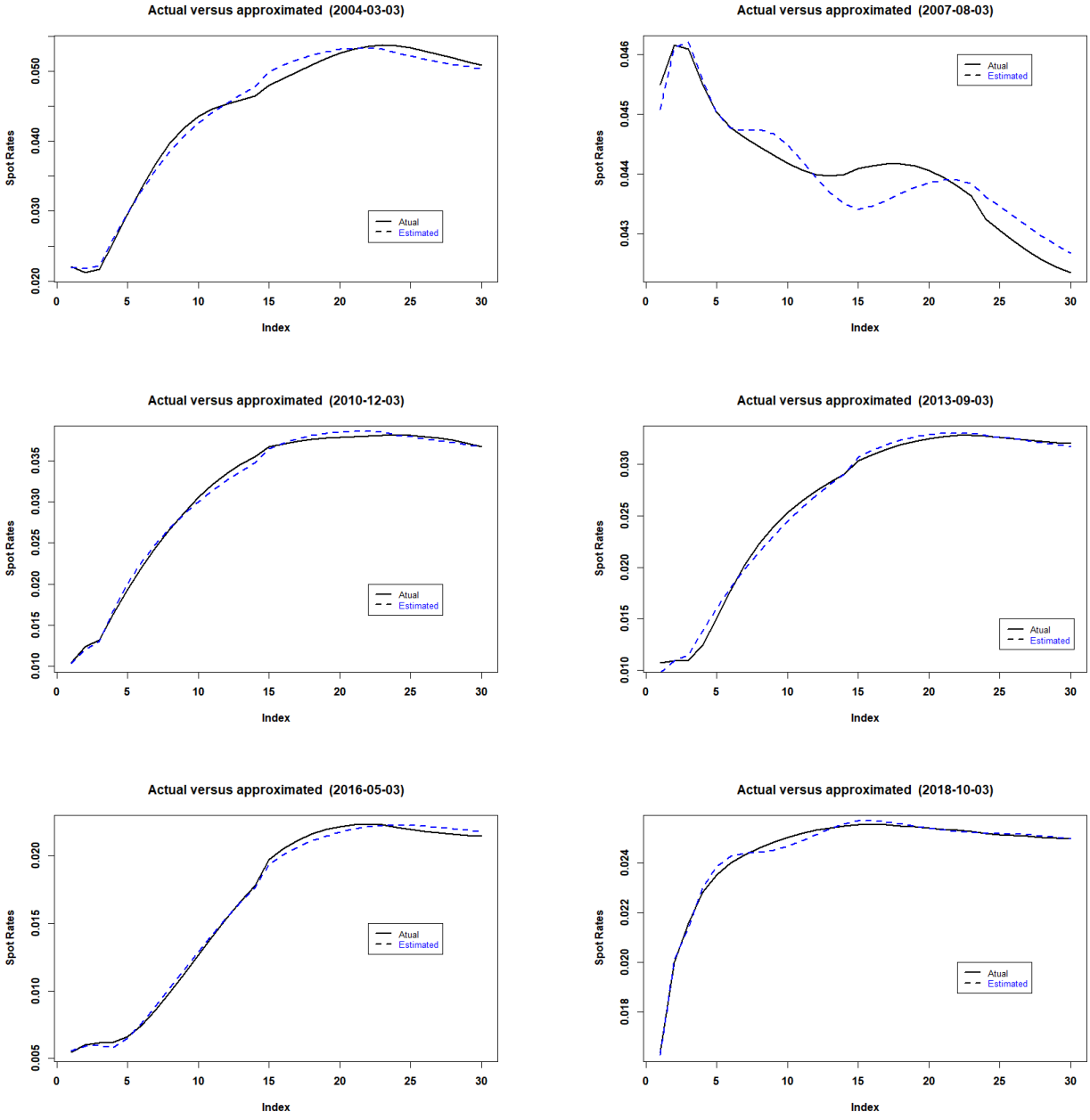
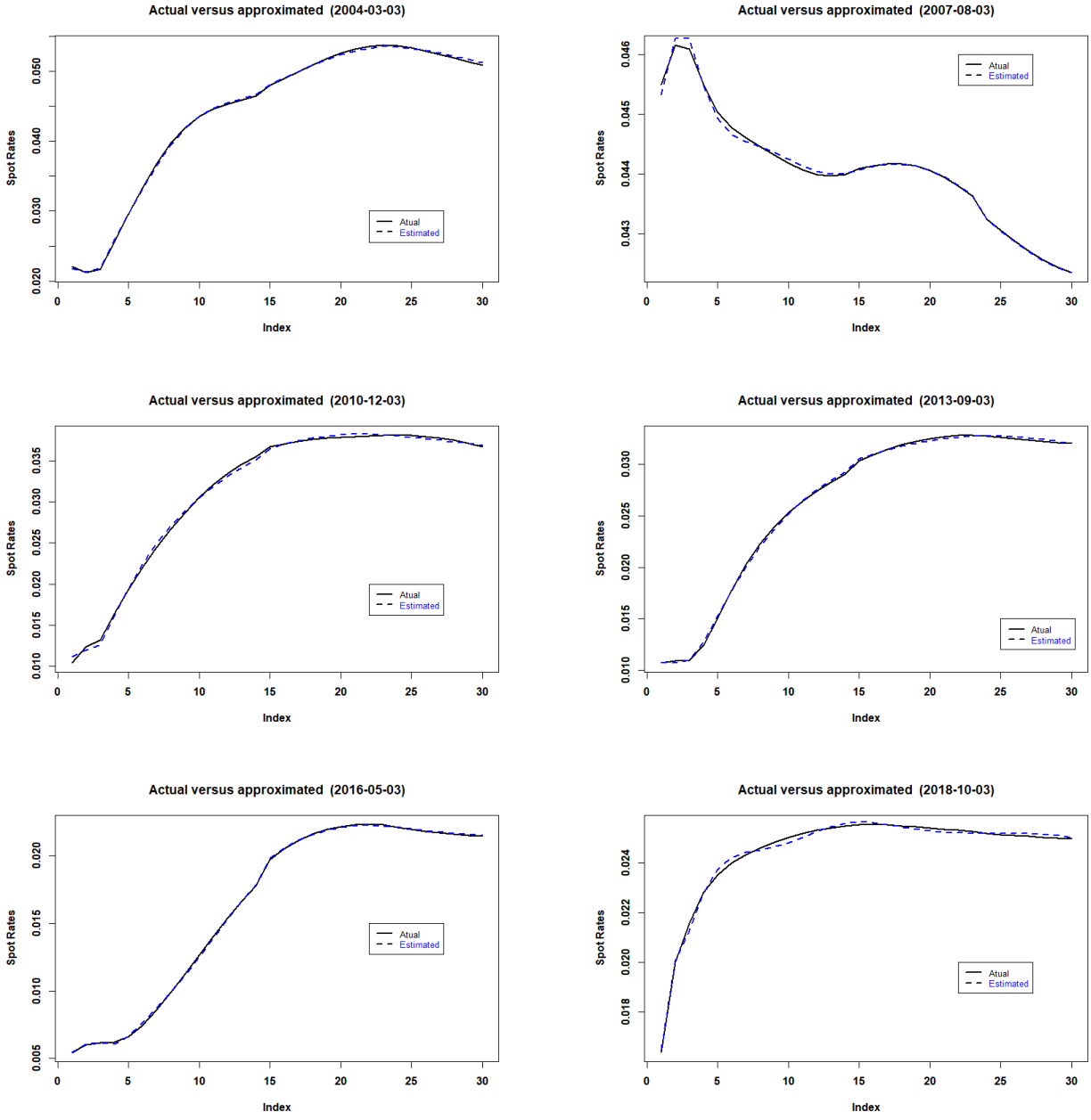


Figure 3.4 **Actual Versus PCA First Five Components**



By looking at Figure 3.2, we observe that the yield curves derived from the first factor only does not fit the observed data well as those derived from all three factors. The former is not able to capture the curvature in the yield curves. Of the selected days the estimated curve is an especially poor fit for 3/8/2007 and 3/10/2018. The curves are however quite good fits for two of the days selected. In all cases the curve fits the points in the middle (i.e.

the medium-term bonds) much better compared to the the short-term and long-term bonds.

