Pricing Inflation and Interest Rates Derivatives with Macroeconomic Foundations

Gabriele Luigi Sarais

Imperial College London

Department of Mathematics

A Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in Financial Mathematics

June 4, 2015

I Gabriele Luigi Sarais declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given in the bibliography.

Abstract

I develop a model to price inflation and interest rates derivatives using continuous-time dynamics linked to monetary macroeconomic models: in this approach the reaction function of the central bank, the bond market liquidity, and expectations play an important role. The model explains the effects of non-standard monetary policies (like quantitative easing or its tapering) on derivatives pricing.

A first adaptation of the discrete-time macroeconomic DSGE model is proposed, and some changes are made to use it for pricing: this is respectful of the original model, but it soon becomes clear that moving to continuous time brings significant benefits.

The continuous-time model is built with no-arbitrage assumptions and economic hypotheses that are inspired by the DSGE model. Interestingly, in the proposed model the short rates dynamics follow a time-varying Hull-White model, which simplifies the calibration. This result is significant from a theoretical perspective as it links the new theory proposed to a well-established model. Further, I obtain closed forms for zero-coupon and year-on-year inflation payoffs. The calibration process is fully separable, which means that it is carried out in many simple steps that do not require intensive computation.

The advantages of this approach become apparent when doing risk analysis on inflation derivatives: because the model explicitly takes into account economic variables, a trader can assess the impact of a change in central bank policy on a complex book of fixed income instruments, which is not straightforward when using standard models.

The analytical tractability of the model makes it a candidate to tackle more complex problems, like inflation skew and counterparty/funding valuation adjustments (known by practitioners as XVA): both problems are interesting from a theoretical and an applied point of view, and, given their computational complexity, benefit from a tractable model. In both cases the results are promising.

Contents

List of Figures

Acknowledgments

I feel obliged to thank all the people who have supported me during my PhD, without whom this would have not certainly been possible.

Firstly, I wish to thank my supervisor, Prof. Damiano Brigo, who has guided me in these years and provided very useful feedback in all phases of this project, from the inflation model construction, to the option skew extension, and on the final chapter on XVA. Damiano has also provided invaluable help to decide the overall structure of the thesis and crucially has helped me find the right balance between theoretical development and market data analysis, which I think is a distinctive feature of this thesis.

Secondly, I wish to thank prof. Lane Hughston and prof. Andrea Macrina, with whom I had useful discussions on inflation modelling and pricing kernel techniques. In particular, prof. Lane Hughston has followed the first part of the project (roughly speaking the first two chapters of this dissertation).

Thirdly, I wish to thank my employer, Mediobanca S.p.A. for the financial support and the flexibility they showed in these years. Further, I am grateful to Lorenzo Cornalba, Daniele Perini, and Antonio Calabrese in the Financial Engineering team, with whom I had very useful discussion on XVAs and Marshall-Olkin models.

Fourthly, I wish to thank Denis Mustafa and Steven Allsopp in the Risk team of Morgan Stanley, for the support and flexibility they have given me and for the very useful discussions we had on inflation and interest rates exotic products and their risk profiles.

Finally I wish to thank my parents for their support and my wife Laura, who has efficiently prevented things from falling apart at home while showing support, generosity, patience and true love. She was the first one to suggest and encourage me to start a PhD. She has also tolerated my chaotic desk and bookshelves at home, and many days when I was busy because of my studies.

But probably the best source of inspiration in these years have been the smiles and the gazes of my son Francesco Leone and my daughter Elena Brunilde.

Introduction

The objective of this dissertation is to propose a new model to price inflation derivatives, based on some macroeconomic dynamics and some models of central bank behaviour. The result is a pricing model that is both meaningful from an economic perspective and computationally tractable: closed forms for vanilla payoffs are available and the calibration process is separable. Building on this computational tractability, I investigate two issues that are relevant both from a theoretical and a practical point of view: modelling the inflation options smile and pricing counterparty and funding valuation adjustments.

The thesis is structured as follows: chapter 1 introduces the inflation markets, market participants and traded payoffs. A brief review of the no-arbitrage paradigm is provided. A literature review of the most widely-used pricing models for inflation-linked payoffs concludes the chapter. There is no original contribution in this chapter, but I feel it is important to include all definitions used in the following chapters.

Chapter 2 introduces the discrete-time DSGE macroeconomic model in its baseline version, following standard macroeconomic theory. The model is supplemented with a Taylor rule to model central bank policy, where the central bank sets the short rate as a function of inflation and growth. A first attempt to use this model for pricing purposes is made, essentially by changing the pricing measure and by writing the model volatilities as a function of the model parameters. On a recent article appeared on Bloomberg, Finance professor Noah Smith suggests that financial institutions should use more the DSGE macroeconomic model to fully understand the causality relationships in the economy ("Wall Street Skips Economics Class", www.bloombergview.com/articles/2014-07-23/wall-street-skips-economics-class): this seems to confirm the intuition behind this dissertation.

This first original attempt to bridge the gap between macroeconomic monetary models and pricing theory probably errs on the side of respecting too much the original discrete-time model, that is left almost unchanged: the pricing formulas can be obtained only by making some approximations. With this in mind, I propose some continuous-time dynamics for the macroeconomic variables that are broadly speaking inspired by the discrete-time DSGE macroeconomic model. This is done because, in my view, modelling stochastic processes and financial quantities in continuous time allows more flexibility com-

pared to the discrete-time case. I run some analysis to show that the two approaches can deliver very similar distributions for inflation and nominal rates: the continuous-time model is more tractable and is chosen to develop the theory further. Some evidence shows that the dynamics I propose fit historic data quite well. The reader interested in the core of the dissertation can skip this chapter and move to the following one.

Chapter 3 is the core of the dissertation. Here the continuous-time dynamics proposed in the previous chapter are complemented by a new formulation of the central bank policy (where the money supply is the policy tool), by some bond market liquidity relationships and by the non-arbitrage principle. All these are put together to obtain three pricing conditions: the model is referred to as "Continuous-Time Central Bank" model, or CTCB model. Interestingly, one of these conditions can be differentiated to obtain the dynamics of the short nominal rate, that follows a time-inhomogeneous Hull-White mean-reverting model. This result is extremely relevant both from a theoretical perspective, as it links the new theory proposed here to a well-established short-rate model, and from a practical point of view, given that closed forms for bond options, caps/floors, and swaptions are available in this framework. The analytical tractability of the model also allows to obtain closed forms for the most widely-traded inflation payoffs, like zero-coupon and year-on-year forwards and options.

In chapter 4 I use the closed-forms pricing formulas obtained in the previous chapter to propose a calibration strategy: the idea is to choose *a priori* some structural economic parameters, to calibrate other parameters to the nominal and inflation swap term structure, and to the prices of at-the-money interest rates caps/floors and inflation zero-coupon options. I propose a strategy to model the instantaneous correlations in the model while preserving the marginal distribution of the state variables. A full calibration to market data is carried out. The chapter ends with some examples of applications of this model, highlighting in particular that this model is both a realistic description of the economic environment and calibrates to market observables: this is a distinctive advantage, especially given the recently-introduced economic-based stress tests that banks have to perform regularly.

Chapter 5 is devoted to option skew in inflation markets. First I explain the peculiarities of this market and investigate what models are most often used to quote prices of inflation vanilla options. I propose a new pricing formula for an inflation option assuming that the underlying inflation rate follows a Student's t-distribution, and I compare the fit to market prices of the Student's model, the Gaussian model, the SABR model, and a Gaussian mixture. These models are widely used in the industry to match market prices: as a consequence they do not provide any meaningful insight on the underlying dynamics. After this first section, the objective is to define some meaningful dynamics that also calibrate the option skews observed in the market. To this end, I choose a time-inhomogeneous Merton-type jump-diffusion model, for which I provide two original results: an algorithm to convert the time-varying parameters into

constant parameters by keeping the same terminal distribution, and an algorithm to solve the partial-integro differential equation associated to an uncertain-parameters version of the time-inhomogeneous jump diffusions. Jump-diffusion models with uncertain parameters can be regarded as an extension of the original CTCB model when one assumes uncertain parameters. The chapter is concluded with a calibration of an uncertain-parameters jump-diffusion model to both zero-coupon and year-on-year Euro-area inflation options, showing that these two markets are implying different dynamics. This shows that the inflation options market is not perfectly liquid.

Chapter 6 is were the CTCB model is extended to price counterparty and funding valuation adjustments: this topic is of extreme practical relevance given that financial institutions are in the process to optimise and hedge their counterparty and funding risk profiles. This chapter contains an extension of the original CTCB model to include the dynamics of default probabilities, interest rates basis, and a Marshall-Olkin model to correlate the defaults of the derivative counterparty and of the derivative seller. The credit terms are partially correlated to the economic variables to model wrong-way risk. This extension is possible thanks to the analytical tractability of the model: finally some simulations show that this model is working as expected and that the wrong-way risk component is modelled in a satisfactory way.

Chapter 1

Trading and modelling inflation

In this chapter we present some preparatory material to develop the theory. [1]

Firstly, inflation markets are described by the price indexes, the traders, and the traded payoffs. The features of these items are described here with a market-oriented angle.

Secondly, a brief introduction to arbitrage theory and pricing kernel properties is followed by some explanation on how the no-arbitrage principle implies some model-independent relationships between some inflation payoffs traded in the market. Because some models presented in the following chapters are set in discrete time, we also present a brief review of the non-arbitrage theory in discrete time.

Finally, a short literature review describes some aspects of the most important pricing models that have been proposed so far to price inflation-linked securities.

1.1 Inflation markets

The economic intuition behind inflation-linked securities is that market agents consume real goods and use money only to purchase them: therefore an increase in the price level reduces their purchasing power *ceteris paribus*. This matters in particular over the long term, because households save to defer consumption. Since price levels have generally increased over time, individuals need to protect their purchasing power. However, tax effects need to be taken into account in the formulation of the hedging strategy.

[1] This chapter does not contain any new result, and sources are clearly stated. However, it is helpful to collect all material that we refer to in the following chapters.

1.1.1 Price indices

Price indices quantify the price level evolution in the economy. Such indices are defined by considering the price of a representative basket of goods and services, normalised to 100 at a given past date (the base). The basket is reviewed annually to better capture consumption patterns. These indices are published monthly by government statistics offices and are subject to revision.

United Kingdom. The RPI (Retail Price Index) is the most important UK price index. Unlike the RPIX, the RPI includes mortgage payments and council tax. Its base year is 1987.

Euro area. The HICPxT (Harmonised Index of Consumer Prices excluding Tobacco) published by Eurostat is the most important Euro area price index and is defined as the consumption-weighted average of the CPI indices of the individual members of the Euro area.[2]

USA. The CPI-U (US City Average All Items Consumer Price Index for all Urban Consumers) is published by the Bureau of Labour Statistics and is the main US inflation index. Its housing weight is greater compared to its European equivalents.

1.1.2 Market participants

Inflation markets allow market participants to transfer their inflation risk, either in a funded (bond) or in an unfunded (swap) format. Governments have been issuing inflation-linked bonds (also known as "linkers") since the 1980s: a liquid inflation-linked market has been developing since the year 2000, in particular in Europe. A comprehensive guide to this market can be found in Deacon, Derry and Mirfendereski [48]. Many of the published research notes offered by investment banks provide further market colour, such as those of Barclays Capital [7] and Lehman Brothers [83].

Inflation market participants can be classified into three categories: inflation payers (structurally long inflation), inflation receivers (structurally short inflation), and inflation payer/receivers (inflation traders).

Inflation payers. Governments have been issuing inflation-linked bonds for a long time: the first was issued in 1742, when Massachusetts issued a bond linked to the price of silver (this bond was not linked to a price index, as the bonds issued in the last thirty years or so, but to a commodity that tracks the price level). The United Kingdom issued inflation-indexed Gilts in 1981, when inflation was high due to the oil shocks of the previous decade. At the end of the 1990s the USA and France issued the TIPS (Treasury Inflation Protected Securities) and the OAT-I (Obligations Assimilables du Trésor indexées sur l'inflation) respectively. Other countries have subsequently entered this market: examples are Australia, Canada, Italy and Sweden. Germany is a notable exception between developed countries, because the

[2] Tobacco is excluded from this index due to an old French law preventing indexation to tobacco prices.

Bundesbank has historically been adverse to any types of indexation, due to its hard stance against inflation (this is due to the devastating effects of the hyperinflation during the Weimar Republic)[3].

Supernational AAA-rated issuers have issued inflation-linked debt. Examples are the EIB (European Investment Bank) and the IBRD (International Bank for Reconstruction and Development).

Private companies whose cashflows are linked to the consumer price level, such as utilities or retailers, are naturally suited to pay inflation-linked coupons. This allows a better asset-liability management and saves private companies any inflation risk premium they would pay if they issued long-dated fixed coupon bonds. The inflation risk premium can be significant (around 200 basis points): these savings can be partially offset by the liquidity premium investors may request, since inflation-linked bonds are typically less liquid than their nominal equivalents. By issuing inflation-linked debt, private companies attract investors interested in diversifying their bond portfolios. Examples of inflation-linked corporate issuers are BG Transco, Tesco, Welsh Water, and Anglian Water in the UK.

Finally, the UK market has seen the issuance of inflation-linked debt in the context of PFI (Private Finance Initiative), where private capitals fund public infrastructure projects. Because the government pays the company a cashflow related to inflation, issuing inflation-linked debt is a natural way to match assets and liabilities. Examples of such issuers include Network Rail and South-East Water.

Inflation receivers. Pension funds liabilities are often linked to the cost of living, making this sector a natural inflation receiver. Private pension schemes, started in the UK and now spreading to the main developed countries, use both inflation bonds and inflation swaps to match their liabilities. The same applies to insurance companies that offer pension schemes alongside other types of coverage. OECD data showed that in 2002 UK pension funds and insurance companies held 75% of the inflation-linked Gilts.

Issuers of retail products linked to inflation (for example the Italian Post Office or some French banks) need to cover their liabilities too. Finally, asset managers offer inflation-linked bond mutual funds, and therefore hold significant amounts of linkers.

Inflation payers/receivers. Inflation desks of investment banks trade inflation to recycle the risk arising from inflation-linked structured products sold to clients. Banks are also primary dealers in inflation bonds issuances, and may need to recycle some of the exposure left in their books. Proprietary desks or hedge funds may either take directional views on inflation or exploit relative-value opportunities on the market. Banks and hedge funds have become major players in the inflation market: the inflation swap notional grew from very little in 2000 to 50 billion EUR in 2004. Inflation swaps currently trade with a low bid/offer spread, on the order of 2-3 basis points. The US swap market, which was lagging

[3]There is a debate between the economists on whether a government should issue inflation-linked securities. On one side, some economists argue that this commits the public sector to moderate price growth (otherwise the government pays higher inflation-linked coupons). On the opposite side, other economists argue that any price indexation would involve the risk of spiralling prices.

behind the European one, has tripled in size between 2008 and 2012.

Central Banks. Central Banks monitor the prices of inflation-linked instruments in order to extract market expectation for inflation, to obtain indirect feedback on their perceived credibility in fighting inflation, as explained in Hurd and Relleen [68].

1.1.3 Payoffs

We list several examples of inflation-linked securities, starting from the simplest up to the most exotic and hybrid ones. Here I_t is the level of the underlying price index at time t. At this stage we do not specify if time is a continuous or a discrete variable: time here is an index. The model proposed in chapter 2 is a discrete time model, where the price index at time t_i written as I_i, while in chapter 3 and following we build continuous-time models, where the price index at time t is written $I(t)$.

The inflation rate p_t is defined as the realised annualised percentage growth rate over a certain time lag of the price index I_t.

The contracts listed below can be traded over the counter, with the only exception of the inflation futures, which are traded on the Chicago Mercantile Exchange (albeit volumes are still small).

Inflation-linked bonds. Inflation-linked bonds were the first kind of inflation-linked security to be traded. They are issued by governments, supernational agencies, and private companies. They can be structured as bonds that pay a fixed notional at maturity plus some inflation-linked coupons (Capital Index Bond, or CIB), or as bonds that pay an inflation-adjusted notional at maturity plus some fixed coupons (Interest Indexed Bond, or IIB). A combination is possible: for example both Gilts in the UK and TIPS in the US adjust both notional and coupons to the realised inflation rate. Maturities are generally long. Inflation coupons are normally floored to zero to avoid payments to the issuer in case of deflation (i.e. negative inflation). These floors have been out of the money for long time, but the deflationary crisis of 2008-2009 has made these floors more important.

Since price indices are highly seasonal (we think to January sales or the Christmas season, for example) the inflation index coupons are normally calculated using the year-on-year percentage change of the price index. This avoids using de-seasonalised indices. Coupons are time-lagged, as the price index can be revised by the statistics office. For example, the current lag for inflation-index Gilts is three months. Finally, inflation bond indices have been developed by banks and are used as benchmark for inflation bond mutual funds. The most important ones are currently produced by Barclays and J. P. Morgan.

Another reason for which linkers are traded is the so-called inflation bond arbitrage: Fleckenstein, Longstaff & Lustig [57] describe how nominal bonds are normally more expensive compared to the equivalent inflation-linked bond. An arbitrage can be set-up (essentially a long position on the inflation-

linked bond, a short position on the nominal bond with the same maturity, and a zero-coupon inflation swap to hedge the inflation risk): the authors measure the arbitrage size and propose some explanations, mainly via liquidity and supply-demand arguments.

Inflation-linked zero-coupon bonds. These inflation bonds, that do not pay coupons, have been issued for example by Sweden.

Zero-coupon inflation index swaps (ZCIIS). In a ZCIIS contract the following cash flows are exchanged at maturity (assumed in T years) for a unit notional:

1. Party A pays in cash $(1+X)^T - 1$ to party B;

2. Party B pays the price index percentage change $(I_T/I_0) - 1$ to party A.

The strike X is often informally regarded as a market expectation of the (average) inflation rate over the next T years. It seems unlikely that this interpretation can be accepted literally. In any case, as it is impossible to forecast inflation over the long term, long-dated ZCIIS strikes are driven more by supply and demand rather than by actual market views.

These contracts are highly liquid in inter-dealer markets and are quoted for all maturities: they are the prototypical inflation derivative thanks to their simplicity and are used both to hedge exposures and to take a view. Another significant advantage is that they do not require any balance sheet at inception, because they are structured as swaps. Normally the unrevised version of the underlying index is used to determine the final payoff.

ZCIIS are important from a theoretical viewpoint: as shown by Brigo & Mercurio [22] their price is model-independent and is the difference between a real and nominal zero-coupon bond prices (the real zero-coupon bond pays at maturity one unit of the price index—this will be better defined in the following sections).

Finally, ZCIIS are normally more expensive than their theoretical inflation-linked bond equivalents since they are more liquid and do not require any balance sheet (at least at inception). As a consequence, they are a popular way to take a view on inflation, especially for the Euro area and the UK. Not surprisingly, the theoretical yield of linkers (also known as "breakeven") differs from the ZCIIS strikes: this basis may become important in stressed financial conditions, as it happened during the 2008-2009 crisis[4]. Because standard ISDA (International Swaps and Derivative Association) agreements are usually in place, collateral is posted to mitigate counterparty risk. This fact implies that there are important funding aspects of inflation derivatives, that will be discussed in full detail in chapter 6. The US swap

[4]Campell, Shiller and Viceira [35] provide detailed time series of the TIPS basis during the Lehman crisis, and argue that during this period the bond-implied breakeven collapsed more than the inflation-swaps breakevens as financing long positions in TIPS became more expensive due to the tensions in the interbank markets. This spread reached approximately 110 bps in early 2009. Therefore inflation expectations where better captured by inflation swaps.

market is much less developed and is still dominated by the TIPS, showing that a strong inflation-linked bond market does not necessarily imply the existence of a highly liquid inflation swap market.

Inflation asset swaps. In an inflation asset swap, a dealer buys an inflation-linked bond by borrowing money (therefore it pays an amount linked to the floating interest rate) and receives the inflation linked coupon. To hedge this payment the dealer enters an inflation swap where it pays inflation and receives a fixed amount at maturity. The dealer finally hedges the interest rate risk (it receives fixed and pays floating) by entering a standard interest rates swap. The dealer makes a profit if the inflation swap breakeven (which he receives) is higher than the one implied by the inflation-linked bond (which he buys with borrowed cash or from a repo transaction). It is worth noting that past realised inflation (which is accrued in the inflation-linked coupon) enters into the valuation of the inflation-linked asset swap, as noted in [94]. Clearly such trades leaves the trader with open refinancing risk, because the repo market maturities are normally shorter that the bond maturities.

Year-on-Year inflation index swaps (YYIIS). The inflation leg of a year-on-year inflation index swap (YYIIS) pays each year the percentage change of the price index for a unit notional (i.e. $(I_Y/I_{Y-1})-1$). As shown in Brigo & Mercurio [22] the pricing of a YYIIS is model-dependent because the correlation between the discount rate and the inflation rate comes into play: this gives rise to a convexity adjustment. The other leg of a YYIIS either pays a fixed or floating rate. YYIIS are natural hedges against inflation caps/floors, which will be introduced shortly after. Due to the model dependence mentioned above, these products are traded in much lower volumes than their zero-coupon counterparties.

Pay-as-you-go structures. These are inflation swaps where the inflation payments happen throughout the life of the trade to avoid a balloon payment at maturity, to reduce the counterparty risk and make the deal less credit-intensive. This format is often chosen by investment banks when dealing with corporate counterparties that do not post collateral.

Inflation futures. Inflation futures are a more recent product and are still not very liquid (they were introduced in 2004). These contracts are exchange-traded at the CME: currently there are twelve quarterly contracts available to trade up to three years. Their price is defined by: $100 -$ Contract Price. For example, a price of 98 for the March 2016 contract implies an annualised inflation rate of 2% between November 2015 and February 2016.

Dual currency swaps. In some countries of Central and Latin America the central bank publishes the price of a parallel unit of account, adjusted by the price index appreciation. An example are the UDI (Unidades de Inversion) in Mexico. A swap where one floating leg pays the nominal rate and the other one pays the UDI can therefore be regarded as a swap on the real rate.

Inflation zero-coupon options. Inflation zero-coupon options pay the realised inflation rate capped

or floored. For example the payoff of an inflation call is:

$$\max\left(\frac{I_T}{I_0} - 1 - X, 0\right).$$

This is equivalent to an option on the price index level at time T with strike $I_0(1+X)$ and notional I_0. Inflation zero-coupon options are the natural building blocks to calibrate the model to the smile, as shown in chapters 4 and 5.

Inflation caps/floors. As their nominal equivalents, inflation caps/floors are defined as the sum of caplets/floorlets on the inflation rate over a certain period. For example, the payoff of an inflation caplet with strike X is:

$$\max\left(\frac{I_T}{I_{T-1}} - 1 - X, 0\right).$$

These payoffs are also know as "year-on-year".

Inflation caps/floors and zero-coupon options are quoted in prices by investment banks and brokers, and prices have been available on Bloomberg since 2007. Prices are available for a wide range of strikes, commonly from -1% to 6%. Inflation caps/floors can be used as building blocks of smile calibration, as shown in chapters 4 and 5: this said, their lower liquidity (compared to zero-coupon trades) has to be taken into account.

1.1.4 Inflation exotics and hybrids

LPI (Limited Price Index)

In 1995 the UK Pensions Act required pension schemes to link payments to the RPI index capped at 5% and floored at 0%. This requirement has given rise to Limited Price Index securities (LPIs) that pay the following premium:

$$\max\left(\min\left(\frac{I_Y}{I_{Y-1}} - 1,\ \text{cap}\right),\ \text{floor}\right).$$

Pricing this inflation call spread requires a model to deal with the option market skew.

These securities are widely traded in the UK market only, and can be packaged as swaps where one party pays RPI and the other pays LPI plus a spread.

Inflation range accruals. Inflation range accruals pay either a fixed or a floating coupon depending on the number of periods in which inflation trades within a given range. They are effectively as a strip of digital options on the inflation index, therefore requiring a model to capture the skew. For example, a fixed coupon range accrual pays a strip of coupons defined as:

$$\text{Fixed Coupon} \times \frac{n}{12}.$$

Here n is the number of months where:

$$\text{Lower Bound} < \frac{I_m}{I_{m-12}} - 1 < \text{Upper Bound}.$$

Inflation spread options. Inflation spread options pay the difference between the appreciation of two different price indices, capped or floored at X. For example a spread cap pays:

$$\max\left(\frac{I_T^1}{I_0^1} - \frac{I_T^2}{I_0^2} - X, 0\right).$$

This product may be used to hedge the basis risk between two price indices and requires modelling their joint dynamics. For example, a bank that is short Belgian inflation via a client trade and that hedges it with European CPI swaps may need to cover its basis risk buying such spread option from a hedge fund.

Real rate derivatives. The real rate r_t is sometimes loosely defined as the difference between the nominal rate n_t and the inflation rate p_t. This definition is in fact not quite correct (see next section). Nevertheless, investment banks have tailored some client solutions by selling structures with a general payoff of the form $f\left(g\left(n_t\right) - h\left(p_t\right)\right)$. Here $f(x)$, $g(x)$, and $h(x)$ are generic real scalar functions of a real scalar variable. For example, this could be a strip of caplets/floorlets having as an underlying a leveraged difference between the nominal rate and the inflation rate. Another interesting example is the so-called "real-Bermudan", which is a callable swap where one party pays nominal rate (Libor, for example) and receives inflation.

Path dependent contracts. Investment banks have structured path dependent inflation trades for clients concerned about deflation risk. These trades have as an underlying a function of the price index, such that if the price index decreases in any period the underlying does not change. It is the same as embedding a ratchet strike in the structure, making these trades very similar to the cliquets equity trades. Therefore a forward volatility skew model is needed to correctly price these structures, as explained in Gatheral [58]. More recently the Republic of Italy has issued a path-dependent inflation bond ("BTP Italia").

Inflation equity/FX hybrids. Inflation hybrids often involve equity or FX derivatives payoffs with a condition depending on the evolution of a pre-specified price index. The simplest structure is an equity payoff (the performance of the stock S_t) floored by inflation, whose payoff at maturity is:

$$\max\left(\frac{S_T}{S_0} - 1, \frac{I_T}{I_0} - 1\right).$$

These hybrid trades are difficult to price since it is necessary to mark the correlation between the main underlying and the inflation index. Furthermore, the model used to simulate the equity paths is

conceptually different from the inflation model (for example a local volatility model can be used for S_t): therefore this model inconsistency has to be addressed by making some assumptions and simplifications. Brigo & Mercurio [22] show a possible pricing strategy involving an *ad hoc* measure change.

Inflation/IR hybrids. A popular payoff is a nominal interest rate caplet/floorlet (here $f(T,S)$ denotes the Libor rate between times T and S) whose strike is related to the realised inflation rate leveraged by a factor L:

$$\max\left(f(T-1,T)+X-L\left(\frac{I_T}{I_T-1}-1\right),0\right).$$

Dodgson & Kainth [50] show how to obtain approximated closed formulas for this payoff in the Hull-White model, introduced later in this chapter.

1.2 Arbitrage pricing

1.2.1 Main facts

Here we provide a short summary of the arbitrage pricing theory in continuous time with a special focus on fixed income securities. From an historical perspective, the seminal papers are Harrison & Kreps [64] and Harrison & Pliska [65], that appeared in 1979 and 1981 respectively: one of the most important ideas in these papers is the link between the economic concept of no-arbitrage and the existence of a specific probability measure. This generalises the results found by Black & Scholes [14] in 1973. The reader interested in more details can refer to Brigo & Mercurio [22] or to Björk [33]; following closely the former source, the theory is built as follows:

1. Assuming continuous time (indexed by a positive real number) one considers a **probability space** $(\Omega,\mathcal{F},\mathbb{P})$ equipped by a right-continuous filtration $\{\mathcal{F}_t\}_{t\geq 0}$. It is assumed that the economy starts at $t=0$.

2. The **economy** is defined by a $K+1$-dimensional positive process $\{S(t)\}_{t\geq 0}$, whose first component $S_0(t)$ is the bank account process. The **bank account** dynamics are described by $dS_0(t) = S_0(t)n(t)dt$, with initial condition $S_0(0)=1$. Here $n(t)$ is the short-term nominal rate: it should be noted that the bank account can also be referred to using the notation $B(t)$. The remaining assets are non-dividend paying: this assumption can be relaxed, however this is not relevant here as we are interested to price fixed-income products.

3. **Discounting** a time t a payment of one unit of currency taking place at future time T is done via the discount factor $D(t,T)$, defined by $D(t,T)=B(t)/B(T)$. Clearly $D(0,t)=1/B(t)=1/S_0(t)$.

4. A **trading strategy** is a $K+1$-dimensional process $\{\phi(t)\}_{t\geq 0}$ that is locally bounded and predictable. This process represents the weights (both positive and negative) of the **trading strategy** whose **value** $V_\phi(t)$ is defined by: $V_\phi(t) = S(t)\cdot\phi(t) = \sum_{i=0}^{K} S_i(t)\phi_i(t)$. The associated **gains** $G_\phi(t)$ are defined by: $G_\phi(t) = \int_0^t dS(u)\cdot\phi(u) = \sum_{i=0}^{K}\int_0^t dS_i(u)\phi_i(u)$. Clearly we have $G_\phi(0) = 0$.

   The predictability request for the process $\{\phi(t)\}_{t\geq 0}$ is needed to impose that the trading position is established immediately before time t, to prevent the trader from adjusting the portfolio in case of a jump: this request can be ignored if randomness is introduced via a continuous process like a Brownian motion.

   To rule out doubling strategies, one requests also that a trading strategy must be bounded from below, i.e. $V_\phi(t) > -c\ \forall t$, where c is a positive real constant.

5. A trading strategy $V_\phi(t)$ is **self-financing** if $V_\phi(t) = G_\phi(t) + V_\phi(0)$. The meaning of this definition is that the value of the trading strategy is only driven by the positions $\phi(t)$ taken by the trader and by the evolution of the securities prices $S(t)$. If one considers the value of the **discounted trading strategy** $V_\phi(t)D(0,t)$, one finds that its value, if the strategy $V_\phi(t)$ is self-financing, is $V_\phi(t)D(0,t) = V_\phi(0) + \int_0^t d[D(0,u)S(u)]\cdot\phi(u)$.

6. An **arbitrage** is defined as a self-financing strategy $\hat{\phi}$ such that $V_{\hat{\phi}}(0) = 0$, $\mathbb{P}(V_{\hat{\phi}}(T) \geq 0) = 1$, and $\mathbb{P}(V_{\hat{\phi}}(T) > 0) > 0$. This means starting a trading strategy with no money ($V_{\hat{\phi}}(0) = 0$), having some positive probability to make some strictly positive gains ($\mathbb{P}(V_{\hat{\phi}}(T) > 0) > 0$), while being certain not to incur any losses ($\mathbb{P}(V_{\hat{\phi}}(T) \geq 0) = 1$).

7. Arbitrage theory is about finding conditions under which arbitrages can be ruled out. The reason for ruling out arbitrage is that if there is arbitrage, two rational parties cannot agree on a price, because one party is incurring in a loss or in no gain with probability one. No rational agent would close such trade. [5] The main result in arbitrage theory is the **first fundamental theorem**, that states that there is no arbitrage if there exists an equivalent martingale measure.

8. An **equivalent martingale measure** $\mathbb{Q}$ is a probability measure on the space $(\Omega, \mathcal{F})$ with the following properties:

   (a) The measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, which means that, $\forall A \in \mathcal{F}$, $\mathbb{P}(A) = 0 \Longleftrightarrow \mathbb{Q}(A) = 0$. The two measures may attribute different probabilities to some events, but share the same set of zero-probability events.

[5] Standard economic theory assumes that all agents are rational (i.e. they maximise their expected utility, and gaining money creates positive utility) and that all agents have full information.

(b) The Radon-Nikodym derivative $d\mathbb{Q}/d\mathbb{P}$ belongs to $L^2(\Omega, \mathcal{F}, \mathbb{P})$, i.e. it is square-integrable with respect to $\mathbb{P}$. This means that $\int_\Omega X[d\mathbb{Q}/d\mathbb{P}]^2 d\mathbb{P}$ always exists and is a finite quantity.

(c) The discounted asset price process $D(0,t)S(t)$ is a $\mathbb{Q}$-martingale, i.e. $D(0,t)S(t) = \mathbb{E}_t^{\mathbb{Q}}[D(0,T)S(T)]$, where $0 \leqslant t \leqslant T$. Here $\mathbb{E}_t^{\mathbb{Q}}[\cdot]$ is the expectation taken under the probability measure $\mathbb{Q}$.

9. A **contingent claim** H is a square-integrable random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The claim is **attainable** if there exists a self-financing strategy ϕ_H such that $V_{\phi_H}(T) = H(T)$, $\forall \omega \in \Omega$. If this holds, the no-arbitrage principle requires that the value of the strategy ϕ_H and the price of the claim H at time t, defined as $h(t)$, are the same $\forall t \leqslant T$. This results reads: $V_{\phi_H}(t) = h(t)$.

10. If there exists an equivalent martingale measure $\mathbb{Q}$ we write that the unique price of the attainable contingent claim at time t is $h(t) = \mathbb{E}_t^{\mathbb{Q}}[D(t,T)H(T)]$.

11. If every contingent claim is attainable, the market is **complete**. This means that every derivative can be replicated via a trading strategy that builds a replicating portfolio using the fundamental securities. The **second fundamental theorem** states that the market is complete if and only if the equivalent martingale measure is unique. Arbitrage can be ruled out in an incomplete market too, as it happens for example in fixed income markets, where the fundamental securities (like the short rate) are not traded assets. The concept of market price of risk comes into play to price derivatives in a consistent way in an incomplete market.

The no-arbitrage paradigm sketched above can be further extended to use different pricing measures that make the calculations simpler. In fact, the risk-neutral pricing measure $\mathbb{Q}$ may not be the best tool to solve all pricing problems across different payoffs. Geman at al. [59] have extended the theory introducing the concept of numeraire.

A **numeraire** is any strictly positive non-dividend paying asset, that is used to rescale other assets to martingales: this yields arbitrage-free dynamics under a different measure. It should be noted that the numeraire technique is an extension of the first fundamental theorem (in that case the numeraire is the bank account $B(t)$). Choosing a different numeraire yields a different pricing measure.

Because the arbitrage-free price is unique under the pricing measure, to avoid re-introducing arbitrage one must impose that the price of any contingent claim is independent from the measure that one chooses. This immediately yields the measure change process value at time t: in fact, assuming that there are two different pricing measures, $\mathbb{M}$ and $\mathbb{N}$, each with numeraire $M(t)$ and $N(t)$ respectively, the Radon-Nikodym derivative at time t is: $d\mathbb{M}/d\mathbb{N}|_{\mathcal{F}_t} = [M(t)N(0)]/[M(0)N(t)]$. This rule is an essential tool in many practical calculations.

If randomness is modelled via a Brownian motion $W(t)$, as it commonly happens in financial mathematics, there are even more powerful tools. For simplicity we assume that the Brownian motion is scalar (Brigo & Mercurio [22] show a more general result for multidimensional processes). One defines the dynamics of some quantity using a stochastic differential equation (SDE) in the form: $dX(t) = \mu_X^{\mathbb{S}}(t)dt + \sigma_X(t)dW^{\mathbb{S}}(t)$, where $\mu_X^{\mathbb{S}}(t)$ is the drift (assumed deterministic), $\sigma_X(t)$ the diffusion term (assumed deterministic), and, crucially, $W^{\mathbb{S}}(t)$ is a Brownian motion under the measure $\mathbb{S}$, with numeraire $S(t)$. Let us assume that we want to change the measure from $\mathbb{S}$ to $\mathbb{U}$, which is another measure whose numeraire is $U(t)$. If we know the diffusion terms of the two numeraire processes, i.e. we know that they are modelled via the SDEs $dS(t) = (...)dt + \sigma_S(t)dW^{\mathbb{S}}(t)$ and $dU = (...)dt + \sigma_U(t)dW^{\mathbb{S}}(t)$ respectively, we know that $dW^{\mathbb{U}}(t) = dW^{\mathbb{S}}(t) - [\sigma_U(t)/U(t) - \sigma_S(t)/S(t)]dt$.

The above result can be regarded as a generalisation of the Girsanov theorem, where one writes $dW^{\mathbb{Q}}(t) = dW^{\mathbb{P}}(t) - [0/B(t) + \lambda(t)\psi(t)/\psi(t)]dt$. In fact, the numeraire associated with the risk-neutral measure $\mathbb{Q}$ is the Bank account, with dynamics $dB(t) = B(t)[n(t)dt + 0dW(t)]$, and the numeraire of the real-world measure $\mathbb{P}$ is the inverse pricing kernel $\psi(t)$, with dynamics $d\psi(t) = \psi(t)[-n(t)dt - \lambda(t)dW(t)]$. The numeraire dynamics are obtained as $d(1/\psi(t)) = (...)dt - \lambda(t)\psi(t)(-1/\psi(t))^2 dW(t)$. By doing the calculations one gets the usual result $dW^{\mathbb{Q}}(t) = dW^{\mathbb{P}}(t) - \lambda(t)dt$. See Hughston [69] for more details of pricing kernel dynamics.

Finally one notes that arbitrage-free pricing of the contingent claim $H(T)$ can be carried out under the measure $\mathbb{P}$ by taking the following expectations: $h(t) = \mathbb{E}_t^{\mathbb{Q}}[H(T)D(t,T)] = \mathbb{E}_t^{\mathbb{P}}[H(T)D(t,T)d\mathbb{Q}/d\mathbb{P}] = \mathbb{E}_t^{\mathbb{P}}[H(T)\psi(T)/\psi(t)]$, where $\psi(t) = D(0,t)d\mathbb{Q}/d\mathbb{P}|_{\mathcal{F}_t}$. This is the idea underlying the pricing kernel, whose properties in discrete time are reviewed in the following section. We focus on the discrete-time case as the model proposed in chapter 2 is set in discrete time. Further, given that we will start from a real-world macroeconomic model, the use of a measure change to move to a pricing measure is an essential step, thus making the pricing kernel an essential tool.

1.2.2 Pricing kernels in discrete time

Some authors take the view that arbitrage-freeness is embodied in assuming the existence of a pricing kernel process (see for example Cochrane [42], Duffie [52] or Björk [33]). The pricing kernel is used in the next section to build a discrete-time equivalent of the continuous-time Black-Scholes dynamics, therefore it can be useful to list its main properties. Further, pricing kernel techniques are used in Hughston & Macrina [71] to build an inflation model based on economic theory, which we present and the end of this chapter and from which we take inspiration to build a discrete time inflation model based on a macroeconomic model in chapter 2.

In the following sections we assume that time is discrete, and that we observe the variables in times $t_0, t_1, \ldots$. To make notation lighter we write the value of variable x at time t_i either x_{t_i} or x_i. In discrete time the pricing kernel process $\{\psi_i\}_{i=0,1,\ldots}$ has the following properties:

1. The process $\{\psi_i\}_{i=0,1,\ldots}$ is a strictly positive process, with $\psi_0 = 1$.[6]

2. The pricing kernel at time t_i is defined as $\psi_i = \prod_{j=1}^{i} (1+\tau_j n_j)^{-1} \mu_i$, where μ_i is the Radon-Nikodym derivative $(d\mathbb{Q}/d\mathbb{P})\,|_{t_i}$. Equivalently it is possible to write: $\psi_i = \mu_i / B_i$[7] if one wants to involve the bank account B_i.

3. For each non-dividend paying security h paying the single cash flow H_N at time T_N, we have: $h_i \psi_i = \mathbb{E}^{\mathbb{P}}[\psi_N H_N | \mathcal{F}_i]$, i.e. $h_i \psi_i$ is a $\mathbb{P}$-martingale.

4. In case of an asset paying the dividend stream D_i the above relationship becomes $h_i \psi_i + \sum_{j=0}^{i} \psi_j D_j = \mathbb{E}^{\mathbb{P}}[\psi_N H_N + \sum_{j=i+1}^{N} \psi_j D_j | \mathcal{F}_i]$.

5. The pricing kernel ψ_i is the inverse of the numeraire chosen to rescale the asset processes to $\mathbb{P}$-martingales.

6. The pricing kernel is related to nominal bond prices via the following: $P(t_i, t_{i+k}) = \mathbb{E}_i^{\mathbb{P}}[\psi_{i+k}/\psi_i]$, $\forall k \in \mathbb{N}$.

7. The pricing kernel is related to the index bond via the following: $P^I(t_i, t_M) = \mathbb{E}_i^{\mathbb{P}}\left[\psi_M I_M\right] / \psi_i$.

8. The pricing kernel is sometimes described as a potential (i.e. a positive supermartingale), as explained in Rogers [104]. The same comment on positivity of interest rates made above should be stressed here.

9. The floating rate note $N(t_0, t_j)$ – which pays in each period from t_0 to t_j the floating interest rate n_i – can be written as a function of the pricing kernel as: $N(t_0, t_j) = \psi_j + \sum_{i=1}^{j} \psi_i n_i$.

Further properties can be found in Hughston [69], Hughston & Rafailidis [73] or Björk [33].

To develop intuition, we note that the first two properties listed above make the pricing kernel a positive process since it is the product of two positive processes (the discount factor and the Radon-Nikodym derivative). The third property is the most important: the pricing kernel allows one to price derivatives

[6] Some authors, including Hughston [69], require the pricing kernel to be a supermartingale. This request is related to the positivity of interest rates. However, after the Lehman crisis there have been some instances where nominal interest rates have been negative: the financial meaning is that market participant become so risk-adverse that they prefer to be charged a negative interest rate by banks instead of investing their money in other assets. German government yields have been negative in 2012 (up to 2 years maturity), and the ECB president Mario Draghi has also hinted at negative policy rates in 2013. Euro deposit rates have been cut to negative levels in June 2014.

