

Kyriakos Chourdakis

# FINANCIAL ENGINEERING

A brief introduction using the Matlab system

Fall 2008

[theponytail.net](http://theponytail.net)



---

## Contents

<b>1 Elements of stochastic calculus .....</b>	<b>1</b>
1.1 The sample space .....	1
$\sigma$ -algebras and Borel sets .....	3
Generated $\sigma$ -algebras .....	4
$\sigma$ -algebras and information .....	5
1.2 Measures and probability .....	5
Measurable functions .....	6
Probability measures .....	7
Equivalent probability measures .....	8
Conditional probability .....	9
Expectations .....	10
Independence .....	11
1.3 Stochastic process .....	11
Filtration .....	13
Distributions of a process .....	14
1.4 Brownian motion and diffusions .....	15
Properties of the Brownian motion .....	17
The Brownian motion is a martingale .....	17
The Brownian Motion is Gaussian, Markov and continuous .....	18
The Brownian motion is a diffusion .....	19
The Brownian motion is wild .....	19
Dealing with diffusions .....	20
1.5 Stochastic differential equations .....	21
Variation processes and the Itô integral .....	21
The Stratonovich integral .....	23
Itô diffusions and Itô processes .....	23
Itô's formula .....	24
1.6 The partial differential equation approach .....	27
Generators .....	27
Stopping times .....	29
1.7 The Feynman-Kac formula .....	31

1.8	Girsanov's transformation .....	32
2	<b>The Black-Scholes world</b> .....	35
2.1	The original derivation.....	35
	The Black-Scholes assumptions .....	35
	The replicating portfolio .....	36
	Arbitrage opportunities .....	37
	The Black-Scholes partial differential equation .....	39
2.2	The fundamental theorem of asset pricing .....	40
	The fundamental theorem of asset pricing and Girsanov's theorem	40
	A second derivation of the Black-Scholes formula.....	42
	Expectation under the true measure $\mathcal{P}$ .....	43
	Expectation under the risk neutral measure $\mathcal{Q}$ .....	43
	The Feynman-Kac form .....	44
2.3	Exotic options .....	44
	Exercise timing .....	44
	Payoff structures .....	45
	Path dependence .....	46
2.4	The Greeks.....	47
	The Delta .....	49
	Dynamic Delta hedging.....	50
	Gamma.....	52
	Dynamic Delta-Gamma hedging .....	55
	Gamma and uncertain volatility .....	58
	Vega .....	61
	Dividends and foreign exchange options .....	63
2.5	Implied volatilities .....	64
2.6	Stylized facts.....	66
	Leptokurtosis .....	67
	Skewness .....	67
	Volatility features .....	68
	Price discontinuities .....	68
3	<b>Finite difference methods</b> .....	69
3.1	Derivative approximations .....	70
3.2	Parabolic PDEs .....	73
	A PDE as a system of ODEs .....	74
	The grid.....	74
	Explicit finite differences .....	76
	Stability and convergence .....	77
	Implicit finite differences .....	79
	The Crank-Nicolson and the $\theta$ -method .....	80
	Boundaries .....	81
3.3	A PDE solver in Matlab .....	82
	Plain vanilla options.....	82

---

Early exercise features .....	83
Barrier features .....	87
Computing the Greeks .....	90
3.4 Multidimensional PDEs .....	93
Finite difference approaches .....	96
Boundary conditions .....	97
Alternative direction implicit methods .....	99
3.5 A two-dimensional solver in Matlab .....	100
3.6 Extensions .....	102
<b>4 Transform methods .....</b>	<b>107</b>
4.1 The setup .....	107
Fourier transforms .....	108
Characteristic functions .....	109
The "dampened" cumulative density .....	110
4.2 Option pricing using transforms .....	111
The Delta-Probability decomposition .....	113
The Fourier transform of the modified call .....	114
4.3 An example in Matlab .....	116
The characteristic functions .....	116
Numerical Fourier inversion .....	117
4.4 Applying Fast Fourier Transform methods .....	121
FFT inversion for the probability density function .....	123
FFT inversion for European call option pricing .....	123
4.5 The fractional FFT .....	124
4.6 Adaptive FFT methods and other tricks .....	127
4.7 Summary .....	129
<b>5 Historical estimation and filtering .....</b>	<b>131</b>
5.1 The likelihood function .....	132
5.2 Properties of the ML estimators .....	134
The score and the information matrix .....	135
Consistency and asymptotic normality .....	135
Hypothesis testing and confidence intervals .....	136
5.3 Some examples .....	137
Linear ARMA models .....	139
Lévy models .....	140
5.4 Likelihood ratio tests .....	141
5.5 The Kalman filter .....	141
The filtering procedure .....	141
Maximum likelihood estimation .....	144
Generalizations and extensions .....	145
Multivariate systems .....	145
Extended Kalman filter .....	147
Unscented Kalman filter .....	149

<b>6 Volatility</b> .....	153
6.1 Some general features .....	153
Historical volatility .....	154
Implied volatility .....	155
The implied volatility surface .....	158
Two modeling approaches .....	159
6.2 Autoregressive conditional heteroscedasticity .....	159
The Arch model .....	160
The Garch model .....	161
The Garch likelihood .....	162
Estimation examples .....	164
Other extensions .....	166
Garch option pricing .....	171
Utility based option pricing .....	171
Distribution based .....	172
The Heston and Nandi model .....	173
6.3 The stochastic volatility framework .....	173
The Hull and White model .....	175
The Stein and Stein model .....	175
The Heston model .....	176
Girsanov's theorem and option pricing .....	177
Example: The Heston model .....	180
The PDE approach .....	180
The Feynman-Kac link .....	183
Example: The Heston model .....	184
Estimation and filtering .....	185
Calibration .....	185
Calibration example .....	187
6.4 The local volatility model .....	191
Interpolation methods .....	192
Implied densities .....	197
Local volatilities .....	198
<b>7 Fixed income securities</b> .....	203
7.1 Yields and compounding .....	203
7.2 The yield curve .....	205
The Nelson-Siegel-Svensson parametrization .....	205
The dynamics of the yield curve .....	207
The forward curve .....	208
7.3 The short rate .....	210
Short rate and bond pricing .....	211
The hedging portfolio .....	212
The price of risk .....	214
7.4 One-factor short rate models .....	215
The Vasicek model .....	216

---

Lognormal models .....	218
The CIR model .....	218
7.5 Models with time varying parameters .....	220
The Ho-Lee model .....	221
The Hull-White model .....	222
Interest rate trees .....	223
Calibration of interest rate trees .....	223
The first stage .....	223
The second stage .....	226
Pricing and price paths .....	228
The Black-Karasinski model .....	229
Calibration issues .....	233
7.6 Multi-factor models .....	235
Factors and principal component analysis .....	237
Kalman filtering .....	241
A multi-factor Gaussian example .....	243
7.7 Forward rate models .....	243
Calibration of HJM models .....	246
Short versus forward rate models .....	247
7.8 Bond derivatives .....	247
The Black-76 formula .....	247
7.9 Changes of numéraire .....	249
7.10 The Libor market model .....	250
8 Credit risk .....	251
A Using Matlab with Microsoft Excel .....	253
A.1 Setting up Matlab with the C/C++ compiler .....	254
A.2 Writing the Matlab functions .....	255
A.3 Writing the VBA code .....	257
A.4 The Excel add-in .....	259
A.5 Invoking and packaging .....	259
References .....	263
Index .....	271



---

## Figures

1.1	Construction of a Brownian motion trajectory.....	16
1.2	Zooming into a Brownian motion sample path.....	20
1.3	A sample path of an Itô integral. ....	26
1.4	Asset price trajectories.....	27
2.1	Behavior of a Call option Delta .....	50
2.2	Dynamic Delta hedging .....	53
2.3	Sample output of the dynamic Delta hedging procedure. A call option is sold at time $t = 0$ and is subsequently Delta hedged to maturity .....	54
2.4	Behavior of a call option Gamma .....	55
2.5	Dynamic Delta-Gamma hedging .....	56
2.6	Histograms for Delta and Delta-Gamma hedging .....	57
2.7	Delta hedging with uncertain volatility .....	61
2.8	Behavior of a call option Vega .....	63
3.1	Finite difference approximation schemes. ....	71
3.2	A two-dimensional grid.....	75
3.3	The Explicit FDM.....	76
3.4	The Implicit FDM.....	79
3.5	The Crank-Nicolson FDM.....	80
3.6	Early exercise region for American options.....	85
3.7	European versus American option prices. ....	89
3.8	Greeks for American and European puts. ....	94
3.9	Oscillations of the Greeks in FDM. ....	95
3.10	The structure of the Q-matrix that approximates a two-dimensional diffusion.....	96
3.11	Two-dimensional PDE grid.....	101
4.1	“Damping” the transform. ....	112
4.2	Finite difference approximation schemes. ....	115

4.3	Numerical Fourier inversion using quadrature .....	118
4.4	Normal density function using Fourier inversion .....	120
4.5	Normal inverse Gaussian density function using Fourier inversion	121
4.6	Comparison of the FFT and the fractional FFT .....	127
5.1	Example of likelihood functions .....	132
5.2	Bias and asymptotic normality .....	138
5.3	Kalman filtering .....	146
6.1	Historical volatility .....	154
6.2	Mixtures of normals .....	155
6.3	The VIX volatility index.....	156
6.4	Implied volatility surface .....	158
6.5	Filtered volatility for DJIA and SPX.....	166
6.6	Calibrated option prices for Heston's model.....	189
6.7	The ill-posed inverse problem .....	190
6.8	Implied and local volatilities .....	194
6.9	Static arbitrage tests .....	195
7.1	Yields curves using the Nelson-Siegel-Svensson parametrization .	206
7.2	Historical yield curves .....	208
7.3	Historical Nelson-Siegel parameters .....	209
7.4	Simulation of CIR yield curves .....	220
7.5	Calibration of the Black-Karasinski model .....	231
7.6	Price path for a ten year bond .....	232
7.7	Price path for bond options .....	233
7.8	Yield curve factor loadings .....	240
7.9	Yield and one-year forward curves.....	245
7.10	Pull-to-par and bond options.....	248
7.11	Cash flows for interest rate caplets and caps.....	248
7.12	Typical Black volatilities for caplets and caps.....	249
A.1	Screenshots of the Matlab Excel Builder .....	256
A.2	The folders created by mxtool .....	257
A.3	Screenshot of the BSPricer add-in.....	262

---

## Listings

1.1	<code>bm_simul.m</code> .....	17
2.1	<code>bs_greeks.m</code> : Black-Scholes Greeks. ....	49
2.2	<code>bs_D_hedge.m</code> : Dynamic Delta hedging. ....	51
3.1	<code>G_call.m</code> : Payoff and boundaries for a call. ....	83
3.2	<code>G_put.m</code> : Payoff and boundaries for a put. ....	83
3.3	<code>pde_bs.m</code> : $\theta$ -method solver for the Black-Scholes PDE. ....	84
3.4	<code>pde_bs_impl.m</code> : Implementation of the $\theta$ -method solver. ....	85
3.5	<code>psor.m</code> : PSOR method. ....	88
3.6	<code>pde_bs_amer.m</code> : $\theta$ -method solver with early exercise. ....	88
3.7	<code>pde_bs_amer_impl.m</code> : Implementation of PSOR for an American put. ....	89
3.8	<code>pde_bs_barr.m</code> : Solver with barrier features. ....	91
3.9	<code>pde_bs_barr_impl.m</code> : Implementation for a discretely monitored barrier option. ....	92
3.10	<code>pde_bs_greeks_impl.m</code> : PDE approximations for the Greeks. ....	93
3.11	<code>callmin.m</code> : Payoff and boundaries for a two-asset option. ....	101
3.12	<code>pde_bs_2d.m</code> : Solver for a two dimensional PDE (part I). ....	103
3.13	<code>pde_bs_2d.m</code> : Solver for a two dimensional PDE (part II). ....	104
3.14	<code>pde_bs_2d_impl.m</code> : Implementation of the two dimensional solver. ....	105
4.1	<code>phi_normal.m</code> : Characteristic function of the normal distribution.	116
4.2	<code>phi_nig.m</code> : Characteristic function of the normal inverse Gaussian distribution. ....	117
4.3	<code>cf_int.m</code> : Trapezoidal integration of a characteristic function. ....	119
4.4	<code>fft_call.m</code> : Call pricing using the FFT. ....	124
4.5	<code>frft.m</code> : Fractional FFT. ....	125
4.6	<code>frft_call.m</code> : Call pricing using the FRFT. ....	126
4.7	<code>frft_integrate.m</code> : Integration over an integral using the FRFT. ....	129
4.8	<code>cf2cdf.m</code> : Transform a characteristic function into a cumulative density function. ....	130
5.1	<code>arma_sim.m</code> , <code>arma_sim.m</code> and <code>arma_sim.m</code> : Simulation and maximum likelihood estimation of ARMA models .....	137

5.2	<code>kalman_filter_1D.m</code> : One dimensional Kalman filter .....	144
5.3	<code>kalman_filter.m</code> : The $N$ -dimensional Kalman filter .....	148
5.4	<code>unscented_filter.m</code> : The unscented Kalman filter .....	151
6.1	<code>garch11_liik.m</code> : Garch likelihood function .....	163
6.2	<code>garch11_impl.m</code> : Estimation of a Garch model .....	165
6.3	<code>egarch11_liik.m</code> : Egarch likelihood function.....	168
6.4	<code>phi_heston.m</code> : Characteristic function of the Heston model .....	177
6.5	<code>ssq_heston.m</code> : Sum of squares for the Heston model. ....	188
6.6	<code>calib_heston.m</code> : Calibration of the Heston model.....	189
6.7	<code>nadwat2.m</code> : Nadaraya-Watson smoother. ....	192
6.8	<code>imp_vol.m</code> : Implied volatility surface smoothing.....	193
6.9	<code>test_vol.m</code> : Tests for static arbitrage. ....	196
6.10	<code>loc_vol.m</code> : Construction of implied densities and the local volatility surface. ....	198
7.1	<code>nelson_siegel_svensson.m</code> : Yields based on the Nelson-Siegel-Svensson parametrization.....	207
7.2	<code>calibrate_ns.m</code> : Calibration of the Nelson-Siegel formula to a yield curve.....	208
7.3	<code>hw_create.m</code> : Create Hull-White trees for the short rate. ....	224
7.4	<code>hw_path.m</code> : Compute the price path of a payoff based on a Hull-White tree for the short rate.....	228
7.5	<code>hw_path_amer.m</code> : The price path of a payoff based on the Hull-White tree when American features are present. ....	229
7.6	<code>hw_path_Impl.m</code> : Implementation of the Black-Karasinski model using a Hull-White interest rate tree. ....	230
7.7	<code>princcomp.m</code> : Correlation structure and principal component analysis of yield curve movements. ....	239
7.8	<code>kf_wrapper.m</code> : A Kalman filter wrapper for the multi-factor Gaussian model.....	244
A.1	Matlab file <code>xl_bs_call.m</code> .....	255
A.2	Matlab file <code>xl_bs_put.m</code> .....	255
A.3	VBA module ( <code>PricerMain</code> ) .....	258
A.4	VBA Activation Handlers ( <code>PricerForm</code> ) .....	259
A.5	VBA User Input Handlers I ( <code>PricerForm</code> ) .....	260
A.6	VBA User Input Handlers II ( <code>PricerForm</code> ) .....	261
A.7	VBA Add-in installation ( <code>thisWorkbook</code> ) .....	261

## Elements of stochastic calculus

In order to understand the evolution of asset prices in continuous time in general, and in the Black-Scholes framework in particular, one needs some tools from stochastic calculus. In this part we will give an overview of the main ideas from a probabilistic point of view. In the next chapter we will put these ideas in developing derivative pricing within the Black-Scholes paradigm.

Our objective at this stage is to construct stochastic processes in continuous time that have the potential to capture the probabilistic/dynamic behavior of assets. We want to be as rigorous as possible in our definitions without leaving any exploitable loopholes, but we don't want to be too abstract. The theory is covered (in increasing mathematical complexity) in [Øksendal \(2003\)](#), the two volumes of [Rogers and Williams \(1994a,b\)](#) and [Protter \(2004\)](#). Expositions that have some elements of maths relating to finance are (again in increasing complexity) [Hull \(2003\)](#), [Neftci \(2000\)](#), [Bingham and Kiesel \(2000\)](#), and [Shreve \(2004a,b\)](#)

### 1.1 THE SAMPLE SPACE

We like to think of stochastic processes (or asset prices in our case) as the outcome of an *experiment* or as the result of the *state of nature*. Each state of the world is a configuration that potentially affects the value of the stochastic process.

**Definition 1.** *We will denote the set of all states of the world with  $\Omega$  and we call it the state space or sample space. The elements  $\omega \in \Omega$  are called the states of the world, sample points or sample paths.*

Of course these states of the world are very complicated multidimensional configurations, and are typically not even numerical. In most cases they are not directly revealed to us (the observer).

**Definition 2.** *A random variable quantifies the outcome of the experiment, by mapping events to real numbers (or vectors), and this is what we actually observe.*

Therefore, a random variable  $X$  is just a function

$$X : \Omega \longrightarrow \mathbb{R}^n : \omega \longrightarrow X(\omega)$$

*Example 1.* As an example, Say that we toss a coin three times, the sample space will be the set (with  $2^3 = 8$  elements)

$$\Omega = \{HHH, HHT, HTH, \dots, TTH, TTT\}$$

This sample space is not numerical, but we can define the random variables  $X$  and  $Y$  in the following way

1.  $X = X(\omega) = \{h : h = \text{number of } H \text{ in } \omega\}$
2.  $Y = Y(\omega) = \{|h - t| : h, t = \text{number of } H, T \text{ in } \omega\}$

In the following table we summarize the possible sample space outcomes, together with the corresponding values of two different random variables  $X$  and  $Y$

$\omega$	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(\omega)$	3	2	2	2	1	1	1	0
$Y(\omega)$	3	1	1	1	1	1	1	3

Apparently, the random variable  $X$  counts the number of heads thrown, while the random variable  $Y$  counts the absolute difference between the heads and tails throws.

Of course this implies that the probabilistic behavior of the random variable will depend solely on the probabilistic behavior of the states of the world. In particular, we can write the probability (although we have not formally defined yet what a probability is)

$$\Pr[X(\omega) = x] = \Pr[\{\omega : X(\omega) = x\}]$$

In the example above the sample space was small and discrete, and for that reason the analysis was pretty much straightforward. Unfortunately, this is not usually the case, and a typical sample space is not discrete. If the sample space is continuous, expressions like  $\Pr[\omega]$  for elements  $\omega \in \Omega$  will be mostly zero, and therefore not of much interest.

Therefore, rather than assigning probabilities to *elements* of  $\Omega$  we need to assign them to *subsets* of  $\Omega$ . A natural question that follows is the following: can we really assign probabilities to *any* subset of  $\Omega$ , no matter how weird and complicated it is? The answer to this question is generally *no*. We can construct sets, like the Vitali set, for which we cannot define probabilities.<sup>1</sup> Subsets of

---

<sup>1</sup> Note that this does not mean that the probability is zero, it means that even if we assume that the probability is zero we are driven to paradoxes. In fact, the probability of such a set can not exist, and we cannot allow such a set to be considered for that purpose.

the sample space  $\Omega$  that are nice enough to allow us to define probabilities on them are called *sigma algebras*.

### $\sigma$ -ALGEBRAS AND BOREL SETS

**Definition 3.** A subset of the power set  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$  is called a  $\sigma$ -algebra on  $\Omega$  if complements and countable unions also belong to the set  $\mathcal{F}$ :

1.  $F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}$  (Complements)
2.  $F_1, F_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} F_i \in \mathcal{F}$  (Countable unions)

It turns out that  $\sigma$ -algebras are just the families of sets we need to define probabilities upon, as they are nice enough not to lead us to complications and paradoxes. Probabilities will be well defined on elements of  $\mathcal{F}$ . The elements of a  $\sigma$ -algebra are therefore called *events*.

As we will see, probabilities are just special cases of a large and very important class of set functions called *measures*. It is measures in general that are defined on  $\sigma$ -algebras. The pair  $(\Omega, \mathcal{F})$  is called a *measurable space*, to indicate the fact that it is prepared to "be measured".

*Example 2.* For a sample space  $\Omega$  there will exist many  $\sigma$ -algebras, and some will be larger than others. For the specific sample space of the previous example, where a coin is tossed three times, some  $\sigma$ -algebras that we may define are the following

1. The *minimal*  $\sigma$ -algebra is  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ . It is apparent that this is the smallest possible set that will satisfy the conditions.
2.  $\mathcal{F}_1 = \{\emptyset, \{HHH, HHT, HTH, HTT\}, \{TTH, THT, THH\}, \Omega\}$  is another  $\sigma$ -algebra on  $\Omega$ . Apparently  $\mathcal{F}_0 \subseteq \mathcal{F}_1$ .
3. The powerset of  $\Omega$  is also a  $\sigma$ -algebra,  $\mathcal{F}_{\infty} = \mathcal{P}(\Omega)$ . In fact this is the largest (or *maximal*)  $\sigma$ -algebra that we can define on a discrete set. If  $\Omega$  was continuous, say the closed interval  $\Omega = \mathbb{R}$ , the powerset is not a sigma algebra as it includes elements like the Vitali set. The largest useful  $\sigma$ -algebra that we use in this case is the *Borel*  $\sigma$ -algebra.

An example of a set that is *not* a  $\sigma$ -algebra is

$$\mathcal{E} = \{\emptyset, \{HHH\}, \{HTH\}, \{THH\}, \{TTH\}, \Omega\}$$

since the complement  $\mathcal{E} \ni \{HHH\}^c = \Omega \setminus \{HHH\} \notin \mathcal{E}$ .

We usually work with subsets of the real numbers, and, as we hinted in the previous subsection,  $\sigma$ -algebras that are defined on the real numbers (or any other Euclidean space) are very important. We saw that when we want to define a large  $\sigma$ -algebra, the powerset is not an option, since it includes pathological cases. The Borel algebra takes its place for such sets.

**Definition 4.** More formally, the Borel ( $\sigma$ -)algebra is the smallest  $\sigma$ -algebra that contains all open sets of  $\mathbb{R}$  (or  $\mathbb{R}^n$ ).

Roughly speaking, Borel sets are constructed from open intervals in  $\mathbb{R}$ , by taking in addition all possible unions, intersections and complements. We denote the Borel algebra with  $\mathcal{B} = \mathcal{B}(\mathbb{R}^n)$ .

In fact, it is very difficult to find a set that does not belong to the Borel algebra, and the ones that don't are so complicated that we cannot enumerate their elements.

### GENERATED $\sigma$ -ALGEBRAS

So far we have defined  $\sigma$ -algebras and we have shown ways to describe them by expressing some property of their elements. We can also define a  $\sigma$ -algebra based on a collection of reference subsets of  $\Omega$ .

**Definition 5.** In particular, given a family  $\mathcal{G}$  of subsets of  $\Omega$ , there is a  $\sigma$ -algebra which is the smallest one that contains  $\mathcal{G}$ . This is the  $\sigma$ -algebra generated by  $\mathcal{G}$ , and we denote it with  $\mathcal{F}(\mathcal{G}) = \mathcal{F}_{\mathcal{G}}$ .

The generated  $\sigma$ -algebra will be equal to the intersection of all  $\sigma$ -algebras that contain  $\mathcal{G}$  (since it is the smallest one with that property)

$$\mathcal{F}(\mathcal{G}) = \bigcap \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-algebra on } \Omega, \text{ and } \mathcal{G} \subset \mathcal{F}\}$$

A random variable can also create a  $\sigma$ -algebra. Given a random variable  $X$ , there is a  $\sigma$ -algebra which is the smallest one that contains the pre-image of  $X$

$$X^{-1}(G) : G \subset \mathbb{R}^n, \text{ and } G \text{ is open}$$

This is the  $\sigma$ -algebra generated by  $X$ , and it is denoted by

$$\mathcal{F}(X) = \mathcal{F}_X = \{X^{-1}(B) : B \in \mathcal{B}\}$$

*Example 3.* Following our coin example, the random variable  $X$  will generate the  $\sigma$ -algebra

$$\begin{aligned} \mathcal{F}_X = & \{\emptyset, \{HHH\}, \{HHT, HTH, THH\}, \{HTT, THT, TTH\}, \\ & \{TTT\}, \text{all complements, all unions, all intersections}\} \end{aligned}$$

It should be straightforward to verify that  $\mathcal{F}_X$

1. is a  $\sigma$ -algebra
2. contains all sets  $X^{-1}(G)$ , for  $G \in \mathcal{B}$
3. is the smallest such set

For  $Y$ , the generated  $\sigma$ -algebra is

$$\begin{aligned} \mathcal{F}_Y = & \{\emptyset, \{HHT, HTH, THH, HTT, THT, TTH\}, \\ & \{HHH, TTT\}, \text{all complements, all unions, all intersections}\} \end{aligned}$$

### $\sigma$ -ALGEBRAS AND INFORMATION

In the theory of stochastic processes  $\sigma$ -algebras are closely linked with information.

Intuitively, the generated  $\sigma$ -algebra  $\mathcal{F}_X$  captures the information we acquire by observing realizations of the random variable  $X$ . Knowing the realization  $X = x$ , allows us to decide in which element of the  $\sigma$ -algebra  $\mathcal{F}_X$  the sample point  $\omega$  belongs. The sample point  $\omega$  is the one that created the realization  $X(\omega) = x$ .

If we have two random variables  $X, Y$  on the same measure space  $(\Omega, \mathcal{FF})$ , if  $\mathcal{F}_Y \subseteq \mathcal{F}_X$  then knowing the realization of  $X$  gives us enough information to determine what the realization of  $Y$  is, without observing it directly. In particular there exists a function such that  $Y = f(X)$

If in addition the inverse set relationship does not hold, and  $\mathcal{F}_X \not\subseteq \mathcal{F}_Y$ , then this function is not invertible, and knowledge of the realization of  $Y$  does not determine  $X$  uniquely. That is to say observing  $Y$  does not offer us enough information to infer the value of  $X$ .

In the case where the  $\sigma$ -algebras are the same,  $\mathcal{F}_Y = \mathcal{F}_X$ , then the two variables contain exactly the same information: observing one is the same as observing the other.

*Example 4.* In our coin example it is easy to confirm that  $\mathcal{F}_Y \subseteq \mathcal{F}_X$ , but  $\mathcal{F}_X \not\subseteq \mathcal{F}_Y$ . This will mean that if we are given the realization of the random variable  $X$ , then we should be able to uniquely determine the realization of the random variable  $Y$ , but not vice versa.

As an example, say that we observe  $X = 2$ . Then we know that the sample point  $\omega$  that was selected from the sample space will belong to set  $F = \{HHT, HTH, THH\}$ , since only for these points  $X(\omega) = 2$ . It is easy to verify that  $F \in \mathcal{F}_X$  and also  $F \in \mathcal{F}_Y$ . In fact, for all  $\omega \in F$  we have  $Y(\omega) = 1$ . Therefore observing  $X = 2$  uniquely determines the value of  $Y = 1$ . On the other hand, say that we observe the random variable  $Y$ , and we have the realization  $Y = 3$ , indicating then the sample point selected belongs in the set  $F^* = \{HHH, TTT\}$ . Now, of course,  $F^* \in \mathcal{F}_Y$  but it does not belong to the  $\sigma$ -algebra generated by  $X$ , that is  $F^* \notin \mathcal{F}_X$ . In fact, if are given  $Y = 3$  are not given enough information to decide if  $X = 3$  or  $X = 0$ .

## 1.2 MEASURES AND PROBABILITY

In the previous section we paved the way for the introduction of the *measure* function. We introduced the sample space and the sets of its subsets that form  $\sigma$ -algebras, which are exactly the sets that are well-behaved enough to be measured. We saw that things are relatively straightforward when the sample space is discrete, and a natural  $\sigma$ -algebra is the powerset. When the sample space is continuous we have to be more careful when constructing  $\sigma$ -algebras,

as there are sets that we need to exclude (like the Vitali set). The Borel algebra is here the natural choice.

Definitions of the probability date back as far as Carneades of Cyrene (214-129BC), a prominent student at Plato's academy. More recently, Abraham De Moivre (1711) and Pierre-Simon Laplace (1812) have also attempted to formalize the everyday notion of "probability". Modern probability took off in the 1930s, largely inspired by the axiomatic foundations on measure theory by Andrei Nikolayevich Kolmogorov.<sup>2</sup>

**Definition 6.** Given a measurable space  $(\Omega, \mathcal{F})$  we can define a measure  $\mu$  as a function that maps elements of the  $\sigma$ -algebra to the real numbers,  $\mu : \mathcal{F} \rightarrow \mathbb{R}$ , and also has the following two properties:

1. the measure of the empty set is zero:  $\mu(\emptyset) = 0$
2. the measure of disjoint sets is the sum of their measures, also called  $\sigma$ -additivity:  $F_1, F_2, \dots \in \mathcal{F}$ , and  $F_i \cap F_j = \emptyset$  for all  $i \neq j \Rightarrow \mu(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} \mu(F_i)$

The measure can be thought of as the mathematical equivalent of our everyday notion of "measure", as in the length of line segments, the volume of solid bodies, the probability of events, the time needed to travel between points, and so on.

After augmenting the measurable space with a measure, the triplet  $(\Omega, \mathcal{F}, \mu)$  is called a *measure space*. The subsets of  $\Omega$  that are elements of  $\mathcal{F}$  are called *measurable sets*, indicating that they can be potentially measured by  $\mu$ . Note that we can define more than one measure on the same measurable space, creating a whole array of measure spaces  $(\Omega, \mathcal{F}, \mu_1)$ ,  $(\Omega, \mathcal{F}, \mu_2)$ , and so on.

## MEASURABLE FUNCTIONS

Based on the notion of measurable sets, we can turn to functions that map from one measurable space  $(\Omega, \mathcal{F})$  to another measurable space  $(\Psi, \mathcal{G})$ . Functions that have the property that their pre-images of measurable sets in the destination measure space  $\Psi$  are also measurable sets in the departure set  $\Omega$ .

**Definition 7.** A function  $f$  that maps from a measure space  $(\Omega, \mathcal{F})$  to  $(\Psi, \mathcal{G})$

$$f : \Omega \longrightarrow \Psi$$

is a  $(\mathcal{F}, \mathcal{G})$ -measurable function if

$$\text{for all } G \in \mathcal{G} \text{ we have } f^{-1}(G) \in \mathcal{F}$$

---

<sup>2</sup> In the words of Kolmogorov: "The theory of probability as [a] mathematical discipline can and should be developed from axioms in exactly the same way as geometry and algebra."

If the function  $f$  maps from  $\Omega$  to the Euclidean space  $\Psi = \mathbb{R}^n$ , augmented with the Borel  $\sigma$ -algebra  $\mathcal{G} = \mathcal{B}(\mathbb{R}^n)$ , then we call the function just  $\mathcal{F}$ -measurable (shortened to  $\mathcal{F}$ -meas)

A random variable  $X$  is indeed a function  $X : \Omega \rightarrow \mathbb{R}^n$ , therefore we can talk of *measurable random variables*. In particular, by the definition of the generated  $\sigma$ -algebra, a random variable will always be measurable to the  $\sigma$ -algebra it generates,  $X$  is  $\mathcal{F}_X$ -meas.

If in addition  $f$  maps to a Euclidean space  $\Omega = \mathbb{R}^m$ , also augmented with the corresponding Borel algebra,  $\mathcal{F} = \mathcal{B}(\mathbb{R}^m)$ , then the function is called just *measurable*.

*Example 5.* In our coin example we defined two random variables,  $X$  and  $Y$ , from the sample space of three coin tosses

$$X, Y : \Omega \longrightarrow \mathbb{R}$$

Each one of these random variable will induce a measurable space on  $\Omega$ , by the  $\sigma$ -algebras it generates

$$X \rightsquigarrow (\Omega, \mathcal{F}_X), \text{ and } Y \rightsquigarrow (\Omega, \mathcal{F}_Y)$$

By construction,  $X$  is  $\mathcal{F}_Y$ -meas and  $Y$  is  $\mathcal{F}_X$ -meas. On the other hand, while  $Y$  is  $\mathcal{F}_X$ -meas,  $X$  is not a  $\mathcal{F}_Y$ -meas random variable. In terms of information  $Y$  being  $\mathcal{F}_X$ -meas means that knowing  $X$  will determine  $Y$ , but not the other way round

## PROBABILITY MEASURES

As we indicated in the last subsections, measures are the mathematical equivalent of our everyday notion of “measure”. In the context of stochastic processes we are not interested in general measures, but in a small subset: the *probability measures*.

**Definition 8.** A probability measure  $\mathcal{P}$  is just a measure on  $(\Omega, \mathcal{F})$ , with the added property that

$$\mathcal{P}(\Omega) = 1$$

The measure space  $(\Omega, \mathcal{F}, \mathcal{P})$  is called a probability space

Therefore, for a function  $P : \Omega \rightarrow \mathbb{R}$  to be a probability measure there are three requirements

1.  $\mathcal{P}(\emptyset) = 0$
2.  $\mathcal{P}(\Omega) = 1$
3. for all  $F_1, F_2, \dots \in \mathcal{F}$ , with  $F_i \cap F_j = \emptyset$  for all  $i \neq j \Rightarrow \mathcal{P}(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} \mathcal{P}(F_i)$

It is obvious that probability measures are not unique on a measurable space. Given  $(\Omega, \mathcal{F})$  we can define different probability spaces  $(\Omega, \mathcal{F}, \mathcal{P}_1)$ ,  $(\Omega, \mathcal{F}, \mathcal{P}_2)$ , and so on.

Given a probability space and a random variable  $X : \Omega \rightarrow \mathbb{R}^n$ , we can define a probability measure on the Euclidean space  $\mathbb{R}^n$  endowed with its Borel algebra,  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ , in the following way

$$\mathcal{P}_X : \mathcal{B}(\mathbb{R}^n) \longrightarrow [0, 1] : \mathcal{P}_X(B) = \mathcal{P}(X^{-1}(B))$$

It is straightforward to verify that  $\mathcal{P}_X$  is a probability measure on  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ .

It is important to remember that the probability measure is defined on events of the sample space, but it induces a probability measure on the real numbers through random variables. That means that the same random variable  $X$  can induce different probability measures on  $\mathbb{R}^n$ , based on different probability spaces. For the same  $B \in \mathcal{B}(\mathbb{R}^n)$

$$\begin{aligned} (\Omega, \mathcal{F}, \mathcal{P}) &\rightsquigarrow \mathcal{P}_X(B) = \mathcal{P}(X^{-1}(B)) \\ (\Omega, \mathcal{F}, \mathcal{Q}) &\rightsquigarrow \mathcal{Q}_X(B) = \mathcal{Q}(X^{-1}(B)) \end{aligned}$$

In practice we cannot manipulate the sample space directly, since we typically we might not even know what the sample space is. Instead, we assume that the sample space exists and a measurable space is well defined, but we work with the induced probability measure. Furthermore, with some abuse of notation, we also denote this induced measure with  $\mathcal{P}$ .

Different induced probability measures  $\mathcal{P}_X, \mathcal{Q}_X, \dots$  will be then due to different measures  $\mathcal{P}, \mathcal{Q}, \dots$  on the measurable space  $(\Omega, \mathcal{F})$ . These different measures can be associated with differences of *beliefs*, differences of *behavior*, or other issues.

For example, in finance investors are interested for the probabilistic behavior of a speculative asset price, which is a random variable  $S : \Omega \rightarrow \mathbb{R}_+$ . In a simple setting, all investors might know and agree on the true (induced) probability measure  $\mathcal{P}$ , but they might behave as if the probability measure was a different one, say  $\mathcal{Q}$ . There can be many ways that this discrepancy can be theoretically explained: we will see that it can be a consequence of the risk aversion of investors, market frictions like transaction costs or liquidity constraints, or other causes.

## EQUIVALENT PROBABILITY MEASURES

As we pointed out in the previous subsection, each random variable can induce a multitude of different probability measures. We can categorize different probability measures according to some of their properties. It turns out that the most important of these classifications is the one that looks at the sets that probability measures assign zero probability. Measures that agree on these sets are called *equivalent*.

**Definition 9.** Given two probability measures  $\mathcal{P}, \mathcal{Q}$  on a measurable space  $(\Omega, \mathcal{F})$ , we say that  $\mathcal{Q}$  is absolutely continuous with respect to  $\mathcal{P}$ , and we write  $\mathcal{P} \ll \mathcal{Q}$ , if

$$(\mathcal{P}(F) = 0 \Rightarrow \mathcal{Q}(F) = 0 \text{ for all } F \in \mathcal{F})$$

The Radon-Nikodym derivative of  $\mathcal{Q}$  with respect to  $\mathcal{P}$  is defined as

$$\mathcal{M} = \frac{d\mathcal{Q}}{d\mathcal{P}}$$

which makes sense since both  $\mathcal{P}$  and  $\mathcal{Q}$  are real-valued functions. If  $\mathcal{Q} \ll \mathcal{P}$  and  $\mathcal{P} \ll \mathcal{Q}$  the probability measures are called equivalent, and we write  $\mathcal{P} \sim \mathcal{Q}$ .

Absolute continuity implies that impossible events under  $\mathcal{P}$  will also be impossible under  $\mathcal{Q}$ . If the measures are equivalent then they agree on the subsets of  $\Omega$  that have zero probability.

## CONDITIONAL PROBABILITY

The conditional probability is one of the main building blocks of probability theory, and deals with situations where some partial knowledge about the outcome of the experiment “shrinks” the sample space.

Given a probability space  $(\Omega, \mathcal{F}, \mathcal{P})$  and two events  $A, F \in \mathcal{F}$ . If we assume that a randomly selected sample  $\omega \in A$ , we want to investigate the probability that  $\omega \in F$ . Since we know that  $\omega \in A$ , the sample space has shrunk to  $A \subseteq \Omega$ , and the appropriate sigma algebra is constructed as  $\mathcal{F}_A = \{F \in \Omega : F = G \cap A, G \in \Omega\}$ . The members of the  $\mathcal{F}_A$  are conditional events, that is to say event  $F$  is the event  $G$  conditional on event  $A$ . We denote the conditional events as  $F = G|A$ .

It is not hard to verify that  $\mathcal{F}_A$  is indeed a  $\sigma$ -algebra on  $A$ .

1. The empty set  $\emptyset = (\emptyset \cap A) \in \mathcal{F}_A$  trivially.
2. Also, for an element  $(G|A) \in \mathcal{F}_A$  the complement (in the set  $A$ )  $(G|A)^c = G^c \cap A \in \mathcal{F}_A$ , since  $G^c \in \mathcal{F}$ .
3. Finally, the countable union  $\bigcup_{i \in I} (G_i|A) = \bigcup_{i \in I} (G_i \cap A) = (\bigcup_{i \in I} G_i) \cap A \in \mathcal{F}_A$ .

Therefore  $(A, \mathcal{F}_A)$  is a measurable space.

**Definition 10.** Consider a probability space  $(\Omega, \mathcal{F}, \mathcal{P})$  and an event  $A \in \mathcal{F}$  with  $\mathcal{P}(A) > 0$ . The conditional probability is defined, for all  $F \in \mathcal{F}$ , as

$$\mathcal{P}_A(F) = \mathcal{P}(F|A) = \frac{\mathcal{P}(F \cap A)}{\mathcal{P}(A)}$$

We can verify easily that  $\mathcal{P}_A$  is a probability measure on  $(\Omega, \mathcal{F})$ , which makes  $(\Omega, \mathcal{F}, \mathcal{P}_A)$  a probability space.<sup>3</sup> For all events  $F \in \mathcal{F}$  where  $\mathcal{P}(F \cap A) = 0$ , the

---

<sup>3</sup> This is indeed an example of different probability measures defined on the same measurable space.

conditional probability  $\mathcal{P}(F|A) = 0$ . This means that these two events cannot happen at the same time.

We argued above that by conditioning on the event  $A$  we shrink the measurable space  $(\Omega, \mathcal{F})$  to the smaller measurable space  $(A, \mathcal{F}_A)$ . In fact, equipped with the measure  $\mathcal{P}_A$ , the latter becomes a probability space. It is easy to verify that  $\mathcal{P}_A$  is a probability measure on  $(A, \mathcal{F}_A)$ , since  $\mathcal{P}(A|A) = 1$ . Thus, we can claim that by conditioning on  $A$  the probability space  $(\Omega, \mathcal{F}, \mathcal{P})$  shrinks to the probability space  $(A, \mathcal{F}_A, \mathcal{P}_A)$ .

We can also successively condition on a family of events  $A_1, A_2, \dots, A_n$ . In fact, we can derive the following useful identity

$$\mathcal{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathcal{P}(A_1) \cdot \mathcal{P}(A_2 | A_1) \cdot \dots \cdot \mathcal{P}(A_n | A_1, \dots, A_{n-1})$$

Another consequence of the definition is the celebrated *Bayes' theorem*, that states that if  $F_i \in \mathcal{F}$ ,  $i \in I$  is a collection of events with  $\bigcup_{i \in I} F_i = \Omega$ , and  $A \in \mathcal{F}$  is another event, then

$$\mathcal{P}(F_\ell | A) = \frac{\mathcal{P}(F_\ell) \mathcal{P}(A | F_\ell)}{\sum_{i \in I} \mathcal{P}(F_i) \mathcal{P}(A | F_i)}$$

Bayes' theorem is extensively used to update expectations and forecasts based on new evidence as this is gathered. This is an example of the filtering problem.

## EXPECTATIONS

Given a probability space  $(\Omega, \mathcal{F}, \mathcal{P})$ , consider an  $\mathcal{F}$ -meas random variable  $X$ , and assume that the random variable is integrable

$$\int_{\Omega} |X(\omega)| \cdot d\mathcal{P}(\omega) < \infty$$

A very important quantity is the expectation of  $X$ .

**Definition 11.** The expectation of  $X$  with respect to the probability measure  $\mathcal{P}$  is given by the integral

$$E[X] = \int_{\Omega} X(\omega) \cdot d\mathcal{P}(\omega) = \int_{\mathbb{R}^n} x \cdot d\mathcal{P}_X(x)$$

The *conditional expectation* given a sub- $\sigma$ -algebra  $\mathcal{G} \subset \mathcal{F}$  is a random variable  $E[X|\mathcal{G}]$  that has the properties

1.  $E[X|\mathcal{G}]$  is  $\mathcal{G}$ -measurable
2. For all  $G \in \mathcal{G}$   $\int_G E[X|\mathcal{G}] d\mathcal{P} = \int_G X d\mathcal{P}$

The conditional expectation is a random variable, since for different  $\omega \in \Omega$  the quantity  $E[X|\mathcal{G}]$  will be different.

One can use the Radon-Nikodym derivative to compute expectations under different equivalent probability measures. In particular, expectations under  $\mathcal{Q}$  are written as

$$E_{\mathcal{Q}} X = \int_{\Omega} x d\mathcal{Q}(x) = \int_{\mathbb{R}} x \frac{d\mathcal{Q}}{d\mathcal{P}} d\mathcal{P}(x) = \int_{\Omega} x \mathcal{M}(x) d\mathcal{P}(x) = E_{\mathcal{P}}[\mathcal{M}(X)X]$$

### INDEPENDENCE

- Two events  $F_1, F_2 \in \mathcal{F}$  are *independent* if

$$P(F_1 \cap F_2) = P(F_1) \cdot P(F_2)$$

- Two  $\sigma$ -algebras  $\mathcal{F}_1, \mathcal{F}_2$  are independent if all pairs  $F_1 \in \mathcal{F}_1$  and  $F_2 \in \mathcal{F}_2$  are independent
- Two random variables  $X_1$  and  $X_2$  are independent if the corresponding generated  $\sigma$ -algebras  $\mathcal{F}_{X_1}$  and  $\mathcal{F}_{X_2}$  are independent
- If  $X$  is  $\mathcal{G}$ -measurable, then  $E[X|\mathcal{G}] = X$
- If  $X$  is independent of  $\mathcal{G}$ , then  $E[X|\mathcal{G}] = EX$
- If  $\mathcal{H} \subset \mathcal{G}$  then  $E[E[X|\mathcal{G}]|\mathcal{G}] = E[X|\mathcal{G}]$  (tower property)

## 1.3 STOCHASTIC PROCESS

Of course a random variable is sufficient if we want to describe uncertainty at a single point in time. For example, we can assume that an asset price at a future date is a random variable that depends on the state of the world on that date. But typically we are interested not only on this static profile of the asset price, but also on the dynamics that might lead there.

Therefore, by collecting a number of random variables, resembling the asset price at different times, we construct a stochastic process.

**Definition 12.** A stochastic process is a parameterized family of random variables  $\{X(t)\}_{t \in T}$ , where all random variables are defined on the same probability space  $(\Omega, \mathcal{F}, P)$

$$X_t : \Omega \longrightarrow \mathbb{R}^n$$

In our setting the subscript  $t$  denotes time, but it could well be a spatial coordinate. The set  $T$  will determine if the stochastic process is defined in continuous or in discrete time. In particular, if  $T = \{0, 1, 2, \dots\}$  then we have a discrete time processes, while if  $T = [0, \infty)$  the process is cast in continuous time.

There are two different ways to look at the realizations of a stochastic process.

1. If we fix time we have a random variable

$$\omega \longrightarrow X(t, \omega), \text{ for all } \omega \in \Omega$$

2. If we fix a state of the world  $\omega$  we have the *trajectory* or *path*

$$t \longrightarrow X(t, \omega), \text{ for all } t \in T$$

There are also different ways to denote a stochastic process, and we use the one that clarifies the way we view it at the time, for example  $X_t$ ,  $X(t, \omega)$ ,  $X_t(\omega)$ , or  $X(\omega)(t)$ .

*Example 6.* Let us revisit our coin experiment, where we flip 3 times. The state space will collect all possible outcomes

$$\Omega = \{HHH, HHT, HTH, \dots, TTT\}$$

We we define the collection of random variables

$$X(t, \omega) = \text{number of } H \text{ in the first } t \text{ throws}$$

These random variables define a stochastic process on  $T = \{0, 1, 2, 3\}$ . In this simple case we can tabulate them and keep track of its behavior for all times and sample points

$\omega$	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(0, \omega)$	0	0	0	0	0	0	0	0
$X(1, \omega)$	1	1	1	0	1	0	0	0
$X(2, \omega)$	2	2	1	1	1	1	0	0
$X(3, \omega)$	3	2	2	2	1	1	1	0

We can fix time, say  $t = 2$ , and concentrate on the random variable  $X_2(\omega)$  which is given by the horizontal slice of the table above

$\omega$	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(2, \omega)$	2	2	1	1	1	1	0	0

Alternatively we can fix the sample point, say  $\omega = THT$ , and concentrate on the function  $X_t(THT)$

$t$	0	1	2	3
$X(t, THT)$	0	0	1	1

*Example 7.* On the same probability space we can define another stochastic process, say  $Y(t)$ , where  $Y(t, \omega) = 1$  if we roll an even number of  $H$  up to time  $t$ , 0 otherwise (where zero is considered an even number). In that case the possible values of the process are given in the following table

$\omega$	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$Y(0, \omega)$	1	1	1	1	1	1	1	1
$Y(1, \omega)$	0	0	0	1	0	1	1	1
$Y(2, \omega)$	1	1	0	0	0	0	1	1
$Y(3, \omega)$	0	1	1	1	0	0	0	1

## FILTRATION

We discussed in the previous section how  $\sigma$ -algebras can be associated with information. In particular, we noted that if a random variable  $X$  is  $\mathcal{F}$ -measurable, then we can determine the value of  $X(\omega)$  without knowing the exact value of the sample point  $\omega$ , but by merely knowing in which sets  $F \in \mathcal{F}$  the sample point belongs to. In the context of stochastic processes information changes: typically information is accumulated and a filtration is defined, but sometimes information can also be destroyed. Therefore, the sigma algebra with respect to the random variables  $X(t)$  are measurable must evolve to reflect that.

**Definition 13.** Consider a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . A filtration is a collection of non-decreasing  $\sigma$ -algebras on  $\Omega$

$$\mathbb{F} = \{\mathcal{F}_t\}_{t \in T} \text{ with } \mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2} \text{ for all } t_1, t_2 \in T, t_1 \leq t_2$$

where of course  $\mathcal{F}_t \subseteq \mathcal{F}$  for all  $t \in T$

The quadruple  $(\Omega, \mathbb{F}, \mathcal{F}, \mathbb{P})$  is called a filtered space.

A stochastic process  $X_t$  is called *adapted* (or  $\mathcal{F}_t$ -adapted if the filtration is ambiguous) if all random variables  $X_t$  are  $\mathcal{F}_t$ -measurable. In the previous section we discussed how a random variable generates a  $\sigma$ -algebra which keeps the information gather by observing the realization of this random variable. Here each collection of random variables  $\{X_s\}_{s \leq t}$  will generate a  $\sigma$ -algebra (for each  $t \in T$ ). This collection of  $\sigma$ -algebras is called the *natural filtration* of the stochastic process  $X_t$ .

We denote this filtration by  $\mathcal{F}_t = \sigma(X_s : 0 \leq s \leq t)$ , and in fact it is the smallest filtration that makes  $X_t$  adapted. It represents the accumulated information we gather by observing the process  $X_t$  up to time  $t$ . Note that this is in fact different from the  $\sigma$ -algebra generated by the random variable  $X_t$  alone, in fact  $\mathcal{F}(X_t) \subseteq \mathcal{F}_t$ .

Intuitively when  $\omega \in \Omega$  was chosen, the complete path  $\{X_t\}_{t \in T}$  was chosen as well, but this path has not been completely revealed to us. Our information consists only of the part  $\{X_s\}_{0 \leq s \leq t}$ . Based on this information we cannot pinpoint precisely which  $\omega$  has selected, but we can tell with certainty if  $\omega$  belongs on some specific subsets of  $\Omega$  that form the natural filtration  $\mathcal{F}_t$ .

Another process  $Y_t(\omega)$  will be  $\mathcal{F}_t$ -adapted if we can ascertain with certainty the value  $Y_t$  by observing  $X_t$ . There are two ways of looking at this dependence

1. There exist functions  $\{f_t\}_{t \in T}$  such that  $Y_t = f_t(\{X_s\}_{0 \leq s \leq t})$  for all  $t$ . The value  $Y_t$  is a deterministic function of the history of  $X_t$  up to time  $t$ .
2. The natural filtration of  $Y_t$  is subsumed in the natural filtration of  $X_t$ , that is to say  $\mathcal{F}(Y_t) \subseteq \mathcal{F}(X_t)$  for all  $t \in T$ .

*Example 8.* In our coin example the  $\sigma$ -algebras generated by the random variables  $X_t$  are the following

$$\begin{aligned}\mathcal{F}(X_0) &= \{\emptyset, \Omega\} \\ \mathcal{F}(X_1) &= \{\emptyset, \{HHH, HHT, HTT, HTH\}, \{THH, THT, TTH, TTT\}, \Omega\} \\ \mathcal{F}(X_2) &= \{\emptyset, \{HHH, HHT\}, \{TTH, TTT\}, \{HTH, HTT, THH, THT\}, \\ &\quad \text{all complements, all unions, all intersections}\}\end{aligned}$$

The corresponding filtrations will include all unions, intersections and complements of the individual algebras, namely

$$\begin{aligned}\mathcal{F}_0 &= \mathcal{F}(X_0) \\ \mathcal{F}_1 &= \mathcal{F}(X_0) \otimes \mathcal{F}(X_1) \\ \mathcal{F}_2 &= \mathcal{F}(X_0) \otimes \mathcal{F}(X_1) \otimes \mathcal{F}(X_2)\end{aligned}$$

For example the set  $\{HTH, HTT\}$  does not belong in either  $\mathcal{F}(X_1)$  nor  $\mathcal{F}(X_2)$ , but it belongs in  $\mathcal{F}_2$ , since

$$\begin{aligned}\{HTH, HTT\} &= \{HHH, HHT, HTT, HTH\} \cap \{HTH, HTT, THH, THT\} \\ \text{where } &\{HHH, HHT, HTT, HTH\} \in \mathcal{F}(X_1) \\ &\{HTH, HTT, THH, THT\} \in \mathcal{F}(X_2)\end{aligned}$$

Intuitively the element represents the event “first toss is a head and second toss is a tail”. Since  $X_t$  measures the number of heads, this event cannot be decided upon by just observing  $X_1$  or by just observing  $X_2$ , but it can be deduced by observing both. In particular, it is equivalent with the event “one head up to time  $t = 1$ ” (the event in  $\mathcal{F}(X_1)$ ) and (intersection) “one head up to time  $t = 2$ ” (the event in  $\mathcal{F}(X_2)$ ).

## DISTRIBUTIONS OF A PROCESS

Based on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  we can define the finite-dimensional distributions of the process  $X_t$ . For any collection of times  $\{t_i\}_{i=1}^m$ , and Borel events  $\{F_i\}_{i=1}^m$  in  $\mathcal{B}(\mathbb{R}^n)$ , the distribution

$$\mathbb{P}(X_{t_1} \in F_1, X_{t_2} \in F_2, \dots, X_{t_m} \in F_m)$$

characterizes the process and determines many important (but not all) properties. The inverse question is of importance too: Given a set of distributions, is there a stochastic process that exhibits them?

*Kolmogorov's extension theorem* gives an answer to that question. Suppose that for all  $\ell = 1, 2, \dots$ , and for all finite set of times  $\{t_i\}_{i=1}^\ell$  in  $T$ , we can provide a probability measure  $\mu_{t_1, t_2, \dots, t_\ell}$  on  $(\mathbb{R}^{n\ell}, \mathcal{B}(\mathbb{R}^{n\ell}))$  that satisfies the following consistency conditions

1. For all Borel sets  $\{F_i\}_{i=1}^\ell$  in  $\mathbb{R}^n$  the recursive extension

$$\mu_{t_1, t_2, \dots, t_\ell}(F_1 \times F_2 \times \dots \times F_\ell) = \mu_{t_1, t_2, \dots, t_\ell, t_{\ell+1}}(F_1 \times F_2 \times \dots \times F_k \times \mathbb{R})$$

2. For all Borel sets  $\{F_i\}_{i=1}^\ell$  in  $\mathbb{R}^n$ , and for all permutations  $\rho$  on the set  $\{1, 2, \dots, \ell\}$

$$\begin{aligned}\mu_{t_{\rho(1)}, t_{\rho(2)}, \dots, t_{\rho(\ell)}}(F_1 \times F_2 \times \dots \times F_\ell) \\ = \mu_{t_1, t_2, \dots, t_\ell}(F_{\rho^{-1}(1)} \times F_{\rho^{-1}(2)} \times \dots \times F_{\rho^{-1}(\ell)})\end{aligned}$$

Then, there exists a probability space  $(\Omega, \mathcal{F}, \mathcal{P})$  and a stochastic process  $X_t$  from  $\Omega$  to  $\mathbb{R}^n$ , which has the measures  $\mu$  as its finite distributions, that is to say for all  $\ell = 1, 2, \dots$

$$\mathcal{P}(X_{t_1} \in F_1, X_{t_2} \in F_2, \dots, X_{t_\ell} \in F_\ell) = \mu_{t_1, t_2, \dots, t_\ell}(F_1 \times F_2 \times \dots \times F_\ell)$$

Kolmogorov's extension theorem gives a very small set of conditions that can lead to the existence of stochastic processes. This can be very useful as we do not need to explicitly construct a process from scratch. Indeed, it is easy to prove the existence of many processes, like ones with infinitely divisible finite distributions, based on this theorem. The most important stochastic process is undoubtedly the Brownian motion.

## 1.4 BROWNIAN MOTION AND DIFFUSIONS

There are many different definitions and characterizations for the *Brownian motion*.<sup>4</sup> Here, in order to utilize Kolmogorov's extension theorem, we will define Brownian motion by invoking its transition density.

For simplicity we will only consider the one-dimensional case, but it is straightforward to see the generalization to more dimensions. We define the Gaussian transition density with parameter  $t$  for all  $x, y \in \text{Real}$ , which essentially describes the probability mass of moving from point  $x$  to  $y$  over a time interval of length  $t$

$$p(x, y; t) = \frac{1}{\sqrt{2\pi t}} \exp \left\{ -\frac{(x-y)^2}{2t} \right\}$$

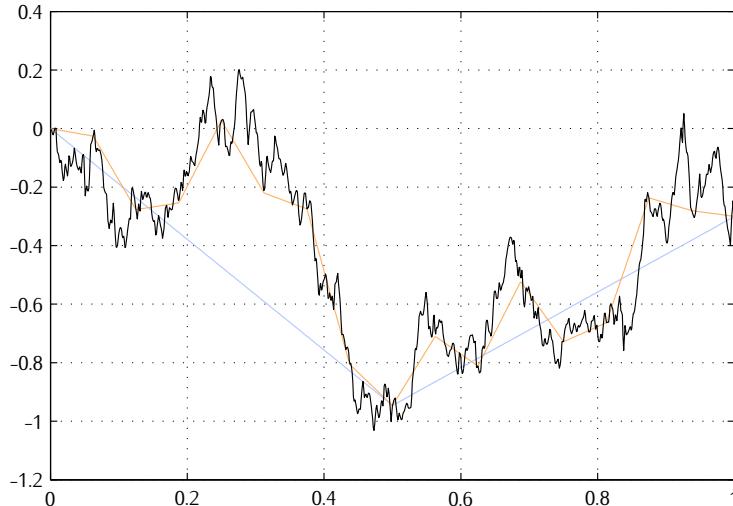
For any  $\ell = 1, 2, \dots$ , all times  $t_1, t_2, \dots, t_\ell$ , and all Borel sets  $F_1, F_2, \dots, F_\ell$  in  $\mathbb{R}$  we also define the probability measures  $\mu$  on  $\mathbb{R}^\ell$  in the following way

$$\begin{aligned}\mu_{t_1, t_2, \dots, t_\ell}(F_1 \times F_2 \times \dots \times F_\ell) \\ = \int_{F_1 \times F_2 \times \dots \times F_\ell} \cdots \int p(x, x_1; t_1) \cdot p(x_1, x_2; t_2 - t_1) \cdots \\ \cdot p(x_{\ell-1}, x_\ell; t_{\ell-1} - t_\ell) \cdot dx_1 dx_2 \cdots dx_\ell\end{aligned}$$

---

<sup>4</sup> The name Brownian motion is in honor of the botanist Robert Brown (1773–1858) who did extensive botanic research in Australia and observed the random movements of particles within pollen, the first well documented example of Brownian motion. The stochastic process is also called Wiener process, in honor of Norbert Wiener (1894–1964) who studied extensively the properties of the process.

FIGURE 1.1: Construction of a Brownian motion trajectory.



It is easy to verify that the assumptions of Kolmogorov's extension theorem are satisfied, which means that there exists a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , and a process with Gaussian increments, which we denote with  $\{B_t\}_{t \geq 0}$  and we define as a Brownian motion [BM] (started at  $B_0 = x$ )

We can also construct the Brownian motion directly, as the infinite sum of random "tent shaped" functions. To this end we will need an infinite collection of standard normal random variables  $B_k^n$ , for all natural numbers  $n = 0, 1, \dots$ , and for all odd numbers  $k$ , where  $k \leq 2^n$ . We need to define the auxiliary function  $g_k^n(u)$  which are piecewise constant

$$g_1^0 = 1, \quad g_k^n(u) = \begin{cases} 2^{(n-1)/2} & \text{for } 2^{-n}(k-1) < u \leq 2^{-n}k \\ -2^{(n-1)/2} & \text{for } 2^{-n}k < u \leq 2^{-n}(k+1) \\ 0 & \text{elsewhere} \end{cases}$$

The "tent shaped" functions ( $f_k^n$  for each  $n$  and  $k$ ) are the following integrals over the interval  $[0, 1]$

$$f_k^n(t) = \int_0^t g_k^n(u) du$$

Finally, the Brownian motion is defined as the sum over all appropriate  $n$  and  $k$ , that is to say

$$B(t) = \sum_{n=0}^{\infty} \sum_{\substack{k \text{ odd} \\ k \leq 2^n}} f_k^n(t) \cdot B_k^n$$

This is of course a function  $B : [0, 1] \rightarrow \mathbb{R}$ . Essentially, at each level  $n$  a finer refinement is added to the existing function, with an impact which falls as  $2^{-n}$

LISTING 1.1: bm\_simul.m

```
% bm_simul.m
function [x, y] = bm_simul(nMax)
x = 0:0.001:1'; % the function support
y = zeros(length(x), nMax+1);
5 y0 = 0;
for n = 0:nMax
    kMax = 2^n; % max number of tents
    for k = 1:2:kMax % loop through odd ones
        y0 = y0 + f(x, n, k)*randn;
    end
    y(:,n+1) = y0;
end

function y = f(u, n, k)
15 if n==0 % the 45o line for n = 0
    y = u;
    return;
end
% for n <> 0
20 d = 2^(-n);
a1 = (u<=k*d).*(u>(k-1)*d);
a2 = (u<=(k+1)*d).*(u>k*d);
y1 = 2^(0.5*n-0.5)*(u - (k-1)*d);
y2 = 2^(0.5*n-0.5)*((k+1)*d - u);
25 y = (a1.*y1) + (a2.*y2);
```

as the functions  $g_k^n$  show. Listing 1.1 shows how the function is implemented (a function call gives the support of the Brownian motion over  $[0, 1]$  as the vector  $x$ , and also the first  $nMax$  levels of approximation as the columns of the matrix  $y$ ). The construction of the Brownian motion in this way is given schematically in figure 1.1, where the construction for levels  $n = 2$ ,  $n = 5$  and  $n = 10$  are illustrated.

## PROPERTIES OF THE BROWNIAN MOTION

Having defined and constructed the Brownian motion, we now turn to investigating some important properties. We will assume that  $\{B_t\}_{t \geq 0}$  is a Brownian motion and  $\mathcal{F}_t$  is its natural filtration.

### The Brownian motion is a martingale

Given a filtered space  $(\Omega, \mathbb{F}, \mathcal{F}, \mathbb{P})$ , an stochastic process  $X_t$  is a martingale if

1.  $X_t$  is adapted to the filtration

for copies, comments, help etc. visit <http://www.theponytail.net/>

2. The process is integrable,  $E[|X_t|] < \infty$ , for all  $t \geq 0$
3. The conditional expectation  $E[X_s | \mathcal{F}_t] = X_t$ , for all  $s \geq t \geq 0$

It is not hard to verify that the Brownian motion is a martingale with respect to its natural filtration. This means that the conditional expected increments of a Brownian motion are zero, or that the best forecast one can provide is just the current value

$$E[B_s - B_t | \mathcal{F}_t] = 0, \text{ or } E[B_s | \mathcal{F}_t] = B_t$$

Also, one can easily show that  $B_t^2 - t$  is a martingale.

*Lévy's theorem* also states the converse: Given a filtered space, if  $\{X_t\}_{t \geq 0}$  is a continuous martingale, and  $X_t^2 - t$  is also a martingale, then  $X_t$  is a Brownian motion. If we drop the second part, and we put instead that  $E[X_s^2 | \mathcal{F}_t] = \beta(s-t)$ , for an adapted function  $\beta$ , then the time-changed process  $X_{\beta(t)}$  will satisfy the requirements of Lévy's theorem. We can then conclude that every continuous martingale can be represented as a time-changed Brownian motion.

Another important martingale is the *exponential martingale* process, given by  $M_t = \exp(\theta B_t - \frac{1}{2}\theta^2 t)$  for any parameter value  $\theta \in \mathbb{R}$ .

### The Brownian Motion is Gaussian, Markov and continuous

By its definition, the Brownian motion is *Gaussian*, that is for all times  $t_1, t_2, \dots, t_\ell \in T$  the random variable  $B = (B_1, B_2, \dots, B_\ell)$  has a multi-normal distribution

A *Markov process* has the property that

$$P(X_s \in F | \mathcal{F}_t) = P(X_s \in F | \mathcal{F}(X_t))$$

for all  $s \geq t \geq 0$ , which means that the conditional distribution depends only on the latest value of the process, and not on the whole history. Remember the difference between the  $\sigma$ -algebras  $\mathcal{F}_t$  which belongs to the filtration of the process and therefore includes the history, and  $\mathcal{F}(X_t)$  which is generated by a single observation at time  $t$ . For that reason Markov process are coined "memory-less". The Brownian motion is Markov, once again by its definition.

A *Feller semigroup* is a family of linear mappings indexed by  $t \geq 0$

$$P_t : \mathcal{C}(\mathbb{R}) \longrightarrow \mathcal{C}(\mathbb{R})$$

where  $\mathcal{C}(\mathbb{R})$  is the family of continuous functions that vanish at infinity, such that

1.  $P_0$  is the identity map
2.  $P_t$  are contraction mappings,  $\|P_t\| \leq 1$  for all  $t \geq 0$
3.  $P_t$  has the *semigroup property*,  $P_{t+s} = P_t \circ P_s$  for all  $t, s \geq 0$ , and
4. The limit  $\lim_{t \downarrow 0} \|P_t f - f\| = 0$ , for all  $f \in \mathcal{C}(\mathbb{R})$

A *Feller transition density* is a density that is associated with a Feller semigroup. A Markov process with a Feller transition function is called a *Feller*

process. One can verify that the Brownian motion is indeed a Feller process, for the Feller semigroup which is defined as

$$P_t f(x) = \int_{\mathbb{R}} p(x, y; t) f(y) dy$$

Based on the Feller semigroup expectations of functions of the Brownian motion will be given by

$$\mathbb{E}[f(B_{t+s})|\mathcal{F}_t] = P_s f(B_t) = \int_{\mathbb{R}} p(B_t, y; s) f(y) dy$$

The Brownian motion is also a process that has almost surely continuous sample paths. This is due to *Kolmogorov's continuity theorem*, which states that if for all  $t \in T$  we can find constants  $\alpha, \beta, y > 0$  such that

$$\mathbb{E}|X_{t_1} - X_{t_2}|^\alpha \leq y|t_1 - t_2|^{1+\beta}, \text{ for all } 0 \leq t_1, t_2 \leq t$$

then  $X_t$  has continuous paths (or at least a version). For the Brownian motion  $\mathbb{E}|B_{t_1} - B_{t_2}|^4 = 3|t_1 - t_2|^2$  and therefore  $B_t$  will have continuous sample paths.

### The Brownian motion is a diffusion

A Markov process with continuous sample paths is called a *diffusion*. A diffusion  $\{X_t\}_{t \geq 0}$  is “characterized” by its local *drift*  $\mu$  and *volatility*  $\sigma$ . Loosely speaking, for small  $\Delta t$  we write the instantaneous drift and volatility

$$\begin{aligned} \mathbb{E}[X_{t+\Delta t} - X_t | \mathcal{F}_t] &= \mu(X_t) \cdot \Delta t + o(\Delta t) \\ \mathbb{E}[(X_{t+\Delta t} - X_t - \mu(X_t)\Delta t)^2 | \mathcal{F}_t] &= \sigma^2(X_t) \cdot \Delta t + o(\Delta t) \end{aligned}$$

If the drift and volatility is constant, the process  $X_t = \mu t + \sigma B_t$  for a Brownian motion  $\{B_t\}_{t \geq 0}$  will be a diffusion. More generally the instantaneous drift and volatility do not have to be constant, but can depend on the location  $X_t$  and the time  $t$ . Diffusions are then given as solutions to *stochastic differential equations*.

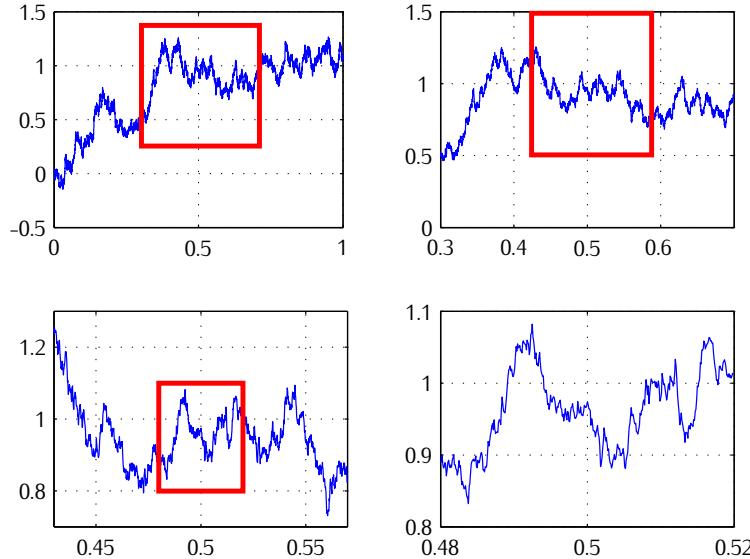
### The Brownian motion is wild

If we fix the sample point  $\omega \in \Omega$ , a Brownian motion as a function of time  $t \rightarrow B(t, \omega)$  is a lot wilder than many “normal” functions. We have shown already, using Kolmogorov’s continuity theorem, that a sample path of the Brownian motion is almost everywhere continuous, but it turns out that it is *nowhere differentiable*.

The *total variation* of a Brownian motion trajectory is unbounded, and the *quadratic variation* is non-zero.

$$\begin{aligned} \sum |B_{t_{k+1}} - B_{t_k}| &\longrightarrow \infty \\ \sum |B_{t_{k+1}} - B_{t_k}|^2 &\longrightarrow t \end{aligned}$$

FIGURE 1.2: Zooming into a Brownian motion sample path.



for partitions  $\{t_i\}$  of the time interval  $[0, t]$ , where  $\sup |t_{k+1} - t_k| \rightarrow 0$ . For “normal” functions the total variation would be the length of the curve; this means that to draw a Brownian motion trajectory on a finite interval we will need an infinite amount of ink. Also, the quadratic variation of “normal” functions is zero, since they will not be infinitely volatile in arbitrarily small intervals.

When we consider a Brownian motion path, it is impossible to find an interval that is monotonic, no matter how much we zoom in the trajectory. Therefore we cannot split a Brownian motion path in two parts with a line that is not vertical. Figure 1.2 gives a trajectory of a Brownian motion and illustrates how wild the path is by successively zooming in the process.

### DEALING WITH DIFFUSIONS

As we mentioned in the previous subsection, diffusions arise as solutions to stochastic differential equations. In finance we typically use diffusions to model factors such as stocks prices, interest rates, volatility, and others, that affect the value of financial contracts. There are three techniques for solving problems that relate to diffusions:

1. The stochastic differential equation [SDE] approach
2. The partial differential equation [PDE] approach
3. The martingale approach

All approaches are in principle interchangeable, but in practice some are more suited for particular problems. As a matter of fact, in finance we use all

three to tackle different situations. PDEs offer a “global” view of the problem in hand, while the other two approaches offer a more probabilistic “local” view.

## 1.5 STOCHASTIC DIFFERENTIAL EQUATIONS

A stochastic differential equation [SDE] resembles a normal differential equation, but some parts or some parameters are assumed random. Therefore, the solution is not a deterministic function but some sort of a generalized, “stochastic” function. The calculus of such functions is called *Itô calculus*, in honor of Kiyoshi Itô (1915–). Loosely speaking, one can represent a SDE as

$$\frac{dX_t}{dt} = \mu(t, X_t) + \text{“noise terms”}$$

The solution of such a differential equation could be represented, once again loosely, as

$$X_t = X_0 + \int_0^t \mu(s, X_s) ds + \int_0^t \text{“noise terms”} ds$$

If we write the “noise terms” in terms of a Brownian motion, say  $B_t$ , we have a process that has given drift and volatility, called an *Itô diffusion*

$$X_t = X_0 + \int_0^t \mu(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s$$

The last integral, called an *Itô integral with respect to a Brownian motion*, is not readily defined, and must clarify what we actually mean by it. Before we do so, note that we usually write the above expression in a shorthand “differential” form as

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dB_t$$

It is obvious that unlike normal (Riemann or Lebesgue), Itô integrals have a probabilistic interpretation, since they depend on a stochastic process. To give a simple motivating example, a Brownian motion can be represented as an Itô integral as

$$B_t = \int_0^t dB_s$$

### VARIATION PROCESSES AND THE ITÔ INTEGRAL

Before we turn to the definition of the Itô integral, we need to give some more information on the variation processes, some of which we have already encountered when discussing the properties of the Brownian motion. For any process  $X_t(\omega)$  we define the *p-th order variation process*, which we denote with  $\langle X, X \rangle_t^{(p)}$ , as the probability limit

$$\langle X, X \rangle_t^{(p)} = \text{plim} \sum_{t_k \leq t} |X_{t_{k+1}}(\omega) - X_{t_k}(\omega)|^p \text{ as } \Delta t_k \rightarrow 0$$

for a dyadic partition  $t_k$  of  $[0, t]$ .

Therefore, the *quadratic variation* of the Brownian motion will be the (probability) limit  $\langle B, B \rangle_t = \langle B, B \rangle_t^{(2)} = \text{plim} \sum |\Delta B_{t_k}|^2$ . We have already seen that the quadratic variation  $\langle B, B \rangle_t = t$ , since

$$\begin{aligned} \mathbb{E} \left[ \sum (\Delta B_{t_k})^2 - t \right] &= 0 \\ \mathbb{E} \left[ \sum (\Delta B_{t_k})^2 - t \right]^2 &= 2 \sum (\Delta t_k)^2 \rightarrow 0 \end{aligned}$$

We usually write the above expression in shorthand as  $[\mathrm{d}B_t(\omega)]^2 = dt$ .

If  $B_t$  was a "normal" function, then the Itô integral could be written in its Riemann sense, by using the derivative of  $B_t$

$$\int_0^t \sigma(s, X_s) dB_s = \int_0^t \sigma(s, X_s) \frac{\mathrm{d}B_u}{\mathrm{d}u} \Big|_{u=s} \mathrm{d}s$$

Here the function  $t \rightarrow B_t(\omega)$  is nowhere differentiable, and therefore we cannot express the integral in such a simple form, but we can think of it as the limit of Riemann sums. When we fix the sample point  $\omega \in \Omega$ , the random variable  $X_t = X_t(\omega)$  becomes a function over time  $t$  (albeit a wild and weird one), allowing these Riemann sums to be defined. This indicates that stochastic integrals will also be random variables, since in that sense they are mappings

$$\omega \longrightarrow \int_0^t f(s, \omega) dB_s(\omega) \in \mathbb{R}$$

For the dyadic partitions  $t_k$  of the time interval  $[0, t]$ , we define the Itô integral as the limit of the random variables

$$\omega \longrightarrow \sum_{k \geq 0} f(t_k, \omega) [B_{t_{k+1}} - B_{t_k}](\omega)$$

The limit is taken with respect to the  $L^2$ -norm  $\|f\|_2 = \sqrt{\int |f(s)|^2 ds}$ . More precisely we first define the integral for simple, step-like functions, then extend it to bounded functions  $\phi_n$ , and finally move to more general functions  $f$ , such that

1.  $(t, \omega) \rightarrow f(t, \omega)$  is  $\mathcal{B} \otimes \mathcal{F}$ -meas
2.  $f(t, \omega)$  is  $\mathcal{F}_t$ -adapted
3.  $\mathbb{E} \int_0^t f^2(s, \omega) ds < \infty$

The final property ensures that the function is  $L^2$ -integrable, and allows the required limits to be well defined using the *Itô isometry* which states

$$\mathbb{E} \left( \int_0^t f^2(s, \omega) dB_s(\omega) \right)^2 = \mathbb{E} \int_0^t f^2(s, \omega) ds$$

In particular, for a sequence of bounded functions  $\phi_n$  that converges (in  $L^2$ ) to  $f$

$$\mathbb{E} \int_0^t [f(s, \omega) - \phi_n(t, \omega)]^2 ds \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

the corresponding stochastic integrals will also converge.

### THE STRATONOVICH INTEGRAL

One important observation when computing the Itô integral is that the left endpoints of each subinterval in the partition were used to evaluate the integrand  $f$ . If the integrand was well behaved, then it would not make any difference if we took the right point or the midpoint instead, but since the integrand has infinite variation it matters.

If we used the midpoint instead, then the resulting random variable is called the *Stratonovic* stochastic integral, denoted with  $\int_0^t \sigma(s, X_s) \circ dB_s$ , which is the limit of the random variables

$$\omega \longrightarrow \sum_{k \geq 0} \frac{f(t_{k+1}, \omega) + f(t_k, \omega)}{2} [B_{t_{k+1}} - B_{t_k}](\omega)$$

It is easy to see that the Stratonovic integral is not an  $\mathcal{F}_{t_k}$ -adapted random variable, since we need to know the value of the process at the future time point  $t+k+1$ , in order to ascertain the value of the Riemann sums. For that reason it is not used as often as the Itô representation in financial mathematics,<sup>5</sup> but it has better convergence properties (due to the midpoint approximation of the integral) and it is used when one needs to simulate stochastic processes. In fact, when the process is an Itô diffusion (see below) the two stochastic integrals are related

$$\int_0^t \sigma(s, X_s) \circ dB_s = \int_0^t \sigma(s, X_s) dB_s + \frac{1}{2} \int_0^t \frac{\partial \sigma(s, X_s)}{\partial x} \sigma(s, X_s) ds$$

### ITÔ DIFFUSIONS AND ITÔ PROCESSES

Consider a Brownian motion  $B_t$  on a filtered space  $(\Omega, \mathbb{F}, \mathcal{F}, P)$ , the filtration it generates  $\{\mathcal{F}_t\}_{t \geq 0}$ , and two  $\mathcal{F}_t$ -adapted functions  $\mu$  and  $\sigma$ .

As we noted before, an *Itô diffusion* is a stochastic process on  $(\Omega, \mathcal{F}, P)$  of the form

$$X_t = X_0 + \int_0^t \mu(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s$$

We need a few conditions that ensure regularity and that solutions for the SDE exist and do not explode in finite time

---

<sup>5</sup> But it can be used for example if  $t$  is a spatial rather than a time coordinate, since then we could actually observe the complete realization in one go.

1. The Itô isometry  $E \int_0^t \sigma^2(s, X_s) ds < \infty$ , for all times  $t \geq 0$
2. There exist constants  $A, B$  such that for  $x, y \in \mathbb{R}$ ,  $s \in [0, t]$

$$|\mu(s, x)| + |\sigma(s, x)| \leq A(1 + |x|)$$

$$|\mu(s, x) - \mu(s, y)| + |\sigma(s, x) - \sigma(s, y)| \leq B(1 + |x - y|)$$

An *Itô process* is a stochastic process on the same filtered space, of the form

$$X_t = X_0 + \int_0^t \mu(s, \omega) ds + \int_0^t \sigma(s, \omega) dB_s$$

In Itô diffusions the information is generated by the Brownian motion, now we let the information to be more general, as long as the Brownian motion remains a martingale.

We consider a filtration  $\mathcal{G}_t$  that makes  $B_t$  a martingale, and assume the  $\mu$  and  $\sigma$  are  $\mathcal{G}_t$ -adapted. Instead of the integrability and isometry assumptions we need instead

$$P\left[\int_0^t |\mu(s, \omega)| ds < \infty \text{ for all } t \geq 0\right] = 1$$

$$P\left[\int_0^t \sigma^2(s, \omega) ds < \infty \text{ for all } t \geq 0\right] = 1$$

Itô processes generalize Itô diffusions in two ways.

1. Information: we can have more information than just the one we gather by observing the SDE, but this information should not make the Brownian motion predictable.
2. Dependence: drift and volatility can depend on the whole history, rather than the latest value of the process,  $X_t$ .

Unlike Itô diffusions, Itô processes are not always Markov. A diffusion  $dX = \mu(t, X)dt + \sigma(t, X)dB$  will coincide in law with a process  $dY = \mu_*(t, \omega)dt + \sigma_*(t, \omega)dB$  if

$$E^x[\mu_*(t, \omega) | \mathcal{F}_t^Y] = \mu(t, Y_t^x), \text{ and } \sigma_*^2(t, \omega) = \sigma^2(t, Y_t^x)$$

which essentially states that the process is Markov.

## ITÔ'S FORMULA

*Itô's formula* or *Itô's lemma* is one of the fundamental tools that we have in stochastic calculus. It plays the rôle that the chain rule plays in normal calculus. Just like the chain rule is used to solve ODEs or PDEs, a clever application of Itô's formula can significantly simplify a SDE. We consider an Itô process

$$dX_t = \mu(t, \omega)dt + \sigma(t, \omega)dB_t$$

A function  $g(t, x)$  in  $\mathcal{C}^{(1,2)}$  (differentiable in  $t$  and twice differentiable in  $x$ ) will define a new Itō process as the transformation  $Y_t = g(t, X_t)$ . Itō's formula describes the dynamics of  $Y_t$  in terms of the drift and volatility of  $X_t$ , and the derivatives of the transformation  $g$ . In particular, the SDE for  $Y_t$  is given by

$$dY_t = \frac{\partial}{\partial t}g(t, X_t)dt + \frac{\partial}{\partial x}g(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2}{\partial x^2}g(t, X_t)(dX_t)^2$$

The “trick” is that the square  $(dX_t)^2$  is computed using the rules

$$dt \cdot dt = dt \cdot dB_t = 0, \text{ and } (dB_t)^2 = dt$$

One can easily prove Itō's formula based on a Taylor's expansion of the function  $g$ .<sup>6</sup> In particular, one can write for  $\Delta t > 0$  the quantity  $\Delta X_t = X_{t+\Delta t} - X_t = \mu(t, X_t)\Delta t + \sigma(t, X_t)(B_{t+\Delta t} - B_t) + o(\Delta t)$ . Taking powers of the Brownian increments  $\Delta B_t = B_{t+\Delta t} - B_t$  yield

$$\begin{aligned} E\Delta B_t &= 0 \\ E(\Delta B_t)^2 &= \Delta t \\ E(\Delta B_t)^n &= o(\Delta t) \text{ for all } n \geq 3 \end{aligned}$$

This implies that the random variable  $(\Delta B_t)^2$  will have expected value equal to  $\Delta t$  and variance of order  $o(\Delta t)$ . A consequence is that in the limit  $\Delta B_t \rightarrow dB_t$ , since the variance goes to zero. Now the Taylor's expansion for  $\Delta Y_t = g(t + \Delta t, X_{t+\Delta t}) - g(t, X_t)$  will give

$$\Delta Y_t = \frac{\partial g(t, X_t)}{\partial t}\Delta t + \frac{\partial g(t, X_t)}{\partial x}\Delta X_t + \frac{1}{2}\frac{\partial^2 g(t, X_t)}{\partial x^2}(\Delta X_t)^2 + o(\Delta t)$$

Passing to the limit yields Itō's formula.

*Example 9.* Itō's formula can be used to simplify SDEs and cast them in a form that is easier to explicitly solve. Say for example that we are interested in the stochastic integral

$$\int_0^t B_s dB_s$$

where  $B_t$  is a standard Brownian motion. We will consider the function  $g(t, x) = \frac{x^2}{2}$ , and define the process  $Y_t = g(t, X_t) = \frac{1}{2}B_t^2$ . Using Itō's formula we can specify the dynamics of this process, namely

$$dY_t = 0 \cdot dt + B_t dB_t + \frac{1}{2} \cdot 1 \cdot (dB_t)^2 = B_t dB_t + \frac{dt}{2}$$

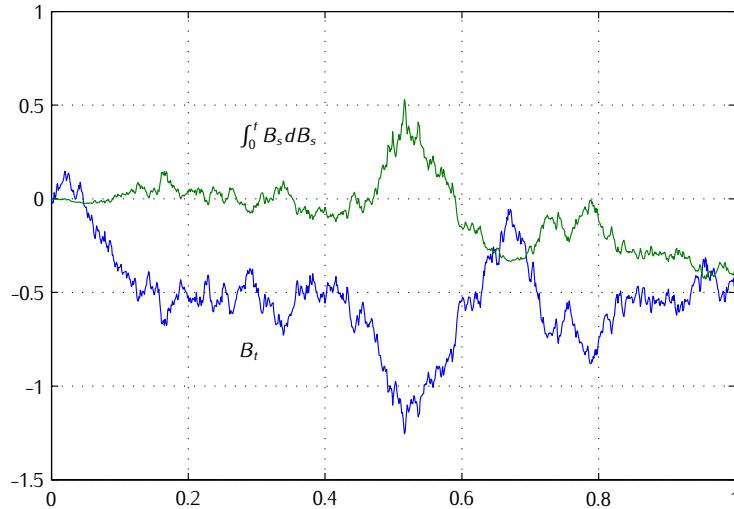
In other words we can write

$$d\left(\frac{1}{2}B_t^2\right) = \frac{1}{2}dt + B_t dB_t$$

---

<sup>6</sup> This Taylor's expansion is valid, since  $g$  is a function that is sufficiently smooth.

FIGURE 1.3: A sample path of an Itô integral.



By taking integrals of both sides, and recognizing that  $\int_0^t d\left(\frac{1}{2}B_s^2\right) = \frac{1}{2}B_t^2$ , we can solve for the Itô integral in question

$$\int_0^t B_s dB_s = \frac{1}{2}B_t^2 - \frac{1}{2}t$$

A trajectory  $B_t(\omega)$  for an element  $\omega \in \Omega$ , and the corresponding solution  $\int_0^t B_s(\omega) dB_s(\omega)$  is given in figure 1.3.

*Example 10.* The most widely used model for a stock price, say  $S_t$ , satisfies the SDE for a *geometric Brownian motion* with constant expected return  $\mu$  and volatility  $\sigma$ , given by

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

This corresponds loosely to the ODE  $\frac{dS_t}{dt} = \alpha S_t$ , which grows exponentially. Motivated by this exponential growth of the ODE we consider the logarithmic function  $g(t, x) = \log x$ . Using Itô's formula we can construct the SDE for  $s_t = \log S_t$

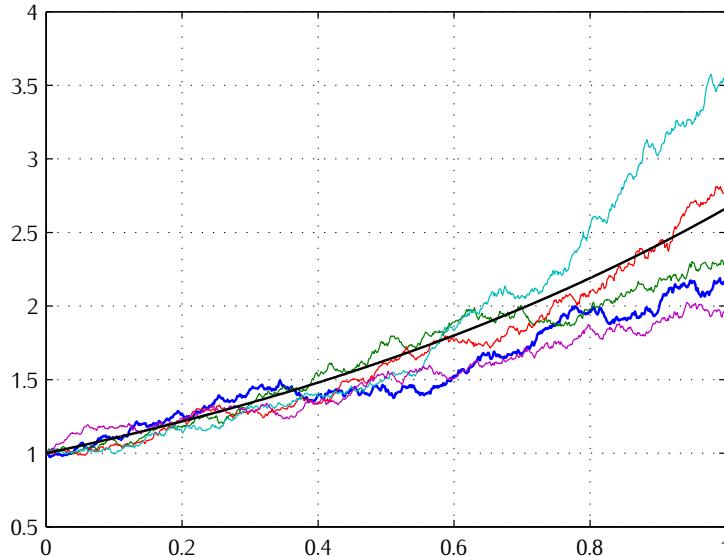
$$ds_t = \left( \mu - \frac{1}{2}\sigma^2 \right) dt + \sigma dB_t$$

This SDE has constant coefficients and can be readily integrated to give

$$s_t = s_0 + \left( \mu - \frac{1}{2}\sigma^2 \right) t + \sigma B_t$$

We can cast this expression back to the asset price itself

FIGURE 1.4: Asset price trajectories



$$S_t = S_0 \exp \left\{ \left( \mu - \frac{1}{2} \sigma^2 \right) t + \sigma B_t \right\}$$

Note that under the geometric Brownian assumption the price of the asset is always positive, an attractive feature in line with the property of *limited liability* of stocks. Some stock price trajectories for different  $\omega \in \Omega$  are given in figure 1.4.

## 1.6 THE PARTIAL DIFFERENTIAL EQUATION APPROACH

In the stochastic differential equation approach we typically consider a function of an Itô diffusion, and construct the dynamics of the process under this transformation. In many applications we are not interested in the actual paths of the process, but for some expectation of some function at a future date. In the PDE approach we want to investigate the transition mechanics of the process, and based on that the transition mechanics of the expectation in hand.

### GENERATORS

Say we are given a Brownian motion  $B_t$  on a filtered space. For a SDE that describes the motion of a stochastic process  $X_t$ , say

for copies, comments, help etc. visit <http://www.theponytail.net/>

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t$$

we can define an *elliptic operator* which is applied to twice-differentiable functions  $f \in \mathcal{C}^{(2)}$

$$\mathcal{L}f(x) = \left[ \mu(t, x) \frac{d}{dx} + \frac{1}{2} \sigma^2(t, x) \frac{d^2}{dx^2} \right] f(x)$$

This elliptic operator is also the infinitesimal generator of the process, which is given formally in the following definition.

