

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

A Thesis Submitted for the Degree of PhD at the University of Warwick

<http://go.warwick.ac.uk/wrap/63584>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.



Term Structure Modelling: Pricing and Risk Management

by

Peter Weigel

Thesis

Submitted in partial fulfilment of the
requirements for admission to the degree of

Doctor of Philosophy

**University of Warwick
Warwick Business School**

May 2003



THE LIBRARY
Tel: +44 24 76523523
Fax: +44 24 76524211

AUTHOR: Peter Weigel DEGREE: Ph.D.

TITLE: Term Structure Modelling:
Pricing and Risk Management

DATE OF DEPOSIT: 28/08/03

I agree that this thesis shall be available in accordance with the regulations governing the University of Warwick theses.

I agree that the summary of this thesis may be submitted for publication.

I agree that the thesis may be photocopied (single copies for study purposes only).

Theses with no restriction on photocopying will also be made available to the British Library for microfilming. The British Library may supply copies to individuals or libraries. subject to a statement from them that the copy is supplied for non-publishing purposes. All copies supplied by the British Library will carry the following statement:

“Attention is drawn to the fact that the copyright of this thesis rests with its author. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author’s written consent.”

AUTHOR’S SIGNATURE: P. Weigel

USER’S DECLARATION

- 1. I undertake not to quote or make use of any information from this thesis without making acknowledgement to the author.
- 2. I further undertake to allow no-one else to use this thesis while it is in my care.

DATE	SIGNATURE	ADDRESS
.....
.....
.....
.....
.....

Contents

List of Tables	vii
List of Figures	x
Acknowledgments	xiv
Declarations	xv
Abstract	xvii
Chapter 1 Introduction to This Thesis	1
Chapter 2 Review of Term Structure Modelling	8
2.1 Introduction	8
2.2 General Framework	13
2.2.1 Pricing Kernel	15
2.2.2 Short Rate Models	25
2.2.3 HJM Approach	31
2.2.4 Alternative Models and Term Structure Modelling Issues . .	37
2.3 Interest Rate Risk Management	41
2.3.1 Traditional Measures of Interest Rate Risk: Duration and Convexity	42
2.3.2 Risk Measures in Pricing Models	46

2.3.3	Empirical Results and General Remarks	53
2.4	Conclusions	54
Chapter 3	Modelling Framework	57
3.1	Introduction	57
3.2	Related Research	62
3.3	Model Specification	67
3.3.1	Infinite-Factor Gaussian Field Models	68
3.3.2	Smoothness of the Field	69
3.4	Conclusions	72
3.5	Appendix	74
Chapter 4	SPD Covariance and Correlation Functions via SPD Ap-	
	proximations	76
4.1	Introduction	76
4.2	Data	79
4.3	SPD Approximation of the Covariance Function	80
4.3.1	Matrix Nearness Solution	81
4.3.2	Surface Approximation, Numerical Analysis, and Empirical Results	84
4.4	SPD Approximation of the Correlation Matrix	91
4.4.1	Method of Alternating Projection	91
4.4.2	Numerical Analysis and Empirical Results	97
4.5	Conclusions	100
4.6	Appendices	103
4.6.1	Figures	103
4.6.2	Tables	113
Chapter 5	SPD Correlation Function with Functional Deformation	
	of Maturity	123

5.1	Introduction	123
5.2	Modelling Framework	125
5.3	Deformation Mappings and Strictly Positive Definite Functions . . .	130
5.4	Data	136
5.5	Sampson and Guttorp Approach	137
5.5.1	Description of the Method	137
5.5.2	Numerical Results	139
5.6	Minimisation of an Objective Function	140
5.6.1	Description of the Methodology	140
5.6.2	Implementation Details and Numerical Results	141
5.6.3	Extensions	143
5.7	Iterative Two-Step Implementation: MNM-LM Algorithm	145
5.7.1	Description of the MNM-LM Algorithm	146
5.7.2	Numerical Results	148
5.8	Alternative Estimation Methods	150
5.9	Conclusions and Discussions	152
5.10	Appendices	155
5.10.1	Simar's Algorithm	155
5.10.2	B-Spline Basis	156
5.10.3	SPD Functions and Kernels	156
5.10.4	Figures	157
5.10.5	Tables	166
Chapter 6	SPD Covariance Functions via Superposition	175
6.1	Introduction	175
6.2	Modelling Framework	176
6.3	Implementation Methodology	181
6.4	Numerical and Empirical Studies	183
6.5	Conclusions and Further Research	189

6.6	Appendices	191
6.6.1	Figures	191
6.6.2	Tables	205
Chapter 7 Implied Kernel Models		213
7.1	Introduction	213
7.2	Related Research	215
7.2.1	Kernel Models	216
7.2.2	Modelling Non-Flat Volatilities	219
7.3	Modelling Framework	221
7.3.1	The Underlying Markov Process	221
7.3.2	Approximation by Radial and Ridge Functions	223
7.3.3	Formulae for Pricing Zero-Coupon Bonds, Caps, and Swaptions	227
7.4	Implementation and Numerical Study of the Model	228
7.4.1	Yield Curve and ATM Caps	228
7.4.2	The Yield Curve and Caplet Volatility Surface	230
7.4.3	The Yield Curve and ATM Swaptions	231
7.4.4	Properties of the Model	231
7.5	Conclusions	232
7.6	Appendices	234
7.6.1	Recovering Forward-Forward Volatilities	234
7.6.2	Proof of Lemma 7.6.1	235
7.6.3	Computing Integrals of Matrix Exponentials	236
7.6.4	Gauss-Hermite Quadrature	238
7.6.5	Figures	239
7.6.6	Tables	251
Chapter 8 Conclusions and Further Research		259

List of Tables

4.1	Comparison between strictly positive definite approximant with the bound $1e-15$ and interpolated covariance.	86
4.2	Comparison between strictly positive definite approximant with the bound b and interpolated covariance.	89
4.3	Comparison between strictly positive definite approximant with the bound b and interpolated covariance.	90
4.4	Comparison between strictly positive definite approximant with a bound and interpolated correlation function, using two-dimensional cubic splines.	97
4.5	Comparison between strictly positive definite approximant with a bound and interpolated correlation function, using two-dimensional interpolating surface splines.	98
4.6	Comparison between strictly positive definite approximant with a bound and interpolated correlation function, using Shepard Interpolation.	100
4.7	Descriptive statistics for the yield differences.	114
4.8	Sample covariance matrix for the first half of the sample, multiplied by a factor $1e + 7$	115
4.9	Sample correlation matrix for the first half of the sample.	116
4.10	Approximation of the correlation surface with two-dimensional cubic splines. Approximation based on yearly grid.	117
4.11	Approximation of the correlation surface with two-dimensional cubic splines. Approximation based on monthly grid.	118
4.12	Approximation of the correlation surface with two-dimensional surface splines. Approximation based on yearly grid.	119
4.13	Approximation of the correlation surface with two-dimensional surface splines. Approximation based on monthly grid.	120
4.14	Approximation of the correlation surface with Shepard interpolation. Approximation based on yearly grid.	121
4.15	Approximation of the correlation surface with Shepard interpolation. Approximation based on monthly grid.	122

5.1	Strictly positive definite functions on \mathbb{R}	131
5.2	Descriptive statistics for the yield differences. Currency: Yen. . . .	167
5.3	Sample covariance matrix, multiplied by a factor $1e + 7$. Currency: Yen.	168
5.4	Sample correlation matrix. Currency: Yen.	169
5.5	Values of the correlation function for selected maturities together with percentage errors resulting from Sampson-Guttorp approach with Gaussian correlation function. Currency: Yen.	170
5.6	Values of the correlation function for selected maturities together with percentage errors resulting from Sampson-Guttorp approach with Matérn correlation function. Currency: Yen.	171
5.7	Currency: Yen. Fitted Exponential-Power correlation function using B-splines, together with percentage errors.	172
5.8	MNM-LM method. Currency: Yen. Fitted Exponential-Power corre- lation function, together with percentage errors.	173
5.9	MNM-LM method. Currency: Yen. Fitted Matérn correlation func- tion, together with percentage errors.	174
6.1	Eigenvalues of the residual covarinace matrix R using stationary Exponential- Power SPD function. Currency: Yen.	206
6.2	Eigenvalues of the residual covarinace matrix R using stationary Matérn SPD function. Currency: Yen.	206
6.3	Eigenvalues of the residual covarinace matrix R using stationary Exponential- Power SPD function. Currency: Dollar.	206
6.4	Eigenvalues of the residual covarinace matrix R using stationary Matérn SPD function. Currency: Dollar.	206
6.5	Values of the correlation function for observed maturities together with percentage errors resulting from kernel approach with Exponential- Power SPD function. Currency: Yen.	207
6.6	Values of the covariance function for selected maturities together with percentage errors resulting from kernel approach with Exponential- Power SPD function. Currency: Yen.	208
6.7	Values of the correlation function for selected maturities together with percentage errors resulting from kernel approach with Matérn SPD function. Currency: Yen.	209
6.8	Values of the covariance function for selected maturities together with percentage errors resulting from kernel approach with Matérn SPD function. Currency: Yen.	210
6.9	Values of the correlation function for selected maturities together with percentage errors resulting from kernel approach with Exponential- Power SPD function. Currency: Dollar.	211

6.10	Values of the covariance function for selected maturities together with percentage errors resulting from kernel approach with Exponential-Power SPD function. Currency: Dollar.	212
7.1	Zero Coupon Discount Function (ZCDF) for 3 Feb 1995.	251
7.2	Market Data for 3 Feb 1995.	251
7.3	Discount bond price for 3 Feb 1995 together with model price and percentage error for the fit.	252
7.4	Results from the model calibration to the ATM caps and zero coupon bond prices. Date: 03-Feb-1995.	253
7.5	Model and market swaption prices, resulting from calibration to the ATM caps and zero coupon bond prices. Date: 03-Feb-1995.	253
7.6	Implied black swaption volatility, resulting from calibration to the ATM caps and zero coupon bond prices. Date: 03-Feb-1995.	253
7.7	Zero coupon bond prices together with calibrated model prices and percentage error. The model was calibrated to the caplet and bond prices. Date: 04-Aug-00.	254
7.8	Calibrated model and market caplets prices for a range of strikes and maturities, together with errors expressed in basis points and percentage. Date: 04-Aug-00.	255
7.9	Mean, standard deviation, skewness and kurtosis for LIBOR rates, resulting from the calibration of the model to the zero-coupon bond and caplets. Date: 04-Aug-00.	256
7.10	Zero coupon bond prices together with calibrated model prices and percentage error. The model was calibrated to the ATM swaptions and bond prices. Date: 04-Aug-00.	257
7.11	Swaptions contracts expressed as maturity \times length. Model fit to the ATM swaption prices in basis point units together with market prices and the percentage errors. Date: 04-Aug-00.	258

List of Figures

4.1	Alternating Projection on closed subsets of \mathbb{R}^2	103
4.2	The positive semidefinite matrix cone K	103
4.3	Sample covariance matrix (above) and sample correlation matrix. .	104
4.4	Approximation of covariance using two-dimensional cubic splines (above) and two-dimensional interpolating surface splines (below).	105
4.5	Approximation of covariance using two-dimensional smoothing surface splines, with smoothing coefficient $\rho = 10$	106
4.6	Two-dimensional cubic splines: Eigenvalues of interpolated matrix (left), and eigenvalues of the nearest covariance matrix, with eigenvalues bounded by bound = $1e - 15$	107
4.7	Two-dimensional surface splines: Eigenvalues of interpolated matrix (left), and eigenvalues of the nearest covariance matrix, with eigenvalues bounded by bound = $1e - 15$	108
4.8	Two-dimensional surface smoothing splines with smoothing parameter $\rho = 10$: Eigenvalues of interpolated matrix (left), and eigenvalues of the nearest covariance matrix, with eigenvalues bounded by bound = $1e - 15$	109
4.9	The data variances together with variances resulting from positive-definite approximants. Two-dimensional cubic splines (above) and two-dimensional smoothing surface splines (below), with smoothing coefficient $\rho = 10$	110
4.10	Two-dimensional cubic splines: Eigenvalues of interpolated matrix (left), and eigenvalues of the nearest correlation matrix, with eigenvalues bounded by bound as in Table 4.4.	111
4.11	Two-dimensional interpolating surface spline: Eigenvalues of the interpolated matrix (left), and eigenvalues of the nearest correlation matrix, with eigenvalues bounded by bound as in Table 4.5.	112
4.12	Shepard interpolation: Eigenvalues of interpolated matrix (left), and eigenvalues of the nearest correlation matrix, with eigenvalues bounded by bound as in Table 4.6.	113

5.1	Examples of stationary correlation functions. (a) Exponential, with $a = 0.5, 1, 2$. (b) Exponential-Power, with $a = 1, b = 0.5, 1, 2$. (c) Triangular, with $a = 0.5, 1, 2$. (d) Power law, with $b = 1$ and $a = 0.5, 1, 1.5$. (e) Matérn, with $\sigma = 1, \nu = 1$, and $\rho = 1, 2, 4$. (f) Matérn with $\sigma = 1, \rho = 2$, and $\nu = 0.8, 2, 5$	158
5.2	Sample covariance matrix (above) and sample correlation matrix (below). Currency: Yen.	159
5.3	Deformation coordinates produced by MDS algorithm (Top). Sampson-Guttorp approach with Gaussian correlation function (Bottom). Currency: Yen.	160
5.4	Deformation function represented by B-splines, for the exponential-power correlation function with coefficients $a = 0.0059$ and $b = 1.575$ (above). Fitted deformed exponential-power correlation functions with the same parameters (below).	161
5.5	Eigenvalues of the matrix resulting from the deformed Exponential-Power correlation function with parameters $a = 0.0059$ and $b = 1.575$. Deformation is represented by B-splines.	162
5.6	MNM-LM method. Deformation points for the exponential-power correlation structure (above), with the coefficients $a = 1$ and $b = 1.56$. Deformation points for the Matern correlation structure (below), with the coefficients $\sigma = 1, \nu = 1.3$, and $\rho = 3.57$	163
5.7	Eigenvalues of the matrices resulting from exponential-power (above), and Matérn correlation functions.	164
5.8	Deformed exponential-power correlation function (above). Deformed Matérn correlation function (below).	165
5.9	Plot of the exponential-power correlation function with the parameters $a = 1$ and $b = 1.56$ (above). Plot of the Matérn correlation function with the parameters $\sigma = 1, \nu = 1.3$, and $\rho = 3.57$ (below).	166
6.1	SPD covariance function (above) and SPD correlation function resulting from Exponential-Power SPD function with the parameters $a = 0.0294$ and $p = 0.537$, and the stationarity parameter $\alpha = 0.1$. Currency: Yen.	191
6.2	Eigenvalues of covariance and correlation functions (above) resulting from the use of exponential-power SPD function. First three empirical basis functions (below). Currency: Yen.	192
6.3	Second three empirical basis functions (above). Third three empirical basis functions (below). Currency: Yen.	193
6.4	SPD covariance function (above) and SPD correlation function resulting from Matérn SPD function with the parameters $\nu = 109.341$ and $\rho = 51.975$, and the stationarity parameter $\alpha = 0.1$. Currency: Yen.	194

6.5	Eigenvalues of covariance and correlation functions (above) resulting from Matérn SPD function. First three empirical basis functions (below). Currency: Yen.	195
6.6	Second three empirical basis functions (above). Third three empirical basis functions (below). Currency: Yen.	196
6.7	SPD covariance function (above) and SPD correlation function resulting from Exponential-Power SPD function with the parameters $a = 0.1035$ and $p = 0.4315$, and the stationarity parameter $\alpha = 0.1$. Currency: Dollar.	197
6.8	Eigenvalues of covariance and correlation functions (above). First three empirical basis functions (below). Currency: Dollar.	198
6.9	Second three empirical basis functions (above). Third three empirical basis functions (below). Currency: Dollar.	199
6.10	SPD covariance function (above) and SPD correlation function resulting from Matérn SPD function with the parameters $\nu = 106.33$ and $\rho = 31.429$, and the stationarity parameter $\alpha = 0.1$. Currency: Dollar.	200
6.11	Eigenvalues of covariance and correlation functions (above). First three empirical basis functions (below). Currency: Dollar.	201
6.12	Second three empirical basis functions (above). Third three empirical basis functions (below). Currency: Dollar.	202
6.13	Exponential-Power SPD function with paramters $a = 0.0294$ and $p = 0.537$, and Matérn with parameter $\nu = 109.341$ and $\rho = 51.975$. Currency: Yen (above). Exponential-Power SPD function with paramters $a = 0.1035$ and $p = 0.4315$, and Matérn with parameter $\nu = 106.33$ and $\rho = 31.429$. Currency: Dollar (below).	203
6.14	Alpha against percentage errors in correlation (above). Alpha against percentage errors in correlation (below). Currency: Dollar.	204
6.15	Correlation function with Exponential-Power SPD function (above). Correlation function with Matérn SPD function (below).	205
7.1	A radial basis function of the form $\exp(-a\ x - b\)$, with $a = 1$ and $b = (3, 3)$	240
7.2	A radial basis function of the form $1/\sqrt{1 + a\ x - b\ }$, with $a = 0.5$ and $b = (3, 3)$	240
7.3	A radial basis function of the form $\log(1.1 + -a\ x - b\)$, with $a = 200$ and $b = (3, 3)$	241
7.4	A radial basis function of the form $\sqrt{0.1 + a\ x - b\ }$, with $a = 0.1$ and $b = (3, 3)$	241
7.5	A typical ridge function.	242
7.6	A logistic sigmoid with the parameter $a = 2$	242
7.7	A sigmoid ridge function.	243

7.8	Yield and forward rate curves. Date: 03-Feb-95.	244
7.9	Forward forward volatility curve. Date: 03-Feb-95.	244
7.10	Yield and forward rate curves. Date: 04-Aug-00.	245
7.11	Black caplet implied volatility surface. Date: 04-Aug-00.	245
7.12	Market caplet prices. Date: 04-Aug-00.	246
7.13	Model caplet prices. Date: 04-Aug-00.	246
7.14	Histogram for the forward LIBOR rates with maturity 2.5 years, together with normal and log-normal densities. The LIBOR rate distribution results from the model fit to the bond and caplet prices. Date: 04-Aug-00.	247
7.15	Histogram for the forward LIBOR rates with maturity 5 years, together with normal and log-normal densities. The LIBOR rate distribution results from the model fit to the bond and caplet prices. Date: 04-Aug-00.	247
7.16	Histogram for the forward LIBOR rates with maturity 7.5 years, together with normal and log-normal densities. The LIBOR rate distribution results from the model fit to the bond and caplet prices. Date: 04-Aug-00.	248
7.17	Histogram for the forward LIBOR rates with maturity 10 years, together with normal and log-normal densities. The LIBOR rate distribution results from the model fit to the bond and caplet prices. Date: 04-Aug-00.	248
7.18	Model and caplet prices with 2.5 years to maturity. Date: 04-Aug-00.	249
7.19	Model and caplet prices with 5 years to maturity. Date: 04-Aug-00.	249
7.20	Model and caplet prices with 7.5 years to maturity. Date: 04-Aug-00.	250
7.21	Model and caplet prices with 10 years to maturity. Date: 04-Aug-00.	250

Acknowledgments

I am grateful to my supervisor Professor Stewart Hodges for his advise and stimulating discussions.

I would like to thank Diana for her love, support, and encouragement during the research and writing process, and above all for becoming a very special person in my life.

Special thanks to my family and all my friends for making my life so pleasant while working on this thesis.

Finally, I would like to acknowledge my viva examiners, Professor Doug Kennedy from Cambridge University and Professor Stephen Schaefer from London Business School, for their time and comments on this thesis.

Declarations

This thesis is my own original work. No part of it has been submitted for a degree at another university.

In Erinnerung an meine Mutter

Abstract

This thesis is about interest rate modelling with applications in pricing and risk management of interest rate derivatives and portfolios.

The first part of the thesis is developed within the random field framework suggested by Kennedy (1994). The framework is rich enough to be used for both pricing and risk management, but we believe its real value lies in the latter. Our main objective is to construct infinite-factor Gaussian field models that can fit the sample covariance matrices observed in the market. This task has not previously been addressed by the work on field methodology. We develop three methodologies for constructing strictly positive definite covariance functions, characterising infinite-factor Gaussian fields. We test all three constructions on the sample covariance and correlation matrices obtained from US and Japanese bond market data. The empirical and numerical tests suggest that these classes of field models present very satisfactory solutions to the posed problem. The models we develop make the random field methodology a much more practical tool. They allow calibration of field models to key market information, namely the covariation of the yields.

The second part of the thesis deals with pricing kernel (potential) models of the term structure. These were first introduced by Constantinides (1992), but were subsequently overshadowed by the market models, which were developed by Miltersen *et al.* (1997), and Brace *et al.* (1997), and are very popular among the practitioners. Our objective is to construct a class of arbitrage-free term structure models that enjoy the same ease of calibration as the market models, but do not suffer from non-Markov evolution as is the case with the market models. We develop a class of models the within pricing kernel framework. I.e., we model the pricing kernel directly, and not a particular interest rate or a set of rates. The construction of the kernel is explicitly linked to the calibrating set of instruments. Thus, once the kernel is constructed it will price correctly the chosen set of instruments, and have a low-dimensional Markov structure. We test our model on yield, at-the-money cap, caplet implied volatility surface, and swaption data. We achieve a very good quality of fit.

Chapter 1

Introduction to This Thesis

Modelling the dynamics of interest rates has become an important as well as extensively studied area in mathematical finance, financial economics and econometrics. This thesis is about interest rate modelling with applications in pricing and risk management of interest rate derivatives and portfolios.

The first part of the thesis is developed within the random field framework suggested by Kennedy (1994). Random field models of the term structure are generalisations of the finite factor models which have dominated the financial economics literature. In a field model, each point along the term structure is a distinct random variable with its own dynamics. The framework is rich enough to be used for both pricing and risk management, however we believe its real value lies in the latter. In contrast with finite-factor models, random field models are consistent with both the current yield curve and any term structure innovation.

What is more important, especially for interest rate risk management applications, Gaussian field models have the potential to exactly fit the empirical covariance structure. Furthermore, N -factor models predict that any, for example, long-term maturity bond can be perfectly hedged with an appropriate position in N -short-term bonds, while a random field model predicts that a better hedging instrument

for a long-term bond is another of similar maturity. This is in agreement with the common hedging practice in fixed income markets.

Our main objective is to construct infinite-factor Gaussian field models that can fit the sample covariance matrices observed in the market. To the best of our knowledge, this task has not previously been addressed by the work on field methodology. We develop three methodologies for constructing strictly positive definite covariance functions, characterising infinite-factor Gaussian fields. We test all three constructions on the sample covariance and correlation matrices obtained from the US and Japanese bond market data. The empirical and numerical tests suggest that these classes of field models present very satisfactory solutions to the posed problem. The models we develop make the random field methodology a much more practical tool. They allow calibration of field models to the key market information, the covariation of the yields.

The second part of this thesis deals with pricing kernel (potential) models of the term structure. These were first introduced by Constantinides (1992), but were subsequently overshadowed by the market models, developed by Miltersen *et al.* (1997), and Brace *et al.* (1997). The market models are very popular among the practitioners, as they allow almost instantaneous calibration to the liquid market prices. However, the main disadvantage of market models turns out to be their non-Markovian property in low-dimensions. This fact inhibits the use of these models for pricing exotic products, where the short rate approach is still the preferred choice.

Thus it is desirable to find a class of models that enjoy the same ease of calibration to the liquid market prices with the market models, and at the same time has a low-dimensional Markov structure as in the case of short rate models. The HJM and the short rate approaches to term structure modelling have been studied extensively. Thus it is unlikely they will yield any class of models that has the desired properties. On the other hand the pricing kernel framework has been studied less extensively.

We feel that it has enough scope and flexibility to achieve our goal.

We develop a class of models within the pricing kernel framework. I.e. we model the pricing kernel directly, and not a particular interest rate or a set of rates. The construction of the kernel is explicitly linked to the calibrating set of instruments. Thus, once the kernel is constructed it prices correctly the chosen set of instruments and has a low-dimensional Markov structure. We demonstrate that the model calibrates well to data. The data used comprised of the yield curve and at-the-money caps for GB pound on February 3, 1995, the yield curve and the caplet black implied volatility surface for GB pound on August 4, 2000, yield curve and at the money black implied swaption matrix, for 4th August, 2000. We achieve a very good quality of fit, measured as the percentage difference between the model and market prices.

Thesis Structure

We review the term structure pricing literature in Chapter 2. The common way to describe the literature is to start with the historically oldest short rate type models, e.g. Vasicek (1977). Then one usually continues with the HJM framework, market models, and so on. Instead of following this standard route, we discuss the literature from the top down. I.e. we will start with the most general arbitrage-free framework, the pricing kernel, then we specialise this framework to the less general classes of models, the short rate models: models of the instantaneous forward rates, etc. This will allow us to place all the model classes within the general pricing theory and understand better their advantages, shortcomings and relations to each other. For example, knowing that the class of market models is a subfamily of the HJM class will help us to understand why market models lack the Markov property.

Part of an input in field models is motivated by issues of fixed income risk measurement and management. In Chapter 2 we also review this literature. Unfortunately, there is no unifying theory as in the case of arbitrage-free pricing. Thus, this chapter

is mainly a collection of techniques, tools, and ideas that can help to understand some of the risks involved in fixed income portfolios. This literature is surprisingly undeveloped and unsatisfying.

In Chapter 3 we describe the Gaussian random field framework. We describe in some detail results and analysis on the field models obtained by Kennedy (1994), Goldstein (2000), and Santa-Clara and Sornette (2001). As distinct from previous authors, we work with the yield curve as fundamental, and model it as a Gaussian random field. This is not essential, as the results will be equally applicable to other term structure parameterisations. We require that the covariance function of the Gaussian field be strictly positive definite (SPD) so that the evolution of the yield is driven by infinite-factor structure. Additionally we discuss several smoothness properties of the yield curve and its evolution. We derive sufficient conditions on the covariance function that are needed to produce such degree of smoothness.

Chapter 4 is the first of three chapters in which we develop methodologies for constructing strictly positive definite instantaneous covariance and correlation functions. In this chapter we provide two alternative non-parametric methodologies. The first methodology, in which we derive an estimate of the instantaneous covariance function, consists of two steps. In the first step we approximate the covariance function on the unobserved grid points, i.e. we exploit the smoothness property of the covariance function. In the second step, we evaluate this approximating function on a grid represented by the required set of maturities. This generates a matrix, which among other entries includes the observed covariance values. Next, we use the *Matrix Nearness* approach to approximate this estimate matrix with the closest positive definite matrix in some norm. Thus, we obtain a strictly positive definite approximant, and we choose this approximant as the estimate of the covariance function on the chosen grid.

The second methodology, in which we derive an estimate of the instantaneous cor-

relation, again consists of two steps. The first step is identical to the first step of the previous methodology, except that we add an additional interpolation technique that takes into account the unit diagonal property of a correlation function. The second step, however, is more involved: the approximation technique used in the previous methodology is not applicable in this case, as the approximating positive definite matrix does not necessarily have unit diagonal. We then apply a modified *Alternating Projection Method* to find an approximating positive definite correlation matrix in the intersection of two convex sets: symmetric matrices with unit diagonal and positive definite matrices. Thus, we obtain a positive definite approximant with a unit diagonal. We choose this approximant as the estimates of the correlation functions on the chosen grid. We test both methodologies numerically on the US interest rate data.

In Chapter 5 we turn to parametric constructions of the correlation functions. We first identify a large class of strictly positive definite functions¹, i.e. a functions of the form $f(x-y)$. Unfortunately, SPD functions produce only stationary fields, i.e. fields with covariance functions depending only on distance between maturities. In our case this means that the covariance function of the increments of the yields will be a function of differences of maturities, i.e. function of $\text{Corr}(T_1, T_2) = g(|T_1 - T_2|)$. Unfortunately, such models are misspecified: the empirical covariance matrices strongly indicate that the covariance function is a function of T_1 and T_2 separately. We adjust for non-stationarity by deforming time to maturity and obtain SPD kernels of the form $\text{Corr}(T_1, T_2) = g(|f(T_1) - f(T_2)|)$, where g is an SPD function, and f a strictly increasing function. We test this class of models on the Japanese Yen bond data. We do not achieve exact fit, but the errors are reasonably small and could be attributed to the sampling error.

The class of models developed in Chapter 5 is quite restrictive, and might not be

¹Usually, a function of the form $f(x-y)$ with the SPD property is referred to as an SPD function. An SPD function of the form $f(x, u)$ is referred to as an SPD kernel.

justified in practice. We approach the problem from a different angle in Chapter 6. The idea is based on superposition: we model the covariance functions as a sum of SPD functions developed in Chapter 5 and PD kernels. The PD kernels correct the SPD functions for non-stationarity and enable exact fit to the sample covariance matrix. We first fit an SPD function to the correlation matrix, this gives us the stationary part of the covariance function. We obtain the PD correction kernel by applying Principal Component Analysis to the residual matrix obtained from the previous fit. As a sum of an SPD function and a PD kernel, the covariance function is an SPD kernel. This construction satisfies the modelling assumptions. We test the models on the sample covariance matrices obtained from US and Japanese bond market data. In all tests we achieve an exact fit. This approach seems to present the best solution for constructing infinite-factor Gaussian field models if one wants to achieve an exact fit to the sample covariance matrix.

In Chapter 7 we study the pricing kernel framework. We develop a class of arbitrage-free multi-factor models within this framework. We model the pricing kernel directly, and not a particular interest rate or a set of rates. In particular, the kernel is represented by a series of radial basis functions, used in Approximation Theory. The construction of the kernel is explicitly linked to the calibrating set of instruments. Thus, once the kernel is constructed it prices correctly the chosen set of instruments and has a low-dimensional Markov structure. We conduct several calibration studies. We fit the model to the yield curve and at-the-money cap prices, to the yield and the caplet surface across a wide range of strikes and maturities, and to the yield curve and at-the-money swaption matrix. We achieve a very good quality of fit, measured in terms of percentage difference between the model prices and market prices.

We conclude and present some ideas on further research in Chapter 8. It should be noted, that although we achieved some very good results in constructing and fitting both implied kernel and infinite-factor Gaussian field models, more extensive empirical testing needs to be conducted. Comparison studies should reveal if these

classes of models are superior to alternative model specifications in pricing and risk management of fixed income instruments and portfolios.

Equivalently, our research suggests that implied pricing kernels may provide an alternative to the standard market models. They are Markov by construction, calibration for a small number of factor is relatively easy, they can deal with American type options, and are flexible in fitting to smiles and skews. Market models have non-Markov dynamics, but are easy to calibrate to certain² European-type instruments even with a large number of factors. Extensions for smiles and skews are quite complex. Further research needs to address the extent of the advantage of the kernel models over other model specifications. Better implementation and calibration techniques need to be developed to exploit the potential of the pricing kernel framework fully.

²Usually, only to instruments for which they were specifically designed.

Chapter 2

Review of Term Structure Modelling

2.1 Introduction

In this chapter we review the term structure pricing and risk management literature. In the present section we introduce some fundamentals commonly used in the construction of term structure models. Then we present a short discussion of some basic concepts of arbitrage-free pricing. In Section 2.2, we proceed with a detailed discussion of the three modelling frameworks used in pricing literature: the pricing kernel, the short rate approach, and the HJM family. We do not follow the conventional route, i.e. a chronological description of the developments in the literature. Instead we discuss the literature from the top down. I.e. we start with the most general arbitrage-free framework, the pricing kernel. In this general setting the short rate or instantaneous forward rates do not necessarily exist. Then we specialise this framework to the less general classes of models: the short rate models, models of the instantaneous forward rates, etc. This will allow us a better understanding of term structure pricing theory and the relation among different modelling approaches. We review the literature for each of these frameworks, discuss their interrelation.

We also describe term structure models that do not naturally fit into the above description, such as infinite-factor models and models for futures prices. However, we defer a detailed discussion of infinite-factor models to Chapter 3. Furthermore, we discuss the advantages and disadvantages of different modelling approaches, parameter estimation issues versus model calibration, and model risk problems.

In Section 2.3.2 we review the literature on fixed income risk measurement and management. Unlike the pricing literature, the risk management literature is rather undeveloped. Currently, there is no dominating theory in this field. Consequently, we present the main ideas and techniques currently in use, point out their shortcomings and suggest possible avenues for further research. We also comment on how the literature might benefit from the research in this thesis. We conclude in Section 2.4.

We first introduce some basic terminology. A *zero coupon bond with maturity T* is a security that pays at time T the amount 1. Let $B(t, T)$ denote the price prevailing at time t of a zero coupon bond with maturity T . The family of bond price processes $B(t, T)$, is called a *term structure model*; we denote the family $(B(t, T))_{0 \leq t \leq T}$ by B . We denote by $B(0, T)$ the *initial term structure*, i.e. the bond prices observed in the market¹ at time 0. For B sufficiently smooth in T , the *instantaneous forward rate* $f(t, T)$ is then defined by

$$f(t, T) := -\frac{\partial}{\partial T} \log B(t, T),$$

and the *short rate* r_t

$$r_t := f(t, t).$$

The instantaneous forward rate $f(t, T)$ can be interpreted as the interest rate prevailing at time t for instantaneous risk free borrowing and lending at the later time T .

¹The estimation of the initial term structure function is not a trivial issue. There is a considerable body of literature dealing with this problem. The techniques vary from relatively simple bootstrapping procedures to some sophisticated approaches using non parametric families of curves such as B-spline (Steely (1991)), exponential spline (Vasicek and Fong (1982)), smoothing splines (Fisher *et al.* (1995)), weighted penalty roughness splines (Waggoner (1997)), and kernel smoothing methods (Linton *et al.* (1999)). Other popular choices are parametric families such as those used by Nelson and Siegel (1987).

This means that an investor can commit himself at time t to instantaneous risk free borrowing and lending at time T at the rate $f(t, T)$. The short rate is then the interest rate at time t for instantaneous risk-free borrowing and lending at the same time t .

The *savings account* β is defined as

$$\beta_t := \exp \left(\int_0^t r_s ds \right)$$

and can be viewed as the value of one currency unit permanently reinvested at the short rate r_t . Note that the short rate is by definition a *compound rate*. A compound interest rate applies to both the principal (here one currency unit) and the accumulated interest. In contrast, the LIBOR rate (London Inter-Bank Offer Rate) is a simple rate in the sense that it applies only to the principal amount. Mathematically, the LIBOR rate $L(t, T; \delta)$ for fixed $\delta > 0$ is given by

$$1 + \delta L(t, T; \delta) := \frac{B(t, T)}{B(t, T + \delta)}. \quad (2.1.1)$$

Because δ is fixed, for ease of notation we write simply $L(t, T)$ instead of $L(t, T; \delta)$. The LIBOR rate $L(t, T)$ can be regarded as the simple interest rate prevailing at time t for an investment at time T of time length δ . It is important to notice that the choice of a particular form for the interest rates is not purely a matter of taste or practical requirements at hand. In some cases, for example when the rates are log-normal, the continuously compound interest rates have a serious drawback: rates explode² and expected rollover returns are infinite even if the rollover period is arbitrarily short. Sandmann and Sondermann (1997) realised that serious problems result from modelling, for example, the short rate as log-normal, and they disappear if one instead models the rate³ for any fixed tenor in this way. This idea has led to

²This has been shown within Eurodollar Futures pricing framework by Hogan and Weintraub (1993).

³These ideas were further developed by Goldys *et al.* (1994), Musiela (1994), Sandmann *et al.* (1995). Their implementation, use, and several extensions are extensively described in a recent book by Rebonato (2002).

a major breakthrough in the last decade in interest rate derivative pricing, namely the creation of so called *market models*⁴.

The theory of term structure models is concerned with modelling the bond prices, LIBOR rates, forward rates or the short rate as stochastic processes $B(\cdot, T)$, $L(\cdot, T)$, $f(\cdot, T)$, or r_t respectively, on a filtered probability space. Several approaches have been studied in the literature. We can classify the variety of models into interest-rate based models and bond price based models. In the *bond price* approach the bond prices are treated as fundamental and the various rates are then obtained from the bond prices. In *interest-rate* models one starts with a model for some rate such as the short rate or the forward rate and then derives the bond prices. A major drawback of these interest-rate models is the fact that getting the dynamics of asset prices and rates one is interested in (e.g. bonds and LIBOR rates) requires a non-trivial computational effort.

Next, we give a short summary of the relation between the absence of arbitrage opportunities in a financial market and the existence of an equivalent martingale measure. This concept, together with a notion of completeness, is at the heart of a model of a financial market. The former allows us to give a well-defined price to a derivative and the latter to define the set of derivatives which can be priced in the market. This theory was initiated by Harrison and Kreps (1979). In a finite discrete-time setting they considered a stochastic process $S_t := (S_t^0, S_t^1, \dots, S_t^n)$ on a finite probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\mathbb{F} := (\mathcal{F}_t)_{t=0,1,\dots,T}$; S^i here represents the price of the i -th security. Let $S_t^0 \equiv 1$. This means that the price of the securities are quoted in units of S^0 . Hence S^0 is also called *numéraire*. A *trading strategy* is a predictable process $\phi_t := (\phi_t^0, \phi_t^1, \dots, \phi_t^n)$ and its *value process* V^ϕ is given by $V_t^\phi := V_0^\phi + \sum_{i=0}^n \phi_t^i S_t^i$. The i -th component of ϕ can be viewed as the quantity of the i -th security held at time t . A portfolio is called *self-financing* if $dV_t^\phi = \sum_{i=0}^n \phi_t^i dS_t^i$, i.e., the changes of its value process are only due to changes

⁴These are models created by Miltersen *et al.* (1997), Brace *et al.* (1997), and Jamshidian (1997).

in the prices of securities. This means that no funds are withdrawn from or added to the portfolio. An *arbitrage opportunity* ϕ is a self-financing portfolio ϕ which delivers a profit with positive probability at zero cost. To be precise, this is a portfolio ϕ with $V_0^\phi = 0$, $V^\phi \geq 0$, $dV^\phi = \sum_{i=0}^n \phi_i^i dS_i^i$ and yet $\mathbb{E}[V_T^\phi > 0] > 0$. Then the Fundamental Theorem of Asset Pricing states that the absence of arbitrage opportunities is equivalent to the existence of a measure \mathbb{Q} equivalent to \mathbb{P} , such that the security price vector S is a martingale under \mathbb{Q} . The measure \mathbb{Q} is called an *equivalent martingale measure* for S with respect to the numéraire S^0 .

The ability to give a price to a derivative is based on *replication*. The value of the derivative is mimicked by a portfolio of assets in the market and so its price is defined as the value of this portfolio. Assets of the economy can replicate or span a certain set of derivatives. Completeness investigates this set of derivatives.

The equivalence of the absence of arbitrage opportunities to the existence of an equivalent martingale measure also essentially holds in a continuous time setting on a general probability space, but becomes considerably harder to state and prove. In particular, it can be shown that on a finite horizon period, absence of arbitrage is equivalent to the existence of a numéraire pair, i.e., a strictly positive continuous semimartingale which is a traded asset and an equivalent martingale measure. For example, if one chooses the savings account as a numéraire then the corresponding equivalent martingale measure is referred to as the *risk-neutral measure*. For the case when a T -maturity bond $B(t, T)$ is chosen as a numéraire the corresponding equivalent martingale measure is referred to as the *T -forward martingale measure*. The assets discounted with this numéraires pair are martingales under this probability measure. This in turn is equivalent to the existence of a *pricing kernel*, some strictly positive continuous semimartingale under some probability measure equivalent to the objective measure. This observation is central to this thesis, as we will develop models based on this idea. Furthermore, completeness refers to uniqueness

of equivalent martingale measure, in which case the price of a derivative is uniquely defined⁵.

In Chapter 7 we will construct a general class of term structure models of the form

$$B(t, T) = \mathbb{E}^{\mathbb{P}_K} \left[\frac{K_T}{K_t} \middle| \mathcal{F}_t \right], \quad (2.1.2)$$

where K_t , is a positive pricing kernel and \mathbb{P}_K some probability measure equivalent to \mathbb{P} . In particular, for a well-defined pricing kernel K we have

$$\frac{B(t, T)}{B(t, S)} = \frac{\mathbb{E}_{\mathbb{P}_K}[\mathbb{E}_{\mathbb{P}_K}[K_T | \mathcal{F}_S] | \mathcal{F}_t]}{\mathbb{E}_{\mathbb{P}_K}[K_S | \mathcal{F}_t]},$$

It follows that for arbitrary but fixed $S \leq T$, the property $B(t, S) \geq B(t, T)$ is equivalent to $\mathbb{E}_{\mathbb{P}_K}[K_T | \mathcal{F}_S] \leq K_S$. Because $B(t, S) \geq B(t, T)$ for all $S \leq T$ corresponds to interest rates being nonnegative, one can conclude that term structure models with nonnegative interest rates are just those generated by supermartingales. This observation is subject to the existence of forward rates.

2.2 General Framework

In this section we want to introduce and examine different modelling frameworks and their relation to each other. Usually, the term structure models are defined in three different ways:

$$B(t, T) = \mathbb{E}_{\mathbb{P}_K} \left[\frac{K_T}{K_t} \middle| \mathcal{F}_t \right], \quad (2.2.1)$$

for some strictly positive, adapted semimartingale, K_t , and \mathbb{P}_K martingale measure equivalent to \mathbb{P} ;

$$B(t, T) = \mathbb{E}_{\mathbb{P}^*} \left[\exp \left(- \int_t^T r_u du \right) \middle| \mathcal{F}_t \right]$$

for some adapted short-rate process r_t and \mathbb{P}^* risk-neutral measure equivalent to \mathbb{P} ; and for sufficiently regular bond price processes $B(t, T)$,

$$B(t, T) = \exp \left(- \int_t^T f_{t,u} du \right) \quad (2.2.2)$$

⁵Points raised here are explained and proved to various degrees of generality in Artzner and Delbaen (1989), Duffie (1996), Delbaen and Schachermayer (1994), (1997), Hunt and Kennedy (2000).

defines implicitly a *family of forward rate processes* $f(t, T)$. Additionally, one requires that a term structure model should satisfy the following axioms, which we will call the term structure axioms⁶:

- A1. For all $T \geq 0$, the process $(B(t, T))$ is a continuous semimartingale adapted to filtration \mathcal{F}_t .
- A2. $B(t, T) \geq 0$ for all $t \leq T$.
- A3. $B(T, T) = 1$.
- A4. The model admits a pricing kernel. I.e. there exists some strictly positive semimartingale K such that $K_t P(t, T)$ is an $(\mathcal{F}_t, \mathbb{P}_K)$ martingale for all $0 \leq t \leq T$.

Next we give interpretations of the above axioms. In our continuous time models, in order that the gain process is well defined, we need to be able to integrate against the assets in the economy. Axiom (A1) is required for this integration to be well defined⁷ The bond holder's limited liability is represented by the positivity of the bond at all times, i.e. the bond holder shouldn't incur any cost from holding the security. Axiom (A3) represents the assumption that the bond cannot default, i.e. credit risk does not exist in our economy. We have discussed above that existence of a pricing kernel over a finite trading horizon implies the absence of arbitrage in the set of well-defined trading strategies⁸. Therefore Axiom (A4) will ensure the absence of arbitrage in our models.

In the next sections we introduce⁹ three term structure modelling frameworks: Pricing Kernel (PK), Heath, Jarrow and Morton (HJM) and Short Rate (SR) which all

⁶See Baxter (1997) for further discussion of these axioms.

⁷See Rogers and Williams (1987) pp.52-63.

⁸i.e. the set which does not have "doubling strategies", see Duffie (1996).

⁹A fairly complete literature review on term structure modelling can be found in Gibson *et al.* (1999), a very interesting summary is given by Rogers (1995), and a general summary of continuous time modelling in finance is in Sundaresan (2000).

satisfy the above term structure axioms. We briefly discuss the literature and provide examples in each of these frameworks. We examine the equivalence between the frameworks and show through the discussion of examples that the following strict inclusion holds

$$HJM \subset SR \subset PK.$$

2.2.1 Pricing Kernel

In this section we discuss literature concerned with term structure modelling within the pricing kernel framework. This way of modelling is relatively new and the amount of literature is not as large as in the other approaches. As the contribution of this thesis is mainly within this modelling framework we discuss the literature in some detail. The term “pricing kernel” is synonymous with the terms “state price density”, “state price deflator” and “stochastic discount factor”. Constantinides (1992) introduces this framework for the first time to arbitrage-free interest modelling and describes how it is related to the expected utility theory (see example below). Zheng (1993) derived bond option formulae in Constantinides’ framework. A discrete approach to this framework has been discussed by Backus and Zin (1994). Dillen (1997) considers a three-factor model in a macroeconomic framework, where he considers the foreign real interest rate, the domestic inflation rate and the real exchange rate. Rogers (1997b) relates the pricing kernel approach to the general theory of Markov processes and provides a number of interesting and tractable examples. Cairns (1999) develops a multi-factor model for long term risk management within the Rogers framework. Jin and Glasserman (2001) discuss a general positive interest rates framework based on the pricing kernel and relate it to the HJM and Flesaker and Hughston (1996a, 1996b) models. Döberlein (1999) provides technical details and proofs for the pricing kernel approach. Hunt *et al.* (2000), and Balland and Hughston (2000) model an appropriate numéraire, which is equivalent to modelling the kernel.

Supporting Equilibrium

Pricing kernels arise naturally in the equilibrium approach to interest rate modelling. The equilibrium approach¹⁰ pioneered by Cox *et al.* (1981b), (1985a), (1985b), starts from a description of the underlying economy and from assumptions about the stochastic evolution of one or more exogenous factors or state variables in the economy, and about the preferences of a representative investor. General equilibrium considerations are used to derive the interest rate and the price of all contingent claims endogenously. Both the term structure and its dynamics are endogenously determined in the equilibrium framework. Furthermore, the functional form of the factor risk premiums (marked price of risk) is also obtained as part of the equilibrium.

We illustrate the meaning of the pricing kernel in the representative-consumer economy in which the consumer has Von Neumann-Morgenstern preferences U defined over the consumption path $\{C(t) : 0 \leq t\}$. In equilibrium the first order condition states

$$\mathbb{E}_t \left[-\frac{\partial U}{\partial C(t)} \frac{P(t)}{\Pi(t)} + \frac{\partial U}{\partial C(T)} \frac{X(T)}{\Pi(T)} \right] = 0, \quad (2.2.3)$$

where $\Pi(t)$ is the price level, $P(t)$ is the nominal price at time t of a claim to a nominal payoff $X(T)$ at some future date T . Defining the pricing kernel p_t as

$$p_t = \Pi^{-1} \mathbb{E}_t \left[\frac{\partial U}{\partial C(t)} \right],$$

the first-order condition (2.2.3) yields our pricing formula (2.1.2). In the case of CIR, the pricing kernel and the interest rate process take the form

$$dp_t = (\epsilon^2 - h)Y_t p_t dt - p_t \epsilon \sqrt{Y_t} dB_t \quad (2.2.4)$$

$$dr_t = \kappa(r^* - r_t)dt + \sigma \sqrt{r_t} dB_t. \quad (2.2.5)$$

¹⁰For further examples of equilibrium models see Longstaff (1989), Longstaff and Schwartz (1992a), Chen and Scott (1992).

The density process of the underlying physical probability measure with respect to the risk-neutral measure can be defined by

$$\frac{d\mathbb{P}}{d\mathbb{P}^*}|_t = \exp\left(\int_0^t r_s ds\right) \frac{p_t}{p_0} = \exp\left(\frac{\epsilon^2/2}{\epsilon^2 - h} \int_0^t r_s ds - \frac{\epsilon}{\sqrt{h - \epsilon}} \int_0^t \sqrt{r_s} dB_s\right).$$

Using this density process we can obtain the dynamics of the interest rate process under the risk-neutral measure, and so we are back in the familiar setting¹¹.

Jin and Glasserman (2001) are able to derive a supporting equilibrium for all HJM type models. Working within pricing the kernel framework they show through specific construction that every HJM-type model arises as the equilibrium term structure in a Cox-Ingersoll-Ross economy. This type of equivalence of between arbitrage-free formulations is widely understood in general terms, see Duffie (1996).

Constantinides (1992) constructs the first pricing kernel model by directly specifying

$$K_t = \exp\left(-at + \int_{i=1}^N (X_{i,t} - \alpha_i)^2\right) \quad (2.2.6)$$

where the state variables $X_{i,t}$, with $i = 1, \dots, N$ are chosen to follow the Ornstein-Uhlenbeck processes,

$$dX_{i,t} = -\lambda_i X_{i,t} dt + \sigma_i dz_{i,t}$$

under the objective measure \mathbb{P} , with a and α_i , constants. The positive rates can be obtained by constraining the parameters of the kernel so that it follows a supermartingale or potential. For (2.2.6) to be a potential¹², the coefficients have to satisfy

$$a - \sum_{i=1}^N \lambda_i \left(\frac{\sigma_i^2}{\lambda_i} - \frac{\alpha_i^2}{2(1 - \nu_i)} \right) > 0 \quad \text{and} \quad \frac{\sigma_i^2}{\lambda_i} < 1$$

(cf: Constantinides (1992), Equations 10 and 11). Explicit formulae can be obtained for zero-coupon bonds and the short rate process has the form

$$r_t = a + \int_{i=1}^N (-b + c(X_{i,t} - d_i)^2),$$

¹¹For further details and examples see Duffie (1996) and Cochrane (2001).

¹²For more details on a potential, see page 20.

where a , b , c and d are obtained from the parameters of the model.

Zheng (1993) extends the result of Constantinides (1992) by including a deterministic function $f(t)$ into the kernel (2.2.6)

$$K_t = \exp \left(-f(t) - \alpha t + \int_{i=1}^N (X_{i,t} - \alpha_i)^2 \right). \quad (2.2.7)$$

This allows him to fit the initial yield curve. He finds that one-factor version of his extension fits the Eurodollar futures options well.

Das and Foresi (1996)

Das and Foresi (1996) consider two models within the pricing kernel approach. The first is based on the assumption that the pricing kernel evolves according to

$$-\frac{dm}{m} = rdt + \lambda dz - \lambda_J h dt + \lambda_J dN_h,$$

with interest rate process r ,

$$dr = a(b - r)dt + \sigma dz + J(\alpha, \psi)dN_h,$$

where λ determines the impact of a diffusion shock to the interest rate on the pricing kernel, and λ_J determines the impact of a jump on the pricing kernel. $J(\alpha, \psi)$ denotes the jump with exponentially distributed jump size and dN_h denotes Poisson arrivals. The parameters a, b and σ have the usual meaning as in short rate models. In this model, the diffusion component of interest rates displays constant conditional variance and jumps are infrequent events that displace interest rates by discrete amounts but do not change their central tendency.

The second model is closely related to the work of Constantinides (1992), and is based on the pricing kernel of the form

$$m(t) = \exp[-y(t) - X(t)],$$

with

$$\begin{aligned}dX &= (x + \lambda^2/2)dt + \lambda dz, \\dx &= a(b - x)dt + \sigma dz, \\y(t) &= y(0) \left[1 - h\mu t + \sum_{j=1}^{N(t)} J_j \right].\end{aligned}$$

In the second model, jumps change the conditional central tendency of interest rates. This model captures the notion that interest rates oscillate smoothly around a central tendency that may change infrequently. In the absence of jumps, this model reduces to the Ornstein-Uhlenbeck model of Vasicek (1977). These models provide more realistic statistical behaviour of interest rates. However, from a practitioner's point of view these models do not fit initial term structures of interest rate and volatilities, and thus are of little interest¹³. In both models, Das and Foresi (1996), are able to derive exact solutions for the prices of pure discount bonds and tractable expressions for bond option prices.

Rogers (1997b)¹⁴ As discussed above, to obtain an arbitrage free model with positive interest rates one needs to specify some strictly positive supermartingale under some measure equivalent to the objective measure \mathbb{P} . Rogers (1997b) utilises this observation and using the theory of Markov processes provides a framework for creating strictly positive supermartingales, thus arbitrage-free term structure models. He starts from the identity

$$P(t, T) = \mathbb{E}_{\mathbb{P}^*} \left[\frac{\beta_t}{\beta_T} | \mathcal{F}_t \right] = \mathbb{E}_{\mathbb{P}_K} \left[\frac{K_T}{K_t} | \mathcal{F}_t \right],$$

where β_t is the savings account and

$$K_t \equiv \frac{1}{\beta_t} \frac{d\mathbb{P}^*}{d\mathbb{P}} \Big|_t = \frac{1}{\beta_t} Z_t.$$

¹³However, it may be possible to extend these models to fit the initial term structure.

¹⁴See also Rogers (1997a).

The economic condition $B(0, t) \equiv \mathbb{E}K_t \rightarrow 0$ as $t \rightarrow \infty$ allows us to identify K as what is known as a *potential*¹⁵.

Rogers (1997b) suggested to use a resolvent of some Markov process to define a potential. In particular, for some bounded, measurable $g : \mathcal{S} \rightarrow [0, \infty)$ he defines the pricing kernel by

$$K_t = \frac{e^{-\alpha t} R_\alpha g(X_t)}{R_\alpha g(X_0)}.$$

I.e. the pricing kernel is a product of the resolvent of some process X_t with function $\exp(-\alpha t)$, scaled by a constant. Next we explain, why the process K_t is a supermartingale. Because of the tower property of conditional expectation the process

$$M_t \equiv \mathbb{E} \left[\int_0^\infty e^{-\alpha s} g(X_s) ds \mid \mathcal{F}_t \right]$$

is a martingale. Then we have

$$\begin{aligned} M_t &= \mathbb{E} \left[\int_0^t e^{-\alpha s} g(X_s) ds \mid \mathcal{F}_t \right] + \mathbb{E} \left[\int_t^\infty e^{-\alpha s} g(X_s) ds \mid \mathcal{F}_t \right] \\ &= \int_0^t e^{-\alpha s} g(X_s) ds + e^{-\alpha t} \mathbb{E}^{X_t} \left[\int_0^\infty e^{-\alpha y} g(X_y) dy \right] \\ &= \int_0^t e^{-\alpha s} g(X_s) ds + e^{-\alpha t} R_\alpha g(X_t). \end{aligned}$$

We observe that process $A \equiv \int_0^t e^{-\alpha s} g(X_s) ds$ is positive and increasing. We can express the pricing kernel as the difference between a martingale and positive increasing process,

$$M_t - A_t = e^{-\alpha t} R_\alpha g(X_t) = R_\alpha g(X_0) K_t. \quad (2.2.8)$$

Thus the pricing kernel K_t is a supermartingale. Since g is assumed to be non-negative and the resolvent is a positive operator, K_t is positive.

Next we identify the short rate process in the potential framework. Applying Itô formula to process $K_t = \frac{1}{\beta} Z_t$, we obtain dynamics

$$dK_t = K_t(-r_t dt + Z_t^{-1} dZ_t), \quad (2.2.9)$$

¹⁵cf: Karatzas and Shreve (1991), Definition 1.3.17.

We use the SDE (2.2.9) as a guide. The differential form of (2.2.8) is given by

$$dK_t = \frac{1}{R_\alpha g(X_0)}(dM_t - dA_t).$$

Multiplying the inverse of K_t with the above equation, and rearranging it, we obtain

$$K^{-1}dK_t = \frac{K^{-1}}{R_\alpha g(X_0)}dM_t - \frac{g(X_t)}{R_\alpha g(X_t)}dt. \quad (2.2.10)$$

Applying the Itô formula to the expression $Z_t \equiv K_t \exp\left(\int_0^t \frac{g(X_s)}{R_\alpha g(X_s)}ds\right)$ we obtain

$$dZ_t = \frac{\exp\left(\int_0^t \frac{g(X_s)}{R_\alpha g(X_s)}ds\right)}{R_\alpha g(X_0)}dM_t. \quad (2.2.11)$$

The properties of the stochastic integral (cf: Karatzas and Shreve(1991) p.147) show (2.2.11) to be a local martingale. We substitute the dM_t in (2.2.10) using (2.2.11), and obtain

$$K^{-1}dK_t = \frac{1}{K_t \exp\left(\int_0^t \frac{g(X_s)}{R_\alpha g(X_s)}ds\right)}dZ_t - \frac{g(X_t)}{R_\alpha g(X_t)}dt \quad (2.2.12)$$

If we compare the above SDE with the one in (2.2.9), we can identify the short rate process in potential framework as

$$r_t \equiv \frac{g(X_t)}{R_\alpha g(X_t)}. \quad (2.2.13)$$

Thus, if we model the pricing kernel process relative to the probability space of the Markov process X , the spot rate process is given by (2.2.13).

When it comes to applying the above concept, it may be not so simple to give the resolvent of a given Markov process in a closed form. In order to circumvent this problem, we make use of the inverse relation between the resolvent R_λ and the generator A . Instead of specifying the resolvent directly we take some positive function $f : \mathcal{S} \rightarrow (0, \infty)$ which will turn out to be the resolvent, and take the inverse operator

$$(\alpha - A)f = g.$$

If we apply the resolvent R_α to g then we have a closed form expression for $R_\alpha g = f$. Using this idea we can express the spot rate process as

$$r_t \equiv \frac{g(X_t)}{R_\alpha g(X_t)} = \frac{(\alpha - A)f(X_t)}{f(X_t)} \quad (2.2.14)$$

To compute the last expression in (2.2.14) is far more straightforward than evaluating the resolvent itself.

To summarise, in order to create a term structure model we can use the following scheme:

1. Fix a Markov Process X , a number α and a nonnegative function f .
2. Define g by $g \equiv (\alpha - A)f$
3. Restrict parameters so that g is nonnegative.
4. Obtain $f \equiv R_\alpha g$. Then the short rate is given by $r_t \equiv \frac{(\alpha - A)f(X_t)}{f(X_t)}$.

This procedure works only for functions in the range of the operator A . If the underlying Markov process X is a diffusion, then A is well defined for the class $C_0^2(\mathbb{R}^n)^{16}$.

Since the initial publication of these ideas, their development has remained rather limited in the research community. This might be because of the relatively high technical barrier one needs to overcome to access this methodology. Some empirical research has been conducted by Rogers and Zane (1996). They fit a two-factor potential model to yield curve data in the US and the UK and to the exchange rates between them. They used a filtering approach and obtained a reasonably good though not perfect fit. Jalali and Kazemi (1997) and Yousaf (2001) use a Markov Chain instead of continuous time process to construct the pricing kernels. Their results indicate that one can obtain a reasonable fit to the initial term structure

¹⁶cf: Øksendal (1995), p. 113.

in the finite-state framework. Unfortunately they stop short of investigation of the options market, where the real application of these methodologies should be tested.

Flesaker and Hughston (1996a) and (1996b) suggested an alternative way to model the term structure. For a given family of positive martingales under some measure \mathbb{Q} equivalent to \mathbb{P} , they specified a term structure of the form

$$D_{tT} = \frac{\int_T^\infty M_{ts} ds}{\int_t^\infty M_{ts} ds}. \quad (2.2.15)$$

In order (2.2.15) be well-defined, we have to assume that the family M is jointly measurable and the integrals are (a.s.) finite. Flesaker and Hughston showed that in this framework interest rates are positive. As Rogers (1997b) and Rutkowski (1997b) observe the Flesaker and Hughston specification is equivalent to

$$K_t = E[A_\infty | \mathcal{F}_t] - A_t, \quad (2.2.16)$$

where $A_t = \int_0^t M_{ss} ds$. This is just another way of representing a potential¹⁷.

A well known special case of the Flesaker and Hughston (1996a) approach called the *rational log-normal* model is given by the pricing kernel

$$K_t = f(t) + g(t)M_t, \quad \forall t \in [0, T'], \quad (2.2.17)$$

where $f, g : [0, T'] \rightarrow \mathbb{R}_+$ are strictly positive decreasing functions, and M is a strictly-positive martingale, with $M_0 = 1$. It follows then that the bond price process is given by

$$B(t, T) = \frac{f(T) + g(T)M_t}{f(t) + g(t)M_t}, \quad \forall t \in [0, T],$$

for any maturity T . In particular the initial term structure is given by

$$B(0, T) = \frac{f(T) + g(T)}{f(0) + g(0)}, \quad \forall T \in [0, T'],$$

¹⁷cf: Protter (1990).

which enables the fit of the initial term structure by construction. If, in addition, one assumes that the martingale M is log-normal, i.e. $dM_t = \sigma_t M_t dW_t$ for some deterministic function σ_t , then prices of caplets and swaptions can be readily computed.

Goldberg (1998) analysed the one-factor rational log-normal model, her findings are that this model has undesirable feature of predicting that the asymptotic value of the short rate volatility is zero¹⁸.

Burnetas and Ritchken (1997) extend the rational log-normal model of Flesaker and Hughston (1996a) by considering a pricing kernel of the form

$$K_t = f(t) + g(t)M_t,$$

where functions $f(t)$ and $g(t)$ are as in (2.2.17). The positive martingale $M(t)$ is given by

$$dM_t/M_t = h(1 - \hat{J})dt + \sigma dW_t + (J - 1)d\pi_t \quad (2.2.18)$$

with $M_0 = 1$ and $d\pi_t$ is an independent Poisson increment with density h . $J - 1$ is the random percentage jump conditional on jump's occurring, with a time invariant distribution. The expected value of J is \hat{J} . The process M_t is a martingale and thus K_t is a potential. As in the Flesaker and Hughston (1996a) approach, the initial term structure is fitted by construction. They derive analytical solutions for European-type options and show that, due to the jump term in (2.2.18), the yield curve distribution is more flexible then the one in the rational log-normal model.

Jin and Glasserman (2001), consider again a potential of the form

$$K_t = E[A_\infty | \mathcal{F}_t] - A_t. \quad \forall 0 \leq t \leq \infty, \quad (2.2.19)$$

¹⁸As pointed out by Rebonato (1996), a similar effect can be observed if one tries to induce decorrelation among forward rates using volatility of the short rate.

To obtain a valid potential, they model process A_t directly. In particular, they assume that there exists an \mathcal{F}_t -adapted positive continuous process $-\mu_t$ such that

$$A_t \equiv - \int_0^t \mu_s ds.$$

Under some technical conditions needed to justify representation (2.2.19), the pricing kernel is defined as in (2.2.19). If one drops the assumption that $-\mu_t$ be positive then the pricing kernel defined by (2.2.19) is a positive semimartingale, assuming that $E[A_\infty | \mathcal{F}_t]$ exists and is bigger than A_t . In this case we arrive at the general pricing kernel definition. Jin and Glasserman systematically analyse relations between the HJM approach with pricing kernel and the Flesaker and Hughtson framework.

Markov-Functional Models As we discussed above, the specification of a pricing kernel is equivalent to specifying a numéraire. In several papers, Hunt *et al.* (2000), Hunt and Kennedy (1998b), and Hunt *et al.* (1996) construct several term structure models. The choice of the numéraire is motivated by the product to be priced. Furthermore one assumes that this numéraire has a functional dependence on some low-dimensional Markov process. The functional form is then chosen so that the model prices some liquid instruments in the market. Similar ideas were independently developed by Balland and Hughston (2000).

2.2.2 Short Rate Models

In this section we review the short rate (SR) approach to the term structure modelling. Historically, this methodology is the oldest and there is a vast amount of literature which explores this approach. These models include Vasicek (1997), Cox *et al.* (1985b), Black *et al.* (1990), Duffie and Kan (1994), etc. By a short rate model we mean a model such that the zero-coupon bond can be represented as

$$B(t, T) = \mathbb{E}_{\mathbb{P}^*} \left[\exp \left(- \int_t^T r_u du \right) \mid \mathcal{F}_t \right], \quad (2.2.20)$$

where \mathbb{P}^* is the risk-neutral measure equivalent to \mathbb{P} , and r_t is a short rate process. In the rest of this section we first present the short rate model under the objective

measure. This specification has been historically the first. Then the model specification under the risk-neutral measure. This is more common in practice. We then describe several classes of short rate models, and mentioned their advantages and disadvantages. Furthermore, we discuss several issues relating to calibration, statistical fitting, and Markov properties of the short rate models. Finally, we relate the short rate framework, to the more general pricing kernel approach, and present an example of a model which is a pricing kernel but not a short rate model.

To construct a short rate model, traditionally one would specify the dynamics of the short rate process under the objective measure \mathbb{P} , say by means of an Itô process,

$$dr_t = \mu_t dt + \sigma_t dW_t, \quad (2.2.21)$$

where μ and σ are adapted stochastic processes, satisfying suitable conditions to ensure the process (2.2.21) is well defined. Then one needs to make an assumption about the market price of risk process λ that is used to define a martingale measure by means of Doléans exponential,

$$\frac{d\mathbb{P}^*}{d\mathbb{P}}|_t = \exp \left(\int_0^t \lambda_u dW_u - \int_0^t |\lambda_u|^2 du \right). \quad (2.2.22)$$

Application of Girsanov's theorem gives us a process for the Brownian motion under the martingale measure. To evaluate the bond price, we simply need to evaluate the expectation operator in (2.2.20), or alternatively, using results of Feynman-Kac, solve the associated parabolic PDE.

In practice, however, one usually, starts directly with short rate dynamics under the martingale measure \mathbb{P} , which renders the change of measure unnecessary. Should one want to calibrate the parameters of the short rate process to the historical data, then again we would need to change the underlying martingale measure to the objective \mathbb{P} . We demonstrate this idea on a very simple example, described originally by Vasicek (1977). The dynamics of the short rate process under the

martingale measure \mathbb{P}^* are given by

$$dr_t = k(\theta - r_t)dt + \sigma dW_t^*, \quad (2.2.23)$$

with r_0 , k , θ and σ positive constants. The process (2.2.23) has Gaussian distribution and can be integrated explicitly. Furthermore, it is relatively easy to compute the conditional expectation in (2.2.20), and to obtain analytical solutions for the bond prices $B(t, T)$. This analytical tractability explains the wide use of Gaussian processes in term structure modelling. Closed-form solutions for options can be obtained just as easily by using a forward measure technique, as was shown by Jamshidian (1989), who pioneered this approach. Next, we assume that the marked price of risk has the form¹⁹ $\lambda_t = \lambda r_t$. The objective measure can be defined just as in (2.2.22) and the new dynamics of r under the objective measure \mathbb{P} are given by

$$dr_t = (k\theta - (k + \lambda\sigma)r_t)dt + \sigma dW_t. \quad (2.2.24)$$

as we see, the structure of the SDE for the short rate in (2.2.24) has not changed. This observation motivates this particular choice of the market price of risk.

There is a large number of short rate models, and we will not be able to give even a brief description of them. Thus, we just sketch some model classes and outline motivations behind their development. Most of the models have their multi-factor version, so we will not go into the details of this aspect. These classes are obtained by imposing restrictions on the drift and diffusion parameters of the short rate process.

As usual, because of analytical tractability, *Gaussian* processes were a common choice for the dynamics of the short rate. Early examples include, Vasicek (1977), Langetieg (1980). To prevent interest rates going negative, *log-normal* dynamics are specified as in Black *et al.* (1990), Black and Karasinski (1991), Jamshidian (1991a). To evaluate path-dependent derivatives, the use of lattices has become a popular

¹⁹Note this is not the only possible specification. The process of the form $\lambda_t = \lambda$, i.e. constant, will have the same effect.

tool. One can think of a lattice as a Markov Chain approximation to some stochastic process, see James and Webber (2000) for a thorough discussion of the lattice models. Evaluation on lattices can be sped up by using Forward Induction (introduced by Jamshidian (1991b)), and Brownian Bridge technique (Gandhi and Hunt (1997)), together with alternative sampling constructions (McCarthy and Webber (1999)).

Non-linear models usually include those with non-linearities in the drift term, such as Longstaff (1989). Many of them exhibit pathologies, though some are quite tractable, such as in Jamshidian (1995), Ahn and Gao (1999), and Ahn *et al.* (2002). The *affine* class of short rate models is probably the most popular. First described by Duffie and Kan²⁰ (1994), who developed a general theory, it was further investigated by Dai and Singleton (2000). This class of models is very flexible and tractable, and provides analytical solutions to a wide range of options. To explain the deviation from log-normality/normality of the yield returns, researchers include jumps into short rate dynamics, as in Ahn and Thompson (1988), and Das (1999b).

One issue which arises in terms structure modelling is parameter estimation versus calibration. A number of methods have been developed for estimating the parameter of continuous-time Markov processes, such as used in term structure modelling. However, when we use parameters estimated through statistical procedures, the model prices do not fit observed bond and derivatives prices. Thus, in practice derivative pricing people tend to use a procedure called *calibration*. This consists of finding model parameters that fit liquid market prices. This discrepancy is, of course, not satisfactory, and is probably due to misspecification of the models. There exists a considerably large body of literature about statistical parameter estimation of term structure models. We mention just a few important works. *Nonparametric density matching* has been considered by Ait-Sahalia (1996b), and Stanton (1997). A more standard method of *Generalised Method of Moments* have been applied by Chan *et al.* (1992). *Efficient Method of Moments* has been used

²⁰see also Duffie and Kan (1996).

by Duffie and Singleton (1993). *Maximum Likelihood* methods have been applied by Brown and Dybvig (1986). *Approximate and simulated maximum-likelihood* can be found in Pedersen (1995), and simulated likelihood estimation in Brandt and Santa-Clara (1999). Recently, *filtering approach* become popular, see Lund (1997).