[7] Assuming the existence of the pricing kernel makes irrelevant whether or not the market is complete, given that the existence of the pricing kernel implies the existence of the risk-neutral martingale measure used to uniquely price contingent claims.

by using the real-world probability measure $\mathbb{P}$ in a no-arbitrage framework, because it "contains" the measure change μ_i. The generalisation of these concepts to continuous time is straightforward.

1.2.3 Discrete-time asset pricing in the lognormal case

Shreve [109] shows how the Girsanov theorem can be adapted to the discrete-time case: an exponential martingale is built using a sufficiently regular market price of risk process and a zero-mean Gaussian process $\{V_i\}_{i=0,1,\ldots}$ (defined by: $V_i \equiv \sum_{k=1}^{i} \lambda_k N_k$, where the N_k here are independent standard Gaussian random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$, and $\{\lambda_k\}_{k=0,1,\ldots}$ is a $\{\mathcal{F}_{k-1}\}$-adapted process). Let us introduce the risk-neutral measure $\mathbb{Q}$.

Setting the measure change process $\mu_i = \mu_0 e^{-\sum_{k=1}^{i} \lambda_k N_k - \frac{1}{2}\sum_{k=1}^{i} \lambda_k^2}$, where $\mu_i = (d\mathbb{Q}/d\mathbb{P}) \mid_{t_i}$ and $\mu_0 = 1$, one shows that the process $\{Y_i\}_{i=0,1,\ldots}$ defined as $Y_i = N_i + \lambda_i$ is a zero-mean Gaussian process under the measure $\mathbb{Q}$ and the variables Y_i are independent from each others. This result is used in chapter 2.

With this result in mind, we can show that the pricing kernel lets one write some discrete-time asset dynamics that recall the more familiar continuous-time dynamics used for example by Black & Scholes [14]. To simplify, we assume that securities do not pay any dividend, shocks are normally distributed and the measure change process is the discretely sampled exponential Brownian martingale $\{\mu_i\}_{i=0,1,\ldots}$: some of these assumptions can be relaxed if needed.

As stated in the previous sections, absence of arbitrage implies that the product of the pricing kernel and the generic asset S_i is a $\mathbb{P}$-martingale

$$\psi_i S_i = \mathbb{E}_i \left[\psi_{i+j} S_{i+j}\right] \quad j \geq 1.$$

We use the notation $\{M_i\}_{i=0,1,\ldots}$ for a generic martingale, which we request to be positive since we are dealing with financial assets. Therefore we write $M_i = \psi_i S_i$. This may be rewritten as the product between another positive martingale $\mathcal{M}_i$ and a generic growth factor G_i; we aim to represent each asset as $S_i = \mathcal{M}_i G_i$. For simplicity we assume that times are equally spaced and that the time lag is 1.

To show this, we make the assumption that the measure change process $\{\mu_i\}_{i=0,1,\ldots}$ and the martingale $\{M_i\}_{i=0,1,\ldots}$ are both exponential martingales generated by the Gaussian variables N_i:

$$M_i = M_0 e^{\sum_{k=1}^{i} \theta_k N_k - \frac{1}{2}\sum_{k=1}^{i} \theta_k^2}$$

$$\mu_i = \mu_0 e^{-\sum_{k=1}^{i} \lambda_k N_k - \frac{1}{2}\sum_{k=1}^{i} \lambda_k^2}.$$

Both processes $\{\mu_i\}_{i=0,1,\ldots}$ and $\{M_i\}_{i=0,1,\ldots}$ are valued 1 at inception.

Their ratio can be written as

$$\frac{M_i}{\mu_i} = \frac{M_0}{\mu_0} e^{\sum_{k=1}^{i}(\theta_k+\lambda_k)N_k - \frac{1}{2}\sum_{k=1}^{i}(\theta_k^2-\lambda_k^2)}.$$

Some algebraic manipulations (adding and subtracting $2\theta_k\lambda_k$ and $2\lambda_k^2$ in the exponent) yield

$$\frac{M_i}{\mu_i} = \frac{M_0}{\mu_0} e^{\sum_{k=1}^{i}(\theta_k+\lambda_k)N_k - \frac{1}{2}\sum_{k=1}^{i}(\theta_k+\lambda_k)^2 + \sum_{k=1}^{i}(\theta_k+\lambda_k)\lambda_k}.$$

By defining:

$$\mathcal{M}_i = \frac{M_0}{\mu_0} e^{\sum_{k=1}^{i}(\theta_k+\lambda_k)N_k - \frac{1}{2}\sum_{k=1}^{i}(\theta_k+\lambda_k)^2}$$

$$G_i = G_0 e^{\sum_{k=1}^{i}(n_k+(\theta_k+\lambda_k)\lambda_k)}$$

$$\sigma_k = \theta_k + \lambda_k$$

$$S_0 = \frac{M_0 G_0}{\mu_0}$$

we retrieve the desired formula for the asset price

$$S_i = \mathcal{M}_i G_i = S_0 e^{\sum_{k=1}^{i}\sigma_k N_k - \frac{1}{2}\sum_{k=1}^{i}\sigma_k^2 + \sum_{k=1}^{i}(n_k+\sigma_k\lambda_k)}. \tag{1.1}$$

If the interest rate, volatility and market price of risk are constant and deterministic we obtain the discrete-time Black-Scholes equivalent dynamics.

1.2.4 Model-independent inflation relationships

The payoffs presented in the previous sections are related by some model-independent relationships which are extremely useful when calibrating a model. Useful references are Brigo & Mercurio [22] and Jäckel & Bonneton [77]. Here $\mathbb{E}_t^{\mathbb{Q}}[X]$ means the expected value of X under $\mathbb{Q}$ conditional to the information available at time t, where $\mathbb{Q}$ is the risk-neutral measure. To stress that the results here are independent on whether one works in continuous or in discrete time, we use the generic notation x_t in this section. For simplicity, in this section we assume that all inflation payments are not time-lagged.

Zero-coupon inflation bond and zero-coupon inflation swap. In a market without liquidity concerns the inflation breakevens implied by the inflation bonds are the same as the ones driving inflation swap market. The zero-coupon inflation bond price is $P^I(t,T) = \mathbb{E}_t^{\mathbb{Q}}[I_T B_t / B_T]$, where B_T is the money market account numeraire. In an inflation swap the strike X is such that the expected value at maturity of the swap is zero: $\mathbb{E}_t^{\mathbb{Q}}[(I_T/I_t - (1+X_{t,T})^{T-t})B_t/B_T)] = 0$. If a measure change is applied from

the risk-neutral $\mathbb{Q}$ to the T-forward measure $\mathbb{Q}^T$, it is possible to rewrite the previous relationships as $P^I(t,T) = P(t,T)\mathbb{E}_t^{\mathbb{Q}^T}[I_T]$ and $P(t,T)\mathbb{E}_t^{\mathbb{Q}^T}[I_T/I_t - (1+X_{t,T})^{T-t})] = 0$ respectively. This point shows that both payoffs depend on the same expectation $\mathbb{E}_t^{\mathbb{Q}^T}[I_T]$, given that all other quantities are known at time t.

Inflation bond and real bond. The zero-coupon real bond – which is not a traded asset – is defined as $P^R(t,T) = P^I(t,T)/I_t$. The real bond is a bond that pays interest in goods and services, whereas the inflation linked bond pays the cash equivalent of the appreciation of the price index. The real economy will be defined fully in the end of this chapter.

Nominal bonds, real bonds and zero-coupon inflation swaps. The relationship

$$PV_{t,T}^{ZC} = \mathbb{E}_t^{\mathbb{Q}}[(I_T/I_t - (1+X_{t,T})^{T-t})B_t/B_T)] = 0$$

can be rewritten as $\mathbb{E}_t^{\mathbb{Q}}[(I_T/I_t)(B_t/B_T)] = \mathbb{E}_t^{\mathbb{Q}}[(1+X_{t,T})^{T-t})B_t/B_T)]$. Here $PV_{t,T}^{ZC}$ means the present value of the at-the-money zero-coupon swap at time t with maturity T and strike $X_{t,T}$. The left-hand side term is further simplified into:

$$\mathbb{E}_t^{\mathbb{Q}}[(I_T/I_t)(B_t/B_T)] = \mathbb{E}_t^{\mathbb{Q}}[(I_T)(B_t/B_T)]/I_t = P^I(t,T)/I_t = P^R(t,T).$$

The right-hand side is simplified into:

$$\mathbb{E}_t^{\mathbb{Q}}[(1+X_{t,T})^{T-t})B_t/B_T)] = (1+X_{t,T})^{T-t}\mathbb{E}_t^{\mathbb{Q}}[B_t/B_T] = (1+X_{t,T})^{T-t}P(t,T).$$

This finally yields the relationship: $(1+X_{t,T})^{T-t} = P^R(t,T)/P(t,T)$. This may be further rewritten in a more expressive way by subtracting 1 on both sides: $(1+X_{t,T})^{T-t} - 1 = P^R(t,T)/P(t,T) - 1$. This relationship means that the present value of the inflation swap fixed payment between times t and T (which is $P(t,T)((1+X_{t,T})^{T-t} - 1)$) is equivalent the difference between the real and the nominal bond prices at time t for maturity T: we can write $P^R(t,T) - P(t,T) = P(t,T)[(1+X_{t,T})^{T-t} - 1]$. One finds in the market quotes for $P(t,T)$ and for $X_{t,T}$, and therefore the quantity $P^R(t,T)$ can be deduced.

Forward price index. The forward price index definition follows the previous result $I^*(t,T) = (1+X_{t,T})^{T-t} = P^R(t,T)/P(t,T)$. This is the expected growth of the price index between times t and T under the T-forward measure: $\mathbb{E}_t^{\mathbb{Q}^T}[I_T] = I_t I^*(t,T)$.

Year-on-year inflation swap. We can write the following relationship involving the year-on-year inflation swap strike $X_{t,T-1,T}^{YoY}$: $PV_{t,T-1,T}^{YoY} = \mathbb{E}_t^{\mathbb{Q}}[(I_T/I_{T-1} - (1+X_{t,T-1,T}^{YoY}))B_t/B_T]$, with $t < T-1 < T$. This is rewritten by highlighting a zero-coupon inflation swap between times $T-1$ and T inside the expectation using the tower law: $PV_{t,T-1,T}^{YoY} = \mathbb{E}_t^{\mathbb{Q}}[B_t/B_{T-1}\mathbb{E}_{T-1}^{\mathbb{Q}}(I_T/I_{T-1} - (1+X_{t,T-1,T}^{YoY}))B_{T-1}/B_T]$.

By making use of the relationship between the present value on a zero-coupon swap, the nominal and the real bonds, it is possible to write $PV^{YoY}_{t,T-1,T} = \mathbb{E}^{\mathbb{Q}}_t[B_t/B_{T-1}(P^R(T-1,T)-P(T-1,T))]$, further simplified into $PV^{YoY}_{t,T-1,T}= \mathbb{E}^{\mathbb{Q}}_t[B_t/B_{T-1}(P^R(T-1,T))] -P(t,T)$. This formulation makes explicit that the price of a year-on-year swap depends on the covariance of nominal interest rates and inflation, which usually takes the form of a model-dependent convexity adjustment if nominal rates are stochastic. In the T-forward measure the previous relationship becomes $PV^{YoY}_{t,T-1,T} = P(t,T)\mathbb{E}^{\mathbb{Q}^T}_t[(I_T/I_{T-1} - (1 + X^{YoY}_{t,T-1,T})]$.

Inflation options. The payoff of an inflation call is:

$$\mathbb{E}^{\mathbb{Q}}_t[B_t/B_T \max(I_T/I_t - X, 0)] = P(t,T)\mathbb{E}^{\mathbb{Q}^T}_t[\max(I_T/I_t - X, 0)].$$

Its at-the-money forward is given by the zero-coupon inflation swap strike at maturity T.

Inflation caps/floors. The payoff of an inflation caplet is:

$$\mathbb{E}^{\mathbb{Q}}_t[B_t/B_T \max(I_T/I_{T-1} - X, 0)] = P(t,T)\mathbb{E}^{\mathbb{Q}^T}_t[\max(I_T/I_{T-1} - X, 0)].$$

Its at-the-money forward is given by the year-on-year swap strike between times $T-1$ and T.

1.3 Inflation pricing models

In this section we review the main inflation pricing models and frameworks proposed in literature and used in the industry.

1.3.1 The foreign-exchange analogy

The foreign-exchange (Forex) analogy, presented in Hughston [70], has been perhaps the main theoretical tool developed so far to price inflation derivatives. Given an economy where both interest rates and inflation-linked instruments are quoted, two interest rates system are introduced: the real system (i.e., "foreign"), and the nominal system (i.e., "domestic"). The price level (e.g., the consumer price index) plays the role of the foreign-exchange rate: therefore the problem of pricing an inflation derivative boils down to pricing a cross-currency interest rates payoff.

We assume the existence of a nominal short rate n_t and a real short rate r_t. We introduce a nominal bond with price $P(0,T)$ that pays a unit notional at maturity T in the domestic (nominal) currency, and a real bond with price $P^R(0,T)$ that pays a unit notional in the foreign (real) currency, i.e. in goods and services. To clarify, the nominal bond pays a certain nominal amount but a random real amount, as the purchasing power at maturity is uncertain. Conversely, a real bond pays a certain real amount (i.e.

goods and services) but has a random nominal value, since the future price of the goods and services paid by the real bond is uncertain. Hughston [70] derives the continuous-time no-arbitrage dynamics of the price index and the bonds (both nominal and real) with standard stochastic calculus arguments. The result is essentially a general HJM-type theory for the inflation rate along with the nominal and real interest rates systems.

The Forex analogy has roots in the macroeconomic theory: the Fisher equation (see Fisher [55]) defines the inflation rate as the difference between the nominal and the real rates plus some correction terms. Further details can be found for example in Dornbusch & Fischer [51].

1.3.2 The Jarrow-Yildirim model

The Jarrow-Yildirim model, presented in [79] and explained further in [22], has been one of the first models proposed to price inflation derivatives and has become the benchmark approach thanks to its mathematical tractability. The Jarrow-Yildirim approach is based on the Forex analogy: therefore there exist two interest rates, nominal and real, modelled by a pair of correlated Hull-White short-rate processes. The price index is shown to be a lognormal process, the drift of which is given by the difference between the nominal and the real rate in the risk-neutral measure.

A third Brownian source of randomness can be added in the equation for the price index level, allowing more flexibility for calibration. The Jarrow-Yildirim approach exploits advantages of the Hull-White model, namely the fact that it automatically calibrates to the term structure and can be implemented both in a tree and in Monte Carlo. What's more, caplet prices are derived from a model-independent relationship involving bond option prices: and the Hull-White model also provides closed form expressions for the latter. Further, the year-on-year convexity adjustment is calculated easily in this model (see Brigo & Mercurio [23]).

From a practitioner perspective, a disadvantage of models based on the Forex analogy is the somewhat unrealistic idea of modelling a real interest rates system: real rates are not quoted as such in the market and therefore marking their volatility is not an altogether straightforward matter. Similar observations can be made concerning the correlation parameters of such models. Both real volatilities and nominal-real correlations can be numerically calibrated against traded derivatives such as inflation caps/floors. There is no clear recipe regarding how to split the volatility implied by inflation option prices into the Jarrow-Yildirim model volatilities for the real rate and the price index.

Brody, Crosby & Li [29] propose a multi-dimensional extension of this model and derive some closed forms for convexity adjustments and LPI prices.

1.3.3 The RBS model

Dodgson & Kainth [50] describe a two-processes Hull-White model, where the state variables are the nominal interest rate and inflation: the real rate is not modelled, making this model easier to calibrate to observed market data. Both state variables are modelled with Ornstein-Uhlenbeck processes with time-varying mean reversion level: this model automatically fits the initial term structure and the inflation swaps curve. Closed forms for year-on-year inflation swaps, inflation options and inflation caps/floors are obtained.

1.3.4 The HJM approach

The HJM methodology has been used to price inflation-linked derivatives as a natural extension of the single-factor models like the Jarrow-Yildirim or the one presented in the RBS document. For example Leung & Wu [87] show how to derive closed forms for the most common inflation-linked payoff in the HJM framework.

1.3.5 The BGM-I approach

The use of the BGM methodology to model inflation derivatives has been proposed by Mercurio [95] and follows the path of the well-known BGM model for nominal interest rates: the dynamics of the nominal and real forward Libor rates are assumed to be lognormal martingales under their respective T-forward measures. The dynamics of the real rates include a quanto adjustment, following the Forex analogy. The drift is adjusted to translate the dynamics of all nominal forward Libor rates into the common terminal measure. This model automatically calibrates to the nominal and real term structure. Mercurio proposes an approximated semi-closed form for the year-on-year inflation index swap. This extension of the BGM model captures a richer set of curve dynamics but brings together all the problems associated to this model, i.e. how to model instantaneous volatilities and correlations or whether to calibrate the terminal correlations or not. An analysis of these problems can be found in Brigo & Mercurio [22].

1.3.6 Stochastic volatility approaches

Stochastic volatility allows flexibility to capture fat tails and skewness of the market-implied distribution. Mercurio & Moreni [96] propose to model the forward inflation as a geometric Brownian motion where the diffusion coefficient follows a CIR square root process. They obtain approximated prices for inflation caplet/floorlet by making use of Fourier inversion techniques, similar to the general Heston case: in addition, they show some calibration results. Kenyon [81] considers many possible alternatives to model fat tails and skew, including Gaussian models and Gaussian models with stochastic volatility. He fits

all these models to market data and shows the calibration results. Andersen [3] uses time-changed Lévy processes to propose a stochastic volatility version of the Jarrow-Yildirim model, and shows a closed form solution for the inflation floorlet.

1.3.7 The Hughston-Macrina (HM) model

Hughston & Macrina [71] propose an original framework to price inflation derivatives: starting from the Forex analogy, they assume the existence of a nominal and a real pricing kernel and use a microeconomic approach based on the convenience yield offered by the money supply to determine the continuous-time dynamics of the price index I_t.

Absence of arbitrage is ensured by the existence of the nominal and real pricing kernels. The price index is defined as the ratio of the real and nominal pricing kernels, which are denoted ψ_t and ψ_t^R respectively in what follows. We remind the reader that $\psi_t^R = \psi_t I_t$. The expressions for these kernels are derived by assuming that market agents maximize a bivariate utility function of a type similar to that proposed by Sidrauski [110] when deciding how much to consume and how much cash to hold. The use of such utility function to model the liquidity preference is sometimes referred to in the literature as an MIU (money-in-utility) approach (see Walsh [115]).

In more detail, in the HM approach the authors introduce the real rate of consumption process $\{C_t\}_{t\geq 0}$ and the nominal money supply process $\{M_t\}_{t\geq 0}$, and define the real money supply process $\{l_t\}_{t\geq 0}$ by setting $l_t \equiv M_t/I_t$. The consumer problem is modelled by consideration of an inter-temporal expected utility maximization with a budget constraint with a finite time horizon. Utility is gained both from consumption and from the benefit of having cash in hand. The objective function is thus of the form:

$$U(\{C_t\},\{l_t\}) = \mathbb{E}\left[\int_0^T e^{-\gamma t} U(C_t, l_t)dt\right]$$

under the constraint:

$$W_0 = \mathbb{E}\left[\int_0^T \psi_t^R C_t dt + \int_0^T \psi_t^R l_t dt\right].$$

The utility function $U(x,y)$ is twice differentiable, has positive first derivatives and negative second derivatives: these are standard requests for utility functions. The real pricing kernel is used to model the intertemporal real discount factor. The intertemporal preference structure is expressed in the utility function by the "impatience penalty" $e^{-\gamma t}$. Expectations are taken with respect to the real-world measure. As $\psi_t^R = \psi_t I_t$ and $l_t \equiv M_t/I_t$, it is possible to rewrite the constraint (1.3.7) in the form:

$$W_0 = \mathbb{E}\left[\int_0^T \psi_t C_t I_t dt + \int_0^T \psi_t I_t \frac{M_t}{I_t} dt\right] = \mathbb{E}\left[\int_0^T \psi_t C_t I_t dt + \int_0^T \psi_t M_t dt\right].$$

This optimization problem is solved by calculus of variations techniques. In particular we set

$$F(t, x, y) = e^{-\gamma t} U(x, y)$$

and

$$G(t, x, y) = \psi_t x I_t + \psi_t y I_t.$$

The Lagrange function $L = F - \mu G$ is then optimized by setting to zero the first derivatives with respect to x and y. The parameter μ is the auxiliary parameter of the Lagrange function, and the maximisation leads to

$$\frac{\partial U}{\partial x} = \mu e^{\gamma t} \psi_t I_t, \qquad \frac{\partial U}{\partial y} = \mu e^{\gamma t} \psi_t I_t.$$

As an example, let us consider the case where the utility function is log-separable:

$$U(C_t, l_t) = a \log(C_t) + b \log(l_t)$$

where a and b are non-negative constants. The first order conditions are

$$\frac{a}{C_t} = \mu e^{\gamma t} \psi_t I_t$$

$$\frac{b}{l_t} = \mu e^{\gamma t} \psi_t I_t.$$

As the two conditions above are equal to the same quantity $\mu e^{\gamma t} \psi_t I_t$, we equate them. Remembering that $l_t = M_t / I_t$ one gets

$$I_t = \frac{a}{b} \frac{M_t}{C_t}.$$

It follows further that the nominal pricing kernel is given by

$$\psi_t = \frac{b e^{-\gamma t}}{\mu M_t} \tag{1.2}$$

and the real pricing kernel is given by

$$\psi_t^R = \frac{a e^{-\gamma t}}{\mu C_t}. \tag{1.3}$$

The above expressions for ψ_t and ψ_t^R guarantee arbitrage-freeness since they are derived from optimality conditions and therefore provide equilibrium levels for the state variables.[8] The explicit derivation of

[8] Intuitively, if there are arbitrage opportunities the representative agent can not find an optimal portfolio allocation, given that any portfolio can be further improved by adding the arbitrage portfolio.

nominal and real pricing kernels given by (1.2) and (1.3) provides us with an explicit model for the price index process $\{I_t\}_{t\geq 0}$ as we remember the equality $\psi_t^R = \psi_t / I_t$.

Interestingly, the expression of the nominal pricing kernel depends on the money supply and the real pricing kernel is a function of the consumption: this shows well the consistency of this approach.

This model has been further investigated by Alexander [2], who models the money supply and consumption processes as geometric Brownian motions to obtain closed forms for the price index process and for inflation options.

Hughston & Macrina [72] further develop this model in the context of information-based asset pricing, where the market filtration is generated by some information processes. In particular, they obtain closed forms for real and nominal bonds.

The Hughston-Macrina approach uses an economic model to price inflation derivatives and the use of the pricing kernel is pivotal. With this in mind, in the following chapters we develop a framework that incorporates a more sophisticated macroeconomic model; furthermore, we make an explicit assumption regarding the behaviour of the central bank, that sets the short-term nominal rate to finetune the economy. The role of the central bank is not made explicit in the Hughston-Macrina approach, as the money supply M_t is somewhat exogenous: by specifying more details on the monetary policy, it is possible to better explain the co-movements of nominal interest rates and inflation.

Chapter 2

Monetary macroeconomic inflation models

In this chapter we review a standard monetary macroeconomic inflation model in discrete time and propose a strategy to use it to price inflation derivatives: the main advantage of this approach is that the inflation dynamics are not taken exogenously but rather are the result of a well-established macroeconomic model. In particular, the co-movement of inflation and nominal interest rates is not taken as an input or modelled via a correlation process (as it happens in many models currently used in the industry) but is the result of central bank policy, via a well-known macroeconomic relationship (Taylor rule). This chapter should be regarded as a prelude to chapter 3: here we develop intuition in a somewhat loose way by studying and tweaking existing macroeconomic discrete-time models. We will formalise these intuitions in the following chapter by building a new continuous-time model.

The task is not straightforward because most macroeconomic literature is written in a somewhat less formalised way compared to financial mathematics. Expectations are often taken only with respect to the real-world econometric measure (which is known as $\mathbb{P}$ or as "physical measure" in financial mathematics): therefore there is no need to specify the measure with respect to which expectations are taken. Therefore measure changes are not widely used. No mention is made of filtrations, adapted processes, measurability. Distributional assumptions tend to be loose (randomness is usually introduced via some so-called "white noise", defined as a zero-mean process whose realisations are independent from each other over time). Stochastic processes tend to be assumed to reach a "steady state", i.e. to converge to some equilibrium value in the long run: this level is always supposed to exist and to be finite. Sometimes variables are expressed as their percentage deviations with respect to their long term equilibrium level: approximations and linearisations are very common. Securities payoffs may be defined with only a few details.

Despite these issues, this theory is the one that central bankers, economists, researchers, and market operators use and refer to: it can not be ignored. The challenges that one faces to use this theory in financial mathematics to carry out derivatives pricing are manyfold: the aim is to complement the macroeconomic model with all the mathematical machinery that has been originally taken as a given in a way that the kernel of the model is not arbitrarily changed but is rather enhanced by an improved formalism. Further, when one makes changes to the original model assumptions, these changes have to be not invasive and have to bear a clear advantage, especially in the calibration phase. At the same time, some approximations may be needed to derive some results that are essential for pricing (closed forms for nominal and inflation bonds, for example).

The chapter is structured as follows. Firstly, we build a general axiomatic framework around the original macroeconomic model: this entails specifying the feature of the time scale, probability space, and traded instruments. Then, for the benefit of the reader not expert in monetary macroeconomics, all the economic quantities and assumptions are listed and defined. Secondly, we introduce a standard monetary macroeconomics model (the DSGE model, or "Dynamic Stochastic General Equilibrium" model) where some $\mathbb{P}$-dynamics for inflation are derived from optimality conditions and realistic market frictions. Thirdly, we derive the expression for the nominal rate and inflation rate variance and higher order moments based on the DSGE model: they turn out to be linear combinations of the variances and the higher order moments of the random processes used originally in the DSGE model. The advantage is clear, as we choose these parameters to match the moments of the distribution implied by market-traded options on interest rates and inflation. Fourthly, we obtain approximated expressions for the nominal and inflation bonds, to calibrate the model to the observed nominal and inflation term structures.

To sum up, the first part of this chapter proposes a useful attempt to bridge the gap between monetary macroeconomics and financial mathematics. This said, at the end of the chapter we suggest how the framework can be somewhat translated into continuous time to improve its tractability and to take into account some very recent market features, like low interest rates and quantitative easing. Although there is no exact correspondence between the original discrete-time DSGE model and the newly-introduced continuous-time dynamics, the latter are clearly inspired by the former. To make this point more evident, some simulations run at the end of the chapter show that the newly-introduced continuous-time dynamics are broadly consistent both with empirical evidence and with the original DSGE discrete-time model.

2.1 Introduction to the DSGE model

DSGE models are an essential tool for the working macro-economist: they are widely used both in academia and by central banks since they explain the short-term real effects of monetary policy. There

is strong empirical evidence supporting the idea that money has real effects: DSGE models describe this effect by assuming a stochastic environment, optimizing behaviour and nominal rigidities in the economy. Consumers maximize their expected utility, which is based on consumption and real cash balances; firms maximize their expected profit stream but are not able to change in each period the prices they charge. The result is a discrete-time model where the macroeconomic variables are affected by their future expectations and some external shocks. The short-term nominal interest rate ("short rate") is part of these dynamics. A further assumption is that the central bank uses a Taylor rule to set the short rate: this means that the short rate is changed in response to the other macroeconomic variables using a simple linear rule (see Taylor [112]). This approach, albeit simple, has proven to be powerful to explain the central bank behaviour.

Finally, we stress that so far we referred to DSGE models in plural as they can be regarded as a family of models that share the main features listed above: consumer habits, capital, labour market rigidities, government, taxes, lagged variables, different central bank policies can be introduced in this framework, giving rise to more complex dynamics. In this section we describe the baseline version of this model, which offers enough flexibility for our purposes.

We present the assumptions of a basic version of the DSGE macroeconomic model, which explains the behaviour of the inflation rate p_i and the output gap x_i based on a general description of the economy. A complete description of this model can be found in this chapter or in Walsh [115], which we follow to present the model. Before presenting the macroeconomic model, we specify the axiomatic foundations that are implicit in the model and that are usually taken as a given by economists.

2.1.1 Axiomatic foundations

Time scale

The model is set in discrete time, where time is a non-negative variable:

$$t_i \in \mathbb{T} = \{t_0, t_1, t_2, ..., t_n, ...\}, n \in \mathbb{N}$$

where t_0 is the present time. To preserve generality, the discrete-time points are not required to be equally spaced. For a variable y at time t_i we often write y_i to make the notation lighter: similarly, the discrete-time stochastic processes $\{y_{t_i}\}_{i=0,1,...}$ can be denoted by $\{y_i\}$.

Probability space

We work with the probability triplet $\{\Omega, \mathcal{F}, \mathbb{P}\}$ and assume the existence of a market filtration $\{\mathcal{F}_{t_i}\}_{i\geq 0}$, which can also be denoted by $\mathcal{F}_i$ for brevity. In particular $\mathbb{P}$ is the real-world ("physical") probability

measure. All filtration-related concepts (mainly the martingality property) are defined with respect to the market filtration. To simplify the notation, in discrete time we use the following notation for conditional expectations:

$$\mathbb{E}_i x_{i+j} = \mathbb{E}_i[x_{i+j}] = \mathbb{E}^{\mathbb{P}}[x_{t_{i+j}} \mid \mathcal{F}_{t_i}].$$

In this chapter, if no probability measure is specified, the expectation is taken with respect to the real-world measure ($\mathbb{P}$). To perform a measure change from the physical measure $\mathbb{P}$ to the risk-neutral measure $\mathbb{Q}$, we introduce the Radon-Nikodym derivative $(d\mathbb{Q}/d\mathbb{P})\mid_{t_i}$, written μ_i for brevity. All regularity requests for the measure change process to exist and to be an L^2 positive martingale are supposed to hold.

Financial instruments

We assume that the financial market is such that there are no transaction costs nor taxes: investors can take any position (either long or short) in any asset. We assume the existence of the following:

1. The short-term nominal interest rate n_i – set by the central bank – is the interest agreed at time t_{i-1} and paid at time t_i by the bank account on the balance at time t_{i-1}. [1] It is used to discount payments. The short-term nominal interest rate process $\{n_i\}_{i=0,1,\ldots}$ is a previsible process, i.e. the short-term nominal interest rate n_i is $\mathcal{F}_{i-1}$-measurable.

2. The bank account $B_i = \prod_{j=1}^{i}(1+\tau_j n_j)$, with $B_0 = 1$. Here τ_i represents the year fraction between times t_{i-1} and t_i. Since the interest rate n_i is $\mathcal{F}_{i-1}$-measurable, the bank account process $\{B_i\}_{i=0,1,\ldots}$ is a previsible process. At time t_{i-1} the cash flow that occurs at time t_i is already known: this is why the bank account is often referred to as the riskless asset.

3. The first two properties implicitly imply a lower bound on negative short nominal interest rates: in this model rates can be negative (as they have been in 2012, for example German Bunds up to 2 years maturity, or in 2014, when the European Central Bank set the deposit rate to -0.2%), however they can not be lower than -100%, otherwise the nominal bank account would have a negative value, which is not possible. Rational agents would not put money into such account that turns assets into liabilities.

4. A system of discount bonds $P(t_i, t_N)$, that pay one unit of currency at time t_N and have the following properties:

   - $P(t_i, t_N) = \mathbb{E}_i^{\mathbb{Q}}\left[\prod_{j=i+1}^{N}(1+\tau_j n_j)^{-1}\right]$

[1] Here we are assuming that the central bank lends money to the commercial banks at the same interest rate paid by these to the money market account holders. We are making the simplifying assumption that the central bank reviews its interest rate with the same time scale by which the interest are accrued in the money market account. This assumption allows one to include in the pricing model a fairly realistic description of central policy.

- $P(t_i, t_i) = 1, \ \ \forall i$
- $P(t_i, t_N) > 0, \ \ \forall i \leq N$
- $P(t_i, t_{i+1}) = B_i / B_{i+1}$.

5. The seasonality-adjusted price index process $\{I_i\}$ that describes the evolution over time of the price level[2].

6. A system of zero-coupon inflation index swaps (ZCIIS), such that the floating leg pays $(I_{i+M-1}/I_i) - 1$ and the fixed leg pays $(1 + X_M)^{M\tau_i} - 1$ (both payments happen at maturity). The strikes X_i are quoted at time t_i for all maturities $t_M > t_i$. Inflation payments are time-lagged in this model as it happens in reality: the price index is subject to revisions and in practice ZCIIS pay the inflation lagged by one period.

7. A system of index-linked zero-coupon bonds $P^I(t_i, t_M)$, which pay at maturity t_M the cash equivalent of the price index I_{M-1}. These bonds are priced consistently with the zero coupon inflation swaps seen in the previous point. Inflation payments are time-lagged in this model as it happens in reality: the price index is subject to revisions and in practice the inflation bonds pay the inflation lagged by one period. These bonds are quoted at time t_i for all maturities $t_M > t_i$. Here we ignore for pricing purposes the deflation floor.

8. A set of traded shares: we denote the price at time t_i of the k-th share by $S_{i,k}$.[3] We assume that if a share is traded at time t_i the cash settlement happens at time t_i.

Macroeconomic variables. The inflation rate is defined by $p_i = ((I_i/I_{i-1}) - 1)/\tau_i$. This is the annualised percentage growth rate of the price index.

The output gap x_i is defined as the difference between the actual and the potential log-linearised growth rate of the economy: $x_i = \hat{y}_i - \hat{y}_i^f$. [4] To provide a complete definition of the output gap, we introduce the Gross Domestic Product (GDP) Y_i – also known as output – which is the value of all final goods and services produced in the economy between times t_{i-1} and t_i. The GDP annualised growth rate is defined as: $y_i = ((Y_i/Y_{i-1}) - 1)/\tau_i$. The growth rate y_i is assumed to have a long term equilibrium level $\bar{y}$ such that $\mathbb{E}(y_i) \to \bar{y}$ as $i \to +\infty$. The variable $\hat{y}_i$ is defined as the percentage deviation between the GDP growth rate y_i and its long term equilibrium level $\bar{y}$: $\hat{y}_i = ((y_i/\bar{y}) - 1)$. Economists often refer to it as the log-linearised GDP growth rate, as $\hat{y}_i = ((y_i/\bar{y}) - 1) \cong log(y_i/\bar{y})$, when $y_i/\bar{y} \to 0$.

[2] Price indices time series clearly show seasonality, mainly driven by sales in January and July and prices increases around Christmas. We do not model these patterns directly at this stage because a seasonality correction can be easily introduced at the last stage. This can be done by assuming that the monthly inflation rate differs from the seasonality-adjusted inflation rate by a certain percentage. Intuitively, seasonality is more relevant for short-maturity inflation trades.

[3] Although these assets are not needed in the inflation model, at this stage we want to show that this setting is very general.

[4] The reason why we are involving log-linearisation will become clear shortly. More information is also available later in this chapter.

If we assume that the economy is subject to some "inefficiencies", we can introduce the potential GDP Y_i^f, which can be defined as the GDP produced if there is no inefficiency: intuitively these inefficiencies prevent the actual GDP Y_i from reaching the "full employment" GDP Y_i^f. We derive the variables y_i^f, $\bar{y}_i^f$, and $\hat{y}_i^f$ in a similar way. This completes the definition of the output gap x_i.

We assume that the processes $\{Y_i\}_{i=0,1,\ldots}$, $\{Y_i^f\}_{i=0,1,\ldots}$, and $\{I_i\}_{i=0,1,\ldots}$ are adapted, therefore the processes $\{x_i\}_{i=0,1,\ldots}$ and $\{p_i\}_{i=0,1,\ldots}$ are adapted too. To complete the formalisation, one needs to assume that all stochastic processes involved in the model converge to a finite equilibrium level when time tends to infinity. Normally, in macroeconomic models the details of this convergence are not specified. Here we require that the stochastic processes of the economic variables are mean-ergodic, and the equilibrium levels should be regarded as their long-term means, which we require to be finite, together with their second, third and fourth moments.

Economic assumptions. We list the microeconomic and macroeconomic assumptions used to describe the economy. These assumptions will be further presented with full details at a later stage.

1. The economy is closed, i.e. there is no exchange rate nor foreign market.
2. All markets are in equilibrium, i.e. demand matches supply for all goods and services markets.
3. The economy is a monetary one, i.e. there is no barter.
4. The representative consumer maximizes his utility function under an intertemporal budget constraint.
5. The representative consumer draws his utility from consuming and keeping cash balances for safety (money-in-utility approach).
6. There is no public sector, therefore there is no taxation.
7. Labour is the only production factor in the technology: this implies that no capital is required, therefore there are no investments.
8. Savings are invested in bonds that pay a coupon equal to the nominal interest rate.
9. The representative consumer consumes multiple goods, each of which is produced in a monopolistic market.
10. The output coincides with private consumption, as there is no government expenditure, no import/export, no taxes nor investment.
11. Firms maximize profits but are not free to modify in each period the prices they charge (sticky prices).

12. The central bank sets the short rate as a linear function of inflation and output gap (Taylor rule). The short rate moves around its equilibrium level.

13. The short rate can be negative in some circumstances.

14. There exists a system of expectations for the output gap and inflation.

15. There is no credit risk.

2.1.2 Model derivation

We follow Walsh [115] to introduce the main equations of the DSGE baseline model: as the material of this section is standard, we give a high level overview. Another interesting overview can be found in Clarida, Gali & Gertler [40]. A slightly more complete introduction to these models can be found in Cochrane [43].

Economy description

The baseline model we work with represents a simple closed economy, with no government and no tax system. The production function depends only on labour since capital is not considered: therefore there is no investment. From a macroeconomic perspective we can state that the output at time t_i equals the aggregate consumption at time t_i:

$$Y_i = C_i. \tag{2.1}$$

The economy is a monetary one with money M_i and price level I_i.

Consumers

The representative household solves a two-steps optimisation problem. It first decides how to allocate its total consumption between different goods – all produced in monopolistic markets – and then chooses how much to consume in total, how much cash to hold, how much to invest in bond holdings and how many hours to work.

In the first step we assume the existence of a continuum of goods c_j produced by a continuum of monopolistic firms j (by convention $j \in [0, 1]$). At time t_i the household chooses the combination of goods c_{ji} that minimizes the cost of the total consumption:

$$\min \int_0^1 p_{ji} c_{ji} dj$$

by taking into account the constraint

$$\left(\int_0^1 (c_{ji})^{\frac{\theta-1}{\theta}} dj\right)^{\frac{\theta}{\theta-1}} \geq C_i.$$

Here p_{ji} is the price of the good j at time t_i and C_i is the total consumption time t_i. The parameter θ is used to model the price elasticity, i.e. how price-sensitive consumption is. This standard optimisation problem is solved in Walsh [115] (p. 233) and yields the optimal amount of consumption of good j given the general price level I_i, the total consumption C_i (to be determined in the next step) and the price of good j, p_{ji}:

$$c_{ji} = \left(\frac{p_{ji}}{I_i}\right)^{-\theta} C_i. \tag{2.2}$$

The second step is modelled as an intertemporal maximisation of the expected utility under a budget constraint, and yields the usual Euler conditions.

The representative household draws its utility from consuming goods (C_i) and holding real cash balances (M_i/I_i) as insurance against uncertainty: furthermore it has negative utility from supplying labour N_i and can save money and purchase bonds B_i that pay a coupon equal to the nominal interest rate n_i in each period. We assume a power utility function: the problem is to find the sequences C_i, M_i, B_i and N_i that solve the problem

$$\max \sum_{t_i=t_0}^{\infty} \beta^{t_i} \mathbb{E}_0 \left[\frac{C_i^{1-\sigma}-1}{1-\sigma} + \frac{\alpha}{1-d}\left(\frac{M_i}{I_i}\right)^{1-d} - \frac{N_i^{1+\eta}-1}{1+\eta}\right].$$

The parameters σ, d, $\alpha > 0$, η indicate how consumption, real cash balance and labour supply influence the utility function. The expectation $\mathbb{E}[\,\cdot]$ is taken with respect to the physical measure $\mathbb{P}$, as usual in any macroeconomic model: in this chapter when no measure is specified it is assumed that the physical measure is used. The parameter $\beta \in (0, 1]$ represents a subjective discount factor over one period. The parameter σ, that is also known as "relative risk aversion", is used to model elasticity of utility to consumption in a constant relative risk aversion (CRRA) utility function. When σ is very high, the agents are extremely risk-averse, as an increase in consumption creates an smaller increase in utility than the correspondent reduction in utility given the same absolute reduction in consumption. When σ is zero, there is risk-neutrality, i.e. the utility grows linearly with consumption; when σ tends to 1, the utility function becomes a logarithmic utility function, which is moderately risk averse. When σ is negative, the utility function becomes convex, indicating risk-seeking (this case is normally excluded).