**Definition 14.** Given an Itô diffusion  $X_t$ , the (infinitesimal) generator of the process, denoted with  $\mathcal{A}$ , is defined for all functions  $f \in \mathcal{C}^{(2)}$  as the limit

$$\mathcal{A}f(x) = \lim_{\Delta t \downarrow 0} \frac{E^x f(X_{\Delta t}) - f(x)}{\Delta t} = \mathcal{L}f(x)$$

It is *Kolmogorov's backward equation* that gives us the expectation of the function  $f$  at a future date, as the solution of a partial differential equation. In particular, if we denote the expectation with  $g(t, x) = E[f(X_T)|X_t = x]$ , for  $f \in \mathcal{C}^{(2)}$  we have

$$-\frac{\partial}{\partial t}g(t, x) = \mathcal{A}_x g(t, x) = \mu(t, x) \frac{\partial}{\partial x}g(t, x) + \frac{1}{2}\sigma^2(t, x) \frac{\partial^2}{\partial x^2}g(t, x)$$

The subscript  $x$  of the generator in the above expression just indicates that the derivatives are partial and taken with respect to  $x$ . The final condition for Kolmogorov's backward PDE will be  $g(T, x) = f(x)$ .

There is a very intuitive way of viewing Kolmogorov's equation. If we consider the expectation  $E_t = g(t, X_t) = E[f(X_T)|\mathcal{F}_t]$  as a stochastic process, then we can observe (by applying Itô's formula) that Kolmogorov's backward PDE sets the drift of  $E_t$  equal to zero, rendering  $E_t$  a martingale. This means that expectations we form at time  $t$  are unbiased, in the sense that we do not anticipate any special information that will change them.

If we use the indicator function, then the expectation becomes the conditional probability that the diffusion will take values in a set  $F$  at time  $T \geq t$ . In particular, if we denote this conditional probability with  $f(t, x; T, F) = \mathbb{P}(X_T \in F | x_t = x) = E[\mathbb{1}(X_T \in F) | x_t = x]$ , then Kolmogorov's backward PDE takes the form

$$-\frac{\partial}{\partial t}f(t, x; T, F) = \mu(t, x) \frac{\partial}{\partial x}f(t, x; T, F) + \frac{1}{2}\sigma^2(t, x) \frac{\partial^2}{\partial x^2}f(t, x; T, F)$$

for  $t \leq T$ , with a terminal condition  $f(T, x; T, F) = \mathbb{1}(x \in F)$  (which is equal to one if  $x \in F$  and zero otherwise). Therefore this PDE describes the evolution of the probability that we will end up in a certain set of states at some future time  $T$ . It is called a *backward PDE* because we start with a terminal condition at the future time  $T$  and integrate backwards to the present time  $t$ .

*Kolmogorov's forward equation* also known as *Fokker-Planck equation* considers the transition density  $p(t, x; T, y) = \mathbb{P}(X_T \in dy | x_t = x)$ . It postulates that

$$\frac{\partial}{\partial T} f(t, x; T, y) = -\frac{\partial}{\partial x} [\mu(T, x)f(t, x; T, y)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [\sigma^2(T, x)f(t, x; T, y)]$$

for  $T \geq t$ , with an initial condition  $f(t, x; t, y) = \delta(x - y)$  (the *Dirac function*). The PDE gives the evolution of the distribution of  $x_T$  given the current state  $x_t = x$ . It is called a *forward PDE* because we start from the state at the current time  $t$  and integrate forwards towards the future time  $T$ .

## STOPPING TIMES

**Definition 15.** A stopping time is a random variable

$$\tau : \Omega \rightarrow [0, \infty], \text{ such that } \{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0$$

That is to say, a stopping time is defined by an event that is  $\mathcal{F}_t$ -measurable. This means that at any time  $t \geq 0$  we can ascertain with certainty whether or not the event has happened. Examples of stopping times are the first hitting times, first exit times from a set, and so on. Given a stopping time we can define the *stopped process*, which is simply  $X_t = X_{\min(t, \tau)}$ .

Say that  $\tau$  is a stopping time with  $E^x \tau < \infty$ , meaning that the process will be stopped at some point in the future almost surely. *Dynkin's formula* gives expectations at a stopping time  $\tau$ , as

$$E^x f(X_\tau) = f(x) + E^x \int_0^\tau \mathcal{A}f(X_s) ds$$

Here,  $f \in \mathcal{C}^{(2)}$ , and also has compact support. Note that in the above integral the upper bound is a random variable. Dynkin's formula can be used to assess when a process is expected to be stopped, that is to say the expectation  $E^x[\tau(\omega)]$ .

*Example 11.* For example, say that we are holding a stock with current price  $S$ , and dynamics that follow the geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

We want to know how long should we expect to wait, before our asset will be worth at least  $\bar{S} > S$ . Mathematically, we are interested on the first exit time from the set  $[0, \bar{S}]$

$$\tau = \inf\{t \geq 0 : S_t \geq \bar{S}\}$$

We cannot directly apply Dynkin's formula, since we need  $E^s \tau < \infty$  and we are not sure about that. For example the asset might have a negative drift and exponentially drop towards zero. We can define instead the exit times from the set  $[a, \bar{S}]$

$$\tau_a = \inf\{t \geq 0 : S_t \geq \bar{S} \text{ or } S_t \leq a\}$$

The expected exit time from a compact set is indeed bounded, and therefore Dynkin's formula can now be applied.

Here it will be useful to remind us the solution of the SDE for the geometric Brownian motion

$$S_t = S_0 \exp \left\{ \left( \mu - \frac{1}{2}\sigma^2 \right) t + \sigma B_t \right\}$$

Suppose that  $\mu < \sigma^2/2$ , which means that the expected returns of the asset are not large enough for the price to be expected to grow. In this case, as  $t \rightarrow \infty$  the expectation of the asset price  $E^S S_t \rightarrow 0$ , and in fact every trajectory  $S_t \rightarrow 0$ , almost surely. Then, every trajectory will exit the set at least at lower bound  $a$ . Of course, the process might hit the upper bound  $\bar{S}$  first. For that reason, say that the probability  $P(S_{\tau_a} = a) = p_a$ , and of course  $P(S_{\tau_a} = \bar{S}) = 1 - p_a$ .

Consider the function  $f(x) = x^{1-2\mu/\sigma^2}$ . This might appear to be an odd choice, by we have selected this function because when we apply generator  $\mathcal{A}$

$$\mathcal{A}f(x) = \mu \left( 1 - \frac{2\mu}{\sigma^2} \right) x^{1-2\mu/\sigma^2} - \frac{1}{2}\sigma^2 \left( 1 - \frac{2\mu}{\sigma^2} \right) \frac{2\mu}{\sigma^2} x^{1-2\mu/\sigma^2} = 0$$

Therefore Dynkin's formula yields (for the exit times  $\tau_a$ )

$$p_a \cdot \bar{S}^{1-2\mu/\sigma^2} + (1 - p_a) \cdot a^{1-2\mu/\sigma^2} = S^{1-2\mu/\sigma^2}$$

Passing to the limit  $a \rightarrow 0$  we can retrieve the probability of never reaching our target of  $\bar{S}$ , namely  $p_a \rightarrow p = \left(\frac{S}{\bar{S}}\right)^{1-2\mu/\sigma^2}$ . This probability become higher for lower expected returns or higher volatility.

If  $\mu > \sigma^2/2$  then the expected returns are high enough for all sample paths to eventually breach the target  $\bar{S}$ , since  $S_t \rightarrow \infty$  as  $t \rightarrow \infty$ , almost surely. The process will exit with probability 1, but  $E\tau$  we might still be  $\infty$ .

In this case we consider the function  $f(x) = \log x$ . Our objective now it for the generator to be constant, in order to simplify the integral  $\int_0^{\tau_a} \mathcal{A}f(X_s) ds$ . In particular

$$\mathcal{A}f(x) = \mu - \frac{1}{2}\sigma^2$$

Dynkin's formula will yield in this case (once again for the exit times  $\tau_a$ )

$$p_a \cdot \log \bar{S} + (1 - p_a) \cdot \log a = \log S + \left( \mu - \frac{1}{2}\sigma^2 \right) E^S \int_0^{\tau} dt$$

Passing to the limit this time will give the expected stopping time

$$E^S \tau_a \rightarrow E^S \tau = \frac{\log (\bar{S}/S)}{\mu - \sigma^2/2} \text{ as } a \rightarrow 0$$

## 1.7 THE FEYNMAN-KAC FORMULA

The *Feynman-Kac formula* generalizes Kolmogorov's backward equation, and provides the connection that we need between the SDE and PDE approaches. It is named after Richard Feynman (1918–1988) and Marek Kac (1914–1984), but was published by Kac in 1949. It gives expectations not only of a functional of the terminal value of the process, but also some functionals that are computed at intermediate points.

In particular, we start with an Itô diffusion which is associated with a generator  $\mathcal{A}$ , and we also consider two functions  $f \in \mathcal{C}^{(2)}$  and  $g \in \mathcal{C}$ , continuous and lower bounded. We are interested in computing an expectation of the form

$$u(t, x) = E^x \left[ \exp \left\{ - \int_0^t g(X_s) ds \right\} \cdot f(X_t) \right]$$

The Feynman-Kac formula states that this expectation satisfies the partial differential equation

$$\frac{\partial}{\partial t} u(t, x) = \mathcal{A} u(t, x) - g(x) u(t, x)$$

with boundary condition  $u(0, x) = f(x)$ .

The Feynman-Kac formula has been very successful in financial mathematics, as it can represent stochastic discount factors through the exponential  $\exp \left\{ - \int_0^t g(X_s) ds \right\}$ .

*Example 12.* Suppose that the interest rate  $r_t$  follows an Itô diffusion given by

$$dr_t = \theta(\rho - r_t)dt + \sigma \sqrt{r_t} dB_t$$

Also suppose that we have an investment that depends on the level of the interest rate, for example a house, with value given by  $H(r)$ . This implies that the property value will also follow an Itô diffusion, with dynamics given by Itô's formula. At a future time  $T$ , the house price will be  $H(r_T)$ , which is of course unknown today.

We are interested in buying the property at time  $T$ , which means that we are interested in the present value of  $H(r_T)$ , namely

$$u(t, x) = E^x \left[ \exp \left\{ - \int_0^T r_s ds \right\} \cdot H(r_T) \right]$$

Say that we have a project with uncertain payoffs that depend on the evolution of a variable  $X_t$ , which has current value  $X_0 = x$ . The dynamics are  $dX = \mu dt + \sigma dB$ , and the project will pay  $f(X_T) = aX_T^2 + b$ . We are interested in establishing the present value

$$E^x \left[ \exp \{-RT\} \cdot (aX_T^2 + b) \right]$$

This will be equal to  $u(T, x)$ , where  $u$  satisfies the PDE

$$\frac{\partial}{\partial t} u(t, x) = \mu \frac{\partial}{\partial x} u(t, x) + \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial x^2} u(t, x) - R u(t, x)$$

with boundary  $u(0, x) = ax^2 + b$

## 1.8 GIRSANOV'S TRANSFORMATION

As we saw in section 1.2 the same measurable space  $(\Omega, \mathcal{F})$  can support different probability measures. We also saw how the Radon–Nikodym derivative can be used to compute expectations under different equivalent measures. *Girsanov's theorem* gives us the tools to specify this Radon–Nikodym derivative for Itô processes. In particular, consider an Itô process on the filtered space  $(\Omega, \{\mathcal{F}_t\}_{0 \leq t \leq T}, \mathcal{F}, \mathbb{P})$ , that solves the SDE

$$dX_t = \mu(t, \omega)dt + \sigma(t, \omega)dB_t$$

According to Girsanov's theorem, for each equivalent measure  $\mathbb{Q} \sim \mathbb{P}$  there exist  $\mathcal{F}_t$ -adapted processes  $\lambda$  such that the process

$$B_t^{\mathbb{Q}}(\omega) = B_t(\omega) + \int_0^t \lambda(s, \omega)ds$$

is a Brownian motion in  $(\Omega, \{\mathcal{F}_t^{\mathbb{Q}}\}_{0 \leq t \leq T}, \mathcal{F}, \mathbb{Q})$ , where  $\mathcal{F}_t^{\mathbb{Q}} = \sigma(B_s^{\mathbb{Q}} : 0 \leq s \leq t)$  is the  $\sigma$ -algebra generated by  $B_t^{\mathbb{Q}}$ .

If we define the  $\mathcal{F}_t$ -adapted function  $\alpha$  as

$$\alpha(t, \omega) = \mu(t, \omega) - \lambda(t, \omega) \cdot \sigma(t, \omega)$$

then the process  $X_t$  can be written as a stochastic differential equation under  $\mathbb{Q}$  as

$$dX_t = \alpha(t, \omega)dt + \sigma(t, \omega)dB_t^{\mathbb{Q}}$$

This means that if we are given an equivalent measure we can explicitly solve for the function  $\lambda(t, \omega)$  and write down the SDE that  $X_t$  will solve under the new measure.

Girsanov's theorem also allows us the inverse construction: given an adapted function  $\lambda(t, \omega)$  we can explicitly construct an equivalent probability measure under which  $B_t(\omega) + \int_0^t \lambda(s, \omega)ds$  is a Brownian motion. We define the exponential martingale

$$M_t = \exp \left\{ - \int_0^t \lambda(s, \omega)dB_s - \frac{1}{2} \int_0^t \lambda^2(s, \omega)ds \right\}$$

Based on this exponential martingale we define the following measure on  $(\Omega, \mathcal{F}_t)$

$$\mathbb{Q}(F) = \int_F M_t(\omega)d\mathbb{P}(\omega) = E[M_t \cdot \mathbf{1}(X_t \in F)], \text{ for all } F \in \mathcal{F}_t$$

which we represent as the real-valued process  $d\mathbb{Q}(\omega) = M_t(\omega) \cdot d\mathbb{P}(\omega)$ , or in terms of the Radon–Nikodym derivative  $\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} = M_t$ . It follows that

1.  $\mathbb{Q}$  is a probability measure on  $\mathcal{F}_t$ .
2. The process  $B_t^{\mathbb{Q}} = \int_0^t \lambda(s, \omega) ds + B_t$  is a Brownian motion on the filtered space  $(\Omega, \mathbb{F}, \mathcal{F}, \mathbb{Q})$ .
3. The Itô process can be written in SDE form as

$$dX_t = \alpha(t, \omega)dt + \sigma(t, \omega)dB_t^{\mathbb{Q}}$$

Essentially under the new measure the Itô process will have a different drift  $\alpha$ , but the same volatility as the original one: this is the *Girsanov transformation*. The process  $B_t$  will not be a Brownian motion under the new measure, since we select elements  $\omega \in \Omega$  using different probability weights. One the other hand, it turns out that the process  $\int_0^t \lambda(s, \omega) ds + B_t$  will be a Brownian motion.

In practice we are not really interested in the probability distribution or the dynamics on the set  $\Omega$ , but rather the distribution of  $\mathcal{F}_t$ -measurable random variables  $Y_t(\omega)$ . The Radon-Nikodym derivative will allow us to express expectations under different equivalent probability measures. In particular,

$$\mathbb{E}^{\mathbb{Q}}[Y_t] = \mathbb{E}^{\mathbb{P}}[M_t \cdot Y_t]$$

This is the relation that is routinely used in financial economics, as we very often want to changes probability measures as they adjust with respect to the risk aversion profile of the agents, or with respect to different numéraire securities.

*Example 13.* Say that the price of an asset follows a geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

Here  $B_t$  is a Brownian motion under the filtered space  $(\Omega, \mathbb{F}, \mathcal{F}, \mathbb{P})$ . We can construct a new probability measure  $\mathbb{Q}$ , under which the asset price is a martingale. We can write the process above as

$$dS_t = S_t \sigma \left( \frac{\mu}{\sigma} dt + dB_t \right)$$

Therefore, if we set  $\lambda = \frac{\mu}{\sigma}$ , then we are looking for the probability measure that makes the process

$$B_t^{\mathbb{Q}} = \lambda t + B_t$$

a  $\mathbb{Q}$ -martingale. Then the asset price process under  $\mathbb{Q}$  will be given by the SDE

$$dS_t = \sigma S_t dB_t^{\mathbb{Q}}$$

Girsanov's theorem tells us that such a probability measure on  $\Omega$  exists. If we want to take expectations under this equivalent measure we need to construct the exponential martingale

$$M_t = \exp \left( -\frac{1}{2} \lambda^2 t - \lambda B_t \right)$$

Then, for any  $\mathcal{F}_t$ -measurable random variable  $Y_t$  we can write  $\mathbb{E}^{\mathbb{Q}}[Y_t] = \mathbb{E}^{\mathbb{P}}[M_t Y_t]$ .

*Example 14.* Let's say that we want to verify the above claim that  $E^Q[Y_t] = E^P[M_t Y_t]$ , stated by Girsanov's theorem. For example, let's take the random variable  $Y_t = \log S_t$ . Under  $\mathcal{Q}$  the logarithm will be given by Itô's formula as

$$Y_t = \log S_0 - \frac{1}{2}\sigma^2 t + \sigma B_t^Q$$

and since  $B_t^Q$  is a  $\mathcal{Q}$ -martingale, the expectation  $E^Q[Y_t] = \log S_0 - \frac{1}{2}\sigma^2 t$ . Under  $\mathcal{P}$  we have to consider the process

$$Z_t = M_t Y_t$$

Itô's formula (applied on the function  $f(x, y) = x \cdot y$ ) will give us the dynamics for  $Z_t$ , namely

$$dZ_t = M_t dY_t + Y_t dM_t + dY_t dM_t$$

which actually produces

$$\begin{aligned} dZ_t &= \exp\left(-\frac{1}{2}\lambda^2 t - \lambda B_t\right) \left[ \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dB_t \right] \\ &\quad + \left[ \log S_0 + \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma B_t \right] [-\lambda dB_t] \\ &\quad + \left[ \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dB_t \right] [-\lambda dB_t] \end{aligned}$$

The solution to the above SDE is written as

$$\begin{aligned} Z_t &= Z_0 + \int_0^t \left[ \exp\left(-\frac{1}{2}\lambda^2 s - \lambda B_s\right) \left(\mu - \frac{1}{2}\sigma^2\right) - \lambda\sigma \right] ds \\ &\quad + \int_0^t \left[ \exp\left(-\frac{1}{2}\lambda^2 s - \lambda B_s\right) \sigma - \lambda \log S_0 - \lambda \left(\mu - \frac{1}{2}\sigma^2\right) s - \lambda\sigma B_s \right] dB_s \end{aligned}$$

Taking expectations, and using that  $Z_t = M_t Y_t$ ,  $Z_0 = M_0 Y_0 = \log S_0$  and  $\lambda = \frac{\mu}{\sigma}$  will yield

$$E^P[M_t Y_t] = \log S_0 + \int_0^t \left(\mu - \frac{1}{2}\sigma^2 - \lambda\sigma\right) ds = \log S_0 - \frac{1}{2}\sigma^2 t$$

And the two expectations are apparently the same. Observe though how easier it was to compute the expectation under  $\mathcal{Q}$ . Girsanov's theorem can be a valuable tool when one wants to simplify complex expectations, just by casting them under a different measure.

## 2

---

### The Black-Scholes world

In this chapter we will use some of the previous results to establish the Black-Scholes (BS) paradigm. We will assume a frictionless market where assets prices follow geometric Brownian motions, and we will investigate the pricing of derivative contracts. The seminal papers of [Black and Scholes \(1973\)](#) and [Merton \(1973\)](#) (also collected in the excellent volume in [Merton, 1992](#)) defined the area and sparked thousands of research articles on the fair pricing and hedging of a variety of contracts.

The original derivation of the BS formula is based on a replicating portfolio that ensures that no arbitrage opportunities are allowed. Say that we are interested in pricing a claim that has payoffs that depend on the value of the *underlying asset* at some fixed future date  $T$ . The idea is to construct a portfolio, using the underlying asset and the risk free bond, that replicates the price path of that claim, and therefore its payoffs. If we achieve that, then the claim in question is *redundant*, in the sense that we can replicate it exactly. In addition, the value of the claim must equal the value of the portfolio, otherwise *arbitrage opportunities* would arise.

After the

## 2.1 THE ORIGINAL DERIVATION

In this section we lay down the assumptions for the BS formula. We also give some important definitions on trading strategies, market completeness and arbitrage. We conclude by illustrating that the market under these assumptions is complete, by constructing the corresponding replicating portfolio.

### THE BLACK-SCHOLES ASSUMPTIONS

We fix a filtered space  $(\Omega, \mathbb{F}, \mathcal{F}, \mathbb{P})$ , and a Brownian motion on that space, say  $B_t$ . We will maintain the following assumptions:

1. The asset price follows a geometric Brownian motion, that is to say

$$dS_t = \mu S_t dt + \sigma S_t dB_t \quad (2.1)$$

The parameter  $\mu$  gives the expected asset return, while  $\sigma$  is the return volatility.

2. There is a risk free asset which grows at a constant rate  $r$ , which applies for both borrowing and lending. There is no bound to the size of funds that can be invested or borrowed risk-free.
3. Trading is continuous in time, both for the risk free asset, the underlying asset and all derivatives. This means that any portfolios can be dynamically rebalanced continuously.
4. All assets are infinitely divisible and there is an inelastic supply at the spot price, that is to say the assets are infinitely liquid. Therefore, the actions of any investor are not sufficient to cause price moves.
5. There are no taxes or any transaction costs. There are no market makers or bid-ask spreads. The spot price is the single price where an unlimited number of shares can be bought. Short selling is also allowed.

A *derivative security* is a contract that offers some payoffs at a future (*maturity*) time  $T$ , that depend on the value of the *underlying asset* at the time, say  $\Pi(S_T)$ . We are interested in establishing the *fair* value  $P_t$  of such a security at all times before maturity, that is the process  $\{P_t : 0 \leq t \leq T\}$ .

### THE REPLICATING PORTFOLIO

Of course the derivative price at time  $t$  will depend only on information available at the time, that is  $P_t$  must be  $\mathcal{F}_t$ -adapted. Also, the asset price is Markovian, which indicates that  $P_t$  should not depend on the history of the asset price, but only on the latest value  $S_t$ . We can therefore write the price of the derivative as a function  $P_t = f(t, S_t)$ . The function  $f$  is the unknown pricing formula.

If we actually had the functional form of  $f(t, S)$ , an application of Itô's formula would provide us with the derivative price dynamics

$$\begin{aligned} dP_t = & \left[ \frac{\partial}{\partial t} f(t, S_t) + \mu S_t \frac{\partial}{\partial S} f(t, S_t) + \frac{\sigma^2 S_t^2}{2} \frac{\partial^2}{\partial S^2} f(t, S_t) \right] dt \\ & + \sigma S_t \frac{\partial}{\partial S} f(t, S_t) dB_t \end{aligned}$$

Although we don't actually know  $f(t, S)$  explicitly yet, we will later use the dynamics above to construct a partial differential equation that the pricing function has to satisfy. To produce the PDE we will need to introduce some terminology, including trading strategies and arbitrage opportunities.

We will construct portfolios that we rebalance in time, and we will keep track of them using a *trading strategy*  $H_t$ . Since we must make all rebalancing decisions based on the available information, the trading strategy will be an

$\mathcal{F}_t$ -adapted process as well. Our investment instruments are the underlying and the risk free asset, therefore the trading strategy  $H = \{(H_t^S, H_t^F) : t \geq 0\}$  where  $H_t^S$  keeps track of the number of shares held, and  $H_t^F$  is the amount invested in the risk free asset (that is the bank balance) at time  $t$ . The value of the portfolio that is generated by the trading strategy is denoted with  $V_t = V_t(H)$ .

A *self-financing* trading strategy is one where no funds can enter or exit the portfolio. All changes in the value are due to changes in the price of the assets that compose it. In this case we don't really need to keep track of the holdings of both assets, since there are related via

$$H_t^F = V_t - H_t^S S_t$$

Therefore we will only keep the process of the shares held,  $H = \{H_t : t \geq 0\}$ , as the trading strategy. Also, in this case the dynamics of the portfolio value are given by

$$dV_t = H_t dS_t + (V_t - H_t S_t) r dt = (H_t S_t \mu + V_t r - H_t S_t r) dt + \sigma H_t S_t dB_t$$

Say for a minute that we knew the pricing formula for the derivative price,  $P_t = f(t, S_t)$ . We can then define a trading strategy, where the number of shares held at each time  $t$  is given by

$$H_t = \frac{\partial}{\partial S} f(t, S_t)$$

We have selected this particular trading strategy because it sets the volatility of the portfolio value,  $V_t$ , equal to the volatility of the derivative value,  $P_t$ . We call this a *hedging* or *replicating strategy* and the portfolio the *hedging* or *replicating portfolio*.

## ARBITRAGE OPPORTUNITIES

We claim that if the portfolio has the same volatility dynamics it should also offer the same return. Otherwise arbitrage opportunities will emerge.

An *arbitrage opportunity* is a trading strategy  $J$  that has the following four properties (for a stopping time  $T > 0$ )

1. The strategy  $J$  is self-financing, that is there are no external cash inflows or outflows. We can move funds from one asset to another, but we cannot introduce new funds.
2.  $V_t(J) = 0$ , that is we can engage in the portfolio at time  $t$  with no initial investment. This means that we can borrow all funds needed to set up the initial strategy at the risk free rate, without investing any funds of our own.
3.  $V_T(J) \geq 0$ , it is impossible to be losing money at time  $T$ . The worst outcome is that we end up with zero funds, but we did not invest any funds in the first place.
4.  $\mathbb{P}(V_T(J) > 0) > 0$ , there is a positive probability that we will actually be making a profit at time  $T$ .

An arbitrage opportunity is a risk free money making device, since with no initial investment we have a probability to make a profit, without running any risk of realizing losses. Finance theory assumes that exploitable arbitrage opportunities do not exist when pricing claims.

Now say that at time 0 we engage in the following self-financing strategy  $\Theta$ , where

1. we are short (we have sold) one derivative contract,
2. we hold  $H_t$  shares, and
3. we keep an amount  $\phi_t$  in the risk free bank account.

Thus, our holdings at any time  $t \geq 0$  will have value

$$V_t(\Theta) = -P_t + H_t S_t + \phi_t$$

We want to keep the initial gross investment equal to zero,  $V_t = 0$ , and therefore our initial bank balance will be  $\phi_0 = P_0 - H_0 S_0$ . We also want to maintain the strategy self-financing, and therefore all changes in the value of our portfolio must come through changes in the assets themselves,

$$dV_t = -dP_t + H_t dS_t + d\phi_t = -dP_t + H_t dS_t + (V_t + P_t - H_t S_t) r dt$$

Using Itô's formula for  $P_t = f(t, S_t)$  and the stochastic differential equation for  $S_t$  we can write after some algebra (which incidentally cancels the drifts  $\mu$ )

$$dV_t = - \left[ f_t(t, S_t) + r S_t f_S(t, S_t) + \frac{\sigma^2 S_t^2}{2} f_{SS}(t, S_t) - V_t r - P_t r \right] dt \quad (2.2)$$

The trading strategy  $\Theta$  is self-financing, and its initial value is  $V_t(\Theta) = 0$ . Therefore it has two of the four requirements that we set for an arbitrage opportunity. In order to avoid such opportunities, we want to verify that there exists no stopping time  $\tau$ , such that  $V_\tau(\Theta) \geq 0$  and  $\mathcal{P}(V_\tau(\Theta) > 0) > 0$ .

The value of the trading strategy will evolve in a deterministic way, as illustrated in the above relationship where no stochastic term is present. Therefore, if the term in the brackets is equal to zero for all  $t$ , then  $dV_t = 0$  which implies  $V_t = V_0 = 0$  for all  $t$ . Then, apparently, no arbitrage opportunities are present since  $\mathcal{P}(V_\tau(\Theta) > 0) = 0$ .

We can also show that this condition is also necessary. Say that  $\tau > 0$  is the first time that the term in brackets of (2.2) becomes non-zero, and say that it is negative implying a positive  $dV_\tau$ . Since  $f$  is continuous in both arguments, there will be an interval  $(\tau, \tau + \Delta t)$  on which portfolio value will remain positive, and therefore the value of the portfolio  $V_{\tau+(\Delta t/2)}(\Theta) > 0$ , which indicates an arbitrage opportunity. If at  $\tau$  the value of the portfolio becomes negative, then we can implement the inverse trading strategy for which  $V_{\tau+(\Delta t/2)}(-\Theta) > 0$ , and again reach an arbitrage opportunity.

## THE BLACK-SCOLES PARTIAL DIFFERENTIAL EQUATION

In the previous subsection we concluded that the value of the composite portfolio  $\Theta$  must be  $V_t(\Theta) = 0$  for all  $t \geq 0$ , otherwise arbitrage opportunities will be present. Then, equation (2.2) will give the celebrated *Black-Scholes partial differential equation* (BS-PDE), namely

$$\frac{\partial}{\partial t} f(t, S) + rS \frac{\partial}{\partial S} f(t, S) + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2}{\partial S^2} f(t, S) = f(t, S)r \quad (2.3)$$

must be satisfied by the derivative pricing function  $f(t, S)$ . This is one of the fundamental relationships in financial economics, as it has to be obeyed by any derivative contract. It shows that the price of the derivative can be replicated by a dynamically balanced portfolio that consists of the underlying asset and a risk free bank account, and is actually independent of the expected return on the underlying asset  $\mu$ .

As we pointed out, in order to derive the BS-PDE we did not make any assumptions on the nature of the contract, meaning that the PDE will be satisfied by all derivatives. The nature of the particular contract will specify the *terminal condition* of the PDE. Indeed, we know that on the maturity date

$$P_T = \Pi(S_T) \Rightarrow f(T, S) = \Pi(S)$$

In their paper BS present the case of a European call option, a contract that gives the holder the right (but not the obligation) to purchase a share at a fixed price  $K$  on the maturity date. Then, the terminal condition becomes  $f(T, S) = \max(S - K, 0) = (S - K)^+$ . In this case BS show how the PDE can be solved analytically and produce the *Black-Scholes formula*, which is the particular pricing function  $f(t, S)$  for this contract

$$f(t, S) = S \cdot N(d_+) - K \cdot \exp(-r(T-t)) \cdot N(d_-) \quad (2.4)$$

where  $N(\cdot)$  is the cumulative (standardized) normal distribution function, and  $d_{\pm}$  are given by

$$d_{\pm} = \frac{\log\left(\frac{S}{K}\right) + (r \pm \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

Typically we prefer to work with *time to maturity*, and we use the change of variable  $t \rightsquigarrow T - t$  (abusing the notation slightly). It is also convenient to define the *log-prices* setting the variable  $s \rightsquigarrow \log S$ . Some elementary calculus produces the BS-PDE under these variable changes, a differential equation with constant coefficients

$$-\frac{\partial}{\partial t} f(t, s) + \left(r - \frac{1}{2}\sigma^2\right) \frac{\partial}{\partial s} f(t, s) + \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial s^2} f(t, s) = f(t, s)r \quad (2.5)$$

Apart from rendering this expression easier for numerical methods to handle (since the coefficients are constant), we have a PDE with an *initial condition* rather than a terminal one, namely  $f(0, s) = (\exp(s) - K)^+$ . In this form, the BS-PDE is a standard *convection-diffusion* partial differential equation, a form that has been studied extensively in classical and quantum physics.

## 2.2 THE FUNDAMENTAL THEOREM OF ASSET PRICING

Since  $V_t(\Theta) = 0$  for all times  $t \geq 0$ , the relationship

$$P_t = H_t S_t + \phi_t$$

will also hold at all times. This means that using the trading strategy  $H_t$  we create a portfolio that will track (or mimic) the process  $P_t$ . Therefore we do not really need to introduce derivatives in the BS world, as their trajectories and payoffs can be replicated by using a carefully selected trading strategy. For that reason we say that in the BS world derivatives are *redundant securities*. This of course only holds under the strict BS assumption, and does not generally hold in any market. It certainly does not hold in the real world where markets are subject to a number of frictions and imperfections.

As we search for markets and models where securities can be hedged, we need to introduce the notion of market completeness. We will say that a market is called *complete* if all claims can be replicated. A market that is complete will of course be arbitrage-free, but the inverse is not true. There are many markets that are arbitrage free but incomplete. One can speculate that the real world markets fall within this category: claims cannot be perfectly replicated due to market imperfections, and these imperfections also make arbitrage opportunities scarce and short lived.

We set a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , under which the price process is defined. In financial mathematics, an *equivalent martingale measure* (EMM) is a measure  $\mathbb{Q}$  equivalent to the objective one  $\mathbb{P}$ , under which all discounted asset prices form martingales. Therefore for the discounting factor  $B_t$ , any price process  $V_t$  will satisfy

$$V_0 = \mathbb{E}^{\mathbb{Q}}[B_t V_t] \text{ for all } t \geq 0$$

The fundamental theorem of asset pricing states the following two propositions:

There exists an EMM  $\Leftrightarrow$  There are no arbitrage opportunities

There exists a unique EMM  $\Leftrightarrow$  The market is complete

## THE FUNDAMENTAL THEOREM OF ASSET PRICING AND GIRSANOV'S THEOREM

Girsanov's theorem is a very useful companion to the fundamental theorem of asset pricing, as it provides us with the link between different equivalent probability measures. A typical approach would be to assume a process for an asset under the true probability measure. This will specify the true dynamics of an asset or a collection of assets, that is to say the process that we would produce

based on time series of the prices. Then, we use Girsanov's theorem and try to specify the Radon-Nikodym derivative that produces discounted asset prices that form martingales. Of course there might be more than one probability measures with that feature, but if we manage to find one then we can conclude that the system is arbitrage-free. If we show that such a measure does not exist, then we know that the system as it stands offers some arbitrage opportunities, and then we can proceed to find them. Unfortunately, the fundamental theorem of asset pricing does not always guide us towards these opportunities, but sometimes it can offer useful insights to identify them.

Suppose that we are facing a market that offers a risk free rate  $r$ , and a collection of  $M$  stocks. There are  $N$  sources of uncertainty, represented by  $N$  independent Brownian motions  $\mathcal{B}(t) = \{B_t^i : i = 1, \dots, N\}$ . If we collect the asset returns in a  $(M \times 1)$  vector with  $dS(t) = \{S_t^j : j = 1, \dots, M\}$ , then we can write

$$dS(t) = \mu \odot S(t) + [\Sigma \cdot dB(t)] \odot S(t)$$

The  $(M \times N)$  matrix  $\Sigma$  will determine the correlation structure of the assets. In fact, the covariance matrix of the stocks will be given by the product  $\Sigma \cdot \Sigma'$ . Essentially, each asset will satisfy the SDE

$$dS_j(t) = \mu_j S_j(t)dt + \sum_{i=1}^N \sigma_{j,i} S_j(t)dB_i(t)$$

We want to establish whether or not we can find a trading strategy using these  $M$  stocks that will be an arbitrage opportunity. To this end we will examine the probability measures that are equivalent to the true one. In particular, all equivalent measures will have a Radon-Nikodym derivative  $M_t$  that satisfies

$$dM_t = M_t \sum_{i=1}^N \lambda_t^i dB_t^i$$

We are looking for these equivalent probability measures under which the discounted prices will form martingales, which means that under the EMM the dynamics of the assets will be

$$dS_t^j = rS_t^j dt + \sum_{i=1}^N \sigma_{j,i} S_t^j dB_t^{i,\Omega}$$

Using Girsanov's theorem we can actually find the instantaneous drift under  $\Omega$ , which will be given by

$$\begin{aligned} E^\Omega dS_t^j &= E^\mathcal{P} \left[ \left( 1 + \frac{dM_t}{M_t} \right) dS_t^j \right] \\ &= \mu_j S_t^j dt + E^\mathcal{P} \left[ \left( \sum_{i=1}^N \lambda_t^i dB_i(t) \right) \left( \sum_{i=1}^N \sigma_{j,i} dB_t^i \right) \right] S_t^j \end{aligned}$$

Since the Brownian motions are mutually independent, we can simplify the above expression to

$$E^Q dS_t^j = \left[ \mu_j + \sum_{i=1}^N \lambda_t^i \sigma_{j,i} \right] S_t^j dt = r S_t^j dt$$

which has to be satisfied for all  $t \geq 0$  and for all  $j = 1, \dots, M$ . Therefore the parameters  $\lambda = \{\lambda^i : i = 1, \dots, N\}$  will be constant, and they must satisfy the system of  $M$  equations with  $N$  unknowns

$$\Sigma \cdot \lambda = \mu - r\mathbf{1}$$

This system can have no solutions, a unique solution, or an infinite number of solutions, depending on the rank of the matrix  $\Sigma$ . If the rank is lower than the number of unknowns,  $\text{rank}(\Sigma) < N$ , then the system will not admit a solution. This means that there does not exist an equivalent martingale measure, and due to the fundamental theorem of asset pricing it is implied that arbitrage trading strategies can be constructed using a portfolio of the  $M$  stocks. If  $\text{rank}(\Sigma) > N$  then there exists an infinite number of vectors  $\lambda$  that are solutions to the system. Each one of these solutions will define an equivalent martingale measure and the market is arbitrage-free. Finally, if  $\text{rank}(\Sigma) = N$  then the solution to the system is unique. This unique  $\lambda$  will define a unique EMM and the market will be complete. In that case, any other asset that depends on the Brownian motions  $B(t)$  can be replicated using the  $M$  assets in the market.

### A SECOND DERIVATION OF THE BLACK-SCHOLES FORMULA

Let us now consider the simple case where there is only one risky stock in the market, with the dynamics given in equation (2.1). Then, it follows that the coefficient of Girsanov's transformation will solve

$$\sigma \cdot \lambda = \mu - r \Rightarrow \lambda = \frac{\mu - r}{\sigma}$$

Therefore the coefficient  $\lambda$  is the *Sharpe ratio* of the risky asset. The Sharpe ratio is a measure of the risk premium per unit of volatility risk, and represents the compensation that investors demand for holding the stock which has uncertain payoffs. In this case, since  $\lambda$  is unique, the market will be complete.

Girsanov's theorem will define the equivalent martingale probability measure  $Q$  as the one with Radon-Nikodym derivative the exponential martingale

$$\left. \frac{dQ}{dP} \right|_{\mathcal{F}_t} = M_t = \exp \left( -\frac{1}{2} \lambda^2 t + \lambda B_t \right)$$

It follows that the discounted prices process of any other asset must form a  $Q$ -martingale as well. In particular we can consider a European-style contract that delivers an amount  $g(S_T)$  at time  $T$ , a payoff that depends explicitly on the price of the underlying stock at the time. The value of this claim at all times  $0 \leq t \leq T$  will satisfy

$$V_t = \exp(-r(T-t)) E_t^{\Omega} g(S_T) = \exp(-r(T-t)) E_t^{\mathcal{P}}[M_T \cdot g(S_T)]$$

These equalities offer us three options to evaluate the value of the derivative at time  $t = 0$ .

### Expectation under the true measure $\mathcal{P}$

Under  $\mathcal{P}$  the asset price and the Radon-Nikodym derivative at time  $T$  are functions of  $B_T$ , and the price of the derivative at time  $t = 0$  can be written, using the second equality, as

$$V_0 = \exp(-rT) E_t^{\mathcal{P}} \left[ \exp \left( -\frac{\lambda^2 T}{2} + \lambda B_T \right) g \left( S_0 \exp \left[ \left( \mu - \frac{\sigma^2}{2} \right) T + \sigma B_T \right] \right) \right]$$

For general functions  $g(\cdot)$  this expectation can be computed by simply simulating values for  $B_T$  from the normal distribution with mean zero and variance  $T$ .

### Expectation under the risk neutral measure $\Omega$

A much simpler approach is to use the fact that the dynamics of the underlying asset under  $\Omega$  are known, and in fact

$$dS_t = rS_t dt + \sigma S_t dB_t^{\Omega}$$

Therefore using the first equality we can express the price of the derivative as

$$V_0 = \exp(-rT) E_t^{\Omega} \left[ g \left( S_0 \exp \left[ \left( r - \frac{\sigma^2}{2} \right) T + \sigma B_T^{\Omega} \right] \right) \right]$$

Now the process  $\{B_t^{\Omega}\}_{t \geq 0}$  is a Brownian motion under  $\Omega$ , and therefore once again we can draw from the normal distribution with zero mean and variance  $T$  to simulate the values of  $B_T^{\Omega}$ . The two expression will of course yield the same result, but the latter is substantially simpler.

In particular, in the case of a standard European call option, the price will satisfy

$$P_0 = \exp(-rT) E_t^{\Omega} \left[ \left( S_0 \exp \left[ \left( r - \frac{\sigma^2}{2} \right) T + \sigma B_T^{\Omega} \right] - K \right)^+ \right]$$

Since  $B_T^{\Omega}$  is normally distributed, after some algebra the expectation simplifies to

$$\begin{aligned} P_0 &= \frac{\exp(-\sigma^2 T/2)}{\sqrt{2\pi T}} S_0 \int_{-d}^{\infty} \exp \left( \sigma B - \frac{B^2}{2T} \right) dB \\ &\quad - \frac{\exp(-rT)}{\sqrt{2\pi T}} K \int_{-d}^{\infty} \exp \left( -\frac{B^2}{2T} \right) dB \end{aligned}$$

In the above expression  $d = [\log(S_0/K) + (r - \sigma^2/2)T]/\sigma$ . Evaluating the two integrals will eventually lead to the Black-Scholes formula.

### The Feynman-Kac form

A third approach would invoke the Feynman-Kac formula. In particular we can write the first expectation of the valuation formula as

$$V_0 = E_0^Q \left[ \exp \left( - \int_0^T r ds \right) g(S_T) \right]$$

with  $S_t$  following the risk neutral dynamics. We shall also define the function  $v(t, s) = E^Q[\exp(-\int_0^t r ds) g(S_t) | S_0 = s]$ , implying that in fact we are interested in the value  $V_0 = v(T, S_0)$ . Following the Feynman-Kac approach (see section 1.7) the function  $v(t, s)$  solves the parabolic PDE that depends on the dynamics of the asset prices process under  $Q$  (since the expectation is taken under  $Q$ )

$$\frac{\partial}{\partial t} v(t, s) = rS \frac{\partial}{\partial s} v(t, s) + \frac{1}{2} \frac{\partial^2}{\partial s^2} v(t, s) - rv(t, s)$$

with initial condition  $v(0, s) = g(s)$ . This is just the Black-Scholes partial differential equation (2.3), after we change the time variable to the time-to-maturity, which transforms the BS-PDE terminal condition into an initial one.

## 2.3 EXOTIC OPTIONS

The power of the fundamental theorem of asset pricing is unleashed when one considers pricing contracts that are more complicated than the simple European calls and puts. There is a very large and fairly liquid market for contracts that are called *exotic*, in the sense that they exhibit features that are non-standard. In practice, the role of a trader is to create tailor-made contracts for her clients, and the role of the financial engineer is to produce benchmark prices for these contracts that are arbitrage-free, and also present ways to hedge the exposure of the trading book using available liquid contracts, like the underlying assets and standard calls and puts.

The fundamental theorem of asset pricing will dictate that no matter how complicated the payoff structure, the no-arbitrage price will be equal to the discounted expected payoffs under the equivalent martingale measure  $Q$ . Sometimes it is more convenient to simulate these payoffs under  $Q$ , or to evaluate the expectation in closed form, but in other cases solving the PDE might be more efficient.

### Exercise timing

Exotic contracts can be classified with respect to their exercise times and their payoff structure. European-style contracts can be exercised only on the maturity date, while *American* derivatives can be exercised at any point before the maturity date. That is to say, a three-month American put with strike price  $30p$  gives the holder the right to sell the underlying asset for  $30p$  at any point she

wishes in the next three months. In this case the holder will have to determine the *optimal* exercise point. *Bermudan* options are somewhat between the European and the American ones,<sup>1</sup> and allow the holder to exercise at a predefined set of equally spaced points. For example if the put option described above was a Bermudan one, perhaps it could offer weekly exercising at the closing of each Friday during the next three months. Once again every Friday the holder must decide if it is optimal to exercise or to wait for the next exercise point.

The *shout option* is slightly more complicated, as the holder has the option to lock-in one or more prices up to the maturity date (that is by "shouting" to the seller), and use the price they choose to compute the payoffs at maturity. For example, if the put was a *one-shout option*, and after six weeks the underlying price is  $22p$  the holder has the opportunity to "shout" and lock in that price. Therefore, if on the maturity date the price is  $26p$  the holder will choose which value will be used to compute the payoffs, in this case  $22p$  which gives payoffs  $(30p - 22p)^+ = 8p$  per share.

Typically, one computes the prices of contracts with exotic exercise structure using the partial differential equation. In most cases this PDE has to be solved numerically. In chapter 3 we will give an overview of some methods that are used to numerically solve for the price of the option, the optimal exercise strategy and the hedging parameters.

### Payoff structures

Apart from the standard calls and puts there can be a wide range of structures that define the payoffs of the contract. The simplest deviation is the *digital option* (also called *binary* or *all-or-nothing option*), where the payoff is a fixed amount if the underlying is above or below the strike price. For example a two month digital call with strike  $60p$  will pay \$1 if the value of the underlying is above  $60p$  after two months. In that sense it is a standard bet on the future level of the underlying asset price.

Another popular option is the *cross option*, where the underlying asset is quoted in one currency but the payoffs (and the strike price) are denominated in another. For example, British Airways are traded in the London stock exchange and are priced in British pounds, but a US based investor will want the strike price and the payoff in US dollars. Therefore, if  $X_t$  is the USD/GBP exchange rate, and  $S_t$  is the BA price in London (quoted in GBP), then a European call will have payoffs of the form  $(S_T X_T - K)^+$ , where the strike price is quoted in USD. Therefore the writer of this option is also exposed to exchange rate risks and the correlation between the exchange rate and the underlying asset returns. A *quanto option* will address this dependence by setting the exchange rate that will be used for the conversion beforehand, say  $X^*$ . Therefore the payoffs will only depend on the fluctuations of the underlying asset, given by  $(S_T X^* - K)^+$ .

The cross option described above is an example of an option that depends on more than one underlying assets. Other exotics share this feature, like the

---

<sup>1</sup> Just as Bermuda is between Europe and the US.

*exchange option* that allows the holder to exchange one asset for another, a *basket option* that uses a portfolio of assets as the underlying asset, or the *rainbow option* that depends on the performance of a collection of assets. An example of a rainbow option is a European put where the payoffs are computed using the worst of ten stocks.

Other contracts have features that involve other derivatives, like the *compound option* that is an option to buy or sell another option. In this case you can have a call on a call, a put on a call, etc. The *swing option* lets the holder decide if she will use the option as a call or as a put, at a pre-specified number of time points. Typically the holder is not allowed to use all options as calls or puts, and some provisions are in place to ensure that a mix is actually used. The *chooser option* is a variant that allows the holder to decide if the option will pay off as a call or a put. This decision must be made at some point before maturity.

If the option is of the European type, one can retrieve its price by using either the PDE or by simulating the expectation. When the number of underlying assets is small it is usually faster to numerically solve the PDE, but as the number of assets grows these numerical methods become increasingly slower. It is typically stated that if the number of assets is larger than four, then simulation methods become more efficient.

### Path dependence

For options with early exercise features one has to make decisions on the exercise times. This decision will be dependent on the complete price path of the underlying asset, and not only on its value at maturity. Some other option contracts exhibit more explicit or stronger path dependence.

A *barrier option* has one or more predefined price levels (the barriers). Reaching these barrier can either activate ("knock-in" barrier) or deactivate ("knock-out" barrier) the contract. Say, for example, that the current price of the underlying asset is 47p, and consider a six month call option with strike 55p and a knock-in barrier at 35p. In order for payoffs to be realized on maturity, not only the price has to end up higher than the 55p strike price, but the contract must have been activated beforehand, that is the price needs to have fallen below 35p at some point before maturity. Monitoring of barrier options is not usually continuous, but takes place on some predefined time points that are typically equally spaced. The payoff of a *Parisian option* will depend on the time that is spent beyond the corresponding barriers, in order to smooth discontinuities.

*Lookback options* have payoffs that depend not on the terminal value of the underlying asset, but on the maximum or the minimum value over a predefined period. Once again in most cases this maximum or minimum is taken over a discrete set of time points. The special case where the maximum or minimum over the whole price path is considered yields the *Russian option*. An *Asian option* will have payoffs that depend on the average (arithmetic or geometric) of the price over a time period, rather than a single value. Therefore an Asian

option could be a call with payoffs that depend on the average daily prices of the underlying over the month prior to maturity.

Path dependence is easily accommodated using simulation methods, as sample paths of the underlying can be produced and the payoff can be computed over each path. Nevertheless, one would still set up the relevant PDEs if this was possible. Sometimes to specify the PDE one must define some auxiliary variables, for example the “time above the knock out barrier” in the case of a Parisian option.

## 2.4 THE GREEKS

So far we have addressed the problem of finding the no-arbitrage price of a derivative contract, under the assumption that underpin the Black-Scholes paradigm. We showed that in that case the market is complete, and any contract can be replicated, at least in principle. Now we will look at these replicating strategies more closely, and investigate a number of different hedging strategies. We will take two different views, illustrated in the next two settings:

1. A trader at a financial institution ABC wants to give a quote for a derivative, most probably one with exotic features. The trader will investigate the trading strategies that would, in theory at least, replicate the payoffs of this derivative. In theory, following the BS procedure we shall hold  $H_t$  shares at all times, as discussed in section 2.1. In practice, as trading is not continuous and markets are not frictionless, this replication will not be exact. The quote she will produce will be the replicating costs, plus a premium for the risk she runs due to imperfect hedging, plus a fee for her time and bonus.
2. An investor XYZ is holding a portfolio of assets that depend on one or more risk factors. She wants to enter some options positions that will hedge her position against adverse moves of these factors, perhaps in the form of exotic options purchased from the financial institution above. Of course this insurance will come at a premium, and she wants to investigate the cost of different protection levels. For example, if her portfolio is well diversified, the market will be a natural factor she is exposed to. She will consider enhancing her portfolio with derivative contracts that are written on a market index.

It is important to observe that the value of the derivative that ABC has sold will obey the same partial differential equation that the portfolio of XYZ does. This follows from the absence of arbitrage opportunities that would otherwise occur. If we assume that there is a single underlying source of risk, summarized by the asset  $S_t$ , then any portfolio or derivative contract with value  $V_t$  can be expressed as a function of  $S_t$ , namely  $V_t = V(t, S_t)$ . This function will satisfy the Black-Scholes PDE

$$\frac{\partial}{\partial t} V(t, S) + rS \frac{\partial}{\partial S} V(t, S) + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2}{\partial S^2} V(t, S) = rV(t, S)$$

We will use some Greek letters for the derivatives involved, namely  $\Delta = \frac{\partial V}{\partial S}$  (the *Delta*),  $\Gamma = \frac{\partial^2 V}{\partial S^2}$  (the *Gamma*) and  $\Theta = -\frac{\partial V}{\partial t}$  (the *Theta*). Then we can write the BS-PDE as

$$-\Theta + rS\Delta + \frac{1}{2}\sigma^2 S^2\Gamma = rV$$

More importantly, a Taylor's expansion of the value function  $V(t, S)$  over a small time interval  $\Delta t$  and a small price change  $\Delta S$  yields

$$\begin{aligned} \Delta V &= \frac{\partial V}{\partial t}\Delta t + \frac{\partial V}{\partial S}\Delta S + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\Delta S^2 + o(\Delta t, \Delta S^2) \\ &\Rightarrow \Delta V \approx -\Theta\Delta t + \Delta\Delta S + \frac{1}{2}\Gamma\Delta S^2 \end{aligned}$$

The Delta of the derivative or the portfolio will therefore represent its sensitivity with respect to changes in the underlying asset. In continuous time trading, holding  $\Delta$  units of the underlying asset at all times is sufficient to replicate the path and payoffs of the portfolio value.  $\Theta$  will be the *time decay* of this value, representing the changes as we move closer to maturity, even if the underlying asset does not move. When trading takes place in discrete time, there is going to be some misalignment between the two values, and higher order derivatives can be used to correct for that. In addition, the  $\Gamma$  controls the size of the hedging error when one uses the wrong volatility for pricing and/or hedging. This is an important feature, as the volatility is the only parameter in the BS PDE that is not directly observed and has to be estimated.

In the BS framework there also some parameters that are considered constant, namely the volatility  $\sigma$ , the risk free rate  $r$ , and the dividend yield  $q$ . Therefore one can write the value of function as  $V_t = V(t, S; \sigma, r, q)$ , and practitioners use the derivatives of the value functions with respect to these parameters as a proxy of the respective sensitivities. In particular  $v$  or  $\kappa = \frac{\partial V}{\partial \sigma}$  (the *Vega* or *Kappa*<sup>2</sup>),  $\rho = \frac{\partial V}{\partial r}$  (the *Rho*), and  $\phi_V = \frac{\partial V}{\partial q}$  (the *Phi*).

With the increased popularity of exotic contracts that are particularly sensitive to some parameter values a new set of sensitivities is sometimes used, although very rarely. These sensitivities are implemented via higher order Taylor's expansions of the value function. Running out of Greek letters, these sensitivities have taken just odd-sounding names or have borrowed their names from quantum mechanics, like the *Speed*  $\frac{\partial^3 V}{\partial S^3}$ , the *Charm*  $\frac{\partial^2 V}{\partial S \partial t}$ , the *Color*  $\frac{\partial^3 V}{\partial S^2 \partial t}$ , the *Vanna*  $\frac{\partial^2 V}{\partial S \partial \sigma}$ , and the *Volga*  $\frac{\partial^2 V}{\partial \sigma^2}$ .

No matter what Greek or non-Greek letters are used, the objective is the same: to enhance the portfolio with a number of contracts that result in a position that is neutral with respect to some Greek. This turns out to be a simple exercise, as portfolios are linear combinations of assets and this carries through to their sensitivities. Say that we are planning to merge two portfolios with values  $V_t^1$

---

<sup>2</sup> Vega is not a Greek letter, and for that reason this sensitivity is also found in the literature as Kappa.

LISTING 2.1: bs\_greeks.m: Black-Scholes Greeks.

```
% bs_greeks.m
function [P, D, G, V] = bs_greeks(S, K, r, s, t, CP)
t = t + eps;
d1 = log(S./K) + (r + 0.5*s.^2).*t;
5 d1 = d1./s./sqrt(t);
d2 = d1 - s.*sqrt(t);
Nd1 = normcdf(CP.*d1);
Nd2 = normcdf(CP.*d2);
nd1 = normpdf(d1);
10 P = CP.* (S.*Nd1 - K.*exp(-r.*t).*Nd2); % price
D = CP.*Nd1; % delta
G = nd1./S./s./sqrt(t); % gamma
V = nd1.*S.*sqrt(t); % vega
```

and  $V_t^2$  into one with value  $V_t^{1+2}$ , and suppose that we are interested in any sensitivity  $\boxtimes$  (where  $\boxtimes$  could be  $\Delta, \Gamma, \dots$ ). It follows that since  $\boxtimes$  is actually a derivative,

$$\boxtimes_t^{1+2} = \boxtimes_t^1 + \boxtimes_t^2$$

The simplest asset that we can use to enhance our portfolio in order to achieve some immunization is the underlying asset itself,  $S_t$ . Trivially, the valuation function of the asset is  $V(t, S; \sigma, r, q) = S$ , and therefore the Delta of the asset  $\Delta^S = \frac{\partial S}{\partial S} = 1$ , while all other sensitivities are equal to zero. The argument above indicates that by augmenting our portfolio with more units of the underlying we will change the Delta of the composite position. In order to immunize other sensitivities we will need to construct a position that incorporates derivative contracts, with the plain vanilla calls and puts being the most readily available candidates. For that reason we will now investigate the Greeks of these simple options and examine how we can use them to achieve Greek-neutrality. Listing 2.1 gives the Matlab function that produces the price and the major Greeks for the Black-Scholes option pricing model, for both calls and puts.

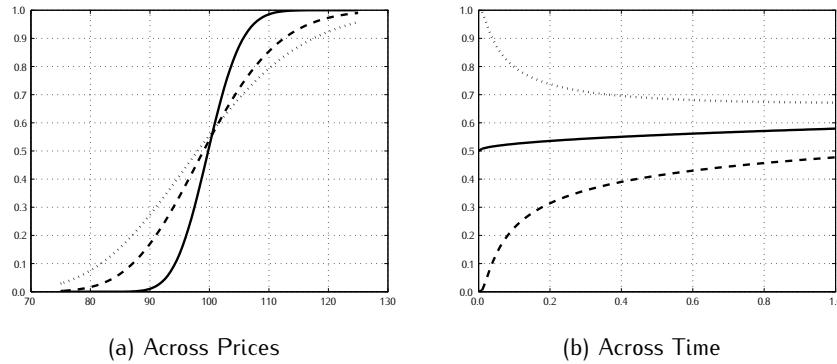
## THE DELTA

Say that start with a portfolio with value  $V$  and Delta  $\Delta^V$ . As we noted above we can adjust the Delta of a portfolio by adding or removing units of the underlying asset. In particular, if we add  $w_S$  units of the asset, the Delta of the portfolio will become

$$\Delta^{V+S} = \Delta^V + w_S \Delta^S = \Delta^V + w_S$$

In order to achieve *Delta-neutrality*,  $\Delta^{V+S} = 0$ , we will need to short  $w_S$  units of the underlying asset. Note that by adding or removing funds from the risk-free bank account does not have any impact on the Greeks. We can therefore adjust the bank balance with the proceedings of this transaction.

FIGURE 2.1: Behavior of a call option Delta. Part (a) gives the behavior of the delta of options with specifications  $\{K, r, \sigma\} = \{100, 0.02, 0.20\}$ , and three different times to maturity:  $t = 0.05$  (solid),  $t = 0.25$  (dashed) and  $t = 0.50$  (dotted). Part (b) gives the behavior of the delta as the time to maturity increases, for a contract which is at-the-money ( $S = 100$ , solid), in-the-money ( $S = 95$ , dashed), and out-of-the-money ( $S = 105$ , dotted).



A position that is Delta-neutral will not change in value for small asset price changes (but it will change due to the time change as  $\theta^V$  dictates). Of course after a small change in the asset price the value of  $\Delta^{V+S}$  will change, as  $\frac{\partial}{\partial S} \Delta^{V+S} = \Gamma^V$ . In order to maintain a Delta neutral portfolio, one has to rebalance it in a continuous fashion, employing a *dynamic Delta hedging* strategy.

In the BS framework European calls (and puts) are priced in closed form as in equation (2.4). Taking the derivative with respect to the price  $S$  yields the Delta for calls and puts

$$\text{Calls: } \Delta^C = N(d_+) \quad \text{Puts: } \Delta^P = 1 - N(d_+)$$

The values of Delta for a European call option, across different spot prices and different maturities is displayed in figure 2.1. The Delta for deep-in-the-money options is equal to one, as exercise appears very likely and the seller of the option will need to hold one unit of asset in order to deliver. For options that are deep-out-of-the-money exercise is unlikely and the seller of the option will not need to carry the asset, making the Delta equal to zero. As the time to maturity increases the Deltas of in- and out-of-the-money contracts converge towards the at-the-money Delta.

### Dynamic Delta hedging

A seller of an option that maintains a Delta-neutral position at all times is replicating the contract and should end up with a zero bank balance, no matter what the path of the underlying asset is. Of course in practice one cannot rebalance

LISTING 2.2: bs\_D\_hedge.m: Dynamic Delta hedging.

```
% bs_D_hedge.m
S0      = 100; % initial price
K       = 100; % strike price
r       = 0.02; % interest
5 mu     = 0.08; % drift
sigma   = 0.10; % volatility
T       = 0.25; % maturity
N       = 250; % number of hedges
dt      = T/N; % time intervals
10 dW    = sqrt(dt)*randn(N,1); % white noise
Tv     = (0:N)'*dt; % keeps track of time
dlnS   = (mu-0.5*sigma^2)*dt + sigma*dW; % dlogS process
lnS    = [0; cumsum(dlnS)]; % logS process
S      = S0*exp(lnS); % price process
15 % option price and Greeks process
[C, D] = bs_greeks(S, K, r, sigma, T-Tv, 1);
PL    = zeros(N+1,1); % bank balance (P&L)
PL(1) = C(1) - D(1)*S(1); % initial P&L
for tndx = 2:N % loop through time
20 INTEREST = PL(tndx-1)*(exp(r*dt)-1); % interest
DDIFF   = D(tndx)-D(tndx-1); % stock needed
BORROW  = DDIFF*S(tndx); % funds needed
PL(tndx) = PL(tndx-1) + INTEREST - BORROW; % P&L
end
25 INTEREST = PL(N)*(exp(r*dt)-1); % final interest payment
DELIVER  = (S(N+1) > K); % stock delivered?
DDIFF   = DELIVER - D(N); % stock needed
BORROW  = DDIFF*S(N+1); % funds needed
PL(N+1) = PL(N) + INTEREST - BORROW + DELIVER*K; % last P&L
```

continuously (even in the ideal case where the markets are frictionless). Figure 2.2 illustrates dynamic Delta hedging in a simulated BS world, while in 2.3 the actual strategy is presented step-by-step. Initially we sell one call option with strike price  $K = 100$  (at-the-money) and four months to maturity for \$2.25. In order to hedge it we need to purchase  $\Delta = 0.55$  shares, and we will need to borrow \$52.72 to carry out this transaction.

As the price of the underlying asset drops, the Delta of the call follows suit. We are therefore selling our holdings gradually, recovering some funds for our bank balance. Eventually the price recovers and we build up the asset holdings once more. In discrete time intervals the option price changes are not matched exactly by changes in our portfolio value. In particular these discrepancies are larger for large moves of the underlying. Overall the hedging portfolio will mimic

the process of the call option to a large extent, but not exactly. In this simulation run we are left with a profit of \$0.12.

Increasing the frequency of trades will decrease the volatility of this hedging error, and of course at the limit the replicating strategy is exact. If from one transaction to the next the Delta does not move a lot, we would expect the impact of discrete hedging to be small. On the other hand, the impact will be most severe in the areas where the Delta itself changes rapidly. The second order sensitivity with respect to the price, the Gamma, is in fact summarizing these effects.

## GAMMA

The gamma of a portfolio is defined as the second order derivative of the portfolio value with respect to the price, or equivalently as the first order sensitivity of the portfolio Delta with respect to the price. As we already mentioned above, we expect the Delta of a portfolio to change across time, as the price of the asset changes. Gamma will give us a quantitative insight on the magnitude of these changes.<sup>3</sup>

We have already analyzed how a portfolio can be made Delta-neutral, by taking a position in the underlying asset. In order to achieve *Gamma-neutrality*, the underlying asset is not sufficient. This is due to the fact that

$$\Gamma^S = \frac{\partial^2 S}{\partial S^2} = 0$$

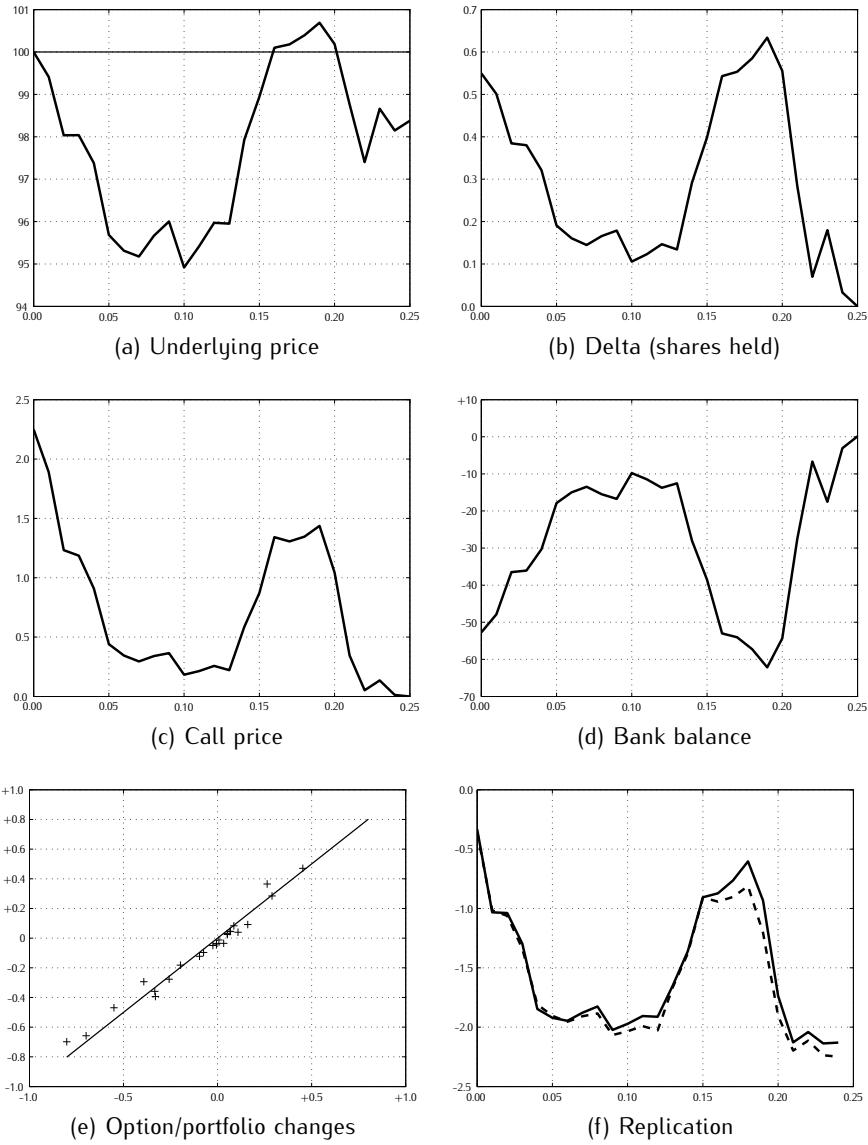
This indicates that we need instruments that are nonlinear with respect to the underlying asset price, in order to achieve Gamma-neutrality. Options are perfect candidates for this job. On the other hand, the fact that  $\Gamma^S = 0$  has some benefits, as it implies that after we have made the portfolio Gamma-neutral we can turn into achieving Delta-neutrality by taking a position in the underlying asset. The zero value of Gamma will not be affected by this position. We call the strategy where we are neutral with respect to both Delta and Gamma simultaneously *dynamic Delta-Gamma hedging*.

Say that we hold a portfolio with value  $V$  and given Delta and Gamma,  $\Delta^V$  and  $\Gamma^V$  respectively. We follow a two step procedure where we first achieve Gamma-neutrality, using a liquid contract with known sensitivities. For instance we can employ a European call option with price  $C$  and known Greeks  $\Delta^C$  and  $\Gamma^C$ . In the second step we will use the underlying asset, which has price  $S$  to achieve delta neutrality (recall that  $\Delta^S = 1$  and  $\Gamma^S = 0$ ). The resulting portfolio will be Delta-Gamma neutral.

---

<sup>3</sup> Delta will also change as time passes, even if the asset price remains the same. The Charm  $\frac{\partial^2 V}{\partial S \partial t}$  would quantify this impact. Generally speaking the impact of asset price changes captured with the Gamma are more significant than the Delta changes captured with the Charm. This happens because the magnitude of the squared Brownian increment (captured by Gamma) is of order  $o(\Delta t)$ , while the Charm captures effects of order  $o(\Delta t^{3/2})$ .

FIGURE 2.2: Dynamic Delta hedging of a call option. At time zero we sell a European call with strike price  $K = 100$ , and we Delta hedge it 25 times over its life. The underlying asset process at the hedging times is given in (a), and the number of shares that we need to hold are given in (b). Subfigure (c) gives the corresponding call price and (d) our bank balance. As the option expires out-of-the money we are not asked to deliver at maturity, and the option expires worthless. In (e) changes in the option price and changes in the hedging portfolio are compared. Subfigure (f) illustrates the replication error between the hedging portfolio (solid) and the option (dashed).

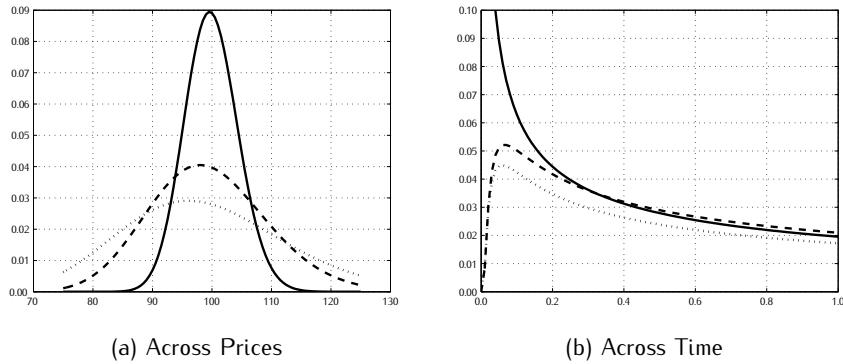


for copies, comments, help etc. visit <http://www.theponytail.net/>

FIGURE 2.3: Sample output of the dynamic Delta hedging procedure. A call option is sold at time  $t = 0$  and is subsequently Delta hedged to maturity

Period	Time	Action	Bank	Shares	Period	Time	Action	Bank	Shares
0	0.000	SEL +1.000 GBP     at £ 2.249    BOR +52.725 GBP    BUY +0.550 SHR     at £100.000       -52.725 0.550			13	0.130	INT -0.003 GBP    BOR -1.205 GBP    BUY -0.013 SHR     at £ 95.951       -12.544 0.134		
1	0.010	INT -0.011 GBP    BOR -4.856 GBP    BUY -0.049 SHR     at £ 99.413       -47.880 0.501			14	0.140	INT -0.003 GBP    BOR +15.478 GBP    BUY +0.158 SHR     at £ 97.933       -28.024 0.292		
2	0.020	INT -0.010 GBP    BOR -11.407 GBP    BUY -0.116 SHR     at £ 98.037       -36.482 0.385			15	0.150	INT -0.006 GBP    BOR +10.517 GBP    BUY +0.106 SHR     at £ 98.943       -38.546 0.399		
3	0.030	INT -0.007 GBP    BOR -0.425 GBP    BUY -0.004 SHR     at £ 98.039       -36.065 0.380			16	0.160	INT -0.008 GBP    BOR +14.469 GBP    BUY +0.145 SHR     at £100.100       -53.023 0.543		
4	0.040	INT -0.007 GBP    BOR -5.757 GBP    BUY -0.059 SHR     at £ 97.381       -30.316 0.321			17	0.170	INT -0.011 GBP    BOR +1.026 GBP    BUY +0.010 SHR     at £100.180       -54.059 0.553		
5	0.050	INT -0.006 GBP    BOR -12.465 GBP    BUY -0.130 SHR     at £ 95.683       -17.857 0.191			18	0.180	INT -0.011 GBP    BOR +3.200 GBP    BUY +0.032 SHR     at £100.395       -57.269 0.585		
6	0.060	INT -0.004 GBP    BOR -2.886 GBP    BUY -0.030 SHR     at £ 95.313       -14.975 0.161			19	0.190	INT -0.011 GBP    BOR +4.883 GBP    BUY +0.048 SHR     at £100.689       -62.164 0.634		
7	0.070	INT -0.003 GBP    BOR -1.502 GBP    BUY -0.016 SHR     at £ 95.177       -13.476 0.145			20	0.200	INT -0.012 GBP    BOR -7.830 GBP    BUY -0.078 SHR     at £100.188       -54.346 0.556		
8	0.080	INT -0.003 GBP    BOR +2.004 GBP    BUY +0.021 SHR     at £ 95.667       -15.482 0.166			21	0.210	INT -0.011 GBP    BOR -26.847 GBP    BUY -0.272 SHR     at £ 98.764       -27.510 0.284		
9	0.090	INT -0.003 GBP    BOR +1.251 GBP    BUY +0.013 SHR     at £ 96.001       -16.736 0.179			22	0.220	INT -0.006 GBP    BOR -20.829 GBP    BUY -0.214 SHR     at £ 97.402       -6.687 0.070		
10	0.100	INT -0.003 GBP    BOR -6.949 GBP    BUY -0.073 SHR     at £ 94.915       -9.791 0.106			23	0.230	INT -0.001 GBP    BOR +10.839 GBP    BUY +0.110 SHR     at £ 98.663       -17.527 0.180		
11	0.110	INT -0.002 GBP    BOR +1.667 GBP    BUY +0.017 SHR     at £ 95.418       -11.460 0.123			24	0.240	INT -0.004 GBP    BOR -14.446 GBP    BUY -0.147 SHR     at £ 98.148       -3.084 0.033		
12	0.120	INT -0.002 GBP    BOR +2.284 GBP    BUY +0.024 SHR     at £ 95.972       -13.745 0.147			25	0.250	INT -0.001 GBP    BOR -3.203 GBP    BUY -0.033 SHR     at £ 98.379    DLV +0.000 SHR     at £100.000       +0.119 0.000		

FIGURE 2.4: Behavior of a call option Gamma. Part (a) gives the behavior of the Gamma of options with specifications  $\{K, r, \sigma\} = \{100, 0.02, 0.20\}$ , and three different times to maturity:  $t = 0.05$  (solid),  $t = 0.25$  (dashed) and  $t = 0.50$  (dotted). Part (b) gives the behavior of the Gamma as the time to maturity increases, for a contract which is at-the-money ( $S = 100$ , solid), in-the-money ( $S = 95$ , dashed), and out-of-the-money ( $S = 105$ , dotted).



We want to buy  $w_C$  units of the option. This makes the value of our composite position equal to  $V + C$ , and most importantly it will have a Gamma equal to  $\Gamma^{V+C} = \Gamma^V + w_C \Gamma^C$ . Therefore, to achieve Gamma-neutrality we need to hold  $w_C = -\frac{\Gamma^V}{\Gamma^C}$  units of the option.

The Delta of the new portfolio is of course  $\Delta^{V+C} = \Delta^V - \frac{\Gamma^V}{\Gamma^C} \Delta^C$ . To make the position Delta-neutral we want to also hold  $w_S = -\Delta^{V+C}$  shares of the underlying asset.

For European call and put options the value of Gamma is given by

$$\Gamma^C = \frac{N'(d_+)}{S\sigma\sqrt{t}}$$

Graphically, figure 2.4 gives Gamma across different moneyness and maturity levels. Apparently the Gamma is significant for contracts that are at-the-money. In particular, the Gamma of at-the-money options goes to infinity as maturity approaches. This is due to the discontinuity of the derivative of the payoff function.

### Dynamic Delta-Gamma hedging

As Gamma is the sensitivity of the Delta with respect to the underlying price  $S$ , we can use a Delta-Gamma neutral strategy to construct a replicating portfolio which is second order accurate in  $S$ . When we Delta hedge over a discrete time interval we introduce replication errors since the Delta of our position will not remain equal to zero as the time changes over this rebalancing interval.

FIGURE 2.5: Dynamic Delta-Gamma hedging of a call option. At time zero we sell a European call with strike price  $K = 100$ , and we Delta-Gamma hedge it 25 times over its life. To do so we use the underlying stock and a call option which has at all points a strike price that is 105% the current spot price. The underlying asset process is the same as in figure 2.2. Subfigure (a) gives the number of options (solid) and shares (dashed) that we need to hold to maintain Delta-Gamma neutrality. The dotted line gives the number of shares that Delta hedge (as in figure 2.2). Subfigure (b) gives the bank balance if we Delta-Gamma hedge (solid) or just Delta hedge (dotted). In (c) changes in the option price and changes in the hedging portfolio are compared. Crosses give the Delta-Gamma hedging deviations, while circles correspond to pure Delta hedging. Finally, subfigure (f) illustrates the replication error between the hedging portfolio (solid) and the option (dashed) which are virtually indistinguishable. The dotted line gives the process of the Delta hedging portfolio.

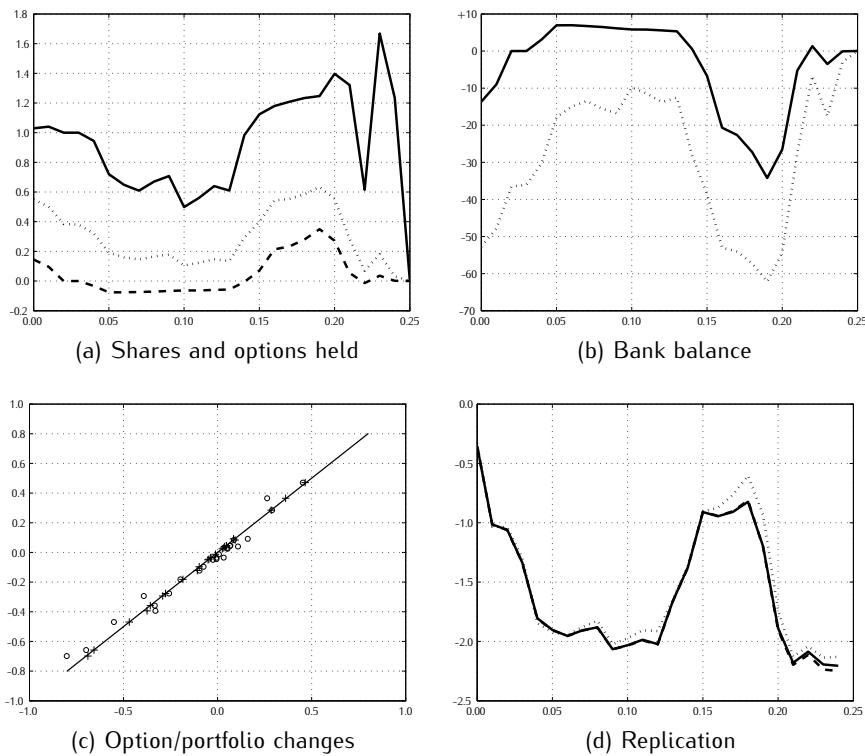
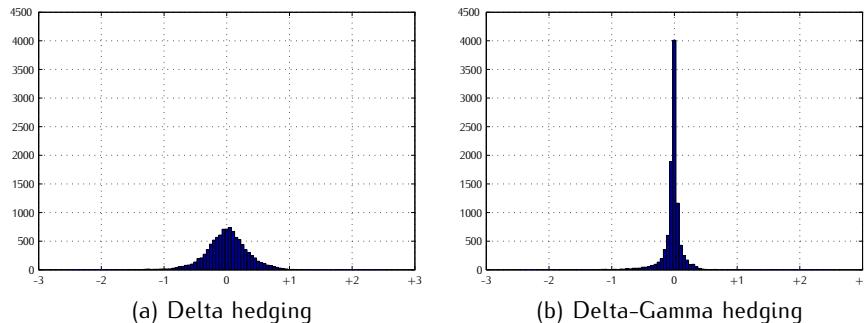


FIGURE 2.6: Comparison of the replication errors for Delta and Delta-Gamma neutral positions. The histograms are based on 10,000 simulations of the underlying asset price. For each run a call option with strike price  $K = 100$  was Delta or Delta-Gamma hedged, as in figures 2.2 and 2.5.



These changes of Delta will be proportional to the derivative  $\frac{\partial \Delta^V}{\partial S} = \frac{\partial^2 V}{\partial S^2} = \Gamma^V$ . Therefore, if we construct a position that has  $\Gamma = \Delta = 0$  we form a portfolio that will maintain a position which is (approximately) neutral for larger price changes, and since the price is diffusive, for longer periods of time.<sup>4</sup>

Of course as we mentioned above we cannot implement such a position using the underlying asset alone, and we will need an instrument that exhibits non-zero Gamma. Typically we use liquid call and put options that are around-the-money to do so. In figure 2.5 we repeat the experiment of figure 2.2 using a Delta-Gamma neutral strategy this time. We sell one call option with strike  $K = 100$  and construct a Delta-Gamma hedge that uses, apart from the underlying asset, a call option. We could use an option with a constant strike price throughout the time to maturity, but there is always the risk that as the underlying price fluctuates this option might become deep-in- or deep-out-of-the-money. Such an option will have  $\Gamma^C \approx 0$  (see figure 2.4), and our position in options  $w_C = -\Gamma^V/\Gamma^C \rightarrow \pm\infty$ . To get around this problem, at each point in time we use a call option that has strike price 105% the value of the underlying asset at this point,  $K^* = 1.05S_t$ . This essentially means that when we rebalance we sell the options we might hold and invest in a brand new contract.<sup>5</sup>

Figure 2.5 gives the processes for this experiment. In subfigure (c) it is easy to see that the Delta-Gamma changes in the portfolio follow the changes of the hedged instrument a lot more closer than the portfolio of figure 2.2 which was only Delta neutral. This improvement in replication accuracy is also illustrated in subfigure (d), where the two processes are virtually indistinguishable.

<sup>4</sup> There is also an error associated with Delta changes as time passes, proportional to the Charm  $\frac{\partial A}{\partial t}$ , but these effects are typically small and deterministic.

<sup>5</sup> Of course if transaction costs were present this would not be the optimal strategy.

We can also repeat the above experiments to assess the *average* performance of simple Delta and Delta-Gamma hedging. Here we create 10,000 simulations of the underlying asset and option prices, and implemented the two hedging strategies. The table below gives the summary statistics for the hedging errors, when we hedge 10, 25 or 50 times during the three-month interval to expiration. Figure 2.6 presents the corresponding histograms for the two hedging strategies, when we rebalance 25 times.

Hedges	Strategy	Mean	St Dev	Min	Max	Skew	Kurt
10	$\Delta$	-0.01	0.52	-3.34	+1.76	-0.47	4.60
	$\Delta\&\Gamma$	-0.12	0.35	-3.76	+1.45	-3.17	19.1
25	$\Delta$	+0.00	0.33	-1.77	+1.30	-0.24	4.30
	$\Delta\&\Gamma$	-0.02	0.17	-2.11	+2.22	-2.63	28.9
50	$\Delta$	-0.00	0.24	-1.50	+1.00	-0.27	4.77
	$\Delta\&\Gamma$	-0.01	0.11	-1.05	+2.28	+0.63	53.3

In comparison, the Delta-Gamma neutral strategy gives pricing errors that are a lot more concentrated around zero, but with significant outliers. This is illustrated in figure 2.6 and the table above. In particular, for all hedging frequencies Delta-Gamma hedging produces half the standard deviation of the errors. On the other hand, for some paths of the underlying asset, implementing Delta-Gamma hedging produces outliers. This is also confirmed by the the table, where the minimum and maximum values and the kurtosis indicate extremely fat tails. Of course, this behavior is dependent on the exact implementation of the hedging strategies, that is to say which instruments are used and how the rebalancing points are selected.

### Gamma and uncertain volatility

The BS-PDE (2.3) depends on two parameters, the risk free rate  $r$ , which is a quantity that is directly observable, and on the volatility  $\sigma$ , which is not. Typically, an options writer will sell contracts based on a conservative estimate of the volatility, say  $\hat{\sigma}$  and subsequently hedge it. It is therefore natural to ask what will the implications be if we hedge our position using a wrong value for  $\sigma$ . It turns out that Gamma has another important role to play, as it will determine the impact of this misspecification. The approach that we follow here is outlined in Carr (2002) and Gatheral (1997, 2006), among others.

To put things concretely, say that the true process for the underlying asset is given by the SDE

$$dS_t = \mu S_t dt + \sigma^A S_t dB_t$$

where the superscript  $\sigma^A$  denotes the actual volatility. We take the position of the writer of a European-style option that offers a payoff  $\Pi(S_T)$  at time  $T$ , and consider its valuation as a function not only of  $(t, S)$ , but also of the volatility  $\sigma$ . For that reason we denote the value of this derivative with  $V(t, S; \sigma)$ . We therefore consider a family of pricing functions for different values of  $\sigma$ , where all satisfy the BS-PDE

$$\frac{\partial}{\partial t} V(t, S; \sigma) + rS \frac{\partial}{\partial S} V(t, S; \sigma) + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2}{\partial S^2} V(t, S; \sigma) = rV(t, S; \sigma)$$

with the appropriate boundary condition  $V(T, S; \sigma) = \Pi(S)$ .

Let us assume that we are asked to sell one such contract, and we quote a price that solves the BS-PDE for a volatility parameter  $\sigma'$ , which we call the *implied volatility*. We can write our quote as  $V'_0 = V(0, S_0; \sigma')$ .

After selling the contract we proceed with the Delta hedging approach. In particular, we implement a self-financing trading strategy  $H = \{(H_t^S, H_t^F) : t \geq 0\}$ , where we  $H_t^S = \Delta_t^H = \frac{\partial}{\partial S} V(t, S_t; \sigma^H)$  units of the underlying asset at each time point, and keep a risk-free bank balance of  $H_t^F$ . Note that when we compute the Delta of the contract we use a third volatility  $\sigma^H$ , that is to say the *hedging volatility*.

The initial bank account balance will be

$$H_0^F = V(0, S_0; \sigma') - \frac{\partial}{\partial S} V(0, S_0; \sigma^H) S_0 \quad (2.6)$$

and the bank account dynamics will be affected at each time by the amount needed to purchase (or sell) stocks to maintain Delta-neutrality, and by interest payments. In particular, at time  $t + dt$  the Delta has changed to  $H_t^S + dH_t^S$ , indicating that we will need to purchase  $dH_t^S$  shares. The price of each share is of course  $S_t + dS_t$  when we make this purchase. Also, over this period we will gain an amount  $rH_t^F dt$  due to the interest on the bank balance. Putting these two together we can write the dynamics of the bank account balance as<sup>6</sup>

$$\begin{aligned} dH_t^F &= -dH_t^S(S_t + dS_t) + rH_t^F dt \\ &= -d(H_t^S S_t) + H_t^S dS_t + rH_t^F dt \\ &= -d(\Delta_t^H S_t) + \Delta_t^H dS_t + rH_t^F dt \end{aligned}$$

The solution of the above SDE can be written as

$$\begin{aligned} \exp(-rT)H_T^F - H_0^F &= -\exp(-rT)\Delta_T^H S_T + \Delta_0^H S_0 \\ &\quad + \int_0^T \exp(-rt) [\Delta_t^H dS_t - r\Delta_t^H S_t dt] \end{aligned} \quad (2.7)$$

Itô's formula will give us the dynamics for the quantity  $V_t^H = V(t, S_t; \sigma^H)$  that gives us the value of Delta that we wish to maintain. In particular,

$$dV_t^H = \left[ \Theta_t^H + \frac{1}{2}(\sigma^A)^2 S_t^2 \Gamma_t^H \right] dt + \Delta_t^H dS_t$$

Since  $V(t, S; \sigma^H)$  satisfies the BS-PDE,  $\Theta^H + \Delta^H rS + \frac{1}{2}(\sigma^H)^2 S^2 \Gamma^H = rV^H$ , we can write

---

<sup>6</sup> The second equation is due to the fact that  $d(X_t Y_t) = X_t dY_t + Y_t dX_t + dX_t dY_t$ , where the others are largely based on  $d(\exp(-rt)X_t) = \exp(-rt)dX_t - r\exp(-rt)X_t$ .

$$dV_t^H = \left[ rV_t^H + \frac{1}{2} \{(\sigma^A)^2 - (\sigma^H)^2\} S_t^2 \Gamma_t^H \right] dt + [\Delta_t^H dS_t - \Delta_t^H rS_t dt]$$

We can solve the above expression for the last square bracket and substitute into the expression for the bank balance dynamics (2.7). This will produce

$$\begin{aligned} \exp(-rT) H_T^F - H_0^F &= -\exp(-rT) \Delta_T^H S_T + \Delta_0^H S_0 \\ &\quad + \int_0^T \exp(-rt) [dV_t^H - rV_t^H dt] \\ &\quad - \frac{1}{2} \{(\sigma^A)^2 - (\sigma^H)^2\} \int_0^T \exp(-rt) S_t^2 \Gamma_t^H dt \end{aligned}$$

Now we can use (2.6) and the fact that  $V_T^H = \Pi(S_T)$  and  $\Delta_T^H = \Pi'(S_T)$  to write the final bank balance in a parsimonious way as

$$\begin{aligned} H_T^F &= \exp(rT) \{ V_0^I - V_0^H \} + \Pi(S_T) - S_T \Pi'(S_T) \\ &\quad - \frac{1}{2} \{(\sigma^A)^2 - (\sigma^H)^2\} \int_0^T \exp(-rt) S_t^2 \Gamma_t^H dt \end{aligned}$$

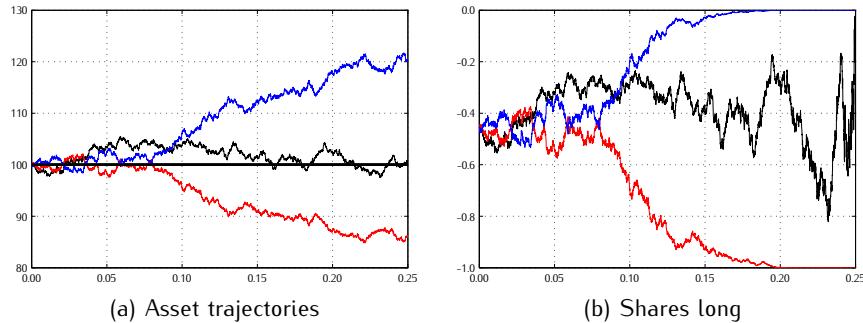
Also, at time  $T$  we are holding  $\Delta_T^H = \Pi'(S_T)$  shares that we will sell, and also deliver the payoff of the derivative contract  $\Pi(S_T)$ . Overall, our profit (or loss) from the Delta hedging strategy will be

$$\begin{aligned} P&L &= \exp(rT) \{ V(0, S_0; \sigma') - V(0, S_0; \sigma^H) \} \\ &\quad + \frac{1}{2} \{(\sigma^H)^2 - (\sigma^A)^2\} \int_0^T \exp(r(T-t)) S_t^2 \Gamma_t^H dt \quad (2.8) \end{aligned}$$

Equation (2.8) is very interesting for a number of reasons. If we happen to know (or be able to estimate fairly accurately) the actual volatility  $\sigma^A$  that we prevail over the life of the contract, then by Delta hedging we can lock in the profit that the difference between the quote  $V(0, S_0; \sigma')$  and the fair value  $V(0, S; \sigma^A)$ , irrespectively of the path of the underlying asset. To do so we should use the actual volatility to compute the Delta of our strategy,  $V_t^H = V_t^A$  for all  $0 \leq t \leq T$ . This happens of course because in this case our dynamic rebalanced portfolio replicates the true payoffs  $\Pi(S_T)$ .