Even within the class of factor models, where the term structure model does not intrinsically imply non-Markovian dynamics, we can nevertheless end up with a non-Markovian process for the short rate. Suppose that we stick to the common practise of writing a low-dimensional dynamic model, for instance a one-factor model for the short rate, when in fact the short rate may be a Markov process, but only as part of a larger system. In typical examples, the system may include yields at longer maturities, as in Brennan and Schwartz (1979):

$$\begin{aligned} dr_t &= \mu(r_t, l_t)dt + \sigma(r_t)dZ_t \\ dl_t &= m(r_t, l_t)dt + s(l_t)dW_t \end{aligned} \tag{2.2.25}$$

where l_t is the yield on a consol bond. In other situations, the short rate is Markovian only when taken with its instantaneous stochastic volatility and/or mean. In all these cases, the short rate process is not individually Markovian.

The question of whether interest rates follow continuous time Markov processes, specifically diffusions, has been addressed by Ait-Sahalia (1997). Here, the author finds, that neither the short rate nor the long rate can be characterised individually as Markov processes, though jointly they form a Markovian system. The slope of the yield curve is a univariate Markov process, even a diffusion. These findings appear to be sensitive to the choice of the dataset.

Next we examine the relation of SR models to the PK framework. If we denote by ζ_t the reciprocal of the bank account β , we can rewrite (2.2.20) as

$$B(t, T) = \frac{\mathbb{E}_{\mathbb{P}} \left[\zeta_T \frac{d\mathbb{P}^*}{d\mathbb{P}} \middle| \mathcal{F}_T \mid \mathcal{F}_t \right]}{\zeta_t \frac{d\mathbb{P}^*}{d\mathbb{P}} \middle| \mathcal{F}_t} = \frac{\mathbb{E}_{\mathbb{P}} [K_T \mid \mathcal{F}_t]}{K_t}, \tag{2.2.26}$$

where the pricing kernel $K_t = \zeta_t \frac{d\mathbb{P}^*}{d\mathbb{P}}|_{\mathcal{F}_t}$ is a product of a positive process of finite variation and a positive martingale, thus a positive continuous semimartingale. The likelihood process $\frac{d\mathbb{P}^*}{d\mathbb{P}}|_{\mathcal{F}_t}$ is well defined by the assumption that $\mathbb{P}^* \sim \mathbb{P}$. This shows that $SR \subseteq PK$, i.e. the class of short rate models is a subset of the class of pricing kernel models. Next, we provide an example showing that this inclusion is strict.

Example 2.2.1. (SR \neq PK). This example comes from Hunt and Kennedy (2000). It introduces a PK model which fails to be an SR model. We define the pricing kernel as a solution of the stochastic differential equation

$$dK_t = K_t^2 dW_t \quad (2.2.27)$$

with initial condition $K_0 = 1$. The solution of this equation is given by $K_t = R_t^{-1}$, where R_t is the Bessel (3) process defined as

$$R_t = ((W_t^{(1)})^2 + (W_t^{(2)})^2 + (W_t^{(3)})^2)^{\frac{1}{2}}$$

with initial conditions $W_0 = ((W_0^{(1)}), (W_0^{(2)}), (W_0^{(3)})) = (1, 0, 0)$, where $W_i(t)$ are independent Brownian motions²¹. The process (2.2.27) is a positive semimartingale, so it defines a valid pricing kernel. Furthermore, it can be shown that K_t is a local martingale but not a proper martingale. Next we show that this is not an SR model. We note that any short rate model can be represented by

$$B(t, T) = \mathbb{E} \left[\exp \left(- \int_t^T r_u du \right) \mid \mathcal{F}_t \right] = K_t^{-1} \mathbb{E}[K_T \mid \mathcal{F}_t],$$

where K_t is a process of finite variation. Thus if the model were an SR model we would be able to find a strictly positive martingale

$$\rho^{-1} = \frac{d\mathbb{P}}{d\mathbb{P}^*}$$

such that $K_t \rho_t^{-1} = \zeta_t$ is of finite variation. It would follow then that $K_t = \rho_t \zeta_t$. Taking the logarithm of both sides we obtain $\log K_t = \log \rho_t + \log \zeta_t$. According to

²¹cf. Karatzas and Shreve (1991), Section 3.3.C.

Meyer the decomposition is unique (cf: Rogers and Williams (1987), Section VI.29) It follows that $\log(\zeta)$ is a constant. Thus K_t is a martingale, which contradicts our result that K_t is a local martingale but not a martingale.

In this section we reviewed the short rate approach to the term structure modelling. We first presented the historical short rate model specification under the objective measure. Then, the model specification under the risk-neutral measure which is more common in practice. We described several classes of short rate models, and mentioned their advantages and disadvantages. Furthermore, we discussed calibration versus statistical fitting issues, and very important Markov properties of the short rate models. Finally, we related the short rate framework, to the more general pricing kernel approach, and presented an example of a model which is a pricing kernel but not a short rate model.

2.2.3 HJM Approach

In this section we introduce the HJM framework²², and relate it to the PK class of models. The idea of Heath, Jarrow and Morton²³ (1992) was to model the uncertainty of the whole yield curve and not just at a single point on it, as modelled by the SR approach. If one assumes that the discount curve $B(t, T)$ is almost everywhere differentiable in the maturity parameter T , and if it is absolutely continuous, then it can be represented as

$$B(t, T) = \exp \left(- \int_t^T f_{tu} du \right), \quad (2.2.28)$$

where

$$f_{tT} = - \frac{\partial \log P(t, T)}{\partial T}. \quad (2.2.29)$$

²²The predecessor of the HJM approach is a discrete time model developed by Ho and Lee (1986). Its continuous properties have been investigated by Sommer 1996. It has the form $df(t, T) = \sigma^2(T - t)dt + \sigma dW_t^*$ with $\sigma(t, T) = \sigma$ constant and the dimension of the underlying Brownian motion equals one.

²³See also earlier work in Heath *et al.* (1990a), Heath *et al.* (1990b).

has the interpretation of the instantaneous forward rate, i.e. the rate for investing at time T for an infinitesimally short time as seen from time t . Heath, Jarrow and Morton directly specified the dynamics of (2.2.29) in terms of a family of stochastic differential equations. For a given T the SDE for the $f(t, T)$ is given by

$$f(t, T) = f(0, T) + \int_0^t \alpha(u, T) du + \int_0^t \sigma(u, T) \cdot dW_u, \quad \forall t \in [0, T], \quad (2.2.30)$$

with $f(\cdot, \cdot)$, $\alpha(\cdot, \cdot)$, and $\sigma(\cdot, \cdot)$ satisfying some technical conditions, which ensures that the solution of SDE (2.2.30) exists and is unique²⁴. The dynamics of the bond process under the physical probability measure \mathbb{P} are given by

$$dP(t, T) = a(t, T)P(t, T)dt + b(t, T)P(t, T) \cdot dW_t,$$

where

$$\begin{aligned} a(t, T) &= f(t, t) - \alpha^*(t, T) + \frac{1}{2}|\sigma^*(t, T)|^2, \quad b(t, T) = -\sigma^*(t, T) \\ \alpha^*(t, T) &= \int_t^T \alpha(t, u) du, \quad \sigma^*(t, T) = \int_t^T \sigma(t, u) du. \end{aligned}$$

To exclude arbitrage across bonds of all maturities among the set of admissible strategies we need to assume the existence of a martingale measure together with some numéraire. We can choose the savings account as numéraire, in this case the equivalent martingale measure, is the risk neutral measure \mathbb{P}^* . Thus we assume there exists an adapted \mathbb{R}^d -valued process λ_t such that

$$\alpha^*(t, T) = \frac{1}{2}|\sigma^*(t, T)|^2 - \sigma^*(t, T) \cdot \lambda_t.$$

The process λ_t subject to some technical conditions defines the risk-neutral measure \mathbb{P}^* . The Brownian motion under the risk-neutral and forward martingale measure takes the form

$$W_t^* = W_t - \int_0^t \lambda_u du.$$

²⁴cf: Heath *et al.* (1992), C1.

The process λ_t is sometimes referred to as the *market price of risk*. The dynamics of the bond price process under the risk-neutral measure \mathbb{P}^* are then given by

$$dP(t, T) = r_t P(t, T) dt - \sigma_t^*(t, T) P(t, T) \cdot dW_t^*,$$

with the dynamics of the forward rates $f(t, T)$,

$$df(t, T) = \sigma_t(t, T) \cdot \sigma_t^*(t, T) dt + \sigma_t(t, T) \cdot dW_t^*.$$

To specify an HJM model one only needs the initial term structure $P(0, T)$ and the volatility function $\sigma(t, T)$. Several volatility specifications have been considered by Amin and Morton (1994). The main problem of HJM methodology is that the evolution of the term structure, which is infinite-dimensional, may not be Markovian with respect to a finite-dimensional state space.

Non-Markovian models are, in practice extremely cumbersome, and often intractable. For example, when pricing derivatives securities with binomial trees, non-Markovian state variables make the tree non-recombining, or “bushy”. The number of nodes in non-recombining trees grows exponentially with the number of time periods, as opposed to linearly. This problem has been addressed by Brace (1996a) who constructs non-bushy trees. Li *et al.* (1995a), construct a lattice where the forward rate is driven by a two-dimensional Markov process. The problem applies to Monte-Carlo simulation as well, here one needs to keep track of a potentially infinite number of points on the forward rate curve. As a result, substantial effort has been devoted to finding restrictions on the volatilities $\nu(\tau, t)$, possibly by extending the state space, such that Heath, Jarrow and Morton model generates Markovian dynamics. For a discussion of necessary and sufficient conditions see Carverhill (1994), Jeffrey (1995), Ritchken and Sankarasubrama (1995), and recent work by Bjork and Svensson (1999).

In the general HJM methodology interest rates can go negative. Sufficient conditions for positive interest rates are discussed in Miltersen (1999a) and Jin and

Glasserman (2001). Extensions of HJM to jump diffusions have been considered by Das (1999a).

Next, we relate the HJM approach to PK. We assume that we are given the volatilities $\sigma(t, T)$ of instantaneous forward rates and the market price of risk λ_t . We define the market price of risk $h_{t,T} = \lambda_t + \int_t^T \sigma(t, \tau) d\tau$. It can be shown that this market price of risk defines the forward martingale measure \mathbb{P}^T . Using Bayes' theorem one can then show that the likelihood ratio process $\frac{d\mathbb{P}^T}{d\mathbb{P}}|_{\mathcal{F}_t}$ that defines the forward measure can be expressed as

$$\frac{d\mathbb{P}^T}{d\mathbb{P}}|_{\mathcal{F}_t} = \frac{\mathbb{E}(K_T|\mathcal{F}_t)}{B(0, T)}.$$

This together with the market price of risk process $h_{t,T}$ gives us the HJM kernel,

$$K_t = B(0, t) \exp \left(- \int_0^t h_{uT} \cdot dW_u - \frac{1}{2} \int_0^t |h_{uT}|^2 du \right).$$

Finally, we present an example which shows that short rate specification is more general than the HJM framework.

Example 2.2.2. (SR \neq HJM). This example comes from Hunt and Kennedy (2000). It displays an SR model which fails to be HJM. Define a function R_n via

$$R_n(t) = \begin{cases} 0, & \text{if } t < 1, \\ 2^n, & \text{if } t - 1 \in [q2^{-n}, (q+1)2^{-n}), q \text{ odd}, \\ -2^n, & \text{if } t - 1 \in [q2^{-n}, (q+1)2^{-n}), q \text{ even}. \end{cases}$$

Define a random variable κ by $\kappa = j$ if $|W_1| \in [L_j, L_{j+1})$. W is a one-dimensional Brownian motion on $(\Omega, \mathcal{F}_t, \mathcal{F}, \mathbb{P})$. The sequence L_j is defined recursively as $L_1 = 0$ and $2(N(L_{j+1}) - N(L_j)) = 2^{-j}$, N is the cumulative normal distribution function. We fix $r \geq 0$, $\alpha < 1$. Next define a process of finite variation ζ_t ,

$$\zeta_t = \exp(-rt) \left(1 + \alpha \int_1^{t \vee 1} R_\kappa(u) du \right).$$

Furthermore, for all $T > 0$, there exists some $\epsilon(T) > 0$ such that ζ_t is \mathcal{F}_T measurable for all $t \in [T, T + \epsilon(T)]$. Thus, if we write $\zeta_t = \exp(-\int_0^t r_u u)$, the short rate exists and is r_t . Thus the model is SR. It can be shown that ζ_t^+ and ζ_t^- are not integrable for $t > 1$ and nowhere differentiable with respect to maturity for $t < 1 \leq T$. This means, the limit from the right is different from the limit from the left, though both are finite. Thus the forward rates do not exist and the model is not HJM.

Market Models

Usually, market practice is to quote the prices of certain standard interest derivatives, such as caps and swaptions, in terms of their implied Black volatility. That is the market implicitly assumes that the rate in question follows a log-normal process with constant volatility. Then the price of an option is obtained from the standard Black formula. In the case of the cap market, this requires the assumption that all rates follow simultaneously log-normal processes. This was thought to be inconsistent with the absence of arbitrage, i.e. with the existence of a numéraire and an equivalent martingale measure. However, it turns out that it is possible for the family of LIBOR rates to follow log-normal process but under different measures. This idea underlines the creation of *market models* that implicitly assume the HJM framework. The first step in this direction was taken by Sandmann and Sondermann (1997), who modelled the discretely-compounded interest rate, rather than the continuously-compounded one. The LIBOR market models were then constructed by Sandmann *et al.* (1995) and Brace *et al.* (1997). Jamshidian (1997) used this idea to create a swap market model²⁵. Next, we sketch the construction of LIBOR market models. We start with a sequence of dates, $T_0 < T_1 < \dots < T_n$, and define the corresponding forward LIBOR rates as in (2.1.1). We work under

²⁵A detailed description of calibration procedures for market models, i.e. choice of correlation matrix and volatility functions, together with test results can be found in Brigo and Mercurio (2001a) and in Rebonato (2002).

the forward martingale measure \mathbb{P}_{T_n} and numéraire $B(t, T_n)$. We define

$$\hat{B}_t^i := \frac{B(t, T_i)}{B(t, T_n)}, \quad \Pi_t^i = \prod_{j=1}^i (1 + \delta L_t^j)$$

for $0 \leq i \leq n$. We assume $\Pi^0 = \hat{B}^i = 1$ and $L^{n+1} = 0$. Furthermore, we have the following relationship

$$\hat{B}_t^i = (1 + \delta L_t^{i+1}) \hat{B}_t^{i+1} = \prod_{j=1}^n (1 + \delta L_t^j) = \frac{\Pi_t^n}{\Pi_t^i}.$$

Assume now that the family of forward rates satisfies

$$dL_t^i = - \left(\sum_{j=i+1}^n \frac{\delta L_t^j}{1 + \delta L_t^j} \sigma_i^i(L_t) \sigma_t^j(L_t) \rho_{ij} \right) L_t^i dt + \sigma_t^i(L_t) L_t^i dW_t^i,$$

then the forward bond price is a martingale,

$$d\hat{B}_t^i = \hat{B}_t^i \sum_{j=i+1}^n \frac{\delta L_t^j}{1 + \delta L_t^j} \sigma_t^j(L_t) dW_t^j.$$

Moreover, we can define the forward martingale measure,

$$\frac{d\mathbb{P}_{T_{i+1}}}{d\mathbb{P}_{T_n}} = \mathcal{E} \left(- \int_0^t \sum_{j=i+1}^n \frac{\delta L_s^j}{1 + \delta L_s^j} \sigma_s^j(L_s) \rho_{ij} ds \right),$$

so that the forward LIBOR rates L_t^i follow a log-normal process under the forward measure $\mathbb{P}_{T_{i+1}}$,

$$dL_t^i = \sigma_t^i(L_t) L_t^i dW_t^i.$$

Thus, we have constructed an arbitrage free term structure model such that all forward LIBOR rates L_t^i follow log-normal processes. As a consequence, caps/floors can be priced consistently with the Black formula.

Unfortunately, like the general HJM framework, market models share the curse of dimensionality. It is a severe impediment to their use for pricing and hedging American type products, or any product that was not covered by the design of the model. A considerable amount of literature attempts to overcome this problem,

sometimes by clever use of simulation, sometimes by approximating the market model by some simpler drift. For example, Brace (1996b), and (1997), develops approximating formulae, for swaption valuation in the LFM, under the assumption of single and then several driving Brownian motions. Brace (1998) and Brace *et al.* (1998) develop simulation algorithms. All of these models admit arbitrage but they do yield curve distributions similar to those of the market model being approximated. Carr and Yang (1997) use Markov Chain Approximation to generate non-Markovian market models in order to price Bermudian interest rate derivatives, and Carr and Yang (1998) use Markov Chain Approximation to approximate the value of American bond options in a general multi-factor HJM framework.

It is common knowledge in the equity and interest rate markets that stocks or rates do not exhibit log-normal distributions under the equivalent martingale measure. Thus the implied volatility does not remain constant across strikes but exhibits dependence qualitatively described as the smile/smirk effect (depending on the shape of the implied volatility curve). Zühldorf (2000) considered extensions to the log-normal forward rate model by assuming quadratic and affine volatility functions. Andersen and Andreasen (1999) suggest the use of CEV process for the forward LIBOR rates, whereas Brigo and Mercurio (2000) use the log-normal mixture. It has to be noted that the extensions of market models to capture smile/smirk effects are not straightforward and quite complex.

2.2.4 Alternative Models and Term Structure Modelling Issues

Infinite-Factor Models

All of the models described above are driven by finitely many Brownian motions, so intrinsically, they are generated by finitely many factors. However, the number of traded bonds, as the number of different maturities, can be considered, at least in principle, infinite. To capture this observation, Kennedy (1994) models the term structure of interest rates as a random field. In particular he suggested a model

where the forward rates follow a continuous Gaussian random field so that they evolve as a continuous random surface. The model is consistent with the initial term structure. He derives a necessary and sufficient condition on the drift to ensure that the discounted prices of zero-coupon bonds are martingales, which allows the pricing of contingent claims.

In a similar work, recently, Goldstein (2000) , and Santa-Clara and Sornette (2001) propose a similar approach to modelling instantaneous forward rates. They simply shock the forward rate curve by a stochastic string which is a solution of some stochastic partial equation. Furthermore, they generalise Kennedy (1994) non-arbitrage condition on the drift to non-Gaussian random fields. The main innovation of this class of models consists in having each instantaneous forward rate driven by its own shock.

Random field models have very appealing properties. In a field model, each point along the term structure is a distinct random variable with its own dynamics. Each point, however, is correlated with the other points in the term structure. In contrast with finite-factor models, random field models are consistent with both the current yield curve and term structure innovation. Gaussian field models have the potential to fit exactly empirical covariance structure. Furthermore, a random field model predicts that a better hedging instrument for a bond is another of similar maturity.

However, to use random field models one needs to find a flexible family of covariance functions, thus a flexible class of field models, that can fit the empirical covariance matrix observed in the market. Functions from this class should be strictly positive definite, thus providing truly infinite factor structure for the model. Neither Kennedy (1994) , nor Goldstein (2000) , nor Santa-Clara and Sornette (2001) identified such a class. Without this class of covariance functions one cannot capitalise on the main advantage of the field models, i.e. capturing the inter-dynamics of movements in the term structure.

In this thesis we will construct Gaussian field models that can fit the sample covariance matrix observed in the market. In particular, in Chapter 3 we will describe the framework in which we will develop Gaussian field models. We will require that the covariance function of Gaussian field is strictly positive definite so that the evolution of the yield is driven by infinite-factor structure. Then in Chapters 4, 5, and 6 we will develop several techniques for constructing Gaussian field with above properties.

It has to be noted that within infinite-factor models one can consider alternative approach to the absence of arbitrage in which infinite portfolios of bonds are allowed. This ideas has been studied by Jacka *et al.* (1999), Bjork *et al.* (1997) and Bjork *et al.* (1997).

Models for Future Prices

Futures contracts are the most frequently traded and most basic securities of an interest rate market. In spite of this observation most interest rate models assume the forward rates as primary elements of the models. The process of futures prices and rates is therefore endogenous to these models. In addition hedging strategies are formulated in terms of forward and/or spot contracts and, to a lesser extent futures. Recently, this problem has attracted some attention in the literature. Heath (1998) models futures rates in the spirit of the HJM framework. He presents a modelling paradigm in which the interest rate futures contracts are taken as basic securities. He shows that his framework is as general as the SR and the HJM. Nielsen and Sandman (2000) use market models to analyse the implied futures price processes, and Stapleton and Subrahmanyam (2001) introduce a model in which any two future prices act as the state variables. It should be pointed out that modelling futures rates has its own advantages and disadvantages and does not seem to be superior to any other modelling framework.

Model Risk

Model risk is becoming an increasingly important concept not only in financial valuation but also for risk management issues and capital adequacy purposes. Model risk arises as a consequence of incorrect modelling, model identification or specification errors, and inadequate estimation procedures, as well as from the application of mathematical and statistical properties of financial models in imperfect financial markets. Measuring model risk is challenging, specifically in the domain of interest rates, where there exists a large number of products and incompatible models simultaneously. Gibson *et al.* (1999) addressed this problem within interest rate pricing and risk management frame work. In particular, Bossy, Gibson, Lhabitant, Pistre, and Talay (1999) analyse model risk for discount bond options within HJM framework. Main findings are that model risk is highly sensitive to the current level of interest rate volatility and the type of position held by the trader, and that it also increases with the time to maturity of the position held. This analysis was further extended by Akgun (2000) to the HJM model with jumps, as described in Jarrow and Madan (1995).

Buhler *et al.* (1995) analyse the empirical quality of one- and two-factor models of HJM type and one- and two-factor models proposed by Uhrig (1995), which is a generalisation of the model of Longstaff and Schwartz (1992a). The analysis was conducted according to estimation, fitting, and valuation problems, and empirical quality. In this respect, the two-factor generalised Longstaff and Schwartz model was superior to the alternative specifications. Recently, Dudenhausen *et al.* (1999) analysed the effect of model and parameter misspecification on the effectiveness of Gaussian hedging strategies withing Gaussian term structure models as well as market models. They argue in favour of the use of Black-type hedging strategies in the case of fixed income security instruments, when the natural hedging instruments are available.

The number of factors can have an important influence on pricing and hedging.

Longstaff *et al.* (2001b) claim that at least a four factor model should be used to avoid mis-pricing of Bermudian swaptions. These findings have been contradicted by Andersen and Andreasen (2001), who claim that one- or two-factor models are sufficient for accurate pricing of the instruments. Furthermore, Rebonato and Cooper (1995) point out that low factor models allow for only limited shapes of correlations between forward rates, which can lead to considerable mispricing. In the next section we discuss somewhat related issues on interest rate risk management.

2.3 Interest Rate Risk Management

Usually, interest rate derivatives are sold to clients over-the-counter by financial institutions or traders. If these instruments are traded at an exchange the financial institution or trader can hedge the position by buying the same instrument at the exchange. The resulting position is then *neutralised* or *covered*. However, most of the OTC instruments are tailored to the specific needs of the clients, and there is no equivalent exchange-traded instrument available for hedging. This significantly complicates the hedging process. The trader has to buy or sell financial instruments to create synthetically the same exposure as the derivative which he or the financial institution has sold to the client.

If both portfolios do not coincide (as is predominantly the case in practice), the trader is exposed to the interest rate risk. This comes from the fact that changes in the yield curve may have different effects on different products. A risk manager has to rely on term structure models to understand his possible exposure to adverse movements in the yield. These term structure models assist him with the process of structuring and restructuring the portfolio according to the desired exposure or level of risk. Furthermore, these models provide the risk manager with risk measures, i.e. numbers which summarise his risk exposure. These measures are important not only for the risk management process but also for many institutional investors who

are requested to produce risk reports for regulatory purposes.

In this chapter we review the main tools of interest rate risk management. We start, in Section 2.3.1, with the traditional, but still widely used, duration and convexity measures. Then, in Section 2.3, we consider risk measures associated with term structure pricing models, such as the short rate and HJM framework. We offer some empirical results and some further discussion in Section 2.3.3.

2.3.1 Traditional Measures of Interest Rate Risk: Duration and Convexity

Until recently, the managing of interest rate risk with bond portfolios has been largely limited to the use of conventional duration. Conventional duration measures an asset's sensitivity to interest rate change under the assumption of infinitesimal parallel shifts in the term structure. This idea goes back at least sixty years. The first attempt to quantify exposure in the fixed income market was done by Macaulay (1938) when he proposed a measure of “duration” to represent the “... essence of the time element of a loan”²⁶. The term “duration” might appear misleading as the prime interest, even of Macaulay, was the risk-proxying properties of this measure. Independently, Hicks (1939) proposed duration as a proxy of basis risk and this has been rediscovered by many later authors. He called it the “average period” measuring the elasticity, with respect to a discount ratio²⁷

Let us consider a bond with payment dates t_1, \dots, t_n , with $t_n = T$. We denote the payment at time t_i by c_i , and the time t value of the bond by $B^c(t, T)$. Furthermore, we define implicitly the yield of the bond y_t^B , using continuous compounding,

$$B^c(t, T) = \sum_{t_i > t} c_i e^{-y_t^B(t_i - t)}.$$

²⁶See Macaulay (1938), p.44.

²⁷(I.e. factor or discount rate + unity.) See Hicks (1939), p.186.

The Macaulay duration D_t^{Mac} of the bond is defined as

$$D_t^{\text{Mac}} = -\frac{1}{B^c(t, T)} \frac{dB^c(t, T)}{dy_t^B} = \frac{\sum_{t_i > t} (t_i - t) c_i e^{-y_t^B(t_i - t)}}{B^c(t, T)} = \sum_{t_i > t} (t_i - t) \omega^{\text{Mac}}(t, t_i),$$

where $\omega^{\text{Mac}}(t, t_i) = c_i e^{-y_t^B(t_i - t)} / B^c(t, T)$. The weights ω^{Mac} can be seen as ratios between the present value of the i 'th payments and the total present value of the bond. We can interpret the Macaulay duration as the sum of these weight scaled by their respective times to maturity. We can express the relative price change of the bond due to infinitesimal change in its yield as,

$$\frac{dB^c(t, T)}{B^c(t, T)} = -D_t^{\text{Mac}} dy_t^B.$$

The Macaulay duration is defined as a measure of the price change induced by an infinitesimal change in the yield of the bond. For a non-infinitesimal change in the yield, a first-order approximation gives that

$$\Delta B^c(t, T) \approx \frac{dB^c(t, T)}{dy_t^B} \Delta y_t^B,$$

and the relative price change of the bond

$$\frac{\Delta B^c(t, T)}{B^c(t, T)} \approx -D_t^{\text{Mac}} \Delta y_t^B.$$

An obvious way to obtain a better approximation is to include a second-order term,

$$\Delta B^c(t, T) \approx \frac{dB^c(t, T)}{dy_t^B} \Delta y_t^B + \frac{1}{2} \frac{d^2 B^c(t, T)}{d(y_t^B)^2} (\Delta y_t^B)^2.$$

Defining the Macaulay convexity by

$$K_t^{\text{Mac}} = \frac{1}{2B^c(t, T)} \frac{d^2 B^c(t, T)}{d(y_t^B)^2} (\Delta y_t^B)^2 = \frac{1}{2} \sum_{t_i > t} (t_i - t)^2 \omega^{\text{Mac}}(t, t_i),$$

we can write the second-order approximation as

$$\frac{\Delta B^c(t, T)}{B^c(t, T)} \approx -D_t^{\text{Mac}} \Delta y_t^B + K_t^{\text{Mac}} (\Delta y_t^B)^2.$$

The Macaulay measures are not directly informative of how the price of a bond is affected by a change in the zero-coupon yield curve and are therefore not a valid basis for comparing the interest rate risk of different bonds. The problem is that the Macaulay measures are defined in terms of the bond's own yield, and a given change in the zero-coupon yield curve will generally result in different changes in the yields of different bonds. It is easy to show ²⁸ that the changes in the yields of different bonds will be the same if and only if the zero-coupon yield curve is flat. In particular, the yield curve is only allowed to move by parallel shifts²⁹.

An alternative duration measure, based on the zero-coupon yield curve rather than the bond's own yield, has been discussed by Fisher and Weil (1971). They defined a duration measure, which we refer to as **Fisher-Weil duration**, by

$$D_t^{FW} = \sum_{t_i > t} (t_i - t) \omega(t, t_i), \quad (2.3.1)$$

where $\omega(t, t_i) = c_i e^{-y_t^{t_i}(t_i - t)} / B^c(t, T)$. Here, $y_t^{t_i}$ is the zero-coupon yield prevailing at time t for the period up to time t_i . Relative to the Macaulay duration, the weights are different. The $\omega(t, t_i)$ are computed using the true present value of the i 'th payment, since the payment is multiplied by the market discount factor $P_t^{t_i} = e^{-y_t^{t_i}(t_i - t)}$. In the weights used in the computation of the Macaulay measures the payments, are discounted using the yield of the bond.

If we think of the bond price as a function of the relevant zero-coupon yields, $y_t^{t_1}, \dots, y_t^{t_n}$,

$$B^c(t, T) = \sum_{t_i > t} c_i e^{-y_t^{t_i}(t_i - t)},$$

we can write the relative price change induced by an instantaneous change in the zero-coupon yields as

$$\frac{dB^c(t, T)}{B^c(t, T)} = \sum_{t_i > t} \frac{1}{B^c(t, T)} \frac{\partial B^c(t, T)}{\partial y_t^{t_i}} - \sum_{t_i > t} w(t, t_i) (t_i - t) dy_t^{t_i}.$$

²⁸ see Ingersoll *et al.* (1978)).

²⁹ As shown by Ingersoll *et al.* (1978) [138], such an assumption conflicts with the no-arbitrage principle.

If the changes in all the zero-coupon yields are identical, the relative price change is proportional to the Fisher-Weil duration. Consequently, the Fisher-Weil duration represents the price sensitivity toward infinitesimal parallel shifts of the zero-coupon yield curve.

We can also define the **Fisher-Weil** convexity as

$$K_t^{FW} = \frac{1}{2} \sum_{t_i > t} (t_i - t)^2 \omega(t, t_i).$$

The relative price change induced by a non-infinitesimal parallel shift of the yield curve can be then approximated by

$$\frac{\Delta B^c(t, T)}{B^c(t, T)} \approx -D_t^{FW} \Delta y_t^* + K_t^{FW} (\Delta y_t^*)^2,$$

where Δy_t^* is the common change in all the zero-coupon yields.

There is a large body of literature focusing on developing better measures of the interest rate risk associated with non-parallel yield curve shifts. First immunisation strategy which explicitly took non-flat term structures into account, was developed by Fisher and Weil (1971). Khang (1979) proposed logarithmic shifts. Several other authors³⁰ have proposed different duration measures for additive, multiplicative, logarithmic, and exponential shifts in the term structure, as well as various combinations of these shifting patterns³¹.

A major weakness of the preceding methodology is that movements in the entire term structure are described by one factor. Whether in a simple model of parallel shifts or in more sophisticated versions, the movement of any specific rate is assumed to predict perfectly movements of any other rate. In practice it is quite possible that short- and long-term zero rates can move in opposite directions, changing the whole slope of the zero-rate curve. Immunising with a model that assumes parallel shifts

³⁰See, for example, Kaufman *et al.* (1981), Kaufman *et al.* (1983b).

³¹Leibowitz, *et al.* (1988) and Klaffky *et al.* (1992) have invented "functional duration".

or some other strong relationship between yields of different maturities will fail to protect against changes in the yield curve, such as flattening, steepening, or other twists.

The general solution to this problem is to construct a model with more than one factor³². Some authors have considered duration measures based on movements of the term structure at different points. For example, Waldman (1992) introduced *partial duration* and Ho (1992) proposed *key rate duration*. These durations examine rather mechanically the sensitivity of fixed income instruments to a chosen set of rates. The choice of rates is ad hoc and is not based on any particular term structure model. In the next section we will look at duration measures arising from the dynamic term structure models.

2.3.2 Risk Measures in Pricing Models

Short Rate Framework

One can use martingale pricing theory to provide a mechanism to generate arbitrary (nonparallel) deformations of the yield curve³³. Similarly to the standard duration, we view the price of any zero-coupon bond as a function of a small number of certain state variables. We can assume, for example, that the short rate $r(t)$ is a Markov process. This implies that the arbitrage-free price at time t of a zero coupon bond with maturity T , $B(t, T)$, can be written as a deterministic function of $r(t)$, i.e.

$$B(t, T) = \phi(t, T, r(t)).$$

Therefore, we can define the duration of a coupon bond as

$$D(r, t) = -\frac{1}{B^c(t, T)} \frac{\partial B^c(t, T)}{\partial r}. \quad (2.3.2)$$

³²General multi-factor duration measures based on empirical studies have been developed by Gultekin and Rogalski (1984), Elton *et al.* (1988), and Elton *et al.* (1990).

³³For a discussion of possible shapes of the yield curve that can be generated within the short rate approach see Schlögel and Sommer (1994) [225] and (1997) [226]

The duration measures we defined in the previous section are measured in units of time and can be interpreted as “effective” time to maturity. The duration defined in (2.3.2) is not measured in time units, but it can be related to such a measure, as we describe next.

Boyle (1978) and, independently, Cox *et al.* (1979), were the first to consider the concept of duration within the context of modern day term structure modelling. Cox *et al.* (1979) They adopted a framework where the short rate follows a time-homogeneous univariate Markov diffusion (e.g. Vasicek (1977), Cox *et al.* (1985b)). This structure is enough to guarantee that at time t the bond price is a function of the short rate r , and time to maturity τ . Cox *et al.* (1979) define measure of duration for a coupon bond as “the [time to] maturity of the discount bond with the same [basis] risk”. The basis risk of a bond with price $B(t, T)$ is defined as $\frac{\partial B(t, T)}{\partial r} / B(t, T)$. That is, Cox *et al.* (1979) define the duration as a number $D^*(r, t)$ which solves,

$$\frac{1}{B^C(t, T)} \frac{\partial B^C(t, T)}{\partial r} = \frac{1}{B(t, t + D^*(r, t))} \frac{\partial B(t, t + D^*(r, t))}{\partial r}.$$

Cox *et al.* (1979) used the term **stochastic duration** for the duration $D^*(r, t)$ to indicate that this duration measure is based on the stochastic evolution of the term structure.

As an example, we calculate both the duration and the stochastic duration for the Vasicek (1977) model. In the time-homogeneous affine one-factor diffusion models, e.g. the Vasicek and CIR models, the zero-coupon bond prices can be written as

$$P(t, T) = e^{-a(T-t) + b(T-t)r}.$$

Consequently, the duration of a coupon bond is

$$D(r, t) = -\frac{1}{B^c(t, T)} \frac{\partial B^c(t, T)}{\partial r} = -\frac{1}{B^c(t, T)} \sum_{t_i > t} b(t_i - t) c_i P(t, t_i) = \sum_{t_i > t} \omega(r, t, t_i) b(t_i - t),$$

where $\omega(r, t, t_i) = c_i P(t, t_i) / B^c(t, T)$ is the i 'th payment's share of the total present value of the bond. Note that the duration of a zero-coupon bond maturing at time T is $b(T - t)$, which is different from $T - t$. The convexity can be computed as

$$K(r, t) = \frac{1}{2} \sum_{t_i > t} \omega(r, t, t_i) b(t_i - t)^2.$$

If b is invertible, we can write the time-denominated duration of a coupon bond explicitly as

$$D^* = b^{-1} \left(\sum_{t_i > t} \omega(r, t, t_i) b(t_i - t) \right).$$

As an example we consider the Vasicek model in which the b -function is given by,

$$b(\tau) = \frac{1}{\kappa} (1 - e^{-\kappa \tau}),$$

so that the duration of a coupon bond is

$$D(r, t) = \sum_{t_i > t} \omega(r, t, t_i) \frac{1}{\kappa} (1 - e^{-\kappa(t_i - t)}) = \frac{1}{\kappa} \left(1 - \sum_{t_i > t} \omega(r, t, t_i) e^{-\kappa(t_i - t)} \right).$$

We have

$$b^{-1}(y) = -\frac{1}{\kappa} \ln(1 - \kappa y),$$

and hence the time-denominated duration of a coupon bond is

$$D^*(r, t) = -\frac{1}{\kappa} \ln \left(\sum_{t_i > t} \omega(r, t, t_i) e^{-\kappa(t_i - t)} \right). \quad (2.3.3)$$

HJM Framework

Alternatively, one may generate yield curves using the HJM arbitrage-free framework, in which

$$B(t, T) = \exp \left(- \int_t^T f(t, u) du \right),$$

where $f(t, u)$ is the instantaneous continuously-compounded forward rate. All of the above concepts of duration have been related to the partial derivatives of the bond price with respect to the short rate, or possibly some other rates. The bond

price $B(t, T)$ is a deterministic function of some stochastic variables only under very restrictive assumptions on the volatility process. This is because the HJM models are naturally Markov in an infinite-dimensional function space, and mappings of infinite-dimensional Markov processes onto finite-dimensional subspaces are not in general Markov.

Within a general HJM framework, however, the bond price depends on the whole instantaneous forward curve. Thus, duration measures defined in terms of partial derivatives with respect to a finite number of state variables do not capture all the richness of the HJM framework, and consequently, the variation in the bond price. The question arises on how to think of duration and convexity in this general setting³⁴.

One could define the concept of duration in the HJM framework by considering the sensitivity of a bond price with respect to all possible interest rates. This would generalise the key-rate duration concept proposed by Ho (1992) which we discussed above. This will create an infinite-dimensional vector and it is not clear how it can be used in practical applications. Next, we discuss two possible generalisations of the duration concept that can be used in the general HJM setting.

In the general HJM framework, the bond value $B(t, T)$ is an \mathcal{F}_t -measurable functional of the Brownian motion $\{W_t; t \geq 0\}$ to which an expansion, into so called Wiener chaos, can be applied. Because Wiener chaos expansions of a Brownian functional can be regarded as the stochastic calculus analogue of Taylors series, duration and convexity can be associated with first and second Wiener chaos³⁵,

³⁴Au and Thurston (1995) defined duration within the HJM framework, under the assumption that the evolution of the term structure depends only on the Markovian short rate. More generally, Jarrow and Turnbull (1994) considered an HJM framework where the evolution of the term structure can be represented by a finite number of Markov state variables. They discussed the delta and gamma hedging of buckets that takes into consideration correlation between the buckets. Hedging within the general HJM model has been investigated by Musiela *et al.* (1993).

³⁵Similarly, Lacoste (1996) associates the delta and gamma of an index option with the first and second Wiener chaos of the option payoff.

respectively.

Next, we briefly discuss the Wiener chaos expansion³⁶. Suppose that W is a one-dimensional Brownian Motion, and define multiple stochastic integrals

$$\int_{\mathbb{S}^n} f(s_1, \dots, s_n) dW(s_1) \dots dW(s_n),$$

where f is a square integrable function on the simplex

$$\mathbb{S}^n = \{(s_1, \dots, s_n) \in \mathbb{R}^n; 0 < s_1 < s_2 < \dots < s_n\}.$$

The set C^n of all such multiple integrals is a closed subspace of $L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$ and is called the n^{th} *Wiener chaos*³⁷. Wiener chaoses of different order are orthogonal and their direct sum is equal to $L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$. Thus, every square-integrable random variable F can be uniquely represented as a sum of its *orthogonal projections* on the successive Wiener chaoses. In this case we can write

$$F = \mathbb{E}F + \sum_{n>0} \int_{\mathbb{S}^n} f_n(s_1, \dots, s_n) dW(s_1) \dots dW(s_n),$$

where the function f_n , square-integrable on \mathbb{S}^n is given by

$$f_n(s_1, \dots, s_n) = \lim_{\epsilon \downarrow 0} \epsilon^{-n} \mathbb{E}F(W(s_1 + \epsilon) - W(s_1)) \dots (W(s_n + \epsilon) - W(s_n)).$$

This is a decomposition of the random variable F into Wiener chaos.

Brace and Musiela (1997) use this technique to derive zero-, first-, and second-order orthogonal projections and suggest employing them for immunisation and hedging purposes. Consider the problem of portfolio immunisation. The aim is to set a portfolio of zero coupon bonds with maturities T_i and amounts a_i , $i = 1, \dots, p$ in such a way that the portfolio value at time T ,

$$V(T) = \sum_{i=1}^p a_i P(T, T_i), \tag{2.3.4}$$

³⁶Elliott and Van Der Hoek (1999) apply the Wiener chaos expansion to investigate the optimal bucketing problem in the case of the two-factor Hull and White model.

³⁷See Øksendal (1997) for more details on Wiener chaos.

will not be exposed to the main sources of uncertainty in this time horizon. The uncertainty at the time horizon T is modelled by the process $\{W_T(t); 0 \leq t \leq T\}$, which is a Brownian motion under the forward measure \mathbb{P}_T . We say that the portfolio has no n 'th exposure if the projection $Proj_{C_T^n} V(T)$ of the random variable $V(T)$ onto the n 'th-Wiener chaos C_T^n of the Brownian motion $W_T(\cdot)$ is zero. In other words, the orthogonal projection of $V(T)$ onto the n 'th-Wiener chaos of $W_T(\cdot)$ corresponds to the n 'th-order exposure of the portfolio at the time horizon T .

The expected value $\mathbb{E}_t B^C(t, T)$ is the projection onto the 0'th-order Wiener chaos. We have

$$\mathbb{E}_t B^C(t, T) = B(0, T)^{-1} \sum_{t_i > t} C_i B(0, t_i).$$

Furthermore, it can be shown, that the variance of the projection onto the first order Wiener chaos is zero if

$$\sum_{i=1}^p a_i B(0, T_i) \mathbb{E}_{T_i} \int_T^{T_i} \sigma_j(t, u) du = 0. \quad (2.3.5)$$

As an example, we want to immunise a liability c at a date T and we are not concerned with exposure of order greater than 1. The objective will be achieved if we choose a portfolio of zero coupon bonds with maturities T_i and amounts a_i , $i = 1, \dots, p$ so that the portfolio $\sum_{i=1}^p c_i P(T, t_i) - c$ has no exposure of orders 0 or 1. The projection onto 0'th-order Wiener chaos is zero if and only if

$$\sum_{i=1}^p c_i P(0, t_i) = c P(0, T), \quad (2.3.6)$$

or equivalently if the present value of the portfolio is equal to the present value of the liability. The projection onto the 1st-order Wiener chaos vanishes if (2.3.5) holds. If (2.3.5) and (2.3.6) hold we say that portfolio (2.3.4) has duration $T \equiv D^{\text{BM}}$.

In the case $\sigma(t, u) = \sigma$, a constant, expression (2.3.5) implies

$$D^{\text{BM}} = \frac{\sum_{i=1}^p c_i T_i B(0, T_i)}{\sum_{i=1}^p c_i B(0, T_i)}.$$

In this case duration coincides with the Fisher-Weil duration we defined in (2.3.1).

As another example we consider the volatility structure $\sigma(t, T) = \sigma e^{-\kappa(T-t)}$ for positive constants σ and κ . This volatility structure corresponds to the Vasicek (1977) model. Again, from expression (2.3.5), we obtain

$$D^{\text{BM}} = -\frac{1}{\kappa} \ln \left(\frac{\sum_{t_i > t} c_i B(t, t_i) e^{-\kappa(t_i - t)}}{\sum_{t_i > t} c_i B(t, t_i)} \right),$$

which is the same expression as the stochastic duration for the Vasicek (1977) model obtained in (2.3.3).

Another approach to defining a notion of duration in the HJM setting has recently been investigated by Jeffrey (2000). He defines the duration of a portfolio containing positive default-free cash flows c_i occurring at respective times t_i , as $D^J \equiv (T - t)$. The parameter T solves

$$\int_t^T \frac{\sigma(t, \nu)}{\psi(t)} d\nu = \frac{\sum_{t_i > t} c_i B(t, t_i) \int_t^{t_i} \frac{\sigma(t, \nu)}{\psi(t)} d\nu}{\sum_{t_i > t} c_i B(t, t_i)}, \quad (2.3.7)$$

where $\sigma(t, T)$ is the forward rate volatility, $\psi(t)$ the volatility of the chosen basis factor $x(t)$, and N the number of Brownian motions in the HJM dynamics. The factors $x(t)$ can be specific interest rates, or more generally a set a Brownian motions under any equivalent probability measure.

For example, consider the case when the forward volatility structure is constant $\sigma(t, T) = \sigma$. If we choose as the basis factor $x(t)$ the Brownian motion $W(t)$, with volatility $\psi(t) = 1$, the duration measure implied by (2.3.7) is

$$D^J = \frac{\sum_{t_i > t} c_i B(t, t_i) (T_i - t)}{\sum_{t_i > t} c_i B(t, t_i)},$$

which is the same expression as the Fisher-Weil duration we defined in (2.3.1).

Next, we consider the case when the forward volatility structure takes the form $\sigma(t, T) = \sigma e^{-\kappa(T-t)}$ for positive constants σ and κ . This volatility structure corresponds to the Vasicek (1977) model. If we choose as the basis factor $x(t)$ the

Brownian motion $W(t)$, with volatility $\psi(t) = 1$, the duration measure implied by (2.3.7) is

$$D^J = -\frac{1}{\kappa} \ln \left(1 - \frac{\sum_{t_i > t} c_i B(t, t_i) (1 - e^{-\kappa(t_i - t)})}{\sum_{t_i > t} c_i B(t, t_i)} \right),$$

which is the same expression as the stochastic duration for the Vasicek (1977) model obtained in (2.3.3). It can be easily shown, that the duration measure D^J with the Brownian motion $W(t)$ as a basis factor, and the duration D^{BM} implied by the Wiener chaos are the same.

2.3.3 Empirical Results and General Remarks

Most of the empirical tests of stochastic duration have not demonstrated any actual superiority to the simple Macaulay duration. Brennan and Schwartz (1983) discuss generalised duration within a two-factor equilibrium setting and compare it with standard definition of duration. They conclude that as long as one deals with simple bond positions, immunisation using standard duration is as good as the generalised version. However, they argue that the latter is better suited to portfolios which include options.

Munk (1999) showed analytically that for any bond the stochastic duration in the Vasicek model is smaller than the Fisher-Weil duration, and the same is the case with Macaulay duration. Recently, Wu (2000) compared empirically the performance of stochastic duration in Vasicek (1977), Cox *et al.* (1979), and Macaulay duration. He finds that Macaulay's duration outperforms the former. This is to be expected, as the hedge based on Macaulay duration is immune to the parallel shifts of the yield curve. The parallel shifts contribute the largest portion of the risk in the yield. This is an empirical observation, as the eigenvector corresponding to the largest eigenvalue in the PCA analysis of the yield differences is approximately flat.

Concepts such as duration give a risk manager rather limited information: they only indicate the sensitivities of his positions to small movements in the specified

risk factors. If we want to understand our risk exposure in a specific instrument or portfolio, we need to know its distribution of returns over a certain period of time. The distribution of returns describes probabilities of all possible outcomes. It can answer all questions about returns: it can be a forecast, or a summary of realised returns. To obtain a distribution of returns a risk manager has to rely on term structure models. Thus, it is of a paramount importance that a model be able to capture the main characteristics of the statistical behaviour of the yield curve.

The distribution of returns is too complicated and detailed in its entirety. Over the past decade other concepts have gained in popularity, such as *downside risk*, *shortfall probability*, and *Value-at-Risk*. These definitions of risk attempt to capture in a single number the essentials of the risk more fully described in the complete distribution. However, their accuracy again depends entirely on the underlying term structure model.

In this section we have reviewed the main tools of interest rate risk management. Unlike the term structure pricing literature, the interest rate literature is rather undeveloped. Whether any of the tools presently in use are superior in terms of hedging performance and risk forecasting, is a matter of further research. However, what can be said at this stage is that the better we can capture the empirical behaviour of the evolution of the term structure, the closer we might be to achieving accurate risk forecasts. In this thesis we develop random field models, which can capture exactly the inter-dynamics of interest rates. We believe that this methodology will help to understand risks involved in fixed income products.

2.4 Conclusions

In this chapter we have reviewed the term structure pricing and risk management literature. We started with the most general arbitrage-free framework, the pricing kernel. Then we specialised this framework to the less general classes of models:

the short rate models, models of the instantaneous forward rates, etc. We found that the pricing kernel framework is the least developed among the approaches to the term structure modelling. We also described term structure models that do not naturally fit into the above description, such as infinite-factor models and models for futures prices. Furthermore, we highlighted the advantages and disadvantages of different modelling approaches, as well as parameter estimation issues versus model calibration, and discussed model risk problems. We also discussed issues in fixed income risk management. We described several techniques currently used to measure interest rate risk. We also pointed out that the literature in this field is rather unsatisfying and it might benefit from the research in this thesis.

Infinite-factor models, or random field models, have very appealing properties. In contrast with finite-factor models, random field models are consistent with both the current yield curve and term structure innovation. Gaussian field models have the potential to fit exactly the empirical covariance structure. Furthermore, a random field model predicts that a better hedging instrument for a bond is another of similar maturity.

We have pointed out that in order to use random field models one needs to find a flexible family of covariance functions, and thus a flexible class of field models. The functions from this class should fit the empirical covariance matrix, and be strictly positive definite. This problem has not been addressed in the field literature. We will address it in the first part of this thesis. In particular, in Chapter 3, we will present a random field framework for the term structure, together with a detailed description of the random field literature. Then, in Chapters 4, 5, and 6 we will develop several techniques for constructing a Gaussian field with the above properties.

In our review, we found that the short rate approach remains the main tool among practitioners for pricing exotic derivatives. The main reason is that this framework allows for a simple Markov representation of the yield curve, so all calculations can

be done in real time. However, the calibration of short rate models to traded instruments is not trivial. Furthermore, we observed that market models are very popular among practitioners, as they allow instantaneous calibration to the instruments for which they were specifically designed. However, they suffer from non-Markov dynamics, so calibration to other instruments is a problem. Furthermore, American type options are difficult to handle with market models, and the extensions to smiles and skews are complex.

To address the above problems, in Chapter 7, we will develop a class of models within the pricing kernel framework. We will model the pricing kernel directly by approximating it with a set of radial basis functions. As the underlying noise in the economy, we will choose a simple multi-factor diffusion. This approach is just as simple in a one-factor as in a multi-factor setting. We will link the approximation of the kernel explicitly to the calibrating set of instruments. Thus, once the kernel is constructed, it will price correctly the chosen set of instruments and has a low-dimensional Markov structure. In summary, this class of kernel models is Markov by construction, easy to calibrate to a variety of instruments, can deal with American-type options, and is flexible in fitting to smiles and skews.

Chapter 3

Modelling Framework

3.1 Introduction

In the last chapter we described methodologies and reviewed literature for pricing and risk management of interest rates derivatives. In this chapter we identify a class of models for the term structure of interest rates, motivated by the following two observations: The number of traded maturities in the fixed income market, though always finite, is generally assumed to be infinite. That is, when modelling interest rates, one assumes the maturity span to be an interval in the real line. Secondly, it is reasonable to assume that each maturity will have its own source of randomness, however small. The logical conclusion from these observations is that a model which claims to explain the relation between different maturities of the term structure should have an infinite number of factors. Despite these observations most, if not all, of the academic literature and industry practice is based on finite-factor models.

Though approaches to modelling term structure have grown in sophistication and complexity over the past 30 years, they are still not able to capture exactly the interrelation between the rate dynamics. The earliest class of models used finite-factor processes with constant coefficients, and then defined the spot rate as a function of

these factors (see e.g., Vasicek (1977), etc¹). Unfortunately, models from this class are not even able to capture the current term structure.

The second class of models either assigned explicit time dependence to the spot rate process, or modelled the forward rate directly, using the initial forward rate curve as an input (see e.g. Heath *et al.* (1992), etc²). Although this class of models allows one to fit the current yield curve, it is not consistent with term structure innovations³. That is, for a specified volatility structure the realisations of the forward curve do not cover all possible observed forward rate curves. Therefore, this class of models is inconsistent with the empirical data.

These shortcomings of the standard models may not be so important for the pricing of certain instruments. However, they fail completely when it comes to the pricing and risk management of instruments or positions that depend on the inter-dynamics of the term structure. The effect of assuming a small number of factors has been noted within the pricing context by Rebonato and Cooper (1995). In particular, they show that the covariance matrix can take only limited shapes, if the number of factors is kept low. This leads to mis-pricing of, for example, swaptions which depend on the correlations between different forward rates.

The consequences of using low factor models are even graver when it comes to risk management of positions that depend on the rates of different maturities. A simple example is when positions are very close to each other. A low factor model will over-estimate the correlation between these positions and hide the actual exposure to risk. This observation is not a deficit of some particular model: to fit correlation fully between, say, N rates, one needs at least an N -factor model. The situation is

¹Other examples include Brennan and Schwartz (1979), Cox *et al.* (1985b), Longstaff, Schwartz (1992a) [175]

²The same problems can be associated with short rate models with time-dependent coefficients such as Black, Derman, and Toy (1990), Hull and White (1993a), Ho and Lee (1986)

³This inconsistency, for example, forces practitioners to recalibrate the parameters of their models.

even more precarious when the number of positions in a portfolio is larger than the number of observed rates. A multi-factor model that will fit the observed correlations will be degenerate when one takes into account un-observed rates.

A further pitfall of finite-factor models lies in the hedging strategies which they predict. In particular, when choosing hedging instruments for a position they do not take into account the maturities of the hedging instruments and the maturity of the position they hedge. For example, in one factor model, a position in bonds maturing in one and two years is a perfect hedge for a bond with thirty years to maturity. This is in stark contrast with fixed income management practice. Practitioners realise that such hedging strategies would fail in practice. Therefore, they typically hedge interest rate risk by estimating the duration of short-, middle-, and long-term commitments separately⁴.

As stated previously, all finite-factor term structure models, taken literally, are incompatible with empirical observation. Therefore for empirical studies⁵ one usually adds error terms to the econometric specification of the model. These errors are then assumed to be independent. This independence assumption is reasonable if the noise is due to non-synchronous trading or observations, or possibly bid-ask spreads. However, if the incompatibility is due to a misspecified model, the error terms will not be independent of the factors and the econometric model will be misspecified.

To solve some of the problems mentioned above, for example, such as that of obtaining complex shapes of the forward curve and realistic correlations between bonds of different maturities, one could introduce a large number of factors. This would make the models very unparsimonious and virtually impossible to estimate.

⁴Ho (1992) attempts to capture this idea by hedging separately "key rate durations" at several different maturities. However, his approach ignores the correlation between different key rate durations.

⁵See, for example, Chen and Scott (1993), and Pearson and Sum (1994).

To overcome these difficulties, Kennedy (1994) models the term structure of interest rates as a random field. In particular he has suggested a model where the forward rates follow a continuous Gaussian random field, so that they evolve as a continuous random surface. The model is consistent with the initial term structure. He derives a necessary and sufficient condition on the drift to ensure that the discounted prices of zero-coupon bonds are martingales, which allows the pricing of contingent claims. Recently, Goldstein (2000), and Santa-Clara and Sornette (2001) have proposed a similar approach to modelling instantaneous forward rates. They simply shock the forward rate curve by a stochastic string which is a solution of some stochastic partial equation. Furthermore, they generalise Kennedy's (1994) non-arbitrage condition on the drift to non-Gaussian random fields. The main innovation of this class of models consists in having each instantaneous forward rate driven by its own shock, while constraining these shocks in a such a way as to keep the forward rate curve continuous.

Since there are infinitely many instantaneous forward rates, in general, random field models are infinite-factor models. In order to price or risk manage derivative securities in a field model, only an estimate of the covariance matrix of the instantaneous forwards is needed. Thus, random fields offer a much more parsimonious description of term structure dynamics than their multi-factor counterparts.

The approach has another advantage, the random field framework naturally accounts for the fact that the best hedging instrument for a bond is another bond of similar maturity. This is in stark contrast to finite factor models. Furthermore, in general it is necessary to use a portfolio with an infinite number of bonds to replicate interest rate contingent claims. However, the pricing of derivatives remains simple: interest rate options can, in general be priced by simulation, and, in some cases, in closed form. As with HJM, random field models do not allow the formulation of a partial differential equation to price derivatives.

Furthermore, models driven by random fields can be estimated without addition of econometric error terms. The models are compatible with any sample of forward rates. This is because, for any parametric specification, there is always a possible path for the random field over a finite interval, that can lead from the forward curve at the beginning of the interval to the forward curve at the end of the interval. The probability of such a realisation of the field might be very low, but it is always possible. Thus there is no need to add observational noise when estimating the model. An estimation procedure would then consist of finding the most likely parameters, given a set of movements of the term structure over time.

However, to use random field models one needs to find a flexible family of covariance functions, and hence a flexible class of field models, that can fit the empirical covariance matrix observed in the market. Functions from this class should be strictly positive definite, thus providing a truly infinite-factor structure for the model. Neither Kennedy (1994), nor Goldstein (2000), nor Santa-Clara and Sornette (2001) addressed this problem. Without this class of covariance functions one cannot capitalise on the main advantage of the field models, i.e. capturing the inter-dynamics of movements in the term structure.

Thus our main objective is to construct Gaussian field models that can fit the sample covariance matrix observed in the market. In this chapter we will describe the framework with which we will develop Gaussian field models. In a departure from previous authors, we work with the yield curve as fundamental, and model it as a Gaussian random field. This is not essential, as the results will be equally applicable to other term structure parameterisations. We will require that the covariance function of the Gaussian field be strictly positive definite, in order for the evolution of the yield to be driven by an infinite-factor structure. Additionally, we will discuss several desirable smoothness properties of the yield curve and its evolution. We will derive sufficient conditions on the covariance function to produce such degrees of smoothness.

The rest of this Chapter is as follows. In Section 3.2, we present a more detailed discussion of the random field models suggested by Kennedy (1994), Goldstein (2000), and Santa-Clara and Sornette (2001). We present our modelling framework in Section 3.3, together with assumptions on the covariance functions that will ensure infinite-factor structure and smoothness of the field realisations. We conclude in Section 3.4.

3.2 Related Research

Kennedy (1994) suggested a model in the spirit of HJM's work, in which the forward rates follow a continuous Gaussian random field so that they evolve as a continuous random surface. In particular, he modelled the instantaneous forward as

$$f_{s,t} = \mu_{s,t} + X_{s,t}, \quad 0 \leq s \leq t,$$

where $X_{s,t}$ is continuous Gaussian random field with covariance structure specified by

$$\text{Cov}(X_{s_1,t_1}, X_{s_2,t_2}) = c(s_1 \wedge s_2, t_1, t_2), \quad 0 \leq s_i \leq t_i, \quad i = 1, 2.$$

The drift function $\mu_{s,t}$ is deterministic and continuous. The model is consistent with the initial term structure, if one sets $\mu_{0,t} \equiv f_{0,t}$.

Furthermore, Kennedy (1994) derived a simple necessary and sufficient condition on the drift surface:

$$\mu_{s,t} = \mu_{0,t} + \int_0^t c(s \wedge \nu, \nu, t) d\nu,$$

which ensures that the discounted prices of zero-coupon bonds are martingales. This allows the pricing of contingent claims.

As a motivating example he suggested a field $X_{s,t}$, is obtained from a deterministic change of the standard Brownian sheet $W_{s,t}$ which is a centred Gaussian random field with covariance function

$$\text{Cov}(W_{s_1,t_1}, W_{s_2,t_2}) = (s_1 \wedge s_2)(t_1 \wedge t_2), \quad s_i \leq 0, t_i \leq 0, i = 1, 2.$$

In this case $X_{s,t}$ may be represented as $X_{s,t} = W_{\sigma(s),\tau(t)}$. This gives correlation of the field for fixed s ,

$$\text{Corr}(X_{s,t_1}, X_{s,t_2}) = \sqrt{\tau(t_1 \vee t_2) / \tau(t_1 \wedge t_2)}.$$

In particular, if $\tau(t) = e^{-\lambda t}$, $\lambda > 0$, then $\text{Corr}(F_{s,t_1}, F_{s,t_2}) = e^{-\lambda - |t_1 - t_2|/2}$.

Goldstein (2000) utilises random fields in a different from the Kennedy (1994) setup. He considers a field labelled with two time indexes, $dZ^T(s)$. The superscript T refers to a maturity date; the index inside the parentheses s refers to the current time. For all dates s , the field describes a realisation of a random function $dZ^T(s)$ for all $T \in (s, \infty)$. Each $dZ^T(s)$ is normally distributed with mean zero and variance ds . Hence, to complete the description of the random field, the correlation structure $\text{Corr}[dZ^{T_1}(s), dZ^{T_2}(s)]$ needs to be specified.

A simple model of a random field is generated by an Ornstein-Uhlenbeck process. This Gaussian field at time s is generated by

$$dZ^T(s) = dZ^s(s)e^{-\rho(T-s)} + \sqrt{2\rho} \int_{u=s}^t dz^s(u)e^{-\rho(T-u)},$$

where $dZ^s(s) \approx \Phi[0, ds]$, and the dz satisfy

$$\mathbb{E}[dz^s(u)] = 0, \quad \text{Cov}[dz^s(u_1)dz^s(u_2)] = \begin{cases} dsdu_1 & \text{if } (u_1 = u_2) \\ 0 & \text{otherwise} \end{cases}$$

The correlation⁶ between the innovations is

$$\text{Corr}(dZ^{T_1}(s), dZ^{T_2}(s)) = e^{-\rho|T_1 - T_2|}. \quad (3.2.1)$$

Note that if $\rho = 0$, the Gaussian field reduces to the one-factor case, since the correlation between innovations of any two forwards becomes unity.

⁶As with traditional Ito calculus, the correlation structure has a rigorous mathematical interpretation only after appropriate integration is performed.

The Ornstein-Uhlenbeck process is continuous but not differentiable. A smoother term structure might be preferred in practice. To accomplish this, Goldstein (2000) suggests integrating the Ornstein-Uhlenbeck process as

$$dV^T(s) = \sqrt{(2\rho^2)} \int_{-\infty}^T du dZ^u(s) e^{-\rho(T-u)}.$$

This results in the correlation structure

$$\text{Corr} [dV^{T_1}(s), dV^{T_2}(s)] = e^{-\rho|T_1-T_2|}|T_1 - T_2|,$$

which is differentiable. Additional integration procedures produce correlation structures which are even smoother.

Goldstein (2000) models the instantaneous forward rates as

$$df^T(s) = \mu_T(s)ds + \sigma_T(s)dZ^T(s).$$

Thus, as in Heath *et al.* (1992), Goldstein (2000) treats the instantaneous forward rates as fundamental. Then assuming the existence of the risk-neutral measure, the dynamics of the instantaneous forwards under this measure become

$$df^T(s) = \mu_T^Q(s)ds + \sigma_T(s)dZ_Q^T(s), \quad (3.2.2)$$

with correlation structure

$$\text{Corr}(df^{T_1}(s), df^{T_2}(s)) \equiv \text{Corr}(dZ_Q^{T_1}(s), dZ_Q^{T_2}(s)) \equiv c(s, T_1, T_2)ds,$$

and drift coefficient

$$\mu_T^Q(s) = \sigma_T(s) \int_s^T du \sigma_u(s) c(s, T, u).$$

This generalises Theorem 1.1 in Kennedy (1994) to non-Gaussian fields.

Santa-Clara and Sornette (2001) work in a similar setting to Goldstein (2000). However, they consider a more general model of the forward rate curve driven by stochastic strings. Again, the dynamics of the forward rates are modelled by

$$d_t f(t, x) = \mu(t, x)dt + \sigma d_t Z(t, x),$$

where the stochastic process $Z(t, x)$ generalises to two dimensions the one-dimensional Brownian motion $W(t)$. The field $Z(t, x)$ depends on time t and on time to maturity⁷ x (not on maturity time T). The notation $d_t Z(t, x)$ denotes a stochastic perturbation to the forward rate curve at time t , with different magnitudes for forward rates with different times to maturity. The subscript t in the differential operator means that the increment is taken with respect to time.

Arbitrage-free dynamics of the instantaneous forward rate under the risk-neutral measure are given by,

$$d_t f(t, x) = \frac{\partial f(t, x)}{\partial x} dt + \sigma(t, x) \int_0^x c(x, y) \sigma(t, y) + \sigma d_t Z_Q(t, x).$$

This is the same as Equation 3.2.2, apart from the additional derivative of the forward with respect to time to maturity. This arises from the alternative parametrisation.

Santa-Clara and Sornette (2001) use stochastic partial differential equations (SPDE) to generate the driving field $d_t Z_Q(t, x)$. The simplest SPDE is

$$\frac{\partial W(t, x)}{\partial t \partial x} = \eta(t, x), \quad (3.2.3)$$

where $\eta(t, x)$ is white noise both in time and time to maturity, characterised by the covariance,

$$\text{Cov}[\eta(t, x), \eta(s, y)] = \delta(s - t) \delta(y - x),$$

where δ denotes the Dirac distribution. The solution of Equation (3.2.3) reads

$$W(t, x) = \int_0^x dy \int_0^t dv \eta(v, y). \quad (3.2.4)$$

This process is known as the Brownian sheet and has the following covariance,

$$\text{Cov}[W(t, x), W(s, y)] = (t \wedge s)(x \wedge y).$$

⁷This specification is sometimes referred to as Musiela parameterisation of the instantness forward rates. Musiela (1993), Brace and Musiela (1994) were the first to use this parametrisation.

The correlation of the increments of this string is

$$\text{Corr}[d_t W(t, x), d_t W(s, y)] = \frac{(x \wedge y)}{\sqrt{xy}}.$$

Another simple formulation that produces many parametric examples is

$$Z(t, x) = Z(0, x) + \int_0^t dv \int_{j(x)}^0 dz \frac{1}{\sqrt{j(x)}} \eta(v, z), \quad (3.2.5)$$

where j is an arbitrary continuous function. We obtain the correlation function of the increments,

$$c(x, y) = \text{Corr}[d_t Z(t, x), d_t Z(t, y)] = \sqrt{\frac{j(x) \wedge j(y)}{j(x) \vee j(y)}}.$$

By making $j(x) = x$ in Equation (3.2.5), we obtain a simple modification of the Brownian sheet. This process is just like the Brownian sheet in Equation (3.2.4), but is rescaled by a factor $\frac{1}{\sqrt{x}}$. It has the correlation function

$$c(x, y) = \text{Corr}[d_t Z(t, x), d_t Z(t, y)] = \sqrt{\frac{x \wedge y}{x \vee y}}.$$

By choosing $j(x) = e^{2\kappa x}$ in Equation (3.2.5), we obtain the Ornstein-Uhlenbeck sheet with correlation structure of the increments

$$c(x, y) = \text{Corr}[d_t Z(t, x), d_t Z(t, y)] = e^{-\kappa|x-y|}.$$

As we see, this correlation is the same as the one obtained by Goldstein (2000) in Equation 3.2.1 above.

The processes studied above are continuous in t and x , but are not differentiable in either t or x . As Goldstein (2000), in order to obtain fields differentiable in time to maturity, Santa-Clara and Sornette (2001) suggest to integrate any of the previously defined strings in x , for example,

$$Y(t, x) = \kappa \sqrt{2} e^{-\kappa x} \int_0^x e^{\kappa y} dy Z(t, y).$$

The correlation of the increments of the integrated Ornstein-Uhlenbeck sheet is

$$c(x, y) = \text{Corr}[d_t Z(t, x), d_t Z(t, y)] = (1 + \kappa|x - y|)e^{-\kappa|x - y|}.$$

Another parametric example is obtained by choosing $j(x) = e^{x^\alpha}$, with an exponent $0 < \alpha < 1$,

$$c(x, y) = \text{Corr}[d_t Z(t, x), d_t Z(t, y)] = e^{-\kappa|x^\alpha - y^\alpha|}.$$

This produces correlations that increase with maturity.

3.3 Model Specification

In this section we set up our modelling framework. Then we describe assumptions on the covariance functions that will ensure infinite-factor structure, and derive necessary conditions to ensure smoothness of the realisations of the field. We parametrise our term structure model in terms of the dynamics of the yield curve, but could equally use instantaneous forward rates, bond prices, etc⁸. This is motivated by the use of yields rather than any other rates in interest rate risk management.

Let $Y_{s,t}$ be the continuously-compounded yield curve for time to maturity t at time s , $0 \leq s \leq t$. We assume that

$$Y_{s,t} = \mu_{s,t} + X_{s,t}, \quad 0 \leq s \leq t,$$

where $X_{s,t}$ is a zero-mean continuous Gaussian random field with covariance structure specified by

$$\text{Cov}(X_{s_1, t_1}, X_{s_2, t_2}) = c(s_1 \wedge s_2, t_1, t_2), \quad 0 \leq s_i \leq t_i, \quad i = 1, 2.$$

We assume that the function c satisfies $c(0, t_1, t_2) \equiv 0$. As a covariance function c has to be symmetric in t_1 and t_2 and positive definite in (s_1, t_1) and (s_2, t_2) . The

⁸For example, Longstaff *et al.* (2001a), (2001b) parametrise their field models in terms of forward swap rate and forward LIBOR rates.

fact that the covariance is specified as a function of $s_1 \wedge s_2$ ensures that the Gaussian random field $X_{s,t}$ has independent increments in the s -direction, in the sense that for any $0 \leq s \leq s' \leq t$, the random variable $X_{s',t} - X_{s,t}$ is independent of the σ -field $\mathcal{F}_t = \sigma\{X_{u,v} : u \leq s, u \leq v\}$; this follows since for $u \leq s, u \leq v$,

$$\begin{aligned} \text{Cov}(X_{s',t} - X_{s,t}, X_{u,v}) &= c(s' \wedge u, t, v) - c(s \wedge u, t, v) \\ &= c(u, t, v) - c(u, t, v) = 0, \end{aligned}$$

which implies independence by the Gaussian assumption. As we are mainly interested in short term risk management we assume that the drift is constant in time and is equal to the initial yield curve $\mu_{0,t} = Y_{0,t}$. This assumption, however, can be easily changed.