The optimisation is carried out under the constraint that the total wealth at time t_i (which is allocated between consumption, real cash balance and bond holdings) has been derived from the previous period

or gained from supplying labour (W_i is the wage gained for 1 unit of labour at time t_i). No wealth is introduced into the system *ex nihilo*:

$$C_i + \frac{M_i}{I_i} + \frac{B_i}{I_i} = \frac{W_i N_i}{I_i} + \frac{M_{i-1}}{I_i} + \frac{B_{i-1}}{I_i}(1 + n_{i-1}).$$

The derivation of the Euler conditions is standard and can be found for example in the second chapter of Walsh [115]. The first order conditions for this problem are the following:

$$C_i^{-\sigma} = (1 + n_i)\beta \mathbb{E}_i \left[\frac{I_i}{I_{i+1}} C_{i+1}^{-\sigma} \right] \tag{2.3}$$

$$\alpha \left(\frac{M_i}{I_i} \right) C_i^{\sigma} = \frac{n_i}{1 + n_i} \tag{2.4}$$

$$\frac{N_i^{\eta}}{C_i^{-\sigma}} = \frac{W_i}{I_i}. \tag{2.5}$$

We focus on C_i. Given the concavity of the utility function, the first order conditions are sufficient for optimality. Since we assume that there is no government, no capital stock (and therefore no investment) and that the economy is closed, we substitute the consumption with the output by (2.1), getting

$$Y_i^{-\sigma} = (1 + n_i)\beta \mathbb{E}_i \left[\frac{I_i}{I_{i+1}} Y_{i+1}^{-\sigma} \right].$$

This condition may be rewritten in log-linearized terms around a zero inflation equilibrium point after some approximations

$$\hat{y}_i = \mathbb{E}_i \hat{y}_{i+1} - \frac{1}{\sigma}(\hat{n}_i - \mathbb{E}_i p_{i+1}). \tag{2.6}$$

This result is shown in Appendix A. The inflation rate p_i is defined as the annualised relative change of the price level I_i from t_{i-1} to t_i.[5]

Firms

The firm profit maximisation problem has to take into account three constraints: the demand curve, the production technology and price stickiness. It involves finding the optimal amount of labour N_i to

[5]It is worth explaining how the log-linearisation used above works. Given a variable F_i at time t_i we assume that its long term equilibrium level is $\bar{F}$ (i.e. that the limit of F_i when time goes to infinity is $\bar{F}$). With the lower case hat $\hat{f}_i$ we indicate the deviation at time t_i of the variable F_i from its long term equilibrium level $\bar{F}$ in percentage terms: this can be approximated with the natural logarithm of their ratio for small deviations. In formulas:

$$\hat{f}_i = \frac{F_i}{\bar{F}} - 1 \cong \log(\frac{F_i}{\bar{F}}).$$

Uhlig [114] gives extensive explanations and examples of this technique: given the analogy between this transformation and the natural logarithm, products can be approximated with sums, powers become multiplicative coefficients and constants disappear as they do not differ from their equilibrium level.

minimize the production cost and the optimal good price p_{ji} to maximize the expected profit stream. The demand curve is given by (2.2). Secondly, technology is such that the output of the j-th firm depends only on labour N_{ji}

$$c_{ji} = Z_i N_{ji}.$$

Here Z_i is a positive random variable with mean 1 that represents a stochastic productivity shock. Thirdly, firms are able to adjust their prices in each period only with probability 1 - ω, with $\omega \in (0, 1]$. This price stickiness assumption is the most interesting one and is essential to define the inflation dynamics of this model.

The first consequence is that the output Y_i will deviate from the output in flexible prices Y_i^f: by making use of (5), we can then define their difference in log-linearized terms as the output gap

$$x_i = \hat{y}_i - \hat{y}_i^f. \tag{2.7}$$

We do not explain the subsequent details: instead we develop some intuition of the inflation mechanics. Since prices are sticky and firms are maximizing their expected profit stream, firms tend to increase their prices not only if production costs rise (which would also happen in a flexible prices framework), but also to compensate for the expected losses they can face as they may not increase prices in the future (with probability ω).

This has two important consequences: firstly, as prices influence output via the demand curve (2.2) and the macroeconomic relation (2.1), inflation is related to the output gap. The output gap increases with inflation. Secondly, if there are inflation expectations, firms will raise prices in the current period because they may not be able to do so in the future. Inflation is therefore a self-fulfilling prophecy.

The result, after some algebraic manipulations, is the so-called neo-Keynesian Phillips curve, which states that the current level of inflation depends both on inflation expectations and the output gap:

$$p_i = \beta \mathbb{E}_i p_{i+1} + k x_i. \tag{2.8}$$

The parameter $k \geqslant 0$ can be regarded as a measure of the market price flexibility and is defined as

$$k = \frac{(1-\omega)(1-\beta\omega)(\sigma+\eta)}{\omega}.$$

It is worth stressing that if prices never change, $\omega = 1$: therefore k equals zero and inflation will only be driven by expectations. As before, the derivation can be found in Walsh [115] (5.4, 5.7).

Putting things together

Equation (2.6) can be rewritten in terms of output gap by using (2.7) by defining

$$u_i = \mathbb{E}_i \hat{y}^f_{i+1} - \hat{y}^f_i \,.$$

We get to a final form for (2.6) that can be put in a system with (2.8)

$$x_i = \mathbb{E}_i x_{i+1} - \frac{1}{\sigma}(\hat{n}_{i+1} - \mathbb{E}_i p_{i+1}) + u_i. \tag{2.9}$$

As we define the rate n_{i+1} as the rate set by the central bank at time t_i and paid at time t_{i+1} we have written $\hat{n}_{i+1}$ rather than $\hat{n}_i$: this proposed change with respect to the original formulation lets one reconcile the original DSGE model with the request that the short rate is a previsible process. This curve can be interpreted as a neo-Keynesian demand curve, where the output gap shows a negative dependency on a function of the real interest rate ($\frac{1}{\sigma}(\hat{n}_{i+1} - \mathbb{E}_i p_{i+1})$). The process $\{u_i\}_{i=0,1,\ldots}$ can be thought as a discrete-time stochastic process that relates the level of the log-linearised flexible price output deviation from its expectations: this difference should depend somehow on the productivity shock seen in (2.1.2), but for our purposes we can think of it as a general stochastic process. Again, we stress that the original macroeconomic model does not make any further assumptions on the shock processes: we take the necessary steps in the following sections, where the DSGE model is used for pricing purposes.

The Taylor rule and the central bank

Equations (2.9) and (2.8) define a discrete-time, bi-dimensional, forward looking stochastic system which is influenced by two exogenous variables: the log-linearized short rate $\hat{n}_{i+1}$ and the process $\{u_i\}_{i=0,1,\ldots}$, related to the productivity shock. We introduce the central bank, which uses the short rate as policy tool. In each period the central bank changes the short rate in response to the inflation and output gap with the following rule:

$$\hat{n}_{i+1} = \delta_\pi p_i + \delta_x x_i + v_i. \tag{2.10}$$

This rule, proposed by Taylor [112], states that the central bank responds to inflation and output gap by setting the short rate: [6] a discrete-time stochastic process $\{v_i\}_{i=0,1,\ldots}$, independent from the process $\{u_i\}_{i=0,1,\ldots}$, is added to increase the flexibility of the model. We remind the reader that the rate n_{i+1} is set by the central bank at time t_i and paid at time t_{i+1}: for this reason we allow a lag in the above form of the Taylor rule. At this stage we also notice that the short rate can be negative in this formulation,

[6] This rule was originally found as an econometric relationship. It has now become a common tool to analyse central bank policy, and its implementation is even available on the Bloomberg software. In chapter 3 we will propose a similar rule based on money supply, which is is probably more useful in the current low rates environment.

which is consistent with the assumptions we have made earlier: values of the nominal rate below -100%, albeit theoretically possible, are to be ruled out under a reasonable model parametrisation and economic assumptions. Finally one notes that the Taylor rule has been defined for $\hat{n}_{i+1}$, which, as explained for the other variables, is the percentage deviation of the nominal rate from its equilibrium level.

Bullard & Mitra [34] analyse similar rules with more realistic timing assumptions (the central bank may be reacting to future expectations of gap and inflation, or may be looking at their lagged values instead). In addition, the short rate can be smoothed as suggested by Woodford [116], essentially by combining (2.10) with an autoregressive process. This framework is somehow simple, as the central bank is not optimizing any objective function: this notwithstanding, the Taylor rule explains well the behaviour of the FED in the last decades, as shown by Clarida, Gali & Gertler [41]. Finally, this linear rule can be regarded as good linear approximation of the optimal policy solution.

2.1.3 System stability

If the Taylor rule (2.10) is plugged into (2.9) and (2.8), we obtain the following system:

$$\begin{bmatrix} x_i \\ p_i \end{bmatrix} = \frac{1}{\sigma + \delta_x + k\delta_\pi} \left(\begin{bmatrix} \sigma & 1 - \beta\delta_\pi \\ k\sigma & k + \beta(\sigma + \delta_x) \end{bmatrix} \mathbb{E}_i \begin{bmatrix} x_{i+1} \\ p_{i+1} \end{bmatrix} + \begin{bmatrix} 1 \\ k \end{bmatrix} (\sigma u_i - v_i) \right). \tag{2.11}$$

The notation is made more compact by defining:

$$A = \frac{1}{\sigma + \delta_x + k\delta_\pi} \begin{bmatrix} \sigma & 1 - \beta\delta_\pi \\ k\sigma & k + \beta(\sigma + \delta_x) \end{bmatrix}$$

$$K = \frac{1}{\sigma + \delta_x + k\delta_\pi} \begin{bmatrix} 1 \\ k \end{bmatrix}$$

$$\xi_i = \begin{bmatrix} x_i \\ p_i \end{bmatrix}$$

$$w_i = (\sigma u_i - v_i).$$

Using the above definitions, we get a more compact expression of the system:

$$\xi_i = A\mathbb{E}_i\xi_{i+1} + Kw_i. \tag{2.12}$$

The proof of this result is given in Appendix A. We investigate the stability conditions, which is equivalent

to asking what reaction function – characterised by the parameters δ_π and δ_x – keeps the economy on a stable path. For example, if the central bank only responds to inflation (i.e. $\delta_x = 0$), we ask whether δ_π has to be greater or lower than one, i.e. if the central bank has to increase the short rate above its equilibrium level more or less than the realised inflation. Clarida, Gali & Gertler [41] show that $\delta_\pi > 1$ is typical of the FED during the Volker tenure (in the early 1980s in the U.S.), which was characterized by lower inflation and output volatility.

The economic intuition is that a reaction parameter close to one means that the nominal rate is increased by the same amount of inflation, thus keeping the real rate unchanged and not stimulating the economy. Bullard & Mitra [34] find that in general the system is stable if and only if

$$k(\delta_\pi - 1) + (1 - \beta)\delta_x > 0. \tag{2.13}$$

The proof of this result is given in Appendix A. They obtain this rule by requiring that both eigenvalues of A lie inside the unit circle. This request is also derived by Blanchard & Khan [15] and used by Flashel & Franke [56] or Walsh [115].

2.2 Using the DSGE model for pricing purposes

2.2.1 Arbitrage-free pricing

The set-up introduced so far lets us use a DSGE macroeconomic model to price inflation derivatives in a no-arbitrage framework with a few minor changes. In general, the pricing kernel properties reviewed in the previous chapter enable one to write the present value at time t_i of a derivative h_i paying the inflation-linked payoff H_N^π at time t_N in the form:

$$h_i = \mathbb{E}^{\mathbb{P}}[\psi_N H_N^\pi | \mathcal{F}_i] \frac{1}{\psi_i}.$$

This framework is very general and can be used to price any payoff. The aim is to price complex inflation trades given the prices of vanilla interest rate and inflation options.

Use of the macroeconomic model: inputs and outputs

Here we make our proposal regarding how to use the DSGE for pricing purposes. We make a distinction between input parameters (the structural parameters of the DSGE model, equilibrium nominal rates, inflation expectations, output-gap expectations), and calibrated parameters (the volatilities and the market prices of risk – to be introduced later in this section).

Calibrating the market prices of risk is not a usual procedure in derivatives pricing, because the real-world drift is not an input in the classic Black-Scholes formula to price contingent claims: however, the DSGE model takes expectations (under the $\mathbb{P}$ measure) as an input. Since these expectations play the role of the drift in (2.12), we need to take both inflation expectations and market implied levels (from the zero-coupon inflation swaps, for example) to calibrate the market prices of risk. The expectation of inflation is a kind of self-fulfilling prophecy: if there are expectations of inflation, then inflation will rise. This exercise is particularly useful for inflation markets, since it is often observed that inflation forecasts and expectations can significantly differ from levels of inflation calculated on a forward basis. Such differences can arise both because of risk aversion and market supply and demand factors: the market can be to a significant extent a "one-way street", overall "short" inflation. In other words market participants on the whole wish to hedge themselves against inflation. In particular, pension funds liabilities have to be covered, while the number and the size of inflation payers is limited.

The idea of using market forecasts as model input, although not commonly used in standard derivatives pricing, lets one use a theoretically consistent macroeconomic model for the pricing of inflation derivatives.

The algorithm we suggest calibrates to both the nominal term structure and the zero-coupon inflation index swaps (ZCIIS), leaving much flexibility to calibrate to market smiles. To achieve this, we explore the statistical properties of the main economic variables, as implied by the DSGE model presented above.

Statistical properties of the inflation rate

From equation (2.12) we write explicitly the dynamics of the inflation rate:

$$p_i = A_{2,1}\mathbb{E}_i x_{i+1} + A_{2,2}\mathbb{E}_i p_{i+1} + K_2 w_i. \tag{2.14}$$

Here $A_{i,j}$ is the (i,j)-th element of the matrix A, and K_i is the i-th element of the vector K. This equation states that the inflation dynamics depend on future expectations of output gap and inflation, plus a stochastic noise term introduced by the dynamics of the output gap and the central bank behaviour: we can safely assume that other factors, such as measurement errors, price index basket rebalancing or any other idiosyncratic factor not directly modelled in this framework may add noise to the inflation dynamics.[7] On the basis of these considerations, we add a further independent source of randomness, modelled with the adapted process $\{z_i\}_{i=0,1,\ldots}$: we require this process to have zero mean, to be independent from its past realisations, to be independent from $\{u_i\}_{i=0,1,\ldots}$ and $\{v_i\}_{i=0,1,\ldots}$, and to have finite

[7] If one takes the view that this third source of randomness is not advisable to include, one assumes that its value is always 0 with probability 1. As one notices in the following developments, this third source of randomness is mainly used in the calibration phase in order to have an additional degree of freedom and has no impact whatsoever on the theoretical development of the model.

variance $\mathrm{Var}(z_i)$, third and fourth moments ($\mathrm{Skew}(z_i)$ and $\mathrm{Kurt}(z_i)$ respectively).

The new expression for the inflation rate becomes:

$$p_i = A_{2,1}\mathbb{E}_i x_{i+1} + A_{2,2}\mathbb{E}_i p_{i+1} + K_2 w_i + z_i. \tag{2.15}$$

Its mean, variance and autocovariance are:

$$\mathbb{E}[p_i] = A_{2,1}\mathbb{E}_i x_{i+1} + A_{2,2}\mathbb{E}_i p_{i+1} \tag{2.16}$$

$$\mathrm{Var}(p_i) = (K_2)^2(\sigma^2\mathrm{Var}(u_i) + \mathrm{Var}(v_i)) + \mathrm{Var}(z_i) \tag{2.17}$$

$$\mathrm{Cov}(p_i, p_{i+j}) = 0, \qquad j \neq 0. \tag{2.18}$$

We note that the variance of the inflation process is a linear combination of the variances of the three processes $\{u_i\}_{i=0,1,\ldots}$, $\{v_i\}_{i=0,1,\ldots}$ and $\{z_i\}_{i=0,1,\ldots}$.

Finally, we calculate the centered third and fourth moments: these may be needed in order to analyse the inflation distribution in a more complete fashion:

$$\mathbb{E}\left[(p_i - \mathbb{E}(p_i))^3\right] = (K_2)^3\sigma^3\mathrm{Skew}(u_i) - (K_2)^3\mathrm{Skew}(v_i) + \mathrm{Skew}(z_i) \tag{2.19}$$

$$\mathbb{E}\left[(p_i - \mathbb{E}(p_i))^4\right] = (K_2)^4\sigma^4\mathrm{Kurt}(u_i) + (K_2)^4\mathrm{Kurt}(v_i) + \mathrm{Kurt}(z_i) +$$

$$+6(K_2)^4\sigma^2\mathrm{Var}(u_i)\mathrm{Var}(v_i) + 6(K_2)^2\mathrm{Var}(v_i)\mathrm{Var}(z_i) + 6(K_2)^2\sigma^2\mathrm{Var}(u_i)\mathrm{Var}(z_i). \tag{2.20}$$

We remind the reader that the first moment of the processes $\{u_i\}_{i=0,1,\ldots}$, $\{v_i\}_{i=0,1,\ldots}$ and $\{z_i\}_{i=0,1,\ldots}$is 0.

Statistical properties of the short-term nominal interest rate

The nominal interest rate n_i is defined as

$$n_i = \bar{n}(1 + \hat{n}_i) \tag{2.21}$$

where $\bar{n}$ is the equilibrium nominal interest rate, which is the short rate that would be chosen by the central bank if the adjustment required by the Taylor rule was zero as $\hat{n}_i$ follows (2.10). This follows by the definition of $\hat{n}_i$ as the log-linearised difference between the actual rate and equilibrium rate.

We take the equilibrium nominal rate $\bar{n}$ as a constant input that can be obtained from research and is therefore not calibrated to any traded asset. We assume that the short rate is used to discount payments

between different counterparties, i.e. it plays the role of the Libor rate: this assumption, albeit strong, simplifies the problem considerably.[8]

If we plug the Taylor rule (2.10) into (2.21) we rewrite the nominal rate as

$$n_{i+1} = \bar{n}(1 + \delta_x x_i + \delta_\pi p_i + v_i). \tag{2.22}$$

We can compact the notation by introducing the vectors

$$\delta = \begin{bmatrix} \delta_x \\ \delta_\pi \end{bmatrix} \qquad \xi_i = \begin{bmatrix} x_i \\ p_i \end{bmatrix}.$$

The interest rate can be rewritten as

$$n_{i+1} = \bar{n}(1 + \delta^T \xi_i + v_i),$$

where $(x)^T$ is the transpose of the vector x. Finally, by making use of (2.12) and (2.15) we get:

$$n_{i+1} = \bar{n}(1 + \delta^T A \mathbb{E}_i \xi_{i+1} + \delta^T K w_i + \delta^T e_2 z_i + v_i),$$

where $e_2^T = \begin{bmatrix} 0 & 1 \end{bmatrix}$.

By making use of this expression we calculate the mean, variance and the autocovariance of the nominal interest rate:

$$\mathbb{E}[n_{i+1}] = \bar{n}(1 + \delta^T A \mathbb{E}\xi_{i+1}) \tag{2.23}$$

$$\mathrm{Var}(n_{i+1}) = (\bar{n})^2(\delta^T K)^2 \sigma^2 \mathrm{Var}(u_i) + (\bar{n})^2(1 - \delta^T K)^2 \mathrm{Var}(v_i) + (\bar{n})^2 \delta_\pi^2 \mathrm{Var}(z_i) \tag{2.24}$$

$$\mathrm{Cov}(n_i, n_{i+j}) = 0, \quad j \neq 0. \tag{2.25}$$

We note that the variance of the interest rate process is a linear combination of the variances of the three processes $\{u_i\}_{i=0,1,\ldots}$, $\{v_i\}_{i=0,1,\ldots}$ and $\{z_i\}_{i=0,1,\ldots}$. We take the equilibrium rates, output gap and inflation expectations as inputs: they may be provided by macroeconomic research or can just be expression of the trader's views.

Similarly to what could be done for the inflation rate, we can also calculate the centered third and fourth moments: these are needed in order to analyse the short rate distribution in a more complete

[8] We recall that in continuous time the short rate $n(t) = f(t,t) = lim_{\Delta T \to 0} F(t, t, t + \Delta T)$ where the forward rate is defined as $F(t, S, T) = (P(t,S)/P(t,T) - 1)/(T - S)$ with $T > S$. In discrete time we define $n_i = f_{i,i} = F(t_i, t_i, t_{i+1})$, therefore getting $n_i = (1/P(t_i, t_{i+1}) - 1)/\tau_{i+1}$. As a consequence, the short rate can be used as the Libor rate, provided that there are no credit concerns in the interbank markets.

fashion:

$$\mathbb{E}\left[(n_i - \mathbb{E}(n_i))^3\right] = [(\delta^T K)^3 \sigma^3 \text{Skew}(u_i) - (1-\delta^T K)^3 \text{Skew}(v_i) + (\delta_\pi)^3 \text{Skew}(z_i)](\bar{n})^3 \tag{2.26}$$

$$\mathbb{E}\left[(n_i - \mathbb{E}(n_i))^4\right] = [(\delta^T K)^4 \sigma^4 \text{Kurt}(u_i) + (1-\delta^T K)^4 \text{Kurt}(v_i) + (\delta_\pi)^4 \text{Kurt}(z_i) +$$

$$+6(\delta^T K)^2 \sigma^2 (1-\delta^T K)^2 \text{Var}(u_i)\text{Var}(v_i) + 6(1-\delta^T K)^2 \delta_\pi^2 \text{Var}(v_i)\text{Var}(z_i) + 6(\delta^T K)^2 (\sigma\delta_\pi)^2 \text{Var}(u_i)\text{Var}(z_i)](\bar{n})^4. \tag{2.27}$$

Finally, we calculate the covariance between the nominal rate n_{i+1} and the inflation p_i, both $\mathcal{F}_i$-measurable:

$$\text{Cov}(p_i, n_{i+1}) = \bar{n} K_2 (\delta^T K) \sigma^2 \text{Var}(u_i) + (\bar{n} K_2 (1-\delta^T K))\text{Var}(v_i) + \bar{n}\delta_\pi \text{Var}(z_i). \tag{2.28}$$

The covariance depends on the Taylor rule parameters vector δ, which makes explicit the philosophy of our modelling approach: any dependence between the nominal interest rate and inflation is not specified exogenously but is a consequence of the central bank reaction function. Furthermore, the correlation becomes one if there is no uncertainty in the Taylor rule, i.e. $\text{Var}(v_i) = 0$ and if $\text{Var}(z_i) = 0$: in this case the central bank reacts deterministically to any change in the economy.

The other interesting limit case is when the output gap evolves deterministically, i.e. $\text{Var}(u_i) = 0$, $1 - \delta^T K < 0$, and $\text{Var}(z_i) = 0$: rates evolve stochastically and correlation becomes -1. As rates increase, the output gap decreases deterministically (because of the demand curve (2.9)), bringing down the inflation according to the Phillips curve (2.8). In this case the only source of randomness is the uncertainty in the short rate evolution due to the Taylor rule.

The DSGE model augmented with the Taylor rule allows for this correlation to take any values between -1 and 1, depending on the central bank reaction function and the specification of the sources of randomness: this can be arguably regarded as an interesting feature of the model, because it does not impose *a priori* any constraint on the correlation range.

Calibrating to rates and inflation smiles: the normal case

The prices of nominal rates and inflation caps/floors across different strikes and maturities are available from brokers or investment banks (for example the Bloomberg pages VOLS or RILO): we can thus deduce the caplet/floorlet prices. Unlike options on other underlyings, inflation options are quoted in prices, not in implied volatilities. By making some distributional assumptions on the nominal rates and inflation, we summarise the distribution using only a few parameters.

For example, we can assume a normal distribution and fit its volatility to the option prices for each

maturity: this assumption is both convenient from an analytical perspective (closed formulas for option prices are obtained) and from a practitioner point of view: if rates are normally (and not lognormally) distributed, the distribution of their relative increments is skewed, (and not Gaussian as in the Black model).[9] In this case we calibrate the variances of the processes $\{u_i\}_{i=0,1,\ldots}$, $\{v_i\}_{i=0,1,\ldots}$, and $\{z_i\}_{i=0,1,\ldots}$ to obtain the market implied variances for nominal rates and inflation. We use conditions (2.17) and (2.24) to calculate the market implied variances for $\{u_i\}_{i=0,1,\ldots}$, $\{v_i\}_{i=0,1,\ldots}$, and $\{z_i\}_{i=0,1,\ldots}$ given the market implied variances of rates/inflation caplets/floorets: a word of caution should be issued, as there is no guarantee to obtain positive variances from this basic algorithm. Negative variances could be floored to zero or more sophisticated root-searching algorithms can be used.

In general, our approach does not rely on any specific distribution: alternative specifications are possible. It is only advisable to use the same distribution for the DSGE shocks $\{u_i\}_{i=0,1,\ldots}$, $\{v_i\}_{i=0,1,\ldots}$, and $\{z_i\}_{i=0,1,\ldots}$, nominal rates and inflation. However, closed formulas are needed to express market prices of options as a function of the model parameters.

Measure change under normality assumptions

At this stage we make explicit the measure change process $\{\mu_i\}_{i=0,1,\ldots}$ to use the real-world expectations to price derivatives in the risk-neutral measure. In this work we will be restricting our choice of measure change processes to those that have the diffusion coefficient equal to the market prices of risk. We define the measure change processes as a discretely-sampled exponential Gaussian martingale: this strategy allows one to obtain a positive martingale. A general introduction to exponential Lévy martingales can be found in Appelbaum [6].

To simplify the notation, we rewrite equation (2.1.3) including the variable z_i in matrix format using the following notation:

$$\xi_i = A\mathbb{E}_i\xi_{i+1} + Kw_i + e_2 z_i = A\mathbb{E}_i\xi_{i+1} + K\sigma u_i - Kv_i + e_2 z_i. \tag{2.29}$$

Defining the matrix C as follows:

$$C = \begin{bmatrix} \sigma K_1 & -K_1 & 0 \\ \sigma K_2 & -K_2 & 1 \end{bmatrix}$$

and compacting all three Gaussian sources of randomness in the three-dimensional vector ε_i defined as:

$$\varepsilon_i = \begin{bmatrix} u_i & v_i & z_i \end{bmatrix}$$

[9] The discussion on what distribution is used to quote inflation option prices is fully developed in chapter 5.

the notation is further simplified into:

$$\xi_i = A\mathbb{E}_i\xi_{i+1} + C\varepsilon_i^T.$$

One notes that the variance-covariance matrix for the vector ε_i is written as:

$$\Sigma_i^\varepsilon = \begin{bmatrix} \mathrm{Var}(u_i) & 0 & 0 \\ 0 & \mathrm{Var}(v_i) & 0 \\ 0 & 0 & \mathrm{Var}(z_i) \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(\varepsilon_i^1) & 0 & 0 \\ 0 & \mathrm{Var}(\varepsilon_i^2) & 0 \\ 0 & 0 & \mathrm{Var}(\varepsilon_i^3) \end{bmatrix}.$$

At this point we introduce the three-dimensional deterministic vector process $\{\lambda_i\}_{i=0,1,\ldots}$defined as:

$$\lambda_i = \begin{bmatrix} \lambda_i^u \\ \lambda_i^v \\ \lambda_i^z \end{bmatrix}.$$

The quantities defined above are used to specify the measure change process $\{\mu_i\}_{i=0,1,\ldots}$, following and generalising Shreve [109]. The measure change process is therefore defined as a multivariate Gaussian exponential martingale in the form:

$$\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_i} = \mu_i = e^{-\epsilon_i\lambda_i - 1/2\lambda_i^T\Sigma_i^\varepsilon\lambda_i}. \tag{2.30}$$

One requires $\mu_0 = 1$ and the market price of risk vector process $\{\lambda_i\}_{i=0,1,\ldots}$to be regular enough for the measure change process $\{\mu_i\}_{i=0,1,\ldots}$ to be a positive and square-integrable martingale. Moving to the risk-neutral measure $\mathbb{Q}$, one obtains that the new process $\nu_i = \varepsilon_i + \lambda_i$ is a zero-mean Gaussian process with independent realisations (both across times and across vector dimensions) under $\mathbb{Q}$. In this measure we also write $u_i^* = u_i + \lambda_i^u$, $v_i^* = v_i + \lambda_i^v$, $z_i^* = z_i + \lambda_i^z$, $w_i^* = \sigma u_i^* - v_i^*$.

We rewrite the expression for the macroeconomic variables (output gap and inflation) once the measure change from $\mathbb{P}$ to $\mathbb{Q}$ has been performed:

$$\xi_i = A\mathbb{E}_i\xi_{i+1} + C\nu_i = A\mathbb{E}_i\xi_{i+1} + C\lambda_i + C\varepsilon_i^T. \tag{2.31}$$

Informally, we can think to the linear function of the market prices of risk λ_i as a "wedge" that is premultiplied by some coefficients in the matrix C and then added to the deterministic linear function of the expectations $A\mathbb{E}\xi_{i+1}$ in order to calibrate the model to the traded prices of nominal bonds and inflation breakevens (through the relationship between nominal bonds, real bonds and inflation index

zero-coupon swaps).

Finally, one finds a compact expression for the nominal short rate n_i and the inflation rate p_i under $\mathbb{Q}$:

$$n_{i+1} = \bar{n}(1+\delta^T\xi_i + v_i^*) = \bar{n}(1+\delta^T(A\mathbb{E}_i\xi_{i+1} + C\nu_i^T) + v_i^*) = \bar{n}(1+\delta^T(A\mathbb{E}_i\xi_{i+1} + C\lambda_i + C\varepsilon_i^T) + v_i + \lambda_i^v)$$

$$p_i = A_{2,1}\mathbb{E}_i x_{i+1} + A_{2,2}\mathbb{E}_i p_{i+1} + K_2 w_i^* + z_i^* = A_{2,1}\mathbb{E}_i x_{i+1} + A_{2,2}\mathbb{E}_i p_{i+1} + \sigma K_2 u_i^* - K_2 v_i^* + z_i^* =$$

$$= A_{2,1}\mathbb{E}_i x_{i+1} + A_{2,2}\mathbb{E}_i p_{i+1} + \sigma K_2(u_i + \lambda_i^u) - K_2(v_i + \lambda_i^v) + (z_i + \lambda_i^z) = A_{2,1}\mathbb{E}_i x_{i+1} + A_{2,2}\mathbb{E}_i p_{i+1} + h\nu_i^T$$

where the vector h has been defined as:

$$h = \begin{bmatrix} \sigma K_2 & -K_2 & 1 \end{bmatrix}.$$

Calibrating to the nominal term structure

We show how to calibrate the model to the nominal interest rates observed in the market by making some approximations. We use market prices of discount factors to provide some expressions to be used in the calibration. We write:

$$P(t_0, t_{i+1}) = \mathbb{E}_0^{\mathbb{P}}[\psi_{i+1}] = \mathbb{E}_0^{\mathbb{Q}}\left[\prod_{j=0}^{i}(1+n_{j+1}\tau_{j+1})^{-1}\right] \cong \mathbb{E}_0^{\mathbb{Q}}\left[e^{-\sum_{j=0}^{i} n_{j+1}\tau_{j+1}}\right]$$

where the last linearisation creates some error that can be reduced by calibrating the model on a finer time grid. The term τ_{i+1} is the year fraction: $\tau_{i+1} = t_{t+1} - t_i$. In practice one does a bootstrapping over each time step, thanks to the fact that the interest rates level is independent from its previous levels.

The following step is to introduce the closed form expression for the nominal rate n_i, as obtained above from the DSGE model:

$$\mathbb{E}_0^{\mathbb{Q}}\left[e^{-\sum_{j=0}^{i} n_{i+1}\tau_{i+1}}\right] = \mathbb{E}_0^{\mathbb{Q}}\left[e^{-\sum_{j=0}^{i} \bar{n}(1+\delta^T(A\mathbb{E}_i\xi_{i+1}+C\nu_i^T)+v_i^*)\tau_{i+1}}\right]$$

$$= e^{-\sum_{j=0}^{i} \bar{n}\tau_{i+1}(1+\delta^T(A\mathbb{E}_i\xi_{i+1}+C\lambda_i)+\lambda_i^v)+\frac{\bar{n}^2\tau_{i+1}^2}{2}\delta^T C\Sigma_i^{\varepsilon}C^T\delta+\frac{\bar{n}^2\tau_{i+1}^2\mathrm{Var}(v_i)}{2})}.$$

By taking the expectations under the normality assumption for the vector ν_i, the one-period discount factors approximated closed form is:

$$\mathbb{E}_0^{\mathbb{Q}}\left[e^{-n_{i+1}\tau_{i+1}}\right] \cong e^{c_1+c_2\mathrm{Var}(u_i)+c_3\mathrm{Var}(v_i)+c_4\mathrm{Var}(z_i)} \tag{2.32}$$

This is useful for bootstrapping. We have:

$$c_1 = -\tau_{i+1}\bar{n}(1 + \delta^T(A\mathbb{E}_i\xi_{i+1} + C\lambda_i) + \lambda_i^v)$$

$$c_2 = \frac{1}{2}\left(\tau_{i+1}\bar{n}\delta^T K\sigma\right)^2$$

$$c_3 = \frac{1}{2}\left(\tau_{i+1}\bar{n}(1 - \delta^T K)\right)^2$$

$$c_4 = \frac{1}{2}(\tau_{i+1}\bar{n}\delta_\pi)^2.$$

Calibrating to the ZCIIS

As shown in Brigo & Mercurio [22] and in chapter 1, the value of the inflation leg of a zero-coupon inflation index swap (ZCIIS) can be regarded as the difference between the real and nominal zero-coupon bond prices with the same maturity date.

We exploit the model-independent relationship between real and nominal bond to write:

$$P^R(t_0, t_{i+1}) = P(t_0, t_{i+1}) + ZCIIS(t_0, t_{i+1}).$$

Since we observe the market prices of nominal bonds and ZCIIS for different maturities, we deduce the value of a real bond, even if these instruments are not traded in the market.

We assume that the real bond pays at maturity t_{i+1} the unit nominal multiplied by the underlying inflation index appreciation between times t_0 and t_i: this is to introduce the inflation publication lag in the formula, which becomes necessary since in reality the inflation rate is only published after a time lag.

The approximated closed form is obtained as follows:

$$P^R(t_0, t_{i+1}) = \mathbb{E}_0^{\mathbb{P}}\left[\frac{I_i}{I_0}\frac{\psi_{i+1}}{\psi_0}\right] = \mathbb{E}_0^{\mathbb{P}}\left[\frac{I_i}{I_0}\psi_{i+1}\right] =$$

$$= \mathbb{E}_0^{\mathbb{Q}}\left[\frac{I_i}{I_0}\prod_{j=1}^{i+1}\frac{1}{1+\tau_j n_j}\right] = \mathbb{E}_0^{\mathbb{Q}}\left[\prod_{j=1}^{i+1}\frac{1+\tau_{j-1}p_{j-1}}{1+\tau_j n_j}\right].$$

By making some straightforward Taylor expansions the last expression can be rewritten as:

$$P^R(t_0, t_{i+1}) = \mathbb{E}_0^{\mathbb{Q}} \left[\prod_{j=1}^{i+1} \frac{e^{log(1+\tau_{j-1}p_{j-1})}}{e^{log(1+\tau_j n_j)}} \right] \cong \mathbb{E}_0^{\mathbb{Q}} \left[\prod_{j=1}^{i+1} \frac{e^{\tau_{j-1}p_{j-1}}}{e^{\tau_j n_j}} \right]$$

$$= \mathbb{E}_0^{\mathbb{Q}} \left[e^{\sum_{j=1}^{i+1}(\tau_{j-1}p_{j-1} - \tau_j n_j)} \right].$$

We assume that $p_0 = 0$ and focus the attention on the one-period real discount factor.

By the same Gaussianity assumptions used above, the following bootstrapping closed formula is obtained by plugging (2.15) into the above expression:

$$\mathbb{E}_0^{\mathbb{Q}}[e^{-\tau_{i+1}n_{i+1}+\tau_i p_i}] = \mathbb{E}_0^{\mathbb{Q}}[e^{-\tau_{i+1}\bar{n}(1+\delta^T(A\mathbb{E}_i\xi_{i+1}+C\nu_i^T)+v_i^*)+\tau_i(A_{2,1}\mathbb{E}_i x_{i+1}+A_{2,2}\mathbb{E}_i p_{i+1}+h\nu_i^T)}] =$$

$$= \mathbb{E}_0^{\mathbb{Q}}[e^{-\tau_{i+1}\bar{n}(1+\delta^T A\mathbb{E}_i\xi_{i+1})+\tau_i(A_{2,1}\mathbb{E}_i x_{i+1}+A_{2,2}\mathbb{E}_i p_{i+1})+\nu_i^T\tau_i h-\tau_{i+1}\bar{n}(\delta^T C+v_i^*)}] = e^{b_1+b_2\mathrm{Var}(u_i)+b_3\mathrm{Var}(v_i)+b_4\mathrm{Var}(z_i)} \tag{2.33}$$

where

$$b_1 = \tau_i A_{2,1}\mathbb{E}_i x_{i+1} + \tau_i A_{2,2}\mathbb{E}_i p_{i+1} + h\lambda_i^T - \tau_{i+1}\bar{n}(1 + \delta^T(A\mathbb{E}_i\xi_{i+1} + C\lambda_i) + \lambda_i^v)$$

$$b_2 = \frac{1}{2}\left(\tau_i K_2\sigma - \tau_{i+1}\bar{n}\delta^T K\sigma\right)^2$$

$$b_3 = \frac{1}{2}\left(\tau_i K_2 - \tau_{i+1}\bar{n}(1 - \delta^T K)\right)^2$$

$$b_4 = \frac{1}{2}(\tau_i - \tau_{i+1}\bar{n}\delta_\pi)^2.$$

We stress that the variances calibrated from option prices are taken as an input in the above expression. The two approximated closed formulas for the nominal (2.32) and the real bond (2.33) can be used to find the values of λ_i that calibrate the model to the market, given the variances of the distributions of the shock factors u_i, v_i, and z_i.

To conclude this section, we observe that the adaptation of the DSGE model to pricing proposed above is extremely respectful of the the original macroeconomic model, but precisely for this reason it is also not straightforward to price derivatives. In fact, to obtain closed forms for the nominal and real bonds one has to resort to approximations and linearisations of exponentials, which are doable but not elegant: the model offers an insight of the macroeconomic forces operating behind the yield curve and the inflation dynamics, but all pricing of more complex derivatives has to happen using Monte Carlo simulations,

which can be cumbersome and time-consuming. Interestingly, the above section shows a first attempt to bridge the gap between two disciplines (monetary macroeconomics and financial mathematics) that are dealing with the same problem (inflation) in two different ways (DSGE modelling versus arbitrage pricing): this represents a step forward in the same direction indicated by Hughston & Macrina [72], who derive some inflation dynamics from a macroeconomic model — even if there is no concept of central bank policy in their work.

With these ideas in mind, in the following section we propose some continuous-time dynamics that are more tractable from a derivatives pricing perspective, while retaining the most significant aspects of the DSGE model presented in the previous sections. It is important to stress that the new dynamics we propose are not a one-to-one translation of the discrete-time DSGE model, but are rather inspired by it. To ensure that the proposed dynamics are meaningful, we bring some empirical evidence that shows that the proposed dynamics are realistic: finally, we show that the discrete-time DSGE model and the continuous-time model proposed generate similar distributions for the main economic variables.

2.3 Building the continuous-time version

In this section we propose a strategy to loosely translate the DSGE model into continuous time by making some assumptions. Therefore we show that some continuous-time dynamics can be derived from a widely-accepted macroeconomic model: they are used in the next chapter to build the inflation pricing model. From this point, the notation for the variable y in continuous time is $y(t)$.

The following assumptions are made:

1. There is no price flexibility for the firms, i.e. $\omega = 1$ and $k = 0$. This assumption is reasonable as markets tend to be far from the perfect competition model, and therefore prices are sticky, especially over a shorter time step.

2. The one-period subjective discount factor is equal to the inverse of the inflation targeting parameter: $\beta\delta_\pi = 1$. This assumption is sensible because, when the central bank fights inflation aggressively (i.e. $\delta_\pi \gg 1$), interest rates increase, pushing down the discount factor β.

3. The GDP growth rate is modelled in the same way as the output gap. In fact, because the output gap is defined as the difference between the actual and the potential GDP growth rate, and because the latter is an abstract concept (in particular their difference can be deemed to be constant over time), this means adding the constant potential growth rate to the output gap.

4. The GDP growth rate is defined as the percentage change of the GDP level from one period to the next one: $x_i = (X_i - X_{i-1})/X_{i-1}$. We change the notation and write: $x_{t_i} = (X_{t_i} - X_{t_{i-1}})/X_{t_{i-1}}$.

Furthermore, we generalise the time step and write: $\Delta t_i = t_i - t_{i-1}$, and we obtain: $x_{t_i} = (X_{t_i} - X_{t_{i-\Delta t_i}})/X_{t_{i-\Delta t_i}} = \Delta X_{t_i}/X_{t_{i-\Delta t_i}}$. Moving to continuous time we write $x(t)$ as $dX(t)/X(t)$. One needs to assume that the positive process $\{X_i\}_{i=0,1,...}$ is regular enough for the limit to exist.

5. A similar line of thought can be followed to show how one moves from the discrete-time definition of inflation, as the percentage change in the price index level (i.e. $p_i = (I_i - I_{i-1})/I_{i-1}$), to the equivalent continuous-time definition (i.e. $p(t)$ is written as $dI(t)/I(t)$). Again we make an obvious request of positivity for the price index process $\{I_i\}_{i=0,1,...}$. One needs to assume that the positive process $\{I_i\}_{i=0,1,...}$ is regular enough for the limit to exist.

6. There are measurement errors and other sources of uncertainty for both inflation and growth rate, modelled by the m-dimensional zero-mean random variable z_i. The m components of this random variable (called z_i^j, with $1, 2, ..., j, ..., m$) are independent from each other. The random variable z_i is also independent from w_i. The effects of the shock z_i^j on x_i and p_i are modelled by the m-dimensional real-valued deterministic processes $\{a_i\}_{i=0,1,...}$ and $\{b_i\}_{i=0,1,...}$, where their single components have notation $a_{i,j}$ and $b_{i,j}$.