It is likely though that we will not know  $\sigma^A$ , and in this case we might choose to hedge using the implied volatility  $\sigma'$ . Then, the first part of  $P&L$  vanishes, and the final profits will depend on the path of the underlying asset, and will therefore be uncertain. In fact, the sign of the profit will depend on the sign of  $\Gamma_t^H$ . For standard calls and puts  $\Gamma \geq 0$ , which implies that we will always realize profits if once again the implied (and here also hedging) volatility is greater than the realized one,  $\sigma' = \sigma^H > \sigma^A$ . The Gamma for standard calls and puts resembles the underlying probability density, and has its peak around-the-money. This means that the realized profits will be maximum if the underlying

FIGURE 2.7: Delta hedging with uncertain volatility. An at-the-money put is sold, and subsequently Delta-hedged using the implied volatility. Different trajectories of the underlying asset will generate different profits, with the highest when the asset does not trend.



price does not trend upwards or downwards (as this would render the option in- or out-of-the-money).

Figure 2.7 gives an example. We are asked to quote an at-the-money European call ( $S_0 = K = \$100$ ) with maturity three months.<sup>7</sup> The actual volatility over the life of the option is  $\sigma^A = 15\%$ , which indicates that the fair value of this contract is  $V_0^A = \$2.74$ . We agree to sell this option at  $V_0^I = \$3.73$ , which implies a volatility  $\sigma^I = 20\%$ . Essentially the option is overpriced by \$0.99. The figure illustrates three possible trajectories, where the underlying asset moves up, down or sideways over the life of the option.

We might know the future actual volatility, in which case we can select  $\sigma^H = 15\%$ . If we do not, we can hedge at the implied volatility  $\sigma^H = 20\%$ . The following table gives the profits realized using each sample path, with 5,000 rebalances over the three month period (about 60 per day). One can observe that in the case where the asset does not trend, using  $\sigma^I$  outperforms  $\sigma^A$ . Also, note that when the asset moves sideways, even such a frequent rehedging strategy is not identical to the continuous one.

	P&L when asset moves		
	up	down	sideways
$\sigma^H = \sigma^A = 15\%$	+\$0.99	+\$0.99	+\$0.92
$\sigma^H = \sigma^I = 20\%$	+\$0.51	+\$0.57	+\$1.44

## VEGA

We have already highlighted the dependence of derivative contracts on the volatility of the underlying asset. The BS methodology makes the assumption

<sup>7</sup> The drift of the underlying asset is  $\mu = 8\%$ , and the risk free rate of interest is  $r = 2\%$ .

that the volatility is constant across time, but practitioners routinely compute the sensitivity of their portfolios with respect to the underlying volatility, and in some cases try to hedge against volatility changes. Of course, in order to be precise one should start with a model that specifies a process for the volatility of the asset, and not the BS framework where the volatility is constant. Then, sensitivities with respect to the spot volatility are in principle computed in a straightforward matter, exactly as we compute the BS Delta.

In practice, practitioners use the Black-Scholes Vega instead. It might appear counterintuitive to use the derivative with respect to a constant, but it offers a good (first order) approximation. Unless the rebalancing intervals are too long, or the volatility behaves in an erratic or discontinuous way, the Vega is fairly robust and easy to compute and use. We follow the last subsection and consider the value of the portfolio as a function of the volatility  $\sigma$  (in addition to  $(t, S)$ ),  $V = V(t, S; \sigma)$ . Then, applying Taylor's expansion yields

$$\Delta V = \Theta \Delta t + \Delta \Delta S + \frac{1}{2} \Gamma \Delta S^2 + v \Delta \sigma + o(\Delta t, \Delta S^2, \Delta \sigma)$$

The underlying asset price does not depend explicitly on the volatility, rendering  $v^S = 0$ . Once more we need to rely on nonlinear contracts, such as options, to make a portfolio *Vega-neutral*. If we want to achieve joint Gamma-Vega-neutrality hedge, we will have to use two *different* derivative securities.

Say that we use two options with prices  $C_1$  and  $C_2$ , with known deltas ( $\Delta^{C_1}$  and  $\Delta^{C_2}$ ), Gammas ( $\Gamma^{C_1}$  and  $\Gamma^{C_2}$ ) and Vegas ( $v^{C_1}$  and  $v^{C_2}$ ). We also use the underlying asset to achieve delta neutrality (of course  $\Delta^S = 1$  and  $\Gamma^S = v^S = 0$ ). We want to buy  $w_{C_1}$  and  $w_{C_2}$  units of the two derivative securities to achieve Gamma-Vega neutrality. We are therefore faced with the system

$$\begin{aligned}\Gamma^{V+C_1+C_2} &= \Gamma^V + w_{C_1} \Gamma^{C_1} + w_{C_2} \Gamma^{C_2} = 0 \\ v^{V+C_1+C_2} &= v^V + w_{C_1} v^{C_1} + w_{C_2} v^{C_2} = 0\end{aligned}$$

This will identify the holdings of the two derivatives

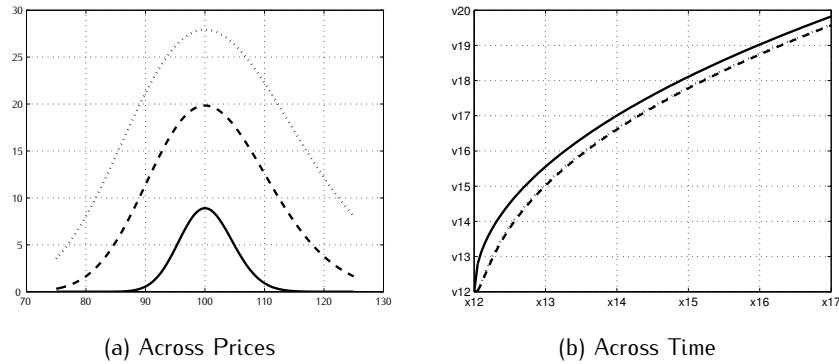
$$w_{C_1} = -\frac{\Gamma^V v_{C_2} - \Gamma^{C_2} v^V}{\Gamma^{C_1} v_{C_2} - \Gamma^{C_2} v^{C_1}} \quad w_{C_2} = -\frac{\Gamma^{C_1} v^V - \Gamma^V v^{C_1}}{\Gamma^{C_1} v^{C_2} - \Gamma^{C_2} v^{C_1}}$$

After that we can adjust our holdings of the underlying asset to make our position Delta-neutral as well. For a European call or put option the BS value of Vega is given by

$$v^C = S \sqrt{t} N'(d_+)$$

Graphically, the Vega across different moneyness and maturity levels is given in figure 2.8. It is straightforward to observe that Vega, like Gamma, is more pronounced for at-the-money options. Unlike Gamma though, the Vega drops as we move closer to the maturity. Thus, to achieve Vega neutrality one should incorporate long dated at-the-money options in her portfolio.

FIGURE 2.8: Behavior of a call option Vega. Part (a) gives the behavior of the Vega of options with specifications  $\{K, r, \sigma\} = \{100, 0.02, 0.20\}$ , and three different times to maturity:  $t = 0.05$  (solid),  $t = 0.25$  (dashed) and  $t = 0.50$  (dotted). Part (b) gives the behavior of the Vega as the time to maturity increases, for a contract which is at-the-money ( $S = 100$ , solid), in-the-money ( $S = 95$ , dashed), and out-of-the-money ( $S = 105$ , dotted).



### DIVIDENDS AND FOREIGN EXCHANGE OPTIONS

In the above analysis we have ignored the impact of dividends, just to keep things simple. When a stock pays a continuous dividend at a constant rate  $q$ , the process of the underlying asset under  $\mathcal{Q}$  is given by the GBM

$$dS_t = (r - q)S_t dt + \sigma S_t dB_t^{\mathcal{Q}}$$

Derivatives will be given once again as expectations  $P_0 = \exp(-rT)E^{\mathcal{Q}}\Pi(S_T)$ , and their pricing function  $P_t = f(t, S_t)$  will satisfy the PDE

$$\frac{\partial}{\partial t}f(t, S) + (r - q)S\frac{\partial}{\partial S}f(t, S) + \frac{1}{2}\sigma^2 S^2\frac{\partial^2}{\partial S^2}f(t, S) = f(t, S)r$$

The prices and the Greeks can be computed easily following the same steps. In particular we can summarize the most useful Greeks in the following catalogue, where  $\lambda = +1$  for calls and  $\lambda = -1$  for puts, and

$$d_{\pm} = \frac{\log(S_0/K) + (r - q \pm \sigma^2/2)(T - t)}{\sigma\sqrt{(T - t)}}$$

- Option price  $P = V(t, S)$

$$P = \lambda S e^{-q(T-t)} N(\lambda d_+) - \lambda K e^{-r(T-t)} N(\lambda d_-)$$

- Delta  $\Delta = \frac{\partial V(t, S)}{\partial S}$

$$\Delta = \lambda e^{-q(T-t)} N(\lambda d_+)$$

- Theta  $\Theta = -\frac{\partial V(t, S)}{\partial t}$

$$\Theta = \frac{SN'(d_+) \sigma e^{-q(T-t)}}{2\sqrt{T-t}} - \lambda q SN(\lambda d_+) e^{-q(T-t)} + \lambda r K e^{-r(T-t)} N(\lambda d_-)$$

- Gamma  $\Gamma = \frac{\partial^2 V(t, S)}{\partial S^2}$

$$\Gamma = \frac{N'(d_+) e^{-q(T-t)}}{S \sigma \sqrt{T-t}}$$

- Vega  $\nu = \frac{\partial V(t, S; \sigma)}{\partial \sigma}$

$$\nu = S \sqrt{T-t} N'(d_+) e^{-q(T-t)}$$

- Rho  $\rho = \frac{\partial V(t, S; r)}{\partial r}$

$$\rho = \lambda K T e^{-r(T-t)} N(\lambda d_-)$$

- Dividend-rho  $\rho^q = \frac{\partial V(t, S; q)}{\partial q}$

$$\rho^q = \lambda S T e^{-q(T-t)} N(\lambda d_+)$$

Similar expressions can be derived for foreign exchange rates. In particular, the exchange rate (denominated in the domestic currency) under risk neutrality is assumed to follow the GBM

$$dS_t = (r^d - r^f) S_t dt + \sigma S_t dB_t^Q$$

In essence holding the foreign currency will depreciate at the risk free rate differential. Therefore it is straightforward to confirm that the formulas for the option prices and their Greeks above will hold, where  $r \rightsquigarrow r^d$  and  $q \rightsquigarrow r^f$ .

## 2.5 IMPLIED VOLATILITIES

As we have pointed out a few times already, the parameters of the BS formula are all considered to be  $\mathcal{F}_0$ -measurable by assumption. In reality though, although the current price, the strike price, the maturity and the interest rate are observed at time  $t = 0$ , the volatility  $\sigma$  of the asset price is not. Since an array of call and put options are also available with prices that are observed at time zero, one will naturally attempt to invert numerically the BS formula and construct a series of *implied volatilities*  $\{\hat{\sigma}(T, K)\}$  across different maturities and strike prices. [Bajeux and Rochet \(1996\)](#) show that there is a one-to-one relationship between implied volatilities and option prices. As pointed out in [Dupire \(1994\)](#), these implied volatilities will indicate how the underlying asset should vibrate in the BS world, in order for the contract to be priced correctly. Following our discussion in the previous section, these volatilities would be natural candidates to compute the Delta that will hedge the option.

If the assumptions underlying the BS formula were correct, then all prices would be priced according to the BS formula, and therefore we should extract the same implied volatility from all options (that is for any maturity and strike price combination). It turns out though, that these implied volatilities are not constant, and in fact exhibit some very clear and persistent patterns. We will see that these patterns can be attributed to actual volatilities that are time varying, to discontinuities in the asset price process, to hedging demands for specific option contracts, and to liquidity premiums for some specific groups of options.

The variation of volatility is a well documented feature of asset returns, and models that incorporate stochastic volatilities, first introduced in Hull and White (1987, HW), give the theoretical background to interpret the implied volatility as the expectation of the average future (realized) volatility over the life of the option.

Say that we are considering the price of a European call, that is  $\Pi(S_T) = (S_T - K)^+$ . We assume, that the future volatility  $\sigma_t$  is stochastic but independent of the stock price, and also has zero price of risk.<sup>8</sup> The main idea of HW is to condition on the average variance over the life of the option, namely the random variable

$$\gamma^2 = \frac{1}{T} \int_0^T \sigma_t^2 dt$$

Then, using the tower property we can write the option price as

$$P_0 = \exp(-rT) E^Q \Pi(S_T) = \exp(-rT) E_\gamma^Q [E^Q[\Pi(S_T)|\gamma]]$$

where the outermost expectation is with respect to all possible realizations of  $\gamma$ .

It turns out that the conditional option prices are equal to their Black-Scholes counterparts, with  $\sigma \rightsquigarrow \gamma$ . Thus, we can write the HW prices as a weighted sum of BS prices

$$f_{HW}(t, S; K, T, r, \dots) = E_\gamma^Q f_{BS}(t, S; K, r, \gamma)$$

In the above expression the dots represent parameters that govern the volatility dynamics, and  $f$ . are the corresponding pricing functions.

If we now consider an at-the-money option, where the strike is set at the forward price  $K_{ATM} = S \exp(rT)$ , then the HW formula will give

$$f_{HW}(t, S; K_{ATM}, r, \dots) = E^Q S \left[ 2N \left( \frac{\gamma}{2} \sqrt{T-t} \right) - 1 \right]$$

On the other hand, if  $P_{ATM}$  is the observed price, the ATM implied volatility will solve

$$P_{ATM} = f_{BS}(t, S; K, r, \hat{\sigma}_{ATM}) = S \left[ 2N \left( \frac{\hat{\sigma}_{ATM}}{2} \sqrt{T-t} \right) - 1 \right]$$

Assuming that the HW model is the correct model,  $f_{HW}(t, S; K_{ATM}, r, \dots) = P_{ATM}$ , we have the relationship

---

<sup>8</sup> Intuitively this means that the volatility risk is diversifiable, or that investors are indifferent to the level of volatility risk. We will come back to these issues later.

$$N\left(\frac{\hat{\sigma}_{ATM}}{2}\sqrt{T}\right) = E^Q N\left(\frac{\gamma}{2}\sqrt{T}\right)$$

Assuming short maturities,  $t \approx T$ , the cumulative normal density is approximately linear around zero, which yields the approximate relationship

$$\hat{\sigma}_{ATM} \approx E^Q \sqrt{\frac{1}{T-t} \int_t^T \sigma_s^2 ds}$$

Thus the implied ATM volatility is approximately equal to the expected average volatility over the life of the option.

## 2.6 STYLIZED FACTS

The log-normality of the asset price distribution, a result of the GBM that underlies the BS derivation, is not a satisfactory assumption. In fact, it has been documented that equity prices do not follow such a distribution even from the PhD dissertation of [Bachelier \(1900\)](#). Nonetheless, the BS methodology results into a formula that is intuitive and very easy to implement in practice, and therefore it is widely used both for academic and practical purposes. In fact, options in exchanges are actually quoted with their implied volatilities rather than in dollar or sterling terms. In addition, the fact that the volatility of the underlying asset and the risk free rate of return are assumed constant, simplifies the exposition, by forcing the markets to be complete.

Testing the BS model gives rise to many theoretical and practical problems. If we use actual option prices to carry out such tests, we cannot distinguish between the potential mis-specifications of the pricing formula and market inefficiencies. The joint hypothesis that the correct model is used and that the markets are efficient is necessarily tested (for a discussion see for example [Hull, 2003](#)). The fact that at any time a parameter of the BS model is actually unobserved further complicates things, as it is not clear which one to use. A third problem arises from the possible asynchronicity of the equity, bond and option markets. If trading does not take place simultaneously, or the market are very thin, it is questionable if the assumption of completeness is satisfactory. Not having data on synchronous transactions in liquid markets distorts the results.

The patterns of the implied volatilities summarize many of the failures of the BS model, and researchers have been looking at them closely since good quality data became available. An early analysis is the seminal paper of [Rubinstein \(1985\)](#), where different patterns of implied volatilities emerge, depending largely on the particular period that was used, with predominantly a U-shaped pattern with the lowest point at-the-money. In the more recent work of [Rubinstein \(1994\)](#) and [Jackwerth and Rubinstein \(1996\)](#) implied volatilities tend to be higher for out-of-the-money puts and lower for out-of-the-money calls. These emerging pattern of implied volatilities with respect to different measures of moneyness is

often encountered in the literature as the *implied volatility smile*, *skew* or *smirk*. If we create a three-dimensional view of the implied volatility with respect to the moneyness and the time to maturity we construct the *implied volatility surface*. Figure XX presents such a surface based on options data on the FTSE100.

These implied volatility patterns can be attributed to some of the best documented stylized facts of the distribution and dynamics of asset returns (two excellent surveys are [Bollerslev, Engle, and Nelson, 1994](#), and [Ghysels, Harvey, and Renault, 1996](#)). Below we shall give a small overview of these features and discuss how they are reflected on the implied volatility surface.

### Leptokurtosis

It has been long observed that asset returns follow a distribution which is far from normal, in particular one that exhibits a substantial degree of excess kurtosis or fat tails ([Fama, 1965](#)). These fat tails seem to be more pronounced for short investment horizons (ie intraday, daily or weekly returns), and they tend to gradually die out for longer ones (ie monthly, quarterly or annual returns). A distribution with high kurtosis is consistent with the presence of an implied volatility smile, as it attaches higher probabilities to extreme events, compared to the normal distribution. If the at-the-money implied volatility is used, then the BS formula will underprice out-of-the-money puts and calls. A higher implied volatility is needed for the BS formula to match the market prices. [Merton \(1976\)](#) among others, notes that a mixture of normal distributions can exhibit fat tails relative to the normal, and therefore models that result in such distributions can be used in order to improve on the BS option pricing results. Most (if not all) modern option pricing models to some extend do exactly that: expressing calendar returns as a mixture of normal distributions.

### Skewness

Apart from exhibiting fat tails, some asset return series also exhibit significant skewness. For stocks and indices this skewness is typically negative, highlighting the fact that the speed that stock prices drop is higher than the speed they grow (although they tend to grow for longer periods than they decline). For currencies the skew is not generally one sided, swinging from positive to negative and back, over periods of time. The asymmetries of the implied volatility skew can be attributed to the skewness of the underlying asset returns. In prices are more likely to drop by a large amount than rise, one would expect out-of-the-money puts to be relatively more expensive than out-of-the-money calls. [Black \(1972\)](#) suggests that volatilities and asset returns are negatively correlated, naming this phenomenon the *leverage effect* or *Fisher-Black effect*. Falling stock prices imply an increased leverage on firms, which is presumed by agents to entail more uncertainty, and therefore volatility. This asymmetry can generate skewed returns, but is not always sufficient to explain the very steep implied skews we observe in (especially index) options markets. A second component that is

needed is accommodating for market crashes arriving as jumps in the asset price process, or even just fears of such crashes (the *crash-o-phobia* of Bates, 1998).

### Volatility features

The fact that volatility is not constant is well documented, and allowing it to be time varying is perhaps the simplest way to construct models that mix normal distributions. Empirically, it appears that volatility in the market comes in cycles, where low volatility periods are followed by high volatility episodes. This feature is known in the literature as *volatility clustering*. The *Arch*, *Garch* and *Egarch* families<sup>9</sup>, as well as models with *stochastic volatility* have been used in the literature to model the time variation of volatility and model volatility clustering. The survey of Ghysels et al. (1996) gives a good overview of volatility models from a modeling perspective. *Local volatility models* take a completely different approach, as they focus solely on the pricing and hedging of derivatives, preferring to keep volatility time-varying but deterministic rather than stochastic (Dupire, 1994). We will discuss these extensions in chapter 6.

The variation of volatility can be linked to the arrivals of information, and high trading volume (Mandelbrot and Taylor, 1967; Karpoff, 1987, among others). One can argue that trading does not take place in a uniform fashion across time: new information will result in a more dense trading pattern with higher trading volumes, which in turn result in higher volatilities.

### Price discontinuities

Even allowing the volatility to be time varying cannot accommodate for very sharp changes in the stock price, typically crashes, which although are very rare events, have a significant impact on the behavior of the market. On October 19th, 1987, the S&P500 index lost about 20% of its value within a day and without any significant warnings. If the market was to follow the Black-Scholes assumption of a GBM with constant volatility, such an event should happen once in  $10^{87}$  years.<sup>10</sup> Even if we allow the volatility to vary wildly, a model with continuous sample paths that will exhibit such a behavior is not plausible.

Starting with Merton (1976), researchers have been augmenting the diffusive part of the price process with

---

<sup>9</sup> Arch here stands for autoregressive conditional heteroscedasticity (Engle, 1982), Garch stands for generalized Arch (Bollerslev, 1986), and Egarch for exponential Garch (Nelson, 1991)

<sup>10</sup> This is a very long time. For a comparison, the age of our universe is estimated to be about  $10^{24}$  years.

# 3

---

## Finite difference methods

The [Black and Scholes \(1973\)](#), BS partial differential equation (PDE) is, as we saw, one of the most fundamental relationships in finance. It is as close to a *law* as we can get in a discipline that deals with human activities. The importance of the expression stems from the fact that it must be satisfied by *all* derivative contracts, independently of their contractual features. In some special cases, for example when the contract in question is a European-style option, the solution of the PDE can be computed in closed-form, but this is not the general case. In many real situations we will have to approximate the solution of the PDE numerically.

If  $t$  denotes time and  $S = S(t)$  is the value of the underlying asset, the BS model assumes that  $S$  follows a geometric Brownian motion

$$dS(t) = \mu S(t)dt + \sigma S(t)dB(t)$$

It follows that, for any derivative contract, the pricing function  $f = f(t, S)$  will satisfy the BS PDE

$$-\frac{\partial f(t, S)}{\partial t} + \mu S \frac{\partial f(t, S)}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 f(t, S)}{\partial S^2} = rf(t, S) \quad (3.1)$$

where  $S$  is the price of the asset and  $t$  is the time to maturity. Equation (3.1) is not sufficient to uniquely specify  $f$ , initial and perhaps a number of boundary conditions are also needed for (3.1) to admit a unique solution. In fact, different derivative contracts will impose different initial and boundary conditions, but (3.1) must be satisfied by all of them. For example, the standard call option will impose the initial condition

$$f(0, S) = \max(S - K, 0)$$

*Finite difference methods* (FDMs) is the generic term for a large number of procedures that can be used for solving a (partial) differential equation, which have as a common denominator some discretization scheme that approximates the

required derivatives. In this chapter we will give an overview of these methods and also examine some examples that illustrate the methodology in financial engineering. Thomas (1995) gives a detailed overview of different approaches, together with exhaustive analysis of the consistency, convergence and stability issues. Wilmott, Dewynne, and Howison (1993) present FDMs within an option pricing framework.

## 3.1 DERIVATIVE APPROXIMATIONS

Before we turn to the fully fledged PDE (3.1), let us assume for a moment that we are given a one-dimensional function  $h = h(x)$ . Our goal is to provide some estimate of the derivative of  $h$  at the point  $\bar{x}$ , namely  $h'(\bar{x}) = \frac{dh(\bar{x})}{dx}$ . We can express the derivative using three different expressions that involve limits towards  $\bar{x}$ :

$$\frac{dh(\bar{x})}{dx} = \begin{cases} \lim_{\Delta x \rightarrow 0} \frac{h(\bar{x} + \Delta x) - h(\bar{x})}{\Delta x} \\ \lim_{\Delta x \rightarrow 0} \frac{h(\bar{x}) - h(\bar{x} - \Delta x)}{\Delta x} \\ \lim_{\Delta x \rightarrow 0} \frac{h(\bar{x} + \Delta x) - h(\bar{x} - \Delta x)}{2\Delta x} \end{cases}$$

For a differentiable function all three limits are equal, and suggest three candidates for discrete approximations for the derivative. In particular we can construct:

1. The *right limit* yields the *forward differences approximation scheme*

$$\frac{dh(\bar{x})}{dx} \approx \frac{h(\bar{x} + \Delta x) - h(\bar{x})}{\Delta x}$$

2. The *left limit* yields the *backward differences approximation scheme*

$$\frac{dh(\bar{x})}{dx} \approx \frac{h(\bar{x}) - h(\bar{x} - \Delta x)}{\Delta x}$$

3. The *central limit* yields the *central differences approximation scheme*

$$\frac{dh(\bar{x})}{dx} \approx \frac{h(\bar{x} + \Delta x) - h(\bar{x} - \Delta x)}{2\Delta x}$$

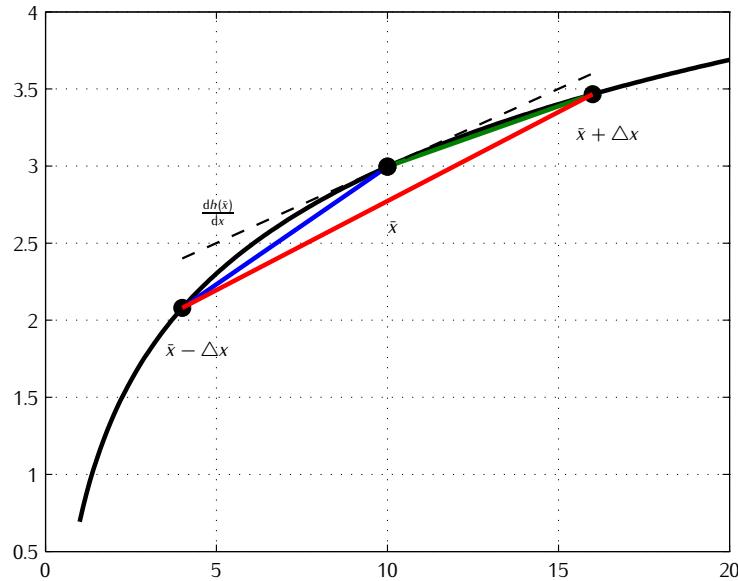
These schemes are illustrated in figure 3.1, where the true derivative is also given for comparisons. Of course the approximation quality will depend on the salient features of the particular function, and in fact, it turns out to be closely related to the behaviour of higher order derivatives.

Let us now assume that we have discretized the support of  $h$  using a uniform grid,  $\{x_i\}_{i=-\infty}^{\infty}$  with  $x_i = x_0 + i \cdot \Delta x$ , and define the values of the function  $h_i = h(x_i)$ . Then, we can introduce the corresponding *difference operators*  $\mathcal{D}_+$ ,  $\mathcal{D}_-$  and  $\mathcal{D}_0$ , and rewrite the difference approximations in shorthand<sup>1</sup> as

---

<sup>1</sup> For us these operators serve as a neat shorthand for the derivative approximations, but there is, in fact, a whole area of *difference calculus* that investigates and exploits their properties.

FIGURE 3.1: Finite difference approximation schemes. The forward (green), backward (blue) and central (red) differences approximation schemes, together with the true derivative (dashed).



$$\text{Forward: } \mathcal{D}_+ h_i = \frac{h_{i+1} - h_i}{\Delta x}$$

$$\text{Backward: } \mathcal{D}_- h_i = \frac{h_i - h_{i-1}}{\Delta x}$$

$$\text{Central: } \mathcal{D}_0 h_i = \frac{h_{i+1} - h_{i-1}}{2\Delta x}$$

What are the properties of these schemes and which one is more accurately representing the true derivative? A first inspection of figure 3.1 reveals that the central differences approximation is closer to the true derivative, but is this generally true? In order to formally assess the quality of the approximations we will use Taylor expansions of  $h$  around the point  $x_i$ , that is to say the expansions of the points  $h_{i\pm 1}$ :

$$h_{i+1} = h_i + \frac{dh(x_i)}{dx} \Delta x + \frac{1}{2} \frac{d^2h(x_i)}{dx^2} \Delta x^2 + \frac{1}{6} \frac{d^3h(x_i)}{dx^3} \Delta x^3 + \dots$$

$$h_{i-1} = h_i - \frac{dh(x_i)}{dx} \Delta x + \frac{1}{2} \frac{d^2h(x_i)}{dx^2} \Delta x^2 - \frac{1}{6} \frac{d^3h(x_i)}{dx^3} \Delta x^3 + \dots$$

Substituting the corresponding values in the approximation schemes will yield the important relationships

$$\begin{aligned}\mathcal{D}_+ h_i &= \frac{dh(x_i)}{dx} + o(\Delta x) \\ \mathcal{D}_- h_i &= \frac{dh(x_i)}{dx} + o(\Delta x) \\ \mathcal{D}_0 h_i &= \frac{dh(x_i)}{dx} + o(\Delta x^2)\end{aligned}$$

In the above expressions we introduce the *big-O* notation, where  $o(\Delta x^n)$  includes all terms of order  $\Delta x^n$  and smaller.<sup>2</sup> Now since  $|\Delta x^2| \ll |\Delta x|$  around zero, it follows that the terms  $|o(\Delta x^2)| \ll |o(\Delta x)|$ , which means that central differences are more accurate than forward or backward differences. We say that central differences are second order accurate while forward and backward differences are first order accurate.

Therefore, without any further information on the function, we should use central differences where possible. If we have some extra information, perhaps using one-sided derivatives might be beneficial. Such cases could arise when the drift term dominates the PDE, or alternatively when the volatility is very small. In our setting though we will concentrate on approximations that use central differences as their backbone.

The BS PDE also involves second order derivatives, on top of the first order ones. We therefore need to establish an approximation scheme for these second derivatives. When we achieved that we will be able to proceed to the actual discretization of the BS PDE (3.1). Since we are trying to establish second order accuracy, we are looking for a scheme that approximates the second derivatives using central differences.

It turns out that an excellent choice is an approximation that takes central differences twice over a half-step  $\frac{\Delta x}{2}$ .

$$\mathcal{D}^2 h_i = \frac{\frac{dh(x_{i+1/2})}{dx} - \frac{dh(x_{i-1/2})}{dx}}{\Delta x} = \frac{\frac{h_{i+1}-h_i}{\Delta x} - \frac{h_i-h_{i-1}}{\Delta x}}{\Delta x} = \frac{h_{i+1} - 2h_i + h_{i-1}}{\Delta x^2}$$

Using the same substitutions from the Taylor expansions as above yields

$$\mathcal{D}^2 h_i = \frac{d^2 h(x_i)}{dx^2} + o(\Delta x^2)$$

Therefore, we conclude that the operator  $\mathcal{D}^2$  is second order accurate. In addition  $\mathcal{D}^2$  has the advantage that in order to compute it we use the same values that were needed for the first difference  $\mathcal{D}_0$ , namely  $h_{i\pm 1}$ , and the value  $h_i$ .

---

<sup>2</sup> Formally, if a function  $g = g(x)$  is  $o(\Delta x^n)$  then the limit of the ratio  $\frac{|g(x)|}{|\Delta x^n|} < C < \infty$  (meaning that it is bounded) as  $x \rightarrow 0$ . Intuitively,  $g(x)$  approaches zero at the same speed as  $\Delta x^n$ . We say that  $g$  is of order  $n$ .

## 3.2 PARABOLIC PDEs

The BS PDE belongs to a wide and well documented class of PDEs called *parabolic partial differential equations*. Many important natural phenomena are associated with parabolic PDEs, ranging from Einstein's heat equation to Schrödinger's description of quantum mechanics. In order to simplify the subsequent notation we use the shorthand *elliptic operator*

$$\mathcal{L}f(t, x) = \alpha(t, x) \frac{\partial f(t, x)}{\partial x} + \beta(t, x) \frac{\partial^2 f(t, x)}{\partial x^2} + \gamma(t, x) f(t, x)$$

for general functionals  $\alpha$ ,  $\beta$  and  $\gamma$ . Therefore the BS PDE will be of the general form

$$\frac{\partial f(t, S)}{\partial t} = \mathcal{L}f(t, S) \quad (3.2)$$

Suppose that we work on a grid  $x = \{x_j\}_{j=-\infty}^{+\infty}$ , with constant grid spacing equal to  $\Delta x$ . We will concentrate on an initial value problem, and therefore assume that the function extends over the whole real line. The problem of boundary conditions will be addressed later in this chapter. We will also define the value function at the grid points,  $f_j(t) = f(t, x_j)$ ,  $j = -\infty, \dots, +\infty$ . We construct the discretized operator by applying the differences  $\mathcal{D}_0$  and  $\mathcal{D}^2$

$$\mathcal{L}f_j(t) = \alpha_j(t) \cdot \mathcal{D}_0 f_j(t) + \beta_j(t) \cdot \mathcal{D}^2 f_j(t) + \gamma_j(t) \cdot f_j(t)$$

In the above expression the functionals  $\alpha_j$ ,  $\beta_j$  and  $\gamma_j$  are just the restrictions of  $\alpha$ ,  $\beta$  and  $\gamma$  on the grid point  $x_i$ . Substituting the difference operators gives

$$\begin{aligned} \mathcal{L}f_j(t) &= \alpha_j(t) \cdot \frac{f_{j+1}(t) - f_{j-1}(t)}{2\Delta x} \\ &\quad + \beta_j(t) \cdot \frac{f_{j+1}(t) - 2f_j(t) + f_{j-1}(t)}{\Delta x^2} + \gamma_j(t) \cdot f_j(t) \end{aligned}$$

Our goal was to construct a discretized operator that, in some sense, converges to the actual operator as the discretization becomes finer, or somehow " $\mathcal{L} \rightarrow \mathcal{L}'$ ". Essentially, since we want to establish convergence we will need a measure of *distance* between the operators. We will discuss these issues in more detail in section 3.2, following the introduction to the explicit method. After establishing this convergence we will move forward and approximate the PDE itself at the point  $x_j$  with

$$\begin{aligned} \frac{\partial f_j(t)}{\partial t} &= \mathcal{L}f_j(t) \Leftrightarrow \\ \frac{\partial f_j(t)}{\partial t} &= q_j^+(t) \cdot f_{j+1}(t) + q_j^0(t) \cdot f_j(t) + q_j^-(t) \cdot f_{j-1}(t) \quad (3.3) \end{aligned}$$

The functionals  $q_j^\pm(t)$  and  $q_j^0(t)$  depend on the structure of the PDE and are given by

$$q_j^\pm(t) = \pm \alpha_j(t) \frac{1}{2\Delta x} + \beta_j(t) \frac{1}{\Delta x^2}$$

$$q_j^0(t) = \gamma_j(t) - 2\beta_j(t) \frac{1}{\Delta x^2}$$

## A PDE AS A SYSTEM OF ODEs

Since (3.3) will hold for all grid points  $\{x_j\}_{j=-\infty}^{+\infty}$ , we have represented the discretized problem (3.3) as a *system of ODEs*, which can be cast in matrix form for  $\mathbf{f}(t) = \{f_j(t)\}_{j=-\infty}^{+\infty}$

$$\frac{\partial \mathbf{f}(t)}{\partial t} = \mathbf{Q}(t) \cdot \mathbf{f}(t) \quad (3.4)$$

subject to the initial condition  $\mathbf{f}(0) = \{f(x_j, 0)\}_{j=-\infty}^{+\infty}$ . Equation (3.4) describes the evolution of the pricing function  $f$  as the time to maturity increases. From this point on we will make the additional assumption that the matrix  $\mathbf{Q}(t)$  is time-invariant,  $\mathbf{Q}(t) = \mathbf{Q}$ . Time dependence can be accommodated in a straightforward way in the numerical implementation. The matrix  $\mathbf{Q}$  is tridiagonal, in particular

$$\mathbf{Q} = \begin{pmatrix} \ddots & \ddots & \ddots & & & \\ & q_{-1}^- & q_0^0 & q_1^+ & 0 & 0 \\ & 0 & q_0^- & q_0^0 & q_1^+ & 0 \\ & 0 & 0 & q_{+1}^- & q_{+1}^0 & q_{+1}^+ \\ & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

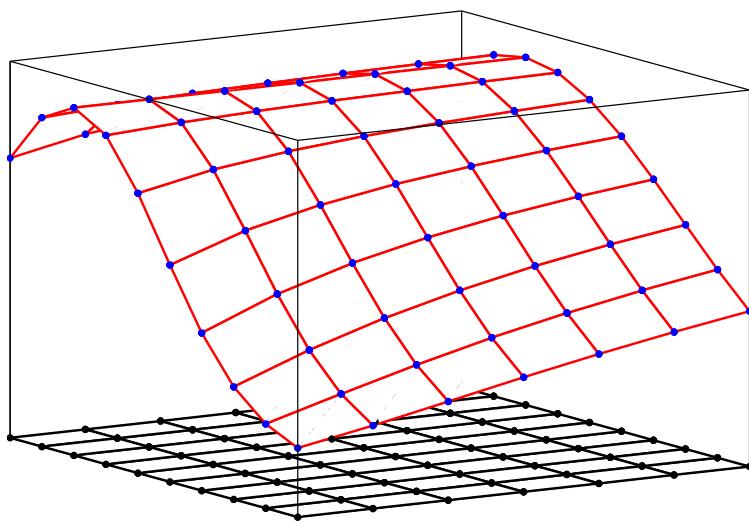
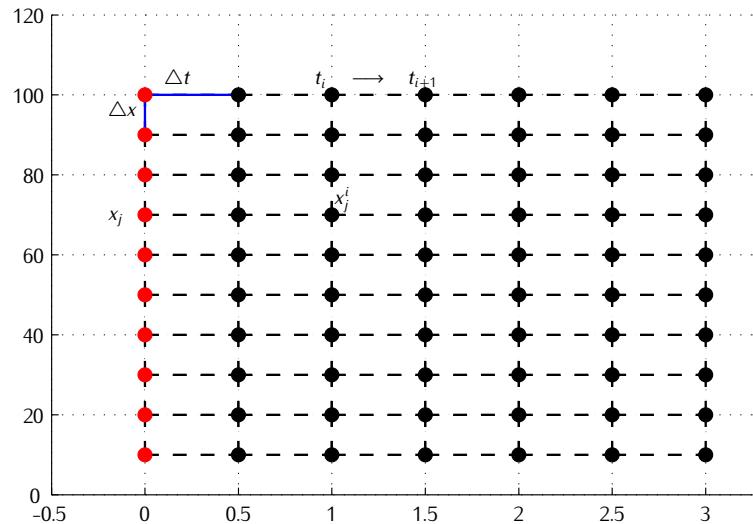
There is a large number of solvers for such systems. We will consider methods that apply time discretization as well, and therefore work on a two-dimensional grid.

## THE GRID

In equation (3.4) we converted the PDE in question into a system of infinite ODEs. Apparently it is not feasible in practice to numerically solve systems with an infinite number of equations. We will therefore need to truncate the grid and consider a subset with  $N_x$  elements  $\mathbf{x} = \{x_j\}_{j=1}^{N_x}$ . This means that we will need to take special care on the treatment of the numerical approximations at the artificial boundaries  $x_1$  and  $x_{N_x}$ . We will discuss these issues in detail in section 3.2.

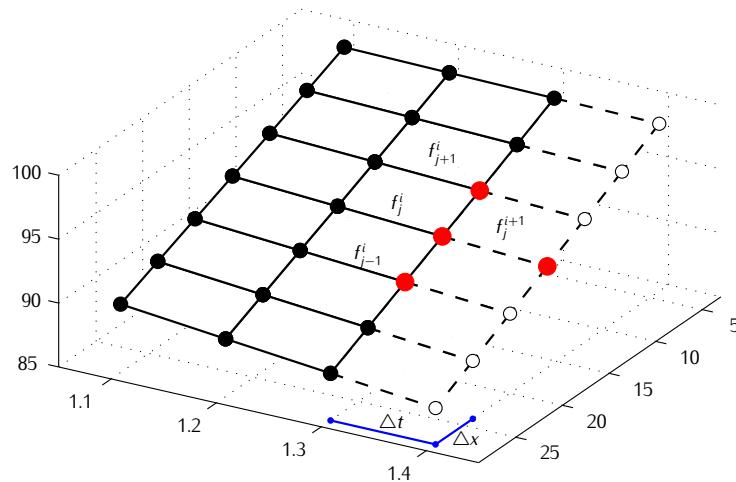
Also, to construct a two-dimensional grid we need to discretize across time as well, using  $N_t$  points that define subintervals of constant width  $\Delta t$ ,  $\{t_i\}_{i=1}^{N_t}$ . Figure 3.2 illustrates such a grid, together with a view of a function surface that we could reconstruct over that grid. It is important to note that neither the space nor the time grid have to be uniform. One can, and in some cases should, consider non-uniform grids based on some qualitative properties of the PDE in hand.

FIGURE 3.2: A two-dimensional grid.



## EXPLICIT FINITE DIFFERENCES

FIGURE 3.3: The Explicit FDM.



As we noted, equation (3.4) describes the dynamic evolution of derivative prices, subject to initial and perhaps boundary conditions. Our time discretization has that objective as well: given the pricing function values at time  $t_i$  we should be able to determine the function values at time  $t_{i+1}$ . Therefore, starting from the initial values at time  $t_0 = 0$ , we recursively produce the values at  $t_1$ ,  $t_2$ , and so on.

At first glance using central differences is not feasible, since the central difference at the time point  $t_0$  needs the values at  $t_1$  and  $t_{-1}$  to be determined, but the values at  $t_{-1}$  are unavailable. On the other hand forward differences in time will do, as we only need the values at  $t_0$  and  $t_1$  to form them. In order to condense notation we will use  $f_j^i$  to denote the value  $f(t_i, x_j)$ . We also assume a uniform grid with spacings  $\Delta x$  and  $\Delta t$ , although it is not a lot harder to work over non-uniform grids.<sup>3</sup> Then, by approximating  $\frac{\partial f(t_i, x_j)}{\partial t} \approx \frac{f_j^{i+1} - f_j^i}{\Delta t}$ , we derive the *explicit finite difference method*

$$\frac{f_j^{i+1} - f_j^i}{\Delta t} = \mathcal{L}f_j(t_i) = q_j^+ f_{j+1}^i + q_j^0 f_j^i + q_j^- f_{j-1}^i \quad (3.5)$$

We can explicitly solve<sup>4</sup> the above expression for  $f_j^{i+1}$ , which yields the recursive relationship

<sup>3</sup> Just a lot more messier.

<sup>4</sup> Hence the name!

$$f_j^{i+1} = q_j^+ \Delta t f_{j+1}^i + (1 + q_j^0 \Delta t) f_j^i + q_j^- \Delta t f_{j-1}^i$$

Essentially, the values  $f_{j\pm 1}^i$  and  $f_j^i$  determine the next period's value  $f_t^{i+1}$ . This is schematically depicted in figure 3.3. In matrix form, the updating takes place as

$$\mathbf{f}^{i+1} = (\mathbf{I} + \mathbf{Q} \Delta t) \cdot \mathbf{f}^i \quad (3.6)$$

Now we turn to the BS PDE (3.1) and apply this discretization scheme. To simplify the expressions we perform the change of variable  $x = \log S$ . This will transform the PDE into one with constant coefficients, namely

$$-\frac{\partial f(t, x)}{\partial t} + \alpha \frac{\partial f(t, x)}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f(t, x)}{\partial x^2} = rf(t, x)$$

with  $\alpha = r - \frac{1}{2}\sigma^2$ . The coefficients  $q_i^\pm$  and  $q_i^0$  in the system of ODEs (3.4), which also determine the explicit scheme (3.6), become

$$q_i^\pm = \pm \frac{\alpha}{2\Delta x} + \frac{\sigma^2}{2\Delta x^2}$$

$$q_i^0 = -r - \frac{\sigma^2}{\Delta x^2}$$

## STABILITY AND CONVERGENCE

By constructing a FDM, like the explicit scheme, we use derivative approximations to reconstruct the true, but unknown, pricing function  $f(t, S)$ . The outcome of the FDM is a set of prices at time  $t_i$ , namely  $\mathbf{f}^i = \{f_j^i\}_{j=1}^{N_x}$ , for all different  $i = 1, \dots, N_t$ . The natural question is of course *how close are the values  $f_j^i$  to the true prices  $f(t_i, x_j)$ ?* If we are to use such a scheme in practice we need to be convinced that somehow " $f_j^i \rightarrow f(t_i, x_j)$ " as the discretization becomes finer. Also, if we are to put some trust in this approximation we should have an idea about the order of this convergence.

One straightforward way would be to examine how the *pointwise errors* between the true prices and their approximation behave. If we denote the true prices with  $\tilde{\mathbf{f}}^i = \{f(t_i, x_j)\}_{j=1}^{N_x}$ , then the errors in question would be the differences

$$\epsilon^i = \mathbf{f}^i - \tilde{\mathbf{f}}^i$$

We can investigate the convergence  $\mathbf{f}^i \rightarrow \tilde{\mathbf{f}}^i$  by inspecting the  $\ell_\infty$ -norm, namely<sup>5</sup>  $\|\epsilon\| = \max_{j=1\dots N_x} |f_j^i - f(t_i, x_j)|$ . Apparently, if the maximum (absolute) value converges to zero, then all other values will do as well, and the FDM prices will converge to the true ones.

Before we move to the inspection of the global errors, we first examine the *local truncation error*, defined as the discrepancy between the true parabolic

---

<sup>5</sup> In some cases it is more convenient to work with the  $\ell_1$ -,  $\ell_2$ - or  $\ell_p$ -norm. The choice largely depends on the problem in hand. See XXXX for details.

PDE (3.2) and the approximated one (3.5), evaluated at the *true pricing function* at the point  $(t_i, x_j)$

$$\tau_j^i = \left( \frac{\partial}{\partial t} f(t_j, x_i) - \mathcal{L}f(t_j, x_i) \right) - \left( \frac{f(t_{i+1}, x_j) - f(t_i, x_j)}{\Delta t} - \mathcal{L}f(t_i, x_j) \right)$$

The definitions and the properties of the difference operators yield that the truncation error  $\tau_j^i = o(\Delta t, \Delta x^2)$ . We therefore say that the explicit method is *first order accurate in time and second order accurate in space*. Intuitively, this truncation error would tell us how errors will be created over one step, if we start from the correct function values. Any scheme that offers order of accuracy greater than zero is called *consistent*.

Of course, even if small errors are created over a given time step, they can still accumulate as we move from one time step to the next. It is possible that they produce feedback effects, producing errors that grow exponentially in time, destroying the approximate solutions and creating oscillatory or explosive behaviour. On the other hand we might construct a FDM that has errors that behave in a “nice” way, without feedback effects. The notion of *stability* captures these ideas.

One intuitive way of looking at stability is through the Courant–Friedrichs–Lewy (CFL) condition<sup>6</sup>, which is based on the notion of the *domain of dependence*. If we have a function  $f(t, S)$ , then the domain of dependence of the point  $(t^*, S^*)$  is the set of points

$$F(t^*, S^*) = \{(t, S) : t \geq t^* \text{ and } f(t, S) \text{ depends on the value } f(t^*, S^*)\}$$

The CFL criterion states that if a numerical scheme is stable, then the true domain of dependence must be smaller than the domain of dependence of the approximating scheme.

In parabolic PDEs the domain of dependence of the process is unbounded, since information travels instantaneously across all values. The domain of dependence of the explicit FDM is bounded, since each value at time  $t_{i+1}$  will only depend on three of its neighbouring values at time  $t_i$ . Therefore, according to CFL criterion in order for the scheme to be stable the condition  $\Delta t = o(\Delta x^2)$  must be satisfied.<sup>7</sup> Therefore the explicit scheme will not be unconditionally stable, and will need very small time discretization steps to offer stability.

The connection between local errors, global errors and stability is given by the *Lax equivalence theorem* which states that a FDM which is consistent and stable will be convergent. This means that the explicit method is not (always) convergent.

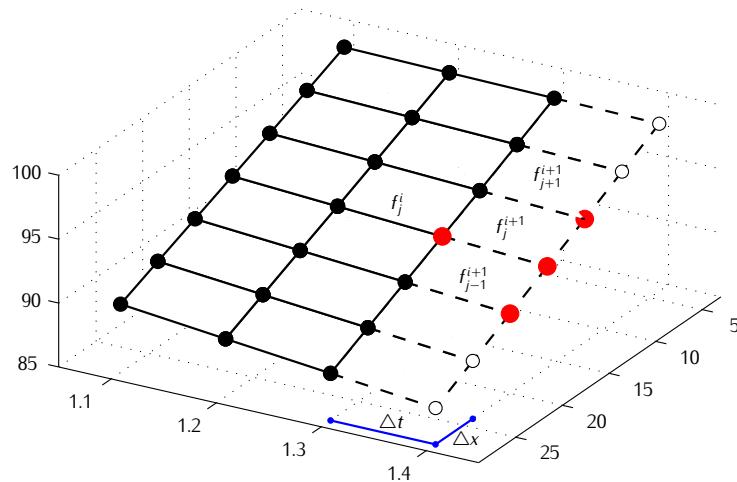
---

<sup>6</sup> Stated in 1928, long before any stability issues were discussed in this context. Richardson initiated FDM schemes as back as 1922 for weather prediction, but did not discover any stability problems.

<sup>7</sup> This means that the time grid must become finer a lot faster than the space grid, for the information to rapidly reach remote values.

## IMPLICIT FINITE DIFFERENCES

FIGURE 3.4: The Implicit FDM.



One way to overcome the stability issues is to use a backward time step. Rather than taking a forward time step at time  $t_i$ , we take a backward step from time  $t_{i+1}$ . This is equivalent to computing the space derivatives at time  $t_{i+1}$  as shown below

$$\frac{f_j^{i+1} - f_j^i}{\Delta t} = q_j^+ f_{j+1}^{i+1} + q_j^0 f_j^{i+1} + q_j^- f_{j-1}^{i+1}$$

This equation relates three quantities at time  $t_{i+1}$  and one quantity at time  $t_i$ , which is schematically given in figure 3.4.

Since we are facing one equation with three unknowns we cannot explicitly give a solution, but we can form a system.

$$-q_j^+ \Delta t f_{j+1}^{i+1} + (1 - q_j^0 \Delta t) f_j^{i+1} - q_j^- \Delta t f_{j-1}^{i+1} = f_j^i$$

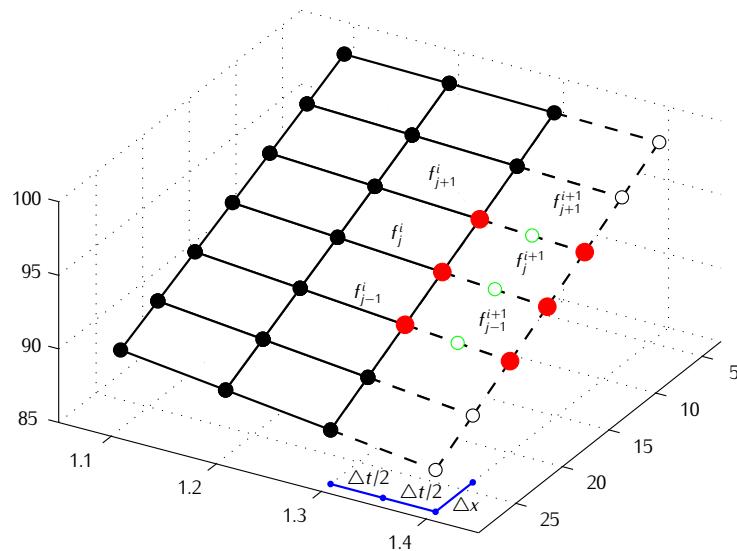
Note that the number of system equations will be equal to the number of unknowns. In matrix form, the system can be written as

$$\mathbf{f}^i = (\mathbf{I} - \mathbf{Q} \Delta t) \cdot \mathbf{f}^{i+1} \Leftrightarrow \mathbf{f}^{i+1} = (\mathbf{I} - \mathbf{Q} \Delta t)^{-1} \cdot \mathbf{f}^i \quad (3.7)$$

The same line of argument we used for the explicit method will give us the order of accuracy of the implicit scheme, the errors being again  $\mathcal{o}(\Delta t, \Delta x)$ . On the other hand, since the value at time  $t_{i+1}$  depends on the whole set of prices  $\mathbf{f}^i$  at time  $t_i$ , the domain of dependence of the implicit scheme is unbounded. From the CFL criterion it follows that the implicit scheme is unconditionally stable.

## THE CRANK-NICOLSON AND THE $\theta$ -METHOD

FIGURE 3.5: The Crank-Nicolson FDM.



Although the implicit scheme is unconditionally stable, it still offers convergence of order  $o(\Delta t, \Delta x^2)$ . The first order convergence in time is due to the nature of the derivative approximation. We can increase this order to two by setting up a central difference scheme in time. We will use a time step of  $\frac{\Delta t}{2}$ , as we did in the approximation of the second order derivative.

This is equivalent in taking the space derivatives at the midpoint between  $t_i$  and  $t_{i+1}$ . This yields the *Crank-Nicolson scheme*

$$\frac{f_j^{i+1} - f_j^i}{\Delta t} = q_j^+ \cdot \frac{f_{j+1}^i + f_{j+1}^{i+1}}{2} + q_j^0 \cdot \frac{f_j^i + f_j^{i+1}}{2} + q_j^- \cdot \frac{f_{j-1}^i + f_{j-1}^{i+1}}{2}$$

Apparently the Crank-Nicolson scheme will relate six points, illustrated in figure 3.5. Another approach is to simply add up one-half times equation (3.6) and one-half times the first of equations (3.7). This yields again the Crank-Nicolson scheme, in matrix form

$$\left( I - \frac{1}{2} Q \Delta t \right) \cdot f^{i+1} = \left( I + \frac{1}{2} Q \Delta t \right) \cdot f^i$$

Since it uses centered differences to approximate all derivatives, the errors in the Crank-Nicolson scheme are  $o(\Delta t^2, \Delta x^2)$ . Therefore, the Crank-Nicolson scheme is *second order accurate both in time and space*. In addition, like the implicit

scheme, the Crank-Nicolson scheme has unbounded domain of dependence, and is therefore unconditionally stable.

Rather than using weights of  $\frac{1}{2}$  to balance the explicit and implicit schemes, one can use different values. This gives rise to the  $\theta$ -method, which encompasses all schemes described so far. In particular, the  $\theta$ -method in matrix form will be

$$(\mathbf{I} - \theta \mathbf{Q} \Delta t) \cdot \mathbf{f}^{i+1} = (\mathbf{I} + (1 - \theta) \mathbf{Q} \Delta t) \cdot \mathbf{f}^i$$

It is straightforward to verify that  $\theta = 0$  yields the explicit scheme,  $\theta = 1$  yields the implicit scheme, and  $\theta = \frac{1}{2}$  yields the Crank-Nicolson scheme.

## BOUNDARIES

The above treatment of parabolic PDEs assumed that the space extends over the real line. Essentially this implies that the matrices involved are of infinite dimensions. Of course in practice we will be faced with finite grids. Sometimes, as is the case with barrier options, boundary conditions will be explicitly imposed by the nature of the derivative contract. In other cases, when the derivative has early exercise features, the boundary is not explicitly defined and is *free* in the sense that it is determined simultaneously with the solution of the PDE.

There are two different kinds of fixed boundary conditions: *Dirichlet conditions* set  $f(t_B, x_B) = f_B$ , that is the *value* of the function is known on the boundary. *Neumann conditions* set  $\frac{\partial f(t_B, x_B)}{\partial x} = \varphi_B$ , that is the *derivative* is known on the boundary. In the second case we will need to devise an approximation scheme that exhibits  $o(\Delta x^2)$  accuracy; if we do not achieve that, then contaminated values will diffuse and eventually corrupt the function values at all grid points. This means that we must use finite difference schemes that achieve  $o(\Delta x^2)$ , like central differences.

Say that we construct a finite space grid  $\{x_i\}_{i=0}^{N_x}$ , which essentially discretizes the interval  $[x_0, x_{N_x}]$ . Most of the elements of the matrix  $\mathbf{Q}$  are not affected by the boundary conditions, and the matrix is still tridiagonal. The only parts that are determined by the fixed boundary conditions are the first and last rows. Thus  $\mathbf{Q}$  will have the form

$$\mathbf{Q} = \begin{pmatrix} \star & \star & 0 & & \\ q_1^- & q_1^0 & q_1^+ & 0 & \\ 0 & q_2^- & q_2^0 & q_2^+ & 0 \\ \ddots & \ddots & \ddots & & \\ 0 & q_j^- & q_j^0 & q_j^+ & 0 \\ \ddots & \ddots & \ddots & & \\ 0 & q_{N_x-1}^- & q_{N_x-1}^0 & q_{N_x-1}^+ & 0 \\ 0 & \star & \star & \star & \end{pmatrix}$$

where the values at  $\star$  are determined by the boundary conditions.

We start with a Dirichlet condition at  $f_{N_x+1}^i = f(t_i, x_{N_x+1}) = f_B^i$ . This point is utilized in the explicit scheme when the value  $f_{N_x}^{i+1}$  at time  $t_{i+1}$  is calculated. In particular

$$f_{N_x}^{i+1} = q_j^+ \Delta t f_B^i + (1 + q_j^0 \Delta t) f_{N_x}^i + q_j^- \Delta t f_{N_x-1}^i$$

Therefore, in matrix form, the updating equation for the explicit scheme becomes

$$\mathbf{f}^{i+1} = (\mathbf{I} + \mathbf{Q} \Delta t) \cdot \mathbf{f}^i + \mathbf{g}^i \Delta t$$

where the last row of  $\mathbf{Q}$  is  $(0, \dots, 0, q_{N_x+1}^-, q_{N_x+1}^0)$ , and  $\mathbf{g}^i$  is an  $(N_x + 1) \times 1$  vector of zeros, with the last element equal to  $q_{N_x+1}^+ f_B^i$ . Similarly, a Dirichlet boundary condition at  $x_0^i$  will set the first row of  $\mathbf{Q}$  is  $(q_0^0, q_0^+, 0, \dots, 0)$ , and the first element of  $\mathbf{g}^i$  is  $q_0^- f_B^i$ .

Within the implicit scheme this boundary would appear in the  $N_x + 1$  system that determines function values at time  $t_i$

$$f_{N_x}^{i-1} = -q_j^+ \Delta t f_B^i + (1 - q_j^0 \Delta t) f_{N_x}^i - q_j^- \Delta t f_{N_x-1}^i$$

Therefore when the Crank-Nicolson method is implemented  $f_B^i$  will affect both pricing formulas at  $t_i$  and  $t_{i+1}$ . Similar formulas can be easily computed for the lower boundary  $x_0$ , where the ghost point  $x_{-1}$  is introduced.

When a Neumann condition is imposed at  $x_{N_x}$ , we apply central differences at the point  $x_{N_x}$  to approximate  $\frac{\partial f(t_i, x_i)}{\partial x} = \varphi_B^i$ , which yields

$$f_{N_x+1}^i = f_{N_x-1}^i + 2\varphi_B^i \Delta x$$

These values can be used in the approximation schemes to set up the last row of  $\mathbf{Q}$ , namely  $(0, \dots, 0, q_{N_x+1}^+ + q_{N_x+1}^-, q_{N_x+1}^0)$ , and the last element of  $\mathbf{g}^i$  equal to  $2q_{N_x+1}^+ \varphi_B^i \Delta x$ . Similarly, a Neumann boundary condition at  $x_0^i$  will set the first row of  $\mathbf{Q}$  to  $(q_0^0, q_0^+ + q_0^-, 0, \dots, 0)$ , and the first element of  $\mathbf{g}^i$  to  $-2q_0^- \varphi_B^i \Delta x$ .

### 3.3 A PDE SOLVER IN MATLAB

#### PLAIN VANILLA OPTIONS

In this section we will build a Matlab example that implements the  $\theta$ -method. We assume that the dynamics of the underlying asset are the ones that govern the BS paradigm. We will need payoff function  $G$ . If we also assume Dirichlet boundary conditions, then the same function will determine the Dirichlet boundaries. If Neumann conditions are specified, then the same function should also give the derivatives on the boundaries. Since we have set up the PDE in terms of the log-price, which we denote in the solver with  $x$ , the derivatives will be equal to  $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial S} \frac{\partial S}{\partial x} = \frac{\partial f}{\partial S} \exp(x)$ . A call option will be implemented by the function in listing 3.1. A put option is implemented in 3.2.<sup>8</sup>

LISTING 3.1: G\_call.m: Payoff and boundaries for a call.

```
% G_call.m
function [y, by, dy] = G_call(x, p)
K = p.K;
dx = x(2)-x(1);
5 y = max(exp(x) - K, 0); % payoff function
by = NaN; % boundary values
dy = [0, exp(x(end))]; % boundary derivative
```

LISTING 3.2: G\_put.m: Payoff and boundaries for a put.

```
% G_put.m
function [y, by, dy] = G_put(x, p)
K = p.K;
dx = x(2)-x(1);
5 y = max(K - exp(x), 0); % payoff function
by = NaN; % boundary values
dy = [-exp(x(1)), 0]; % boundary derivative
```

The initialization part of the PDE solver just decomposes the structure `p` and constructs the log-price and the time grids. The function `f` returns the payoff values, the boundary values and the derivatives on the boundaries. Therefore both Dirichlet and Neumann conditions can be accommodated for. The tridiagonal `Q` matrix will be constructed according to whether we have specified Dirichlet or Neumann boundary conditions. We use the switch `boundtype` that keeps the boundary type as a two-element vector. At this stage we assume that the same boundary applies to all time steps. The Matlab code for the PDE solver is given in listing 3.3.

Here we use the Matlab backslash operator `A\B = inv(A) * B`. The snippet in 3.4 illustrates how the function can be called to compute the price of a European put, and plots the pricing function. Setting `p.boundtype = 1` will implement the PDE solver with Dirichlet boundary conditions.

### EARLY EXERCISE FEATURES

In many cases the derivative in hand has early exercise features, either American (where the option can be exercised at any point prior to maturity), or Bermudan

---

<sup>8</sup> We will implement the solver using Neumann conditions, and therefore we pass the boundary values as `NaN`. Actually, for the put price the corresponding Dirichlet boundary condition is not time homogeneous, and our solver will need slight modifications to accommodate time inhomogeneous boundaries.

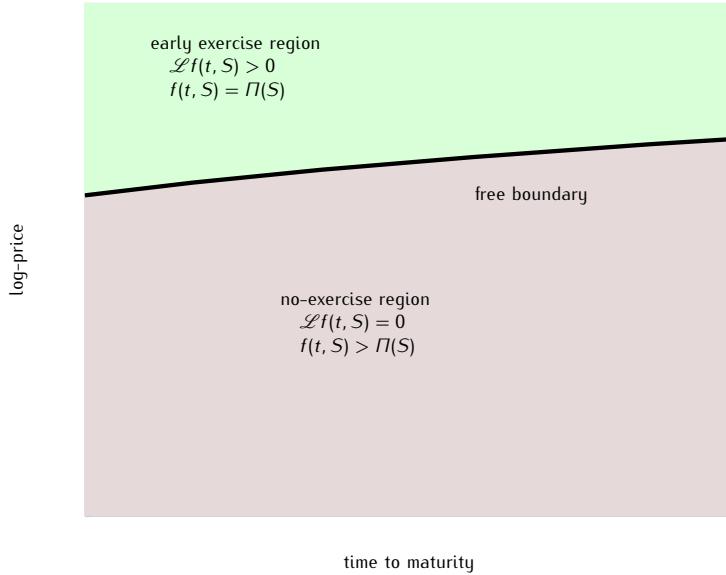
LISTING 3.3: pde\_bs.m:  $\theta$ -method solver for the Black-Scholes PDE.

```
% pde_bs.m
function [xv, tv, FT] = pde_bs(f, p)
theta = p.theta; % for theta-method
r = p.r; % risk free rate
sigma = p.sigma; % volatility
a = r - .5*sigma*sigma;
T = p.t; % maturity
Nt = p.tnumber; % time intervals
dt = T/Nt; % time grid size
tv = [0:dt:T]; % time grid
bx = p.xboundary; % max log-price
Nx = p.xnumber; % log-price intervals
dx = 2*bx/Nx; % log-price grid size
xv = [-bx:dx:bx]; % log-price grid
15 boundtype = p.boundtype;
[f0, bf0, df0] = feval(f, xv', p); % f call
qp = .5*a/dx + .5*sigma*sigma/dx/dx; % q-plus
qm = -.5*a/dx + .5*sigma*sigma/dx/dx; % q-minus
q0 = -r - sigma*sigma/dx/dx; % q-zero
20 % matrix Q
Q = diag(q0*ones(Nx+1,1)) + diag(qp*ones(Nx,1), +1) + ...
    diag(qm*ones(Nx,1), -1);
g = zeros(Nx+1, 1); % vector of constants
% boundary conditions
if boundtype % Dirichlet conditions
    g(1) = qm * bf0(1); % top constant
    g(end) = qp * bf0(2); % bottom constant
else % Neumann conditions
    Q(1, 2) = qm + qp; % top Q-matrix
    Q(Nx+1, Nx) = qm + qp; % bottom Q-matrix
    g(1) = -2*dx*qm * df0(1); % top constant
    g(end) = 2*dx*qp * df0(2); % bottom constant
end
FT = zeros(Nx+1, Nt+1); % grids of results
FT(:,1) = f0; % initial condition
35 for tndx = 2:Nt+1; % loop theta-method through time
    FT(:,tndx) = (eye(Nx+1) - theta*Q*dt) \ ( (eye(Nx+1) ...
        + (1-theta)*Q*dt) * FT(:,tndx-1) + g*dt );
end
```

LISTING 3.4: pde\_bs\_impl.m: Implementation of the  $\theta$ -method solver.

```
% pde_bs_impl.m
clear;
clc;
5 p.theta      = 0.50;      % for theta-method
p.r           = 0.04;       % risk free rate
p.sigma       = 0.30;       % volatility
p.t            = 0.25;       % maturity
p.K            = 0.95;       % strike price
p.tnumber     = 60;         % time intervals
10 p.xboundary = 0.40;       % max log-price
p.xnumber     = 80;         % log-price intervals
p.boundtype   = 0;          % boundary type
% call the PDE solver
[xv, tv, FT] = pde_bs(@G_put, p);
15 % make a 3D plot of the results
surf(exp(xv), tv, FT');
```

FIGURE 3.6: Early exercise region for an American put. The time-price space is separated into two parts. If the boundary is crossed then exercise becomes optimal.



(where the option can be exercised at a predefined set of times). With small changes the PDE solver we constructed can take care of these features.

Essentially, the holder of the option has to make a decision at these time points: exercise early and receive the intrinsic value, or wait and continue holding the option. In terms of PDE jargon, the problem is now a *free-boundary* problem. There is a boundary, which is at the point unknown for us, which separates the region of  $(t, S)$  where early exercise is optimal and the region of where it is optimal to wait. Figure 3.6 illustrates these regions. Thus, within the “waiting optimal” region the BS PDE is satisfied, while outside the boundary  $f(t, S)$  will be equal to the payoff function  $\Pi(S)$ .

The boundary function is unknown, but it has a known property: it will be the first point at which  $f(t, S) = \Pi(S)$ . This follows from a no-arbitrage argument that gives that the pricing function has to be smooth and not exhibit discontinuities. In terms of the pricing function, it will satisfy

$$\mathcal{L}f(t, S) \geq 0 \quad (3.8)$$

$$f(t, S) \geq \Pi(S) \quad (3.9)$$

$$\mathcal{L}f(t, S) \cdot (f(t, S) - \Pi(S)) = 0 \quad (3.10)$$

The BS PDE is satisfied within the no-exercise region, while the pricing function is satisfied within the exercise region. Equation (3.10) reflects that. Within the exercise region  $\mathcal{L}f(t, S) > 0$ , while within the no-exercise region  $f(t, S) > \Pi(S)$ . Equations (3.8–3.9) cover these possibilities.

This indicates that a strategy to compute the option price whenearly exercise is allowed will be to set

$$f_j^i = \max(\hat{f}_j^i, \Pi(S_j))$$

where  $\hat{f}$  is the price if no exercise takes place. Therefore the option holder’s strategy is implemented: the holder will compare the value of the option if she did not exercise with the price if she does; the option value will be the maximum of the two. Although the above approach is straightforward in the explicit method case, it is not so in the other methods where a system has to be solved. In these cases we are looking for solutions of a system subject to a set of inequality conditions. In the most general  $\theta$ -scheme, the system has the form

$$\begin{aligned} (\mathbf{I} - \theta \mathbf{Q} \Delta t) \cdot \mathbf{f}^{i+1} &\geq (\mathbf{I} + (1 - \theta) \mathbf{Q} \Delta t) \cdot \mathbf{f}^i \\ \mathbf{f}^{i+1} &\geq \Pi(\mathbf{S}) \\ [(\mathbf{I} - \theta \mathbf{Q} \Delta t) \cdot \mathbf{f}^{i+1} - (\mathbf{I} + (1 - \theta) \mathbf{Q} \Delta t) \cdot \mathbf{f}^i] \odot [\mathbf{f}^{i+1} - \Pi(\mathbf{S})] &= 0 \end{aligned}$$

where  $\mathbf{S}$  is the vector of the grid prices of the underlying asset, and the inequality is taken element-wise.

Such systems can not be explicitly solved, but there are iterative methods, like the *projected successive over-relaxation* or PSOR method. Given a system

$$\begin{aligned} \mathbf{A} \cdot \mathbf{x} &\geq \mathbf{b} \\ \mathbf{x} &\geq \mathbf{c} \\ [\mathbf{A} \cdot \mathbf{x} - \mathbf{b}] \odot [\mathbf{x} - \mathbf{c}] &= \mathbf{0} \end{aligned}$$

a starting value  $\mathbf{x}^{(0)}$ , and a *relaxation parameter*  $\omega \in (0, 2)$ , the PSOR method updates<sup>9</sup>

$$x_i^{(k+1)} = \max \left( c_i, (1 - \omega)x_i^{(k)} + \frac{\omega}{a_{ii}} \left[ b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)} \right] \right)$$

The PSOR procedure is implemented in the short code given in 3.5. The programme will solve  $\mathbf{A} * \mathbf{x} \geq \mathbf{b}$  and  $\mathbf{x} \geq \mathbf{c}$ , while one of the two equalities will strictly hold for each element. The initial value is  $\mathbf{x}_{\text{init}}$ . The function returns the solution vector  $\mathbf{x}$ , and an indicator vector  $\mathbf{ex}$  of the elements where the second equality holds; in our case the early exercise points. The solver has to be adjusted to accommodate for the early exercise, and the code is given in 3.6. We introduce `omega` and call the PSOR procedure. We demand accuracy of  $10^{-6}$ , while we allow for 100 iterations to achieve that.

The snippet in listing 3.7 implements the pricing of a European and an American put and examines the results. The strike price is \$1.05. To make the differences clearer the interest rate is set to 10%. The results are given in figure 3.7; the American option prices approach the payoff function for small values of the spot price, while European prices cross. Early exercise will be optimal if the spot price is below \$0.90, where the American prices touch the payoff function.

## BARRIER FEATURES

European vanilla options (calls and puts) are exercised on maturity, and have payoffs that depend on the final value of the underlying asset. Barrier options have an extra feature: the option might not be active to maturity, depending on whether or not the barrier has been triggered. Denote the barrier level with  $B$ . The jargon for barrier options specifies the impact of the barrier as follows

- *Up*: there is an upper barrier, or *Down*: there is a lower barrier
- *In*: the contract is not activated before the barrier is triggered, or *Out*: if the barrier is breached the contract is cancelled

Therefore we can have eight standard combinations

$$\left. \begin{array}{l} \text{Up} \\ \text{Down} \end{array} \right\} \text{-and-} \left. \begin{array}{l} \text{In} \\ \text{Out} \end{array} \right\} \left\{ \begin{array}{l} \text{Calls} \\ \text{Puts} \end{array} \right\}$$

---

<sup>9</sup> A value  $0 < \omega < 1$  corresponds to *under-relaxation*,  $\omega = 1$  is the Gauss-Seidel algorithm, while  $1 < \omega < 2$  corresponds to *over-relaxation*. In our case we want to use a value that implements over-relaxation.

LISTING 3.5: psor.m: PSOR method.

```
% psor.m
function [x,ex] = psor(A, b, c, xinit, omega, tol, jmax)
n      = length(b);           % length of vectors
x      = xinit;              % initialize x
5     j      = 0;               % number of iterations
flag   = 1;                  % stopping flag
while flag
    j      = j + 1;          % next iteration
    xinit = x;                % old value
10   for i = 1:n             % update new value
        x(i) = max(c(i), x(i) +...
                     omega*( b(i)-A(i,:)*x ) / A(i,i));
    end
    % change small enough or too many iterations
15   if (norm(x-xinit) < tol)|(j > jmax)
        flag = 0;
    end
end
ex   = (x==c) & (x>0);    % the early exercise region
```

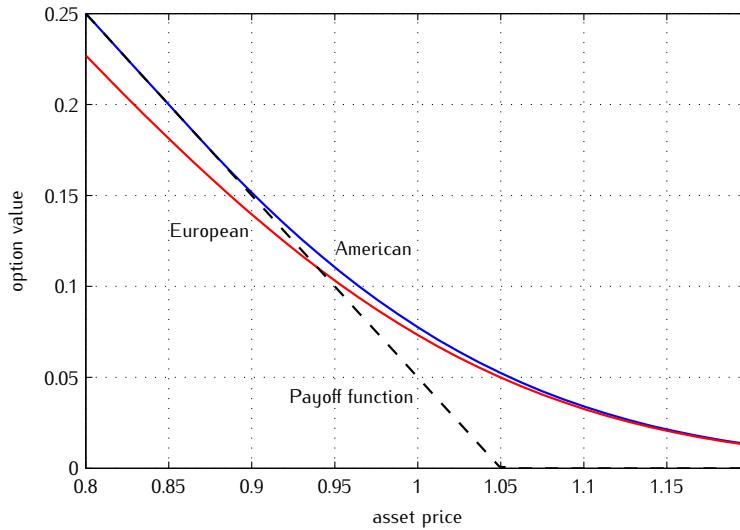
LISTING 3.6: pde\_bs\_amer.m:  $\theta$ -method solver with early exercise.

```
% pde_bs_amer.m
function [xv, tv, FT, EX] = pde_bs_amer(f, p)
omega = p.omega;                 % for PSOR

5   lines 3-37 of pde_bs.m

% exercise region
EX      = FT;
EX(:,1) = (f0>0);
10  % loop through time
for tndx = 2:Nt+1;
    % theta-method
    A = eye(Nx+1) - theta*Q*dt;
    b = ( eye(Nx+1) + (1-theta)*Q*dt ) * FT(:,tndx-1) + ...
        g*dt;
15  [w, ex] = psor( A, b, f0, FT(:, tndx-1), omega, 1e...
        -6, 100 );
    FT(:,tndx) = w;
    EX(:,tndx) = ex;
end
```

FIGURE 3.7: European versus American option prices. The American option will reach the payoff function, while the price of the European contract can cross below that level.



LISTING 3.7: pde\_bs\_amer\_impl.m: Implementation of PSOR for an American put.

```
% pde_bs_amer_impl.m
p.omega      = 1.50; % for PSOR

lines 2-12 of pde_bs_impl.m

% call the PDE solver for American options
[xx, tv, FAT, EX] = pde_bs_amer(@G_put, p);
% call the PDE solver for European options
[xx, tv, FET]     = pde_bs(@G_put, p);
FAT = FAT(:,end); % last American prices
FET = FET(:,end); % last European prices
ST  = exp(xx);    % asset prices
% plot option value and the payoff function
plot(ST,FAT, ST,FET, ST,max(p.K-ST,0));
```

Barrier options are examples of *path dependent* contracts, since the final payoffs depend on the price path before maturity. This path dependence is considered *mild*, since we are not interested on the actual levels, but only on the behavior relative to the barrier, i.e. if the barrier is triggered.

For example, consider an *up-and-out call*, where the spot price of the underlying is  $S_0 = \$85$ , the strike price is  $K = \$105$  and the barrier is at  $B = \$120$ . This contract will pay off only if the price of the underlying remains below \$120 for the life of the option. If  $S_t \geq B$  at any  $t$ , then the payoffs (and the value of the contract) become zero. One can see that we should expect this contract to have some strange behaviour when the price is around the barrier level.

Contrast an *up-and-in call* with the same specifications. For the contract to pay anything, the price has to reach at least  $S_t = \$120$  for some  $t$  (but might drop in later times).

Now suppose that an investor holds both contracts, and observe that (for any sample path) the barrier can either be triggered or not. Thus, when one is active the other one is not. Holding both of them replicates the *vanilla call*. Therefore,

$$P_{U\&O} + P_{U\&I} = P_{\text{Call}}, \text{ and } P_{D\&O} + P_{D\&I} = P_{\text{Put}}$$

In the above examples the barrier contract was monitored *continuously*. For such contracts closed-form solutions exist. In practice though, barrier options are monitored discretely, that is to say one examines where the underlying spot price is with respect to the barrier at a discrete set of points. For example a barrier contract might be monitored on the closing of each Friday. Monitoring can have a substantial impact on the pricing of barrier options. For that reason numerical methods are employed to price barrier options.

An up-and-out option will follow the BS PDE, where a boundary will exist at  $S_t = B$  (in fact  $f(t, S) = 0$  for all  $S \geq B$ ). This feature can be very easily implemented in the finite difference schemes that we discussed. In particular, the barrier will be active only on the monitoring dates, and a PDE with no barriers<sup>10</sup> will be solved. Essentially, we can compute the updated values  $\mathbf{f}^{i+1}$  normally, and then impose the condition  $f(t_i, x_j) = 0$  if  $t_i$  is a monitoring date and  $\exp(x_i) \geq B$ .

A Matlab listing that implements pricing of up- and down-and-out calls and put is given in 3.8. The snippet that calls this function is given in 3.9.

## COMPUTING THE GREEKS

Using finite differences is very useful when one is looking for the hedge parameters, in particular the option's Delta and Gamma. Given that they are quantities that are defined as derivatives with respect to the price, one can compute them rapidly over the given grid. Some care has to be taken here, as we have implemented the discretization in log-prices. The Delta and Gamma of the option will

---

<sup>10</sup> Of there will be barriers imposed at extreme values that are necessary to discretize the state space.

LISTING 3.8: pde\_bs\_barr.m: Solver with barrier features.

```
% pde_bs_barr.m
function [xv, tv, FT] = pde_bs_barr(f, p)
barrT = p.barriertimes; % barrier times
barrB = p.barrierlevel; % barrier level
5 barrU = p.barrierdirection; % up (1) or down (0)

other stuff

% boundary conditions
10 if boundtype % Dirichlet conditions
    g(1) = qm * bf0(1) * (1-barrU);
    g(end) = qp * bf0(2) * barrU;
else % Neumann conditions
    Q(1,2) = qm + qp;
15 Q(Nx+1, Nx) = qm + qp;
    g(1) = -2*dx*qm * df0(1) * (1-barrU);
    g(end) = 2*dx*qp * df0(2) * barrU;
end
% up or down multiplier
20 if barrU % up-and-out
    barrM = (exp(xv') >= barrB);
else % down-and-out
    barrM = (exp(xv') <= barrB);
end
25 % grids of results
FT = zeros(Nx+1, Nt+1);
FT(:,1) = f0.*barrM; % initial condition
if barrT == 0 % continuous monitoring
    barrT = tv;
30 else % adjust monitoring points
    barrTa = interp1(tv, tv, barrT, 'nearest');
end
% loop through time
for tndx = 2:Nt+1;
35 Ft = (eye(Nx+1) - theta*Q*dt) \ ( (eye(Nx+1) + (1-
theta)*Q*dt) * FT(:,tndx-1) + g*dt );
    if sum( tv(tndx)==barrTa ) % if monitoring date
        FT(:,tndx) = Ft .* barrM;
    else
        FT(:,tndx) = Ft;
40 end
end
```

for copies, comments, help etc. visit <http://www.theponytail.net/>

LISTING 3.9: pde\_bs\_barr\_impl.m: Implementation for a discretely monitored barrier option.

```
% pde_bs_barr_impl.m
p.barriertimes = [0:0.075:0.25]; % barrier times
p.barrierlevel = 0.90; % barrier level
p.barrierdirection = 1; % up (1) or down (0)
5
lines 2-12 of pde_bs_impl.m

% call the PDE solver for barrier options
[xv, tv, FT] = pde_bs_barr(@G_put, p);
10 % create a surface plot
surf(exp(xv), tv, FT');
```

be equal to

$$\Delta = \frac{\partial f}{\partial S} = \frac{\partial f}{\partial x} \exp(-x)$$

$$\Gamma = \frac{\partial^2 f}{\partial S^2} = \left( \frac{\partial^2 f}{\partial x^2} - \frac{\partial f}{\partial x} \right) \exp(-2x)$$

The derivatives with respect to the log-price  $x$  can be computed using finite differences on the grid (in fact they have been computed already when solving the PDE). Note that, since we approximate all quantities using central differences, the first and last grid points will be lost.

The snippet in 3.10 shows how the Greeks can be computed over a grid, while figure 3.8 gives the output. In order to make clear the effect of early exercise we use a relatively high interest rate of 10%. We also implement a relatively dense ( $100 \times 100$ ) grid over  $(t, S)$  to ensure that the derivatives are accurate. Observe that the Deltas of both options approach their minimum values of  $-1$  in a continuous way. The Gammas, on the other hand, show different patterns with the American Gamma jumping to zero.

Even if we use a stable FDM method, like the Crank–Nicolson, computing the greeks does not always give stable results. For example figure 3.9 presents the Greeks for the same American and European put options as 3.8, but with the time steps decreased to 10. The Delta is apparently computed with errors, which are magnified when the Gamma is numerically approximated. Note that the instability is introduced by reducing the time steps; the log-price grid is still based on 100 subintervals. In other cases explosive Greeks are an outcome of the contract specifications. For instance a barrier option will exhibit Deltas that behave very erratically around the barrier, since the pricing function is not differentiable there.

LISTING 3.10: pde\_bs\_greeks\_impl.m: PDE approximations for the Greeks.

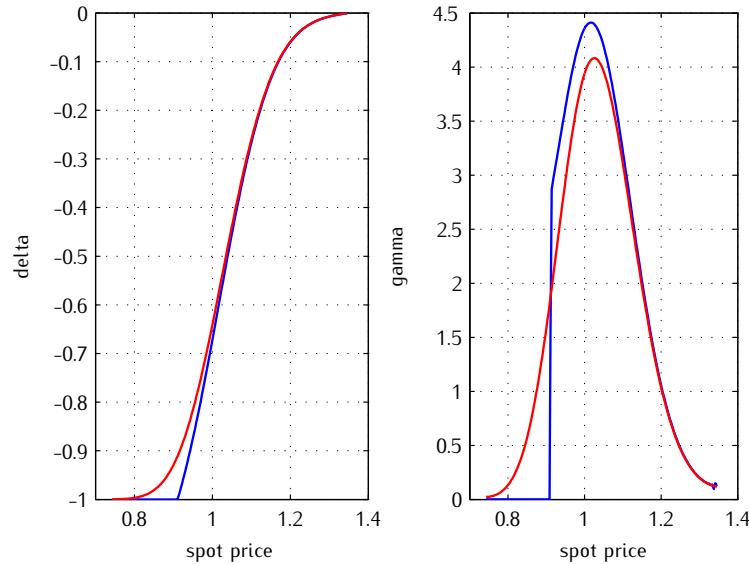
```
% pde_bs_greeks_impl.m
p.theta      = 0.50; % for theta-method
p.omega      = 1.50; % for PSOR
p.r          = 0.10; % risk free rate
5 p.sigma      = 0.30; % volatility
p.t          = 0.10; % maturity
p.K          = 1.05; % strike price
p.tnumber    = 100; % time intervals
p.xboundary  = 0.30; % max log-price
10 p.xnumber   = 100; % log-price intervals
p.boundtype  = 0; % boundary type
[xv, tv, FAT, EX] = pde_bs_amer(@G_put, p);
[xv, tv, FET] = pde_bs(@G_put, p);
A = FAT(:,end)'; % current American prices
15 E = FET(:,end)'; % current European prices
x0 = xv(2:end-1); % truncated grid
S0 = exp(x0); % spot
Dx = xv(2)-xv(1); % log-price grid step
% first derivatives with respect to log-price
20 A1 = (A(3:end)-A(1:end-2))/(2*Dx);
E1 = (E(3:end)-E(1:end-2))/(2*Dx);
% second derivatives with respect to log-price
A2 = (A(3:end)-2*A(2:end-1)+A(1:end-2))/(Dx^2);
E2 = (E(3:end)-2*E(2:end-1)+E(1:end-2))/(Dx^2);
25 DA = A1./S0; % American delta
DE = E1./S0; % European delta
GA = (A2 - A1)./S0./S0; % American gamma
GE = (E2 - E1)./S0./S0; % European gamma
% plots of deltas and gammas
30 subplot(1,2,1); plot(exp(x0), DA, exp(x0), DE);
subplot(1,2,2); plot(exp(x0), GA, exp(x0), GE);
```

## 3.4 MULTIDIMENSIONAL PDEs

In many cases the problem in hand can only be cast in a PDE form that has more than one space dimensions. This can be the case of a derivative that depends on more than one asset, or a derivative that depends on a single that exhibits stochastic volatility, or even a derivative in a BS world that is strongly path-dependent.