There is a significant improvement in the fit of the curves for all days when all three factors are used. This is especially true for the days that were not fit well by the one-factor curves. We also fit the curve using the first five components. We observed little difference in the fit of the five-factor curves and the three-factor curves.

Table 3.2 **RMSE for The Selected Days (Bps)**

	2004-03-03	**2007-08-03**	**2010-12-03**	**2013-09-03**	**2016-05-03**	**2018-10-03**
PC1	0.463	0.675	0.269	0.138	0.071	0.500
PC3	0.099	0.035	0.048	0.053	0.031	0.016
PC5	0.020	0.006	0.032	0.017	0.008	0.013

Table 3.2 also shows the root mean square errors between the observed and the estimated spot rates using the principal components. The errors are negligible for the three factor and five factor model. The error decreases with an increasing number of factors. The root mean square errors for the five factor model for the 03-12-2010 and the 03-10-2018 are very close to the three factor model. The minimum number of factors required to capture the shape of the yield curve is three as was identified in the previous working papers—representing the level, steepness and curvature.

3.2 Modelling Data With The Vasicek Model

In this section the Kalman filter algorithm described in Section (1.1.3) is applied to the Canadian zero-coupon data for the period 1991 to 2018 using spot-rate data for only the third day of the month. We obtain our interest rates using real world data as discussed earlier. Tables 3.3 to 3.5 present estimates for the one factor, two factor and, three factor models respectively. The standard deviations for these parameters are presented in brackets

in the tables.

κ, the mean reversion speed, characterizes the mean half-life for the model. The mean half-life represents the expected time for the process to return halfway to its long term mean, and is computed as half life = $ln(2)/\kappa$. The mean half-life for the one factor model is 1732.87yrs. For the two factor model it is 990.2yrs for the first state variable and 34.5yrs for the second state variable. The mean half-life for the second state variable is lower than that of the first. For the three factor model the mean half-lives are 1386.3yrs, 33yrs and 10yrs for the first, second and third state variables respectively. The mean half-life decreases with increasing number of factors. The one factor model has the biggest mean half-life and this indicates weaker mean reversion. According to Chen and Scott [1993], mean reversion is useful for capturing underlying trends. Slow mean reversion in particular is useful for predicting long-term trends such as long-term bond prices. These trends tend to be sensitive to the κ and λ parameters.

Table 3.3 **Estimates From The One Factor Model**

Parameter	**Estimate**	
κ	0.0004 $(2.97e^{-}5)$	Log L= 29776.8
θ	0.8027 $(4.66e^{-}5)$	BIC= -59511.42
σ	0.0014 $(9.7e^{-}5)$	
λ	-0.0001 $(2.04e^{-}5)$	

To assess which of the models fits the term structure best, we consider the log likelihood

Table 3.4 **Estimates From The Two Factor Model**

Parameter	**State Variable 1**	**State Variable 2**	
κ	0.0007	0.0201	Log L= 33741.2
	$(4.201e^{-}4)$	$(4.5e^{-}5)$	
θ	0.6031	0.1002	BIC= -67384.48
	$(1.807e^{-}1)$	$(1.81e^{-}1)$	
σ	0.0045	0.0016	
	$(1.19e^{-}4)$	$(2.72e^{-}5)$	
λ	-0.0001	-0.0001	
	$(4.85e^{-}2)$	$(2.45e^{-}2)$	

and the BIC. The log likelihoods for the one-factor model, two-factor model and three-factor model are 29776.8, 33741.2 and 35074.59 respectively. The log likelihood increases when increasing number of factors with that of the three factor being the highest. The BICs for the one-factor, two-factor and three-factor models are -59511.42, -67384.48 and -70017.19 respectively. The BIC decreases by 12% from the one factor model to the two factor model and by 3.8% from the two factor to the three factor model. The results from the log likelihood and the BIC indicate that the three-factor model provides the best fit for the term structure. The standard deviations of the measurement errors shown in Appendix A also show that that the results from the three-factor model are more precise than those from the one-factor and two-factor model. The measurement errors decrease with increasing number of factors for most of the bonds.

Using the estimated parameters we computed the estimated model spot rates and compare these to the actual (observed) spot rates. The unobservable factors used for the estimation were obtained from the filtered estimates based on information available at time t. Seven days were selected out of the sample. Table 4.6 shows the root mean square errors

Table 3.5 **Estimates From The Three Factor Model**

Parameter	**State Variable 1**	**State Variable 2**	**State Variable 3**	
κ	0.0005	0.0209	0.0649	Log L= 35074.59
θ	0.0649	0.1697	0.7280	BIC=-70017.19
σ	0.00152	0.00706	0.00786	
λ	-0.0001	-0.0001	0.1315	

for the one-factor, two-factor and three-factor models for the days selected and Figures 3.6 to 3.12 show the plots of the actual spot rates compared to the estimated spot rates. In Table 4.6 we observe that the sum of squared errors decreases with an increasing number of parameters for all the days selected. From Figures 3.6 to 3.12 we observe that the one-factor model performs poorly for all the days selected. It fits the data poorly at the extremes of the yield curve and only passes through a few points in the middle of the curve. The two-factor and three-factor models are very close in fit for most of the days except for 2018-10-03 and 2002-06-03 when the three-factor model outperforms the two-factor model. The three-factor model in that case is able to capture the curvature of the yield curve and this is consistent with the results obtained by Litterman and Scheinkman [1991]–i.e. that the first factor captures the level of the curve, the second factor the steepness, and the third factor the curvature. The average spot rates for each bond for the one-factor, two-factor and three-factor models are shown in Figure 3.5. In this plot we observe that the two-factor and three-factor models perform similarly with both fitting the short end of the curve much better than the long end.

Table 3.6 **RMSE for The Selected Days (Bps)**

	1997-03-03	2002-06-03	2004-03-03	2007-08-03	2013-09-03	2016-05-03	2018-10-03
One Factor	0.884	0.403	0.527	0.799	0.174	0.133	0.598
Two Factor	0.245	0.392	0.121	0.047	0.087	0.272	0.212
Three Factor	0.059	0.055	0.139	0.035	0.092	0.094	0.073