7. The product of the $\mathbb{P}$-expectation terms by some constants that appear in the DSGE model can be written as $\sigma/(\sigma+\delta_x+k\delta_\pi))\mathbb{E}_i x_{i+1} = m_X(t_i)(t_{i+1}-t_i)$ and $(k+\beta(\sigma+\delta_x))/(\sigma+\delta_x+k\delta_\pi)\mathbb{E}_i p_{i+1} = m_I(t_i)(t_{i+1}-t_i)$ respectively. We assume that the quantities $m_X(t_i)$ and $m_I(t_i)$ are realisations of adapted stochastic processes. This means that these expectations are not dependent on the chosen time lag, and can be written as the product by a real function of time ($m_X(t_i)$ and $m_I(t_i)$ respectively) and the chosen time lag. We generalise the time lag by writing $\sigma/(\sigma+\delta_x+k\delta_\pi))\mathbb{E}_{t_i} x_{t_i+\Delta t_i} = m_X(t_i)\Delta t_i$ and $(k+\beta(\sigma+\delta_x))/(\sigma+\delta_x+k\delta_\pi)\mathbb{E}_{t_i} p_{t_i+\Delta t_i} = m_I(t_i)\Delta t_i$ respectively. When one moves to continuous time, $\Delta t_i \to dt$, and the real quantities $m_X(t)$ and $m_I(t)$ do not change. Therefore we write the products of expectation terms and constants as a continuous time drift ($m_X(t)dt$ and $m_I(t)dt$ respectively). We require that both $m_X(t)dt$ and $m_I(t)dt$ are bounded functions.

8. The random variables u_i and z_i^j are independent and normally distributed, with mean 0 and unit variance.

9. The random variables u_i and z_i^j are independent from their previous levels. For example, taken u_i, we write $\mathrm{Cov}(u_i, u_l) = \delta_{i,l}$. In this context $\delta_{i,l}$ is the Kronecker's delta sign, taking value 0 in all cases where $i \neq l$ and 1 when $i = l$.

10. If we take the standard normal random variable w_i, we introduce the random variable U_i, defined as $U_i = \sum_{k=1}^{i} w_k$, with $U_0 = 0$. Based on all the assumptions made, $U_i \sim N(0, i)$. By construction,

the process $\{U_i\}_{i=0,1,...}$ has zero mean, independent increments and $U_i - U_l \sim N(0, i-l),\ i > l$. The increment $U_{i+k} - U_{l+k}$ has the same normal distribution as the increment $U_i - U_l$, for each k.

11. By generalising the time lag, one considers that $\Delta U_{t_i} = U_{t_i} - U_{t_i - \Delta t_i} \sim N(0, \Delta t_i)$. Moving to continuous time one gets a Brownian motion. A similar discussion can be held for the m-dimensional random variable z_i, which becomes an m-dimensional Brownian motion with independent components. Shreve [109] gives full details of this procedure to build the Brownian motion starting from a discrete-time Gaussian process.

12. To compact notation, one introduces the *m+1*-dimensional (or alternatively n-dimensional) vectors, defined as $s^X_{t_i} = [\sigma, a^1_{t_i}, ..., a^M_{t_i}]$, and $s^I_{t_i} = [0, b^1_{t_i}, ..., b^M_{t_i}]$. The idea is to compact all the random terms to express them using a lighter notation.

13. To move to continuous time, one assumes that the processes $\{s^X_{t_i}\}_{t_i=0,1,...}$ and $\{s^I_{t_i}\}_{t_i=0,1,...}$are regular enough for the limits $s^X_{t_i} \rightarrow\ s_X(t)$ and $s^I_{t_i} \rightarrow\ s_I(t)$ when $dt \rightarrow 0$ to exist and for the total variance to be the same. These functions are both bounded across all components.

The system (2.1.3) can be rewritten in discrete time using a generic time step Δt_i as:

$$\begin{bmatrix} x_{t_i} \\ p_{t_i} \end{bmatrix} = \left(\begin{bmatrix} m_X(t_i) \\ m_I(t_i) \end{bmatrix} \Delta t_i + \begin{bmatrix} \sigma \\ 0 \end{bmatrix} (w_{t_i})\, \Delta t_i^{1/2} + \sum_{j=1}^{M} \begin{bmatrix} a_{i,j} \\ b_{i,j} \end{bmatrix} \left(z^j_{t_i} \right) \Delta t_i^{1/2} \right) \tag{2.34}$$

From the assumptions made above, the two above equations can be translated in continuous time as follows:

$$dX(t)/X(t) = m_X(t)dt\ + s_X(t) \cdot\ dW(t) \tag{2.35}$$

$$dI(t)/I(t) = m_I(t)dt\ + s_I(t) \cdot\ dW(t), \tag{2.36}$$

where $\{W(t)\}_{t \geqslant 0}$ is an n-dimensional $\mathbb{P}$-Brownian motion. Here the notation $\cdot$ is used to refer to the vector product. The functions $m_X(t)$, $m_I(t)$, $s_X(t)$, and $s_I(t)$ are regular enough for the above SDEs to have a unique strong solution.

At this stage we complement this model with some dynamics for the expectations of the drift: in fact, as shown in the following section, empirical evidence suggests that expectations themselves are subject to frequent revisions (as the economic agents process new information and data) and therefore are themselves stochastic. A possible expression for the dynamics of the expectations is the following:

$$dm_X(t) = a_X(t)dt + b_X(t) \cdot\ dW(t) \tag{2.37}$$

$$dm_I(t) = a_I(t)dt + b_I(t) \cdot dW(t). \tag{2.38}$$

where the scalar processes $\{a_X(t)\}_{t\geq 0}$ and $\{a_I(t)\}_{t\geq 0}$, and the $m+1$-dimensional processes $\{b_X(t)\}_{t\geq 0}$ and $\{b_I(t)\}_{t\geq 0}$ are deterministic processes regular enough for the SDEs to be integrated and to have a unique strong solution. To conclude, the above stochastic differential equations are derived from a well-established macroeconomic model. They are consistent with empirical evidence (as shown in the next section) and are used in the following chapter as a part of a wider setup to build a structural pricing continuous-time model for inflation derivatives, based on macroeconomic assumptions.

2.3.1 Testing the dynamics against empirical evidence

In this section we show some economic time series to confirm that, over time, the growth rate of real GDP and of the price index are stationary processes that show some randomness. The aim is not to propose any econometric analysis but to show some graphs to develop intuition on the behaviour of the economic variables we want to model. The levels of real GDP and price index are growing in an exponential fashion over time: these two observations confirm that the choice of a Brownian motion with time-changing coefficients and stochastic drift is a sensible option.

Evidence is shown for the US and the UK economy, and similar results hold for most economies. All data are sourced from Bloomberg.

Fact 1 - Over time both price indexes and real GDP have grown steadily, as shown by the first four figures.

Fact 2 - Over time their growth rate has been subject to some randomness, as shown by the fifth to the eighth figure of this section.

Further, we show some evidence of expectations (or forecast) of UK GDP growth rate (compiled by Bloomberg) and of the US inflation rate (compiled by the University of Michigan): both series show that the expectations themselves are stochastic, which suggests that the assumption of assuming stochastic expectations is sensible and consistent with empirical evidence.

Fact 3 - Growth rate and inflation expectations are subject to randomness: this is shown by the last two figures of this section.

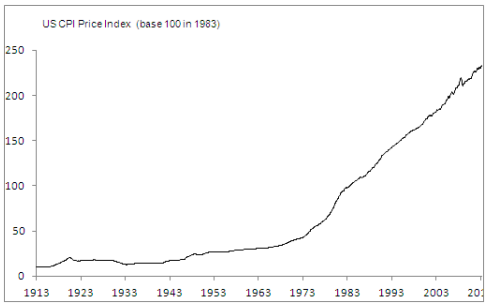


Figure 2.1: Time series of US CPI price index

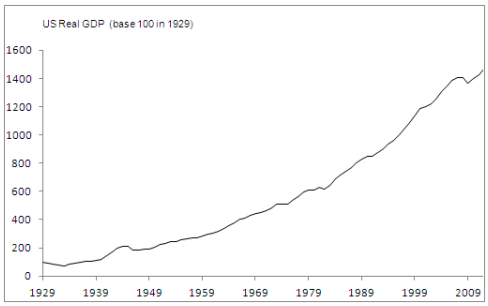


Figure 2.2: Time series of US real GDP.

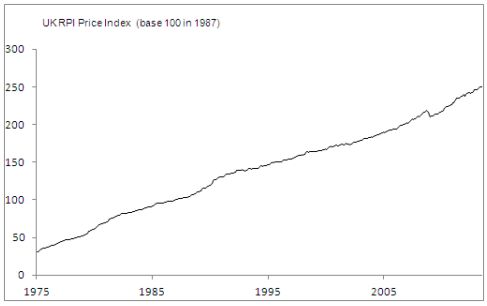


Figure 2.3: Time series of UK RPI price index.

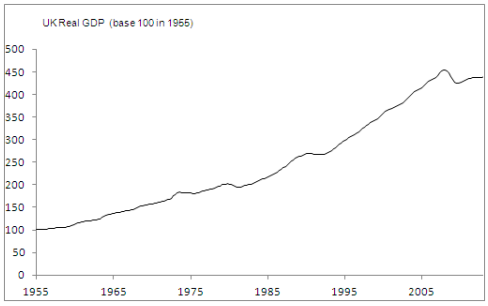


Figure 2.4: Time series of UK real GDP.

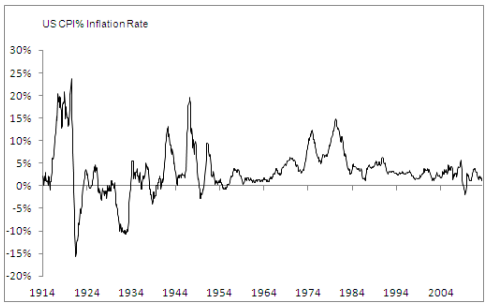


Figure 2.5: Time series of US CPI inflation.

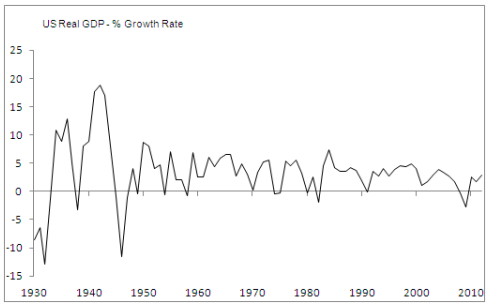


Figure 2.6: Time series of US real GDP growth rate.

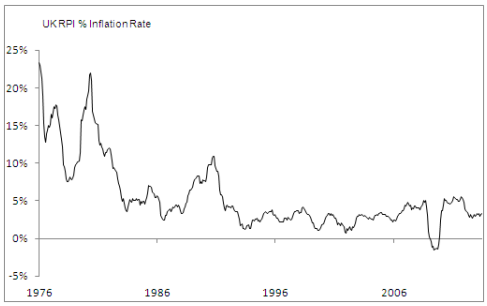


Figure 2.7: Time series of UK RPI inflation.

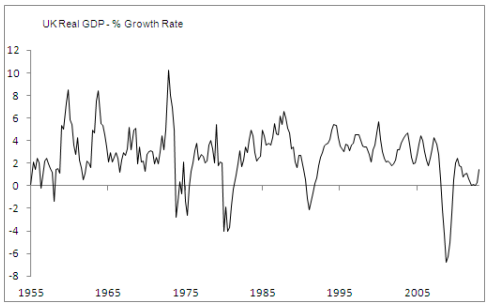


Figure 2.8: Time series of UK real GDP growth rate.

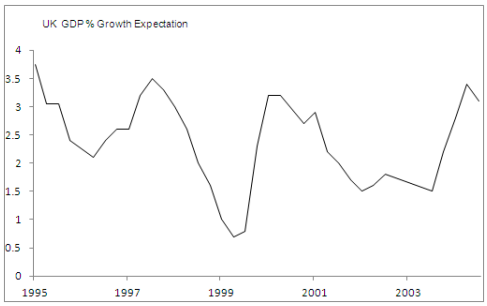


Figure 2.9: Time series of UK real GDP growth expectations (survey by Bloomberg).

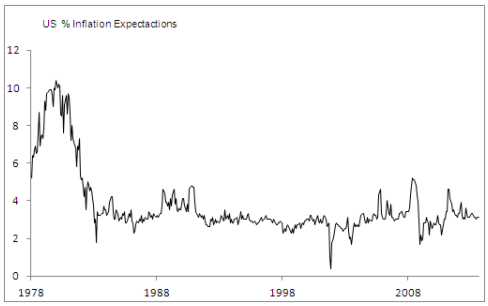


Figure 2.10: Time series of US inflation expectations (survey by University of Michigan).

2.3.2 Comparing the DSGE model with the continuous-time model

This section shows that the discrete-time DSGE model and the continuous-time model we propose can deliver similar distributions for the main economic variables if one parametrizes them in a consistent way. Therefore in the following chapter we choose the continuous-time model to develop the theory as it is superior compared to the DSGE model as far as its analytical tractability is concerned. In fact, stochastic calculus in continuous time is an extremely powerful tool and is somewhat more developed compared to the discrete-time case. For example, the continuous time machinery lets us obtain a closed form for the year-on-year convexity adjustment in chapter 3.

Further, in the following chapter we show that we find closed form expressions in the continuous-time model for both the nominal and inflation term structure, for both nominal rates and inflation options, and for year-on-year inflation forwards, without having to resort to the linearisations and approximations used earlier in this chapter when dealing with the discrete-time DSGE model.

In order to obtain similar distributions for the most relevant financial quantities, one applies a moment-matching technique across both models. We assume that all parameters in the continuous-time model are expressed as right-continuous step functions and that the dimensionality of the Brownian motion is 3. The use of step functions in the continuous time model imposes no practical constraint, as this model is calibrated to a finite set of market observable, as discussed in chapter 4. We focus our attention on second order moments, as the first order moments are straightforward to match. All analysis is done to match the distributions over the first year, with the subsequent years following exactly the same algorithm.

Inflation rate. In the discrete-time DSGE model, the variance of the inflation rate (formula (2.17)) is:

$$\mathrm{Var}(p_i) = (K_2)^2(\sigma^2\mathrm{Var}(u_i) + \mathrm{Var}(v_i)) + \mathrm{Var}(z_i).$$

In the next chapter (see 3.51 on page 95) we show the diffusion term of the inflation rate (approximated

by the ratio $dI(t)/I(t)$):

$$[b_I(t)(T-t)+s_I(t)].$$

This implies that the total variance over the first year ($t=0$ and $T=1$) is:

$$\sum_{i=1}^{3}[b_I(0)+s_I(0)]^2.$$

In discrete time we have $t_{i-1}=0$ and $t_i=1$. Therefore it makes sense to match the two conditions by requesting that:

$$\mathrm{Var}(p_i)/(t_i-t_{i-1})=[(K_2)^2(\sigma^2\mathrm{Var}(u_i)+\mathrm{Var}(v_i))+\mathrm{Var}(z_i)]/(t_i-t_{i-1})=\sum_{i=1}^{3}[b_I(0)+s_I(0)]^2. \quad (2.39)$$

Short rate. A similar method can be applied to the variance of the nominal short rate (formula (2.24)), that in the DSGE set-up is calculated as:

$$\mathrm{Var}(n_{i+1})=(\bar{n})^2(\delta^T K)^2\sigma^2\mathrm{Var}(u_i)+(\bar{n})^2(1-\delta^T K)^2\mathrm{Var}(v_i)+(\bar{n})^2\delta_\pi^2\mathrm{Var}(z_i).$$

Because the short-term nominal rate level at time t_{i+1} is independent from its level at the previous time t_i (this follows because the nominal rate is a linear combination of the output gap and inflation, both of which are driven by Gaussian processes that are independent from their own realisations over time), we write the variance of the change in the nominal rate as:

$$\mathrm{Var}(n_{i+1}-n_i)=\mathrm{Var}(n_{i+1})+\mathrm{Var}(n_i)=$$

$$(\bar{n})^2[(\delta^T K)^2\sigma^2(\mathrm{Var}(u_i)+\mathrm{Var}(u_{i-1}))+(1-\delta^T K)^2(\mathrm{Var}(v_i)+\mathrm{Var}(v_{i-1}))+\delta_\pi^2(\mathrm{Var}(z_i)+\mathrm{Var}(z_{i-1}))].$$

In the next chapter (see 3.28 on page 82) we show that the diffusion term of the nominal short rate differential $dn(t)$ is:

$$-\frac{h_x b_X(t)+h_p b_I(t)}{\zeta(t)}.$$

The matching condition is:

$$[(\bar{n})^2[(\delta^T K)^2\sigma^2(\mathrm{Var}(u_i)+\mathrm{Var}(u_{i-1}))+(1-\delta^T K)^2(\mathrm{Var}(v_i)+\mathrm{Var}(v_{i-1}))+\delta_\pi^2(\mathrm{Var}(z_i)+\mathrm{Var}(z_{i-1}))]]/(t_i-t_{i-1})= \quad (2.40)$$

$$= -\frac{h_x b_X(0) + h_p b_I(0)}{\zeta(0)}.$$

For the moment we only assume that $\zeta(t)$ is a positive calibration real scalar function, and that the real positive parameters h_x and h_p are taken exogenously. In fact, as shown in the next chapter, they have a precise financial meaning. This said, the only purpose of this exercise at this stage is to show that some statistical properties in two different models can be matched.

Covariance between nominal short rate and inflation rate. The covariance in the DSGE model is:

$$\mathrm{Cov}(p_i, n_{i+1}) = \bar{n}K_2(\delta^T K)\sigma^2 \mathrm{Var}(u_i) + (\bar{n}K_2(1-\delta^T K))\mathrm{Var}(v_i) + \bar{n}\delta_\pi \mathrm{Var}(z_i).$$

Because the short-term nominal rate level at time t_i is independent from the inflation level at the same time t_i (as discussed above), this covariance can be interpreted also as

$$\mathrm{Cov}(p_i, n_{i+1}) = \mathrm{Cov}(p_i, n_{i+1} - n_i).$$

The correlation is calculated as follows:

$$\mathrm{Corr}(p_i, n_{i+1} - n_i) = \frac{\bar{n}K_2(\delta^T K)\sigma^2 \mathrm{Var}(u_i) + (\bar{n}K_2(1-\delta^T K))\mathrm{Var}(v_i) + \bar{n}\delta_\pi \mathrm{Var}(z_i)}{(K_2)^2(\sigma^2 \mathrm{Var}(u_i) + \mathrm{Var}(v_i)) + \mathrm{Var}(z_i))^{1/2}(\mathrm{Var}(n_{i+1}) + \mathrm{Var}(n_i))^{1/2}}.$$

By doing some basic calculations, and by taking into account results 3.28 on page 82 and 3.51 on page 95, and result 2 on page 96, we show that the instantaneous correlation of the nominal short rate change and the inflation rate between times t and T is:

$$-[b_I(t)(T-t) + s_I(t)] \cdot [\zeta(t)^{-1}(h_x b_X(t) + h_p b_I(t))].$$

The matching condition is:

$$\mathrm{Cov}(p_i, n_{i+1} - n_i)/(t_i - t_{i-1}) = -[b_I(0)(1-0) + s_I(0)] \cdot [\zeta(0)^{-1}(h_x b_X(0) + h_p b_I(0))]. \tag{2.41}$$

Example. To show the application of the above methodology, we simulate over the first year the GDP growth rate, the inflation rate, and the short nominal interest rate over 5,000 Monte Carlo trials. The parametrisation proposed below has no specific financial meaning and is provided only as an example. We assume that the shocks in the DSGE model are normally distributed with some variances that are

calibrated below.

The assumptions made on the economy and the financial market are the following:

1. Market agents are risk-neutral. This translates in a σ parameter of 0 in the DSGE model, and in zero market prices of risk, both in the discrete-time and in the continuous-time Gaussian processes.

2. The inflation rate is expected to be 3% in the first year, with a standard deviation of 1.1%. This standard deviation can be either an empirical estimate or can be inferred from traded derivatives markets. The source of this standard deviation is not relevant in this exercise.

3. The output gap is expected to be -2% in the first year.

4. The potential growth rate of the economy is 2%.

5. The equilibrium level of the short-term nominal rate is 4%.

6. The standard deviation of the nominal short rate is 0.45%. The short rate is currently at 2.1%.

7. The central bank is attaching three times more importance to fighting inflation than to stimulating growth.

8. The correlation between the nominal rate change and the inflation rate is positive (given that the central bank is targeting inflation in a very aggressive way), and is 65%.

The time index at 1 means that the parameter is relative to the first year, that is the point in time that we are simulating. The parametrisation we choose for the DSGE model for this example is the following:

Parameter	**Value**	**Comment**
σ	0	Agents are risk-neutral
k	0.01	Prices are sticky
δ_π	3	The central bank fights inflation aggressively
δ_x	1	The central bank is not targeting growth aggressively
β	0.95	Standard subjective discount factor
$\mathrm{Var}(u_1)$	0.01	
$\mathrm{Var}(v_1)$	0.01	
$\mathrm{Var}(z_1)$	0.0001	

Moving to the continuous time model, we propose the following parametrisation based on the same economic assumptions proposed above, applying the moment matching conditions stated above. The time indexes are either 0 (initial condition) or 1 (final condition). Between these two times we think to the parametric functions like $a_X(t)$, $a_I(t)$, $b_X(t)$, $b_I(t)$, $s_I(t)$, and $s_X(t)$ as right-continuous step functions: because the dimensionality of the Brownian motion is 3, there are three values for the volatility vectors specified below. The Monte Carlo simulation has been run in one time step equal to one year. Finally, some of the model parameters that appear in the below table have not been explained in the continuous-time construction reviewed above, but are introduced in the following chapter.

The aim of this section is to show that the two models provide results that are broadly in line: the full explanation of the continuous-time model and of the meaning of its parameters is given in the following chapter. We stress that the example proposed above is not a general algorithm to move from a discrete-time DSGE model to a continuous-time equivalent.

Parameter	**Value**	**Value (2)**	**Value (3)**	**Comment**
h_x	1			Sensitivity of central bank to growth
h_p	3			Sensitivity of central bank to inflation
$\bar{x}$	2%			Central bank target growth
$\bar{p}$	2%			Central bank target inflation
$\zeta(0)$	2.015			Please refer to next chapter for this parameter
$a_X(0)$	0.5%			
$a_I(0)$	-1.5%			
$\mu_X(0)$	-0.5%			
$\mu_I(0)$	4.5%			
$s_X(0)$	0.01	0.01	0.01	
$s_I(0)$	0.01	0.01	0.01	
$b_X(0)$	0.004	-0.01	-0.00005	
$b_I(0)$	-0.001	0.0005	-0.0005	

Results Here we show a table comparing the target levels, the results in the DSGE model, and the results in the continuous-time model for the short rate change and the inflation rate in the first year.

Statistic	Target	DSGE simulation	Continuous-time simulation
$\mathbb{E}[n_{i+1} - n_i]$	2%	2.01%	1.98%
$\mathbb{E}[p_i]$	3%	3.01%	2.97%
$\text{StDev}[n_{i+1} - n_i]$	0.45%	0.42%	0.44%
$\text{StDev}[p_i]$	1.1%	1.08%	1.07%
$\text{Corr}[n_{i+1} - n_i, p_i]$	65%	64.16%	69.07%

Here we show the scatter-plot of the two variables together. This shows that both the marginal distributions and the joint distribution are matched in a satisfactory way.

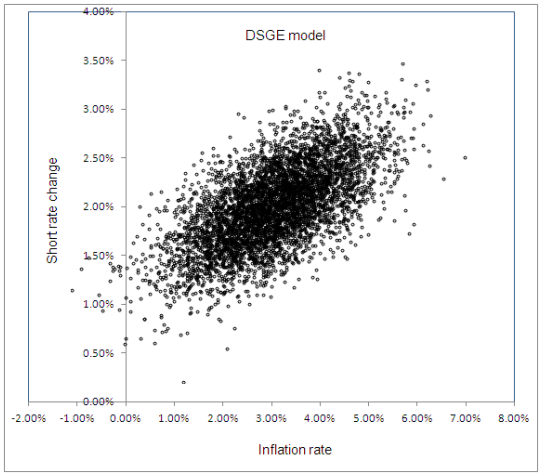


Figure 2.11: Scatter plot of 5,000 Monte Carlo simulations of the DSGE model. Change in nominal short rate and inflation rate over one year.

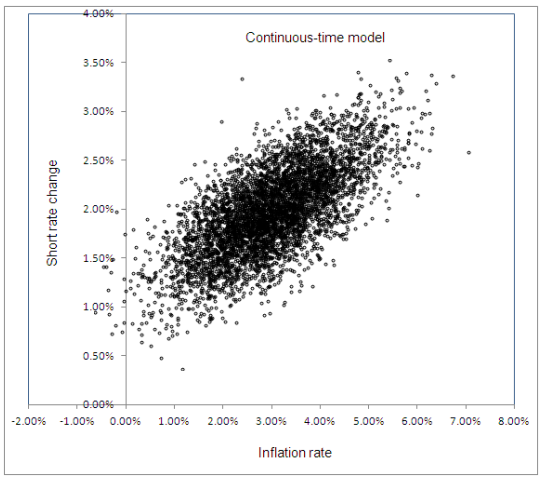


Figure 2.12: Scatter plot of 5,000 Monte Carlo simulations of the continuous-time model. Change in nominal short rate and inflation rate over one year.

Chapter 3

Inflation derivatives pricing with a central bank reaction function

In this chapter we propose a continuous-time model to price inflation-linked and fixed-income derivatives by use of a model that explicitly takes into account the economic dynamics and the central bank behaviour. As it happened in the discrete-time DSGE model analysed in the previous chapter, the co-movement of interest rates and inflation is not specified exogenously but rather is the result of central bank policy.

To achieve this, we make some standard assumptions regarding the structure of the financial market (absence of arbitrage) and model the relative changes of both real GDP (Gross Domestic Product) and Price Index using Brownian motions with stochastic drifts. Furthermore the central bank trades nominal bonds to change the money supply in the economy, to keep the growth rate and inflation around some pre-specified targets (see for example Walsh [115]). These bond trades have an impact on the nominal bond prices, and on the term structure of interest rates. Normally inflation-linked pricing models model the co-movement of inflation and nominal interest rates exogenously, without specifying the economic rationale behind this: we think that bridging the gap between economics and finance can be beneficial for both disciplines.[1]

The advantages of this approach are manifold. Firstly, the dynamics assumed in this model appear to be consistent with the behaviour of central banks in recent years, when significant purchases of bonds (the so-called "quantitative easing") have been made since short-term interest rates have reached (and in some cases crossed) the zero lower bound in many developed economies. One can ask why it is important

[1] Another example of inflation-linked pricing model based on sound economic assumptions can be found in Hughston & Macrina [71], Hughston & Macrina [72], and Alexander [2]: the spirit of these papers has been a source of inspiration for the current model, as they use a microeconomic approach based on Sidrauski [110] to determine the continuous-time dynamics of the price index (this said, they do not model the central bank reaction function).

that a pricing model generates asset prices using realistic dynamics: after all, this could be irrelevant once the model has calibrated to a set of market observables. We think that such model would fail to minimise hedging profit and loss volatility if a dynamic hedging simulation was run and would make realistic stress testing difficult.[2] Further, the same model can be used for pricing and risk applications.

Secondly, this approach does not rely on the so-called "Forex Analogy", which assumes the existence of the "real" economy (see Hughston [70] and chapter 2). The Forex analogy has roots in the economic theory (see Fisher [55]). We use this only as a calculation device in one occasion, and the quantities we model are all market observables: this makes this model different from the Jarrow-Yildirim model (see Jarrow & Yildirim [79] or Brody, Crosby & Li [29]). The main advantage is that the model parameters are calibrated in a transparent way to liquid market observables (nominal bonds, inflation swaps, nominal and inflation caps and floors), as opposed to using and estimating a real rate volatility which is hardly observable in the market. In practice, one avoids taking costly uncertain-parameter reserves or valuation adjustments, as required by accounting principles. Examples of models that do not rely on the Forex Analogy can be found in Dodgson & Kainth [50], in Mercurio [95] or in Brigo & Mercurio [22]: however all these models are not based on macroeconomic foundations.

Thirdly, although the model is complex and takes into account many market features, we show that the dynamics of the short-term nominal rate can be reconciled with a well-established short interest rate model (the generalised Hull-White model, which is a time-varying parameters version of the Ornstein-Uhlenbeck process). This provides both an elegant link with the established theory and some closed forms of nominal rates derivatives that are useful for the calibration. Interestingly, we find that the request that we make on the function $Z(T)$, that models the impact of monetary policy on bond prices, has consistent implications on the mean reversion property of the short rate and the modelling of intra-curve correlations.

Fourthly, zero-coupon and year-on-year inflation forwards and option prices are derived in closed forms: the model remains tractable even if is based on realistic assumptions.

Fifthly, the extension of this model to the open economy (and so to cross-currency or quanto inflation derivatives) is straightforward.

Sixthly, the calibration of this model is computationally not intensive, which allows fast pricing of all kind of trades, from inflation options, to year-on-year caps and floors, to more complex inflation structures as LPI (Limited Price Index). The main reason for this computational simplicity is that we propose a separable calibration strategy. This point is fully developed in the following chapter.

The reader interested in the inflation derivatives market can refer to some marketing notes edited

[2] Dynamic hedging simulations can be used to assess the quality of a model: the idea is to generate some "real world" dynamics and assess how the delta hedging done through a model performs, in terms of reducing hedging profit and loss volatility ("slippages"). Examples of these techniques can be found, for example, in Rebonato [103].

by investment banks, like Barclays [7] or Lehman Brothers [83], or alternatively Deacon, Derry & Mirfendereski [48], Benaben [10], Campbell, Shiller & Viceira [35], McGrath & Windle [94], and Jäckel & Bonneton [77].

3.1 Inflation model assumptions

3.1.1 Probabilistic set-up

1. The model is set in continuous time t. Time is a positive real number and is expressed in years. From this point the notation for the continuous-time variable y at time t is $y(t)$.

2. Randomness is modelled via an n-dimensional $\mathbb{P}$-Brownian motion $\{W(t)\}_{t\geq 0}$. The probability measure $\mathbb{P}$ is fully defined in the next point. The n components of the Brownian motion $W(t)$ are independent.

3. We work with the probability triplet $\{\Omega, \mathcal{F}, \mathbb{P}\}$ equipped with the natural filtration $\{\mathcal{F}_t\}_{t\geq 0}$ generated by the Brownian motion $\{W(t)\}_{t\geq 0}$. All filtration-related concepts are defined with respect to this filtration. In particular $\mathbb{P}$ is the real-world ("physical") probability measure. If no probability measure is specified, the expectation is taken with respect to the real-world measure ($\mathbb{P}$).

4. Derivatives pricing can be carried out in the $\mathbb{P}$ measure via the pricing kernel (defined both below and in chapter 1), or in the risk-neutral measure $\mathbb{Q}$ (defined as the pricing measure that uses the money market account $B(t)$ as numeraire), or finally in the T-forward measure $\mathbb{Q}^T$, defined as the pricing measure that uses the bond price $P(t,T)$ as numeraire. The bonds $P(t,T)$ and the money market account $B(t)$ are defined in detail in the next section. Expectations of a payoff Π taken under the generic measure $\mathbb{M}$ given the information available at time t are denoted as $\mathbb{E}^{\mathbb{M}}[\Pi|\mathcal{F}_t]$ or alternatively $\mathbb{E}_t^{\mathbb{M}}[\Pi]$.

3.1.2 Financial instruments

All instruments listed below and their related quantities are modelled as $\{\mathcal{F}_t\}_{t\geq 0}$-adapted stochastic processes and are regular enough to ensure the existence of the expectations introduced later. The list of instruments is not exhaustive but only contains the ones needed to build the model.

1. **Nominal zero-coupon bonds**, that pay with certainty (i.e. risk-free) one unit of currency at time T, have price $P(t,T)$ at time t. There exists a continuum of bond prices, i.e. $T \in [t, +\infty)$.[3]

[3] When calibrating the model to market observables, this assumption is relaxed because only a finite set of bond maturities are quoted on the market.

From the nominal bond prices one derives all kind of rates, for example instantaneous forward rates $f(t,T) = -\partial \log(P(t,T))/\partial T$ and the short rate $n(t) = f(t,t)$. See Brigo & Mercurio [22] for further details.

2. **Money market account** $B(t)$, with dynamics $dB(t) = n(t)B(t)dt$, $B(0) = 1$.

3. **Price Index** $I(t)$, which is a positive stochastic process that reflects the price level of the economy. Its dynamics are specified later in the economy set-up. This index is de-seasonalised.

4. **Zero-coupon Inflation Index Swaps** $ZCIIS(t,T)$: the inflation leg pays $I(T)/I(t)-1$, while the fixed leg pays $(1 + K(t,T))^{T-t} - 1$. Both payments happen at maturity T and there is no time lag. The inflation breakeven $K(t,T)$ is agreed at time t: there exists a continuum of inflation breakevens, i.e. $T \in [t, +\infty)$.[4] In a zero-coupon inflation swap the strike $K(t,T)$ is such that at inception the expected value at maturity of the swap is zero: $\mathbb{E}_t^{\mathbb{Q}}[((I(T)/I(t) - (1+K(t,T))^{T-t})B(t)/B(T))] = 0$.

5. **Inflation bonds** $P^I(t,T)$, which pay at time T the level of the price index $I(T)$, with no time lag. There exists a continuum of inflation bond prices, i.e. $T \in [t, +\infty)$. They are also known as "linkers". Because we are working in a market without liquidity concerns, the inflation dynamics implied by the inflation bond prices are the same as the ones implied by the inflation swap market. The zero-coupon linker price is $P^I(t,T) = \mathbb{E}_t^{\mathbb{Q}}[I(T)B(t)/B(T)]$. Normally these bonds have an implicit deflation floor: the bond holder does not pay the issuer in case of deflation. This feature is ignored in the model we present here: this has no practical consequences on the consistency of our approach.

3.1.3 Financial market

The assumptions regarding the financial market are standard:

1. There is no credit risk in the economy.

2. The financial market is arbitrage-free. A thorough treatment of absence of arbitrage and its implications can be found in Björk [33], Cochrane [42], Hughston & Rafailidis [73], or Duffie [52].

3. Assuming that we use the money market account $B(t)$ as numeraire, we are working in the risk-neutral measure $\mathbb{Q}$. This implies that the bond price $\mathbb{Q}$-dynamics are given by

$$dP(t,T)/P(t,T) = n(t)dt \ + \sigma_P(t,T) \cdot dW^{\mathbb{Q}}(t) \tag{3.1}$$

[4] When calibrating the model to market observables, this assumption is relaxed because only a finite amount of inflation swaps maturities are quoted on the market.

where the bond volatility $\sigma_P(t,T)$ is an n-dimensional deterministic process.[5] These volatilities are referred to as "model" volatilities in the calibration section, as opposed to volatilities implied by market prices of options. The form of these bond volatilities is left general at this point, but at a later stage in this chapter we fully characterise them in terms of other model parameters.

4. The Radon-Nikodym derivative $L(t) = d\mathbb{Q}/d\mathbb{P}|_{\mathcal{F}_t}$ has the dynamics: $dL(t) = -\ L(t)\lambda(t) \cdot dW^{\mathbb{P}}(t)$, where $\lambda(t)$ is an n-dimensional deterministic process. The process $\{\lambda(t)\}_{t\geq 0}$ is called "market price of risk".

5. The pricing kernel $\psi(t)$ defined as $\psi(t) = L(t)/B(t)$ has dynamics: $d\psi(t)/\psi(t) = -n(t)dt\ -\ \lambda(t) \cdot dW^{\mathbb{P}}(t)$. The pricing kernel has many useful properties: for our purposes we remember here that $P(t,T) = \mathbb{E}_t^{\mathbb{P}}[\psi(T)/\psi(t)]$. Further analysis of the pricing kernel properties can be found in chapter 1, or in Constantinides [44], Hughston [69], Leippold & Wu [86], Shefrin [108], and Rogers [104].

6. The real bond, which is not an asset traded on the market, is defined as the ratio between the inflation bond and the current price index level: $P^R(t,T) = P^I(t,T)/I(t)$.[6] Both $P^I(t,T)$ and $I(t)$ have been defined previously. Using the same logic as above, from the real bond prices one extracts a real term structure of interest rates: in particular we define the real short rate $r(t) = f^R(t,t)$, where $f^R(t,T) = -\partial \log(P^R(t,T))/\partial T$. The process $\{r(t)\}_{t\geq 0}$ is $\{\mathcal{F}_t\}_{t\geq 0}$-adapted and can be used to define the real money market account $B^R(t)$, with dynamics $dB^R(t) = r(t)B^R(t)dt$: $B^R(t)$ is locally riskless in the real risk-neutral measure $\mathbb{Q}^R$ (as $B(t)$ is in the $\mathbb{Q}$ measure). We also define the real pricing kernel $\psi^R(t) = I(t)\psi(t)$: similarly we show that $P^R(t,T) = \mathbb{E}_t^{\mathbb{P}}[\psi^R(T)/\psi^R(t)]$. Furthermore, we introduce the $\{\mathcal{F}_t\}_{t\geq 0}$-adapted process $\left\{\lambda^R(t)\right\}_{t\geq 0}$, called "real market price of risk" and obtain the dynamics: $d\psi^R(t) = -r(t)\psi^R(t)dt\ -\ \psi^R(t)\lambda^R(t) \cdot dW^{\mathbb{P}}(t)$.

7. The definition of the real bond, real rates and real pricing kernel is sufficient to define another economy, labelled "real" economy: this is introduced in chapter 1. This is the cornerstone of the so-called "Forex Analogy" (see Hughston [70], Hughston [69], or Brigo & Mercurio [22]): because we write $I(t) = \psi^R(t)/\psi(t)$, we see the price index $I(t)$ as the exchange rate between the real and the nominal economy. This allows one to obtain, in analogy with the FX spot rate drift, the risk-neutral drift for the price index: $\mathbb{E}_t^{\mathbb{Q}}[I(t+dt) - I(t)] = (n(t) - r(t))I(t)dt$.

[5] Given two n-dimensional vectors a, b, with components $a_1, ..., a_n$ and $b_1, ..., b_n$ respectively, the notation $a \cdot b$ is equivalent to $\sum_{i=1}^n a_i b_i$. This notation is used extensively in this work. Under no circumstances this notation has to be confused with a Stratonovich integral.

[6] It is worth stressing that, although the model proposed in this paper does not require the concept of real bond and real rates, these are often found in the literature: therefore it is useful to show how these quantities can be recovered in the present set-up. It is used in this chapter as a calculation tool for the year-on-year inflation forward.

3.1.4 Economy dynamics and central bank role

We make some assumptions regarding the economy. These assumptions let us make a realistic description of the economy based on a continuous-time model inspired by a widely-used discrete-time macroeconomic model, called "Dynamic Stochastic General Equilibrium" model (referred to as DSGE—presented and discussed in the previous chapter).

1. At time t, the economy is described by three positive quantities $X(t), I(t)$, and $M(t)$, that represent the real output of the economy, the price level in the economy, and the money supply respectively. [7] The real output is an alternative expression for the real Gross Domestic Product, also referred to as real GDP. Money supply is defined as the total amount of cash available in the economy. The processes $\{X(t)\}_{t\geq 0}$, $\{I(t)\}_{t\geq 0}$, and $\{M(t)\}_{t\geq 0}$ are positive $\{\mathcal{F}_t\}_{t\geq 0}$-adapted processes. The economy is closed.

2. The $\mathbb{P}$-dynamics of instantaneous output and price index are defined as follows:

$$dX(t) = X(t)[m_X(t)dt \ + s_X(t) \cdot \ dW^{\mathbb{P}}(t)] \tag{3.2}$$

$$dI(t) = I(t)[m_I(t)dt \ + s_I(t) \cdot \ dW^{\mathbb{P}}(t)] \tag{3.3}$$

   where $m_X(t)$ and $m_I(t)$ are one-dimensional stochastic $\{\mathcal{F}_t\}_{t\geq 0}$-adapted processes whose dynamics are to be defined below, and $s_X(t)$ and $s_I(t)$ are n-dimensional deterministic processes. These volatilities are referred to as "model" volatilities in the calibration section, as opposed to volatilities implied by market prices of options. In particular we showed in the previous chapter that the above two equations for the growth and inflation rate can be derived from a well-specified macroeconomic model and are consistent with historic data.

3. The dynamics of the expectations are modelled using the SDEs:

$$dm_X(t) = a_X(t)dt + b_X(t) \cdot \ dW^{\mathbb{P}}(t) \tag{3.4}$$

$$dm_I(t) = a_I(t)dt + b_I(t) \cdot \ dW^{\mathbb{P}}(t) \tag{3.5}$$

   where the processes $a_X(t)$ and $a_I(t)$ are one-dimensional deterministic processes and $b_X(t)$ and $b_I(t)$ are n-dimensional deterministic processes. In the previous chapter we have provided evidence that shows that growth and inflation expectations are themselves stochastic, which supports

[7] For the purposes of this thesis it is not worthwhile analysing in a different way the different monetary aggregates like M1, M2, and M3.

our modelling proposal stated above (we have used a very general SDE driven by a Brownian differential).