Typically the PDE will be still a parabolic one, with a multidimensional elliptic operator. For example in the two-dimensional case the operator on the function  $f = f(t, x, y)$  will be

FIGURE 3.8: Greeks for American and European puts. A European and an American put are priced using the Crank-Nicolson method on a (100, 100) grid over  $(t, S)$ , and the Greeks are computed using finite differences. The Greeks for the European put are given in red and for the American put in blue



$$\mathcal{L}f = \alpha_x \frac{\partial f}{\partial x} + \alpha_y \frac{\partial f}{\partial y} + \beta_x \frac{\partial^2 f}{\partial x^2} + \beta_y \frac{\partial^2 f}{\partial y^2} + \beta_{xy} \frac{\partial^2 f}{\partial x \partial y} + \gamma f$$

Apparently we will need to discretise both dimensions to approximate the elliptic operator. The price function at the typical grid point will be now  $f_{j,k}(t) = f(t, x_j, y_k)$ . The single-variable derivatives pose no real problem, we just need to take some care when computing the cross derivative approximation. For example, one can use the Taylor's expansion of the values  $f_{j\pm 1, k\pm 1}$  and  $f_{j\pm 1, k\pm 1}$

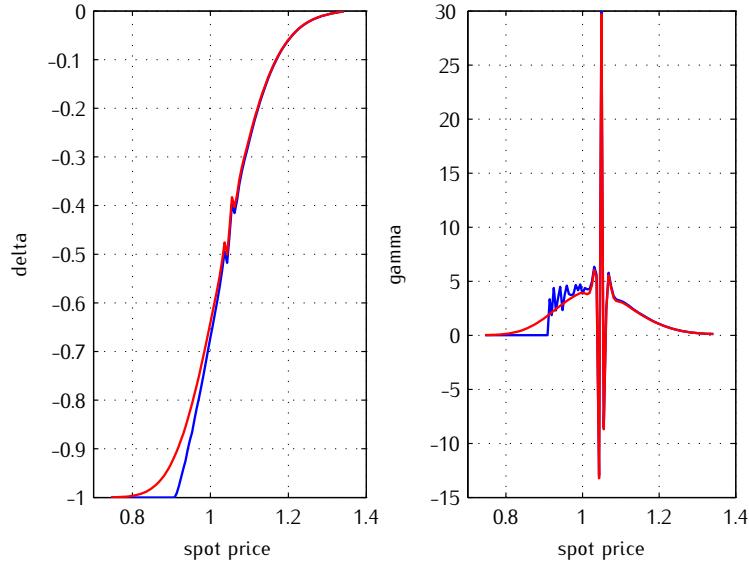
$$f_{j\pm 1, k\pm 1} = f_{j,k} + \frac{\partial f_{j,k}}{\partial x}(\pm \Delta x) + \frac{\partial f_{j,k}}{\partial y}(\pm \Delta y) + \frac{1}{2} \frac{\partial^2 f_{j,k}}{\partial x^2} \Delta x^2 + \frac{1}{2} \frac{\partial^2 f_{j,k}}{\partial y^2} \Delta y^2 + \frac{\partial^2 f_{j,k}}{\partial x \partial y}(\pm \Delta x)(\pm \Delta y) + o(\Delta x^3, \Delta y^3)$$

The operator

$$\mathcal{D}_{xy}^2 f_{j,k} \approx \frac{f_{j+1,k+1} + f_{j-1,k-1} - f_{j+1,k-1} - f_{j-1,k+1}}{4 \Delta x \Delta y}$$

will approximate the cross derivative  $\frac{\partial^2 f(t, x_j, y_k)}{\partial x \partial y}$

FIGURE 3.9: Oscillations of the Greeks in FDM. A European and an American put are priced using the Crank-Nicolson method on a  $(10, 100)$  grid over  $(t, S)$ , and the Greeks are computed using finite differences. The Greeks for the European put are given in red and for the American put in blue



This uses four points to approximate the cross derivative, but it is not the only way to do so.<sup>11</sup> In any case we can write the discretized operator

$$\mathcal{L} = \alpha_x \mathcal{D}_x + \alpha_y \mathcal{D}_y + \beta_x \mathcal{D}_x^2 + \beta_y \mathcal{D}_y^2 + \beta_{xy} \mathcal{D}_{xy}^2 + \gamma$$

If we consider an  $(N_x, N_y)$ -point grid over  $(x, y)$ , then we can construct the matrix  $\mathbf{Q}$  which will be  $(N_x \times N_y, N_x \times N_y)$ . The prices  $f(t, x_j, y_k)$  actually form a matrix  $\mathbf{F}(t)$  for a given  $t$ , but we prefer to think of them as a vector  $\mathbf{f} = \mathbf{f}(t)$  produced by stacking the columns of this matrix. Therefore, the price  $f(t, x_j, y_k)$  will be mapped to the  $(k-1)N_x + j$  element of  $\mathbf{f}$

$$\mathbf{f} = \begin{pmatrix} f(t, x_1, y_1) \\ f(t, x_2, y_1) \\ \vdots \\ f(t, x_{N_x}, y_1) \\ f(t, x_1, y_2) \\ \vdots \\ f(t, x_{N_x}, y_{N_y}) \end{pmatrix}$$

---

<sup>11</sup> For example Ikonen and Toivanen (2004) give an alternative.

FIGURE 3.10: The structure of the Q-matrix that approximates a two-dimensional diffusion.

$$\left( \begin{array}{|c|c|c|c|} \hline & \star \star 0 0 0 0 & \star \star 0 0 0 0 & 0 0 0 0 0 0 & 0 0 0 0 0 0 \\ \hline & \star \star \star 0 0 0 & \star \star \star 0 0 0 & 0 0 0 0 0 0 & 0 0 0 0 0 0 \\ \hline 0 & \star \star \star 0 0 & 0 \star \star \star 0 0 & 0 0 0 0 0 0 & 0 0 0 0 0 0 \\ \hline 0 & 0 \star \star \star 0 & 0 0 \star \star \star 0 & 0 0 0 0 0 0 & 0 0 0 0 0 0 \\ \hline 0 & 0 0 \star \star \star 0 & 0 0 0 \star \star \star 0 & 0 0 0 0 0 0 & 0 0 0 0 0 0 \\ \hline 0 & 0 0 0 \star \star \star & 0 0 0 0 \star \star \star & 0 0 0 0 0 0 & 0 0 0 0 0 0 \\ \hline \hline & \star \star 0 0 0 0 & \star \star 0 0 0 0 & \star \star 0 0 0 0 & 0 0 0 0 0 0 \\ \hline & \star \star \star 0 0 0 & \star \star \star 0 0 0 & \star \star \star 0 0 0 & 0 0 0 0 0 0 \\ \hline 0 & \star \star \star 0 0 & 0 \star \star \star 0 0 & 0 \star \star \star 0 0 & 0 0 0 0 0 0 \\ \hline 0 & 0 \star \star \star 0 & 0 0 \star \star \star 0 & 0 0 \star \star \star 0 & 0 0 0 0 0 0 \\ \hline 0 & 0 0 \star \star \star 0 & 0 0 0 \star \star \star 0 & 0 0 0 \star \star \star 0 & 0 0 0 0 0 0 \\ \hline 0 & 0 0 0 \star \star \star & 0 0 0 0 \star \star \star & 0 0 0 0 \star \star \star & 0 0 0 0 0 0 \\ \hline \hline 0 & 0 0 0 0 0 & \star \star 0 0 0 0 & \star \star 0 0 0 0 & \star \star 0 0 0 0 \\ \hline 0 & 0 0 0 0 0 & \star \star \star 0 0 0 & \star \star \star 0 0 0 & \star \star \star 0 0 0 \\ \hline 0 & 0 0 0 0 0 & 0 \spadesuit \clubsuit \diamondsuit 0 0 & 0 \clubsuit \spadesuit \clubsuit \diamondsuit 0 0 & 0 \diamondsuit \spadesuit \clubsuit \diamondsuit 0 0 \\ \hline 0 & 0 0 0 0 0 & 0 0 \star \star \star 0 0 & 0 0 \star \star \star 0 0 & 0 0 \star \star \star 0 0 \\ \hline 0 & 0 0 0 0 0 & 0 0 0 \star \star \star 0 & 0 0 0 \star \star \star 0 & 0 0 0 \star \star \star 0 \\ \hline 0 & 0 0 0 0 0 & 0 0 0 0 \star \star \star & 0 0 0 0 \star \star \star & 0 0 0 0 \star \star \star \\ \hline \hline 0 & 0 0 0 0 0 & 0 0 0 0 0 0 & \star \star 0 0 0 0 & \star \star 0 0 0 0 \\ \hline 0 & 0 0 0 0 0 & 0 0 0 0 0 0 & \star \star \star 0 0 0 & \star \star \star 0 0 0 \\ \hline 0 & 0 0 0 0 0 & 0 0 0 0 0 0 & 0 \star \star \star 0 0 0 & 0 \star \star \star 0 0 0 \\ \hline 0 & 0 0 0 0 0 & 0 0 0 0 0 0 & 0 0 \star \star \star 0 0 & 0 0 \star \star \star 0 0 \\ \hline 0 & 0 0 0 0 0 & 0 0 0 0 0 0 & 0 0 0 \star \star \star 0 & 0 0 0 \star \star \star 0 \\ \hline 0 & 0 0 0 0 0 & 0 0 0 0 0 0 & 0 0 0 0 \star \star \star & 0 0 0 0 \star \star \star \\ \hline \end{array} \right)$$

Matrix  $\mathbf{Q}$  will now be a block-tridiagonal matrix, with block elements that are tridiagonal themselves. Also  $\mathbf{Q}$  is a *banded* matrix, meaning that all elements can be included within a band around the main diagonal. The structure is given in figure 3.10 for an approximation that uses six points to discretize  $x$  and four points to discretize  $y$ . The elements  $\spadesuit$  give the elements that reflect moves to  $j \pm 1$ , used for the derivatives with respect to  $x$ ; the elements  $\clubsuit$  reflect moves to  $k \pm 1$ , used for the derivatives with respect to  $y$ ; and the elements  $\diamondsuit$  reflect moves to  $(j \pm 1, k \pm 1)$ , used for the cross derivative.

### FINITE DIFFERENCE APPROACHES

We have managed to represent the solution of the two-dimensional BS PDE as a system of ODEs, which is very similar to the approach we took when we first discussed the one-dimensional problem. The difference is of course that now the size of the matrix  $\mathbf{Q}$  is larger by one order of magnitude. Nevertheless, we can represent the problem as

$$\frac{\partial f(t)}{\partial t} = \mathbf{Q} \cdot \mathbf{f}(t)$$

By using the same arguments we can once again one can construct explicit, implicit and  $\theta$ -methods. For example, the Crank-Nicolson scheme will be of the usual form

$$\left(\mathbf{I} - \frac{1}{2}\mathbf{Q}\Delta t\right) \cdot \mathbf{f}^{i+1} = \left(\mathbf{I} + \frac{1}{2}\mathbf{Q}\Delta t\right) \cdot \mathbf{f}^i$$

As an example consider an option with payoffs that depend on two correlated assets that follow geometric Brownian motions. The BS PDE in terms of the log-prices  $x$  and  $y$  will be

$$\begin{aligned} \frac{\partial f(t, x, y)}{\partial t} &= \alpha_x \frac{\partial f(t, x, y)}{\partial x} + \alpha_y \frac{\partial f(t, x, y)}{\partial y} \\ &+ \frac{1}{2}\sigma_x^2 \frac{\partial^2 f(t, x, y)}{\partial x^2} + \frac{1}{2}\sigma_y^2 \frac{\partial^2 f(t, x, y)}{\partial y^2} + \rho\sigma_x\sigma_y \frac{\partial^2 f(t, x, y)}{\partial x \partial y} - rf(t, x, y) \end{aligned}$$

We also assume that a set of Neumann conditions is specified at each boundary, namely the corresponding derivatives  $\frac{\partial f}{\partial x}$  ( $\frac{\partial f}{\partial y}$  resp.) being equal to  $\varphi_{x,1}$  and  $\varphi_{x,N_x}$  ( $\varphi_{y,1}$  and  $\varphi_{y,N_y}$  resp.).

To build  $\mathbf{Q}$  we essentially consider an  $(N_y, N_y)$  tridiagonal matrix, where changes in  $y$  are captured, which has elements that are  $(N_x, N_x)$  matrices where changes in  $x$  are captured. We can represent in the following form where all submatrices have dimensions  $(N_x, N_x)$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{D} & \mathbf{B} & & & \\ \mathbf{C} & \mathbf{D} & \mathbf{E} & & \\ & \mathbf{C} & \mathbf{D} & \mathbf{E} & \\ & & \ddots & \ddots & \ddots & \\ & & & \mathbf{C} & \mathbf{D} & \mathbf{E} \\ & & & & \mathbf{C} & \mathbf{D} & \mathbf{E} \\ & & & & & \mathbf{F} & \mathbf{D} \end{pmatrix} \quad (3.11)$$

## BOUNDARY CONDITIONS

The boundary conditions will have an effect on these matrices. In particular, the first and last rows of all matrices will depend on boundary conditions on  $x$ . In addition, all elements of the block matrices  $\mathbf{B}$  and  $\mathbf{F}$  will depend on the boundary conditions imposed on  $y$ . The generic sum that  $\mathbf{Q}$  implements is given by

$$\begin{aligned} \frac{\partial f_{j,k}}{\partial t} &= q_{(-,-)}f_{j-1,k-1} + q_{(0,-)}f_{j,k-1} + q_{(+,-)}f_{j+1,k-1} \\ &+ q_{(-,0)}f_{j-1,k} + q_{(0,0)}f_{j,k} + q_{(+,0)}f_{j+1,k} \\ &+ q_{(-,+)}f_{j-1,k+1} + q_{(0,+)}f_{j,k+1} + q_{(+,+)}f_{j+1,k+1} \end{aligned}$$

where the coefficients are given by the following quantities (with the elements that correspond to figure 3.10 also indicated)

$$\begin{aligned}
 (\spadesuit): \quad q_{(\pm,0)} &= \pm\alpha_x/(2\Delta x) + \sigma_x^2/(2\Delta x^2) \\
 (\heartsuit): \quad q_{(0,\pm)} &= \pm\alpha_y/(2\Delta y) + \sigma_y^2/(2\Delta y^2) \\
 (\star): \quad q_{(0,0)} &= -r - \sigma_x^2/(\Delta x^2) - \sigma_y^2/(\Delta y^2) \\
 (\diamondsuit): \quad q_{(+,+)} &= q_{(-,-)} = +\rho\sigma_x\sigma_y/(4\Delta x\Delta y) \\
 (\clubsuit): \quad q_{(-,+)} &= q_{(+,-)} = -\rho\sigma_x\sigma_y/(4\Delta x\Delta y)
 \end{aligned}$$

Boundary conditions will influence the first and last rows of each block, as this is where the boundaries of  $x$  are positioned. The whole first and last blocks will be also affected, since this is where the boundaries of  $y$  are positioned. The first and last rows of these particular blocks will correspond to the corner boundaries. Also, the boundaries will specify a matrix of constants  $\mathbf{G}$ , just like the vector of constants we constructed in the univariate case.

For Neumann conditions the elements  $(1,2)$  and  $(N_x, N_x - 1)$  of each block are given by  $q_{(+,+)} + q_{(-,-)}$ . Of course a similar relationship will hold for all elements of the  $(1,2)$  and  $(N_y, N_y - 1)$  block, which will have elements given by  $q_{(+,+)} + q_{(-,-)}$ . Apparently the  $(1,2)$  and  $(N_x, N_x - 1)$  elements of these particular blocks will be dependent on both boundary conditions, and also the boundary condition across the diagonal. The values for these elements will be given by  $q_{(+,+)} + q_{(-,+)} + q_{(+,-)} + q_{(-,-)}$ .

The elements of the matrix  $\mathbf{G}$  will be also determined by the Neumann conditions for  $k = 2, \dots, N_y - 1$  and  $j = 2, \dots, N_x - 1$ . Say that  $\varphi_{x,(j,k)} = \frac{\partial f(x_j, y_k)}{\partial x}$ , and  $\varphi_{y,(j,k)} = \frac{\partial f(x_j, y_k)}{\partial y}$ .

$$\begin{aligned}
 \mathbf{G}(1, k) &= -2q_{(-,0)}\varphi_{x,(1,k)}\Delta x, & \mathbf{G}(N_x, k) &= +2q_{(+,0)}\varphi_{x,(N_x,k)}\Delta x \\
 \mathbf{G}(j, 1) &= -2q_{(0,-)}\varphi_{y,(j,1)}\Delta y, & \mathbf{G}(1, N_y) &= +2q_{(0,+)}\varphi_{y,(1,N_y)}\Delta y
 \end{aligned}$$

The corner elements of  $\mathbf{G}$  will be determined by both boundary conditions, as well as the boundary across the diagonal. For example, the element  $(1, 1)$  will be

$$\begin{aligned}
 \mathbf{G}(1, 1) &= -2q_{(-,0)}\varphi_{x,(1,1)}\Delta x - 2q_{(0,-)}\varphi_{y,(1,1)}\Delta y \\
 &\quad + q_{(-,-)}[-\varphi_{x,(1,1)} - \varphi_{y,(1,1)}]\sqrt{\Delta x^2 + \Delta y^2}
 \end{aligned}$$

The other four points have similar expressions. We will vectorize the constraints by stacking the columns of  $\mathbf{G}$  into the vector  $\mathbf{g}$ .

If we include the impact of the boundary conditions (and keep in mind that they might be time varying), the system of ODEs that will give us an approximate solution to the two-dimensional PDE is now given by

$$\frac{\partial \mathbf{f}(t)}{\partial t} = \mathbf{Q} \cdot \mathbf{f}(t) + \mathbf{g}(t) \tag{3.12}$$

If the boundary conditions are homogeneous,  $\mathbf{g}(t) = \mathbf{g}$ , then the solution of the system is

$$\mathbf{f}(t) = \exp(\mathbf{Q}t) \cdot \mathbf{f}(0) + \mathbf{Q}^{-1} \cdot [\exp(\mathbf{Q}t) - \mathbf{I}] \cdot \mathbf{g}$$

We also cast system (3.12) in the  $\theta$ -form, and approximate it as the solution of the updating scheme

$$(\mathbf{I} - \theta \cdot \mathbf{Q}\Delta t) \cdot \mathbf{f}^{i+1} = (\mathbf{I} + (1 - \theta) \cdot \mathbf{Q}\Delta t) \cdot \mathbf{f}^i + \theta \cdot \mathbf{g}^{i+1} + (1 - \theta) \cdot \mathbf{g}^i$$

Once again if the boundaries are homogeneous in time the scheme can be written as

$$(\mathbf{I} - \theta \cdot \mathbf{Q}\Delta t) \cdot \mathbf{f}^{i+1} = (\mathbf{I} + (1 - \theta) \cdot \mathbf{Q}\Delta t) \cdot \mathbf{f}^i + \mathbf{g}$$

In theory solving this system does not present any differences, but in practice it might not be feasible since  $\mathbf{Q}$  is not tridiagonal. For that reason a number of *alternating direction implicit* (ADI) and *local one-dimensional* (LOD, also known as *Soviet splitting*) schemes are typically used. Such schemes do not solve over all dimensions simultaneously, but instead split each time step into substeps, and assume that over each substep the system moves across a single direction. Therefore at each substep one has to solve a system that is indeed tridiagonal.

### ALTERNATIVE DIRECTION IMPLICIT METHODS

To understand the ADI methods it is intuitive to write down the Crank-Nicolson system in terms of operators

$$\begin{aligned} & (1 - \alpha_x \mathcal{D}_x - \alpha_y \mathcal{D}_y - \beta_x^* \mathcal{D}_x^2 - \beta_y^* \mathcal{D}_y^2 - \beta_{xy} \mathcal{D}_{xy}^2 - \gamma) \mathbf{f}^{i+1} \\ &= (1 + \alpha_x \mathcal{D}_x + \alpha_y \mathcal{D}_y + \beta_x^* \mathcal{D}_x^2 + \beta_y^* \mathcal{D}_y^2 + \beta_{xy} \mathcal{D}_{xy}^2 + \gamma) \mathbf{f}^i \end{aligned}$$

where we have defined  $\beta_x^* = \beta_x - \frac{\beta_{xy}}{2} \frac{\Delta x}{\Delta y}$  and  $\beta_y^* = \beta_y - \frac{\beta_{xy}}{2} \frac{\Delta y}{\Delta x}$ , to save some space. Now we can put down the approximation

$$\begin{aligned} & \left[ 1 - \alpha_x \mathcal{D}_x - \beta_x^* \mathcal{D}_x^2 - \frac{\gamma}{3} \right] \left[ 1 - \beta_{xy} \mathcal{D}_{xy}^2 - \frac{\gamma}{3} \right] \left[ 1 - \alpha_y \mathcal{D}_y - \beta_y^* \mathcal{D}_y^2 - \frac{\gamma}{3} \right] \mathbf{f}^{i+1} \\ &= \left[ 1 + \alpha_x \mathcal{D}_x + \beta_x^* \mathcal{D}_x^2 + \frac{\gamma}{3} \right] \left[ 1 + \beta_{xy} \mathcal{D}_{xy}^2 + \frac{\gamma}{3} \right] \left[ 1 + \alpha_y \mathcal{D}_y + \beta_y^* \mathcal{D}_y^2 + \frac{\gamma}{3} \right] \mathbf{f}^i \end{aligned}$$

It is tedious to go through the algebra, but one can show that the approximation of the operators is at least of second order in time and both directions. Therefore the results are not expected to deteriorate due to this operator splitting. In the [Peaceman and H. H. Rachford \(1955\)](#) scheme we implement the following three steps, solving for auxiliary values  $\mathbf{f}^*$  and  $\mathbf{f}^{**}$

$$\begin{aligned} & \left[ 1 - \alpha_y \mathcal{D}_y - \beta_y^* \mathcal{D}_y^2 - \frac{\gamma}{3} \right] \mathbf{f}^* = \left[ 1 + \alpha_y \mathcal{D}_y + \beta_y^* \mathcal{D}_y^2 + \frac{\gamma}{3} \right] \mathbf{f}^i \\ & \left[ 1 - \beta_{xy} \mathcal{D}_{xy}^2 - \frac{\gamma}{3} \right] \mathbf{f}^{**} = \left[ 1 + \beta_{xy} \mathcal{D}_{xy}^2 + \frac{\gamma}{3} \right] \mathbf{f}^* \\ & \left[ 1 - \alpha_x \mathcal{D}_x - \beta_x^* \mathcal{D}_x^2 - \frac{\gamma}{3} \right] \mathbf{f}^{i+1} = \left[ 1 + \alpha_x \mathcal{D}_x + \beta_x^* \mathcal{D}_x^2 + \frac{\gamma}{3} \right] \mathbf{f}^{**} \end{aligned}$$

For the D'yakonov scheme (see [Marchuk, 1990; McKee, Wall, and Wilson, 1996](#)) we use a slightly different splitting where at the first step we produce the complete right-hand-side

$$\begin{aligned} \left[1 - \alpha_y \mathcal{D}_y - \beta_y^* \mathcal{D}_y^2 - \frac{\gamma}{3}\right] \mathbf{f}^* &= \left[1 + \alpha_x \mathcal{D}_x + \beta_x^* \mathcal{D}_x^2 + \frac{\gamma}{3}\right] \left[1 + \beta_{xy} \mathcal{D}_{xy}^2 + \frac{\gamma}{3}\right] \\ &\quad \left[1 + \alpha_y \mathcal{D}_y + \beta_y^* \mathcal{D}_y^2 + \frac{\gamma}{3}\right] \mathbf{f}^i \\ \left[1 - \beta_{xy} \mathcal{D}_{xy}^2 - \frac{\gamma}{3}\right] \mathbf{f}^{**} &= \mathbf{f}^* \\ \left[1 - \alpha_x \mathcal{D}_x - \beta_x^* \mathcal{D}_x^2 - \frac{\gamma}{3}\right] \mathbf{f}^{i+1} &= \mathbf{f}^{**} \end{aligned}$$

In both cases the operations are implemented using matrices that can be cast in tridiagonal form by permutations of their elements. In the multidimensional PDE problems, one has to take special care when dealing with the boundary conditions, as it may be confusing. Also, some decisions have to be made on the corners, which are affected by boundary conditions on more than one dimension.

## 3.5 A TWO-DIMENSIONAL SOLVER IN MATLAB

We now turn into the implementation of a PDE solver in two space dimensions, using the  $\theta$ -method. Essentially this is more of a book-keeping exercise, where we need to consider the structure of matrix  $\mathbf{Q}$ , and especially the boundary conditions. Here we will focus on boundary conditions of the Neumann type.

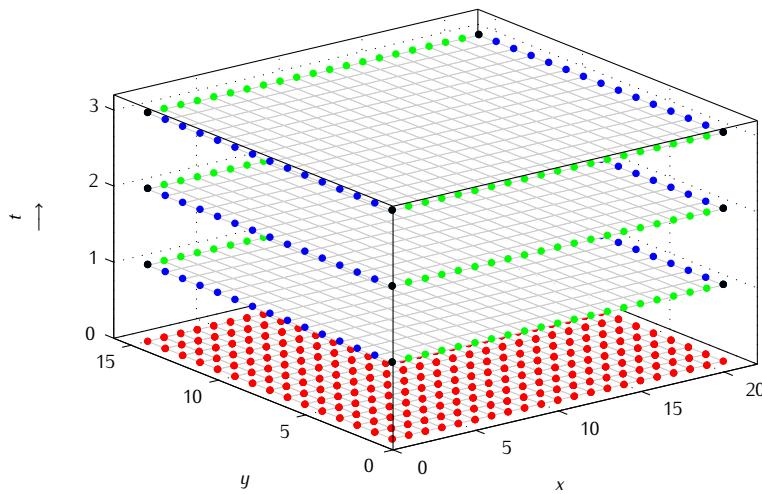
We will assume that the payoff function returns not only the function values, but also the derivative over the boundaries, together with the derivatives at the corner points across the diagonal directions. Figure 3.11 shows the positions of these boundaries. Each horizontal slice gives the  $(x, y)$ -grid at a different time point. The colored points denote the boundary and initial values that are necessary to solve the PDE numerically. In particular the bottom slice, at  $t = 0$ , gives the set of initial conditions that need to be specified to start the algorithm. In the next time periods the boundaries are illustrated. The blue points show the boundaries at  $x = x_1$  and  $x = x_{N_x}$ , while the green points show the boundaries at  $y = y_1$  and  $y = y_{N_y}$ . At the black (corner) point both boundary conditions will have an impact. Essentially these point illustrate where matrix  $\mathbf{G}$  has potentially non-zero elements. The elements of matrix  $\mathbf{Q}$  that are affected lie just within these points.

As an example we will use a *correlation option*, which is essentially a European call option on the minimum price of two underlying assets. The payoff function of this derivative is

$$\Pi(S_1, S_2) = \max(\min(S_1, S_2) - K)$$

We will make the assumption that both assets follow geometric Brownian motions, with correlation parameter  $\rho$ . The pricing function will satisfy the two-

FIGURE 3.11: Schematic representation of the evolution of a two-dimensional PDE. The points where initial and boundary conditions have to be specified are also illustrated. red points: initial conditions ( $t = 0$ ) active; blue points: boundary conditions on  $x$  active; green points: boundary conditions on  $y$  active; and black points: both boundary conditions active.



LISTING 3.11: callmin.m: Payoff and boundaries for a two-asset option.

```
% callmin.m
function [y, yD] = callmin(xv, yv, p)
K = p.K; % strike price
[sx, sy] = meshgrid(exp(xv), exp(yv)); % grid
5 dx = xv(2)-xv(1); dy = yv(2)-yv(1); % differentials
y = max(min(sx,sy)-1, 0); % payoff function
% partial derivatives
yD.Dx0 = (y(2,:)-y(1,:))/dy; % (1 ,k )
yD.DxN = (y(end,:)-y(end-1,:))/dy; % (Nx,k )
10 yD.Dy0 = (y(:,2)-y(:,1))/dx; % (j ,1 )
yD.DyN = (y(:,end)-y(:,end-1))/dx; % (j , Ny)
% diagonal derivatives
dz = sqrt(dx^2+dy^2); % diag differential
yD.D00 = (y(1,1)-y(2,2))/dz; % (1 ,1 )
15 yD.D0N = (y(1,end)-y(2,end-1))/dz; % (Nx,1 )
yD.DN0 = (y(end,1)-y(end-1,2))/dz; % (1 ,Ny)
yD.DNN = (y(end,end)-y(end-1,end-1))/dz; % (Nx,Ny)
```

dimensional Black-Scholes PDE, which in terms of the log-prices  $x_1 = \log S_1$  and  $x_2 = \log S_2$  can be written as

$$\begin{aligned} -\frac{\partial f(t, x_1, x_2)}{\partial t} + \alpha_1 \frac{\partial f(t, x_1, x_2)}{\partial x_1} + \alpha_2 \frac{\partial f(t, x_1, x_2)}{\partial x_2} + \frac{1}{2} \sigma_1^2 \frac{\partial^2 f(t, x_1, x_2)}{\partial x_1^2} \\ + \frac{1}{2} \sigma_2^2 \frac{\partial^2 f(t, x_1, x_2)}{\partial x_2^2} + \rho \sigma_1 \sigma_2 \frac{\partial^2 f(t, x_1, x_2)}{\partial x_1 \partial x_2} - r f(t, x_1, x_2) = 0 \end{aligned}$$

for  $\alpha_1 = r - \frac{1}{2}\sigma_1^2$  and  $\alpha_2 = r - \frac{1}{2}\sigma_2^2$ . As this is a European style contract, the Neumann boundaries across  $x_1$  and  $x_2$  will be such that the derivative at each one of these points is the same through time. Therefore, and since the function is piecewise linear, there is no point in explicitly computing the partial derivative at each point, since we can do that numerically. Listing 3.11 gives the Matlab code that returns the payoff function and vectors of derivatives. One can verify that the derivatives are computed numerically rather explicitly in all directions.

Listings 3.12-3.13 give the Matlab code that implements the  $\theta$ -method to solve the two-dimensional BS PDE. In the first part we setup the matrices that will serve as the blocks that make up the Q-matrix. Given the small number of nonzero elements, all matrix definitions and manipulations are done using sparse matrix commands. Matrices B to F corresponds to the blocks  $B \dots F$  in equation 3.11. Matrix G keeps the constraints, as discussed in section 3.4, while the reshaped (stacked) form g corresponds to vector  $g$ .

The Matlab code that actually implements the Crank-Nicolson method to price the correlation is given in listing 3.14. Two assets are considered that exhibit different volatilities. The discretization grid across the two dimensions is constructed using  $(51 \times 51)$  points. The call has half year to maturity, and we use 30 time steps to compute the price. Therefore we will need to solve 30 systems of 2601 equations with 2601 unknowns to arrive to the result: a substantial computational demand.

## 3.6 EXTENSIONS

Apart from the contracts and the techniques we discussed, there is a very large number of exotic options with features that can be implemented within the PDE framework. Sometimes we will need to extend the dimensionality of the problem to accommodate for these special features. For example, in many cases a rebate is offered when the barrier it triggered. This will make sure that breaching the barrier will not leave you empty handed. It is straightforward to handle such rebates in the finite differences procedure.

Other contracts attempt to *cushion* the barrier effect and the discontinuities it creates. For example, in Parisian options the barrier is triggered only if the barrier remains breached for a given (cumulative) time. To solve for this option we need to introduce an extra variable, namely the *cumulative time that barrier has been breached*, say  $\tau$ .

LISTING 3.12: pde\_bs\_2d.m: Solver for a two dimensional PDE (part I).

```
% pde_bs_2d.m
function [Sx, Sy, fM] = pde_bs_2d(f, p)
% model parameters
r = p.r; rho = p.rho; theta = p.theta;
5 T = p.T; Nt = p.Nt;
sx = p.sigmax; sy = p.sigmay;
ax = r - 0.5*sx^2; ay = r - 0.5*sy^2;
% discretization parameters
bx = p.bx; Nx = p.Nx; by = p.by; Ny = p.Ny;
10 dx = 2*bx/(Nx-1); xv = -bx:dx:bx; % grid over log-S1
dy = 2*by/(Ny-1); yv = -by:dy:by; % grid over log-S2
dz = sqrt(dx^2+dy^2); % diagonal grid size
[f1, yD] = feval(f, xv, yv, p); % call payoff function
% derivatives over the boundaries
15 Dx0 = yD.Dx0; DxN = yD.DxN; Dy0 = yD.Dy0; DyN = yD.DyN;
DOO = yD.DOO; DON = yD.DON; DNO = yD.DNO; DNN = yD.DNN;
% stack columns
f0 = reshape(f1, Ny*Nx, 1);
Sx = exp(xv); Sy = exp(yv); % actual prices
20 qp0 = 0.5*ax/dx + 0.5*sx^2/dx^2; % (+,0)
qm0 = -0.5*ax/dx + 0.5*sx^2/dx^2; % (-,0)
q0p = 0.5*ay/dx + 0.5*sy^2/dy^2; % (0,+)
q0m = -0.5*ay/dx + 0.5*sy^2/dy^2; % (0,-)
q00 = -r - sx^2/dx^2 - sy^2/dy^2; % (0,0)
25 qpp = 0.25*rho*sx*sy/dx/dy; qmm = qpp; % (+,+), (-,-)
qpm = -qpp; qmp = -qpp; % (+,-), (-,+)
I = ones(Nx, 1); IO = ones(Nx-1, 1);
% matrices for x-boundaries
D = spdiags(q00*I, 0, Nx, Nx) + ...
30 spdiags(qm0*I, -1, Nx, Nx) + spdiags(qp0*I, +1, Nx, Nx);
C = spdiags(q0m*I, 0, Nx, Nx) + ...
spdiags(qmm*I, -1, Nx, Nx) + spdiags(qpm*I, +1, Nx, Nx);
E = spdiags(q0p*I, 0, Nx, Nx) + ...
spdiags(qmp*I, -1, Nx, Nx) + spdiags(qpp*I, +1, Nx, Nx);
35 % boundary corrections for x
D(1,2) = qp0+qm0; D(Nx,Nx-1) = qp0+qmm;
C(1,2) = qpm+qmm; C(Nx,Nx-1) = qpm+qmm;
E(1,2) = qpp+qmp; E(Nx,Nx-1) = qpp+qmp;
% matrices for y-boundaries
40 B = E + spdiags(q0m*I, 0, Nx, Nx) + ...
spdiags(qmm*I, -1, Nx, Nx) + spdiags(qpm*I, +1, Nx, Nx);
F = C + spdiags(q0p*I, 0, Nx, Nx) + ...
spdiags(qmp*I, -1, Nx, Nx) + spdiags(qpp*I, +1, Nx, Nx);
45 % boundary corrections for y
B(1,2) = B(1,2) + qmm; B(Nx,Nx-1) = B(Nx,Nx-1) + qpm;
F(1,2) = F(1,2) + qmp; F(Nx,Nx-1) = F(Nx,Nx-1) + qpp;
```

for copies, comments, help etc. visit <http://www.theponytail.net/>

LISTING 3.13: pde\_bs\_2d.m: Solver for a two dimensional PDE (part II).

```

Q = sparse (Nx*Ny , Nx*Ny); % Q-matrix
G = sparse (Nx,Ny);           % matrix of constants
% first block row
50 Q(1:Nx,1:Nx) = D;
Q(1:Nx,(Nx+1):(2*Nx)) = B;
G(:,1) = -2*(qmm+q0m+qpm)*Dy0*dx;
G(1,1) = H(1,2) -2*(qmm+qm0+qmp)*Dx0(1)...
    *dx -2*qmm*D00*dz;
G(Nx,1) = H(Nx,1)+2*(qpm+qp0+qpp)*DxN(1)...
    *dx -2*qpm*D00*dz;
for jy=2:Ny-1 % loop through middle block rows
    Q((jy-1)*Nx+1:jy*Nx,(jy-1)*Nx+1:jy*Nx) = D;
    Q((jy-1)*Nx+1:jy*Nx,(jy-2)*Nx+1:(jy-1)*Nx) = C;
60    Q((jy-1)*Nx+1:jy*Nx,jy*Nx+1:(jy+1)*Nx) = E;
    G(1,jy) = -2*(qmm+qm0+qmp)*Dx0(jy)*dx;
    G(Nx,jy)= 2*(qpm+qp0+qpp)*DxN(jy)*dx;
end
% last block row
65 Q((Ny-1)*Nx+1:Ny*Nx,(Ny-1)*Nx+1:Ny*Nx) = D;
Q((Ny-1)*Nx+1:Ny*Nx,(Ny-2)*Nx+1:(Ny-1)*Nx) = F;
G(:,Ny) = 2*(qmm+q0m+qpm)*DyN*dy;
G(1,Ny) = G(1,Ny) -2*(qmm+qm0+qmp)*Dx0(Ny)*dx...
    -2*qmp*D0N*dz;
70 G(Nx,Ny)= G(Nx,Ny)+2*(qpm+qp0+qpp)*DxN(Ny)*dx...
    -2*qpp*DNN*dz;
g = reshape(G, Ny*Nx, 1); % stack constant columns
dt = T/(Nt-1); % time step
QI = speye(Ny*Nx) - theta*dt*Q; % implicit part
75 QE = speye(Ny*Nx) + (1-theta)*dt*Q; % explicit part
f = f0;
for m=1:Nt-1 % loop through time steps
    f = QI \ (QE*f + g*dt); % solve theta-system
end
80 fM = reshape(f,Nx,Ny); % re-stack for output

```

Apparently, the derivative price will now be a function  $f(S, t, \tau)$ . Also,  $\tau$  will evolve as an ODE  $d\tau = dt$  if  $s_t > \log B$  and  $d\tau = 0$ , otherwise. The price will satisfy a different PDE within each domain

$$\begin{aligned}
 S_t < B : f_t(S, t, \tau) + rSf_S(S, t, \tau) + \frac{1}{2}\sigma^2S^2f_{SS}(S, t, \tau) &= rf(S, t, \tau) \\
 S_t \geq B : f_t(S, t, \tau) + rSf_S(S, t, \tau) + f_\tau(S, t, \tau) + \frac{1}{2}\sigma^2S^2f_{SS}(S, t, \tau) &= rf(S, t, \tau)
 \end{aligned}$$

LISTING 3.14: pde\_bs\_2d\_impl.m: Implementation of the two dimensional solver.

```
% pde_bs_2d_impl.m
p.r      = 0.04; p.rho     = 0.30; % intrst / corr
p.sigmax = 0.20; p.sigmay   = 0.30; % volatilities
p.K      = 1.00; p.theta    = 0.50; % strike / theta
5 p.T     = 0.50; p.Nt     = 30; % time steps
p.bx    = 0.50; p.Nx     = 51; % log-S1 grid
p.by    = 0.50; p.Ny     = 51; % log-S2 grid
[Sx, Sy, fM] = pde_bs_2d(@callmin, p); % call pricer
ix = find((Sx>.8).* (Sy<1.2));
10 iy = find((Sy>.8).* (Sy<1.2));
contour(Sx(ix), Sy(iy), fM(iy,ix));
```

To solve for this contract we would need a grid over a 3-D region, and of course a more complex set of boundary conditions needs to be specified.

Another group of problems that can be attacked using PDEs arises when single asset models with more than one factors are considered. For example one might want to price derivative contracts under the [Heston \(1993\)](#) stochastic volatility model, where

$$\begin{aligned} dS(t) &= \mu S(t)dt + \sqrt{v(t)}S(t)dB_s(t) \\ dv(t) &= \kappa [\bar{v} - v(t)]dt + \phi \sqrt{v(t)}dB_v(t) \\ dB_s(t)dB_v(t) &= \rho dt \end{aligned}$$

Here a derivative will apparently depend on the current volatility as well as the price, having a pricing function  $f(t, S, v)$ . Therefore, the PDE that will be satisfied by such a contract will be a two-dimensional one.

Finally, in modeling fixed-income or credit related securities (and their derivatives) one might need to resort to multi-factor specifications, for example a corporate bond being a function of an  $M$ -dimensional *state* vector  $x(t)$  that has dynamics express via a stochastic differential equation

$$dx(t) = \mu(t, x(t))dt + \Sigma(t, x(t)) \cdot dB(t)$$

The PDE approach can also be applied in such a setting, although as the dimensionality increases implementation become infeasible (and simulation-based methods are typically preferred).



## Transform methods

Following the success of the [Black and Scholes \(1973\)](#) model on pricing and hedging derivative contracts, there has been a surge of research on models that can capture the stylized facts of asset and derivative markets. Although the BS paradigm is elegant and intuitive, it still maintains a number of assumptions that are too restrictive. In particular, the assumption of identically distributed and independent Gaussian innovations clearly contradicts empirical evidence.

When developing specifications that relax these assumptions, academics and practitioners alike discovered that apart from the BS case, very few models offer European option prices in closed form. Being able to rapidly compute European call and put prices is paramount, since typically a theoretical model will be calibrated on a set of prices that come from options markets. The parameter values retrieved from this calibration will be used to price and devise hedging strategies for more exotic contracts.

It turned out that, in many interesting cases, even though derivative prices or the risk-neutral density cannot be explicitly computed, the characteristic function of the log-returns is tractable. Based on this quantity, researchers did indeed link the characteristic function to the European call and put price, via an application of Fourier transforms (see [Heston, 1993](#); [Bates, 1998](#); [Madan, Carr, and Chang, 1998](#); [Carr and Madan, 1999](#); [Duffie, Pan, and Singleton, 2000](#); [Bakshi and Madan, 2000](#), *inter alia* for different modeling approaches).

### 4.1 THE SETUP

Assume that we are interested in an economy where there exists an asset with price process  $S(t)$ . We also assume that there is a risk-free asset, offering a deterministic rate of return  $r(t)$ , implying a set of bond prices  $B(t) = \exp \int_0^t r(s)ds$ . We start our analysis with the logarithmic return over a maturity  $T$ , say  $X(T) = \log \frac{S(T)}{S(0)}$ . We understand that as a random variable,  $X(T)$  will be distributed according to a probability measure  $\mathcal{P}$ , the *true* or *objective* probability

measure. Also, we assume that there exists an *equivalent* probability measure  $\mathbb{Q}$ , under which the discounted price will form a martingale

$$B(T) \cdot E_{\mathbb{Q}} S(T) = S(0) \quad (4.1)$$

This is called the *risk-neutral* or *risk adjusted* probability measure. This measure need not be unique, given the current set of bond and asset prices, unless the market is complete, but all derivative contracts will have a no-arbitrage price that is equal to their discounted expected payoffs under this measure. That is to say, a European call option will satisfy

$$P_{\text{call}} = B(T) \cdot E_{\mathbb{Q}} \max(S(T) - K, 0)$$

Under the BS assumptions  $\mathbb{Q}$  will be unique, and  $X(T)$  will follow a Gaussian distribution under both  $\mathcal{P}$  and  $\mathbb{Q}$ . Under more general assumptions this need not be the case. Since we are interested in the pricing of derivatives we are going to ignore the true probability measure from now on, and focus instead on the qualities and characteristics of the risk-neutral measure. Therefore all expectations are assumed to be under  $\mathbb{Q}$ , unless explicitly stated otherwise.

## FOURIER TRANSFORMS

One of the most important tools for solving PDEs is the *Fourier transform* of a function  $f(x)$ . In particular, we define as the Fourier transform of  $f(x)$  a new function  $\phi(u)$ , such that

$$\mathcal{F}[f](u) = \phi(u) = \int_{\mathbb{R}} \exp(iux) f(x) dx$$

where  $i = \sqrt{-1}$  is the imaginary unit. It turns out that each function  $f$  defines a *unique* transform  $\phi$ , and this transform is invertible: if we are given  $\phi$  we can retrieve the original function  $f$ , using the *inverse Fourier transform*

$$\mathcal{F}^{-1}[\phi](x) = f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-iux) \phi(u) du$$

There can be some confusion, as different disciplines define the Fourier transform slightly different, setting  $\exp(\pm iux)$  the other way round, or multiplying both integrals with  $\frac{1}{\sqrt{2\pi}}$  to result in symmetric expressions. Here we use the definition that Matlab implements, but one has to always verify what a computer language offers.

Fourier transforms have some properties that make them invaluable tools for solutions of differential equations, the most important being that the transform is a linear operator

$$\mathcal{F}[af_1 + bf_2](u) = a\mathcal{F}[f_1](u) + b\mathcal{F}[f_2](u)$$

and that the Fourier transform of a derivative is given by

$$\mathcal{F}\left[\frac{d^n f(x)}{dx^n}\right](u) = (iu)^n \mathcal{F}[f](u)$$

if all derivatives up to order  $n$  decay to zero for large  $|x|$ .

To illustrate the point, consider the BS PDE in terms of the logarithms  $x = \log S$ , namely

$$\frac{\partial f(t, x)}{\partial t} = \alpha \frac{\partial f(t, x)}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f(t, x)}{\partial x^2} - rf(t, x)$$

If we apply the Fourier transform (with respect to  $x$ ) on both sides, then the left-hand-side becomes

$$\begin{aligned} \mathcal{F}\left[\frac{\partial}{\partial t} f(t, x)\right](u) &= \int_{\mathbb{R}} \exp(iu) \frac{\partial}{\partial t} f(t, x) dx \\ &= \frac{\partial}{\partial t} \left\{ \int_{\mathbb{R}} \exp(iu) f(t, x) dx \right\} = \frac{\partial}{\partial t} \phi(t, u) \end{aligned}$$

while the right-hand-side yields

$$\mathcal{F}\left[\left(\alpha \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial x^2} - r\right) f(t, x)\right](u) = \left[\alpha(iu) + \frac{1}{2}(iu)^2 - r\right] \phi(t, u)$$

Therefore, by applying the Fourier transform we actually transformed a complicated second order PDE into a simple first order ODE which has a straightforward solution

$$\begin{aligned} \frac{\partial}{\partial t} \phi(t, u) &= \left[\alpha iu - \frac{1}{2}u^2 - r\right] \phi(t, u) \\ \implies \phi(t, u) &= \phi(0, u) \cdot \exp\left(\alpha iut - \frac{1}{2}u^2t - rt\right) \end{aligned}$$

with  $\phi(0, u)$  the initial condition.

## CHARACTERISTIC FUNCTIONS

If  $f$  is a probability density function that measures a random variable, say  $X(t)$ , then its Fourier transform is called the *characteristic function* of the random variable. It is also convenient to represent the characteristic function as an expectation, namely

$$\phi(t, u) = E \exp(iuX(t))$$

Characteristic functions are typically covered in most statistics textbooks.<sup>1</sup> Since functions and their Fourier transforms uniquely define each other, the

---

<sup>1</sup> A good reference for characteristic functions and their properties is [Kendall and Stuart \(1977, ch 4\)](#).

characteristic function will have enough information to uniquely define the probability distribution of the random variable. In particular, the inverse Fourier transform will determine the probability density function.

In many cases it is tractable to solve for the characteristic function of a random variable or a process, rather than the probability density itself. A large and very flexible class of processes, the Lévy processes, are in fact defined through their characteristic functions.

Characteristic functions have more important properties. By taking derivatives at the origin  $u = 0$ , one can retrieve successive moments of the random variable, as

$$\mathbb{E}[X(t)]^n = (\mathrm{i}^{-n}) \left. \frac{\partial^n \phi(t, u)}{\partial u^n} \right|_{u=0}$$

This means that qualitative properties of the distribution, such as the volatility, skewness and kurtosis can be ascertained directly from the characteristic function. In addition, it becomes straightforward to implement calibration methods that are based on the moments.

The characteristic function has the property  $\phi(t, -u) = \bar{\phi}(t, u)$ , with  $\bar{z}$  the complex conjugate. Thus, the real part is an even function over  $u$ , while the imaginary part is odd. This is in line with the fact that the probability density is a real valued function, since to achieve that when integrating over the real line the imaginary parts must cancel out. One can use this property to write the Fourier inversion that recovers the probability density function as

$$f(t, x) = \frac{1}{\pi} \int_0^\infty \operatorname{Re}[\exp(-\mathrm{i}ux)\phi(t, u)] du$$

The cumulative density function is of course (the function  $\mathbb{N}(\cdot)$  is the indicator function)

$$F(t, x) = \mathbb{P}[X(t) \leq x] = \mathbb{E}[\mathbb{N}(X \leq x)] = \int_{-\infty}^x f(s) ds$$

It is also possible to recover the cumulative density function, from the characteristic function

$$\begin{aligned} F(t, x) &= \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{\exp(\mathrm{i}ux)\phi(t, -u) - \exp(-\mathrm{i}ux)\phi(t, u)}{\mathrm{i}u} du \\ &= \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \operatorname{Re} \left[ \frac{\exp(-\mathrm{i}ux)\phi(t, u)}{\mathrm{i}u} \right] du \end{aligned}$$

### THE “DAMPENED” CUMULATIVE DENSITY

Computing the cumulative density as above can be very cumbersome, and the approach does not lend itself naturally to the FFT techniques that we will discuss later. One main drawback is the fact that the integrand diverges at zero, rendering the numerical integration unstable in many cases. Here we present a technique which allows us to rewrite the cumulative density as a Fourier

transform that is well defined. We will use exactly the same trick in the next section, in order to compute a call option price as a single and numerically tractable Fourier transform.

We introduce the *damping factor*  $\eta > 0$ , and define the damped cumulative probability as

$$F^\eta(t, x) = \exp(-\eta x) P[X(t) \leq x]$$

It is possible to derive the characteristic function of this function, say  $\phi^\eta(t, u)$  as follows

$$\begin{aligned} \phi^\eta(t, u) &= \int_{\mathbb{R}} \exp(iux) F^\eta(t, x) dx = \int_{\mathbb{R}} \exp(iux) \exp(-\eta x) P[X(t) \leq x] dx \\ &= \int_{\mathbb{R}} \int_{-\infty}^x \exp(iux - \eta z) f(t, z) dz dx \end{aligned}$$

The order of integration can be reversed as follows (details on how exactly this is carried out can be found in the next section where the same approach is implemented in an option pricing framework)

$$\begin{aligned} \phi^\eta(t, u) &= \int_{\mathbb{R}} \int_z^\infty \exp((iu - \eta)z) f(t, z) dz dx \\ &= \int_{\mathbb{R}} \frac{\exp((iu - \eta)z)}{\eta - iu} f(t, z) dz = \frac{1}{\eta - iu} \phi(t, \eta + iu) \end{aligned}$$

We can therefore compute the cumulative probability by “un-damping” this characteristic function, in effect computing

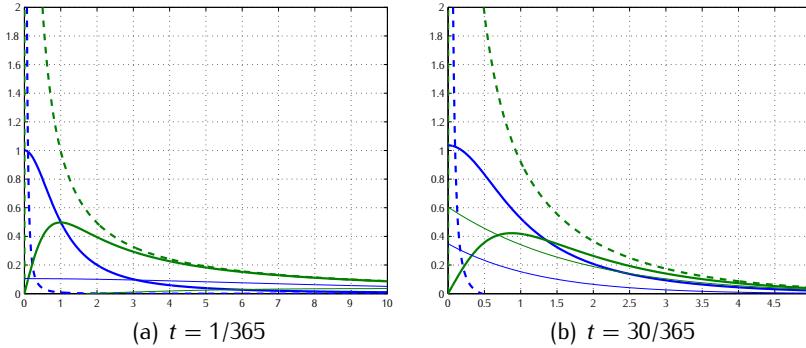
$$F(t, x) = \exp(\eta x) F^\eta(t, x) = \exp(\eta x) \int_{\mathbb{R}} \exp(-iux) \phi^\eta(t, u) du$$

The choice of  $\eta$  is important, as it will determine how accurate numerical implementations will be. A small value for  $\eta$  will eliminate the singularity theoretically, but it might not reduce its impact around zero sufficiently for numerical purposes. If  $\eta$  is too large, then the characteristic function can be pushed towards zero, which will not allow us to accurately reconstruct its shape and integrate it with precision. Typically, a value of  $\eta$  in the region of 1 to 5 gives satisfactory results. In figure 4.1 one can see the impact of the damping parameter  $\eta$  on the function to be integrated. As we move from  $\eta \approx 0$  to  $\eta = 1$  the function becomes progressively better behaved, and therefore easier to numerically integrate. But if we keep damping we run the risk of pushing the whole integrand close to zero, as it is illustrated when we set  $\eta = 10$ .

## 4.2 OPTION PRICING USING TRANSFORMS

It is also possible to recover European option prices from the characteristic function of the log-return. We assume that under the risk neutral measure, the logarithm of the price satisfies

FIGURE 4.1: Damping the Fourier transform to avoid the singularity at the origin. The integrand for the normal inverse Gaussian distribution with parameters  $\{\mu, \alpha, \beta, \delta\} = \{8\%, 7.00, -3.50, 0.25\}$  is presented for different values of the damping parameter  $\eta$ . The real (imaginary) part is given in blue (green). The dashed thick line gives the integrand for  $\eta = 0.01 \approx 0$ , which diverges at zero. The solid thick line presents the integrand for  $\eta = 1$ , while the solid thin line assumes  $\eta = 10$ . Two different horizons of one day and one month are presented, to illustrate the change in the tail behavior as the maturity is decreased.



$$\log S(T) = \log S(0) + X(T) \quad (4.2)$$

We specify that this relationship holds under the risk neutral measure because it is most likely that the market we are working on is incomplete. If this is the case, then we are not able to specify the no-arbitrage prices of options solely on the information embedded in (4.2) that is specified under  $\mathcal{P}$ . We need a change of measure technique,<sup>2</sup> as well as a number of preferences parameters that will allow us to determine the equivalent measure  $\mathcal{Q}$ . in order to sidestep these issues we can assume that the process in (4.2) is defined under  $\mathcal{Q}$ . The only constraints that must be imposed is that the expectation

$$B(T) \cdot E S(T) = S(0) \implies E \exp X(T) = \frac{1}{B(T)}$$

If we assume that the characteristic function  $\phi(T, u)$  of  $\log S(T)$  is given to us, then the above constraint can be expressed as a constraint on the characteristic function, that is to say

$$\phi(T, -i) = \frac{1}{B(T)}$$

There have been two methods that compute European calls and puts through the characteristic function. Following the seminal work of Bakshi and Madan

<sup>2</sup> For example Girsanov's theorem (Øksendal, 2003), or the Esscher transform (Gerber and Shiu, 1994), can be used to define equivalent martingale measures.

(2000) the call/put option price is expressed in a form that resembles the Black-Scholes formula, for example

$$P_{\text{call}} = S(0)\Pi_1 - KB(T)\Pi_2$$

where the quantities  $\Pi_1$  and  $\Pi_2$  depend on the particular characteristic function. This approach offers the same intuition as the Black-Scholes formula, where  $\Pi_1$  is the option delta, and  $\Pi_2$  is the risk neutral probability of exercise.

Although the above expression is elegant and intuitive, it does not lend itself for numerical implementations. More recently, Carr and Madan (1999) develop the Fourier transform of the (modified) European option price directly, by expressing (with  $k = \log K$  the log-strike)

$$P_{\text{call}} = \exp(-\eta k) \cdot \mathcal{F}^{-1}[\psi(T, u; \eta)](k)$$

where  $\psi(t, u)$  is a function of the characteristic function  $\phi(t, u)$ , and  $\eta$  is an auxiliary *dampening* parameter.

### THE DELTA-PROBABILITY DECOMPOSITION

The Delta-Probability decomposition of the European price has its roots in the work of Heston (1993) on stochastic volatility, although in Heston's original paper the decomposition is not proved in its generality. Bakshi and Madan (2000) provide a general approach where derivative payoffs are spanned using trigonometric functions. Here we will provide a heuristic proof for the special case of a European call option (see also Heston and Nandi, 2000, for details)

Assuming that the probability density function under the risk neutral measure  $\mathbb{Q}$  for the time  $t$  log-price is  $f(t, x)$ , we can write the European call option price as the expected value of its payoffs, as

$$\begin{aligned} P_{\text{call}} &= B(T) \cdot \mathbb{E} \max(S(T) - K, 0) \\ &= B(T) \int_{\log K}^{\infty} (\exp(x) - K)f(T, x)dx \\ &= B(T) \int_{\log K}^{\infty} \exp(x)f(t, x)dx - B(T)K \int_{\log K}^{\infty} f(T, x)dx \end{aligned}$$

The second integral is just the probability  $P[\log S(T) \geq \log K]$ , and since  $\phi(t, u)$  is the characteristic function of  $\log S(T)$  this will be equal to

$$\Pi_2 = \int_{\log K}^{\infty} f(T, x)dx = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \left[ \frac{\exp(-iu)}{iu} \phi(T, u) \right] du$$

To compute the second integral we use the trick of multiplying and dividing the expression as follows

$$\int_{\log K}^{\infty} \exp(x)f(T, x)dx = \frac{\int_{\log K}^{\infty} \exp(x)f(T, x)dx}{\int_{-\infty}^{\infty} \exp(x)f(T, x)dx} \cdot \int_{-\infty}^{\infty} \exp(x)f(T, x)dx \quad (4.3)$$

Note that the quantity  $\int_{-\infty}^{\infty} \exp(x) f(T, x) dx = \frac{S(0)}{B(T)}$  due to the risk-neutrality restriction (4.1). Also, the fraction in the above expression is by construction between zero and one, therefore it can be interpreted as some probability. In particular, if we define

$$f^*(T, x) = \frac{\exp(x) f(T, x)}{\int_{-\infty}^{\infty} \exp(x) f(T, x) dx}$$

then the fraction in (4.3) can be expressed as  $\int_{\log K}^{\infty} f^*(T, x) dx$ . The Fourier transform of  $f^*(T, x)$  is given by

$$\phi^*(T, u) = \int_{\mathbb{R}} \exp(iux) f^*(T, x) dx = \frac{\phi(T, u - i)}{\phi(T, -i)}$$

We can now define the quantity

$$\Pi_1 = \int_{\log K}^{\infty} f^*(T, x) dx = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \left[ \frac{\exp(-iux) \phi(T, u - i)}{iu \phi(T, -i)} \right] du$$

Putting everything together will yield the European call option price, which has the same structure as the Black-Scholes formula, where instead of the cumulative normal values we have  $\Pi_1$  and  $\Pi_2$ . To summarize

$$P_{\text{call}} = S(0)\Pi_1 - KB(T)\Pi_2$$

where

$$\Pi_1 = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \left[ \frac{\exp(-iux) \phi(T, u - i)}{iu \phi(T, -i)} \right] du$$

$$\Pi_2 = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \left[ \frac{\exp(-iux) \phi(T, u)}{iu} \right] du$$

### THE FOURIER TRANSFORM OF THE MODIFIED CALL

The Delta-probability decomposition gives an intuitive expression for the value of the European call, but is not very efficient operationally since the integrals required are not defined at the point  $u = 0$ . More recently, Carr and Madan (1999) developed the characteristic function of a modified price itself. In particular, if we introduce a parameter  $\eta$ , we can define the modified call price as<sup>3</sup>

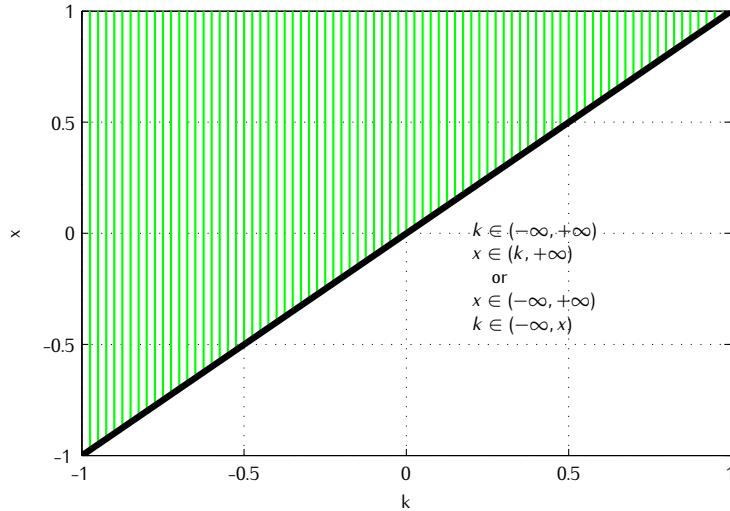
$$P_{\text{call}}^{\eta}(k) = \exp(-\eta k) P_{\text{call}}(k)$$

which we consider a function of the log-strike price  $k = \log K$ . We assume that the maturity of the option is fixed at  $T$ . The Fourier transform of the modified call, say  $\psi^{\eta}(t, u) = \mathcal{F}[P_{\text{call}}^{\eta}](u)$ , is given by

---

<sup>3</sup> We need the parameter  $\eta$  to modify the original call price, since the original call price is not square integrable.

FIGURE 4.2: Finite difference approximation schemes. The forward (green), backward (blue) and central (red) differences approximation schemes, together with the true derivative (dashed).



$$\begin{aligned}\psi^\eta(T, u) &= B(T) \int_{\mathbb{R}} \exp(iuk) P_{\text{call}}^\eta(k) dk \\ &= \int_{\mathbb{R}} \exp(iuk) \left[ \exp(\eta k) \int_k^\infty (\exp(x) - \exp(k)) f(T, k) dx \right] dk\end{aligned}$$

We will change the order of integration, and therefore the integration limits will change from  $(k, x) \in (-\infty, +\infty) \times (k, +\infty)$  to  $(x, k) \in (-\infty, +\infty) \times (-\infty, x)$ , as shown in figure 4.2. Then

$$\begin{aligned}\psi^\eta(T, u) &= B(T) \int_{\mathbb{R}} \int_{-\infty}^x \exp(iuk + \eta k + x) f(T, x) dk dx \\ &\quad - B(T) \int_{\mathbb{R}} \int_{-\infty}^x \exp(iuk + \eta k + k) f(T, x) dk dx\end{aligned}$$

Now since  $\eta \neq 0$  both inner integrals will vanish at  $k \rightarrow -\infty$  (precisely the reason we introduced this parameter), and the Fourier transform becomes

$$\begin{aligned}\psi^\eta(T, u) &= B(T) \int_{\mathbb{R}} \left[ \frac{1}{iu + \eta} - \frac{1}{iu + \eta + 1} \right] \phi(T, u - i(\eta + 1)) dx \\ &= \frac{B(T)}{(iu + \eta)(iu + \eta + 1)} \phi(T, u - i(\eta + 1))\end{aligned}$$

This is a closed form expression of the Fourier transform of the modified call price, in terms of the characteristic function of the log-price. Therefore, to retrieve

for copies, comments, help etc. visit <http://www.theponytail.net/>

LISTING 4.1: phi\_normal.m: Characteristic function of the normal distribution.

```
% phi_normal.m
function y = phi_normal(u, p)
t      = p.t;
r      = p.r;
5 sigma = p.sigma;
a      = r - 0.5*sigma^2;
y = exp(i*t*a*u - .5*t*sigma^2*u.^2);
```

the original call price we just need to apply the inverse Fourier transform on  $\psi^\eta(T, u)$

$$P_{\text{call}}(k) = \mathcal{F}^{-1}[\psi](k) = \frac{\exp(-\eta k)}{2\pi} \int_{\mathbb{R}} \exp(-iuk) \psi^\eta(T, u) du$$

Option prices are of course real numbers, and that implies that the Fourier transform  $\psi^\eta(T, u)$  must have odd imaginary and even real parts. Therefore we can simplify the pricing formula to

$$P_{\text{call}}(k) = \frac{\exp(-\eta k)}{\pi} \int_0^\infty \text{Re}[\exp(-iuk) \psi^\eta(T, u)] du \quad (4.4)$$

The choice of the parameter  $\eta$  determines how fast the integrand approaches zero. Admissible values for  $\eta$  are the ones for which  $|\psi^\eta(T, 0)| < \infty$ , which in turn implies that  $|\mathbb{E}[S(T)]^{\eta+1}| < \infty$ , or equivalently that the  $(\eta + 1)$ -th moment exists and is finite. For more information for the choice of  $\eta$  see Carr and Madan (1999) and Lee (2004b).

## 4.3 AN EXAMPLE IN MATLAB

In the following subsections we will give examples of transform methods that are based on the normal and the normal inverse Gaussian (NIG) distribution (see for example Barndorff-Nielsen, 1998, for details).

### THE CHARACTERISTIC FUNCTIONS

We want to set up a model under the risk neutral measure  $\mathcal{Q}$  of the form

$$S(T) = S(0) \exp\{aT + X(T)\}$$

where  $X(T)$  is a random variable with a given characteristic function. If we assume that the interest rate is constant, then under  $\mathcal{Q}$

$$\mathbb{E}[\exp\{aT + X(T)\} | \mathcal{F}(0)] = \exp\{rT\} \Rightarrow a = r - \frac{1}{T} \log \mathbb{E}[\exp\{X(T)\} | \mathcal{F}(0)]$$

for copies, comments, help etc. visit <http://www.theponytail.net/>

LISTING 4.2: `phi_nig.m`: Characteristic function of the normal inverse Gaussian distribution.

```
% phi_nig.m
function y = phi_nig(u, p)
t      = p.t;
r      = p.r;
5    delta = p.delta;
alpha = p.alpha;
beta  = p.beta;
a = r - delta*sqrt(alpha^2-beta^2) + ...
     delta*sqrt(alpha^2-(beta+1)^2);
10   y = exp(i*t*a*u + t*delta*sqrt(alpha^2-beta^2) - ...
              t*delta*sqrt(alpha^2-(beta+i*u).^2));
```

The expectation can be cast in terms of the characteristic function of  $X(T)$ , giving the constraint

$$a = r - \frac{1}{T} \log \phi(T, -i)$$

This constraint will ensure that the under risk-neutrality the asset will grow at the same rate as the risk free asset.

The characteristic function for the normal distribution, implemented in listing 4.1, is given by

$$\phi(t, u) = \exp \left\{ itau - \frac{1}{2} t \sigma^2 u^2 \right\}$$

for  $a = r - \frac{1}{2} \sigma^2$ . The characteristic function of the NIG distribution is given in listing 4.2

$$\phi(t, u) = \exp \left\{ itau + t \delta \sqrt{\alpha^2 - \beta^2} - t \delta \sqrt{\alpha^2 - (\beta + iu)^2} \right\}$$

In this case the parameter  $a = r - \delta \sqrt{\alpha^2 - \beta^2} + \delta \sqrt{\alpha^2 - (\beta + 1)^2}$ .

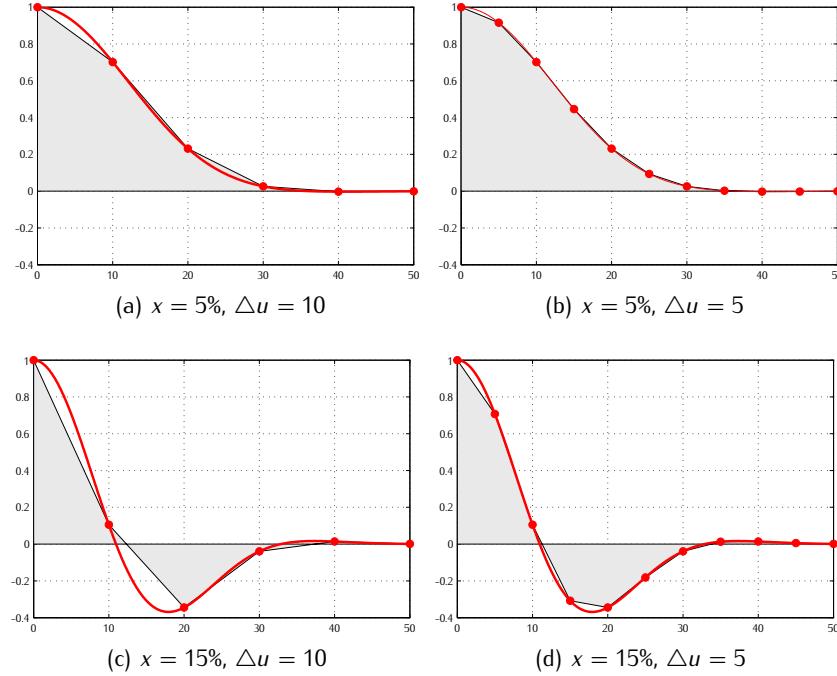
## NUMERICAL FOURIER INVERSION

Say that we are interested in inverting the characteristic function to produce the probability density function or to compute option prices. To do so we need the value, at the point  $x$ , of an integral of the form

$$f(t, x) = \int_0^\infty \exp(-ixu) h(u) du$$

The integral will be approximated with a quadrature, and here we will use the trapezoidal rule.

FIGURE 4.3: Numerical Fourier inversion using quadrature. The integral  $\int_0^\infty \text{Re}[\exp(-iux)\phi(T, u)]du$  is approximated where  $\phi(T, u)$  is the characteristic function of the normal distribution, with  $\mu = 8\%$ ,  $\sigma = 25\%$  and  $T = 30/365$ . The upper integration bound is  $\bar{u} = 50$ . Results for  $x = 5\%$  and  $x = 15\%$ , as well as  $\Delta u = 10$  and  $\Delta u = 5$  are given.



In particular, we start with a truncating the interval  $[0, \infty)$ , over which the characteristic function is integrated. We select a point  $\bar{u}$  that is large enough for the contribution of the integral after this point to be negligible. Then we discretize the interval  $[0, \bar{u}]$  into  $N - 1$  subintervals with spacing  $\Delta u$ , that is we produce the points  $u = \{u_j = j\Delta u : j = 0, \dots, N\}$ . For a given maturity  $T$  we denote the integrand with  $h^*(x, u) = \exp(-iux)h(u)$ , and produce the values at the grid points  $h_j(x) = h^*(x, u_j)$ . Then, the trapezoidal approximation to the integral is given by

$$\int_0^\infty h(x, u)du \approx \sum_{j=0}^N h_j(x)\Delta u - \frac{1}{2}(h_0(x) + h_N(x))\Delta u$$

Therefore, in order to carry out the numerical integration, one has to make two *ad hoc* choices, namely the upper integration bound  $\bar{u}$  and the grid spacing  $\Delta u$ . Selecting  $\bar{u}$  can be guided by the speed of decay of the characteristic

LISTING 4.3: cf\_int.m: Trapezoidal integration of a characteristic function.

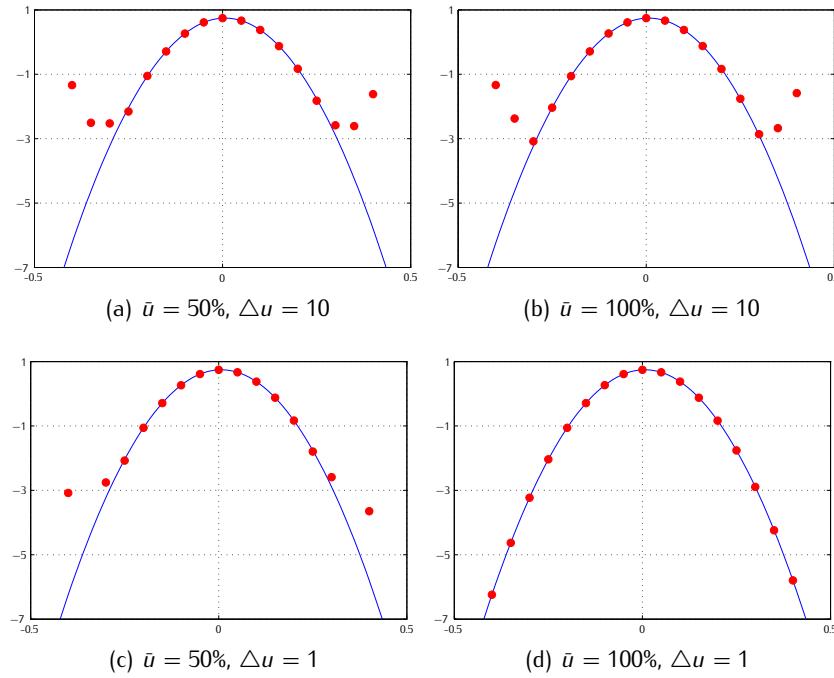
```
% cf_int.m
p.t      = 30/365;
p.r      = 0.08;
p.sigma  = 0.25;
5 D = 1;                      % grid size over the char func
theta = [0:D:100]';           % char function discretization
x = [-0.40:0.05:0.40]';       % output grid for density
Nx = length(x);
pdf = zeros(Nx,1);            % keeps density values pdf(x)
10 for k=1:Nx                 % loop over x-values
    % the integrand values
    y = exp(-i*x(k)*theta).*phi_normal(theta, p);
    y = real(y);
    I = sum(y)-0.5*(y(1)+y(end)); % trapezoidal rule
    pdf(k) = I*D/pi;             % the output
15 end
```

function. A good choice of  $\Delta u$  on the other hand can be a little bit trickier, since the quantity  $\exp\{iu_x\} = \cos(u_x) + i\sin(u_x)$  will be oscillatory with frequency that increases with  $x$ . Figure 4.3 gives the quadrature approximations for the case of the normal distribution. The characteristic function corresponds to a density with mean  $\mu = 8\%$  p.a. and volatility  $\sigma = 20\%$  p.a., and the maturity is one month,  $T = 30/365$ . We have set the upper quadrature bound to  $\bar{u} = 50$ , and use two different grid sizes, a “coarse” one  $\Delta u = 10$  and a “fine” one  $\Delta u = 5$ . We investigate the integration for  $x = 5\%$  and for  $x = 15\%$ . In the first case the integrand is not oscillatory and even the “coarse” approximation captures the integration fairly accurately. When  $x = 15\%$  the integrand oscillates and a “finer” grid is required. This example illustrates that one must be careful and cautious when setting up numerical integration procedures that automatically select the values for  $\bar{u}$  and  $\Delta u$ .

In order to reconstruct the probability density function we need to repeat the procedure outlined above for different values of  $x$ . This is carried out in listing 4.3 for the normal distribution. The results are plotted in figure 4.4 in logarithmic scale for different values of  $\bar{u}$  and  $\Delta u$ . One can verify that if we are interested in the central part of the distribution, then a coarse grid with is sufficient while the results are not particular sensitive to the choice of the upper integration bound. On the other hand, if we want to compute the probability density at the tails, then we need to implement a very fine grid over a large support interval.

There is a distinct and very important relationship between the fat tails and the decay of the characteristic function. In particular, the higher the kurtosis of the distribution, the slower the characteristic function decays towards zero as  $u$  increases. This has some implications on the implementation of the nu-

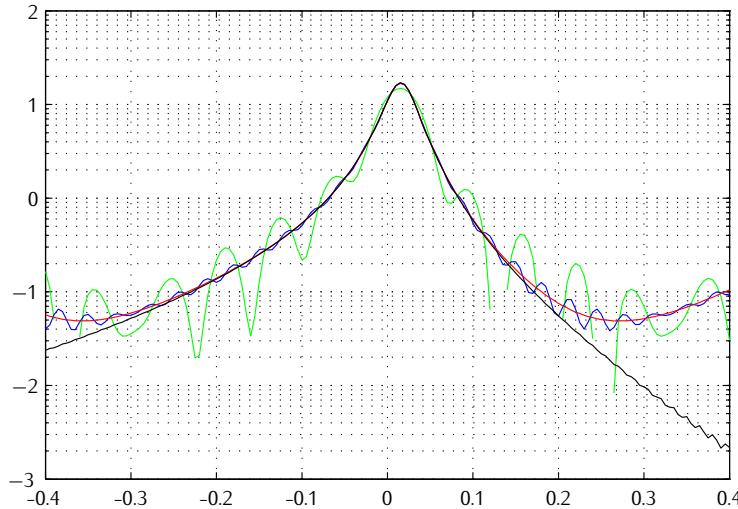
FIGURE 4.4: Probability density function using Fourier inversion. Logarithmic plots of the integral  $f(T, x) = \frac{1}{\pi} \int_0^\infty \text{Re}[\exp(-iux)\phi(T, u)]du$ , where  $\phi$  is the normal characteristic function, together with the true normal density. The parameters set is  $\{\mu, \sigma, T\} = \{8\%, 25\%, 30/365\}$ . Results for  $\bar{u} = 50$  and  $\bar{u} = 100$ , as well as  $\Delta u = 10$  and  $\Delta u = 1$  are given.



merical integration, as we must be ready to integrate over a very long support. On the other hand, the oscillations introduced by the complex exponential do not depend on the characteristic function itself, therefore we will need a fine grid to accurately compute the density around the tails. This indicates that we will potentially have to carry out numerical quadratures with many hundreds of thousands of grid points.

This is illustrated with the NIG distribution. We consider a parameter set  $\{\mu, T, \alpha, \beta, \delta\} = \{8\%, 30/365, 7.00, -3.50, 0.25\}$ . Taking derivatives of the characteristic function gives the volatility, skewness and excess kurtosis, which are 23.4%, -1.21 and 3.96 respectively. Therefore the distribution exhibits negative skewness and excess kurtosis of a magnitude that is observed in asset returns. We want to investigate the stability of the Fourier inversion, and figure 4.5 gives the results for different combinations of  $\Delta u$  and  $\bar{u}$ . Here we can clearly see the effect of different choices of these parameters. As we increase the integration interval the oscillations in the probability function are reduced, but the function

FIGURE 4.5: Probability density function using Fourier inversion. Logarithmic plots of the integral  $f(T, x) = \frac{1}{\pi} \int_0^\infty \text{Re}[\exp(-iux)\phi(T, u)]du$ , where  $\phi$  is the normal inverse Gaussian characteristic function. The parameters set is  $\{\mu, \alpha, \beta, \delta, T\} = \{8\%, 7.00, -3.50, 0.25, 30/365\}$ . Results for  $\Delta u = 10$  and  $\bar{u} = 100$  (green),  $\bar{u} = 200$  (blue) and  $\bar{u} = 400$  (red). The density for  $\Delta u = 1$  and  $\bar{u} = 400$  (black) is also given.



can still be inaccurate around the tails. We need to reduce the grid size to increase the overall accuracy. Observe that the right tail is slightly oscillatory even when  $\{\Delta u, \bar{u}\} = \{1, 400\}$ .

## 4.4 APPLYING FAST FOURIER TRANSFORM METHODS

In the previous section we implemented a numerical integration method that approximates an integral of the form

$$y(x) = \int_0^\infty \exp(-iux)h(u)du$$

This integral can then be used to retrieve the probability density function at the point  $x$ , or a European call option price with log-strike price  $x$ . Typically, we want to compute the integral for many different values of the parameter  $x$ , in order to reconstruct the probability density function or the implied volatility smile. Using the approach we outline above, we must perform as many numerical integrations as the number of abscissas over  $x$ .

The Fast Fourier Transform (FFT) is a numerical procedure that simultaneously computes  $N$  sums of the form

$$z_k^* = \sum_{j=1}^N \exp\left(-\frac{2\pi}{N}(j-1)(k-1)\right) \cdot z_j \quad (4.5)$$

for all  $k = 1, \dots, N$ . The number of operations needed for the FFT is of order  $\Theta(N \log N)$ . For comparison, if we wanted to compute the above sums separately and independently it would take  $\Theta(N^2)$  operations, meaning that in order to double the number of points the number of operations will increase fourfold. With the FFT the computational burden increases a bit more than two times. This substantial speedup is the reason that has made FFT popular in computational finance, since we typically need to evaluate thousands of Fourier inversions when calibrating models to observed volatility surfaces.

The input of the FFT is a vector  $z \in \mathbb{C}^N$ , and the output is a new vector  $z^* \in \mathbb{C}^N$ . Each element of  $z^*$  will keep the sum for the corresponding value of  $k$ . Our task is therefore to cast the integral approximation in a form that can be computed using FFT.

The first step is of course to discretize the interval  $[0, \bar{u}]$  using  $N$  equidistant points, and say that we set  $u = \{(j-1)\Delta u : j = 1, \dots, N\}$ . Therefore the trapezoidal approximation to the integral is given by

$$\begin{aligned} & \frac{1}{2} \exp(-iu_1x)h(u_1)\Delta u + \exp(-iu_2x)h(u_2)\Delta u + \exp(-iu_3x)h(u_3)\Delta u + \dots \\ & + \exp(-iu_{N-1}x)h(u_{N-1})\Delta u + \frac{1}{2} \exp(-iu_Nx)h(u_N)\Delta u \end{aligned}$$

Thus, if we set  $\alpha = (\frac{1}{2}, 1, 1, \dots, 1, \frac{1}{2})'$  and  $h_j = h(u_j)$ , we can write the approximation as the sum

$$\sum_{j=1}^N \exp(-iu_jx)\alpha_j h_j \Delta u$$

Since the FFT will also return an  $(N \times 1)$  vector, we should set the procedure to produce values for a set  $x = \{x_1 + (k-1)\Delta x : k = 1, \dots, N\}$ . We typically want these values to be symmetric around zero,<sup>4</sup> and therefore we can set  $x_1 = -\frac{N}{2}\Delta x$ . The approximating sum for these values of  $x$  will therefore be given by

---

<sup>4</sup> When we invert to construct a probability density we typically interested in the density at log-returns symmetric around the peak which will be close to zero. If we invert for option pricing purposes, we can normalize the current price to one. Then each value of  $y$  will correspond to a log-strike price, and we typically want to retrieve option prices which are in-, at- and in-the-money. The at-the-money level will be around the current log-price which is of course zero.

$$\begin{aligned}
y_k &= \sum_{j=1}^N \exp(-iu_j x_k) \alpha_j h_j \Delta u \\
&= \sum_{j=1}^N \exp(-i(j-1)(k-1)\Delta u \Delta x) \exp(-i(j-1)x_1 \Delta u) \alpha_j h_j \Delta u \\
&= \sum_{j=1}^N \exp(-i(j-1)(k-1)\Delta u \Delta x) z_j
\end{aligned}$$

where  $z_j = \exp(-ix_1 u_j) \alpha_j h_j \Delta u$ . The sum above will be of the FFT form (4.5) only if  $\Delta u \Delta x = \frac{2\pi}{N}$ . This set a constraint on the relationship between the characteristic function input grid size, and the output log-return or log-strike grid size. This completes the integral approximation; it is now straightforward to invert for the probability density function or for call prices.

### FFT INVERSION FOR THE PROBABILITY DENSITY FUNCTION

To summarize, in order to invert the characteristic function we need to take the following steps (with  $\odot$  we denote element-by-element vector multiplication):

1. Input the grid sizes  $\Delta u$  and  $\Delta x$ , as well as the number of integration points  $N$ . Make sure that they satisfy  $\Delta u \Delta x = \frac{2\pi}{N}$ .
2. Construct the vectors  $u = \{(j-1)\Delta u : j = 1, \dots, N\}$  and  $x = \{-\frac{N}{2}\Delta x + (k-1)\Delta x : k = 1, \dots, N\}$ .
3. Compute the vector  $z = \exp(-ix_1 u) \odot \phi(T, u)$ .
4. For the trapezoidal rule set  $z_1 = \frac{z_1}{2}$  and  $z_N = \frac{z_N}{2}$ .
5. Run the FFT on  $z$ , that is  $z^* = \text{FFT}(z)$ .
6. Compute the density function values  $y = \frac{1}{\pi} \text{Re}[z^*]$ .
7. Output the pair  $(x, y)$ : the value  $y_k$  is the probability density for the log-return  $x_k$ , for  $k = 1, \dots, N$ .

### FFT INVERSION FOR EUROPEAN CALL OPTION PRICING

The inversion of the characteristic function to compute options is very similar. We just need to also compute the Fourier transform of the modified call before invoking the FFT. The steps are the following (with  $\odot$  we denote element-by-element vector multiplication and with  $\oslash$  element-by-element division):

1. Input the grid sizes  $\Delta u$  and  $\Delta x$ , as well as the number of integration points  $N$ . Make sure that they satisfy  $\Delta u \Delta x = \frac{2\pi}{N}$ . Also input the “dampening parameter” for the modified call  $\eta$ .
2. Construct the vectors  $u = \{(j-1)\Delta u : j = 1, \dots, N\}$  and  $x = \{-\frac{N}{2}\Delta x + (k-1)\Delta x : k = 1, \dots, N\}$ .
3. Construct the Fourier transform of the modified call

$$\psi = \exp(-rT) \phi(T, u - i(\eta + 1)) \oslash [(iu + \eta) \odot (iu + \eta + 1)]$$

LISTING 4.4: fft\_call.m: Call pricing using the FFT.

```
% fft_call.m
function [k, y] = fft_call(cf, pcf, pf)
% values for FFT implementation
h = pf.eta; N = pf.N; uBar = pf.uBar;
5 % parameter values
r = pcf.r; t = pcf.t;
kBar = -N*Dk/2;
Du = uBar/(N-1); Dk = 2*pi/N/Du;% grid sizes
u = [0:N-1]*Du; k = [0:N-1]*Dk + kBar;
10 % compute the psi-function
z = feval(cf, u-i*(h+1), pcf);
z = exp(-r*t)*z./(h^2+h-u.^2+i*(2*h+1)*u);
z = exp(-i*u*kBar).*z.*Du; % integrand
% trapezoidal rule
z(1) = 0.5*z(1); z(end) = 0.5*z(end);
w = real(fft(x)); % FFT
y = exp(-h*k).*w/pi; % output
```

4. Compute the vector  $z = \exp(-ix_1u) \odot \psi$ .
5. For the trapezoidal rule set  $z_1 = \frac{z_1}{2}$  and  $z_N = \frac{z_N}{2}$ .
6. Run the FFT on  $z$ , that is  $z^* = \text{FFT}(z)$ .
7. Compute option values  $y = \frac{1}{\pi} \exp(-\eta x) \odot \text{Re}[z^*]$ .
8. Output the pair  $(x, y)$ : the value  $y_k$  is call option that corresponds to an option with log-strike price  $x_k$ , for  $k = 1, \dots, N$ .

The inversion of the Fourier transform for the modified call is implemented in listing 4.4. One needs to specify the corresponding characteristic function, for example  $[k, y] = \text{fft\_call}(@phi_nig, pcf, pf)$  in order to retrieve a set of log-strike prices and the corresponding call prices. The parameters for the characteristic function are passed through the structure pcf, while parameters for the FFT inversion are passed through pf.

## 4.5 THE FRACTIONAL FFT

The restriction  $\Delta u \Delta x = \frac{2\pi}{N}$  that has to be satisfied when applying the FFT hampers the flexibility of the method. One naturally wants a fine grid when integrating over the characteristic function, but a small  $\Delta u$  can result in very coarse output grids. For example, a 512 point integration over the interval  $[0, 100]$  would offer a good approximation to invert for the normal distribution of figure 4.4. This implies  $\Delta u = 0.1957$ , and in order for the FFT to be applied we have to set  $\Delta x = \frac{2\pi}{N\Delta u} = 0.0627$ , with  $x_1 = -16.05\%$ . This means that only a very

LISTING 4.5: frft.m: Fractional FFT.

```
% frft.m
function f = frft(x, a)
N = size(x, 2);
e1 = exp(-pi*i*a * (0:(N-1)).^2);
5 e2 = exp(pi*i*a *(N:-1:1).^2);
z1 = [x.*e1, zeros(1,N)];
z2 = [1./e1, e2];
fz1 = fft(z1);
fz2 = fft(z2);
10 fz = fz1 .* fz2;
ifz = ifft(fz);
f = e1.*ifz(1,1:N);
```

small number of the 512 output values are actually within the  $\pm 30\%$  which we might be interested in.

One way that can result in smaller output grids is increasing the FFT size  $N$ . We have chosen the upper integration bound in a way that the characteristic function is virtually zero outside the interval. Therefore, when we increase  $N$  we just “pad with zeros” the input vector  $z$ . For example, if we append the 512 vector with 7680 zeros we will implement a 8192-point FFT, which will return a more acceptable output grid of 0.0039. But of course applying an FFT which is 16 times longer will have a serious impact on the speed of the method.

The fractional FFT method, outlined in Chourdakis (2005), addresses this issue. The fractional FFT (FRFT) with parameter  $\alpha$  will compute the more general expression

$$z_k^* = \sum_{j=1}^N \exp(-2\pi\alpha(j-1)(k-1)) \cdot z_j \quad (4.6)$$

for all  $k = 1, \dots, N$ . In order to implement a  $N$ -point FRFT one needs to invoke three times a standard  $2N$ -point FFT,<sup>5</sup> but the freedom of selecting  $\Delta u$  and  $\Delta x$  independently can actually improve the speed for a given degree of accuracy.

In particular, the following steps implement an  $N$ -point FRFT, coded in listing 4.5

1. Create the  $(N \times 1)$  vectors  $\boldsymbol{\varepsilon}$  and  $\tilde{\boldsymbol{\varepsilon}}$

$$\begin{aligned}\boldsymbol{\varepsilon}_1 &= \{\exp(-i\pi\alpha(j-1)^2) : j = 1, \dots, N\} \\ \boldsymbol{\varepsilon}_2 &= \{\exp(i\pi\alpha(N-j+1)^2) : j = 1, \dots, N\}\end{aligned}$$

2. Based on these auxiliary vectors create the two  $(2N \times 1)$  vectors  $z_1$  and  $z_2$

---

<sup>5</sup> For proofs and discussion on the FRFT also see Bailey and Swarztrauber (1991, 1994).