Figure 3.5 **Average spot rates for each bond**

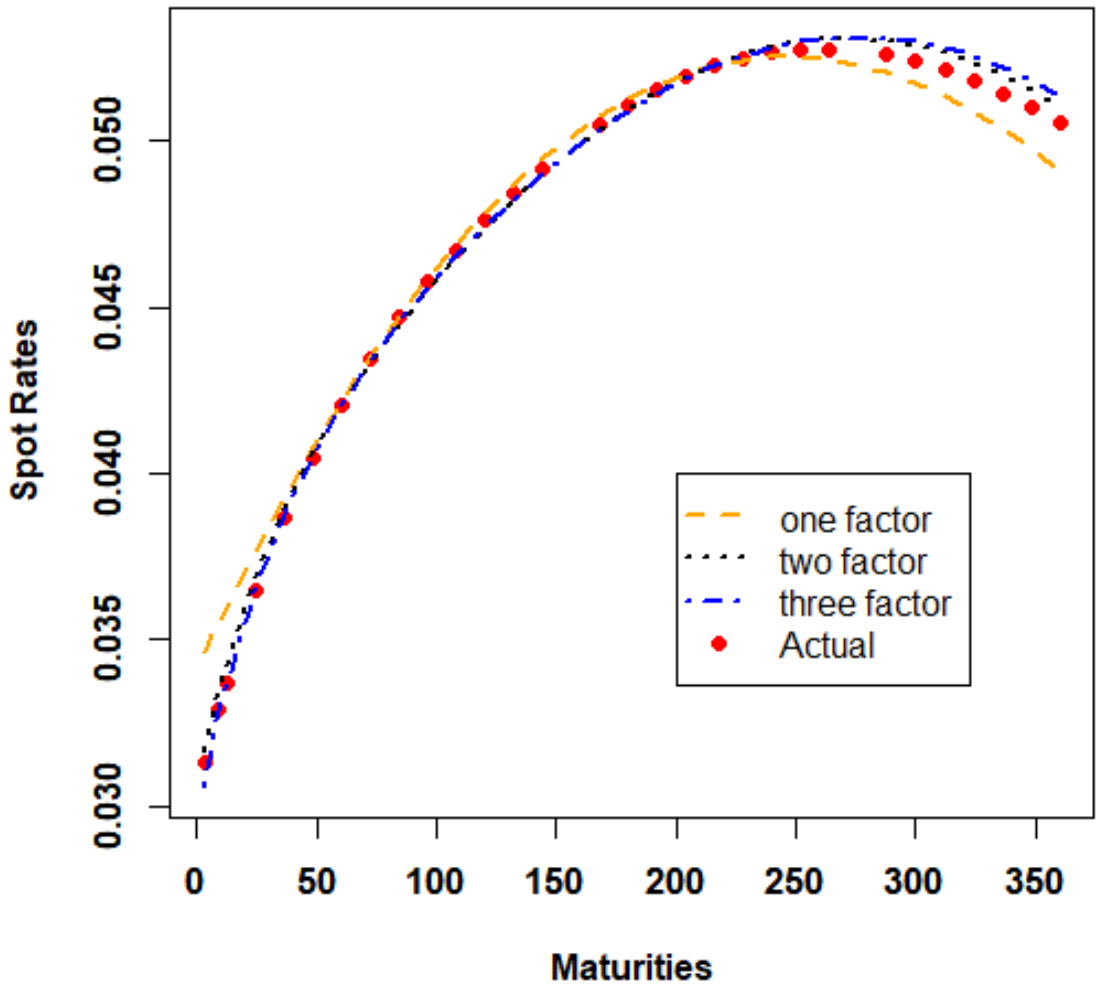


Figure 3.6 **Actual Yield Curve Versus Vasicek(1997-03-03)**

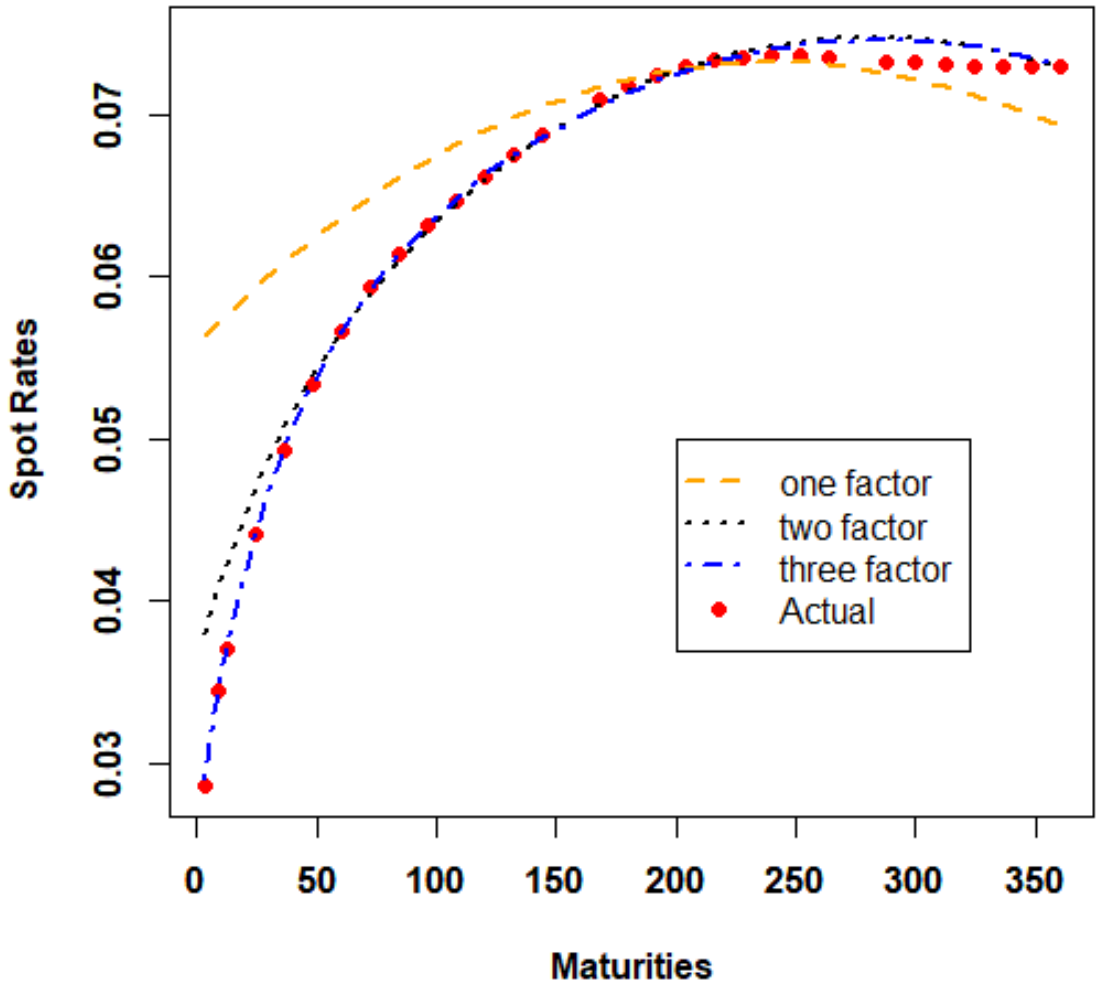


Figure 3.7 **Actual Yield Curve Versus Vasicek (2002-06-03)**

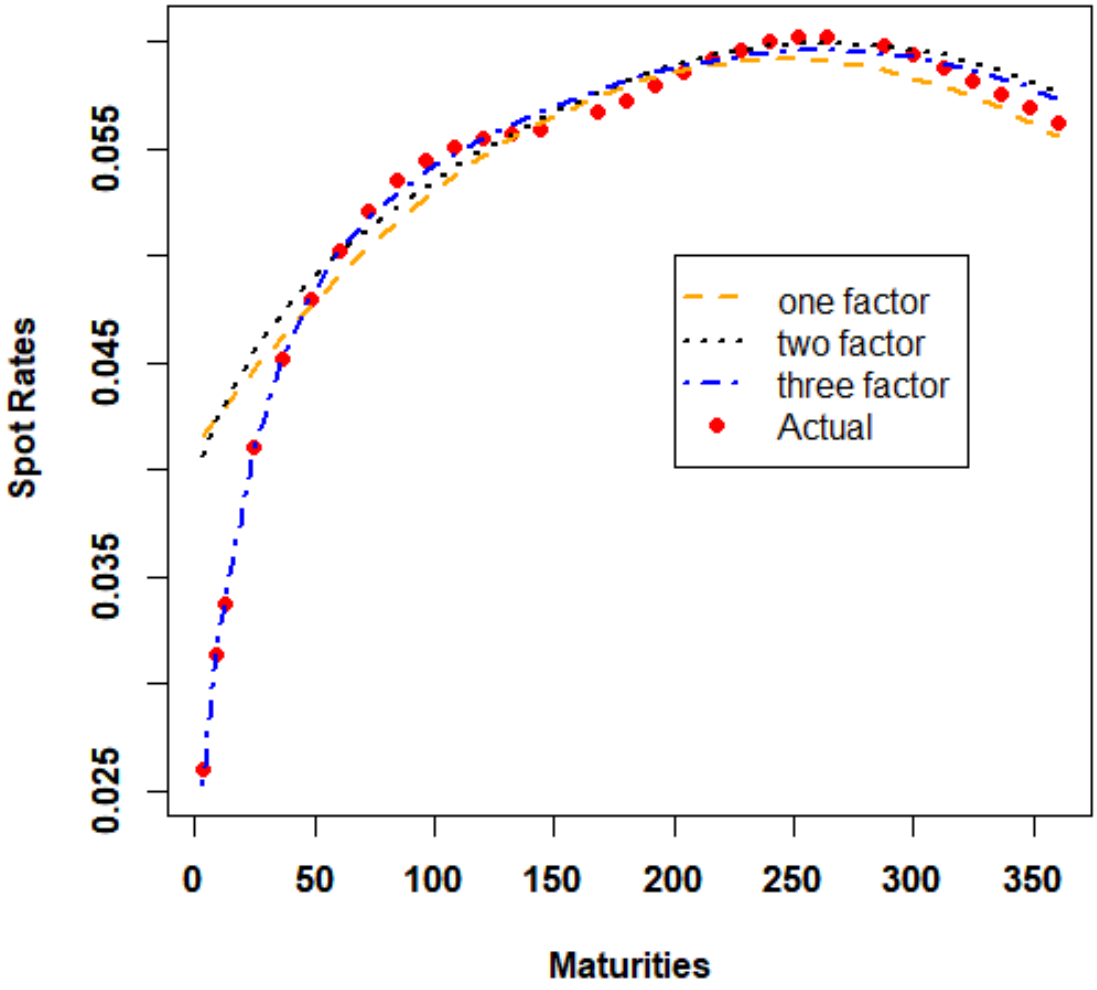


Figure 3.8 **Actual Yield Curve Versus Vasicek (2004-03-03)**

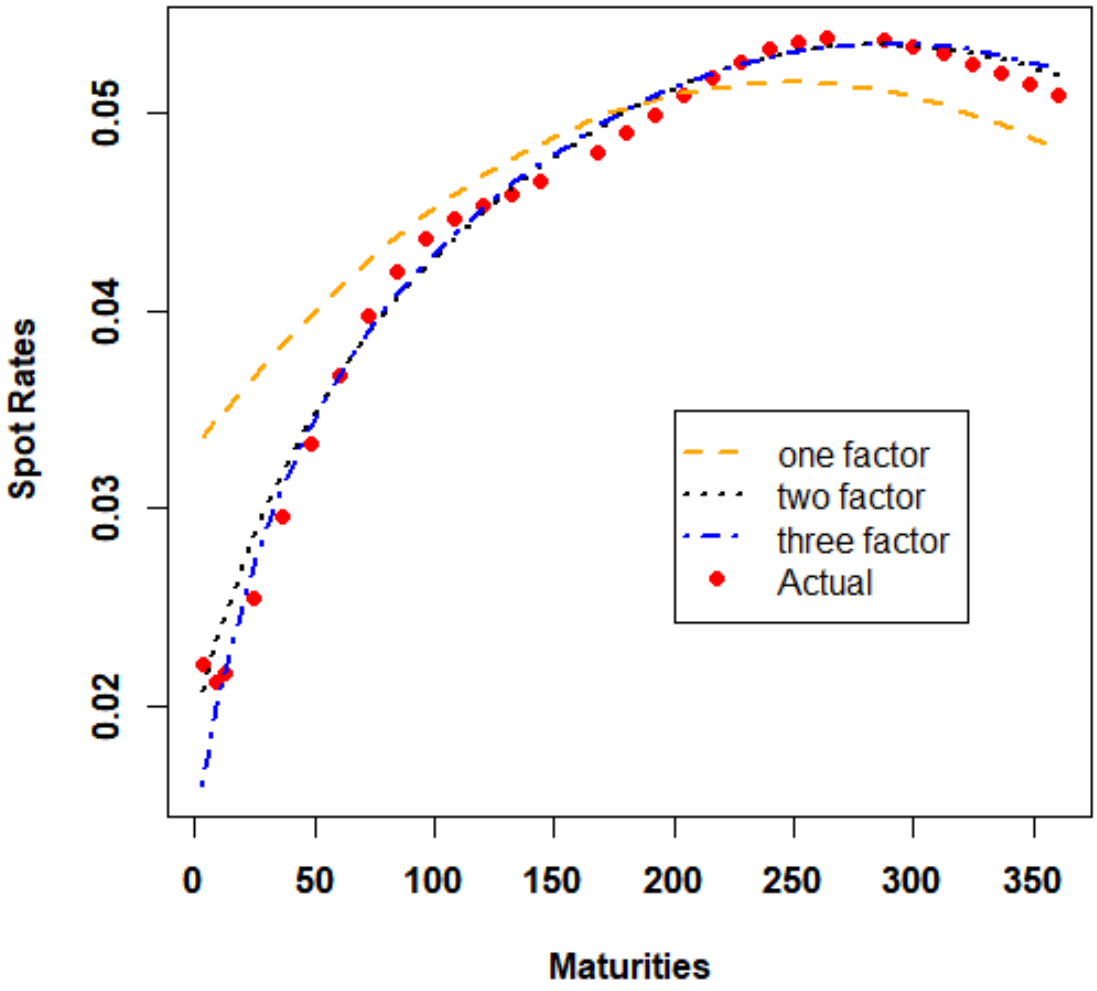