4. We assume that the central bank is the only institution responsible for money supply. The central bank uses the money supply as a policy tool and tries to keep the economy close to a target annual growth rate $\bar{x}$ and to a target annual inflation rate $\bar{p}$. The targets $\bar{x}$ and $\bar{p}$ are constant real numbers. According to standard macroeconomic theory, an increase in money supply can increase both the price level and the output: the central bank can attach more importance to the growth target or to price stability. The relative importance of these two goals is modelled with the two real positive constants h_x and h_p. To summarise the above assumptions, we propose an original model to explain the central bank policy using the $\mathbb{P}$-dynamics

$$dM(t)/M(t) = -h_p(dI(t)/I(t) - \bar{p}dt) - h_x(dX(t)/X(t) - \bar{x}dt) + s_M(t) \cdot \, dW^{\mathbb{P}}(t). \tag{3.6}$$

This function is also known as "central bank reaction function".[8] Here $s_M(t)$ is an n-dimensional deterministic process that measures the uncertainty around the central bank policy. These volatilities are referred to as "model" volatilities in the calibration section, as opposed to volatilities implied by market prices of options. The above equation can be read as follows: modulo some uncertainty (modelled by the term $s_M(t) \cdot \, dW(t)$), the central bank reduces the money supply (both $-h_p$ and $-h_x$ are negative real numbers [9]) when inflation or output growth are above their targets. It should be noted that the above specification for the central bank policy is similar to the Taylor rule (see Walsh [115], Woodford [116], Taylor [112], Clarida, Dali & Gertler [41], or Clarida, Dali & Gertler [40]. The rule has also been discussed in the second chapter). Because the Taylor rule assumes that the short-term interest rate is the monetary policy tool (as opposed to the money supply), the Taylor rule can lead to negative policy rates, while in a low rates environment central banks tend to use open market operations as policy tools.[10]

5. The central bank changes the money supply by trading in the secondary bond market, which has some feedback effects on bond prices. These effects are known by market participants and are

[8] The above expression for the reaction function attaches more importance to intuition than to formal correctness: if one wanted to write an expression containing only stochastic differentials and not involving ratios of stochastic differentials (like $dM(t)/M(t)$, $dI(t)/I(t)$, or $dX(t)/X(t)$) one can define the reaction function in logarithmic differential terms and adjust the equilibrium levels from $\bar{p}$ and $\bar{x}$ to $\bar{p}^*$ and $\bar{x}^*$ respectively for the change in drifts:

$$d\log M(t) = [-h_p(d\log I(t) - \bar{p}^* dt) - h_x(d\log X(t) - \bar{x}^* dt) + s_M(t) \cdot \, dW^{\mathbb{P}}(t)]. \tag{3.7}$$

In the rest of the paper we will not be using the above expression and will develop our theory using (3.6).

[9] Because the so-called "quantitative easing" has been implemented only in the last few years by some central banks, it is not possible to provide data-based estimates of these parameters for the moment.

[10] This consideration is even more relevant in the current low rates environment, where the main option left to the central banks in the USA, UK, Japan and the Euro area is to purchase bonds to stimulate and reflate the economy (so-called "quantitative easing").

priced in the market. The central bank can also target some specific sectors of the yield curve, for example it can decide to sell short maturity bonds and buy longer maturities bonds to make the curve flatter while not inflating its balance sheet.[11] We assume that the relative increase in the money supply has a linear relationship with the relative increase in the bond prices, weighted for each maturity T by a weight function $Z(T)$. These effects are priced by the market and are modelled by the $\mathbb{Q}$-dynamics for the money supply:

$$dM(t)/M(t) = \gamma dt + \int_t^{t+\Omega} Z(T)[dP(t,T)/P(t,T)]dT + s_L(t) \cdot dW^{\mathbb{Q}}(t). \tag{3.8}$$

Here γ is a real constant that models the natural growth of the money supply; $Z(T)$ is a continous, real, positive and increasing deterministic scalar function of the bond maturity T. The request that the function $Z(T)$ is always positive comes from economic considerations: if bond prices increase, nominal rates decrease, which is equivalent to saying that the money supply goes up. In this framework the interest rate itself is not the policy tool, because all monetary policy is modelled via the money supply $M(t)$. The integral in the above expression is a deterministic one, given that at time t the quantity $dP(t,T)/P(t,T)$ is known for all maturities T: the integral in the above expression has to be regarded as a way to weight the impact of relative changes in the bond prices across the different maturities $T \in [t, t+\Omega]$ of the term structure.[12] The real positive constant $\Omega > 0$ represents the time horizon used by the central bank to purchase or sell nominal bonds in order to influence the money supply $M(t)$. For example, if the central bank is buying or selling bonds up to the 30 years maturity, one sets the parameter Ω to 30. We assume that the integral in the above expression always exists and is a finite quantity.

Uncertainty around this relationship is captured by the stochastic differential, multiplied by a liquidity volatility deterministic n-dimensional process $\{s_L(t)\}_{t\geq 0}$.

6. We finally require the following relationship to hold:

$$h_p s_I(t) + h_x s_X(t) - s_M(t) = 0. \tag{3.10}$$

This condition is equivalent to asking that the central bank reaction function is locally riskless. It should be noted that the reaction function is still stochastic as the drifts are stochastic, and

[11] The "operation twist" implemented by the FED in 2011 is a good example.

[12] An observation similar to the comment made on the reaction function can be done at this stage: the liquidity relationship can be rewritten as:

$$d\log M(t) = [\gamma dt + \int_t^{t+\Omega} Z(T)[d\log P(t,T) + 1/2[\sigma_P(t,T)\cdot\sigma_P(t,T)]]dT + s_L(t)\cdot dW^{\mathbb{Q}}(t)]. \tag{3.9}$$

We will not be using the above expression in the model theory development, but rather 3.8.

that the liquidity relationship is stochastic as the short rate is stochastic. The short rate is the risk-neutral drift of the bond prices dynamics. We also note that the above condition, together with condition (3.22) introduced later, ensures that the diffusion term for the $\mathbb{P}$-dynamics (3.6) is the same as the diffusion term for the $\mathbb{Q}$-dynamics (3.8), which satisfies Girsanov's theorem.

3.2 CTCB Model construction

We build the pricing model: this is referred to as "Continuous Time Central Bank" (CTCB) model in the following sections. The assumptions made so far can be regarded as standard no-arbitrage assumptions in the financial market, in conjunction with some reasonable assumptions on growth and inflation rate (modelled as Brownian motions with some stochastic drifts — historically GDP and price levels have shown an upward trend with some noise). Furthermore, the central bank trades nominal bonds to keep the economy around some target levels, and this has some (wanted) effects on bond prices, and hence on the yield curve.

The model construction that follows puts together the financial market and macroeconomic assumptions to obtain a pricing framework that is consistent both with the economic theory and the no-arbitrage principle.

Step 1 - The risk-neutral dynamics for the economic variables and their expectations are calculated using Girsanov theorem:

$$dm_X(t) = (a_X(t) - \lambda(t) \cdot b_X(t))dt + b_X(t) \cdot \ dW^{\mathbb{Q}}(t) \tag{3.11}$$

$$dm_I(t) = (a_I(t) - \lambda(t) \cdot b_I(t))dt + b_I(t) \cdot \ dW^{\mathbb{Q}}(t) \tag{3.12}$$

$$dX(t)/X(t) = (m_X(t) - \lambda(t) \cdot s_X(t))dt \ + s_X(t) \cdot \ dW^{\mathbb{Q}}(t) \tag{3.13}$$

$$dI(t)/I(t) = (m_I(t) - \lambda(t) \cdot s_I(t))dt \ + s_I(t) \cdot \ dW^{\mathbb{Q}}(t). \tag{3.14}$$

Step 2 - Similarly, the risk-neutral dynamics for the central bank policy are obtained using Girsanov theorem:

$$dM(t)/M(t) = -h_p(dI(t)/I(t) - \bar{p}dt) - h_x(dX(t)/X(t) - \bar{x}dt) - \lambda(t) \cdot s_M(t)dt + s_M(t) \cdot \ dW^{\mathbb{Q}}(t). \tag{3.15}$$

Step 3 - Putting together the central bank policy equation (3.15) and the economy dynamics (equa-

tions (3.13) and (3.14)) in the risk-neutral measure we get:

$$dM(t)/M(t) = -h_p((m_I(t) - \lambda(t) \cdot s_I(t))dt \ + s_I(t) \cdot \ dW^{\mathbb{Q}}(t) - \bar{p}dt)$$

$$-h_x((m_X(t) - \lambda(t) \cdot s_X(t))dt \ + s_X(t) \cdot \ dW^{\mathbb{Q}}(t) - \bar{x}dt) - \lambda(t) \cdot s_M(t)dt + s_M(t) \cdot \ dW^{\mathbb{Q}}(t). \qquad (3.16)$$

Step 4 - Equating the central bank policy equation (3.16) and equation (3.8), which models the impact of central bank policy on bond prices, we obtain:

$$\gamma dt + \int_t^{t+\Omega} Z(T)[dP(t,T)/P(t,T)]dT + s_L(t) \cdot dW^{\mathbb{Q}}(t) = -h_p((m_I(t) - \lambda(t) \cdot s_I(t))dt + s_I(t) \cdot dW^{\mathbb{Q}}(t) - \bar{p}dt)$$

$$-h_x((m_X(t) - \lambda(t) \cdot s_X(t))dt \ + s_X(t) \cdot \ dW^{\mathbb{Q}}(t) - \bar{x}dt) - \lambda(t) \cdot s_M(t)dt + s_M(t) \cdot \ dW^{\mathbb{Q}}(t). \qquad (3.17)$$

Step 5 - Putting together the above equation (3.17) (which contains the economy dynamics, the central bank policy, and its effect on bond prices) and the no arbitrage condition for the bond price dynamics (equation (3.1)), we obtain:

$$\gamma dt + \int_t^{t+\Omega} Z(T)[n(t)dt + \sigma_P(t,T) \cdot dW^{\mathbb{Q}}(t)]dT + s_L(t) \cdot dW^{\mathbb{Q}}(t) = -h_p((m_I(t) - \lambda(t) \cdot s_I(t))dt + s_I(t) \cdot dW^{\mathbb{Q}}(t) - \bar{p}dt)$$

$$-h_x((m_X(t) - \lambda(t) \cdot s_X(t))dt \ + s_X(t) \cdot \ dW^{\mathbb{Q}}(t) - \bar{x}dt) - \lambda(t) \cdot s_M(t)dt + s_M(t) \cdot \ dW^{\mathbb{Q}}(t). \qquad (3.18)$$

We compact the notation by introducing the following functions:

$$\zeta(t) = \int_t^{t+\Omega} Z(T)dT$$

$$\Sigma_P(t) = \int_t^{t+\Omega} Z(T)\sigma_P(t,T)dT.$$

We note that the function $\zeta(t)$ is always strictly positive because the function $Z(T)$ and the constant Ω are requested to be strictly positive.

With the above definitions the model equation becomes:

$$\zeta(t)n(t)dt + \Sigma_P(t) \cdot dW^{\mathbb{Q}}(t) =$$

$$-h_p[(m_I(t) - \lambda(t) \cdot s_I(t))dt \ + s_I(t) \cdot \ dW^{\mathbb{Q}}(t) - \bar{p}dt]$$

$$-h_x[(m_X(t) - \lambda(t) \cdot s_X(t))dt + s_X(t) \cdot \ dW^{\mathbb{Q}}(t) - \bar{x}dt]$$

$$-\lambda(t) \cdot s_M(t)dt + s_M(t) \cdot \ dW^{\mathbb{Q}}(t)$$

$$-\gamma dt - s_L(t) \cdot \ dW^{\mathbb{Q}}(t).$$

Step 6 - The no-arbitrage conditions are obtained from the above equation by collecting the terms multiplied by dt and $dW(t)$ in the following way:

$$\gamma = h_p\bar{p} + h_x\bar{x} \tag{3.19}$$

$$\zeta(t)n(t) \ = -h_p[m_I(t) - \lambda(t) \cdot s_I(t)] - h_x[m_X(t) - \lambda(t) \cdot s_X(t)] - \lambda(t) \cdot s_M(t) \tag{3.20}$$

$$\Sigma_P(t) = \int_t^{t+\Omega} Z(T)\sigma_P(t,T)dT = -h_p s_I(t) - h_x s_X(t) + s_M(t) - s_L(t). \tag{3.21}$$

We note that we can equate the equation above to $s_L(t)$ thanks to condition (3.8):

$$\Sigma_P(t) = \int_t^{t+\Omega} Z(T)\sigma_P(t,T)dT = -h_p s_I(t) - h_x s_X(t) + s_M(t) - s_L(t) = -s_L(t). \tag{3.22}$$

Some observations can be made regarding these conditions:

1. The first condition can be regarded as the risk-neutral drift for the money supply assuming that there is no uncertainty and no monetary policy. Here we group all deterministic terms multiplied by dt. Therefore we can refer to the constant γ as the natural money supply growth rate. We also note that the constant γ is likely to be positive, given that the central bank reaction function parameters h_x and h_P are positive by construction and that the target levels $\bar{x}$ and $\bar{p}$ are usually positive numbers. This matches the intuition that over time the money supply tends to grow, unless the central bank tries to reduce it.

2. The second calibration conditions gives us a closed-form expression for the short rate that is used in the following section to get the short rate dynamics. If one remembers the condition (3.10)

$$h_p s_I(t) + h_x s_X(t) - s_M(t) = 0$$

the calibration condition simplifies into

$$\zeta(t)n(t) \ = -h_p m_I(t) - h_x m_X(t) \tag{3.23}$$

which shows that the second calibration condition contains all stochastic terms multiplied by dt. In fact the expectation drifts $m_I(t)$ and $m_X(t)$ are stochastic.

3. The bond volatilities $\sigma_P(t,T)$ are determined by other model parameters and can be regarded as a combination of the economic factor volatilities. This is clear from the third calibration condition, which contains all terms multiplied by $dW(t)$. There are constraints on $s_M(t)$ (see condition (3.10)) for which one has to write $\Sigma_P(t) = -s_L(t)$.

Step 7 - To price inflation derivatives, it can be useful to work with the T^*-forward measure. By using the techniques detailed for example in Brigo & Mercurio [22], one obtains the inflation and bond price dynamics under this measure:

$$dI(t)/I(t) = (m_I(t) - \lambda(t) \cdot s_I(t) + \sigma_P(t,T^*) \cdot s_I(t))dt + s_I(t) \cdot dW^{T^*}(t) \tag{3.24}$$

$$dP(t,T)/P(t,T) = (n(t) + s_P(t,T) \cdot \sigma_P(t,T^*))dt + s_P(t,T) \cdot dW^{T^*}(t). \tag{3.25}$$

Step 8 - Because a closed expression for the short rate $n(t)$ has been found in Step 6, it is possible to make explicit the pricing kernel $\psi(t) = L(t)/B(t)$ in this model:

$$\psi(t) = L(t)e^{-\int_0^t n(s)ds}.$$

Having an explicit form for the pricing kernel allows one to price derivatives using the real world measure, if needed.

We close this section with an observation on instantaneous correlations. In this model the same n-dimensional Brownian motion is the source of randomness for all variables. Given two stochastic differential equations $dX(t) = a\,dW^1(t) + b\,dW^2(t)$ and $dY(t) = c\,dW^1(t) + f\,dW^2(t)$, where $\{W^1(t)\}_{t\geq 0}$ and $\{W^2(t)\}_{t\geq 0}$ are two independent one-dimensional Brownian motions and a, b, c, and f are deterministic real constants, one obtains $dX(t)dY(t) = (ac + bf)dt$. Perhaps we can write a general formula for the instantaneous correlation ρ_t using quadratic variations and covariation:

$$\rho_t = \frac{d\langle X,Y\rangle_t}{\sqrt{d\langle X,X\rangle_t\, d\langle Y,Y\rangle_t}} = \frac{(ac+bf)}{\sqrt{a^2+b^2}\sqrt{c^2+d^2}}.$$

Therefore one can use the model volatilities to calibrate also market-implied instantaneous correlations between the macroeconomic variables: this is proposed the next chapter.

3.3 Equivalent interest rates model

Here we show that the model presented in the previous section, although is a completely new model and is derived from macroeconomic assumptions, gives some dynamics for the short rate that are consistent with a mean-reverting Hull-White model. The Hull-White model, presented in Hull & White [74] and further analysed in Brigo & Mercurio [22], is a widely-used model for the short rate $n(t)$ that has the properties to be mean-reverting and to calibrate to any given term structure of interest rates. In this section we show how this model is derived within the macroeconomic framework and study its mean-reverting property as a function of the economy parameters.

The derivation is carried out as follows. The second calibration condition (3.20) gives an expression containing the short-term interest rate $n(t)$:

$$\zeta(t)n(t) \ = -h_p[m_I(t) - \lambda(t) \cdot s_I(t)] - h_x[m_X(t) - \lambda(t) \cdot s_X(t)] - \lambda(t) \cdot s_M(t).$$

If one differentiates this condition and remembers the condition (3.10)

$$h_p s_I(t) + h_x s_X(t) - s_M(t) = 0$$

one gets:

$$d[\zeta(t)n(t)] \ = -h_p[dm_I(t)] - h_x[dm_X(t)]$$

$$d\zeta(t)n(t) + \zeta(t)dn(t) = -h_p dm_I(t) - h_x dm_X(t).$$

There is no covariance term in the left-hand side of the above differential given that $\zeta(t)$ is a deterministic function. We remember the expressions for the drift differentials (3.11) and (3.12) and substitute them in the above expression, obtaining:

$$d\zeta(t)n(t) + \zeta(t)dn(t) = -h_p[[a_I(t) - \lambda(t) \cdot b_I(t)]dt + b_I(t) \cdot dW^{\mathbb{Q}}(t)]+$$

$$-h_x[[a_X(t) - \lambda(t) \cdot b_X(t)]dt + b_X(t) \cdot dW^{\mathbb{Q}}(t)].$$

Further, one needs to calculate the differential of $\zeta(t)$:

$$d\zeta(t) = \left(\frac{\partial \int_t^{t+\Omega} Z(T)dT}{\partial t} \right) dt = [Z(t+\Omega) - Z(t)]dt.$$

One substitutes the above result in the differential, and after rearranging one obtains:

$$\zeta(t)dn(t) = -[Z(t+\Omega) - Z(t)]n(t)dt - h_p[[a_I(t) - \lambda(t) \cdot b_I(t)]dt + b_I(t) \cdot dW^{\mathbb{Q}}(t)]+$$

$$-h_x[[a_X(t) - \lambda(t) \cdot b_X(t)]dt + b_X(t) \cdot dW^{\mathbb{Q}}(t)].$$

We can highlight the differential of the short rate $n(t)$:

$$dn(t) = -[Z(t+\Omega) - Z(t)]/\zeta(t)n(t)dt - h_p/\zeta(t)[[a_I(t) - \lambda(t) \cdot b_I(t)]dt$$

$$+b_I(t) \cdot dW^{\mathbb{Q}}(t)] - h_x/\zeta(t)[[a_X(t) - \lambda(t) \cdot b_X(t)]dt + b_X(t) \cdot dW^{\mathbb{Q}}(t)].$$

To compact notation one defines the following terms:

$$f_1(t) = [-h_p a_I(t) - h_x a_X(t)]/\zeta(t) \tag{3.26}$$

$$f_2(t) = [Z(t+\Omega) - Z(t)]/\zeta(t) \tag{3.27}$$

$$\sigma_n(t) = [-h_x b_X(t) - h_p b_I(t)]/\zeta(t). \tag{3.28}$$

Clearly one can find different parameterisations of the CTCB model (i.e. different functions $Z(T)$, $a_I(t)$, $a_X(t)$, $b_I(t)$, $b_X(t)$, or different parameters h_p, h_x) that yield the same Hull-White parameters. We remind the reader that the assumptions made on the function $Z(T)$ ensure that the function $\zeta(t)$ is always strictly positive and bounded, and therefore can be safely used as a denominator. This shows that the model implies some short nominal interest rates dynamics that are similar to the ones assumed by the generalised Vasicek model:

$$dn(t) = [f_1(t) - f_2(t)n(t) - \lambda(t) \cdot \sigma_n(t)]dt + \sigma_n(t) \cdot dW^{\mathbb{Q}}(t).$$

It is important to notice that the requests made on the function $Z(T)$ (to be an increasing and positive function), besides making sense from an economic perspective (as explained in 3.1.4, point 5), also imply that $f_2(t)$ is always positive, i.e. that the nominal short rate is mean-reverting.

Before doing some further analysis, we notice that the source of randomness in the CTCB model is n-dimensional, and the volatility function $\sigma_n(t)$ is n-dimensional accordingly. To stress the difference against the original Hull-White model, where the driving Brownian motion is scalar, we write the scalar Hull-White volatility as $\sigma_n^*(t)$: we link the two processes by asking that the total instantaneous variance of the source of randomness of the CTCB model is the same as the total variance of the Hull-White

model. The relationship is:

$$[\sigma_n^*(t)]^2 = \sum_{i=1}^{n} [\sigma_n^i(t)]^2 \tag{3.29}$$

where $\sigma_n^i(t)$ is the i-th component of the n-dimensional model volatility function $\sigma_n(t)$.

Consistently with the Hull-White model, the distribution of the short rate is Gaussian and can generate negative short nominal rates: in the current low rates environment, when central banks are setting negative deposit rates, we don't think this is a theoretical problem, rather we think that this model is probably better suited than other positive rates models to deal with the current market conditions. In Denmark and Eurozone the short interest rate in 2014 was negative at -0.2%: in practice central banks can set a negative short rate to stimulate commercial banks to lend to consumers and firms in times of economic distress.

The only differences w.r.t. to the original extended Vasicek model is that the volatility is a multidimensional function of time, and that the driving source of randomness is a multidimensional Brownian motion: as explained in detail in point 4 below, we find the volatility vector components to target a certain level of total volatility, and therefore the marginal distribution. This said, these differences do not prevent us from reaching the following conclusions:

1. If we only want to use this model to price interest rates derivatives, one calibrates the function $f_1(t)$ to the nominal forward rates observed in the market, as suggested in the original Hull-White paper, modulo some changes. Alternatively, we use the calibration condition $\mathbb{E}_t^{T^*}[n(T)] = f(t,T)$.

2. The nominal bond price is such that the volatility of the relative moves is a deterministic volatility function. This is no surprise given the original assumptions. This is important because it can simplify the calculation of the year-on-year convexity adjustment, as it is shown in the following sections.

3. Thanks to the above fact we use Black-Scholes formulas to price European bond options. Moving to the relevant forward measure allows one to carry out discounting by multiplying by the market bond prices. Because it is trivial to price bond options in this model, we write Black-type formulas for bond options. These gives closed forms for nominal caps/floors and swaptions (using the method presented in Jamshidian [78]): this intuition is developed in the following sections.

4. Because the process for the short rate is normal, trees can be easily constructed. In fact the n-dimensional Brownian motion can be treated as a one-dimensional process for this purpose (this technique is also called "flattening", where the independent components of the Brownian motion are "summed" and considered as a single Brownian motion with the appropriate diffusion term).

3.4 Further analysis on the mean reversion property

3.4.1 General case

Here we analyse the mean reversion coefficient $f_2(t)$ found in the previous section (result (3.27)) and we link it to the general theory of mean-reverting Gaussian models developed by Hull & White [74]. Here we refer to the original formulation of the model, where the driving Brownian motion $\sigma_n^*(t)$ is a scalar process: we translate it into a vector by using (3.29). As we have shown, this is not a major problem: one can "flatten" the vector volatility into an equivalent scalar volatility that leaves the total instantaneous variance unchanged.

In particular, in [74] Hull & White present a version of the model with time-dependent coefficients, where the dynamics of the short rate $n(t)$ are governed by the SDE:

$$dn(t) = [\theta(t) - a(t)n(t) + b - \lambda(t)\sigma_n^*(t)]dt + \sigma_n^*(t)dW^{\mathbb{Q}}(t). \tag{3.30}$$

In this formulation for simplicity we are setting the original parameter $b = 0$, compared to the original formulation in the paper [74]. Clearly $\lambda(t)$ is one-dimensional in the above equation. They suggest a calibration strategy that yields the model parameters as functions of the two functions used to fit the term structure of interest rates using the Ansatz:

$$P(t,T) = A(t,T)e^{-n(t)B(t,T)}. \tag{3.31}$$

At the initial time $t = 0$ the positive functions $A(0,T)$ and $B(0,T)$ are numerically calibrated to match the observed term structure $P(0,T)$.

In particular, Hull & White find that the mean reversion speed $a(t)$ has to satisfy the condition:

$$a(t) = -\frac{\frac{\partial B^2(0,t)}{\partial t^2}}{\frac{\partial B(0,t)}{\partial t}} = -\frac{\beta''(t)}{\beta'(t)} \tag{3.32}$$

where we have made the notation lighter by defining: $\beta(t) = B(0,t)$.

Further, the authors prove a calibration condition for the mean reversion level parameter $\theta(t)$:

$$\theta(t) = \lambda(t)\sigma_n^*(t) - a(t)\frac{\partial \log A(0,t)}{\partial t} - \frac{\partial^2 \log A(0,t)}{\partial t^2} + \left[\frac{\partial B(0,t)}{\partial t}\right]^2 \int_0^t \left[\frac{\sigma_n^*(s)}{\frac{\partial B(0,s)}{\partial s}}\right]^2 ds. \tag{3.33}$$

At this stage we observe that the time-dependent version of the Hull-White model does not necessarily

imply mean reversion: in fact, the mean reversion speed coefficient $a(t)$ is positive (i.e. there is mean reversion) only if the sign of the first derivative $\beta^{'}(t)$ is different from the sign of the second derivative $\beta^{''}(t)$.

For example, if one takes $A(0,t) = 1$, which is a legitimate choice, the function $B(0,t)$ is increasing with t where the term structure is not downward-sloping and $n(0) > 0$. Let us introduce the compounded spot rate for maturity T observed at time t and denote it by $Y(t,T)$: it is defined as the flat interest rate such that: $P(t,T) = e^{-Y(t,T)(T-t)}$. If $A(0,t) = 1$, this expression has to be equal to $P(0,T) = e^{-n(0)B(0,T)}$. Equating the two terms one gets $Y(0,T) = n(0)\frac{B(0,T)}{T}$.

We make some stylised examples to gain intuition in some extreme cases:

1. If $B(0,T)$ is equal to the maturity, i.e. $B(0,T) = T$, its second derivative is zero. This implies no mean reversion, which is consistent to the fact that the term structure is flat at $n(0)$, i.e. $Y(0,T) = n(0)\frac{T}{T} = n(0)$. There is no mean reversion because there is no need of it if rates are constant in maturity.

2. If $B(0,T)$ is quadratic in maturity, i.e. $B(0,T) = T^2$, its first derivative is linear in maturity and its second derivative is constant. The mean reversion speed is therefore a negative number, pointing to an explosion of interest rates. This is consistent with the fact that the term structure is very steep (linear in maturity), i.e. $Y(0,T) = n(0)\frac{T^2}{T} = n(0)T$.

3. If $B(0,T)$ is a square root function of the maturity i.e. $B(0,T) = T^{1/2}$, its first derivative would be $\frac{1}{2}T^{-1/2}$, its second derivative would crucially be the negative $-\frac{1}{4}T^{-3/2}$, therefore the ratio $-\frac{\beta^{''}(t)}{\beta^{'}(t)}$ would be a positive number, indicating reversion to the mean. This is consistent with the fact that the term structure is very inverted, i.e. $Y(0,T) = n(0)\frac{T^{1/2}}{T} = n(0)T^{-1/2}$. Clearly an infinite short rate at time $t = 0$ is a somewhat idealised situation, which is perhaps useful only to understand how this model works in theory.

We claim that for non-pathological term structures of interest rates the model should ensure mean reversion. Clearly this example is simplistic as we assumed for simplicity that $A(0,t) = 1$.

We focus our attention on the mean-reversion speed, and equate the result from Hull & White [74] to the expression found in the previous section for the CTCB model. This allows one to draw some conclusions on the function $Z(T)$. By equating (3.27) and (3.32) one gets:

$$-\frac{\beta^{''}(t)}{\beta^{'}(t)} = [Z(t+\Omega) - Z(t)]/\zeta(t)$$

$$-\frac{\beta^{''}(t)}{\beta^{'}(t)} = \frac{[Z(t+\Omega) - Z(t)]}{\int_t^{t+\Omega} Z(T)dT}$$

$$\int_t^{t+\Omega} Z(T)dT = -[Z(t+\Omega) - Z(t)]\frac{\beta^{'}(t)}{\beta^{''}(t)}.$$

We take a derivative of the above expression w.r.t. t:

$$[Z(t+\Omega) - Z(t)] = -[Z^{'}(t+\Omega) - Z^{'}(t)]\frac{\beta^{'}(t)}{\beta^{''}(t)} - [Z(t+\Omega) - Z(t)]\frac{(\beta^{''}(t))^2 - \beta^{'}(t)\beta^{'''}(t))}{(\beta^{''}(t))^2}.$$

We rearrange the above expression as:

$$[Z(t+\Omega) - Z(t)]\left(\frac{2(\beta^{''}(t))^2 - \beta^{'}(t)\beta^{'''}(t)}{(\beta^{''}(t))^2}\right) = -[Z^{'}(t+\Omega) - Z^{'}(t)]\frac{\beta^{'}(t)}{\beta^{''}(t)}$$

$$\frac{[Z(t+\Omega) - Z(t)]}{[Z'(t+\Omega) - Z'(t)]} = -\frac{\frac{\beta^{'}(t)}{\beta^{''}(t)}}{\left(\frac{2(\beta^{''}(t))^2 - \beta^{'}(t)\beta^{'''}(t)}{(\beta^{''}(t))^2}\right)} = -\frac{\beta^{'}(t)\beta^{''}(t)}{2(\beta^{''}(t))^2 - \beta^{'}(t)\beta^{'''}(t)}.$$

By defining

$$u(t) = Z(t+\Omega) - Z(t)$$

and

$$\alpha(t) = -\frac{\beta^{'}(t)\beta^{''}(t)}{2(\beta^{''}(t))^2 - \beta^{'}(t)\beta^{'''}(t)}$$

and requiring that $\beta^{'}(t) \neq 0$, $\beta^{''}(t) \neq 0$, one rewrites the above expression as:

$$\frac{u'(t)}{u(t)} = \frac{d\log\ u(t)}{dt} = \alpha^{-1}(t).$$

This linear ODE is solved in $t > t_0$ to yield:

$$u(t) = u(t_0)e^{\int_{t_0}^{t} \alpha^{-1}(s)ds}.$$

The meaning of the above result is a relationship between the functions $Z(t)$ and $\beta(t)$:

$$[Z(t+\Omega) - Z(t)] = [Z(t_0+\Omega) - Z(t_0)]e^{\int_{t_0}^{t} -\frac{2(\beta^{''}(s))^2 - \beta^{'}(s)\beta^{'''}(s)}{\beta^{'}(s)\beta^{''}(s)}ds}. \tag{3.34}$$

We make two observations based on this expression:

1. To ensure mean reversion, one has to require that the function $Z(t)$ is positive and increasing: this request is made in the previous section and is confirmed by looking at the properties of the Hull-White model, in particular (3.34). In fact, if $Z(t)$ is positive and increasing, $Z(t_0+\Omega) > Z(t_0)$, as $\Omega > 0$. Further, an exponential is always positive, and therefore $Z(t+\Omega) > Z(t)$.

2. We use the relationship above to calibrate the function $Z(t)$ as a function of $\beta(t)$ and $A(0,t)$ (i.e. the term structure of interest rates): alternatively we set the functions $A(0,t)$ and $Z(t)$ using some functional forms, and obtain the function $\beta(t)$ by calibrating to the term structure. The latter seems more appropriate in the context we are working in, as we may want to make some assumptions on the market liquidity function $Z(t)$.

The analysis done so far lets one impose a further calibration constraint on the model bond volatilities. In the previous section we have introduced the bond volatilities $\sigma_P(t,T)$ without specifying more details: taking in consideration result (3.28) and result (3.39) (that is proved in the following section) we impose a further calibration condition:

$$\sigma_P(t,T) = [h_x b_X(t) + h_p b_I(t)]/\zeta(t) \left[\frac{\beta(T) - \beta(t)}{\beta'(t)}\right]. \tag{3.35}$$

3.4.2 Constant mean reversion speed

We conclude this section with an observation regarding the constant mean reversion speed of the Hull-White model, that is used in many applications. The main result we find is that if one imposes that the function $Z(T)$ is an exponential in the form $Z(T) = e^{\delta T}$ (with $\delta > 0$ to ensure that $Z(T)$ is increasing in T), one immediately shows that the mean reversion speed is constant:

$$[Z(t+\Omega) - Z(t)]/\zeta(t) = \frac{Z(t+\Omega) - Z(t)}{\int_t^{t+\Omega} Z(T)dT} = \frac{e^{\delta(t+\Omega)} - e^{\delta t}}{\int_t^{t+\Omega} e^{\delta T} dT} = \frac{e^{\delta(t+\Omega)} - e^{\delta t}}{\frac{1}{\delta}(e^{\delta(t+\Omega)} - e^{\delta t})} = \delta. \tag{3.36}$$

This result is interesting from a theoretical perspective, because a higher mean reversion speed decreases the intra-curve rates correlation, as it is well known in literature: intuitively the short end of the curve moves faster than longer maturities, as the short rate reverts to its mean. This has a meaning similar to increasing the parameter δ: a higher parameter δ means that the longer maturities of the curve react more strongly to monetary policy compared to the short end of the curve. This increases the intra-curve decorrelation.

This result has a very practical implication too: the relationship (3.34) can be difficult to implement numerically, as one wants to impose the liquidity function $Z(t)$ and imply $\beta(t)$ (the converse would be trivial): this could reduce the flexibility of the model.

We remember result (3.32) and obtain the differential equation $\delta = -\frac{\beta''(t)}{\beta'(t)}$, which is solved by $\beta(t) = e^{-\delta t}$. Therefore, known $Z(t)$, one obtains $B(0,t) = \beta(t)$. We get the function $B(t,T) = [B(0,T) - B(0,t)]/(\partial B(0,t)/\partial t) = (e^{-\delta T} - e^{-\delta t})/(-\delta e^{-\delta t}) = (\delta^{-1})(1 - e^{-\delta^{(T-t)}})$: the function $B(t,T)$ is positive in this parametrisation. By observing from the market the short rate $n(0)$ and the term structure $P(0,t)$,

one uses the Ansatz $P(0,T) = A(0,T)e^{-n(0)B(0,T)}$ to obtain $A(0,t) = \frac{P(0,t)}{e^{-n(0)e^{-\delta t}}}$, which fully calibrates the model to the nominal term structure. This result is exploited in the calibration process in the following chapter. From this point we assume that, unless otherwise stated, $Z(T) = e^{\delta T}$.

By doing some calculations one shows how the form $Z(T) = e^{\delta T}$ compares with the relationship (3.34):

$$[Z(t+\Omega) - Z(t)] = [Z(t_0+\Omega) - Z(t_0)]e^{\int_{t_0}^{t} -\frac{2(\beta''(s))^2 - \beta'(s)\beta'''(s)}{\beta'(s)\beta''(s)}ds}$$

$$e^{\delta(t+\Omega)} - e^{\delta t} = [e^{\delta(t_0+\Omega)} - e^{\delta t_0}]e^{\int_{t_0}^{t} \frac{2\delta^4 - \delta^4}{\delta^3}ds} = [e^{\delta(t_0+\Omega)} - e^{\delta t_0}](e^{\delta(t-t_0)}).$$

Finally it is useful to show how the choice of $Z(T) = e^{\delta T}$ shows perfect consistency with the classical integration of the Ornstein-Uhlenbeck process: in fact, given the SDE $dn(t) = [\theta(t) - an(t)]dt + \sigma_n^*(t)dW(t)$ with known initial condition $n(s)$, one writes the differential for $n(t)e^{at}$, integrates and obtains the standard result:

$$n(t) = n(s)e^{-a(t-s)} + \int_s^t e^{-a(t-u)}\theta(u)du + \int_s^t e^{-a(t-u)}\sigma_n^*(u)dW(u).$$

In the above formula $\sigma_n^*(u)$ refers to the Hull-White scalar short rate volatility. If one remembers the calibration condition (3.20) we substitute the expression for $\zeta(t)$ inside it and confirm that one gets the same result stated in the above formula. In fact:

$$\zeta(t) = \int_t^{t+\Omega} Z(T)dT = \int_t^{t+\Omega} e^{\delta T}dT = \frac{e^{\delta(t+\Omega)} - e^{\delta t}}{\delta} = \frac{e^{\delta t}(e^{\delta\Omega} - 1)}{\delta}.$$

If one recalls the calibration condition (3.20) and the regularity condition (3.10) one writes:

$$\zeta(t)n(t) \ = -h_p[m_I(t) - \lambda(t)\cdot s_I(t)] - h_x[m_X(t) - \lambda(t)\cdot s_X(t)] - \lambda(t)\cdot s_M(t) = -h_p m_I(t) - h_x m_X(t).$$

Further, one integrates the SDEs (3.5) and (3.4):

$$\frac{e^{\delta t}(e^{\delta\Omega} - 1)}{\delta}n(t) \ = -[h_p m_I(s) + h_x m_X(s)] - \int_s^t [h_p a_I(u) + h_x a_X(u)]du - \int_s^t [h_p b_I(u) + h_x b_X(u)]\cdot dW(u)$$

$$n(t) = \frac{-[h_p m_I(s) + h_x m_X(s)]\delta e^{-\delta t}}{(e^{\delta\Omega} - 1)}\frac{e^{\delta s}}{e^{\delta s}} - \int_s^t \frac{[h_p a_I(u) + h_x a_X(u)]\delta e^{-\delta t}}{(e^{\delta\Omega} - 1)}\frac{e^{\delta u}}{e^{\delta u}}du - \int_s^t \frac{[h_p b_I(u) + h_x b_X(u)]\delta e^{-\delta t}}{(e^{\delta\Omega} - 1)}\frac{e^{\delta s}}{e^{\delta s}}\cdot dW(u)$$

$$n(t) = n(s)e^{-\delta(t-s)} - \int_s^t \frac{[h_p a_I(u) + h_x a_X(u)]\delta e^{-\delta(t-u)}}{(e^{\delta\Omega} - 1)}\frac{1}{e^{\delta u}}du - \int_s^t \frac{[h_p b_I(u) + h_x b_X(u)]\delta e^{-\delta(t-u)}}{(e^{\delta\Omega} - 1)}\frac{1}{e^{\delta u}}\cdot dW(u)$$

where we recall the calibration condition (3.20) rewritten as $n(s) = \frac{-[h_p m_I(s) + h_x m_X(s)]}{\zeta(s)}$.

Further calculations yield:

$$n(t) = n(s)e^{-\delta(t-s)} - \int_s^t \frac{[h_p a_I(u) + h_x a_X(u)]e^{-\delta(t-u)}}{\zeta(u)} du - \int_s^t \frac{[h_p b_I(u) + h_x b_X(u)]\delta e^{-\delta(t-u)}}{\zeta(u)} \cdot dW(u).$$

If one remembers the definitions (3.26) and (3.28) one finally obtains the desired result:

$$n(t) = n(s)e^{-\delta(t-s)} + \int_s^t e^{-\delta(t-u)}\theta(u)du + \int_s^t e^{-\delta(t-u)}\sigma_n(u) \cdot dW(u).$$

In the above formula we stress that the function $\sigma_n(u)$ is now the CTCB vector short rate volatility. The parameter δ in the CTCB model has the same meaning of the parameter a, using the notation of the original Hull-White model.

3.5 Pricing of vanilla interest rates derivatives

We found that our macroeconomic-based inflation model yields some short rate dynamics that are consistent with the Hull-White model: this makes the pricing of interest rates derivatives much simpler, as the theory has been extensively developed for this model. Because in the Hull-White model the bond prices are lognormally distributed, we use Black-type formulae to price bond options. Bond options are used also to find the prices of caplets and floorlets, letting one price caps and floors: this is explained for example in Brigo & Mercurio [22]. For swaptions, the method suggested by Jamshidian [78] can be followed. Closed forms allow faster pricing of vanilla interest rates derivatives, which speeds up the calibration. Further, if an uncertain-parameter version of the model is used to better match the market-observed skews, the pricing of vanillas is a simple linear combination of the closed forms found in the base case. This approach is proposed in chapter 5.

Before starting, we quote some results that are useful and that can be found for example in Hull & White [74] and Brigo & Mercurio [22]. In the following calculations we use the quantity $P(t, T_1, T_2)$, defined as a portfolio containing a long position in the bond $P(t, T_2)$ and a short position in the bond $P(t, T_1)$, with $T_1 < T_2$.

Lemma 1 *The undiscounted price of a European vanilla option on a lognormally distributed asset X(T) with strike K, whose logarithm has expectation* $\mathbb{E}[\log X(T)] = M$ *and variance* $Var[\log X(T)] = V^2$*, is given by:*

$$\mathbb{E}[\omega(X-K)^+] = \omega e^{M+1/2V^2} N(\omega(M - \log(K) + V^2)/V) - \omega K N(\omega(M - \log(K))/V) \tag{3.37}$$

where the function N(x) is the cumulative standard Gaussian distribution, i.e. $N(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} e^{-\frac{t^2}{2}} dt$

and $\omega \in \{-1, 1\}$ *for puts and calls respectively.*

Proof. Brigo & Mercurio [22] provide this proof in Appendix D.