LISTING 4.6: frft\_call.m: Call pricing using the FRFT.

```
% frft_call.m
function [k, y] = frft_call(cf, pcf, pf)
% values for FRFT implementation
h = pf.eta; N = pf.N; uBar = pf.uBar; kBar = pf.kBar;
% parameter values
r = pcf.r; t = pcf.t;
kBar = -N*Dk/2;
Du = uBar/(N-1); Dk = 2*kBar/N; % grid sizes
alpha = Du*Dk/2/pi; % fractional parameter
u = [0:N-1]*Du; k = [0:N-1]*Dk + kBar;
% compute the psi-function
z = feval(cf, u-i*(h+1), pcf);
z = exp(-r*t)*z./(h^2+h-u.^2+i*(2*h+1)*u);
z = exp(-i*u*kBar).*z.*Du; % integrand
% trapezoidal rule
z(1) = 0.5*z(1); z(end) = 0.5*z(end);
w = real(frft(x, alpha)); % FRFT
y = exp(-h*k).*w/pi; % output
```

$$z_1 = \begin{pmatrix} z \odot \epsilon_1 \\ 0 \end{pmatrix} \text{ and } z_2 = \begin{pmatrix} 1 \odot \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

3. Apply the FFTs on these vectors

$$z_1^* = \text{FFT}(z_1) \text{ and } z_2^* = \text{FFT}(z_2)$$

4. The  $N$ -point FRFT will be the first  $N$  elements of the inverse FFT

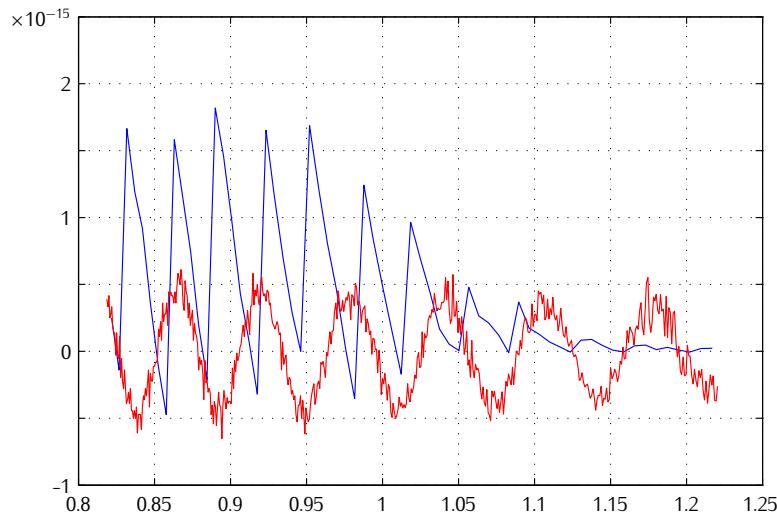
$$z^* = \epsilon_1 \odot \text{IFFT}(z_1^* \odot z_2^*)$$

We can now easily adapt the recipes of the previous section to accommodate the fractional FFT. We can now choose the two grid sizes freely, and set the fractional parameter  $\alpha = \frac{\Delta u \Delta x}{2\pi}$ . Thus, we need to change the corresponding steps of the recipes to:

Run the fractional FFT on  $z$  with fractional parameter  $\frac{\Delta u \Delta x}{2\pi}$ , that is  $z^* = \text{FRFT}(z, \frac{\Delta u \Delta x}{2\pi})$ .

Listing 4.6 implement the fractional FFT based option pricing. Chourdakis (2005) gives details on the accuracy of this method for option pricing based on a number of experiments that compares the fractional to the standard FFT. Figure 4.6 gives an example that is based on the normal distribution. One can observe the exceptional accuracy of both methods: a 8192-point FFT is contrasted to a 512-point FRFT.

FIGURE 4.6: Comparison of the FFT and the fractional FFT, based on the Black-Scholes model. The figure shows the errors between option prices computed using the transform methods and their closed form values, for different strike prices. The blue line gives the errors of the standard FFT method, while the red line gives the errors of the fractional FFT. All values are  $\times 10^{-15}$ .



## 4.6 ADAPTIVE FFT METHODS AND OTHER TRICKS

There is a number of ways in which adaptive integration can be employed within the fractional FFT framework. In particular, as the fractional FFT allows us to integrate over an arbitrary region, it is natural to consider splitting the support of the characteristic function into subintervals and apply the transform sequentially.

Also, we might consider improving the accuracy of each integration segment. In the previous sections we worked with the trapezoidal rule, essentially approximating the function

$$\int_{u_1}^{u_N} \exp(-iu_x) h(u) du \approx \sum_{j=1}^N \exp(-iu_j x) \alpha_j h_j \Delta u$$

for  $h_j = h(u_j)$ , and  $\alpha = (\frac{1}{2}, 1, 1, \dots, 1, \frac{1}{2})'$ . What we have done is approximating the whole integrand as a piecewise linear function. This integrand is the product of two terms: the first,  $\exp(-iu_x)$ , is a combination of trigonometric functions,

and will be highly oscillatory, especially for large values of  $|x|$ ; the second,  $h(u)$ , is also oscillatory but typically very mildly and also independent of  $x$ .

It therefore makes sense to approximate only the second component as a piecewise linear function, and leave the first part intact. We therefore split the integral into  $N - 1$  sub-integrals

$$\int_{u_1}^{u_N} \exp(-iux) h(u) du = \sum_{j=1}^{N-1} \int_{u_j}^{u_{j+1}} \exp(-iux) h(u) du$$

We then use the linear approximation within each subinterval

$$h(u) \approx h_j + (u - u_j) \frac{h_{j+1} - h_j}{u_{j+1} - u_j} = a_j + b_j u \frac{\Delta h_j}{\Delta u_j} \text{ for } u_j \leq u \leq u_{j+1}$$

Thus, each subintegral can be computed as

$$\begin{aligned} \int_{u_j}^{u_{j+1}} \exp(-iux) [a_j + b_j u] du &= \exp(-iux) \left[ \frac{a_j + b_j u}{-ix} - \frac{b_j}{(-ix)^2} \right] \Big|_{u=u_j}^{u_{j+1}} \\ &= \frac{i}{x} [h_{j+1} \exp(-iu_{j+1}x) - h_j \exp(-iu_jx)] \\ &\quad + \frac{b_j}{x^2} [\exp(-iu_{j+1}x) - \exp(-iu_jx)] \end{aligned}$$

Luckily, the first square brackets will cancel out sequentially, as we sum over the sub-integrals, which will give us the result after some straightforward algebra

$$\begin{aligned} \int_{u_1}^{u_N} \exp(-iux) h(u) du &\approx \frac{i}{x} [h_N \exp(-iu_Nx) - h_1 \exp(-iu_1x)] + \frac{1}{x^2} \sum_{j=1}^N c_j \exp(-iu_jx) \end{aligned}$$

with  $c = (0 - b_1, b_1 - b_2, b_2 - b_3, \dots, b_{N-2} - b_{N-1}, b_{N-1} - 0)$ . Recall that  $b_j = \frac{h_{j+1} - h_j}{\Delta u}$ . The sum in the above expression can now be computed using the fractional FFT procedure.

It might seem initially that the above expression will diverge as  $x \rightarrow 0$ . This is not the case. In fact, by twice applying l'Hôpital's rule as needed, one can show that the expression converges to the trapezoidal rule we obtained in the previous section.

Listing 4.7 shows an implementation of this integration using the fractional FFT. This method is directly implemented in listing 4.8, which shows how an adaptive integration technique can be used to invert the characteristic function and recover the cumulative density.

LISTING 4.7: frft\_integrate.m: Integration over an integral using the FRFT.

```
% frft_integrate.m
function y = KC_frft_integrate(f, u, x)
du = u(2)-u(1); % step of the support
dx = x(2)-x(1); % step of the results
5 f0 = f(1); fE = f(end); % first and last elements
x0 = x(1);
u0 = u(1); uE = u(end);
b = diff(f)/du; % b_j points
c = diff([0 b 0]); % c_j points
10 h = c.*exp(-i*x0*u); % vector passed to FRFT
a = 0.5*du*dx/pi; % FRFT parameter
h1 = KC_frft(h, a);
y = i*(f0*exp(-i*u0*x)-fE*exp(-i*uE*x))./x;
y = y + h1.*exp(-i*u0*(x-x0))./(x.^2);
15 end
```

## 4.7 SUMMARY

To summarize, for large classes of models closed form solutions even for European style options are not available but their characteristic function is available in closed form. For example, models where the logarithmic price is Lévy (Madan et al., 1998; Carr, Geman, Madan, and Yor, 2002), Garch models (Heston and Nandi, 2000), affine models (Heston, 1993; Duffie et al., 2000; Bates, 2000, 1998), regime switching models (Chourdakis, 2002) or stochastic volatility Lévy models (Carr, Geman, Madan, and Yor, 2003; Carr and Wu, 2004) fall within this category.

Fourier transform methods can be applied to recover numerically European call and put prices from the Fourier transform of the modified call. Therefore such models can be rapidly calibrated to a set of observed options contracts, as we will investigate in the next chapter on volatility. The FFT method or its fractional variant are well suited to perform this inversion. Also, one can use these methods to invert the characteristic function itself, thus recovering numerically the probability density function. This can in turn be used to set up numerical procedures for pricing American style or other exotic contracts, for example as in Andricopoulos, Widdicks, Duck, and Newton (2003).

LISTING 4.8: cf2cdf.m: Transform a characteristic function into a cumulative density function.

```
% cf2cdf.m
function [k, y, yv] = cf2cdf(cf, p, x, varargin)
ErrTol = 1e-8; % acceptable CDF improvement
MaxIter = 1000; % max number of iterations
5 a = 2; % damping factor
K = 8; N = 2^K; % FRFT points
kL = linspace(-x, 0, N+1); kL = kL(1:end-1);
kG = linspace(0, x, N+1); kG = kG(2:end);
flag = 1; iter = 1; % stopping flag
10 ucr = 0; umx = 0.25; % integration interval
y = 0; yv = [];
while flag
    yGe= Gint(ucr, ucr+umx); % get density over [0,+inf]
    yLe= Lint(ucr, ucr+umx); % get density over [-inf,0]
15    ye = [yLe, yGe];
    y = y + ye;
    yv = [yv; y];
    if ( max(ye)<ErrTol && iter>3 ) || iter>MaxIter
        flag = 0;
20    end
    ucr = ucr +umx;
    iter = iter+1;
    umx = umx*2;
end
25 y((N+1):end) = 1-y((N+1):end);
k = [kL, kG];
% dampen and integrate over negatives
function z = Lint(x1, x2)
    u = linspace(x1, x2, N);
30    ua = u+i*a;
    fa = feval(cf, ua, p);
    fb = fa./(a-i*u);
    z = frft_integrate(fb, u, kL);
    z = -exp(a*kL).*real(z)/pi;
end
% dampen and integrate over positives
function z = Gint(x1, x2)
    u = linspace(x1, x2, N);
    ua = u-i*a;
40    fa = feval(cf, ua, p);
    fb = fa./(a+i*u);
    z = frft_integrate(fb, u, kG);
    z = -exp(-a*kG).*real(z)/pi;
end
45 end
```

## Historical estimation and filtering

It is typical in many, if not all, financial application to face models that depend on one or more parameter values, which have to be somehow determined. For example, if we are making the assumption that the stock price we are investigating follows a homogeneous geometric Brownian motion, then we would be interested in estimating the expected return and the corresponding volatility. Then we could produce forecasts, option prices, confidence intervals and risk measures for an investment on this asset.

At this point we must remind ourselves that not all of the above operations are carried out under the same measure. This fact will largely determine which data will be appropriate to facilitate a calibration method. Some parameters, such as the drift in the Black-Scholes framework, are not the same under the objective and the pricing measure, while some others, such as the volatility, are.

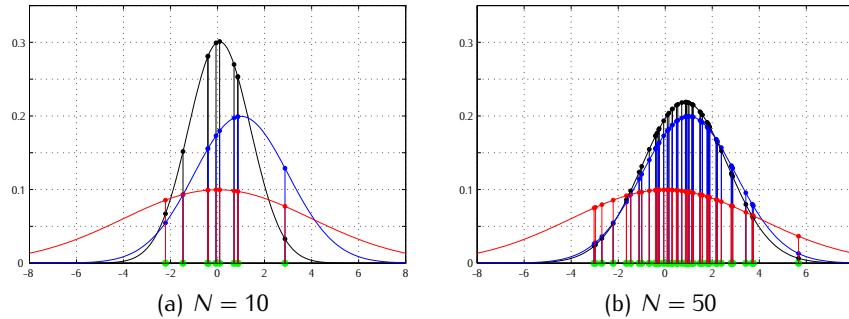
In particular, if our ultimate goal is pricing, we must place ourselves under the pricing measure and use instruments that are also determined under the same measure. In this way the prices that we produce will be consistent with the prices that we use as inputs, and we will not leave any room for arbitrage. The dynamics recovered under this data set will not be the real dynamics of the underlying asset: instead, they will be consistent with the attitude of investors against risk, and thus modified accordingly. In general, drifts will be lower, volatilities will be higher, and jumps will be more frequent and more severe. When pricing assets, investors behave as if this is the, precisely because these are the scenarios that they dislike.

On the other hand, if our goal is forecasting or risk management, we are interested in the real asset dynamics. We do not want the parameters to be contaminated by risk aversion, and the appropriate data in this case would be actual asset prices. Based on the real historical movements of assets we will base our forecasts for their future behaviour.

Nevertheless, there are situations where we might want (or have to) use both probability measures jointly. As derivative prices are forward looking we might want to augment our information set with their prices, in order to produce more accurate forecasts. From an "academic" point of view, since the distance between

FIGURE 5.1: Examples of density and likelihood functions. A sample of size  $N$  is drawn from the  $N(1.00, 2.00)$  distribution, presented with the green points on the horizontal axis. Three densities are also presented, together with the corresponding sample values. The blue curve gives the true  $N(1.00, 2.00)$  which gives a log-likelihood  $L(N = 10) = -19.37$  and  $L(N = 50) = -101.54$ ; the red curve gives the curve for  $N(0.00, 4.00)$  which far from the true density and has a low log-likelihood  $L(N = 10) = -23.60$  and  $L(N = 50) = -121.55$ ; finally the black curve is the one that maximizes the likelihood,  $N(0.08, 1.32)$  with  $L = -17.00$  for  $N = 10$ , and  $N(0.83, 1.82)$  with  $L = -100.98$  for  $N = 50$ .

[  lik\_example.m ]



the two probability measures depend on the risk premiums, we might want to identify these premiums for different risk components. For instance, we might want to quantify the price of volatility risk versus the price of jump risk. Finally, in some situations we do not observe the underlying asset directly. This is the case in fixed income markets, where we can attempt to identify the true dynamics using time series of bonds which are evaluated under the pricing measure.

In this chapter we will focus on the case where calibration is carried out using a time series of historical values. There is a plethora of methods available, but we will focus on the most popular one, the *maximum likelihood estimation* (MLE) technique. We will not focus on deriving the properties of MLE, but will rather refer to [Davidson and MacKinnon \(1985\)](#) and [Hamilton \(1994\)](#). These books also give a detailed analysis of variants of MLE, as well as alternative *method's of moments*. For an introduction to Bayesian techniques, a good starting point is [Zellner \(1995\)](#).

## 5.1 THE LIKELIHOOD FUNCTION

Suppose that we have in our disposal a time series of observations, say  $x = \{x_1, \dots, x_T\}$ , and a model which we assume has produced these observations. We

will denote with large  $X$  the random variables that are produced by the model,<sup>1</sup> and with small  $x$  the realizations that make up our sample. We will collect all parameters of this model in a  $(K \times 1)$  vector  $\vartheta$ . Our objective is twofold: we want to find an estimator  $\hat{\vartheta}$  of  $\vartheta$  which is based on our data set, but we also want to produce some confidence intervals on  $\hat{\vartheta}$ , acknowledging the fact that our data set is finite and thus our produced estimators are not equal to the true parameters of the data generating process. The likelihood function is a *measure of fit* that will allow us to fulfill these objectives.

To implement the likelihood function we scan through the sample, and pretend that we are standing at each point in time. Suppose that we are currently at the  $t$ -th observation, with  $1 \leq t \leq T$ . Given a value of the parameter set,  $\vartheta$ , we produce the conditional density  $f_{X_{t+1}|t}(x|\vartheta, x_1, \dots, x_t)$ , which we abbreviate with  $f_{t+1|t}(x)$ . Notice that we only use the information that was available at time  $t$ . Essentially we are asking the question: when we were at time  $t$  in the past, how would our forecast for time  $t+1$  look like? Then we go to the next time period  $t+1$  and see how good our forecasting density was: if we forecasted rather well, then the value of the density  $f_{t+1|t}(x_{t+1})$  will be high; if our forecasting density was poor, then  $f_{t+1|t}(x_{t+1})$  would be close to zero.

The value  $f_{t+1|t}(x_{t+1})$  is the likelihood of the point  $x_{t+1}$ , seen as a function of the parameter vector. To make this more explicit we introduce the notation  $f_{t+1|t}(\vartheta; x)$  for this likelihood. In order to construct the likelihood of the sample, we take the product  $\prod_{t=1}^{T-1} f_{t+1|t}(\vartheta; x)$ . The maximum likelihood estimator  $\hat{\vartheta}$  will be the one that maximizes the sample likelihood. Since this maximum will be the same under an increasing transformation, and for some other properties that we will discuss shortly, we typically work with the *log-likelihood* of the sample<sup>2</sup>

$$L(\vartheta; x) = \sum_{t=1}^{T-1} \log f(x_{t+1}|\vartheta, x_1, \dots, x_t)$$

To select the maximum log-likelihood we need to set the first order conditions, namely that

$$\frac{\partial}{\partial \vartheta} L(\hat{\vartheta}; x) = 0$$

The second order conditions will dictate that for the likelihood to be actually maximized, the  $K \times K$  Hessian matrix

---

<sup>1</sup> To be more precise,  $X$  contains the random variables that are conditional on their history. That is to say, the random variable  $X_t$  is conditional on the realizations of all values that preceded it, namely  $\{X_{t-1}, X_{t-2}, \dots, X_1\}$ .

<sup>2</sup> There are practical as well as theoretical reasons for doing so. Imagine having a sample of 1000 observations from the red density of figure 5.1(b) on page 132, where each has a likelihood of about  $0.1 = 10^{-1}$ . Then the likelihood of the sample would be of the order  $10^{-1000}$ , small enough to confuse the best of computers. But the log-likelihood (with base 10 for simplicity) is  $-1000$ , a much more manageable figure. This is a practical issue; the theoretical benefits include the computation of standard errors as described in the text.

$$\mathbf{H} = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} L(\hat{\boldsymbol{\theta}}; \mathbf{x}) \text{ is negative definite}$$

The maximization of the log-likelihood function can be carried out analytically in some special cases, but we typically employ some algorithm to produce  $\hat{\boldsymbol{\theta}}$  numerically. The choice of the appropriate algorithm will depend on the nature of the likelihood function: if it is relatively well behaved, then a standard hill climbing algorithm will be sufficient. In more complex cases, where the likelihood exhibits local maxima or is even undefined for specific parameter sets, one needs to resort to other techniques such as genetic algorithms or other simulation based methods.

Figure 5.1 illustrates the intuition behind the likelihood function. Samples are drawn from the blue distribution (for simplicity we assume that the sample elements are independent and identically distributed) of lengths  $N = 10$  and  $N = 50$ . To compute the corresponding likelihood values, one has to compute the density value at the sample points as shown. The red curves give a density that is far away from the true one, and we can see that overall the function values are lower. We numerically maximize the log-likelihood and estimate the density that has produced the data, which is given in black. When the sample is small, the estimated density is not close to the true data generating process, but it will converge as the sample size increases.

## 5.2 PROPERTIES OF THE ML ESTIMATORS

Maximum likelihood estimators share some very appealing asymptotic properties. Asymptotic in this context means that these properties hold at the limit, when the sample size approaches infinity. Therefore, one would tend to consider them more valid for large data samples. Unfortunately, how large a “large” sample should be is not set in stone, and depends on data generating process. For that reason it is always to verify the validity of any claims that are based on these properties via a small simulation experiment. Here we will go through some fundamental properties and will see how they can be used to make inference on the quality of the estimators.

We make the fundamental assumption that our model is correctly specified. This means that there is a data generating process which we guess correctly up to the parameter values. If our model is misspecified all properties go out of the window, even asymptotically. By using so-called *bootstrap techniques* we can take some steps towards testing our hypotheses and constructing confidence intervals, while taking into account possible misspecification.

We will denote the true value of the parameter set with  $\boldsymbol{\theta}^*$ . This is the set that has actually generated the series we observe. For us though, this is a random variable as we do not observe it directly. In fact, it is the qualities of this random variable that we intend to quantify.

### THE SCORE AND THE INFORMATION MATRIX

As we said, we produce the maximum likelihood estimator by setting  $\partial L(\boldsymbol{\vartheta}; \mathbf{x})/\partial \boldsymbol{\vartheta} = 0$ . This derivative is also called the *score* function, and the first order condition corresponds to an important property of the score, namely that its expectation, at the true parameter set is zero

$$\mathbb{E} \frac{\partial L(\boldsymbol{\vartheta}^*; \mathbf{X})}{\partial \boldsymbol{\vartheta}} = \mathbf{0}$$

Note that the random variable in the above expectation is the data sample  $\mathbf{X}$ . For IID processes, maximum likelihood estimation can be viewed as setting the empirical expectation of this score to zero.

In the same light we define the (*Fisher*) *information matrix* as minus the expectation of the second derivative of the log-likelihood, evaluated again at the true parameter point

$$\mathcal{I}(\hat{\boldsymbol{\vartheta}}) = -\mathbb{E} \frac{\partial^2 L(\boldsymbol{\vartheta}^*; \mathbf{X})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'}$$

As before, the Hessian matrix produces an estimate of the information matrix which is based on the sample. The information matrix will be by construction positive definite, and therefore invertible.

It turns out that we can also say something on the covariance matrix of the score. In fact, it will be equal to the information matrix

$$\sqrt{\frac{\partial L(\hat{\boldsymbol{\vartheta}}; \mathbf{X})}{\partial \boldsymbol{\vartheta}}} = \mathbb{E} \left( \frac{\partial L(\hat{\boldsymbol{\vartheta}}; \mathbf{X})}{\partial \boldsymbol{\vartheta}} \right)^2 = \mathcal{I}(\hat{\boldsymbol{\vartheta}})$$

What is the correct way to view these expectations and variances? Say that we knew the true parameter set, and we constructed a zillion sample paths based on these parameters, each one of length  $T$ . If we compute the score vector based on each one of these samples, we would find that the average of each element is zero and that the covariance matrix is given by the information matrix.

The information matrix plays another important role, as its positive definiteness is a necessary condition for all other asymptotic properties to carry through.

### CONSISTENCY AND ASYMPTOTIC NORMALITY

Lets say that we have a method of producing estimators, not necessarily maximum likelihood. If we produce different samples with the true parameter set, we will obviously end up with a different estimated value each time, and let us denote with  $\hat{\boldsymbol{\vartheta}}(\mathbf{x})$  the estimated parameter set that is generated by the sample  $\mathbf{x}$ . Of course we cannot carry this experiment out, since we do not know the true parameter set, but we can pose the question: if we produce a zillion alternative

samples, do we expect the average of their estimators to be equal to the true one? An estimation method is called *unbiased* if this is true, namely that

$$E\hat{\vartheta}(X) = \vartheta^*$$

The maximum likelihood estimator is not generally unbiased, and this apparently is not a good thing. But the maximum likelihood estimator is *consistent*, which means that as the sample size increases the bias drops to zero. Furthermore, the variance of the estimator's distribution also drops to zero, indicating that the maximum likelihood estimator will converge to the true value as the sample size increases, or more formally that

$$\text{plim}\hat{\vartheta}(X) = \vartheta^* \text{ as the sample size increases}$$

It also turns out that the distribution of the MLE is Gaussian, with covariance matrix equal to the inverse of the Fisher information matrix evaluated at the true parameter value. We can therefore write

$$\hat{\vartheta}(X) \sim N(\vartheta^*, \mathcal{I}(\vartheta^*)^{-1})$$

Furthermore, the variance  $\mathcal{I}(\vartheta^*)^{-1}$  of the MLE is equal to the so called *Cramér-Rao lower bound*, which states that no other unbiased estimator will have smaller variance than the MLE. This also makes the maximum likelihood estimator asymptotically efficient. In practice we do not know the value of  $\mathcal{I}(\vartheta^*)$  and use an estimate instead, for example one based on the Hessian of the log-likelihood.

### HYPOTHESIS TESTING AND CONFIDENCE INTERVALS

For large samples we can utilize the efficiency and asymptotic normality of the maximum likelihood estimator, and produce confidence intervals that are based on the normal distribution. In particular, based on the sample we can test the hypothesis

$$H_0 : \vartheta^* = \vartheta^\dagger$$

against the alternative that  $\vartheta^* \neq \vartheta^\dagger$ . Under the null the maximum likelihood estimates will be asymptotically distributed as

$$\hat{\vartheta}(X) \sim N(\vartheta^\dagger, \mathcal{I}(\vartheta^\dagger)^{-1})$$

We are therefore naturally led to the statistic

$$Z(x) = \frac{\hat{\vartheta}(x) - \vartheta^\dagger}{\sqrt{\mathcal{I}(\vartheta^\dagger)^{-1}}}$$

which is distributed as a standardized normal. We would reject the null hypothesis at, say, the 95% confidence level if  $|Z(x)| > 1.96$ .

LISTING 5.1: `arma_sim.m`, `arma_sim.m` and `arma_sim.m`: Simulation and maximum likelihood estimation of ARMA models

```
%arma_sim.m
function x = arma_sim(c, ar, ma, s, N)
p = length(ar); q = length(ma); a = max(p,q);
T = N(1) + a; K = N(2);
5 u = s*randn(T, K); x = zeros(T, K);
x(1:a,:) = c/(1-sum(ar));
x0 = 0; u0 = 0;
for z=a:(T-1)
    if p>0, x0 = ar*x((z-p+1):z,:); end
10 if q>0, u0 = ma*u((z-q+1):z,:); end
    x(z+1,:) = c + x0 + u0 + u(z+1,:);
end

%arma_lik.m
15 function y = arma_lik(c, ar, ma, s, x)
p = length(ar); q = length(ma); a = max(p,q);
T = length(x) + a; u = zeros(T, 1);
x = [c/(1-sum(ar))*ones(a,1); x];
x0 = 0; u0 = 0; y = 0;
20 for z=a:(T-1)
    if p>0, x0 = ar*x((z-p+1):z); end
    if q>0, u0 = ma*u((z-q+1):z); end
    u(z-p+2) = x(z-p+2) - (c + x0 + u0);
    y = y - log(normpdf(u(z-p+2), 0, s));
25 end

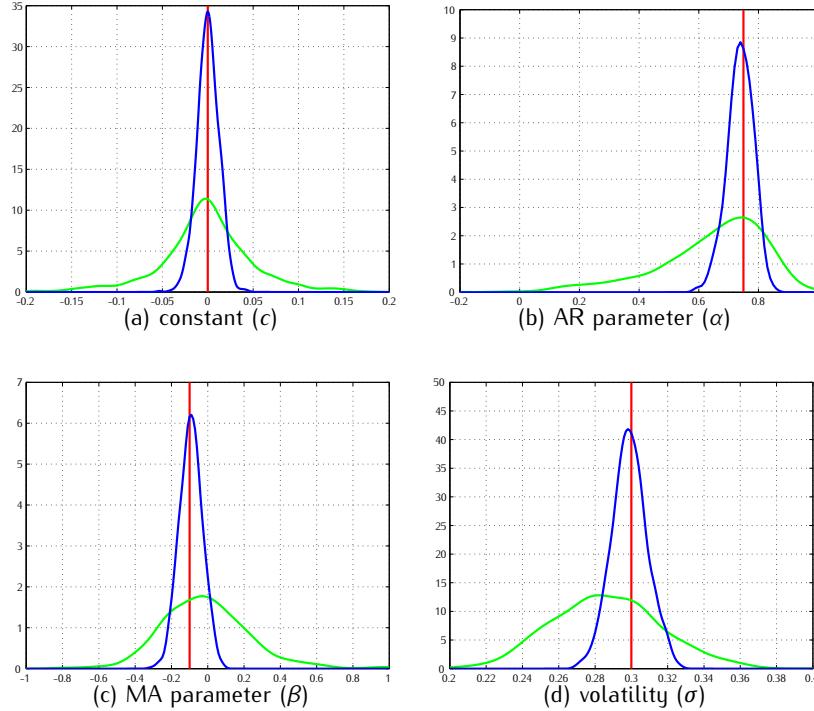
%arma_example.m
N = 1000; T = 50;
y = arma_sim(0.00, [0.75], [-0.10], 0.3, [T, N]);
30 x = zeros(N, 4);
opt = optimset('LargeScale','off','Display','off');
for in=1:N
    f = @(v) arma_lik(v(1), v(2), v(3), v(4), y(:,in));
    z = fmincon(f, [0.00 0.75 -0.10 0.3], [], [], ...
35 [], [], [-Inf -0.999 -0.999 0.01], ...
[Inf 0.999 0.999 0.50], [], opt);
    x(in,:) = z;
end
```

## 5.3 SOME EXAMPLES

for copies, comments, help etc. visit <http://www.theponytail.net/>

FIGURE 5.2: Bias and asymptotic normality of maximum likelihood estimators. Sample paths of an ARMA(1,1) model  $x_t = c + \alpha x_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}$  are simulated, and the parameters are subsequently estimated. 1,000 short samples ( $T = 50$ ) and 1,000 longer samples ( $T = 500$ ) were generated. The graphs give the distribution of the estimators for the short sample size in green, and for the longer sample size in blue. The true parameter values are in red. The table presents some summary statistics that correspond to the distributions.

[  `arma_example.m` ]



	$T$	$c$	$\alpha$	$\beta$	$\sigma$
True		0.00	0.75	-0.10	0.30
Mean	50	-0.002	0.64	-0.013	0.27
	500	0.000	0.74	-0.093	0.30
Volatility	50	0.056	0.20	0.24	0.030
	500	0.012	0.044	0.063	0.010
Skewness	50	0.001	-1.8	0.57	0.076
	500	-0.12	-0.34	0.067	0.025
Kurtosis	50	8.1	8.8	4.7	2.9
	500	3.5	3.1	3.0	3.0

## LINEAR ARMA MODELS

Suppose that the data generating process has autoregressive and moving average terms, and for simplicity assume that both effects are of the first order. Then, we can write the process as

$$x_t = c + \alpha x_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}, \quad \varepsilon_t \sim N(0, \sigma^2)$$

Since our sample is finite, we have to make an extra assumption on the values just before the starting date, namely  $x_0$  and  $\varepsilon_0$ . We will set them equal to their expected values, that is to say  $x_0 = \frac{c}{1-\alpha}$  and  $\varepsilon_0 = 0$ .

The parameter set in this case is  $\vartheta = \{c, \alpha, \beta, \sigma\}$ . Each observation is Gaussian conditional on its past, and in particular

$$x_t | \vartheta, x_{t-1}, \dots, x_0 \sim N(c + \alpha x_{t-1} + \beta \varepsilon_{t-1}, \sigma^2)$$

Therefore we can compute the log-likelihood function of the sample as

$$L(\vartheta; x) = -\frac{T}{2} \log(2\pi\sigma^2) - \sum_{t=1}^T \frac{(x_t - c - \alpha x_{t-1} - \beta \varepsilon_{t-1})^2}{2\sigma^2}$$

If we have enough time in our hands we could in principle maximize this function analytically, but it is more convenient to carry out the maximization numerically. Listing 5.1 shows how this can be done in a very simple way. In fact, listing 5.1 conducts a small experiment that illustrates the bias and non-normality of maximum likelihood estimators in small samples.

We assume a known parameter set,  $\vartheta = \{0, 0.75, -0.10, 0.30\}$ , and simulate paths of length  $T = 50$  and  $T = 500$ . In total we simulate 1000 paths for each length. We then estimate the parameters using maximum likelihood. *Consistency* will indicate that although in small samples the estimators might be biased, as the sample grows the mean should converge to the true value, while the estimator variance should decrease. Due to the *asymptotic normality* the distribution should become gradually closer to the Gaussian one.

Figure 5.2 gives the results of this experiment, and we can observe that this is indeed the case. The bias is more pronounced for the autoregressive and the moving average parameters, with their means biased towards zero for the smaller sample. The non-Gaussian nature of the estimators is also apparent, with the kurtosis being consistently high. As the sample size increases, the estimator densities become tighter and markedly more symmetric. The volatility, although slightly biased for the smaller sample, is very accurately estimated, exhibiting very small standard deviation. This is a more general feature: drifts are very sensitive to the actual path, as they largely depend on the first and the last observation. Volatilities on the other hand take more information over the variability of the path, and their estimators converge a lot faster.

Note that the fact that asymptotically the parameters are Gaussian does not imply that they are independent. Indeed, the correlation matrices of the parameter estimates are

$$\begin{pmatrix} 1 & 0.048 & 0.000 & -0.038 \\ 0.048 & 1 & -0.71 & 0.081 \\ 0.000 & -0.71 & 1 & -0.044 \\ -0.038 & 0.081 & -0.044 & 1 \end{pmatrix}, \text{ and } \begin{pmatrix} 1 & 0.052 & -0.015 & -0.042 \\ 0.052 & 1 & -0.77 & 0.099 \\ -0.015 & -0.77 & 1 & -0.030 \\ -0.042 & 0.099 & -0.030 & 1 \end{pmatrix}$$

for the sample sizes  $T = 50$  and  $T = 500$  respectively. Observe the high negative correlation between the estimator of the autoregressive and the moving average terms. As these two parameters compete to capture the same features of the data,<sup>3</sup> the estimates parameters tend to come in high/low pairs.

## LÉVY MODELS

Lévy models can be easily estimated using the MLE approach, by inverting the characteristic function using the FFT or fractional FFT methods of chapter 4. We can invert the characteristic function directly to produce the probability density, or we can invert for the cumulative density and then use numerical differentiation. Although the second method appears to be more cumbersome, it is often more stable. This happens in the case of Lévy models because the density typically exhibits a very sharp peak, which the direct transform might fail to capture.

Irrespective of the method we choose to construct the density, the maximization of the log-likelihood should be straightforward. As Lévy models are time-homogeneous, the returns are identically distributed. If we denote with  $f(x; \theta)$  the probability density of the Lévy model, then the log-likelihood can be easily computed over a series of returns  $\{x_1, \dots, x_T\}$  as

$$L(\vartheta; x) = \sum_{t=1}^T \log f(x_t | \vartheta)$$

Generally speaking, as the FFT method will produce a dense grid for the probability density function we only have to call the Fourier inversion once at each likelihood evaluation and interpolate between those points. This renders MLE quite an efficient method for the estimation of Lévy processes.<sup>4</sup>

In our example we will be using the cumulative density function, recovered with the code of listing 4.8. We use data of the S&P500 index.

---

<sup>3</sup> An AR(1) process can be written as an MA( $\infty$ ) one and vice versa. Therefore a series that is generated by an AR(1) data generating process will produce MA(1) estimators as a first order approximation, if the estimated model is misspecified.

<sup>4</sup> Some popular Lévy models admit closed form expressions for the probability density function. In principle this means that one can avoid the FFT step altogether and use the closed form instead. It turns out that in the majority of cases these densities are expressed in terms of special functions, which can be more expensive to compute (over the data set) than a single FFT!

## 5.4 LIKELIHOOD RATIO TESTS

## 5.5 THE KALMAN FILTER

Many financial quantities can (and will) be dependent on factors that are unobserved. Market sentiment, asset volatility, liquidity, the instantaneous risk free rate and the position on the business cycle are just a few examples of such unobserved factors which have an impact on the behaviour of other series. For some of these quantities one can develop proxies, but they themselves will serve at best as an unbiased but noisy manifestation of the real underlying quantity.

To deal with such problems various optimal filters have been developed, with the Kalman filter being the most popular of all. Essentially, the Kalman filter is described by a pair of linear equations, which specify the dynamics of the latent and the observed variables. In particular, in its simplest form we can write

$$\begin{aligned} \text{observation eq: } X_t &= a_x + b_x Y_t + \epsilon_t \\ \text{transition eq: } Y_t &= a_y + b_y Y_{t-1} + \eta_t \end{aligned}$$

with Gaussian error terms  $\epsilon_t$  and  $\eta_t$ .

The information set in our disposal consists of observations of  $\mathcal{F}_t = \{X_s : s \leq t\}$ , but we are really interested on the dynamics of the process  $y_t$ . The filter provides us with estimates  $\mu_{t|s} = E[Y_t | \mathcal{F}_s]$ , together with the conditional variances  $v_{t|s} = E[(Y_t - \mu_{t|s})^2 | \mathcal{F}_s]$ . As it turns out the conditional distributions are Gaussian, and are therefore summarized by these two moments.

### THE FILTERING PROCEDURE

Kalman filtering consists of two steps, which are commonly referred to as the *prediction* and *correction* steps. A recursive algorithm produces the conditional distribution of  $Y_t | \mathcal{F}_t$ , based on the conditional distribution  $Y_t | \mathcal{F}_{t-1}$ . Therefore, given an initial Gaussian distribution for  $Y_0$ , one can move forward and produce the filtered conditional distributions for all times  $t$ .

We begin by some notation. The conditional distribution of  $Y_t$  given the information at time  $s$  is denoted with  $F_{t|s}(y) = P[Y_t \leq y | \mathcal{F}_s]$ , and the conditional density  $f_{t|s}(y) = P[Y_t \in dy | \mathcal{F}_s] = \frac{d}{dy} F_{t|s}(y)$ . Using this notation, the prediction step with provide us with the conditional density  $f_{t|t-1}$ , as a function of the filtered density  $f_{t-1|t-1}$ ; the correction step will incorporate the new observation  $x_t$  and update the density  $f_{t|t-1}$  to produce the new filtered density  $f_{t|t}$ . We will use  $g_{t|s}(x)$  for the corresponding density of  $X_t$ ,  $g_{t|s}(x) = P[X_t \in dx | \mathcal{F}_s]$ . With  $\phi(z; \mu, \sigma^2)$  we denote the Gaussian density with mean  $\mu$  and variance  $\sigma^2$ .

For the prediction step we employing the formula for conditional probabilities, conditioning on the value of  $Y_{t-1}$

$$\begin{aligned}
 f_{t|t-1}(y_t) &= P[Y_t \in dy_t | \mathcal{F}_{t-1}] \\
 &= \int_{\mathbb{R}} P[Y_t \in dy_t | Y_{t-1} \in dy_{t-1}, \mathcal{F}_{t-1}] P[Y_{t-1} \in dy_{t-1} | \mathcal{F}_{t-1}] dy_{t-1} \\
 &= \int_{\mathbb{R}} \phi(y_t; a_y + b_y y_{t-1}, \sigma_\eta) f_{t-1|t-1}(y_{t-1}) dy_{t-1}
 \end{aligned}$$

Observe that if the filtered density  $f_{t-1|t-1}$  is Gaussian, then the prediction for  $Y_t$  will also follow a Gaussian distribution, as the convolution of normals.

In the correction step we incorporate the new observation  $X_t = x_t$  which updates the information set  $\mathcal{F}_t = \{X_t = x_t\} \cup \mathcal{F}_{t-1}$ . We then write

$$f_{t|t}(y_t) = P[Y_t \in dy_t | \mathcal{F}_t] = P[Y_t \in dy_t | \{X_t \in dx_t\} \cup \mathcal{F}_{t-1}]$$

Bayes' formula will then provide us with the alternative representation

$$\begin{aligned}
 f_{t|t}(y_t) &= P[X_t \in dx_t | \{Y_t \in dy_t\} \cup \mathcal{F}_{t-1}] \frac{P[Y_t \in dy_t | \cup \mathcal{F}_{t-1}]}{P[X_t \in dx_t | \cup \mathcal{F}_{t-1}]} \\
 &= \phi(x_t; a_x + b_x y_t, \sigma_\epsilon) \frac{f_{t|t-1}(y_t)}{g_{t|t-1}(x_t)}
 \end{aligned}$$

In this expression  $g_{t|t-1}(x_t)$  is treated as a constant, as it is not dependent on the variable  $y_t$ .<sup>5</sup> Once again, if  $f_{t|t-1}$  is Gaussian, it follows that  $f_{t|t}$  will also be normally distributed, as the product of two Gaussian densities.

We have in fact shown that if the distribution of the initial state variable  $Y_1 | \mathcal{F}_0$  is Gaussian, th will also be Gaussian, which implies that  $Y_1 | \mathcal{F}_1$  is Gaussian, which in turn implies that  $Y_2 | \mathcal{F}_1$ ,  $Y_2 | \mathcal{F}_2$ ,  $Y_3 | \mathcal{F}_2$ , ...,  $Y_{t+1} | \mathcal{F}_t$ ,  $Y_{t+1} | \mathcal{F}_{t+1}$ , and so on, are all normally distributed.

This implies that our filtering procedure can be implemented by keeping track of the conditional means and variances, rather than the whole distribution. In fact, the prediction step will give the quantities  $\mu_{t|t-1}$  and  $\nu_{t|t-1}$  in terms of  $\mu_{t-1|t-1}$  and  $\nu_{t-1|t-1}$  using iterated expectations on the transition equation. In particular, we can write

$$\mu_{t|t-1} = E[Y_t | \mathcal{F}_{t-1}] = E[E[Y_t | Y_{t-1}, \mathcal{F}_{t-1}]]$$

with the outer expectation integrating over  $Y_{t-1}$ . The transition equation yields the inner expectation,  $E[Y_t | Y_{t-1}, \mathcal{F}_{t-1}] = a_y + b_y Y_{t-1}$ . The same procedure can be applied to the variance, and the resulting predicted moments are given by

$$\begin{aligned}
 \mu_{t|t-1} &= a_y + b_y \mu_{t-1|t-1} \\
 \nu_{t|t-1} &= b_y^2 \nu_{t-1|t-1} + \sigma_\eta^2
 \end{aligned}$$

---

<sup>5</sup> Aficionados of Bayesian statistical inference will recognize  $g_{t|t-1}(x_t)$  as the normalization constant which would be probably ignored. But in our setting it is not ignored, in fact it facilitates the maximum likelihood estimation of the parameters.

The correction step will update the means and variances using the observation  $X_t = x_t$ . In particular, computing the product of the two Gaussian densities gives<sup>6</sup>

$$\begin{aligned} f_{t|t}(y_t) &\propto \exp \left\{ -\frac{(x_t - a_x - b_x y_t)^2}{2\sigma_\epsilon^2} \right\} \exp \left\{ -\frac{(y_t - \mu_{t|t-1})^2}{2\sigma_{t|t-1}^2} \right\} \\ &\propto \exp \left\{ -\frac{(y_t - \mu_{t|t})^2}{2\nu_{t|t}} \right\} \end{aligned}$$

$$\text{for the quantities } \mu_{t|t} = \frac{\nu_{t|t-1} b_x (x_t - a_x) + \sigma_\epsilon^2 \mu_{t|t-1}}{\nu_{t|t-1} b_x^2 + \sigma_\epsilon^2}$$

$$\nu_{t|t} = \frac{\nu_{t|t-1} \sigma_\epsilon^2}{\nu_{t|t-1} b_x^2 + \sigma_\epsilon^2}$$

In the literature the above expressions are more compactly written in terms of the *Kalman gain*  $K_t$ , namely

$$\begin{aligned} K_t &= \frac{\nu_{t|t-1} b_x}{\nu_{t|t-1} b_x^2 + \sigma_\epsilon^2} \\ \mu_{t|t} &= \mu_{t|t-1} + K_t (x_t - a_x - b_x \mu_{t|t-1}) \\ \nu_{t|t} &= \nu_{t|t-1} - K_t \nu_{t|t-1} b_x \end{aligned}$$

This concludes the recursive scheme. Given the values for the parameter set  $\vartheta = \{a_x, b_x, a_y, b_y, \sigma_\epsilon, \sigma_\eta\}$  and an initial Gaussian distribution for  $Y_1|F_0$ , we can filter the expected path for  $X_t$ , together with its estimated variance. This initial distribution is typically taken as the unconditional distribution of the latent variable, that is to say a Gaussian with mean  $\mu_{1|0} = a_y/(1 - b_y)$  and  $\nu_{1|0} = \sigma_\eta^2/(1 - b_y^2)$ .

For completeness we give the filtering equations in the box below

#### Prediction

$$\begin{aligned} \mu_{t|t-1} &= a_y + b_y \mu_{t-1|t-1} \\ \nu_{t|t-1} &= b_y^2 \nu_{t-1|t-1} + \sigma_\eta^2 \end{aligned}$$

#### Correction

$$\begin{aligned} K_t &= \frac{\nu_{t|t-1} b_x}{\nu_{t|t-1} b_x^2 + \sigma_\epsilon^2} \\ \mu_{t|t} &= \mu_{t|t-1} + K_t (x_t - a_x - b_x \mu_{t|t-1}) \\ \nu_{t|t} &= \nu_{t|t-1} - K_t \nu_{t|t-1} b_x \end{aligned}$$

<sup>6</sup> Regarding the notation, “ $\propto$ ” stands for “proportional to”; that is  $x \propto y$  means that  $x = Cy$  for some constant  $C$ . Here we know that the resulting expression is a Gaussian density, and therefore we are just interested in the structure of the exponential rather than the constant that ensures that the total probability is equal to one.

LISTING 5.2: kalman\_filter\_1D.m: One dimensional Kalman filter

```
%kalman_filter_1D.m
function [L, m, s] = kalman_filter_1D(p, x)
aX = p(1); bX = p(2); aY = p(3); bY = p(4);
vX = p(5)^2; vY = p(6)^2;
5 T = length(x);
m0 = aY/(1-bY^2); v0 = vX/(1-bY^2); % predicted moments
m = zeros(T,1); v = zeros(T,1); L = 0;
for t=1:T
    % likelihood evaluation
10    m1 = a_x + bX*m0;
    v1 = bX^2*v0 + vX;
    L = L + log(v1) + (x(t)-m1)/v1;
    % correction step
    K = v0*bX/(v0*bX^2+vX); % Kalman gain
15    m(t) = m0 + K*(x(t) - aX - bX*m0); % update mean
    v(t) = v0 - K*v0*bX; % update variance
    % prediction step
    m0 = aY + bY*m(t); % update mean
    v0 = v(t)*bY^2 + vY; % update variance
20 end
s = sqrt(v);
```

## MAXIMUM LIKELIHOOD ESTIMATION

The quantity  $g_{t|t-1}(x_t)$  is very useful in this setting, since it represents the likelihood of observation  $x_t$ , given the information at the previous point in time,  $\mathcal{F}_{t-1}$ . With that, we can devise a maximum likelihood estimation strategy for the parameters  $\vartheta$ , by searching for the parameter set that optimized the log-likelihood of the sample

$$\hat{\vartheta} = \arg \max_{\vartheta} \left\{ \sum_{t=1}^T \log g_{t|t-1}(x_t; \vartheta) \right\}$$

In particular, we can write the conditional likelihood of each observation by conditioning on  $Y_t = y_t$

$$g_{t|t-1}(x_t) = \int_{\mathbb{R}} \phi(x_t; a_x + b_x y_t, \sigma_e) \phi(y_t; \mu_{t|t-1}, \sqrt{v_{t|t-1}}) dy_t$$

which once again as a convolution is indeed Gaussian. Iterated expectation of the observation equation will provide us with the mean ( $\mu_{t|t-1}^x$ ) and variance ( $v_{t|t-1}^x$ ) of the conditional likelihood

$$\begin{aligned}\mu_{t|t-1}^x &= a_x + b_x \mu_{t|t-1} \\ v_{t|t-1}^x &= b_x^2 v_{t|t-1} + \sigma_\epsilon^2\end{aligned}$$

Listing 5.2 gives an implementation of the one dimensional Kalman filter, together with the log-likelihood computation. In figure 5.3 we present an example of the filter output. The system of equations that have been implemented are

$$\begin{aligned}\text{observation: } X_t &= 0.00 + 1.00 Y_t + \epsilon_t, \sigma_\epsilon = 0.50 \\ \text{transition: } Y_t &= 0.00 + 0.90 Y_{t-1} + \eta_t, \sigma_\eta = 0.20\end{aligned}$$

The series of interest is  $Y_t$ , and its dynamics are known, but we cannot observe  $Y_t$  directly. Instead, we observe a noisy version  $X_t$ , which we use to filter out the path of  $Y_t$ . In the figure, the observed series are given by the blue crosses, and based on these we filter out  $\mu_{t|t}$ , which is given in red. With green we give the true path of  $Y_t$ , which we want to reconstruct. We also give the two standard deviation shaded area  $\mu_{t|t} \pm 2\sqrt{v_{t|t}}$ .

However, in this experiment we have assumed that the parameters of the dynamic system are known, which is not the case in practice. In a real life situation, we would be given the series of observations  $X_t$ , and we would be asked to estimate the parameter vector  $\vartheta$ , and then filter out the latent component.

## GENERALIZATIONS AND EXTENSIONS

This paradigm is now extended in two directions: first we introduce multivariate filtering systems, and we then consider a more general nonlinear framework.

### Multivariate systems

The filter described above can be easily extended to vector processes; the matrix algebra is a bit more involved, but the ideas remain the same.<sup>7</sup> Exogenous explanatory variables can be also included to the observation equation. In its most general form, the Kalman filter equations are given by

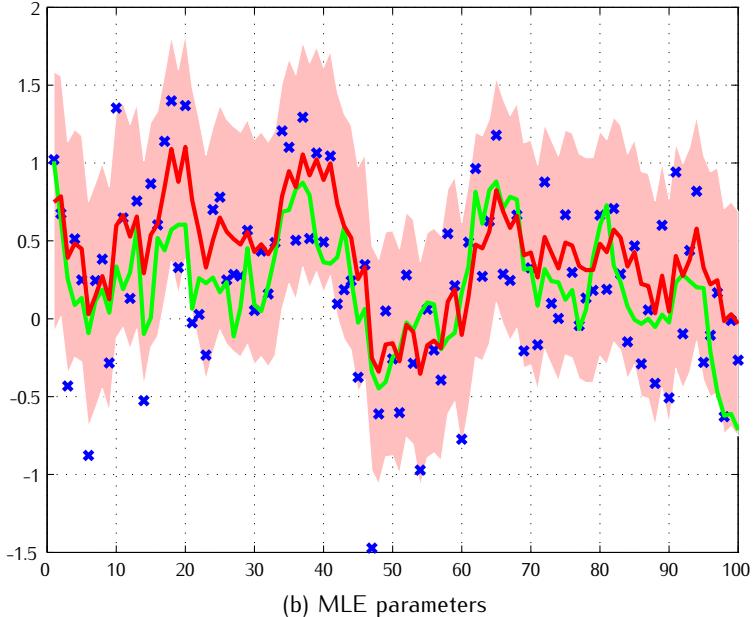
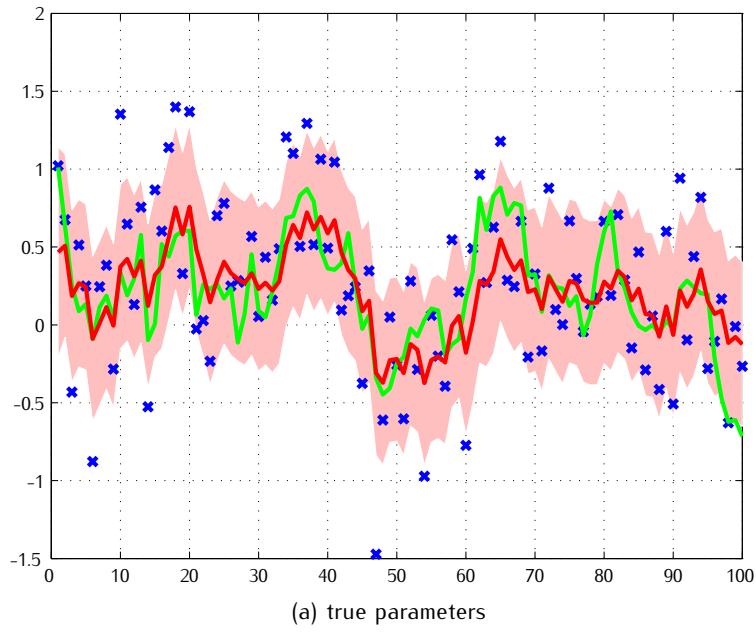
$$\begin{aligned}\text{observation eq: } X_t &= A Z_t + B_x Y_t + \epsilon_t \\ \text{transition eq: } Y_t &= B_y Y_{t-1} + \eta_t\end{aligned}$$

Here,  $X_t$  is a  $(N_x \times 1)$  vector of the observed variables,  $Y_t$  collects the  $N_y$  latent factors, and  $Z_t$  keeps the  $N_z$  explanatory variables. In addition, the white noise processes  $\epsilon_t$  and  $\eta_t$  possess covariance matrices  $\Sigma_\epsilon$  and  $\Sigma_\eta$  respectively.

Once again the filtering equations will be Gaussian, but now in a multivariate setting. We will denote with  $\mu_{t|s} = E[Y_t | \mathcal{F}_\tau]$ , with corresponding covariance matrix  $V_{t|s} = E[(Y_t - \mu_{t|s})(Y_t - \mu_{t|s})' | \mathcal{F}_\tau]$ . The Kalman filter updating equations are given in the box below as the counterparts to the univariate case

<sup>7</sup> See Hamilton (1994) for an excellent exposition, and Caines (1988) for a more detailed treatment of linear stochastic systems.

FIGURE 5.3: Kalman filtering example. The latent series  $X_t$  that we want to reconstruct is given in green, and the observed series  $Y_t$  is given by the blue crosses. The filtered series  $\mu_{t|t}$  is given in red, together with the two standard deviation shaded area  $\mu_{t|t} \pm 2\sqrt{v_{t|t}}$ .  
[  kalman\_filter\_example.m ]



for copies, comments, help etc. visit <http://www.theponytail.net/>

**Prediction**

$$\begin{aligned}\boldsymbol{\mu}_{t|t-1} &= \mathbf{B}_y \boldsymbol{\mu}_{t-1|t-1} \\ \mathbf{V}_{t|t-1} &= \mathbf{B}_y \mathbf{V}_{t-1|t-1} \mathbf{B}'_y + \boldsymbol{\Sigma}_\eta\end{aligned}$$

**Correction**

$$\begin{aligned}\mathbf{K}_t &= \mathbf{V}_{t|t-1} \mathbf{B}'_x (\mathbf{B}_x \mathbf{V}_{t|t-1} \mathbf{B}'_x + \boldsymbol{\Sigma}_\epsilon)^{-1} \\ \boldsymbol{\mu}_{t|t} &= \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t (x_t - \mathbf{A} z_t - \mathbf{B}_x \boldsymbol{\mu}_{t|t-1}) \\ \mathbf{V}_{t|t} &= \mathbf{V}_{t|t-1} - \mathbf{K}_t \mathbf{B}_x \mathbf{V}_{t|t-1}\end{aligned}$$

If we set the innovation process

$$\mathbf{W}_t = \mathbf{X}_t - \mathbf{A} Z_t - \mathbf{B}_x \boldsymbol{\mu}_{t|t-1}$$

and denote with  $w_t$  its realization, then we can write the correction step above in an alternative, sometimes more intuitive form, as

**Correction**

$$\begin{aligned}\mathbf{K}_t &= \langle \mathbf{X}_t, Y_t | \mathcal{F}_{t-1} \rangle \langle \mathbf{W}_t, \mathbf{W}_t | \mathcal{F}_{t-1} \rangle^{-1} \\ \boldsymbol{\mu}_{t|t} &= \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t w_t \\ \mathbf{V}_{t|t} &= \mathbf{V}_{t|t-1} - \mathbf{K}_t \langle \mathbf{W}_t, \mathbf{W}_t | \mathcal{F}_{t-1} \rangle \mathbf{K}'_t\end{aligned}$$

with  $\langle \cdot, \cdot | \cdot \rangle$  a shorthand for the conditional covariance matrix.

Listing 5.3 implements the general form of the multivariate Kalman filter. Such filters are typically used in yield curve modeling. For that reason, an implementation example of the multivariate Kalman filter can be found in section 7.6.

The standard Kalman filter system is linear in nature and Gaussian. Another important feature (or limitation) is that the latent variable enters the observation through the drift. A number of extensions have been proposed in the literature which attempt to relax some of these restrictions. As a rule, these extensions involve approximations at one point or the other, and therefore their performance is as good as this approximation. A nonlinear version of the system can be written as follows, where we ignore the impact of exogenous variables for simplicity

$$\begin{aligned}\text{observation eq: } \mathbf{X}_t &= h(Y_t, \epsilon_t) \\ \text{transition eq: } Y_t &= f(Y_{t-1}, \eta_t)\end{aligned}$$

The idea is to propagate the mean and covariance matrix by prediction and correction steps, just like the standard Kalman filter. The differences between the approaches is in the way these moments are computed.

**Extended Kalman filter**

The *extended Kalman filter* takes a simple linearization approach, by approximating a nonlinear function with its first order approximation. This means that

LISTING 5.3: *kalman\_filter.m*: The  $N$ -dimensional Kalman filter

```
%kalman_filter.m
function [L, M, S] = kalman_filter(p, X)
CX = p.CX; BX = p.BX; SE = p.SE; % observations
CY = p.CY; BY = p.BY; SH = p.SH; % factors
5 MC = p.M0; VC = p.V0;
T = size(X,1);
M = zeros(T, size(SH,1)); S = M; L = 0;
for t=1:T
    % prediction step
10    MP = CY + BY*MC; % update mean
    VP = BY*VC*BY' + SH; % update variance
    %% likelihood evaluation
    Mx = CX + BX*MP;
    Vx = BX*VP*BX' + SE;
15    Xt = X(t,:)' - Mx;
    L = L + log(det(Vx)) + Xt'/Vx*Xt;
    % correction step
    K = VP*BX'/Vx; % Kalman gain
    MC = MP + K*(X(t,:)' - CX - BX*MP); % update mean
20    VC = VP - K*BX*VP; % update variance
    % store
    M(t,:) = MC';
    S(t,:) = sqrt(diag(VC))';
end
```

the linearization error will be an effect of Jensen's inequality. For instance, the transition equation is linearized around the point  $(\mu_{t-1|t-1}, \mathbf{0})$ , which yields the approximation

$$f(Y_t, \eta_t) \approx f(\mu_{t-1|t-1}, \mathbf{0}) + F_y [Y_{t-1} - \mu_{t-1|t-1}] + F_\eta \eta_t$$

Where  $F_y$  and  $F_\eta$  are the matrices of first derivatives (the Jacobians) of the function  $f$ , with respect to the corresponding elements. Now the system has a linear form, and taking the expectation and the covariance produces the set of prediction equations

$$\begin{aligned}\mu_{t|t-1} &= f(\mu_{t-1|t-1}, \mathbf{0}) \\ V_{t|t-1} &= F_y V_{t-1|t-1} F_y' + F_\eta \Sigma_\eta F_\eta'\end{aligned}$$

We can apply the same idea to produce the correction step, namely

$$\begin{aligned}\mathbf{K}_t &= \mathbf{V}_{t|t-1} \mathbf{H}'_x (\mathbf{H}_x \mathbf{V}_{t|t-1} \mathbf{H}'_x + \mathbf{H}'_\epsilon \boldsymbol{\Sigma}_\epsilon \mathbf{H}_\epsilon)^{-1} \\ \boldsymbol{\mu}_{t|t} &= \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t (\mathbf{x}_t - \mathbf{H}_x \boldsymbol{\mu}_{t|t-1}) \\ \mathbf{v}_{t|t} &= \mathbf{v}_{t|t-1} - \mathbf{K}_t \mathbf{H}_x \mathbf{V}_{t|t-1}\end{aligned}$$

Once again the matrices  $\mathbf{F}_y$  and  $\mathbf{F}_\eta$  denote the appropriate Jacobians, in this case of the function  $\mathbf{h}$ .

### Unscented Kalman filter

Although the standard Kalman filter is optimal in the space of linear models, the extended Kalman filter is suboptimal in the much wider space of nonlinear models. The *unscented Kalman filter* takes a different route in producing estimates for the mean and the covariance of the relevant random variables. This was introduced in [Julier and Uhlmann \(1996, 1997\)](#) and has since gained popularity in nonlinear filtering.

The main idea behind the unscented filter can be summarized in contrast to the extended Kalman filter approach: Both transition and measurement equations are nonlinear functions of random variables. In the extended Kalman filter, the means and the covariances of these nonlinear functions are approximated by linearizing the functions themselves. In the unscented Kalman filter, the means and covariances are computed by creating first a small sample of points from the appropriate multivariate Gaussian distribution, and then transform these points via these nonlinear functions. These sample points are called *sigma points* in the unscented filtering jargon. It turns out that in many cases this approach provides a much more robust approximation to the true dynamics. In addition, the unscented implementation does not require taking derivatives, which can be a cumbersome procedure and requires sufficiently smooth functions.

It is clear from the Kalman filtering equations that all needed for the prediction and correction steps are the following steps

1. For the prediction step produce  $\boldsymbol{\mu}_{t|t-1}$  and  $\mathbf{V}_{t|t-1}$ .
2. For the correction step compute the covariance between the next observation and the latent variable  $\langle \mathbf{X}_t, Y_t | \mathcal{F}_{t-1} \rangle$ .
3. For the correction step compute the variance of the innovation  $\langle \mathbf{W}_t, \mathbf{W}_t | \mathcal{F}_{t-1} \rangle$ .

The unscented filter produces these quantities in an intuitive and simple to implement way. To do so, we need to introduce the *augmented state vector*  $\mathbf{Y}_t^a$ , which concatenates the state process and the innovations, and has length  $N^a$

$$\mathbf{Y}_t^a = \begin{pmatrix} \mathbf{Y}_t \\ \boldsymbol{\epsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix}$$

We will denote with  $\boldsymbol{\mu}_{t|s}^a$  and  $\mathbf{V}_{t|s}^a$  the mean and covariance matrix of this augmented state vector. The unscented Kalman filter will iterate these two moments forward; to retrieve the moments of the latent variable  $Y_t$ , one has to extract the corresponding first rows of  $\boldsymbol{\mu}_{t|s}^a$ , and the top left submatrix of  $\mathbf{V}_{t|s}^a$ , respectively.

The augmented state vector moments are initialized with

$$\boldsymbol{\mu}_{0|0}^a = \begin{pmatrix} \boldsymbol{\mu}_{0|0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \text{ and } \mathbf{V}_{0|0}^a = \begin{pmatrix} \mathbf{V}_{0|0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\epsilon & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_\eta \end{pmatrix}$$

The unscented Kalman filter will iterate through the subsequent times  $t = 1, 2, \dots, T$ , applying a prediction and correction step.

For the prediction step we compute  $M = 2N^a + 1$  sigma points, based on columns of the Cholesky decomposition of the covariance matrix  $\mathbf{V}_{t-1|t-1}^a$  (which we denote here with a square root). The points are computed by horizontally concatenating as follows

$$\boldsymbol{\varsigma}_{t-1|t-1}^a = \left( \boldsymbol{\mu}_{t-1|t-1}^a; \boldsymbol{\mu}_{t-1|t-1}^a + \gamma \sqrt{\mathbf{V}_{t-1|t-1}^a}; \boldsymbol{\mu}_{t-1|t-1}^a - \gamma \sqrt{\mathbf{V}_{t-1|t-1}^a} \right)$$

This will be a  $(N^a \times M)$  matrix. The idea behind this computation is that the  $M$ -point sample that is produced by the columns of  $\boldsymbol{\varsigma}_{t-1|t-1}^a$  exhibits mean and covariance of  $\boldsymbol{\mu}_{t-1|t-1}^a$  and  $\mathbf{V}_{t-1|t-1}^a$ , respectively. It can be viewed as a minimal Monte-Carlo simulation of a sample with given first two moments. Thus we can view the sigma points as

$$\boldsymbol{\varsigma}_{t-1|t-1}^a = \begin{pmatrix} \boldsymbol{\varsigma}_{t-1|t-1}^{[Y,1]} & \boldsymbol{\varsigma}_{t-1|t-1}^{[Y,2]} & \cdots & \boldsymbol{\varsigma}_{t-1|t-1}^{[Y,M]} \\ \boldsymbol{\varsigma}_{t-1|t-1}^{[\epsilon,1]} & \boldsymbol{\varsigma}_{t-1|t-1}^{[\epsilon,2]} & \cdots & \boldsymbol{\varsigma}_{t-1|t-1}^{[\epsilon,M]} \\ \boldsymbol{\varsigma}_{t-1|t-1}^{[\eta,1]} & \boldsymbol{\varsigma}_{t-1|t-1}^{[\eta,2]} & \cdots & \boldsymbol{\varsigma}_{t-1|t-1}^{[\eta,M]} \end{pmatrix}$$

a concatenation of samples from the state variable and the error terms, given the information at time  $t-1$ .

Based on these sigma points one can produce now the predictions for the state variable and its covariance, by taking the sample moments of the function  $f$ , applied at the sigma points

$$\begin{aligned} \boldsymbol{\varsigma}_{t|t-1}^{[Y,m]} &= f(\boldsymbol{\varsigma}_{t-1|t-1}^{[Y,m]}, \boldsymbol{\varsigma}_{t-1|t-1}^{[\eta,m]}) \text{ for } m = 1, 2, \dots, M \\ \boldsymbol{\mu}_{t|t-1} &= \frac{1}{M} \sum_{m=1}^M \boldsymbol{\varsigma}_{t|t-1}^{[Y,m]} \\ \mathbf{V}_{t|t-1} &= \frac{1}{M} \sum_{m=1}^M [\boldsymbol{\varsigma}_{t|t-1}^{[Y,m]} - \boldsymbol{\mu}_{t|t-1}] [\boldsymbol{\varsigma}_{t|t-1}^{[Y,m]} - \boldsymbol{\mu}_{t|t-1}]' \end{aligned}$$

This completes the prediction step.

Applying the function  $h$  at the sigma points will produce the forecast for the observation at time  $t$ . Using these forecasts allows us to form the innovations at the forecasted sigma points, and thus the covariance matrices  $\langle X_t, Y_t | \mathcal{F}_{t-1} \rangle$  and  $\langle W_t, W_t | \mathcal{F}_{t-1} \rangle$ . In particular

LISTING 5.4: unscented\_filter.m: The unscented Kalman filter

```
%unscented_filter.m
function [Mt, St] = unscented_filter(p, data)
f = p.f; % measurement function
h = p.h; % transition function
5 Se = p.Sigma_epsilon; % measurement error covariance
Sh = p.Sigma_eta; % transition error covariance
M0 = p.M0; V0 = p.V0; % initial moments

Ny = size(M0,1); Ne = size(Se,1);
10 Nh = size(Sh,1); M = Ny+Ne+Nh;
W = [0.5/M*ones(2*M,1)];
W = diag(W);
Ma = [M0; zeros(Ne,1); zeros(Nh,1)];
15 Va = [V0 zeros(Ny,Ne) zeros(Ny,Nh)
zeros(Ne,Ny) Se zeros(Ne,Nh)
zeros(Nh,Ny) zeros(Nh,Ne) Sh];
Mt = []; St = [];
for t = 1:length(data)
    % construct sigma points
20 C = sqrt(M)*chol(Va);
SP = repmat(Ma,[1, 2*M]) + [C - C];
SPy = SP(1:Ny,:);
SPe = SP(Ny+1:Ny+Nh,:);
SPh = SP(Ny+Nh+1:end,:);
25 % projection
SPyP = f(SPy, SPe);
My = repmat(sum(SPyp*W, 2), [1,2*M]);
Vyy = (SPyP-My)*W*(SPyP-My)';
% correction
30 SPxP = h(SPyp, SPh);
SPwP = data(t) - SPxP;
Mx = repmat(sum(SPxP*W, 2), [1,2*M]);
Mw = repmat(sum(SPwP*W, 2), [1,2*M]);
Vxy = (SPyP-My)*W*(SPxP-Mx)';
Vww = (SPwP-Mw)*W*(SPwP-Mw)';
35 G = Vxy*inv(Vww);
Ma(1:Ny) = My(:,1) + G*Mw(:,1);
Va(1:Ny,1:Ny) = Vyy - G*Vww*G';
Mt = [Mt; Ma(1:Ny)'];
40 St = [St; sqrt(diag(Va(1:Ny,1:Ny))))'];
end
```

for copies, comments, help etc. visit <http://www.theponytail.net/>

$$\begin{aligned}
 \zeta_{t|t-1}^{[X,m]} &= h(\zeta_{t|t-1}^{[Y,m]}, \zeta_{t-1|t-1}^{[\epsilon,m]}) \text{ for } m = 1, 2, \dots, M \\
 \zeta_{t|t-1}^{[W,m]} &= x_t - \zeta_{t|t-1}^{[X,m]} \\
 \mu_{t|t-1}^X &= \frac{1}{M} \sum_{m=1}^M \zeta_{t|t-1}^{[X,m]} \\
 \mu_{t|t-1}^W &= \frac{1}{M} \sum_{m=1}^M \zeta_{t|t-1}^{[W,m]} \\
 \langle X_t, Y_t | \mathcal{F}_{t-1} \rangle &= \frac{1}{M} \sum_{m=1}^M [\zeta_{t|t-1}^{[X,m]} - \mu_{t|t-1}^X] [\zeta_{t|t-1}^{[Y,m]} - \mu_{t|t-1}]' \\
 \langle W_t, Y_t | \mathcal{F}_{t-1} \rangle &= \frac{1}{M} \sum_{m=1}^M [\zeta_{t|t-1}^{[W,m]} - \mu_{t|t-1}^W] [\zeta_{t|t-1}^{[Y,m]} - \mu_{t|t-1}]'
 \end{aligned}$$

Based on these quantities we can now update the mean and covariance of the latent variable, incorporating the new information as follows

$$\begin{aligned}
 \mathbf{K}_t &= \langle X_t, Y_t | \mathcal{F}_{t-1} \rangle \langle W_t, Y_t | \mathcal{F}_{t-1} \rangle^{-1} \\
 \mu_{t|t} &= \mu_{t|t-1} + \mathbf{K}_t w_t \\
 \mathbf{V}_{t|t} &= \mathbf{V}_{t|t-1} - \mathbf{K}_t \langle W_t, Y_t | \mathcal{F}_{t-1} \rangle \mathbf{K}_t'
 \end{aligned}$$

This completes the correction step and the recursion algorithm.

## Volatility

In this chapter we will investigate the modeling of volatility, and its implications on derivative pricing. We will start with some stylized facts of the historical and implied volatility, which will benchmark any forecasting or pricing methodology. We will then give an overview of *Garch*-type volatility filters and discuss how the parameters can be estimated using maximum likelihood. We will see that although Garch filters do a very good job in filtering and forecasting volatility, they fall somewhat short in the derivatives pricing arena. These shortcomings stem from the fact that Garch, by construction, is set up in discrete time, while modern pricing theory is set up under continuous time assumptions.

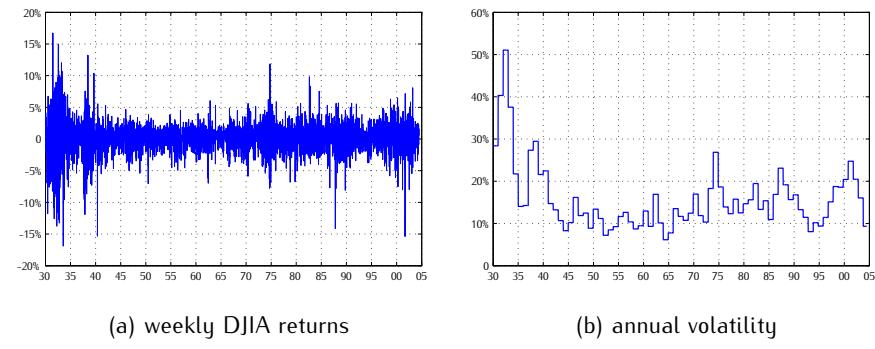
Two families of volatility models will be introduced for pricing and hedging. *Stochastic volatility* models extend the Black-Scholes methodology by introducing an extra diffusion that models volatility. *Local volatility* models, on the other hand, take a different point of view, and make volatility a non-linear function of time and the underlying asset. Of course each approach has some benefits but also some limitations, and for that reason we contrast and compare these methods.

It is important to note that this chapter deals exclusively with equity volatility, and to some extend exchange rate volatility. These processes are typically represented using some variants of random walk models. Fixed income securities models, and their volatility structures, will be covered in a later chapter.

### 6.1 SOME GENERAL FEATURES

This first section will cover some stylized features of volatility. We will differentiate between historical and implied volatilities. Although the qualitative properties of these two are similar, their quantitative aspects might differ substantially, as they are specified under two different (but nevertheless equivalent) probability measures.

FIGURE 6.1: Dow Jones industrial average (DJIA) weekly returns and yearly historical volatility. The (annualized) volatility is computed over non-overlapping 52 week periods from the beginning of 1930 to 2005.



## HISTORICAL VOLATILITY

Volatility in financial markets varies over time. This is one of the most documented stylized facts of asset prices. For example, figure 6.1(a) gives a very long series of weekly returns<sup>1</sup> on the Dow Jones industrial average index (DJIA, or just “the Dow”). Subfigure 6.1(b) presents the (annualized) standard deviation of consecutive and non-overlapping 52-week intervals, a proxy of the realized DJIA volatility over yearly periods. One can readily observe this time variability of the realized volatility, and in fact we can easily associate it with distinct events, like the Great Depression (early 30s), the Second World War (late 30s/early 40s), the Oil Crisis (mid 70s), and the Russian Crisis (late 90s).

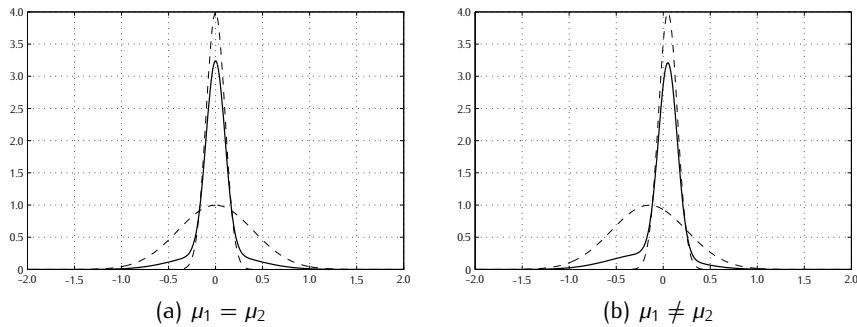
If we compute the summary statistics of the DJIA returns, we will find that the unconditional distribution exhibits fat tails (high kurtosis). In particular, the kurtosis of this sample is  $k = 8.61$ . The variability of volatility can cause fat tails in the *unconditional* distribution, even if the *conditional* returns are normally distributed. To illustrate this point, consider a simple example where the volatility can take only two values,  $\sigma_t = \sigma_1 = 10\%$  or  $\sigma_t = \sigma_2 = 40\%$ , and both means are zero. Say that we denote with  $f_N(x; \mu, \sigma)$  the corresponding normal probability density functions.

Also, suppose that  $p_1 = 75\%$  of the time returns are drawn from a normal<sup>2</sup>  $r \propto f_N(r; 0, \sigma_1)$ , and in the other  $p_2 = 25\%$  of the time they are drawn from a second normal  $r \propto f_N(r; 0, \sigma_2)$ . If we consider the unconditional distribution, its probability density function will be a *mixture* of the two normal distributions, and in fact

<sup>1</sup> Here by returns we actually mean log-returns, that is if  $S_t$  is the time-series of DJIA values,  $r_t = \log S_{t-1} - \log S_t$ .

<sup>2</sup> Here the notation  $x \propto f(x; \dots)$  means that  $x$  is distributed as a random variable that has a probability density function given by  $f(x; \dots)$ .

FIGURE 6.2: This figure illustrates the different kurtosis and skewness patterns that can be generated by mixing two normal distributions. In both figures  $\sigma_1 = 10\%$  and  $\sigma_2 = 40\%$ . In subfigure (a) the two means are equal  $\mu_1 = \mu_2 = 0$ , a setting that can generate fat tails but not skewness. In subfigure (b)  $\mu_1 = 5\%$  and  $\mu_2 = -15\%$ , generating negative skewness in addition to the fat tails.



$$r \propto p_1 f_N(r; 0, \sigma_1) + p_2 f_N(r; 0, \sigma_2)$$

Figure 6.2(a) illustrates exactly this point, and gives the two conditional normals and the unconditional distribution. One can easily compute the statistics for the unconditional returns, and in particular the unconditional volatility  $\sigma = 21.7\%$ , and the kurtosis  $k = 8.7 > 3$ .

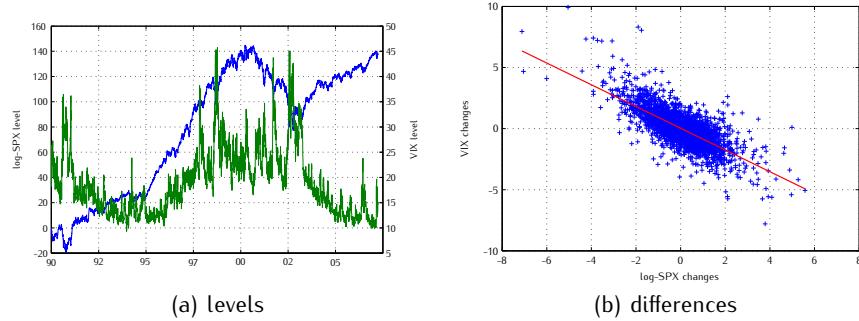
Inspecting figure 6.1(b), one can also observe that the historical realized volatility does not swing wildly, but exhibits a cyclical pattern. In particular, it appears that volatility exhibits *high autocorrelation*, with low (high) volatility periods more likely to be followed by more low (high) volatility periods. In the literature these patterns are often described as *volatility clusters*. Having said that, the volatility process appears to be *stationary*, in the sense that it remains between bounds, an intuitive feature.<sup>3</sup> We can imagine that there is some *long run volatility* that serves as an attractor, with the spot volatility hovering around this level.

### IMPLIED VOLATILITY

In chapter 2 we gave a quick introduction to the notion of the *implied volatility* (IV), denoted with  $\hat{\sigma}$ . In particular, given an observed European call or put option price  $P_{\text{obs}}$ , the IV will equate it to the theoretical Black-Scholes value, solving the equation

<sup>3</sup> The intuition stems from the fact that, unlike prices themselves, market volatility can not increase without bounds. Even if we are asked to provide some estimate for the volatility of DJIA in 1,000 years, we would probably come up with a value that reflects current volatility bounds. If we are asked to estimate the level of DJIA in 1,000 years' time, we would produce a very large number.

FIGURE 6.3: The S&P500 index (SPX, in blue) and the implied volatility index (VIX, in green) are given in subfigure (a). Subfigure (b) presents a scatterplot of the corresponding differences, illustrating the return/volatility correlation.



$$\hat{\sigma} \in \mathbb{R}_+ : P_{\text{obs}} = f_{\text{BS}}(S, K, T, r, \hat{\sigma})$$

We also showed that short at-the-money IV reflects expected volatility (under the equivalent martingale measure that prices the corresponding option).

As expected, implied volatility shows similar patterns to the historical realized volatility. In particular, time series of IV for a particular contract with fixed maturity exhibit an autoregressive structure and clusters. Figure 6.3(a) gives the S&P500 index as well as the implied volatility index VIX, released by the Chicago Board Options Exchange (CBOE). The VIX is computed as a weighted average of option prices that bracket 30 days to maturity, with more weight given to options that are at-the-money.<sup>4</sup> These options are written on the S&P500 index, and the data span a period from January 1990 (when the VIX index was first released) up to April 2007. The VIX has been coined as investors' "fear gauge", and figure 6.3(a) certainly illustrates that. Just like the realized volatility (discussed in the previous subsection), the VIX increases in periods where significant events cause the market to go into turmoil. We can clearly see the first Gulf war (8/90-2/91), the East Asian crisis (5-8/97), the Russian crisis and the collapse of Long-Term Capital Management (5-9/98), the 9/11 attacks (11/01), and the buildup to the invasion of Iraq (3/03). In all these episodes the market level declined.

Based on our catalogue of high volatility episodes that we devised above (using the VIX or the historical realized volatility), it is apparent that they were accompanied by periods of low or negative returns. Each one of these clusters is a chapter of market turmoil. This suggests that there might be some negative correlation between the market returns and their contemporaneous volatility. Periods of high volatility are accompanied by low returns, while returns are higher when volatility is low. These two market *regimes* reflect the *bad* and

<sup>4</sup> The step-by-step construction of the VIX index is given in the White Paper CBOE (2003).

*good times* in the market.<sup>5</sup> It is easy to investigate the validity of this claim, by a simple scatterplot of the realized volatility against market returns, as in figure 6.3(b). The negative relationship is apparent, and is verified by a simple regression that indicates a relationship of the form  $\Delta\hat{\sigma} = 0.03\% - 0.89 \cdot \Delta\hat{\mu}$ . This indicates that a 1% drop in the market index is accompanied (on average) by a 0.89% rise of the market volatility.<sup>6</sup>

This negative correlation is often coined the *leverage effect*, as it can be theoretically explained by the degree of leverage that underlies the capital structure of the firm. In particular, as the firm value is the sum of debt and equity, a shock to the value of the firm will have an impact on the stock price that depends on the leverage. If the firm has been financed by issuing stock alone, then a 1% increase in the firm value will result in a 1% in the stock price; on the other hand, if the firm is levered, the impact on the stock price can be a lot higher than 1%, depending on the leverage.<sup>7</sup> Thus, higher leverage will produce higher stock price volatility. In addition, a negative stock price shock will increase leverage, implying that negative returns imply higher volatility, and hence the negative correlation. Early research (for example Christie, 1982) indicate that there is indeed a relationship between this correlation and the balance sheet, but more recent evidence indicates that this effect cannot really explain the magnitude of the asymmetry that is observed or implied from options markets (Figlewski and Wang, 2000). It appears that this negative relationship might be better attributed to the erratic behavior of market participants during market downturns.

Whatever its reasons, this negative relationship between asset returns and their volatilities manifests itself as negative skewness in the unconditional distribution. Coming back to our toy model that we used to investigate kurtosis, assume that now returns can come from two normals with parameters  $(\mu_1, \sigma_1) = (5\%, 10\%)$  (the good times), or  $(\mu_2, \sigma_2) = (-15\%, 40\%)$  (the bad times). Figure 6.2(b) presents the unconditional distribution for this setting. Straightforward calculations can reveal that the skewness of the unconditional returns is  $s = -1.37$ , while kurtosis is similar,  $k = 8.28$ .

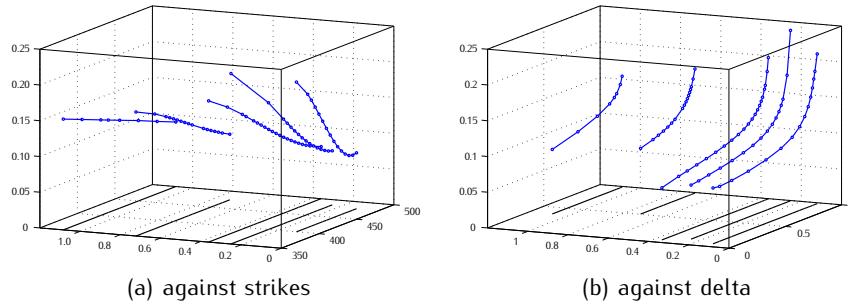
---

<sup>5</sup> Also found in the literature as *bust* and *boom*, *bear* and *bull*, or *recession* and *expansion*, depending on the journal or publication one is reading.

<sup>6</sup> Of course this is a very crude method. Bouchaud and Potters (2001) give a formal empirical investigation based on a large number of stocks and indices, and find that this negative correlation is more pronounced for indices, but more persistent for individual stocks.

<sup>7</sup> We follow here the standard Modigliani and Miller (1958) capital structure approach, where stocks and bonds represent different ways of splitting firm ownership. Say that a company is worth  $\$100m$ , with  $\$10m$  in stock and the rest ( $\$90m$ ) in bonds. If the value of the firm increases by 1% up to  $\$101m$ , the value of the stock will increase to  $\$11m$  to reflect that increase (since the debt value cannot change). This will imply a 10% rise in the stock price.

FIGURE 6.4: Implied volatilities  $\hat{\sigma}$ , plotted against maturity  $T$  and either strike prices  $K$  (subfigure a) or against the corresponding Deltas  $\Delta = N(\log(F/K)/(\hat{\sigma}\sqrt{T}) + \hat{\sigma}\sqrt{T}/2)$  (subfigure b).



### The implied volatility surface

Given an underlying asset, at any point in time there will exist a number of options, spanning a range of strike prices for different times to maturity. Each one of these options can be inverted to deliver an implied volatility  $\hat{\sigma}(K, T)$ . A three-dimensional scatterplot of these implied volatilities gives the *implied volatility surface*, which has some distinct and very interesting features. Such a scatter plot is given in figure 6.4(a), for options computed on the S&P500 index.

One can readily observe an *implied volatility skew* for each maturity level. IV is higher for small strikes, which correspond to in-the-money calls or out-of-the-money puts, and declines as we move towards higher strikes. Another observation is that this skew is more pronounced for options with shorter maturities, and flattens out for long dated options. The monotonic relation between the BS price and volatility indicates that out-of-the-money puts appear to carry a higher premium than the corresponding out-of-the-money calls. More recently, the volatility skew is presented against other forms of moneyness that remove some of the maturity effect. Figure 6.4(b) gives such an example, where the same IV surface is re-parameterized with respect to the Delta of the appropriate call. In that case at-the-money contracts will be mapped to a Delta of  $\Delta = 1/2$ .

If the BS model was the correct one, that is to say if log-returns were normally distributed with constant volatility, IV surfaces would be flat. The shape of the IV surface can point towards these deviations from normality, and in fact it can reveal the risk-neutral distribution that is consistent with the observed implied volatilities. In particular, across moneyness the skew pattern we outline above is typical of index options, and to some extend stock options. Currency options can also exhibit a U-shaped pattern of implied volatilities, coined the *volatility smile*. Such a pattern was also encountered in stock and index options before the 1987 stock market crash (see Rubinstein, 1985, 1994; Jackwerth and Rubinstein, 1996, for details). A volatility smile will be consistent with a distribution that exhibits fat tails, since in that case it would be more likely to

exercise out-of-the-money puts or calls. To reproduce a volatility skew, one will need a distribution that is not only leptokurtic but also skewed.

The volatility surface itself is not stable across time. The dynamics of the surface are investigated in [Skiadopoulos, Hodges, and Cleow \(2000\)](#), and more recently in [Cont and da Fonseca \(2002\)](#). Assumptions on these dynamics are going to affect the Delta hedging schemes that can be employed. [Derman \(1999\)](#) discusses such hedging rules, namely the *sticky strike*, the *sticky Delta* or the *sticky local volatility* strategy.

One challenge is to construct a theoretical model that can replicate the shape and the dynamics of the IV surface.

## TWO MODELING APPROACHES

To model the time varying nature of the asset return volatility, typically one has to choose between a Garch and a SV approach. Each one has its benefits, but also some shortfalls and peculiarities. Generally speaking, the Garch family is more suited for historical estimation and risk management purposes, while the stochastic volatility is better adapted towards derivative pricing and hedging. The following table gives a quick comparison of the two families. In the next sections we will give more details.

	Garch	SV
current volatility	known	unknown
conditional volatility	computable	unknown
volatility randomness	no extra source	extra source
volatility price of risk	set internally	set externally
time frame	discrete	continuous
incompleteness	discrete time	extra diffusions
option pricing	very limited	available
historical calibration	maximum likelihood	hard
calibration to options	hard	transforms

## 6.2 AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY

In the previous section we pointed out that a mixture of normal distribution has the potential to produce distributions that exhibit skewness and excess kurtosis. Also, by investigating historical realized returns and implied volatilities, we concluded that market volatility is time varying and cyclical. Autoregressive conditional heteroscedasticity models build exactly on these points. The definitive reference is [Hamilton \(1994\)](#).

Assume a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , and say that we are interested in modeling a series of returns  $r_t = r_t(\omega)$  for  $t = 1, \dots, T$  and  $\omega \in \Omega$ . The

information that is gathered up to period  $t$  is represented by the filtration  $\mathcal{F}_t = \sigma(r_s : 0 \leq s \leq t)$ . The conditional distribution is normal

$$r_t | \mathcal{F}_{t-1} \sim f_N(r_t; \mu_t, \sigma_t)$$

but having a different volatility  $\sigma_t$ , and possibly a different mean  $\mu_t$ . This volatility is updated using a mechanism that ensures that at each period  $t - 1$  we can ascertain the parameters of next period's returns,  $\sigma_t$  and  $\mu_t$ , based on past returns alone. In probability jargon we say that both  $\sigma_t$  and  $\mu_t$  are  $\mathcal{F}_{t-1}$ -adapted.