3.3.1 Infinite-Factor Gaussian Field Models

Random field models can accommodate both finite- and infinite-factor models. The main motivation for introducing field models is their ability to capture the infinite-factor nature of yield curve dynamics. Thus it is important to identify conditions on the field that ensure this property.

For Gaussian fields these conditions can be expressed in terms of their covariance functions. A yield curve is generated by an infinite-factor Gaussian field when for any choice of (distinct) maturities and any (positive) time horizon the covariance matrix for these set of yields is positive definite. Any positive definite function can be used to define a covariance function. However, from a modelling point of view we are only interested in functions with the above properties. Next, we formally define such classes of functions.

Definition 3.3.1. Let S be a subset of \mathbb{R}^n . A function $P : S \times S \rightarrow \mathbb{R}$ is called a *positive definite kernel* or *PD kernel* if the matrix $M = (P(x_i, x_j))_{i,j=1,\dots,n}$ is *positive semi-definite (PSD) matrix*, i.e.,

$$\sum_{i,j=1}^n c_i c_j P(x_i, x_j) \geq 0 \tag{3.3.1}$$

for all $c = (c_1, \dots, c_n) \in \mathbb{R}^n$ and for all choices of finite subsets $X = \{x_1, \dots, x_n\} \subseteq S$. A function P is called a *strictly positive kernel* or *SPD kernel* if the strict inequality holds in (3.3.1) for all $c \in \mathbb{R}^n$ with $c \neq 0$ and all choices of X . Then M is called a *positive definite matrix* (PD matrix). A function f on \mathbb{R}^n is said to be *positive definite* if $f(x - y)$, $x, y \in \mathbb{R}^n$, is a positive definite kernel. A function f on \mathbb{R}^n is said to be *strictly positive definite* if $f(x - y)$, $x, y \in \mathbb{R}^n$, is a strictly positive definite kernel.

Thus the assumption that the covariance function of a Gaussian field is a strictly positive kernel is equivalent to the assumption that the field is generated by an infinite factor structure. The main challenge is therefore to construct strictly positive kernels that are consistent with the observed sample covariance and correlation matrices.

If we assume that our random field $X(t, \tau)$ is independent of time t , then a simple example satisfying our modelling assumptions is the well-known Brownian motion, i.e. the process defined by $\{X_\tau, \tau \geq 0\}$, with $\mathbb{E}X_\tau = 0$ and $\mathbb{E}X_{\tau_1}X_{\tau_2} = t \wedge s$. If $0 < \tau_1 < \tau_2 < \dots < \tau_N$, the matrix $R = [\tau_k \wedge \tau_l]$ is positive definite. This process can be used for modelling the changes of the yield curve. However, as one easily observes, the shape of the covariance function for Brownian motion is rather limited.

3.3.2 Smoothness of the Field

To apply the Gaussian field model we need to decide on the type of covariance function. In the case of Gaussian fields, properties of covariance functions relate to the smoothness properties of the field. Thus the choice of the covariance function $c(s, u, v)$ is essential in ensuring that the fitted model will be well behaved in any desired sense. This section examines the conditions that will ensure that the yield curve driven by the field is continuous and smooth.

In general, financial economics has little to say on to the smoothness of yield curves. However, one could think of the yields as being represented by integrals over instantaneous forward rates. This will induce differentiability on the yield curve. Unfortunately this issue cannot be resolved empirically, the reason being that differentiability of the yield curve in x relates to smoothness of the yield curve between two very close maturities x and y . Therefore in order to determine whether the yield curve is differentiable we would need to observe yields that are contiguous. As there are only finitely many traded fixed income instruments, such yields cannot be observed. The decision to use fields that produce continuous or differentiable curves⁹ is thus fundamentally a matter of taste.

We will assume that the volatility structures are time-stationary and discuss how to allow for time-dependence later.

Assumption 3.3.2. $c(s, t_1, t_2) = \int_0^s g(t_1 - u, t_2 - u)$ for all $t_1, t_2 \geq s \geq 0$ with some non-negative definite function $g(u, v)$ satisfying the conditions of Lemma 3.3.3.

This assumption implies $c(0, t_1, t_2) = 0$ for all t_1 and t_2 , i.e. \mathcal{F}_0 is a trivial σ -field¹⁰. Next we show the motivation behind this assumption. Consider the covariance between increments of the yields,

$$\begin{aligned} \text{Cov}[dY(s, t_1), dY(s, t_2)] &= \lim_{t \rightarrow s^+} \frac{\text{Cov}[Y(t, t_1) - Y(s, t_1), Y(t, t_2) - Y(s, t_2)]}{t - s} ds \\ &= \lim_{t \rightarrow s^+} \frac{c(t, t_1, t_2) - c(s, t_1, t_2)}{t - s} ds = g(t_1 - s, t_2 - s) ds \end{aligned}$$

It follows from the above equation that our choice of covariance function guarantees time-stationary covariance for increments of the yield curve. Note, however, that it

⁹A random field is sample-path continuous if $\mathbb{P}\{\omega : \|X(t_n, \omega) - X(t^*, \omega)\| \rightarrow 0, \text{ as } n \rightarrow \infty\} = 1$, and sample path-differentiable if $\mathbb{P}\{\omega : X_j(t^*) \text{ exists}\} = 1$, where $X_j(t^*)$ denotes j th first-order partial derivative. Almost-sure property in this case means that all of the paths except the ones on the set of \mathbb{P} -measure zero have the above sample behaviour.

¹⁰We could have assumed the covariance of the form $c(s, t_1, t_2) = \int_0^s g(t_1 - u, t_2 - u)du + h(t_1, t_2)$, with $h(t_1, t_2) \neq 0$. This would correspond to the assumption that \mathcal{F}_0 is not a trivial σ -field. Kennedy (1997) examines such cases in detail.

does not imply stationarity of increments, as this would require covariance of the form $g(t_2 - t_1) ds$, i.e. a function of one argument only.

Lemma 3.3.3 provides a sufficient condition for the continuity of the yield curve.

Lemma 3.3.3. For $\text{Cov}[F(s_1, t_1), F(s_2, t_2)] = c(s_1 \wedge s_2, t_1, t_2) = \int_0^{s_1 \wedge s_2} g(t_1 - u, t_2 - u) du$, $0 \leq s_1 \leq t_1 < \infty$, $0 \leq s_2 \leq t_2 < \infty$, the Gaussian random field has continuous sample functions, with probability one, when $g(u, v)$ is continuous and bounded.

For the proof of Lemma 3.3.3 see Pang (1997). The forward rate surface is continuous when covariance function of the yield increment $g(u, v)$ is continuous and bounded. The conditions of Proposition 3.3.3 are very natural, and they impose no significant constraints on the type of covariance function we may want to use. Lemma 3.3.4 provides a sufficient condition for the smoothness of the yield curve.

Lemma 3.3.4. For $\text{Cov}[F(s_1, t_1), F(s_2, t_2)] = c(s_1 \wedge s_2, t_1, t_2) = \int_0^{s_1 \wedge s_2} g(t_1 - u, t_2 - u) du$, $0 \leq s_1 \leq t_1 < \infty$, $0 \leq s_2 \leq t_2 < \infty$, the Gaussian random field is smooth, with probability one, when $\frac{\partial^2 g(u, v)}{\partial u \partial v}$ is continuous and bounded.

Lemma 3.3.4 is proved in Appendix 3.5.

Lemmas 3.3.3 and 3.3.4 provide constraints on the function $g(u, v)$. Later in this chapter we will discuss further constraints on the function $g(u, v)$ which will ensure its positive definiteness.

It is arguable that our time-stationary covariance function is unsatisfactory because empirical data suggests that volatilities do vary through time. We can address this issue if the covariance function is assumed to be

$$c(s, t_1, t_2) = \int_0^s f(u) g(t_1 - u, t_2 - u) du.$$

where $f(u)$ is a non-negative function. Then it follows that the covariance of increments of the forward rate is given by,

$$\text{Cov}[dF(s, t_1), dF(s, t_2)] = f(s) g(t_1 - s, t_2 - s).$$

The above covariance function allows for time-dependent volatilities, or heteroscedasticity. The covariance surface, as a function of the times to maturities, maintains its shape through time but its level is permitted to shift up and down to reflect changes in volatility levels.

3.4 Conclusions

Random field models of the term structure are generalisations of the finite-factor models which have dominated the financial economics literature. In a field model, each point along the term structure is a distinct random variable with its own dynamics. Each point, however, is correlated with the other points in the term structure. In contrast with finite-factor models, random field models are consistent with both the current yield curve and term structure innovation.

What is more important, especially for interest rate risk management applications, is that Gaussian field models have the potential to fit empirical covariance structure exactly. Furthermore, N -factor models predict that any long-term bond can be perfectly hedged with an appropriate position in N short-term bonds, where a random field model predicts that a better hedging instrument for a long-term bond is another of similar maturity. This is common knowledge among participants in fixed-income markets, but it is not supported by standard models.

Random field models have been applied to interest rate derivative pricing by Kennedy (1994) who models the instantaneous forward rate as a Gaussian field. In slightly a different setting, Goldstein (2000), and Santa-Clara and Sornette (2001), have incorporated random field shocks which are solutions of stochastic partial differential equations into the dynamics of the instantaneous forward rates. However, to use any of these approaches one need to find a flexible family of covariance functions, and thus a flexible class of field models, that can fit the empirical covariance matrix observed in the market. Functions from this class should be strictly positive defi-

nite, thus providing a truly infinite-factor structure for the model. Neither Kennedy (1994), nor Goldstein (2000), nor Santa-Clara and Sornette (2001) have addressed this question. Without this class of covariance functions one cannot capitalise on the main advantage of the field models, i.e. capturing the inter-dynamics of movements in the term structure.

We believe that true value of random field models lies in application to risk management of interest rate sensitive portfolios. Field models are unique in capturing the infinite-dimensional structure of the yield curve, and explaining the interdependence of the rates. Thus field's ability to fit the observed covariance function exactly, or at least very closely, becomes paramount in risk management applications.

Our main objective is to construct Gaussian field models that can fit the sample covariance matrix observed in the market. In this chapter we described the modelling framework. Unlike previous authors, we work with the yield curve as fundamental, and model it as a Gaussian random field. This is not essential, as the results will be equally applicable to other term structure parameterisations. We require that the covariance function of the Gaussian field is strictly positive definite, so that the evolution of the yield is driven by infinite-factor structure. Additionally, we have discussed several desirable smoothness properties of the yield curve and its evolution that one might desire. We have derived sufficient conditions on the covariance function that are needed to produce such a degree of smoothness.

To achieve our objective we will develop three methodologies for constructing strictly positive definite functions, and thus infinite-factor Gaussian fields. In particular, in Chapter 4, we develop a method that produces a covariance function on a grid of a desired size. The method consists of two steps. In the first step we use surface interpolation and approximation techniques to construct a covariance or correlation function approximation based on the sample covariance or correlation matrices. Of course, the constructed surfaces will lack the SPD properties. In the second step

we evaluate the surface on the chosen grid, which gives us a matrix. Then we approximate this matrix by a closest PD covariance or correlation matrix.

In Chapter 5, we identify a large class of strictly positive definite functions. Unfortunately, SPD functions produce only stationary fields. In our case this means that the covariance function of the increments of the yields will be a function of differences of maturities, i.e. a function of $g(x) = g(|T_1 - T_2|)$. Unfortunately, such a model is misspecified: the empirical covariance matrices strongly indicate that the covariance function is a function of T_1 and T_2 separately. We adjust for non-stationarity by deforming time to maturity and obtain an SPD kernel of the form $g(x) = g(|f(T_1) - f(T_2)|)$, where g is an SPD function. This idea resembles the change of time technique in Diffusion Theory.

The class of models developed in Chapter 5 might prove to be too restrictive. Thus In Chapter 6, we approach our problem from a different angle. The idea is based on superposition: we model the covariance function as a sum of SPD functions developed in Chapter 5 and a PD kernel that corrects the function for non-stationarity and enables an exact fit to the sample covariance matrix. The correction term is obtained from Principal Component Analysis of the error matrix, which in turn is obtained from the fit of the SPD functions to the covariance matrix. The sum of SPD and PD kernels is again SPD, thus this construction satisfies the modelling assumptions. We test all three construction on the sample covariance and correlation matrices obtained from US and Japanese bond market data.

3.5 Appendix

Proof. (Lemma 3.3.3)

We use Lemma (3.3.3) together with mean-square differentiability to obtain sample path differentiability of the Gaussian field $F_{s,t}$ with respect to maturity t . From Theorem 2.2.2, Adler (1981), it follows that if the derivative $\partial^2 c(s_1 \wedge s_2, t_1, t_2) / \partial t_1 \partial t_2$

exists and is finite, then the mean-square partial derivative of $F_{s,t}$ with respect to maturity time, $\frac{dF_{s,t}}{dt}$, exists. We only need to show that this derived field is sample path continuous. Standard calculations yield,

$$\text{Cov} \left[\frac{dF_{s_1,t_1}}{dt_1}, \frac{dF_{s_2,t_2}}{dt_2} \right] = \frac{\partial^2 c(s_1 \wedge s_2, t_1, t_2)}{\partial t_1 \partial t_2} = \int_0^{s_1 \wedge s_2} \frac{\partial^2 g(t_1 - u, t_2 - u)}{\partial t_1 \partial t_2} du$$

The functional form of the covariance function of the derived field $\frac{dF_{s,t}}{dt}$ is the same as the covariance function of the original field. Thus, Lemma 3.3.3 guarantees sample path continuity of the derived field $\frac{dF_{s,t}}{dt}$. \square

Chapter 4

SPD Covariance and Correlation Functions via SPD Approximations

4.1 Introduction

In Chapter 3 we introduced the Gaussian random field framework for modelling the term structure of interest rates. As we observed, the only missing information in specifying a Gaussian field model is a covariance function. Furthermore, we argued that the term structures are generated by infinite-factor driving processes. To capture this property we need to model the term structure as an infinite-factor random field. In the case of a Gaussian random field, this requirement is equivalent to the strictly positive definite (SPD) property of the covariance or correlation function of the random field.

This chapter is the first of three in which we develop techniques for constructing functions with this property. In this chapter we provide two alternative non-parametric methodologies. The first is designed to estimate the SPD covariance functions, the

second for the SPD correlation functions. In both case we assume that the instantaneous functions depend on the time to maturities. The empirical sample matrix¹ can be interpreted as an estimator of the continuous function sampled on a finite grid. The grid represents the maturities of the sampled interest rates. Our goal is to find estimates of that function on a grid that is finer than the observed maturities and is compatible with the requirement that the function be SPD. Moreover the estimate must be consistent with the observed sample matrix. We do not construct continuous SPD functions, but take the view that an estimate of that function on a finite grid is sufficient for all practical purposes. We are able to provide positive definite estimations of the covariance and correlation functions for any number of maturities, consistent with the observed sample covariance and correlation matrices and possible smoothness requirements of the yield curve.

The first methodology, in which we derive an estimate of the instantaneous covariance function, consists of two steps. In the first step we approximate the covariance function on the unobserved grid points, i.e. we exploit the smoothness property of the covariance function. We employ surface interpolation procedures compatible with the smoothness requirements of the Gaussian field to make the first order approximation to the unobserved covariance function. In the second step, we evaluate this approximating function on a grid represented by the required set of maturities. This generates a matrix, which among other entries includes the observed covariance values. However, this matrix will not be positive definite in general, though we might expect that the true values of the covariance function for these maturities are not too far from this estimate matrix. Next, we approximate this estimate matrix with the closest positive definite matrix in some norm. We use the *Matrix Nearness* approach to solve this problem. Thus, we obtain a strictly positive definite approximant, and we choose this approximant as the estimate of the covariance function

¹When we speak about empirical sample matrices, we mean empirical sample covariance or correlation matrices.

on the chosen grid. We provide several numerical studies of the methodology based on US interest rate data and evaluate its performance.

The second methodology, in which we derive an estimate of the instantaneous correlation, again consists of two steps. The first step is identical to the first step of the previous methodology, except that we add an additional interpolation technique that takes into account the unit diagonal property of a correlation function. The second step, however, is more involved: the approximation technique used in the previous methodology is not applicable in this case, as the approximating positive definite matrix does not necessarily have unit diagonal. To find a solution in the case of the correlation matrix, we specify a Hilbert space setting for the problem. We then apply a modified *Alternating Projection Method* to find an approximating positive definite correlation matrix in the intersection of two convex sets: symmetric matrices with unit diagonal and positive definite matrices². Thus, we obtain a positive definite approximant with a unit diagonal. We choose this approximant as the estimates of the correlation functions on the chosen grid. Again, we test our methodology numerically on the US interest rate data.

As our procedure for estimating the covariance and correlation functions is numerical in nature, we need to investigate consistency questions. The first question is consistency with the observed data: does the SPD approximant coincide with or approximate sufficiently closely the observed sample covariance or correlation matrix? We measure the distance between the matrices in terms of Frobenius norm. I.e. sum of squared distance of corresponding matrix entries. This means that we approximate the individual variances and covariances, and consequently, the aggregated exposure of the portfolio. The Frobenius norm is also a common choice in statistical applications.

The second question is consistency for a different choice of grid. I.e. if we choose two

²Here the problem is that the set of positive eigenvalues is not closed, so we will restrict the set to the positive definite matrices with eigenvalues larger than some fixed small value.

grids, that have some maturities in common, do the SPD approximants coincide, or are they at least reasonably close to each other for the common maturities? We address these questions in the numerical study of the methodologies.

The organisation of this chapter is as follows. In Section 4.2 we discuss the data set used in the numerical and empirical analysis. In Section 4.3 we develop the first methodology for the estimation of the covariance function. The second methodology for the estimation of the correlation function is developed in Section 4.4. Section 4.5 offers some concluding remarks and discussions.

4.2 Data

For this study we use the yields coming from the Treasury STRIPS (Separate Trading of Registered Interest and Principal Securities) program, which have been traded since 1986. These instruments were created to meet the demand for zero coupon obligations. They are not special issues: the Treasury merely declares that specific notes and bonds (and not others) are eligible for the STRIPS program, and the *stripping* of these issues is done by government bond dealers who give a special security identification number for these issues. The data used for this study is obtained from coupon STRIPS.

In Table 4.7, we provide the descriptive statistics of the whole sample, daily data on 13 maturities, in the period from 01/07/91 to 23/02/01, for the yield differences. Observe that the distribution of the yield differences for each maturity is centred at zero, and it has a very small standard deviation. However, they are not normal, since they are skewed with very fat tails. Skewness is negative for the first 11 maturities and it generally decreases for higher maturities. The yield differences show a high degree of kurtosis that increases dramatically for long maturities.

We plotted in Figure 4.3 the sample covariance and correlation matrices for the first half of the data sample. The plot of the sample correlation shows an expected

decreasing behaviour: the further distance between any two maturities the lower the correlation. Around the centre correlation decreases quickly showing a convex pattern. The shape of the correlation becomes concave for maturities away from the centre.

The actual values of sample covariance and correlation are in the Tables 4.8 and 4.9. Observe the high correlation between maturities. The figures fluctuate between 0.486 and 0.953. Also observe that the drop in correlation closer to the diagonal is more significant than those correlations further away from the diagonal.

We admit that the data used in this study is probably of not very good quality. We point out that our objective is not to conduct an empirical study but merely to test the flexibility of the proposed methodology. Thus, we believe the quality of the data is sufficient for our purpose.

4.3 SPD Approximation of the Covariance Function

As discussed in the introduction, in this section we develop the non-parametric methodology for the estimation of the covariance function. This is done in two steps. In the first step we interpolate sample covariance matrix to obtain approximation of the the covariance function on the unobserved grid points. In the second step, we evaluate this approximating function on a grid represented by the required set of maturities. This generates a matrix, which we approximate with the closest positive definite matrix in some norm. We use *Matrix Nearness* approach to solve this problem. Thus, we obtain a strictly positive definite approximant, and we choose this approximant as the estimate of the covariance function on the chosen grid.

We start this section with the second step. I.e. we develop a strictly positive definite approximation for any given matrix. Then we discuss several surface interpolation

and approximation procedures used in the first step of the methodology. Simultaneously with the discussion we conduct several case studies of the methodology on the US data and assess its performance.

4.3.1 Matrix Nearness Solution

In this section we present a solution to the following problem. For a given matrix, we are interested in finding the closest positive definite matrix. One possible way to solve this problem is to use a *Matrix Nearness* approach.

In general, a matrix nearness problem consists of finding, for an arbitrary matrix A , the nearest member of some given class of matrices. The distance between matrix A and the approximant is measured in a matrix norm. A survey of nearness problems can be found in Higham (1989). One considers a distance function

$$\delta(A) = \min \{ \|E\| : A + E \in S \text{ has property } P \} \quad A \in S, \quad (4.3.1)$$

where S stands for $\mathbb{C}^{m \times n}$ or $\mathbb{R}^{m \times n}$. The property P has to define a subspace or compact subset of S , otherwise the function (4.3.1) is not well-defined. The problem consists then of finding a formula for $\delta(A)$, or at least some characterisation, and determining $X = A + E_{min}$, where E_{min} is a matrix for which the minimum in (4.3.1) is attained. The uniqueness of X is of interest as well. In the case that the property P stands for positive semi-definiteness, we call any X satisfying $\|A - X\| = \delta(A)$ *positive-approximant* of A . Even though the first two questions can be answered positively, in practice one will be interested in efficient algorithms for computing $\delta(A)$ and X .

Frobenius norm and 2-norm³ yield usually most tractable results for nearness prob-

³For matrices of fixed dimension one can introduce norms $\|A\|$, with $A \in \mathbb{C}^{m \times n}$ with the usual norm axioms, see for example Stoer and Bulirsch (1980). Some of examples are $\|A\| = \max_i \sum_k |a_{ik}|$ row-sum norm, $\|A\|_F = \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2} = \text{trace}(A^* A)^{1/2}$ Frobenius or Schur norm, and $\|A\|_2 = \rho(A^* A)^{1/2}$ 2-norm, where ρ stands for *spectral radius*: $\rho(A) = \max\{|\lambda| : \det(A - \lambda I) = 0\}$. The last two norms are unitary invariant, that is, $\|UAV\| = \|A\|$ for any unitary U and V .

lem. The Frobenius norm is more natural in our case as it measures the Euclidean distance between a matrix A and its approximant. It can be viewed as a least-square approximant, which is common in statistical applications.

Next we state Higham's (1988) result that deals with positive definite approximations. Furthermore, we provide a generalisation of this result for the case of strictly positive definite approximants. The following theorem has been proved in Higham (1988):

Theorem 4.3.1 (Higham 1988). *Let $A \in \mathbb{R}^{n \times n}$ and let $A_H = UH$ be a polar decomposition⁴. Then*

$$X_F = (A_H + H)/2 \quad (4.3.2)$$

is the unique positive approximant of A in the Frobenius norm, and

$$\delta_F(A)^2 = \sum_{\lambda_i(A_H) < 0} \lambda_i(A_H)^2 + \|A_K\|_F^2$$

The proof of the Theorem 4.3.1 is simple. Let X be positive semi-definite. From the decomposition

$$\|A - X\|_F^2 = \|A_K + A_H - X\|_F^2 = \|A_H - X\|_F^2 + \|A_K\|_F^2$$

follows that one needs to approximate only the hermitian part of A . From the unitary invariance of the Frobenius norm and spectral decomposition of $A_H = \Pi \Lambda \Pi^T$ follows

$$\|A_H - X\|_F^2 = \|\Lambda - Y\|_F^2 = \sum_{i \neq j} y_{ij}^2 + \sum_i (\lambda_i - y_{ii})^2 \geq \sum_{\lambda_i < 0} \lambda_i^2 \quad (4.3.3)$$

where $Y = \Pi^T X \Pi$. The unique minimum and the equality in the last inequality is

⁴Polar decomposition of a matrix $A \in \mathbb{C}^{n \times n}$ is given by UH where U is a unitary matrix ($U^*U = I$), and H is a positive semidefinite matrix of the same rank as that of A . Any matrix $A \in \mathbb{C}^{n \times n}$ may be expressed in the form $A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*) \equiv A_H + A_K$. A_H is called the hermitian part of A ($A = A^*$) and A_K the skew-hermitian part.

attained for $Y = \text{diag}(d_i)$, with

$$d_i = \begin{cases} \lambda_i, & \lambda_i \geq 0, \\ 0, & \lambda_i < 0 \end{cases} \quad (4.3.4)$$

Thus, $X_F = \Pi \text{diag}(d_i) \Pi^T$. The representation (4.3.2) follows, since $H = \Pi \text{diag}(|\lambda_i|) \Pi^T$.

For the class of normal matrices⁵, Halmos (1972) showed that Frobenius and 2-norm positive approximants are the same.

In our problem, we are interested in approximating a given matrix A with positive definite matrices and not merely with positive semi-definite. Unfortunately, the set of positive-definite matrices is not closed so the distance function (4.3.1) is not well defined. To circumvent this problem, we constrain the set of positive definite matrices by bounding the smallest eigenvalue to be a positive number. In this case, simple modification of the above Theorem 4.3.1 would apply. We term this approximant as a *strictly positive approximant with the bound b* and denote it with X^b . Next Corollary 4.3.2 states this modification:

Corollary 4.3.2. The unique strictly positive approximant for a matrix $A \in \mathbb{R}^{n \times n}$ with the positive lower bound b in Frobenius norm is given by

$$X_F^b = \Pi \text{diag}(d_i^b) \Pi^T,$$

and

$$\delta_F^b(A)^2 = \sum_{\lambda_i(A_H) < b} (\lambda_i(A_H) - b)^2 + \|A_K\|_F^2.$$

The diagonal matrix $\text{diag}(d_i^b)$ is defined similarly to (4.3.4) with

$$d_i^b = \begin{cases} \lambda_i & \lambda_i \geq b, \\ b, & \lambda_i < b. \end{cases}$$

⁵A matrix is called *normal* when $AA^T = A^T A$. For example, symmetric matrices, in particular covariance and correlation matrices, are normal.

The result follows then from (4.3.3). For the case $b = 0$, we obtain the same result as in the Theorem 4.3.2. The question arising in this context is the importance of the bound b . In financial terms, this means, what is the variance of the smallest disturbance factor on the term structure can have. This is an empirical issue, so we do not follow this matter any further at this stage.

Computation of the strictly positive approximant X_F^b is straightforward. We only need to compute the spectral decomposition of A_H . To find the distance between the original matrix and the strictly positive approximant $\delta_F^b(A)$, we only need to sum up all $(\lambda_i(A_H) - b)^2$ over all i , with $\lambda_i(A_H) < b$, and find $\|A_K\|_F^2$. The latter part of the calculation can be omitted if A is symmetric.

A review of the methods for finding the spectral decomposition, i.e. eigenvalues and eigenvectors for matrices, can be found, for example, in Stoer and Bulirsch (1980). The best method, currently known, for the eigenvalue problems is the QR algorithm of Francis (1961/62). This is an iterative method with the number of operations proportional to $O(n^3)$ for a general matrix A , but only to $O(n)$ for a symmetric tridiagonal matrix, per iteration step. Thus for QR algorithm to be efficient, one has to apply first the Housholder algorithm which reduces a given matrix to a tridiagonal one. This method requires for a symmetric matrix $O(\frac{2}{3}n^3)$ operations. Disregarding exceptions, QR -algorithm has a cubic convergence rate.

4.3.2 Surface Approximation, Numerical Analysis, and Empirical Results

In this section we discuss several surface interpolation and approximation techniques, and present an analysis of our methodology applied to covariance function estimation.

Interpolating Two-Dimensional Cubic Splines for Constructing Smooth Surfaces

This algorithm works on a specified rectangular grid $G = [a, b] \times [c, d]$, in the x, y -plane

$$G = \{(x_i, y_i) : a = x_0 < \dots < x_n = b, c = y_0 < \dots < y_n = d\}$$

with heights $u_{ij} := u(x_i, y_j)$, $i = 0, \dots, m$, $j = 0, \dots, n$. We want to find an interpolating smooth surface for the ordinates u_{ij} , which is described by a two-dimensional spline function $S = S(x, y)$ for (x, y) in the rectangle $R = \{(x, y) : a \leq x \leq b, c \leq y \leq d\}$. The bi-cubic spline function $S = S(x, y)$ is defined for $(x, y) \in R$ by the following properties:

1. S fulfils the interpolating conditions: $S(x_i, y_j) = u_{ij}$, $i = 0, \dots, m$, $j = 0, \dots, n$.
2. S is continuously differentiable on R and $\partial^2 / \partial x \partial y$ is continuous on R .
3. In each sub-rectangle R_{ij} ,

$$R_{ij} := \{(x, y) : x_i \leq x \leq x_{i+1}, y_i \leq y \leq y_{i+1}\},$$

S is defined as a bi-cubic polynomial $f_{ij} = f_{ij}(x, y)$.

4. S fulfils certain boundary conditions.

Due to property (3), the bi-cubic spline function S has the form

$$S(x, y) \equiv f_{ij}(x, y) = \sum_{k=0}^3 \sum_{s=0}^3 a_{ijk s} (x - x_i)^k (y - y_j)^s$$

for $(x, y) \in R_{ij}$, $i = 0, \dots, m-1$, $j = 0, \dots, n-1$. This problem can always be solved, as shown by Boor (1962).

Results

To test our methodology we devised the following experiment. Starting from the sample covariance matrix for 13 maturities, we estimate covariance for intermediate maturities, for which yields are not observed. We estimate the covariances using interpolating two-dimensional cubic splines. The idea is to mimic a portfolio with the appropriate asset and liability cash flows. Thus, we considered three matrices corresponding to yearly, monthly, and weekly cash flows. Correspondingly, the dimensions of the interpolated matrices are 30 for yearly, 349 for monthly, and 1509 for weekly cash flows. A portfolio consisting of 1500 asset and liability positions is quite appropriate for a middle size bank or a hedge fund. Table 4.1 summarises the results. The first row shows the Frobenius distance between the interpolated

Table 4.1: Comparison between strictly positive definite approximant with the bound $1e-15$ and interpolated covariance.

Grid	Yearly	Monthly	Weekly
$d_F(A - X_F)$	3.66e-12	8.77e-11	4.92e-10
Max. Diff.	7.95e-13	2.25e-11	9.10e-11
Max. % Diff.	0.00034	0.0092	0.037

covariance matrix and the closest positive definite approximant with a bound $1e-15$, for three portfolios. The approximation worsens, when we go from yearly to monthly and weekly estimations. However, it is still very good. The second and third lines, shows the maximum absolute and maximum percentage distance between the matrices. It is quite remarkable, that maximum percentage distance between the interpolated matrix of dimension 1509 and the closest positive definite approximant is as small as 0.037%. Additionally, we can look at the behaviour of the eigenvalues of the interpolated and the positive-definite approximant. Figure 4.6 displays the sorted eigenvalues in the natural log-scale for three portfolios. The natural logarithm is taken on the absolute value of the sorted eigenvalues. The values on the graph first decline, then when they become negative because of the absolute value they rise again. Thus we can see that roughly half of the eigenvalues of the interpolated

matrices are negative. However, the negative eigenvalues are of order less than 10^{-10} . This explains why the positive definite approximant approximates the matrices with such a good fit.

Unfortunately, the two-dimensional cubic spline shares the same problems with their one-dimensional counterpart and with general polynomial interpolations. The sample path behaviour of the surface though smooth is very irregular. As can be seen from Figure 4.4, the interpolated covariance surface has many bumps that are not in the original data. This effect can be seen in the plot of the variance function resulting from the two-dimensional cubic interpolation in Figure 4.9. In this plot we displayed the original data together with variances resulting from the three positive-definite approximants (yearly, monthly, weekly). As one can see the approximants are consistent. The common points from three different approximants are very close to each other and the data. However, the variance function is very “wiggly” and is not believable from a practical point of view. In the fixed income markets, one would expect a monotonic function, with a bump around two years to maturity. This example shows, that the choice of the interpolating surface is very important. In the next section we describe an alternative interpolation technique and compare results with the two-dimensional cubic interpolation.

Two-Dimensional Interpolating Surface Splines

As in the Section 4.3.2, we want to compute a smooth two-dimensional interpolating surface for a set of prescribed points $(x_i, y_i, f_i) \in \mathbb{R}^3$, $i = 1, \dots, NX$, with $f_i = f(x_i, y_i)$. However instead of using one-dimensional splines, the surface shall be interpreted in a physical way as described by the bending of a thin plate that is deflected at several independent points vertically in such a way as to minimise the bending energy. This approach generalises natural splines from one dimension to

two dimensions. We want to find a function

$$\Phi(x, y) = \sum_{j=1}^n c_j \phi_j(x, y)$$

so that the interpolating conditions

$$\Phi(x_i, y_i) = \sum_{j=1}^n c_j \phi_j(x_i, y_i) \stackrel{!}{=} f_i,$$

are fulfilled, for $i = 1, \dots, NX$, as well as certain minimal properties. We define the functions ϕ_j as follows:

1. $\phi_{j+NX}(x, y) = p_j^{M-1}(x, y)$, for $j = 1, \dots, MM$, monomials up to degree $M-1$,
2. $\phi_j(x, y) = E(x - x_j, y - y_j)$ for $j = 1, \dots, NX$ with $E(x, y) := (r^2)^{M-1} \ln r^2$ for $r^2 = x^2 + y^2$ (kernel function).

This conditions lead to a linear system with $NX + MM$ rows and columns.

$$\begin{pmatrix} G & P \\ P^T & 0 \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}, \quad (4.3.5)$$

where $G = (E(x_i - x_j, y_i - y_j))$, for $i, j = 1, \dots, NX$, and $P = (p_j^{M-1}(x_i, x_j))$, $i = 1, \dots, NX$, $j = 1, \dots, MM$. Unfortunately, the linear systems that arise in this context are generally ill conditioned⁶. If the diagonal elements of G are zero one speaks of interpolating surface splines.

Results

Here, we make the same analysis as with the two-dimensional cubic splines. Table 4.2 summarises the results. Compared with the previous section the fit between the interpolated matrix and the positive-definite approximants is less good. We can see in the third line of the table that maximum percentage distance reaches 1.6% for the monthly cash flows matrix. This result can be confirmed from the plot of

⁶For more details see, for example, Engeln-Müllges and Uhlig (1996)

Table 4.2: Comparison between strictly positive definite approximant with the bound b and interpolated covariance.

Grid	Yearly	Monthly	Weekly
$d_F(A - X_F)$	1.55e-08	1.74e-07	5.72e-8
Max. Diff.	4.16e-09	4.85e-09	3.95e-9
Max. % Diff.	1.36	1.61	1.56

the absolute values of the sorted eigenvalues in Figure 4.7. The negative eigenvalues are of higher order than in the case of two-dimensional cubic splines. Another interesting feature is that decay of positive eigenvalues is slower than in the first cubic splines case. However the two-dimensional interpolating surface splines share the same problems with the cubic splines: the interpolated surface is not monotonic enough. This can be seen in Figure 4.4. The way forward is to use the regularisation techniques to smooth out the “wiggly” surface. In this case one would not have interpolating surface but will achieve a smoother surface. The argument for interpolating surface is questionable, as we do not work with the true covariance matrix but merely with its estimate.

Two-Dimensional Smoothing Surface Splines

Given the same set up as in the interpolation case one can relax the condition that the surfaces passes through the observed points. So achieved degree of freedom can be used to obtain more smooth surfaces. More precisely, we can define the penalised residual sum of squares of a surface Φ by

$$S(\Phi) = \sum_i (f_i - \Phi(x_i, y_i))^2 + \rho J(\Phi).$$

The parameter ρ is a smoothing parameter, and the function $S(\Phi)$ combines a term quantifying the lack of fit of Φ to data with a roughness penalty term. The penalty functional $J(\Phi)$ measures the overall roughness or “wiggleness” of Φ . A common choice for J is

$$J(\Phi) = \int \int_{\mathbb{R}^2} \left\{ \left(\frac{\partial^2 \Phi}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 \Phi}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 \Phi}{\partial y^2} \right)^2 \right\} dx dy$$

For $\rho = 0$ we obtain interpolating surface, and for $\rho \rightarrow \infty$ Φ becomes a plane. It can be shown⁷ that to solve this smoothing problem one need to solve the same linear system as in (4.3.5). The only difference is that matrix G needs to be replaced by matrix $G + \rho I$.

Results

To test the smoothing spline case we choose the smoothing parameter ρ equal to ten. According to our experience this value is quite high and we cannot expect a good fit to the sample covariance matrix. This rather extreme choice is made to test the behaviour of the approximants. We performed the same analysis for the smoothing splines as for the interpolating case. The summary of the results is in the Table 4.3. It can be seen that the fit between the matrix obtained from the smoothing surface

Table 4.3: Comparison between strictly positive definite approximant with the bound b and interpolated covariance.

Grid	Yearly	Monthly	Weekly
$d_F(A - X_F)$	6.19e-09	6.40e-08	2.7e-07
Max. Diff.	8.49e-10	1.10e-09	1.2e-09
Max. % Diff.	0.38	0.58	0.6

and the positive definite approximant is very good. The percentage error of the maximum distance is less than one. However, as can be seen in Figure 4.5, the smoothing surface, is much smoother then in the interpolating case. This difference is even more pronounced if we compare variance of the data points with the variances of the approximants in Figure 4.9. The variances of the approximants exhibit little variation, however the fit to the data is rather poor. The truth lies somewhere in between. One needs to make a choice between the degree of smoothness and the goodness of fit to the data. The behaviour of eigenvalues can be seen in Figure 4.8. Though the surface has quite different shape from the interpolating case, eigenvalues are quite similar.

⁷See, for example, Green and Silverman (1994).

4.4 SPD Approximation of the Correlation Matrix

In the previous Section 4.3 we approximated the covariance function on a designated grid by a positive definite matrix. Often, researchers and practitioners prefer to work with the correlation rather than covariance. The above approach would fail, if we wanted to approximate correlation function on a grid. The reason is that the positive definite approximant from above analysis is not guaranteed to have unit diagonal.

In this section we develop the second non-parametric methodology for the estimation of the covariance function. This, again, is done in two steps. In the first step we interpolate sample correlation matrix to obtain an approximation of the correlation function on the unobserved grid points. We use the same techniques as for the previous methodology, one exception being that we add an additional interpolation technique, Shepard interpolation, that takes into account the unit diagonal property of a correlation function. In the second step, we evaluate this approximating function on a grid represented by the required set of maturities. We then apply a version of the *Alternating Projection Method* to find an approximating positive definite correlation matrix. As we deal with sets and not spaces, the original Alternating Projection Method is not applicable in our case.

We start this section with the second step. I.e. we develop a strictly positive definite approximation with a unit diagonal. Then we apply conduct several numerical studies on the US interest data using surface interpolation and approximation procedures used in the first step of the methodology.

4.4.1 Method of Alternating Projection

Recently, Higham (2002) developed an iterative procedure for finding the nearest correlation matrix. We modify his method and use this modification in the construction of a strictly positive definite approximant with a unit diagonal.

To simplify the reading of this section, we briefly sketch its content. We start with the statement of the problem, then we restate it in a Hilbert space setting. Hilbert space structure allows us to show the existence of the solution. Next, we state the general result on the Alternating Projection Method. We then note that it is not directly applicable in our case, so we introduce a modification of the result. We finish the section with the the algorithm for finding the solution.

First, we state the problem. For a given correlation function on a grid, with the matrix $A \in \mathbb{R}^{n \times n}$ we want to find the positive definite matrix X which solves

$$d(A) = \min\{\|A - X\|_F : X \text{ is a } \mathcal{P}^b, \text{ correlation matrix}\} \quad (4.4.1)$$

where \mathcal{P}^b stands for the set of positive definite matrices with the smallest eigenvalue greater than or equal to $b \in \mathbb{R}_+$. We consider sets

$$P = \{Y = Y^T \in \mathbb{R}^{n \times n} : \mathcal{P}^b\},$$

$$U = \{Y = Y^T \in \mathbb{R}^{n \times n} : y_{ii} = 1, i = 1 : n\}.$$

The solution to problem (4.4.1) should be in the intersection of the sets P and U .

Next we restate the problem within Hilbert space setting. On the linear space $\mathbb{R}^{n \times n}$ we define an inner product function $\langle \cdot, \cdot \rangle$:

$$\langle A, B \rangle = \text{trace}(A^T B), \quad \text{for all } A, B \in \mathbb{R}^{n \times n}$$

The space $(\mathbb{R}^{n \times n}, \langle \cdot, \cdot \rangle)$ is complete, that is, every Cauchy sequence in $\mathbb{R}^{n \times n}$ converges with respect to the norm induced by the inner product $\|A\| := \sqrt{\langle A, A \rangle}$, $A \in \mathbb{R}^{n \times n}$. In particular, this is the norm we are using in defining the minimum norm optimisation problem (4.4.1). In following, this additional geometrical structure will be exploited to find the solution to (4.4.1). We observe, that both sets, P and U , are convex and closed. It follows then, Luenberger (1968), Proposition 2.4.1, that the intersection, $S \cap P$, is a closed convex set, as well. From Hilbert space theory we have following standard result about projection on closed convex sets, Luenberger (1968), Theorem 3.12.1,

Theorem 4.4.1. *Let x be a vector in a Hilbert space H and let K be a closed convex subset of H . Then there is a unique vector $k_0 \in K$ such that*

$$\|x - k_0\| \leq \|x - k\|$$

for all $k \in K$. Furthermore, a necessary and sufficient condition that k_0 be unique minimising vector is that

$$(x - k_0 | k - k_0) \leq 0 \quad (4.4.2)$$

for all $k \in K$.

Though Theorem 4.4.1 guarantees existence of a unique solution in the intersection of P and U , it does not give any means for finding it and the condition (4.4.2) on the angle of the solution is the only characterisation one has.

Next step in finding the solution is a result originally proved by von Neumann (1950), Theorem 13.7, for $k = 2$ and, independently, by Wiener (1955). The case $k \geq 2$ has been generalised by Halperin (1962). It is generally known as the *Method of Alternating Projection*, which is an iterative scheme for finding the best approximation to any given point in a Hilbert space from the intersection of a finite collection of closed subspaces.

Let M_1, M_2, \dots, M_k be $k \geq 2$ closed subspaces in the Hilbert space X . The mapping in Theorem 4.4.1, which associates with every element x of the Hilbert space X the closest element in a convex subset M is termed *orthogonal projection* if M is a closed subspace. We denote this projection by P_M , i.e. $P_M(x)$ is the best approximation of nearest point in M to any $x \in X$. We can state now the

Theorem 4.4.2 (von Neumann-Halperin). *For each $x \in X$.*

$$\lim_{n \rightarrow \infty} \|(P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^n(x) - P_{\cap_1^k M_i}(x)\| = 0.$$

That is, $P_{\cap_1^k M_i}(x)$ is the limit of the sequence $\{x_n\}$ defined inductively by $x_1 = x$ and

$$x_{n+1} = (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})(x_n) \quad (n = 1, 2, \dots)$$

which is obtained by determining best approximation onto the individual subspaces M_i in a cyclic manner. The expression $(P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^n$ stands for the n -th cycle of the projection $P_{M_k} P_{M_{k-1}} \cdots P_{M_1}$. Unfortunately, in our problem, the intersection set $P \cap U$ is not a subspace, thus $P_{P \cap U}$ is not a projection.

Example 4.4.3. We borrow the following example from Shih-Ping Han (1988). Consider space \mathbb{R}^2 equipped with the Euclidean inner product. Let $M_1 := \{(\zeta_1, \zeta_2) | \zeta_2 \leq 0\}$ and $M_2 := \{(\zeta_1, \zeta_2) | \zeta_1 + \zeta_2 \leq 0\}$. The projection method of Theorem 4.4.2 does not work for any vector x outside M_1 and M_2 . This is demonstrated in Figure 4.1. The first drawing represents application of the standard projection method on the vector x . The vector $x_{12} = P_{M_2} P_{M_1}$ is clearly suboptimal. The iterative projection terminates at x_{12} since the vector is in the intersection of M_1 and M_2 .

To overcome this constraint, Dykstra (1983), developed, within a least-square estimation setting, an algorithm for projecting an element in a finite-dimensional inner-product space onto a closed convex cone M , when K can be written as $M_1 \cap \cdots \cap M_r$, is also a closed convex cone. This result has been generalised in Boyle and Dykstra (1985). In particular, it is shown that closed convex cones can be replaced by arbitrary closed convex sets C_i and the finite-dimensional inner-product space by a general Hilbert space.

The idea of Dykstra (1983) was to introduce a correction procedure which will prevent the pathological cases such as in the above example. Let M_1, \dots, M_k be closed convex subsets of a Hilbert space H . Given a vector $x \in H$ one applies the

first cycle of the standard projection method, obtaining so the sequence:

$$\begin{aligned} x_{11} &= P_{M_1}(x), & I_{11} &= x_{11} - x, \\ &\vdots & &\vdots \\ x_{1k} &= P_{M_1}(x_{1(k-1)}), & I_{1k} &= x_{1k} - x_{1(k-1)}. \end{aligned}$$

In the second cycle, the operator P_{M_1} is applied on $x_{1k} - I_{11}$, instead of on x_{1k} . The increment I_{11} can be interpreted as the normal to the orthogonal projection x_{11} . Repeating this procedure with following P_{M_i} , one obtains

$$\begin{aligned} x_{21} &= P_{M_1}(x_{1k} - I_{11}), & I_{21} &= x_{21} - (x_{1k} - I_{11}), \\ &\vdots & &\vdots \\ x_{2k} &= P_{M_1}(x_{1(k-1)} - I_{1k}), & I_{2k} &= x_{2k} - x_{2(k-1)}. \end{aligned}$$

Continuing this routine of removing the increment in the previous cycle associated with M_i , before projecting onto M_i , generates the infinite arrays $\{x_{ni}\}$ and $\{I_{ni}\}$, where $n \geq 1$ and $1 \leq i \leq r$. The sequence $\{x_{ni}\}$ satisfies following

Theorem 4.4.4 (Boyle and Dykstra (1985), Theorem 2). *For any $1 \leq i \leq r$ the sequence $\{x_{ni}\}$ converges strongly to x^* , i.e., $\|x_{ni} - x^*\| \rightarrow 0$.*

The second drawing in Figure 4.1 represents the effect of correction by normal I_{11} . The vector x_{12} is effectively taken out of the intersection $M_1 \cap M_2$ and the iterative procedure will proceed with the limit as the projection of x onto $M_1 \cap M_2$.

Thus, to find the closest strictly positive definite correlation matrix for a given matrix A , we can apply the following algorithm,

Algorithm 4.4.5. $\Delta S_0 = 0, Y_0 = A$

for $k = 1, 2, \dots$

$R_k = Y_{k-1} - \Delta S_{k-1}$, where ΔS_{k-1} is Dykstra's correction.

$X_k = P_P(R_k)$

$$\Delta S_k = X_k - R_k$$

$$Y_k = P_U(X_k)$$

end

Projections P_P and P_U

To apply above algorithm to the problem (4.4.1) we only need to find the projections P_P and P_U on sets P and U . The projection of a matrix $A \in \mathbb{R}^{n \times n}$ on the closed set of all positive definite matrices with the smallest eigenvalue bounded by a positive constant b has been found in Section 4.3. In particular, for a symmetric matrix $A = \Pi \text{diag}(\lambda_i) \Pi^T$, we have the projection $P_P(A) = \Pi \text{diag}(\max(\lambda_i, b)) \Pi^T$.

It is simple to show that the projection $P_U(A)$ is of the form,

$$P_U(A) = \begin{cases} a_{ij} & i \neq j, \\ 1, & i = j. \end{cases}$$

I.e., we simply need to set the diagonal entries of A to 1.

Remark 4.4.6. In a recent working paper Zhang and Wu (2003) developed a procedure for a low-rank approximation of a correlation matrix using Lagrange multipliers. Wu (2003) applied this approximation withing pricing framework in the context of calibration of LIBOR market models.

Recently, the author of this thesis, in Weigel (2003) applied the above developed procedure to the calibration of LIBOR market models to historic correlations. The preliminary results suggest that this method presents a more simple and easier to implements solution than that of Zhang and Wu (2003).

Additionally, Brace and Womersley (2000), performed an implicit calibration to the implied correlation matrix, within swaption pricing framework, using semidefinite

programming. However, the relations to the above mentioned procedure is rather involved.

4.4.2 Numerical Analysis and Empirical Results

Interpolating Two-Dimensional Cubic Splines for Constructing Smooth Surfaces

Here we employ the same interpolating methodology as in the case Section 4.3.2. The summary results can be seen in the Table 4.4. We considered only yearly and

Table 4.4: Comparison between strictly positive definite approximant with a bound and interpolated correlation function, using two-dimensional cubic splines.

Grid	Yearly	Monthly
$d_F(A - X_F)$	0.43	4.27
Max. Diff.	0.17	0.16
Max. % Diff.	14.20	16.52
Data % Diff.	1.88	1.96
Bound	1e-5	1e-2
Tolerance	1e-6	1e-3
Number of Iter.	42	55

monthly maturity spans, as the numerical computations were quite expensive. Line 4 provides maximum percentage difference between the unit diagonal SPD approximant and the data. For both yearly and monthly holding it is almost two percent and is substantially larger then in the covariance approximation case. Moreover, the line 3 describes the maximum percentage difference between the interpolating surface and the approximant. This error is as large as 16%. This indicates, that unit diagonal approximants are not as flexible as the simple SPD approximants. The plots of ordered eigenvalues in the log scale can be seen in the Figure 4.10. Tables 4.10 and 4.11 present the unit diagonal SPD approximant for yearly and monthly maturities together with the percentage error fit relative to the sample correlation matrix.

Two-Dimensional Interpolating Surface Splines

Here, we consider the interpolating surface splines. The percentage distances be-

Table 4.5: Comparison between strictly positive definite approximant with a bound and interpolated correlation function, using two-dimensional interpolating surface splines.

Grid	Yearly	Monthly
$d_F(A - X_F)$	0.17	1.05
Max. Diff.	0.076	0.145
Max. % Diff.	8.25791	16.98
Data % Diff.	2.25	2.22
Bound	1e-5	1e-2
Tolerance	1e-6	1e-4
Number of Iter.	25	9

tween the data and the unit diagonal SPD approximants are even larger then in the case of two-dimensional cubic splines.

The plots of ordered eigenvalues in the log scale can be seen in the Figure 4.11. Tables 4.12 and 4.13 present the unit diagonal SPD approximant for yearly and monthly maturities together with the percentage error fit relative to the sample correlation matrix.

Shepard Interpolation

The disadvantage of the previous two interpolation techniques when evaluated on the diagonal they do not have unit values. I.e. they do not have the properties of the correlation functions. One could impose this by adding unit diagonal as additional interpolation points. This, however, will produce a grid that is not rectangular. Next we analyse a technique that can deal with this type of problems.

If the nodes (x_j, y_j) , $j = 0, \dots, N$, with $(x_j, y_j) \in B \subset \mathbb{R}^2$, do not form a rectangular grid but are arranged in a completely arbitrary and unordered way, common method of choice is method Shepard (1968). This method has proved to be well suited

for the graphic representation of surfaces. Its interpolating function Φ is uniquely determined independently from the ordering of the nodes (x_j, y_j) . The function $f : z = f(x, y)$ for $(x, y) \in B$, where B is an arbitrary region of the x, y plane, is interpolated for the given nodes (x_j, y_j) by the function

$$\phi(x, y) = \sum_{j=0}^N \omega_j(x, y) f_j. \quad (4.4.3)$$

Here $\phi(x_j, y_j) = f(x_j, y_j)$ for $j = 0, \dots, N$, where the f_j are the given functional values $f(x_j, y_j)$ at the nodes (x_j, y_j) , $j = 0, \dots, N$, and $\omega_j = \omega_j(x, y)$ is a weight function. Setting

$$r_j(x, y) = \sqrt{(x - x_j)^2 + (y - y_j^2)} \quad j = 0, \dots, N,$$

we can define ω_j as

$$\omega_j(x, y) = \frac{r_j^{-\mu}}{\sum_{i=0}^N r_i^{-\mu}}, \quad 0 < \mu < \infty. \quad (4.4.4)$$

With (4.4.3) the Shepard function ϕ has the representation

$$\phi(x, y) = \frac{\sum_{j=0}^N r_j^{-\mu} f_j}{\sum_{i=0}^N r_i^{-\mu}}.$$

The exponent μ in (4.4.4) can be chosen arbitrarily, depending on how smooth one wants the interpolating surface to be. If $0 < \mu \leq 1$, the function ϕ has peaks at the nodes. If $\mu > 1$, the function is level at the nodes.

Results

The summary of result for Shepard interpolation can be seen in the Table 4.6. As in two previous cases the fit is quite poor. Plot of eigenvalues in Figure 4.12 reveals that negative eigenvalues of the matrix resulting from interpolation has very high negative eigenvalues, this might explain the performance of the unit diagonal SPD approximant. Tables 4.14 and 4.15 present the unit diagonal SPD approximant for yearly and monthly maturities together with the percentage error fit relative to the sample correlation matrix.

Table 4.6: Comparison between strictly positive definite approximant with a bound and interpolated correlation function, using Shepard Interpolation.

Grid	Yearly	Monthly
$d_F(A - X_F)$	0.42	4.12
Max. Diff.	0.067	0.16
Max. % Diff.	6.73	19.28
Data % Diff.	1.66	1.86
Bound	1e-5	1e-2
Tolerance	1e-6	1e-3
Number of Iter.	65	20

4.5 Conclusions

In this chapter we developed two alternative non-parametric methodologies for the estimation of strictly positive definite covariance and correlation functions. Both methodologies consist of two steps. In the first step we interpolate or approximate the sample matrix to obtain the values of the covariance or correlation function on the unobserved grid points. In the second step, we evaluate this approximating function on a grid represented by the required set of maturities. This generates a matrix. However, this matrix will not be positive definite in general. However, we might expect that the true values of the SPD functions for these maturities are not too far from this estimate matrix.

In the second step, we approximate the estimate matrix by its closest positive definite matrix in some norm. This approximant is then chosen to be the estimate of the SPD function on chosen grid. At this point the two methodologies differ. For the SPD covariance function approximation we use the *Matrix Nearness* approach. This yields the closest positive definite matrix in the Frobenius norm. This is sufficient, as any positive definite matrix is covariance function. For the SPD correlation function we apply a modification of the *Alternating Projection Method* to find an approximating positive definite correlation matrix in the intersection of two convex sets: symmetric matrices with unit diagonal and positive definite matrices.

To test the methodologies, we conducted several numerical studies on US interest data. Both methodologies work quite well. However, we feel that the methodology for the estimation of an SPD covariance function is slightly superior to the methodology for the estimation of an SPD correlation function, the reason being that the algorithm for approximation with positive definite matrices is simpler to implement than that for with positive definite matrices with unit diagonal, i.e. correlation matrices. Moreover, our numerical results suggest that the quality of fit is better when we estimate an SPD covariance function.

Although we achieved very good overall results, the methodologies we developed have serious drawbacks. They depend crucially on the choice of the interpolating procedure. In general, it is difficult to decide what method to choose. Additionally, it must be noted that we do not know exactly the sample matrices even on a finite grid. Instead, we deal with sample estimators. The interpolation procedures tend to pick up all the noise in the estimate and produce “wiggly” surfaces which are not justified economically. Smoothing instead of interpolating is a natural way to proceed in this context. Thus we introduced a smoothing approach based on two-dimensional smoothing splines. This allowed us to get rid of “wiggleness” in the interpolating surface. However, the smoothing parameter becomes an issue in this context. It can be chosen by the researcher or practitioner at their discretion. Alternatively, one can apply a cross-validation methodology to estimate a smoothing parameter.

Two types of consistency problems arise in this context. The first is to do with consistency with the data. Once an interpolation procedure is chosen, it is necessary that the approximants have similar covariances and correlations to the sample covariance and correlation. We performed several experiments and found that the differences in the case of covariance are negligible. However, correlation approximants in some cases have shown considerable errors. The second type of consistency concerns different choices of maturities. Given two approximants, arising from two

different choices of maturities, do covariances and correlations differ for common maturities? Again, we found that the estimators are consistent in the case of covariances. Correlation approximants exhibited quite large errors.

The methods for estimating SPD functions developed in this section may seem to be awkward. They do not produce an actual function, only its values on some grid. In the next two chapters we will develop semi-parametric and fully parametric approaches for estimation of SPD covariance and correlation functions. These approaches will yield proper functional forms of the SPD covariance and correlation functions. The degrees of freedom arising in the present context will be reduced to the estimation of a small number of parameters.

4.6 Appendices

4.6.1 Figures

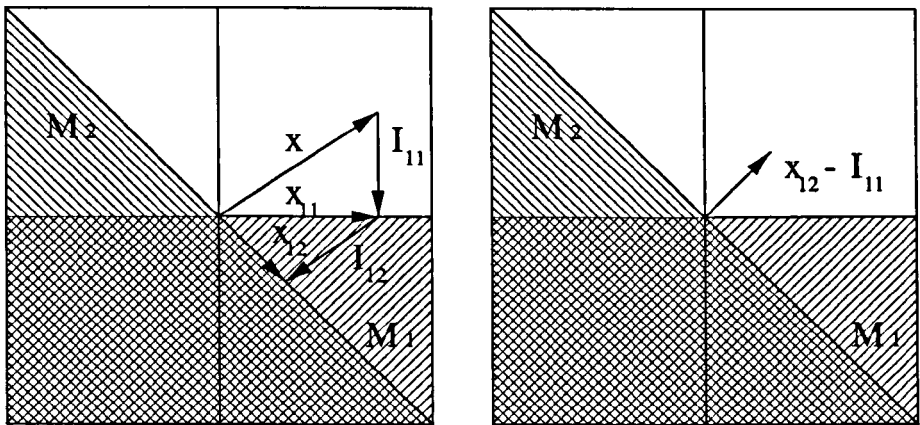


Figure 4.1: Alternating Projection on closed subsets of \mathbb{R}^2 .

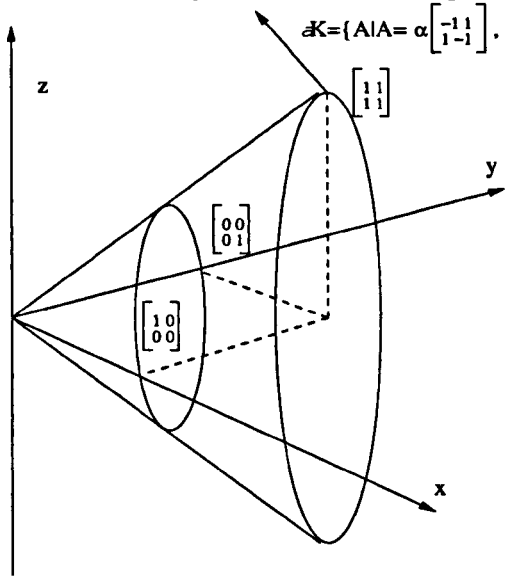


Figure 4.2: The positive semidefinite matrix cone K .

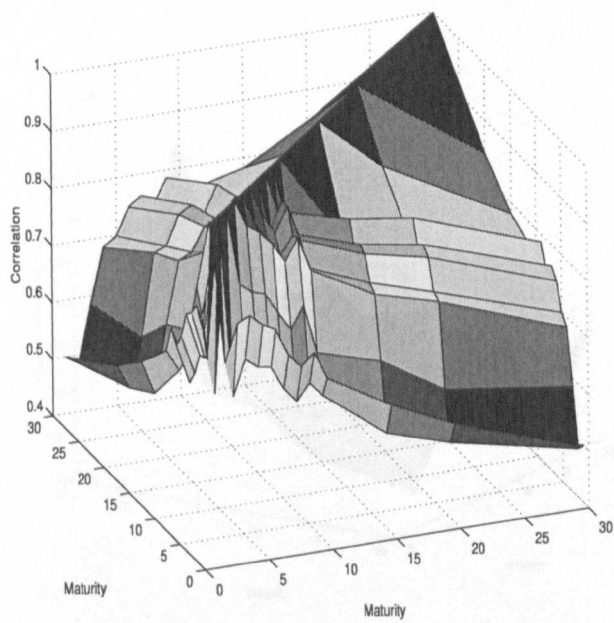
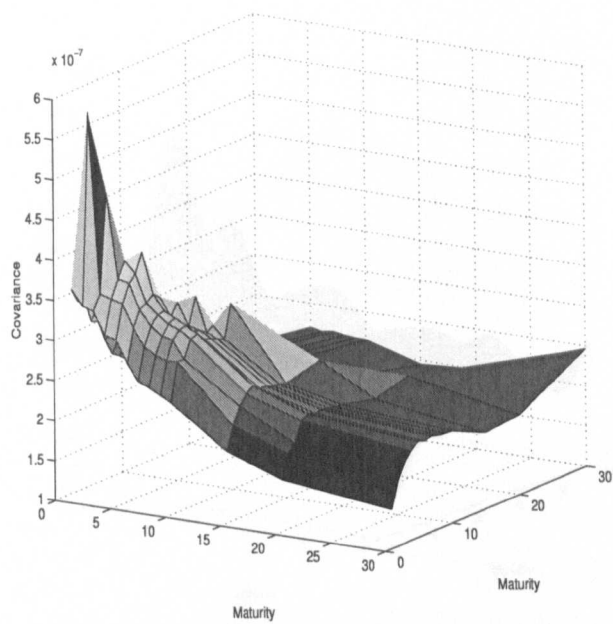


Figure 4.3: Sample covariance matrix (above) and sample correlation matrix.

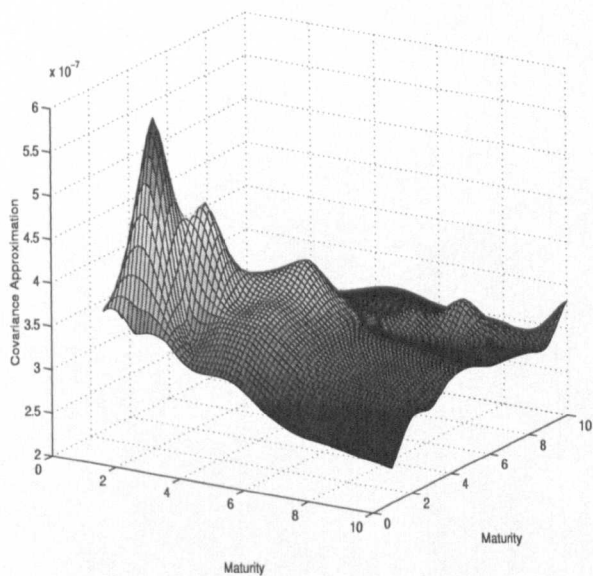
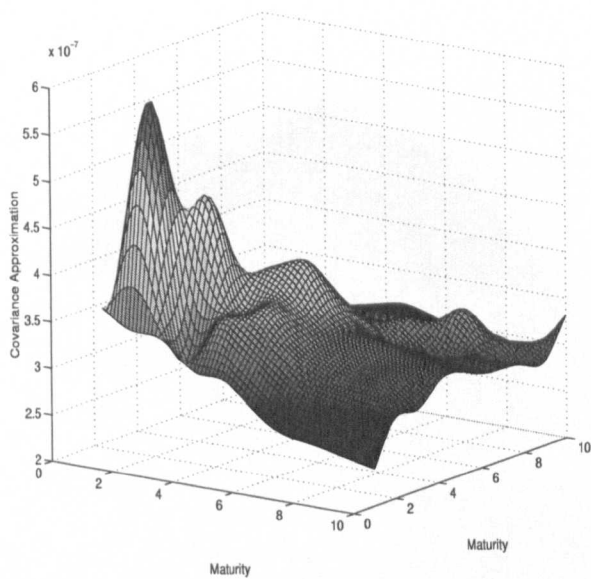


Figure 4.4: Approximation of covariance using two-dimensional cubic splines (above) and two-dimensional interpolating surface splines (below).

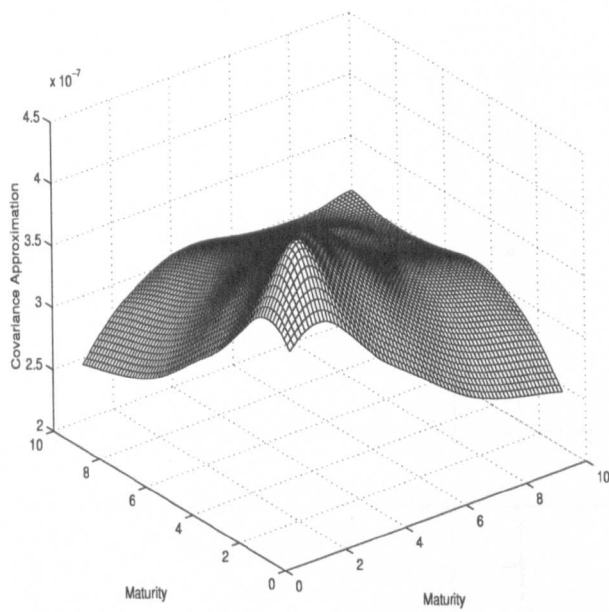


Figure 4.5: Approximation of covariance using two-dimensional smoothing surface splines, with smoothing coefficient $\rho = 10$.

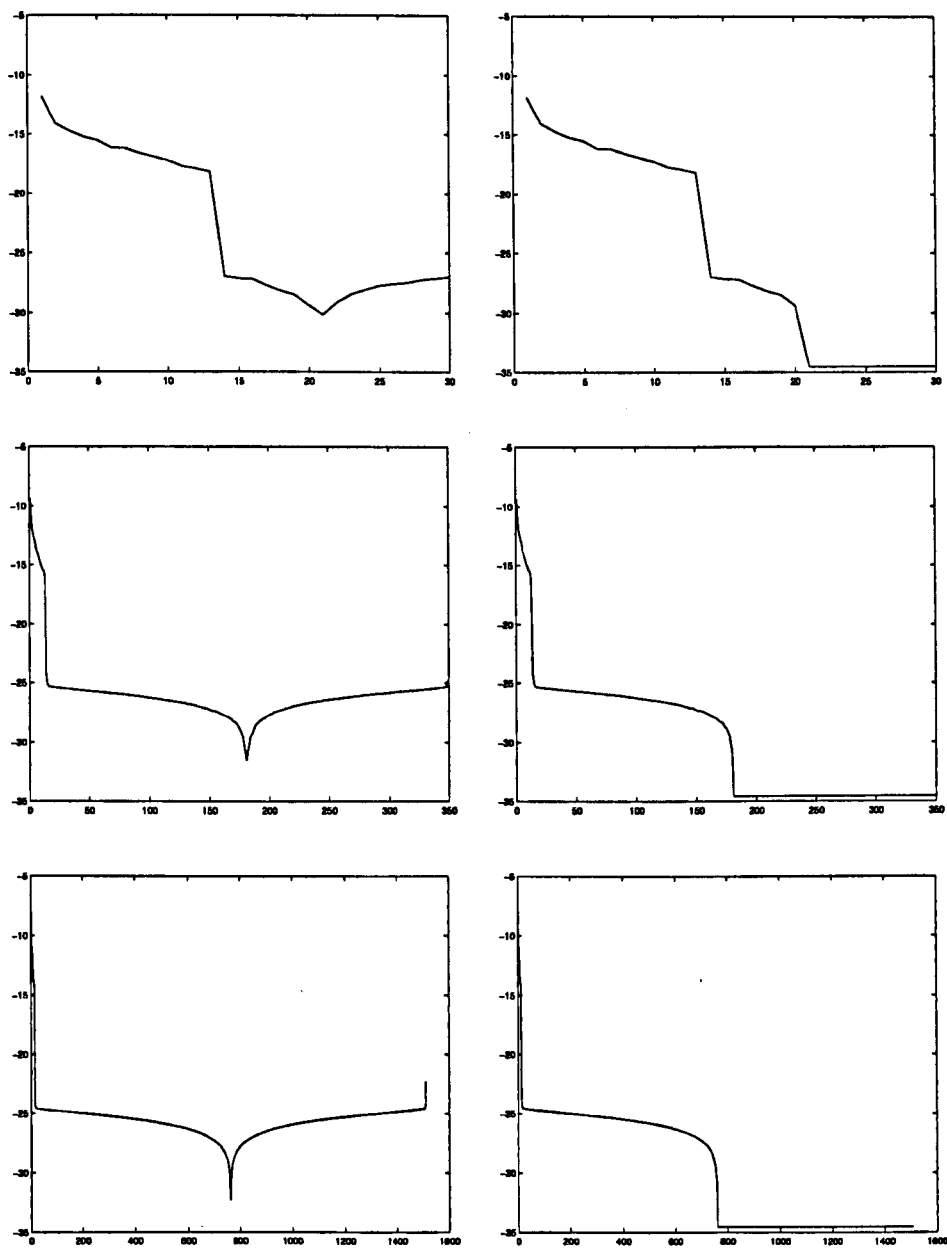


Figure 4.6: Two-dimensional cubic splines: Eigenvalues of interpolated matrix (left), and eigenvalues of the nearest covariance matrix, with eigenvalues bounded by bound $= 1e - 15$.