Lemma 2 *The variance between times* t *and* T_1 *of the return of the quantity* $P(t, T_1, T_2)$*, with* $t < T_1 < T_2$*, in the Hull-White model is*

$$V_P(t, T_1, T_2) = [\beta(T_2) - \beta(T_1)]^2 \int_t^{T_1} \left[\frac{\sigma_n^*(u)}{\beta'(u)}\right]^2 du. \tag{3.38}$$

Proof. Hull & White [74] provide this result in section 2. One defines the bond price $P(t, T) = A(t, T)e^{-n(t)B(t,T)}$, and remembers the Hull-White SDE that describes the evolution of the short rate: $dn(t) = [\theta(t) - a(t)n(t) - \lambda(t)\sigma_n^*(t)]dt + \sigma_n^*(t)dW^{\mathbb{Q}}(t)$. Using Ito's lemma we obtain the risk-neutral dynamics of the bond price process, which are:

$$dP(t, T) = P(t, T)[n(t)dt - B(t, T)\sigma_n^*(t)dW^{\mathbb{Q}}(t)].$$

Therefore the bond volatility is $-B(t, T)\sigma_n^*(t)$: for the sake of clarity we stress again that here $\sigma_n^*(t)$ is the scalar short rate volatility of the dual Hull-White model; the last lemma of this section shows how this is related to the CTCB short rate model volatilities. The instantaneous variance of the return of the bond portfolio $P(t, T_1, T_2)$ would be $(B(t, T_1)\sigma_n^*(t))^2 + (B(t, T_2)\sigma_n^*(t))^2 - 2\rho(t, T_1, T_2)B(t, T_1)\sigma_n^*(t)B(t, T_2)\sigma_n^*(t)$, where $\rho(t, T_1, T_2)$ is the instantaneous correlation between the two bond prices. Because the model is a one-factor model, $\rho(t, T_1, T_2) = 1$, yielding an instantaneous variance of $(B(t, T_1)\sigma_n^*(t))^2 + (B(t, T_2)\sigma_n^*(t))^2 - 2B(t, T_1)\sigma_n^*(t)B(t, T_2)\sigma_n^*(t)$ which is equivalent to

$$[B(t, T_2)\sigma_n^*(t) - B(t, T_1)\sigma_n^*(t)]^2 = (\sigma_n^*(t))^2[B(t, T_2) - B(t, T_1)]^2. \tag{3.39}$$

If one recalls that in the time-varying version of the Hull-White model one calibration condition is:

$$B(t, T) = \frac{B(0, T) - B(0, t)}{\frac{\partial B(0,t)}{\partial t}} = \frac{\beta(T) - \beta(t)}{\beta'(t)}$$

we plug this result into the formula, obtaining an instantaneous variance of

$$(\sigma_n^*(t))^2 \left[\frac{\beta(T_2) - \beta(t)}{\beta'(t)} - \frac{\beta(T_1) - \beta(t)}{\beta'(t)}\right]^2.$$

Integrating between times t and T_1 yields the final result:

$$V_P(t,T_1,T_2) = \int_t^{T_1} (\sigma_n^*(u))^2 \left[\frac{\beta(T_2)-\beta(u)}{\beta'(u)} - \frac{\beta(T_1)-\beta(u)}{\beta'(u)}\right]^2 du = [\beta(T_2)-\beta(T_1)]^2 \int_t^{T_1} \left[\frac{\sigma_n^*(u)}{\beta'(u)}\right]^2 du.$$

Lemma 3 *In the Hull-White model, the price at time t of a zero-coupon bond option on the quantity $P(t,T_1,T_2)$ with option maturity T_1, strike K, and with $t < T_1 < T_2$ is*

$$ZBO(call,t,T_1,T_2,K) = P(t,T_2)N(h) - KP(t,T_1)N(h - (V_P(t,T_1,T_2))^{\frac{1}{2}}) \tag{3.40}$$

$$ZBO(put,t,T_1,T_2,K) = -P(t,T_2)N(-h) + KP(t,T_1)N(V_P(t,T_1,T_2))^{\frac{1}{2}} - h) \tag{3.41}$$

where $h = \frac{1}{V_P(t,T_1,T_2))^{\frac{1}{2}}} \log\left[\frac{P(t,T_2)}{P(t,T_1)K}\right] + \frac{V_P(t,T_1,T_2))^{\frac{1}{2}}}{2}$.

Proof. Hull & White [74] provide this result in section 2. A similar proof, for time-independent coefficients is available in Brigo & Mercurio [22] in section 3.3.2. The proof is based on the result of the previous lemma coupled with a Black-type option pricing formula.

Lemma 4 *The price at time t of a caplet (floorlet) with maturity T_1, strike K, and notional M, on the forward rate between times T_1 and T_2, denoted as $F(t,T_1,T_2)$, is the price of a put (call) option with strike $(1+K(T_2-T_1))^{-1}$, notional $M(1+K(T_2-T_1))$, maturity T_1 on the quantity $P(t,T_1,T_2)$. The result is model independent and we assume $t < T_1 < T_2$.*

$$Caplet(t,T_1,T_2,K,N) = M(1+K(T_2-T_1))ZBO(put,t,T_1,T_2,(1+K(T_2-T_1))^{-1}) \tag{3.42}$$

$$Floorlet(t,T_1,T_2,K,N) = M(1+K(T_2-T_1))ZBO(call,t,T_1,T_2,(1+K(T_2-T_1))^{-1}). \tag{3.43}$$

Proof. Brigo & Mercurio [22] provide this proof in section 2.6.1. The proof is carried out remembering that the forward rate can be written as $F(t,T_1,T_2) = \frac{1}{T_2-T_1}\left(\frac{P(t,T_1)}{P(t,T_2)} - 1\right)$. ZBO prices were found in the previous lemma.

Lemma 5 *The price of a unit-notional option on a coupon-bearing bond with maturity T in the Hull-White model is equivalent to pricing a portfolio of zero-coupon bond options using the special nominal rate n^*. Coupons paid at time $T_i > T$ are denoted by c_i, and the last (M-th) coupon includes the notional repayment.*

$$CBO(call,t,T,T_1...T_M,c_1...c_M,K) = \sum_{i=1}^{M} c_i ZBO(call,t,T,T_i,P(t,T_i,n^*)) \tag{3.44}$$

$$CBO(put, t, T, T_1...T_M, c_1...c_M, K) = \sum_{i=1}^{M} c_i ZBO(put, t, T, T_i, P(t, T_i, n^*)). \tag{3.45}$$

Proof. The proof is in Brigo & Mercurio [22] in section 3.11.1, based on Jamshidian [78]. In the following proof we denote the price of a zero-coupon bond $P(t,T)$ in the Hull-White model as a function of the short rate as $P(t,T,n(t))$. For example, a put option with strike K on a coupon-bearing bond with M coupons c_i paid at times T_i would have payoff $\left[K - \sum_{i=1}^{M} P(t,T_i,n(t))c_i\right]^+$. We assume we can find a special nominal rate n^* such that $K = \sum_{i=1}^{M} P(t,T_i,n^*)c_i$. Therefore the option payoff can be rewritten as $\left[\sum_{i=1}^{M}[P(t,T_i,n^*) - P(t,T_i,n(t))]c_i\right]^+$: we want to rewrite the positive part of this sum as the sum of the single positive components $\left[\sum_{i=1}^{M}[P(t,T_i,n^*) - P(t,T_i,n(t))]^+ c_i\right]$. We note that this is possible only if the bond price is a monotonic function of the short rate $n(t)$, which implies that all terms in the previous sum have the same sign. This property is satisfied by the Hull-White model given their assumption $P(t,T) = A(t,T)e^{-n(t)B(t,T)}$ and that $B(t,T) > 0$ (we found that $B(t,T) = (1-e^{-\delta(T-t)})/\delta$).

Lemma 6 *The price at time t of a payer (P) swaption (that gives the right to enter at time T into a payer swap with fixed rate K and M payment dates $T_i > T$) is equivalent to the price of an option on a coupon-bearing bond. The result is model-independent. Receiver (R) swaptions are recovered via call-put parity.*

$$Swtpn(P, t, T, T_1...T_M, K) = CBO(put, t, T, T_1...T_M, c_1...c_M, K) = \sum_{i=1}^{M} c_i ZBO(put, t, T, T_i, X_i) \tag{3.46}$$

$$Swtpn(R, t, T, T_1...T_M, K) = CBO(call, t, T, T_1...T_M, c_1...c_M, K) = \sum_{i=1}^{M} c_i ZBO(call, t, T, T_i, X_i) \tag{3.47}$$

$$X_i = A(T, T_i)e^{-n^* B(T,T_i)}.$$

Proof. The proof is in Brigo & Mercurio [22] in section 3.3.2.
The special rate n^* such that $K = \sum_{i=1}^{M} P(t,T_i,n^*)c_i$ is found numerically. Here c_i is the fixed rate $K(T_i - T_{i-1})$ except for the last maturity when also the notional is paid back. Therefore one writes the swaption as an option on a coupon-bearing bond, which is itself equivalent to a portfolio of options with maturity $T < T_i$ on zero-coupon bonds with strikes $X_i = A(T,T_i)e^{-n^* B(T,T_i)}$.

Lemma 7 *The price of a swaption with strike K, maturity T and payment dates $T_i > T$ in the Hull-White model is:*

$$S(P, t, T, T_1...T_M, K) = \sum_{i=1}^{M} c_i ZBO(put, t, T, T_i, X_i) = \sum_{i=1}^{M} c_i[-P(t,T_i)N(-h_i) + X_i P(t,T_{i-1})N((V_P(t,T_{i-1},T_i))^{\frac{1}{2}} - h_i)] \tag{3.48}$$

$$S(R,t,T,T_1...T_M,K) = \sum_{i=1}^{M} c_i ZBO(call,t,T,T_i,X_i) = \sum_{i=1}^{M} c_i P(t,T_i)N(h_i) - X_i P(t,T_{i-1})N(h_i - (V_P(t,T_{i-1},T_i))^{\frac{1}{2}}) \tag{3.49}$$

$$h_i = \frac{1}{V_P(t,T_i,T_{i-1}))^{\frac{1}{2}}} \log\left[\frac{P(t,T_i)}{P(t,T_{i-1})X_i}\right] + \frac{V_P(t,T_{i-1},T_i))^{\frac{1}{2}}}{2}$$

$$X_i = A(T,T_i)e^{-n^* B(T,T_i)}.$$

Proof. The proof is in Brigo & Mercurio [22] in section 3.3.2. based on Jamshidian [78].

We can price derivatives based on the above results using the macroeconomic model defined in the previous sections and leveraging on the equivalent short rate model, for which all the previous results are well known.

Lemma 8 *The prices of bond options, caplets and floorlets, and swaptions in the CTCB model follow the formulas proposed above with the following parametrisation*

$$[\sigma_n^*(t)]^2 = \sum_{i=1}^{n} \{[-h_x b_X^i(t) - h_p b_I^i(t)]/\zeta(t)\}^2$$

where n is the dimensionality of the Brownian motion $W(t)$. Here $b_X^i(t)$ is the i-th component of the volatility vector $b_X(t)$, and $b_I^i(t)$ is the i-th component of the volatility vector $b_I(t)$.

Proof. Follows from the lemmas proved above and from relationship (3.28).

3.6 Pricing zero-coupon inflation swaps and options

In this section we calculate the full expression for the price index $I(t)$: its conditional lognormality translates into closed forms ("Black type") for zero-coupon inflation options. This makes the model calibration much faster. The price index dynamics in the forward measure can be used to simplify the problem by discounting via multiplication by the zero-coupon bond.

To do these analyses, we calculate the closed form dynamics of $I(t)$ taking into account the stochastic dynamics of its drift $m_I(t)$.
We start by obtaining their T^*-forward dynamics:

$$dI(t)/I(t) = (m_I(t) - \lambda(t) \cdot s_I(t) + \sigma_P(t,T^*) \cdot s_I(t))dt + s_I(t) \cdot dW^{T^*}(t)$$

$$dm_I(t) = [a_I(t) - \lambda(t) \cdot b_I(t) + \sigma_P(t,T^*) \cdot b_I(t)]dt + b_I(t) \cdot dW^{T^*}(t).$$

We compact the notation by defining:

$$g_1(t) = -s_I(t) \cdot (\lambda(t) - \sigma_P(t, T^*))$$

$$g_2(t) = a_I(t) - b_I(t) \cdot (\lambda(t) - \sigma_P(t, T^*)).$$

We notice that $g_1(t)$ and $g_2(t)$ are deterministic as all the quantities used to build them are deterministic. At this stage we recall that in the CTCB model we assuming $Z(T) = e^{\delta T}$: therefore the bond option volatilities are expressed as:

$$\sigma_P(t, T) = \frac{[h_x b_X(t) + h_p b_I(t)]}{(e^{\delta(t+\Omega)} - e^{\delta t})}(1 - e^{\delta^{(T-t)}}).$$

The dynamics are rewritten in a more compact form as:

$$dI(t)/I(t) = (m_I(t) + g_1(t))dt + s_I(t) \cdot \ dW^{T^*}(t)$$

$$dm_I(t) = g_2(t)dt + b_I(t) \cdot dW^{T^*}(t).$$

We are in a position to integrate the drift of the price index over time:

$$dm_I(s) = g_2(s)ds + b_I(s) \cdot dW^{T^*}(s)$$

$$m_I(s) - m_I(t) = \int_t^s g_2(u)du + \int_t^s b_I(u) \cdot dW^{T^*}(u).$$

We integrate the expression for $m_I(s)$ between times t and T:

$$\int_t^T m_I(s)ds = m_I(t)(T - t) + \int_t^T \int_t^s g_2(u)duds + \int_t^T \int_t^s b_I(u) \cdot dW^{T^*}(u)ds.$$

Applying Fubini's theorem, we recall that

$$\int_t^T \int_t^s g_2(u)duds = \int_t^T \int_u^T ds\, g_2(u)du = \int_t^T (T - u)g_2(u)du$$

and that

$$\int_t^T \int_t^s b_I(u) \cdot dW^{T^*}(u)ds = \int_t^T \int_u^T ds\, b_I(u) \cdot dW^{T^*}(u) = \int_t^T (T - u)b_I(u) \cdot dW^{T^*}(u).$$

We write the integral of the price index drift in a simpler form:

$$\int_t^T m_I(s)ds = m_I(t)(T-t) + \int_t^T (T-s)g_2(s)ds + \int_t^T (T-s)b_I(s)\cdot dW^{T^*}(s).$$

We write the normal distribution of the integral of the drift using Ito isometry:

$$\int_t^T m_I(s)ds \sim \mathcal{N}\left(m_I(t)(T-t) + \int_t^T (T-s)g_2(s)ds, \int_t^T (T-s)^2 b_I(s)\cdot b_I(s)ds\right).$$

With the above results in mind we derive the expression for the price index level $I(t)$:

$$I(T) = I(t)e^{\int_t^T m_I(s)ds + \int_t^T (g_1(s) - \frac{1}{2}s_I(s)\cdot\ s_I(s))ds + \int_t^T s_I(s)\cdot\ dW^{T^*}(s)}$$

$$I(T) = I(t)e^{\int_t^T (m_I(t) + (T-s)g_2(s) + g_1(s) - \frac{1}{2}s_I(s)\cdot\ s_I(s))ds + \int_t^T ((T-s)b_I(s) + s_I(s))\cdot\ dW^{T^*}(s)}.$$

To achieve a lighter notation, we define:

$$g_3(s) = m_I(t) + (T-s)g_2(s) + g_1(s) - \frac{1}{2}s_I(s)\cdot\ s_I(s) \tag{3.50}$$

$$g_4(s) = (T-s)b_I(s) + s_I(s). \tag{3.51}$$

We note that the functions $g_3(t)$ and $g_4(t)$ are deterministic. Based on the above, we obtain the following expression for the T^*-dynamics and the terminal distribution of $I(t)$:

$$d\log I(t) = g_3(t)dt + g_4(t)\cdot dW^{T^*}(t) \tag{3.52}$$

$$dI(t)/I(t) = [g_3(t) + \frac{1}{2}g_4(t)\cdot g_4(t)]dt + g_4(t)\cdot dW^{T^*}(t) \tag{3.53}$$

$$\log\ \frac{I(T)}{I(t)} = \int_t^T g_3(s)ds + \int_t^T g_4(s)\cdot dW^{T^*}(s) \sim \mathcal{N}\left(\int_t^T g_3(s)ds, \int_t^T g_4(s)\cdot g_4(s)ds\right) \tag{3.54}$$

$$I(T) = I(t)e^{\int_t^T g_3(s)ds + \int_t^T g_4(s)\cdot dW^{T^*}(s)}. \tag{3.55}$$

Before moving to price zero-coupon inflation options, we make two observations that leverage on the previous results.

1. It should be noted that the same machinery can be used to find the distribution of the real GDP $X(t)$, which, although is not needed for pricing at this stage, can still be useful to backtest the model. It should be noted that if one wanted to price a growth-linked bond these dynamics would be needed: such bonds have been discussed as a potential way to restructure the Greek public

debt. This means that we write:

$$h_1(t) = -s_X(t) \cdot (\lambda(t) - s_P(t, T^*))$$

$$h_2(t) = a_X(t) - b_X(t) \cdot (\lambda(t) - s_P(t, T^*))$$

$$h_3(t) = m_X(t) + (T-t)h_2(t) + h_1(t) - \frac{1}{2}s_X(t) \cdot \ s_X(t)$$

$$h_4(t) = (T-t)b_X(t) + s_X(t)$$

$$d \log X(t) = h_3(t)dt + h_4(t) \cdot dW^{T^*}(t) \tag{3.56}$$

$$dX(t)/X(t) = [h_3(t) + \frac{1}{2}h_4(t) \cdot h_4(t)]dt + h_4(t) \cdot dW^{T^*}(t) \tag{3.57}$$

$$\log \ \frac{X(T)}{X(t)} = \int_t^T h_3(s)ds + \int_t^T h_4(s) \cdot dW^{T^*}(s) \sim \mathcal{N}\left(\int_t^T h_3(s)ds, \int_t^T h_4(s) \cdot h_4(s)ds\right) \tag{3.58}$$

$$X(T) = X(t)e^{\int_t^T h_3(s)ds + \int_t^T h_4(s) \cdot dW^{T^*}(s)}. \tag{3.59}$$

2. We calculate the instantaneous covariance between the price index relative changes and the short interest rate absolute changes. In fact:

$$\text{Cov}[[I(t+dt) - I(t)]/I(t), n(t+dt) - n(t)] = [(T-t)b_I(t) + s_I(t)] \cdot [-h_x b_X(t) - h_p b_I(t)]/\zeta(t)dt.$$

This result should be compared with the result (2.3.2) in the previous chapter, where we show that the correlation between inflation and nominal rates depends crucially on the central bank reaction function parameters. The formula above has a similar meaning: *ceteris paribus*, the higher the reaction function parameters h_x and h_p are, the higher (in absolute value) the above correlation is (barring compensations coming from the terms $b_X(t)$ and $b_I(t)$). If these parameters are zero, there is no correlation between the two variables. The main result is that this correlation is not taken as a given but is rather a consequence of how the central bank reacts to economic data.

Finally, thanks to the above results we calculate the prices of zero-coupon inflation options in this model.

Lemma 9 *The undiscounted price of a zero-coupon inflation option priced at time t with maturity T and strike K in the CTCB model is*

$$\omega e^{M+1/2V^2} N(\omega(M - (1+K)^{T-t} + V^2)/V) - \omega K N(\omega(M - (1+K)^{T-t})/V) \tag{3.60}$$

where $N(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} e^{-\frac{s^2}{2}} ds$ *and* $\omega \in \{-1, 1\}$ *for puts and calls respectively. Further,*

$$M = \int_t^T g_3(s) ds$$

$$V^2 = \int_t^T g_4(s) \cdot g_4(s) ds.$$

Proof. Using result (3.37) and the distribution of the logarithm of the price index obtained in (3.54) one obtains the above result.

3.7 Pricing year-on-year inflation swaps and options

In this section we discuss year-on-year payoffs, that are model dependent. A convexity adjustment has to be introduced to take into account the co-movement of the nominal interest rate (used for discounting between times t and T_i) and the price index: here we calculate it for the CTCB model.
The calculation of the year-on-year forward follows these steps:

1. Calculation of the real bond volatility in the CTCB model;
2. Calculation of the inflation forward volatility in the CTCB model;
3. Derivation of the dynamics for the ratio of two geometric Brownian motions;
4. Change of measure in the inflation forwards dynamics to ensure they are expressed in the same forward measure;
5. Derivation of the dynamics of the ratio of the two inflation forwards, thanks to the results obtained in the above two steps;
6. Proof that the year-on-year inflation can be represented as the expectation of the ratio of two inflation forwards, whose dynamics are obtained in the above step.

Step 1 – The real economy, as defined at the beginning of this chapter, is used in this analysis as a calculation device and is not an essential feature of the model. We obtain the dynamics of the real bond, defined as:

$$P^r(t,T) = \mathbb{E}_t^{\mathbb{Q}}[I(T)/I(t) e^{-\int_t^T n(s) ds}] = P(t,T) \mathbb{E}_t^{\mathbb{Q}^T}[I(T)/I(t)] =$$

$$= P(t,T) e^{\int_t^T g_3(s) + \frac{1}{2} g_4(s) \cdot g_4(s) ds} = P(t,T) e^{\int_t^T [m_I(t) + g_5(s)] ds} = P(t,T) e^{m_I(t)(T-t) + \int_t^T g_5(s)\, ds}$$

where we define $g_5(s) = g_3(s) + \frac{1}{2}g_4(s) \cdot g_4(s) - m_I(t)(T-t) = g_1(s) + (T-s)g_2(s) + \frac{1}{2}g_4(s) \cdot g_4(s) - \frac{1}{2}$ $s_I(s) \cdot \ s_I(s)$.

The function $g_5(s)$ is deterministic. Therefore we can say that the real bond price $P^r(t,T)$ is a function of the inflation expectation drift $m_I(t)$ and the nominal bond $P(t,T)$: both quantities are stochastic and their SDEs are known. By applying Ito's lemma, taking into account the dynamics of $P(t,T)$ and $m_I(t)$, we obtain:

$$dP^r(t,T) = P^r(t,T)[(...)dt \ + \sigma_{P^r}(t,T) \cdot dW^{\mathbb{Q}^T}(t)]$$

where $\sigma_{P^r}(t,T) = \sigma_P(t,T) + b_I(t)(T-t)$. We are not interested in the drift component, but only in the diffusion term. This result is obtained by explicitly calculating the diffusion term:

$$\left(\frac{\partial P^r(t,T)}{\partial P(t,T)} P(t,T)\sigma_P(t,T) + \frac{\partial P^r(t,T)}{\partial m_I(t)} b_I(t)\right) \cdot dW^{\mathbb{Q}^T}(t) =$$

$$\left(e^{m_I(t)(T-t)+\int_t^T g_5(s)\,ds} P(t,T)\sigma_P(t,T) + (T-t)P(t,T)e^{m_I(t)(T-t)+\int_t^T g_5(s)\,ds} b_I(t)\right) \cdot dW^{\mathbb{Q}^T}(t) =$$

$$P^r(t,T)(\sigma_P(t,T) + b_I(t)(T-t)) \cdot dW^{\mathbb{Q}^T}(t).$$

Step 2 – We build a T-forward martingale by defining a portfolio with a zero-coupon inflation swap with notional *I(t)* and maturity T, and divide by the numeraire, i.e. the nominal bond $P(t,T)$. We recall a model-independent result that states that the present value (PV) of a zero-coupon inflation swap is the difference between the real and nominal bond of the same maturity (see for reference 1.2.4). We get:

$$I(t)(P^r(t,T) - P(t,T))/P(t,T) = I(t)(P^r(t,T)/P(t,T) - 1).$$

We focus our attention on the quantity $I(t)P^r(t,T)/P(t,T)$, known as the forward price index: $\hat{I}(t,T) = I(t)P^r(t,T)/P(t,T)$.

The reason why this is called forward price index is clear if one makes the following observation:

$$P^r(t,T) = \mathbb{E}_t^{\mathbb{Q}}[I(T)/I(t)e^{-\int_t^T n(s)ds}] = P(t,T)/I(t)\mathbb{E}_t^{\mathbb{Q}^T}[I(T)].$$

One obtains:

$$\hat{I}(t,T) = I(t)P^r(t,T)/P(t,T) = \mathbb{E}_t^{\mathbb{Q}^T}[I(T)].$$

Obviously $\hat{I}(T,T) = I(T)$.

By using Ito's Lemma on $\hat{I}(t,T) = I(t)P^r(t,T)/P(t,T)$, we obtain its risk-neutral dynamics. Again, we confirm that in the T-forward dynamics the forward price index has to be a positive martingale:

$$d\hat{I}(t,T) = \hat{I}(t,T)s_{\hat{I}}(t,T) \cdot dW^{\mathbb{Q}^T}(t)$$

where $s_{\hat{I}}(t,T)$ is determined from the other model volatilities via Ito's Lemma in the way showed below.

In particular one obtains:

$$s_{\hat{I}}(t,T) = s_I(t) + b_I(t)(T-t).$$

To show this, one applies Ito's lemma to get the diffusion part of the forward price index $\hat{I}(t,T) = I(t)P^r(t,T)/P(t,T)$:

$$\left(\frac{\partial \hat{I}(t,T)}{\partial I(t)} I(t)s_I(t) + \frac{\partial \hat{I}(t,T)}{\partial P^r(t,T)} P^r(t,T)\sigma_{P^r}(t,T) + \frac{\partial \hat{I}(t,T)}{\partial P(t,T)} P(t,T)\sigma_P(t,T)\right) \cdot dW^{\mathbb{Q}^T}(t) =$$

$$\left(\frac{P^r(t,T)}{P(t,T)} I(t)s_I(t) + \frac{I(t)}{P(t,T)} P^r(t,T)\sigma_{P^r}(t,T) - \frac{I(t)P^r(t,T)}{(P(t,T))^2} P(t,T)\sigma_P(t,T)\right) \cdot dW^{\mathbb{Q}^T}(t) =$$

$$\left(\hat{I}(t,T)s_I(t) + \hat{I}(t,T)\sigma_{P^r}(t,T) - \hat{I}(t,T)\sigma_P(t,T)\right) \cdot dW^{\mathbb{Q}^T}(t) =$$

$$\left(s_I(t) + \sigma_P(t,T) + b_I(t)(T-t) - \sigma_P(t,T)\right) \hat{I}(t,T) \cdot dW^{\mathbb{Q}^T}(t) = s_I(t) + b_I(t)(T-t).$$

This final step is possible thanks to the expression of the diffusion term of the real bond found in step 1: $\sigma_{P^r}(t,T) = \sigma_P(t,T) + b_I(t)(T-t)$.

Step 3 – By a simple application of Ito's lemma we show a general result: taken some deterministic and regular functions a, b, and s, (here a is a scalar function, b and s are vectorial functions with the same dimension of the driving Brownian motion $W(t)$), if one has two SDEs defined as $dX(t) = X(t)s \cdot dW(t)$ and $dY(t) = Y(t)[a\,dt + b \cdot dW(t)]$, the ratio $Z(t) = X(t)/Y(t)$ has dynamics:

$$dZ(t) = Z(t)[(-a + b \cdot b - s \cdot b)dt + (s-b) \cdot dW(t)].$$

The above result is used in step 5.

Step 4 – Similarly to what is done for the BGM model, one chooses a reference tenor T^* and changes the dynamics of the inflation forwards to the same forward measure (see for example Belgrade & Benhamou [9]). The dynamics of the inflation forwards were found in step 2: therefore we know explicitly the dynamics of $\hat{I}(t,T_i)$ and $\hat{I}(t,T_j)$. For example, if the reference tenor is T_i one obtains the following

dynamics for the inflation forwards at tenors T_i and T_j ($T_i > T_j$):

$$d\hat{I}(t,T_i) = \hat{I}(t,T_i)s_{\hat{I}}(t,T_i) \cdot dW^{\mathbb{Q}^{T_i}}(t)$$

$$d\hat{I}(t,T_j) = \hat{I}(t,T_j)(-(\sigma_P(t,T_i) - \sigma_P(t,T_j)) \cdot s_{\hat{I}}(t,T_j)dt + s_{\hat{I}}(t,T_j) \cdot dW^{\mathbb{Q}^{T_i}}(t)).$$

Step 5 – One introduces the price index ratio process $\mathcal{I}(t,T_j,T_i) = \hat{I}(t,T_i)/\hat{I}(t,T_j)$: using Ito's lemma and the results found at step 3 above one obtains its dynamics.

$$d\mathcal{I}(t,T_j,T_i) = \mathcal{I}(t,T_j,T_i)[((\sigma_P(t,T_i) - \sigma_P(t,T_j)) \cdot \hat{s}_I(t,T_j)$$

$$+s_{\hat{I}}(t,T_j) \cdot s_{\hat{I}}(t,T_j) - s_{\hat{I}}(t,T_i) \cdot s_{\hat{I}}(t,T_j))dt$$

$$+(s_{\hat{I}}(t,T_i) - s_{\hat{I}}(t,T_j)) \cdot dW^{\mathbb{Q}^{T_i}}(t)].$$

By assuming $t \leq t_h < T_j < T_i$ we can write the expectation of the ratio as:

$$\mathbb{E}_t^{\mathbb{Q}^{T_i}}[\mathcal{I}(t_h,T_j,T_i] = \mathcal{I}(t,T_j,T_i)e^{\int_t^{T_j}((\sigma_P(u,T_i)-\sigma_P(u,T_j))\cdot\hat{s}_I(u,T_j)+s_{\hat{I}}(u,T_j)\cdot s_{\hat{I}}(u,T_j)-s_{\hat{I}}(u,T_i)\cdot s_{\hat{I}}(u,T_j))du}.$$

Step 6 – We link the price index ratio to the year-on-year payoff: the year-on-year forward can be expressed as an expectation of $\mathcal{I}$:

$$\mathbb{E}_t^{\mathbb{Q}^{T_i}}[I(T_i)/I(T_j)] = \mathbb{E}_t^{\mathbb{Q}^{T_i}}[\hat{I}(T_i,T_i)/\hat{I}(T_j,T_j)] =$$

$$\mathbb{E}_t^{\mathbb{Q}^{T_i}}[\mathbb{E}_{T_j}^{\mathbb{Q}^{T_i}}[\hat{I}(T_i,T_i)/\hat{I}(T_j,T_j)]] = \mathbb{E}_t^{\mathbb{Q}^{T_i}}[\hat{I}(T_j,T_i)/\hat{I}(T_j,T_j)] =$$

$$\mathbb{E}_t^{\mathbb{Q}^{T_i}}[\mathcal{I}(T_j,T_j,T_i] = \mathcal{I}(t,T_j,T_i)e^{\int_t^{T_j}((\sigma_P(u,T_j)-\sigma_P(u,T_i))\cdot\hat{s}_{\hat{I}}(u,T_j)+s_{\hat{I}}(u,T_j)\cdot s_{\hat{I}}(u,T_j)-s_{\hat{I}}(u,T_i)\cdot s_{\hat{I}}(u,T_j))du} =$$

$$= \frac{\hat{I}(t,T_i)}{\hat{I}(t,T_j)}e^{\int_t^{T_j}((\sigma_P(u,T_i)-\sigma_P(u,T_j))\cdot s_{\hat{I}}(u,T_j)+s_{\hat{I}}(u,T_j)\cdot s_{\hat{I}}(u,T_j)-s_{\hat{I}}(u,T_i)\cdot s_{\hat{I}}(u,T_j))du}.$$

We also write the distribution of the logarithm of the price index ratio process $\mathcal{I}(t,T_j,T_i) = \hat{I}(t,T_i)/\hat{I}(t,T_j)$:

$$\log\mathcal{I}(t,T_j,T_i) \sim \mathcal{N}(\int_t^{T_j}((\sigma_P(u,T_i) - \sigma_P(u,T_j)) \cdot \hat{s}_I(u,T_j) + s_{\hat{I}}(u,T_j) \cdot s_{\hat{I}}(u,T_j) - s_{\hat{I}}(u,T_i) \cdot s_{\hat{I}}(u,T_j))du+ \tag{3.61}$$

$$\int_{T_j}^{T_i} g_3(s)ds, \int_{T_j}^{T_i}((\sigma_{\hat{I}}(u,T_i) - \sigma_{\hat{I}}(u,T_j)) \cdot ((\sigma_{\hat{I}}(u,T_i) - \sigma_{\hat{I}}(u,T_j))du).$$

Finally, thanks to the lognormality result shown above, we calculate the prices in this model of a year-on-year caplet.

Lemma 10 *The undiscounted price of a year-on-year inflation caplet/floorlet priced at time t with strike K in the CTCB model is*

$$\omega e^{M+1/2V^2} N(\omega(M-(1+K)+V^2)/V) - \omega K N(\omega(M-(1+K))/V) \tag{3.62}$$

where $N(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} e^{-\frac{t^2}{2}} dt$ *and* $\omega \in \{-1, 1\}$ *for floorlets and caplets respectively. The year-on-year inflation is calculated between times* T_j *and* T_i*. Further,*

$$M = \int_t^{T_j} ((\sigma_P(u,T_i) - \sigma_P(u,T_j)) \cdot s_{\hat{I}}(u,T_j) + s_{\hat{I}}(u,T_j) \cdot s_{\hat{I}}(u,T_j) - s_{\hat{I}}(u,T_i) \cdot s_{\hat{I}}(u,T_j))du + \int_{T_j}^{T_i} g_3(u)du$$

$$V^2 = \int_{T_j}^{T_i} ((\sigma_{\hat{I}}(u,T_i) - \sigma_{\hat{I}}(u,T_j)) \cdot ((\sigma_{\hat{I}}(u,T_i) - \sigma_{\hat{I}}(u,T_j))du$$

Proof. Using result (3.37), the result found in step 6, and the distribution of the logarithm of the price index ratio obtained in (3.7) one obtains the above result.

3.8 Single currency derivatives pricing simulation

To test the results found in the previous sections, we implement a Monte Carlo simulation to check the closed forms for zero-coupon and year-on-year inflation options. We have run 20,000 simulations over 10 years, and here we show the results, the standard error and the closed form results. We price caps with strikes 0, 1, 2, 3, 4, 5 percent with maturities from 1 to 10 years. We assume that the dimensionality of the driving Brownian motion is 3.

For this simulation, we assume the following set of parameters, that are constant over time: $a_X(t) = 0\%$, $a_I(t) = 0.5\%$, $b_X(t) = 0\%$, $b_I(t) = 0.3\%$, $s_X(t) = 0\%$, $s_I(t) = 0.3\%$, $\sigma_P(t,T) = 1\%$, $\lambda(t) = 0\%$, $\mu_I(0) = 0\%$; in case of vector functions, like the volatilities, we assume that the value is the same for all 3 components. For this analysis we have only presented the parameters that are directly relevant for the pricing of inflation derivatives: a full calibration exercise is presented in the following chapter.

The data show that there is good agreement between the Monte Carlo simulation (below referred to as "MC PV") and the closed forms (below referred to as "PV - form"), and that the number of simulations is high enough to control the numerical error (below referred to as "MC error"). The results for zero-coupon

options are the following (strikes in columns, maturities in rows):

MC PV	0%	1%	2%	3%	4%	5%
1	0.00209	0.00006	0	0	0	0
2	0.00772	0.00058	0	0	0	0
3	0.01804	0.00268	0.00009	0	0	0
4	0.03359	0.00784	0.00056	0.00001	0	0
5	0.05477	0.01731	0.00219	0.00009	0	0
6	0.08208	0.03225	0.00625	0.00046	0.00002	0
7	0.11599	0.05352	0.01421	0.00167	0.00009	0
8	0.15721	0.08208	0.02753	0.00475	0.00038	0.00002
9	0.20641	0.1188	0.04775	0.01112	0.00129	0.00008
10	0.26444	0.16451	0.07619	0.02256	0.00363	0.00032

MC Error	0%	1%	2%	3%	4%	5%
1	0.00002	0	0	0	0	0
2	0.00006	0.00002	0	0	0	0
3	0.00012	0.00005	0.00001	0	0	0
4	0.00018	0.0001	0.00003	0	0	0
5	0.00027	0.00018	0.00006	0.00001	0	0
6	0.00036	0.00027	0.00012	0.00003	0	0
7	0.00047	0.00039	0.00021	0.00007	0.00002	0
8	0.0006	0.00052	0.00033	0.00014	0.00004	0.00001
9	0.00074	0.00068	0.00049	0.00024	0.00008	0.00002
10	0.00091	0.00086	0.00067	0.00038	0.00015	0.00004

PV - form	0%	1%	2%	3%	4%	5%
1	0.00212	0.00006	0	0	0	0
2	0.00779	0.00059	0.00001	0	0	0
3	0.01814	0.00271	0.00009	0	0	0
4	0.03374	0.00786	0.00057	0.00001	0	0
5	0.05502	0.0174	0.00224	0.00009	0	0
6	0.08244	0.03244	0.00631	0.00047	0.00001	0
7	0.11649	0.05391	0.01427	0.00172	0.00008	0
8	0.15781	0.08265	0.02766	0.00483	0.00038	0.00001
9	0.20712	0.11952	0.04799	0.01121	0.00133	0.00007
10	0.26534	0.16545	0.07674	0.02262	0.00371	0.00031

Difference: PV - form, MC sim.	0	1%	2%	3%	4%	5%
1	-0.00003	0	0	0	0	0
2	-0.00006	-0.00001	0	0	0	0
3	-0.0001	-0.00003	0	0	0	0
4	-0.00015	-0.00003	-0.00001	0	0	0
5	-0.00025	-0.00009	-0.00005	0	0	0
6	-0.00036	-0.00019	-0.00007	-0.00002	0	0
7	-0.0005	-0.00038	-0.00006	-0.00005	0.00001	0
8	-0.0006	-0.00057	-0.00013	-0.00008	0	0
9	-0.00071	-0.00073	-0.00025	-0.00009	-0.00004	0.00001
10	-0.0009	-0.00095	-0.00055	-0.00006	-0.00008	0.00001

The results for year-on-year options are the following:

MC PV	0%	1%	2%	3%	4%	5%
1	0.0021	0.00006	0	0	0	0
2	0.00621	0.00114	0.00007	0	0	0
3	0.01081	0.00375	0.00066	0.00005	0	0
4	0.01562	0.00736	0.00229	0.00041	0.00004	0
5	0.02055	0.01159	0.0049	0.00141	0.00025	0.00003
6	0.02549	0.01611	0.0083	0.00322	0.00087	0.00015
7	0.03051	0.0209	0.01233	0.00588	0.00212	0.00055
8	0.03558	0.02581	0.01672	0.00922	0.0041	0.00142
9	0.04066	0.0308	0.02139	0.01311	0.00681	0.00288
10	0.04566	0.03575	0.02613	0.01731	0.01005	0.00493

MC Error	0%	1%	2%	3%	4%	5%
1	0.00002	0	0	0	0	0
2	0.00004	0.00002	0	0	0	0
3	0.00006	0.00004	0.00002	0	0	0
4	0.00007	0.00006	0.00003	0.00001	0	0
5	0.00008	0.00007	0.00005	0.00003	0.00001	0
6	0.00009	0.00008	0.00007	0.00004	0.00002	0.00001
7	0.0001	0.00009	0.00008	0.00006	0.00004	0.00002
8	0.00011	0.0001	0.00009	0.00008	0.00005	0.00003
9	0.00011	0.00011	0.00011	0.00009	0.00007	0.00004
10	0.00012	0.00012	0.00012	0.0001	0.00008	0.00006

PV - form	0%	1%	2%	3%	4%	5%
1	0.00207	0.00006	0	0	0	0
2	0.00553	0.00049	0	0	0	0
3	0.01045	0.00304	0.00033	0.00001	0	0
4	0.01542	0.00685	0.00176	0.00021	0.00001	0
5	0.02044	0.01125	0.0044	0.00106	0.00014	0.00001
6	0.02548	0.01595	0.00793	0.00282	0.00065	0.00009
7	0.03054	0.02081	0.01205	0.00548	0.00182	0.00041
8	0.0356	0.02576	0.01654	0.00888	0.00375	0.00118
9	0.04066	0.03076	0.02125	0.01282	0.00643	0.00255
10	0.04573	0.03579	0.02609	0.01714	0.00974	0.0046

Difference: PV - form, MC sim.	0%	1%	2%	3%	4%	5%
1	0.00003	0	0	0	0	0
2	0.00068	0.00065	0.00006	0	0	0
3	0.00036	0.0007	0.00034	0.00004	0	0
4	0.00019	0.00051	0.00053	0.0002	0.00003	0
5	0.00011	0.00034	0.0005	0.00035	0.00011	0.00002
6	0	0.00016	0.00036	0.00041	0.00022	0.00006
7	-0.00002	0.00008	0.00028	0.0004	0.00031	0.00014
8	-0.00002	0.00005	0.00018	0.00033	0.00035	0.00024
9	0	0.00004	0.00014	0.00029	0.00038	0.00032
10	-0.00006	-0.00004	0.00004	0.00017	0.0003	0.00033

3.9 Extension to the open economy

The framework we propose also allows to price inflation derivatives that are struck in a different currency. To do this, one defines the quantities reviewed in the previous sections also for the foreign economy and then introduces the domestic risk-neutral process for the FX rate $\{Y(t)\}_{t\geq 0}$, expressed in the FORDOM convention (i.e. one unit of foreign currency buys FORDOM units of domestic currency). One assumes that the foreign economy works in a similar way, that there is a foreign central bank and that there is a liquidity relationship in the foreign bond market between foreign bond prices and foreign money supply. All parameters for the foreign economy variables are denoted in a way similar to the one used in the

domestic one, with an index f. The dynamics of the foreign assets and other quantities are:

$$dX^f(t)/X^f(t) = (m_{X^f}(t) - \lambda^f(t) \cdot s_{X^f}(t)\ )dt\ + s_I(t) \cdot\ dW^{\mathbb{Q}^f}(t)$$

$$dI^f(t)/I^f(t) = (m_{I^f}(t) - \lambda^f(t) \cdot s_{I^f}(t)\ )dt\ + s_I(t) \cdot\ dW^{\mathbb{Q}^f}(t)$$

$$dm_{X^f}(t) = (a_{X^f}(t) - \lambda^f(t) \cdot b_{X^f}(t))dt + b_{X^f}(t) \cdot\ dW^{\mathbb{Q}^f}(t)$$

$$dm_{I^f}(t) = (a_{I^f}(t) - \lambda^f(t) \cdot b_{I^f}(t))dt + b_{I^f}(t) \cdot\ dW^{\mathbb{Q}^f}(t)$$

$$dP^f(t,T)/P^f(t,T) = n^f(t)dt\ + \sigma_{P^f}(t,T) \cdot\ dW^{\mathbb{Q}^f}(t)$$

$$dn^f(t) = [f_1^f(t) - f_2^f(t)n^f(t) - \lambda^f(t) \cdot \sigma_{n^f}(t)]dt + \sigma_{n^f}(t) \cdot dW^{\mathbb{Q}^f}(t)$$

$$dY(t)/Y(t) = (n(t) - n^f(t))dt\ + s_Y(t) \cdot\ dW^{\mathbb{Q}}(t).$$

We are assuming that the same Brownian motion drives both the domestic and the foreign economy. The parameters for the foreign short rate dynamics are defined in the same way the domestic ones were defined:

$$f_2^f(t) = [Z^f(t+\Omega) - Z^f(t)]/\zeta^f(t) \tag{3.63}$$

$$f_1^f(t) = [-h_p^f a_{I^f}(t) - h_x^f a_{X^f}(t)]/\zeta^f(t) \tag{3.64}$$

$$\sigma_n^f(t) = [-h_x^f b_{X^f}(t) - h_p^f b_{I^f}(t)]/\zeta^f(t). \tag{3.65}$$

By changing the numeraire in the foreign economy from $B^f(t)$ to $Y(t)B^f(t)$, one achieves the domestic risk-neutral dynamics for the foreign economic variables. This translates into a change of drift of $s_{(\cdot)}(t) \cdot s_Y(t)$, where $s_{(\cdot)}(t)$ is the Brownian volatility for a generic model variable.