Figure 3.9 **Actual Yield Curve Versus Vasicek (2007-08-03)**

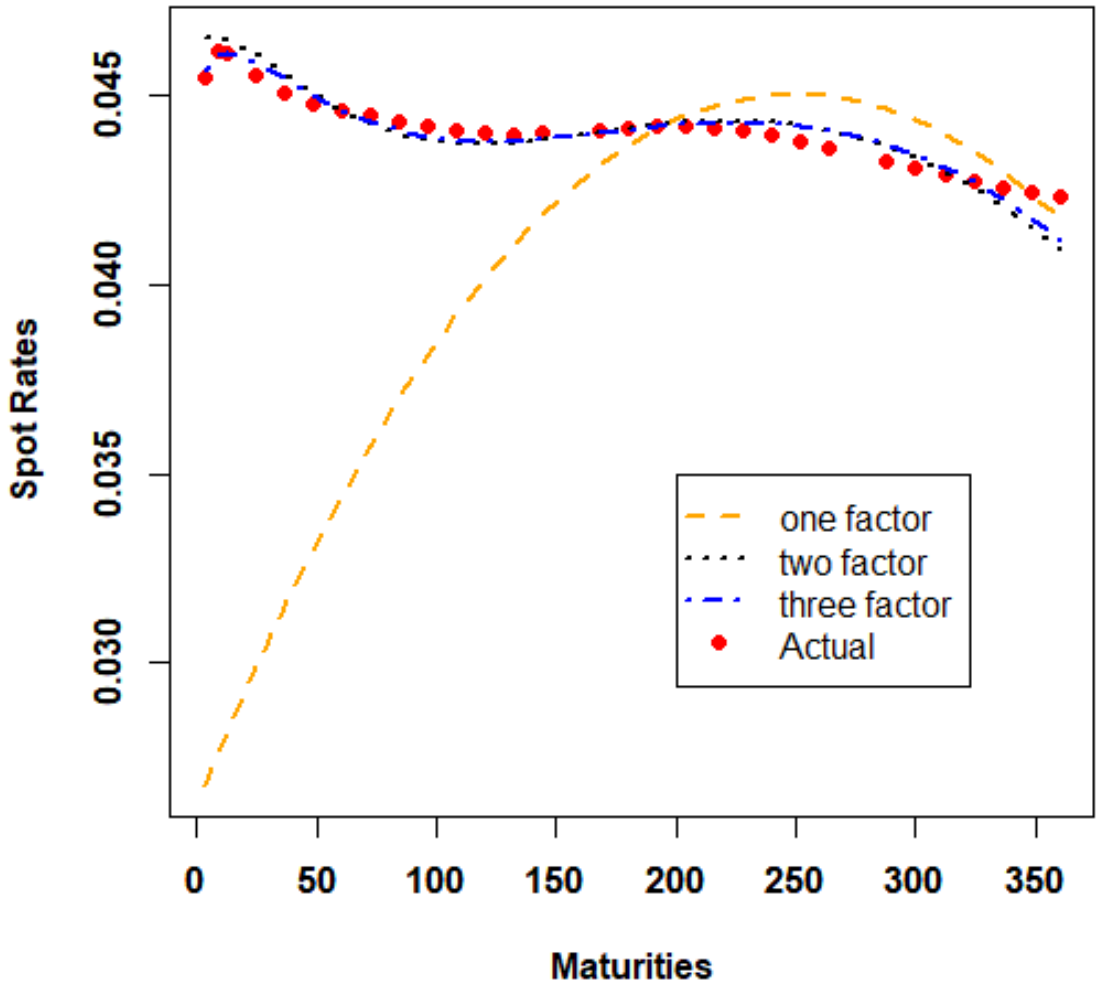


Figure 3.10 **Actual Yield Curve Versus Vasicek (2013-09-03)**

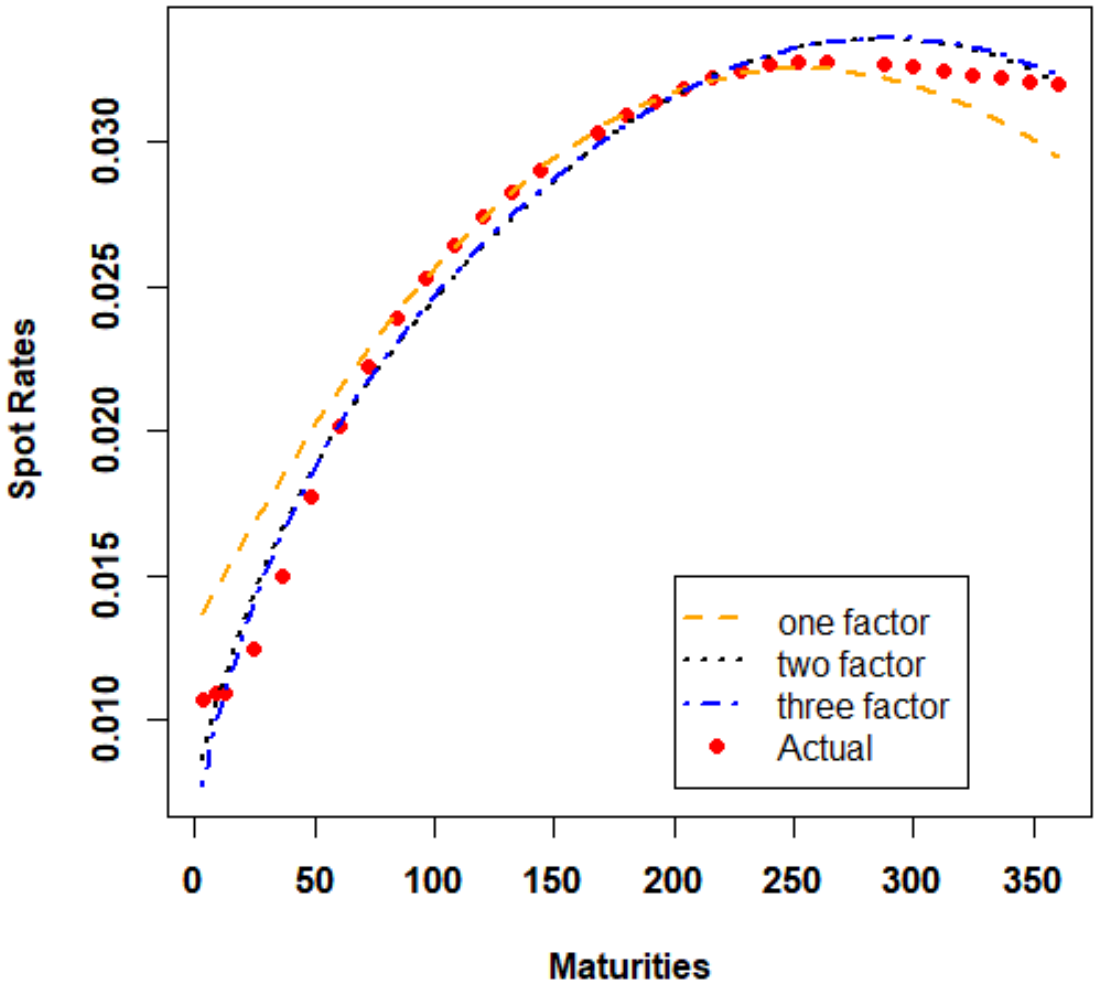


Figure 3.11 **Actual Yield Curve Versus Vasicek (2016-05-03)**

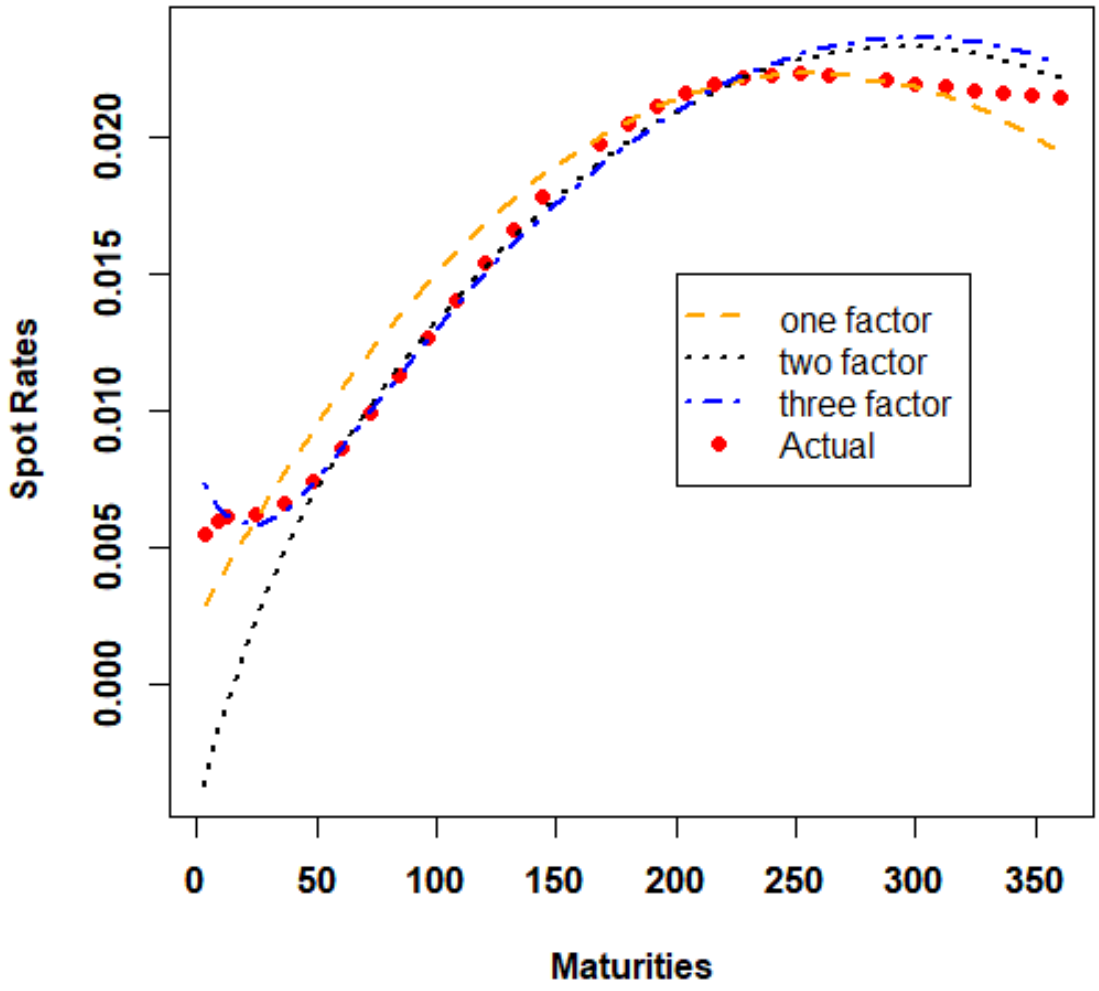
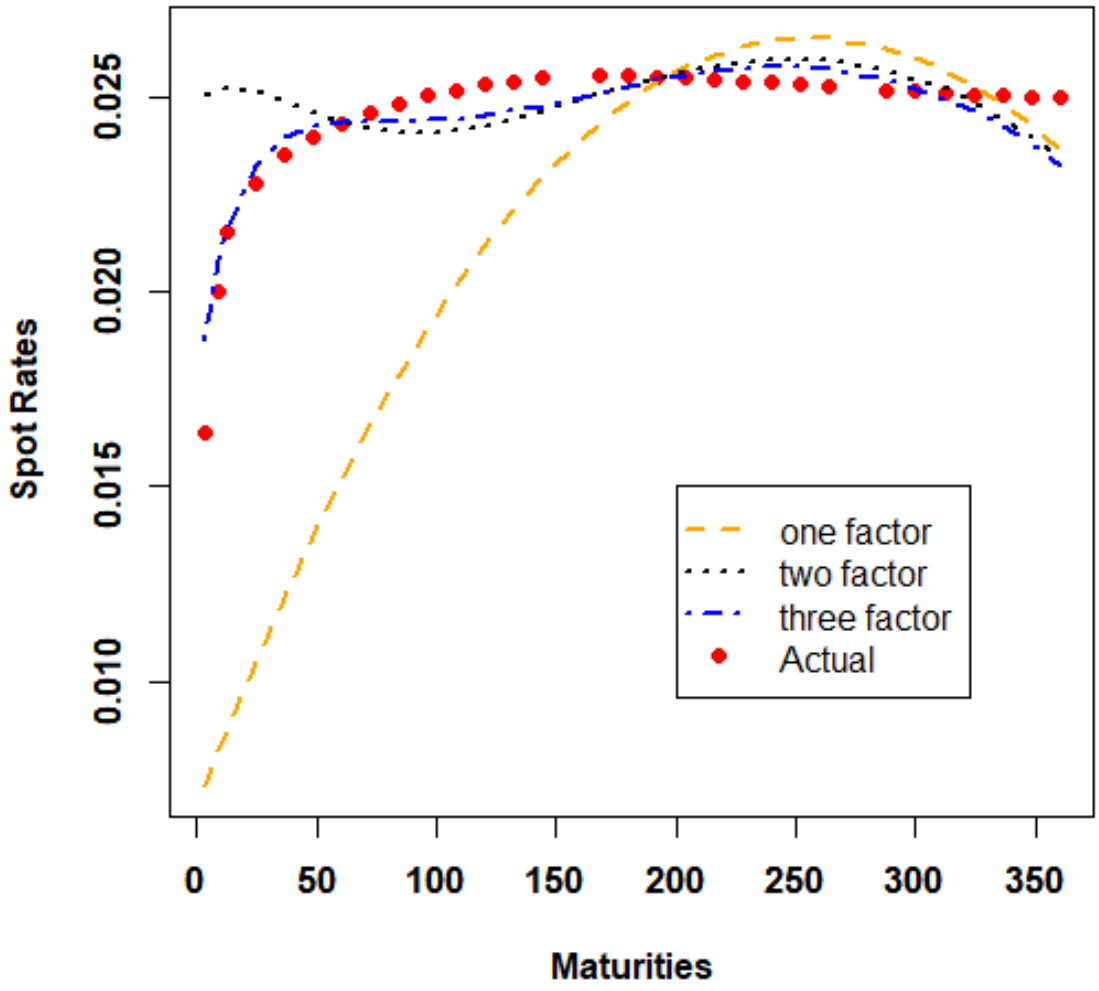


Figure 3.12 **Actual Yield Curve Versus Vasicek (2018-10-03)**



In Figures 3.13 and 3.14 we plotted the estimated time series for the one-factor, two-factor and three-factor models over the sample period and compared them to the actual rates for the 3-month spot rate representing the short end and the 30-year rate representing the long end. The fit of the estimated curve to the data gets better with increasing number of factors for both the short end rate and the long end rate. The three-factor model outperforms the two-factor model and the one-factor model for the short end of the yield curve.

In all the cases presented above the three factor model fits the general shape of the yield as well as the slope. This indicates that increasing the number of factors increases the ability of the model to capture the shape of the yield curve. Also three factors are sufficient to capture the level, the slope and curvature of the yield curve.