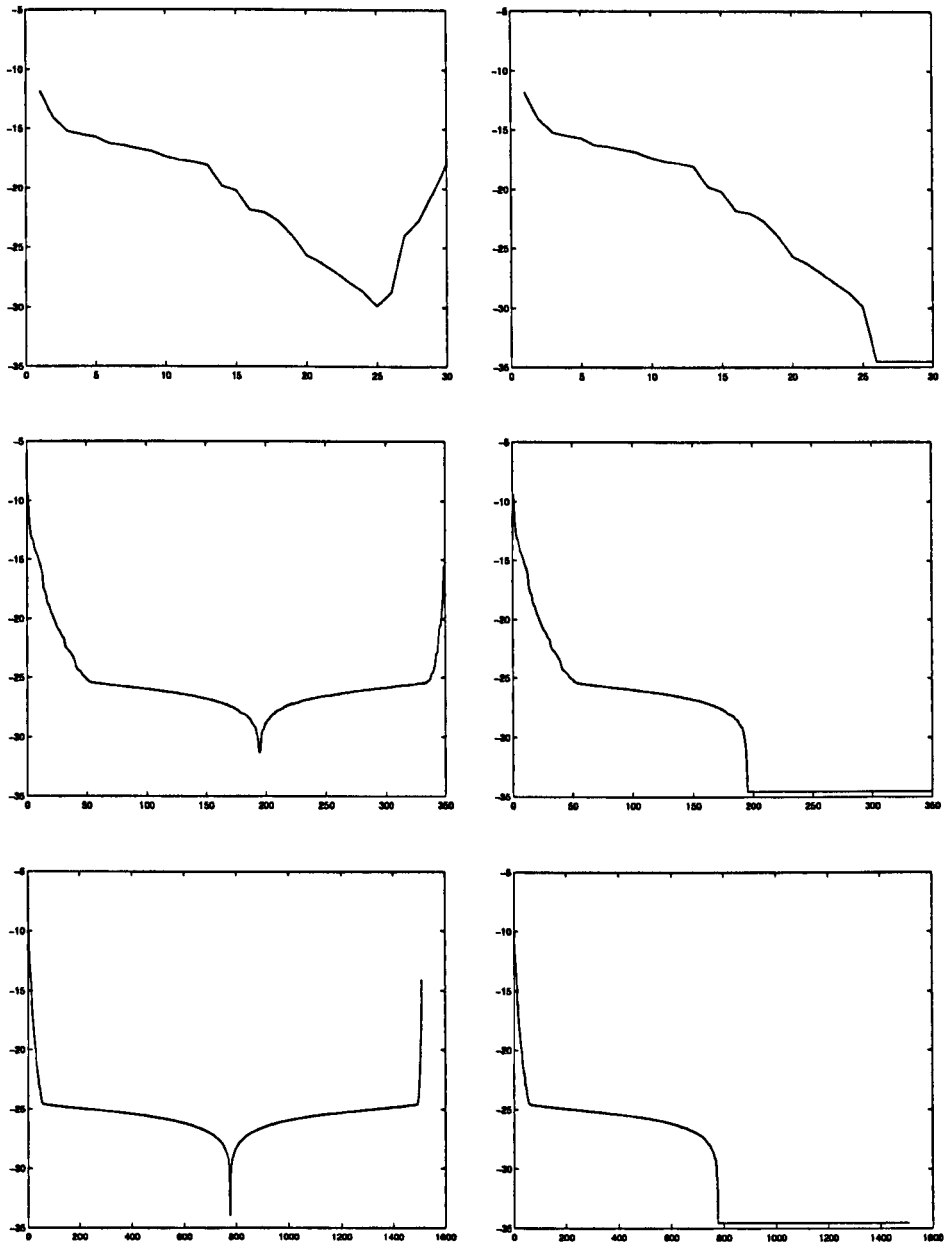


Figure 4.7: Two-dimensional surface splines: Eigenvalues of interpolated matrix (left), and eigenvalues of the nearest covariance matrix, with eigenvalues bounded by $\text{bound} = 1e - 15$.

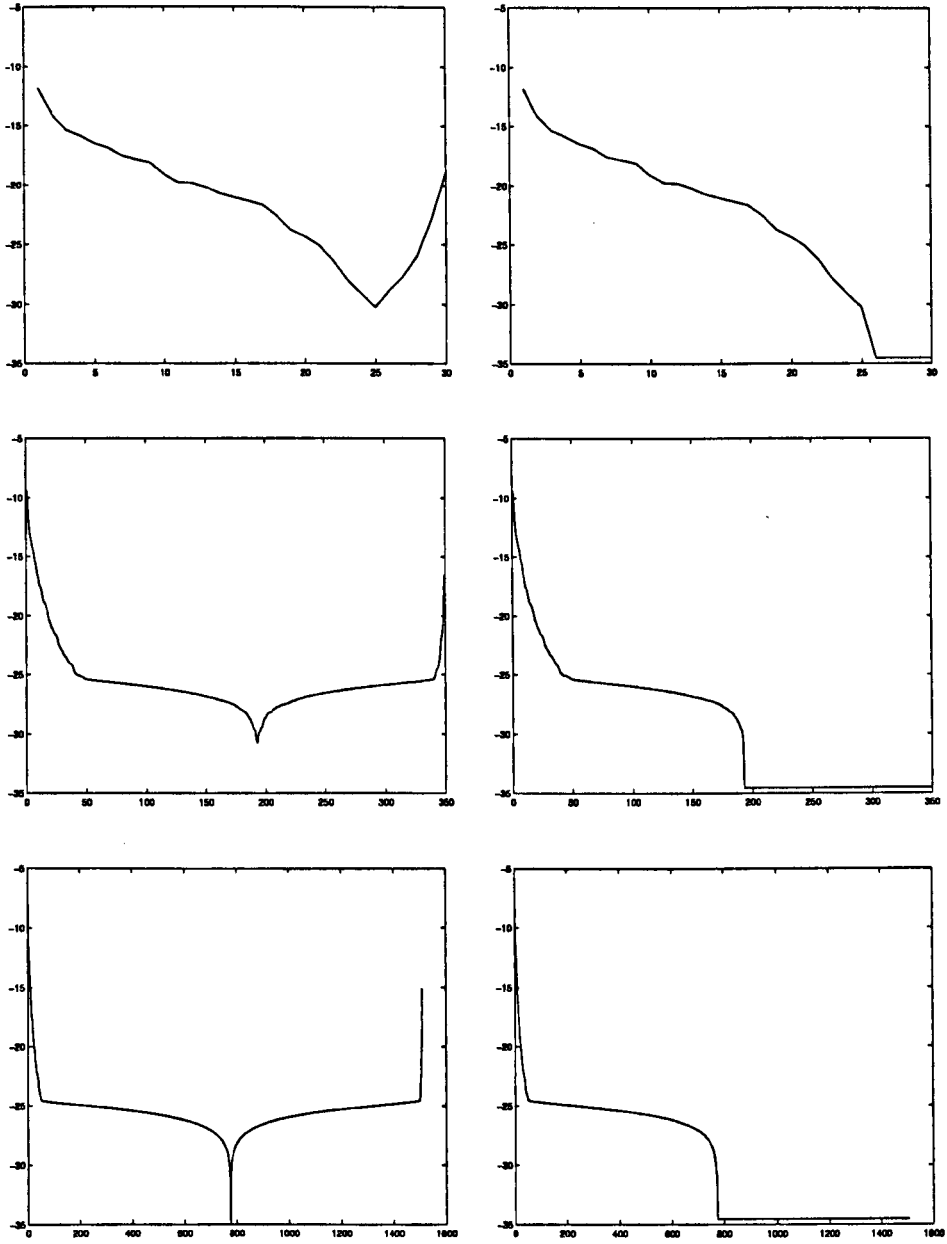


Figure 4.8: Two-dimensional surface smoothing splines with smoothing parameter $\rho = 10$: Eigenvalues of interpolated matrix (left), and eigenvalues of the nearest covariance matrix, with eigenvalues bounded by bound = $1e - 15$.

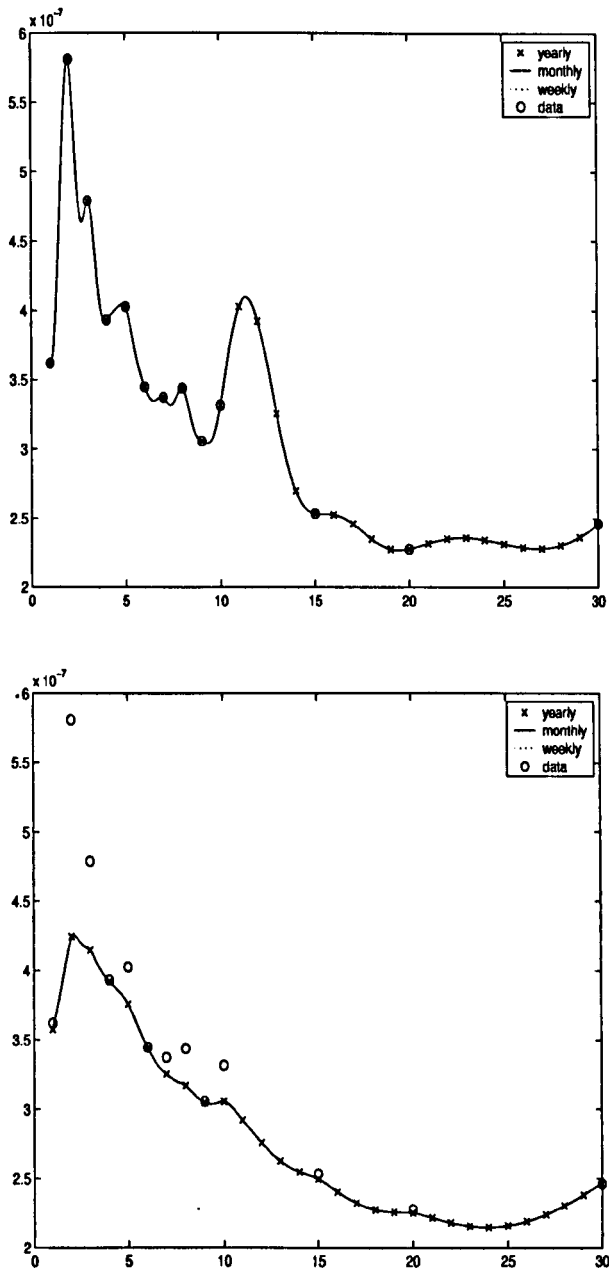


Figure 4.9: The data variances together with variances resulting from positive-definite approximants. Two-dimensional cubic splines (above) and two-dimensional smoothing surface splines (below), with smoothing coefficient $\rho = 10$.

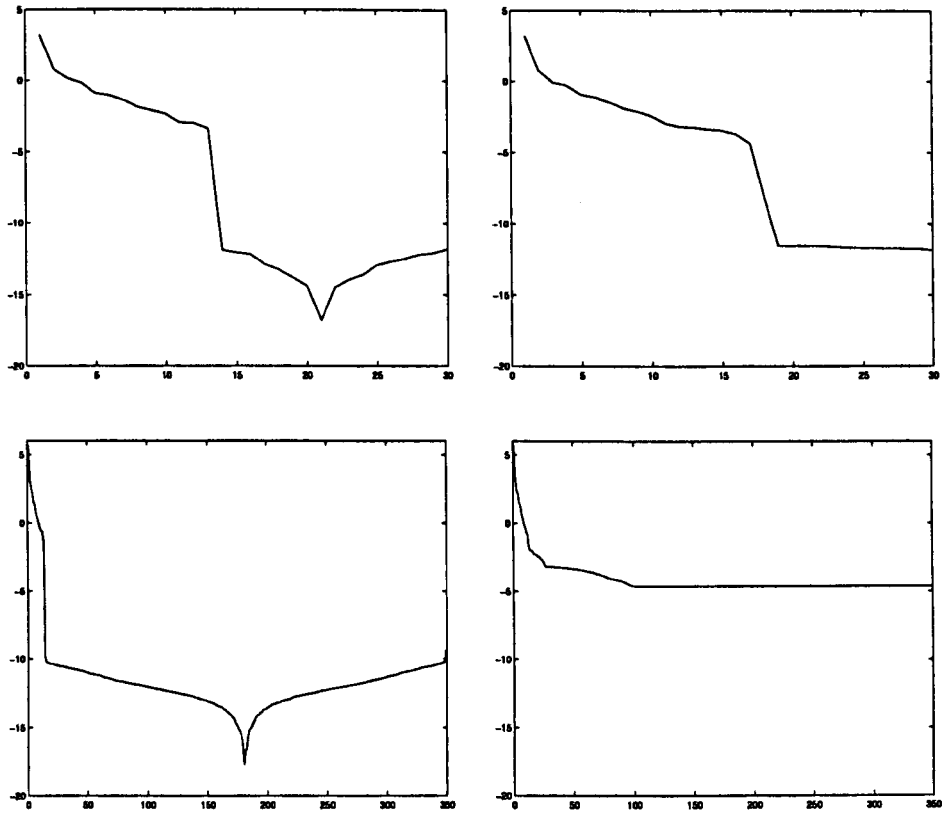


Figure 4.10: Two-dimensional cubic splines: Eigenvalues of interpolated matrix (left), and eigenvalues of the nearest correlation matrix, with eigenvalues bounded by bound as in Table 4.4.

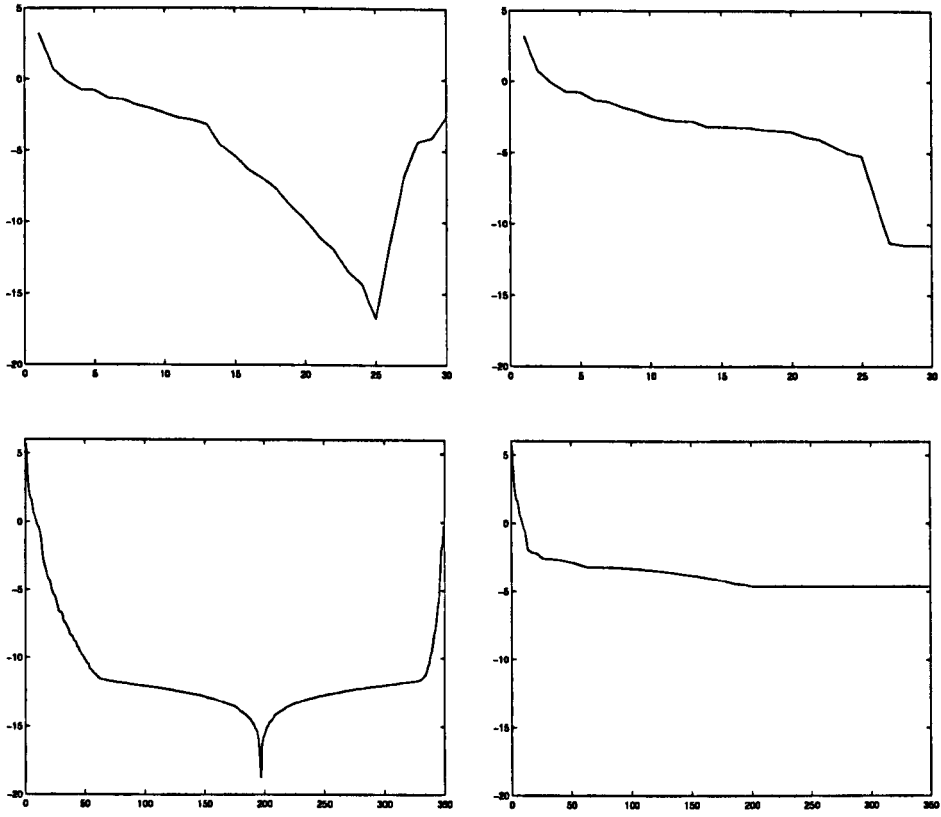


Figure 4.11: Two-dimesional interpolating surface spline: Eigenvalues of the interpolated matrix (left), and eigenvalues of the nearest correlation matrix, with eigenvalues bounded by bound as in Table 4.5.

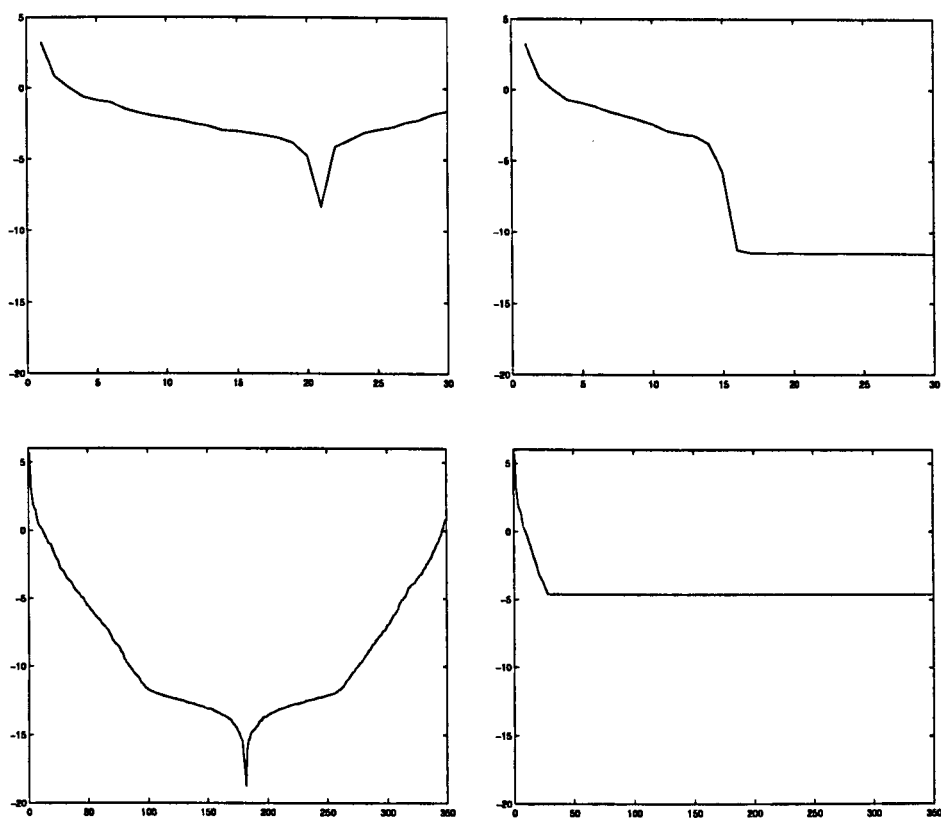


Figure 4.12: Shepard interpolation: Eigenvalues of interpolated matrix (left), and eigenvalues of the nearest correlation matrix, with eigenvalues bounded by bound as in Table 4.6.

4.6.2 Tables

Table 4.7: Descriptive statistics for the yield differences.

Yield	Mean	St.Dev.	Min	Median	Max	Skewness	Kurtosis	Kol.-Smir.	Sig.
1	8.026e-06	0.0005881	-0.004192	0	0.005550	-0.0081	8.853	0.105	0
2	9.665e-06	0.0006700	-0.009466	0	0.009183	-0.1774	33.39	0.095	0
3	1.064e-05	0.0006463	-0.005105	0	0.005011	-0.1068	7.663	0.086	0
4	1.171e-05	0.0006092	-0.005575	0	0.005481	-0.2679	7.386	0.075	0
5	1.191e-05	0.0006215	-0.005666	0	0.005572	-0.2483	7.787	0.081	0
6	1.232e-05	0.0006173	-0.006062	0	0.005945	-0.4134	14.32	0.080	0
7	1.231e-05	0.0006055	-0.008597	0	0.006317	-0.8962	22.22	0.084	0
8	1.226e-05	0.0006078	-0.008495	0	0.006785	-0.7141	24.74	0.081	0
9	1.223e-05	0.0006116	-0.008585	0	0.006593	-0.8266	26.93	0.089	0
10	1.204e-05	0.0005994	-0.008674	0	0.006683	-0.9600	26.54	0.084	0
15	1.119e-05	0.0005704	-0.008652	0	0.007420	-0.5779	35.84	0.089	0
20	1.061e-05	0.0005674	-0.007881	0	0.007881	0.1255	38.84	0.088	0
30	9.215e-06	0.0006070	-0.009235	0	0.009516	0.2146	55.03	0.100	0

Table 4.8: Sample covariance matrix for the first half of the sample, multiplied by a factor $1e+7$.

	1	2	3	4	5	6	7	8	9	10	15	20	30
1	3.62												
2	3.43	5.81											
3	3.39	3.57	4.79										
4	3.12	3.53	3.74	3.93									
5	3.08	3.49	3.71	3.79	4.03								
6	2.80	3.21	3.40	3.45	3.53	3.45							
7	2.60	3.00	3.22	3.33	3.34	3.23	3.38						
8	2.54	2.91	3.12	3.23	3.27	3.14	3.11	3.44					
9	2.47	2.86	3.09	3.18	3.20	3.10	3.07	3.01	3.06				
10	2.40	2.87	2.87	3.09	3.12	3.01	2.99	2.94	2.91	3.32			
15	1.90	2.18	2.39	2.51	2.58	2.51	2.50	2.47	2.46	2.43	2.54		
20	1.66	1.94	2.04	2.28	2.34	2.31	2.32	2.30	2.28	2.28	2.17	2.28	
30	1.49	1.84	2.02	2.12	2.20	2.17	2.20	2.17	2.16	2.15	1.95	2.00	2.46

Table 4.9: Sample correlation matrix for the first half of the sample.

	1	2	3	4	5	6	7	8	9	10	15	20
2	0.749											
3	0.815	0.677										
4	0.826	0.739	0.862									
5	0.806	0.721	0.845	0.953								
6	0.794	0.717	0.837	0.937	0.947							
7	0.745	0.677	0.802	0.913	0.906	0.946						
8	0.721	0.651	0.768	0.879	0.879	0.911	0.911					
9	0.742	0.679	0.807	0.916	0.912	0.953	0.955	0.928				
10	0.693	0.654	0.720	0.855	0.852	0.891	0.892	0.871	0.913			
15	0.628	0.569	0.686	0.795	0.806	0.848	0.856	0.836	0.882	0.836		
20	0.577	0.533	0.618	0.763	0.771	0.823	0.836	0.820	0.865	0.828	0.902	
30	0.500	0.486	0.587	0.682	0.699	0.745	0.762	0.746	0.787	0.754	0.782	0.843

Table 4.10: Approximation of the correlation surface with two-dimensional cubic splines. Approximation based on yearly grid.

	1	2	3	4	5	6	7	8	9	10	15	20
2	0.749 (0.00)											
3	0.815 (0.00)	0.677 (0.01)										
4	0.826 (0.00)	0.739 (0.01)	0.862 (0.00)									
5	0.806 (0.00)	0.721 (0.01)	0.845 (0.00)	0.953 (0.01)								
6	0.794 (0.00)	0.717 (0.01)	0.838 (0.01)	0.938 (0.02)	0.947 (0.01)							
7	0.745 (0.00)	0.677 (0.01)	0.802 (0.02)	0.913 (0.03)	0.906 (0.02)	0.946 (0.04)						
8	0.721 (0.02)	0.651 (0.02)	0.768 (0.03)	0.879 (0.00)	0.879 (0.01)	0.911 (0.01)	0.911 (0.03)					
9	0.742 (0.02)	0.679 (0.00)	0.808 (0.08)	0.917 (0.07)	0.912 (0.04)	0.954 (0.11)	0.957 (0.19)	0.927 (0.14)				
10	0.694 (0.15)	0.655 (0.17)	0.717 (0.38)	0.854 (0.11)	0.852 (0.01)	0.889 (0.21)	0.888 (0.49)	0.878 (0.79)	0.895 (1.88)			
15	0.628 (0.01)	0.569 (0.02)	0.687 (0.03)	0.795 (0.03)	0.807 (0.07)	0.849 (0.14)	0.858 (0.18)	0.837 (0.10)	0.886 (0.38)	0.837 (0.14)		
20	0.577 (0.01)	0.533 (0.02)	0.618 (0.00)	0.764 (0.02)	0.772 (0.02)	0.824 (0.05)	0.837 (0.07)	0.820 (0.00)	0.867 (0.18)	0.826 (0.27)	0.907 (0.52)	
30	0.500 (0.01)	0.486 (0.01)	0.587 (0.02)	0.682 (0.01)	0.699 (0.00)	0.745 (0.00)	0.762 (0.01)	0.746 (0.02)	0.787 (0.04)	0.752 (0.20)	0.781 (0.10)	0.844 (0.04)

Table 4.11: Approximation of the correlation surface with two-dimensional cubic splines. Approximation based on monthly grid.

	1	2	3	4	5	6	7	8	9	10	15	20
2	0.750 (0.24)											
3	0.816 (0.18)	0.675 (0.28)										
4	0.826 (0.05)	0.739 (0.04)	0.865 (0.30)									
5	0.805 (0.07)	0.722 (0.09)	0.843 (0.22)	0.956 (0.35)								
6	0.794 (0.02)	0.717 (0.02)	0.838 (0.10)	0.937 (0.08)	0.947 (0.03)							
7	0.745 (0.02)	0.677 (0.04)	0.801 (0.03)	0.914 (0.11)	0.905 (0.09)	0.949 (0.35)						
8	0.721 (0.02)	0.651 (0.01)	0.768 (0.00)	0.879 (0.03)	0.879 (0.03)	0.911 (0.10)	0.911 (0.06)					
9	0.742 (0.00)	0.679 (0.00)	0.807 (0.06)	0.917 (0.04)	0.912 (0.03)	0.954 (0.07)	0.956 (0.13)	0.927 (0.15)				
10	0.694 (0.24)	0.655 (0.13)	0.718 (0.29)	0.855 (0.05)	0.852 (0.04)	0.890 (0.11)	0.888 (0.51)	0.879 (0.94)	0.895 (1.97)			
15	0.628 (0.01)	0.569 (0.02)	0.687 (0.03)	0.795 (0.02)	0.807 (0.05)	0.849 (0.09)	0.857 (0.14)	0.836 (0.04)	0.885 (0.25)	0.837 (0.09)		
20	0.577 (0.01)	0.533 (0.01)	0.618 (0.01)	0.763 (0.01)	0.771 (0.02)	0.823 (0.03)	0.837 (0.06)	0.820 (0.02)	0.866 (0.13)	0.826 (0.21)	0.905 (0.33)	
30	0.500 (0.01)	0.486 (0.02)	0.587 (0.05)	0.682 (0.02)	0.699 (0.00)	0.745 (0.00)	0.762 (0.01)	0.746 (0.03)	0.787 (0.01)	0.752 (0.20)	0.780 (0.22)	0.846 (0.28)

Table 4.12: Approximation of the correlation surface with two-dimensional surface splines. Approximation based on yearly grid.

	1	2	3	4	5	6	7	8	9	10	15	20
2	0.749 (0.06)											
3	0.815 (0.04)	0.679 (0.30)										
4	0.824 (0.13)	0.733 (0.83)	0.858 (0.50)									
5	0.806 (0.00)	0.722 (0.12)	0.845 (0.07)	0.951 (0.24)								
6	0.793 (0.09)	0.715 (0.32)	0.836 (0.19)	0.943 (0.55)	0.947 (0.06)							
7	0.744 (0.03)	0.676 (0.04)	0.802 (0.02)	0.913 (0.08)	0.906 (0.00)	0.947 (0.06)						
8	0.721 (0.05)	0.653 (0.41)	0.770 (0.24)	0.873 (0.65)	0.880 (0.10)	0.909 (0.22)	0.911 (0.02)					
9	0.742 (0.05)	0.677 (0.21)	0.806 (0.13)	0.918 (0.20)	0.912 (0.06)	0.953 (0.04)	0.954 (0.08)	0.927 (0.16)				
10	0.693 (0.12)	0.657 (0.46)	0.722 (0.28)	0.848 (0.88)	0.853 (0.14)	0.888 (0.37)	0.892 (0.09)	0.874 (0.31)	0.913 (0.09)			
15	0.628 (0.06)	0.573 (0.75)	0.689 (0.41)	0.786 (1.19)	0.808 (0.25)	0.845 (0.33)	0.856 (0.01)	0.840 (0.50)	0.879 (0.32)	0.841 (0.56)		
20	0.577 (0.10)	0.541 (1.51)	0.624 (0.96)	0.746 (2.25)	0.775 (0.43)	0.818 (0.60)	0.836 (0.01)	0.828 (0.93)	0.859 (0.72)	0.835 (0.83)	0.914 (1.24)	
30	0.501 (0.23)	0.491 (1.15)	0.591 (0.66)	0.671 (1.62)	0.700 (0.13)	0.740 (0.62)	0.762 (0.08)	0.751 (0.67)	0.784 (0.35)	0.759 (0.70)	0.789 (0.97)	0.856 (1.52)

Table 4.13: Approximation of the correlation surface with two-dimensional surface splines. Approximation based on monthly grid.

	1	2	3	4	5	6	7	8	9	10	15	20
2	0.751 (0.37)											
3	0.816 (0.11)	0.682 (0.80)										
4	0.823 (0.37)	0.733 (0.81)	0.859 (0.32)									
5	0.805 (0.10)	0.723 (0.17)	0.845 (0.01)	0.952 (0.11)								
6	0.790 (0.40)	0.715 (0.30)	0.835 (0.32)	0.943 (0.54)	0.948 (0.12)							
7	0.743 (0.16)	0.677 (0.00)	0.801 (0.05)	0.913 (0.00)	0.906 (0.05)	0.949 (0.33)						
8	0.721 (0.02)	0.654 (0.53)	0.771 (0.36)	0.873 (0.73)	0.882 (0.27)	0.909 (0.24)	0.913 (0.23)					
9	0.745 (0.41)	0.678 (0.17)	0.805 (0.28)	0.915 (0.16)	0.912 (0.00)	0.950 (0.33)	0.953 (0.22)	0.929 (0.05)				
10	0.695 (0.41)	0.657 (0.54)	0.723 (0.44)	0.844 (1.33)	0.856 (0.43)	0.886 (0.51)	0.891 (0.12)	0.875 (0.42)	0.918 (0.61)			
15	0.627 (0.18)	0.574 (0.96)	0.690 (0.52)	0.783 (1.51)	0.811 (0.54)	0.845 (0.34)	0.856 (0.01)	0.842 (0.77)	0.878 (0.51)	0.843 (0.86)		
20	0.576 (0.15)	0.541 (1.41)	0.625 (1.13)	0.747 (2.22)	0.776 (0.64)	0.819 (0.57)	0.836 (0.01)	0.829 (1.02)	0.858 (0.85)	0.835 (0.82)	0.912 (1.10)	
30	0.501 (0.28)	0.493 (1.51)	0.593 (0.99)	0.669 (1.93)	0.701 (0.28)	0.739 (0.75)	0.761 (0.17)	0.753 (1.02)	0.781 (0.73)	0.760 (0.81)	0.789 (0.99)	0.856 (1.56)

Table 4.14: Approximation of the correlation surface with Shepard interpolation. Approximation based on yearly grid.

	1	2	3	4	5	6	7	8	9	10	15	20
1												
2	0.749 (0.06)											
3	0.815 (0.02)	0.680 (0.46)										
4	0.826 (0.08)	0.732 (0.87)	0.858 (0.44)									
5	0.805 (0.13)	0.722 (0.04)	0.845 (0.03)	0.951 (0.17)								
6	0.793 (0.05)	0.714 (0.45)	0.836 (0.22)	0.941 (0.40)	0.947 (0.02)							
7	0.744 (0.05)	0.675 (0.28)	0.801 (0.13)	0.915 (0.22)	0.906 (0.05)	0.947 (0.15)						
8	0.720 (0.09)	0.652 (0.28)	0.769 (0.14)	0.876 (0.35)	0.880 (0.14)	0.911 (0.08)	0.911 (0.02)					
9	0.742 (0.02)	0.676 (0.49)	0.805 (0.30)	0.921 (0.51)	0.912 (0.04)	0.956 (0.24)	0.956 (0.15)	0.927 (0.11)				
10	0.692 (0.10)	0.658 (0.57)	0.722 (0.24)	0.849 (0.70)	0.855 (0.33)	0.890 (0.13)	0.893 (0.02)	0.875 (0.40)	0.914 (0.17)			
15	0.626 (0.40)	0.572 (0.53)	0.689 (0.34)	0.788 (0.87)	0.809 (0.37)	0.846 (0.24)	0.855 (0.11)	0.839 (0.33)	0.877 (0.58)	0.835 (0.10)		
20	0.577 (0.02)	0.542 (1.66)	0.624 (0.94)	0.751 (1.65)	0.773 (0.16)	0.818 (0.66)	0.834 (0.34)	0.824 (0.41)	0.858 (0.89)	0.832 (0.55)	0.906 (0.38)	
30	0.502 (0.49)	0.492 (1.26)	0.590 (0.58)	0.676 (0.86)	0.698 (0.15)	0.740 (0.65)	0.759 (0.47)	0.746 (0.07)	0.780 (0.80)	0.756 (0.28)	0.781 (0.09)	0.846 (0.30)

Table 4.15: Approximation of the correlation surface with Shepard interpolation. Approximation based on monthly grid.

	1	2	3	4	5	6	7	8	9	10	15	20
2	0.747 (0.24)											
3	0.816 (0.15)	0.677 (0.10)										
4	0.827 (0.16)	0.731 (1.01)	0.858 (0.48)									
5	0.805 (0.08)	0.720 (0.11)	0.845 (0.03)	0.952 (0.07)								
6	0.794 (0.01)	0.713 (0.60)	0.835 (0.24)	0.942 (0.51)	0.948 (0.06)							
7	0.745 (0.01)	0.674 (0.45)	0.801 (0.14)	0.916 (0.33)	0.906 (0.06)	0.948 (0.22)						
8	0.720 (0.09)	0.651 (0.14)	0.769 (0.13)	0.876 (0.31)	0.880 (0.12)	0.911 (0.09)	0.911 (0.07)					
9	0.742 (0.04)	0.675 (0.58)	0.805 (0.26)	0.921 (0.51)	0.913 (0.06)	0.956 (0.28)	0.957 (0.19)	0.927 (0.10)				
10	0.691 (0.22)	0.658 (0.57)	0.721 (0.14)	0.849 (0.77)	0.854 (0.27)	0.889 (0.23)	0.892 (0.07)	0.874 (0.37)	0.915 (0.29)			
15	0.626 (0.38)	0.570 (0.18)	0.689 (0.40)	0.789 (0.75)	0.809 (0.31)	0.846 (0.18)	0.856 (0.05)	0.839 (0.35)	0.880 (0.28)	0.839 (0.30)		
20	0.578 (0.22)	0.540 (1.17)	0.624 (0.85)	0.754 (1.29)	0.772 (0.15)	0.819 (0.51)	0.835 (0.19)	0.823 (0.38)	0.860 (0.62)	0.832 (0.55)	0.905 (0.30)	
30	0.503 (0.53)	0.490 (0.81)	0.591 (0.72)	0.678 (0.57)	0.698 (0.12)	0.741 (0.56)	0.760 (0.36)	0.746 (0.07)	0.782 (0.59)	0.756 (0.26)	0.782 (0.02)	0.847 (0.44)

Chapter 5

SPD Correlation Function with Functional Deformation of Maturity

5.1 Introduction

In this chapter we continue the study of strictly positive definite instantaneous covariance and correlation functions. In the Chapter 4 we developed two alternative non-parametric methodologies for constructing SPD covariance and correlation functions for Gaussian random fields. We now turn our attention to semi-parametric or fully-parametric forms of these functions. In particular, we develop a general class of Gaussian random field models that satisfies our modelling assumption. I.e. the covariance and correlation functions of this class are strictly positive definite and fit the observed sample covariance and correlation matrices.

We suggest a class of correlation functions of the form,

$$\text{Corr}(s, t) = C(g(s) - g(t)), \quad (5.1.1)$$

where $g : \mathbb{R} \rightarrow \mathbb{R}$ some strictly increasing function and C is an SPD function.

The map g acts as a deformation from the original “maturity” to an alternative “deformation” space. This approach will yield a large, sufficiently flexible class of non-stationary Gaussian fields with strictly positive definite covariance function. I.e. this field will exhibit an infinite-dimensional factor structure.

To make this approach practical we identify a class of strictly positive definite functions. Furthermore, we find sufficient conditions on the deformation functions g , so that these deformations leave the SPD property of the function invariant. Furthermore, we implement our methodology and test it on Japanese Yen interest data, We also provide a discussion of the advantages and disadvantages of this methodology.

Here is a summary of the chapter. In Section 5.2 we describe the modelling framework and give some motivation for the class of models introduced. A large class of SPD functions together with the maturity deformation functions are characterised in Section 5.3. We also provide several parametric examples of SPD functions and discuss their properties. The data set used for the empirical analysis is discussed in Section 5.4. In the following three sections we develop procedures for implementation and parameter estimation of the SPD functions and deformation mappings. In particular, in Section 5.5 we develop a two-step implementation procedure. In the first step we estimate the deformation functions using the non-metric Shepard-Kruskal MDS-Algorithm to find the deformation function, and in the second we use a very general SPD function to fit the resulting correlation function to the correlation sample matrix. In Section 5.6 we represent the deformation function as a linear combination of some basis functions. This allows a simultaneous fit to the sample correlation function using the coefficients of the basis functions and the parameters of the SPD functions. In Section 5.7 we develop an iterative two step procedure. The first step is based on a modified downhill simplex which provides a very efficient search tool for the deformation function, and the second step is based on minimisation of the SPD function parameters. Additionally, we test all three approaches on Japanese Yen interest rate data. In Section 5.8 we discuss alternative estimation

techniques, such as maximum likelihood and Bayesian methods, that can be used in our framework. We conclude and provide a discussion for further research in Section 5.9.

5.2 Modelling Framework

We start this section by highlighting the difficulties in modelling functional forms of instantaneous SPD covariance and correlation functions. We present several examples of SPD functions which we use as a motivation for our modelling framework. We then present the modelling framework itself in terms of a generic SPD correlation function. Finally, we discuss other approaches for constructing correlation functions and matrices which have appeared in the research literature and can be related to our framework.

It is well known from empirical observations of the historical sample and implied correlation matrices that they exhibit quite complex shapes. For example, the changes in yields are more highly correlated for long maturities than they are for short maturities. The so called “decorrelation”, i.e. the decline of the correlation off the diagonal, is faster for maturities close to each other and slows down for maturities farther apart, etc. These observations indicate that examples of field models, suggested by Kennedy (1994), Goldstein (2000), and Santa-Clara and Sornette (2001), which we reviewed in Chapter 3, are not flexible enough to take into account this behaviour of the instantaneous correlation function.

Ideally, we would like to have an SPD kernel with sufficient flexibility in parameters so that it fits the sample correlation matrix and satisfies some smoothness requirements. Unfortunately, there is a limited number of known SPD kernels. Moreover, it is difficult, in general, to show whether a given function is SPD or not. There has been considerable interest throughout last century from both pure and applied mathematicians in positive and strictly positive definite functions and kernels. An

extensive treatment of these properties can be found in a seminal work by Karlin (1968). The recent book by Sasvári (1994) contains a theoretical exposition of positive definite functions. For a historical review of positive definite functions and kernels see Stewart (1976). We provide a short discussion of the original study of these properties in Appendix 5.10.3.

A simple example of an SPD function on $(-\infty, \infty)$ is

$$K(s, t) = \exp(st). \quad (5.2.1)$$

The matrix resulting from this function is a generalised Vandermonde matrix, the determinant of which is always positive. Since any minor of a generalised Vandermonde matrix is also a generalised Vandermonde matrix, the function $K(s, t)$ is SPD. Moreover, for $\phi(x)$ strictly increasing, the function

$$K(s, t) = \exp(\phi(s), \phi(t))$$

is SPD as well. We will justify this observation later in the chapter. This example provides us with a general strategy: find a suitable kernel and then deform the coordinates using a strictly positive function. If we assume that function (5.2.1) is a covariance function then the corresponding correlation function is of the form

$$\text{Corr}(s, t) = \exp\left(-\frac{1}{2}(s - t)^2\right).$$

This correlation function has been obtained from the covariance function by scaling with variance functions. Thus we observe that an SPD function remains SPD after scaling with a positive function. Note, the last correlation is a function of s and t only via $|s - t|$. We will see later, that SPD functions with this property are easier to find and to characterise than those which depend on both arguments separately.

Other examples of SPD functions are the covariance and correlation of standard Brownian motion,

$$\text{Cov}(s, t) = \min(s, t), \quad \text{Corr}(s, t) = \frac{\min(s, t)}{\sqrt{st}},$$

and of the Ornstein-Uhlenbeck process, with initial random variable having a normal distribution with mean zero and variance $\sigma^2/2\alpha$,

$$\text{Cov}(s, t) = (\sigma^2/2\alpha) \exp(-|t - s|), \quad \text{Corr}(s, t) = \exp(-|t - s|).$$

Note again, in the last example the function correlation and covariance functions depend on s and t only via $|t - s|$.

In the above examples we have noted the frequent appearance of SPD functions that depend on their argument only through their difference. This is not accidental. The situation with positive definite (PD) functions which depend on their arguments through their difference is not as bad as with general PD kernels. This class of functions can be characterised either in terms of Fourier transforms of finite Borel measures, or using the theory of completely monotone functions. Furthermore, one can derive sufficient condition to restrict this class further to include only SPD functions. We will say more on this in the next section. We also have observed that one can depart from the dependence of the arguments on their difference through deformation of the coordinates via a strictly increasing transformation.

Drawing on these observations we postulate the following generic form for the instantaneous SPD covariance function,

$$\text{Cov}(s, t) = \sigma(s)\sigma(t)C(g(s) - g(t)), \quad (5.2.2)$$

where σ^2 is a strictly positive variance function and C an SPD function, with the function g acting as a deformation mapping. We assume that we have methods for constructing the variance function σ^2 . This can be done by various means. E.g., one can use some parametric family of functions, or non-parametric techniques such as splines¹. Thus we restrict our attention only to the instantaneous correlation function. This is because it is well known in fixed income markets that the variance

¹For a summary of possible variance function construction techniques see, for example, James and Webber (2000).

depends on maturity. Consequently, by modelling the correlation function we reduce the amount of non-stationarity we need to capture, and can concentrate on capturing purely spatial dependence.

Remark 5.2.1. Rebonato (1999b), within the pricing framework of calibration of market models, suggested a procedure for constructing a correlation matrix. A particular case of his procedure, the two factor model², can be seen as a degenerate example of our approach. In particular, his two-factor correlation matrix is given by

$$\rho_{ij} = \cos(\theta_i - \theta_j),$$

for some angles θ_i , $i = 1, \dots, n$. The continuous time-equivalent of this matrix is a correlation function,

$$\rho(t, s) = \cos(f(t) - f(s)),$$

with the deformation function $f(t)$ being a continuous time-equivalent of the discrete mapping $i \rightarrow \theta_i$.

Remark 5.2.2. Kurbanmuradov *et al.* (2000), again within the pricing framework of calibration market models suggested³ the following form for the implied correlation matrix,

$$\rho_{ij} = \frac{\min(b_i, b_j)}{\max(b_i, b_j)}, \quad (5.2.3)$$

with

$$b_i = \exp[\beta(i-1)^\alpha], \quad \beta > 0, 0 < \alpha < 1. \quad (5.2.4)$$

One can arrive at this function by considering the covariance function of the standard Brownian sheet W_{st} ,

$$\text{Cov}(W_{s_1, t_1}, W_{s_2, t_2}) = \min(s_1, s_2) \min(t_1, t_2), \quad s_i \geq 0, t_i \geq 0, i = 1, 2.$$

²More details on this approach can be found in Chapter 6, Remark 6.2.1.

³See also Coffey and Schoenmakers (1999) and Coffey and Schoenmakers (2000) for further details.

If we restrict the field on the diagonal, i.e. for $s = t$ in W_{st} , then the covariance and correlation of W_{tt} become

$$\text{Cov}(W_{ss}, W_{tt}) = \min^2(s, t), \quad \text{Corr}(W_{ss}, W_{tt}) = \frac{\min(s, t)}{\max(s, t)}. \quad (5.2.5)$$

Furthermore, deforming the coordinates of the last correlation function by a strictly monotone mapping $f(\cdot)$ leads to the function

$$K(t, s) = \frac{\min(f(s), f(t))}{\max(f(s), f(t))},$$

which is the continuous time-equivalent of the function (5.2.3). The deformation function f is just the continuous time-equivalent of the discrete time deformation (5.2.4). The function (5.2.5) can also be obtained by a square of the covariance function of the standard Brownian motion,

$$K(t, s) = \min(s, t).$$

The powers of the function can be seen as a special case of the iteration⁴ of functions over a sigma-finite measure. The iterated function is SPD, provided that the starting function is. In the matrix formulation, function iteration corresponds to the Hadamard product of two matrices. The SPD property of the iterated function is then a simple consequence of the Schur⁵ Product Theorem.

In this section we have introduced a general framework for constructing SPD correlation functions. In the next section we characterise a large class of SPD functions which can be used in the construction of the generic SPD correlation function (5.2.2). Furthermore, we provide sufficient conditions on the deformation mapping so that our generic correlation function is well defined.

⁴For more details on iterated functions in particular, and composed functions in general, see Karlin (1968), Chapter 3.

⁵See, for example, Horn and Johnson (1985), Chapter 7.

5.3 Deformation Mappings and Strictly Positive Definite Functions

In the previous section we suggested a generic correlation function of the form,

$$\text{Corr}(t, s) = C(g(t) - g(s)),$$

where C is an SPD function and g is a maturity deformation mapping. For this approach to be useful in practice we need to find examples of such functions. Additionally, we need to derive sufficient conditions on the deformation mapping so that it does not destroy the SPD property of the initial function. In this section we address both of these problems: we characterise a class of SPD functions that depend on their arguments via their difference. We also discuss several examples of SPD functions which have been used in spatial statistics to represent homogeneous random fields. Finally, we justify the use of strictly increasing transformations in our construction.

As we mentioned in the previous section, positive definite (PD) functions which depend on their arguments through their difference are quite common. This class of functions can be characterised in terms of *finite Borel measures*. Consider the n -dimension Borel-measurable space $(\mathbb{R}^n, \mathcal{B}^n)$. We call every finite measure⁶ μ on \mathcal{B}^n , a *finite Borel measure*, and denote the set of these measures by $\mathcal{M}_+^b(\mathbb{R}^n)$. The *Fourier transform*⁷ of a measure $\mu \in \mathcal{M}_+^b(\mathbb{R}^n)$ is the complex function $\hat{\mu}$ on \mathbb{R}^n defined by

$$\hat{\mu}(x) = \int e^{i\langle x, y \rangle} \mu(dy), \quad (x \in \mathbb{R}^n).$$

Bochner (1932) and (1933) characterised all positive definite and continuous functions in terms of the set $\mathcal{M}_+^b(\mathbb{R}^n)$. His results states, that in order a function

⁶A measure μ is called finite if its *total mass*, in our case $\mu(\mathbb{R}^n)$ is finite.

⁷In probability theory one is interested in Fourier transforms of distributions \mathbb{P}_X of random variables X with values in \mathbb{R}^n . In particular, $\mathbb{P}_X \in \mathcal{M}_+^b(\mathbb{R}^n)$ for $b = 1$, and $\hat{\mathbb{P}}_X$ is called the *characteristic function* of X . It follows then that continuous positive definite functions are basically characteristic functions.

$f : \mathbb{R}^n \rightarrow \mathbb{C}$ be positive definite and continuous, it is necessary and sufficient that it be the Fourier transform of a nonnegative finite-valued Borel measure on \mathbb{R}^n .

For our construction (5.2.2) we need a subclass of the PD functions, namely the SPD functions. Recently Chang (1996) and Cheney and Light (2000), pp. 87-92, have strengthened the result of Bochner and derived sufficient conditions for the functions on \mathbb{R}^n to be strictly positive definite. For the case $n = 1$, their result states for f a nonnegative Borel function on \mathbb{R} , such that $0 < \int_{\mathbb{R}} f < \infty$, the \hat{f} is strictly positive definite.

Thus one can find strictly positive definite functions by using the above result and a table of Fourier transforms. The pairs shown in Table 5.1, from Oberhettinger (1973), give some examples⁸.

Table 5.1: Strictly positive definite functions on \mathbb{R}

$f(x) = \begin{cases} 1/2 & x \leq 1 \\ 0 & x > 1 \end{cases}$	$\hat{f} = x^{-1} \sin x$
$f = (1 + x^2)^{-2}/\pi$	$\hat{f} = e^{- x }$
$f = e^{- x }/2$	$\hat{f} = (1 + x^2)^{-1}$
$f = \pi^{-1/2} e^{-x^2}$	$\hat{f} = e^{-x^2/4}$
$f = (1/2\pi)(1 + x^{-2})$	$\hat{f} = x ^{-1}(1 - e^{- x })$
$f = \operatorname{sech}(\pi x)$	$\hat{f} = \operatorname{sech}(y/2)$
$f = (2x)^{-1}(x^{-1} - \operatorname{csch} x)$	$\hat{f} = \log(1 + e^{-\pi/ x })$

Another way to obtain SPD functions is to use the theory of complete monotone functions. A function is said to be *completely monotone* on $[0, \infty)$ if three conditions are satisfied:

1. $f \in C[0, \infty)$,

⁸In that volume the Fourier transform is defined by $\hat{f} = \int_{-\infty}^{\infty} f(y)e^{ixy}dy$.

2. $f \in C^\infty[0, \infty)$,

3. $(-1)^k f^{(k)} \geq 0$ for $t > 0$ and $k = 0, 1, 2, \dots$

Such functions exist in great abundance. Here are some examples that can be quickly verified directly from the definition:

1. $f(t) = a$, where $a \geq 0$,

2. $f(t) = (t + a)^b$, where $a > 0 \geq b$,

3. $f(t) = e^{-at}$, where $a \geq 0$.

The class of complete monotone functions can be characterised in terms of the Laplace transforms. I.e. a function is completely monotone if and only if it is the Laplace transform of a nonnegative bounded Borel measure⁹. This result can be formulated in terms of the Riemann-Stieltjes integral: A function $f : [0, \infty) \rightarrow [0, \infty)$ is completely monotone if and only if there is a nondecreasing bounded function γ such that $f(t) = \int_0^\infty e^{-st} d\gamma(s)$.

The complete monotone functions in turn can be used to find the strictly positive definite functions. This result is due to Schoenberg (1938). It states that if f is completely monotone but not constant on $[0, \infty)$, then the function $x \mapsto f(\|x\|^2)$ is strictly positive definite function.

Next, we discuss several examples of SPD functions. These examples come from the spatial statistics literature. They are usually used to model the correlation functions of isotropic and homogeneous processes¹⁰.

Exponential:

$$C(h) = \exp(-a^2 \|h\|), \quad h \in \mathbb{R}^n.$$

⁹This is the Bernstein and Widder theorem.

¹⁰A comprehensive review of this type of models can be found, for example, in Cressie (1993).

Figure 5.10.4 (a) plots the exponential correlation function for several parameters a . For lower values of the parameter a the correlation exhibit a fatter tail, but it maintains a convex shape.

Exponential-Power:

$$C(h) = \exp(-a\|h\|^b), \quad h \in \mathbb{R}^n,$$

where $0 < b \leq 2$ and $a \geq 0$. This class of correlation functions generalises the exponential form. Figure 5.10.4 (b) plots the Exponential-Power correlation function for fixed a and several parameters b . For smaller values of the exponent the correlation function exhibits a larger tail and is increasingly convex at the origin. However, for higher values of the exponent, the tail tends to be smaller and the function becomes concave at the origin.

Gaussian:

$$C(h) = \exp(-a\|h\|^2), \quad h \in \mathbb{R}^n.$$

This correlation is a special case of Exponential-Power with $b = 2$. It is called a *Gaussian correlation function* because of its functional form. The corresponding spectral density is given by

$$g(u) = cu^{d-1} \exp(-u^2/4a^2).$$

Triangular: has the form

$$C(h) = (1 - a\|h\|)^+, \quad h \in \mathbb{R}^n,$$

and its plot can be seen in Figure 5.10.4 (c).

Power Law: has the form

$$C(h) = (1 - b\|h\|^a)^+, \quad h \in \mathbb{R}^n,$$

Figure 5.10.4 (d) plots this function for several parameters of the exponent. The function changes shape from convex to concave as the parameter goes from 0 to 2. Positive definiteness requires $0 \leq a < 2$.

The Matérn class: This class of isotropic covariance functions was originally given by Matérn (1986). There are several parameterisations of this family of covariance functions. We use a parametrisation given by Handcock and Wallis (1994). They write the isotropic spectral density as

$$g_{\eta}(u) = \frac{\sigma c(\nu, \rho)}{\left(\frac{4\nu}{\rho^2} + u^2\right)^{\nu+d/2}},$$

where $\eta = (\sigma, \nu, \rho)$, with $\sigma, \nu, \rho > 0$, and

$$c(\nu, \rho) = \frac{\Gamma(\nu + \frac{n}{2})(4\nu)^{\nu}}{\pi^{d/2}\Gamma(\nu)\rho^{2\nu}}.$$

The corresponding covariance function is

$$C^{\eta}(h) = \frac{\sigma}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\nu^{1/2}h}{\rho}\right)^{\nu} \mathcal{K}_{\nu}\left(\frac{2\nu^{1/2}h}{\rho}\right), \quad h \in \mathbb{R}^n, \quad (5.3.1)$$

which has the nice property that it does not depend on n . The function $\Gamma(\cdot)$ is the usual gamma function and \mathcal{K}_{ν} is the modified Bessel function of order ν . The parameter σ is just variance of the field, and ρ measures how quickly the correlations of the random field decay with distance. Figure 5.10.4 (e) plots Matérn covariance functions for $\sigma = 1$, $\nu = 1$ and several values of ρ , and shows that decorrelation occurs more slowly as ρ increases. The influence of the parameter ν , for fixed σ and ρ , is shown in Figure 5.10.4 (f). The special case $\nu = 0.5$ corresponds to the exponential correlation function, and the limit as $\nu \rightarrow \infty$ results in the Gaussian form.

Additionally, it is important to mention for later use, that the class of covariance functions is closed under addition, multiplication, and passages to the limit; see

Loève (1960), p.468, for the proof and several examples. Through use of the closure property one obtains, for example, a general Gaussian correlation function,

$$C(h) = \int_{-\infty}^{\infty} \exp(-h^2 t^2) dF(t), \quad (5.3.2)$$

where $F(t)$ is an arbitrary one-dimensional distribution function. This correlation function is positive definite in every dimension simultaneously.

Next we want to determine what kind of maturity transformations are allowed. Starting from an SPD function, we are only interested in deformations that do not distort the SPD property. The next theorem can be seen as a special case of Theorem 2.1, in Karlin (1968).

Theorem 5.3.1. *Let $K(x, y)$ be an SPD-function and $x, y \in [0, T]$.*

1. *Let $\phi(x)$ be positive on $[0, T]$, then $L(x, y) = \phi(x)\phi(y)K(x, y)$ is an SPD-function.*
2. *Additionally, let ϕ^{-1} be a strictly increasing function mapping $[0, T]$ onto $[0, S]$, where ϕ^{-1} is the inverse function of ϕ . Denote $u = \phi^{-1}(x)$ and $v = \phi^{-1}(y)$. Then $L(u, v) = K[\phi(x), \phi(y)]$ is SPD.*

Theorem 5.3.1 shows, in particular, that we are allowed to deform the maturity by a strictly monotone function. The correlation function $C(|f(s) - f(t)|)$ will remain SPD.

To summarise, in this section we have described sufficient conditions for the functions to be strictly positive definite, in the case when they depend on the arguments via their difference. Moreover, we have shown that if the deformation functions are strictly monotonic they will not destroy the SPD property of a function. Furthermore, we have provided several examples of PD and SPD functions that have been used in spatial statistics applications. We will use some of these examples in the implementation and testing of our methodology. In the next section we discuss the data set used for the testing.

5.4 Data

Later in the chapter we will study the implementation procedures for our generic SPD correlation function (5.2.2). To test the flexibility of this framework we use the sample correlation matrix arising from Japanese Yen for the period 2/12/96 to 25/10/01. The data is courtesy of LCH, London, and has been obtained by stripping the Japanese government issues.

In Table 5.2, we provide the descriptive statistics of the whole sample of yield differences. Observe that the distribution of the yield differences for each maturity is centred at zero, and it has a very small standard deviation. However, they are not normal, since they are skewed with very fat tails. Skewness is positive for all maturities and it generally decreases for higher maturities. The yield differences show a high degree of kurtosis that is roughly constant, with the exception of three and four years to maturity.

In Figure 5.2 we plotted the sample covariance and correlation matrices. The plot of the sample correlation shows an expected decreasing behaviour: the further the distance between any two maturities, the lower the correlation. The actual values of sample covariance and correlation are in Tables 5.3 and 5.4. Observe the high correlation between maturities. The figures fluctuate between 0.659 and 0.998. Also observe that the drop in correlation closer to the diagonal is more significant than those correlations further away from the diagonal.

In the next three sections we study several implementation procedures for our generic SPD correlation function. Their flexibility in fitting sample correlation matrices will be tested on the Japanese Yen interest rate data discussed above.

5.5 Sampson and Guttorp Approach

This section is the first of three in which we implement the framework suggested above for the construction of instantaneous SPD correlation functions. The implementation idea comes from the work pioneered by Sampson and Guttorp (1992) in the field of spatial statistics. Traditionally, models in spatial statistics were based on homogeneous fields, i.e. fields based on correlation functions which depend on their arguments only via their difference. Sampson and Guttorp (1992) suggested an extension to non-homogeneous fields. We adapt their procedure to our problem. In the first part of this section we describe the implementation. Then, in the second part we test the methodology on the Japanese Yen interest rate data and discuss the results.

5.5.1 Description of the Method

The generic form of the instantaneous correlation function (5.1.1) which we study in this chapter consists of two ingredients: an SPD function and a strictly increasing deformation mapping. The idea of the implementation method is to decouple the choice of these two ingredients into two steps: i.e. first find a deformation mapping under the assumption of a very general correlation function, then, for a fixed deformation mapping, find the best correlation function.

First we find the deformation mapping of n maturities, T_1, \dots, T_n , by applying the non-metric Shepard-Kruskal MDS-Algorithm, as described, for example, in Mardia *et al.* (1979), or Cox and Cox (2001). This algorithm determines a monotone function δ such that the matrix of $\delta_{ij} = \delta(d_{ij})$ can be accurately represented by inter-point distances in a one-dimensional metric rescaling. That is, it determines a configuration of maturities T_i so that

$$\delta(d_{ij}) = \delta_{ij} \approx |T_i - T_j|.$$

Solving this relationship for d_{ij} we obtain a function g that serves a role analogous to the usual variogram function. For a homogeneous field a variogram function γ can be written as $\gamma(h) = 1 - \rho(h)$, where ρ is the correlation of the field. The function g can be thought of as applied to distances in the deformed maturity space D ,

$$d_{ij}^2 = (\delta^{-1})^2 \approx g(|T_i - T_j|).$$

The multidimensional scaling algorithm proceeds by searching a configuration of maturities T_i , so that the distances $h_{ij} = |T_i - T_j|$ minimise a stress criterion defined as

$$\min_{\delta} \frac{\sum_{i < j} [\delta(d_{jk}) - h_{ij}]^2}{\sum_{i < j} h_{ij}^2},$$

where the minimum is taken over all monotone functions. We obtain a monotonic maturity deformation mapping on a discrete set of n maturities, T_1, \dots, T_n . To obtain the mapping for a continuum of maturities, we can extend the deformation to the entire maturity space M . This can be done in various ways, but we must bear in mind that this mapping has to be strictly increasing.

In the first step we found a maturity deformation mapping under the assumption of a general monotonic function g which played the role of variogram, or equivalently correlation function in the deformed space. To finish our construction we need to find an SPD function to replace g . Thus, the second step consists in finding an SPD function which depends on its arguments only through their difference. To do this, we choose a general Gaussian correlation function, which we described in Section 5.3,

$$C(h) = \int_{-\infty}^{\infty} \exp(-h^2 t^2) dF(t),$$

where $F(t)$ is an arbitrary one-dimensional distribution function.

Next, for a given vector (T_1, \dots, T_n) of deformation we determine a least squares fit for the observed sample correlation matrix in terms of a discrete distribution F , spec-

ified by its support t_1, \dots, t_n , and corresponding nonnegative numbers $\alpha_1, \dots, \alpha_n$. We provide the details in Appendix 5.10.1.

In the next section we test our procedure on a sample correlation matrix calculated from the Japanese Yen interest rate data.

5.5.2 Numerical Results

To test the methodology described above we used the sample correlation matrix obtained from the data discussed in Section 5.4. The original Sampson-Guttorp approach using a general correlation function of the type (5.3.2) is rather disappointing. Table 5.5 shows the function values of the fit evaluated on the observed maturities together with percentage errors compared with the sample correlation matrix, as in Table 5.4. However, as can be seen from Figure 5.3 (below), it does capture the general shape of the sample correlation matrix, as compared with Figure 5.2. We also plotted in Figure 5.3 (top) the deformation mapping obtained with the MDS algorithm.

By way of comparison, instead of the general Gaussian correlation function (5.3.2), we fitted the Matérn correlation function (5.3.1). We achieved much better results. The best fit to the sample correlation matrix is obtained for the parameter vector $\eta = (1, 1.397, 6.386)$. Table 5.6 shows the values of our constructed SPD correlation function, together with percentage errors compared with the sample correlation matrix. The maximum percentage error is 7.5 for the maturities of 2 and 9 years. The error is still quite large, but it is encouraging that the fit can be substantially improved by a better choice of correlation function.

In this section we have described a general procedure for implementation of the generic instantaneous SPD correlation functions of the form (5.1.1). It consists of two steps: First, using the Multidimensional scaling find a strictly increasing deformation mapping. Then choose a general SPD correlation that provides the best

fit to the sample correlation matrix. Furthermore, we have implemented the method and tested it on a sample correlation matrix from Japanese Yen. We achieved a rather disappointing quality of fit for a general Gaussian SPD correlation function. However, the fit could be considerably improved by a different choice of the SPD function.

In the following sections we will further investigate the possibility of improving the fit by investigating alternative implementations of the methodology.

5.6 Minimisation of an Objective Function

In the last section, we implemented our generic SPD correlation function,

$$\text{Corr}(t, s) = C(|g(t) - g(s)|),$$

using a two-step procedure, consisting first of finding a deformation function and then an SPD correlation function. This approach did not result in a good quality of fit to the sample correlation matrix. The obvious problem with this construction is that the choice of deformation mapping was independent of the fitting of the correlation function. We would need to perform a search among all possible correlation functions to find the best fit. We believe that an alternative method would perform better if we combined this two steps into one: i.e. simultaneous fit of both the deformation mapping and an SPD function is bound to improve the quality of fit. In the present section we will study this approach.

5.6.1 Description of the Methodology

In this section we propose an alternative approach. We choose a particular parametric SPD correlation function together with some set of functions which is used as a basis for the deformation mapping. We then search among the parameters of the correlation function and the coefficients of the basis functions for the best fit to

the sample correlation matrix. Let $\hat{\rho}_{ij}$, $i, j = 1, 2, \dots, n$ denote sample correlations for n -maturities. Then the criterion to minimise is, for example,

$$\sum_{i < j} \left(\frac{\hat{\rho}_{ij} - \rho_{\alpha}(h_{\beta}(T_i), h_{\beta}(T_j))}{\rho_{\alpha}(h_{\beta}(T_i), h_{\beta}(T_j))} \right)^2, \quad (5.6.1)$$

where α denotes the set of parameters of the SPD correlation function ρ and β the set of parameters of some family of functions. The best fit is found by minimising this criterion over the parameter sets α and β .

Intuitively, we should obtain better results in this case than in the method we used in the last section. However, one problem here is that the deformation mapping has to be strictly increasing. To achieve this we can use a strictly increasing family of functions as a basis for the deformation mapping. Alternatively, we can choose some general set of functions and then introduce some restrictions into the minimisation procedure to keep the function monotonic. Other methods for achieving this are also available and we discuss them later in the section.

5.6.2 Implementation Details and Numerical Results

To test our methodology, we assume that the deformation function can be approximated by a cubic spline. A cubic spline is a piecewise-cubic polynomial joined at so-called knot points¹¹. At each knot point, the polynomials that meet are restricted so that the level and the first two derivatives of each cubic are identical. Each additional knot point in the spline adds one independent parameter, as three of the four parameters of the additional cubic polynomial are constrained by the above restriction. By increasing the number of knots, cubic splines provide increasingly flexible functional forms. A simple, numerically stable parametrisation of a cubic spline is provided by a cubic B-spline basis. The cubic B-spline basis can be represented by

¹¹In Finance, several spline techniques have been used for extracting the term structure of interest rates from a cross-section of coupon bond prices. Among others, McCulloch (1975) uses regression cubic splines, and Vasicek and Fong (1982) use exponential splines to estimate the discount function.

the row vector

$$\phi(\tau) := (\phi_1(\tau), \dots, \phi_\kappa(\tau)).$$

Over any interval between adjacent knot points s_k and s_{k+1} , there are four non-zero B-splines, with adjacent intervals sharing three of them. This gives $\phi(\tau)$ a semi-orthogonal structure from which it gets its numerical stability. Any cubic spline can be constructed from linear combinations of the B-splines, $\phi(\tau)\beta$, where $\beta := (\beta_1, \dots, \beta_\kappa)^T$ is a vector of coefficients. We provide a short description of the B-spline basis in Appendix 5.10.2.

As our test SPD function we chose the Exponential-Power correlation function,

$$C(h) = \exp(-a^2|h|^b).$$

We minimised criterion (5.6.1), using the choice of the SPD function and B-spline basis. The parameters we obtained for the SPD function are $a = 0.0059$ and $b = 1.575$. We have plotted the deformation mapping in Figure 5.4 (above). We can see clearly that the function is strictly increasing. Its shape is similar to the deformation mapping we obtained in the previous section, Figure 5.3 (above). Additionally, we have plotted the fitted deformed Exponential-Power correlation function in Figure 5.4 (below). It clearly has the shape of the sample correlation matrix, Figure 5.2.

Table 5.7 shows the values of our constructed SPD correlation function. We evaluated them at the observed maturities together. Table 5.7 also shows percentage errors of the fit compared with the sample correlation matrix. We can see that the fit is very good, with the largest percentage error 1.34, for the maturities of 2 and 4 years.

Next, we tested our SPD correlation function for the SPD property. We evaluated the deformed correlation function on a semi-annual grid for a the maturities up to 30 years. We then calculated the the eigenvalues of this matrix. We plotted the natural logarithm of the eigenvalues in Figure 5.5. As we can see, they are

all positive, confirming our result. Next, we discuss a possible extension of our implementation methodology which can reduce the number of coefficients in the deformation mapping representation.

5.6.3 Extensions

In this section we discuss one possible extension of the above implementation procedure which may allow smoothing out the non-bijectivity in the spline representation. Moreover, it can effectively reduce the number of parameters.

In our methodology, the number of parameters is determined by the number of knot points. Too few knot points can lead to a poor estimation and too many to over-fitting. To overcome this problem we use smoothing splines which incorporate a “roughness” penalty. An increase in penalty reduces the effective number of parameters, the reason being that splines resulting from a higher roughness penalty are smoother and this forces an implicit relationship between the parameters of the spline. We consider a penalty of the form,

$$\lambda \int_0^T h''(\tau)^2 d\tau,$$

i.e. a constant times the integral of the squared second derivative of the function being splined. We assume that the measure of the degree of smoothness λ is constant, though this assumption can be relaxed. Our problem now consists of minimising the residual sums of squares plus the penalty:

$$\min_{h(\tau) \in \mathcal{H}} \left[\sum_{i,j} \left(\frac{\hat{\rho}_{ij} - \rho(h(T_i) - h(T_j))}{\rho(h(T_i) - h(T_j))} \right) + \lambda \int_0^T h''(\tau)^2 d\tau \right], \quad (5.6.2)$$

where \mathcal{H} is the space of all functions defined on \mathbb{R}_+ with squared second derivative which integrate to a finite value.

The primary factor influencing the choice of the smoothing parameter λ is the apparent smoothness of the mapping itself. In the limit, as $\lambda \rightarrow \infty$, the mapping

becomes affine - a mapping corresponding to a stationary or homogeneous correlation model. By choosing λ sufficiently large we can guarantee that the mapping will become bijective. For small values of λ we may get over-fitting of the sample correlation and a mapping that is not bijective. One choice of the parameter λ can be made by using the cross-validation technique¹².

In terms of the spline, $h_s(\tau, \beta)$, the penalty can be written as follows:

$$\lambda \int_0^T \left(\frac{\partial^2 h_s(\tau, \beta)}{\partial \tau^2} d\tau \right)^2 d\tau = \lambda \beta^T \int_0^T \left(\phi''(\tau)^T \phi''(\tau) d\tau \right) \beta = \lambda \beta^T H \beta,$$

where H is a $\kappa \times \kappa$ band-diagonal matrix. Since any β that makes $h_s(\tau, \beta)$ linear in τ is not penalised, H has two zero eigenvalues. The matrix H is completely determined by the knot points.

For a fixed value of λ the minimisation problem can be stated as follows:

$$\min_{\beta(\lambda)} \left[\sum_{i,j} \left(\frac{\hat{\rho}_{ij} - \rho(h_s(T_i; \beta) - h_s(T_j; \beta))}{\rho(h_s(T_i) - h_s(T_j))} \right) + \lambda \beta^T H \beta \right]. \quad (5.6.3)$$

Criterion (5.6.3) can be minimised using a Newton-type iterative algorithm. Minimisation is done iteratively in two stages. The first stage involves minimisation of the criterion with respect to the parameters underlying the SPD function ρ_α , for fixed co-ordinates. The second stage involves minimisation of the criterion with respect to the co-ordinates, for fixed correlation parameters.