3.10 Uncertain-parameters extension

The model presented in this chapter gets its randomness from an n-dimensional Brownian motion $W(t)$. We extend the theory proposed to the Merton jump-diffusion (JD) case, which adds flexibility to model the inflation options skew. Here we show that the Merton equation can be obtained in the framework proposed above if one assumes that the model has uncertain parameters. An uncertain-parameters model is a model whose parameters can take random values that are known at inception. Normally one assumes that there is a finite number of possible levels for the parameters and that the parameter set is determined, loosely speaking, one instant before the process starts. The distributions of the state variables are mixtures of distributions: more details are given in Appendix C and chapter 5. For an

introduction one can also see Brigo & Mercurio [22].

In the Merton model the source of randomness is the process $\Lambda(t) = s_I(t)W(t) + \sum_{i=1}^{N(t)}(J_i - 1)$, where:

1. $N(t)$ is a Poisson process with intensity h independent from the Brownian motion $W(t)$ and from the jump sizes $J_1, J_2, \dots$.

2. The logarithm of the jump size J follows a normal distribution with constant mean μ_J and variance $(\delta_J)^2$. Therefore the expected jump size is $\mathbb{E}[J-1] = e^{\mu_J + \frac{1}{2}(\delta_J)^2} - 1 = k$. The logarithm of the jump size is also independent from the Brownian motion $W(t)$.

3. The drift of the process has been adjusted to take into account the compensator: $\mu_I(t) - hk$.

For simplicity, here we consider a one-dimensional source of randomness.

For example, the equation governing the evolution of the price index would read:

$$dI(t) = I(t)[(\mu_I(t) - hk)dt + d\Lambda^{\mathbb{Q}}(t)] =$$

$$= I(t)[(\mu_I(t) - hk)dt + s_I(t)dW^{\mathbb{Q}}(t) + (J-1)dN^{\mathbb{Q}}(t)].$$

As Merton has showed, because the distribution of J is lognormal, the distribution of $\log[I(T)/I(t_0)]$ is still normal conditional to the event $\{N(T) = n\}$.

Therefore we regard such model as an uncertain-parameters model, where, with probability $\mathbb{Q}(N(T) = n) = e^{-hT}(hT)^n/n!$, the SDE for $I(t)$ is:

$$dI(t) = I(t)[(\mu_I(t) - hk + n(\mu_j + \frac{1}{2}\delta_J^2)/(T-t_0))dt + ((s_I(t))^2 + n(\delta_J)^2/(T-t_0))^{\frac{1}{2}}dW^{\mathbb{Q}}(t)].$$

Therefore the theory developed so far for the Brownian case is extended to the Merton case by making some assumptions regarding uncertain model parameters.

Finally, one notes that if the Merton model parameters are deterministic functions of time, by carefully manipulating the parameters we rewrite a time-varying JD model as a classic (constant parameters) Merton JD. This model captures the inflation options skew in a more satisfactory way, as shown in chapter 5.

Chapter 4

Model calibration and applications

In the previous chapter we built the CTCB model, deduced some no-arbitrage conditions derived from some assumptions on the economy dynamics, central bank policy, policy impact on bond prices, and the no-arbitrage principle in the financial market: we deduced some closed form expressions of inflation zero-coupons and year-on-year options. Further, the model implies some short rates dynamics that are consistent with the Hull-White model, and thus one obtains closed form expressions for interest rates caps and floors.

In this chapter we propose a strategy to calibrate the model to market observables by finding suitable parameters, and show some practical applications. Two main advantages become apparent: firstly, the CTCB model is analytically tractable and therefore the calibration process is separable and does not require intensive computation. Secondly, because the model is based on economic theory, we run some economic stress scenarios and obtain the answers directly from the model itself, without having to make assumptions on how an economic shock would impact on financial quantities such as inflation and rates volatilities.

From a numerical perspective, the at-the-money calibration only requires some zero-finding routines, which are not computationally intensive: one searches for some model parameters such that the difference between market observables and model prices is zero.

4.1 At-the-money calibration strategy

Here we detail the steps to calibrate the CTCB model proposed in the previous chapter: further, we make some practical assumptions on some functions. These have no impact on the theoretical construction of the model but let one use it in practice. We are calibrating the model at time $t_0 = 0$: we are still making the assumption that the market observables are continuous functions of the maturity, to keep the

notation light (this assumption is removed in the next sections). When calibrating vector parameters, like $b_I(t)$ or $s_I(t)$ to name a few, we do not make any assumptions regarding how the total quantity needed to calibrate the model is split across the single components: this topic is analysed later in this chapter (in 4.1.3) to model correlations. In general, in the calibration process we have minimised the absolute difference between market prices and model prices of the calibration instruments.

4.1.1 Calibration steps: a first strategy

1. One makes explicit the structural parameters of the model, namely the reaction function parameters h_x and h_p, the reaction function targets $\bar{x}$ and $\bar{p}$, the liquidity horizon of the central bank Ω, and the function $Z(T) = e^{\delta T}$ by choosing the parameter $\delta > 0$. In practice, these parameters are to be regarded not as a target for the calibration, but as an input from economic research that is expected to stay constant over time. The GDP volatility $s_X(t)$ can be regarded as an input of the model, and therefore can be estimated using historic data.

2. Once one knows the parameter $\delta > 0$, we write $Z(T) = e^{\delta T}$ and $\zeta(t) = \int_t^{t+\Omega} Z(T)dt = (\delta^{-1})(e^{\delta(t+\Omega)} - e^{\delta t})$. Further one finds the function $\beta(T) = B(0,T) = e^{-\delta T}$: we remind the reader that this function is the one used in the Ansatz $P(t,T) = A(t,T)e^{-n(t)B(t,T)}$ that characterises the nominal bond prices $P(t,T)$ as a function of the short rate $n(t)$. Once $B(0,t)$ is found, from the market bond prices $P(0,T)$ and the market quote for $n(0)$ we deduce the function $A(0,T) = P(0,T)/e^{-n(0)e^{-\delta T}}$. One should remember that the functions $A(t,T)$ and $B(t,T)$ are not core functions of the CTCB model, but are only relevant to its equivalent Hull-White model. Finally, if needed we get the function $B(t,T) = [B(0,T) - B(0,t)]/(\partial B(0,t)/\partial t) = (e^{-\delta T} - e^{-\delta t})/(-\delta e^{-\delta t}) = (\delta^{-1})(1 - e^{-\delta^{(T-t)}})$: this is a standard result of the Hull-White model.

3. By exploiting the fact that a CTCB model implies an equivalent Hull-White model for the short rate $n(t)$, we immediately calculate the mean reversion speed $a(t)$: as proved in the previous chapter, the parametrisation $Z(T) = e^{\delta T}$ implies that the mean reversion speed is constant and equivalent to δ; we recall that

$$a(t) = [Z(t+\Omega) - Z(t)]/\zeta(t) = \frac{Z(t+\Omega) - Z(t)}{\int_t^{t+\Omega} Z(T)dT} = \delta. \tag{4.1}$$

We remind the reader that the Hull-White SDE for the short rate is: $dn(t) = [\theta(t) - a(t)n(t) - \lambda(t)\sigma_n(t)]dt + \sigma_n^*(t)dW^{\mathbb{Q}}(t)$. We notice that we are still missing the short rate volatility $\sigma_n^*(t)$ and the market price of risk $\lambda(t)$ to get the mean reversion level function $\theta(t)$. This function is found in the following steps.

4. One takes the market quotes of at-the-money caps and floors: from these it is straightforward to get the single at-the-money caplets and floorlets. These are sensitive to the interest rate volatility, and can be used to calibrate some CTCB model volatilities: we recall that at time t the price of a caplet with strike K on the Libor between times T_{i-1} and T_i is equivalent to the price of a put option with expiry T_{i-1} on a zero-coupon bond with maturity $T_i > T_{i-1}$. As explained in the previous chapter, the price of such option can be obtained in closed form via a Black-type formula in the Hull-White model, where the total variance used for pricing is:

$$V^2(t, T_{i-1}, T_i) = [\beta(T_i) - \beta(T_{i-1})]^2 \int_t^{T_{i-1}} \left[\frac{\sigma_n^*(u)}{\beta'(u)}\right]^2 du. \tag{4.2}$$

Because we know a closed form for $\beta(T) = B(0,T) = e^{-\delta T}$ and $\beta'(T) = \partial B(0,T)/\partial T = -\delta e^{-\delta T}$, the above formula is written as:

$$V^2(t, T_{i-1}, T_i) = [e^{-\delta T_i} - e^{-\delta T_{i-1}}]^2 \int_t^{T_{i-1}} \left[\frac{\sigma_n^*(u)}{-\delta e^{-\delta u}}\right]^2 du. \tag{4.3}$$

Finally, we recall that in the equivalent Hull-White model the short rate volatility is expressed as:

$$\sigma_n(t) = [-h_x b_X(t) - h_p b_I(t)]/\zeta(t) = -\delta[h_x b_X(t) + h_p b_I(t)]/(e^{\delta(t+\Omega)} - e^{\delta t}). \tag{4.4}$$

One should refer to (3.29) to see how one moves from the scalar original Hull-White volatility $\sigma_n^*(t)$ to the vector short rate volatility in the CTCB model: $[\sigma_n^*(t)]^2 = \sum_{i=1}^n [\sigma_n^i(t)]^2$.

Therefore, at the end of this step we have fully calibrated the CTCB model to the nominal term structure and at-the-money caps/floors volatilities, and found the economic expectation volatility functions $b_X(t)$ and $b_I(t)$: one chooses these functions to ensure that the model at-the-money caps/floors prices match the ones observed in the market. Alternatively, one can specify the function $b_X(t)$ based on historic data and only calibrate $b_I(t)$. If one wanted to use the CTCB model to price nominal rate derivatives, the calibration process could be ended here.

5. A first consequence of the above result is that, exploiting the standard result $\sigma_P(t,T) = -\sigma_n(t)B(t,T)$ in the Hull-White model, we can write explicitly the bond volatilities: these are needed either if one needs to simulate Libor rates $F(t, T_{i-1}, T_i) = (P(t, T_{i-1})/P(t, T_i) - 1)/(T_i - T_{i-1})$ or when building the drift adjustment to move to the T^*-forward measure. Making all dependencies explicit we write

$$\sigma_P(t,T) = -\delta \frac{[-h_x b_X(t) - h_p b_I(t)]}{(e^{\delta(t+\Omega)} - e^{\delta t})}(\delta^{-1})(1 - e^{-\delta^{(T-t)}}) = \frac{[h_x b_X(t) + h_p b_I(t)]}{(e^{\delta(t+\Omega)} - e^{\delta t})}(1 - e^{-\delta^{(T-t)}}). \tag{4.5}$$

6. To conclude the calibration of the nominal rates part of the model, the function $A(t,T)$ can be made explicit: in fact a standard Hull-White result gives the following relationship involving the function $A(t,T)$:

$$\log A(t,T) = \log A(0,T) - \log A(0,t) - B(t,T)\frac{\partial \log A(0,t)}{\partial t} - \frac{1}{2}\left[B(t,T)\frac{\partial B(0,t)}{\partial t}\right]^2 \int_0^t \left[\frac{\sigma_n(s)}{\frac{\partial B(0,s)}{\partial s}}\right]^2 ds. \tag{4.6}$$

This result is extremely helpful to simulate the forward Libor rates without simulating bond prices, but only the short rate. In practice one may need to smooth the function $\log A(0,t)$ by using splines.

7. We calibrate to the inflation volatilities implied by the market. We recall that the total variance of the quantity $\log\left(I(T)/I(0)\right)$ is

$$\int_0^T g_4(s) \cdot g_4(s) ds \tag{4.7}$$

where $g_4(t) = (T-t)b_I(t) + s_I(t)$. Therefore one can find the function $s_I(t)$, under the constraint that we know already the function $b_I(t)$ from the nominal short rate volatility calibration, using the closed forms for inflation zero-coupon options that we found in the previous chapter.

8. At this stage one has enough information to calibrate the model to the inflation breakeven strikes from zero-coupon inflation swaps, remembering that the expectation of the quantity $\log\left(I(T)/I(0)\right)$ is:

$$\int_0^T g_3(s) ds \tag{4.8}$$

where we recall the definitions of $g_1(t) = -s_I(t) \cdot (\lambda(t) - s_P(t,T^*))$, $g_2(t) = a_I(t) - b_I(t) \cdot (\lambda(t) - s_P(t,T^*))$, and $g_3(t) = m_I(0) + (T-t)g_2(t) + g_1(t) - \frac{1}{2}\, s_I(t) \cdot\ s_I(t)$. We remind the reader that these results where found under the T^*-forward measure, hence the term $s_P(t,T^*)$. Therefore we found the market prices of risk function $\lambda(t)$ and the inflation expectation drift function $a_I(t)$. Alternatively, one can specify the market price of risk $\lambda(t)$ based on historic data and calibrate only $a_I(t)$. The former alternative is well suited for relative value analysis, i.e. the trader, based on a view of the economy and the observed market prices, obtains the market prices of risks implied by market prices, and gauges the illiquidity spots or the inconsistencies between the market participants' risk preferences. The latter alternative is more suited to replicate market prices, i.e. the trader makes an assumption on the market risk aversion and obtains the implied paths for the price index and GDP growth expectations.

9. We can write the Hull-White equivalent mean reversion level by exploiting the standard result:

$$\theta(t) = \lambda(t)\sigma_n(t) - \delta\frac{\partial \log A(0,t)}{\partial t} - \frac{\partial^2 \log A(0,t)}{\partial t^2} + \left[\frac{\partial B(0,t)}{\partial t}\right]^2 \int_0^t \left[\frac{\sigma_n(s)}{\frac{\partial B(0,s)}{\partial s}}\right]^2 ds. \tag{4.9}$$

This result is used to find the growth expectation drift function $a_X(t)$ remembering that the mean reversion level in the Hull-White equivalent model is given by:

$$\theta(t) = [-h_p a_I(t) - h_x a_X(t)]/\zeta(t). \tag{4.10}$$

10. The previous two points can be compacted into one, if one assumes to know the expectation drifts $a_X(t)$ and $a_I(t)$ and therefore calibrates only the market price of risk $\lambda(t)$.

11. Finally, one recalls the volatility condition (3.10) to calculate $s_M(t) = h_p s_I(t) + h_x s_X(t)$: these volatilities may be needed to run a full simulation of the model but are not needed to price derivatives.

4.1.2 Calibration steps: an alternative strategy

Here we propose a minor change to the calibration strategy proposed above: this can be introduced to ensure full calibration of the model. We are calibrating the functions $b_X(t)$ and $b_I(t)$ first, based on the market prices of nominal caps and floors (step 4): in a second step (step 7) we find the function $s_I(t)$ that calibrates the market prices of inflation zero-coupon options. In this step there can be a problem, given that the total variance is $\int_0^T g_4(s) \cdot g_4(s) ds$, where $g_4(t) = (T-t)b_I(t) + s_I(t)$: the function $s_I(t)$ in some cases can only increase the total variance given $b_I(t)$, and the function $b_I(t)$ is multiplied by $T-t$, which can lead to excessive implied variance at long maturities. In practice, in some cases the model may not calibrate to inflation zero-coupon options, because it can not reduce the implied variance below a certain threshold. If inflation volatilities are too low, full calibration to the inflation option prices may not be achieved. If one does not need to calibrate the model to inflation options this is not a problem.

To overcome this problem, we suggest to calibrate the functions $b_I(t)$ and $s_I(t)$ to inflation zero-coupon options across all maturities as a first step, and then to use the function $b_X(t)$ to calibrate the nominal caps and floors: the advantage is that calibration is guaranteed in both inflation and caps and floors volatilities. The trader can not mark freely the output expectation volatilities $b_X(t)$, which was possible in the approach proposed originally. We do not have an explicit preference for either approach: the choice depends on whether one wants to control the output expectation volatilities $b_X(t)$ or guarantee a full calibration to inflation option prices.

4.1.3 Variance split and calibration to correlations

In the calibration strategy proposed in the previous section we found some model volatilities, namely $b_I(t)$, $b_X(t)$, $s_I(t)$, and $s_X(t)$. Because these processes are multidimensional with dimension n, given the market quotes to match, there are multiple ways to split the total variance into its components. We regard this fact as an opportunity to calibrate (at least approximately) to an instantaneous correlation structure that the trader can choose.

Let us take the model volatility process $b_I(t)$: all we say for it can be exactly extended to the remaining three processes. We introduce some weights, called $w^i_{b_I(t)}$ with $i = 1, 2, ..., n$ and such that $\sum_{i=1}^n [w^i_{b_I(t)}]^2 = 1$. In theory one can write $v_{b_I(t)} = \sum_{i=1}^n [b^i_I(t)]^2$, where $b^i_I(t)$ is the i-th component of $b_I(t)$: in practice the calibration process proposed above only yields the total variance $v_{b_I(t)}$ that fits observed option prices. One defines $[b^i_I(t)]^2 = v_{b_I(t)}[w^i_{b_I(t)}]^2$: thus the total variance is split according to some pre-defined weights. This is done for all four model volatilities, yielding the total variances $v_{b_I(t)}$, $v_{b_X(t)}$, $v_{s_I(t)}$, and $v_{s_X(t)}$, assuming that one knows the weights $w^i_{b_I(t)}$, $w^i_{b_X(t)}$, $w^i_{s_I(t)}$, and $w^i_{s_X(t)}$.

These four sets of weights can be determined in a way to target a given instantaneous correlation level. Let us define the list variables for which we want to impose a correlation structure. They are the changes in the short rate $dn(t)$ and the relative changes in the price index $dI(t)/I(t)$. Perhaps one may also be interested to impose a correlation structure that includes the relative changes of the real GDP $dX(t)/X(t)$. We assume we know the market-implied 3×3 correlation matrix.

We want to find the weights $w^i_{b_I(t)}$, $w^i_{b_X(t)}$, $w^i_{s_I(t)}$, and $w^i_{s_X(t)}$ such that the instantaneous model correlations are as close as possible to the market-implied correlations: clearly there is a trade-off between the accuracy of this fit and the dimensionality n of the Brownian motion $\{W(t)\}_{t\geq 0}$. A high enough dimensionality n can ensure an exact fit, but this would make the model overparametrised and difficult to manage. The accuracy is measured as the square difference between the market implied correlation $\rho^{MKT}_{a(t),b(t)}(t)$ of the generic variables $a(t)$ and $b(t)$ and the model correlations $\rho^{MOD}_{a(t),b(t)}(t)$: here $a(t) \in \mathcal{V} = \{dn(t), dI(t)/I(t), dX(t)/X(t)\}$ and $b(t) \in \mathcal{V}$.

We know the model volatility functions (i.e. the diffusion terms) for the 3 variables in closed form from the previous chapter, for the short rate change, for the price index relative change, and for the

output relative change respectively:

$$Diffusion(n(t)) = \frac{-\delta[h_x b_X(t) + h_p b_I(t)]}{(e^{\delta(t+\Omega)} - e^{\delta t})}$$

$$Diffusion(I(t)) = (T-t)b_I(t) + s_I(t)$$

$$Diffusion(X(t)) = (T-t)b_X(t) + s_X(t).$$

In general, for two generic driftless scalar real processes $Y(t)$ and $Z(t)$, four real constants a, b, c, f, two n-dimensional vector volatility deterministic real processes $\{s_1(t)\}_{t\geq 0}$ and $\{s_2(t)\}_{t\geq 0}$, and for an n-dimensional Brownian motion $\{W(t)\}_{t\geq 0}$ with independent components, we can assume the following dynamic equations:

$$dY(t) = (as_1(t) + bs_2(t)) \cdot dW(t)$$

$$dZ(t) = (cs_1(t) + fs_2(t)) \cdot dW(t).$$

We drop the time dependency to make the notation lighter and write the above as a sum of component-by-component products. The Brownian motion differential components are denoted by dW_i, while the single volatility components are denoted by s_1^i and s_2^i:

$$dY = a\sum_{i=1}^{n} s_1^i dW_i + b\sum_{i=1}^{n} s_2^i dW_i$$

$$dZ = c\sum_{i=1}^{n} s_1^i dW_i + f\sum_{i=1}^{n} s_2^i dW_i.$$

We substitute the single components using the total variance technique proposed above, in practice by

writing $[s_1^i]^2 = v_1[w_1^i]^2$ and $[s_2^i]^2 = v_2[w_2^i]^2$:

$$dY = a\sum_{i=1}^{n} v_1^{\frac{1}{2}}[w_1^i]dW_i + b\sum_{i=1}^{n} v_2^{\frac{1}{2}}[w_2^i]dW_i$$
$$dZ = c\sum_{i=1}^{n} v_1^{\frac{1}{2}}[w_1^i]dW_i + f\sum_{i=1}^{n} v_2^{\frac{1}{2}}[w_2^i]dW_i.$$

We want to write the instantaneous correlation between $dY(t)$ and $dZ(t)$, written as

$$\rho_{dY(t),dZ(t)}(t) = \frac{\langle dY(t), dZ(t)\rangle}{[\langle dY(t), dY(t)\rangle \langle dZ(t), dZ(t)\rangle]^{\frac{1}{2}}}.$$

By doing the calculations and thanks to the independence of the components of the Brownian motion, one gets:

$$\langle dY, dZ\rangle = \left[a\sum_{i=1}^{n} v_1^{\frac{1}{2}}[w_1^i]dW_i + b\sum_{i=1}^{n} v_2^{\frac{1}{2}}[w_2^i]dW_i\right]\left[c\sum_{i=1}^{n} v_1^{\frac{1}{2}}[w_1^i]dW_i + f\sum_{i=1}^{n} v_2^{\frac{1}{2}}[w_2^i]dW_i\right] =$$

$$(af+bc)v_1^{\frac{1}{2}}v_2^{\frac{1}{2}}\sum_{i=1}^{n}[w_1^i]dW_i\sum_{i=1}^{n}[w_2^i]dW_i + acv_1\sum_{i=1}^{n}[w_1^i]dW_i\sum_{i=1}^{n}[w_1^i]dW_i + bfv_2\sum_{i=1}^{n}[w_2^i]dW_i\sum_{i=1}^{n}[w_2^i]dW_i =$$

$$(af+bc)v_1^{\frac{1}{2}}v_2^{\frac{1}{2}}\sum_{i=1}^{n} w_1^i w_2^i dt + acv_1\sum_{i=1}^{n}[w_1^i]^2 dt + bfv_2\sum_{i=1}^{n}[w_2^i]^2 dt = \left\{(af+bc)v_1^{\frac{1}{2}}v_2^{\frac{1}{2}}\sum_{i=1}^{n} w_1^i w_2^i + acv_1 + bfv_2\right\}dt.$$

For the denominator terms one writes similarly:

$$\langle dY, dY\rangle = \left[a\sum_{i=1}^{n} v_1^{\frac{1}{2}}[w_1^i]dW_i + b\sum_{i=1}^{n} v_2^{\frac{1}{2}}[w_2^i]dW_i\right]\left[a\sum_{i=1}^{n} v_1^{\frac{1}{2}}[w_1^i]dW_i + b\sum_{i=1}^{n} v_2^{\frac{1}{2}}[w_2^i]dW_i\right] =$$

$$(a^2v_1 + b^2v_2)dt + 2abv_1^{\frac{1}{2}}v_2^{\frac{1}{2}}\sum_{i=1}^{n} w_1^i w_2^i dt$$

and

$$\langle dZ, dZ\rangle = \left[c\sum_{i=1}^{n} v_1^{\frac{1}{2}}[w_1^i]dW_i + f\sum_{i=1}^{n} v_2^{\frac{1}{2}}[w_2^i]dW_i\right]\left[c\sum_{i=1}^{n} v_1^{\frac{1}{2}}[w_1^i]dW_i + f\sum_{i=1}^{n} v_2^{\frac{1}{2}}[w_2^i]dW_i\right] =$$

$$(c^2v_1 + f^2v_2)dt + 2cfv_1^{\frac{1}{2}}v_2^{\frac{1}{2}}\sum_{i=1}^{n} w_1^i w_2^i dt.$$

We finally write:

$$\rho_{dY(t),dZ(t)}(t) = \frac{(af+bc)v_1^{\frac{1}{2}}v_2^{\frac{1}{2}}\sum_{i=1}^n w_1^i w_2^i + acv_1 + bfv_2}{[(a^2v_1+b^2v_2)+2abv_1^{\frac{1}{2}}v_2^{\frac{1}{2}}\sum_{i=1}^n w_1^i w_2^i]^{\frac{1}{2}}[(c^2v_1+f^2v_2)+2cfv_1^{\frac{1}{2}}v_2^{\frac{1}{2}}\sum_{i=1}^n w_1^i w_2^i]^{\frac{1}{2}}}.$$

It is clear that the above generic parametrisation is a slight simplification of the format of all SDEs in the CTCB model, and therefore can be used as a general framework (the simplification is that in the above example for clarity we have assumed only two model volatility functions $s_1(t)$ and $s_2(t)$, while the CTCB has four, namely $b_I(t)$, $b_X(t)$, $s_I(t)$, and $s_X(t)$). Here we notice that, if we know some model parameters a, b, c, f and the total variances v_1 and v_2 from the previous calibration step, we choose the weights w_1^i and w_2^i to target a specific correlation level.

For example, to get $\rho^{MKT}_{dn(t),dI(t)/I(t)}(t)$, one writes:

$$\rho^{MKT}_{dn(t),dI(t)/I(t)}(t) = \frac{\langle \sigma_n(t)\cdot dW(t), [(T-t)b_I(t)+s_I(t,T)]\cdot dW(t)\rangle}{[\langle \sigma_n(t)\cdot dW(t), \langle \sigma_n(t)\cdot dW(t)\rangle \langle [(T-t)b_I(t)+s_I(t,T)]\cdot dW(t), [(T-t)b_I(t)+s_I(t,T)]\cdot dW(t)\rangle]^{\frac{1}{2}}}$$

In this framework one writes:

$$\sigma_n(t)\cdot dW(t) = \sum_{i=1}^n -\frac{h_p}{\zeta(t)}b_I^i(t)dW_i(t) - \frac{h_x}{\zeta(t)}b_X^i(t)dW_i(t) = \sum_{i=1}^n -\frac{h_p}{\zeta(t)}w^i_{b_I(t)}\sqrt{v_{b_I}(t)}dW_i(t) - \frac{h_x}{\zeta(t)}w^i_{b_X(t)}\sqrt{v_{b_X}(t)}dW_i(t)$$

$$[(T-t)b_I(t)+s_I(t,T)]\cdot dW(t) = \sum_{i=1}^n (T-t)b_I^i(t)dW_i(t) + s_I^i(t)dW_i(t) =$$

$$= \sum_{i=1}^n (T-t)w^i_{b_I(t)}\sqrt{v_{b_I}(t)}dW_i(t) + w^i_{s_I(t)}\sqrt{v_{s_I}(t)}dW_i(t).$$

By doing the calculations one gets to the final result.

This example shows that all model correlations can be computed in closed form as a function of the known model parameters and the unknown model volatilities weights $w^i_{b_I(t)}$, $w^i_{b_X(t)}$, $w^i_{s_I(t)}$, and $w^i_{s_X(t)}$.

The non-linear optimisation problem can be formalised as follows:

$$\min \sum_{a(t)\in\mathcal{V}} \sum_{b(t)\in\mathcal{V}, b(t)\neq a(t)} [\rho^{MKT}_{a(t),b(t)}(t, w^i_{b_I(t)}, w^i_{b_X(t)}, w^i_{s_I(t)}, w^i_{s_X(t)}) - \rho^{MOD}_{a(t),b(t)}(t)]^2 \tag{4.11}$$

under the constraints: $\sum_{i=1}^n [w^i_{b_I(t)}]^2 = 1$, $\sum_{i=1}^n [w^i_{b_X(t)}]^2 = 1$, $\sum_{i=1}^n [w^i_{s_I(t)}]^2 = 1$, $\sum_{i=1}^n [w^i_{s_X(t)}]^2 = 1$.

4.1.4 The trade-off between smoothness and calibration accuracy

Another aspect to be considered for the calibration is that the calibrated parameter curves (as a function of time t) may not be smooth, which can be a problem if one is after a realistic description of the market.

In fact it may be hard to justify kinks or irregular shapes for some maturities.

If one is prepared to accept a less accurate fit, one can find a parametric curve that minimises the distance between the calibrated time series of one component of the parameter (for example, $b_I^i(t)$) and an idealised and smooth function, for example with shape:

$$(a + bt + ct^2)e^{g(t-s)+h(t-u)^2}$$

where a, b, c, g, h, s, u are real constants that are found to minimise the fitting error.

4.2 At-the-money calibration results

4.2.1 Technical assumptions

We make some operational assumptions to deal with the data, not to be considered part of the core model construction; they are made explicit here. In general, when making choices, we assume we want to maximise the calibration accuracy for pricing purposes, as a market maker would do.

1. We assume that all model functions $b_I(t)$, $s_I(t)$, $a_I(t)$, $b_X(t)$, $s_X(t)$, and $a_X(t)$ as step functions, where the discontinuities are located at the quoted maturities.

2. We linearly interpolate the market observables at equally spaced time steps, where the time interval is one year. The market observables are the nominal interest curve, the inflation zero-coupon curve, the prices of at-the-money caplets, and the prices of at-the-money zero-coupon inflation options.

3. At-the-money caplets are not directly traded in the market, but are recovered as differences between the PV of the at-the-money caps of two maturities.

4. The prices of zero-coupon inflation options are not quoted for at-the-money strikes but for fixed strikes, therefore a second linear interpolation across strikes is done for each maturity.

5. We assume that the market prices of risk are constant and equal to zero for all components: therefore one obtains the risk-neutral paths for the expected inflation and growth rate.

6. We calibrate inflation options first and then nominal caplets, by keeping the function $b_X(t)$ constant. Therefore we use the "alternative strategy" detailed in section 4.1.2. We decide not to smooth the curves of the model parameters as detailed in section 4.1,4, to maximise the calibration precision.

7. The dimensionality of the driving Brownian motion is 3. The choice appears to be a good compromise between model simplicity and calibration flexibility.

8. For the zero-finding routine, we used Newton's method with maximum 5,000 iterations and absolute price difference tolerance of 0.00000001.

9. The integrals such as the ones in (4.3) are approximated using the rectangles method with a time interval of 0.01 years.

10. The weights used in the correlation targeting step to allocate the variance between the different components of the noise source are assumed to be constant over time.

11. The instantaneous correlations assumed are: -60% for interest rates/inflation, -60% for interest rates/growth, and 70% for inflation/growth; they are chosen by following standard economic theory and market sentiment. Higher interest rates reduce growth and reduce inflation. Higher growth normally brings about higher inflation, as the economy is overheating.

4.2.2 Economic assumptions

We make the following assumptions regarding the static model parameters.

Parameter	Level
δ	0.05
h_P	1.75
h_X	2.5
$\bar{p}$	2%
$\bar{x}$	2%
Ω	5

This model parametrisation is certainly subjective, and reflects our view that in 2012 the European Central Bank (ECB) has been attaching more importance to reviving growth than to subduing inflation (therefore $h_X > h_P$). The ECB's official inflation target is 2%, and it is consensus between economists that the long term growth rate of a developed economy should be around 2%: hence we set $\bar{p} = 0.02$ and $\bar{x} = 0.02$.

Finally, up to 2012 the ECB had no tradition of quantitative easing on long maturities (like, for example, the FED): therefore we cap the maturity of the instruments used for monetary policy to 5 years ($\Omega = 5$). The choice of the parameter δ has been made as follows: in the Hull-White model, this parameter is the product between the long-term equilibrium level for the short interest rate and the adjustment speed. Because interest rates in 2012 were low by historic standards, we assume a much higher equilibrium level at 4%: further, an acceptable adjustment speed is 1.25, which yields $\delta = 0.05$. To

check the stability of the calibration, these parameters have been shocked and the model has recalibrated satisfactorily.

4.2.3 Market data

We calibrate the model to the European inflation market as of 7th December 2012, using market data up to the 10 years maturity. Below we show the nominal curve, the inflation zero-coupon breakeven swap curve, the prices of at-the-money caplets, and the prices of at-the-money zero-coupon inflation options.

Maturity (years)	Nominal IR	Inflation ZC B/E	ATM Caplet PV	ATM ZC Infl. Option PV
1	0.0022	0.0152	0.0007	0.0039
2	0.0026	0.016	0.0017	0.0086
3	0.0045	0.0163	0.0044	0.0147
4	0.0063	0.0166	0.0055	0.0234
5	0.0081	0.017	0.0076	0.0317
6	0.01	0.0173	0.0094	0.0402
7	0.0118	0.0176	0.0108	0.0483
8	0.0136	0.0182	0.0119	0.0594
9	0.0152	0.0189	0.0127	0.0696
10	0.0168	0.0195	0.0134	0.079

4.2.4 Correlation targeting

Correlation targeting has been has been achieved by finding the variance weights $w^i_{b_I(t)}$, $w^i_{b_X(t)}$, $w^i_{s_I(t)}$, and $w^i_{s_X(t)}$ under constraints. In practice, given that the dimensionality of the driving Brownian motion is 3, we have to find 8 weights (2 weights for each function, given that the third is calculated using the constraint that the sum of their squares has to be 1). We remind the reader that we assumed that the weights are constant over time. The result of the numerical optimisation is shown below.

	i=1	i=2	i=3
$w^i_{b_I(t)}$	0.20285	0.13219	0.97024
$w^i_{b_X(t)}$	-0.95101	-0.02865	0.30781
$w^i_{s_I(t)}$	0.14035	0.10000	0.98503
$w^i_{s_X(t)}$	0.85195	-0.07168	0.51868

Interestingly the weights for $b_I(t)$ and $s_I(t)$ have the same sign across all the 3 components, which

is consistent with the original idea of the DSGE macroeconomic model, i.e. inflation depends heavily on inflation expectations. Instead, the first term shows different signs for $b_X(t)$ and $s_X(t)$, which is consistent with productivity shocks, i.e. the growth expectations can differ from realised growth rates.

4.2.5 Results

The following model parameters have been found for the price index processes:

Maturity (years)	$b_I^1(t)$	$b_I^2(t)$	$b_I^3(t)$	$s_I^1(t)$	$s_I^2(t)$	$s_I^3(t)$	$a_I(t)$
1	-0.000269	0.000314	0.000911	0.000273	0.005481	0.007732	-0.000059
2	-0.000269	0.000314	0.000911	0.000512	0.010292	0.014518	0.001543
3	-0.000269	0.000314	0.000911	0.000783	0.015726	0.022182	-0.000174
4	-0.000269	0.000314	0.000911	0.001167	0.023437	0.03306	0.000697
5	-0.000269	0.000314	0.000911	0.001317	0.026467	0.037333	0.00026
6	-0.000269	0.000314	0.000911	0.001461	0.029347	0.041396	0.00016
7	-0.000269	0.000314	0.000911	0.001517	0.030486	0.043003	0.00043
8	-0.000269	0.000314	0.000911	0.001919	0.038553	0.054382	0.003361
9	-0.000269	0.000314	0.000911	0.001895	0.038063	0.053691	0.000545
10	-0.000269	0.000314	0.000911	0.001803	0.036218	0.051088	0.00151

The following model parameters have been found for the GDP processes:

Maturity (years)	$b_X^1(t)$	$b_X^2(t)$	$b_X^3(t)$	$s_X^1(t)$	$s_X^2(t)$	$s_X^3(t)$	$a_X(t)$
1	-0.003463	-0.000807	0.001562	0.009875	0.000464	0.001507	-0.009337
2	-0.008657	-0.002016	0.003905	0.009875	0.000464	0.001507	-0.015211
3	-0.024641	-0.005739	0.011115	0.009875	0.000464	0.001507	-0.008325
4	-0.022408	-0.005219	0.010107	0.009875	0.000464	0.001507	-0.013803
5	-0.038067	-0.008867	0.01717	0.009875	0.000464	0.001507	-0.013651
6	-0.045838	-0.010677	0.020675	0.009875	0.000464	0.001507	-0.013985
7	-0.049767	-0.011592	0.022448	0.009875	0.000464	0.001507	-0.013004
8	-0.053907	-0.012556	0.024315	0.009875	0.000464	0.001507	-0.015206
9	-0.05733	-0.013353	0.025859	0.009875	0.000464	0.001507	-0.012411
10	-0.05733	-0.013353	0.025859	0.009875	0.000464	0.001507	-0.009437

In all cases the absolute calibration error has been below the threshold of 0.0000001.

4.3 Applications

In this final section we show how the CTCB model can be used in practice, how it behaves, and why it can be regarded as a better choice in some cases. We start by building intuition on how a different central bank reaction function impacts on inflation first order risk and cross gammas. Further, because we have a single model to explain the dynamics of both inflation and interest rates derivatives, we propose a concrete example of inflation book macro hedging by using interest rates derivatives. Finally, we make comment on stress tests.

4.3.1 Derivatives risk as a function of the central bank reaction function

We price a 2% zero-coupon inflation cap with 10 years maturity and 1 EUR notional in the CTCB model and then shock the central bank reaction function parameters. In this way we assess the impact of a sudden (and not hedgeable) change in the central bank reaction function (or, more practically, of a new president of the central bank who may have different views and attitudes compared to the previous one).

In particular we find that inflation delta (defined as the change in PV when the inflation curve is shifted up by 1 basis point) is not sensitive to the central bank reaction function parameters (we shock separately the parameters h_P and h_X by 0.5 and in both cases the inflation delta stays at 0.04447): this is expected as the sensitivity of an inflation claim to inflation should mainly depend on the inflation level and the payoff and not by the central bank reaction function.

4.3.2 Cross gammas as a function of the central bank reaction function

Let us consider a long at-the-money zero-coupon inflation option with strike K, where at maturity T we receive the performance of the price index if above $(1 + K)^T - 1$. This trade has sensitivity to the inflation curve (long sensitivity, the higher the inflation curve the higher the final payoff). Because we receive a positive sum at maturity, there is sensitivity to the nominal rates curve; higher rates reduce the present value of a fixed payment in the future (referred to as being "long bond" by traders).

The cross gamma between inflation and nominal rates is the rate of change of the nominal rates delta when the inflation curve moves up (or, alternatively, the rate of change of the inflation delta when the nominal interest rates curve moves up): intuitively, if the inflation curve moves up, we receive more money from the inflation option. Therefore we claim that higher inflation brings about a longer bond position (i.e., when rates move up, the position is worth less as a positive future payment is discounted at a higher rate): the cross gamma is a negative one.

We explore how this cross-gamma is affected by a different central bank reaction function for a 10 years 2% inflation zero-coupon cap with 1 EUR notional. Similarly to what is happening for the inflation delta, our results show that the negative cross gamma stays constant at -0.000044 when the central bank reaction function parameters h_P and h_X are shocked up by 0.5 separately. This is expected as sensitivities should not depend on the central bank reaction function, but only on the payoff.

4.3.3 Inflation book macro-hedging in the CTCB model

Let us assume that an investment bank has sold a low strike inflation floor, which is a popular hedge against deflation: for example a macro hedge fund may want to buy protection against a deflation scenario. This trade would probably make a good margin for the bank, given the relative low liquidity of low strike inflation options. However, this would expose the bank to a significant downside risk that is difficult to recycle. An option for the bank would be to buy a nominal interest rates floor, as a macro hedge given that this market is more liquid than the inflation options market. In a low inflation environment interest rates would go down, making money on the long nominal interest rates hedge while losing on the short inflation client trade. Normally investment banks use different models to price nominal rates and inflation trades, and the decision on the amount of nominal hedge to buy to offset the short inflation position is taken in a very informal and imprecise way. This can lead to significant losses due to model risk. We argue that one of the key advantages of the CTCB model is that it offers a global representation of the economy and allows consistent pricing of interest rates and inflation trades with no ambiguity: this is because this hedging problem boils down to how the central bank can affect the nominal yield curve given a deflationary scenario.