Figure 3.13 **Observed And Estimated -3-Month Bond**

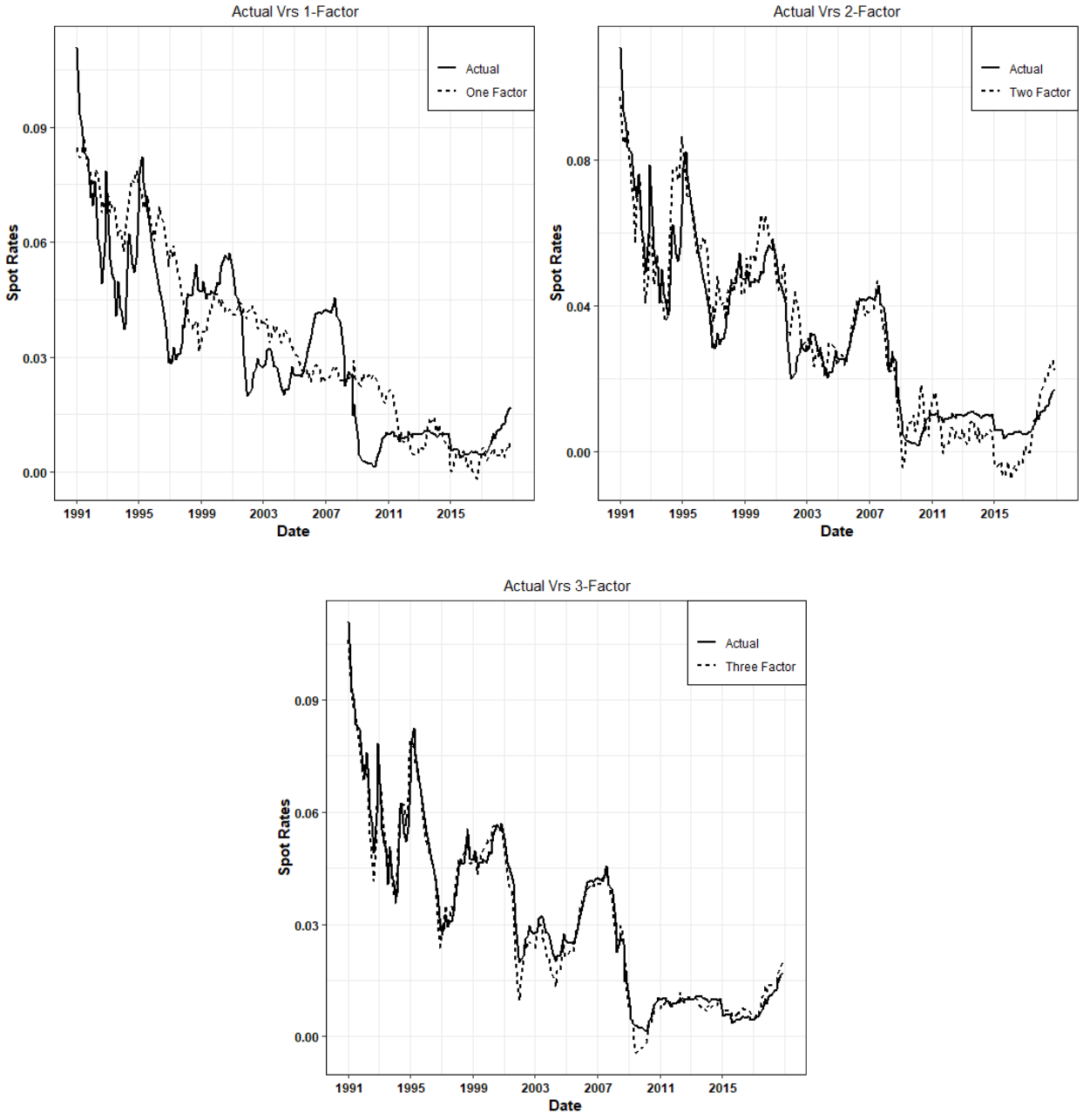
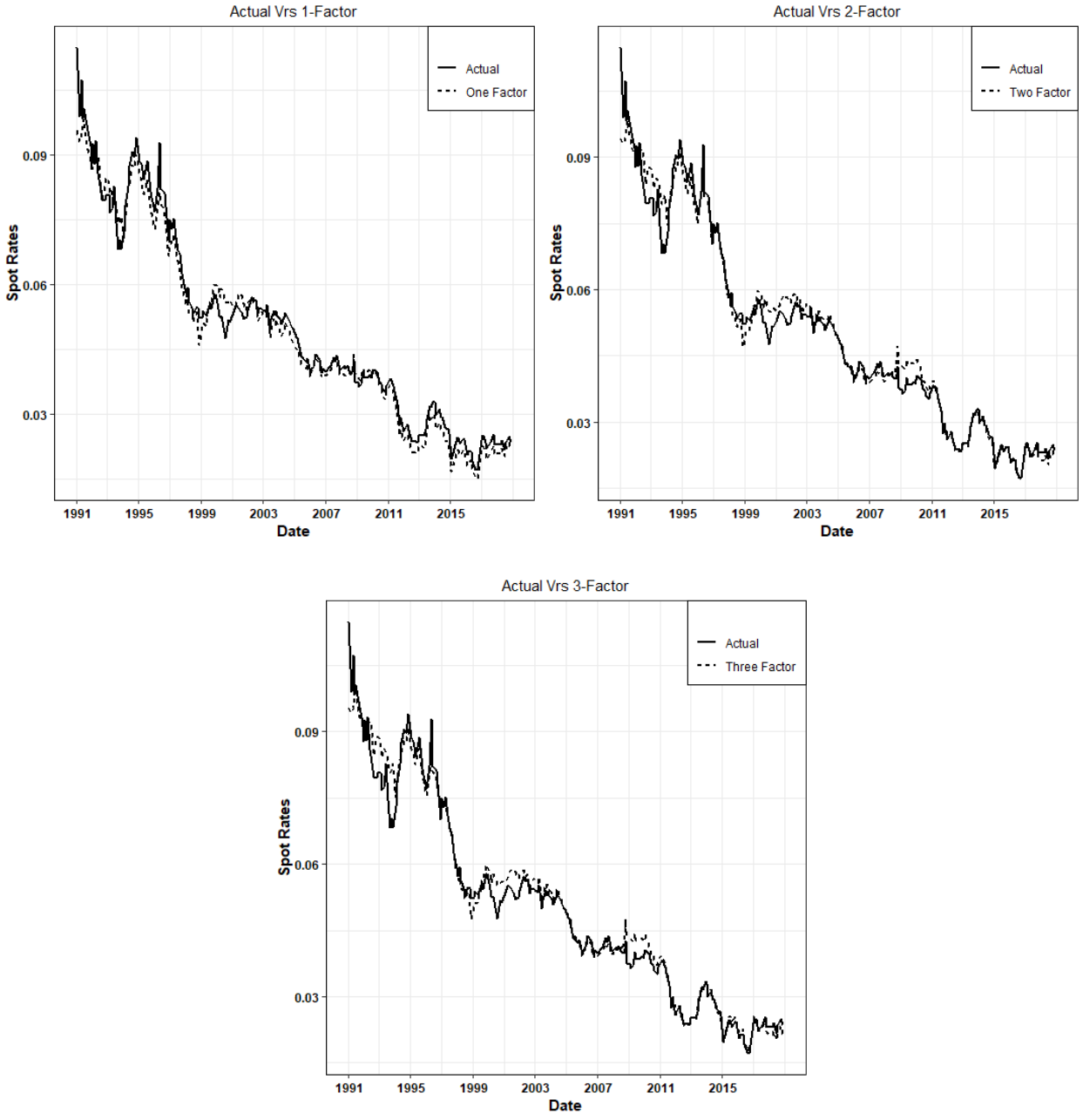


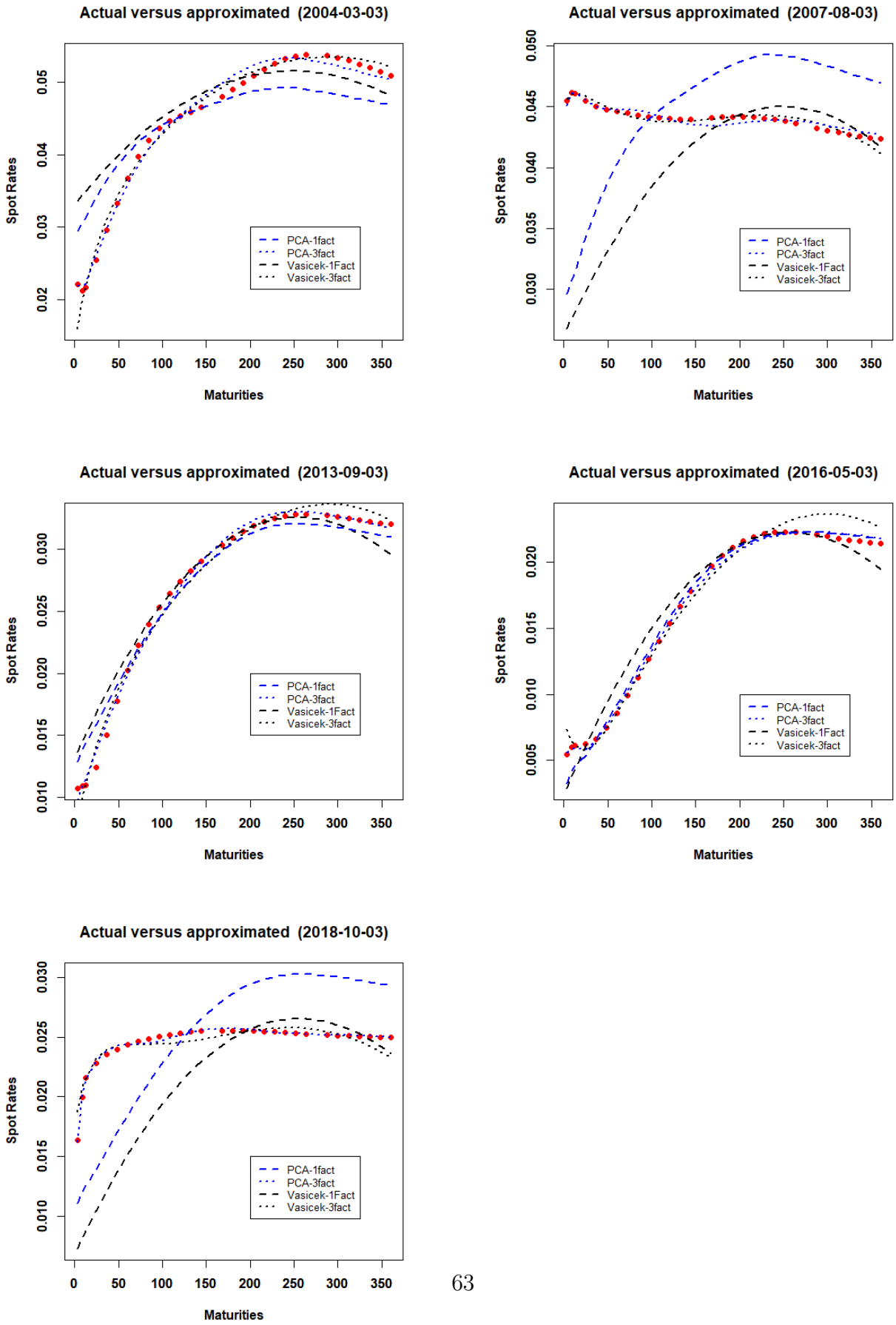
Figure 3.14 **Observed And Estimated -Longest Term Bond(360 months)**