In this section we have developed a simple procedure for the implementation of our generic SPD correlation function. The method is simple to understand and to implement. It is very parsimonious as it involves only small number of parameters. The number of parameters can be controlled by the user, depending on particular choice of SPD function, deformation mapping representation, and the form of the sample correlation matrix. We also have discussed a possible extension of the methodology

¹²See, for example, Wahba (1990) for details.

by introduction of a “roughness penalty”. This can further reduce the number of parameters and smooth out possible non-bijectivities in the deformation mapping.

Our numerical results indicate that this method is superior in terms of quality of fit to the two-stage procedure introduced in the previous section. In the next section we introduce a third implementation method, based on a modification of the downhill simplex algorithm of Nelder and Mead (1965).

5.7 Iterative Two-Step Implementation: MNM-LM Algorithm

In this section we introduce a third method for the implementation of our generic SPD correlation function (5.1.1). Like the first method introduced in Section 5.5, this one is implemented in two steps. However, unlike the first it is possible to iterate these two steps, and so considerably improve its performance.

We denote the sample correlation estimates by $\hat{\rho}(T_i, T_j)$ for each pair of maturities (T_i, T_j) . Then the method consists in minimising the objective function,

$$U(T, \alpha) = \sum_{i < j} [\hat{\rho}(T_i, T_j) - \rho_\alpha(\|T_i - T_j\|)]^2, \quad (5.7.1)$$

with respect to the deformation vector $T = (T_1, T_2, \dots, T_n)$ subject to a strict monotonicity condition, and with respect to the parameter vector α of the SPD function.

To obtain the optimal deformation vector we introduce a modification of the downhill simplex algorithm of Nelder and Mead (1965), which we abbreviate as the MNM algorithm. This modification allows a very efficient search for the strictly increasing deformation vector T . In the second step, we perform a search for the parameter vector α of the SPD function using Levenberg-Marquardt (LM). We then iterate these two steps. We abbreviate our implementation method as the MNM-LM algorithm. Next, we describe the method in detail.

5.7.1 Description of the MNM-LM Algorithm

Our algorithm consists of two steps: the first is minimisation of the objective function (5.7.1) with respect to the deformation vector T , and the second is minimisation of the objective function with respect to the parameter vector α .

To perform the first step we modify the downhill simplex algorithm originally proposed by Nelder and Mead (1965). The Nelder-Mead algorithm is widely used for non-linear unconstrained optimisation. It should not be confused with the simplex algorithm for linear programming: both algorithms employ a sequence of simplices, but are otherwise completely different. The Nelder-Mead algorithm attempts to minimise a scalar-valued non-linear function of n real variables using only function values, without any derivative information. The Nelder-Mead algorithm maintains at each step a non-degenerate *simplex*, a geometric figure in n dimensions of non-zero volume that is the convex hull of $n + 1$ vertices.

Each iteration of a simplex-based direct search method begins with a simplex, specified by its $n + 1$ vertices and the associated function values. One or more test points are computed, along with their function values, and the iteration terminates with a new simplex such that the function values at its vertices satisfy some form of descent condition compared to the previous simplex. In the standard Nelder-Mead algorithm, at the end of the iteration step there can be a *reflection* away from the high point, a *reflection and expansion* away from the high point, a *contraction* along *one* dimension from the high point, or a *contraction* along *all* dimensions toward the low point. Four scalar parameters control the size of each possible movement of the simplex. The NM algorithm is particularly parsimonious in function evaluations per iteration, since in practice it typically requires only one or two function evaluations to construct a new simplex. It should be pointed out that there has been almost no published theoretical analysis treating the NM algorithm. One exception is Lagarias *et al.* (1998), who consider the convergence properties of the NM algorithm in low

dimensions.

To incorporate a constraint optimisation into the NM algorithm we propose the following modification of the algorithm. We start with a simplex that satisfies the strict monotonicity conditions. I.e., we start with a simplex such that all of its vertices have coordinates of the type $T = (T_1, T_2, \dots, T_n)$, with $T_1 < T_2 < \dots < T_n$. Then, at the end of each iteration step, once the movement of a simplex is performed, we perform a check on the new chosen vertex. If this vertex satisfies the monotonicity condition, we accept this movement; otherwise we assign this vertex new coordinates, by simply reordering the coordinates to form an increasing sequence.

The second part of the algorithm is done by minimisation of the objective function (5.7.1) with respect to the parameter vector η by application of the Levenberg-Marquardt (LM) method. This method has become the standard on nonlinear least-square routines in the case where the derivatives with respect to the minimisation are available. The description of the method can be found, for example, in Press *et al.* (1992).

Summary of the minimisation algorithm:

1. Starting step: Define initial simplex. Set $T^0 = (T_1, T_2, \dots, T_n)$, for the first vertex. Then $T^i = T^0 + \lambda e_i$, for the remaining n vertices. Choose starting value for the parameter vector η .
2. Iteration step: Perform modified Nelder-Mead algorithm. Perform Levenberg-Marquardt algorithm on the parameter vector η .
3. Reinitialisation step: Define new initial simplex. Take the solution of the modified Nelder-Mead algorithm as the first vertex T^0 . Define the vertices T^i for $i = 1, \dots, n$ as in starting step, with λ half of the minimum distance of the coordinates in T^0 .

4. Stopping criterion: Terminate the iteration if the objective function changes by less than a specified tolerance for two consecutive iteration steps.

The MNM-LM algorithm provides us with only a discrete set of data for the deformation function, i.e. with $\hat{\tau}_1, \hat{\tau}_2, \dots, \hat{\tau}_n$. For our application we require knowledge of the deformation for all points in the specified interval. There are several approaches one can use to find the deformation function Φ . The choice should be made from case to case depending on the set $\hat{\tau}_1, \hat{\tau}_2, \dots, \hat{\tau}_n$. One could apply a least-square procedure together with some parametric family of functions, interpolation, or smoothing procedure. One should bear in mind that whatever the choice is, the resulting function has to be strictly increasing, thus preserving strict positive definiteness of the correlation function. Regularisation techniques are particularly appealing in this framework, as they allow the choice of the degree of smoothness of Φ . The smoothing parameter can be used to smooth out non-bijection in the function Φ .

In our experience, the proposed MNM-ML algorithm works extremely well. However, it is quite simple to generalise the method to provide it with more flexibility. In general, the objective function (5.7.1) may have several local minima. In such situations, the simulated annealing method, which is probabilistic in nature, has proved to be very useful. Simulated annealing can be used in combination with the MNM-ML algorithm to improve its efficiency. This is, however, a matter for further research. In the next section we present results from the numerical implementation of the MNM-ML algorithm.

5.7.2 Numerical Results

For the first test of the methodology we used the Exponential-Power correlation function. We achieved a maximum percentage error of 1.46 at the maturities of 2 and 4 years. The parameters of the stationary correlation function were $a = 1$

and $b = 1.56$. We have plotted the Exponential-Power correlation function with these parameters in Figure 5.9 (above). The deformation mapping is plotted in Figure 5.6 (above). The values of the deformed Exponential-Power correlation function together with percentage errors for the observed maturities are in Table 5.8.

For the second test we used the Matérn correlation function. We achieved a maximum percentage error of 1.3 at the maturities of 2 and 4 years. The minimisation procedure yielded the following parameters for the correlation function: $\sigma = 1$, $\nu = 1.3$, and $\rho = 3.57$. The deformation mapping is plotted in Figure 5.6 (below). Note the slope of this map is much steeper at the short end of the curve. We plotted the Matérn correlation function with these parameters in Figure 5.9 (below). The fit is slightly better than in the case of the Exponential-Power correlation function. We expected this result as the Matérn correlation function is more flexible. The values of the deformed Matérn correlation function together with the percentage errors for the observed maturities are in Table 5.9.

Additionally, we evaluated the deformed correlation functions on a grid consisting of 250×250 equidistant points. The log of the absolute values of the eigenvalues of these matrices are plotted in Figure 5.7. It is interesting to compare the behaviour of the eigenvalues in both cases. In the case of the Exponential-Power correlation function, the eigenvalues collapse very fast toward zero. However, in the case of the Matérn correlation function, the descent is slower. This behaviour may be explained by the behaviour on the diagonal of the deformed correlation functions in Figure 5.8. It is quite flat in the first case and shows some decorrelation in the second. This decorrelation is caused by the steep slope of the corresponding deformation function as mentioned above. Thus, we observe that fast decorrelation causes slow decay in the eigenvalues. This will become more apparent in Chapter 6.

In this section we have introduced a third method for the implementation of our generic SPD correlation function (5.1.1). The method is iterative in nature, com-

binning two minimisation steps. The numerical implementation and fitting results indicate that the method is very simple to use and provides a very good quality of fit. In the next section we discuss alternative estimation techniques such as maximum likelihood and Bayesian methods, which can be used to fit our generic SPD correlation function (5.2.2).

5.8 Alternative Estimation Methods

All three implementation methods we presented above could be classified in terms of statistical methodology as method of moments. They involved fitting the SPD correlation function directly to the sample correlation matrix. These methods could be expected to be inferior in terms of statistical efficiency compared with the maximum likelihood and Bayesian approaches. In this section we discuss a possible implementation method for the maximum likelihood approach and briefly discuss the advantages of the Bayesian framework.

First we discuss the maximum likelihood approach¹³. We start with N realisations Z_1, \dots, Z_N of the random field for n observed maturities. Thus each observed realisation k is a vector $Z_k = (Z_k(T_1), \dots, Z_k(T_n))$, where T_1, \dots, T_n are observed maturities, for $k = 1, 2, \dots, N$. We assume that these realisations are independent with

$$Z_k \sim N_n = (\mu, \Sigma),$$

where N_n denotes the n -dimensional normal distribution, μ the n -vector of means, and Σ an $n \times n$ covariance matrix.

Furthermore, we assume $\sigma_{ij} = \text{Cov}\{Z_k(T_i), Z_k(T_j)\}$ is of the form,

$$\sigma_{ij} = C(h(T_i) - h(T_j)),$$

¹³In geostatistics, the possibility of fitting models of spatial type with a maximum likelihood approach has been considered by Mardia and Goodall (1993), and Smith (1996).

where C is an SPD function and h is represented by a linear combination of some function sequence.

Next we assume that the data is normalised, or that the variances have already been estimated by some methods; thus we concentrate on the estimation of the correlation functions. Under these assumptions the negative log-likelihood based on Z_1, \dots, Z_N becomes

$$L = \frac{M}{2} \log |\Sigma| + \frac{M-1}{2} \text{tr} \left(\Sigma^{-1} \hat{\Sigma} \right),$$

where $\hat{\Sigma}$ is the usual $N \times N$ sample correlation matrix.

In our case, as in Section 5.6, we might consider cubic splines to represent the function h . h can be defined in terms of constants $a_1, a_2, \delta_1, \delta_2, \dots, \delta_n$, and $t_1 < t_2 < \dots < t_n$ by the expression

$$h(t) = a_1 + a_2 t + \frac{1}{12} \sum_1^n \delta_i \eta_i(t), \quad (5.8.1)$$

subject to the constraints,

$$\sum_{i=1}^n \delta_i = \sum_{i=1}^n \delta_i t_i = 0, \text{ and } \eta_i(t) = |t - t_i|^3. \quad (5.8.2)$$

Thus (5.8.1) represents h as a sum of linear terms and n *radial basis functions* η_i with *centres* at the observed maturities T_i . It can be shown that h is a cubic spline, with knots at the points T_1, T_2, \dots, T_n ; furthermore the constraints (5.8.2) imply that h'' and h''' are both zero outside $[t_1, \dots, t_n]$, so the curve h is a *natural* cubic spline¹⁴.

As we discussed in Section 5.6, one way to reduce implicitly the number of parameters in the regression splines is to use a penalty functional (see equation (5.6.2)). Alternatively one can simply restrict the representation (5.8.1) to a subset of radial basis functions. Thus we assume

$$\delta_i = 0 \text{ for } i \notin \{i_1, \dots, i_m\} \quad (5.8.3)$$

¹⁴For further details on this characterisation see, for example, Green and Silverman (1994).

where i_1, \dots, i_m are some subset of indexes to be determined. This formulation reduces the number of parameters in the log-likelihood minimisation problem and can make the computational burden manageable. The set of indexes in (5.8.3) can be found by some criteria, for example by a cross-validation study.

As an alternative to the maximum likelihood approach one can consider the Bayesian¹⁵ point of view, as it also involves calculation of the likelihood function. This method involves sampling from the posterior distribution. When applied with efficient Markov Chain Monte Carlo methods this may prove to be not much more difficult computationally than maximisation of the likelihood function with respect to a high number of parameters. Bayesian methods have the advantage that when used in prediction, they correctly allow for the uncertainty in the parameters being estimated.

In this section we have briefly described the implementation of alternative estimation approaches that can be used to estimate our generic SPD correlation function (5.2.2). Whether these methods achieve a better statistical fit of our Gaussian field to the data can only be answered by empirical studies. In the next section we present conclusions of this chapter and discuss some ideas for further research.

5.9 Conclusions and Discussions

In this chapter we have continued the study of SPD instantaneous correlation functions for the construction of infinite-factor random field models. We suggested a generic SPD correlation function of the form

$$\text{Corr}(s, t) = C(g(s) - g(t)),$$

¹⁵In geostatistics, recent papers by Damien *et al.* (2001) and Schmidt and O'Hagan (2000) have adapted Bayesian modelling within the deformation approach. The difficulties in their approach lie in specification of a suitable parametric model with associated prior distribution. These choices are often arbitrary.

where C is a SPD functions and g strictly increasing maturity deformation mapping. We characterised a large class of SPD function and derived sufficient conditions on the deformation mapping so that our generic correlation function would have the SPD property.

Furthermore, we provided three implementation methods. The first method is based on a two-step implementation. In the first step we find the deformation mapping using multidimensional scaling algorithm under the assumption of a very general SPD function. Then in the second step we fit the SPD function under a fixed deformation mapping.

The second method is based on choosing some parametric SPD function and some function basis to represent the deformation mapping. The SPD correlation function is then found by minimising some objective function over the parameters of the SPD function and the coefficients of the function basis.

The third implementation method is an iterative two-step procedure. Both steps minimise in turn an objective function, the first step finding an optimal deformation vector, and the second by minimising the parameter vector of the SPD function. We are able to loop over these two steps and so considerably improve the quality of the fit.

We have implemented and tested all three methods on the Japanese Yen interest rate data. We found that the first method is rather disappointing. It produced instantaneous SPD correlation function that could only capture the overall shape of the sample correlation matrix. The actual fit however was not very good. The second and third methods produced much better results. Not only did they capture the shape of the sample correlation matrix, but they had an excellent fit, with maximum percentage error less than 1.5.

The main advantage of the methodology developed in this chapter is that it provides

a parsimonious functional form for the SPD correlation functions. The number of parameters is relatively small and can be further reduced by various means discussed in the chapter. Thus, we can relatively easily construct random field models that fit empirical sample correlation matrices and have infinite-factor structure. Thus, the class of field models we developed in this chapter shows that random field models are able to capture the properties of the instantaneous correlation matrix observed in the market. Furthermore, the implementation methodology we develop should make them more accessible for practical use in the pricing and risk-management of interest rate products.

There are two remaining problems for further research. The first is concerned with alternative estimation methods. We briefly discussed the maximum likelihood and Bayesian estimation approaches in this chapter. It remains to be found whether they provide a better statistical fit to the data. The second problem is to find out how flexible the class of field models we suggested in this chapter is. Obviously, the functional form of the generic correlation function is quite restrictive despite the maturity deformation mapping. However, only empirical testing on a wider set of interest rate data coming from different currencies can validate this methodology.

In Chapter 6 we will develop alternative methodology for the construction of a functional forms of the SPD covariance function. This construction consists of superposition of an SPD function with a PD function. The PD function can be thought of as capturing the main driving factors in the yield changes and the SPD function as providing general correlated noise affecting all maturities. The SPD correlation function resulting from this method fits exactly the sample correlation matrix. Furthermore, the random field resulting from this methodology has a very intuitive financial interpretation.

5.10 Appendices

5.10.1 Simar's Algorithm

In this appendix we explain how we fitted a general Gaussian correlation function,

$$\rho(h) = \int_{-\infty}^{\infty} \exp(-h^2 t^2) dF(t), \quad (5.10.1)$$

to the sample correlation matrix.

In our application we want to minimise the functional $\phi(F)$ given by

$$\phi(F) = \sum_{i < j} (\hat{\rho}_{ij} - \rho(|h_{ij}|; F))^2. \quad (5.10.2)$$

That is, we want to compute a least-square fit of the sample correlations ρ_{ij} to the correlation function evaluated at D -space distance h_{ij} . We apply an algorithm suggested by Simar (1976). The result of this minimisation will be a discrete measure, supported on points t_1, \dots, t_k , with corresponding masses $\alpha_1, \dots, \alpha_k$, where $k \leq n/2$, and $n = N(N - 1)/2 - N$, the number of maturities pairs minus the number of maturities.

We start with the first step of the algorithm, i.e. with a distribution F_{t_1} that puts mass 1 at t_1 and has $\phi(F) < \infty$. In the second step of the algorithm we select the next point in the support of F by minimising the directional derivative of the functional (5.10.2) in the direction of the point mass at t . This is defined as

$$\lim_{\epsilon} \frac{1}{\epsilon} (\phi((1 - \epsilon)F + \epsilon F_t) - \phi(F)),$$

and simplifies in our case to

$$-2 \sum_{i < j} (\rho(|h_{ij}|; F_t) - \rho(|h_{ij}|; F)) (\hat{\rho}_{ij} - \rho(|h_{ij}|; F)).$$

This derivative is minimised at the solution to

$$\sum_{i < j} |h_{ij}|^2 \exp(-|h_{ij}|^2 t^2) (\hat{\rho}_{ij} - \rho(|h_{ij}|; F)) = 0.$$

Having added this value of t to the support of F , in the third step we minimise $\phi(F)$ as a function of the masses α_i , constrained to be non-negative and adding up to 1, so that the diagonal of the correlation function is 1. We iterate the whole procedure, choosing first a new t , and then a new set of α_i , until either we have $n/2$ points in the support, or the directional derivative for adding a new point to the support is less than a specified amount.

5.10.2 B-Spline Basis

In this Appendix we provide a short description of a B-spline basis. Let $\{s_k\}_1^K$ denote the knot points, with $s_k < s_{k+1}$, $s_1 = 0$, and $s_K = M$, the maximum maturity of any bond in the sample. The knot points define $K - 1$ intervals over the domain of the spline, $[0, T]$. For the purpose of defining a B-spline basis, it is convenient to define an augmented set of knot points, $\{d_k\}_{k=1}^{K+6}$, where $d_1 = d_2 = d_3 = s_1$, $d_{K+4} = d_{K+5} = d_{K+6} = s_K$, and $d_{k+3} = s_k$ for $1 < k \leq K$.

A cubic B-spline basis is a vector of $\kappa = K + 2$ cubic B-splines defined over the domain. A B-spline is defined by the following recursion, where $r = 4$ for a cubic B-spline and $1 < k \leq \kappa$:

$$\phi_k^r(\tau) = \frac{\phi_k^{r-1}(\tau)(\tau - d_k)}{d_{k+r-1} - d_k} + \frac{\phi_k^{r-1}(\tau)(d_{k+r} - \tau)}{d_{k+r} - d_{k+1}}$$

for τ in $[0, T]$, with

$$\phi_k^1(\tau) = \begin{cases} 1, & \text{if } d_k \leq \tau < d_{k+1}, \\ 0, & \text{otherwise.} \end{cases}$$

5.10.3 SPD Functions and Kernels

We start this appendix with a definition of positive definite functions and kernels in a way they are usually presented in integral equation and approximation theories. We give a short historical discussion of their earliest treatments.

If S is a subset of \mathbb{R}^n , then a function $P : S \times S \rightarrow \mathbb{R}$ is called a *positive definite kernel* or *PD kernel* if the matrix $M = (P(x_i, x_j))_{i,j=1,\dots,n}$ is *positive semi-definite*; i.e.,

$$\sum_{i,j=1}^n c_i c_j P(x_i, x_j) \geq 0 \quad (5.10.3)$$

for all $c = (c_1, \dots, c_n) \in \mathbb{R}^n$ and for all choices of finite subsets $X = \{x_1, \dots, x_n\} \subseteq S$. A function P is called a *strictly positive kernel* or *SPD kernel* if the strict inequality holds in (5.10.3) for all $c \in \mathbb{R}^n$ with $c \neq 0$ and all choices of X . When a PD/SPD kernel depends on s and t only via the difference $t-s$, i.e. $K(s, t) = f(t-s)$, then it is referred to as a PD/SPD function.

General interest in positive definite kernels and functions goes back at least one hundred years. PD kernels, as defined by (5.10.3), arose for the first time in a paper by Mercer(1909) on integral equations. Mercer's work itself followed from a paper by Hilbert (1904) on Fredholm integral equations of the second kind:

$$f(s) = \phi(s) - \lambda \int_a^b K(s, t) \phi(t) dt,$$

where K is a continuous real symmetric ($K(s, t) = K(t, s)$) kernel. Mercer(1909) defined a continuous real symmetric kernel $K(s, t)$ to be of *positive type* if

$$J(x) = \int_a^b \int_a^b K(s, t) x(s) x(t) ds dt \geq 0$$

for all real continuous functions x on $[a, b]$, and he proved that (5.10.3) is a necessary and sufficient condition for a kernel¹⁶ to be of positive type. At about the same time, Young (1909), motivated by a different question in the theory of integral equations, showed that for continuous kernels, (5.10.3) is equivalent to $J(x) \geq 0$ for all $x \in L^1[a, b]$. Note however, that to date there has been no equivalent characterisation result for SPD kernels.

5.10.4 Figures

¹⁶In Finance, integral kernels have been used by Heaney and Cheng (1984) in the framework of mean-variance portfolio selection with an infinite number of securities.

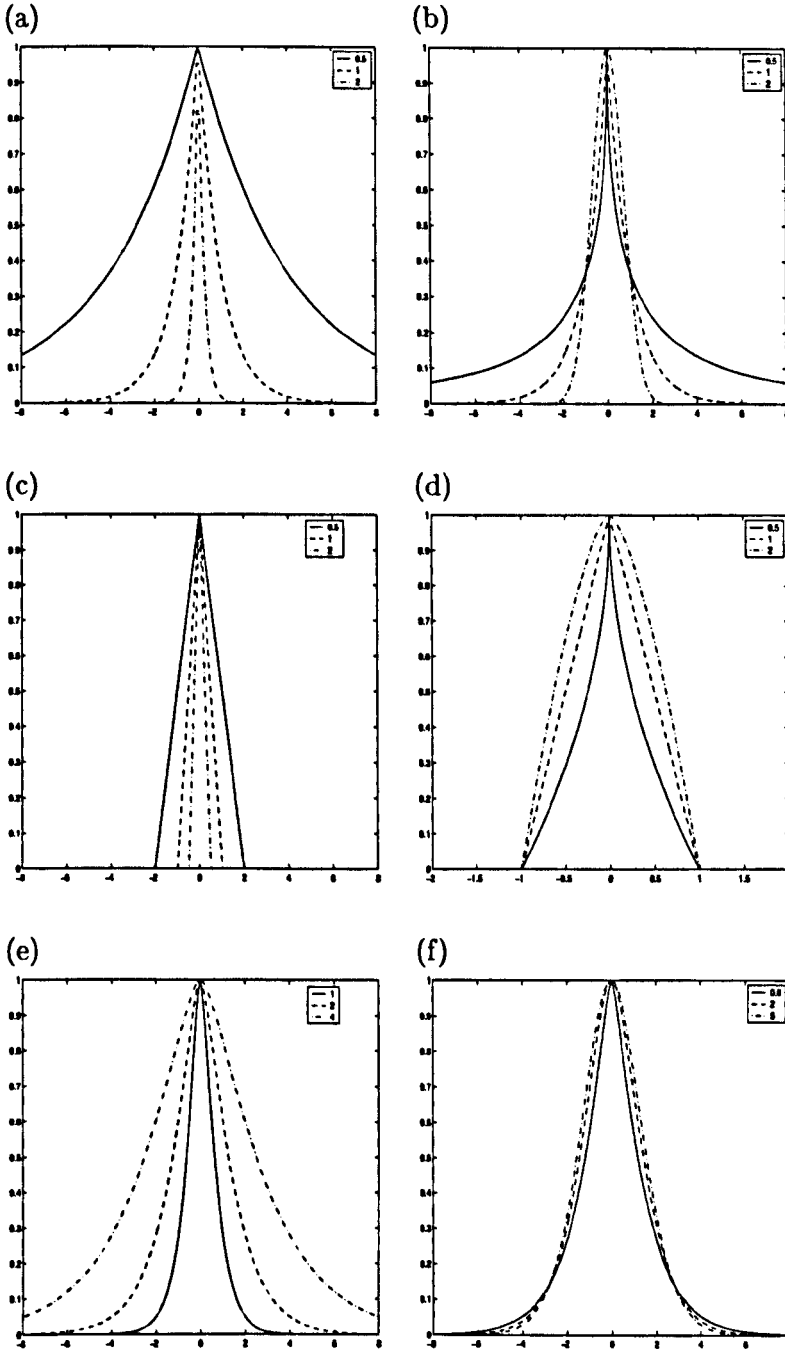


Figure 5.1: Examples of stationary correlation functions. (a) Exponential, with $a = 0.5, 1, 2$. (b) Exponential-Power, with $a = 1, b = 0.5, 1, 2$. (c) Triangular, with $a = 0.5, 1, 2$. (d) Power law, with $b = 1$ and $a = 0.5, 1, 1.5$. (e) Matérn, with $\sigma = 1$, $\nu = 1$, and $\rho = 1, 2, 4$. (f) Matérn with $\sigma = 1$, $\rho = 2$, and $\nu = 0.8, 2, 5$.

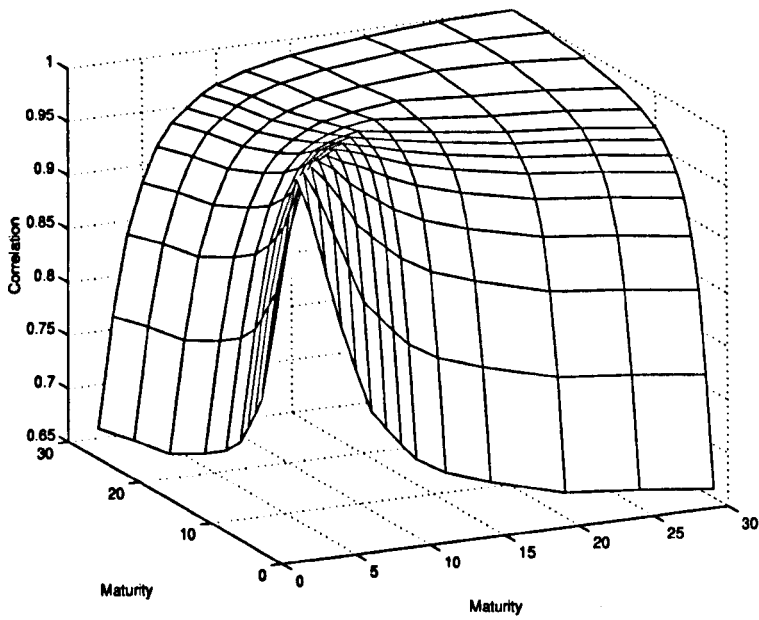
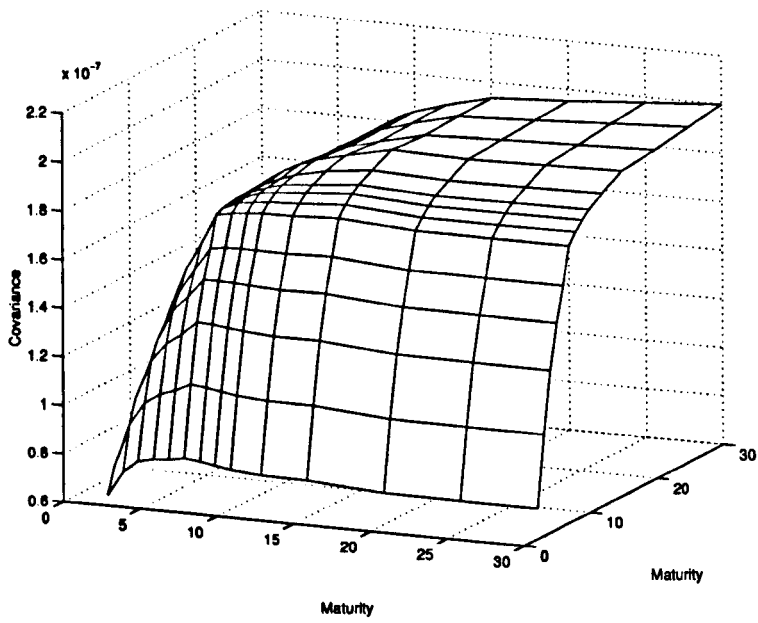


Figure 5.2: Sample covariance matrix (above) and sample correlation matrix (below). Currency: Yen.

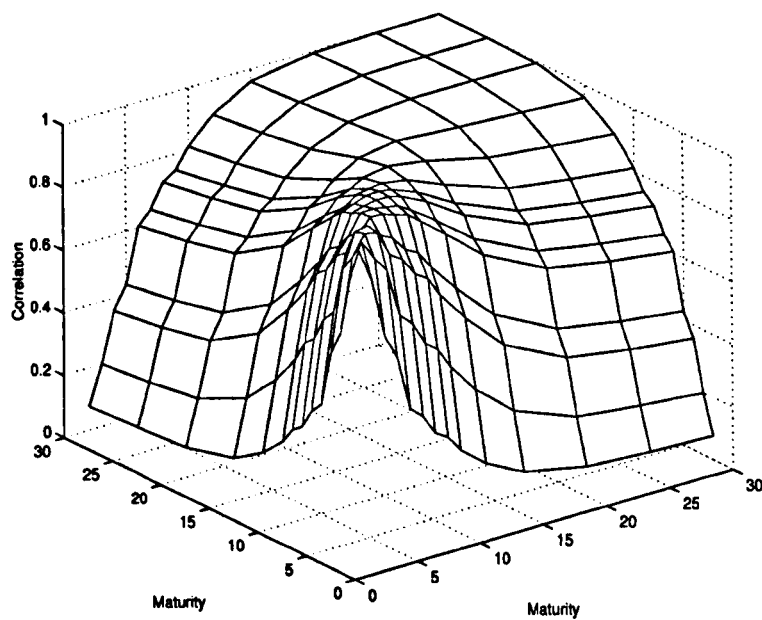
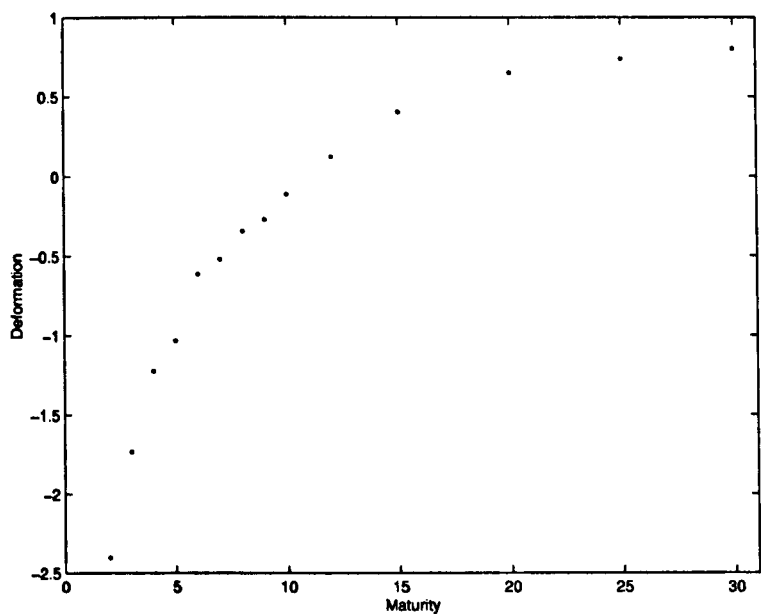


Figure 5.3: Deformation coordinates produced by MDS algorithm (Top). Sampson-Guttrop approach with Gaussian correlation function (Bottom). Currency: Yen.

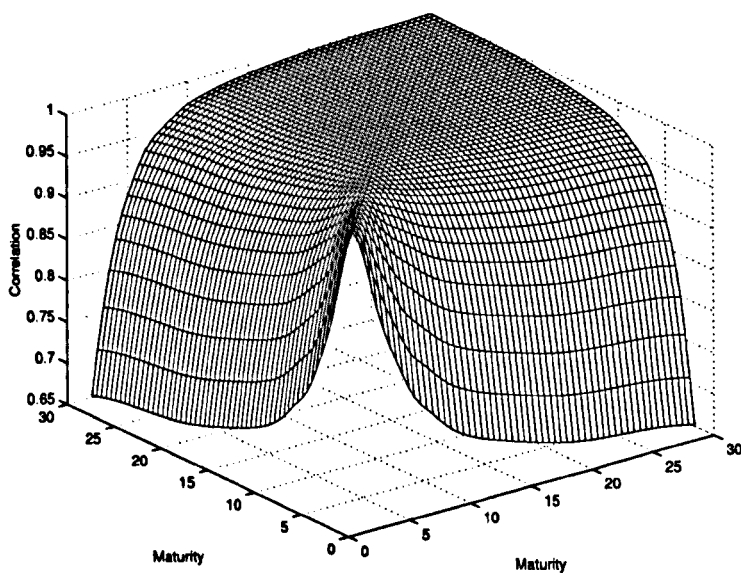
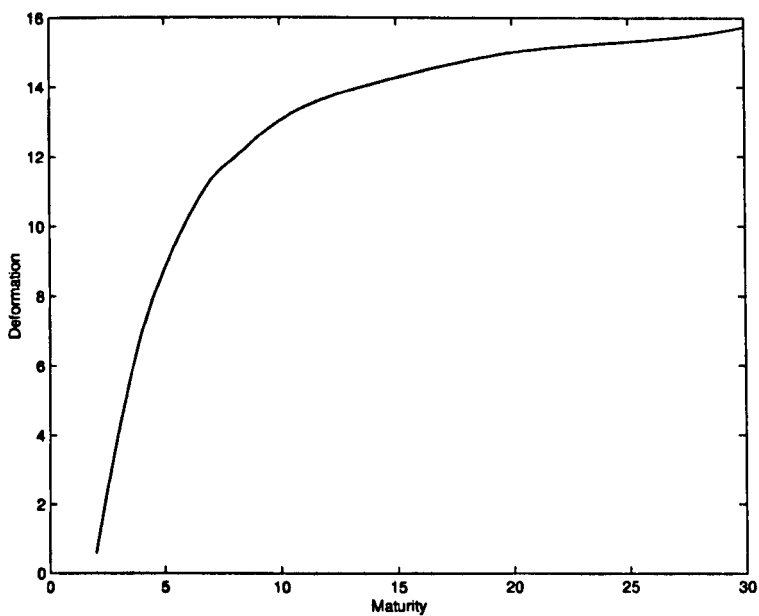


Figure 5.4: Deformation function represented by B-splines, for the exponential-power correlation function with coefficients $a = 0.0059$ and $b = 1.575$ (above). Fitted deformed exponential-power correlation functions with the same parameters (below).

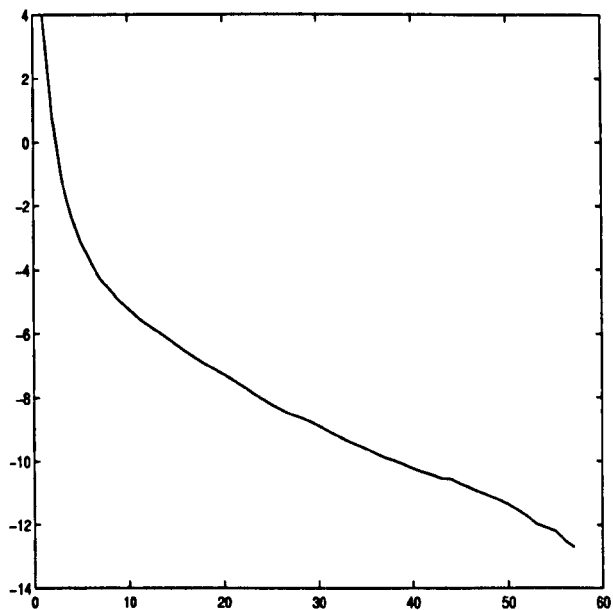


Figure 5.5: Eigenvalues of the matrix resulting from the deformed Exponential-Power correlation function with parameters $a = 0.0059$ and $b = 1.575$. Deformation is represented by B-splines.

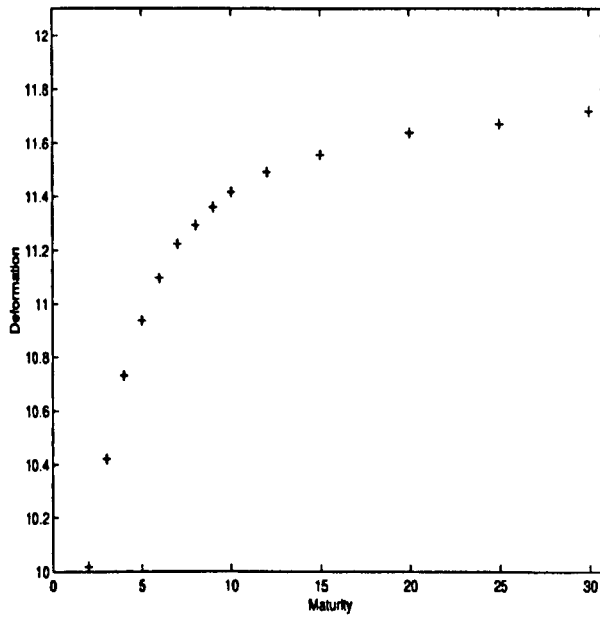
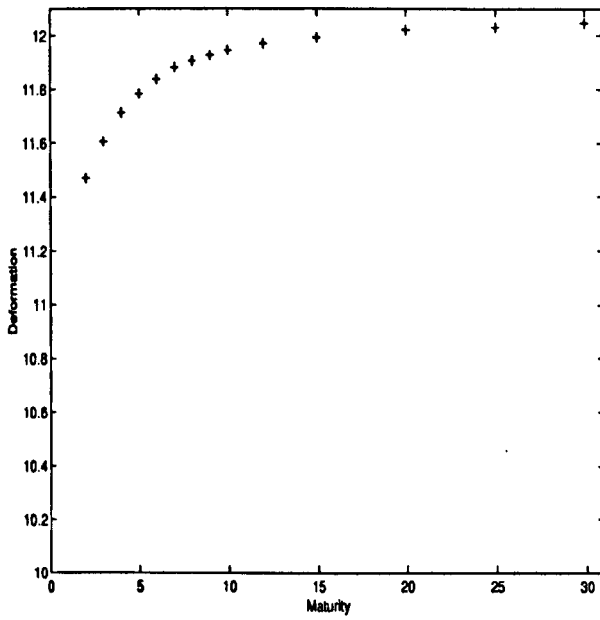


Figure 5.6: MNM-LM method. Deformation points for the exponential-power correlation structure (above), with the coefficients $a = 1$ and $b = 1.56$. Deformation points for the Matern correlation structure (below), with the coefficients $\sigma = 1$, $\nu = 1.3$, and $\rho = 3.57$.

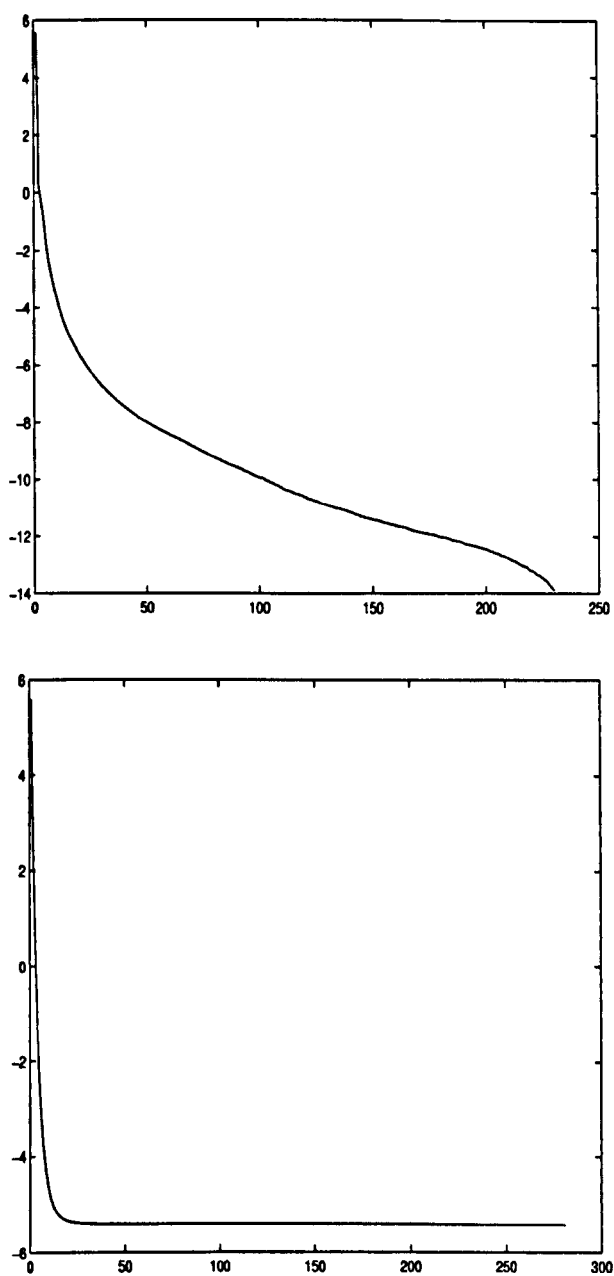


Figure 5.7: Eigenvalues of the matrices resulting from exponential-power (above), and Matérn correlation functions.

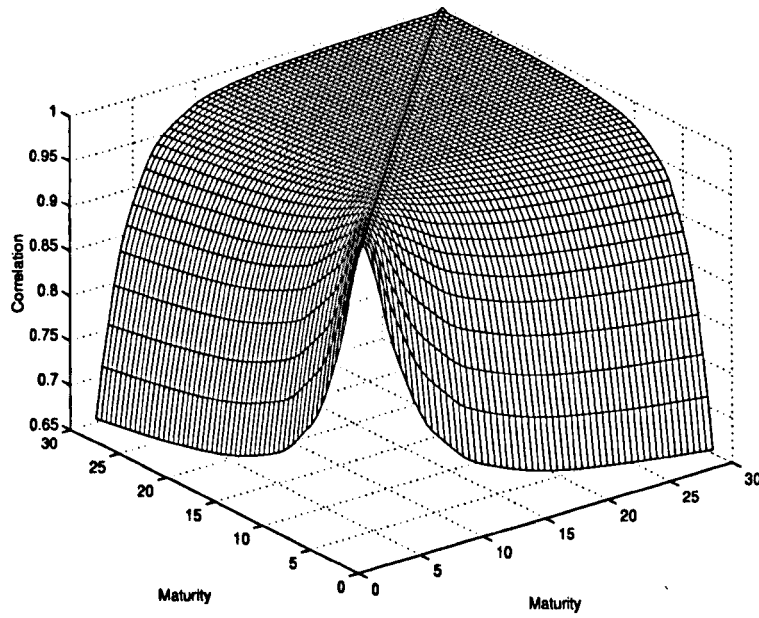
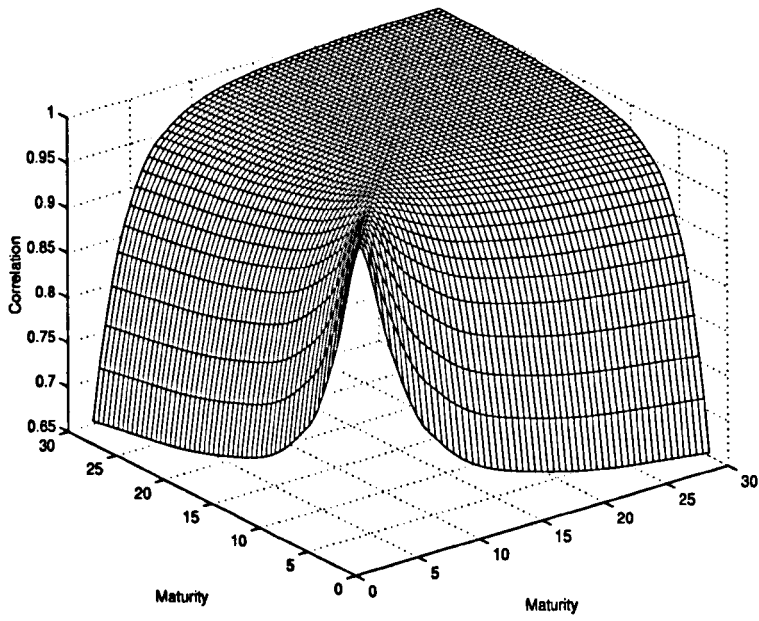


Figure 5.8: Deformed exponential-power correlation function (above). Deformed Matérn correlation function (below).

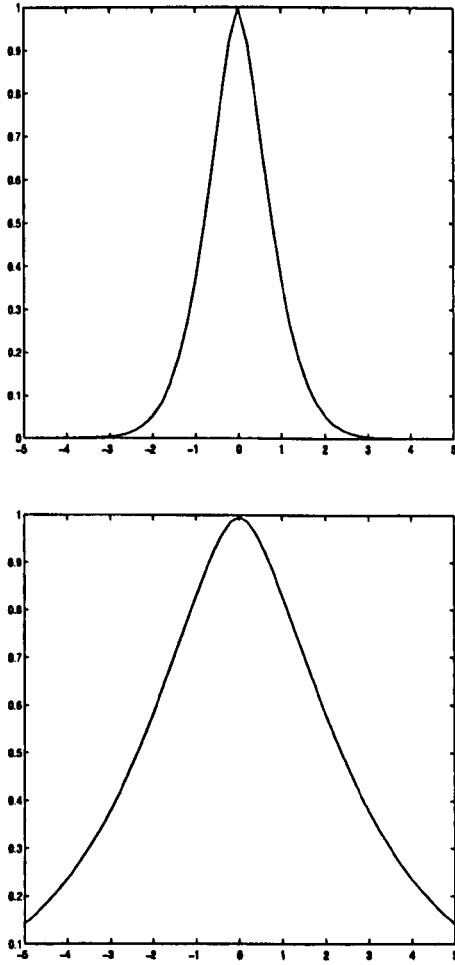


Figure 5.9: Plot of the exponential-power correlation function with the parameters $a = 1$ and $b = 1.56$ (above). Plot of the Matérn correlation function with the parameters $\sigma = 1$, $\nu = 1.3$, and $\rho = 3.57$ (below).

5.10.5 Tables

Table 5.2: Descriptive statistics for the yield differences. Currency: Yen.

Yield	Mean	St.Dev.	Min	Median	Max	Skewness	Kurtosis
2	-6.00758E-06	0.000247	-0.001511	0	0.001757	0.478	8.531
3	-8.09891E-06	0.000299	-0.001599	-9.90099E-06	0.002782	0.918	11.574
4	-9.91868E-06	0.000339	-0.001897	-1.97398E-05	0.003172	0.979	11.064
5	-1.13224E-05	0.000367	-0.002189	-1.98386E-05	0.003255	0.845	9.407
6	-1.22605E-05	0.000388	-0.002303	-2.95779E-05	0.003545	0.906	9.394
7	-1.26426E-05	0.000416	-0.002515	-2.95162E-05	0.003805	0.871	9.214
8	-1.25512E-05	0.000420	-0.002511	-2.93606E-05	0.003987	0.914	9.783
9	-1.22469E-05	0.000424	-0.002557	-2.92741E-05	0.003991	0.873	9.603
10	-1.19379E-05	0.000427	-0.002691	-1.97355E-05	0.003967	0.786	9.294
12	-1.13376E-05	0.000432	-0.002783	-1.96865E-05	0.004056	0.739	9.504
15	-1.10098E-05	0.000439	-0.002931	-1.95401E-05	0.004162	0.688	9.833
20	-1.0376E-05	0.000440	-0.003023	-2.43039E-05	0.004153	0.625	9.926
25	-9.12917E-06	0.000443	-0.003069	-2.19308E-05	0.004150	0.585	9.743
30	-7.88415E-06	0.000447	-0.003116	-2.89924E-05	0.004147	0.549	9.448

Table 5.3: Sample covariance matrix, multiplied by a factor $1e + 7$. Currency: Yen.

	2	3	4	5	6	7	8	9	10	12	15	20	25	30
2	0.611													
3	0.712	0.894												
4	0.762	0.989	1.15											
5	0.776	1.03	1.23	1.35										
6	0.781	1.05	1.27	1.41	1.51									
7	0.793	1.08	1.33	1.49	1.60	1.73								
8	0.785	1.08	1.32	1.49	1.61	1.74	1.77							
9	0.774	1.07	1.32	1.49	1.61	1.75	1.78	1.80						
10	0.763	1.06	1.31	1.48	1.61	1.75	1.78	1.81	1.83					
12	0.754	1.05	1.30	1.48	1.61	1.75	1.79	1.82	1.84	1.87				
15	0.752	1.05	1.30	1.48	1.62	1.76	1.80	1.83	1.86	1.89	1.93			
20	0.731	1.02	1.27	1.46	1.59	1.74	1.78	1.82	1.84	1.88	1.92	1.93		
25	0.730	1.02	1.27	1.45	1.59	1.74	1.78	1.82	1.85	1.88	1.93	1.94	1.96	
30	0.728	1.02	1.27	1.45	1.59	1.74	1.78	1.82	1.85	1.89	1.94	1.95	1.98	2

Table 5.4: Sample correlation matrix. Currency: Yen.

	2	3	4	5	6	7	8	9	10	12	15	20	25
3	0.964												
4	0.909	0.975											
5	0.856	0.940	0.985										
6	0.813	0.907	0.965	0.991									
7	0.773	0.872	0.942	0.976	0.993								
8	0.755	0.857	0.928	0.966	0.987	0.996							
9	0.739	0.841	0.915	0.955	0.980	0.992	0.996						
10	0.723	0.827	0.904	0.946	0.972	0.987	0.993	0.997					
12	0.707	0.811	0.888	0.932	0.961	0.977	0.985	0.991	0.996				
15	0.692	0.797	0.873	0.918	0.948	0.966	0.976	0.984	0.990	0.995			
20	0.673	0.778	0.855	0.902	0.933	0.953	0.964	0.974	0.981	0.988	0.995		
25	0.667	0.771	0.847	0.894	0.926	0.946	0.958	0.968	0.975	0.984	0.992	0.998	
30	0.659	0.762	0.837	0.884	0.916	0.936	0.948	0.959	0.967	0.976	0.985	0.993	0.998

Table 5.5: Values of the correlation function for selected maturities together with percentage errors resulting from Sampson-Guttorp approach with Gaussian correlation function. Currency: Yen.

	2	3	4	5	6	7	8	9	10	12	15	20	25
3	0.893												
	7.37												
4	0.705	0.937											
	22.41	3.87											
5	0.624	0.885	0.991										
	27.06	5.89	0.57										
6	0.448	0.731	0.911	0.957									
	44.97	19.39	5.63	3.43									
7	0.410	0.692	0.883	0.936	0.998								
	46.89	20.71	6.23	4.10	0.51								
8	0.345	0.617	0.823	0.888	0.982	0.992							
	54.35	28.02	11.31	8.07	0.52	0.39							
9	0.320	0.586	0.797	0.865	0.971	0.985	0.999						
	56.69	30.34	12.96	9.48	0.87	0.70	0.24						
10	0.268	0.518	0.733	0.808	0.939	0.959	0.987	0.994					
	62.94	37.37	18.83	14.56	3.45	2.80	0.64	0.35					
12	0.202	0.423	0.635	0.716	0.874	0.902	0.947	0.962	0.986				
	71.39	47.84	28.42	23.13	9.07	7.61	3.86	2.97	0.91				
15	0.139	0.319	0.515	0.596	0.772	0.808	0.870	0.892	0.936	0.980			
	79.98	59.98	41.04	35.06	18.63	16.36	10.94	9.36	5.46	1.50			
20	0.097	0.241	0.415	0.492	0.670	0.710	0.781	0.808	0.865	0.932	0.985		
	85.65	68.99	51.46	45.46	28.15	25.47	19.03	16.99	11.81	5.64	1.01		
25	0.084	0.217	0.382	0.456	0.634	0.674	0.747	0.775	0.835	0.910	0.973	0.998	
	87.34	71.83	54.94	48.96	31.58	28.77	22.05	19.91	14.36	7.54	1.91	0.01	
30	0.076	0.200	0.358	0.431	0.606	0.646	0.720	0.750	0.812	0.891	0.961	0.994	0.999
	88.43	73.71	57.25	51.28	33.84	30.96	24.04	21.84	16.03	8.76	2.40	0.11	0.06

Table 5.6: Values of the correlation function for selected maturities together with percentage errors resulting from Sampson-Guttorp approach with Matérn correlation function. Currency: Yen.

	2	3	4	5	6	7	8	9	10	12	15	20	25
3	0.969												
	0.57												
4	0.919	0.982											
	1.07	0.67											
5	0.896	0.967	0.997										
	4.69	2.88	1.20										
6	0.841	0.926	0.974	0.987									
	3.44	2.10	0.96	0.37									
7	0.828	0.915	0.967	0.981	0.999								
	7.22	4.88	2.67	0.54	0.66								
8	0.804	0.894	0.951	0.968	0.994	0.998							
	6.43	4.31	2.40	0.25	0.75	0.13							
9	0.794	0.885	0.943	0.962	0.991	0.995	1.000						
	7.51	5.17	3.08	0.67	1.17	0.35	0.33						
10	0.772	0.864	0.926	0.947	0.982	0.988	0.996	0.998					
	6.75	4.49	2.53	0.05	0.99	0.09	0.28	0.08					
12	0.738	0.833	0.899	0.922	0.964	0.972	0.984	0.989	0.996				
	4.53	2.74	1.30	1.07	0.37	0.48	0.10	0.29	0.02				
15	0.699	0.794	0.863	0.888	0.937	0.947	0.963	0.969	0.981	0.994			
	0.88	0.39	1.17	3.31	1.23	2.01	1.35	1.52	0.86	0.13			
20	0.664	0.759	0.830	0.856	0.909	0.920	0.939	0.947	0.962	0.980	0.995		
	1.30	2.46	2.88	5.08	2.59	3.44	2.61	2.78	1.91	0.80	0.03		
25	0.652	0.747	0.818	0.844	0.899	0.910	0.930	0.938	0.954	0.974	0.992	0.999	
	2.21	3.20	3.45	5.59	2.97	3.81	2.92	3.13	2.19	1.00	0.01	0.11	
30	0.643	0.738	0.809	0.835	0.891	0.902	0.923	0.931	0.948	0.969	0.988	0.998	1.000
	2.35	3.25	3.37	5.47	2.77	3.63	2.70	2.95	1.98	0.76	0.35	0.50	0.13

Table 5.7: Currency: Yen. Fitted Exponential-Power correlation function using B-splines, together with percentage errors.

	2	3	4	5	6	7	8	9	10	12	15	20	25
3	0.957												
	0.71												
4	0.897	0.971											
	1.34	0.45											
5	0.849	0.936	0.984										
	0.74	0.48	0.09										
6	0.811	0.904	0.963	0.990									
	0.26	0.28	0.26	0.08									
7	0.779	0.876	0.941	0.975	0.993								
	0.86	0.45	0.08	0.13	0.02								
8	0.761	0.860	0.928	0.964	0.986	0.997							
	0.79	0.39	0.08	0.12	0.13	0.10							
9	0.745	0.845	0.915	0.954	0.978	0.992	0.997						
	0.82	0.41	0.09	0.17	0.21	0.04	0.12						
10	0.730	0.831	0.903	0.944	0.970	0.986	0.993	0.998					
	1.00	0.49	0.10	0.24	0.27	0.06	0.02	0.09					
12	0.711	0.813	0.887	0.930	0.958	0.977	0.986	0.993	0.997				
	0.66	0.33	0.07	0.14	0.22	0.07	0.06	0.12	0.15				
15	0.694	0.797	0.872	0.917	0.947	0.968	0.978	0.986	0.992	0.997			
	0.21	0.00	0.15	0.10	0.13	0.22	0.16	0.18	0.21	0.23			
20	0.673	0.776	0.853	0.900	0.932	0.955	0.966	0.976	0.983	0.991	0.996		
	0.01	0.24	0.17	0.16	0.10	0.24	0.18	0.20	0.26	0.28	0.15		
25	0.664	0.768	0.845	0.893	0.926	0.949	0.961	0.971	0.979	0.987	0.994	0.999	
	0.42	0.49	0.26	0.14	0.04	0.35	0.31	0.31	0.37	0.37	0.23	0.08	
30	0.652	0.756	0.834	0.883	0.916	0.941	0.953	0.964	0.972	0.982	0.989	0.996	0.998
	1.05	0.89	0.43	0.15	0.03	0.50	0.50	0.49	0.59	0.58	0.46	0.33	0.02

Table 5.8: MNM-LM method. Currency: Yen. Fitted Exponential-Power correlation function, together with percentage errors.

	2	3	4	5	6	7	8	9	10	12	15	20	25
3	0.956												
	0.79												
4	0.896	0.970											
	1.46	0.51											
5	0.848	0.935	0.984										
	0.86	0.56	0.12										
6	0.810	0.903	0.962	0.990									
	0.35	0.35	0.31	0.10									
7	0.779	0.876	0.940	0.974	0.993								
	0.81	0.39	0.12	0.16	0.00								
8	0.761	0.860	0.927	0.964	0.986	0.997							
	0.77	0.35	0.12	0.15	0.15	0.09							
9	0.745	0.845	0.914	0.953	0.977	0.992	0.997						
	0.83	0.39	0.12	0.20	0.23	0.02	0.11						
10	0.730	0.831	0.902	0.944	0.969	0.986	0.993	0.998					
	1.04	0.48	0.12	0.27	0.30	0.08	0.01	0.08					
12	0.712	0.813	0.887	0.930	0.958	0.977	0.986	0.992	0.997				
	0.72	0.34	0.08	0.16	0.25	0.05	0.04	0.10	0.14				
15	0.695	0.797	0.872	0.917	0.947	0.968	0.978	0.986	0.992	0.997			
	0.30	0.02	0.15	0.12	0.16	0.18	0.13	0.16	0.20	0.22			
20	0.674	0.777	0.853	0.900	0.932	0.955	0.966	0.975	0.983	0.991	0.996		
	0.11	0.19	0.17	0.18	0.14	0.20	0.14	0.17	0.22	0.26	0.14		
25	0.665	0.768	0.845	0.893	0.925	0.949	0.961	0.971	0.978	0.987	0.994	0.999	
	0.29	0.44	0.25	0.16	0.07	0.31	0.27	0.26	0.33	0.34	0.21	0.08	
30	0.653	0.756	0.834	0.882	0.916	0.940	0.953	0.963	0.972	0.982	0.989	0.996	0.998
	0.89	0.82	0.41	0.17	0.00	0.45	0.45	0.44	0.54	0.55	0.43	0.31	0.01

Table 5.9: MNM-LM method. Currency: Yen. Fitted Matérn correlation function, together with percentage errors.

	2	3	4	5	6	7	8	9	10	12	15	20	25
3	0.958												
	0.63												
4	0.897	0.972											
	1.30	0.35											
5	0.849	0.938	0.984										
	0.76	0.28	0.11										
6	0.810	0.906	0.964	0.989									
	0.34	0.11	0.12	0.21									
7	0.778	0.877	0.942	0.975	0.991								
	0.73	0.52	0.07	0.09	0.18								
8	0.760	0.860	0.929	0.965	0.985	0.994							
	0.66	0.42	0.04	0.04	0.21	0.21							
9	0.744	0.845	0.915	0.955	0.978	0.990	0.994						
	0.70	0.39	0.00	0.08	0.21	0.14	0.20						
10	0.729	0.831	0.903	0.945	0.970	0.985	0.991	0.995					
	0.89	0.43	0.05	0.17	0.24	0.15	0.18	0.26					
12	0.711	0.812	0.887	0.931	0.959	0.977	0.985	0.991	0.994				
	0.59	0.23	0.07	0.09	0.16	0.07	0.02	0.07	0.16				
15	0.694	0.796	0.872	0.917	0.948	0.968	0.978	0.985	0.990	0.994			
	0.22	0.11	0.19	0.07	0.06	0.27	0.16	0.10	0.05	0.09			
20	0.674	0.775	0.852	0.900	0.933	0.956	0.967	0.976	0.983	0.989	0.994		
	0.10	0.35	0.27	0.18	0.06	0.32	0.25	0.23	0.22	0.13	0.14		
25	0.665	0.767	0.844	0.893	0.926	0.950	0.962	0.971	0.979	0.987	0.992	0.995	
	0.23	0.59	0.36	0.17	0.00	0.44	0.40	0.36	0.38	0.28	0.02	0.31	
30	0.654	0.755	0.833	0.882	0.916	0.942	0.954	0.965	0.973	0.982	0.988	0.994	0.995
	0.77	0.97	0.54	0.21	0.06	0.59	0.60	0.58	0.64	0.57	0.34	0.04	0.35

Chapter 6

SPD Covariance Functions via Superposition

6.1 Introduction

This chapter is the last of three in which we construct strictly positive definite covariance functions, and thus infinite-factor Gaussian fields. In the last chapter we developed a procedure for the construction of functional forms of strictly positive definite (SPD) correlation functions. This involved taking an SPD stationary function and deforming its coordinates in a strictly increasing fashion. Though this methodology provided us with some flexibility in fitting instantaneous correlation and covariance functions, it might prove to be too restrictive.

In this chapter we suggest an alternative methodology for the construction of SPD covariance functions. The idea is very intuitive and is close to the way participants in fixed income markets think about the structure of yield curve changes. We model the instantaneous covariance function as a superposition of two components. The first component captures the covariation of the main driving factors. The second part of the covariance function captures the infinite-factor noise. We implement this

methodology and study it on the US Dollar and Japanese Yen interest rate data. Our results suggest that this methodology provides the best solution for construction SPD correlation and covariance functions.

The organisation of this chapter is as follows. In Section 6.2 we introduce the modelling framework. The implementation methodology is presented in Section 6.3. In Section 6.4 we test the methodology in a number of numerical experiments and discuss the results. In Section 6.5 we set out conclusions.

6.2 Modelling Framework

In this section we present our modelling framework. We start the section by constructing a generic SPD covariance function. Then, we remark on other constructions that have appeared in the research literature and can be related to our methodology.

We suggest covariance functions of the form,

$$C(s, t) = \sum_{n=1}^N \lambda_n \phi_n(s) \phi_n(t) + \alpha \sigma(s) \sigma(t) \rho(s - t), \quad (6.2.1)$$

where ϕ_n are basis functions, $\lambda_n > 0$. The function σ^2 is strictly positive, the function ρ is SPD, and $0 \leq \alpha \leq 1$ is a scaling factor. It is a simple consequence of the results in linear algebra that the function (6.2.1) is SPD. The main ingredient in this construction is the SPD function ρ , as it carries the infinite-factor structure. We have characterised a large class of such SPD function in the previous chapter. This should provide us with a sufficient number of function that we can use in construction (6.2.1). Of crucial importance is the parameter α in (6.2.1). It defines the proportion of variance explained by the stationary part of the covariance function. For high values of α the correlation function of the process becomes nearly stationary. For low values of α the process becomes degenerate, as it is only driven by the finite factor structure in the second part of (6.2.1).

If we assume that the changes in the yield curve come from a Gaussian field, as in our case, the covariance function (6.2.1) can be thought of as coming from the field

$$R(t) = \sqrt{\alpha}S(t) + \sum_{n=1}^N X_n \sqrt{\lambda_n} \phi_n(t),$$

where S is a field with a strictly positive stationary correlation function ρ and positive variance function σ^2 . The random coefficients X_n are uncorrelated random variables with zero mean and unit variance.

Thus, the field is a sum of N random factors X_n and a stationary infinite-factor noise structure represented by the field S . The fraction of variance contributed by the stationary part of the covariance function is α . We refer to the parameter α as the stationarity parameter.

This type of construction is analogous to the N -factor APT model, Ross (1976), where the return of each security depends on N market risks plus an idiosyncratic risk. Similar ideas can also be found in the multiple-factor risk models used in fixed income markets. They are based on the notion that the returns of a bond can be explained by a linear combination of common factors plus an idiosyncratic element that pertains to that particular bond. That is if we denote by X the vector of spot rate changes, the objective is to approximate X as a linear combination of a small set of fundamental shifts. Thus,

$$X = \hat{X} + \sum_{k=1}^K h_k U_k + E,$$

where \hat{X} is the vector of average yield changes, h_k the random factors, U_k the eigenvectors, or the directions of the changes, and E is the error vector. Note, these models usually assume that errors are independent. To determine the set of vectors U_1, \dots, U_K we can perform a Principal Component Analysis. In this case, the set U_1, \dots, U_K is the set of the first K eigenvectors of the covariance matrix ranked by the corresponding eigenvalue.

It is common to find that the largest part of the variation of the term structure can be explained by only three factors¹. The largest impact comes from the parallel move or *level* factor, followed by *steepness* and *curvature*. These names come from the shape of the first three eigenvectors.

Remark 6.2.1. Within the pricing framework, Rebonato (1999b) suggested a procedure for constructing a correlation matrix for calibration of LIBOR market models². Next, we will show how his procedure can be seen as a degenerate example of our approach. In particular, Rebonato (1999b) suggested approximating the implied correlation matrix \hat{C} by a matrix of the form BB^T , where the matrix B is constructed via the following algorithm,

$$\begin{aligned} b_{ij} &= \cos \theta_{ij} \prod_{k=1}^{j-1} \sin \theta_{ik} & : \quad j = 1, \dots, n-1, \\ b_{ij} &= \prod_{k=1}^{j-1} \sin \theta_{ik} & : \quad j = n, \end{aligned}$$

for an arbitrary set of angles θ_{ij} . To embed his procedure into our framework, we start with the sequence of functions,

$$\begin{aligned} \psi_j(t) &= \cos(t) \sin^{j-1}(t) & : \quad j = 1, \dots, n-1, \\ \psi_j(t) &= \sin^{j-1}(t) & : \quad j = n, \end{aligned}$$

which are linearly independent in some function space. Trivially, we obtain a new sequence by a simple transformation of the old one, i.e.

$$\begin{aligned} \phi_j(t) &= \psi_j(f_j(t)) & : \quad j = 1, \dots, n-1, \\ \phi_j(t) &= \psi_j(f_j(t)) & : \quad j = n. \end{aligned}$$

The transformations $f_j(t)$ are continuous time-equivalents of the discrete mapping $ij \rightarrow \theta_{ij}$. The functional sequence $\phi_j(t)$, $j = 1, \dots, n$, is a continuous time-equivalent of the sequence b_{ij} that Rebonato used in his constructions. The sequence $\phi_j(t)$ in turn can be used in the construction of a degenerate (not SPD) covariance function (6.2.1), with $\alpha = 0$ and $\lambda_n = 1$, for $n = 1, \dots, N$.

¹See, for example, Litterman and Scheinkman (1988), Barber and Copper (1996), Knez *et al.* (1994), Reitano (1992) and (1996).

²These are models developed by Miltersen *et al.* (1997), and Brace *et al.* (1997).

Remark 6.2.2. Here, we note how the covariance construction of type (6.2.1) arise naturally in the theory of stochastic processes.

As a first example, we consider the Karhunen-Loève expansion for a Brownian motion on an interval $[0, T]$. To obtain the bi-orthogonal series we need to solve the integral equation,

$$\int_0^T K(t, s)\phi(s)ds = \lambda\phi(t) \quad 0 \leq t \leq T, \quad (6.2.2)$$

where the function $K(t, s) = \min(t, s)$ is the covariance function of a Brownian motion. Differentiating twice with respect to t yields

$$-\phi(t) = \lambda\ddot{\phi}(t) \quad 0 \leq t \leq T, \quad (6.2.3)$$

with boundary conditions $\phi(0)$ and $\dot{\phi}(T) = 0$. With the help of equation (6.2.3) we obtain the eigenvalues³ of (6.2.2),

$$\lambda_n \frac{T^2}{(n + \frac{1}{2})^2 \pi^2} \quad n = 0, 1, 2, \dots$$

with normalised eigenfunctions

$$\phi_n(t) = \sqrt{\frac{2}{T}} \sin \left[\left(n + \frac{1}{2}\right) \pi \left(\frac{t}{T}\right) \right].$$

It then follows from Mercer's theorem that

$$\min(t, s) = \frac{2}{T} \sum_{n=0}^{\infty} \frac{T^2}{\pi^2 (n + \frac{1}{2})^2} \sin \left[\left(n + \frac{1}{2}\right) \pi \left(\frac{t}{T}\right) \right] \sin \left[\left(n + \frac{1}{2}\right) \pi \left(\frac{s}{T}\right) \right],$$

uniformly on $[0, T]^2$. In particular, all the eigenvalues λ_n are positive, so the function $\min(t, s)$ is definite in the sense of Hilbert. Moreover, the Karhunen-Loève expansion of the Brownian motion is

$$W_t = \sum_{n=0}^{\infty} \sqrt{\lambda_n} \phi_n(t) Z_n, \quad (6.2.4)$$

³For a detailed discussion of this example, see Wong (1971).

where Z_0, Z_1, \dots are independent standard normals. The eigenvalues λ_n decrease with n and the eigenfunctions ϕ_n are of increasing frequency, so the Z_n fill in increasingly fine details in this representation of the Brownian path.