### THE ARCH MODEL

Engle (1982) set up a process which he coined Arch(1), standing for autoregressive conditional heteroscedasticity of order one. In particular

$$\begin{aligned} r_t &= \mu + \epsilon_t \\ \epsilon_t &\sim N(0, h_t) \\ h_t &= \omega + \gamma \epsilon_{t-1}^2 \end{aligned}$$

In this model the conditional return is indeed normally distributed,  $r_t | \mathcal{F}_{t-1} \sim f_N(r_t; \mu, \sqrt{h_t})$ , and the volatility is  $\mathcal{F}_{t-1}$ -adapted since it is a function of  $\epsilon_{t-1} = r_{t-1} - \mu$  which is known at time  $t - 1$ . Also, if the volatility at time  $t - 1$  is large, then it will be more likely to draw a large (in absolute terms)  $\epsilon_t$ . Therefore an Arch(1) will exhibit some autocorrelation in the volatility. In order to ensure that the volatility is positive we need to impose the restrictions  $\omega, \gamma \geq 0$ .

We can write volatility forecasts  $h_{t+s|t} = E[\epsilon_{t+s}^2 | \mathcal{F}_t] = E_t \epsilon_{t+s}^2$  by backward substitution as

$$h_{t+s|t} = E_t \epsilon_{t+s}^2 = \omega + \gamma E_t \epsilon_{t+s-1}^2 = \omega + \gamma h_{t+s-1|t}$$

which yields the forecasts (using also  $h_{t+1|t} = h_{t+1}$  which is known at time  $t$ )

$$\begin{aligned} h_{t+s|t} &= \omega(1 + \gamma + \cdots + \gamma^{s-2}) + \gamma^{s-1} h_{t+1|t} \\ &= \omega \frac{1 - \gamma^{s-1}}{1 - \gamma} + \gamma^{s-1} h_{t+1} = \frac{\omega}{1 - \gamma} + \gamma^{s-1} \left( h_{t+1} - \frac{\omega}{1 - \gamma} \right) \end{aligned}$$

The above expression also indicates that the constraint  $\gamma \leq 1$  is needed to ensure that the volatility process is not explosive. In that case, the long run expectation for the volatility is  $h^* = \frac{\omega}{1-\gamma}$ . The expected *integrated variance*,  $H_{t,s} = E[\sum_{k=1}^s \epsilon_{t+k|t}^2 | \mathcal{F}_t] = \sum_{k=1}^s h_{t+k|t}$  will be given by

$$H_{t,s} = (s-1) \frac{\omega}{1-\gamma} + \frac{\gamma - \gamma^s}{1-\gamma} \left( h_{t+1} - \frac{\omega}{1-\gamma} \right)$$

The Arch(1) model can be easily extended to one of order  $p$  (an Arch( $p$ ) model), by allowing the variance to depend on more lagged values of  $\epsilon$

$$h_t = \omega + \gamma_1 \epsilon_{t-1}^2 + \gamma_2 \epsilon_{t-2}^2 + \cdots + \gamma_p \epsilon_{t-p}^2$$

For this process to avoid explosive volatility we need the constraint  $\sum \gamma_n \leq 1$ . The Arch( $\infty$ ) is the natural extension where the whole history of error terms affects our volatility forecast. Actually, early research on Arch models indicated that a large number of lags are required to capture the dynamics of asset volatility, pointing towards some Arch( $\infty$ ) structure. This gave eventually rise to the Garch extension.

### THE GARCH MODEL

The Garch model (generalized Arch) of [Bollerslev \(1986\)](#) extends the Arch family by adding dependence on past variances. For example, the popular Garch(1,1) specifies

$$\begin{aligned} r_t &= \mu + \epsilon_t \\ \epsilon_t &\sim N(0, h_t) \\ h_t &= \omega + \beta h_{t-1} + \gamma \epsilon_{t-1}^2 \end{aligned}$$

The additional constraint  $\beta \geq 0$  is sufficient to keep the variance positive. This seemingly small addition is equivalent to an Arch( $\infty$ ) structure, which is clear if we back-substitute the conditional variances which yields for  $s$  lags

$$h_t = \omega \frac{1 - \beta^s}{1 - \beta} + \beta^s h_{t-s} + \gamma \epsilon_{t-1}^2 + \gamma \beta \epsilon_{t-2}^2 + \cdots + \gamma \beta^{s-1} \epsilon_{t-s}^2$$

If  $\beta \leq 1$ , then we can let  $s \rightarrow \infty$ , giving the Arch( $\infty$ ) form of the Garch(1,1) model

$$h_t = \frac{\omega}{1 - \beta} + \gamma \epsilon_{t-1}^2 + \gamma \beta \epsilon_{t-2}^2 + \gamma \beta^2 \epsilon_{t-3}^2 + \cdots$$

The impact of lagged errors decays exponentially as we move further back in the past of the series. The Garch(1,1) model has been extremely popular amongst econometricians and practitioners that need to either filter or forecast volatility. The natural generalization Garch( $p,q$ ) includes  $p$  lags of the squared error terms and  $q$  lagged variances.

Once again we can derive the volatility forecasts using forward substitution, in particular

$$h_{t+s|t} = E_t \epsilon_{t+s}^2 = \omega + \beta E_t h_{t+s-1} + \gamma E_t \epsilon_{t+s-1}^2 = \omega + (\beta + \gamma) h_{t+s-1|t}$$

which is the same form we encountered in the Arch case for  $\gamma \rightsquigarrow \beta + \gamma$ . Therefore we can compute forecasts for the variance and the integrated variance if we denote  $\kappa = \beta + \gamma$  (the so called *persistence parameter*)

$$\begin{aligned} h_{t+s|t} &= \frac{\omega}{1 - \kappa} + \kappa^{s-1} \left( h_{t+1} - \frac{\omega}{1 - \kappa} \right) \\ H_{t,s} &= (s-1) \frac{\omega}{1 - \kappa} + \frac{\kappa - \kappa^s}{1 - \kappa} \left( h_{t+1} - \frac{\omega}{1 - \kappa} \right) \end{aligned}$$

The long run (or unconditional) variance is now given by  $h^* = \frac{\omega}{1-\beta-\gamma}$ . In order for the variance to remain well defined we need to impose the constraint  $\beta + \gamma \leq 0$ .

## THE GARCH LIKELIHOOD

In order to use a Garch model we need to know the parameters of the process, namely  $\{\mu, \beta, \gamma, \omega\}$ . We can estimate these parameters based on a time series of historical returns  $r = \{r_1, \dots, r_T\}$ . If we denote with  $f_{t-1}(r_t) = \mathbb{P}[r_t \in dr | \mathcal{F}_{t-1}]$  the conditional density, then the *likelihood* of the sample is given by the product  $\mathcal{L}(r) = \prod_{t=1}^T f_{t-1}(r_t)$ . We are usually employ the logarithm of this expression, the *log-likelihood*

$$\log \mathcal{L}(r) = \sum_{t=1}^T \log f_{t-1}(r_t)$$

The fact that *conditionally* the random variables  $r_t | \mathcal{F}_{t-1}$  are normally distributed, allows one to compute the likelihood for a given set of parameters  $\vartheta = \{\mu, \omega, \beta, \gamma\}$ . Often we set the long run variance  $h^*$  equal to the sample variance  $\bar{\sigma}^2$ , and therefore set  $\omega = \bar{\sigma}^2(1 - \beta - \gamma)$ . This makes sense if our sample is fairly long, and can significantly help the numerical optimization algorithm. In that case the parameter vector to be estimated is  $\vartheta = \{\mu, \beta, \gamma\}$ . In order to start the recursive algorithm that computes the Garch variance we also need an initial value for  $h_0$ . We can also use  $h_0 = \bar{\sigma}^2$ , or we can add  $h_0$  to the parameter vector and let it be estimated.

In the Garch process we defined above, the parameter  $\mu$  is not the expected rate of return. In particular, as the asset price is lognormally distributed,  $S_t = S_{t-1} \exp(r_t)$ , the expected return is  $E_{t-1} S_t = S_{t+1} \exp(\mu - \frac{1}{2}h_t)$ . Therefore, if we want  $\mu$  to denote the constant expected return, then we need to set up the Garch equation as

$$r_t = \mu - \frac{1}{2}h_t + \epsilon_t, \quad \epsilon_t \sim N(0, h_t)$$

The next steps, implemented in listing 6.1, show how the likelihood can be computed for a given set of parameters  $\vartheta$  and a sample  $r = \{r_t\}$ . The popularity of the Garch model stems from the fact that this likelihood is computed rapidly and can be easily and quickly maximized. The ideas behind maximum likelihood estimation were covered in detail in chapter 5.

1. If they are not part of  $\vartheta$ , we set the parameters  $\omega = \bar{\sigma}^2(1 - \beta - \gamma)$  and  $h_0 = \bar{\sigma}^2$ .
2. Based on the parameters  $\{\mu, \omega, \beta, \gamma\}$  and the initial value  $h_0$ , we filter the volatility series, applying the Garch(1,1) recursion

$$\begin{aligned} \epsilon_t &= r_t - \mu + \frac{1}{2}h_t \\ h_{t+1} &= \omega + \beta h_t + \gamma \epsilon_t^2 \end{aligned}$$

LISTING 6.1: garch11\_lik.m: Garch likelihood function.

```
% garch11_lik.m
function [lik, V] = garch11_lik(par, data)
vBar = var(data);
m = par(1); % mu
5 b = par(2); % beta
c = par(3); % gamma
a = vBar*(1-b-c); % omega (restricted)
N = size(data);
V = zeros(N+1);
10 spotV = vBar; % current variance
V(1) = spotV; % initial variance
lik = 0.0; % the likelihood
for indx=1:N
    % epsilon
    error = data(indx) - m + 0.5*spotV;
    % update likelihood
    lik = lik + (error^2)/spotV + log(spotV);
    % update conditional variance
    spotV = a + b*spotV + c*error^2;
    % store conditional variance
    20 V(indx+1) = spotV;
end
% discard last variance forecast
V = V(1:N)
```

3. Now we have the variance series which allows us to compute the log-likelihood of each observation  $r_t$ . Since  $r_t|\mathcal{F}_{t-1} \sim N(\mu, h_t)$

$$\log \mathcal{L}(r_t|\vartheta) = -\frac{(r_t - \mu)^2}{2h_t} - \frac{1}{2} \log h_t - \frac{1}{2} \log 2\pi$$

4. Finally adding up will give the log-likelihood of the sample

$$\log \mathcal{L}(r|\vartheta) = \sum_{t=1}^T \log \mathcal{L}(r_t|\vartheta)$$

The maximization of the log-likelihood is typically numerically, using a hill climbing algorithm. Press, Flannery, Teukolsky, and Vetterling (1992) describe a number of such algorithms. We will denote with  $\hat{\vartheta}$  the parameter vector that maximizes the sample log-likelihood. Essentially the first order conditions set the Jacobian equal to zero

$$\left. \frac{\partial \log \mathcal{L}(r|\vartheta)}{\partial \vartheta} \right|_{\vartheta=\hat{\vartheta}} = 0$$

for copies, comments, help etc. visit <http://www.theponytail.net/>

The Hessian matrix of second derivatives can help us produce the *asymptotic* standard errors

$$\hat{\mathbf{H}} = \frac{\partial^2 \log \mathcal{L}(r|\vartheta)}{\partial \vartheta' \partial} \Big|_{\vartheta=\hat{\vartheta}}$$

The covariance matrix of  $\hat{\vartheta}$  is given by the inverse of the Hessian (Hamilton, 1994, gives methods to estimate  $\mathbf{H}$ ).

### Estimation examples

As an example, we will estimate two time series using the Garch(1,1) process for the volatility. We start with the long DJIA index sampled weekly from 1930 to 2004 (plotted in figure 6.1(a)), and then move to the shorter SPX index sampled daily from 1990 to mid-2007 (plotted in figure 6.3). Listing 6.2 shows how the log-likelihood can be optimized.

The estimation is done using the Optimization Toolbox in Matlab, although any hill climbing algorithm will do in that simple case. We use constrained optimization to ensure that  $\beta$  and  $\gamma$  are bounded between zero and one. Also we want to ensure that  $\kappa = \beta + \gamma < 1$ . The standard errors are produced using the Hessian matrix that is estimated by the toolbox.<sup>8</sup> We also use the restriction on the long run variance, and set the initial variance equal to the sample variance. The maximum likelihood parameters are given below (all in percentage terms), with standard errors in parentheses.

	DJIA	SPX
$\mu$	0.19 (0.02)	0.05 (0.01)
$\beta$	91.32 (0.94)	93.93 (0.53)
$\gamma$	7.66 (0.76)	5.42 (0.45)
$\kappa = \beta + \gamma$	98.98	99.45

Both times give similar estimated values. If we write the error term  $\epsilon_t = \sqrt{h_t} \eta_t$  for  $\eta_t \sim N(0, 1)$ , then  $\epsilon_t^2 = h_t \eta_t^2$  and since  $E\eta_t^2 = 1$  we can write  $\epsilon_t^2 = h_t(1+u_t)$  where now  $Eu_t = 0$  (but of course  $u_t$  is not normal). The the Garch(1,1) variance process can be cast in an autoregressive AR(1) form

$$h_t = \omega + \kappa h_{t-1} + \gamma h_t u_t$$

The importance of the coefficient  $\kappa$  becomes now apparent, as it will determine the decay of variance shocks. In both time series  $\kappa \approx 1$ , which indicates

---

<sup>8</sup> The optimization toolbox actually updates estimates of the Hessian and the output is not always reliable. Some care has to taken here, and the standard errors should be taken with a pinch of salt. Hamilton (1994) gives a number of superior methods such as the score, or outer product method, etc.

LISTING 6.2: garch11\_impl.m: Estimation of a Garch model.

```
% garch11_impl.m
% load data
%data = xlsread('DJIweekly.xls'); T0 = 52; % DJIA
data = xlsread('SPX_VIX.xls'); T0 = 252; % SPX
5 T = x2mdate(data(2:end,1)); % time
y = diff(log(data(:,2))); % log-returns
% set initial values
mm = mean(y); % mu
b = 0.7; % beta
10 c = 0.1; % gamma
par = [mm b c];
% constraints
parL = [-0.10 0.001 0.001]; % positive
parU = [ 0.10 0.999 0.999]; % less than one
15 cA = [0 1 1]; % beta+gamma<1
cB = [0.999];
% optimization options
optopt = optimset('MaxIter',100,'Display','iter',...
    LargeScale','off');
[x, fval, exitflag, output, lambda, jacob, hess] = ...
20 fmincon(@garch11_lik, par, cA, cB, [], [], ...
    parL, parU, [], optopt, y);
par = x; % estimates
stderrs = diag(sqrt(inv(hess))), % standard errors
[lik,V] = garch11_lik(par, y); % filter at estimates
25 % output variance time series
figure(1);
plot(T, 100*sqrt(V*T0));
grid on;
datetick('x','yy');
```

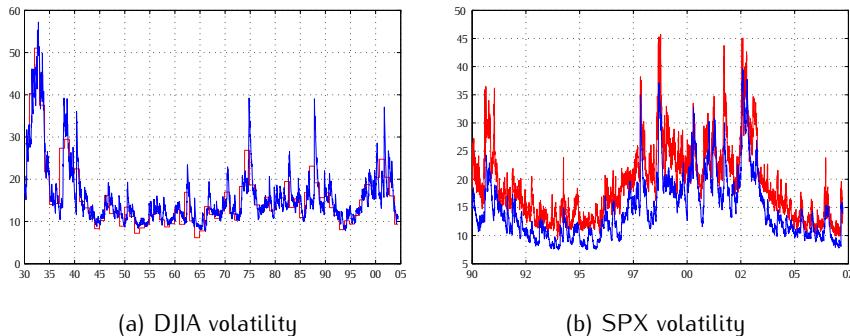
that volatility behaves as a near unit root process.<sup>9</sup> In such a process shocks to the volatility are near permanent, and the process is reverting very slowly towards the long run variance.<sup>10</sup>

These parameter estimates are typical of Garch estimations, and the near integrated behavior has been the topic of substantial research through the 80s and the 90s. A number of researchers introduced Garch variants that exhibit long memory, such as the fractionally integrated Garch (Figarch) of Baillie, Bollerslev,

<sup>9</sup> In fact, if we trust the standard errors we are not able to reject the hypothesis  $\kappa = 1$ .

<sup>10</sup> A Garch process with  $\beta + \gamma = 1$  is called integrated Garch (Igarch), and is equivalent to the exponentially weighted moving average (EWMA) specification, where the variance is updated as  $\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1 - \lambda)\epsilon_{t-1}^2$ . In this case the volatility behaves as a random walk.

FIGURE 6.5: Filtered volatility for the DJIA and the SPX index. In subfigure (a) the Garch variance (blue) of weekly DJIA returns is plotted with the historical realized volatility (red). In (b) the Garch variance (blue) of daily SPX returns is plotted with the implied volatility VIX index (red)



and Mikkelsen (1993). Others acknowledge that models with structural breaks in the variance process can exhibit spuriously high persistence (Lamoureux and Lastrapes, 1990), and produce models that exhibit large swings in the long run variance attractor (Hamilton and Susmel, 1994; Dueker, 1997).

Figure 6.5 gives the filtered volatility for both cases. This is a by-product of the likelihood evaluation. For comparison, the historical volatility (of figure 6.1) and the implied volatility VIX index (of figure 6.3) are also presented. The filtered volatilities are computed using the maximum likelihood parameter estimates. On point worth making is that the implied volatility overestimates the true volatility, illustrated in subfigure (b), where the VIX index is above the filtered volatility for most of time. This due to the fact that implied volatility can be thought as a volatility forecast under an equivalent martingale measure, rather than a true forecast. There will be different risk premiums embedded in the implied volatility, rendering it a biased estimator or forecast of the true volatility.

## OTHER EXTENSIONS

Apart from the simple Garch(1,1) model that we already presented, there have been scores of modifications and extensions, tailor made to fit the stylized facts of asset prices. We will give here a few useful alternatives.

In the standard Garch model we assumed that conditional returns are normally distributed, and write  $\epsilon_t = \sqrt{h_t} \eta_t$ , with  $\eta_t \sim N(0, 1)$ . The likelihood function was based on this assumption. It is straightforward to use another distribution for  $\eta_t$ ; if it has a density function that is known in closed form, then it is straightforward to modify the likelihood function appropriately. Of course it might be necessary to normalize the distribution to ensure that  $E\eta_t = 0$  and  $E\eta_t^2 = 1$ . A popular choice is the Student-t distribution which can accommodate

conditional fat tails. The density function of the Student-t distribution with  $v$  degrees of freedom is

$$f_t(x; v) = \frac{\Gamma((v+1)/2)}{\sqrt{\pi v} \Gamma(v/2)} \left(1 + \frac{x^2}{v}\right)^{-(v+1)/2}$$

As the t distribution has variance  $\frac{v}{v-2}$ , we can set the density of  $\eta_t$  equal to

$$\eta_t \propto \sqrt{\frac{v-2}{v}} f_t \left( \eta_t \sqrt{\frac{v-2}{v}}; v \right)$$

We can augment the parameter vector  $\vartheta$  with  $v$ , and the third step of the likelihood evaluation will now become

3\* Now we have the variance series which allows us to compute the log-likelihood of each observation

$$\begin{aligned} \log \mathcal{L}(r_t | \vartheta) &= \log \left( \frac{\sqrt{v-2} \Gamma((v+1)/2)}{v \sqrt{\pi} \Gamma(v/2)} \right) - \frac{1}{2} \log h_t \\ &\quad - \frac{v+1}{2} \log \left( 1 + \frac{(r_t - \mu)^2}{h_t} \frac{v-2}{v^2} \right) \end{aligned}$$

Garch models based on normal or t distributed errors do not exhibit skewness. Nelson (1991) considers the generalized error distribution (GED) which can potentially capture skewed errors. Having said that, this approach does not model the leverage effect directly. The GJR-Garch model, introduced in Glosten, Jagannathan, and Runkle (1993), uses a dummy variable to assume different impact of positive and negative news on the variance process. In particular

$$h_t = \omega + \beta h_{t-1} + \gamma \epsilon_{t-1}^2 + \gamma^* I(\epsilon_{t-1} \leq 0) \epsilon_{t-1}^2$$

The function  $I(x)$  is the indicator function. Therefore, if  $\gamma^* > 0$  a negative return will increase the conditional variance more than a positive one ( $\gamma + \gamma^*$  instead of  $\gamma$ ).<sup>11</sup> But even with the GJR approach we will not have the situation illustrated in figure 6.3(b), where positive returns will actually have a negative impact on the volatility.

The Egarch model of Nelson (1991) takes a more direct approach, as it uses raw rather than squared returns. This implies that the sign is not lost and will have an impact. In order to get around the non-negativity issue he models the logarithm of the variance

$$\log h_t = \omega + \beta \log h_{t-1} + \gamma (|\eta_t| + \theta \eta_t), \quad \eta_t = \frac{\epsilon_t}{\sqrt{h_t}}$$

---

<sup>11</sup> Other asymmetric extensions include the threshold model of Zakoian (1994) and the quadratic Garch of Sentana (1995).

LISTING 6.3: egarch11\_lik.m: Egarch likelihood function.

```
%egarch11_lik.m
function [lik, V] = egarch11_lik(par, data)
vBar = var(data);
m = par(1); % mu
5 a = par(2); % omega
b = par(3); % beta
c = par(4); % gamma
w = par(5); % theta
N = size(data);
10 V = zeros(N+1);
spotV = vBar; % current variance
V(1) = spotV; % initial variance
lik = 0.; % likelihood
for indx=1:N
15 spotV = spotV + eps;
% epsilon
error = data(indx) - m + 0.5*spotV;
% update likelihood
lik = lik + (error^2)/spotV + log(spotV);
20 % normalize for eta
eta = error/sqrt(spotV);
% update variance
logV = log(spotV);
logV = a + b*logV + c*(abs(eta) + w*eta);
25 spotV = exp(logV);
V(indx+1) = spotV;
end
V = V(1:N);
% if variance failed set a large value
30 % (perhaps due to absurd parameter values)
if isnan(lik)
lik = 1e8;
end
```

In the Egarch approach  $\gamma\theta < 0$  will be consistent with figure 6.3(b), as higher returns will lower volatility. Listing 6.3 shows an implementation of the Egarch likelihood function. As there are no constraints in the Egarch maximization, the hill climbing algorithm might attempt to compute the likelihood for absurd parameter values as it tries to find the optimum. There are a couple of tricks in the code that ensure that a likelihood value will be returned. The implementation for the optimization resembles listing 6.2, but we shall use unconstrained optimization. The maximum likelihood parameters are given below for the two time series

	DJIA	SPX
$\mu$	0.14% (0.02%)	0.04% (0.01%)
$\omega$	-0.2968 (0.0204)	-0.2612 (0.0128)
$\beta$	0.9785 (0.0021)	0.9826 (0.0013)
$\gamma$	0.1678 (0.0109)	0.1249 (0.0075)
$\theta$	-0.4093 (0.0402)	-0.6145 (0.0587)

As expected, the product  $\gamma\theta < 0$ , supporting the negative returns/volatility relationship. The filtered variances are similar to the ones in figure 6.5.

Asset pricing models typically assert that market volatility is a measure of systematic risk, and that the expected return should be adjusted accordingly. If  $r$  is the risk free rate of return, then popular modifications to the Garch equation are the so called Garch-in-mean models

$$\begin{aligned} r_t &= r + \lambda\sqrt{h_t} - \frac{1}{2}h_t + \epsilon_t, \quad \epsilon_t \sim N(0, h_t) \\ r_t &= r + \lambda h_t - \frac{1}{2}h_t + \epsilon_t, \quad \epsilon_t \sim N(0, h_t) \end{aligned}$$

The parameter  $\lambda$  in the above expressions denotes the price of risk. Note that in the first alternative the asset exhibits constant Sharpe ratio.

Garch models can also be extended to more dimensions. In that case the covariance matrix is updated at each time step. In the univariate case we needed to take some care to ensure that the variance remained positive; now, in an analogous fashion, we must make sure that the covariance matrix is *positive definite*. This is not a trivial task. Also, in the general case a large number of parameters have to be estimated, and we usually estimate restricted versions in order to reduce the dimensionality.<sup>12</sup> In general, a multivariate Garch(1,1) will be of the form

$$\begin{aligned} r_t &= \mu_t + \epsilon_t \\ \epsilon_t &= H_t^{1/2} \eta_t \\ \eta_t &\sim N(\mathbf{0}, I) \end{aligned}$$

The matrix  $H_t^{1/2}$  can be thought of as the one obtained from the Cholesky factorization of the covariance matrix  $H_t$ . The covariance matrix can be updated in a form that is analogous to the univariate Garch(1,1)

<sup>12</sup> The most widely used forms are the VEC specification of Bollerslev, Engle, and Wooldridge (1988), and the BEKK specification of Engle and Kroner (1995). A recent survey of different approaches and methods is Bauwens, Laurent, and Rombouts (2006).

$$\mathbf{H}_t = \boldsymbol{\Omega} + \mathbf{B} \odot \mathbf{H}_{t-1} + \alpha \odot (\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t)$$

In this case the  $(i, j)$ -th element of the covariance matrix will depend on its lagged value and on the product  $\epsilon_{t-1}^{(i)} \epsilon_{t-1}^{(j)}$ . Of course more general forms are possible, with covariances that depend on different lagged covariances or error products.

To illustrate the multivariate Garch, we will use an example that is based on the Capital Asset Pricing Model (CAPM). In particular, asset returns will depend on the covariance with the market and the market premium, which in turn will depend on the market variance. If we denote with  $r_t^A$ ,  $r_t^M$  and  $r_t^F$  the asset, market and risk free rates of return, then we can write the CAPM relationships as

$$\begin{aligned} r_t^A &= r_t^F + \frac{E_{t-1}(\boldsymbol{\epsilon}_t^A \boldsymbol{\epsilon}_t^M)}{E_{t-1}(\boldsymbol{\epsilon}_t^M)^2} (E_{t-1}r_t^M - r_t^F) + \boldsymbol{\epsilon}_t^A \\ r_t^M &= r_t^F + \lambda E_{t-1}(\boldsymbol{\epsilon}_t^M)^2 + \boldsymbol{\epsilon}_t^M \end{aligned}$$

Since  $E_{t-1}r_t^M - r_t^F = \lambda E_{t-1}(\boldsymbol{\epsilon}_t^M)^2$ , the above system simplifies to

$$\begin{aligned} r_t^A - r_t^F &= \lambda E_{t-1}(\boldsymbol{\epsilon}_t^A \boldsymbol{\epsilon}_t^M) + \boldsymbol{\epsilon}_t^A \\ r_t^M - r_t^F &= \lambda E_{t-1}(\boldsymbol{\epsilon}_t^M)^2 + \boldsymbol{\epsilon}_t^M \end{aligned}$$

We can estimate the above specification using a multivariate Garch approach, taking into account that the covariance and the variances can be time varying. If we define

$$\mathbf{r}_t^* = \begin{pmatrix} r_t^A - r_t^F \\ r_t^M - r_t^F \end{pmatrix}, \quad \boldsymbol{\epsilon}_t = \begin{pmatrix} \boldsymbol{\epsilon}_t^A \\ \boldsymbol{\epsilon}_t^M \end{pmatrix}, \quad \mathbf{H}_t = E_{t-1}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t)$$

then we can estimate the process (with 17 parameters)

$$\begin{aligned} \mathbf{r}_t^* &= \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \lambda_{1,1} & \lambda_{1,2} & \lambda_{1,3} \\ \lambda_{2,1} & \lambda_{2,2} & \lambda_{2,3} \end{pmatrix} \cdot g(\mathbf{H}_t) + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{H}_t) \\ \mathbf{H}_t &= \begin{pmatrix} \omega_{1,1} & \omega_{1,2} \\ \omega_{1,2} & \omega_{2,2} \end{pmatrix} + \begin{pmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{1,2} & \beta_{2,2} \end{pmatrix} \odot \mathbf{H}_{t-1} + \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} \\ \gamma_{1,2} & \gamma_{2,2} \end{pmatrix} \odot (\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t) \end{aligned}$$

The function  $g(\mathbf{H}_t) = (\mathbf{H}_t^{(1,1)}, \mathbf{H}_t^{(2,2)}, \mathbf{H}_t^{(1,2)})'$  takes the unique elements of the covariance matrix and puts them in a vector form.

If the conditional CAPM with time varying risk premiums is sufficient to explain the asset and market returns, then the following restrictions should be satisfied

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \lambda_{1,1} & \lambda_{1,2} & \lambda_{1,3} \\ \lambda_{2,1} & \lambda_{2,2} & \lambda_{2,3} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

The restrictions can be tested with a likelihood ratio test.

## GARCH OPTION PRICING

The Garch family of models has been the workhorse of volatility modeling and has had many applications in testing, forecasting, and risk management. Applications within a pricing framework one the other hand have been very limited. The reason is that Garch models are set-up in discrete time, and for that reason the underlying market is incomplete.

This means that replicating portfolios do not exist for derivative assets. Intuitively, this is due to the fact that the state-space is too dense compared to the time-space (where rebalancing takes place). Over a time step the asset price can ‘jump’ to an infinite number of values, and it is impossible to construct a position that will hedge against all possibilities. In contrast, when trading takes place continuously the asset price diffuses from one level to the next, giving us the opportunity to create a dynamic Delta hedging strategy.

This is not a feature of Garch models alone; all models that are set up in discrete time and have continuous support will share the same drawback. Even in the simple model where the asset log-price follows a random walk model in discrete time the market is incomplete. This implies that there is not a unique way to identify the risk adjusted probability measure in discrete time models. For example, there is nothing to stop us from specifying

$$\begin{aligned} S_{t+1} &= S_t \exp(\mu + \epsilon_{t+1}), \quad \epsilon_{t+1} \sim N(0, \sigma^2), \quad \text{under } \mathcal{P} \\ S_{t+1} &= S_t \exp(\mu^Q + \epsilon_{t+1}^Q), \quad \epsilon_{t+1} \sim t(v), \quad \text{under } \mathcal{Q} \end{aligned}$$

for  $\mu^Q$  chosen in a way that makes the discounted price a martingale under  $\mathcal{Q}$ ,  $S_t = E_t^Q[\exp(-r\Delta t)S_{t+1}]$ . But not all is lost: we just need to impose some more structure that will eventually constrain our choices for  $\mathcal{Q}$ . Here we will outline two methods to achieve that, but since derivative pricing typically takes place in a continuous time setting, we will not dwell into details.

1. We might impose assumptions on the utility structure. Assuming a certain utility form will set the family of equivalent measures. In particular, the parameters of the utility function may be recovered from the true stock and risk free expected returns.
2. We can assume that the density structure has to be maintained, that is to say if errors are normally distributed under  $\mathcal{P}$ , then they must be normally distributed under  $\mathcal{Q}$ .

### Utility based option pricing

In our first approach will will assume a utility function  $U(t, W_t)$ , which measures utility of wealth  $W_t$  realized at time  $t$ . We will also need the relationship between wealth  $W_t$  and the underlying asset price  $S_t$  ((an early source for this approach is [Brennan, 1979](#))). For example, if the underlying asset is a wide index, then one might assume that investor’s wealth is very correlated with this index. If the underlying asset is a small stock, then the correlation will be smaller. This

resembles the impact of the idiosyncratic versus the systematic risk in asset pricing models.

Option prices can be computed from the *Euler equations*, which state that the price at time  $t$  of a random claim that is realized at time  $T > t$ , say  $X_T$ , is given by (see for example Barone-Adesi, Engle, and Mancini, 2004)

$$X_t = E_t \left( \frac{U_W(T, W_T)}{U_W(t, W_t)} X_T \right)$$

Essentially, the Euler equation weights each outcome with its impact on the marginal rate of substitution, before taking expectations. The price of a European call option would be then equal to

$$P_t = E_t \left( \frac{U_W(T, W_T)}{U_W(t, W_t)} (S_T - K)^+ \right)$$

Note that in the above expression there is no talk of equivalent measures. All expectation are taken directly under  $\mathcal{P}$ . Nevertheless, if we think of the marginal rate of substitution as a Radon-Nikodym derivative, then we can define the equivalent probability measure.

Of course, in general it is not straightforward neither to specify the appropriate utility nor to compute the expectation in closed form, but things are substantially simplified if we consider power utility functions. In fact , we will arrive to the Esscher transform, which has been very successful in actuarial sciences. This is described in detail in Gerber and Shiu (1994).

### Distribution based

The second method takes a more direct approach. Suppose that the log price follows the standard Garch(1,1) model

$$\begin{aligned} \Delta \log S_t &= \mu - \frac{1}{2} h_t + \sqrt{h_t} \eta_t \\ h_t &= \omega + \beta h_{t-1} + \gamma h_{t-1} \eta_{t-1}^2 \end{aligned}$$

Rather than trying to derive, we *define* the risk neutral probability measure as the one under which the random variable

$$\eta_t^Q = \eta_t - \frac{r - \mu}{\sqrt{h_t}}$$

is a martingale. Then under risk neutrality the asset log price follows

$$\begin{aligned} \Delta \log S_t &= r - \frac{1}{2} h_t + \sqrt{h_t} \eta_t^Q \\ h_t &= \omega + \beta h_{t-1} + \gamma h_{t-1} \left( \eta_{t-1}^Q + \frac{r - \mu}{\sqrt{h_t}} \right)^2 \end{aligned}$$

This approach is pretty much described in Duan (1995). Derivative are computed, in the usual way, as the expectation under risk neutrality. The benefit of this approach is that standardized errors remain normally distributed even after the probability measure change. Equivalently, we can say that the Black-Scholes formula holds for options with one period to maturity.

One major drawback of the standard Garch model is that the expectation that prices derivatives is not generally computable in closed form. Of course simulation based techniques can be employed, but they will be time consuming. An alternative, presented in Duan, Gauthier, and Simonato (1999) can be used. In this approach the state space is discretized, and a Markov chain is used to approximate the Garch dynamics.

The Garch variant introduced in Heston and Nandi (2000) circumvents the computability issue, and we present their approach in the following subsection.

### The Heston and Nandi model

Heston and Nandi (2000) propose a similar class of Garch-type processes

$$\begin{aligned}\Delta \log S_t &= r + \lambda h_t - \frac{1}{2} h_t + \sqrt{h_t} \eta_t \\ h_t &= \omega + \beta h_{t-1} + \gamma (\eta_{t-1} - \delta \sqrt{h_t})^2\end{aligned}$$

Here the bilinearity in the variance process is broken. That is to say, the product  $\sqrt{h_{t-1}} \eta_{t-1}$  is not present and  $\eta_{t-1}$ , which is a standardized normal series, appears in the variance update alone.

We set  $\eta_t^Q = \eta_t + \lambda \sqrt{h_t}$ , and define the probability measure  $\mathcal{Q}$  as one that is equivalent to  $\mathcal{P}$ , and also  $\eta_t^Q \sim N(0, 1)$  under this measure. Then, the asset price process under  $\mathcal{Q}$  will satisfy

$$\begin{aligned}\Delta \log S_t &= r - \frac{1}{2} h_t + \sqrt{h_t} \eta_t^Q \\ h_t &= \omega + \beta h_{t-1} + \gamma (\eta_{t-1}^Q - \delta^Q \sqrt{h_t})^2\end{aligned}$$

for  $\delta^Q = \delta + \lambda$ .

Unlike the standard Garch model, the Heston and Nandi (2000) modification allows one to compute the characteristic function as a closed form recursion. Then, option prices or risk neutral can be easily computed using the methods described in chapter 4.

## 6.3 THE STOCHASTIC VOLATILITY FRAMEWORK

As we pointed out in the previous section, for option pricing purposes continuous time stochastic volatility models are immensely more popular than model set-up

in discrete time.<sup>13</sup> The generic stochastic volatility process is described by a system of two SDEs

$$\begin{aligned} dS_t &= \mu_t S_t dt + \sqrt{\nu_t} S_t dB_t \\ dv_t &= \alpha(v_t) dt + \beta(v_t) dB_t^v \end{aligned}$$

The leverage effect is accommodated by allowing the asset return and the volatility innovations to be correlated

$$E_t dB_t dB_t^v = \rho dt$$

Derivative prices will have a pricing function that will depend on the volatility, on top of time and the underlying asset price

$$P_t = f(t, S_t, v_t)$$

When we introduce stochastic volatilities we move away from the Black-Scholes paradigm, where there was a single Brownian motion that generates uncertainty and markets are complete. Stochastic volatility models are driven by *two* Brownian motions, and for that reason the market that is constructed by the risk free and the underlying asset is not complete. This means that there is an infinite number of derivative prices that rule out arbitrage.

Having said that, if we augment the hedging instruments with one derivative contract the market becomes complete. Therefore derivatives will now have to be priced in relation to each other, as well as the underlying asset. This is one of the reasons that models with stochastic volatilities are calibrated to observed derivative prices.

Another point of viewing this issue is through the notion of volatility risk, introduced by the second BM  $B_t^v$  (and in particular the part  $\bar{B}_t$  of  $B_t^v$  that is orthogonal to  $B_t$ , since we can write  $B_t = \rho B_t^v + \sqrt{1 - \rho^2} \bar{B}_t$ ). The underlying asset does not depend on this BM, and therefore the risk generated by this BM is not actually priced within the asset price. On the other hand, of course, the risk of  $B_t$  is embedded in the risk premium  $\mu_t - r_t$ . Investor's might be risk averse towards this risk, and although this risk aversion is not manifested in the market for the underlying asset, it will be manifested in the options market as these contracts depend on  $v_t$  directly. Using one derivative we can identify the risk premium, and then we can price all other derivatives accordingly.

We will describe two approaches that reach derivative prices, one that implements Girsanov's theorem and one that constructs a hedging portfolio in the spirit of BS. Before we do that, we will go through some standard stochastic volatility models that have been proposed, discussing some of their properties and features. We will just present a selected few here, to illustrate the motivation as they try to capture they stylized features of volatility processes.

---

<sup>13</sup> Stochastic volatility models can also be set in discrete time, like the specification described in [Harvey, Ruiz, and Shephard \(1994\)](#). They are used for historical estimation but are not popular for derivative pricing, just like their Garch-type counterparts.

## THE HULL AND WHITE MODEL

The first stochastic volatility model was introduced in Hull and White (1987, HW). HW recognized that if investors are indifferent towards volatility risk (that is  $v_t$  is not correlated with the consumption that enters their utility function), and volatility is independent of the underlying asset price process, then one can integrate out the volatility path, and write the price of an option as a weighted average. In particular, if we are pricing a European call option, then it is sufficient to condition on the average variance over the life of the option

$$P_{HW} = \int_0^\infty f_{BS}(t, S_t; \bar{v}) f(\bar{v}) d\bar{v}$$

where the average variance over the life of the derivative in question is defined as

$$\bar{v} = \frac{1}{T-t} \int_t^T v_s ds$$

and  $f(\bar{v})$  is the probability density of the average variance process.

For example, in the original Hull and White (1987) article the variance is assumed to follow a geometric Brownian motion, which is uncorrelated with the asset price process.

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dB_t \\ dv_t &= \theta v_t dt + \phi v_t dB_t^\nu \end{aligned}$$

In this case, HW give a series approximation for the option price, which is based on the moments of the average variance.

The HW model was the first approach (together with Wiggins, 1987) towards a pricing formula for SV models, but the model they propose does not capture the desired features of realized volatilities. In particular, under the geometric Brownian motion dynamics, variance will be lognormally distributed. In the long run, the volatility paths will either explode towards infinity, or they will fall to zero, depending on the parameter values. Volatility in the HW model does not exhibit mean reversion and is not stationary. As maturities increase, the variance of out volatility forecasts increases without bound.

## THE STEIN AND STEIN MODEL

The Stein and Stein (1991, SS) model remedies the long run behavior of the HW specification. In particular, rather than a geometric Brownian motion, SS use an Ohrnstein-Uhlenbeck (OU) process. The OU process exhibits mean reversion, and for that reason has a long run stationary distribution, which is actually normal. SS model the volatility rather than the variance

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma_t S_t dB_t \\ d\sigma_t &= \theta(\bar{\sigma} - \sigma_t) dt + \xi dB_t^\sigma \end{aligned}$$

This process was later extended in [Schöbel and Zhu \(1999\)](#) by allowing the two BM processes to be correlated. The volatility process follows a normal distribution for each maturity, and therefore can cross zero. This implied that the true correlation (that is  $E_t dS_t d\sigma_t$ ) changes sign when this happens. This can be an undesirable property of the model.

[Schöbel and Zhu \(1999\)](#) compute the characteristic function of the log-price

$$\begin{aligned}\phi(T, u) = & \exp \left\{ iu \log(S_0) + iu\mu T - \frac{1}{2}iu \frac{\rho}{\xi} (\sigma_0^2 + \xi^2) \right\} \\ & \times \exp \left\{ \frac{1}{2}D(T; s_1, s_3)\sigma_0^2 + B(T; s_1, s_2, s_3)\sigma_0 + C(T; s_1, s_2, s_3) \right\}\end{aligned}$$

The functions  $D$ ,  $B$  and  $C$  are solutions of a system of ODEs, and are given in a closed (but complicated) form in the appendix of [Schöbel and Zhu \(1999\)](#).

### THE HESTON MODEL

By far the most popular model with stochastic volatility is [Heston \(1993\)](#). The variance follows the square root process of (also called a Feller process, developed in [Feller, 1951](#)), also used as the building block for the [Cox, Ingersoll, and Ross \(1985\)](#) model for the term structure of interest rates. The dynamics are given by

$$\begin{aligned}dS_t &= \mu S_t dt + \sqrt{v_t} S_t dB_t \\ dv_t &= \theta(\bar{v} - v_t)dt + \xi \sqrt{v_t} dB_t^\nu \\ E_t dB_t dB_t^\nu &= \rho dt\end{aligned}$$

The Heston model has a number of attractive features and a convenient parameterization. In particular, the variance process is always non-negative, and is actually strictly positive if  $2\theta\bar{v} > \xi^2$ . The volatility-of-volatility parameter  $\xi$  controls the kurtosis, while the correlation parameter  $\rho$  can be used to set the skewness of the density of asset returns. The variance process exhibits mean reversion, having as an attractor the long run variance parameter  $\bar{v}$ . The parameter  $\theta$  defines the strength of mean reversion, and dictates how quickly the volatility skew flattens out.

This model belongs to the more general class of *affine* models of [Duffie et al. \(2000\)](#), and the characteristic function of the log-price is given in closed form. In particular it has an *exponential-affine* form<sup>14</sup>

---

<sup>14</sup> We use the negative square root in  $d$ , found in [Gatheral \(2006\)](#), unlike the original formulation in [Heston \(1993\)](#). [Albrecher, Mayer, Schoutens, and Tistaert \(2007\)](#) discuss this choice and show that the two are equivalent, but using the negative root offers higher stability for long maturities. The problem arises due to the branch cuts of the complex logarithm in  $C(u, T)$ . A description of the problem and a different approach can be found in [Kahl and Jäckel \(2005\)](#).

LISTING 6.4: phi\_heston.m: Characteristic function of the Heston model.

```
% phi_heston.m
function y = phi_heston(u, p)
t      = p.t;
r      = p.r;
5 v0    = p.v0;
vBar   = p.vBar;
theta  = p.theta;
xi     = p.xi;
rho    = p.rho;
10 d   = -sqrt((i*rho*xi*u-theta).^2 + (xi.^2)*u.* (i+u));
h   = theta - i*rho*xi*u + d;
g   = h./ (h - 2*d);
ed  = exp(t*d);
15 LG = log((1 - g.*ed)./(1 - g));
C   = i*r*t*u + theta*vBar/(xi.^2)*(h*t - 2*LG);
D   = h/(xi.^2).* (1 - ed)./(1 - g.*ed);
y   = exp(C + v0*D);
```

$$\phi(u, T) = \exp \{ C(u, T) + D(u, T)v_0 + iu \log S_0 \}$$

with

$$\begin{aligned} C(u, T) &= iu\mu T + \frac{\theta\bar{V}}{\xi^2} \left\{ (\theta - iu\rho\xi + d)T - 2 \log \frac{1 - g \exp(dT)}{1 - g} \right\} \\ D(u, T) &= \frac{\theta - iu\rho\xi + d}{\xi^2} \cdot \frac{1 - \exp(dT)}{1 - g \exp(dT)} \\ g &= \frac{\theta - iu\rho\xi + d}{\theta - iu\rho\xi - d} \\ d &= -\sqrt{(iu\rho\xi - \theta)^2 + \xi^2 u(i + u)} \end{aligned}$$

The characteristic function of the Heston model is given in 6.4. This can be used to compute European style vanilla calls and puts using the transform methods outlined in chapter 4. We will be using this approach later in this chapter to calibrate the Heston model to a set of observed option prices.

### GIRSANOV'S THEOREM AND OPTION PRICING

We will now turn to the problem of option pricing, and discuss the two main methods. We start with an implementation of Girsanov's theorem, and then we will investigate the hedging structure that will give us the corresponding PDE. We set up a filtered space  $\{\Omega, \mathbb{F}, \mathcal{F}, \mathbb{P}\}$  and two correlated Brownian motions with respect to  $\mathbb{P}$ ,  $B_t$  and  $B_t^\nu$ . Based on these BMs we now consider a general stochastic volatility specification

for copies, comments, help etc. visit <http://www.theponytail.net/>

$$\begin{aligned} dS_t &= \mu_t S_t dt + \sqrt{\nu_t} S_t dB_t \\ dv_t &= \alpha(v_t) dt + \beta(v_t) dB_t^\nu \\ E_t^P dB_t dB_t^\nu &= \rho dt \end{aligned}$$

In order to apply Girsanov's theorem we define the process  $M_t$  via the stochastic differential equation

$$dM_t = \phi_t M_t dB_t + M_t \psi_t dB_t^\nu$$

with initial value  $M_0 = 1$ . The solution of this SDE is the exponential martingale (with respect to  $\mathcal{P}$ ), which has the form

$$M_t = \exp \left\{ \int_t^T \phi_s dB_s - \frac{1}{2} \int_t^T \phi_s^2 ds + \int_t^T \psi_s dB_s^\nu - \frac{1}{2} \int_t^T \psi_s^2 ds \right\}$$

The processes  $\phi_t$  and  $\psi_t$  are  $\mathcal{F}_t$ -adapted, and therefore they can be functions of  $(t, S_t, v_t)$ . Based on this exponential martingale we can define a probability measure  $\mathcal{Q}$ , which is equivalent to  $\mathcal{P}$ . In fact, every choice of processes  $\phi_t$  and  $\psi_t$  will produce a different equivalent measure. The only constraint we need to impose on these processes is that the discounted underlying asset price must form a martingale under  $\mathcal{Q}$ , which then becomes an equivalent martingale measure (EMM). The fundamental theorem of asset pricing postulates that if this is the case, then there will be no arbitrage opportunities in the market. It turns out that this constraint is not sufficient to identify both processes, something that we should anticipate since the market is incomplete and there will not be a unique EMM.

The EMM will be defined via its Radon-Nikodym derivative with respect to the true measure,

$$\left. \frac{d\mathcal{Q}}{d\mathcal{P}} \right|_t = M_t$$

If  $Y_T$  is a  $\mathcal{F}_T$ -measurable random variable, then expectations under the equivalent measure will be given as

$$E_t^Q Y_T = E_t^P \frac{M_T}{M_t} Y_T$$

It is useful to compute the expectations over an infinitesimal interval  $dt$ , as this will help us compute the drifts and volatilities under  $\mathcal{Q}$ . In particular we will have

$$\begin{aligned} E_t^Q dY_t &= E_t^P \frac{M_t + dM_t}{M_t} dY_T \\ &= E_t^P \left( 1 + \frac{dM_t}{M_t} \right) dY_T = E_t^P dY_t + E_t^P (\phi_t dB_t + \psi_t dB_t^\nu) dY_T \end{aligned}$$

We can employ the above relationship to compute the drifts and the volatilities of the asset returns under  $\mathcal{Q}$

$$E_t^Q \left( \frac{dS_t}{S_t} \right) = (\mu + \phi_t \sqrt{v_t} + \rho \psi_t \sqrt{v_t}) dt, \text{ and } E_t^Q \left( \frac{dS_t}{S_t} \right)^2 = v_t dt$$

The drift and volatility of the variance process are

$$E_t^Q dv_t = (\alpha(v_t) + \rho \phi_t \beta(v_t) + \psi_t \beta(v_t)) dt, \text{ and } E_t^Q (dv_t)^2 = \beta^2(v_t) dt$$

This verifies that under equivalent probability measures the drifts are adjusted but volatilities are not. Now an EMM will be one that satisfies

$$E_t^Q (dS_t/S_t) = r dt$$

This constraint yields a relationship between  $\phi_t$  and  $\psi_t$

$$\phi_t + \rho \psi_t = -\frac{\mu - r}{\sqrt{v_t}} = -\Xi^S(t, S_t, v_t)$$

The function  $\Xi^S(t, S, v)$  is the market price of risk, the Sharpe ratio of the underlying asset. In order to construct a system we need a second equation, and essentially we have the freedom to choose the market price of volatility risk. Thus if we select a function  $E_t^Q dv_t = \alpha^Q(t, S, v)$ , which will be the variance drift under risk neutrality, we can set up a second equation

$$\rho \phi_t + \psi_t = -\frac{\alpha(v_t) - \alpha^Q(t, S_t, v_t)}{\beta(v_t)} = -\Xi^V(t, S_t, v_t)$$

where  $\Xi^V(t, S, v)$  will be the price of volatility risk.

The market risk premium  $\Xi^S$  will be typically positive, as the underlying asset will offer expected returns that are higher than the risk free rate. This reflects the fact that investors prefer higher returns, but are risk averse against declining prices. When it comes to volatility, we would expect investors to prefer *lower* volatility, and be risk averse against *volatility increases*. This indicates that it would make sense to select  $\alpha^Q$  in a way that implies a negative risk premium  $\Xi^V$ , and one that does not increase with volatility. Essentially this means that  $\alpha^Q \geq \alpha$ . In practice we will have to find a convenient parameterization for  $\alpha^Q$  or  $\Xi^V$  that leads to expressions that admit solutions, and at the same time restrict the family of admissible EMMs. The parameter values cannot be determined from the dynamics of the underlying asset, but they can be recovered from observed derivative prices.

If we solve the above system we can find the processes  $\phi_t$  and  $\psi_t$ , and through them the appropriate EMM, as follows

$$\begin{aligned} \phi_t &= \frac{1}{1 - \rho^2} (-\Xi^S + \rho \Xi^V) \\ \psi_t &= \frac{1}{1 - \rho^2} (-\Xi^V + \rho \Xi^S) \end{aligned}$$

Finally, derivative prices can be written as expectations under  $Q$ , where the asset dynamics are

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{v_t} S_t dB_t^Q \\ dv_t &= \alpha^Q(t, S_t, v_t) dt + \beta(v_t) dB_t^{Q,v} \\ E_t^Q dB_t^Q dB_t^{Q,v} &= \rho dt \end{aligned}$$

where  $\alpha^Q(t, S_t, v_t) = \alpha(v_t) - \Xi^v(t, S_t, v_t)\beta(v_t)$ .

### Example: The Heston model

In Heston's model the variance drift and volatility are given by

$$\begin{aligned} \alpha(v) &= \theta(\bar{v} - v) \\ \beta(v) &= \xi\sqrt{v} \end{aligned}$$

The price of risk is determined by the risk free rate and the asset price dynamics

$$\Xi^S(t, S, v) = \frac{\mu - r}{\sqrt{v}}$$

We are free to select the price of volatility risk. Say we set it equal to

$$\Xi^v(t, S, v) = \frac{\kappa}{\xi}\sqrt{v_t}$$

for a parameter  $\kappa \leq 0$  (to conform with agents that are averse towards *higher* volatility). Then, the risk premium will be positive and increasing with volatility. In addition, such a risk premium will lead to risk neutral dynamics that have the same form as the dynamics under  $\mathcal{P}$ .

Girsanov's theorem will give the process under  $\mathcal{Q}$

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{v_t} S_t dB_t^Q \\ dv_t &= \alpha^Q(v_t) dt + \xi\sqrt{v_t} dB_t^{v,Q} \end{aligned}$$

The risk neutral variance drift  $\alpha^Q(v_t) = \theta(\bar{v} - v_t) - \Xi^v(t, S_t, v_t)\xi\sqrt{v_t}$ . Then we can rewrite the dynamics

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{v_t} S_t dB_t^Q \\ dv_t &= \theta^Q(\bar{v}^Q - v_t) dt + \xi\sqrt{v_t} dB_t^{Q,v} \end{aligned}$$

for the parameters  $\theta^Q = \theta + \kappa$  and  $\bar{v}^Q = \frac{\theta\bar{v}}{\theta+\kappa}$ . Due to their risk aversion, manifested through the parameter  $\kappa \leq 0$ , investors behave as if the long run volatility is *higher* than it really is, and as if volatility exhibits *higher* persistence.

### THE PDE APPROACH

Alternatively, we can take a route that follows the BS methodology, where a derivative is shorted and subsequently hedged. This will give rise to the PDE representation of the price. In the BS world with constant volatility, it was

sufficient to use the underlying asset and the money market account to achieve the hedge. Here, as we have one more source of risk, these two instruments will not be sufficient to eliminate volatility risk. To hedge our short derivative we will use the money market account, the underlying asset *and* one extra derivative.

Consider a derivative  $X$ , and denote its pricing function with  $f(t, S, v)$ . The functional form of  $f$  will depend on the particulars of the contract, such as maturity, payoff structure, optionality, etc. Therefore, the process for this derivative will be given by  $X_t = f(t, S_t, v_t)$ . Following the BS argument, if we knew the functional form of  $f$ , we could compute the dynamics of the derivative price using Ito's formula (for two dimensions)

$$dX_t = \left( \alpha_t^X + \mu S_t \frac{\partial f}{\partial S} \right) dt + \beta_t^{X,S} dB_t + \beta_t^{X,v} dB_t^v$$

where

$$\begin{aligned} \alpha_t^X &= \frac{\partial f}{\partial t} + \alpha(v_t) \frac{\partial f}{\partial v} + \frac{1}{2} v_t S_t^2 \frac{\partial^2 f}{\partial S^2} \\ &\quad + \frac{1}{2} \beta^2(v_t) \frac{\partial^2 f}{\partial v^2} + \rho \sqrt{v_t} S_t \beta(v_t) \frac{\partial^2 f}{\partial S \partial v} \\ \beta_t^{X,S} &= \sigma S_t \frac{\partial f}{\partial S} \\ \beta_t^{X,v} &= \beta(v_t) \frac{\partial f}{\partial S} \end{aligned}$$

From this expression it is apparent that if we construct a portfolio using only the underlying asset and the bank account, we will not be able to replicate the price process  $X_t$ , since the risk source  $B_t^v$  cannot be reproduced. The market that is based only on these instruments is incomplete, since the derivative cannot be replicated.

But we can dynamically complete the market using another derivative  $X^*$ , with pricing function  $f^*(t, S, v)$ . This will work of course if  $X^*$  actually depends on the BM  $B_t^v$ , which is typically the case.<sup>15</sup> In practice we would perhaps replicate  $X$  (say a barrier option), using the risk free asset, the underlying asset and a liquid derivative  $X^*$  (say a vanilla at the money option).

Following BS, we short  $X$  and construct a portfolio of the underlying stock and the other derivative. We want to select the weights of this portfolio in a way that makes it risk free. Then it should grow at the risk free rate.

Say that at each point in time we hold  $\Delta_t$  units of the underlying asset and  $\Delta_t^*$  units of the derivative. Then the change in our portfolio value  $\Pi_t$  will be

$$d\Pi_t = dX_t - \Delta_t dS_t - \Delta_t^* dX_t^*$$

---

<sup>15</sup> It is sufficient that the pricing function depends on  $v$ , in order for the derivative  $\frac{\partial f}{\partial v} \neq 0$ , as we will see below. For example a forward contract is a derivative but it would not depend on  $B_t^v$ , since its pricing function  $f(t, S, v) = S \exp(r(T - t))$  does not depend on  $v$ .

Substituting for the dynamics of  $dX_t$ ,  $dS_t$  and  $dX_t^*$  will give the portfolio dynamics

$$\begin{aligned} d\Pi_t &= (\dots)dt \\ &\quad + \left( \sqrt{v_t} S_t \frac{\partial f}{\partial S} - \Delta_t \sqrt{v_t} S_t - \Delta_t^* \sqrt{v_t} S_t \frac{\partial f^*}{\partial S} \right) dW_t \\ &\quad + \left( \beta(v_t) \frac{\partial f}{\partial v} - \Delta_t^* \beta(v_t) \frac{\partial f^*}{\partial v} \right) dZ_t \end{aligned}$$

If we select the portfolio weights that make the parentheses equal to zero, then we have constructed a risk free portfolio. The solution is obviously<sup>16</sup>

$$\begin{aligned} \Delta_t^* &= \frac{\partial f}{\partial v} \left( \frac{\partial f^*}{\partial v} \right)^{-1} \\ \Delta_t &= \frac{\partial f}{\partial S} - \Delta_t^* \frac{\partial f^*}{\partial S} \end{aligned}$$

And since the portfolio will be risk now free, it will also have to grow at the risk free rate of return

$$d\Pi_t = r\Pi_t = r(X_t - \Delta_t S_t - \Delta_t^* X_t^*)$$

We should expect that the drifts will give the PDE that we are looking for, but at the moment we have a medley of partial derivatives of both pricing functions  $f$  and  $f^*$ . Nevertheless, we can carry on setting the portfolio drifts equal, which yields

$$\alpha^X + \mu S \frac{\partial f}{\partial S} - \Delta \mu S - \Delta^* \alpha^{X^*} - \Delta^* \mu S \frac{\partial f^*}{\partial S} = r(f - \Delta S - \Delta^* f^*)$$

Since  $\Delta + \Delta^* \frac{\partial f^*}{\partial S} = \frac{\partial f}{\partial S}$  the drift of the underlying asset  $\mu$  will cancel out, resembling the BS scenario. Furthermore, if we substitute the hedging weights  $\Delta$  and  $\Delta^*$  and rearrange to separate the starred from the non-starred elements

$$\lambda = \frac{\alpha^X + rS \frac{\partial f}{\partial S} - rf}{\frac{\partial f}{\partial v}} = \frac{\alpha^{X^*} + rS \frac{\partial f^*}{\partial S} - rf^*}{\frac{\partial f^*}{\partial v}} = \lambda^*$$

The following line of argument is the most important part of the derivation, and the most tricky to understand at first reading: In the above expression the RHS ratio  $\lambda$  (which depends only on  $f$ ) is equal to the LHS ratio  $\lambda^*$  (which depends only on  $f^*$ ). Recall that  $f$  and  $f^*$  are the pricing functions of two *arbitrary* derivatives, which means that the above ratio will be the same for all derivative contracts. If we selected another derivative contract  $X^{**}$ , then for its pricing function  $\lambda = \lambda^{**}$ , which implies  $\lambda = \lambda^* = \lambda^{**}$ , etc. This means that although

---

<sup>16</sup> Apparently, for the solution to exist we need  $\frac{\partial f^*}{\partial v} \neq 0$ . This corresponds to our previous remark that a forward contract cannot serve as the hedging instrument.

$\lambda$  can depend on  $(t, S, v)$ , it cannot depend on the particular features of each derivative contract (since if it did, it wouldn't be equal for all of them). We therefore conclude that

$$\lambda = \lambda^* = \lambda^{**} = \dots = \lambda(t, S, v)$$

That is very important, because it means that all derivatives will satisfy the same ratio, which can be rewritten as a single PDE<sup>17</sup>

$$\begin{aligned} \frac{\alpha^X + rS \frac{\partial f}{\partial S} - rf}{\frac{\partial f}{\partial v}} &= \lambda(t, S, v) \\ &\Rightarrow \frac{\partial f}{\partial t} + \{\alpha(v) - \lambda(t, S, v)\} \frac{\partial f}{\partial v} + \frac{1}{2} v S^2 \frac{\partial^2 f}{\partial S^2} \\ &\quad + \frac{1}{2} \beta^2(v) \frac{\partial^2 f}{\partial v^2} + \rho \sqrt{v} S_t \beta(v) \frac{\partial^2 f}{\partial S \partial v} + r S \frac{\partial f}{\partial S} = rf \end{aligned}$$

As always, the boundary conditions of this PDE will define which contract is actually priced. In particular, the terminal condition  $f(T, S, v) = \Pi(S)$ , with  $\Pi(S)$  the payoff of the derivative.

### The Feynman-Kac link

It is very instructive to pause at this point and verify the links that connect the two approaches. Using Girsanov's theorem we built the EMM and we concluded that a derivative, say with payoff  $X_T = f(T, S_T, v_T) = \Pi(S_T)$ , will be priced as the expectation under the EMM

$$X_0 = \exp(-rT) E_0^\Omega X_T = \exp(-rT) E_0^\Omega \Pi(S_T)$$

where the dynamics of the underlying asset and its volatility are given by the SDEs

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{v_t} S_t dB_t^\Omega \\ dv_t &= \alpha^\Omega(t, S_t, v_t) dt + \beta(v_t) dB_t^{\Omega,v} \\ E_t^\Omega dB_t^\Omega dB_t^{\Omega,v} &= \rho dt \end{aligned}$$

with the drift of the variance process given by  $\alpha^\Omega(t, S, v) = \alpha(v) - \Xi^v(t, S, v)\beta(v)$ . The price of volatility risk is  $\Xi^v$ .

Using the PDE approach we concluded that the pricing function  $f(t, S, v)$  will solve the PDE

$$\begin{aligned} \frac{\partial f}{\partial t} + \alpha^*(t, S, v) \frac{\partial f}{\partial v} + \frac{1}{2} v S^2 \frac{\partial^2 f}{\partial S^2} \\ + \frac{1}{2} \beta^2(v) \frac{\partial^2 f}{\partial v^2} + \rho \sqrt{v} S_t \beta(v) \frac{\partial^2 f}{\partial S \partial v} + r S \frac{\partial f}{\partial S} = rf \end{aligned}$$

---

<sup>17</sup> An identical line of argument is used in fixed income securities, which we will follow in chapter XX.

with  $\alpha^*(t, S, v) = \alpha(v) - \lambda(t, S, v)$  boundary condition  $f(T, S, v) = \Pi(S)$ .

The Feynman-Kac formula links the two approaches, as it casts the solution of the PDE as an expectation under the dynamics dictated by Girsanov's theorem. In fact, it follows that  $\alpha^Q(t, S, v) = \alpha^*(t, S, v)$ , which implies that

$$\lambda(t, S, v) = \Xi^v(t, S, v)\beta(v) = \alpha(v) - \alpha^Q(t, S, v)$$

The free functional  $\lambda$  that we introduced in the PDE approach can be interpreted as the total volatility risk premium. For investors that are averse towards high volatility  $\lambda \leq 0$ .

### Example: The Heston model

If we implement the PDE using the [Heston \(1993\)](#) dynamics, the derivative pricing function will satisfy

$$\begin{aligned} \frac{\partial f}{\partial t} + \{\theta(\bar{v} - v) - \lambda(t, S, v)\} \frac{\partial f}{\partial v} + \frac{1}{2}vS^2 \frac{\partial^2 f}{\partial S^2} \\ + \frac{1}{2}\xi^2 v \frac{\partial^2 f}{\partial v^2} + \rho v S \xi \frac{\partial^2 f}{\partial S \partial v} + rS \frac{\partial f}{\partial S} = rf \end{aligned}$$

In his original paper, Heston assumes  $\lambda(t, S, v)$  to be proportional to the variance  $v$

$$\lambda(t, S, v) = \lambda v$$

Essentially, following our previous discussion, this indicates that the equivalent function  $\Xi^v$  in the EMM approach will be

$$\Xi^v(t, S, v) = \frac{\lambda(t, S, v)}{\beta(v)} = \frac{\lambda}{\xi} \sqrt{v}$$

This means that the parameter  $\lambda$  of the PDE approach has exactly the same interpretation as  $\kappa$ . This choice for  $\lambda$  sets the PDE

$$\begin{aligned} \frac{\partial f}{\partial t} + \theta^Q(\bar{v}^Q - v) \frac{\partial f}{\partial v} + \frac{1}{2}vS^2 \frac{\partial^2 f}{\partial S^2} \\ + \frac{1}{2}\xi^2 v \frac{\partial^2 f}{\partial v^2} + \rho v S \xi \frac{\partial^2 f}{\partial S \partial v} + rS \frac{\partial f}{\partial S} = rf \end{aligned}$$

The boundary conditions are also need to specified. Following [Heston \(1993\)](#), for a European call option

$$\begin{aligned}
f(S, v, T) &= (S - K)^+ \\
\frac{\partial f}{\partial S}(\infty, v, t) &= 1 \\
f(0, v, t) &= 0 \\
f(S, \infty, t) &= S \\
\frac{\partial f}{\partial t}(S, 0, t) + \theta^Q \bar{v}^Q \frac{\partial f}{\partial v}(S, 0, t) \\
&+ rS \frac{\partial f}{\partial S}(S, 0, t) = rf(S, 0, t)
\end{aligned}$$

## ESTIMATION AND FILTERING

Since in stochastic volatility models the volatility is unobserved, it is generally *very hard* to estimate the parameters based on historical asset returns, and filter the unobserved volatility process. People have used a number of approaches, for some of which we give references below. For more details see the surveys of Ghysels et al. (1996) and Javaheri (2005).

1. Indirect inference: Estimating a deterministic model, for example via Arch or Egarch, and then studying the dynamics of the filtered volatility (Nelson, 1990, 1991).
2. Simulation based methods: Although the conditional moments or the likelihood are not available in closed form, they can be simulated. Of course, a simulation has to be run between all time steps, which makes these procedures computationally intensive and very time consuming. Examples include
  - Efficient Method of Moments (EMM), e.g. Gallant and Tauchen (1993)
  - Simulated Maximum Likelihood (SMM), e.g. Sandmann and Koopman (1998)
  - Markov Chain Monte Carlo (MCMC), e.g. (Eraker, Johannes, and Polson, 2001)
  - (Unscented-) Particle Filter, (PF, UPF), e.g. van der Merwe, de Freitas, Doucet, and Wan (2001)
3. Kalman filter methods: The classical Kalman filter is not directly applicable, but it can be used in some cases after a transformation. Versions of the extended Kalman filter have also been employed.
4. Likelihood Approximation Methods: The likelihood can be approximated for the affine class of models, constructing an updating procedure for the characteristic function (Bates, 2005). Alternatively, the volatility process itself can be approximated using a Markov chain, as in Chourdakis (2002).

## CALIBRATION

Even if we estimate the parameters of a stochastic volatility models using historical time series of asset returns, not all parameters would be useful for the purpose of derivative pricing. This happens because the estimated parameters

would be the ones under the true probability measure, while investors will use some adjusted parameters to price derivatives. In particular, for stochastic volatility models the drift of the variance will be a modification of the true one, which is done by setting the price of volatility risk. To recover this price of risk, one should consult some existing derivative prices.

For that reason, practitioners and (to some extend) academics prefer to use only derivative prices, and calibrate the model based on a set of liquid options. A standard setting is where a derivatives desk wants to sell an exotic option, and then hedge its exposure, and say that a stochastic volatility model is employed. The desk would look at the market prices of liquid European calls and puts, and would calibrate the pricing function to these prices. Such parameters are the risk neutral ones, and therefore can be used unmodified to price and hedge the exotic option. In a sense, they are a generalization of the BS implied volatilities. In a way, practitioners want to price the exotic contract in a way that is consistent with the observed vanillas.

If the calibrated model was the one that actually generated the data, then these implied parameters should be stationary through time, and their variability should be due to measurement errors alone. In practice of course this is not the case, and practitioners tend to recalibrate some parameters every day (and sometimes more often).

To implement this calibration we will need to minimize some measure of distance between the theoretical model prices and the prices of observed options. Say that we have a pricing function  $P(\tau, K; \vartheta) = P(\tau, K; S_0, r; \vartheta)$ , where  $\vartheta$  denotes the set of unobserved parameters that we need to extract. Also denote with  $\sigma(\tau, K; S_0, r; \vartheta)$  the implied volatility of that theoretical price, and with  $P^*(\tau, K)$  and  $\sigma^*(\tau, K)$  the observed market price and implied volatility. For example, in Heston's case  $\vartheta = \{v_0, \theta, \bar{v}, \xi, \rho\}$ . There are many objective functions that one can use for the minimization, the most popular having a weighted sum of squares form

$$G(\vartheta) = \sum_i \sum_j w_{i,j} |P(\tau_i, K_j; \vartheta) - P^*(\tau_i, K_j)|^2$$

The weights  $w_{i,j}$  can be used to different ends. Sometimes the choice of  $w_{i,j}$  reflect the liquidity of different options using a measure such as the bid-ask spread. In other cases one wants to give more weight to options that are near-the-money (using for example the Gamma), or to options with shorter maturities. In other cases one might want to implement a weighting scheme based on the options' Vega, in order to mimic an objective function that is cast in the implied volatility space.

Recovering the parameter set  $\vartheta$  is not a trivial problem, as the objective function can (and in many cases does) exhibit multiple local minima. This is a common feature of *inverse problems* like this calibration exercise. Typically some *regularization* is implemented, in order to make the problem well posed for standard hill climbing algorithms. A popular example is *Tikhonov-Phillips regularization* (see [Lagnado and Osher, 1997](#); [Crépey, 2003](#), for an illustration), where the objective function is replaced by

$$\tilde{G}(\vartheta) = G(\vartheta) + \alpha \cdot g^2(\vartheta, \vartheta_0)$$

for a regularization parameter  $\alpha$ . The role of the penalty function  $g(\vartheta, \vartheta_0)$  is to keep the parameter vector  $\vartheta$  as close as possible to a vector that is based on some *prior* information  $\vartheta_0$ . Depending on the particular pricing form, sometimes non-smoothness penalties are also sometimes imposed.<sup>18</sup>

From a finance point of view, the issue of multiple optima highlights the existence of model risk (Ben Hamida and Cont, 2005). Based on the finite set of option prices, different model parameters are indistinguishable. Using one set over another to price an exotic contract which might be sensitive to them can introduce losses. More generally, given the increasing arsenal of theoretical pricing models, different model classes can give identical fit for vanilla contracts (for details see Schoutens, Simons, and Tistaert, 2004).

### Calibration example

As an example we will fit Heston's stochastic volatility model to a set of observed option prices. In particular, we are going to use contracts on the SP500 index written on April 24, 2007. The objective function that we will use is just the sum of squared differences between model and observed prices. Listing 6.5 gives the code that computes the objective function. The prices are computed using the fractional FFT (see chapter 4), and the integration bounds are automatically selected to reflect the decay of the integrand  $\psi(u, T)$ .<sup>19</sup>

The snippet 6.6 shows how this objective function can be implemented to calibrate Heston's model using a set of observed put prices on the SP500 index. There are eight different maturities in the data set, ranging from 13 to 594 days. The sum of squared differences between the theoretical and observed prices is minimized, and for this example we did not use any weighting scheme. Figure 6.6 shows the observed option prices and the corresponding fitted values. The table below gives the calibrated parameters  $\hat{\vartheta} = \{v_0, \theta, \bar{v}, \xi, \rho\}$

$v_0$	0.0219
$\theta$	5.5292
$\bar{v}$	0.0229
$\xi$	1.0895
$\rho$	-0.6459

<sup>18</sup> This is particularly true for calibrating local volatility models which have a large number of parameters. We will discuss this family of models in the next section.

<sup>19</sup> As Kahl and Jäckel (2005) show, the characteristic function of the Heston model for large arguments decays as  $A \exp(-uC)/u^2$  times a cosine (where  $A = \psi(0, T)$  and  $C = \sqrt{1 - \rho^2}(v_0 + \theta\bar{v}T)/\xi$ ). We can therefore bound the integral  $\int_z^\infty |\psi(u, T)|du$  by  $|A| \exp(-Cz)/z$ . The solution of  $\exp(w)w = x$  is *Lambert's W function* which is implemented in Matlab through the Symbolic Math Toolbox. If this toolbox is not available we have to devise a different strategy to set the upper integration bound, for example using the moments expansion for the characteristic function. If everything else fails we can just set a 'large value' for the upper integration bound, or set up an adaptive integration scheme.

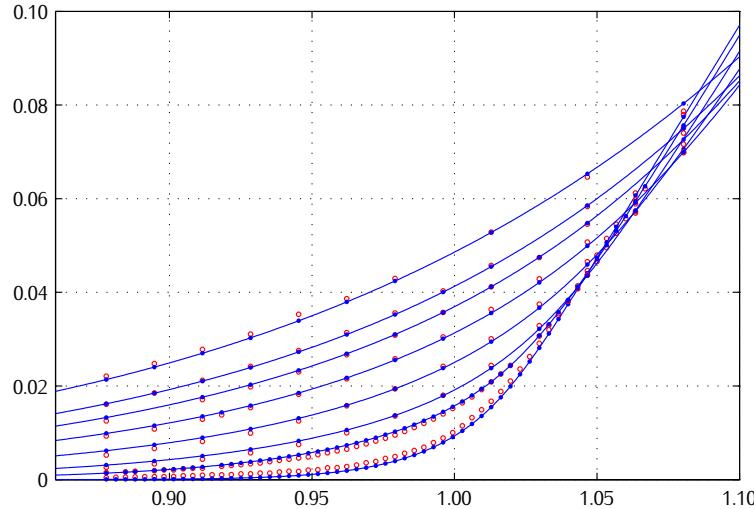
LISTING 6.5: ssq\_heston.m: Sum of squares for the Heston model.

```
%ssq_heston.m
function y = ssq_heston(par, pf, data)
ps.v0      = par(1); ps.vBar    = par(2);
ps.theta   = par(3); ps.xi     = par(4);
5 ps.rho    = par(5);
eta = pf.eta; k   = 10^(-pf.kappa);
CP = data(:,1); T  = data(:,2)/365; K = data(:,3);
P  = data(:,4); S0 = data(:,5); r  = data(:,6)/100;
% normalize prices and strikes for S0 = 1
10 Pn = P./S0; Kn = K./S0;
% select different maturities
[Tu, Iu] = unique(T); ru = r(Iu); Nu = length(Tu);
y = 0;                                     % will keep output SSQ
for n = 1:Nu                                % loop through T
15   ps.t = Tu(n);                         % set Heston params
   ps.r = ru(n);
   IC = (T==Tu(n)) & (CP > 0);          % select calls
   IP = (T==Tu(n)) & (CP < 0);          % select puts
   % set parameters for FRFT
20   % #if the function lambertw(.) is not available set
   % #pf.uBar to a 'large value' for the upper bound
   a0 = real(phi_heston(-i*(eta+1)/eta/(eta+1), ps));
   a1 = sqrt(1-(ps.rho)^2)*...
         ((ps.v0)+(ps.theta)*(ps.vBar)*(ps.t))/(ps.xi);
25   pf.uBar = lambertw(a0*a1/k)/a1;
   pf.kBar = 1.20*max(abs(log(Kn(IC | IP)))); % run FRFT pricing engine for Heston
   [Kv, Cv] = frft_call(@phi_heston, ps, pf);
   % construct strikes and put prices (from parity)
30   Kv = exp(Kv); Pv = Cv + exp(-ru(n)*Tu(n))*Kv - 1;
   Cf = interp1(Kv, Cv, Kn(IC)); % interpolate sample
   Pf = interp1(Kv, Pv, Kn(IP));
   % update SSQ
   y = y + sum((Cf-Pn(IC)).^2) + sum((Pf-Pn(IP)).^2);
35 end
y = log(y); % take log to help optimization
```

LISTING 6.6: calib\_heston.m: Calibration of the Heston model.

```
%calib_heston.m
% import option prices data
data = xlsread('SPX_24_04_07_a.xls');
% set up parameters for the FRFT
5 pf.eta = 1.50; % Carr-Madan parameter
pf.N = 512; % number of FFT points
pf.kappa = 6; % upper integration parameter
% initial parameter set
par = [0.02 0.02 5.00 1.00 -0.60];
10 % options for the optimization
opt = optimset('LargeScale', 'off', 'Display', 'iter');
par = fmincon(@ssq_heston, par, [], [], [], [], ...
    [0.005 0.005 0.10 0.10 -0.99], ...
    [0.500 0.500 20.0 5.00 -0.40], ...
    [], opt, pf, data, 0);
```

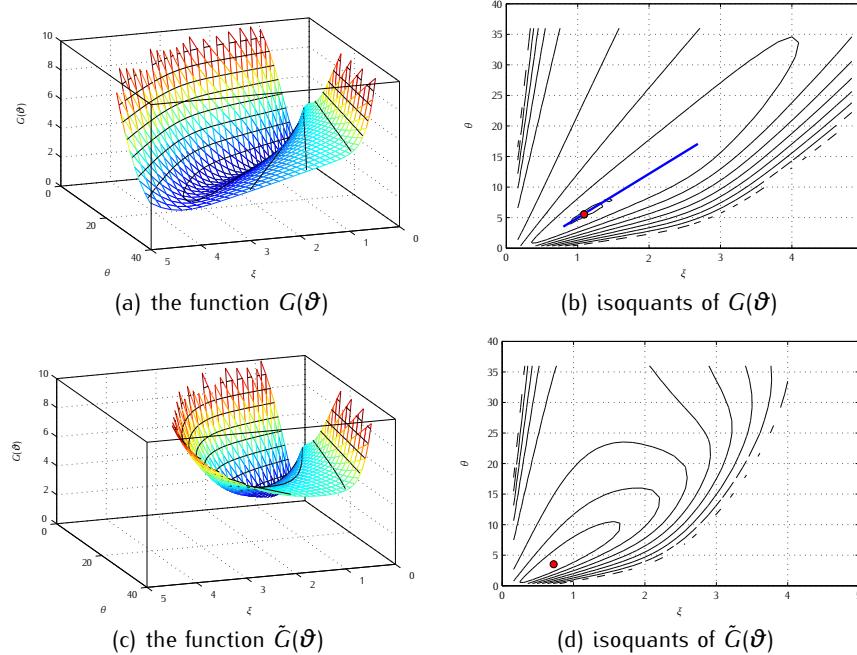
FIGURE 6.6: Calibrated option prices for Heston's model. The red circles give the observed put prices, while the blue dots are the theoretical prices based on Heston's model that minimize the squared errors.



These parameters are typical of a calibration procedure, and give an objective function value of  $G(\hat{\theta}) = 0.0090$ . The question is of course whether or not these parameters are unique, well defined and stable. In figure 6.7(a) we show the function  $G(\theta)$  for different combinations of  $(\theta, \xi)$ , keeping the rest of the parameters at their estimated values. There appears to be a "valley" across a

FIGURE 6.7: The ill-posed inverse problem in Heston's case. Subfigure (a) gives the objective function that is minimized to calibrate the parameters. Subfigure (b) presents the isoquants of this function, together with the minimum point attained using numerical optimization. Observe that all points that are roughly across the red line are indistinguishable. The regularized function is given in (c), while (d) shows its isoquants. Observe that the regularized function is better behaved than the original one.

[  check\_heston.m ]



set of values, indicating that it is very hard to precisely identify the optimal parameter combination. It is apparent that combinations of values across the red line in 6.7.b will give values for the objective function that are very close. This means that based on this set of vanilla options the combinations  $(\theta, \xi) = (5.0, 1.0), (10.0, 1.7)$  or  $(15.0, 2.5)$  are pretty much indistinguishable.

One way around this problem would be to enhance the information, by including more contracts such as *forward starting options* or *cliquet options*.<sup>20</sup>

<sup>20</sup> A *forward starting option* is an option that has some features that are not determined until a future time. For example, one could buy (and pay today for) a put option with three years maturity, but where the strike price will be determined as the level of SP500 after one year. Essentially one buys today what is going to be an ATM put in a years' time. A *cliquet* or *ratchet option* is somewhat similar, resembling a basket of forward starting options. For example I could have a contract where every year the

These are contracts that depend on the *dynamics* of the transition densities for the volatility, and not only on the densities themselves as vanillas do. For example, a forward starting option would depend on the joint distribution of the volatilities at the starting and the maturity times. Alternatively, if such exotic contracts do not exist or they are not liquid enough to offer uncontaminated values, one could stick with vanilla options and use a regularization technique. This demands some prior view on some parameter values, which could be based on historical evidence or analysts' forecasts. As an example, in Heston's model the parameter  $\xi$  is the same under both the objective and the risk neutral measure. Based on an estimate  $\xi_0$  using historical series of returns and/or option values, one can set up the objective function

$$\tilde{G}(\vartheta) = G(\vartheta) + \alpha(\xi - \xi_0)^2$$

In that way estimates will be biased towards combination where the prior value is  $\xi_0$ . For example, the estimation results of [Bakshi, Cao, and Chen \(1997\)](#) based on option prices and the joint estimation of returns and volatility in [Pan \(1997\)](#), indicate a value of  $\xi \approx 0.40$ . Therefore, if we set  $\xi_0 = 0.40$  and  $\alpha = 0.005$ , the objective function to be minimized is the one given in figure 6.7(c,d). The optimal values are now given in the following table

$v_0$	0.0200
$\theta$	3.5260
$\bar{v}$	0.0232
$\xi$	0.7310
$\rho$	-0.7048

The new objective function at the optimal is  $\tilde{G}(\hat{\vartheta}) = 0.0099$  which implies a sum of squares value  $G(\hat{\vartheta}) = 0.0094$ , which is not far from the unconditional optimization result.

## 6.4 THE LOCAL VOLATILITY MODEL

Stochastic volatility models take the view that there is an extra Brownian motion that is responsible for volatility changes. This extra source of randomness creates a market that is incomplete, where options are not redundant securities. Practically, this means that in order to hedge a position one needs to hedge against volatility risk as well as market risk. Local volatility models take a completely different view. No extra source of randomness is introduced, and the markets remain complete. In order to account for the implied volatility skew there is a nonlinear (but deterministic) volatility structure

---

payoff is determined and paid, and the strike price is readjusted according to the new SP500 level.

LISTING 6.7: nadwat2.m: Nadaraya-Watson smoother.

```
%nadwat2.m
function zi = nadwat2(x, y, z, w, xi, yi, hx, hy)
N = length(x);
if isempty(w)
    w = ones(N, 1);
end
zi1 = 0;
zi2 = 0;
for j = 1:N
    xe = exp(-0.5/hx^2*(xi-x(j)).^2)/sqrt(2*pi)/hx;
    ye = exp(-0.5/hy^2*(yi-y(j)).^2)/sqrt(2*pi)/hy;
    zi1 = zi1 + z(j)*w(j)*xe.*ye;
    zi2 = zi2 + w(j)*xe.*ye;
end
zi = zi1./zi2;
```

$$dS_t = rS_t dt + \sigma(t, S_t) S_t dB_t$$

As vanilla options are expressed via the risk neutral expectation of the random variable  $S_T$ , local volatility models attempt to construct the function  $\sigma(t, S)$  that is consistent with the implied risk neutral densities for different maturities. The methodology of local volatility models follows the one on implied risk neutral densities, originating in the pioneering work of [Breeden and Litzenberger \(1978\)](#).

These methods are inherently nonparametric, and rely on a large number of option contracts that span different strikes and maturities. In reality there is only a relatively small set of observed option prices that is traded, and for that reason some *interpolation* or *smoothing* techniques must be employed to artificially reconstruct the true pricing function or the volatility surface. Of course this implies that the results will be sensitive to the particular method that is used. Also, care has to be taken to ensure that the resulting prices are arbitrage free.

## INTERPOLATION METHODS

There are many interpolation methods that one can use on the implied volatility surface. As second order derivatives of the corresponding pricing function are required, it is paramount that the surface is sufficiently smooth. In fact, it is common practice to sacrifice the perfect fit in order to ensure smoothness, which suggests that we are actually implementing an implied volatility *smoother* rather than an *interpolator*. Within this obvious tradeoff we have to selecting the degree of fit versus smoothness, which is more of an art than a science.

One popular approach is to use a family of known functions, and reconstruct the volatility surface as a weighted sum of them. As an example we can use

for copies, comments, help etc. visit <http://www.theponytail.net/>

LISTING 6.8: imp\_vol.m: Implied volatility surface smoothing.

```
%imp_vol.m
data = xlsread('SPX_24_04_07_a.xls');
CP = data(:,1); T = data(:,2)/365; K = data(:,3);
P = data(:,4); S0 = data(:,5); r = data(:,6)/100;
5 [IV, IVi] = bs_iv(S0, K, r, T, CP, P);
[Tu, Iu] = unique(T);
% create output grids
dK = 5; dT = 0.02;
Ko = (1300:dK:1600)';
10 To = (min(0.90*T):dT:max(1.10*T))';
NKo = length(Ko); NTo = length(To);
[Kgo, Tgo] = meshgrid(Ko, To);
Kvo = reshape(Kgo, [NKo*NTo, 1]); % vectorize
Tvo = reshape(Tgo, [NKo*NTo, 1]);
15 Tmo = log(Tvo); % transform
Kmo = log(S0(1)./Kvo)./sqrt(Tvo);
% prepare actual data
Tm = log(T); % transform
20 Km = log(S0(1)./K)./sqrt(T);
% Nadaraya-Watson smoother
% IVvo = nadwat2(Km, Tm, [], Kmo, Tmo, 0.05, 0.10);
% Radial Basis Function smoother
coef = rbfcreate([Km'; Tm'], IV', ...
25 'RBFFunction', 'multiquadric', 'RBFSmooth', 0.25);
IVvo = rbfinterp([Kmo'; Tmo'], coef');

% interpolate risk free rate for output
rvo = interp1(Tu, r(Iu), Tvo, 'linear', 'extrap');
30 % compute prices and reshape to matrices
Pvo = bs_greeks(S0(1), Kvo, rvo, IVvo, Tvo, 1);
IVo = reshape(IVvo, [NTo, NKo]);
Po = reshape(Pvo, [NTo, NKo]);
ro = reshape(rvo, [NTo, NKo]);
```

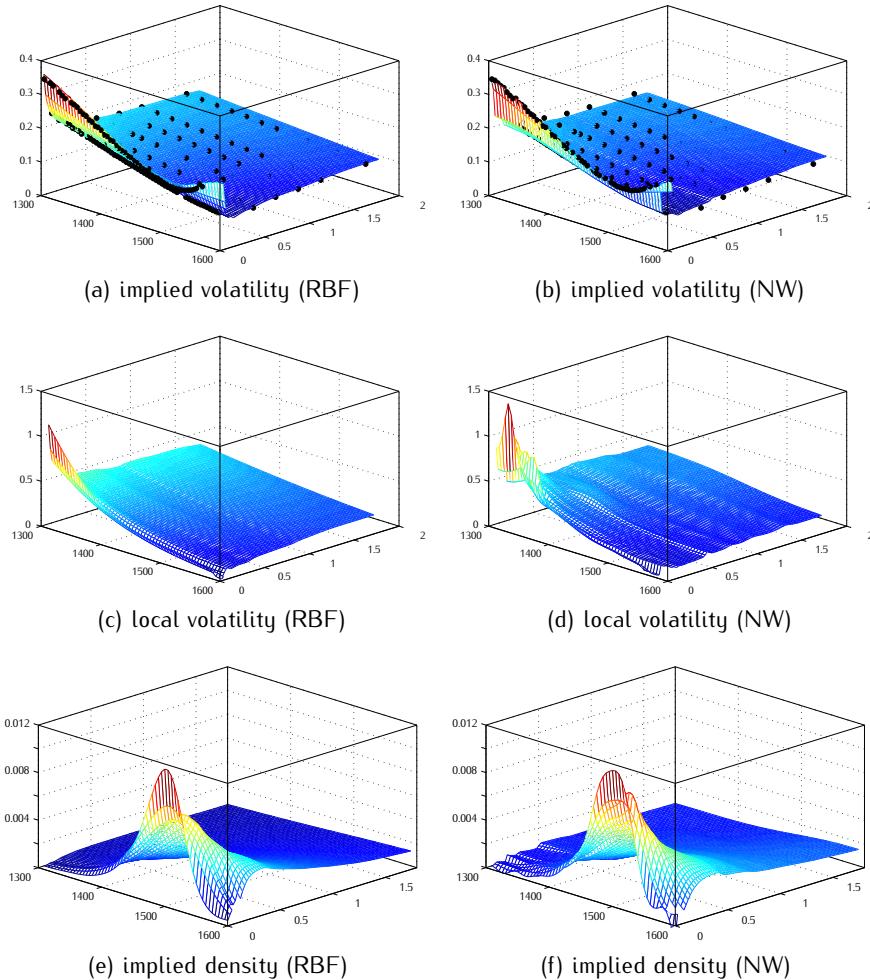
the *radial basis function* (RBF) interpolation, where we reconstruct an unknown function using the form

$$f(x) = c_0 + \mathbf{c}'x + \sum_{n=1}^N \lambda_n \phi(\|x - x_n\|)$$

The points that we observe are given at the nodes  $x_n$ , for  $n = 1, \dots, N$ . The radial function  $\phi(x)$  will determine how the impact of the value at each node behaves. Common radial functions include the Gaussian  $\phi(x) = \exp(-x^2/(2\sigma^2))$  and the

for copies, comments, help etc. visit <http://www.theponytail.net/>

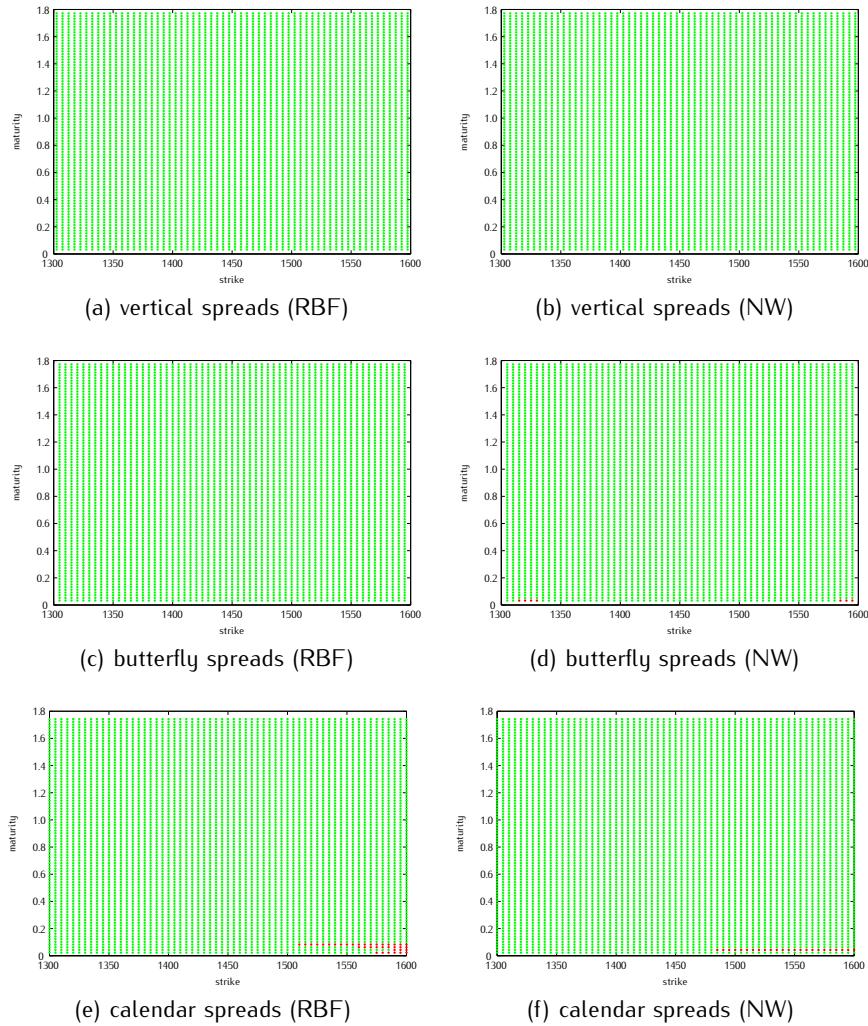
FIGURE 6.8: Implied volatilities smoothed with the radial basis function (RBF, left) and the Nadaraya-Watson (NW, right) methods. The corresponding local volatility surfaces and the implied probability density functions for different horizons are also presented.



multiquadratic function  $\phi(x) = \sqrt{1 + (x/\sigma)^2}$ , among others.<sup>21</sup> The values of the parameters  $c_0, c$  and  $\lambda_n$  are determined using the observed value function at the nodes  $x_n$  and the required degree of smoothness. Figure 6.8(a) presents a set

<sup>21</sup> The parameter  $\sigma$  is user defined. In Matlab the RBF interpolation is implemented in the package of Alex Chirokov that can be download at <http://www.mathworks.com/matlabcentral/>.