3.3 Comparison of the Vasicek Model to the PCA Model

Section 3.1 presented results obtained from the PCA model and Section 3.2 results from the Vasicek model. In both models more than one factor was used in fitting the term structure of the yield curve to assess the number of factors necessary in characterising the yield curve. Data from the same days were used to fit both the PCA and the Vasicek model when conducting the analysis. With both models we observe that increasing the number of factors improved the fit of the estimated yield curve to the observed spot-rate data. In Figure 3.15 we present the term structure plot of the one-factor and three-factor Vasicek model, and the PCA with one and three principal components on the same graph for the days selected. Comparing the one factor model to the estimated spot rate for the first principal component in the PCA model we observe that the both models perform very poorly for two of the selected days—03/08/2007 and 03/10/2018. However the PCA model performs better than the Vasicek model for the remaining days. For the three factor Vasicek model and and the PCA with three principal components, we observe that both models outperforms the one factor Vasicek model or the single principal component. For the days that the one factor or the single principal component performed poorly, there was significant improvement with the three-factor model and the PCA model with three principal components. Also from the RMSE in Tables 3.2 and 3.6 we observe that the errors for the days shown in the PCA model are lower than that of the Vasicek model. The PCA model performs better than the Vasicek model. The PCA model is able to capture the short end of the yield curve better than the Vasicek model for the 03-03-2004, 03-05-2016 and 03-10-2018. The plot for the first three components for the PCA model outperforms the Vasicek model.

Figure 3.15 **PCA versus Vasicek**



Conclusion

The main objective of this thesis was to compare the multifactor Vasicek and PCA approaches to estimating the yield curve. The affine term structure model under the discrete time Vasicek model was discussed and developed in Chapter 1 together with the methodology to be used in the parameter estimation, i.e. the Kalman filter. This methodology was mainly selected because it uses all the present and past information to determine the current estimate and the yield curve and does not require that the state variable is observable. A machine learning technique, principal component analysis (PCA) was also used in this thesis to model the term structure.

These models were fit to the Canadian zero-coupon data. A descriptive analytics of this data was performed between two sub sample periods that is before the financial crisis and after the financial crisis. The numerical results section presents the results from the implementation of the Kalman filter algorithm on the Canadian zero-coupon and that of the PCA on the same data. It then compares the performance of the two in fitting the yield curve. The parameters for the one-factor, two-factor and three-factor discrete time Vasicek models were estimated using the Kalman filter algorithm and this was used to estimate the spot rates. A principal component analysis was also performed and the first principal component, first three principal component and the first five principal component were used to fit the term structure of interest rates. In both cases it was observed that more than one factor is required to account for the changes over time in the slope and shape of the yield curve however three factors were found to be sufficient to capture the curvature of the yield

curve. The first factor is responsible for the level of the yield curve and the second factor the slope of the yield curve. For all the days selected for fitting of the observed to the estimated data, it was observed that the third factor is able to capture the curvature of the yield curve with the second factor performing as well as the third for some days.

The results from the work performed above suggest that the principal component analysis performs better than the Vasicek model. In conclusion the machine learning method chosen outperforms the traditional econometric approach in fitting the yield curve but however does not give the estimates of the parameters driving the interest model.

Having applied this concept with the first model developed to describe the short rate process future research can be done with models that do not assume negative interest rates so this can be useful in applied work.

Bibliography

Simon H Babbs and K Ben Nowman. An application of generalized Vasicek term structure models to the UK gilt-edged market: a Kalman filtering analysis. *Applied Financial Economics*, 8(6):637–644, 1998.

Bank of Canada. Yield curves for zero-coupon bonds, 2018. URL `https://www.bankofcanada.ca/rates/interest-rates/bond-yield-curves/`.

Tomas Björk. *Arbitrage theory in continuous time*. Oxford university press, 2009.

David Jamieson Bolder. Affine term-structure models: Theory and implementation. *Available at SSRN 1082826*, 2001.

David Jamieson Bolder and Scott Gusba. Exponentials, polynomials, and Fourier series: More yield curve modelling at the Bank of Canada. *Available at SSRN 1082835*, 2002.

David Jamieson Bolder, Grahame Johnson, and Adam Metzler. An empirical analysis of the Canadian term structure of zero-coupon interest rates. *Available at SSRN 1082833*, 2004.

Alan Brace, Dariusz Gątarek, and Marek Musiela. The market model of interest rate dynamics. *Mathematical Finance*, 7(2):127–155, 1997.

Damiano Brigo and Fabio Mercurio. *Interest rate models-theory and practice: with smile, inflation and credit*. Springer Science & Business Media, 2007.

Kalok C Chan, G Andrew Karolyi, Francis A Longstaff, and Anthony B Sanders. An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance*, 47(3):1209–1227, 1992.

Ren-Raw Chen and Louis Scott. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *The Journal of Fixed Income*, 3(3): 14–31, 1993.

Ren-Raw Chen and Louis Scott. Multi-factor cox-ingersoll-ross models of the term structure: Estimates and tests from a Kalman filter model. *The Journal of Real Estate Finance and Economics*, 27(2):143–172, 2003.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 50(4):1029–1054, 1982.

Andrew C Harvey. *Forecasting, structural time series models and theKalman filter*. Cambridge university press, 1990.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

Will Kenton. The great recession, 2018. URL `https://www.investopedia.com/terms/g/great-recession.asp`.

B Li, E DeWetering, G Lucas, R Brenner, and A Shapiro. Merrill Lynch exponential spline model. Technical report, Merrill Lynch working paper, 2001.

Robert Litterman and Jose Scheinkman. Common factors affecting bond returns. *Journal of Fixed Income*, 1(1):54–61, 1991.

Kristian R Miltersen, Klaus Sandmann, and Dieter Sondermann. Closed form solutions for term structure derivatives with log-normal interest rates. *The Journal of Finance*, 52(1): 409–430, 1997.

Neil D Pearson and Tong-Sheng Sun. A test of the Cox, Ingersoll, Ross model of the term structure of interest rates using the method of maximum likelihood. Technical report, Working paper, MIT, 1989.

Richard J Rendleman and Brit J Bartter. The pricing of options on debt securities. *Journal of Financial and Quantitative Analysis*, 15(1):11–24, 1980.

Anthony B Sanders and Haluk Unal. On the intertemporal behavior of the short-term rate of interest. *Journal of Financial and Quantitative Analysis*, 23(4):417–423, 1988.

Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.

Martin Vojtek. Calibration of interest rate models-transition market case. *Available at SSRN 843425*, 2004.

Mario V Wüthrich and Michael Merz. *Financial modeling, actuarial valuation and solvency in insurance*. Springer, 2013.

Appendix A

A.1 Standard Deviation of The Measurement Errors

Table A.1

Bond type (Months)	One Factor	Two Factor	Three Factor Factor
3	0.1105	0.8718	0.057
9	0.1058	0.0742	0.026
12	0.103	0.0686	0.010
24	0.092	0.051	0.028
36	0.084	0.037	0.026
48	0.0768	0.024	0.020
60	0.0714	0.010	0.010
72	0.0663	0.020	0.020
84	0.0608	0.026	0.024
96	0.057	0.028	0.024
108	0.052	0.028	0.026
120	0.047	0.028	0.028
132	0.042	0.032	0.032
144	0.040	0.033	0.033
168	0.033	0.033	0.035
180	0.028	0.030	0.032
192	0.022	0.026	0.028
204	0.017	0.022	0.022
216	0.022	0.020	0.020
228	0.030	0.024	0.024
240	0.036	0.030	0.028
252	0.040	0.035	0.033
264	0.042	0.037	0.036
288	0.046	0.0412	0.040
300	0.046	0.042	0.040
312	0.046	0.042	0.041
324	0.045	0.044	0.042
336	0.045	0.046	0.045
348	0.048	0.050	0.050
360	0.054	0.057	0.057