Representation (6.2.4) in the last equation can be seen as an example of our construction (6.2.1) with $\alpha = 0$ and $N = \infty$, i.e. an infinite number of basis functions $\phi_n(\cdot)$, and weights λ_n . Alternatively, we could have cut off the series after finitely many basis functions and used the remaining part of the function as the field in analogy to the field S in construction (6.2.1).

Other expansions of the Brownian motion covariance function $\min(t, s)$ are possible. The counterpart of the discrete Brownian bridge construction for continuous paths is the Lévy-Ciessilski expansion⁴. We define Haar functions on $[0, 1]$ by setting $H_1(t) \equiv 1$, $H_2(t) = 1$ on $[0, 1/2)$ and -1 on $[1/2, 1]$,

$$H_{2^{n+1}}(t) = \begin{cases} 2^{n/2}, & 0 \leq t < 2^{-(n+1)}, \\ -2^{n/2}, & 2^{-(n+1)} \leq t < 2^{-n}, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$H_{2^n+j}(t) = H_{2^n+1}(t - \frac{j-1}{2^n}), \quad j = 1, \dots, 2^n - 1.$$

The Haar functions form a complete orthonormal basis in $L^2[0, 1]$. From the Haar functions we define the Schauder functions

$$F_m(t) = \int_0^t H_m(u) du.$$

With the help of the Schauder functions the continuous Brownian path can be represented as

$$W_t = \sum_{n=0}^{\infty} F_n(t) Z_n. \tag{6.2.5}$$

⁴For a detailed exposition of this construction see, for example, Janicki and Weron (1994).

Again, representation (6.2.5) can be embedded in construction (6.2.1), with $\alpha = 0$ and $N = \infty$, i.e. an infinite number of basis functions $\phi_n(\cdot) = F_n(\cdot)$ with weights $\lambda_n = 1$.

The Karhunen-Loève expansion and Principal Components construction are optimal⁵ in the sense that they allocate maximum variability to each initial portion of the driving sequence Z_0, Z_1, \dots

6.3 Implementation Methodology

In this section we explain how we can implement the methodology suggested above. The implementation consists of several steps. First we need to decide which SPD function we want to use. Next we need a means for choosing the parameter vector for the SPD function ρ , the stationarity parameter α , and the weights λ_n . Furthermore, we need to find appropriate basis functions ϕ_n , and the variance function σ^2 .

We start with the variance function $\sigma(\cdot)$. This is relatively easy. Starting from the sample variance, the problem consists in interpolating the variance for the unknown maturities. There is a wide range of interpolation techniques that can be used for this problem. However, whatever the choice, the resulting function should reflect the commonly observed properties of the sample variance, i.e. a hump around early maturities and monotonic behaviour elsewhere. In particular, the function shouldn't exhibit "wiggleness", as commonly observed in polynomial interpolations⁷.

Next, we choose an SPD function and fit it to the sample correlation matrix. We are not concerned with achieving a perfect fit at this stage. We mainly want to capture

⁵These properties of the expansions are usually used in the context of Monte Carlo simulations of second-order stochastic processes. Two of the most powerful techniques are the *Brownian bridge* and *Principal Component*. For a description of these techniques see Acworth *et al.* (1998), and references therein. The generalisation of principal component analysis for general second-order processes is the Karhunen-Loève ⁶ expansion. This procedure allows an expansion of a process in the quadratic mean in terms of the bi-orthogonal series.

⁷For a discussion of a possible choice of approximation or interpolation technique for the variance functions see, for example, James and Webber (2000)

the general shape of the sample correlation matrix. To complete the construction of the second part of the SPD covariance function (6.2.1), we choose the stationarity parameter $0 < \alpha < 1$.

The remaining part of the implementation is to construct the the first part of the SPD covariance function (6.2.1), the driving finite-factor structure. We calculate the residual matrix $R_{i,j}$,

$$R_{i,j} = C^s(T_i, T_j) - \alpha \sigma(T_i) \sigma(T_j) \rho(T_i - T_j).$$

This residual matrix $R_{i,j}$ captures the main factor structure in the covariance matrix. We then perform the PCA analysis on this residual matrix. This gives us a set of eigenvalues and corresponding eigenvectors.

To obtain a perfect fit of our SPD covariance function (6.2.1) to the sample covariance matrix, all eigenvalues of the residual matrix should be positive. However, if they are not, we need to calculate the residual matrix with a lower choice of the stationarity parameter α . We will say more on the choice of this parameter in the next section.

Once we are satisfied with the choice of eigenvalues λ , we need to find “basis” functions⁸ ϕ_n . A natural choice would be just the interpolants or approximants of the eigenvectors obtained from the PCA. The same comments apply to the interpolation/approximation of eigenvectors as to the case of sample variance discussed at the beginning of the section.

This completes the construction of our SPD covariance function (6.2.1). We summarise the implementation algorithm:

1. Construct the variance function σ^2 .

⁸These functions are not basis functions in the usual sense, as used in Functional Analysis or elsewhere in mathematics. They are sometimes referred to as empirical basis functions.

2. Fit an SPD function to the sample correlation matrix.
3. Choose the variance contribution parameter α .
4. Obtain the residual sample covariance

$$R_{i,j} = C^s(T_i, T_j) - \alpha \sigma(T_i) \sigma(T_j) \rho(T_i - T_j).$$

5. Perform Principal Component Analysis (PCA) on the residual matrix $R(T_i, T_j)$.
6. If all eigenvalues obtained from the PCA are nonnegative, set λ_n equal to the first N positive eigenvalues from the PCA. Otherwise go to step 3 with a lower parameter α .
7. Construct the basis functions ϕ_n .

In this section we have discussed the implementation procedure of the methodology presented in the previous section. The procedure almost completely avoids optimisation, and is mainly based on the interpolation of several quantities. Several numerical tests have shown that the procedure is very robust. In particular, the search for the optimal parameters for the SPD function ρ , by far the most difficult part of the algorithm, appears to be very stable. This procedure provides an exact fit of the SPD covariance function (6.2.1) to the sample covariance and matrices. In the next section we present numerical tests of the methodology on several data sets.

6.4 Numerical and Empirical Studies

We applied the methodology to the data sets used in the two previous chapters, specifically to US Dollars and Japanese Yen. For this purpose we chose two examples of SPD functions described in Chapter 5, Section 5.3: the Exponential-Power

$$C(h) = \exp(-a^2|h|^p),$$

and the Matérn correlation functions,

$$C^\eta(h) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\nu^{1/2}h}{\rho} \right)^\nu \mathcal{K}_\nu \left(\frac{2\nu^{1/2}h}{\rho} \right), \quad h \in \mathbb{R}^n,$$

where the function $\Gamma(\cdot)$ is the usual gamma function and \mathcal{K}_ν is the modified Bessel function of order ν . Both functions depend on two parameters, but the latter is more flexible.

We conducted four numerical studies based on these two SPD functions and the two data sets. Next, we describe the results and then conclude the section with a general discussion of the different aspects of the methodology.

Exponential-Power SPD function. Currency: Yen.

We began the numerical investigation of our methodology by fitting the SPD covariance function (6.2.1) to the sample covariance matrix coming from Japanese Yen interest rate data.

As stationarity parameter we chose $\alpha = 0.1$. We performed one step of the above suggested algorithm, i.e. we did not change the values of α . Not all of the eigenvalues of the residual matrix were positive, so we did not achieve a perfect fit.

Table 6.5 presents the values of the correlation function resulting from our construction together with the percentage errors. Table 6.6 presents the same values, but for the covariance function. As can be seen, though the fit is not perfect, it is very good. In the case of the covariance function we achieved a maximum percentage error of the 0.42, and in the case of correlation function we obtained a slightly larger maximum percentage error of 0.55 for the maturities 25 and 30.

Observe that the fit for most of the maturities is exact, with the exception of distant maturities. The parameters for the SPD function are $a = 0.0294$ and $p = 0.537$. We plotted the resulting SPD covariance and correlation functions in Figure 6.1. Furthermore, we presented the ordered set of eigenvalues of the residual covariance

matrix R in Table 6.1. There are six negative eigenvalues. This is the source of the error we observed in our fit. However, they are relatively small, so the percentage error is less than 0.5 %.

As a test, we evaluated covariance and correlation functions on the equidistant grid with intervals of half a year. The logs of the resulting eigenvalues are plotted in Figure 6.2 (above). As can be seen, this confirms our theoretical result that all eigenvalues should be positive. Additionally, we plotted in Figures 6.2 (below) and 6.3 the first nine empirical basis functions ϕ_n .

The first three eigenfunctions exhibit the well-known shapes that are usually referred to as level, steepness, and curvature. It should be noted, however, that Principal Component Analysis was performed on the residual sample covariance matrix and not on the sample covariance matrix itself, as is usually the case. The higher-order empirical basis functions exhibit higher frequencies and are harder to interpret. The Exponential-Power SPD function with fitted parameters is plotted in Figure 6.13 (above). Note the spiky shape at the origin. We have also plotted the constructed covariance and correlation functions (Figure 6.1). Note that the spike of the SPD function is reflected in the spike along the diagonal of the correlation function.

Matérn correlation function. Currency: Yen. In this experiment, we investigated the Matérn SPD function on the same data set, Japanese Yen, as in the first exercise. The stationarity parameter $\alpha = 0.1$ as before.

Table 6.7 presents the values of the correlation function resulting from our construction, together with the percentage errors. Table 6.8 presents the same values but for the covariance function. As expected all percentage errors are zero. In both cases, the maximum percentage error of the covariance function is practically nil. The same is true of the correlation function. The quality of the fit is not surprising, as the Matérn SPD function is more flexible than the Exponential-Power.

We have plotted the Matérn SPD function with the fitted parameters $\nu = 109.341$ and $\rho = 51.975$ in Figure 6.13 (above). Note the rounded shape at the origin. We also plotted the constructed SPD covariance and correlation functions in Figure 6.4. Note, that the diagonal of the correlation function is smooth. This can be explained by the shape of the Matérn correlation function in Figure 6.13 (above).

The eigenvalues of the residual covariance matrix are presented in Table 6.2. There is only one negative eigenvalue, which explains the quality of fit of our covariance and correlation functions. Again, as a test, we evaluated the covariance and correlation functions on an equidistant grid with intervals of half a year. The logs of the resulting eigenvalues are plotted in Figure 6.5 (above). All eigenvalues are positive. We have plotted the first nine empirical basis functions ϕ_n in Figures 6.5 (below) and 6.6. The shapes of the basis functions are similar to those in the previous exercise.

Exponential-Power SPD function. Currency: Dollar. We continued our numerical analysis with construction of the SPD covariance functions based on the sample covariance matrix coming from US Dollar interest rate data. As the SPD function we chose the Exponential-Power. The stationarity parameter was again $\alpha = 0.1$.

This construction resulted in the exact fit of covariance and correlation functions to the sample correlation and covariance matrices. Tables 6.9 and 6.10 present the values of the correlation and covariance functions resulting from our construction together with the percentage errors. The ordered set of eigenvalues of the residual covariance matrix R can be seen in Table 6.3. All eigenvalues are positive, so the fit is exact.

The constructed SPD covariance and correlation functions can be seen in Figure 6.7. Note the fast decorrelation and non-smoothness of the correlation function on the diagonal. This effect results from the spiky stationary component of the covariance

function. To show this, we have plotted the Exponential-Power SPD function with the fitted parameters $a = 0.1035$ and $p = 0.4315$. in Figure 6.13 (below). Note the spiky shape at the origin.

As in the exercises above, we evaluated the covariance and correlation functions on a equidistant grid with intervals of half a year, and plotted the logs of the resulting eigenvalues in Figure 6.8 (above). Again, all of the eigenvalues are positive. The first nine empirical basis function ϕ_n are plotted in Figures 6.8 (below), and 6.9. The first three empirical basis functions do not exhibit any clear shapes. This might be due to poor data quality.

Matérn correlation function. Currency: Dollar.

For this exercise, we constructed the SPD covariance function, based again on sample covariance functions coming from the US Dollar interest data. We used Matérn SPD function as the stationary component of the covariance function, and the stationarity parameter $\alpha = 0.1$ We achieved an an exact fit of the SPD covariance and correlation functions to the sample matrices.

We have plotted the Matérn correlation function with the fitted parameters $\nu = 106.33$ and $\rho = 31.429$ in Figure 6.13 (below). Again, note the rounded shape at the origin. The correlation and covariance functions can be seen in Figure 6.10. Note that the diagonal of the correlation function is smooth. This can be explained by the shape of the Matérn correlation function in Figure 6.13 (below). The eigenvalues of the residual covariance matrix R are again positive as can be seen in Table 6.4.

The logs of the eigenvalues of the function evaluated on a semi-annual grid are plotted in Figure 6.11 (above). All eigenvalues are positive. We have plotted the first nine empirical basis functions ϕ_n in Figures 6.11 (below) and 6.12. The shapes of the basis functions are similar to those of the previous exercise.

General Remarks. It is interesting to understand the influence of the stationarity parameter α on the SPD covariance construction. To see this, we constructed the SPD covariance function for several different values of α , for both types of SPD functions, the Exponential-Power and the Matérn functions. We have plotted the percentage errors of the fit against α in Figure 6.14.

Observe that the percentage error crucially depends on the type of SPD function used. In the case of the Exponential-Power function a significant error can be seen for α larger than 0.25. However, the increase in the error with increasing α is linear with a small slope. The behaviour of the percentage error is entirely different in the case of the Matérn SPD function. It becomes significant for α larger than 0.5. However, the increase is sharp and non-linear.

Next we have investigated the role of the underlying SPD functions. We plotted both constructed correlation functions for the high value $\alpha = 0.9$ in Figure 6.15. This is, of course, a very high value, but it should highlight the importance of the choice of underlying SPD function. The correlation functions in Figure 6.15 show entirely different behaviour on the diagonal, though they have been fitted to the same data. Both functions have infinite-dimensional structure, but only in the case of the stationary Exponential-Power function do we achieve fast decorrelation. This raises an important point: when dealing with functions the speed of decorrelation cannot be explained simply by the dimensionality of the underlying eigenvalue structure.

We believe that the behaviour of eigenvalues plays an important role in the speed of decorrelation. To see this, compare Figure 6.2 (above) with Figure 6.5 (above). Both figures present eigenvalues of the fitted SPD covariance and correlation functions to the sample covariance matrix of the Japanese Yen evaluated on an equidistant grid. However, they differ in the type of the underlying SPD functions. The first figure results from the Exponential-Power and the second from the Matérn function. Though the fit is equivalently good for both types of stationary correlation functions,

the behaviour of eigenvalues is markedly different.

In the case of the Exponential-Power function we have a smooth decay in the log values of the eigenvalues. In the case of the Matérn function, the decay shows a jump toward extremely small values of the eigenvalues. This shows that there is only a finite number of significant factors that drives this covariance function. This behaviour is reflected in the shape of the correlation function. The correlation function resulting from the use of the Exponential-Power stationary function exhibits a spiky form on the diagonal in Figure 6.1. The behaviour of the correlation function resulting from the Matérn function is smooth on the diagonal in Figure 6.10. We observe exactly the same phenomenon if we compare equivalent figures for the US Dollar. The log decay of eigenvalues is smooth in Figure 6.8, but exhibits discontinuity in Figure 6.11. The shape of the correlation function is spiky in Figure 6.7 and smooth in Figure 6.10. From these observations we conclude that correlation functions exhibiting fast decorrelation have a smooth and slow decay of log eigenvalues.

6.5 Conclusions and Further Research

In this chapter we have developed a further methodology for constructing of the instantaneous SPD covariance and correlation functions. It consists of a superposition of an SPD function, a large class of which we described in the previous chapter, with a PD function. One possible interpretation of this construction is that the PD function captures the main finite-dimensional structure, whereas the SPD function captures the general stationary infinite-factor noise.

Furthermore, we have described an implementation algorithm, and conducted extensive numerical studies based on the sample covariance matrices coming from the Japanese Yen and US Dollar interest rate data. The numerical results show that this construction is very flexible and can handle sample correlation matrices with

both monotonic and non-monotonic decay. In most cases we achieved a perfect fit to the sample covariance and correlation matrices.

We believe that the methodology developed in this chapter provides the best solution for our problem: the construction of instantaneous SPD covariance and correlation functions for use in random field models of term structure of interest rates. The SPD covariance and correlation functions constructed here are extremely flexible, and will probably fit any term structure sample covariance and correlation matrices. The computational burden is very small. It involves, mainly, function interpolation and calculation of the eigenvalues and eigenvectors of relatively low-dimensional matrices.

In Chapter 7 we turn our attention to the study of finite-factor models. Within the general framework of the pricing kernel, we develop a class of arbitrage-free multi-factor term structure models. This class of models is mainly intended for pricing interest rate derivatives.

6.6 Appendices

6.6.1 Figures

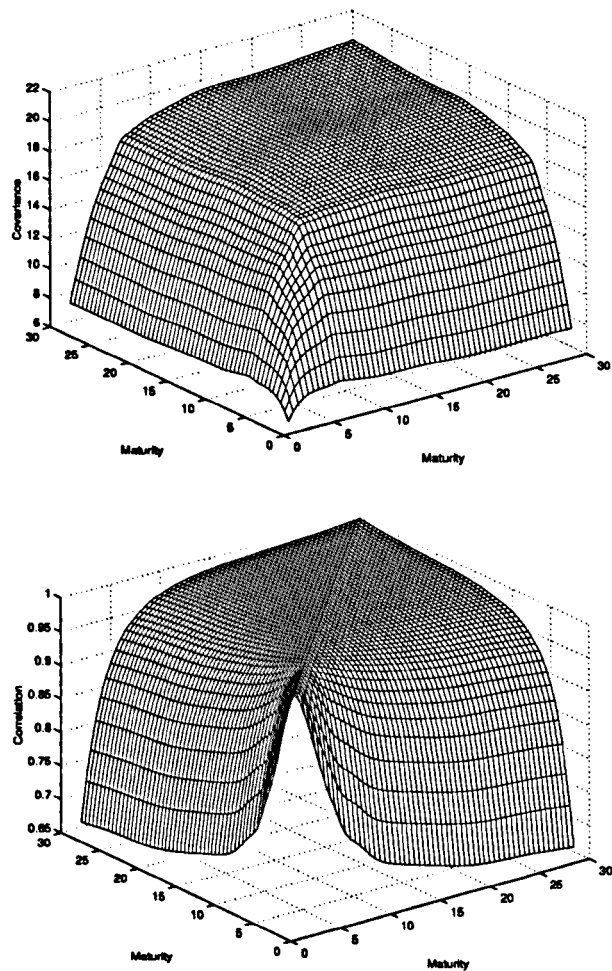


Figure 6.1: SPD covariance function (above) and SPD correlation function resulting from Exponential-Power SPD function with the parameters $a = 0.0294$ and $p = 0.537$, and the stationarity parameter $\alpha = 0.1$. Currency: Yen.

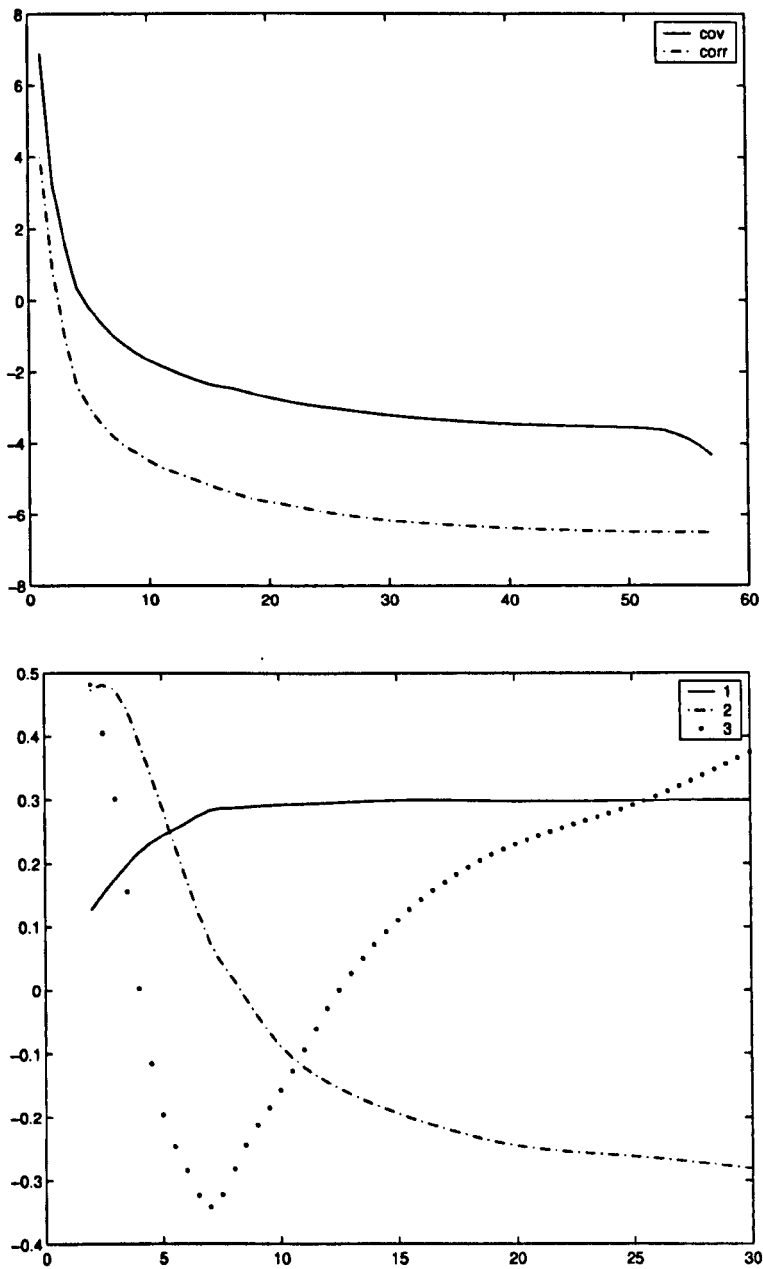


Figure 6.2: Eigenvalues of covariance and correlation functions (above) resulting from the use of exponential-power SPD function. First three empirical basis functions (below). Currency: Yen.

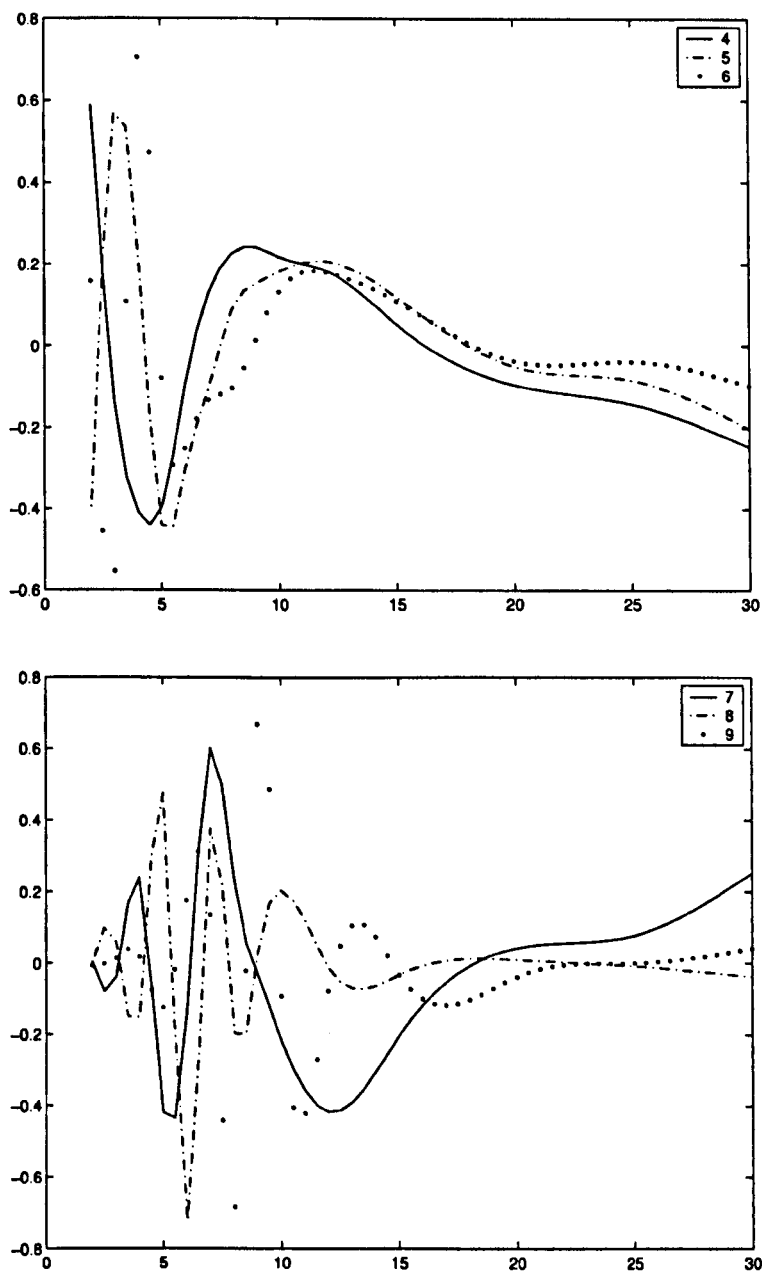


Figure 6.3: Second three empirical basis functions (above). Third three empirical basis functions (below). Currency: Yen.

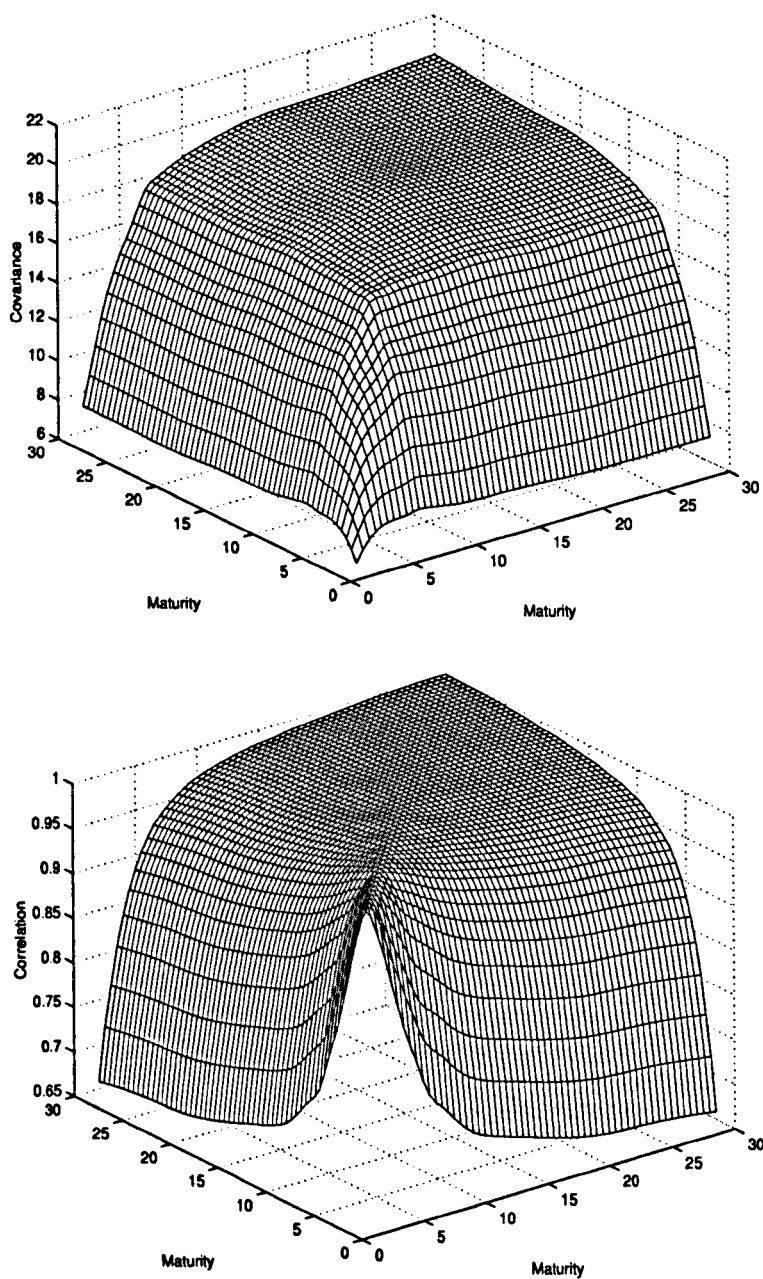


Figure 6.4: SPD covariance function (above) and SPD correlation function resulting from Matérn SPD function with the parameters $\nu = 109.341$ and $\rho = 51.975$, and the stationarity parameter $\alpha = 0.1$. Currency: Yen.

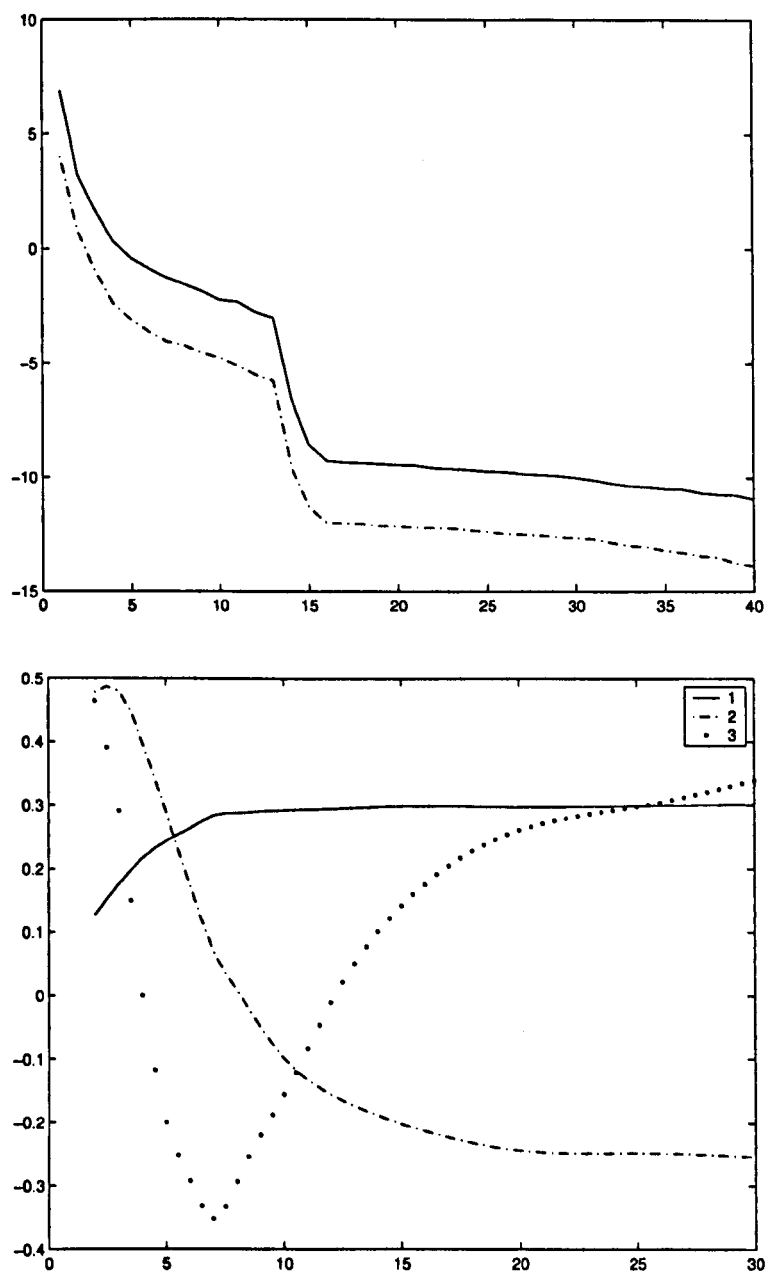


Figure 6.5: Eigenvalues of covariance and correlation functions (above) resulting from Matérn SPD function. First three empirical basis functions (below). Currency: Yen.

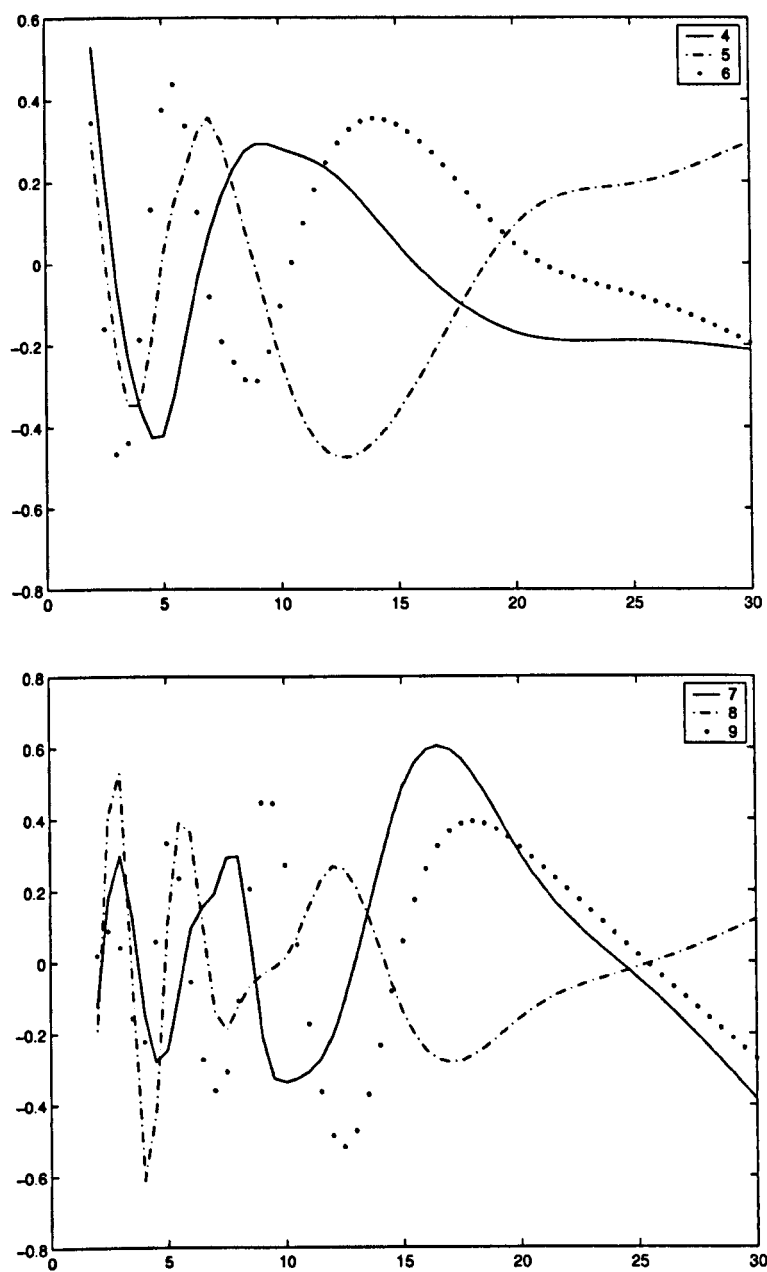


Figure 6.6: Second three empirical basis functions (above). Third three empirical basis functions (below). Currency: Yen.

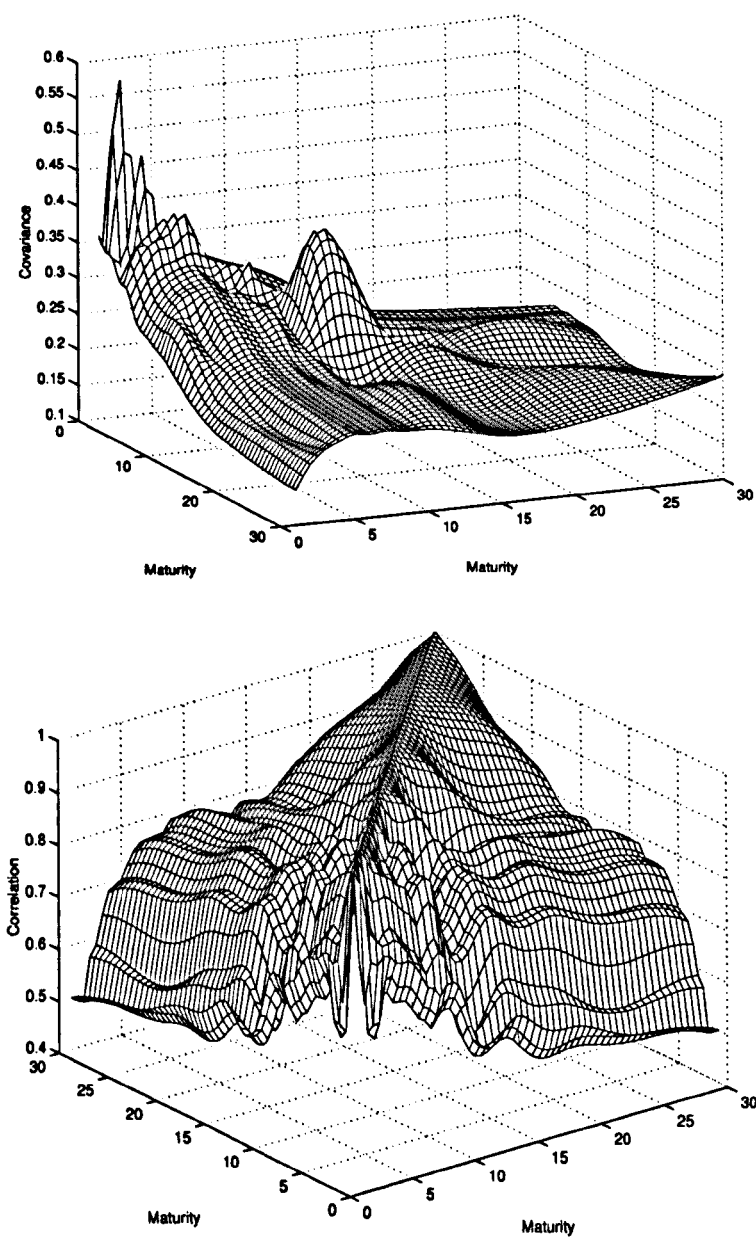


Figure 6.7: SPD covariance function (above) and SPD correlation function resulting from Exponential-Power SPD function with the parameters $a = 0.1035$ and $p = 0.4315$, and the stationarity parameter $\alpha = 0.1$. Currency: Dollar.

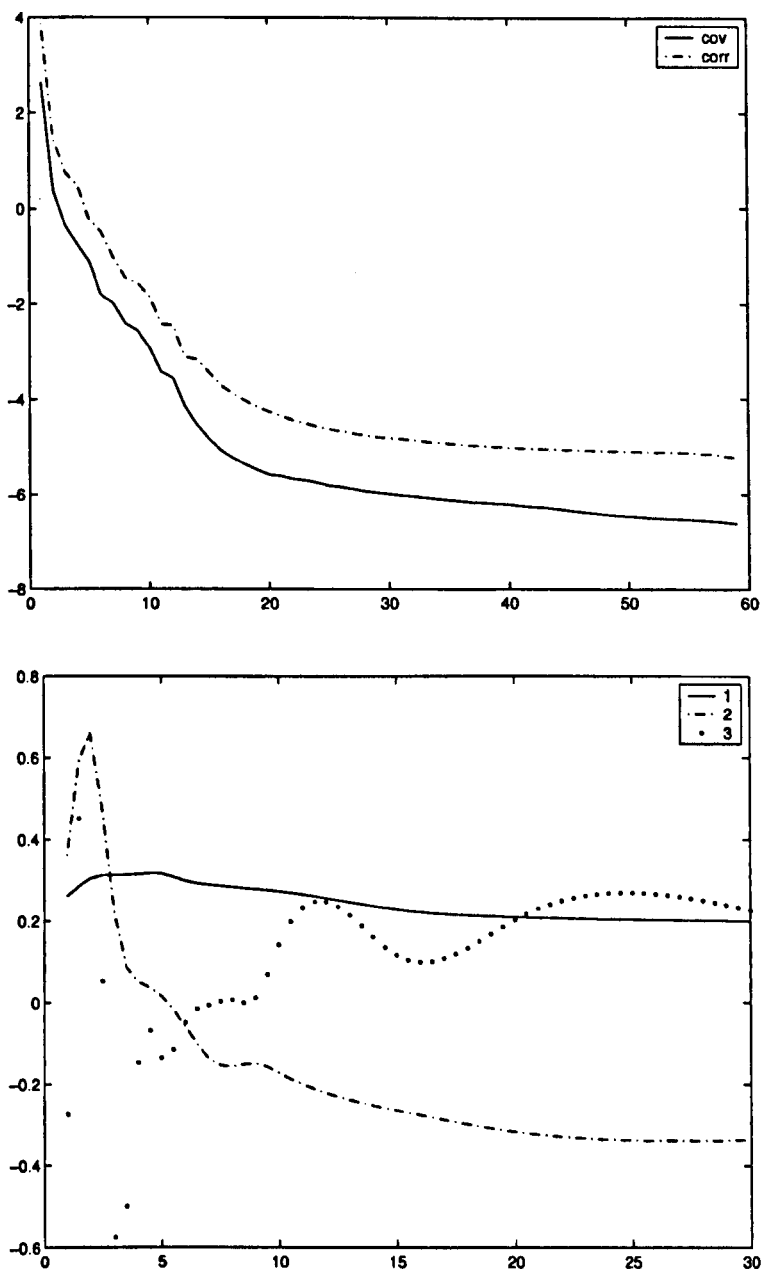


Figure 6.8: Eigenvalues of covariance and correlation functions (above). First three empirical basis functions (below). Currency: Dollar.

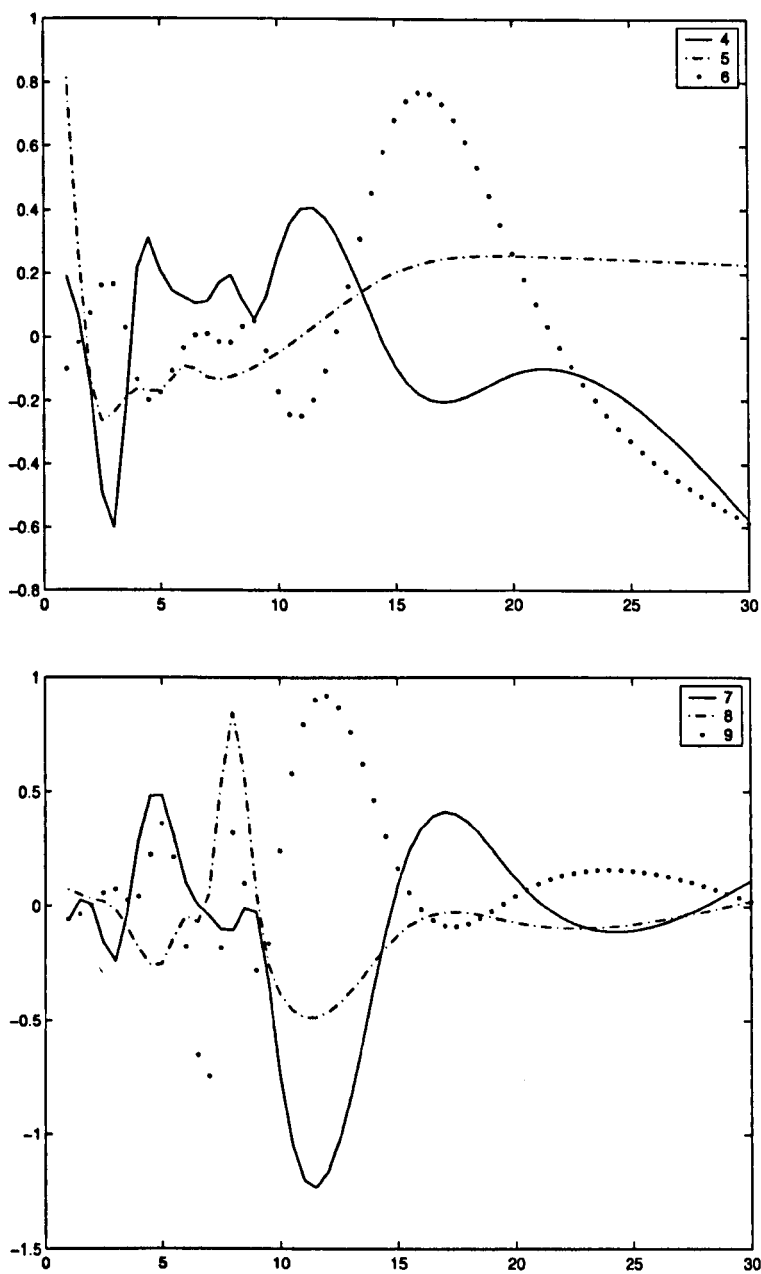


Figure 6.9: Second three empirical basis functions (above). Third three empirical basis functions (below). Currency: Dollar.

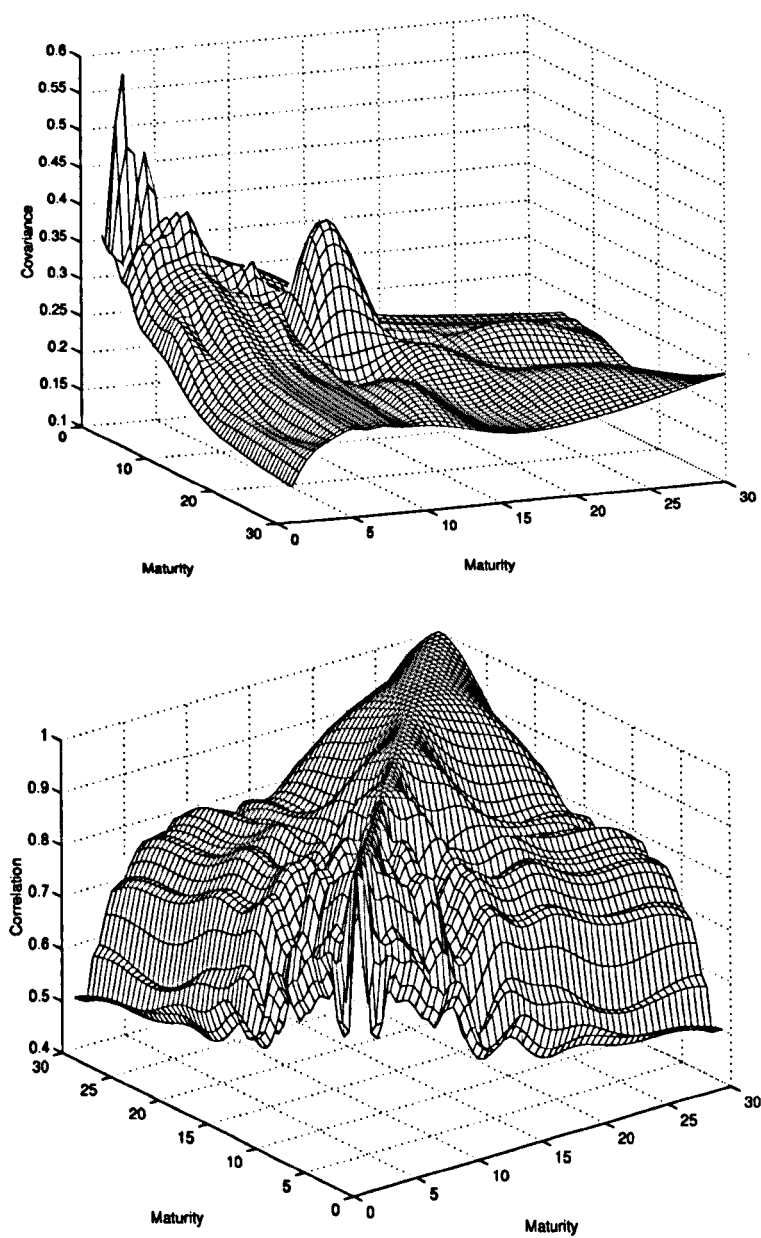


Figure 6.10: SPD covariance function (above) and SPD correlation function resulting from Matérn SPD function with the parameters $\nu = 106.33$ and $\rho = 31.429$, and the stationarity parameter $\alpha = 0.1$. Currency: Dollar.

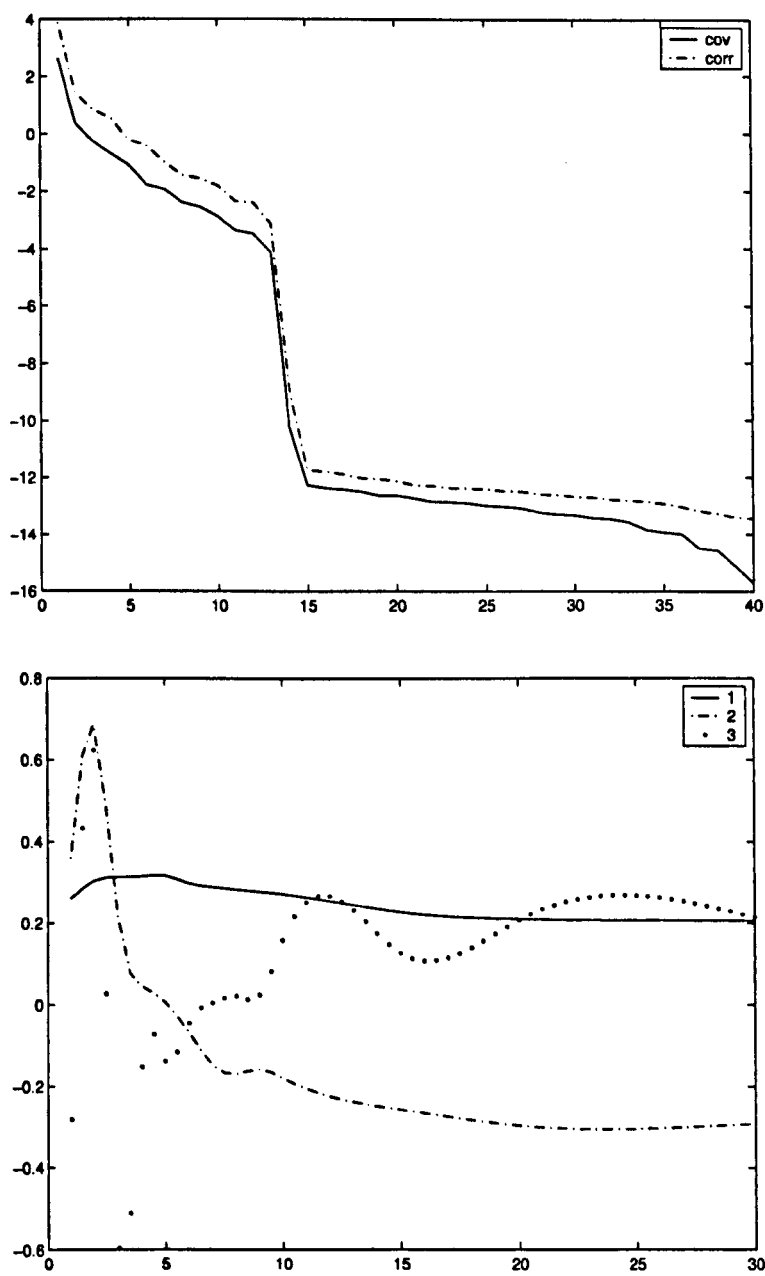


Figure 6.11: Eigenvalues of covariance and correlation functions (above). First three empirical basis functions (below). Currency: Dollar.

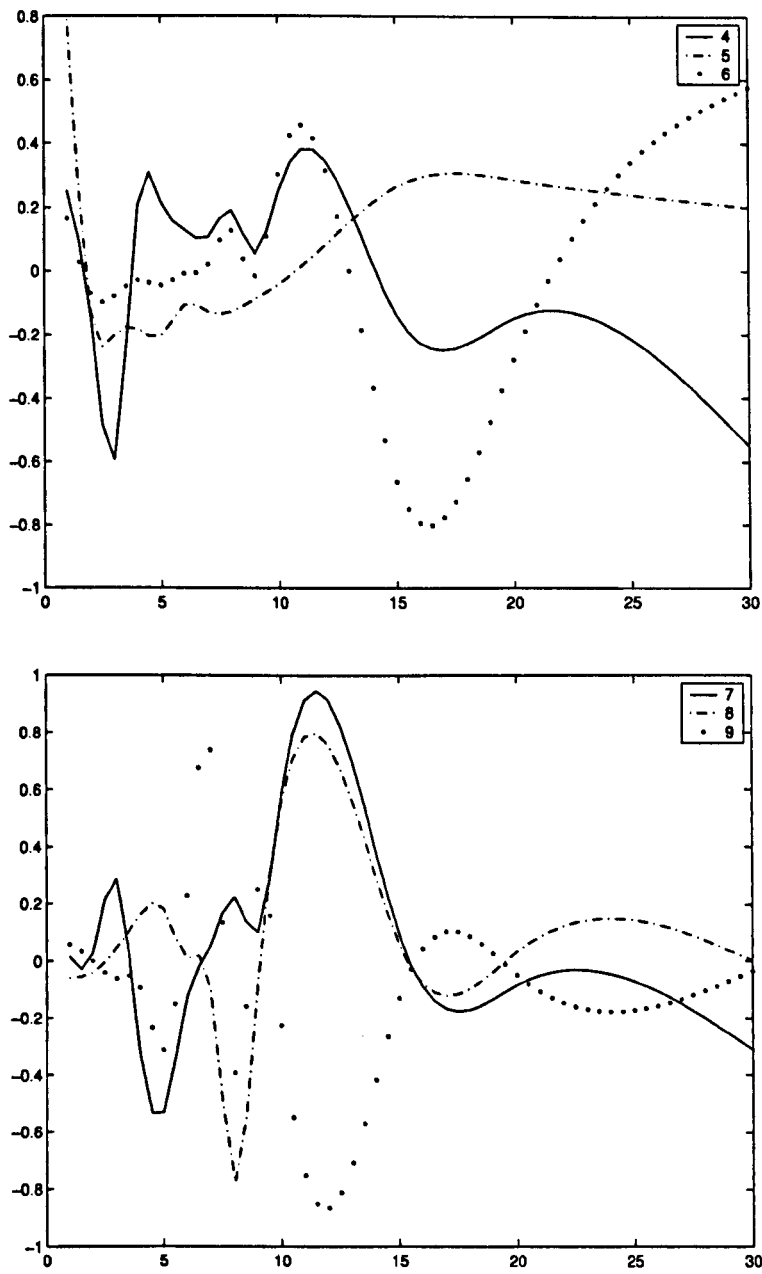


Figure 6.12: Second three empirical basis functions (above). Third three empirical basis functions (below). Currency: Dollar.

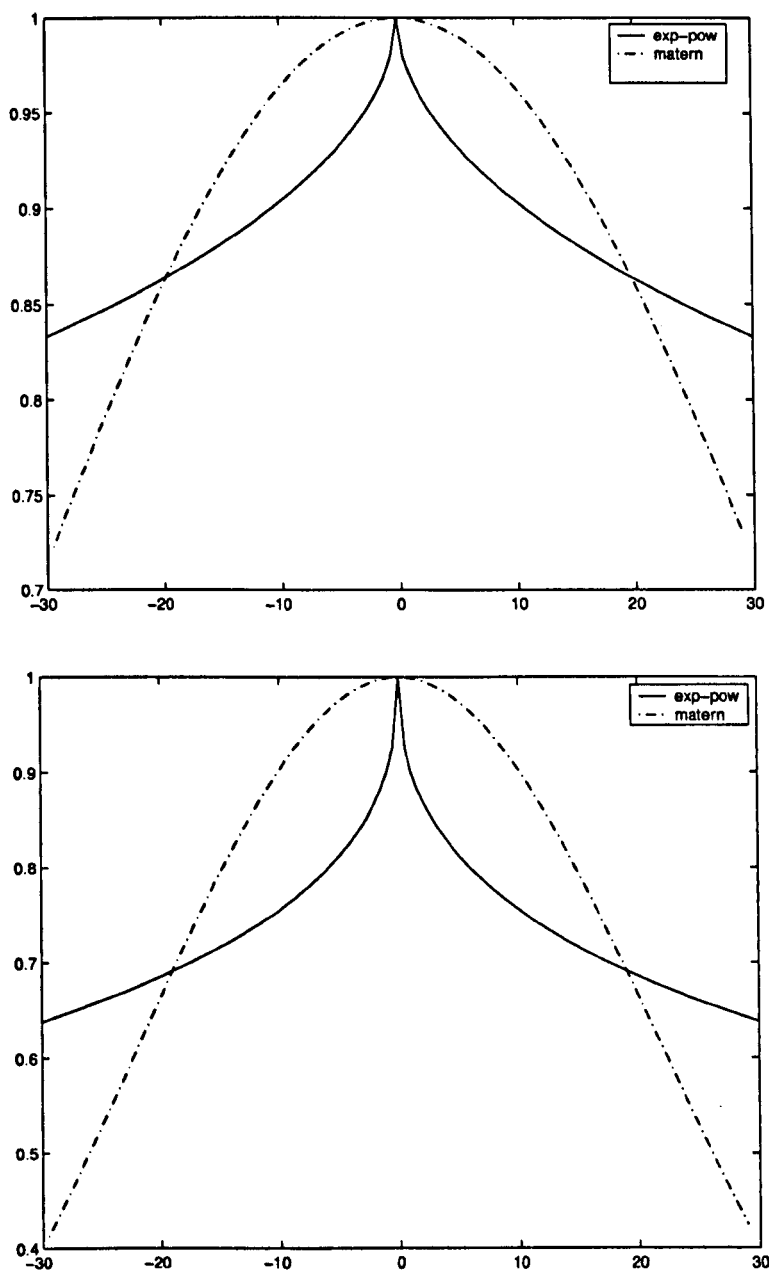


Figure 6.13: Exponential-Power SPD function with parameters $a = 0.0294$ and $p = 0.537$, and Matérn with parameter $\nu = 109.341$ and $\rho = 51.975$. Currency: Yen (above). Exponential-Power SPD function with parameters $a = 0.1035$ and $p = 0.4315$, and Matérn with parameter $\nu = 106.33$ and $\rho = 31.429$. Currency: Dollar (below).

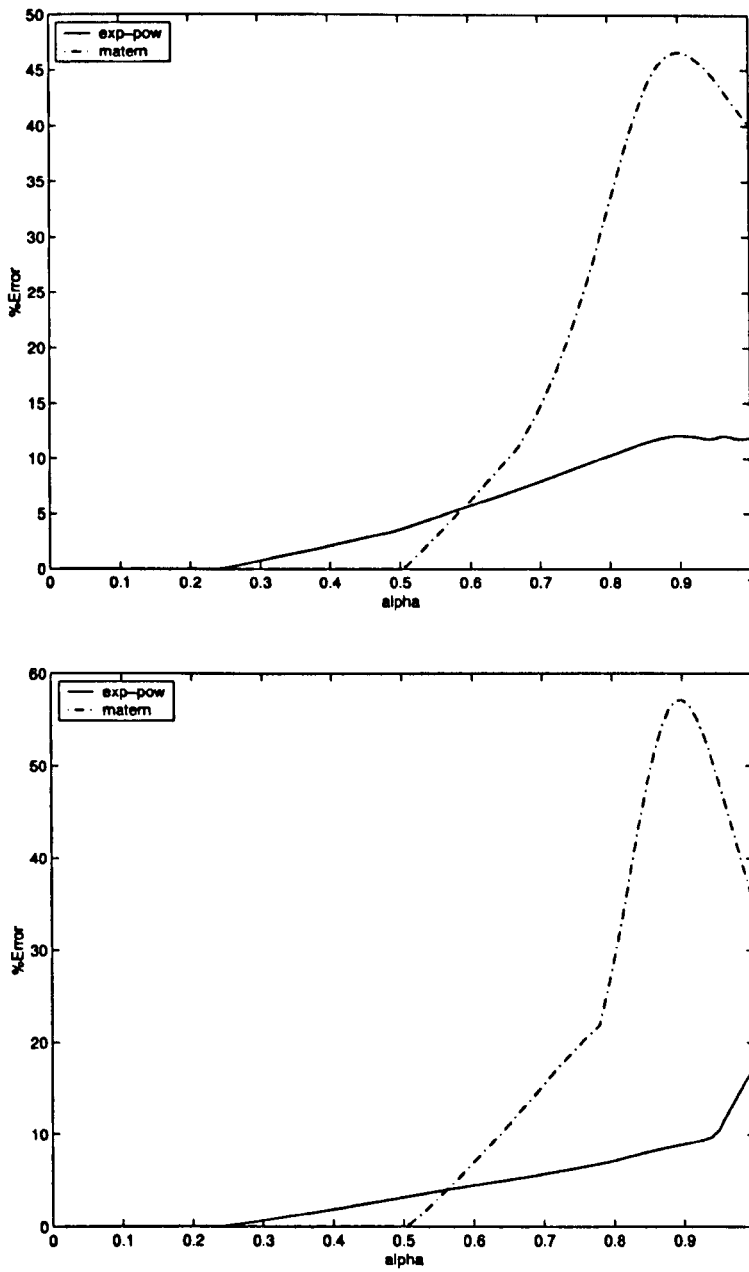


Figure 6.14: Alpha against percentage errors in correlation (above). Alpha against percentage errors in correlation (below). Currency: Dollar.

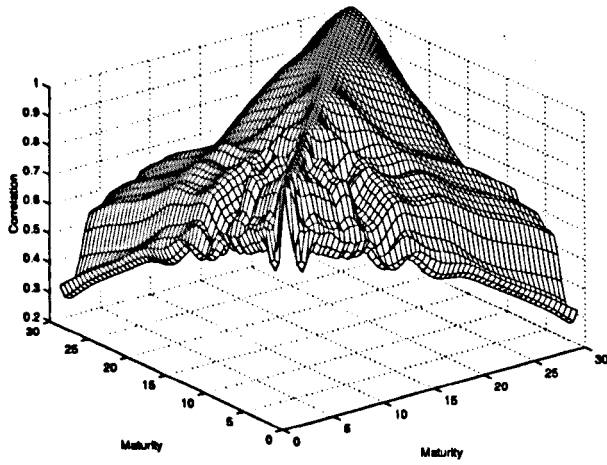
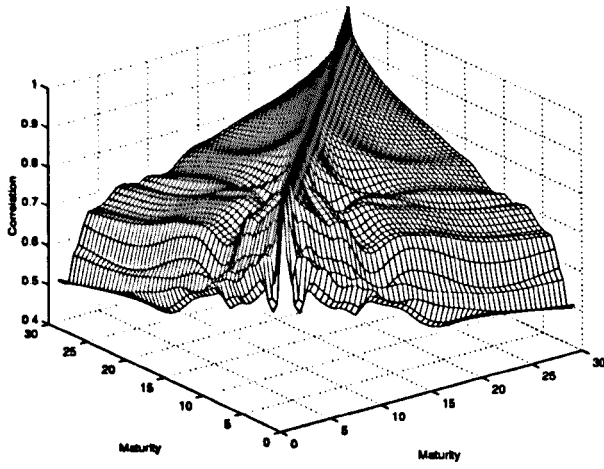


Figure 6.15: Correlation function with Exponential-Power SPD function (above). Correlation function with Matérn SPD function (below).

6.6.2 Tables

Table 6.1: Eigenvalues of the residual covarinace matrix R using stationary Exponential-Power SPD function. Currency: Yen.

1	190.567	6	0.041	11	-0.024
2	8.715	7	0.018	12	-0.044
3	1.489	8	0.008	13	-0.085
4	0.275	9	-0.005	14	-0.117
5	0.074	10	-0.014		

Table 6.2: Eigenvalues of the residual covarinace matrix R using stationary Matérn SPD function. Currency: Yen.

1	189.821	6	0.106	11	0.041
2	8.549	7	0.076	12	0.035
3	1.553	8	0.065	13	0.031
4	0.356	9	0.056	14	-0.0003
5	0.164	10	0.045		

Table 6.3: Eigenvalues of the residual covarinace matrix R using stationary Exponential-Power SPD function. Currency: Dollar.

1	3.333	6	0.043	11	0.015
2	0.320	7	0.040	12	0.010
3	0.168	8	0.037	13	0.007
4	0.073	9	0.024		
5	0.062	10	0.016		

Table 6.4: Eigenvalues of the residual covarinace matrix R using stationary Matérn SPD function. Currency: Dollar.

1	3.287	6	0.047	11	0.019
2	0.319	7	0.047	12	0.013
3	0.172	8	0.042	13	0.010
4	0.076	9	0.028		
5	0.068	10	0.021		

Table 6.5: Values of the correlation function for observed maturities together with percentage errors resulting from kernel approach with Exponential-Power SPD function. Currency: Yen.

	2	3	4	5	6	7	8	9	10	12	15	20	25
3	0.964												
	0.00												
4	0.909	0.975											
	0.00	0.00											
5	0.856	0.940	0.985										
	0.00	0.00	0.00										
6	0.813	0.907	0.965	0.991									
	0.00	0.00	0.00	0.01									
7	0.773	0.872	0.941	0.976	0.993								
	0.02	0.01	0.02	0.02	0.01								
8	0.755	0.857	0.928	0.965	0.987	0.996							
	0.01	0.02	0.02	0.01	0.03	0.05							
9	0.738	0.841	0.915	0.955	0.979	0.991	0.996						
	0.02	0.02	0.02	0.01	0.03	0.06	0.02						
10	0.722	0.827	0.903	0.946	0.972	0.986	0.992	0.996					
	0.05	0.05	0.05	0.06	0.04	0.09	0.07	0.07					
12	0.706	0.810	0.887	0.931	0.960	0.976	0.985	0.991	0.994				
	0.04	0.04	0.04	0.05	0.05	0.04	0.06	0.07	0.12				
15	0.692	0.796	0.872	0.917	0.947	0.965	0.975	0.983	0.988	0.993			
	0.10	0.09	0.11	0.09	0.10	0.08	0.12	0.13	0.19	0.18			
20	0.672	0.777	0.853	0.900	0.932	0.951	0.962	0.972	0.978	0.986	0.992		
	0.16	0.18	0.16	0.19	0.17	0.18	0.21	0.20	0.24	0.24	0.31		
25	0.665	0.770	0.846	0.892	0.924	0.944	0.956	0.966	0.973	0.981	0.988	0.993	
	0.21	0.22	0.21	0.22	0.21	0.23	0.22	0.23	0.26	0.28	0.38	0.52	
30	0.658	0.761	0.836	0.883	0.915	0.935	0.947	0.958	0.965	0.975	0.983	0.989	0.993
	0.13	0.11	0.13	0.12	0.14	0.16	0.12	0.13	0.13	0.12	0.19	0.41	0.55

Table 6.6: Values of the covariance function for selected maturities together with percentage errors resulting from kernel approach with Exponential-Power SPD function. Currency: Yen.

	2	3	4	5	6	7	8	9	10	12	15	20	25
2	6.11												
3	0.00	8.94											
4	0.00	0.00	11.50										
5	0.00	0.00	0.00	13.47									
6	0.00	0.00	0.00	0.01	15.09								
7	0.00	0.00	0.00	0.01	0.01	17.27							
8	0.00	0.00	0.00	0.01	0.01	0.03	17.67						
9	0.00	0.00	0.00	0.01	0.01	0.02	0.03	17.99					
10	0.00	0.00	0.00	0.01	0.01	0.02	0.01	0.04	18.28				
12	0.00	0.00	0.00	0.01	0.01	0.03	0.00	0.00	0.10	18.69			
15	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.03	0.08	19.34		
20	0.00	0.00	0.01	0.01	0.00	0.03	0.01	0.02	0.05	0.04	0.19	19.41	
25	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.02	0.02	0.04	0.35	19.69
	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.02	0.08	0.13	0.43

Table 6.7: Values of the correlation function for selected maturities together with percentage errors resulting from kernel approach with Matérn SPD function. Currency: Yen.

	2	3	4	5	6	7	8	9	10	12	15	20	25
3	0.964 0.00												
4	0.909	0.975											
	0.00	0.00											
5	0.856	0.940	0.985										
	0.00	0.00	0.00										
6	0.813	0.907	0.965	0.991									
	0.00	0.00	0.00	0.00									
7	0.773	0.872	0.942	0.976	0.993								
	0.00	0.00	0.00	0.00	0.00								
8	0.755	0.857	0.928	0.966	0.987	0.996							
	0.00	0.00	0.00	0.00	0.00	0.00							
9	0.739	0.841	0.915	0.955	0.980	0.992	0.996						
	0.00	0.00	0.00	0.00	0.00	0.00	0.00						
10	0.723	0.827	0.904	0.946	0.972	0.987	0.993	0.997					
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
12	0.707	0.811	0.888	0.932	0.961	0.977	0.985	0.991	0.996				
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
15	0.692	0.797	0.873	0.918	0.948	0.966	0.976	0.984	0.990	0.995			
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
20	0.673	0.778	0.855	0.902	0.933	0.953	0.964	0.974	0.981	0.988	0.995		
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
25	0.667	0.771	0.847	0.894	0.926	0.946	0.958	0.968	0.975	0.984	0.992	0.998	
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
30	0.659	0.762	0.837	0.884	0.916	0.936	0.948	0.959	0.967	0.976	0.985	0.993	0.998
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6.8: Values of the covariance function for selected maturities together with percentage errors resulting from kernel approach with Matérn SPD function. Currency: Yen.