FIGURE 6.g: Static arbitrage tests for the smoothed implied volatility functions of figure 6.8. Vertical, butterfly and calendar spreads are constructed and their prices are examined. Green dots represent spreads that have admissible prices, while red dots indicate spreads that offer arbitrage opportunities as they are violating the corresponding bounds.



of observed implied volatilities together with the smoothed surface constructed using the RBF interpolation method. The implementation is given in listing 6.8.

The Nadaraya-Watson (NW) smoother is another popular choice. Here the approximating function takes the form

for copies, comments, help etc. visit <http://www.theponytail.net/>

LISTING 6.9: test\_vol.m: Tests for static arbitrage.

```
%test_vol.m
imp_vol; % load data and smooth volatility surface
% vertical spreads
VS = (Po(:,1:end-1)-Po(:,2:end))/dK;
5 VS = (VS>=0)&(VS<=1);
% butterfly spreads
BS = Po(:,3:end)-2*Po(:,2:end-1)+Po(:,1:end-2);
BS = (BS>=0);
% calendar spreads
10 CS = Po(2:end,:)-Po(1:end-1,:);
CS = (CS>=0);
```

$$f(x) = \frac{\sum_{n=1}^N w_n y_n \exp(-x' \mathbf{H} x)}{\sum_{n=1}^N w_n \exp(-x' \mathbf{H} x)}$$

where  $y_n$  is the observed value at the point  $x_n$ , and the matrix  $\mathbf{H} = \text{diag}(h_1, \dots, h_n)$  is user defined. This is implemented for the two-dimensional case in listing 6.7. Figure 6.8(b) gives the implied volatility surface smoothed using the Nadaraya-Watson method.

Of course the smoothed or interpolated volatility surface can be mapped to call and put prices using the Black-Scholes formula. There is also a number of restrictions that one needs to take into account when constructing the volatility surface. In particular, it is important to verify that the resulting prices do not permit arbitrage opportunities. As shown in Carr and Madan (2005) it is straightforward to rule out static arbitrage by checking the prices of simple vertical spreads, butterflies and calendar spreads. More precisely, having constructed a grid of call prices for different strikes  $0 = K_0, K_1, K_2, \dots$  and maturities  $0 = T_0, T_1, T_2, \dots$ , with  $C_{i,j} = f_{BS}(t, S; K_i, T_j, r, \sigma(K_i, T_j))$ , we need to construct the following quantities

1. Vertical spreads  $VS_{i,j} = \frac{C_{i-1,j} - C_{i,j}}{K_i - K_{i-1}}$ . There should be  $0 \leq VS_{i,j} \leq 1$  for all  $i, j = 0, 1, \dots$
2. Butterfly spreads  $BS_{i,j} = C_{i-1,j} - \frac{K_{i+1} - K_{i-1}}{K_{i+1} - K_i} C_{i,j} + \frac{K_i - K_{i-1}}{K_{i+1} - K_i} C_{i+1,j}$ . There should be  $BS_{i,j} \geq 0$  for all  $i, j = 0, 1, \dots$
3. Calendar spreads  $CS_{i,j} = C_{i,j+1} - C_{i,j}$ . There should be  $CS_{i,j} \geq 0$  for all  $i, j = 0, 1, \dots$

In figure 6.9 we construct these tests for the resulting volatility surfaces based on the two smoothing methods, implemented in listing 6.9. With green dots we denote the points where no arbitrage opportunities exist, while red dots represent arbitrage opportunities. Both RBF and NW methods yield prices that pass the vertical spread tests. The NW smoother produces a very small number of very short away-from-the-money prices that allow the setup of butterfly spreads

with negative value. Both methods fail the calendar spread test for far-out-of-the-money calls with very short maturities. Nevertheless, the bid-ask spreads in these areas are wide enough to ensure that these opportunities are not actually exploitable. Overall the results are very good, but if needed one can incorporate these tests within the fitting procedures, and thus find smoothed volatility surfaces that by construction pass all three arbitrage tests.

Another important feature of the implied volatility is that it should behave in a linear fashion for extreme log-strikes (Lee, 2004a; Gatheral, 2004). This indicates that it makes sense to extrapolate the implied volatility linearly to extend outside the region of observed prices.

Apart from these nonparametric methods one can set up parametric curves to fit the implied volatility skew. A parametric form might be less accurate, but it can offer more a robust fit where the resulting prices are by construction free of arbitrage. Gatheral (2004) proposes an implied variance function for each maturity horizon, coined the *stochastic volatility inspired* (SVI) parameterization, of the form

$$v(k; \alpha, \beta, \sigma, \rho, \mu) = \alpha + \beta \left( \rho(k - \mu) + \sqrt{(k - \mu)^2 + \sigma^2} \right)$$

where  $k = \log(K/F)$ . This form always remains positive and ensures that it grows in a linear fashion for extreme log-strikes. In particular Gatheral (2004) shows that  $\alpha$  controls for the variance level,  $\beta$  controls the angle between the two asymptotes,  $\sigma$  controls the smoothness around the turning point,  $\rho$  controls the orientation of the skew, and  $\mu$  shifts the skew across the moneyness level.

### IMPLIED DENSITIES

Based on the implied volatility function  $\hat{\sigma}(T, K)$  the empirical pricing function is easily determined via the Black-Scholes formula

$$P(T, K) = f_{BS}(t, S_0; T, K, r, \hat{\sigma}(T, K))$$

It has been recognized, since Breeden and Litzenberger (1978), that the empirical pricing function can reveal information on the risk neutral probability density that is *implied* by the market. In particular, if  $\mathbb{Q}_t(S)$  is this risk neutral probability measure of the underlying asset with horizon  $t$ , then the call price can be written as the expectation

$$P(T, K) = \exp(-rT) \int_K^\infty (S - K) d\mathbb{Q}_T(S)$$

If we differentiate twice with respect to the strike price, using the Leibniz rule

$$\frac{\partial}{\partial t} \int_{\alpha(t)}^{\beta(t)} g(x, t) dx = g(\beta(t), t) \frac{d\beta(t)}{dt} - g(\alpha(t), t) \frac{d\alpha(t)}{dt} + \int_{\alpha(t)}^{\beta(t)} \left[ \frac{\partial}{\partial t} g(x, t) \right] dx$$

LISTING 6.10: `loc_vol.m`: Construction of implied densities and the local volatility surface.

```
%loc_vol.m
imp_vol; % load data and smooth volatility surface
% 1st deriv wrt T
D1T = (Po(3:end,:)-Po(1:end-2,:))/(2*dT);
% 1st deriv wrt K
D1K = (Po(:,3:end)-Po(:,1:end-2))/(2*dK);
% 2nd deriv wrt K
D2K = (Po(:,3:end)-2*Po(:,2:end-1)+Po(:,1:end-2))/dK^2;
% implied probability density function
F = exp(-ro(:,2:end-1).*Tgo(:,2:end-1)).*D2K;
% local volatility function
LV = Po(2:end-1,2:end-1) - ...
    Kgo(2:end-1,2:end-1).*D1K(2:end-1,:);
LV = D1T(:,2:end-1) + ro(2:end-1,2:end-1).*LV;
LV = LV./(0.5*Kgo(2:end-1,2:end-1).^2.*D2K(2:end-1,:));
```

we obtain the [Breeden and Litzenberger \(1978\)](#) expression for the implied probability density function

$$d\Omega_T(S) = \exp(rT) \left. \frac{\partial^2 P(T, K)}{\partial K^2} \right|_{K=S} \quad (6.1)$$

It is easy to compute this derivative numerically, and therefore approximate the implied density using central differences. In particular

$$d\Omega_T(S) \approx \exp(rT) \frac{P(T, S - \Delta K) - 2P(T, S) + P(T, S + \Delta K)}{(\Delta K)^2}$$

One can recognize that the above expression is the price of  $1/(\Delta K)^2$  units of a very tight butterfly spread around  $S$ , like the one used in the static arbitrage tests above. The relation between the butterfly spread and the risk neutral probability density is well known amongst practitioners, and can be used to isolate the exposure to specific ranges of the underlying. We carry out this approximation in listing 6.10, and the resulting densities are presented in figures 6.8(e,f).

## LOCAL VOLATILITIES

A natural question that follows is whether or not a process exists that is consistent with the sequence of implied risk neutral densities. After all, Kolmogorov's extension theorem 1.3 postulates that given a collection of transition densities such a process might exist. [Dupire \(1994\)](#) recognized that one might be able to find a *diffusion* which is consistent with the observed option prices, constructing

the so called *local volatility model*, where the return volatility is a deterministic function of time and the underlying asset. In a series of papers Derman and Kani (1994), Derman, Kani, and Chriss (1996), and Derman, Kani, and Zou (1996) outline the use of the local volatility function for pricing and hedging, while Barle and Cakici (1998) present a method of constructing an implied trinomial tree that is consistent with observed option prices.

The dynamics of the underlying asset (under the risk neutral measure) are given by

$$dS_t = rS_t dt + \sigma(t, S_t) S_t dB_t \quad (6.2)$$

The popularity of the local volatility approach stems from the fact that the steps taken in the derivation of the Black-Scholes PDE can be replicated, since the *local volatility function*  $\sigma(t, S)$  is deterministic. In particular, the markets remain complete as there is only one source of uncertainty that can be hedged out using the underlying asset and the risk free bank account.

The pricing function for any derivative under the local volatility dynamics will therefore satisfy a PDE that resembles the Black-Scholes one

$$\frac{\partial}{\partial t} f(t, S) + rS \frac{\partial}{\partial S} f(t, S) + \frac{1}{2} \sigma^2(t, S) S^2 \frac{\partial^2}{\partial S^2} f(t, S) = rf(t, S)$$

Of course, having a functional form for the volatility will mean that closed form expressions are unattainable even for plain vanilla contracts. Nevertheless, it is straightforward to modify the finite difference methods that we outlined in chapter 4 (for example the  $\theta$ -method in listing 3.3) to account for the local volatility structure.

Dupire (1993) notes that if the diffusion (6.2) is consistent with the risk neutral densities (6.1), then the risk neutral densities must satisfy the forward Kolmogorov equation (see section 1.6). In particular, if we denote the transition density with  $f^Q(t, K; T, S) = Q(S_T \in dS | S_t = K)$ , then the forward Kolmogorov equation will take the form

$$\begin{aligned} \frac{\partial}{\partial T} f^Q(t, K; T, S) \\ = -\frac{\partial}{\partial K} [rK f^Q(t, K; T, S)] + \frac{1}{2} \frac{\partial^2}{\partial K^2} [\sigma^2(T, K) f^Q(t, K; T, S)] \end{aligned}$$

Given the Breeden-Litzenberger representation of the densities (6.1), we can write

$$f^Q(t, K; T, S) = \exp(rT) \frac{\partial^2 P(T, K)}{\partial K^2}$$

By taking the derivative with respect to  $T$ , and substituting in the forward equation we have, after some simplifications, the following

$$\begin{aligned} r \frac{\partial^2 P(T, K)}{\partial K^2} + \frac{\partial^3 P(T, K)}{\partial T \partial K^2} + \frac{\partial}{\partial K} \left[ rK \frac{\partial^2 P(T, K)}{\partial K^2} \right] \\ - \frac{1}{2} \frac{\partial^2}{\partial K^2} \left[ \sigma^2(T, K) K^2 \frac{\partial^2 P(T, K)}{\partial K^2} \right] = 0 \end{aligned}$$

We can integrate the above expression twice with respect to  $K$  which will eventually yield the PDE<sup>22</sup>

$$\frac{\partial P(T, K)}{\partial T} + rK \frac{\partial P(T, K)}{\partial K} - \frac{1}{2} \sigma^2(T, K) K^2 \frac{\partial^2 P(T, K)}{\partial K^2} = c_0(T)K + c_1(T)$$

The functionals  $c_0(T)$  and  $c_1(T)$  appear as integration constants with respect to  $K$ , and need to be identified using some boundary behavior. In particular, we can use that as the strike price increases,  $K \rightarrow \infty$ , the call prices and all derivatives will decay to zero. This will happen because the risk neutral probability density  $f^Q(t, K; T, S)$  decays as  $S \rightarrow \infty$ . In that case the left hand side that involves the derivatives will equal to zero, which implies that  $c_0(T) = c_1(T) = 0$  for all maturities  $T$ . The Dupire PDE is therefore

$$\frac{\partial P(T, K)}{\partial T} + rK \frac{\partial P(T, K)}{\partial K} - \frac{1}{2} \sigma^2(T, K) K^2 \frac{\partial^2 P(T, K)}{\partial K^2} = 0$$

This partial differential equation resembles the Black-Scholes PDE, and is actually its *adjoint* in the sense that Kolmogorov's backward and forward equations are. The Black-Scholes PDE will give the evolution of the call price as we approach maturity and as the underlying asset changes, keeping the strike and maturity constant. The Dupire PDE is satisfied by a call option as the maturity and the strike price change, keeping the current time and current spot price constant.

We can solve the above expression for the local volatility function

$$\sigma(T, K) = \sqrt{\frac{\frac{\partial P(T, K)}{\partial T} + rK \frac{\partial P(T, K)}{\partial K}}{\frac{1}{2} K^2 \frac{\partial^2 P(T, K)}{\partial K^2}}}$$

The above links the local volatility model with prices of observed call options<sup>23</sup> and in principle it could be used to extract the local volatility function  $\sigma(t, S)$  from a set of observed contracts. Unfortunately, there is a number of practical problems with this approach, which stem from the fact that the local volatility is a function of first and second derivatives of the pricing function  $P(T, K)$ . For a start, there is only a relatively small number of calls and puts available at any point in time, which means that we will need to set up some interpolation before we carry out the necessary numerical differentiation using finite differences. Therefore, our results will be dependent on the interpolation scheme that we use.

In addition, the observed option prices are "noisy", and interpolating through their values will cause its own problems. Numerical differentiation is unstable at the interpolating nodes, and attempting to take the second derivative is a guarantee for disaster, with the resulting local volatility surfaces varying

---

<sup>22</sup> During the second integration we use the identity  $K \frac{\partial^2 P(T, K)}{\partial K^2} = \frac{\partial}{\partial K} \left[ K \frac{\partial P(T, K)}{\partial K} \right] - \frac{\partial P(T, K)}{\partial K}$ .

<sup>23</sup> And also put options through the put-call parity.

wildly. For these reasons practitioners prefer to use a smoothing method, like the Nadaraya-Watson or the Radial Basis Function that we outlined above. The construction of the local volatility is given in listing 6.10, together with the density extraction. Figures 6.8(c,d) give the local volatility surface for the two smoothing procedures. One can observe that the two surfaces look very similar, with the RBF producing somewhat smoother time derivatives. Of course this will depend largely on the parameters that define the smoothing procedure, which are chosen *ad hoc*.

The local volatility function can also play the role of the risk neutral estimator of the instantaneous future volatility at time  $T$ , if the underlying asset level at this future time is equal to  $K$ . As shown in Derman and Kani (1998)

$$\sigma^2(T, K) = E^Q((dS_T)^2 | S_T = K) = E^Q \left( \lim_{\Delta t \downarrow 0} (S_{T+\Delta t} - K)^2 | S_T = K \right)$$

If one assumes a form for the implied volatility function, either using an interpolator or a smoother, it is possible to express the local volatility  $\sigma(t, S)$  in terms of the implied volatility  $\hat{\sigma}(T, K)|_{T=t, K=S}$ . This is of course feasible since the pricing function

$$P(T, K) = f_{BS}(t, S_0; T, K, r, \hat{\sigma}(T, K))$$

which can be differentiated analytically with respect to the strike  $K$  and the maturity  $T$ . It is actually more convenient to work with the moneyness  $y = \log(K/F) = \log(K/S) + rT$ , and also consider the *implied total variance* as a function of the maturity and the moneyness  $w(T, y) = T\hat{\sigma}^2(T, K)$ . Then, as shown in Gatheral (2006) the local variance can be easily computed as

$$\sigma^2(T, y) = \frac{\frac{1}{T} \frac{\partial w(T, y)}{\partial T}}{1 - \frac{y}{w(T, y)} \frac{\partial w(T, y)}{\partial y} + \frac{4y^2 - 4w(T, y) - w^2(T, y)}{16w^2(T, y)} \left( \frac{\partial w(T, y)}{\partial y} \right)^2 + \frac{1}{2} \frac{\partial^2 w(T, y)}{\partial y^2}}$$



## Fixed income securities

Fixed income securities<sup>1</sup> promise to pay a stream of fixed amount at predefined points in time. Typically, the issuers of these securities are either governments (*sovereign bonds*) or corporations (*corporate bonds*). Bonds are debt instruments, used by governments or corporations to borrow money from investors.

*Zero-coupon bonds* offer a single fixed payment, called the *face value* of the bond  $\Pi$ , on the maturity date  $T$ . *Coupon bearing bonds* also promise to pay a stream of cash flows, the *coupons*, in addition to the face value. In particular, a  $c\%$ -coupon bond will pay an amount equal to  $\frac{c}{100}\Pi$  of the face value per year. Typically payments are made in two semi-annual installments of  $\frac{1}{2}\frac{c}{100}\Pi$  each.

Instruments with short maturities, like the US Treasury bills, are typically zero coupon. Longer maturity instruments, like the US Treasury bonds, are typically coupon bearing. Corporate bonds also typically bear coupons. In most cases, when a new coupon bearing instrument is introduced, the coupons are chosen as for the instrument to sell *at par*. This means that its initial price is approximately equal to its face value. Then, the coupon reflects the rate of interest: for example if a sovereign 6% 30 year bond will face value \$100 is issued and sells at par, then the buyer will lend the government today \$100, and will receive  $\frac{1}{2} \times \frac{6}{100} \times \$100 = \$3$  every six months for the next 30 years, plus \$100 on the bond maturity.

### 7.1 YIELDS AND COMPOUNDING

Of course, the bond will almost never sell at exactly its par value. The *yield* of the bond is the equivalent constant rate of interest that is able to replicate all cashflows to maturity, when investing an amount equal to the current bond price. It is obvious that the frequency that one reinvests the proceeds will be a

---

<sup>1</sup> In this chapter we will call all fixed income securities *bonds*, although in reality the word "bond" is reserved for instruments with relatively long maturities. Shorter instruments in the US are called "bills" or "notes".

factor that will affect the bond yield. When the yield of an instrument is quoted, it is important to know what compounding method has been used, in order to truly compare bonds.

In particular, let  $P_t$  denote the price of an instrument at time  $t$  (measured in years). The *simple yield*  $y_1(t_1, t_2)$ , between two dates  $t_1$  and  $t_2$ , satisfies

$$\frac{P_{t_2}}{P_{t_1}} = 1 + y_1(t_1, t_2)(t_2 - t_1) \Rightarrow y_1(t_1, t_2) = \frac{1}{t_2 - t_1} \left( \frac{P_{t_2}}{P_{t_1}} - 1 \right)$$

The simple yield is the return of an investment equal to  $P_{t_1}$  that is initiated at time  $t_1$ , and is then liquidated at time  $t_2$  for a price  $P_{t_2}$ . There is no intermediate reinvestment of any possible proceedings.

Of course one could sell the instrument at the intermediate time  $t^* = \frac{t_1+t_2}{2}$  for a price  $P_{t^*}$ , and reinvest this amount for the remaining time to  $t_2$ . Say that the yield of this strategy is denoted  $y_2(t_1, t_2)$ . In that case the two simple investments will satisfy

$$\begin{aligned} \frac{P_{t_2}}{P_{t^*}} &= 1 + y_2(t_1, t_2)(t_2 - t^*) \\ \frac{P_{t^*}}{P_{t_1}} &= 1 + y_2(t_1, t_2)(t^* - t_1) \end{aligned}$$

Multiplying the will give the yield if we compound twice during the life of the bond, namely

$$\frac{P_{t_2}}{P_{t_1}} = \left( 1 + y_2(t_1, t_2) \frac{t_2 - t_1}{2} \right)^2 \Rightarrow y_2(t_1, t_2) = \frac{2}{t_2 - t_1} \left[ \left( \frac{P_{t_2}}{P_{t_1}} \right)^{1/2} - 1 \right]$$

More generally, if we compound  $m$  times over the life of the bond we can follow the same procedure to deduce the yield  $y_m(t_1, t_2)$

$$\frac{P_{t_2}}{P_{t_1}} = \left( 1 + y_m(t_1, t_2) \frac{t_2 - t_1}{m} \right)^m \Rightarrow y_m(t_1, t_2) = \frac{m}{t_2 - t_1} \left[ \left( \frac{P_{t_2}}{P_{t_1}} \right)^{1/m} - 1 \right]$$

If we pass to the limit  $m \rightarrow \infty$  we can recover the *continuously compounded return*  $y_\infty(y_1, y_2)$  using the fact that  $\lim_{m \rightarrow \infty} (1 + a/m)^m = \exp(a)$

$$\begin{aligned} \frac{P_{t_2}}{P_{t_1}} &= \lim_{m \rightarrow \infty} \left( 1 + y_\infty(t_1, t_2) \frac{t_2 - t_1}{m} \right)^m = \exp(y_\infty(t_1, t_2)(t_2 - t_1)) \\ &\Rightarrow y_\infty(t_1, t_2) = \frac{1}{t_2 - t_1} \log \left( \frac{P_{t_2}}{P_{t_1}} \right) \end{aligned}$$

We will work with the continuous compounded return from now on, and we will drop the subscript  $\infty$ , writing instead  $y_\infty(t, T) = y(t, T) = y_t(T)$ . We will also denote with  $P(t, T) = P_t(T)$  the price at time  $t$  that matures at time  $T$ .

Different instruments are quoted using different market conventions. In addition to the compounding, there are also conventions in the way time intervals are computed. For example, some instruments are quoted assuming that the month has 30 days and the year 360 (the 30/360 convention). This means that in order to convert months to days we assume that each month has 30 days, and to convert days to years we assume that each year has 360 days. Other instruments might be quoted using the ACT/365, ACT/ACT or more rarely the conventions.<sup>2</sup>

If we are dealing with coupon bearing bonds, then we would need to decompose the coupon payments and take their present values individually.

As an example, say that we are interested in a zero-coupon bond that has 22 months and 12 days to maturity, and we are quoted a yield of 5.5%. What is the value of this bond today, assuming that the face value is \$100? If this bond is compounded semiannually and the 30/360 convention is used, then we would compute the time interval as

$$22m + 12d = 1y + 10m + 12d = 1y + 300d + 12d = 1y + 312d = 1.866y$$

Compounding will take place at the points  $t_0 = 0.366y$ ,  $t_1 = 0.866y$ ,  $t_2 = 1.366y$  and  $t_3 = 1.866y$ . Therefore, if we denote with bond prices will satisfy

$$\frac{100}{P_{t_1}} = (1 + 0.055 \times 0.366)(1 + 0.055 \times 0.5)^3 =$$

The financial toolbox of Matlab has a number of functions that convert between different day count conventions, and the appropriate discount factors. Here, since the markets are set up in continuous time we will use continuous compounding.

## 7.2 THE YIELD CURVE

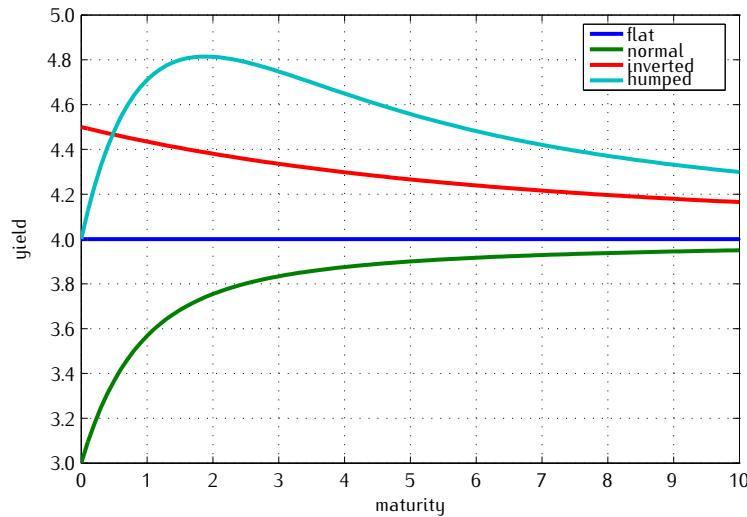
At each point in time  $t$  we have the opportunity to invest in instruments of different maturities  $\tau$ , each offering a particular yield  $y(t, \tau)$ . The mapping  $\tau \rightarrow y(t, t+\tau)$  is called the *yield curve*. Essentially it represents the annualized return that is guaranteed by a zero-coupon bond with maturity  $\tau$ . Observed yield curves are typically upward sloping, with the yields for long maturities being higher than the short ones. Such yield curves are called *normal* to illustrate that this pattern is the most common. Having said that, *flat* or *inverted* (downward sloping) yield curves are also occasionally observed. A *humped* yield curve pattern is rarely encountered.

### THE NELSON-SIEGEL-SVENSSON PARAMETRIZATION

---

<sup>2</sup> More information can be found at the International Swaps and Derivatives Association website ([www.isda.org](http://www.isda.org)), and in particular in ISDA (1998).

FIGURE 7.1: Examples of yield curves using the Nelson-Siegel-Svensson parametrization. The parametric form is able to produce curves that exhibit the basic yield curve shapes.



Nelson and Siegel (1987) and Svensson (1994), collectively denoted with NSS, discuss various parametric forms of the yield curve, summarized in the form

$$\tau \rightarrow \beta_0 + \beta_1 \frac{1 - e^{-\tau/\tau_1}}{\tau/\tau_1} + \beta_2 \frac{e^{-\tau/\tau_2}(1 - e^{-\tau/\tau_2})}{\tau/\tau_2}$$

There are five parameters in the NSS expression, which control different aspects of the yield curve shape. In particular,  $\beta_0$  can be used to shift the yield curve up and down, therefore defining the yields for long maturities.  $\beta_1$  controls the amount of curvature that the curve exhibits, while  $\beta_3$  is responsible for a potential hump. The parameters  $\tau_1$  and  $\tau_2$  will determine at which maturities the curvature and the hump are most pronounced. Figure 7.1 shows some basic yield curves patterns that can be produced using the NSS approach. The original paper of Nelson and Siegel (1987) used only the first two components, allowing for level and curvature effects. Svensson (1994) augmented the formula with the hump component.

In reality we only observe yields for a relatively small number of maturities, and the quotes can be contaminated with noise. This can be due to non synchronous trading, illiquidity and other microstructure issues. The NSS functionals can be used to smooth the observed yield curve, and to interpolate for maturities that are not directly traded. Also, generalized versions of the NSS approach can be used to study the *dynamics* of the yield curve in time.

LISTING 7.1: `nelson_siegel_svensson.m`: Yields based on the Nelson-Siegel-Svensson parametrization.

```
% nelson_siegel_svensson.m
function y = nelson_siegel_svensson(T, par)
b0 = par.beta0;
b1 = par.beta1;
5 b2 = par.beta2;
t1 = par.tau1;
t2 = par.tau2 + eps;
y = b0 + b1*(1-exp(-T./t1))./(T./t1) ...
+ b2*exp(-T./t2).* (1-exp(-T./t2))./(T./t2);
```

Listing 7.1 gives a simple Matlab code that implements the NSS formula. Individual yield curves can be used to calibrate this formula and retrieve the corresponding parameters.

In this chapter we are interested in the construction of mathematical models that have two desirable and very significant features. On one hand, they should have the potential to reproduce observed yield curves. In addition, they must be able to capture the evolution of the yield curve through time, in order to offer reliable prices for derivative contracts that are based on future yields or bond prices.

### THE DYNAMICS OF THE YIELD CURVE

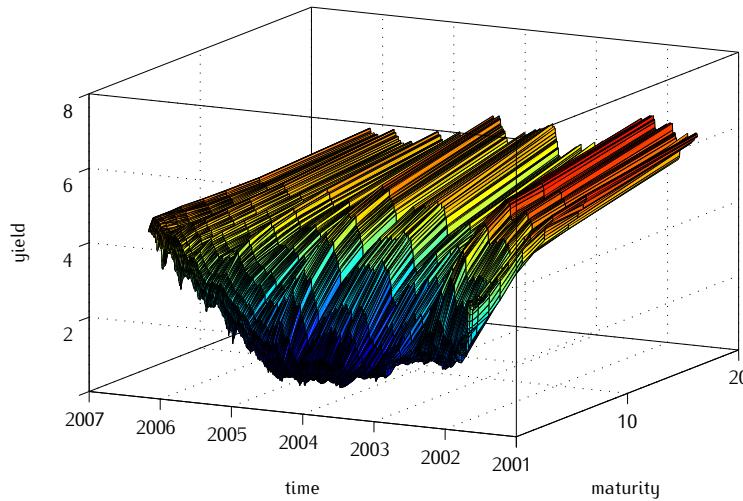
Figure 7.2 gives historical yield curves over the period 2001-07. A few casual observations can be made, which can offer valuable insight on the stylized facts that a fixed income model should adhere to. It appears that the yield curve is indeed typically upward sloping, with a few instances where it is flat or slightly inverted. There is no significant “hump-ness” in this particular dataset. The short end of the yield curve appears to be a lot more volatile than the long end, which is relatively stable. Also, yields of different maturities do not tend to move in opposite directions. In contrast they seem to be quite strongly correlated.

We can use these yields to recover the parameters of the NSS formula. In this particular instance we assume that  $\beta_2 = \tau_2 = 0$ , as the yields in the dataset do not exhibit a humped pattern. An example of how one can calibrate these parameters is given in listing 7.2. Figure 7.3 shows these parameter estimates through time.

The parameter  $\beta_0$  corresponds to the maximum yield across different maturities. As the long term bond yields gradually dropped throughout the sample period,  $\beta_0$  also decreases to reflect that. As illustrated by the time path of the parameter  $\beta_1$ , the yield curve became slightly more convex in the first half of the period, flattening quite rapidly afterwards. Parameter  $\tau_1$  shows that the short end of the yield curve rose steeply between mid-2002 to mid-2003.

for copies, comments, help etc. visit <http://www.theponytail.net/>

FIGURE 7.2: Historical yield curve dynamics. The corresponding Matlab code can be found in listing 7.7.



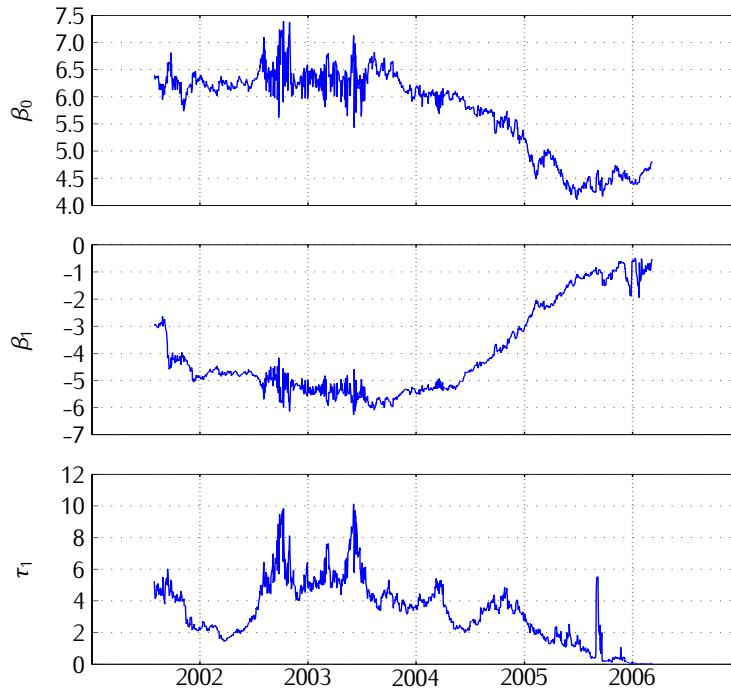
LISTING 7.2: `calibrate_ns.m`: Calibration of the Nelson-Siegel formula to a yield curve.

```
% calibrate_ns.m
function [z, par] = calibrate_ns(T, Y)
par = 0;
opt = optimset('Display', 'none');
5 z = lsqnonlin(@ssq, [mean(Y), 0.1, 0.1], ...
[], [], opt);
    function sq = ssq(p)
        par.beta0 = p(1);
        par.beta1 = p(2);
10       par.beta2 = 0;
        par.tau1 = p(3);
        par.tau2 = 0;
        sq = nelson_siegel_svensson(T, par) - Y;
    end
15 end
```

## THE FORWARD CURVE

Different points on the yield curve provides us with the risk free rates of return, for investments that commence 'now' (that is time  $t$  in our notation), and mature at different times in the future. The yield curve also defines the *forward rates*,

FIGURE 7.3: Historical Nelson-Siegel parameters. The [Nelson and Siegel \(1987\)](#) formula  $\beta_0 + \beta_1(1 - \exp(-\tau/\tau_1)) / (\tau/\tau_1)$  is calibrated to the yields of figure 7.2 and the parameters are presented below. The level parameter  $\beta_0$ , the convexity magnitude parameter  $\beta_1$  and the convexity steepness parameter  $\tau_1$  are given.



which are the fixed rates of return that are set and reserved at time  $t$ , but will be applicable over a future time period.

In particular, say that we select two points on the yield curve, for bonds that mature at times  $T^*$  and  $T$ , with  $T > T^* > t$ . The prices of these bonds will be  $P_t(T)$  and  $P_t(T^*)$ , respectively. Now assume that we are interested in setting the forward (continuously compounded) rate of interest for an investment that will commence at time  $T^*$  and will mature at  $T$ , which we will denote with  $f_t(T^*, T)$ .

Consider the following two investments over the period  $[t, T]$ :

1. Buy one risk free bond that matures at time  $T$ . This will cost  $P_t(T)$  today, and will deliver one pound at time  $T$ .
2. Buy  $P_t(T)/P_t(T^*)$  units of the risk free bond that matures at time  $T^*$ . Also enter a forward contract to invest risk-free over the period  $[T^*, T]$ , at the rate  $f_t(T^*, T)$ . This strategy will also cost  $P_t(T)$  today, as it is free to enter a forward contract. The first leg will deliver  $P_t(T)/P_t(T^*)$  pounds at time  $T^*$ , which will be invested at the forward rate. Therefore at time  $T$  this strategy will deliver  $\exp\{f_t(T^*, T) \cdot (T - T^*)\} \cdot P_t(T)/P_t(T^*)$ .

These two strategies have the same initial cost to set up, the same maturity, and are both risk free. Therefore they should deliver the same amount on the maturity date  $T$ , otherwise arbitrage opportunities would arise. For example, if the second strategy was delivering more than one pound at time  $T$ , then one would borrow  $P_t(T)$  at the risk free rate to enter the second strategy with zero cost at time zero.

Therefore, the arbitrage free forward rate will satisfy

$$\begin{aligned} P_t(T) &= P_t(T^*) \cdot \exp \{-f_t(T^*, T) \cdot (T - T^*)\} \\ \Rightarrow f_t(T^*, T) &= -\frac{1}{T - T^*} \log \frac{P_t(T)}{P_t(T^*)} \end{aligned}$$

If we let the time between the two maturities shrink down to zero, by letting for example  $T^* \rightarrow T$ , we define the (*instantaneous*) *forward rate*. This is essentially the short rate that we can reserve today, but will applied at time  $T$

$$\begin{aligned} f_t(T) &= -\lim_{T^* \uparrow T} \frac{\log P_t(T) - \log P_t(T^*)}{T - T^*} \\ &= -\frac{\partial \log P_t(T)}{\partial T} = y_t(T) + (T - t) \frac{\partial y_t(T)}{\partial T} \end{aligned}$$

Forward rates for different maturities define the *forward curve*. There is a correspondence between the yield and forward curves, and knowing one leads to the other.

## 7.3 THE SHORT RATE

Historically, the first family of models introduced in the fixed income literature were the so called *short rate* or *one-factor models*. The main underlying assumption is that there is a unique Brownian motion that is responsible for the uncertainty in the economy. More formally, we start with a filtered probability space  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ , and consider a Brownian motion  $B_t$  with respect to this probability measure.

The main ingredient of the one-factor model is the *short rate process* that is to say the process of the instantaneous risk free rate  $r_t$ . Essentially, this is the rate offered by the *bank account* or *current account*, which is not fixed for any period of time, but is nevertheless risk free during the infinitesimal period  $(t, t + dt)$ . The investor is not bound for any maturity and can withdraw funds from (or add funds to) this account freely, without incurring any penalties. This is in contrast to other financial assets, like the ones introduced in the Black-Scholes paradigm, where the return over this infinitesimal period is random. Of course, investing in the bank account over a longer period of time is not a risk free investment, since the short rate will change. Having said that, one can show that if the short rate is process that drives the economy, then bonds

with different maturities can be priced in a consistent way that does not permit arbitrage opportunities. This means that eventually we will show that all bonds can be priced relative to each other in a unique way. Intuitively, one can think of bonds as *derivatives* which are contingent on the future realizations of the short rate.

To put things more concretely, the current account will satisfy the ordinary differential equation

$$dB_t = B_t r_t dt \Rightarrow B_t = B_0 \exp \left\{ \int_0^t r_s ds \right\}$$

### SHORT RATE AND BOND PRICING

As we argued above, the short rate process evolves in a stochastic way, and the uncertainty is described by a Brownian motion  $B_t$ . We can therefore cast the short rate process as a SDE

$$dr_t = \mu(t, r_t) dt + \sigma(t, r_t) dB_t$$

Our objective is to establish prices for bonds with different maturities. The only constraint that we need to take into account is that the prices of these bonds must rule out any arbitrage opportunities. In all generality, the price (at time  $t$ ) of a bond with maturity  $T$  can depend at most on the time  $t$  and the short rate level  $r_t$ , that is to say

$$P_t(T) = g(t, r_t; T)$$

This formalizes the statement we made above, that the bond is a derivative on the short rate. It appears that the setting is similar to the one in equity derivative pricing, if we consider the short rate as the underlying asset. In particular we can see the analogy

$$\begin{aligned} \text{equity price: } dS_t &= \mu(t, S_t) dt + \sigma(t, S_t) dB_t \\ \text{interest rate: } dr_t &= \mu(t, r_t) dt + \sigma(t, r_t) dB_t \end{aligned}$$

In both cases we want to establish a derivative pricing relationship

$$\begin{aligned} \text{equity derivative: } P_t &= g(t, S_t) \\ \text{bond price: } P_t(T) &= g(t, r_t; T) \end{aligned}$$

Although the two settings appear to be very similar, there is a very significant difference: *Unlike equities, the short rate is not a traded asset*. This means that we cannot buy or sell the short rate, and therefore we cannot construct the necessary risk free positions that produced the Black-Scholes PDE. The market, as we constructed it, is incomplete.

In fact, the pricing of bonds has more common features with the pricing of options under stochastic volatility, where again we introduced a non-traded factor (the volatility of the equity returns). Then (section 6.3) we constructed a

portfolio of two options, in order to solve for the price of volatility risk. Here we will use the same trick, namely to construct a portfolio of two bonds with different maturities, and investigate the conditions that would make it (instantaneously) risk free. This will naturally introduce the *price of short rate risk* that will be unknown; we will be able to determine this price of risk by calibrating the model on the observed yield curve, in the same spirit as the calibration of SV models on the implied volatility surface. These are summarized in the following table

	equity SV	fixed income
non-traded asset:	volatility	short rate
used to hedge:	2 options	2 bonds
calibrate on:	IV surface	yield curve

### THE HEDGING PORTFOLIO

Let us consider two bonds with maturities  $T_1$  and  $T_2$ , and say that their prices are given by the functions  $P_t(T_j) = g(t, r_t; T_j) = g_j(t, r_t)$ , for  $j = 1, 2$ . Applying Ito's formula to the pricing function will give the dynamics of the bond prices, namely

$$dP_t(T_j) = \alpha_t(T_j)dt + \beta_t(T_j)dB_t$$

with

$$\begin{aligned}\alpha_t(T_j) &= \frac{\partial g_j(t, r_t)}{\partial t} + \mu(t, r_t) \frac{\partial g_j(t, r_t)}{\partial r} + \frac{1}{2} \sigma^2(t, r_t) \frac{\partial^2 g_j(t, r_t)}{\partial r^2} \\ \beta_t(T_j) &= \sigma(t, r_t) \frac{\partial g_j(t, r_t)}{\partial r}\end{aligned}$$

Note that both bonds will depend on the same Brownian motion  $B_t$ , as this is the only source of uncertainty that affects the bond dynamics through the short rate.

Say that we sell the first bond and buy  $\Delta_t$  units of the second one. The portfolio will have value  $\Pi_t = P_t(T_1) - \Delta_t P_t(T_2)$ , and will obey the SDE

$$d\Pi_t = dP_t(T_1) - \Delta_t dP_t(T_2)$$

Our aim is to construct a risk free portfolio; therefore, to eliminate dependence on  $dB_t$  we choose the portfolio as

$$\Delta_t = \frac{\beta_t(T_1)}{\beta_t(T_2)} = \frac{\partial g_1(t, r_t)/\partial r}{\partial g_2(t, r_t)/\partial r}$$

Then the portfolio will evolve according to the ordinary differential equation

$$d\Pi_t = \left( \alpha_t(T_1) - \frac{\beta_t(T_1)}{\beta_t(T_2)} \alpha_t(T_2) \right) dt \quad (7.1)$$

Since the portfolio is now risk free it must grow as the current account, at rate  $r_t$ . If that were not the case, arbitrage opportunities would appear. This means that

$$d\Pi_t = \Pi_t r_t dt = \left( g_1(t, r_t) - \frac{\beta_t(T_1)}{\beta_t(T_2)} g_2(t, r_t) \right) r_t dt \quad (7.2)$$

Equating (7.1) and (7.2) yields the consistency relationship

$$\frac{\alpha_t(T_1) - g(t, r_t; T_1)r_t}{\beta_t(T_1)} = \frac{\alpha_t(T_2) - g(t, r_t; T_2)r_t}{\beta_t(T_2)}$$

Now we invoke the same line of argument that we used in section 6.3. In order to set up the above relationship we did not explicitly specify a particular pair of bonds, and it will therefore hold for any pair of maturities. Thus, for any set of maturities  $T_1, T_2, T_3, T_4, \dots$  we can write

$$\begin{aligned} \frac{\alpha_t(T_1) - g(t, r_t; T_1)r_t}{\beta_t(T_1)} &= \frac{\alpha_t(T_2) - g(t, r_t; T_2)r_t}{\beta_t(T_2)} \\ &= \frac{\alpha_t(T_3) - g(t, r_t; T_3)r_t}{\beta_t(T_3)} = \frac{\alpha_t(T_4) - g(t, r_t; T_4)r_t}{\beta_t(T_4)} = \dots \end{aligned}$$

Therefore the ratio cannot depend on the particular bond maturities, it can at most depend on  $(t, r_t)$ , say that it is equal to  $\lambda(t, r_t)$ . This means that we can write

$$\frac{\alpha_t(T) - g(t, r_t; T)r_t}{\beta_t(T)} = \lambda(t, r_t)$$

for any maturity  $T$ . Essentially we have managed to derive the PDE that the bond pricing formula has to satisfy, in order to rule out arbitrage opportunities. We can thus drop the maturity  $T$ , as it is not affecting the PDE in any way, and write

$$\begin{aligned} \frac{\partial g(t, r)}{\partial t} + \{\mu(t, r) - \lambda(t, r)\sigma(t, r)\} \frac{\partial g(t, r)}{\partial r} \\ + \frac{1}{2}\sigma^2(t, r) \frac{\partial^2 g(t, r)}{\partial r^2} = g(t, r)r \end{aligned}$$

This PDE is called the *term structure PDE*, and a boundary condition is needed in order to solve it analytically or numerically. For a zero-coupon bond that matures at time  $T$  the boundary condition for this PDE will be  $g(T, r; T) = 1$ . Although the PDE is called the term structure PDE, we never used the fact the the instruments are actually bonds. The quantities  $T_j$  can be thought as indices for different interest rate sensitive instruments: bond options, caps, floors or swaptions will all satisfy the term structure PDE. In general, any contingent claim that promises to pay  $\Phi(r(T))$  at time  $T$  will satisfy the same PDE, with boundary condition

$$g(T, r) = \Phi(r)$$

### THE PRICE OF RISK

The price of risk functional  $\lambda(t, r)$  can be freely selected, as long as it does not permit arbitrage opportunities. Intuitively, it seems to be a good idea to ensure that the function of the spot rate  $r \rightarrow \lambda(t, r)$  remains bounded for all times  $t$ . This would ensure that the coefficients of the PDE will not explode at any finite time, and therefore a solution will exist.

Another way of viewing this kind of restriction is by considering the equivalent probability measure, under which pricing takes place. Essentially, if we denote with  $\mu^Q(t, r) = \mu(t, r) - \lambda(t, r)\sigma(r, t)$ , then the PDE becomes

$$\frac{\partial g(t, r)}{\partial t} + \mu^Q(t, r) \frac{\partial g(t, r)}{\partial r} + \frac{1}{2} \sigma^2(t, r) \frac{\partial^2 g(t, r)}{\partial r^2} = g(t, r)r$$

This PDE will be solved subject to the boundary condition  $g(T, r) = \phi(r)$ . For example, for zero-coupon bonds that mature at  $T$  we will have  $\phi(r) = 1$ .

The Feynman-Kac theorem postulates that the solution of this PDE can be expressed as an expectation

$$g(t, r) = E^Q \left[ \exp \left( - \int_t^T r_s ds \right) \phi(r_T) \middle| r_t = r \right]$$

If  $B^Q$  is a Brownian motion under  $Q$ , then the process for  $r_t$  is given by

$$\begin{aligned} dr_t &= \mu^Q(t, r_t) dt + \sigma(t, r_t) dB_t^Q \\ &= \mu(t, r_t) \sigma(t, r) dt + \sigma(t, r_t) \{dB_t^Q - \lambda(t, r_t)dt\} \end{aligned}$$

The probability measure  $Q$  should be equivalent to the true measure  $P$ , otherwise arbitrage opportunities would be possible (this is due to the fundamental theorem of asset pricing). We can also write  $B_t^Q = B_t + \int_0^t \lambda(s, r_s) ds$ , which suggests that in fact

$$\lambda(t, r_t) = \left. \frac{dQ}{dP} \right|_{\mathcal{F}_t}$$

That is to say the process  $\lambda_t = \lambda(t, r_t)$  is the Radon-Nikodym derivative of the risk adjusted measure with respect to the true one. In order for this to be a valid measure, the Novikov condition must be satisfied, namely that the following expectation is finite for all  $t$

$$E \exp \left( \int_0^t \lambda^2(s, r_s) ds \right) < \infty$$

It is apparent that if we require  $\lambda(t, r)$  to be bounded for all  $t$ , then the above expectation will also be bounded. This is a feature that is shared by most models for the short rate.

Since we are observing bonds which are priced under  $Q$ , it is impossible to explicitly decompose  $\lambda$  from the true short rate drift  $\mu$ . The best we can do,

given this information, is calibrating the short rate model under risk neutrality, and therefore recovering  $\mu^Q$ . If our purpose it to price interest rate sensitive securities this does not pose a problem, as pricing will also take place under  $Q$ . Having said that, we might be interested in the true short rate process, perhaps for risk management which takes place under the objective probability measure. In that case we can recover the price of risk and the true drift using filtering methods, for example a version of the Kalman filter.

There is a very extensive literature that investigates the determinants of the yield curve, trying to explain why it takes its various shapes and what makes it evolve, for example from a normal to an inverted one. The same factors will of course also influence the risk premium  $\lambda(r, t)$ . Some of the standard term structure theories include the following

1. The *pure expectations* hypothesis assumes that bonds are perfect substitutes. Bond prices are determined from the expectations of future short rates. As the short rate evolves and these expectations vary, the yield curve will shift to accommodate them. Very high spot rates could therefore imply an inverted yield curve.
2. *Market segmentation* takes an opposite view. Short and long bonds are not substitutes, due to taxation and different investor objectives. For example pension funds might only be interested in the long end of the curve, while hedge funds could be willing to invest in short maturity instruments. The prices for bonds of different yield ranges are determined independently.
3. Somewhat between the above two extreme points lies the theory of *preferred habitat*. Investors forecast future rates, but also have a set investment horizon, demanding an extra premium to invest in bonds outside their preferred maturity ranges. As short term investors outnumber long term ones, prices of long maturity bonds will be relatively lower, rendering a normal term structure. This will be inverted if expectations change sufficiently.
4. The *liquidity preferences* theory goes one step further and states that investors will demand an extra premium for having their money tied up for a longer period. Long maturity bonds will therefore have to offer higher yields to reflect this premium.

As it is naturally expected, all factor will influence the term structure behavior to some extent at each point in time.

## 7.4 ONE-FACTOR SHORT RATE MODELS

Following the discussion of the previous section, we are looking for specifications for the short rate under  $Q$ . From now one we will be working only under the risk neutral measure, therefore unless otherwise stated we will drop the superscript  $Q$ . Our objective is to consider parametric forms  $\mu(t, r)$  and  $\sigma(t, r)$  that define the dynamics of the SDE for the short rate

$$dr_t = \mu(t, r_t)dt + \sigma(t, r_t)dB_t$$

In selecting  $\mu$  and  $\sigma$  we need to keep in mind some stylized facts of interest rates, and some desirable properties of interest rate models

- The short rate and yields for all maturities are always positive.
- The process is stationary, in the sense that there is a long run distribution for the short rate. This indicates that the short rate should be allowed to increase without bound, and some sort of mean reversion should be present.
- As interest rates increase they become more volatile.
- The term structure of interest rates can be upward sloping, downward sloping or humped. The model should be capable of producing different yield curve shapes.
- The short end of the yield curve is substantially more volatile than the long end. The long end appears to evolve in a much smoother way.
- Yields for different maturities are correlated (and ones for adjacent maturities very strongly correlated), but not perfectly so.
- Finally, for an interest rate model to be operational, it should offer bond and bond derivatives in closed form (or at least in a form that is readily computable).

### THE VASICEK MODEL

The first generation of one-factor models assumed a time-homogeneous structure that lead to tractable expressions for bond prices. The [Vasicek \(1977\)](#) model casts the short rate as an Ornstein-Uhlenbeck process, namely

$$dr_t = \theta(\bar{r} - r_t)dt + \sigma dB_t$$

In the Vasicek framework the short rate is Gaussian, a feature that leads to closed form solutions for a number of instruments. For that reason the Vasicek specification is still used by some practitioners today. In particular

$$r_t | r_0 \sim N(\exp(xxxxx), \exp(xxxxx))$$

If we assume a constant price of risk  $\lambda(t, r) = \lambda$ , then under risk neutrality the dynamics of the short rate are

$$dr_t = \theta(\bar{r}^Q - r_t)dt + \sigma dB_t^Q$$

for  $\bar{r}^Q = \bar{r} + \frac{\lambda}{\theta}$ . This indicates that as investors are risk averse, they behave as if the long run attractor of the short rate is higher than what it actually is. The pricing functions of interest rate sensitive securities will satisfy the PDE

$$\frac{\partial g(t, r)}{\partial t} + \theta(\bar{r}^Q - r) \frac{\partial g(t, r)}{\partial r} + \frac{1}{2}\sigma^2 \frac{\partial^2 g(t, r)}{\partial r^2} = g(t, r)r$$

For example, in the case of a bond that matures at time  $T$  the terminal condition will be  $g(T, r; T) = 1$ , and we guess the solution of the PDE to be of the *exponential affine*

$$g(t, r; T) = \exp(C(t; T) + D(t; T) \cdot r)$$

If we substitute this expression in the PDE we can write

$$[C_t + \theta\bar{r}^Q D + \sigma^2 D^2/2] + [D_t - \theta D - 1]r = 0$$

As the PDE has to be satisfied for all initial spot rates  $r$ , we conclude that both square brackets must be equal to zero, and that  $C(T; T) = D(T; T) = 0$ . Therefore we recover a system of ODEs for the functionals  $C$  and  $D$ , namely<sup>3</sup>

$$\begin{aligned} C_t(t; T) + \theta\bar{r}^Q D(t; T) + \frac{1}{2}\sigma^2 D^2(t; T) &= 0 \\ D_t(t; T) - \theta D(t; T) &= 1 \\ C(T; T) &= 0 \\ D(T; T) &= 0 \end{aligned}$$

The solution of the above system will give the Vasicek bond pricing formula, namely

$$\begin{aligned} D(t; T) &= -\frac{1 - \exp\{\theta(T - t)\}}{\theta} \\ C(t; T) &= \frac{[D(t; T) - (T - t)] \cdot [\theta^2\bar{r}^Q - \sigma^2/2]}{\theta^2} - \frac{\sigma^2 D^2(t; T)}{4\theta} \end{aligned}$$

One important feature of the Vasicek model is the mean reversion it exhibits. In particular, the short rate of interest is attracted towards a long run value  $\bar{r}$ . The strength of this mean reversion is controlled by the parameter  $\theta$ . Intuitively, the half life of the conditional expectation is  $1/\theta$ , which means that if the short rate is at level  $r_t$  at time  $t$ , then it is expected to cover half its distance from the long run value in  $1/\theta$  years. The main shortfall of the Vasicek model is that it permits the short rate to take negative values. This happens because the short rate is normally distributed, and therefore can take values over the real line.

As bond prices are exponentially affine with the short rate, and future short rates are normally distributed, it is easy to infer that future bond prices will follow the lognormal distribution. Therefore bond options will be priced with formulas similar to the Black-Scholes one for equity options. In particular, the price of a call option with strike price  $K$  that matures at time  $\tau$ , written on a zero coupon bond that pays one pound at time  $T > \tau$  will be equal to

---

<sup>3</sup> Here we follow the approach outlined in [Duffie and Kan \(1996\)](#) for general affine structures. Such systems of ODEs that are 'linear-quadratic' are known as Riccati equations.

$$C_t(\tau, K; T) = P_t(T)N(d_+) - KP_t(\tau)N(d_-)$$

where  $d_{\pm} = \frac{1}{\sigma^*} \log \frac{P_t(T)}{K \cdot P_t(\tau)} \pm \frac{\sigma^*}{2}$

and  $\sigma^* = \frac{\sigma}{\theta} \sqrt{\frac{1 - \exp\{-2\theta\tau\}}{2\theta}} [1 - \exp\{-\theta(T - \tau)\}]$

### LOGNORMAL MODELS

The main shortcoming of the Vasicek model is that it permits negative nominal interest rates. One straightforward way around this problem is to cast the problem in terms of the logarithm of the short rate. The first application of this idea can be found in [Dothan \(1978\)](#) model, which specifies

$$dr_t = \theta r_t dt + \sigma r_t dB_t, \text{ or } d \log r_t = \left( \theta - \frac{1}{2} \sigma^2 \right) dt + \sigma dB_t$$

Here the short rate follows the geometric Brownian motion, just like the underlying stock in the Black-Scholes paradigm. The short rate is log-normally distributed, and therefore takes only positive values. On the other hand, there is no mean reversion present, and the long run forecast for the short rate will either be explosive (if  $\theta > \sigma^2/2$ ) or zero (if  $\theta < \sigma^2/2$ ). For that reason the Dothan model is not popular for modeling purposes.

Another approach is casting the logarithm of the short rate to follow the Ornstein-Uhlenbeck process, giving rise to the exponential Vasicek model

$$d \log r_t = \theta(\log \bar{r} - \log r_t)dt + \sigma dB_t$$

In section 7.5 we will discuss the numerical implementation of a popular extension of this model, due to [Black and Karasinski \(1991\)](#).

An important feature of all lognormal models is the so-called “explosive” behavior of the bank account (see for example the discussion in [Brigo and Mercurio, 2001](#); [Sandmann and Sondermann, 1997](#)). Loosely speaking, if the yield is log-normally distributed, then the expected bank account is given by an expression of the form

$$EB_t = E \exp\{\exp\{Z\}\}, \text{ with } Z \sim N(\mu_Z, \sigma_Z^2)$$

It turns out that this expectation is infinite for all values of  $\mu_Z$  and  $\sigma_Z$ . That means that, according to lognormal models, even investing for a very short horizon (where the yield is approximately normal) offers infinite expected returns. Technically speaking, the right tail of the lognormal distribution does not decay fast enough, and this is the reason of the infinite expectation.

### THE CIR MODEL

The most popular member across the one factor model family is without doubt the one proposed in [Cox et al. \(1985, CIR\)](#). The short rate follows the “square

root" or Feller process.<sup>4</sup>

$$dr_t = \theta(\bar{r} - r_t) dt + \sigma\sqrt{r_t} dB_t$$

The CIR model is able to capture most of the desired properties of short rate models. The process is mean reverting, with the long run attractor equal to  $\bar{r}$ . The speed of mean reversion is controlled by the parameter  $\theta$ . As the short rate increases, its volatility also increases, at a degree which is dictated by  $\sigma$ . CIR show that the transition density of the process is a non-central chi-square. In particular, for

$$(2cr_T)|r_t \sim \chi^2 \left( 4\frac{\theta\bar{r}}{\sigma^2}, 2r_t c \exp\{-\theta(T-t)\} \right)$$

with  $c = \frac{2\theta}{\sigma^2(1 - \exp\{-\theta(T-t)\})}$

Having the transition density in closed form allows us to calibrate the parameters to a set of historical data. Unfortunately, the short rate is not directly observed, but practitioners use yields of bonds with short maturities as a proxy for the dynamics. More elaborate methods involve (Kalman) filtering and are discussed later.

One can readily compute the expected value and the variance of the short rate process, in particular

$$\begin{aligned} E[r_T|r_t] &= \bar{r} + \exp\{-\theta(T-t)\}(r_t - \bar{r}) \\ V[r_T|r_t] &= \frac{\sigma^2}{2\theta}(1 - \exp\{-\theta(T-t)\})[\bar{r} + (2r_t - \bar{r})\exp\{-\theta(T-t)\}] \end{aligned}$$

Also, as the forecasting horizon increases, the stationary (unconditional) distribution of the short rate is Gamma

$$r_t \sim \alpha \left( \frac{2\theta}{\sigma^2}, \frac{2\theta\bar{r}}{\sigma^2} \right)$$

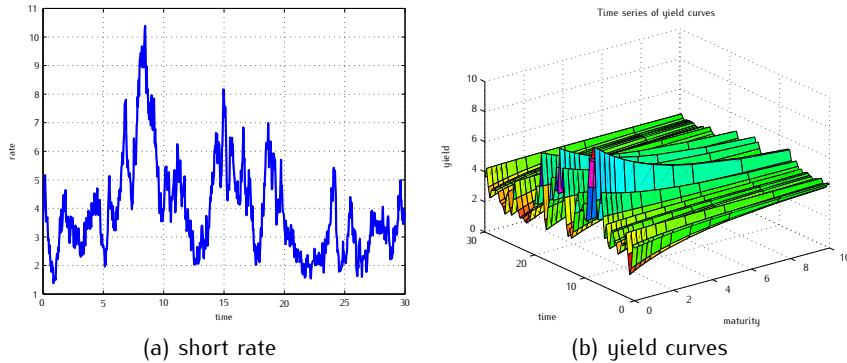
The instantaneous variance of the square root process is proportional to its level. For that reason, if the short rate reaches zero the stochastic component disappears, and the process will revert towards its positive long run mean. Therefore the CIR model rules out negative short rates,  $r_t \geq 0$  for all  $t$ . In particular, [Feller \(1951\)](#) shows that if the condition  $2\theta\bar{r} > \sigma^2$  is satisfied, then the mean reversion is strong enough for the process never to reach zero. In that case the inequality is strict,  $r_t > 0$  for all  $t$ .

CIR provide a bond pricing formula which also takes the exponentially affine form

---

<sup>4</sup> Discussed in detail in [Feller \(1951\)](#).

FIGURE 7.4: Simulation of CIR yield curves.



$$\begin{aligned}
 P_t(T) &= \exp(C(t; T) + D(t; T) \cdot r_0), \text{ with} \\
 C(t; T) &= -\frac{2\theta\bar{r}}{\sigma^2} \log \left[ \frac{2\gamma \exp\{(\theta + \gamma)(T - t)/2\}}{(\theta + \gamma)(\exp\{\gamma(T - t)\} - 1) + 2\gamma} \right] \\
 D(t; T) &= -\frac{2(\exp\{\gamma(T - t)\} - 1)}{(\theta + \gamma)(\exp\{\gamma(T - t)\} - 1) + 2\gamma} \\
 \gamma &= \sqrt{\theta^2 + 2\sigma^2}
 \end{aligned}$$

Option prices also take a (relatively) simple form, being dependent on the cumulative densities of non-central chi-square distributions

$$C_t(\tau, K; T) = P_t(T)\chi^2(d_1; v_1, v_2) - K P_t(\tau)\chi^2(d_2; v_1, v_3)$$

with

$$\begin{aligned}
 d_1 &= 2r^*[\varphi_1 + \varphi_2 + D(\tau; T)], \quad d_2 = 2r^*[\varphi_1 + \varphi_2] \\
 v_1 &= \frac{4\theta\bar{r}}{\sigma^2}, \quad v_2 = \frac{2\varphi_1^2 r_0 \exp\{\gamma(\tau - t)\}}{\varphi_1 + \varphi_2 + D(\tau; T)}, \quad v_3 = \frac{2\varphi_1^2 r_0 \exp\{\gamma(\tau - t)\}}{\varphi_1 + \varphi_2} \\
 \varphi_1 &= \frac{2\gamma}{\sigma^2(\exp\{\gamma(\tau - t)\} - 1)}, \quad \varphi_2 = \frac{\theta + \gamma}{\sigma^2} \\
 \gamma &= \sqrt{\theta^2 + 2\sigma^2}, \quad r^* = \frac{1}{D(\tau; T)} \log \left[ \frac{C(\tau; T)}{K} \right]
 \end{aligned}$$

## 7.5 MODELS WITH TIME VARYING PARAMETERS

The one factor models we described above have a finite number of parameters. Although some models can give flexible yield curve shapes, and conform to the stylized facts (for example CIR), they cannot match the observed yield curve

exactly. The problem is of course that a large (or infinite if we decide to interpolate) number of bonds have to be matched using a finite number of parameters and a given parametric form: the distance between model and observed prices can be minimized but not set to zero.

This means that for (practically) all maturities the bonds will be mispriced. This might appear to be a critical drawback, as one is not required to trade at the model price. It becomes a more serious flaw when one considers derivatives, where small discrepancies will be magnified, and in fact arbitrage opportunities will emerge.

Models with *time varying parameters* were set up to create that capture any initial yield curve, and therefore are at least arbitrage-free when it comes to pricing fixed income derivatives. Such models will assume that one or more parameters are deterministic functions of time, carefully chosen in a way that ensures that a term structure is perfectly replicated. Therefore a standard input for such models would be the current observed yield curve. State-of-the-art variants also be calibrated on implied volatility curves (from caplets, caps or swaption prices).

### THE HO-LEE MODEL

The first one-factor models with varying parameters proposed in the literature was [Ho and Lee \(1986\)](#). The underlying assumption was the short rate of interest is a simple random walk with drift

$$dr_t = \theta_t dt + \sigma dB_t$$

The drift  $t \mapsto \theta_t$  is a deterministic function of time. In particular, the bond prices will satisfy

$$\begin{aligned} P_t(T) &= E_t \exp \left\{ - \int_t^T r_s ds \right\} \\ &= E_t \exp \left\{ - \int_t^T \left[ r_t + \int_t^s \theta_u du + \sigma \int_t^s dB_u \right] ds \right\} \end{aligned}$$

Changing the order of integration we conclude that

$$\begin{aligned} P_t(T) &= E_t \exp \left\{ -(T-t)r_t - \int_t^T \int_t^s \theta_u du ds - \sigma \int_t^T \int_t^s dB_u ds \right\} \\ &= E_t \exp \left\{ -(T-t)r_t - \int_t^T \int_u^T \theta_u ds du - \sigma \int_t^T \int_u^T ds dB_u \right\} \\ &= E_t \exp \left\{ -(T-t)r_t - \int_t^T (T-u)\theta_u du - \sigma \int_t^T (T-u)dB_u \right\} \end{aligned}$$

The last integral is actually a normally distributed random variable, following Ito's isometry

$$\sigma \int_t^T (T-u) dB_u \sim N \left( 0, \sigma^2 \int_t^T (T-u)^2 du \right) = N \left( 0, \frac{\sigma^2}{3} (T-t)^3 \right)$$

Therefore the expectation can be computed in closed form, implying a yield to maturity

$$P_t(T) = \exp \left\{ -(T-t)r_t - \int_t^T (T-u)\theta_u du - \frac{\sigma^2}{6}(T-t)^3 \right\}$$

$$\Rightarrow y_t(T) = r_t + \int_t^T \frac{T-u}{T-t} \theta_u du + \frac{\sigma^2}{6}(T-t)^2$$

The above expression in fact is the functional form for the yield curve, if the time varying drift functional  $\theta_t$  was known. It turns out that it is more convenient to use the forward curve instead. In particular, applying the Leibnitz rule for differentiation, yields

$$f_t(T) = -\frac{\partial \log P_t(T)}{\partial T} = r_t + \int_t^T \theta_u du + \frac{\sigma^2}{2}(T-t)^2$$

If we differentiate the forward curve with respect to maturity we can achieve an expression for  $\theta_t$

$$\theta_T = \frac{\partial f_t(T)}{\partial T} + \sigma^2(T-t)$$

Knowing the drift functional can lead to prices for bonds and bond options that are easy to compute. Using the above relationship we can write

$$P_t(T) = \exp \left\{ -(T-t)r_t - \int_t^T f_t(s) ds - \frac{\sigma^2}{3}(T-t)^3 \right\}$$

As bond prices are lognormally distributed, options on these bond can be priced using a formula that is analogous to the Black-Scholes one

### THE HULL-WHITE MODEL

The breakthrough of the Ho-Lee model was that it provided a structure where the observed yield curve is perfectly matched, not allowing for arbitrage opportunities between model and observed prices. Having said that, it has two significant drawbacks, as there is no mean reversion present and the normality assumption allows negative nominal interest rates. In particular, not exhibiting mean reversion means that the distribution for the short rate widens with the time horizon, and the probability of negative rates increases. [Hull and White \(1990\)](#) take the Ho-Lee model one step further, and construct a model that exhibits mean reversion in the spirit of the Vasicek framework. For that reason the Hull-White model is also known as the *extended Vasicek* model.

The short rate is given by

$$dr_t = (\theta_t - \alpha r_t) dt + \sigma dB_t$$

Now the short rate will revert towards  $\theta_t/\alpha$ , with  $t \mapsto \theta_t$  a deterministic function of time. Although negative rates are permitted, in many cases the presence of mean reversion ensures that their probabilities are fairly small.

Using exactly the same arguments as the ones in the Ho-Lee case, we can solve for the functional  $\theta_T$  in terms of the forward curve  $f_t(T)$

### INTEREST RATE TREES

Models with time varying parameters give bond and bond option prices that are expressed as integrals of the forward curve. In practice, such models and their extensions are implemented through trees. In particular, the seminal papers of Hull and White (1994, 1996, henceforth HW) show how one can produce trinomial trees that will approximate a generic model of the form

$$\begin{aligned} d\xi_t &= (\theta_t - \alpha_t \xi_t) dt + \sigma_t dB_t \\ r_t &= \varphi_t(\xi_t) \end{aligned}$$

The state variable  $\xi_t$  follows a generalized Ornstein-Uhlenbeck process. The mean reversion level, the speed of mean reversion and the volatility are allowed to be deterministic functions of time. The short rate process is given as transformation of this state variable. Typical transformations are the identity  $\varphi_t(\xi) = \xi$  and the exponential  $\varphi_t(\xi) = \exp\{\xi\}$ .

### CALIBRATION OF INTEREST RATE TREES

The calibration of an interest rate tree using the HW methodology is carried out in two stages, first building an auxiliary tree and then adjusting it to match the observed yield curve.

#### The first stage

In the first stage a trinomial tree is built that reverts to zero, approximating the diffusion

$$d\xi_t = -\alpha_t \xi_t dt + \sigma_t dB_t$$

As an example we will assume that the mean reversion parameter and the volatility are constant, but extensions are straightforward, if one wishes to render them time-varying.

The tree that approximates the process  $\xi_t$  is constructed recursively. Let us assume that the tree has been constructed up to time  $t$ , and denote with its discretized values  $\bar{\xi}_i$ , for  $i = -\bar{m}, \dots, \bar{m}$ . We will show how to select the nodes and the transition probabilities that will grow this tree to time  $t + \Delta t$ . The first step is to select the grid spacing across the state space, for which HW suggest

$$\Delta \zeta = 3\sigma\sqrt{\Delta t}$$

LISTING 7.3: hw\_create.m: Create Hull-White trees for the short rate.

```
% hw_create.m
function HW = hw_create(T, B, f1, alpha, sigma)
DT = diff(T);
Dx = feval(sigma, T);
5 Dx = Dx(1:end-1).*sqrt(3*DT);
V = (Dx.^2)./3;
x = {[0]}; P = {}; Cx = {};
for ix = 1:length(DT)
    xe = cell2mat(x(end));
    10 Dxi= Dx(ix);
    i = length(x); i0 = length(xe);
    M = -DT(ix)*feval(alpha, T(ix))*xe;
    K = round((xe+M)./Dxi);
    m = max(K)+1;
    A = (xe +M -K*Dxi)./Dxi;
    P{ix} = [V(ix)/(2*Dxi^2)+A.*(A+1)./2, ...
              1-V(ix)/Dxi^2-A.^2, ...
              V(ix)/(2*Dxi^2)+A.*(A-1)./2];
    Cx{ix} = [K-1, K, K+1] +m+1;
    x{ix+1} = Dxi*(-m:m)';
20 end
Q = {[1]}; r = {};
for kx = 1:length(x)-1
    Ckx = Cx{kx};
    25 g = fzero(@(z) (Q{kx})'...
                 *exp(-feval(f1,x{kx}+z)*DT(kx))-B(kx), 0.01);
    r{kx} = feval(f1, g +x{kx});
    xx = [];
    for j = 1:max(max(Ckx))
        30 xx(j,1) = sum(sum( (Ckx==j).*P{kx}, 2)...
                     .*exp(-r{kx}*DT(kx)).*Q{kx} );
    end
    Q{kx+1} = xx;
end
35 HW.P = P; HW.Q = Q;
HW.C = Cx; HW.x = x;
HW.T = T; HW.r = r;
```

We will also assume that the discretization across time is done in equal time steps, and therefore the space step  $\Delta\zeta$  is also the same through time. The implementation in listing 7.3 relaxes all these assumptions and constructs a tree with time varying  $\alpha_t$  and  $\sigma_t$ , and also allows for variable time steps.

We then construct the grid at time  $t$  which extends across  $2m + 1$  elements (the choice for  $m$  will be discussed shortly)

$$\zeta = \{i\Delta\zeta : i = -m, \dots, 0, \dots, m\}$$

Typically, from the point  $\bar{\zeta}_i = i\Delta\zeta$  the process can move to the nodes  $\{\zeta_{i+1}, \zeta_i, \zeta_{i-1}\}$ . Then, one can solve a system that matches the instantaneous drift and volatility for the probabilities  $\{p_+, p_0, p_-\}$

$$\begin{aligned} p_+\Delta\zeta - p_-\Delta\zeta &= -\alpha i\Delta\zeta\Delta t \\ p_+(\Delta\zeta)^2 + p_-(\Delta\zeta)^2 &= \sigma^2\Delta t + \alpha^2 i^2(\Delta\zeta)^2(\Delta t)^2 \\ p_+ + p_0 + p_- &= 1 \end{aligned}$$

The solution sets

$$\begin{aligned} p_+ &= \frac{1}{6} + \frac{1}{2}\alpha i\Delta t[1 - \alpha i\Delta t] \\ p_0 &= \frac{2}{3} - \alpha^2 i^2(\Delta t)^2 \\ p_- &= \frac{1}{6} + \frac{1}{2}\alpha i\Delta t[1 + \alpha i\Delta t] \end{aligned}$$

for all  $j = -m, \dots, m$ . If all those probabilities are positive, then the tree will grow and will have  $2(m + 1) + 1$  elements in the next time period.

Encountering negative probabilities indicates that the mean reversion of the tree is too strong at these nodes for this particular transition structure. The geometry of the tree will then change, and the tree will stop growing. For example, having  $p_+ < 0$  indicates that the mean reversion is pushing the process towards zero quite strongly, and we therefore have to change the geometry of the tree and consider transitions towards the nodes  $\{\zeta_i, \zeta_{i-1}, \zeta_{i-2}\}$ . Of course, due to symmetry we will encounter negative  $p_-$  on the other end of the grid, suggesting transitions towards the nodes  $\{\zeta_{i+2}, \zeta_{i+1}, \zeta_i\}$ .

Solving for these alternative transition geometries yields

$$\begin{aligned} p_0 &= \frac{7}{6} + \frac{1}{2}\alpha i\Delta t[\alpha i\Delta t - 3] \\ p_- &= -\frac{1}{3} - \alpha i\Delta t[\alpha i\Delta t - 2] \\ p_{--} &= \frac{1}{6} + \frac{1}{2}\alpha i\Delta t[\alpha i\Delta t - 1] \end{aligned}$$

and

$$\begin{aligned}
 p_{++} &= \frac{1}{6} + \frac{1}{2}\alpha i\Delta t[\alpha i\Delta t + 1] \\
 p_+ &= -\frac{1}{3} - \alpha i\Delta t[\alpha i\Delta t + 2] \\
 p_0 &= \frac{7}{6} + \frac{1}{2}\alpha i\Delta t[\alpha i\Delta t + 3]
 \end{aligned}$$

As we noted, in such cases the tree will not grow and the next set of nodes will also have  $m+1$  elements. The top half of listing 7.3 implements this method for a more general setting, when the time steps, the mean reversion and the volatility are all time varying. It makes sense to select the value of  $m$  as the first one for which the volatility geometry changes from the standard one to the ones that force mean reversion:

$$m = \max \left\{ \text{round} \left[ \frac{\bar{\zeta}_i}{\Delta \zeta} (1 - \alpha \Delta t) \right] \right\} + 1$$

### The second stage

In the second stage the nodes of the tree that replicates the process  $\zeta_t$  are shifted up or down in order to match the dynamics that price bonds exactly. This creates the trinomial tree that will be approximating  $\xi_t$ . To this end, when calibrating a HW tree we make use of the so called *Arrow-Debreu (AD) state prices*, which we define now.

An *Arrow-Debreu security* is a generic contingent claim that will pay one monetary unit if a certain event is realized at a particular point in time. Otherwise the AD security pays nothing. The AD state price is the price of this security. It is easy to see that AD securities can be used as building blocks to construct more complex payoffs.

In models with a continuous state space, the Arrow-Debreu security is a European style contract that pays off the Dirac delta function on its maturity. In the context of HW trees the state space is discretized, as at time  $T$  the short rate can take one out of  $K_T$  possible values. Then, an AD security will pay one pound if the short rate is at its  $k$ -th value at time  $T$ , and zero otherwise. We denote with  $Q_t(k, T)$  the price of this AD security at time zero, the AD state price.

One can readily observe that if we purchase all Arrow-Debreu securities that mature at time  $T$ , then we are sure to receive one pound on that date. Effectively we have constructed the payoff of the risk free bond. Arbitrage arguments will then indicate that the sum of all AD state prices across states will equal the price of a zero coupon bond.

$$\sum_{k=1}^{K_T} Q_t(k, T) = P_t(T)$$

We can also construct an inductive relationship that links AD securities with successive maturities  $T$  and  $T+1$ . In particular, like any other security, under the

risk neutral probability measure, discounted AD securities will form martingales. Suppose that the tree can take one of  $K_t$  different values at time  $t$ , and denote with  $c_t$  the actual level of the tree at that time, with  $c_t \in \{1, 2, \dots, K_t\}$ . Then we can write

$$Q_t(k, T) = \mathbb{E}_t^{\mathbb{Q}}[Q_T(k, T)] = \mathbb{E}_t^{\mathbb{Q}}\left[\frac{I(c_T = k)}{B_T}\right] = \mathbb{E}_t^{\mathbb{Q}}\left[\frac{1}{B_T} \middle| c_T = k\right] P_t^{\mathbb{Q}}[c_T = k]$$

Conditioning on the state at time  $T - 1$ , and using the definition of the bank account process, allows us to expand the conditional expectation as

$$\begin{aligned} & \mathbb{E}_t^{\mathbb{Q}}\left[\frac{1}{B_T} \middle| c_T = k\right] \\ &= \sum_{j=1}^{K_{T-1}} \mathbb{E}_t^{\mathbb{Q}}\left[\frac{1}{B_{T-1} e^{r_{T-1} \Delta t}} \middle| c_T = k, c_{T-1} = j\right] P_t^{\mathbb{Q}}[c_{T-1} = j | c_T = k] \\ &= \sum_{j=1}^{K_{T-1}} e^{-r_j \Delta t} \mathbb{E}_t^{\mathbb{Q}}\left[\frac{1}{B_{T-1}} \middle| c_T = k, c_{T-1} = j\right] P_t^{\mathbb{Q}}[c_{T-1} = j | c_T = k] \\ &= \sum_{j=1}^{K_{T-1}} e^{-r_j \Delta t} \mathbb{E}_t^{\mathbb{Q}}\left[\frac{1}{B_{T-1}} \middle| c_{T-1} = j\right] P_t^{\mathbb{Q}}[c_{T-1} = j | c_T = k] \end{aligned}$$

Bayes' rule will provide us with

$$P_t^{\mathbb{Q}}[c_{T-1} = j | c_T = k] = P_t^{\mathbb{Q}}[c_T = k | c_{T-1} = j] \frac{P_t^{\mathbb{Q}}[c_{T-1} = j]}{P_t^{\mathbb{Q}}[c_T = k]}$$

The quantity  $p_t(j, k) = P_t^{\mathbb{Q}}[c_T = k | c_{T-1} = j]$  is just the (risk neutral) transitional probability of moving from state  $j$  to state  $k$  at time  $t$ . The AD state price is then simplified to

$$\begin{aligned} Q_t(k, T) &= \sum_{j=1}^{K_{T-1}} e^{-r_j \Delta t} p_t(j, k) \mathbb{E}_t^{\mathbb{Q}}\left[\frac{1}{B_{T-1}} \middle| c_{T-1} = j\right] P_t^{\mathbb{Q}}[c_{T-1} = j] \\ &= \sum_{j=1}^{K_{T-1}} e^{-r_j \Delta t} p_t(j, k) Q_t(j, T-1) \end{aligned}$$

In the above expression  $r_j = \varphi(\xi_j)$ , and  $\xi_j = \zeta_j + \omega$ .

Essentially, in order to fit the observed yield curve one has to solve numerically for the value of  $\omega$  at each maturity horizon  $T$ . If one also renders the volatility and/or the speed of mean reversion time varying, then the parameters have to be calibrated on a richer set of data that will identify these parameter values. Typically, deterministic volatility functions are chosen as to match implied volatilities that are derived from caps or swaptions.

LISTING 7.4: hw\_path.m: Compute the price path of a payoff based on a Hull-White tree for the short rate.

```
% hw_path.m
function [PO, PP] = hw_path(HW, PF)
r = HW.r; Cx = HW.C;
T = HW.T; P = HW.P;
5 NT = length(PF);
DT = diff(T);
PO = PF{end};
PP{NT} = PO;
for kx = (NT-1):-1:2
    PPk = PO(Cx{kx});
    PO = exp(-r{kx}*DT(kx)).*sum(PPk.*P{kx},2) + PF{kx};
    PP{kx} = PO;
end
PO = exp(-r{1}*DT(1))*P{1}*PO + PF{1};
15 PP{1} = PO;
```

Overall, the construction of an interest rate tree resembles the local volatility models for equity derivatives. In both frameworks we attempt to exactly replicate a market implied curve or surface. One has to keep in mind the dangers of overfitting, which would introduce spurious qualities into the model. In many cases market quotes of illiquid instruments can severely distort the model behavior.

### Pricing and price paths

After the tree has been fitted to the yield curve, we can proceed to pricing various interest rate sensitive instruments, such as bond options, interest rate caps, floor, swaps or swaptions. Essentially we can find the fair value of a given stream of contingent cashflows, in a way that is consistent with the prices of risk free bonds.

If there are no early exercise features, prices of contingent claims can be computed by summing up the corresponding AD security prices. In many cases we are not only interested in the fair value of the contract, but also on its price path. For example, in order to find the fair value of a put option with three year maturity, which is written on a ten-year bond, we need to consider the price paths of the ten-year bond, in order to ascertain the option payoffs. Price paths can be easily computed by iterating backwards through the tree, starting from the terminal date. Listing 7.4 shows how this can be easily implemented. To allow for early exercise one just has to check if early exercise is optimal at each tree node (implemented in listing 7.5).

LISTING 7.5: `hw_path_amer.m`: The price path of a payoff based on the Hull-White tree when American features are present.

```
% hw_path_amer.m
function [PO, PP, XR] = hw_path_amer(HW, PF, PX)
r = HW.r; Cx = HW.C;
T = HW.T; P = HW.P;
NT = length(PF);
5 DT = diff(T);
PO = PF{NT};
PO = max(PO, PX{NT});
PP{NT} = PO; XR{NT} = (PO==PX{NT} & PO>0);
10 for kx = (NT-1):-1:2
    PPk = PO(Cx{kx});
    PO = exp(-r{kx}*DT(kx)).*sum(PPk.*P{kx},2) + PF{kx};
    PO = max(PO, PX{kx});
    PP{kx} = PO; XR{kx} = (PO==PX{kx} & PO>0);
15 end
PO = exp(-r{1}*DT(1))*P{1}*PO;
PO = max(PO, PX{1});
PP{1} = PO; XR{1} = (PO==PX{1} & PO>0);
```

## THE BLACK-KARASINSKI MODEL

The most popular special case of this very general specification is the [Black and Karasinski \(1991\)](#) model, which is in spirit similar to the exponential Vasicek model with time varying parameters. Here

$$\begin{aligned} d\xi_t &= \alpha(\theta_t - \xi_t) dt + \sigma dB_t \\ r_t &= \exp\{\xi_t\} \end{aligned}$$

This specification exhibits mean reversion, and through the exponential transformation ensures that the short rate remains positive. As with all lognormal models, the Black-Karasinski model implies an explosive expectation for the bank account, but since in practice the implementation is done over a finite tree, this drawback is not severe.

LISTING 7.6: hw\_path\_impl.m: Implementation of the Black-Karasinski model using a Hull-White interest rate tree.

```
% hw_impl.m
% observed yield curve
T0 = [1/252, ..., 30]';
Y0 = [ 5.29, ..., 5.94]'; Y0 = Y0/100;
5 T = [0:0.0625:3 3.125:0.125:10 11:0.5:30]';
Y = interp1(T0, Y0, T, 'cubic', 'extrap');
B = exp(-Y.*T);
H = hw_create(T, B(2:end), @(x) exp(x), ...
    @(t) 0.25*ones(size(t)), ...
    @(t) 0.2*ones(size(t)));
10

iB = find(T==10);
TC = [0.5:0.5:10]';
PF = {};
15 for ix = 1:iB
    if ismember(T(ix), TC)
        PF{ix} = 2.75*ones(size(H.Q{ix}));
    else
        PF{ix} = zeros(size(H.Q{ix}));
    end
20 end
PF{end} = PF{end} + 100;
[PO, PP] = hw_path(H, PF); figure(1); hw_graph(PP, H);

25 iE = find(T==2);
PPE = {PP{1:iE}}; figure(2); hw_graph(PPE, H);

PF = {};
for ix = 1:iE
    PF{ix} = zeros(size(H.Q{ix}));
30 end
PF{end} = max(80-PP{iE}, 0);
[PEO, PE] = hw_path(H, PF); figure(3); hw_graph(PE, H);

35 PF = {}; PX = {};
for ix = 1:iE
    PF{ix} = zeros(size(H.Q{ix}));
    PX{ix} = max(80-PP{ix}, 0);
end
40 PF{end} = max(80-PP{iE}, 0);
[PA0, PA, XR] = hw_path_amer(H, PF, PX);
figure(4); hw_graph_amer(PA, H, XR);
figure(5); hw_graph_amer(PPE, H, XR);
```

FIGURE 7.5: Calibration of a tree for the short rate that implements the Black-Karasinski model. The Hull-White framework is implemented.