	2	3	4	5	6	7	8	9	10	12	15	20	25
2	6.11												
3	0.00	8.94											
4	0.00	0.00	11.50										
5	7.62	9.89	0.00										
	0.00	0.00	0.00										
5	7.76	10.32	12.26	13.47									
	0.00	0.00	0.00	0.00									
6	7.81	10.53	12.71	14.12	15.09								
	0.00	0.00	0.00	0.00	0.00								
7	7.93	10.84	13.27	14.88	16.02	17.26							
	0.00	0.00	0.00	0.00	0.00	0.00							
8	7.85	10.77	13.23	14.89	16.12	17.40	17.67						
	0.00	0.00	0.00	0.00	0.00	0.00	0.00						
9	7.74	10.67	13.16	14.87	16.14	17.48	17.76	17.99					
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
10	7.63	10.57	13.09	14.84	16.14	17.52	17.84	18.07	18.26				
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
12	7.54	10.47	13.01	14.77	16.12	17.54	17.90	18.17	18.38	18.67			
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
15	7.52	10.47	13.01	14.80	16.19	17.63	18.03	18.34	18.58	18.90	19.31		
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
20	7.31	10.23	12.75	14.55	15.94	17.41	17.83	18.16	18.43	18.78	19.23	19.34	
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
25	7.30	10.21	12.72	14.53	15.93	17.40	17.83	18.18	18.45	18.82	19.29	19.44	19.61
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6.9: Values of the correlation function for selected maturities together with percentage errors resulting from kernel approach with Exponential-Power SPD function. Currency: Dollar.

	1	2	3	4	5	6	7	8	9	10	15	20
2	0.749											
	0.00											
3	0.815	0.677										
	0.00	0.00										
4	0.826	0.739	0.862									
	0.00	0.00	0.00									
5	0.806	0.721	0.845	0.953								
	0.00	0.00	0.00	0.00								
6	0.794	0.717	0.837	0.937	0.947							
	0.00	0.00	0.00	0.00	0.00							
7	0.745	0.677	0.802	0.913	0.906	0.946						
	0.00	0.00	0.00	0.00	0.00	0.00						
8	0.721	0.651	0.768	0.879	0.879	0.911	0.911					
	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
9	0.742	0.679	0.807	0.916	0.912	0.953	0.955	0.928				
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
10	0.693	0.654	0.720	0.855	0.852	0.891	0.892	0.871	0.913			
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
15	0.628	0.569	0.686	0.795	0.806	0.848	0.856	0.836	0.882	0.836		
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
20	0.577	0.533	0.618	0.763	0.771	0.823	0.836	0.820	0.865	0.828	0.902	
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
30	0.500	0.486	0.587	0.682	0.699	0.745	0.762	0.746	0.787	0.754	0.782	0.843
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6.10: Values of the covariance function for selected maturities together with percentage errors resulting from kernel approach with Exponential-Power SPD function. Currency: Dollar.

1	2	3	4	5	6	7	8	9	10	15	20	25	30
1	0.36												
2	0.00												
3	0.34	0.58											
	0.00	0.00											
4	0.34	0.36	0.48										
5	0.00	0.00	0.00										
6	0.31	0.35	0.37	0.39									
	0.00	0.00	0.00	0.00									
7	0.31	0.35	0.37	0.38	0.40								
	0.00	0.00	0.00	0.00	0.00								
8	0.28	0.32	0.34	0.35	0.35	0.34							
	0.00	0.00	0.00	0.00	0.00	0.00							
9	0.26	0.30	0.32	0.33	0.33	0.32	0.34						
	0.00	0.00	0.00	0.00	0.00	0.00	0.00						
10	0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.34					
	0.25	0.29	0.31	0.32	0.33	0.31	0.31	0.00	0.31				
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
15	0.24	0.29	0.29	0.31	0.32	0.31	0.31	0.30	0.31	0.31			
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
20	0.19	0.22	0.24	0.25	0.26	0.25	0.25	0.25	0.25	0.25	0.33		
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
30	0.17	0.19	0.20	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.22	0.23	
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	0.15	0.18	0.20	0.21	0.22	0.22	0.22	0.22	0.22	0.22	0.20	0.20	0.25
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Chapter 7

Implied Kernel Models

7.1 Introduction

In Chapter 2 we described the state of the art of modern fixed income modelling. We saw that the market models are very popular among practitioners, as they allow almost instantaneous calibration to liquid market prices. The main disadvantage of market models turned out to be their high-dimensional Markov structure. This inhibits the use of these models for pricing exotic products, where the short rate approach is still the preferred choice. Furthermore, extensions of the market models to account for volatility smiles and skews are quite complex.

At this point we ask what we might expect from a good term structure model. We want it to produce explicit solutions for interest rate derivatives. Most of the interest rate derivatives in a portfolio will have quite distinct structures. To obtain explicit solutions for even a small subset of derivatives we will need to make very strong modelling assumptions at the expense of the other derivatives. It is questionable whether we want to sacrifice prudent modelling assumptions to achieve this.

Do we want exact calibration to the prices of all traded derivatives? All derivatives are traded with bid-ask spread. So the price is not uniquely defined. Furthermore,

the prices are not updated continuously and simultaneously. Calibrating the model exactly to the quoted prices will mean calibrating the model to old and not synchronised information. So, it is again questionable whether it is sensible to calibrate the model exactly. However, one might expect that the actual prices will not be too far from the observed ones, so the prices should be reasonably well approximated by the model.

Do we want the model to produce simple numerical solutions? The answer is yes. This is an important property. In a trading environment one needs the price information very fast. So, one expects a model to deliver a price within seconds, even for exotic-type products.

Thus there is a need for finding a class of models that enjoys the same ease of calibration to liquid market prices as the market models, and at the same time has a low-dimensional Markov structure as in the case of short rate models. It should be flexible in fitting volatility smiles and skews. Furthermore, it should produce simple numerical solutions for most of the interest rate derivatives.

To achieve this goal we have chosen the pricing kernel framework, i.e. framework in which the bond price can be represented as

$$B(t, T) = \mathbb{E}^{\mathbb{P}^*} \left[\frac{K_T}{K_t} | \mathcal{F}_t \right], \quad (7.1.1)$$

where K_t is positive pricing kernel and \mathbb{P}^* some probability measure equivalent to the objective \mathbb{P} . In Chapter 2 we described this approach in details and examined its relation with the HJM and short rate frameworks. We found that the latter have been studied heavily and it is unlikely they will yield any class of models with desired properties. On the other hand the pricing kernel framework has been studied far less. We feel that it has enough scope and flexibility to achieve our goal. Only a few authors have exploited this framework directly. The idea of using this framework in arbitrage pricing was pioneered by Constantinides (1992). It was then pursued

by Flesaker and Hughston (1996a), Rogers (1997b), Balland and Hughston (2000), and Hunt *et al* (2000). We describe their ideas in Section 7.2.

Though we model the pricing kernel, we link its construction explicitly with the calibrating set of instruments. Thus, once the kernel is constructed it prices correctly the chosen set of instruments and has a low-dimensional Markov structure. In particular, we assume that a pricing kernel K_t can be approximated by a series of functionals of some underlying Markov process, i.e.,

$$K_t \equiv \sum_i a_i(t) f_i(t, X_t).$$

The coefficients of this approximation are then implied from the set of liquid market instruments, such as bond, caps, swaptions, etc. We call models of this type Implied Pricing Kernel (IPK) models.

The outline of this chapter is as follows. In Section 7.2, we motivate our modelling framework and describe its relation to other pricing kernel models. We also briefly describe the literature on modelling non-flat implied volatilities. In Section 7.3, we introduce several families of approximating functions that can be used to approximate the pricing kernel. We also discuss a general Gaussian diffusion process that we use in the implementation of IPK models. Furthermore, we present pricing formulae for several fixed income instruments within the pricing kernel approach. We discuss implementation issues and conduct several calibrating studies in Section 7.4. We conclude in Section 7.5.

7.2 Related Research

In this section we motivate our class of models and describe its relation with the existing pricing kernel literature. As our approach can also handle non-flat volatilities, we also discuss literature relating to this topic.

7.2.1 Kernel Models

To define a pricing kernel model one needs two ingredients: a model for the underlying noise in the economy and a functional form relating that noise to the pricing kernel. The underlying noise in the economy is usually modelled by a simple stochastic process, such as the Ornstein-Uhlenbeck process. Then the pricing kernel is defined to be some strictly positive function of the process.

For example, Constantinides (1992) chooses the kernel,

$$K_t = \exp \left(-\alpha t + \sum_{i=1}^N (X_{i,t} - \alpha_i)^2 \right)$$

where the stochastic process is a sum of squares of displaced Ornstein-Uhlenbeck processes $X_{i,t}$, and the function relating the process to the kernel is simply the exponential. Das and Foresi (1996) choose as the driving process the sum of two components, the Ornstein-Uhlenbeck process X_t with stochastic mean, and a pure jump process $y(t)$. The function is again exponential,

$$K(t) = \exp(-y_t - X_t).$$

Flesaker and Hughston (1996a) in their *rational log-normal* model use a kernel that can be written

$$K_t = f(t) + \exp(c_t + X_t),$$

where f_t and c_t are deterministic functions. In this kernel the underlying process is a simple Gaussian diffusion and the kernel is again defined as the exponential of this diffusion plus some deterministic functions. Rogers (1997b) considers kernels of the form

$$K_t = \frac{e^{-\alpha t} R_\alpha g(X_t)}{R_\alpha g(X_0)},$$

which is a positive supermartingale, i.e. he defines a positive supermartingale of the resolvent of some process X_t with function $\exp(-\alpha t)$. As examples of his framework,

among others, he considers the exponential-linear kernel

$$K_t = \exp(-\alpha t + a \cdot X_t),$$

the exponential-quadratic

$$K_t = \exp(-\alpha t + \frac{1}{2}(X_t - c)^T Q(X_t - c)),$$

and the quadratic kernel,

$$K_t = \gamma + \frac{1}{2}(X_t - c)^T Q(X_t - c).$$

In all above examples the functional form was a simple positive function such as an exponential or quadratic. An opposite point of view has been taken by Hunt¹ *et al* (2000). They develop a class of pricing kernel models, the Markov-functional interest rate models (MFM), in which the positive functional is constructed from the market price information. As the underlying noise in the economy they choose a simple Gaussian diffusion.

For example, in their LIBOR Markov-functional model they choose the reciprocal of a bond of fixed maturity as a pricing kernel, $K_t \equiv 1/B(t, T_{n+1})$. The market price information, in this case, is given by a set of cap prices. More precisely, they take as given the set of caplet prices for maturities T_i , $i = 1, \dots, n$, and for all possible strikes. They choose as the underlying Markov process Gaussian diffusion of the form

$$dX_t = \sigma_t^n dW_t,$$

where $\sigma_t^n = \sigma \exp(at)$. To specify the model completely they use a backward induction. To start the induction they assumed that the last forward LIBOR rate follows a log-normal diffusion² $dL_t^n = \sigma_t^n L_t^n dW_t$, under the forward measure \mathbb{P}_{n+1} , corresponding to the numéraire $B(t, T_{n+1})$. Thus, the distribution of $B(T_n, T_{n+1})$ can

¹See, also, Hunt *et al* (1996) and Hunt and Kennedy(1998b).

²This assumption is not necessary; we only need to assume that the functional form of the forward LIBOR rate with respect to some underlying Markov process is explicitly known (and is monotone function of the process).

be recovered from its dependence on $L_{T_n}^n$. To determine the model completely one needs to find functional forms of the kernel $K(X_{T_i})$ for the dates $T_i, i = 1, \dots, n-1$. These functional forms can be found by inverting the digital caplet prices into the bond price $B(T_i, T_{n+1})$.

To summarise, the main idea of Markov functional models is that the distribution of a pricing kernel can be recovered from the prices of digital options via an inversion procedure. Thus, by construction, the LIBOR Markov-functional model prices exactly the digital caplets for a given set of maturities and a continuum of strikes, and consequently the initial term structure. However, the main disadvantage of this method is that it imposes a rather rigid distribution on the kernel which is needed to price the digital caplets correctly. Furthermore, the method is rather difficult to generalise to higher dimensions. We have implemented elsewhere a multidimensional version of the Markov functional model and have found that the results are quite similar to the one dimensional case. Thus, it is not clear if this method provides any more generality when working in higher dimensions.

It is worth mentioning a parallel paper by Balland and Hughston (2000). They develop a lattice type model based on exactly the same idea as the MFM framework. However, the implementation of this lattice model is rather cumbersome. They make use of the change of numéraire technique in every time step, which makes the method very tedious to implement. At the heart of the method, as with MFM, is the relation of the digital option prices with the functional form of the kernel.

The pricing kernel models described above can be thought of as belonging to two classes: parametric and non-parametric. Models belonging to the parametric class are pricing kernel which are defined as simple positive functional forms, such as an exponential or quadratic, of some Markov process. In the models belonging to the non-parametric class the functional form of the kernel is implied non-parametrically from the data.

In this chapter we take a “middle way” point of view. We assume that the pricing kernels can be approximated by a series of functions of some Markov process. The actual coefficients and parameters of the series are then implied from the market prices, such as bonds, caps, swaptions, etc. We refer to this class of models as Implied Pricing Kernel models (IPK). This class of models presents several advantages over previous pricing kernel models. It has low-dimensional Markov structure. It is just as simple in a one- as in a multi-dimensional framework. It is flexible in fitting volatility skews. Furthermore, it does not require fitting continuum of price information to construct its functional form. Moreover, this method allows the analytical pricing of certain instruments, as in the case of parametric kernel models.

7.2.2 Modelling Non-Flat Volatilities

When plotting implied volatilities of the caplet prices observed in the interest rate market³ against strikes and maturities, one usually observes non-flat surfaces. This means that the distribution of the forward LIBOR rates is not log-normal as implied by the Black formula. To account for this phenomenon several modelling approaches have been suggested. They range from simple parametric to fully non-parametric methods. We briefly describe them in this section.

The first approach is based on an extension of the standard log-normal LIBOR market model by assuming alternative dynamics for the forward-rate process that lead to volatility smiles or skews. For example, Andersen and Andreasen (2000), used the Cox (1975) CEV process for the forward rate. Zühldorf (2000) considered affine and quadratic volatility functions for the instantaneous volatility function of the forward rate diffusions. All these extensions exhibit quite flexible volatility skews. However, these extensions do not allow for the calibration of the whole volatility surface.

³The same phenomenon occurs in the equity and foreign-exchange markets as well.

The class of models for the dynamics of forward LIBOR rates introduced recently by Brigo and Mercurio (2000), and Brigo *et al* (2002), can be considered as a semi-parametric approach. It is based on the assumption that the forward rate density is given by a mixture of known basic densities. Simple log-normal dynamics as well as mixture dynamics can be generalised by a displace-diffusion technique. This involves shifting a process by a constant. Both, the mixture-diffusion and displace-diffusion techniques enable better, albeit not exact, fit to the implied volatilities.

Yet another way of modelling non-flat volatilities has recently been considered by Joshi and Rebonato (2001). They consider a stochastic volatility extension of the standard LIBOR market model. In particular, they assume a certain parametric form for the instantaneous forward volatility function. The parameters of this volatility function are assumed to follow certain stochastic processes. By a clever choice of implementation procedures they achieve very good results both in fitting the implied volatilities (though not exact), and in rapid pricing of path-dependent options.

Finally, the approach based on the assumption of continuum of traded strikes can be considered as a fully parametric. This goes back to Breeden and Litzenberger (1978), and has been recently applied⁴ in an interest rate framework by Kuan and Weber (1998). This method is commonly referred to as implied pricing. The main problems of this method are numerical instability and the need for interpolation between option prices⁵ between consecutive strikes. The models described above by Hunt *et al* (2000) and Balland and Hughston (2000) can be considered as examples of this approach as well.

In this section we have described several approaches to model non-flat volatilities. They range from parametric to fully non-parametric methods. The characteristic

⁴See also Amin and Ng (1997), and Coutant *et al* (2001).

⁵Alternatively, the interpolation can be performed on Arrow-Debreu prices, the density function, or local drift and volatility functions.

feature of parametric methods is that they are rather limited in their flexibility to fit shapes of the implied volatility structures whereas non-parametric methods fit any implied volatility exactly. As we have noted in the introduction the exact fit is probably not desirable, although a reasonably good fit may be expected from a good term structure model. In this light the semi-parametric model yield the best result. In the next section we describe our implied pricing kernel class of models which, can be considered as a semi-parametric approach.

7.3 Modelling Framework

Assume that the unknown kernel is of the form

$$K_t = f(t, X_t), \quad (7.3.1)$$

where X_t is some Markov process and f_t is a strictly positive continuous function on \mathbb{R}^d for every t . Furthermore, we take the view that the pricing kernel (7.3.1) can be approximated by a kernel of the form,

$$K_t = \sum_i a_i(t) f_i(t, X_t), \quad (7.3.2)$$

where X_t is some Markov process and f_i a family of strictly positive continuous functions.

Next, we describe the underlying process we used in the implementation of the model, and several function families that can be used in approximating kernel (7.3.2).

7.3.1 The Underlying Markov Process

For numerical implementation of the models of the form (7.3.2) we choose a d -dimensional linear stochastic differential equation of the form

$$dx_t = (A(t)x_t + a(t))dt + B(t)dW_t, \quad x_{t_0} = c \quad (7.3.3)$$

where $A(t)$ is a $d \times d$ -matrix valued function, $a(t)$ is \mathbb{R}^d -valued function, and $B(t)$ is $d \times m$ -matrix valued. This process is very tractable and has been a popular choice in interest rate modelling, especially within the short rate framework. Examples are Vasicek (1977), Langetieg (1980), El Karoui and Lacoste (1995).

The solution of the linear SDE (7.3.3) is a *Gaussian* stochastic process x_t , with mean value

$$m_t = \mathbb{E}x_t = \Phi(t) \left(c + \int_{t_0}^t \Phi(u)^{-1} a(s) ds \right) \quad (7.3.4)$$

and covariance matrix

$$\begin{aligned} K(s, t) &= \mathbb{E}(x_s - \mathbb{E}x_s)(x_t - \mathbb{E}x_t)' \\ &= \Phi(s) \left(\int_{t_0}^{\min(s, t)} \Phi(u)^{-1} B(u) B(u)' (\Phi(u)^{-1})' du \right) \Phi(t)', \end{aligned}$$

where the matrix $\Phi(t) = \Phi(t, t_0)$ is the matrix of solutions of the homogeneous equation⁶

$$\dot{x}_t = A(t)x_t,$$

with the unit vectors $c = e_i$ in the x_i -direction as initial value. If, for example, $A(t) \equiv A$ is independent of t , then

$$\Phi(t) = \exp(A(t - t_0)) = \sum_0^{\infty} A^n (t - t_0)^n / n!.$$

Furthermore, we assume that the matrices A and B , and the vector a , are independent of t . In this case, to obtain the moments of (7.3.3), we need to evaluate the integral of the matrix exponentials in (7.3.4) and (7.3.5). We employ the method of diagonal Padé approximation with scaling and squaring as described in Van Loan (1978). See Appendix 7.6.3 for the details of the algorithm.

Most of the integration in the numerical implementation of the model in this chapter has been done with *Gauss-Hermite* quadrature. This type of quadrature is particularly useful when we consider stochastic processes with Gaussian distributions, as

⁶Further details on SDE's of the form (7.3.3) can be found in Arnold (1974).

they approximate integrals of the type

$$\int_{-\infty}^{\infty} F(x)e^{-x^2} dx,$$

and the Gaussian density can be used as the weight function. This method can be generalised to integrals in higher dimensions, see Appendix 7.6.4.

7.3.2 Approximation by Radial and Ridge Functions

In this section we describe two classes of functions which can be used to approximate a pricing kernel. The member functions of both classes are strictly positive. Thus, if we choose only positive coefficients in the approximation then our model will be arbitrage-free. Note, both classes of functions are suitable for multivariate approximation. This allows a simple extension of the IPK models from a one-factor to a multi-factor framework.

The first class of functions is that of positive functions which have radial symmetry. A real-valued function F on an inner-product space is said to be *radial* if $F(x) = F(y)$ whenever $\|x\| = \|y\|$. If this property is present, the value of $F(x)$ depends only on $\|x\|$, and consequently there exists another function $f : [0, \infty) \rightarrow \mathbb{R}$ such that

$$F(x) = f(\|x\|^2).$$

This type of function is referred to in Approximation Theory as a *radial basis function*.

Certain classes of radial basis functions have nice interpolating and approximating properties. Specifically, it can be shown that for, say, a radial function f , for each compact Q in \mathbb{R}^n , the set

$$\{x \rightarrow f(x - y) : y \in Q\}$$

is fundamental in $C(Q)$, where $C(Q)$ is the set of all continuous functions. Put in another way, for each $g \in C(Q)$ and for each $\epsilon > 0$, there is a linear combination $\sum_{y \in Q} c_j f(x - y)$, so that $\|g(x) - \sum_{y \in Q} c_j f(x - y)\| < \epsilon$. (The sum is finite.)

For example, if f is completely monotone⁷ but not constant on $[0, \infty)$, and Q is a compact subset in \mathbb{R}^n , then the set of functions

$$\{x \rightarrow f(\|x - y\|) : y \in Q\}$$

is fundamental in $C(Q)$.

This result provides a rich source of functions that are suitable for approximation of data in the Euclidean spaces $\mathbb{R}^1, \mathbb{R}^2, \dots$. The following functions satisfy the specified conditions

1. $f(t) = (t + 1)^{-1}$,
2. $f(t) = e^{-t}$,
3. $f(t) = (t + 1)^{-1/2}$.

In Figures 7.1 and 7.2 we plot two examples of functions of this class. Note the typical bell-shaped form. These functions can be used to interpolate or approximate arbitrary data by functions of the form,

1. $F(t) = \sum_{j=1}^n \frac{c_j}{\sqrt{1 + \|x - x_j\|^2}}$,
2. $F(t) = \sum_{j=1}^n c_j e^{-\|x - x_j\|^2}$,
3. $F(t) = \sum_{j=1}^n \frac{c_j}{1 + \|x - x_j\|^2}$.

The following belong to another type of radial function, that are not included in the above class, but have similar approximating properties:

⁷This class of function is due to Schoenberg (1938). In particular, he showed, that the function $x \rightarrow f(\|x\|^2)$ is a radial strictly positive definite function on any inner-product space. I.e., for any n distinct points x_1, x_2, \dots, x_n in such a space the matrix $A_{ij} = f(\|x_i - x_j\|^2)$ is positive definite.

1. $f(t) = \sqrt{t}$,
2. $f(t) = \sqrt{t+1}$,
3. $f(t) = \log(t+1)$.

This type of function is quite different from the radial positive definite functions. Figures 7.3 and 7.4 present two examples. These functions⁸ belong to the class with the properties that f' is completely monotone but not constant on $(0, \infty)$, $f(0) \geq 0$, and that $f(\infty) = 0$. In Appendix 7.6.2, Lemma 7.6.1, we show that under certain assumptions radial basis functions can approximate in distribution a pricing kernel.

The second class of functions that we suggest for the approximation of the pricing kernel is the class of so-called *ridge* functions. As the radial basis function, the ridge functions have been used in multivariate approximation theory. We denote a normed linear space by X , and by X^* the space of all continuous linear functionals on X . A function $f : X \rightarrow \mathbb{R}$ is called *ridge function* if it can be represented in the form $f = g \circ \phi$, where $g : \mathbb{R} \rightarrow \mathbb{R}$ and $\phi \in X^*$. Every continuous linear functional on \mathbb{R}^s is of the form

$$\phi(x) = \alpha_1 \zeta_1 + \alpha_2 \zeta_2 + \cdots + \alpha_s \zeta_s.$$

A ridge function on \mathbb{R}^s is then a function of the form

$$f(x) = g(\alpha_1 \zeta_1 + \alpha_2 \zeta_2 + \cdots + \alpha_s \zeta_s),$$

where $x = \zeta_1, \zeta_2, \dots, \zeta_s$.

Figure 7.5 shows the ridge function $z = p(x-y)$, where p is a polynomial of degree 4. A single ridge function is very limited in its capacity to approximate an arbitrary continuous function on X . In particular, the graph of a ridge function is a ruled

⁸This class of functions has been described by Micchelli (1986). Functions from this class have the property that for any n distinct points x_1, x_2, \dots, x_n the matrix $A_{ij} = f(\|x_i - x_j\|^2)$ is non-singular.

surface. For approximation purposes one considers a linear combination of ridge functions of the form

$$f = \sum_{i=1}^m c_i g_i \circ \phi_i \quad (g_i \in C(\mathbb{R}), \phi_i \in X^*).$$

The coefficients c_i are unnecessary in this representation, since they can be absorbed by the functions g_i . It is important to note that not all continuous functions on X are linear combinations of ridge functions. But every continuous function on X can be well approximated by such linear combinations.

For some applications, such as neural networks⁹, it is very desirable to employ a single function g in the ridge functions, which is usually called a *sigmoid function* and denoted σ , and has the property that

$$\lim_{t \rightarrow \infty} \sigma(t) = 1 \text{ and } \lim_{t \rightarrow -\infty} \sigma(t) = 0.$$

The following functions are examples of sigmoids:

1. The logistic sigmoid $f(x) = \frac{1}{1 + \exp(-ax)}$.
2. The Heaviside function $f(x) = 1$ if $x \geq 0$, $f(x) = 0$ if $x < 0$.

A typical example of ridge function is plotted in Figure 7.5, and sigmoid ridge functions in one and two dimensions are shown in Figure 7.6 and 7.7.

In this section we have described two classes of positive functions, the radial basis functions and the ridge functions, that can be used to approximate a pricing kernel. In Section 7.4 we will implement an implied pricing kernel model based on the Gaussian radial basis function e^{-t} . I.e., we will approximate the pricing kernel by a linear combination of the form $F(t) = \sum_{j=1}^n c_j e^{-\|x-x_j\|^2}$.

⁹This type of function have been used by Bansal and Viswanathan (1993), and Bansal *et al* (1993) used the logistic sigmoid, to approximate the pricing kernel within the asset pricing framework.

7.3.3 Formulae for Pricing Zero-Coupon Bonds, Caps, and Swaptions

In this section we derive expressions for pricing formulae for zero-coupon bonds, caplets and swaptions. These expressions will be used in the calibration studies of the model in the next section. Assuming that the pricing formula has the form (7.3.2), the price of a zero coupon bond can be expressed as

$$P(t, T) = \frac{\mathbb{E}[K_T | \mathcal{F}_t]}{K_t} = \frac{\sum_i^N a_i(T) \tilde{\mathbb{E}}[f_i(T, X_T) | \mathcal{F}_t]}{\sum_i a_i(t) f_i(t, X_t)},$$

where $\mathbb{E}[\cdot | \mathcal{F}_t]$, expectation operator conditional on the sigma algebra \mathcal{F}_t , taken with respect to a reference measure, the equivalent to the risk neutral measure. The value of a caplet at time t with maturity date T_{n-1} and payoff at the date T_n , follows from observing that $P(T_{n-1}, T_n) = (1 + L(T_{n-1})\delta)^{-1}$, where $L(T_{n-1})$ is the δ -period LIBOR rate,

$$\begin{aligned} \text{Cpl}(t, T_{n-1}, T_n) &= \mathbb{E} \left[\frac{K_{T_{n-1}}}{K_t} P(T_{n-1}, T_n) \delta (L(T_{n-1}) - K)^+ | \mathcal{F}_t \right] \\ &= \mathbb{E} \left[\frac{K_{T_{n-1}}}{K_t} (1 - \tilde{\delta} P(T_{n-1}, T_n))^+ | \mathcal{F}_t \right] \end{aligned} \quad (7.3.5)$$

where $\tilde{\delta} = 1 + \delta K$, with K the strike of the caplet and δ the accrual period of the underlying LIBOR rate.

Similarly, the price of a (payer) swaption at time t with maturity date T_0 and last payment at time T_N , with the length of the accrual period δ , is given by

$$\text{Swp}(t, T_0, T_N) = \mathbb{E} \left[\frac{K_{T_0}}{K_t} D(T_0, T_N) (\kappa(T_0, T_N) - K)^+ | \mathcal{F}_t \right],$$

where K is the strike, κ is the swap rate defined as

$$\kappa(T_0, T_N) = \frac{(1 - P(T_0, T_N))}{\delta \sum_{i=1}^N P(T_0, T_i)},$$

and $D(T_0, T_N) = \delta \sum_{i=1}^N P(T_0, T_i)$.

For our numerical study of the model in the next section we are going to use one particular family of radial functions.

$$F(t) = \sum_{j=1}^n a_j(t) e^{-\|x - b_j(t)\|^2}$$

This kernel allows analytical solution for the zero-coupon bond price,

$$P(t, T) = \frac{\sum_{j=1}^n a_j(T) e^{-(m_{Tt} - b_j(T))^T (I + 2V_{Tt})^{-1} (m_{Tt} - b_j(T))}}{\sqrt{\det(I + 2V_{Tt})} \sum_{j=1}^n a_j(t) e^{-(X_t - b_j(t))^T (X_t - b_j(t))}},$$

where¹⁰ m_{Tt} and V_{Tt} are conditional mean and variance of the process X_T conditioned on the process at time t .

7.4 Implementation and Numerical Study of the Model

In this section we calibrate the implied kernel model to the data and study its properties. We choose the model based on the Gaussian radial functions and calibrate it to three sets of data: the yield curve and at-the-money caps for the GB pound on February 3, 1995, the yield curve and the caplet black implied volatility surface for the GB pound on August 4, 2000, and the yield curve and at-the-money black implied swaption matrix for 4th August, 2000. For all implementations we use only three basis functions.

7.4.1 Yield Curve and ATM Caps

In the first study of the model, we investigate whether our implied pricing kernel model was able to fit zero-coupon yield curve and ATM caps. To calibrate the models we used data from the UK market on February, 3, 1995, which we obtained from Brace *et al* (1997). The zero coupon discount curve is given in Table 7.1. To obtain the yield and forward rate on other dates than in Table 7.1. we used spline approximation techniques, described in Dierckx (1993), and in the accompanying

¹⁰This expression for the bond price can be easily confirmed by completing the square in the Gaussian density.

Fortran software package *FITPACK*. The approximations by smoothing splines led to oscillations in the forward curve. The best results, were achieved by using the weighted least-squares spline with a judiciously chosen set of knots. The plots of the resulting forward rate and yield curves are in Figure 7.8, and the corresponding values are in Table 7.2.

To calibrate the model to the ATM cap prices we used a similar algorithm to the one used in calibration of the market models. I.e., we recovered the black implied forward-forward volatilities. This gave us prices of individual caplets comprising the caps. The details of the recovery of the forward forward volatilities are described in Appendix 7.6.1. We plotted the forward forward volatilities in Figure 7.9. As we see, it has the typical hump shape around 2 years to maturity and is monotonic elsewhere. We then calibrated our model to these caplet prices and subsequently to the quoted ATM cap prices.

The calibration results of our model are presented in Tables 7.3 and 7.4. The fit to the yield curve is exact. The fit to cap prices is quite good. The largest error is just below 3bp. Measured in percentage terms most of the errors are less than 0.5 percent with the exception of the 3-year cap, which has an error of 1.5 percent. For comparison reasons we also calculated the swaption prices implied by our calibrated model. Then we compared the results with the swaption market prices. We presented the results in Table 7.5. Note that we did not try to calibrate the model to these swaption prices. As can be expected the errors are quite large.

Most of the swaption prices are higher than the market prices. This is in line with other findings in the literature. For example, de Jong *et al* (2001) analyse the LIBOR market models on the US caplet and swaption data. They find that swaptions are overpriced with the average absolute pricing error around 1 volatility point. Similar results have been found by Driessen *et al* (2000) who investigate the performance of HJM multi-factor models on the US data.

We also calculated the implied black swaption volatility matrix resulting from the model's calibration to the ATM caps. This matrix can be seen in Table 7.6. In most of the cases the contracts change smoothly from maturity to maturity, and between different underlying swap lengths.

7.4.2 The Yield Curve and Caplet Volatility Surface

In this study we tested whether the implied pricing kernel model could fit the yield curve and caplet volatility surface. The dataset was the yield curve and the caplet black implied volatility surface for the GB pound on August 4, 2000. In particular, we fitted the yield curve with maturities from 2 to 10 years with quarterly steps, and caplet prices with the same maturities and strikes ranging from 2% to 10%, with 0.01% step size. As in the previous exercise, the forward rate curve implied by the bond data was not smooth and exhibited oscillating behaviour. To overcome this, we applied some smoothing. We have plotted the resulting yield and forward curves in Figure 7.10.

We have presented the zero coupon bond prices together with model bond prices and percentage errors in Table 7.7. The fit is not perfect; but most of the errors are below 0.1%. In Table 7.8, we present the results of the fit to the caplet prices. The first line in each block presents the market prices, and the second the model prices. The third and forth lines present errors expressed in percentage terms and basis points.

The quality of fit of the caplet surface is very good for the longer maturities. In most cases it is less than 1%. However, for maturities less than 5 years and high strikes the quality decreases significantly. The reason for, we believe, lies in the low precision of the implementation programme. We plotted the market caplet prices in Figure 7.12, and the model caplet prices in Figure 7.13. We have also plotted market and model caplet prices together for selected maturities of 2.5, 5, 7.5, 10

years to maturity in Figures 7.18 to 7.21.

7.4.3 The Yield Curve and ATM Swaptions

In this exercise we tested whether the implied pricing kernel model could fit the yield curve and ATM swaption prices. The dataset was again GB pounds on August 4, 2000, comprising the zero coupon bond prices and ATM swaptions. We have presented the zero coupon bond together with the model prices and the percentage error in Table 7.10. As we see the model bond prices fit the market prices almost exactly. The Table 7.11 presents the market and model swaption prices for several maturities and lengths of the underlying swap contracts together with percentage errors of the fit. In many cases the fit is quite good, but in some cases the error is quite large. This is not unexpected, as our model is only one-factor. We would need higher-factor models to achieve a better fit.

7.4.4 Properties of the Model

We also investigated distributional properties of the implied kernel model calibrated to the yield curve and the caplet black implied volatility surface for GB pound on August 4, 2000. We simulated the 3-month LIBOR rates implied by our model using Monte-Carlo technique. In Table 7.9 we present the descriptive statistics for the LIBOR rates for maturities between 2 and 10 years, resulting from our simulation, including the mean, standard deviation, skewness and kurtosis. Observe the presence of skewness and kurtosis which grows for longer maturities.

We have also plotted in Figures 7.14 to 7.17, the histograms of the LIBOR rates with maturities of 2.5, 5, 7.5, and 10 years. In each plot we have also added the density functions of normal and log-normal distributions with the same moments as the simulated distribution of the LIBOR rates. Note the IPK model produced positive interest rates though, it was not constrained to do so.

7.5 Conclusions

In this chapter we have introduced a class of pricing kernel models, the Implied Pricing Kernel (IPK) models, for the term structure of interest rates. This class is mainly intended for pricing interest rate derivatives. The main motivation for its development has been the goal of finding an alternative to the market models. This alternative should combine the advantages of the market models such as the ease of calibration to liquid market prices, with the advantages of the short rate models, such as low-dimensional Markov structure. We also wanted this class to be flexible in fitting other price information such as skews etc.

Unlike previous literature which has utilised a pricing kernel approach, we have approximated the pricing kernel by a functional series. Thus, implicitly, we have acknowledged that we do not know its correct functional form. However, we believe that the pricing kernel can be approximated sufficiently closely by a judiciously chosen functional series and the underlying process. We have suggested two types of approximating series: the first is based on the radial basis functions, and the second on the ridge functions. The basis functions of both types are strictly positive. Thus, by choosing only non-negative coefficients in the series we have achieved a strictly positive pricing kernel which guarantees absence of arbitrage in the model. Furthermore, the coefficients and parameters of the series have been chosen so that the model fits the market price information, such as zero coupon bond prices, caps, caplets, and swaptions. As the underlying noise in the economy we chose a simple multi-factor Gaussian diffusion. Our approach is just as simple in a one-factor as in a multi-factor setting. In a summary, the class of implied kernel models is Markov by construction, calibration for a small number of factors is relatively easy, it can deal with American type options, and is flexible in fitting to skews.

The IPK class models have an advantage over the models with a fixed pricing kernel designed by Constantinides (1992), or Flesaker and Hughston (1996a). They are

more flexible in fitting the market price information. It also has an advantage over the non-parametric type of kernel models designed by Hunt *et al* (2000), and Balland and Hughston (2000). This non-parametric approach relies on inferring the kernel from the digital caplets or swaption prices. For a given maturity such kernels can only fit one type of instrument only, e.g. a cap price or possibly a set of caplets with different strikes. However, they are incapable of dealing with say a set of caplets and swaptions at the same time, or swaptions with different lengths of underlying swaps. The possibility of extending that type of model to a multi-factor setting is questionable. The IPK models can fit simultaneously to any market price information available. The quality of the fit can be improved by adding more basis functions in the approximating series. Moreover, the extension of the underlying noise to a multi-factor structure is trivial.

We conducted several model calibration studies for one-factor IPK models. This comprised of calibrations to the yield curve and at the money cap prices, the yield curve and caplet implied volatility surface, and the yield curve and swaption data. Overall, we achieved a reasonably good quality of fit. We also studied numerically the distributional properties of the forward LIBOR rates implied by a calibrated IPK model. These studies are, of course, only preliminary. More extensive tests of this class of models is needed to assess its full advantages and shortcomings.

There are several questions that can be addressed by further research. We have suggested two classes of strictly positive approximating functions. It is not clear which class of functions and which members of these classes are better suited for approximation of the pricing kernel. We have assumed a Gaussian diffusion as the underlying noise structure. Other processes may be more appropriate in this case. On the implementation side, better evaluation and calibrating techniques may be needed, especially for multi-factor IPK models. We suggest further research directions for the pricing kernel model in concluding Chapter 8.

7.6 Appendices

7.6.1 Recovering Forward-Forward Volatilities

The market typically quotes volatilities for caps with first reset date either in three months (T_0 equal to three months, $\alpha = 0$ and all other T 's equally spaced three months apart) or in six months (T_0 equal to six months, $\alpha = 0$ and all other T 's equally spaced six month apart), and progressively increasing maturities. In our case UK market caps are quarterly while swaptions are semiannual. We set $\mathcal{T}_j = [T_0, \dots, T_j]$ for all j . An equation is considered between the market price $\text{Cap}^{MKT}(0, \mathcal{T}_j, K)$ of the cap with $\alpha = 0$ and $\beta = j$, and the sum of the first j caplet prices,

$$\text{Cap}^{MKT}(0, \mathcal{T}_j, K) = \sum_{i=1}^j \tau_i P(0, T_i) \text{Bl}(K, F_i(0), \sqrt{T_{i-1}} v_{T_j-\text{cap}}) \quad (7.6.1)$$

where a same average-volatility value $v_{T_j-\text{cap}}$ has been substituted in all caplets up to j . The quantities $v_{T_j-\text{cap}}$ are sometimes called *forward volatility*. The market solves the above equation in $v_{T_j-\text{cap}}$ and quotes $v_{T_j-\text{cap}}$ annualised and in percentages.

One can obtain another representation for the value of the cap by using

$$\begin{aligned} \sum_{i=1}^j \tau_i P(0, T_i) \text{Bl}(K, F_i(0), \sqrt{T_{i-1}} v_{T_j-\text{cap}}) = \\ \sum_{i=1}^j \tau_i P(0, T_i) \text{Bl}(K, F_i(0), \sqrt{T_{i-1}} v_{T_j-\text{caplet}}). \end{aligned} \quad (7.6.2)$$

The quantities $v_{T_j-\text{caplet}}$ are called sometime *forward-forward volatilities*. Notice that *different* average volatilities $v_{T_j-\text{caplet}}$ are assumed for different caplets occurring to the T_j -maturity cap.

We used a stripping algorithm to recover the $v_{T_j-\text{caplet}}$'s from the market quoted $v_{T_j-\text{cap}}$'s based on the last equality applied to $j = 1, \dots, 5, 7, 10$. We parametrised the forward forward volatility curve by the family of curves described by Nelson and

Siegel (1987). The curve $f(t)$ is of the form

$$f(t) = \beta_0 + (\beta_1 + \beta_2 t) \exp(-kt)$$

with four parameters $(\beta_0, \beta_1, \beta_2, k)$. We minimised the criterion

$$\sqrt{\frac{1}{n} \sum^n (\text{Cap}^{MKT} - \text{Cap}^{Model})^2},$$

which resulted in the parameter vector $(\beta_0, \beta_1, \beta_2, k) = (0.124, -0.0167, 0.105, 0.553)$ and $MSE = 0.00015$.

7.6.2 Proof of Lemma 7.6.1

Lemma 7.6.1. Let $X \in \mathbb{R}^d$ be a random variable such that its push forward measure \mathbb{P}_X is a measure with density g relative to the λ^d -Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}^d)$. Furthermore, let $\{f_n\}$, be a sequence of real, measurable function on \mathbb{R}^d converging point-wise to a measurable function f . Then the sequence of random variables $f_n(X)$ converges to $f(X)$ in distribution.

Proof. To prove convergence in distribution we need to show only that the characteristic function of $f_n(X)$, $\hat{f}_n(t)$, converges point-wise to the characteristic function of $f(X)$, $\hat{f}(t)$.

$$\lim_{n \rightarrow \infty} \hat{f}_n(t) = \lim_{n \rightarrow \infty} \int e^{-ity} \mathbb{P}_{f_n(X)}(dy) = \lim_{n \rightarrow \infty} \int e^{-itf_n(x)} \mathbb{P}_X(dx),$$

From the point-wise convergence of f_n to f follows point-wise convergence of e^{-itf_n} to e^{-itf} . Furthermore, from

$$|e^{-itf_n} g(x)| \leq g(x)$$

it follows that the integrand is dominated by a function $g \in L^1(\mathcal{B}^d, \lambda^d)$. The interchange of limit and integration is justified by Lebesgue's dominated convergence theorem. (See Rudin (1976), Theorem 11.32.) \square

7.6.3 Computing Integrals of Matrix Exponentials

Here we sketch the algorithm used in this chapter to calculate the integral of the matrix exponential. For a full discussion of the method and the quality of the approximation see Van Loan (1978).

Let A , B , and Q_c be real matrices of dimensions $n \times n$, $n \times p$, and $n \times n$, respectively. Assume that Q_c is symmetric ($Q_c^T = Q_c$) and positive definite ($x^T Q_c x \geq 0$). The methodology involves calculating the following integrals

$$\begin{aligned} H(\Delta) &= \int_0^\Delta e^{As} B ds, \\ Q(\Delta) &= \int_0^\Delta e^{A^T s} Q_c e^{As} ds, \\ M(\Delta) &= \int_0^\Delta e^{A^T s} Q_s H(s) ds, \\ W(\Delta) &= \int_0^\Delta H^T(s) Q_c H(s) ds. \end{aligned} \tag{7.6.3}$$

The method for calculating these integrals involves computing the exponential of a certain block triangular matrix and combining various sub-matrices of the result to obtain (7.6.3)-(7.6.3). In particular, if we apply Theorem 1. in Van Loan (1978) to

$$\hat{C} = \begin{bmatrix} -A^T & I & 0 & 0 \\ 0 & -A^T & Q_c & 0 \\ 0 & 0 & A & B \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

we find

$$e^{\hat{C}} = \begin{bmatrix} \hat{F}_1(t) & \hat{G}_1(t) & \hat{H}_1(t) & \hat{K}_1(t) \\ 0 & \hat{F}_2(t) & \hat{G}_2(t) & \hat{H}_2(t) \\ 0 & 0 & \hat{F}_3(t) & \hat{G}_3(t) \\ 0 & 0 & 0 & \hat{F}_4(t) \end{bmatrix},$$

where

$$\hat{F}_3(t) = e^{At},$$

$$\begin{aligned}
\hat{G}_2(t) &= e^{-A^T t} \int_0^t e^{A^T s} Q_c e^{A s} ds, \\
\hat{G}_3(t) &= \int_0^t e^{A(t-s)} B ds, \\
\hat{H}_2(t) &= e^{-A^T t} \int_0^t \int_0^s e^{A^T s} Q_c e^{A r} B dr ds, \\
\hat{K}_1(t) &= e^{-A^T t} \int_0^t \int_0^s \int_0^r e^{A^T r} Q_c e^{A w} B dw dr ds.
\end{aligned}$$

It turns out that the integrals (7.6.3)-(7.6.3) can be expressed in terms of these sub-matrices of $e^{\hat{C}t}$ when we set $t = \Delta$:

$$H(\Delta) = \hat{G}_3(\Delta), \quad (7.6.4)$$

$$Q(\Delta) = \hat{F}_3(\Delta)^T \hat{G}_2(\Delta), \quad (7.6.5)$$

$$M(\Delta) = \hat{F}_3(\Delta)^T \hat{H}_2(\Delta), \quad (7.6.6)$$

$$W(\Delta) = \left[B^T \hat{F}_3(\Delta)^T \hat{K}_1(\Delta) \right] + \left[B^T \hat{F}_3(\Delta)^T \hat{K}_1(\Delta) \right]^T. \quad (7.6.7)$$

To use (7.6.4)-(7.6.7) for practical applications we need a means for estimating $e^{\hat{C}\Delta}$. One possibility is to use Ward's algorithm, with estimates of the form

$$e^{\hat{C}\Delta} = \left[R_{qq} \left(\frac{\hat{C}\Delta}{2^j} \right) \right]^{2^j}, \quad q, j \geq 0, \quad (7.6.8)$$

where $R_{qq}(z)$ is the (q, q) -Padé approximant to e^z ,

$$R_{qq}(z) = \frac{\sum_{k=0}^q c_k z^k}{\sum_{k=0}^q c_k (-z)^k}, \quad c_k = \frac{(2q-k)!q!}{(2q)!k!(q-k)!}.$$

The scaling by 2^j followed by repeatedly squaring greatly enhances the numerical properties of the ordinary Padé approximation. It is clear that the approximation in (7.6.8) has the form

$$\left[R_{qq} \left(\frac{\hat{C}\Delta}{2^j} \right) \right]^{2^j} = \begin{bmatrix} F_1(t) & G_1(t) & H_1(t) & K_1(t) \\ 0 & F_2(t) & G_2(t) & H_2(t) \\ 0 & 0 & F_3(t) & G_3(t) \\ 0 & 0 & 0 & F_4(t) \end{bmatrix}.$$

Thus, in accordance with (7.6.4)-(7.6.7), we obtain the following approximations to (7.6.3)-(7.6.3):

$$\begin{aligned} H(\Delta) &\approx G_3(\Delta), \\ Q(\Delta) &\approx F_3(\Delta)^T G_2(\Delta), \\ M(\Delta) &\approx F_3(\Delta)^T H_2(\Delta), \\ W(\Delta) &\approx [B^T F_3(\Delta)^T K_1(\Delta)]^T + [B^T F_3(\Delta)^T K_1(\Delta)]. \end{aligned}$$

This procedure is easy to implement. All that is involved is a single call to any Padé matrix exponential subroutine, followed by some elementary matrix computations.

7.6.4 Gauss-Hermite Quadrature

This type of quadrature is particularly useful when we consider stochastic processes with Gaussian distributions, as they approximate integrals of the type

$$\int_{-\infty}^{\infty} F(x) e^{-x^2} dx.$$

The general idea of Gaussian quadrature is to choose not only the weighting coefficients, but also the location of the abscissas at which the function is to be evaluated. Thus one can achieve Gaussian quadrature formulae whose order is twice that of Newton-Cotes formulae with the same number of function evaluations. The useful feature of Gaussian quadrature formulae is that we can arrange the choice of weights and abscissas to make the integral exact for the class of integrands “polynomials times some known function $W(x)$ ” rather than for the usual class of integrands “polynomials”. Given weights ω_i and abscissas x_i the approximation

$$\int_b^a F(x) W(x) dx \approx \sum_{i=1}^N w_i F(x_i),$$

is exact if $F(x)$ is a polynomial. If the integration interval is $(-\infty, \infty)$, and the weight function is Gaussian density $W(x) = e^{-x^2}$, we speak of *Gauss-Hermite*

quadrature. In general, high order quadrature corresponds to high accuracy only if the integrand is very smooth. For further details on Gaussian quadrature see, for example, Stoer and Bulirsch (1980).

This method can be generalised to higher dimensions as well. As an example of its use, let us assume that we want to compute the first order moment of the two-dimensional function $F(x_1, x_2)$, where

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right)$$

We therefore have to compute the integral

$$|\Sigma|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x_1, x_2) \exp \left(-\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right) dx_1 dx_2,$$

where $x = (x_1, x_2)'$, $\mu = (\mu_1, \mu_2)$, $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$. Let Φ be the Cholesky decomposition of Σ such that $\Sigma = \Phi\Phi'$, and make the change of variable

$$y = \Phi^{-1}(x - \mu)/\sqrt{2} \iff x = \sqrt{2}\Phi y + \mu.$$

Then the integral rewrites as

$$\pi^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\sqrt{2}\Phi y + \mu) \exp \left(-\sum_1^s y_i^2 \right) dy_1 dy_2.$$

We then use the product rule, relying on one-dimensional Gauss-Hermite quadrature, so that we approximate the integral by

$$\pi^{-1} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \omega_{i_1}^1 \omega_{i_2}^2 F(\sqrt{2}\phi_{11}y_1 + \mu_1, \sqrt{2}(\phi_{21}y_1 + \phi_{22}y_2) + \mu_2).$$

7.6.5 Figures

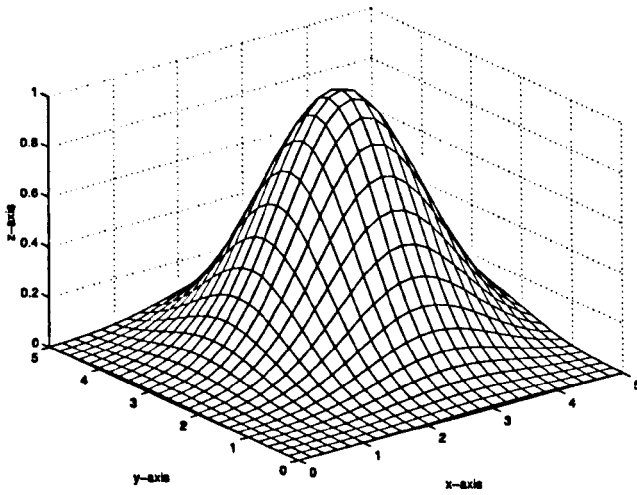


Figure 7.1: A radial basis function of the form $\exp(-a\|x - b\|)$, with $a = 1$ and $b = (3, 3)$.

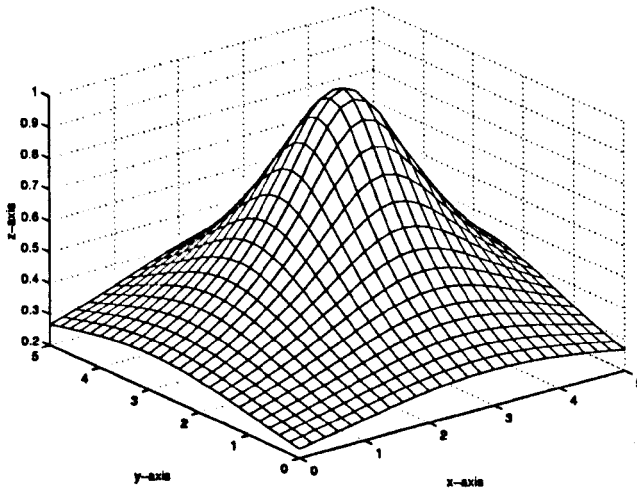


Figure 7.2: A radial basis function of the form $1/\sqrt{1 + a\|x - b\|}$, with $a = 0.5$ and $b = (3, 3)$.

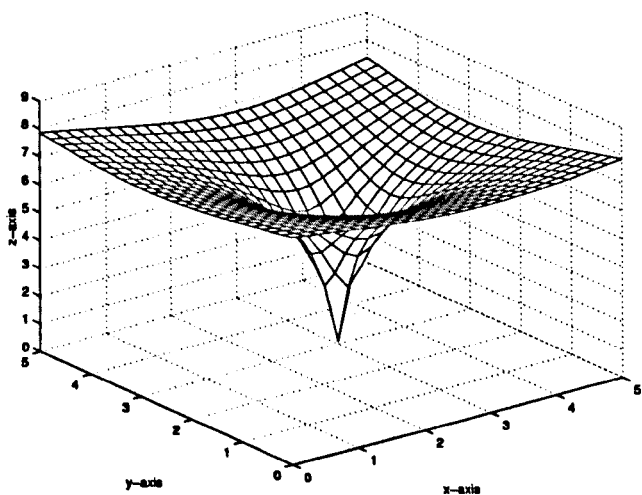


Figure 7.3: A radial basis function of the form $\log(1.1 + -a\|x - b\|)$, with $a = 200$ and $b = (3, 3)$.

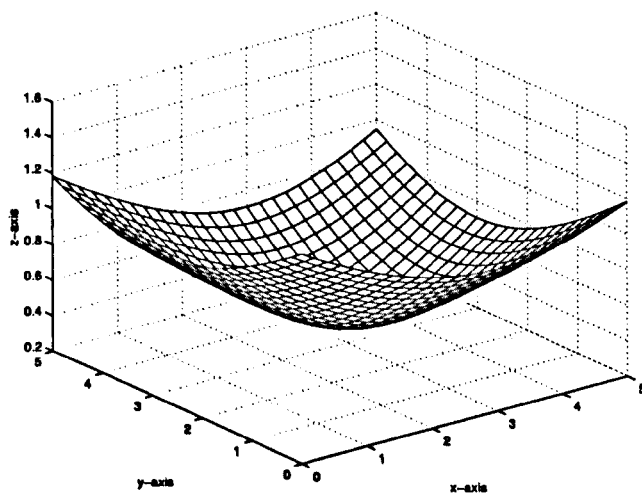


Figure 7.4: A radial basis function of the form $\sqrt{0.1 + a\|x - b\|}$, with $a = 0.1$ and $b = (3, 3)$.

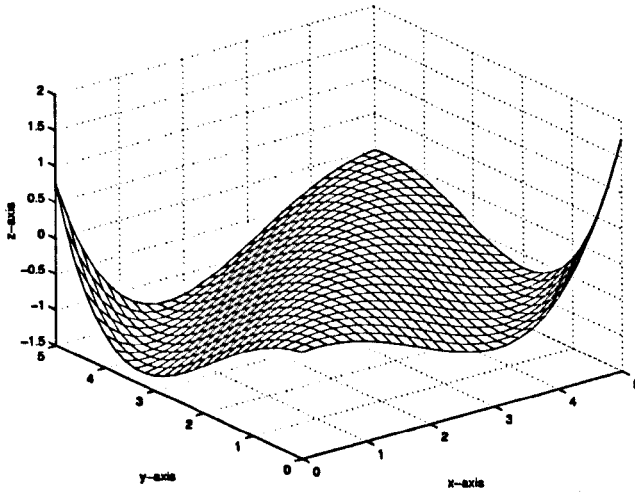


Figure 7.5: A typical ridge function.

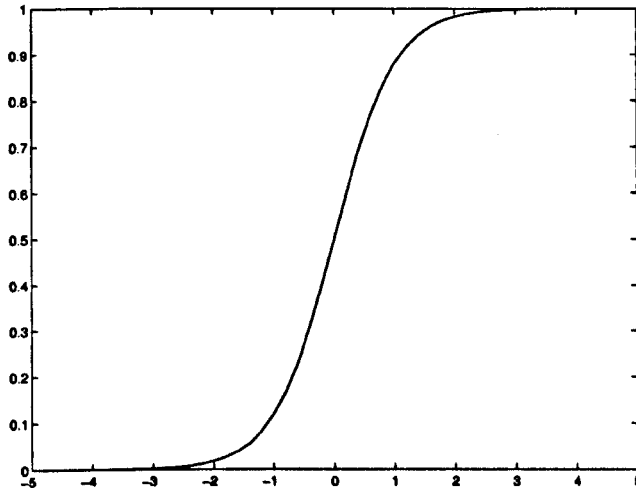


Figure 7.6: A logistic sigmoid with the parameter $a = 2$.

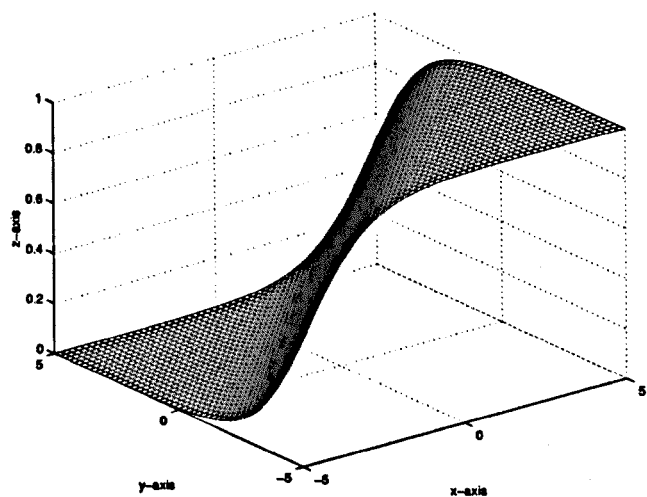


Figure 7.7: A sigmoid ridge function.

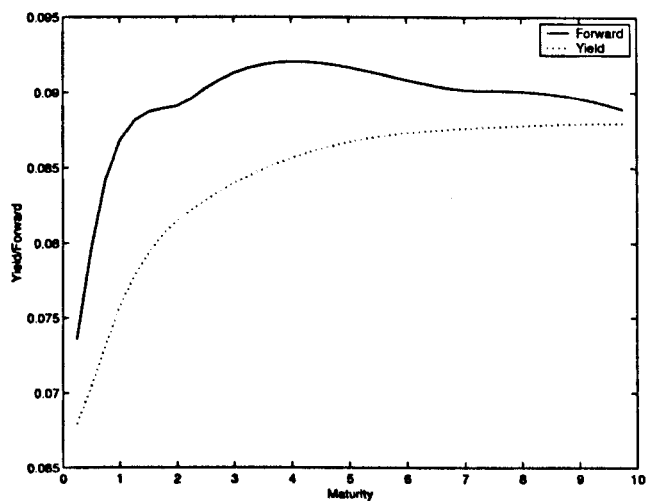


Figure 7.8: Yield and forward rate curves. Date: 03-Feb-95.

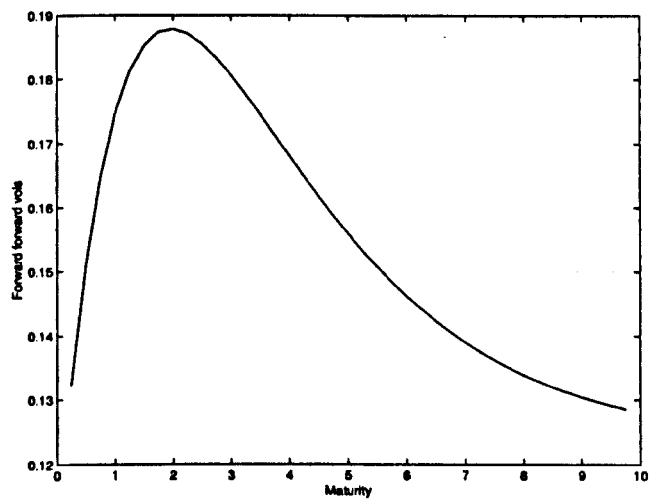


Figure 7.9: Forward forward volatility curve. Date: 03-Feb-95.

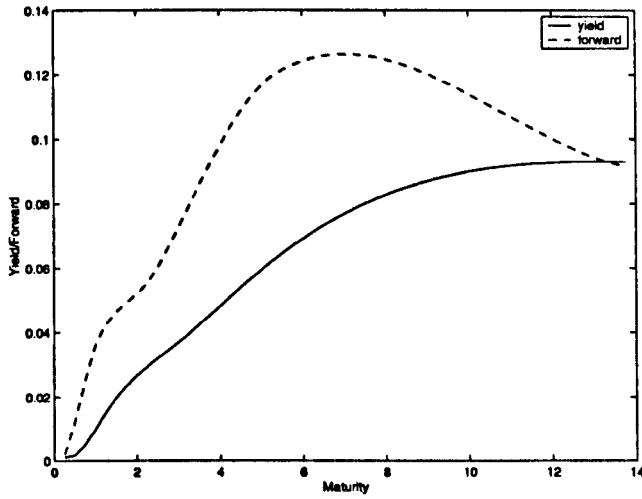


Figure 7.10: Yield and forward rate curves. Date: 04-Aug-00.

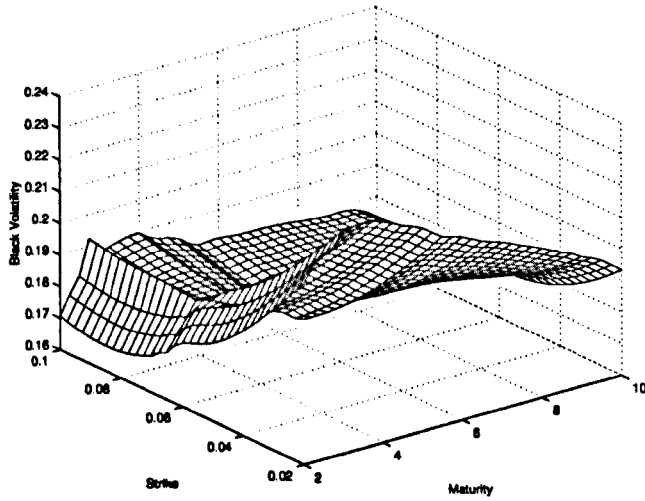


Figure 7.11: Black caplet implied volatility surface. Date: 04-Aug-00.

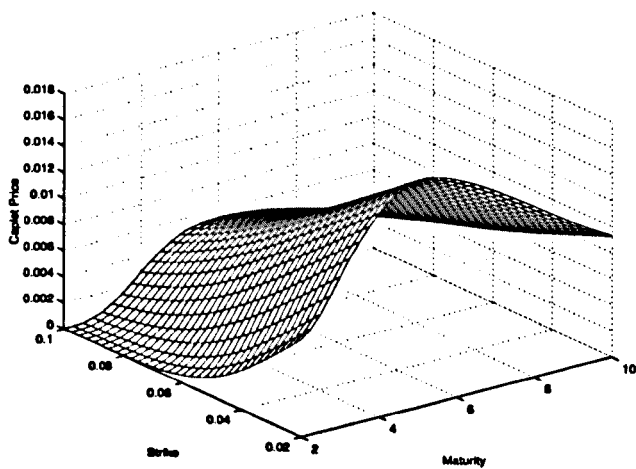


Figure 7.12: Market caplet prices. Date: 04-Aug-00.

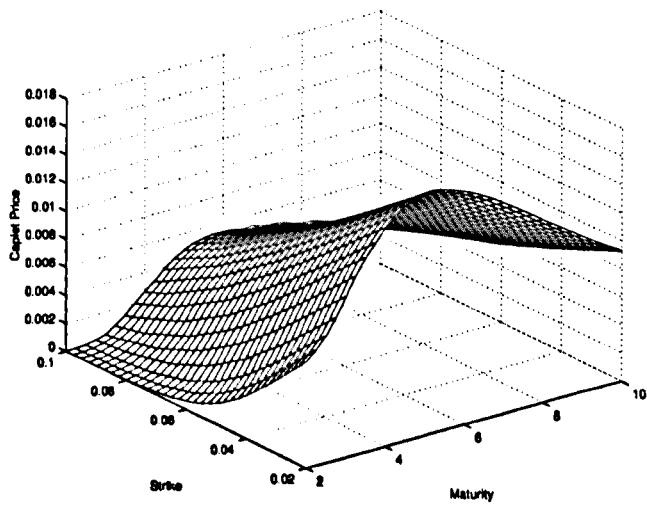


Figure 7.13: Model caplet prices. Date: 04-Aug-00.

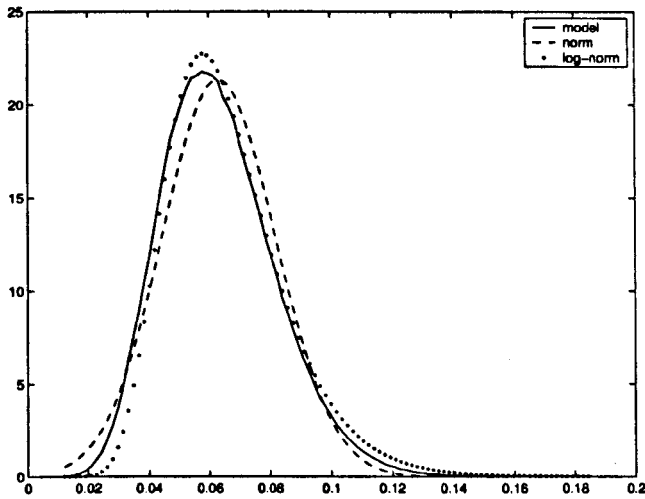


Figure 7.14: Histogram for the forward LIBOR rates with maturity 2.5 years, together with normal and log-normal densities. The LIBOR rate distribution results from the model fit to the bond and caplet prices. Date: 04-Aug-00.

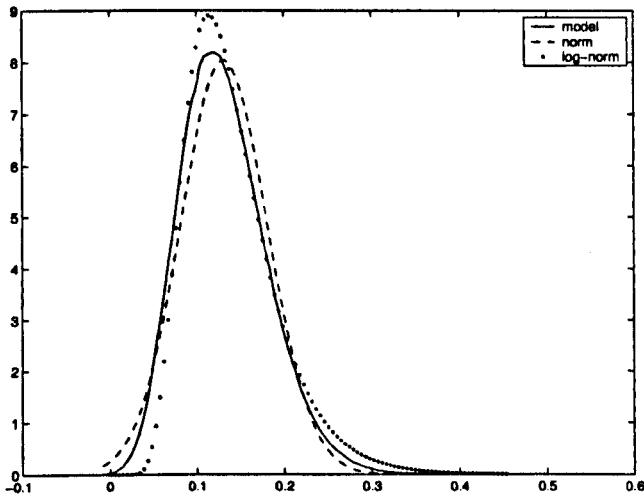


Figure 7.15: Histogram for the forward LIBOR rates with maturity 5 years, together with normal and log-normal densities. The LIBOR rate distribution results from the model fit to the bond and caplet prices. Date: 04-Aug-00.

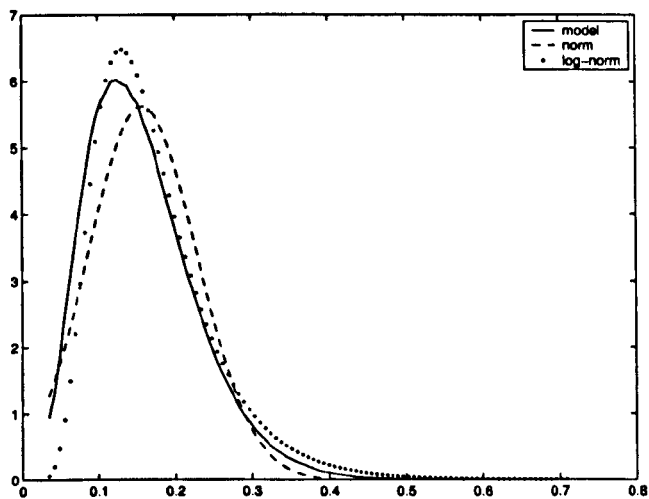


Figure 7.16: Histogram for the forward LIBOR rates with maturity 7.5 years, together with normal and log-normal densities. The LIBOR rate distribution results from the model fit to the bond and caplet prices. Date: 04-Aug-00.

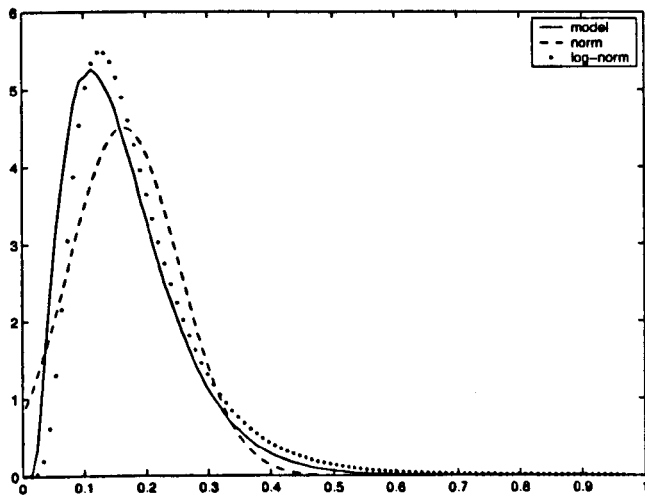


Figure 7.17: Histogram for the forward LIBOR rates with maturity 10 years, together with normal and log-normal densities. The LIBOR rate distribution results from the model fit to the bond and caplet prices. Date: 04-Aug-00.

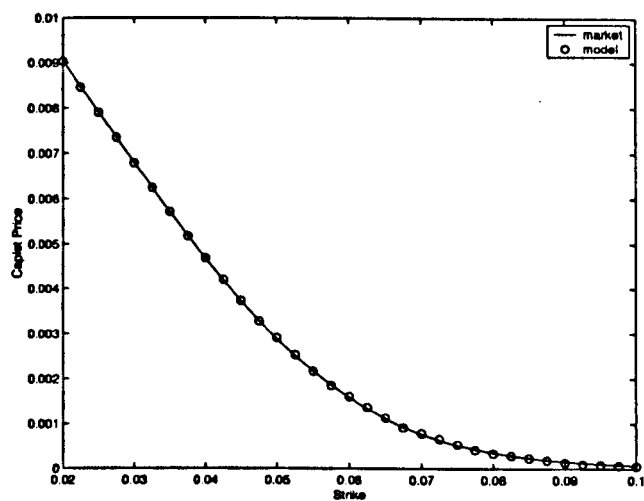


Figure 7.18: Model and caplet prices with 2.5 years to maturity. Date: 04-Aug-00.

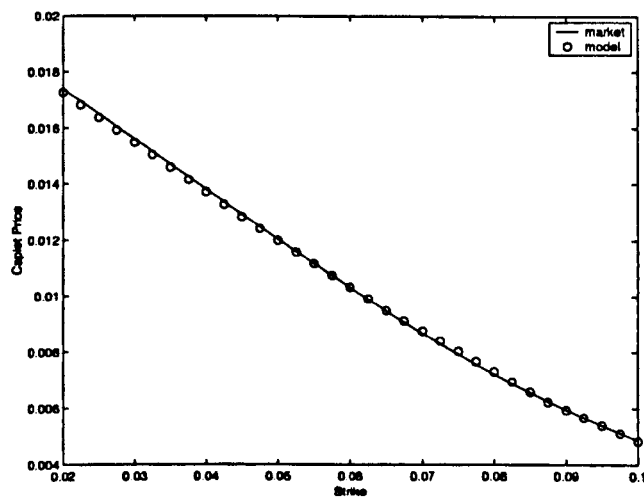


Figure 7.19: Model and caplet prices with 5 years to maturity. Date: 04-Aug-00.

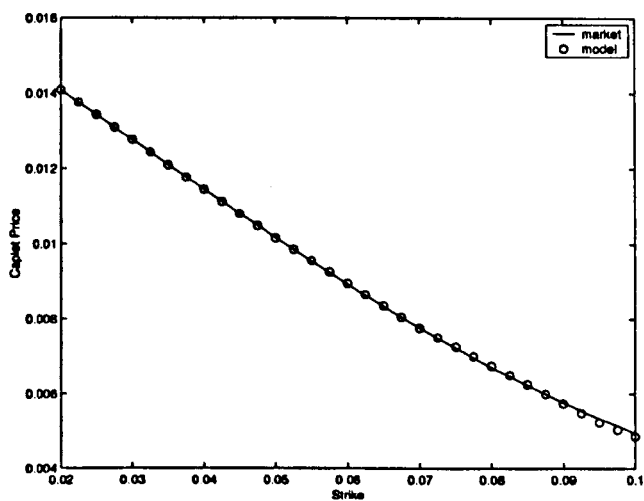


Figure 7.20: Model and caplet prices with 7.5 years to maturity. Date: 04-Aug-00.

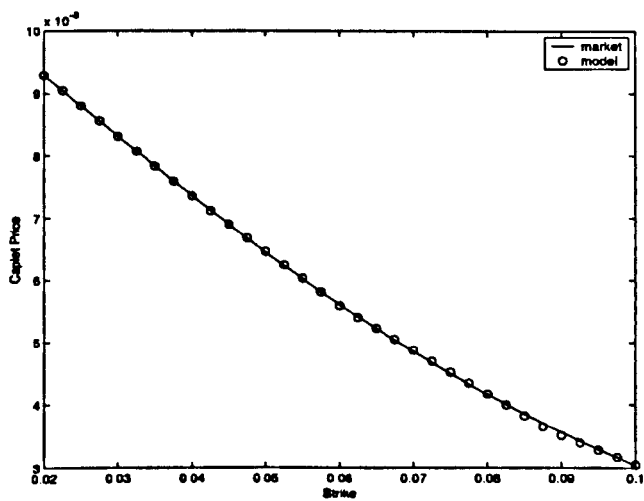


Figure 7.21: Model and caplet prices with 10 years to maturity. Date: 04-Aug-00.

7.6.6 Tables

Table 7.1: Zero Coupon Discount Function (ZCDF) for 3 Feb 1995.

Tenor: x	ZCDF (x)	Tenor: x	ZCDF (x)	Tenor: x	ZCDF (x)
0.07671233	0.99489605	2.37260274	0.82165363	6.50136986	0.56589374
0.10958904	0.99268989	2.62191781	0.80338660	7.00821918	0.53980408
0.37808219	0.97422289	2.87123288	0.78546824	7.50684932	0.51675532
0.62739726	0.95577923	3.12054795	0.76794952	8.00547945	0.49467915
0.87671233	0.93669956	3.49863014	0.74763879	8.50410959	0.47353468
1.12602740	0.91730596	4.00273973	0.71105063	9.00547945	0.45317677
1.37534247	0.89785353	4.49863014	0.67976222	9.50410959	0.43378439
1.62465753	0.87847062	5.00273973	0.64895804	10.00821918	0.41501669
1.87397260	0.85927558	5.50136986	0.62032221	10.50821918	0.39720417
2.12328767	0.84029504	6.01095890	0.59213852	11.00821918	0.38015617

Table 7.2: Market Data for 3 Feb 1995.

Tenor	Yield	Forward	Tenor	Yield	Forward	Tenor	Yield	Forward
0.25	6.794	7.359	3.50	8.488	9.183	6.75	8.755	9.028
0.50	7.043	7.963	3.75	8.527	9.197	7.00	8.761	9.018
0.75	7.324	8.426	4.00	8.563	9.202	7.25	8.766	9.012
1.00	7.577	8.676	4.25	8.594	9.201	7.50	8.771	9.012
1.25	7.779	8.811	4.50	8.622	9.193	7.75	8.776	9.011
1.50	7.935	8.870	4.75	8.647	9.182	8.00	8.780	9.007
1.75	8.055	8.891	5.00	8.666	9.165	8.25	8.784	9.001
2.00	8.147	8.908	5.25	8.687	9.146	8.50	8.787	8.991
2.25	8.221	8.957	5.50	8.703	9.125	8.75	8.790	8.978
2.50	8.284	9.025	5.75	8.717	9.103	9.00	8.793	8.961
2.75	8.343	9.082	6.00	8.729	9.082	9.25	8.794	8.941
3.00	8.396	9.126	6.25	8.739	9.067	9.50	8.796	8.917
3.25	8.444	9.160	6.50	8.747	9.046	9.75	8.796	8.888

Table 7.3: Discount bond price for 3 Feb 1995 together with model price and percentage error for the fit.

Maturity	ZCB	Model	Error (%)	Maturity	ZCB	Model	Error (%)
0.25	0.9832	0.9832	0.00	5.00	0.6483	0.6483	0.00
0.50	0.9654	0.9654	0.00	5.25	0.6338	0.6338	0.00
0.75	0.9466	0.9464	0.00	5.50	0.6196	0.6196	0.00
1.00	0.9270	0.9270	0.00	5.75	0.6058	0.6058	0.00
1.25	0.9073	0.9073	0.00	6.00	0.5923	0.5923	0.00
1.50	0.8878	0.8878	0.00	6.25	0.5791	0.5791	0.00
1.75	0.8685	0.8685	0.00	6.50	0.5663	0.5663	0.00
2.00	0.8496	0.8496	0.00	6.75	0.5538	0.5538	0.00
2.25	0.8311	0.8311	0.00	7.00	0.5416	0.5416	0.00
2.50	0.8129	0.8129	0.00	7.25	0.5296	0.5296	0.00
2.75	0.7950	0.7950	0.00	7.50	0.5180	0.5180	0.00
3.00	0.7773	0.7773	0.00	7.75	0.5065	0.5065	0.00
3.25	0.7600	0.7600	0.00	8.00	0.4954	0.4954	0.00
3.50	0.7430	0.7430	0.00	8.25	0.4845	0.4845	0.00
3.75	0.7263	0.7263	0.00	8.50	0.4738	0.4738	0.00
4.00	0.7100	0.7100	0.00	8.75	0.4634	0.4634	0.00
4.25	0.6940	0.6940	0.00	9.00	0.4532	0.4532	0.00
4.50	0.6784	0.6784	0.00	9.25	0.4433	0.4433	0.00
4.75	0.6632	0.6632	0.00	9.50	0.4336	0.4336	0.00
5.00	0.6483	0.6483	0.00	10.00	0.4149	0.4149	0.00

Table 7.4: Results from the model calibration to the ATM caps and zero coupon bond prices. Date: 03-Feb-1995.

Length	A-T-M Strike (%)	Black Vol (%)	Market Price (bp)	Model Price (bp)	Error (%)	Error (bp)
1	7.91	15.50	27	27	0.04	0.01
2	8.41	17.75	99	99	0.32	0.32
3	8.60	18.00	183	186	1.57	2.92
4	8.74	17.75	272	273	0.40	1.09
5	8.82	17.75	356	356	-0.07	-0.25
7	8.88	16.50	508	507	-0.25	-1.25
10	8.90	15.50	701	699	-0.24	-1.71

Table 7.5: Model and market swaption prices, resulting from calibration to the ATM caps and zero coupon bond prices. Date: 03-Feb-1995.

Maturity × Swap Length	A-T-M Strike (%)	Black Vol (%)	Market Price (bp)	Model Price (bp)	Error (%)	Error (bp)
0.25 × 2	8.56	16.75	51	44	-13.05	-6.6
0.25 × 3	8.74	16.50	73	73	-0.44	-0.3
0.25 × 5	8.93	15.00	105	121	15.58	16.3
0.25 × 7	8.99	13.75	124	158	26.70	33.2
1 × 4	9.13	15.50	172	193	11.99	20.6
1 × 9	9.13	13.25	270	316	16.99	46.0
2 × 8	9.17	12.75	312	352	12.62	39.4

Table 7.6: Implied black swaption volatility, resulting from calibration to the ATM caps and zero coupon bond prices. Date: 03-Feb-1995.

	1	2	3	4	5	6	7	8
2	16.43	14.67	16.92	16.15	15.04	9.87	5.66	9.54
3	17.06	15.63	16.76	15.88	11.94	8.94	7.91	
4	17.39	15.84	16.53	13.52	10.91	9.92		
5	17.40	15.96	14.66	12.52	11.45			
6	17.39	14.62	13.75	12.80				
7	16.14	13.92	13.89					
8	15.45	14.08						
9	15.52							

Table 7.7: Zero coupon bond prices together with calibrated model prices and percentage error. The model was calibrated to the caplet and bond prices. Date: 04-Aug-00.

Maturity	ZCB	Model	Error (%)	Maturity	ZCB	Model	Error (%)
2.00	0.9483	0.9473	0.101	6.25	0.6406	0.6409	-0.048
2.25	0.9361	0.9351	0.107	6.50	0.6212	0.6216	-0.055
2.50	0.9234	0.9223	0.114	6.75	0.6023	0.6027	-0.063
2.75	0.9098	0.9088	0.114	7.00	0.5839	0.5842	-0.063
3.00	0.8950	0.8942	0.098	7.25	0.5660	0.5664	-0.072
3.25	0.8791	0.8783	0.090	7.50	0.5487	0.5491	-0.074
3.50	0.8620	0.8613	0.085	7.75	0.5319	0.5323	-0.075
3.75	0.8439	0.8432	0.088	8.00	0.5158	0.5162	-0.077
4.00	0.8250	0.8244	0.076	8.25	0.5002	0.5006	-0.087
4.25	0.8053	0.8047	0.069	8.50	0.4852	0.4856	-0.086
4.50	0.7850	0.7845	0.061	8.75	0.4707	0.4711	-0.087
4.75	0.7643	0.7639	0.050	9.00	0.4568	0.4572	-0.087
5.00	0.7434	0.7431	0.038	9.25	0.4435	0.4439	-0.087
5.25	0.7223	0.7222	0.019	9.50	0.4307	0.4311	-0.089
5.50	0.7014	0.7014	0.002	9.75	0.4184	0.4188	-0.087
5.75	0.6808	0.6809	-0.016	10.00	0.4067	0.4070	-0.088
6.00	0.6605	0.6608	-0.035				

Table 7.8: Calibrated model and market caplets prices for a range of strikes and maturities, together with errors expressed in basis points and percentage. Date: 04-Aug-00.

	0.02%	0.03%	0.04%	0.05%	0.06%	0.07%	0.08%	0.09%	0.1%
2	75.09	51.79	30.37	14.51	5.95	2.06	0.61	0.15	0.03
M	74.60	51.48	30.42	14.69	5.84	1.77	0.48	0.13	0.04
bp	-0.49	-0.30	0.05	0.18	-0.10	-0.29	-0.13	-0.02	0.01
%	-0.65	-0.59	0.16	1.20	-1.79	-16.18	-26.78	-15.97	24.54
3	114.73	93.01	72.08	52.97	36.62	23.29	14.16	8.31	4.11
M	115.55	93.70	72.45	52.85	36.11	22.93	14.06	8.43	5.07
bp	0.83	0.69	0.37	-0.12	-0.51	-0.37	-0.09	0.13	0.96
%	0.71	0.73	0.51	-0.23	-1.41	-1.60	-0.67	1.53	18.94
4	156.05	136.02	116.27	97.28	79.21	62.76	48.63	36.05	26.61
M	156.70	136.61	116.71	97.29	78.81	62.00	47.98	36.63	27.82
bp	0.64	0.59	0.44	0.01	-0.40	-0.76	-0.65	0.57	1.21
%	0.41	0.43	0.38	0.01	-0.50	-1.23	-1.35	1.57	4.35
5	172.70	154.79	137.11	119.90	103.18	87.53	73.04	59.31	48.43
M	174.06	156.03	138.09	120.38	103.10	86.73	72.14	59.46	48.77
bp	1.36	1.24	0.99	0.48	-0.09	-0.80	-0.90	0.15	0.34
%	0.78	0.79	0.71	0.40	-0.09	-0.92	-1.25	0.26	0.70
6	165.99	150.03	134.27	118.65	103.95	89.30	76.71	64.13	53.54
M	166.74	150.75	134.86	119.17	103.85	89.30	76.13	64.45	54.36
bp	0.75	0.73	0.59	0.52	-0.10	0.00	-0.58	0.32	0.82
%	0.45	0.48	0.44	0.43	-0.10	0.00	-0.77	0.49	1.50
7	150.13	136.14	122.35	108.64	95.78	82.91	71.42	60.56	49.80
M	150.45	136.34	122.35	108.59	95.26	82.69	71.28	61.10	52.22
bp	0.32	0.20	-0.00	-0.06	-0.52	-0.21	-0.14	0.54	2.43
%	0.21	0.14	-0.00	-0.05	-0.54	-0.26	-0.19	0.89	4.65
8	130.67	118.36	106.21	94.33	83.07	71.82	62.37	53.02	44.24
M	130.92	118.48	106.18	94.18	82.69	72.01	62.30	53.64	46.08
bp	0.25	0.12	-0.03	-0.15	-0.38	0.19	-0.07	0.62	1.84
%	0.19	0.10	-0.02	-0.16	-0.46	0.26	-0.12	1.16	3.99
9	111.21	100.11	89.25	79.01	69.14	60.36	52.29	44.23	38.17
M	111.15	100.15	89.35	78.92	69.13	60.16	52.00	44.75	38.44
bp	-0.06	0.04	0.10	-0.09	-0.01	-0.19	-0.29	0.52	0.28
%	-0.05	0.04	0.11	-0.11	-0.02	-0.32	-0.57	1.17	0.72
10	92.96	83.11	73.54	64.62	55.92	48.73	41.75	35.12	30.51
M	92.91	83.13	73.59	64.48	56.08	48.55	41.68	35.63	30.42
bp	-0.05	0.02	0.05	-0.14	0.17	-0.17	-0.07	0.52	-0.09
%	-0.06	0.03	0.07	-0.21	0.29	-0.36	-0.17	1.45	-0.29

Table 7.9: Mean, standard deviation, skewness and kurtosis for LIBOR rates, resulting from the calibration of the model to the zero-coupon bond and caplets. Date: 04-Aug-00.

Maturity	Mean	St.Dev.	Skewness	Kurtosis
2.00	0.054	0.014	0.686	0.646
2.25	0.058	0.016	0.715	0.678
2.50	0.063	0.019	0.533	0.407
2.75	0.070	0.023	0.482	0.332
3.00	0.078	0.025	0.492	0.354
3.25	0.085	0.028	0.491	0.354
3.50	0.093	0.032	0.740	0.733
3.75	0.100	0.036	0.508	0.373
4.00	0.107	0.038	0.557	0.443
4.25	0.114	0.041	0.540	0.422
4.50	0.120	0.044	0.552	0.444
4.75	0.126	0.047	0.556	0.452
5.00	0.131	0.050	0.516	0.398
5.25	0.136	0.053	0.635	0.587
5.50	0.139	0.055	0.620	0.562
5.75	0.143	0.058	0.682	0.677
6.00	0.145	0.059	0.699	0.713
6.25	0.148	0.061	0.739	0.797
6.50	0.149	0.058	0.633	0.694
6.75	0.154	0.067	0.880	1.055
7.00	0.153	0.061	0.627	0.699
7.25	0.156	0.068	0.822	0.975
7.50	0.158	0.071	0.873	1.083
7.75	0.159	0.071	0.862	1.100
8.00	0.158	0.070	0.811	1.086
8.25	0.163	0.080	0.983	1.364
8.50	0.164	0.081	0.999	1.414
8.75	0.164	0.082	1.043	1.619
9.00	0.164	0.082	0.977	1.371
9.25	0.165	0.085	1.035	1.579
9.50	0.166	0.088	1.092	1.734
9.75	0.166	0.090	1.171	2.077
10.00	0.165	0.089	1.153	2.006

Table 7.10: Zero coupon bond prices together with calibrated model prices and percentage error. The model was calibrated to the ATM swaptions and bond prices.
Date: 04-Aug-00.

Maturity	ZCB	Model	Error (%)	Maturity	ZCB	Model	Error (%)
2.00	0.9483	0.9498	-0.160	6.25	0.6406	0.6406	0.000
2.25	0.9361	0.9361	0.000	6.50	0.6212	0.6212	-0.000
2.50	0.9234	0.9234	0.000	6.75	0.6023	0.6023	0.000
2.75	0.9098	0.9098	-0.000	7.00	0.5839	0.5830	0.151
3.00	0.8950	0.8946	0.042	7.25	0.5660	0.5660	0.000
3.25	0.8791	0.8791	0.000	7.50	0.5487	0.5487	-0.000
3.50	0.8620	0.8620	0.000	7.75	0.5319	0.5319	0.000
3.75	0.8439	0.8439	-0.000	8.00	0.5158	0.5112	0.885
4.00	0.8250	0.8248	0.024	8.25	0.5002	0.5002	0.000
4.25	0.8053	0.8053	0.000	8.50	0.4852	0.4852	0.000
4.50	0.7850	0.7850	-0.000	8.75	0.4707	0.4707	-0.000
4.75	0.7643	0.7643	-0.000	9.00	0.4568	0.4579	-0.226
5.00	0.7434	0.7448	-0.196	9.25	0.4435	0.4435	-0.000
5.25	0.7223	0.7223	-0.000	9.50	0.4307	0.4307	-0.000
5.50	0.7014	0.7014	0.000	9.75	0.4184	0.4184	0.000
5.75	0.6808	0.6808	0.000	10.00	0.4067	0.4045	0.536
6.00	0.6605	0.6638	-0.502	10.25	0.3954	0.3954	-0.000

Table 7.11: Swaptions contracts expressed as maturity \times length. Model fit to the ATM swaption prices in basis point units together with market prices and the percentage errors. Date: 04-Aug-00.

	2	3	4	5	6	7	8	9	10
0.5	24	41	55	62	64	64	58	53	47
Mk	24	41	55	62	64	64	58	53	47
%	0.07	-0.00	0.00	0.01	-0.00	-0.00	0.00	0.00	-0.00
1	43	74	113	91	121	117	113	79	80
Mk	51	87	113	124	123	117	106	95	84
%	15.36	15.63	-0.00	26.08	1.99	-0.01	-5.97	17.47	5.38
2	102	174	197	201	225	182	175	157	150
Mk	119	179	220	232	228	212	191	171	150
%	14.50	2.91	10.51	13.33	1.40	14.11	8.75	7.99	0.00
3	184	247	298	302	295	258	240	227	216
Mk	195	269	310	318	307	284	255	227	202
%	5.33	8.05	3.88	4.95	4.14	8.98	5.94	-0.00	-7.12
4	244	339	381	364	369	332	305	291	
Mk	265	346	384	385	369	340	305	272	
%	7.86	2.03	0.61	5.42	-0.00	2.24	0.00	-6.98	
5	323	410	444	435	440	399	368		
Mk	329	410	444	435	415	381	344		
%	1.95	0.00	0.01	0.00	-6.05	-4.77	-6.85		
6	387	467	502	502	505	461			
Mk	385	465	496	484	461	427			
%	-0.45	-0.42	-1.26	-3.65	-9.41	-7.99			
7	428	520	560	563	564				
Mk	428	510	538	523	499				
%	0.00	-1.92	-4.13	-7.68	-13.01				
8	473	571	616	619					
Mk	470	545	569	556					
%	-0.77	-4.81	-8.14	-11.33					
9	515	619	666						
Mk	492	569	597						
%	-4.59	-8.69	-11.62						
10	554	665							
Mk	513	593							
%	-8.06	-12.03							

Chapter 8

Conclusions and Further Research

In this concluding chapter we will simultaneously summarise the main contribution of the thesis, and suggest some interesting areas for further research. In this thesis we investigated interest rate modelling with the emphasis on applications in pricing and risk management of interest rate derivatives and portfolios.

The first part of the thesis was developed within the random field framework suggested by Kennedy (1994). To use the random field approach one needs to find a flexible family of covariance functions, and thus a flexible class of field models, that can fit the empirical covariance matrix observed in the market. Functions from this class should be strictly positive definite, thus providing a truly infinite-factor structure for the model. Neither Kennedy (1994), nor subsequent research by Goldstein (2000), and Santa-Clara and Sornette (2001) have addressed this question. Without this class of covariance functions one cannot capitalise on the main advantage of the field models, i.e. capturing the inter-dynamics of movements in the term structure.

We have developed three methodologies for constructing strictly positive definite

covariance functions, characterising infinite-factor Gaussian fields. We tested all three constructions on the sample covariance and correlation matrices obtained from US and Japanese bond market data. The empirical and numerical test suggest that these classes of field models present a very satisfactory solution to the posed problem. The models we have developed make the random field methodology a much more practical tool, which can fully incorporate the key market information: the covariation of the yield curve. This should allow a better understanding of the risk exposure in the positions sensitive to the interest rates. Furthermore, our research should facilitate the usage of field models within the pricing context, especially for instruments depending on the evolution of the whole yield curve.

We achieved some very good results in constructing and fitting infinite-factor Gaussian field models, but more extensive empirical testing still needs to be conducted. Comparison studies should reveal if these classes of models are superior to alternative model specifications in pricing and risk management of fixed income instruments and portfolios.

A number of further extensions suggest themselves. In this thesis we have not exhausted all possible ways of constructing SPD covariance and correlation functions. Other methods are possible. In fact, inspired by the research in this thesis, in a joint work Johnson and Weigel (2003) investigate a method for the construction such functions using the tools of Matrix Completion¹ Theory. The idea is to consider the observed sample correlation matrix as a sub-matrix of a larger matrix. The missing entries can then be filled subject to constraints, such as positive definiteness, monotonicity, etc. Preliminary results are very promising.

¹A partial matrix is a matrix in which some entries are specified and others are not. A completion of a partial matrix is a specific choice of values for the unspecified entries. A pattern for $n \times n$ matrices is a list of positions of an $n \times n$ matrix, that is, a subset of $\{1, \dots, n\} \times \{1, \dots, n\}$. A partial matrix specifies the pattern if its specified entries are exactly those listed in the pattern. For a particular class P of matrices, the P -matrix completion problem for patterns asks which patterns have the property that any partial P -matrix that specifies the pattern can be completed to a P -matrix. In our case the class P comprises unit diagonal and positive definite matrices.

In the field models we have considered that the instantaneous covariance and correlation functions depend on the time to maturity. This may not be justified in practice. A more general model, such as one that allows a time-varying covariance structure, might be needed. One example of this type of model is Collin-Dufresne and Goldstein (2000). As we have mentioned, such extensions are relatively trivial within our framework.

An area that is closely related to fixed income modelling is credit risk. The random field methodology could be extended to include defaultable term structures as well. Credit risk can be incorporated by adding an additional dimension in the random field that results in a correlation function of the form $\text{Corr}(T, S, R_1, R_2)$, where R_1 and R_2 represent ratings, T and S maturities of defaultable bonds. Thus, changes in yield of a defaultable bond can be modelled by a random field $X_{T,R}$ with $(T, R) \in \mathbb{R}^+ \times \mathbb{R}^+$. In fact, in a recent working paper, Douady and Jeanblanc (2002), have presented a default bond pricing model that is driven by a spread field. This spread field depends on the continuous rating and time to maturity.

The extension of field methodology to credit markets might provide a more integrated approach to risk management. This should have a number of consequences in practice. For example, from the banking supervision point of view, this should allow for more accurate capital requirements, which will release capital that can be invested somewhere else. From risk management point of view, this methodology should allow for a better risk assessment across different divisions of a bank or other financial institutions.

The second part of this thesis deals with pricing kernel (potential) models of the term structure, first introduced by Constantinides (1992). We have observed that market models are very popular among practitioners, as they allow instantaneous calibration to market prices. However they suffer from several restrictions. They have non-Markov dynamics. Though they are easy to calibrate to instruments for

which they were specifically designed, calibration to other instruments may be also a problem. Extensions to smiles and skews are complex.

Motivated by these observations, we set out to find a class of models that enjoys the same ease of calibration to the liquid market prices as the market models, and at the same time has a low-dimensional Markov structure, as in the case of short rate models.

We have developed a class of models within the pricing kernel framework. We have modelled the pricing kernel directly by approximating it with a set of radial basis functions. As the underlying noise in the economy, we have chosen a simple multi-factor diffusion. Our approach is just as simple in a one-factor as in a multi-factor setting. We have linked the construction of the kernel explicitly to the calibrating set of instruments. Once the kernel is constructed it prices correctly the chosen set of instruments and has a low-dimensional Markov structure. We tested our model on yield, at-the-money cap, caplet implied volatility surface, and swaption data. We achieved a very good quality of fit. In a summary, the class of implied kernel models is Markov by construction, calibration for a small number of factors is relatively easy, it can deal with American type options, and is flexible in fitting to smiles and skews.

Again, a number of further extensions suggest themselves. Our modelling framework in particular, and the kernel approach in general, might be extended to defaultable bonds. In a recent working paper, Polenghi (2002a), shows how the potential approach may be extended to handle defaultable bonds. He introduces a theoretical framework to price both risk-free and defaultable bonds. The model allows for multiple sources of dynamics with an unrestricted correlation matrix, and is consistent with positive interest rates.

In another development of the kernel framework, Rogers and Yousaf (2001) began to study what could be achieved with a finite-state Markov chain as the underlying

Markov process. The advantages of such a modelling approach are compelling: pricing of European options reduces to summing over some small finite set. Moreover, a relatively small number of instruments would be required to hedge any given asset, by hedging out the jumps of the asset value at the time when the Markov chain jumps. This replaces the concept of delta-hedging within this model. A further stage of this development would be the incorporation of credit-risky instruments. However, the potential approach would need some modification at a theoretical level to allow for this.

Another possible application of the kernel approach lies in pricing convertible bonds. A convertible bond is a coupon paying corporate bond that can be converted into company stocks at the discretion of the holder. A convertible bond is a challenging instrument to value, because it is both an equity and an interest rate derivative. These two components are subject to different credit risks. The traditional approach to this problem has been to write down stochastic differential equations for, say, the value of the firm and the short term interest rate, and to derive no-arbitrage pricing from solving the resulting partial differential equation. These equations support a pricing kernel which is obtained directly from the solution. Thus, in principal by modelling the kernel directly we could price the convertible securities as well. This would require an extension of the framework to deal with equities as well. These can be modelled within the existing kernel approach, but it may turn out to be easier, especially for applications in practice, to set up a model where the shares are log-Brownian within the overall framework of a kernel model for interest rates.

Beyond possible applications to convertibles, there is considerable scope for testing the kernel approach to modelling interest rates in many countries. A major advantage of the kernel approach over conventional approaches is that adding new countries does not require new sources of noise to be introduced. Rather one simply sets up a new state-price density process for the new country. The exchange rates between the countries are automatically modelled consistently by this. Pilot stud-

ies by Rogers and Zane (1996) of various simple diffusion models resulted in quite favourable fits, and it seemed important to investigate this more deeply. Of special interest are applications of the kernel approach to pricing cross currency derivatives. As the underlying dimension can be kept low this approach should allow for more effective pricing methods.

Further applications of the kernel approach might include the study on sovereign debt. In a recent study, Polenghi (2002b) develops a multi-factor econometric model for the joint dynamics of the sovereign debt issued by the US and by a representative set of developing countries. The potential approach is again used to price both risk-free and defaultable bonds. He finds that the correlations among spreads and the US yield curve changes with the sample period, with the maturity of the instruments and with the ratings of the emerging countries.

Bibliography

- [1] P. Acworth, M. Broadie, and P. Glasserman. A comparison of Some Monte Carlo and Quasi Monte Carlo techniques for Option Pricing. In *Monte Carlo and Quasi-Monte Carlo Methods 1996*, pages 1–18. Springer-Verlag New-York, 1998.
- [2] R.J. Adler. *The Geometry of Random Fields*. John Wiley and Sons, 1981.
- [3] C.M. Ahn and H.E. Thompson. Jump Diffusion Processes and the Term Structure of Interest Rates. *Journal of Finance*, 43(1):155–174, 1988.
- [4] D.H. Ahn and B. Gao. A Parametric Nonlinear Model of Term Structure Dynamics. *The Review of Financial Studies*, 12(4):721–762, Special, 1999.
- [5] D.H. Ahn, R.F. Dittmar, and B. Gao. Quadratic Term Structure Models: Theory and Evidence. *The Review of Financial Studies*, 15(1):243–288, Spring 2002.
- [6] Y. Ait-Sahalia. Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies*, 9:385–426, 1996b.
- [7] Y. Ait-Sahalia. Do Interest Rates Really Follow Continuous-Time Markov Diffusions? Working paper, University of Chicago, 1997.
- [8] A. Akgun. Model Risk with Jump-Diffusion Process. Working paper, University of Lausanne, 2000.
- [9] A.J. Amin, K.I. Morton. Implied Volatility Functions in Arbitrage-Free Term Structure Models. *Journal of Financial Economics*, 35:141–180, 1994.
- [10] K.I. Amin and V.K. Ng. Inferring Future Volatility from the Information in Implied Volatility in Eurodollar Options: A New Approach. *The Review of Financial Studies*, 10(2):333–367, 1997.
- [11] L. Andersen and J. Andreasen. Volatility Skews and Extensions of the Libor Market Model. *Applied Mathematical Finance*, 7:1–32, 2000.

- [12] L. Andersen and J. Andreasen. Factor Dependence of Bermudian Swaptions: Fact or Fiction. *Journal of Financial Economics*, 62:3–37, 2001.
- [13] L. Arnold. *Stochastic Differential Equations: Theory and Applications*. John Wiley and Sons, 1974.
- [14] P. Artzner and F. Delbaen. Term Structure of Interest Rates: The Martingale Approach. *Advances in Applied Mathematics*, 10:95–129, 1989.
- [15] K.T. Au and D.C. Thurston. A New Class of Duration Measures. *Economics Letters*, 47:371–375, 1995.
- [16] D.K. Backus and S.E. Zin. Reverse Engineering the Yield Curve. Working paper, NBER 4676, 1994.
- [17] P. Balland and L.P. Hughston. Markov Market Model Consistent With Cap Smile. *International Journal of Theoretical and Applied Finance*, 3(2):161–181, 2000.
- [18] R. Bansal, D.A. Hsieh, and S. Viswanathan. A New Approach to International Arbitrage Pricing. *The Journal of Finance*, XLVIII(5):1719–1747, December 1993.
- [19] R. Bansal and S. Viswanathan. No Arbitrage and Arbitrage Pricing: A New Approach. *The Journal of Finance*, XLVIII(4):1231–1261, September 1993.
- [20] J.R. Barber and M.L. Copper. Immunizing Using Principal Component Analysis. *The Journal of Portfolio Management*, pages 99–105, Fall 1996.
- [21] M. Baxter. General Interest Rate Models and the Universality of HJM. In M.A.H. Dempster and S.R. Pliska, editors, *Mathematics of Derivative Securities*, pages 315–335. Cambridge University Press, Cambridge, England, 1997.
- [22] G.O. Bierwag, G.G. Kaufman, R. Schweitzer, and A. Toevs. The Art of Risk Management in Bond Portfolios. *Journal of Portfolio Management*, 7(3):27–36, 1981.
- [23] G.O. Bierwag, G.G. Kaufman, and A. Toevs. Duration Analysis and Immunization. In G.O. Bierwag, G.G. Kaufman, and A. Toevs, editors, *Innovations in Bond Portfolio Management*. JAI Press Inc., Greenwich, Connecticut, 1983b.
- [24] T. Bjork, G. Di Masi, Y. Kabanov, and W. Runggaldier. Towards a General Theory of Bond Markets. *Finance and Stochastics*, 1:141–174, 1997.
- [25] T. Bjork, Y. Kabanov, and W. Runggaldier. Bond Market Structure in the Presence of Market Point Processes. *Mathematical Finance*, 7:211–239, 1997.

- [26] T. Bjork and L. Svensson. On the Existence of Finite Dimensional Realizations for Nonlinear Forward Rate Models. Working paper, Department of Finance, Stockholm School of Economics, 1999.
- [27] F. Black, E. Derman, and W. Toy. A One-Factor Model of Interest Rates and Its Application to Treasury Bond Options. *Financial Analysts Journal*, pages 33–39, Jan/Feb 1990.
- [28] F. Black and P. Karasinski. Bond and Option Pricing When Short Rates are Lognormal. *Financial Analysts Journal*, pages 52–59, July/Aug 1991.
- [29] S. Bochner. *Vorlesungen Über Fouriersche Integrale*. Akademische Verlagsgesellschaft, Leipzig, 1932.
- [30] S. Bochner. Monotone Funktionen, Stieltjes Integrale and Harmonisch Analyse. *Math. Ann.*, 108:378–410, 1933.
- [31] C. de Boor. Bicubic Spline Interpolation. *J. Math. Phys.*, 41:215, 1962.
- [32] M. Bossy, R. Gibson, F.S. Lhabitant, N. Pistre, and D. Talay. Model Risk Analysis For Bond Options in a Heath-Jarrow-Morton Framework. Report, RiskLab, 1998.
- [33] J.P. Boyle and R.L. Dykstra. A Method for Finding Projections onto the Intersection of Convex Sets in Hilbert Spaces. In *Advances in Order Restricted Inference*, volume 37 of *Lecture Notes in Statistics*, pages 28–47. Springer-Verlag, 1985.
- [34] P.P. Boyle. Immunization Under Stochastic Models of the Term Structure. *Journal of the Institute of Actuaries*, 105(2):177–187, 1978.
- [35] A. Brace. Non-Bushy Trees for Gaussian HJM and Lognormal Forward Models. Working paper, UNSW, Australia, 1996a.
- [36] A. Brace. Dual Swap and Swaption Formulae in the Normal and Lognormal Models. Working paper, School of Mathematics, UNSW, Australia, 1996b.
- [37] A. Brace. Rank-2 Swaption Formulae. Working paper, School of Mathematics, UNSW, Australia, 1997.
- [38] A. Brace. Simulation in HJM and LFM models. Working paper, School of Mathematics, UNSW, Australia, 1998.
- [39] A. Brace, D. Gatareck, and M. Musiela. The Market Model of Interest Rate Dynamics. *Mathematical Finance*, 7:127–154, 1997.
- [40] A. Brace and M. Musiela. A Multifactor Gauss Markov Implementation of Heath, Arrow, and Morton. *Mathematical Finance*, 4:259–283, 1994.

- [41] A. Brace and M. Musiela. Durtion, Convexity and Wiener Chaos. Technical report, University of New South Wales, 1997.
- [42] A. Brace, M. Musiela, and Schlögl. A Simulation Algorithm Based on Measure Relationships in the Lognormal Market Models. Working paper, FMMA and The University of New South Wales, 1998.
- [43] A. Brace and R.S. Womersley. Exact Fit to the Swaption Volatility Matrix Using Semidefinite Programming. Working paper, National Australia Bank, 2000.
- [44] M.W. Brandt and P. Santa-Clara. Simulated Likelihood Estimation of Multivariate Diffusions with an Application to Interest Rates and Exchange Rates with Stochastic Volatility. Working paper, Whe Wharton School, University of Pennsylvania, 1999.
- [45] D.T. Breeden and R.H. Litzenberger. Prices of State-Contingent Claims Implicit in Option Pricess. *Journal of Business*, 1978.
- [46] M.J. Brennan and E.S. Schwartz. A Continuous Time Approach to the Pricing of Bonds. *Journal of Banking and Finance*, 3:133–155, 1979.
- [47] M.J. Brennan and E.S. Schwartz. Duration, Bond Pricing, and Portfolio Management. In G. Bierwag, G. Kaufman, and A. Toevs, editors, *Innovations in Bond Portfolio Management: Duration Analysis and Immunization*. JAI Press, Greenwich, CT, 1983.
- [48] D. Brigo and F Mercurio. A Mixed-up Smile. *Risk*, 2000.
- [49] D. Brigo and F Mercurio. *Interest Rate Models: Theory and Practice*. Springer, 2001a.
- [50] D. Brigo and F. Mercurio. Lognormal-Mixture Dynamics and Calibration to Market Volatility Smiles. *International Journal of Theoretical and Applied Finance*, 5(4):427–446, 2002.
- [51] D. Brigo, F. Mercurio, and F. Rapisarda. Pricing the Smile in a forward LIBOR Market Model. Working paper, Banca IMI, 2001.
- [52] S. Brown and P. Dybvig. The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates. *The Journal of Finance*, 41:617–630, July 1986.
- [53] W. Buhler, M. Uhrig, U. Walter, and T. Weber. An Empirical Comparison of Alternative Models for Valuing Interest Rate Options. Working paper, University of Mannheim, 1995.

- [54] A.N. Burnetas and P. Ritchken. On Rational Jump Diffusion Models: An Approach Using Potentiall. *Review of Derivatives Research*, 1:325–349, 1997.
- [55] A.J.G. Cairns. A Multifactor Model for the Term Structure and Inflation for Long-Term Risk Management. Working paper, Department of Actuarial Mathematics and Statistics, Heriot-Watt University, 1999.
- [56] P. Carr and G. Yang. Simulating Bermudian Interest Rate Derivatives. Working paper, Morgan Stanley, 1997.
- [57] P. Carr and G. Yang. Simulating American Bond Options in an HJM Framework. Working paper, Morgan Stanley, 1998.
- [58] A. Carverhill. When is the Short Rate Markovian. *Mathematical Finance*, 4(4):305–312, Oktober 1994.
- [59] K.C. Chan, G.A. Karolyi, F.A. Longstaff, and A.B. Sanders. An Emprical Comparison of Alternative Models of the Term Structure of Interest . *The Journal of Finance*, 47:1209–1228, 1992.
- [60] K. Chang. Strictly Positive Definite Functions. *Journal of Approximation Theory*, 87:148–158, 1996.
- [61] R.R. Chen and L. Scott. Pricing Interest Rate Options in a Two-Factor Cox-Ingersoll-Ross Model of the Term Structure. *Review of Financial Studies*, 5(4):613–636, 1992.
- [62] R.R. Chen and L. Scott. ML Estimation for a Multifactor Equilibrium Model of the Term Structure. *Journal of Fixed Income*, 3:14–31, 1993.
- [63] W. Cheney and W. Light. *A Course in Approximation Theory*. Brooks/Cole Series in Advance Mathematics. Brooks Cole, 2000.
- [64] J.H. Cochrane. *Asset Pricing*. Princeton University Press, 2001.
- [65] B. Coffey and J. Schoenmakers. LIBOR Rate Models, Related Derivatives and Model Calibration. Working paper, Weierstrass-Institute, Berlin, 1999.
- [66] B. Coffey and J. Schoenmakers. Stable implied calibration of multi-factor LIBOR Model via a Semi-Parametric Correlation Structure. Working paper, Weierstrass-Institute, Berlin, 2000.
- [67] P. Collin-Dufresne and R. Goldstein. "True" Stochastic Volatility and Generalized Affine Models of the Term Structure. Working paper, Carnegie Mellon University, 2000.
- [68] G.M. Constantinides. A Theory of the Nominal Term Structure of Interest Rates. *The Review of Financial Studies*, 5(4):531–552, 1992.

- [69] S. Coutant, E. Jondeau, and M. Rockinger. Reading PIBOR futures options smiles: The 1997 snap election. *Journal of Banking and Finance*, 25:1957–1987, 2001.
- [70] J.C. Cox. Notes on Option Pricing I: Constant Elasticity of Variance Diffusions. Technical report, Stanford University, 1975.
- [71] J.C. Cox, J.E. Ingersoll, and S.A. Ross. Duration and the Measurement of Basis Risk. *Journal of Business*, 52(1):51–61, January 1979.
- [72] J.C. Cox, J.E. Ingersoll, and S.A. Ross. A reexamination of Traditional Hypotheses about the term structure of interest rates. *Journal of Finance*, XXXVI(4):769–799, September 1981b.
- [73] J.C. Cox, J.E. Ingersoll, and S.A. Ross. An Intertemporal General Equilibrium Model of Asset Prices. *Econometrica*, 53(2):363–384, March 1985a.
- [74] J.C. Cox, J.E. Ingersoll, and S.A. Ross. A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–407, March 1985b.
- [75] T.F. Cox and M.A.A. Cox. *Multidimensional Scaling*. Chapman and Hall, second edition, 2001.
- [76] N.A. Cressie. *Statistic for Spatial Data*. Wiley, New-York, revised edition edition, 1993.
- [77] Q. Dai and K.J. Singleton. Specification Analysis of Affine Term Structure Models. *The Journal of Finance*, 55:385–407, 2000.
- [78] D. Damien, P.D. Sampson, and P. Guttorp. Bayesian Estimation of Semi-Parametric Non-Stationary Spatial Covariance Structures. *Environmetrics*, 12:161–178, 2001.
- [79] S.R. Das. A Direct Discrete-Time Approach to Poisson-Gaussian Bond Option Pricing in the Heath-Jarrow-Morton Model. *Journal of Economic Dynamics and Control*, 23:333–369, 1999a.
- [80] S.R. Das. The Surprise Element: Jumps in Interest Rate Diffusions. Working paper, Harvard University and NBER, 1999b.
- [81] S.R. Das and S. Foresi. Exact Solutions for Bond and Option Prices with Systematic Jump Risk. *Review of Derivatives Research*, 1:1–24, 1996.
- [82] F. de Jong, J. Driessen, and A. Pelsser. Libor Market Models versus Swap Market Models for Pricing Interest Rate Derivatives: An Empirical Analysis. *European Finance Review*, 5:201–237, 2001.

- [83] F. Delbaen and W. Schachermayer. A General Version of the Fundamental Theorem of Asset Pricing. *Mathematische Annalen*, pages 463–520, 1994.
- [84] F. Delbaen and W. Schachermayer. Non-Arbitrage and the Fundamental Theorem of Asset Pricing: Summary of Main Results. *Proc. Symp. Appl. Mathematics*, 1997.
- [85] P. Dierckx. *Curve and Surface Fitting with Splines*. Oxford Science Publications, 1993.
- [86] H. Dillen. A Model of Term Structure of Interest Rates in an Open Economy with Regime Shifts. *Journal of International Money and Finance*, 16:795–819, 1997.
- [87] F. Döblerlein. *On Term Structure Models Generated by Semimartingales*. PhD thesis, Technische Universität Berlin, 1999.
- [88] R. Douady and M. Jeanblanc. A Rating-Based Model for Credit Derivatives. Working paper, Stochastic Finance Software, 2002.
- [89] J. Driessen, P. Klaassen, and B. Melenberg. The Performance of Multi-Factor Term Structure Models for Pricing and Hedging Caps and Swaptions. Working paper, Tilburg University, 2000.
- [90] A. Dudenhausen, E. Schlögel, and L. Schlögel. Robustness of Gaussian Hedges and the Hedging of Fixed Income Derivatives. Working paper, 1999.
- [91] D. Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton (New Jersey), 2nd edition, 1996.
- [92] D. Duffie and R. Kan. Multi-factor term structure models. *Philosoph. Trans. R. Soc. Lond. A*, 347:577–586, 1994.
- [93] D. Duffie and R. Kan. A Yield Factor Model of Interest Rates. *Mathematical Finance*, 6:379–406, 1996.
- [94] D. Duffie and K.J. Singleton. Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, 61:929–952, 1993.
- [95] R.L. Dykstra. An Algorithm for Restricted Least Square Regression. *Journal of the American Statistical Association*, 78(384):837–842, December 1983.
- [96] N. El Karoui and V. Lacoste. Multifactor Models of the Term Structure of Interest Rates. Working Paper DR95010, Ecole Supérieure des Sciences Economiques et Commerciales, 1995.

- [97] R.J. Elliott and J. Van Der Hoek. Using the Hull and White Two Factor Model in Bank Treasury Risk Management. Working paper, Department of Mathematical Sciences, University of Alberta, August 1999.
- [98] E.J. Elton, M.J. Gruber, and R. Michaely. The Structure of Spot Rates and Immunization. *Journal of Finance*, 45(2):629–642, 1990.
- [99] E.J. Elton, M.J. Gruber, and NadarmP.G. Bond Returns, Immunization and the Returning Generating Process. *Studies in Banking and Finance*, 5:125–154, 1988.
- [100] G. Engeln-Müllges and F. Uhlig. *Numerical Algorithms with C*. Springer, 1996.
- [101] L. Fisher and R.L. Weil. Coping with Risk of Interest-Rate Fluctuations. *Journal of Business*, 44(4):408–431, Oktober 1971.
- [102] M. Fisher, D. Nychka, and D. Zervos. Fitting the Term Structure of Interest Rates with Smoothing Splines. Finance and economics discussion paper, Federal Reserve Bank, 1995.
- [103] B. Flesaker and L. Hughston. Positive Interest. *Risk*, 9(1):1–4, January 1996a.
- [104] B. Flesaker and L. Hughston. Positive Interest: Foreign Exchange. In *Vasicek and Beyond*, pages 351–367. Risk Publications, 1996b.
- [105] J.F. Francis. The QR Transformation. A Unitary Analogue to the LR Transformation. *Computer J.*, 4:265–271, 332–345, 1961/62.
- [106] S.K. Gandhi and P. Hunt. Numerical Option Pricing Using Conditioned Diffusions. In M.A.H. Dempster and S.R. Pliska, editors, *Mathematics of Derivative Securities*, pages 457–472. Cambridge University Press, 1997.
- [107] R. Gibson, F.S. Lhabitant, N. Pistre, and D. Talay. Interest Rate Model Risk: an Overview. *Risk*, 1(3):37–62, 1999.
- [108] R. Gibson, F.S. Lhabitant, and D. Talay. Modeling the Term Structure of Interest Rates: a Review of the Literature. Working paper, University of Lausanne, 1999.
- [109] L.R. Goldberg. Volatility of the Short Rate in the Rational Lognormal Model. *Finance and Stochastic*, 2:199–211, 1998.
- [110] R. Goldstein. The Term Structure of Interest Rates as a Random Field. *The Review of Financial Studies*, 13:365–384, 2000.
- [111] B. Goldys, M. Musiela, and D. Sondermann. Lognormality of Rates and Term Structure Models. Working paper, University of New South Wales, 1994.

- [112] P.J. Green and B.W. Silverman. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Monographs on Statistics and Applied Probability 58. Chapman and Hall, London, 1994.
- [113] N.B. Gultekin and R.J. Rogalski. The Alternative Duration Specifications and the Measurement of Basis Risk: Empirical Tests. *Journal of Business*, 57(2):241-264, 1984.
- [114] P.R. Halmos. Positive Approximants of Operators. *Indiana University Mathematical Journal*, 21:951-960, 1972.
- [115] I. Halperin. The Product of Projection Operators. *Acta. Sci. Math. (Szeged)*, 23:96-99, 1962.
- [116] Shih-Ping Han. A Successive Projection Method. *Mathematical Programming*, 40:1-14, 1988.
- [117] M.S. Handcock and J.R. Wallis. An Approach to Statistical Spatial-Temporal Modeling of Meteorological Fields (with discussion). *Journal of American Statistical Association*, 89:368-390, 1994.
- [118] J.M. Harrison and D.M. Kreps. Martingales and Arbitrage in Multiperiod Securities Markets. *Journal of Economic Theory*, 20(30):381-408, 1979.
- [119] W.J. Heaney and P.L. Cheng. Continuous Maturity Diversification of Default-Free Bond Portfolios and a Generalization of Efficient Diversification. *The Journal of Finance*, XXXIX(4):1101-1117, 1984.
- [120] D. Heath. Term Structure Models based on Future Prices. Working paper, Carnegie Mellon University, 1998.
- [121] D. Heath, R. Jarrow, and A. Morton. Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation. *Journal of Financial Quantitative Analysis*, 25:419-440, December 1990a.
- [122] D. Heath, R. Jarrow, and A. Morton. Contingent Claim Valuation with a Random Evolution of Interest Rates. *Review of Futures Markets*, 9:54-76, 1990b.
- [123] D. Heath, R. Jarrow, and A. Morton. Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica*, 60(1):77-105, January 1992.
- [124] J.R. Hicks. *Value and Capital*. Clarendon Press, Oxford, 1939.
- [125] N.J. Higham. Computing Nearest Symmetric Positive Semidefinite Matrix. *Linear Algebra and Applications*, 103:103-118, 1988.

- [126] N.J. Higham. Matrix Nearness Problem and Applications. In M.J.C. Gover and S. Barnett, editors, *Application of Matrix Theory*. Oxford University Press, 1989.
- [127] N.J. Higham. Computing the Nearest Correlation Matrix - A problem from Finance. *IMA Journal of Numerical Analysis*, 22:329–343, 2002.
- [128] D. Hilbert. Grundzüge einer Allgemeinen Theorie der Linearen Integralgleichungen I. *Gött.Nachrichten, math.-phys. K1*, pages 49–91, 1904.
- [129] T.S. Ho. Key Rate Duration. *Journal of Fixed Income*, 2(2):29–44, 1992.
- [130] T.S.Y. Ho and Lee S-B. Term Structure Movements and Pricing Interest Rate Contingent Claims. *The Journal of Finance*, XLI(5):1011–1029, 1986.
- [131] M. Hogan and K. Weintraub. The Lognormal Interest Rate Model and Eurodollar Futures. Working paper, Citibank, 1993.
- [132] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.
- [133] J. Hull and A. White. One Factor Interest Rate Models and the Valuation of Interest Rate Derivative Securities. *Journal of Financial and Quantitative Analysis*, pages 235–254, June 1993a.
- [134] P.J. Hunt and J.E. Kennedy. Implied Interest Rate Pricing Models. *Finance and Stochastics*, 2:275–293, 1998b.
- [135] P.J. Hunt and J.E. Kennedy. *Financial Derivatives in Theory and Practice*. John Wiley and Sons Ltd, 2000.
- [136] P.J. Hunt, J.E. Kennedy, and A. Pelsser. Markov-Functional Interest Rate Models. *Finance and Stochastics*, 4:391–408, 2000.
- [137] P.J. Hunt, J.E. Kennedy, and E.M. Scott. Terminal Swap-Rate Models. Working paper, University of Warwick, 1996.
- [138] J. Ingersoll, J. Skelton, and R. Weil. Duration Forty Years Later. *Journal of Financial and Quantitative Analysis*, 13:627–650, 1978.
- [139] S.D. Jacka, K. Hamza, and F.C. Klebaner. No Arbitrage Condition in SPDE Model for Interest Rates. Preprint, The University of Melbourne, 1999.
- [140] P. Jalali and H. Kazemi. A New Approach to the Valuation of Interest Rate Derivatives: Arrow-Debreu Prices Implicit in the Term Structure of Interest Rates. Working paper, The University of Massachusetts, 1997.
- [141] J. James and N. Webber. *Interest Rate Modelling*. John Wiley and Sons, 2000.

- [142] F. Jamshidian. An Exact Bond Option Pricing Formula. *Journal of Finance*, 44:205-209, March 1989.
- [143] F. Jamshidian. Bond and Options Evaluation in the Gaussian Interest Rate Model. *Res. Finance*, 6:131-170, 1991a.
- [144] F. Jamshidian. Forward Induction and Construction of Yield Curve Diffusion Models. *The Journal of Fixed Income*, pages 62-74, June 1991b.
- [145] F. Jamshidian. A simple class of square-root interest-rate models. *Applied Mathematical Finance*, 2:61-72, 1995.
- [146] F. Jamshidian. Libor and Swap Market Models and Measures. *Finance and Stochastics*, 1:293-330, 1997.
- [147] A. Janicki and Weron.A. *Simulation of Chaotic Behavior of α -Stable Stochastic Processes*. Marcel-Dekker, New-York, 1994.
- [148] R. Jarrow and D. Madan. Option Pricing Using the Term Structure of Interest Rates to Hedge Systematic Discontinuities in Asset Returns. *Mathematical Finance*, 1995.
- [149] R.A. Jarrow and M.S. Turnbull. Delta, Gamma, and Bucket Hedging of Interest Rate Derivatives. *Applied Mathematical Finance*, 1:21-48, 1994.
- [150] A. Jeffrey. Single Factor Heath-Jarrow-Morton Term Structure Models Based on Markov Spot Interest Rate Dynamics. *Journal of Financial and Quantitative Analysis*, 30(4):619-642, December 1995.
- [151] A. Jeffrey. Duration, Convexity and Higher Order Hedging. Working paper, Yale School of Management, 2000.
- [152] Y. Jin and P. Glasserman. Equilibrium Positive Interest Rates: A Unified View. *The Review of Financial Studies*, 14(1):187-214, 2001.
- [153] Y. Johnson and P. Weigel. Construction of a Strictly Positive Correlation Function via Matrix Completion. Working paper, FORC, Warwick Business School, 2003.
- [154] M. Joshi and R. Rebonato. A Stochastic-Volatility, Displaced-Diffusion Extension of the LIBOR Market Model. Technical report, Royal Bank of Scotland Quantitative Research Centre, 2001.
- [155] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, second edition, 1991.
- [156] S. Karlin. *Total Positivity*. Stanford University Press, Stanford, 1968.

- [157] D.P. Kennedy. The Term Structure of Interest Rates as a Gaussian Random Field. *Mathematical Finance*, 4:247–258, 1994.
- [158] D.P. Kennedy. Characterizing Gausssian Models of the Term Structure of Interest Rates. *Mathematical Finance*, 7:107–118, 1997.
- [159] C. Khang. Bond Immunization When Short-Term Interest Rates Fluctuate More Than Long-Term Rates. *Journal of Financial and Quantitative Analysis*, XIV(5):1086–1090, December 1979.
- [160] T.F. Klaffky, Y.Y. Ma, and A. Nozari. Managing Yield Curve Exposure. *Journal of Fixed Income*, 2(3):39–45, 1992.
- [161] P.J. Knez, R. Litterman, and J. Scheinkman. Explorations Into Factors Explaining Money Market Returns. *The Journal of Finance*, XLIX(5):1861–1882, December 1994.
- [162] C.H. Kuan and N.J. Webber. Valuing Interest Rate Derivatives Consistent with a Volatility Smile. Working paper, University of Warwick, 1998.
- [163] O. Kurbanmuradov, K. Sabelfeld, and Schoenmakers. Lognormal Approximations to LIBOR Market Models. Working paper, Weierstrauss Institute, Berlin, 2000.
- [164] V. Lacoste. Wiener Chaos: A New Approach to Option Hedging . *Mathematical Finance*, 6(2):197–213, April 1996.
- [165] J.C. Lagarias, J.A. Reeds, M.H. Wright, and M.P. Wright. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1998.
- [166] T.C. Langetieg. A Multivariate Model of the Term Structure. *The Journal of Finance*, 35(1):71–97, March 1980.
- [167] M.L. Leibowitz, W. Krasker, and A. Nozari. Spread Duration: A New Tool for Bond Portfolio Management. Technical report, Salomon Brother Inc., 1988.
- [168] A. Li, P. Ritchken, and Sankarasubramanian. Lattice Methods for Pricing American Interest Rate Claims. *Journal of Finance*, 50:719–737, 1995a.
- [169] O. Linton, E. Mammen, J. Nielsen, and C. Tanggaard. Yield Curve Estimation by Kernel Smoothing Methods. Technical report, London School of Economics, 1999.
- [170] R. Litterman and J. Scheinkman. Common Factors Affecting Bond Returns. Report, Goldman Sachs, September 1988.

- [171] M. Loève. *Probability theory*. D. Van Nostrand Company, Inc., 3d edition, 1960.
- [172] F.A. Longstaff. A Nonlinear General Equilibrium Model of the Term Structure of Interest Rates. *Journal of Financial Economics*, 23:195–224, 1989.
- [173] F.A. Longstaff, P. Santa-Clara, and E.S. Schwartz. The Relative Valuation of Caps and Swaptions: Theory and Empirical Evidence. *The Journal of Finance*, 2001a.
- [174] F.A. Longstaff, P. Santa-Clara, and E.S. Schwartz. Throwing Away a Billion Dollars: the Cost of Suboptimal Exercise Strategies in the Swaption Market. *Journal of Financial Economics*, 62:39–66, 2001b.
- [175] F.A. Longstaff and E.S. Schwartz. Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model. *Journal of Finance*, 47:1259–1282, 1992a.
- [176] D.G. Luenberger. *Optimization by Vector Space Methods*. John Wiley and Sons, Inc., 1968.
- [177] J. Lund. Non Linear Kalman Filtering Techniques for Term Structure Models. Working paper, University of Aarhus, 1997.
- [178] F.R. Macaulay. *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields, and Stock Prices in the United States since 1856*. Columbia University Press, New York, 1938.
- [179] K.V. Mardia and C.R. Goodall. Spatial-Temporal Analysis of Multivariate Environmental Monitoring Data. In G.P. Patil and C.R. Rao, editors, *Multivariate Environmental Statistics*, pages 347–386. Elsevier Science Publishers, 1993.
- [180] K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Academic Press, New York, 1979.
- [181] B. Matérn. *Spatial Variation*. Lecture Notes in Statistics. Springer Verlag, New York, 2 edition, 1986.
- [182] L.A. McCarthy and N.J. Webber. An Icosahedral Lattice Method for Three-factor Models. Working paper, University of Warwick, 1999.
- [183] J.H. McCulloch. The Tax Adjusted Yield Curve. *Journal of Finance*, 30:811–29, June 1975.
- [184] J. Mercer. Function of Positive and Negative Type and their Connection with the Theory of Integral Equations. *Philos. Trans. Roy. Soc.*, 209:415–446, 1909.

- [185] C.A. Micchelli. Interpolation of Scattered data: Distance Matrices and Conditionally Positive Definite Functions. *Constructive Approximation*, 2:11–22, 1986.
- [186] K. Miltersen, K. Sandmann, and D. Sondermann. Closed form solutions for the Term Structure Derivatives with log-normal interest rates. *Journal of Finance*, 52:409–430, 1997.
- [187] K.R. Miltersen. An Arbitrage Theory of the Term Structure of Interest Rates. Working paper, Odense University, 1999a.
- [188] C. Munk. Stochastic Duration and Fast Coupon Bond Option Pricing in Multi-Factor Models. *Review of Derivative Research*, 3(2):157–181, 1999a.
- [189] M. Musiela. Stochastic PDEs and Term Structure Models. Working paper, The University of New South Wales, 1993.
- [190] M. Musiela. Nominal Annual Rates and Log-Normal Volatility Structures. Working paper, University of New South Wales, 1994.
- [191] M. Musiela, S.M. Turnbull, and L.M. Wakeman. Interest Rate Risk Management. *Review of Futures Markets*, 12:221–261, 1993.
- [192] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.
- [193] C.R. Nelson and A.F. Siegel. Parsimonious Modeling of Yield Curves. *Journal of Business*, 60(4):473–489, 1987.
- [194] J.A. Nielsen and K. Sandman. On the Market Model of Future Rates. Working paper, Johannes Gutenberg-University of Mainz, Germany, 2000.
- [195] F. Oberhettinger. *Fourier Transforms of Distributions and Their Inverses*. Academic Press, 1973.
- [196] B. Øksendal. *Stochastic Differential Equations*. Springer, fourth edition, 1995.
- [197] B. Øksendal. An Introduction to Malliavin Calculus with Applications to Economics. Technical report, Department of Mathematics, University of Oslo, 1997.
- [198] K. Pang. Calibration of Kennedy and Multi-Factor Gaussian HJM to Caps and Swaptions Prices. Working paper, FORC, Warwick Business School, 1997.
- [199] N.D. Pearson and T. Sun. Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, And Ross Model. *Journal of Finance*, 49:1279–1304, 1994.

- [200] M.B. Pedersen. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22:55–71, 1995.
- [201] M. Polenghi. A Kernel Approach To Valuing Defaultible Debt. Preprint, Birkbeck College, 2002.
- [202] M. Polenghi. The Correlation Structure Of Emerging Market Sovereign Spreads: A Kernel Approach. Preprint, Birkbeck College, 2002.
- [203] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 2nd edition, 1992.
- [204] P. Protter. *Stochastic Integration and Differential Equations*. Springer, 1990.
- [205] R. Rebonato. *Interest-Rate Option Models*. John Wiley and Sons, 2 edition, 1996.
- [206] R. Rebonato. On the Simultaneous Calibration of Multifactor Lognormal Interest Rate Models to Black Volatilities and to Correlation Matrix. *Journal of Computational Finance*, 2(4):5–27, Summer 1999.
- [207] R. Rebonato. *Modern Pricing Theory of Interest-Rate Derivatives*. Princeton University Press, 2002.
- [208] R. Rebonato and I. Cooper. Limitations of Simple Two-Factor Interest Rate Models. *Jour. Fin. Eng.*, pages 1–16, 1995.
- [209] R.R. Reitano. Non-Parallel Yield Curve Shifts and Immunization. *Journal of Portfolio Management*, 18(3):36–43, 1992.
- [210] R.R. Reitano. Non-Parallel Yield Curve Shifts and Stochastic Immunization. *Journal of Portfolio Management*, 22(2):71–78, 1992.
- [211] P. Ritchken and L. Sankarasubrama. Volatility Structures of Forward Rates and the Dynamics of the Term Structure. *Mathematical Finance*, 5(1):55–72, January 1995.
- [212] C. Rogers and O. Zane. Fitting Potential Models to Interest Rate and Foreign Exchange Data. In *Vasicek and Beyond*, pages 327–342. Risk Publications, 1996.
- [213] L.C.G. Rogers. Which Model for Term-Structure of Interest Rates Should One Use? In *IMA: Mathematical Finance*, volume 65, pages 93–116. Springer, Berlin Heidelberg New York, 1995.
- [214] L.C.G. Rogers. One For All. *RISK Magazine*, 10, March 1997a.

- [215] L.C.G. Rogers. The Potential Approach to the Term Structure of Interest Rates and Foreign Exchange Rates. *Mathematical Finance*, 7(2):157–176, April 1997b.
- [216] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes, and Martingales*, volume 1. Wiley Series in Probability and Mathematical Statistics, 1994.
- [217] S.A. Ross. Arbitrage Pricing Theory of Capital Asset Pricing. *Journal of Economic Theory*, 13:341–360, 1976.
- [218] W. Rudin. *Principals of Mathematical Analysis*. McGraw-Hill, 1976.
- [219] M. Rutkowsky. A Note on the Flesaker-Hughston Model of Term Structure of Interest Rates. *Appl. Math. Finance*, pages 151–163, 1997b.
- [220] P.D. Sampson and P. Guttorp. On the nonparametric estimation of covariance function. *Journal of American Statistical Association*, 87:108–119, 1992.
- [221] K. Sandmann and D. Sondermann. A Note on the Stability of Lognormal Interest Rate Models and the Pricing of Eurodollar Futures. *Mathematical Finance*, 7(2):119–125, 1997.
- [222] K. Sandmann, D. Sondermann, and K.R. Miltersen. Closed Form Term Structure Derivatives in a Heath-Jarrow-Morton Model with Log-Normal Annually Compounded Interest Rates. In *Proceedings of the Seventh Annual European Futures Research Symposium Bonn, 1994*, pages 145–165. Chicago Board of Trade, 1995.
- [223] P. Santa-Clara and D. Sornette. The Dynamics of the Forward Interest Rate Curve with Stochastic String Shocks. *The Review of Financial Studies*, 14:149–185, Spring 2001.
- [224] Z. Sasvári. *Positive Definite and Definitizable Functions*. Akademie Verlag GmbH, Berlin, 1994.
- [225] E. Schlögel and D. Sommer. On Short Rate Processes and Their Implications for Term Structure Movements. Discussion Paper B-293, Department of Statistics, University of Bonn, 1994.
- [226] E. Schlögel and D. Sommer. Factor Models and the Shape of the Term Structure. Discussion Paper B-395, Department of Statistics, University of Bonn, 1997.
- [227] A.M. Schmidt and A. O'Hagan. Bayesian inference for Nonstationary Spatial Covariance Structure via Spatial Deformations. Preprint, University of Sheffield, 2000.

- [228] I.J. Schoenberg. Metric Spaces and Completely Monotone Functions. *Ann.Math.*, 39:811–841, 1938.
- [229] D. Shepard. A Two Dimensional Interpolation Function for irregularly-spaced data. In *ACM National Conference*, pages 517–524, 1968.
- [230] L. Simar. Maximum Likelihood Estimation of a Compound Poisson Process. *The Annals of Statistics*, 4(6):1200–1209, 1976.
- [231] R.L. Smith. Estimating Nonstationary Spatial Correlations. Working paper, Cambridge University and University of North Carolina, 1996.
- [232] D. Sommer. Continuous-Time Limits in the Generalized Ho-Lee Framework under the Forward Measure. Working paper, University of Bonn, 1996.
- [233] R. Stanton. A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. *Journal of Finance*, 52, 1997.
- [234] R.C. Stapleton and M.G. Subrahmanyam. The Term Structure of Interest Rates Futures Prices. Working paper, Lancaster University, 2001.
- [235] M. Steely, J. Estimating the Gilt-Edged Term Structure: Basis-Splines and Confidence Intervals. *Journal of Business Finance and Accounting*, 1991.
- [236] J. Stewart. Positive Definite Functions and Generalizations, an Historical Survey. *Rocky Mountain Journal of Mathematics*, 6(3):409–434, 1976.
- [237] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Springer-Verlag, 1980.
- [238] S.M. Sundaresan. Continuous-Time Methods in Finance: A Review and an Assessment. *The Journal of Finance*, 2000.
- [239] M. Uhrig. *Bewertung von Zinsoptionen by stochastischer Zinsvolatilität: ein Inversionansatz*. Dissertation, Universty of Mannheim, 1995.
- [240] C.F. Van Loan. Computing Integrals Involving the Matrix Exponential. *IEEE Transactions on Automatic Control*, 33:395–404, 1978.
- [241] O. Vasicek. An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5:177–188, 1977.
- [242] O.A. Vasicek and G.H. Fong. Term Structure Modeling Using Exponential Splines. *The Journal of Finance*, 37:339–348, 1982.
- [243] J. Von Neumann. *Functional Operators. The Geometry of Ortogonal Spaces*, volume II. Princeton University Press, NJ, Princeton, N.J., 1950. This is a reprint of mimeographed lecture notes first distributed in 1933.

- [244] D.F. Waggoner. Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices. Working paper, Federal Reserve Bank of Atlanta, 1997.
- [245] G. Wahba. *Spline Models for Observational Data*. SIAM: Philadelphia, 1990.
- [246] M. Waldman. Beyond Duration: Risk Dimensions of Mortgage Securities. *Journal of Fixed Income*, 2(3):5–15, 1992.
- [247] P. Weigel. A Note on Optimal Calibration of the LIBOR Market Model to the Historic Correlations. Working paper, University of Warwick, Warwick Business School, 2003.
- [248] N. Wiener. On Factorization of Matrices. *Comment Mathematici Helvetici*, 29:97–111, 1955.
- [249] E Wong. *Stochastic Processes in Information and Dynamical Systems*, volume 1. McGraw-Hill, 1971.
- [250] L. Wu. Fast At-The-Money Calibration of the LIBOR Market Model Through Lagrange Multipliers. *The Journal of Computational Finance*, 6(2), Winter 2003.
- [251] X. Wu. A New Stochastic Duration Measure by the Vasicek and CIR Term Structure Theories. mimeo, City University of Hong Kong, 2000.
- [252] W.H. Young. A Note on a Class of symmetric functions and on a Theorem Required in the Theory of Integral Equations. *Phillos. Trans. Roy. Soc. London*, Ser. A, 209, 1909.
- [253] F.A. Yousaf. *Three Mathematical Topics from the Financial Markets*. PhD thesis, The University of Bath, 2001.
- [254] Z. Zhang and L. Wu. Optimal Low-Rank Approximation to a Correlation Matrix. *Linear Algebra and Its Applications*, 364:161–187, 2003.
- [255] C.C. Zheng. An Arbitrage-free SAINTS Model of Interest Rates. Working paper, First National Bank of Chicago, 1993.
- [256] C. Zühldorf. Extended Market Models with Affine and Quadratic Volatility. Working paper, University of Bonn, 2000.

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

A Thesis Submitted for the Degree of PhD at the University of Warwick

<http://go.warwick.ac.uk/wrap/63584>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.