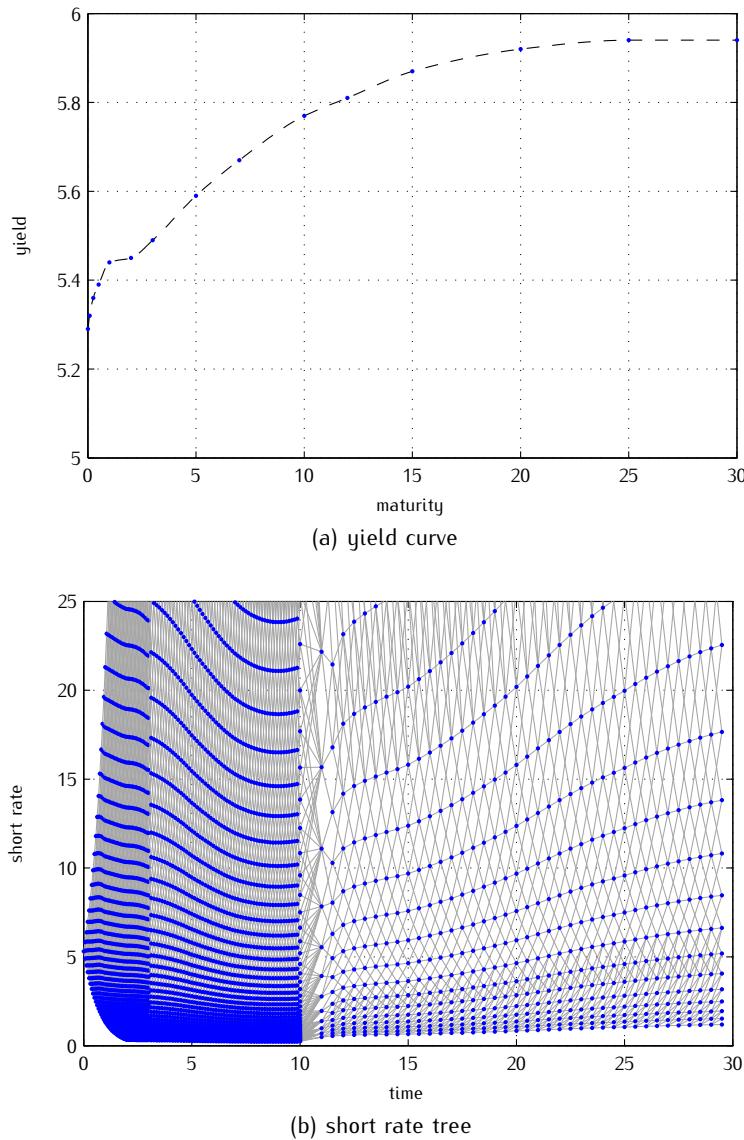
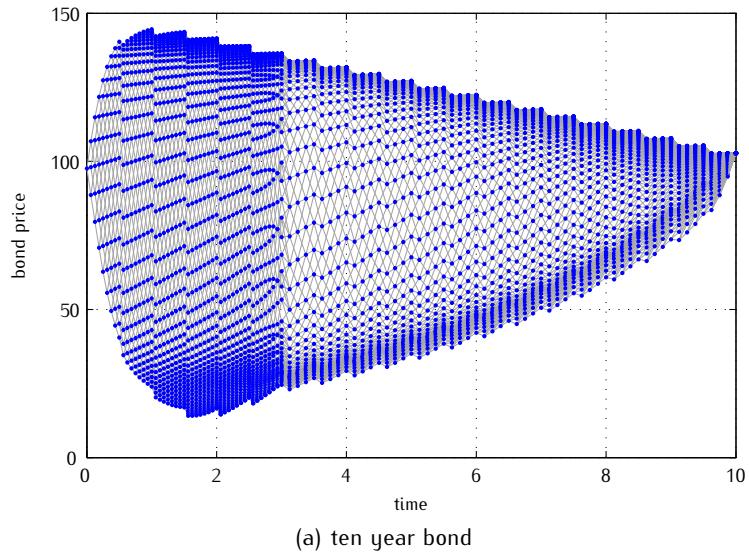
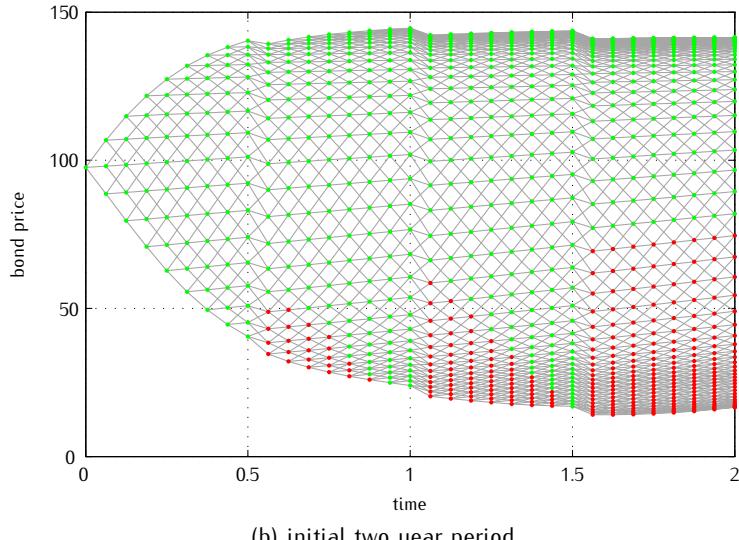


FIGURE 7.6: Price path for a ten year 5.50% coupon bearing bond. The price paths are consistent with the yield curve of figure (7.5), modeled using the Black-Karasinski process.

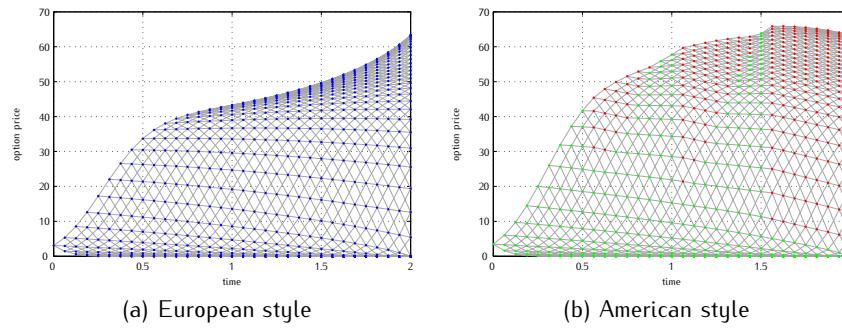


(a) ten year bond



(b) initial two year period

FIGURE 7.7: Price path for a two year put option, written on the ten year coupon bearing bond of figure (7.6). The strike price is set at \$80.



Listing 7.6 shows how the HW tree building methodology is applied in the Black-Karasinski case. The yield curve of figure (7.5.a) is assumed, and a HW is constructed. We use  $\Delta t = 1/16$  over the first three years,  $\Delta t = 1/8$  from the third to the tenth, and  $\Delta t = 1/2$  for remaining twenty years. A view of the resulting interest rate tree is given in figure (7.5.b), where this uneven time discretization is apparent.

Of course, to value such a simple bond we do not need to construct the complete price path, and in fact we do not need to construct a HW tree at all. The fair price can be determined by using the yield curve alone, just by discounting all cashflows. The price path is needed though if we want to value an option on this ten year bond.

As an example we consider a two year put, with strike price  $K = \$80$ . To price this option we need the distribution of the bond price after two years. Figure (7.6.b) gives the possible bond prices and the corresponding price paths for the two year period. Essentially, the put option gives us the right to sell the bond at the strike price if the interest rates after two years are too high. The price paths for a European and an American version are illustrated in figure (7.7); the corresponding prices are  $P_E = \$3.08$  and  $P_A = \$3.46$ , indicating an early exercise premium of \$0.38, which is actually more than 10% of the option price. The red points (in figures 7.6.b and 7.7.b) indicate the scenarios where early exercise is optimal. We can observe how the coupon payments affect the exercise boundary, as we would prefer to exercise immediately after the coupon payment is realized.

### CALIBRATION ISSUES

All models with time varying parameters can be cast in a binomial/trinomial form that approximates the short rate movements. Nevertheless, although the tree will

always adjust to match the current yield curve, following the HW procedure, there are still parameters to be identified.

As an example, in the Black-Karasinski examples above we assumed  $\alpha = 0.25$  and  $\sigma = 0.20$ , but without any justification of these values. There are two options in setting values for such free parameters, based on historical yield curves or based on derivative prices.

Given a set of historical yield curves, one can produce estimates for the speed of mean reversion and the volatility parameters. This can be done by proxying the historical (unobserved) spot rate with a yield of a relatively short maturity, and then applying maximum likelihood estimation methods. Of course one has to keep in mind that such an exercise is carried out under the physical probability measure, which might or might not have an impact, depending on the exact parameterization of the price of risk. Even better, one can maintain the spot rate unobservable, and use Kalman filtering techniques that draw information from the whole yield curve. Such an approach allows one to jointly recover the parameters under both physical and risk adjusted measures.

If derivative prices on interest rate sensitive instruments are available and liquid, then their prices can be used to also provide estimates for the fixed parameters. This is typically done by minimizing the squared price differences, or the differences between model and actual implied volatilities. This seems to be the method of choice amongst practitioners, but care must be taken to avoid pitfalls.

In spirit, the second approach is very similar to the standard calibration of stochastic volatility models to derivative prices, and is also subject to the implementation difficulties that are associated with such models. In particular, one main obstacle is model identification, where different set of parameters produce the same optimal objective function.

Typically, a model will be calibrated on a set of interest rate caps and swaptions, which are instruments that are sensitive to the volatility of the interest rate. In virtually all short rate models the terminal volatility is the outcome of the two quantities we want to retrieve, namely the speed of mean reversion and the volatility of the innovations. Decreasing the speed has more or less the same effect as increasing the volatility, and calibrating these quantities is not a well identified problem. As in the stochastic volatility example, there is a locus of parameter pairs that produce the same optimal fit, and we have no information to distinguish between them. Surprisingly many practitioners choose to ignore this issue, selecting the first set of points that their numerical optimizer returns. But this can be the source of severe mispricing of other, more exotic, contracts that will be valued using the calibrated parameters.

In an ideal world, one would get around this issue by also calibrating to derivatives that are sensitive to future transition densities, such as forward starting swaptions, but unfortunately such contracts are generally not available, and very illiquid when they exist. Another option is to use the same regularization techniques that we outlined in the stochastic volatility case, where the prior pa-

rameters can be selected as the maximum likelihood estimates of the historical ones.

Overall, one has to remember that calibration is more of an art than science (although well disguised as science).

## 7.6 MULTI-FACTOR MODELS

Short rate models are also known as *one factor* models, as there is only one source of uncertainty in the economy, which is represented by the short rate. Although short rate models are very useful for some applications, they are not sufficient for others. In particular, the presence of a single factor implied that yields of all maturities are perfectly correlated (albeit with different volatilities). This is easily illustrated in the case of the exponentially affine models, like Vasicek or CIR, where the yield is a linear function of the current short rate. Yields for two different maturities  $\tau_1$  and  $\tau_2$ , and their dynamics, are given by

$$y_t(t + \tau_i) = A(t; t + \tau_i) + B(t; t + \tau_i) r_t \Rightarrow dy_t(t + \tau_i) = B(t; t + \tau_i) dr_t$$

TABLE 7.1: Correlations of yields for different maturities. Bonds with longer maturities exhibit relatively higher correlation. Listing 7.7 gives the relevant Matlab code.

1m	3m	6m	1y	2y	3y	5y	7y	10y	20y	
1.00	0.56	0.40	0.28	0.21	0.19	0.18	0.16	0.15	0.10	1m
	1.00	0.76	0.55	0.42	0.40	0.36	0.32	0.30	0.24	3m
		1.00	0.85	0.68	0.64	0.59	0.55	0.52	0.44	6m
			1.00	0.87	0.83	0.78	0.74	0.71	0.64	1y
				1.00	0.97	0.92	0.88	0.85	0.77	2y
					1.00	0.96	0.93	0.89	0.83	3y
						1.00	0.98	0.96	0.90	5y
							1.00	0.98	0.94	7y
								1.00	0.96	10y
									1.00	20y

Therefore, the correlation of the two yields is

$$\frac{E_t\{dy_t(t + \tau_1)dy_t(t + \tau_2)\}}{\sqrt{E_t(dy_t(t + \tau_1))^2 E_t(dy_t(t + \tau_2))^2}} = \frac{B(t; t + \tau_1)B(t; t + \tau_2)}{\sqrt{B^2(t; t + \tau_1) \cdot B^2(t; t + \tau_2)}} = 1$$

In practice, this correlation might be high, but is not perfect as the one factor family suggests. For example, table 7.1 presents the historical correlation of various bonds with different maturities. Although the correlation is positive

across the board, its magnitude varies substantially at different horizons. In particular, the long end of the yield curve is much stronger correlated than the short end: the ten and twenty year bonds move pretty much in unison, with correlation over 95%, while the one and three months exhibit about half of this dependence. Also, each maturity exhibits correlations that decay as we consider bonds with increasing maturity differences. For example, the two year bond is stronger correlated with the three year rather than the seven year instruments.

One way of increasing the number of free parameters is by considering multi-factor models. For example, we can consider the interest rate to be the sum of two simple Vasicek processes, by setting

$$r_t = x_t^{(1)} + x_t^{(2)}, \text{ where}$$

$$dx_t^{(j)} = \theta^{(j)}(\bar{x}^{(j)} - x_t^{(j)})dt + \sigma^{(j)}dB_t^{(j)}$$

The two processes  $x_t^{(1)}$  and  $x_t^{(2)}$  are called *factors*, and are in principle unobserved. In the general specification we can also assume the factors to exhibit some correlation  $\rho$ . If we maintain that our model is affine, then we can postulate that the yield is again a linear combination

$$y_t(t + \tau_i) = A(t; t + \tau_i) + B_1(t; t + \tau_i)x_t^{(1)} + B_2(t; t + \tau_i)x_t^{(2)}$$

$$\Rightarrow dy_t(t + \tau_i) = B_1(t; t + \tau_i)dx_t^{(1)} + B_2(t; t + \tau_i)dx_t^{(2)}$$

After some extremely boring calculations we can derive the covariance of the changes of bonds with yields  $\tau_1$  and  $\tau_2$ , as

$$\rho(\tau_1, \tau_2) = \pm \sqrt{1 - \frac{(1-\rho^2)(B_{11}B_{22}-B_{21}B_{12})^2\sigma_1^2\sigma_2^2}{(B_{11}^2\sigma_1^2+B_{21}^2\sigma_2^2+2\rho B_{11}B_{21}\sigma_1\sigma_2)(B_{12}^2\sigma_1^2+B_{22}^2\sigma_2^2+2\rho B_{12}B_{22}\sigma_1\sigma_2)}}$$

We use the shorthand notation  $B_{ij} = B_i(t; t + \tau_j)$ . The sign of the correlation is positive if

$$\rho > -\frac{B_{11}B_{12}\sigma_1^2 + B_{21}B_{22}\sigma_2^2}{(B_{11}B_{22} + B_{21}B_{12})\sigma_1\sigma_2}$$

Therefore from the expression above we can assess the implied correlation of special cases. In particular, setting  $\rho = 1$  will render  $\rho(\tau_1, \tau_2) = 1$  as the two factors are driven by the same Brownian motion. Specifying uncorrelated factors will always produce positive correlation across the maturities. Large negative correlations produce a peculiar effect: as we set  $\rho = -1$  we observe that the yield correlation will be either perfectly positive or perfectly negative, depending on the maturity pair.<sup>5</sup>

---

<sup>5</sup> We mention this peculiarity because it is fairly common for a correlation of  $\rho = -1$  to be calibrated from cap or swaption prices. It is very unlikely that such a value reflects the interest rate dynamics, and it is a feature that is more likely to point towards more complex dynamics for the rate and its volatility that go beyond the affine setting.

Postulating the short rate to be the sum of factors is not the only way to construct multi-factor models. For example, in an early article [Brennan and Schwartz \(1982\)](#) consider a model where the long run interest rate is stochastic, serving as the second factor. Intuitively, there is a slowly mean reverting process which is largely a proxy for the business cycle and determines the long run attractor of the short rate. More recently, [Longstaff and Schwartz \(1992\)](#) propose a model where the second factor is the volatility of the short rate.

Other multi-factor specifications have been also considered in the literature. For example [Brennan and Schwartz \(1982\)](#) consider a model where the long run interest rate is also stochastic, serving as the second factor. Intuitively, long swings in the short rate are determined by the latter process, which exhibits weak mean reversion. This process will determine the behavior of the long end of the yield curve. The short end of the yield curve will be determined by the process that reverts faster towards the long run short rate process. In another paper, [Longstaff and Schwartz \(1992\)](#) propose a model where the second factor is the volatility of the short rate. The mean reverting nature of the short rate implies that the effect of stochastic volatility will have a higher impact on the short end of the curve.

In order to achieve a perfect match to a given yield curve [Brigo and Mercurio \(2001\)](#) describe a method that enhances the multi-factor specification, by adding a deterministic function of time  $\varphi(t)$

$$r_t = x_t^{(1)} + x_t^{(2)} + \varphi(t)$$

[Brigo and Mercurio \(2001\)](#) show how one can retrieve the deterministic function  $\varphi(t)$  for a variety of processes, including multi-factor Vasicek and CIR.

## FACTORS AND PRINCIPAL COMPONENT ANALYSIS

The historical relative moves of yields for various maturities also provide motivation for using multi-factor specifications. In particular, although yields are strongly correlated, they don't always move in the same direction. In fact, there are periods when the yield curve remains relatively flat, and episodes of steeply rising yield curves. A single-factor model is not adequate to reproduce such patterns, as it will always generate yields that move together across all maturities.

*Principal Component Analysis* (PCA) techniques can be employed to explore the variability and correlations of various yields. PCA has as input a set of  $N$  correlated series, and decomposes them into  $N$  uncorrelated components, which are called the *factors*. In order to do so, the covariance structure is computed and its eigen-structure is produced. The eigenvectors with the largest eigenvalues point towards the most important factors, and can be utilized to investigate which proportion of the variability of the original series is explained by individual factors. Typically, one looks for a set of factors that will explain 90-95% of the total variability.

Also, yields are strongly autocorrelated through time. It makes then perfect sense to work with the time-differenced series: in essence the factors will then

explain *changes* in the yield curve behavior through time, rather than the yield curve *level*. We will denote with  $y(t; \tau)$  the yield of bond with maturity  $\tau$ , recorded at time  $t = 1, 2, \dots$

Therefore, each yield change  $\Delta y_j = y(t; \tau_j) - y(t-1; \tau_j)$  is written as the weighted sum of  $n$  factors

$$\Delta y_j = c_j + \ell_{j,1}f_1 + \cdots + \ell_{j,i}f_i + \cdots + \ell_{j,n}f_n$$

The coefficients  $\ell_{j,i}$  are called *factor loadings*, and essentially determine the sensitivity of yield  $y_j$  to factor  $f_i$ , and  $c_j$  is a constant  $c_j = E\Delta y_j$ . If we assume that the factors are uncorrelated, and they are normalized with zero mean and unit variance, then we can write the covariance of different yields as

$$\text{Cov}(\Delta y_j, \Delta y_k) = \sum_{m=1}^n \ell_{j,m} \ell_{k,m}$$

Therefore, if we denote with  $L$  the matrix that collects all factor loadings, then the covariance matrix  $\Sigma$  of the yield changes will be equal to

$$\Sigma = L L'$$

Given that the covariance matrix is not singular, an eigen-decomposition will produce a matrix  $V$  with the linearly independent eigenvectors, together with a diagonal matrix of eigenvalues  $M$ , such that

$$\Sigma = V M V^{-1}$$

But as  $\Sigma$  is symmetric, the eigenvectors form an orthonormal matrix  $V^{-1} = V'$ , which implies essentially that the factor loadings matrix can be expressed as

$$L = V \sqrt{M}$$

Using this representation we can write the yield changes in terms of the elements of the eigenvector matrix  $V$  and the eigenvalues in  $M$

$$\Delta y_j = c_j + v_{j,1}\sqrt{m_1}f_1 + \cdots + v_{j,i}\sqrt{m_i}f_i + \cdots + v_{j,n}\sqrt{m_n}f_n$$

It is therefore intuitive that factors that are associated with higher eigenvalues will contribute more to the total variability of the series. In particular, if we consider the overall variance of all yields, then we can write as the sum of all eigenvalues

$$\sum_{j=1}^n \text{Var}(\Delta y_j) = \sum_{j=1}^n \sum_{i=1}^n v_{j,i}^2 m_i = \sum_{j=1}^n m_i$$

where the last equality follows from the fact that the eigenvectors are normalized to unit length. In factor analysis we select to use only the largest  $\bar{n}$  eigenvalues,

LISTING 7.7: `princcomp.m`: Correlation structure and principal component analysis of yield curve movements.

```
% princcomp.m
mtry = [1/12 3/12 6/12 1 2 3 5 7 10 20]; % maturities
data = xlsread('US_Rates_2001.xls'); % get the data
dates = x2mdate(data(:,1));
5 rate = data(:,2:end);
drate = diff(rate);
% data surface
figure(1);
surf(mtry, dates, rate);
datetick('y',10);
10 % correlation structure of interest rate changes
CV = corrcoef(drate);
disp('Correlation Structure');
for ix=1:length(mtry)
    fprintf('%6.2f ', 100*CV(:,ix));
    fprintf('\n');
end
% principal component analysis
[CF,SC,LT] = princomp(drate);
20 fprintf('\nEigenvalue decomposition\n');
fprintf('factor_eigv1_eigv1_cumul\n');
fprintf('%4d %6.2f %6.2f\n', [(1:length(mtry)), ...
    100*LT/sum(LT) 100*cumsum(LT)/sum(LT)]);
figure(2);
CF = CF(:,1:3);
SC = SC(:,1:3);
plot(CF, '-o', 'LineWidth', 1);
grid on;
xlabel('maturity');
30 ylabel('factor_loading');
legend('the_first_factor', 'the_second_factor', ...
    'the_third_factor');
```

and implicitly approximate the contributions of the rest as being independent across all maturities

$$\Delta y_j = c_j + v_{j,1}\sqrt{m_1}f_1 + \cdots + v_{j,i}\sqrt{m_i}f_i + \cdots + v_{j,\bar{n}}\sqrt{m_{\bar{n}}}f_{\bar{n}} + \eta_j$$

The number of factors that we retain should be used as to make the variance of the remainder component  $\eta_j$  small. As a rule of thumb,  $\bar{n}$  should ensure that at least 95% of the total variance is explained by the corresponding factors. If one finds that such a value of  $\bar{n}$  is large compared to the total number of variables, it is evidence that factor analysis might not appropriate for this case.

factor	1	2	3	4	5	6	7	8	9	10
eigenvalue (% of sum)	81.7	8.8	4.6	2.2	1.1	0.6	0.3	0.3	0.2	0.2
cumulative (% of sum)	81.7	90.6	95.2	97.3	98.4	98.9	99.3	99.6	99.8	100.0

TABLE 7.2: Relative magnitude of the eigenvalues for the decomposition of the correlation matrix. The first three factors are responsible for over 95% of the yield variability.

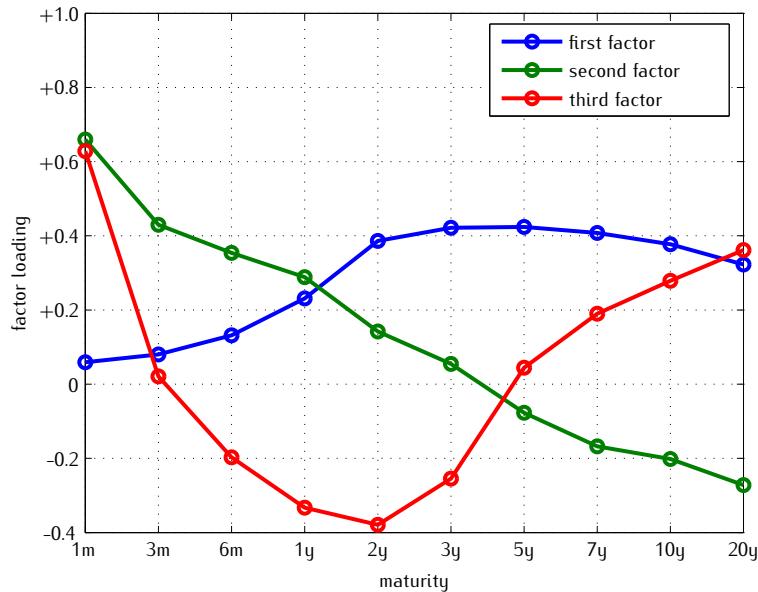


FIGURE 7.8: Yield curve factor loadings. A principal component analysis is applied to changes of yields over different maturities. The three factors that correspond to the level, the slope and the convexity are clearly identified.

The recipe for principal component analysis is illustrated in listing 7.7. The relative contribution of the  $j$ -th factor, together with the cumulative contribution of the first  $j$  factors are given in table 7.2. For example, the third factor explains 4.6% of the variability of interest rate changes, while the first three factors explain 95.2%. We can therefore adopt a three-factor model as an approximation that explains sufficiently well the yield curve dynamics.

One of the benefits of factor analysis, especially in interest rate modeling, is the intuition that it can offer. Figure 7.8 plots the factor loadings for the first three factors, against maturities. One can think of these curves as the impact on the yield curve of  $j$ -th factor shock, with magnitude  $\sqrt{m_j}$ . For example, a shock of the first factor will shift the yields of all maturities in the same direction. This will essentially move the whole yield curve upwards or downwards, and for that reason we coin the first factor as the *level factor*. This of course

is a rough statement, as it is obvious that such shifts will not be parallel, with longer maturities being more responsive. Nevertheless, we can apply the same reasoning for the second factor: now a positive factor shock will shift yields of short maturities<sup>6</sup> upwards, and at the same time will shift longer yields downwards. Such a shock will “rotate” or change the “slope” of the yield curve, and for that reason we call it the *slope factor*. One can now imagine why the third factor is called the *convexity factor*, as it affects the “hump” of the yield curve.

## KALMAN FILTERING

Principal component analysis of yields is a quick-and-dirty method of isolating the factors present in yield curve moves, and quantifying their impact. But in the end of the day it is a statistical technique, with no robust structure behind the dynamics of the factors. When we take the covariance matrix of yield changes, we implicitly make an assumption on their dynamics, namely that these changes are stationary (and therefore yields follow unit root processes).

Kalman filtering techniques can be applied for a substantial class of models, and in particular the relatively large affine family. As an example we will investigate an OU factor setup, which we will calibrate on a set of historical yield curves. As we will now have a complete model to describe the bond yields and their dynamics, the parameters and the corresponding risk premia can be jointly recovered.

To be more concrete, assume that, under the physical measure, the factors are specified as

$$dx_t^{(j)} = -\theta^{(j)}x_t^{(j)}dt + \sigma^{(j)}dB_t^{(j)}$$

That is to say, the factors are behaving as OU processes with mean reversion level at zero. The short rate will be given as the sum of these processes plus a constant term  $r_t = c + \sum_j x_t^{(j)}$ . We can assume a constant price of risk that will eventually after some algebra contribute to the constant term. We can then rewrite the risk adjusted short rate as  $r_t = c + \lambda + \sum_j x_t^{(j)}$ ; the risk adjusted dynamics for the factors are given by the stochastic differential equations

$$dx_t^{(j)} = \theta^{(j)}(\lambda^{(j)} - x_t^{(j)})dt + \sigma^{(j)}dB_t^{\mathcal{Q},(j)}$$

with  $B_t^{\mathcal{Q},(j)}$  being a  $\mathcal{Q}$ -Brownian motion.

With the further assumption that these Brownian motions are independent across the factors, we can write the bond price as the product of Vasicek prices, having the form

---

<sup>6</sup> Less than three years in this case.

$$\begin{aligned} P_t(\tau) &= E^{\Omega} \exp \left\{ \int_t^{t+\tau} r_s ds \right\} \\ &= \exp \left\{ -(c + \lambda)\tau + \sum_j C^{(j)}(t; \tau) + \sum_j D^{(j)}(t; \tau)x_t^{(j)} \right\} \end{aligned}$$

Notice that this is slightly different to the form presented during the discussion of the Vasicek model, as here  $\tau$  denotes the *time to maturity*, while there  $T$  denoted the *maturity date*. We change the notation slightly to economize on space here.

Therefore, the yields for different maturities are given by

$$y_t(\tau) = c\tau + \sum_j \tilde{C}^{(j)}(t; \tau) + \sum_j \tilde{D}^{(j)}(t; \tau)x_t^{(j)}$$

with the functions  $\tilde{C}$  and  $\tilde{D}$  are given below (the superscripts are removed to further ease the notation)

$$\begin{aligned} \tilde{C}(t; \tau) &= \frac{[1 - \tilde{D}(t; \tau)][\theta^2\lambda - \sigma^2/2]}{\theta^2} + \frac{[\sigma\tilde{D}(t; \tau)]^2\tau}{4\theta} \\ \tilde{D}(t; \tau) &= \frac{1 - \exp\{\theta\tau\}}{\theta\tau} \end{aligned}$$

Given that we have a set of observed historical yields in our disposal, we can set up a state space representation of the multi-factor process, which we can estimate using the Kalman filter. For the exposition we will present an example of the two factor case, but generalizations are straightforward.

The transition equations for the factors will take place under the physical measure. Both factors evolve in continuous time and are unobserved, but given the OU structure their distribution over a discrete time interval  $\Delta t$  will be Gaussian. We can therefore write what will be the *transition equations* of the Kalman filter as

$$x_{t+\Delta t}^{(j)} = \beta^{(j)}x_t + \eta_t^{(j)}$$

with the coefficients given by (once again we discard the superscripts)

$$\begin{aligned} \beta &= \exp\{-\theta\Delta t\} \\ \sigma_{\eta}^2 &= \frac{\sigma^2}{2\theta}(1 - \exp\{-2\theta\Delta t\}) \end{aligned}$$

The *measurement equations* are based on the observed yields of different maturities, and will be linear in the factors. Of course the model prices will not match the observed historical yields exactly, and error terms need to be introduced. This will be due to the fact that every model is just an approximation

of reality, and is therefore mis-specified to some extent. We will denote with  $\epsilon_{t,k}$  the error term at time  $t$  that corresponds to the yield of maturity  $\tau_k$ . Thus, we write

$$y_t(\tau_k) = c\tau + \sum_j \tilde{C}^{(j)}(t; \tau_k) + \sum_j \tilde{D}^{(j)}(t; \tau_k)x_t^{(j)} + \epsilon_{t,k}$$

### A MULTI-FACTOR GAUSSIAN EXAMPLE

We illustrate the use of Kalman filtering with an example that implements the multi-factor version of the Vasicek model described above. In particular, we simulate a pair of factors that obey the system of stochastic differential equations

$$\begin{aligned} dx_t^{(1)} &= -0.05 x_t^{(1)} dt + 0.010 dB_t^{(1)} \\ dx_t^{(2)} &= -0.20 x_t^{(2)} dt + 0.030 dB_t^{(2)} \end{aligned}$$

with the two Brownian motions uncorrelated. The prices of risk are assumed zero, and therefore the dynamics under the risk adjusted probability measure remain unaltered. The instantaneous rate will be the sum of the two factors, and the yield curve will be an affine function of them.

Listing 7.8 implements a wrapper that converts the inputs of the Gaussian  $N$ -factor model to the form that is expected by the Kalman filter algorithm of listing 5.3. To implement the filter, there must be some discrepancy between model and observed yields, and for that reason we add a Gaussian noise  $\epsilon$  to each yield, with standard deviation  $\sigma_\epsilon = 0.1\%$ . Otherwise we can solve the yield curve for the factor values, and there would not be much filtering involved! The simulated yield curves that serve as the input are given in figure XXX, together with the simulated and filtered factor paths.

Given a well specified model and the true parameter values, the Kalman filter does an outstanding job in recovering the factor trajectories. Of course for the filter to be of any practical relevance, it will have to provide us with decent parameter estimates as well. We therefore turn in investigating the performance of the Kalman filter, if the parameter set is unknown.

Just like the standard Kalman filter, we will address the estimation problem with the maximum likelihood approach. But unlike the standard

## 7.7 FORWARD RATE MODELS

A more recent family of models utilizes the forward curve for modeling purposes,

The models described so far

Investing over the period  $(t, T)$  can be seen as similar to investing first over  $(0, S)$  and then reinvesting over  $(S, T)$

Of course we don't know at time  $t$  what interest rate will prevail at time  $S$  for a bond that matures at  $T$ ; therefore the second strategy is not risk free

LISTING 7.8: kf\_wrapper.m: A Kalman filter wrapper for the multi-factor Gaussian model.

```
% kf_wrapper.m
function [L, M, S] = kf_wrapper(p, N, dT, T, X)
p = abs(p); NT = length(T); m0 = zeros(N,1);
v0 = m0; b=m0; a=m0; v=m0;
C = zeros(NT,1); D = zeros(NT,N);
5   for j=1:N % loop over factors
        % allocate parameters
        theta = p(1+4*(j-1)); xbar = p(2+4*(j-1));
        sig2 = p(3+4*(j-1))^2; risk = p(4+4*(j-1));
10    % setup model and get bond yields
        m0(j) = xbar;
        v0(j) = 0.5*sig2/theta;
        b(j) = exp(-theta*dT);
        a(j) = xbar*(1-b(j));
15    v(j) = v0(j)*(1-b(j)^2);
        for k=1:NT % loop over maturities
            t = T(k);
            D(k,j) = (1-exp(-theta*t))/theta/t;
            C(k) = C(k) + ...
20            (xbar+risk-0.5*sig2/theta^2)*(1-D(k,j)) + ...
                0.25*sig2*t/theta*D(k,j)^2;
            end
        end
        % construct Kalman filter inputs
25    kf.CY = a; kf.CX = C; kf.BX = D;
    kf.BY = diag(b); kf.SH = diag(v);
    kf.SE = (p(end)^2)*eye(NT);
    kf.M0 = m0; kf.V0 = diag(v0);
    [L, M, S] = kalman_filter(kf, X);

```

But we can lock at time  $t$  an interest rate which will be applied over the interval  $(S, T)$

This will be the *forward rate*  $F(t; S, T)$

At  $S \rightarrow T$  we have the *instantaneous* forward rate  $F(t; T)$

No arbitrage indicates that the (continuously compounded) forward rate is

$$F(t; S, T) = -\frac{\log P(t; T) - \log P(t; S)}{T - S}$$

#### Forward rates and bond prices

The instantaneous forward rates are closely linked to bonds

In the limit, as we split the interval  $(S, T)$  into subintervals we reach

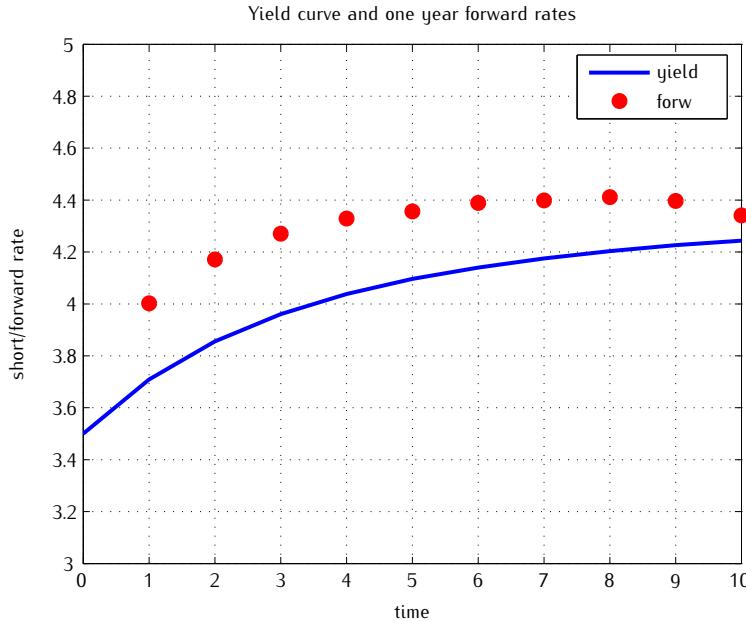


FIGURE 7.9: Yield and one-year forward curves.

$$\log P(t; T) - \log P(t; S) = - \int_S^T F(t; s) ds$$

which yields for  $S \rightarrow t$

$$\begin{aligned} \log P(t; T) &= - \int_t^T F(t; s) ds \\ \Rightarrow P(t; T) &= \exp \left\{ - \int_t^T F(t; s) ds \right\} \\ \Rightarrow F(t; T) &= \frac{\partial \log P(t; T)}{\partial T} \end{aligned}$$

Even though multi-factor models consider a larger set of parameters to be calibrated, they are still finite

Therefore a perfect fit to the initial yield curve cannot be ensured

Also, we are increasingly interested in matching volatility structures implied from cap/swaption prices

Forward rate models have the whole yield curve as an input (and possibly a volatility curve as well)

These models were introduced in [Heath, Jarrow, and Morton \(1992, HJM\)](#): we exploit the link between bond prices and the forward curve

In essence we model each bond maturity with a separate SDE

Therefore we are facing a system of infinite SDEs, with the initial forward curve as a boundary condition

Of course, some relationships will ensure that no arbitrage is permitted

In particular, if the forward rate dynamics are given by

$$dF(t; T) = \mu(t, T)dt + \sigma(t, T)dW(t)$$

Ito's formula gives the bond dynamics

$$\begin{aligned} \frac{dP(t; T)}{P(t; T)} &= \left( r(t) + \mu^*(t, T) + \frac{1}{2} \|\sigma^*(t, T)\|^2 \right) dt \\ &\quad + \sigma^*(t, T)dW(t) \end{aligned}$$

The functions

$$\begin{aligned} \mu^*(t, T) &= - \int_t^T \mu(t, s)ds \\ \sigma^*(t, T) &= - \int_t^T \sigma(t, s)ds \end{aligned}$$

If we use the current account for discounting (as the numéraire), we expect the discounted bonds to form martingales

$$\frac{P(t; T)}{B(t)} = E \frac{P(T; T)}{B(T)} = E \frac{1}{B(T)}$$

HJM show that the no-arbitrage condition is

$$\mu(t, T) = \sigma(t, T) \int_t^T \sigma(t, s)ds$$

## CALIBRATION OF HJM MODELS

HJM models need

The initial forward rateThe volatility structure

The forward curve which follows from the yield curve. This is specified under risk neutrality

Since the forwards are given as derivatives of the yield curve one has to be careful when constructing the yield curve: many instruments on top of bonds are also used for that (e.g. swaps, futures)

The volatility structure can be specified using PCA on past yield curves.

Volatility structures are the same under all measures

Volatilities implied from derivatives can also be used

## SHORT VERSUS FORWARD RATE MODELS

Short rate models

- ⊕ Markovian in nature, easy to model
- ⊕ Many derivative prices in closed form
- ⊕ Tree building, finite differences or simulation: all easy
- ⊖ Arbitrage not ruled out
- ⊖ Volatility can be hard to model

Forward rate models

- ⊕ No arbitrage by nature
- ⊕ Very flexible volatility structures
- ⊕ Easy to include many factors
- ⊖ Short rate non Markovian generally
- ⊖ No Feynman-Kač representations
- ⊖ No trees or finite differences; only simulations

## 7.8 BOND DERIVATIVES

There is a large number of bond and interest rate derivatives with a liquid market

Forwards, swaps, bond options, caplets, floorlets, caps, floors and swaptions are some examples

The pricing of bond options is most important, since prices of caplets and floorlets can be expressed as an option on a zero, while swaptions can be expressed as an option on a coupon paying bond

Unlike equity options, bond options have some distinctive features that arise from the nature of interest rates

For example, bond prices are known both at the current time  $t$  and on maturity  $T$ , a feature known as a *pull to par*

**CIR distribution of 1 year zero price after 0.5 years**

### THE BLACK-76 FORMULA

Black (1976, B76) in an influential paper considered the pricing of options on commodities

Commodities have specific cycles, storage and availability costs that are not captured in the standard BS methodology

With some modifications the B76 formula is used in the market to quote bonds, caps, swaptions, etc

The B76 formula assumes that bonds or interest rates (depending on the instrument) are log-normally distributed, with  $\sigma$  being the measure of volatility

Implied volatility curves are constructed for all these instruments

See Brigo and Mercurio (2001) for all relevant formulas

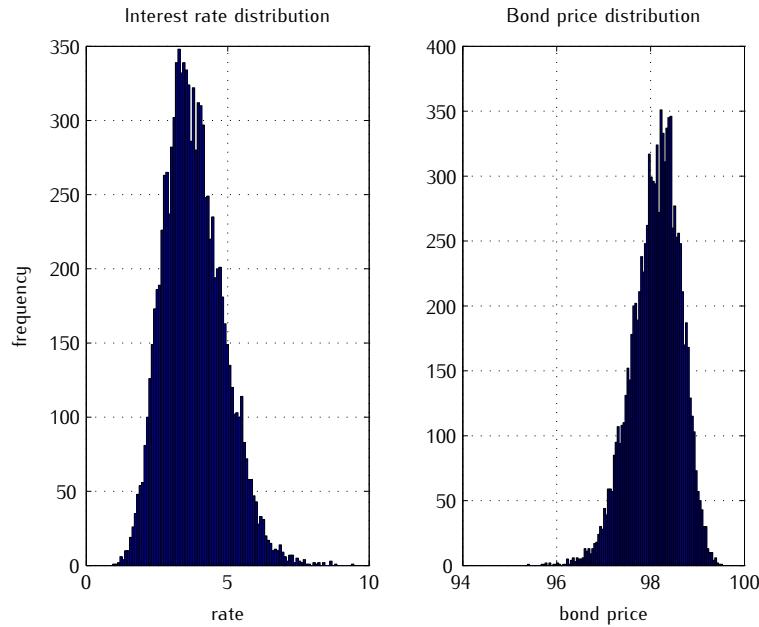


FIGURE 7.10: Pull-to-par and bond options.

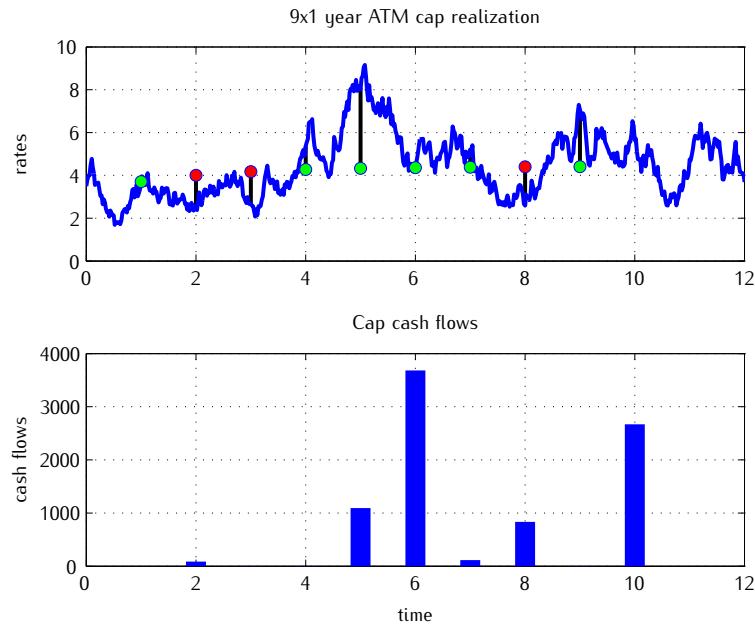


FIGURE 7.11: Cash flows for interest rate caplets and caps.

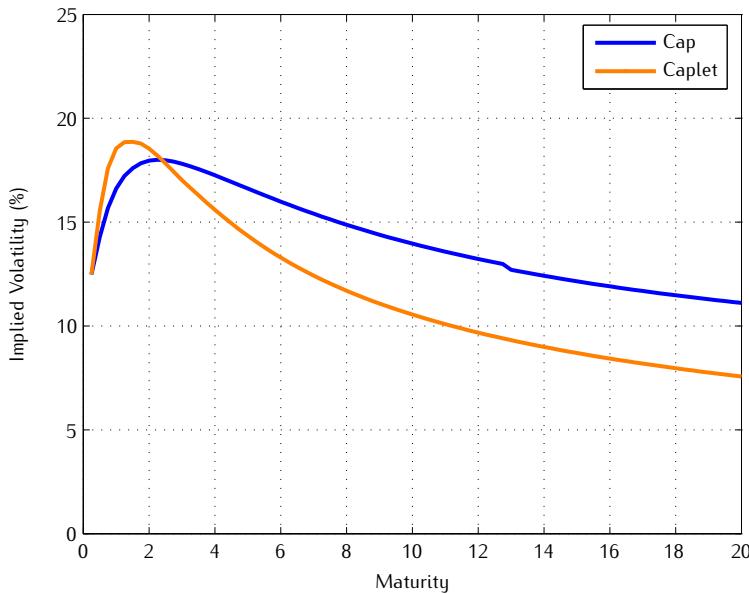


FIGURE 7.12: Typical Black volatilities for caplets and caps.

## 7.9 CHANGES OF NUMÉRAIRE

In general, if the payoffs are given as  $\phi(r(T))$ , then the price at time  $t$  is given by

$$\begin{aligned} f(r(t), t) &= E_t^Q \left\{ \exp \left( - \int_t^T r(s) ds \right) \phi(r(T)) \right\} \\ &= E_t^Q \left\{ \frac{B(t)}{B(T)} \phi(r(T)) \right\} \end{aligned}$$

If the discounting factor  $\frac{B(t)}{B(T)}$  was independent of the payoffs we would be able to split the integral

But since we are dealing with stochastic rates they are not

Implicitly we are using the bank account as the numéraire, but there is nothing special with this choice

As shown in Geman, el Karoui, and Rochet (1995), one can choose any positive asset as the numéraire

For each numéraire there exists an equivalent measure, under which every asset is a martingale

That is to say, if  $N(t)$  is the process of the numéraire, then there exists a measure  $N$  induced by this numéraire, such that

$$\frac{X(t)}{N(t)} = E_t^N \frac{X(T)}{N(T)}$$

for any asset process  $X(t)$ . Then, we can express the value at  $t$  as  $X(t) = N(t)E_t^N \frac{X(T)}{N(T)}$

Given a problem, a good numéraire choice can simplify things enormously

For example, we can use the bond that matures at time  $T$  as the numéraire; then all asset prices are given in terms of this asset (rather than currency units)

If  $T$  is the measure induced by this bond, we can write the payoffs as

$$f(r(t), t) = P(t; T) E_t^T \frac{\phi(r(T))}{P(T; T)} = P(t; T) E_t^T \phi(r(T))$$

To make this approach operational we need to find under which measure  $T$  all bonds discounted with  $P(t; T)$  form martingales

## 7.10 THE LIBOR MARKET MODEL

A variant of the HJM model, which constructs lognormal rates was proposed in [Brace, Gatarek, and Musiela \(1995\)](#) and [Miltersen, Sandmann, and Sondermann \(1997\)](#)

Since it produces prices that agree with the B76 market quotes the model has been coined the “market” model

It uses fixed maturity forward curves, rather than the instantaneous forward rate, since prices become explosive in that case

Typically the 3 month Libor rate is used as the underlying  
This is the model of choice for many practitioners





# A

---

## Using Matlab with Microsoft Excel

In many practical situations one needs to export Matlab functionality to a spreadsheet programme like Microsoft Excel. Fortunately, Microsoft Windows provide the functionality via the COM component objects. Using the Matlab compiler one can build a standalone COM component in the form of a dynamically linked library (DLL) which can be invoked from Visual Basic for Applications (VBA), the programming language used throughout Excel.

By using VBA one can construct an Excel add-in which can be exported to any computer running Excel. One of the main benefits is that all required Matlab functions are exported, and therefore the host computer need not have Matlab installed.<sup>1</sup> Also, Graphical User Interfaces (GUIs) can be constructed easily using VBA. The main reference of this appendix is [MathWorks \(2005, chapter 4.18\)](#).

In this appendix we will describe the procedure, and we will produce functions that implement the [Black and Scholes \(1973\)](#) pricing model for calls and puts. There are four+one steps in creating the Excel-Matlab link.

- 0 Setup the C/C++ compiler to work with Matlab
- 1 Write the functions in Matlab and create the COM component.
- 2 Write the VBA code that communicates with the DLL and performs the operations.
- 3 Create the GUI in Excel.
- 4 Put everything in a package that can be readily installed on any computer with Excel.

---

<sup>1</sup> Computers that don't run Matlab will need the Matlab Component Runtime (MCR) set of libraries which is freely available.

## A.1 SETTING UP MATLAB WITH THE C/C++ COMPILER

Before starting the procedure we have to ensure that the Matlab C compiler is properly set up. Running `mbuild -setup` at the Matlab command prompt will allow us to select the compiler we want to use. Since we need to compile COM components we will need the Microsoft Visual C/C++ compiler in our system. It is truly unbelievable but Microsoft is giving away the compiler for free as part of the Visual C++ 2005 Express Edition (VC). This is recorded as VC version 8.0, and it only compatible with Matlab 7.3.<sup>2</sup> You can download VC at

<http://msdn2.microsoft.com/vstudio/express/visualc/download/>

You should download and run the file `vcsetup.exe`, and install it to the default directory

`C:\Program Files\Microsoft Visual Studio 8\`

The second step is to get the Windows Platform SDK (Windows Server 2003 R2) from the web

<http://msdn.microsoft.com/vstudio/express/visualc/usingpsdk/>

You should download the installer `PSDK-x86.exe`. It is important to select a *custom install* and put as the target directory

`C:\Program Files\Microsoft Visual Studio 8\VC\PlatformSDK\`

This is where Matlab looks for some necessary files. You don't need to install all components. The required ones are the following

1. Microsoft Windows Core SDK ► Build Environment ► Build Environment (x86 32-bit)
2. Microsoft Windows Core SDK ► Redistributable Components
3. Microsoft Data Access Services (MDAS) SDK ► Tools ► Tools (x86 32-bit)
4. Microsoft Data Access Services (MDAS) SDK ► Build Environment ► Build Environment (x86 32-bit)
5. Debugging Tools for Windows

After everything is installed we are ready to setup the Matlab compiler. At the Matlab prompt just input `mbuild -setup` and then select the appropriate compiler.

---

<sup>2</sup> If you run an earlier versions of Matlab, like 7.0.4, you will need VC 6.0, 7.0 or 7.1. The VC 7.1 is shipped with the Visual C++ Toolbox 2003, which is not officially supported but is out there on the net.

LISTING A.1: Matlab file xl\_bs\_call.m

```
% xl_bs_call.m
function [P, D] = xl_bs_call(S, K, r, T, sigma)
d1 = log(S./K) + (r + 0.5*sigma.^2).*T;
d1 = d1./sigma./sqrt(T);
5 d2 = d1 - sigma.*sqrt(T);
N = normcdf(d1);
D = normcdf(d2);
P = S.*N - K.*exp(-r.*T).*D;
```

LISTING A.2: Matlab file xl\_bs\_put.m

```
% xl_bs_put.m
function [P, D] = xl_bs_put(S, K, r, T, sigma)
d1 = log(S./K) + (r + 0.5*sigma.^2).*T;
d1 = d1./sigma./sqrt(T);
5 d2 = d1 - sigma.*sqrt(T);
N = normcdf(-d1);
D = normcdf(-d2);
P = -S.*N + K.*exp(-r.*T).*D;
```

## A.2 WRITING THE MATLAB FUNCTIONS

We will now write the Matlab function that implement the standard BS prices and some hedging parameters. We are interested in passing whole arrays as arguments, and also want arrays to be returned. There are many ways of doing this, but to illustrate how different arrays are passed we will create two functions, shown in listings A.1 and A.2. Both functions should be straightforward. The set of parameters is input (perhaps some in vector form) and the prices and deltas are returned.

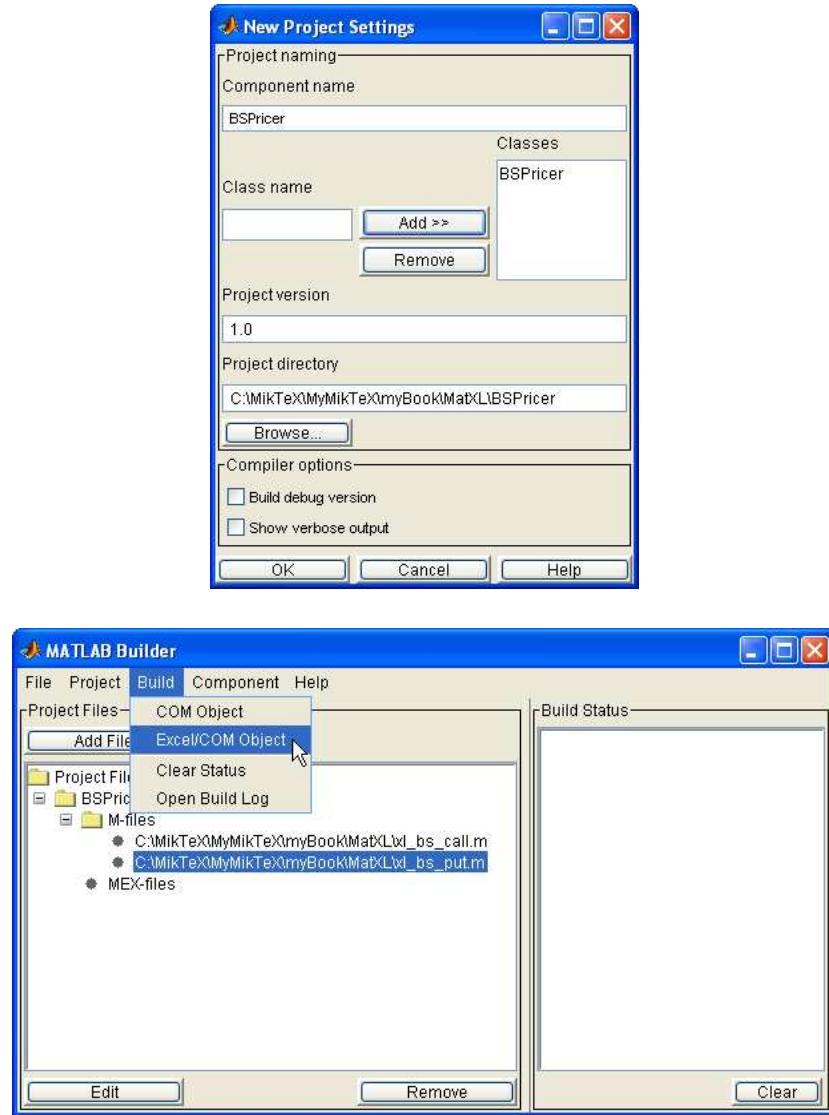
The second step is to create the COM component using the Matlab Excel builder. This is invoked by running `mxltool` at the Matlab command prompt. A window for the Matlab builder will then open. Go to `File ▶ New Project` to start a new Builder project. We set the parameters as shown in figure A.1.

Component name:	BSPricer
Class name:	BSPricerclass
Project version:	1.0

The next step is to add the files `xl_bs_call.m` and `xl_bs_put.m` to the project, by clicking on `Add File`. We must now save the project, and click on `Build ▶ Build COM/Excel Object` to actually build the component. Two sub-folders are created in the `BSPricer` folder, as shown in figure A.2.

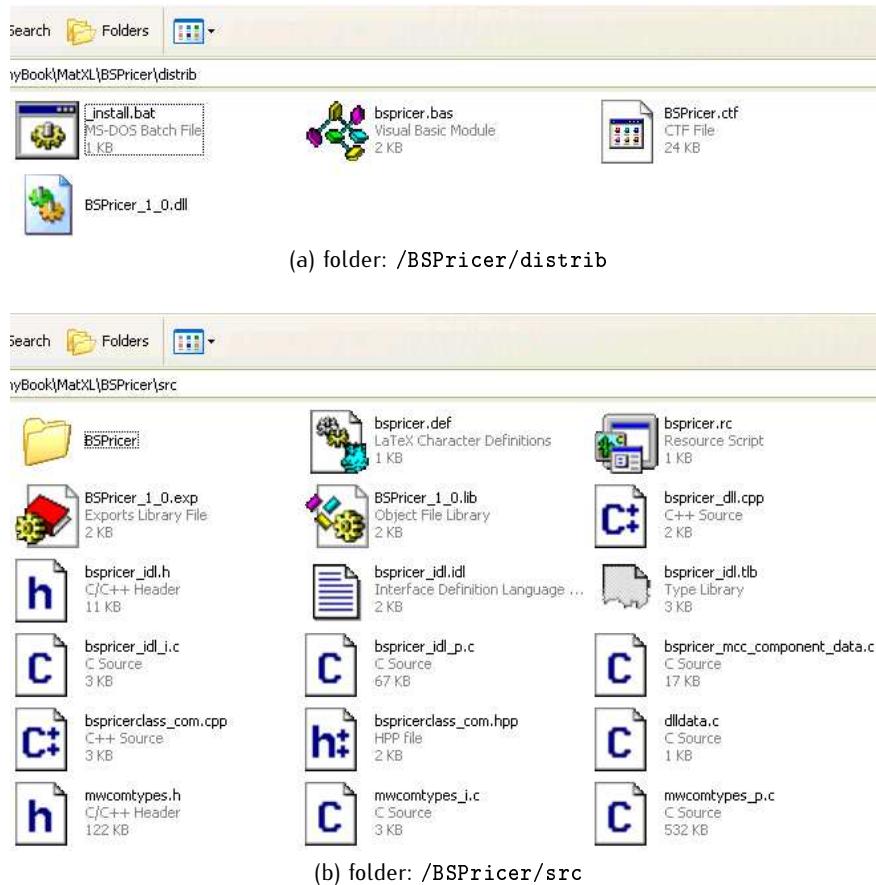
for copies, comments, help etc. visit <http://www.theponytail.net/>

FIGURE A.1: Screenshots of the Matlab Excel Builder



for copies, comments, help etc. visit <http://www.theponytail.net/>

FIGURE A.2: The folders created by mxtool



### A.3 WRITING THE VBA CODE

Open Excel and go to Tools ▶ Macro ▶ Visual Basic Editor. We will now need to tell Excel that the Matlab libraries shall become available. Within VBA go to Tools ▶ References and select the necessary libraries which are now available, the one we wrote and one that contains general Matlab utilities:

BSPricer 1.0 Type Library  
MWComUtil 1.0 Type Library

Now we will need to write some VBA code that initializes the add-in and also define some global variables that will be kept between function calls in Excel. To do that we need a module. Right-click on VBA Project (Book1) ▶ Insert

for copies, comments, help etc. visit <http://www.theponytail.net/>

## LISTING A.3: VBA module (PricerMain)

```
' PricerMain
Public myPricer As BSPricer.BSPricerclass
Public myUtil As MWUtil
Public bModuleInitialized As Boolean
5 Public StrikesRange As Range
Public CallPricesRange As Range
Public DeltasRange As Range
Public S0, sigma, InterestRate, Maturity As Double

10 Private Sub LoadPricer()
    Dim myForm As PricerForm
    On Error GoTo HandleError
    Call InitPricer
    Set myForm = New PricerForm
15    Call myForm.Show
    Exit Sub
HandleError:
    MsgBox (Err.Description)
End Sub

20 Private Sub InitPricer()
    If Not bModuleInitialized Then
        On Error GoTo Handle_Error
        If myUtil Is Nothing Then
            Set myUtil = New MWUtil
            Call myUtil.MWIInitApplication(Application)
25        End If
        If myPricer Is Nothing Then
            Set myPricer = New BSPricer.BSPricerclass
        End If
30        S0 = 100#
        sigma = 0.15
        InterestRate = 0.04
        Maturity = 0.25
35        bModuleInitialized = True
        Exit Sub
Handle_Error:
        bModuleInitialized = False
        End If
40 End Sub
```

- Module. Change the name of this module (at its properties) to PricerMain, and insert the code given in listing A.3.

## LISTING A.4: VBA Activation Handlers (PricerForm)

```

' PricerForm
private Sub UserForm_Activate()
    On Error GoTo HandleError
    SBox.Value = S0
    SigmaBox.Value = sigma
    InterestRateBox.Value = InterestRate
    MaturityBox.Value = Maturity
    If Not StrikesRange Is Nothing Then
        StrikesRangeBox.Text = StrikesRange.Address
    End If
    If Not CallPricesRange Is Nothing Then
        CallPricesRangeBox.Text = _
            CallPricesRange.Address
    End If
    Exit Sub
HandleError:
    MsgBox (Err.Description)
End Sub

```

We now need to turn to the GUI, which can be as in the screenshot A.3. The components need some event handlers, that will respond to the activation of the form and user input. All this code must reside within the PricerForm form code. When the form is activated some initial values must be set, which is done in listing A.4. The user can click either the compute... button (listing A.5) or the about... button (listing A.6).

## A.4 THE EXCEL ADD-IN

The last part is to create the code that puts the add-in in Excel. Right-click on ThisWorkbook ► View Code and add the code of listing A.7. This code will install and uninstall the add-in in the Tools menu item of Excel. A button BSPricer... is added which invokes the LoadPricer subroutine when clicked.

Now we can save the add-in in the /BSPricer/distrib directory, ready for packaging. Note that the Excel file has to be saved as an Excel add-in file (.xla).

## A.5 INVOKING AND PACKAGING

To check that everything is all right, we can close Excel and reopen it. Within Tools ► Add-ins we should be able to locate BSPricer.xla. Then a BS

for copies, comments, help etc. visit <http://www.theponytail.net/>

## LISTING A.5: VBA User Input Handlers | (PricerForm)

```
' PricerForm
Private Sub OkButton_Click()
    Dim Rng As Range
    If myPricer Is Nothing Then GoTo ExitForm
    'On Error Resume Next
    Set Rng = Range(StrikeRangeBox.Text)
    If Err <> 0 Then
        MsgBox ("Invalid range for strike prices")
        Exit Sub
    End If
    Set StrikesRange = Rng
    Set Rng = Range(CallPricesRangeBox.Text)
    If Err <> 0 Then
        MsgBox ("Invalid range for call prices")
        Exit Sub
    End If
    Set CallPricesRange = Rng
    Set Rng = Range(DeltasRangeBox.Text)
    If Err <> 0 Then
        MsgBox ("Invalid range for deltas")
        Exit Sub
    End If
    Set DeltasRange = Rng
    If CallPricesRange.Cells.Count <> _
        StrikesRange.Cells.Count Then
        MsgBox ("Selections must have the same size")
        Exit Sub
    End If
    S0 = CDbl(SBox.Text)
    sigma = CDbl(SigmaBox.Text)
    InterestRate = CDbl(InterestRateBox.Text)
    Maturity = CDbl(MaturityBox.Text)
    Call myPricer.xl_bs_call(2, CallPricesRange, _
        DeltasRange, S0, StrikesRange, _
        InterestRate, Maturity, sigma)
    GoTo ExitForm
HandleError:
    MsgBox (Err.Description)
ExitForm:
    Unload Me
End Sub
```

## LISTING A.6: VBA User Input Handlers II (PricerForm)

```

' PricerForm
Private Sub AboutButton_Click()
    Dim myAbout As AboutForm
    Set myAbout = New AboutForm
    myAbout.Show
End Sub

```

## LISTING A.7: VBA Add-in installation (thisWorkbook)

```

' ThisWorkbook
Private Sub Workbook_AddinInstall()
    Call AddPricerMenuItem
End Sub
5
Private Sub Workbook_AddInUninstall()
    Call RemovePricerMenuItem
End Sub

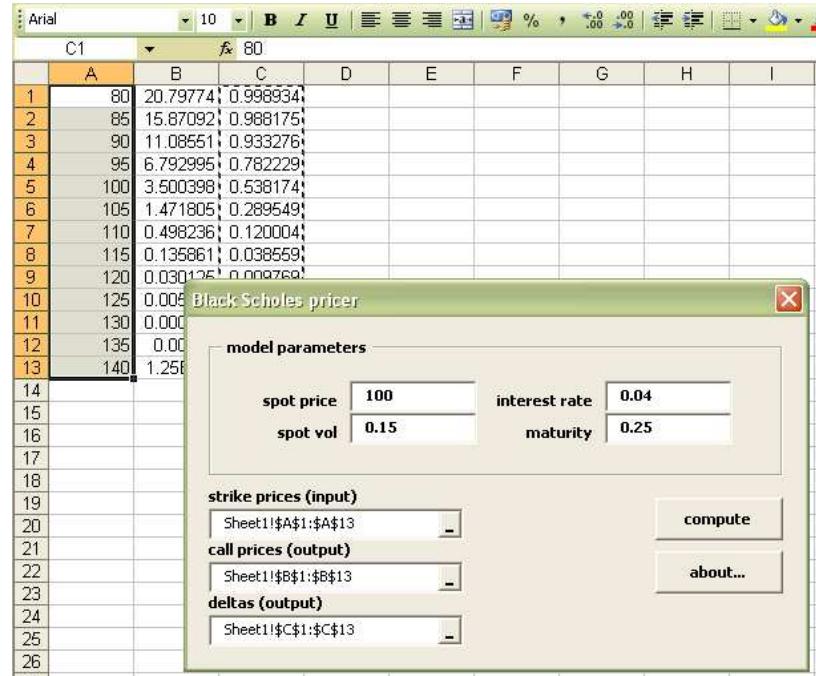
10 Private Sub AddPricerMenuItem()
    Dim ToolsMenu As CommandBarPopup
    Dim NewMenuItem As CommandBarButton
    Call RemovePricerMenuItem
    Set ToolsMenu = _
        Application.CommandBars(1).FindControl(ID:=30007)
    If ToolsMenu Is Nothing Then Exit Sub
    Set NewMenuItem = _
        ToolsMenu.Controls.Add(Type:=msoControlButton)
    NewMenuItem.Caption = "BS\u2022Pricer..."
    NewMenuItem.OnAction = "LoadPricer"
20 End Sub

25 Private Sub RemovePricerMenuItem()
    Dim CmdBar As CommandBar
    Dim Ctrl As CommandBarControl
    On Error Resume Next
    Set CmdBar = Application.CommandBars(1)
    Set Ctrl = CmdBar.FindControl(ID:=30007)
    Call Ctrl.Controls("BS\u2022Pricer\u2022...").Delete
30 End Sub

```

Pricer... item should be now present in the menu. Invoking this item should allow us to run the DLL and compute option prices. A screenshot of the add-in in action is given in figure A.3.

FIGURE A.3: Screenshot of the BSPricer add-in.



Going back to Matlab and the Excel Builder tool, we can package the component. Matlab will put together the DLL and the .xla file, and will create an executable that registers the dynamic library with Windows.

We can now ship the add-in and use with a computer that does not have Matlab installed, but the host computer must have the freely available Matlab Component Runtime (MCR) libraries. The MCR must be the same version as the Matlab that created our file. We can build the MCR with our Matlab installation by using `mcrbuilder`, and then ship the `MCRInstaller.exe` with our add-in. Note that we only need to do this once: after the host computer has MCR properly setup, we can add more add-ins, given that they have been created using the same Matlab version.

---

## References

- Albrecher, H., P. Mayer, W. Schoutens, and J. Tistaert (2007, January). The little Heston trap. *Wilmott Magazine*, 83–92.
- Andricopoulos, A. D., M. Widdicks, P. W. Duck, and D. P. Newton (2003). Universal option valuation using quadrature methods. *Journal of Financial Economics* 67, 447–471.
- Bachelier, L. (1900). *Théorie de la Spéculation*. Gauthier-Villars.
- Bailey, D. H. and P. N. Swarztrauber (1991). The fractional fourier transform and applications. *SIAM Review* 33(3), 389–404.
- Bailey, D. H. and P. N. Swarztrauber (1994). A fast method for the numerical evaluation of continuous fourier and laplace transforms. *SIAM Journal on Scientific Computing* 15(5), 1105–1110.
- Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1993). Fractionally integrated generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*.
- Bajeux, I. and J. C. Rochet (1996). Dynamic spanning: Are options an appropriate instrument? *Mathematical Finance* 6, 1–16.
- Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *The Journal of Finance* 5, 2003–2049.
- Bakshi, G. and D. Madan (2000). Spanning and derivative-security valuation. *Journal of Financial Economics* 55, 205–238.
- Barle, S. and N. Cakici (1998). How to grow a smiling tree. *Journal of Financial Engineering* 7(2), 127–146.
- Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance and Stochastics* 2, 41–68.
- Barone-Adesi, G., R. Engle, and L. Mancini (2004). GARCH options in incomplete markets. Working Paper.
- Bates, D. S. (1998). Pricing options under jump diffusion processes. Technical Report 37/88, The Wharton School, University of Pennsylvania.
- Bates, D. S. (2000). Post-'87 crash fears in S&P500 futures options. *Journal of Econometrics* 94, 181–238.

- Bates, D. S. (2005). Maximum likelihood estimation of latent affine processes. *Review of Financial Studies*, forthcoming.
- Bauwens, L., S. Laurent, and J. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21(1), 79–109.
- Ben Hamida, S. and R. Cont (2005). Recovering volatility from option prices by evolutionary computation. *Journal of Computational Finance* 8(4), XX–XX.
- Bingham, N. H. and R. Kiesel (2000). *Risk-Neutral Valuation*. London, UK: Springer-Verlag.
- Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of Business* 45, 444–455.
- Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics* 3(1), 167–79.
- Black, F. and P. Karasinski (1991). Bond and option prices when short rates are lognormal. *Financial Analyst Journal* (Jul-Aug), 52–59.
- Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637–659.
- Bollerslev, T., R. Engle, and D. Nelson (1994). ARCH models. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, IV. Amsterdam: North Holland.
- Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988, February). A capital asset pricing model with time varying covariances. *Journal of Political Economy* 96(1), 116–131.
- Bollerslev, T. R. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307–327.
- Bouchaud, J.-P. and M. Potters (2001). More stylized facts of financial markets: Leverage effect and downside correlation. *Physica A* 299, 60–70.
- Brace, A., D. Gątarek, and M. Musiela (1995). The market model of interest rate dynamics. Working paper, University of South Wales, Australia.
- Breeden, D. T. and R. Litzenberger (1978). Prices of state contingent claims implicit in option prices. *Journal of Business* 51, 621–51.
- Brennan, M. J. (1979). The pricing of contingent claims in discrete time models. *The Journal of Finance* 34, 53–68.
- Brennan, M. J. and E. Schwartz (1982). An equilibrium model of bond pricing and a test of market efficiency. *Journal of Financial and Quantitative Analysis* 3, 301–329.
- Brigo, D. and F. Mercurio (2001). *Interest Rate Models: Theory and Practice*. New York, NY: Springer Verlag.
- Caines, P. (1988). *Linear Stochastic Systems*. Probability and Mathematical Statistic. New York, NY: John Wiley and Sons.
- Carr, P. (2002). Frequently asked questions in option pricing theory. Technical report, forth. *Journal of Derivatives*.
- Carr, P., H. Geman, D. Madan, and M. Yor (2002). The fine structure of asset returns: An empirical investigation. *Journal of Business* 75(2), 305–332.
- Carr, P., H. Geman, D. Madan, and M. Yor (2003). Stochastic volatility for Lévy processes. *Mathematical Finance* 13(3), 345–382.

- Carr, P. and D. Madan (1999). Option valuation using the Fast Fourier Transform. *Journal of Computational Finance* 3, 463–520.
- Carr, P. and D. Madan (2005). A note on sufficient conditions for no arbitrage. *Finance Research Letters* 2, 125–130.
- Carr, P. and L. Wu (2004). Time-changed Lévy processes and option pricing. *Journal of Financial Economics* 71(1), 113–141.
- CBOE (2003). VIX® CBOE volatility index. White Paper, Chicago Board Options Exchange.
- Chourdakis, K. (2002). Continuous time regime switching models and applications in estimating processes with stochastic volatility and jumps. Technical Report 464, Queen Mary, University of London.
- Chourdakis, K. (2005). Option pricing using the Fractional FFT. *Journal of Computational Finance* 8(2), 1–18.
- Christie, A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics* 3, 407–432.
- Cont, R. and J. da Fonseca (2002). Dynamics of implied volatility surfaces. *Quantitative Finance* 2(1), 45–60.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica* 53, 385–407.
- Crépey, S. (2003). Calibration of the local volatility function in a generalized Black-Scholes model using Tikhonov regularization. *SIAM Journal of Mathematical Analysis* 34, 1183–1206.
- Davidson, R. and J. G. MacKinnon (1985, February). The interpretation of test statistics. *Canadian Journal of Economics* 18(1), 38–57.
- Derman, E. (1999). Regimes of volatility. *RISK* 12(4), 55–59.
- Derman, E. and I. Kani (1994). Riding on a smile. *RISK* 7(2), 32–39.
- Derman, E. and I. Kani (1998). Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure of volatility. *International Journal of Theoretical and Applied Finance* 1(1), 61–110.
- Derman, E., I. Kani, and N. Chriss (1996). Implied trinomial trees of the volatility smile. *Journal of Derivatives* 3(4), 7–22.
- Derman, E., I. Kani, and J. Z. Zou (1996, July). The local volatility surface: Unlocking the information in index options pricing. *Financial Analysts Journal*, 25–36.
- Dothan, U. (1978). On the term structure of interest rates. *Journal of Financial Economics* 6(1), 59–69.
- Duan, J.-C. (1995). The Garch option pricing model. *Mathematical Finance* 5(1), 13–32.
- Duan, J.-C., G. Gauthier, and J.-G. Simonato (1999). An analytical approximation for the Garch option pricing model. *Journal of Computational Finance* 2, 75–116.
- Dueker, M. (1997). Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business and Economic Statistics* 15, 26–34.
- Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance* 6(4), 379–406.

- Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–1376.
- Dupire, B. (1993). Pricing and hedging with smiles. In *Proceedings of the AFFI Conference, La Baule*.
- Dupire, B. (1994). Pricing with a smile. *RISK* 7(1), 18–20.
- Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987–1008.
- Engle, R. and F. K. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.
- Eraker, B., M. Johannes, and N. Polson (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics* 19(2), 177–91.
- Fama, E. F. (1965). The behavior of stock market prices. *Journal of Business* 38, 34–105.
- Feller, W. E. (1951). Two singular diffusion problems. *Annals of Mathematics* 54, 173–182.
- Figlewski, S. and X. Wang (2000). Is the “leverage effect” a leverage effect? Working Paper, SSRN 256109.
- Gallant, A. R. and G. Tauchen (1993). SNP: A program for nonparametric time series analysis. version 8.3 user’s guide. Working Paper, University of North Carolina.
- Gatheral, J. (1997). Delta hedging with uncertain volatility. In I. Nelken (Ed.), *Volatility in the Capital Markets: State-of-the-Art Techniques for Modeling, Managing, and Trading Volatility*. Glenlake Publishing Company.
- Gatheral, J. (2004). A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. In *Global Derivatives and Risk Management*.
- Gatheral, J. (2006). *The Volatility Surface: A Practitioner’s Guide*. New York, NY: Wiley Finance.
- Geman, H., N. el Karoui, and J.-C. Rochet (1995). Changes of numéraire changes of probability measure and option pricing. *Journal of Applied Probability* 32, 443–458.
- Gerber, H. U. and E. S. W. Shiu (1994). Option pricing by Esscher transforms. *Transactions of the Society of Actuaries* XLVI, 99–191.
- Ghysels, E., A. Harvey, and E. Renault (1996). Stochastic volatility. In G. Maddala and C. Rao (Eds.), *Handbook of Statistics, 14, Statistical Methods in Finance*. North Holland.
- Glosten, L. R., R. Jagannathan, and D. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48(5), 1779–1801.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.
- Hamilton, J. D. and R. Susmel (1994). Autoregressive conditional heteroscedasticity and changes in regime. *Journal of Econometrics* 64, 307–333.

- Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies* 61, 247–264.
- Heath, D., R. Jarrow, and A. Morton (1992). Bond pricing and the term structure of interest rates: A new methodology. *Econometrica* 60(1), 77–105.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6, 327–344.
- Heston, S. L. and S. Nandi (2000). A closed-form GARCH option pricing model. *Review of Financial Studies* Frth, Frth.
- Ho, T. S. Y. and S.-B. Lee (1986). Term structure movements and pricing interest rate contingent claims. *Journal of Finance* 41, 1011–1029.
- Hull, J. C. (2003). *Options, Futures and Other Derivatives*. (5th ed.). New Jersey, NJ: Prentice Hall.
- Hull, J. C. and A. White (1987). The pricing of options with stochastic volatilities. *The Journal of Finance* 42, 281–300.
- Hull, J. C. and A. White (1990). Pricing interest rate derivative securities. *Review of Financial Studies* 3(4), 573–592.
- Hull, J. C. and A. White (1994). Numerical procedures for implementing term structure models I. *Journal of Derivatives* 2, 7–16.
- Hull, J. C. and A. White (1996). Using Hull-White interest rate trees. *Journal of Derivatives*, 26–36.
- Ikonen, S. and J. Toivanen (2004). Operator splitting methods for American option pricing. *Applied Mathematics Letters* 17, 809–814.
- ISDA (1998). EMU and market conventions: Recent developments. International Swaps and Derivatives Association document BS:9951.1.
- Jackwerth, J. C. and M. Rubinstein (1996). Recovering probability distributions from options prices. *The Journal of Finance* 51, 1611–1631.
- Jawaheri, A. (2005). *Inside Volatility Arbitrage: The Secrets of Skewness*. Hoboken, NJ: Wiley.
- Julier, S. J. and J. K. Uhlmann (1996). A general method for approximating nonlinear transformations of probability distributions. Technical report.
- Julier, S. J. and J. K. Uhlmann (1997). A new extension of the Kalman filter to nonlinear systems. In *International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, pp. 182–193.
- Kahl, C. and P. Jäckel (2005, September). Not-so-complex logarithms in the Heston model. *Wilmott Magazine*, 94–103.
- Karpoff, J. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis* 22, 109–126.
- Kendal, M. and A. Stuart (1977). *The Advanced Theory of Statistics*. (4th ed.), Volume I. London, U.K.: Charles Griffin and Co.
- Lagnado, R. and S. Osher (1997). A technique for calibrating derivative security pricing models: Numerical solutions of an inverse problem. *Journal of Computational Finance* 1(1), 13–25.

- Lamourex, G. and W. Lastrapes (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business and Economic Statistics* 23, 225–234.
- Lee, R. (2004a). The moment formula for implied volatility at extreme strikes. *Journal of Mathematical Finance* 14(3), 469–480.
- Lee, R. (2004b). Option pricing by transform methods: extensions, unification and error control. *Journal of Computational Finance* 7(3), 51–86.
- Longstaff, F. A. and E. S. Schwartz (1992). Interest rate volatility and the term structure: A two factor general equilibrium model. *Journal of Finance* 47(4), 1259–1282.
- Madan, D., P. Carr, and E. Chang (1998). The variance gamma process and option pricing. *European Finance Review* 2, 79–105.
- Mandelbrot, B. and H. Taylor (1967). On the distribution of stock price differences. *Operations Research* 15, 1057–1062.
- Marchuk, G. I. (1990). Splitting and alternating direction methods. In N. Holland (Ed.), *Handbook of Numerical Analysis*, Volume 1, pp. 197–462. Amsterdam, Holland.
- MathWorks (2005). *Matlab Builder for Excel 1.2.5 (User's Guide)*. The Math-Works.
- McKee, S., D. P. Wall, and S. K. Wilson (1996). An alternating direction implicit scheme for parabolic equations mixed derivative and convective terms. *Journal of Computational Physics* 126(1), 64–76.
- Merton, R. (1976). Option pricing when the underlying stock returns are discontinuous. *Journal of Financial Economics* 4, 125–144.
- Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Sciences* 4, 141–183.
- Merton, R. C. (1992). *Continuous Time Finance*. (2nd ed.). Blackwell Publishing.
- Miltersen, K., K. Sandmann, and D. Sondermann (1997). Closed form solutions for term structure derivatives with lognormal interest rates. *Journal of Finance* 52(1), 409–30.
- Modigliani, F. and M. Miller (1958). The cost of capital, corporation finance, and the theory of investment. *American Economic Review* 48, 261–297.
- Neftci, S. N. (2000). *Introduction to the Mathematics of Financial Derivatives*. (2nd ed.). Academic Press.
- Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curves. *Journal of Business* 60(4), 473–489.
- Nelson, D. B. (1990). Arch models as diffusion approximations. *Journal of Econometrics* 45, 7–39.
- Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.
- Øksendal, B. (2003). *Stochastic Differential Equations*. (6th ed.). New York, NY: Springer-Verlag.
- Pan, J. (1997). Stochastic volatility with reset at jumps. Permanent Working Paper.

- Peaceman, D. W. and J. H. H. Rachford (1955). The numerical solution of parabolic and elliptic differential equations. *Journal of the Society of Industrial and Applied Mathematics* 3, 28–45.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1992). *Numerical Recipes in C: The Art of Scientific Computing* (2nd ed.). Cambridge University Press.
- Protter, P. E. (2004). *Stochastic Integration and Differential Equations*. (2nd ed.). New York, NY: Springer-Verlag.
- Rogers, L. C. G. and D. Williams (1994a). *Diffusions, Markov Processes and Martingales. Volume 1: Foundations* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Rogers, L. C. G. and D. Williams (1994b). *Diffusions, Markov Processes and Martingales. Volume 2: Itô Calculus* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Rubinstein, M. (1985). Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE from August 23, 1976 through August 31, 1978. *The Journal of Finance* 40, 455–480.
- Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance* 49, 771–818.
- Sandmann, G. and S. J. Koopman (1998). Estimation of stochastic volatility models via monte carlo maximum likelihood. *Journal of Econometrics* 87(2), 271–301.
- Sandmann, K. and D. Sondermann (1997). A note on the stability of lognormal interest rate models and the pricing of eurodollar futures. *Mathematical Finance* 7(2), 119–125.
- Schöbel, R. and J. Zhu (1999). Stochastic volatility using an Ornstein-Uhlenbeck process: An extension. *European Finance Review* 3, 23–46.
- Schoutens, W., E. Simons, and J. Tistaert (2004, March). A perfect calibration! Now what? *Wilmott Magazine*, XX–XX.
- Sentana, E. (1995). Quadratic Garch models. *Review of Economic Studies* 62, 639–661.
- Shreve, S. (2004a). *Stochastic Methods in Finance v1: The Binomial Asset Pricing Model*. New York, NY: Springer-Verlag.
- Shreve, S. (2004b). *Stochastic Methods in Finance v2: Continuous Time Models*. New York, NY: Springer-Verlag.
- Skiadopoulos, G., S. Hodges, and L. Clelow (2000). Dynamics of the S&P500 implied volatility surface. *Review of Derivatives Research* 3, 263–282.
- Stein, E. M. and J. C. Stein (1991). Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies* 4, 727–752.
- Svensson, L. (1994). Estimating and interpreting forward interest rates: Sweden 1992-4. Discussion Paper 1051, Centre for Economic Policy Research.
- Thomas, J. W. (1995). *Numerical Partial Differential Equations*. Number 22 in Texts in Applied Mathematics. New York, NY: Springer.
- van der Merwe, R., N. de Freitas, A. Doucet, and E. Wan (2001). The unscented particle filter. In *Advances in Neural Information Processing Systems* 13.

- Vasiček, O. A. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.
- Wiggins, J. B. (1987). Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics* 19, 351–372.
- Wilmott, P., J. Dewynne, and S. Howison (1993). *Option Pricing. Mathematical Models and Computation*. Oxford, UK: Oxford Financial Press.
- Zakoian, M. (1994). Threshold heteroscedastic models. *Journal of Economic Dynamics and Control* 18, 931–955.
- Zellner, A. (1995). *Introduction to Bayesian Inference in Econometrics*. Chichester, UK: John Wiley and Sons.

---

## Index

The hook for the index entries.

- affine, 190, 193
- affine models, 152
- arbitrage, 35, 150
  - static, 169
- Arch model, 136
- Assymmetric Garch model, 143
- Black-Scholes model, 149
- Black-Scholes PDE, 175
- bond
  - coupon, 177
  - face value, 177
  - par, 177
  - yield, 177
- bond option, 187
- Borel algebra, 3
- Brownian motion, 186
- butterfly spread, 169
- calendar spread, 169
- cap, 187
- Capital Asset Pricing Model (CAPM), 145
- capital structure, 133
- CBOE, 132
- characteristic function, 149
- CIR model, 192
- compounding, 178
  - continuously, 178
- corporate bond, 177
- current account, 184
- Dothan model, 191
- Dow Jones index, 130, 140
- Effcient Method of Moments, 160
- Egarch model, 143, 160
- equivalent martingale measure, 142, 154
- equivalent probability measure, 188
- equivalent probability measures, 153
- Esscher transform, 148
- Euler equation, 147
- event, 3
- expectation hypothesis, 188
- explosive bank account, 192
- exponential martingale, 153
- exponential Vasicek model, 191
- exponentially weighted moving average model, 141
- fat tails, 130
- Feller process, 192
- Feynman-Kac formula, 159
- Figarch model, 141
- fixed income security, 177
- floor, 187
- forward curve, 184
- forward rate, 182
  - instantaneous, 184
- fundamental theorem of asset pricing, 188
- Garch
  - and stochastic volatility, 135
- Garch model, 137
  - and Arch( $\infty$ ), 137
  - and incomplete markets, 146
  - and persistent volatility, 137, 141
  - fractionally integrated, 141
  - Garch(1,1), 137
  - GED, 143
  - in-mean, 145
  - maximum likelihood estimation, 138
  - multidimensional, 145
  - non Gaussian, 142
  - skewness, 143
  - standard errors, 139
  - volatility forecasts, 137

generalized error distribution, 143  
Girsanov theorem, 150, 152  
Igarch model, 141  
implied density, 167, 173  
    Breeden-Litzenberger method, 175  
implied tree, 174  
implied volatility, 131  
    and Delta, 134  
    and expected volatility, 132  
    and moneyness, 134  
    and realized volatility, 142  
    skew, 134  
    smile, 134  
    surface, 133  
        dynamics, 134  
        sticky Delta, 135  
        sticky strike, 135  
SVI parameterization, 172  
inverse problem, 162  
Itô formula, 186  
  
Kalman filter, 161, 192  
Kolmogorov backward equation, 175  
Kolmogorov extension theorem, 174  
Kolmogorov forward equation, 175  
  
Leibniz rule, 173  
leverage effect, 133, 143, 149  
liquidity preferences theory, 189  
local volatility, 167, 174  
    function, 174  
    PDE representation, 175  
long memory, 141  
  
marginal rate of substitution, 147  
as Radon-Nikodym derivative, 148  
  
market segmentation, 189  
Markov chain, 148, 161  
Markov Chain Monte Carlo, 160  
maturity date, 35  
maximum likelihood estimation, 127  
    standard errors, 139  
mean reversion, 192  
measurable space, 3  
measure, 3  
mixture of distributions, 130  
model risk, 162  
no-arbitrage tests, 169  
Novikov condition, 188  
Ohrnstein-Uhlenbeck process, 151  
one-factor model, 184  
Ornstein-Uhlenbeck process, 190  
overfitting, 201  
Particle filter, 161  
penalty function, 162  
preferred habitat, 189  
price of interest rate risk, 185  
price of risk, 145  
prior information, 162  
Radon-Nikodym derivative, 154, 188  
    marginal rate of substitution, 148  
random variable, 1  
redundant claim, 35  
regimes  
    of volatility, 132  
regularization, 162  
    Tikhonov-Phillips, 162  
risk aversion, 150  
risk premium, 150  
    time varying, 146  
sample path, 1  
sample point, see sample path  
sample space, *see state space*  
Sharpe ratio, 145, 154  
short rate  
    stylized facts, 189  
short rate model, *see one-factor model*  
 $\sigma$  algebra, 3  
generated, 4  
Simulated Method of Moments, 160  
smoothing, 167, 176  
    Nadaraya-Watson, 168  
radial basis function, 167  
sovereign bond, 177  
SP500 index, 132, 140  
square root process, 151  
"square root" process, 192  
state space, 1  
stochastic volatility, 149, 185  
    and Garch, 135  
calibration, 161, 165  
estimation, 160  
PDE, 156  
replicating portfolio, 157  
Student-t, 142  
swaption, 187  
  
term structure PDE, 187  
transform methods, 152  
  
underlying asset, 35  
utility function, 147  
  
Vasicek model, 190  
Vega, 162  
vertical spread, 169  
Vitali set, 2  
VIX index, 132, 141, 142  
    and financial crises, 132  
    and realized volatility, 132  
volatility  
    and correlation with returns, 132, 143, 149  
    and financial crises, 130  
attractor, 131

clusters, 131                    volatility risk, 150                    shapes, 179  
cyclical, 131                    yield curve, 179                    theories of, 188  
long memory, 141                historical, 181  
persistence, 141                parametric forms, 179              zero-coupon bond, 177  
time varying, 130