For example, we use the calibrated CTCB model to run a Monte Carlo simulation over the maturity of the inflation client trade. One selects the paths where inflation has gone down enough for the short client trade to be in the money, and obtains a conditional distribution for the forward Libor rates given the inflation decrease: by pricing nominal floors in these scenarios, the trader can assess what nominal rates strikes are best used to hedge a deflationary scenario, choose the cheapest strikes, and, most importantly, calculate a scenario-driven hedge ratio. The idea is that a different reaction function has an impact on the co-movement of inflation and interest rates, and therefore makes the proposed macro hedge more or less effective.

We stress that this example is not a pricing application, and therefore there is some profit and loss volatility during the life of the trade, as we are hedging a deflationary scenario that may not materialise in the end. With this in mind, we think that this methodology helps the trader macro-hedge an inflation book in a way that is consistent with some view of the economy and with no model bias, given that the

same model is used to price the inflation client trade and the interest rates macro hedge.

4.3.4 Stress testing in the CTCB model

In recent years, in particular in the wake of the Lehman crisis, regulators have increasingly requested the most systemic financial institutions to run stress tests, i.e. to calculate the impact of a sudden extreme market move on their books. For example, the FED has introduced the CCAR in late 2010 (Comprehensive Capital Analysis and Review): financial institutions calculate the impact of a crisis similar to the second half of 2008. One of the challenges that financial institutions face is to convert the market moves seen in the market and in the economy into a model parametrisation. In some cases regulators can define scenarios based on economic variables, as expectations, growth rates, or inflation rates.

Because the CTCB model takes the economy as an input, the economic shocks can be easily taken as an input and the model itself answers to the questions asked by regulators: there is no need to shock model parameters like for example expectations or their volatilities, given that the model calibration delivers the new set of parameters that fit to the stressed economic conditions.

Chapter 5

Inflation skew modelling

This chapter is about inflation options skew modelling. Volatility skew is a well-known phenomenon: the reader can refer to Gatheral [58] or to Rebonato [103] for comprehensive references across the main asset classes.

The chapter is structured as follows: we put the inflation skew problem into the wider context of volatility skew and discuss the idiosyncracies of the inflation options market. Then we introduce the main "static" inflation models, used by the main investment banks and brokers to price and quote vanilla instruments. We prove a new result: we show an option pricing formula under the assumption that the underlying level follows a Student's t-distribution.

To further develop the theory, we propose a simple strategy to reconcile the inflation skew problem with the classic skew problem for other asset classes, where the underlying can not take negative values: this is the case of equity derivatives, for example. This "equity analogy" opens up the use and the adaptation of all the extensive theory and "dynamic" models that have been developed to price the skew in equity derivatives. We remind the reader that one significant peculiarity of inflation derivatives is that the underlying (the inflation rate) can be negative in case of deflation.

A further source of complexity is that one may want to calibrate the model to the zero-coupon inflation options and/or to year-on-year caps/floors: this means calibrating a model to the terminal and/or to the forward distributions. This problem is equivalent to pricing forward-starting options in equity derivatives, where much theory has been developed to capture market prices of cliquet trades. At the end of the chapter we show that the market prices of year-on-year options are generally inconsistent with zero-coupon options, which confirms that the two markets are somewhat segmented and not liquid.

To better capture these skews we extend the Merton jump-diffusion model (JD) to have time-varying parameters (TV-JD) and uncertain-parameters (UP-TV-JD): this model, with constant and deterministic

parameters, was originally presented in Merton [97] and is fully reviewed in Matsuda [93] and Cheang & Chiarella [39].

We then extend a result relating to uncertain-parameter diffusion models found by Brigo [16] to the jump-diffusion case, showing how an uncertain-parameters time-varying jump-diffusion (UP-TV-JD) model can be compacted into a more tractable time-varying jump-diffusion (TV-JD) model (with no uncertain parameters) by numerically solving an integral partial derivatives equation (IPDE). This last model can be then converted into a constant-parameter jump-diffusion (JD) model, which yields simple closed-forms option prices for vanilla calls and puts, as shown in the original Merton paper. Pricing of more complex structures can be done using IPDE finite-difference techniques or Monte Carlo simulations.

Uncertain-parameter time-varying jump-diffusion models are an extremely powerful tool to calibrate option prices across different strikes and to specify a single consistent dynamic equation across all maturities.

5.1 Putting inflation skew in context

In 1973, the seminal paper by Black & Scholes [14] assumed that the dynamics of the stock price were driven by a geometric Brownian motion with constant and deterministic parameters: this implies that the return distribution is normal and that the price distribution is lognormal. In particular, the diffusion coefficient (also known by market practitioners as "volatility") has always attracted a lot of attention: as a matter of fact, the price of any vanilla European call or put depends positively on the volatility parameter. The intuition is that the option holder has bought some protection that limits his downside and leaves unlimited upside (at least for a call option). If prices become more volatile, the holder can only make more money in the favourable scenarios, while being covered when the market goes against him. Practitioners refer to this fact in many ways, as being "long optionality", or being "long vega", or being "long gamma".

The constant volatility assumption, while allowing Black & Scholes to propose their celebrated formula and to build an extremely elegant framework, becomes untenable as soon as one looks at the time series of returns across any asset class: return data are far from normal and exhibit fat tails and asymmetry, as shown for example in Jondeau, Poon & Rockinger [80]. It can be safely affirmed that the history of option pricing theory, from 1973 to the current days, has been an attempt to extend the elegant delta-hedging, continuous-rebalancing, and no-arbitrage Black & Scholes paradigm while introducing more realistic assumptions, in particular non-Gaussian returns. Some examples include local volatility models, jump-diffusion models, stochastic volatility models, and uncertain-parameters models: more recent attempts include stochastic-local volatility models, stochastic volatility with jumps, or Lèvy models. In all cases

the aim is to obtain closed or semi-closed forms for prices of vanilla options and, ideally, exotic options.

Besides tracking the development of option pricing theory, it should be stressed that in 1973 the Black-Scholes model was used to find the price of an option, which was then regarded as an unknown. The vanilla call was regarded as a derivative, whose price was derived entirely from the properties of its underlying, which is a "primary" security (whose price is instead determined by market demand and supply). Thanks to the important development of option markets that was possible since then (because of the Black & Scholes formula itself), with time traders started pricing options as if they where primary securities as well, i.e. the volatility became a new "primary" asset class and options at different strikes and maturities were offered to counterparties. At that point an interesting phenomenon occurred: if one inverts the Black & Scholes price and obtains its implied volatility, one notices that the low-strikes options tend to be more expensive, i.e. they are priced with a higher volatility, compared to at-the-money options. This phenomenon is known as "volatility skew": the intuition behind this is that market participant are risk-averse and demand much protection against a market crash, thus driving up the prices (and the implied volatilities) of low strikes options. This is more pronounced for short-maturity options. In option markets where the crash can happen in only one direction, like equity, commodity, bond options, the skew tends to be more on the low-strikes side. Interestingly, in the FX options market the skew tends to be more symmetric because in both cases (FX exchange rate moving up or down) one of the two economies is going to suffer a currency shock and therefore some agents need to hedge their downside. If one plots the market-implied volatilities as a function of strike (or moneyness or log-moneyness) one recovers the classical "smile" or "smirk" shape. From this high-level summary it should be clear that the concept of skew is a way to price risk-aversion and non-Gaussianity, and, most importantly, this idea works in a model-dependent way: the implied volatility is the volatility that, assuming volatility is constant and deterministic, gives the market price of options if one plugs it in the Black & Scholes model.

In fixed-income markets, where there is no universally-accepted model to price vanilla bond options, rates future options, caps/floors or swaptions, things become more complicated. Generally, these prices are quoted either in Black-Lognormal volatilities, or in Black-Normal volatilities, or in SABR parameters. These three pricing models are widely known. Operators plug market-implied parameters into these known formulas, and price vanilla instruments: it should be noted that, although in all these cases one can write the dynamics of the asset or the underlying rate, there are different set of parameters for different tenors-maturities. Therefore these models are used only to specify their terminal distributions. Exotic instruments are priced via more complex models, like the Hull-White, the HJM, or the BGM models, whose model parameters are calibrated to match market quotes of selected vanilla instruments.

At this point, to move our focus to inflation derivatives, one should note that, if one uses a Black-

Lognormal model for a forward swap rate, one is assuming that the relative changes of the forward swap rates have a Gaussian distribution. Therefore the level of the forward swap rate has a lognormal distribution: this implies no skew (as returns are Gaussian) and strictly positive rates. Strictly positive rates have been regarded for long time as a desirable feature of any interest rates model, however after the Lehman crisis there have been instances where rates have been negative, initially for short times and in the short end of the curves. Displaced diffusions, presented in Brigo & Mercurio [22] can be used: in this case the underlying can be negative but can not be lower than a certain threshold, which may be difficult to determine.

Some traders also prefer to allow negative rates to introduce some asymmetry and skew into the model. A mathematically simple solution is to assume that the distribution of the rate level is normal: by construction this allows more probability mass to lower rates scenarios, and therefore it introduces downside skew because the return distribution is no longer normal. Further, there are closed forms pricing formulas for options in this case. This said, we show that even a Black-Normal model fails to capture the complexity of the full spectrum of option prices across different strikes. The SABR model, originally presented by Hagan, Kumar, Lesniewski & Woodward [63], offers some additional freedom to calibrate market skews and is widely used in the industry.

Inflation markets are even more complicated, for three reasons: firstly, by construction the underlying can assume negative values (one can have deflation, i.e. a reduction in price levels). Secondly, because there is no widely-accepted model for this asset class, option prices (both zero-coupon options and year-on-year caps/floors) are quoted in price and not in terms of some model parameter(s). Thirdly, data show that different market operators are averse to different types of risks, either deflation or hyperinflation, which drives up the prices of low-strikes and high strikes options and sometimes creating inconsistencies between the cap and floors markets. Liquidity can be a problem and some prices can remain stale for some time. To make this point stronger, we point out how the Bloomberg pricing server allows traders the choice to calibrate their inflation model either to caps or floor, as their prices could be inconsistent.

The following section shows some models used by practitioners to allow negative inflation and calibrate to prices across different strikes for a given maturity. As explained above, these models should be thought as terminal distribution models, even if it is possible to specify an SDE for the inflation rate in some cases, traders use them to fit only the vanilla smiles of the inflation options, each maturity taken separately. Therefore they can not be used to price path-dependent inflation exotics like LPI or real-Bermudan swaptions, but only to quote or interpolate market prices of inflation vanilla instruments.

5.2 Static skew models

In this section we show some models used for the terminal distribution of a given inflation rate. They all assume that:

1. The underlying can assume negative values—therefore these models are suitable to price inflation caplets and floorlets also with negative strikes.

2. Nominal interest rates are deterministic and we operate under the $\mathbb{Q}$-measure or, alternatively, we operate under the T-forward measure: in both cases this makes discounting equivalent to multiplying the expected payoff by the zero-coupon bond price with the same maturity of the option. In both cases the result is the same, as we are writing closed forms for undiscounted payoffs.

3. The forward underlying inflation under the chosen measure is known and equal to $\bar{S}$. This assumption is not always straightforward to make. The at-the-money forward of zero-coupon inflation options is known and is the zero-coupon inflation rate at a given maturity: these curves are liquid and are openly quoted in the market. Problems start when one looks at year-on-year caps/floors. If one assumes that nominal rates are deterministic, the correlation between nominal rate and inflation does not play any role and the year-on-year forward can be calculated off the zero-coupon curve. If one assumes that nominal rates are stochastic, the year-on-year forward is a model-dependent quantity, as shown in Brigo & Mercurio [22] and in chapters 1 and 3: therefore one needs an inflation and nominal rates "dynamic" model to feed the year-on-year forward into the "static" option pricing model, and this loop is contradictory to say the minimum. To tackle this problem, some primary financial institutions use a different dynamic model to obtain by simulation the year-on-year forward inflation: this procedure is quoted here because it used by market practitioners, but it is clearly flawed from a theoretical perspective: two different and potentially inconsistent models can not be used in conjunction to model the same phenomenon.

5.2.1 Inflation option pricing under the t-distribution: a new formula

We derive some closed forms expressions for option prices under the assumptions that the underlying distribution is a Student's t-distribution with scale parameter σ and number of degrees of freedom n. The main advantage of a Student's t-distribution is that it is easier to model fatter tails, as the distribution can even exhibit infinite first moment when there is only one degree of freedom (and the Student's t-distribution becomes a Cauchy distribution). Because option prices are expectations, one needs to request that $n > 1$ so that expectations under the Student's t-distribution exist.

Because the t-distribution converges to the normal distribution when the number of degrees of freedom diverges to infinity, we find helpful to recall the derivation of the undiscounted call price for an option priced at time t with maturity T in the normal case:

$$C(t,T,K) = \mathbb{E}^{\mathbb{Q}}[\max(S-K,0)] = \int_K^{+\infty} (S-K)\frac{1}{\sqrt{2\pi\sigma^2(T-t)}} e^{-\frac{1}{2\sigma^2(T-t)}(S-\bar{S})^2} dS.$$

This integral is solved by changing the variable $\xi = \frac{S-\bar{S}}{\sigma\sqrt{T-t}}$ obtaining the Bachelier formula:

$$C(t,T,K) = \sigma(T-t)^{1/2} f(y) + (\bar{S}-K)(1-F(y)) \tag{5.1}$$

where $y = \frac{K-\bar{S}}{\sigma\sqrt{T-t}}$, $f(S) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(S)^2}$ and $F(S) = \int_{-\infty}^{S} f(t)dt$. It should be noted that, if one wanted to write the dynamics of S, they would be:

$$dS(t) = \sigma dW(t) \quad S(0) = \bar{S}$$

where σ is a real constant and $W(t)$ is a Brownian motion. Clearly, because one starts from the terminal distribution, stating the dynamics is not an essential step in this approach.

The Student's t-distribution case follows the same path but is somewhat more involved as far as the integrals are concerned. The undiscounted call price is written as

$$C(t,T,K) = \mathbb{E}^{\mathbb{Q}}[\max(S-K,0)] = \int_K^{+\infty} (S-K)g(S)dS$$

$$= \int_K^{+\infty} (S-K)\frac{1}{\sigma\sqrt{n\pi(T-t)}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{\left(1+\left(\frac{S-\bar{S}}{\sigma\sqrt{T-t}}\right)^2 \frac{1}{n}\right)^{\frac{n+1}{2}}} dS.$$

Here $\Gamma(n) = \int_0^{+\infty} e^{-x}x^{n-1}dx$. The integral is solved by making two substitutions: $\xi = \frac{S-\bar{S}}{\sigma\sqrt{T-t}}$ and $z = \frac{1}{\left(1+\xi^2\frac{1}{n}\right)^{\frac{n+1}{2}}}$. The final result reads

$$\sigma(T-t)^{1/2} g(y) \frac{n}{n-1}\left(1+\frac{y^2}{n}\right) + (\bar{S}-K)(1-G(y)) \tag{5.2}$$

where $y = \frac{K-\bar{S}}{\sigma\sqrt{T-t}}$, $g(S) = (n\pi)^{-1/2}\Gamma(\frac{n+1}{2})/\Gamma(\frac{n}{2})(1+S^2/n)^{\frac{2}{n+1}}$ and $G(S) = \int_{-\infty}^{S} g(t)dt$. The convergence of the integral is ensured if $n > 1$. The derivations are available in Appendix B.

We also note that the call price in the t-distribution case converges to the price in the normal case, which confirms the validity of this formula: to the best of our knowledge, we are not aware of any previous

similar result in the literature. On a related subject, Cassidy, Hamp & Ouyed [36] provide some pricing formulas in the case where the logarithm of the underlying follows a t-distribution. We do not exploit their result here because we want inflation to take also negative values (to model deflation). More details on the t-distribution can be found in Shaw [107] and Bayley [8]. Prices of put options can be recovered via the call-put parity.

In the last part of this section we discuss some recent theoretical developments that specify some diffusion dynamics that result in the state variable to have a marginal Student's t-distribution. Bibby, Skovgaard & Sørensen [12] show, under some regularity conditions, that the mean-reverting Gaussian process $\{X(t)\}_{t\geq 0}$ that is a solution of the SDE:

$$dX(t) = -\theta(X(t) - \mu)dt + \sqrt{v(X(t))}dW(t)$$

is ergodic and has invariant density $f(x)$ if the local variance function $v(x)$ satisfies the condition:

$$v(x) = \frac{2\theta\mu F(x) - 2\theta \int_l^x yf(y)dy}{f(x)}.$$

The density $f(x)$ is strictly positive in the interval (l, u), has finite variance, is bounded and continuous. Here the positive real constant θ is the mean reversion speed, the real constant $\mu \in (l, u)$ is the mean reversion level, the process $\{W(t)\}_{t\geq 0}$ is a scalar Brownian motion, $F(x)$ is the cumulative density function of $f(x)$.

If one defines the random variable $Y(t)$ by applying a linear transformation to $X(t)$, such that $Y(t) = a + \sigma X(t)$, with $\sigma > 0$ and $a \in \mathbb{R}$, and denotes its invariant density as $g(x)$, one writes:

$$v_g(x) = \sigma^2 v_f(\frac{y-a}{\sigma})$$

where v_f is the squared diffusion term of the process that has $f(x)$ as invariant distribution.

To obtain a mean-reverting process with invariant distribution that follows a Student's t-distribution with n degrees of freedom, if $\mu = 0$ one takes $v(x) = \frac{2\theta}{n-1}(n + x^2)$. Clearly in this case $-l = u = \infty$. We need to require $n > 2$ to ensure that the distribution has finite variance so that it satisfies the regularity conditions detailed above.

So far we have discussed how to find the local variance function $v(x)$ to ensure that the invariant distribution of the process $\{X(t)\}_{t\geq 0}$ is $f(x)$. With these results one specifies the local variance function $v(x)$ to ensure that the marginal distribution of the process $\{X(t)\}_{t\geq 0}$ is $f(x)$. Bibby, Skovgaard &

Sørensen [12] show that this is achieved by requesting

$$v(x) = \frac{2\theta\mu_h f^*(x)}{f(x)}.$$

The meaning of the coefficient μ_h becomes clear later, as it is the expectation of a mixing distribution $h(u)$. This result is obtained by noting that the Student's t-distribution can be regarded as a normal variance mixture with an inverse gamma mixing distribution, which means that we write:

$$f(x) = \int_0^{+\infty} \frac{1}{\sqrt{2\pi u}} e^{-\frac{1}{2}\frac{x^2}{u}} h(u)du.$$

The distribution $f^*(x)$ is the normal variance mixture with mixing density $h^*(u) = uh(u)/\mathbb{E}[u]$. In the case of an inverse gamma mixing distribution, one has $h(u) = \beta^\alpha/\Gamma(\alpha)u^{-\alpha-1}e^{-\frac{\beta}{u}}$, where α and β are some distribution parameters. We remind the reader that, if a random variable Z follows a gamma distribution with parameters α and β^{-1} , its inverse Z^{-1} follows an inverse gamma distribution with parameters α and β.

To use these results in practice one needs to specify the stochastic processes for each the zero-coupon inflation rate or for the year-on-year rate in a way that they satisfy the market calibration results. We stress that at this stage we are building a separate model for each maturity T and that, for each maturity T, we are calibrating the number of degrees of freedom n_T and the scale parameter σ_T to best fit a Student's t-distribution to the traded option prices, using the pricing result found previously.

To specify the dynamics, one writes the inflation rate $p(t)$ as the sum of the random variable $X(t)$ and the deterministic real function $a(t)$: $p(t) = X(t)+a(t)$. The function $a(t)$ is used to calibrate the forward, as shown in Brigo, Morini & Pallavicini [27], and the SDE for $X(t)$ is $dX(t) = -\theta X(t)+\sqrt{v(X(t))}dW(t)$. One needs to specify the initial condition $X(0)$ and the mean-reversion speed $\theta > 0$.

As explained above, one chooses $v(x) = \frac{2\theta\mu_h f^*(x)}{f(x)}$. This ensures that the marginal distribution of $X(t)$ is a Student's t-distribution with the desired number of degrees of freedom and density $f(x)$. If necessary, we apply the linear transformation $Y(t) = a + \sigma X(t)$ proposed above to introduce a scale parameter σ to further improve the fit to the market data.

5.2.2 Inflation option pricing under the SABR model

The SABR model was introduced by Hagan, Kumar, Lesniewski & Woodward [63], and is defined by the dynamics of the underlying $S(t)$ and its volatility $\sigma(t)$ via two SDEs:

$$dS(t) = \sigma(t)S^\beta(t)dW_1(t), \quad S(0) = \bar{S}$$

$$d\sigma(t) = \sigma(t)\alpha dW_2(t), \quad \sigma(0) = \sigma_0 > 0$$

$$dW_1(t)dW_2(t) = \rho dt.$$

Here $W_1(t)$ and $W_2(t)$ are two Brownian motions with instantaneous correlation ρ. The real constant parameters $\alpha \geq 0$, $\beta \in [0,1]$, $\rho \in [-1,1]$ are commonly referred to as "Vol of vol", "beta", and "spot-vol correlation". Brigo & Mercurio [22] fully explain this model and its properties.

For our purposes, it suffices to note that if one sets $\beta = 0$ the underlying can take negative values. Further, if $\alpha = 0$ the SABR model is equivalent to the Black-Normal model reviewed above. A strictly positive α increases the probability of extreme events on both sides of the distribution *ceteris paribus*, while a negative (positive) spot-vol correlation creates negative (positive) skew, attributing more probability mass to low (high) inflation scenarios.

The SABR model is extremely flexible to calibrate smiles, and, thanks to some approximations, we recover the normal Black volatility (to be plugged in formula 5.1 on page 129 and here noted as σ_N) as an approximated function of the option strike K, maturity T, and the parameters σ_0, α, and ρ. This formula is derived in [63] by making use of variational techniques. The formula is:

$$\sigma_N \cong \alpha \frac{\bar{S} - K}{D(\zeta)} \left\{ 1 + \left[\frac{2\gamma_2 - \gamma_1^2}{24} \frac{\sigma_0^2}{\alpha^2} + \frac{\rho\gamma_1\sigma_0}{4\alpha} + \frac{2 - 3\rho^2}{24} \right] T\alpha^2 \right\} \tag{5.3}$$

where one has:

$$\zeta = \frac{\alpha}{\sigma_0(1-\beta)} \left(\bar{S}^{1-\beta} - K^{1-\beta} \right)$$

$$D(\zeta) = \log \left(\frac{(1 - 2\rho\zeta + \zeta^2)^{\frac{1}{2}} + \zeta - \rho}{1 - \rho} \right)$$

$$\gamma_1 = \frac{\beta}{F}$$

$$\gamma_2 = \frac{-\beta(1-\beta)}{F^2}$$

$$F = (\bar{S} + K)/2.$$

To conclude, a word of caution should be issued regarding the above approximations, as they can become unstable under low volatilities and low underlying levels. This problem is well known in the industry.

5.2.3 A first static calibration contest

In the previous sections we presented some models that can be used to statically calibrate the inflation smile for a given maturity. Here we compare their calibration capabilities.

For convenience we list the market data here, so that the reader can reference them when reading the following sections. The data is from Bloomberg for 7th December 2012 for the 2 year maturity:

1. The 2 year riskless bond is priced at 99.98% and the 2 years zero-coupon inflation swap is at 1.6%.

2. Market prices (in basis points) of zero-coupon inflation caps with 2 years maturity:

Strike	2%	3%	4%	6%
Cap Price (bps)	50	16	7	2

3. Market prices (in basis points) of zero-coupon inflation floors with 2 years maturity:

Strike	-1%	0%	1%
Floor Price (bps)	2	9	35

4. Using the call-put parity one obtains the equivalent cap prices:

Strike	-1%	0%	1%	2%	3%	4%	6%
Cap Price (bps)	262	169	85	50	16	7	2

The following graph shows the market option prices and the prices recovered by calibrating three models to the market. Prices are in logarithmic scale. We minimise the sum of squared errors multiplied by 10^{16}. The calibration parameters found are:

1. Black normal: volatility $\sigma = 1.06\%$. Calibration error metric: 32,449.

2. t-distribution: volatility $\sigma = 0.92\%$, degrees of freedom $n = 6$. Calibration error metric: 31,394.

3. SABR normal: volatility $\sigma_0 = 1.04\%$, Vol-of-vol $\alpha = 30\%$, Spot-vol correlation $\rho = 66\%$. Calibration error metric: 9.497.

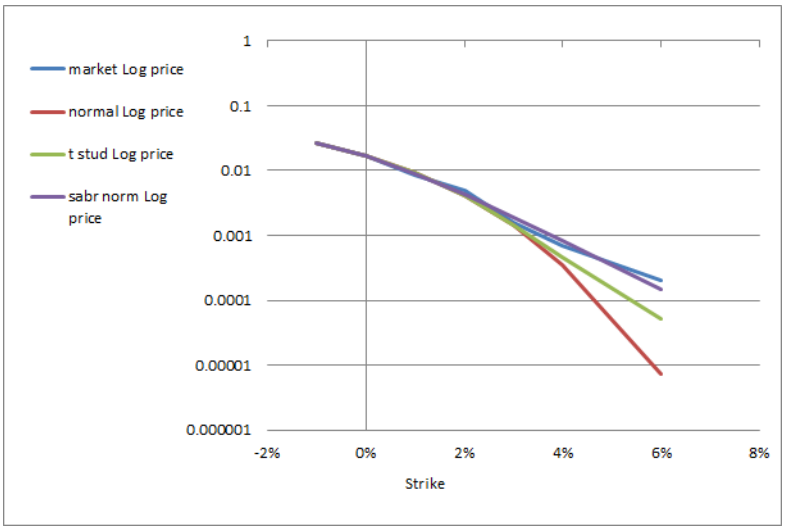


Figure 5.1: Log-option prices, 2Y maturity, across strikes.

From the graph it is clear that the Gaussian model is not the best solution to calibrate the full skew range.

The t-distribution model performs better using only two parameters: the SABR model allows a good calibration but needs the highest number of parameters.

Finally, one notes that in the low strikes region all models seem to be performing well, while there are problems in the high strike region. This may be due to the fact that high strike inflation caps may be bought by different types of users, and therefore the two markets are not perfectly liquid. In the following section we show another and more extreme example of "split" market, and show how a Gaussian mixture can help improving the calibration.

5.2.4 Inflation option pricing under Gaussian mixtures

In this section we show how to calibrate a normal mixture to traded prices of zero-coupon inflation options.

We source from Bloomberg the prices as of 25th July 2012 of options on the European Harmonised Consumer Price index, at different strikes and with 2 year maturity. Floors are available for low strikes (from -1% up to +1%), and caps are available for high strikes (from 2% to 5%).[1] Using the call-put parity, floor prices are converted into equivalent cap prices. Because option prices are available only for a very sparse set of strikes (-1%, 0%, 1%, 2%, 3%, 4%, 5%), we find an interpolating function to fit them. Once this function is estimated, we obtain the probability distribution on the underlying level in 2 years by taking the second derivative of the undiscounted option prices with respect to the strikes, as detailed, for example, in Derman & Kani [49].

This distribution looks bimodal, reflecting the uncertainty of market participants on whether there will be inflation or outright deflation at that time and possibly some illiquidity between the cap and floors markets. From the distribution we fit a bimodal Gaussian mixture by using some numerical optimisation methods (other alternatives are the EF3M algorithm, presented in Lopez de Prado & Foreman [90] or the closed forms presented in Appendix C, due to Bertholon, Monfort & Pegoraro [11]). The bimodality of the distribution makes it particularly interesting to study the market-implied inflation distribution under mixtures.

Below are the data:

1. The 2 year riskless bond is priced at 99% and the 2 year zero-coupon inflation swap is at 1.43%.

2. Market prices (in basis points) of zero-coupon inflation caps with 2 year maturity:

[1] Unlike other markets, the inflation options market quotes prices instead of implied volatilities. Prices are expressed in basis points (abbreviated in bps), where 1 basis point is equivalent to 1%/100.

Strike	2%	3%	4%	5%
Cap Price (bps)	48.5	17.8	7.5	3.5

3. Market prices (in basis points) of zero-coupon inflation floors with 2 year maturity:

Strike	-1%	0%	1%
Floor Price (bps)	1.5	8.5	47.3

4. Using the call-put parity one obtains the equivalent cap prices:

Strike	-1%	0%	1%	2%	3%	4%	5%
Cap Price (bps)	243.5	151.5	91.3	48.5	17.8	7.5	3.5

Below we detail the steps we followed and show the results we obtained:

1. To obtain a continuous function (the option prices as a function of strikes—denoted as $P(K)$), we use the above prices to fit the following functional form for the natural logarithm of the cap prices:

$$\log(P(K)) = a + b\exp((-(K-c)/d)^2) + e\exp((-(K-f)/g)^4)$$

where the parameters a, b, c, d, e, f, and g are real constants. We obtain the following parameters:

Parameter	Value
a	-33.64
b	22.47
c	42.36
d	156.01
e	22.53
f	0.19
g	6.93

The fit is good, as shown in the below table:

Strike	-1%	0%	1%	2%	3%	4%	5%
Cap Price (bps) - MARKET	243.5	151.5	91.3	48.5	17.8	7.5	3.5
Cap Price (bps) - FIT	247.8	153.5	93.4	52.0	19.5	2.9	0.1

The graph below shows the market prices (blue) versus the fitted prices (pink) by strike:

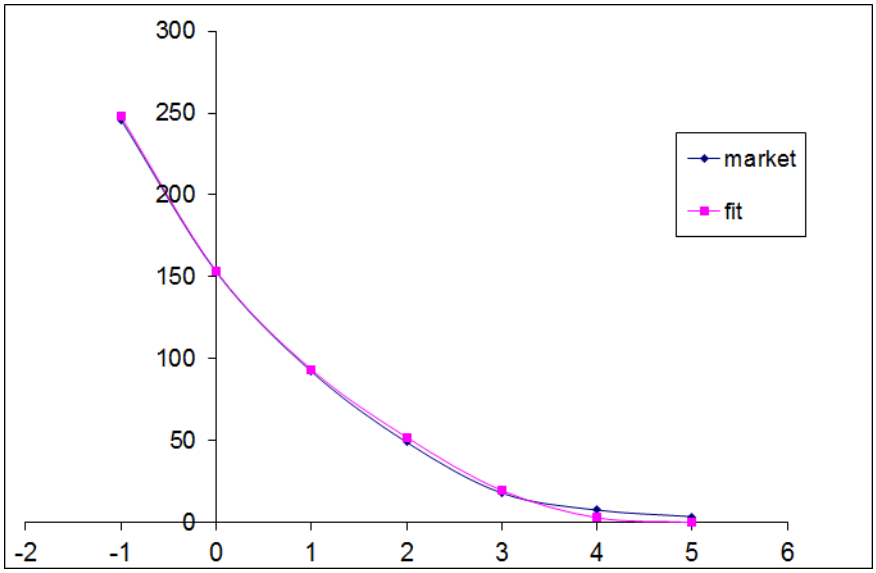


Figure 5.2: Market and fitted prices of 2 years inflation options.

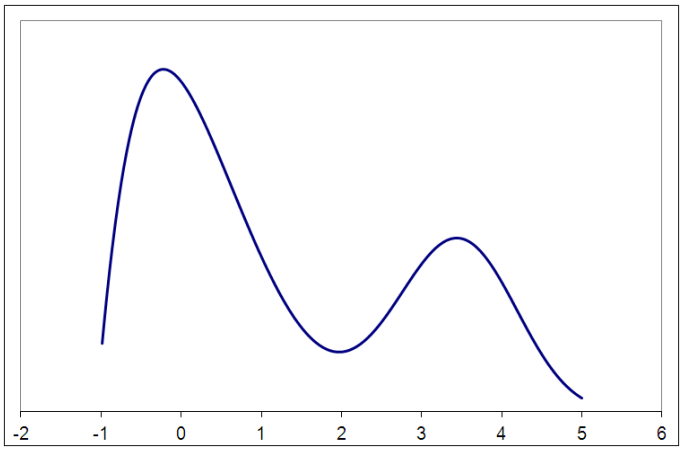


Figure 5.3: Probability distribution implied by option prices in two years.

2. By taking the second derivative of the fit function using finite differences, one obtains the bimodal implied terminal probability distribution in figure 5.3 for the European Harmonised Consumer Price Index in 2 years:

3. From this distribution we calculate the moments:

Moment	Value (%)
1	1.43
2	4.74
3	16.27
4	60.48

4. A numerical optimisation is run to minimise the distance between the actual distribution and the Gaussian mixture, under the first order constraint: $p\mu_1 + (1-p)\mu_2 = \bar{S}$.

Parameter	Level
p	0.413
μ_1	3.431
μ_2	0.017
σ_1	0.912
σ_2	0.646

The distribution fit that one obtains is shown below (dotted pink line, compared against the target distribution in blue solid line):

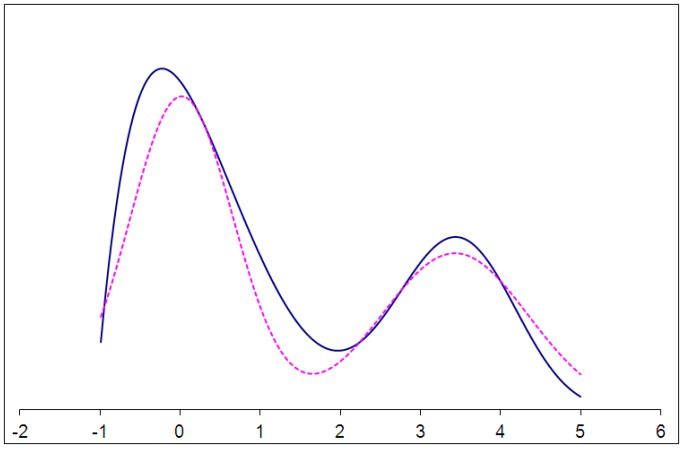


Figure 5.4: Probability distribution implied by option prices in two years — fitted using a bimodal Gaussian mixture.

It should be stressed that, even if the two distributions are not exactly the same, this method offers good computational simplicity and speed, which is a critical factor when pricing books with thousands of positions.

This example shows that the conversion from option prices to a normal mixture can be carried out in a straightforward way. The strong bimodality of this distributions encourages us to use mixtures to model the inflation smile (and, as we will see later in this chapter, uncertain-parameter models, that yield distributions that are mixtures: thus they combine analytical tractability and good fit to the smile).

Finally, one should note the strong bimodality of the distribution: this can be caused either by a genuine market uncertainty regarding what the prevailing scenario will be (either deflation or high inflation) or may mean that the market for high strike caps is somewhat not fully integrated with the market of low strike floors, due to liquidity issues.

5.3 The equity analogy: a new inflation skew metric

The previous sections have shown that there is no scarcity of models that can calibrate a static skew, whereby "static" means "at a given maturity under its terminal distribution". In fact, these models are calibrated only on a specific maturity, and, if rates are stochastic, are used assuming different terminal distributions for each maturity.

This means that these models can not be used to price path-dependent inflation derivatives but only to quote prices for vanilla instruments.

The objective of this section is to propose a different solution to this problem: we write a model used for path-dependent inflation derivatives, leveraging on existing jump-diffusion models. Because most of the skew models have been developed to work with strictly positive underlyings, like equity prices, commodity prices or FX rates, it is natural to convert the inflation underlying $p(t)$ – that can take negative values – into a positive underlying, namely the price index $I(t)$. As we will stress, this is consisent with the CTCB model proposed in chapter 3. To show this, one works with undiscounted option prices for simplicity, as we will do in the following two lemmas. Discounting is then performed by multiplying by the bond price, assuming that either rates are deterministic or that we are working under the T-forward distribution.

Lemma 11 *A zero-coupon inflation cap option with maturity T and strike K, denoted by $ZCC(T,K)$, is equivalent to a European call on the price index $I(T)$ with same maturity T, strike $\kappa = I(t)(1+K)^{T-t}$, and with notional $1/I(t)$, where $I(t)$ is the value of the price index at time t.*

Proof. We write the payoff of the $ZCC(T,K)$ as:

$$ZCC(T,K) = \max[0, (I(T)/I(t) - 1 - ((1+K)^{T-t} - 1)] = \max[0, (I(T)/I(t) - (1+K)^{T-t}] =$$

$$= \frac{1}{I(t)} \max[0, I(T) - I(t)(1+K)^{T-t}].$$

For example, a one year zero-coupon option struck at 5%, when the price index level is at 100, is equivalent to a European call on the price index level with same maturity and strike 105.

Lemma 12 *A year-on-year inflation cap with maturity T_i and strike K, denoted by $YoYC(T_{i-1}, T_i, K)$, is equivalent to a forward-starting European call on the price index $I(T_i)$ and strike $\kappa = (1+K)^{T_i - T_{i-1}}$.*

Proof. we write the payoff of the $YoYC(T_{i-1}, T_i, K)$ as:

$$YoYC(T_{i-1}, T_i, K) = \max[0, (I(T_i)/I(T_{i-1}) - 1 - ((1+K)^{T_i - T_{i-1}} - 1)] = \max[0, (I(T_i)/I(T_{i-1}) - (1+K)^{T_i - T_{i-1}}].$$

One notes that, in the year-on-year option case, the equivalent strike level is still expressed in percentage terms, which is a convention for forward-starting options: because the strike level is not known in advance, the forward starting strike is always expressed as a multiplicative coefficient. Of course this is not happening for zero-coupon trades, where the strike is known since inception.

The equity analogy proposed above lets one choose as skew metric the implied lognormal volatility of the price index process. This is equivalent to postulate that, in the T-forward measure $\mathbb{Q}^T$, the price index process follows the SDE:

$$dI(t) = I(t)[\mu_I^{\mathbb{Q}^T}(t)dt + s_I(t)dW^{\mathbb{Q}^T}(t)]$$

where the drift $\mu_I^{\mathbb{Q}^T}(t)$ and diffusion $s_I(t)$ functions are realisations of some previsible processes. One notes that this approach is also consistent with the continuous-time macroeconomic model presented in chapter 2 and used for derivatives pricing purposes in chapters 3 and 4.

At this point one notes that the inflation model proposed in the previous chapters not only is derived from a sound macroeconomic framework, but is also extremely flexible from a modelling perspective, as it allows one to model path dependent inflation trades.

As an example, the below tables show the market data available on the ICAP Bloomberg page for 7th December 2012, for inflation zero-coupon and year-on-year caps/floors. These data are used in the calibration tests in the rest of this chapter. Floors are taken up to the strike 1%, while caps are considered for strikes above 1%.

1. Nominal OIS curve:

Maturity (years)	1	2	3	5	7	10	12	15	20	30
OIS rate (%)	-0.01	0.01	0.1	0.43	0.81	1.27	1.51	1.76	1.94	2.05

2. Zero-coupon inflation curve:

Maturity (years)	1	2	3	5	7	10	12	15	20	30
ZCIIS (%)	1.52	1.6	1.63	1.7	1.76	1.95	2.02	2.1	2.18	2.34

3. Zero-coupon options prices (prices in in basis points, maturity in years in the columns, inflation strikes in percentage points in the rows)[2]:

Call:1 \ Put:-1	Strike\Mat.	1	2	3	5	7	10	12	15	20	30
1	1	72	159	259	500	748	1224	1502	1872	2439	3808
1	1.5	40	95	163	354	535	898	1110	1378	1762	2732
1	2	20	50	99	243	382	674	833	1049	1358	2158
1	2.5	9	30	60	164	262	467	588	733	904	1336
1	3	4	16	46	114	176	342	422	529	666	1004
1	3.5	2	11	25	76	130	231	299	377	447	608
1	4	1	6	20	58	80	163	214	254	332	411
1	4.5	1	6	13	37	71	114	154	197	228	294
-1	-2	0	1	1	1	1	1	1	1	1	1
-1	-1	1	2	4	5	8	10	10	11	12	14
-1	-0.5	2	5	7	12	21	28	29	28	29	33
-1	0	6	9	16	30	50	65	66	66	66	66
-1	0.5	11	17	31	71	107	133	135	137	131	144
-1	1	19	34	62	141	234	287	262	324	333	350
-1	1.5	40	71	120	255	361	442	471	510	535	508
-1	2	69	143	211	425	597	751	828	938	1018	1144
-1	2.5	107	215	330	595	833	1061	1186	1366	1614	1917
-1	3	152	304	466	823	1142	1503	1697	1998	2492	3324

[2]Some data has been interpolated as it was missing or obviously inconsistent with the rest of the smile: Puts at strike 1% (Maturities: 7, 10, 15, 20, 30 years), Puts at strike 2% (Maturities: 2, 5, 7 , 10, 12, 15 years).

4. Year-on-year options caps/floors (prices in in basis points, maturity in years in the columns, inflation strikes in percentage points in the rows)[3]:

Cap:1 \Floor:-1	Strike\Mat.	1	2	3	5	7	10	12	15	20	30
1	1	72	187	320	630	954	1495	1817	2274	2957	4142
1	1.5	40	120	224	493	759	1186	1425	1774	2280	3265
1	2	19	75	151	373	590	885	1112	1389	1826	2551
1	2.5	9	49	101	283	460	684	835	1077	1367	1955
1	3	4	29	69	182	363	508	644	795	1044	1481
1	3.5	2	19	50	146	245	400	484	616	799	1120
1	4	1	14	37	138	238	371	386	468	607	855
1	4.5	1	9	28	113	197	304	303	405	485	663
1	5	0	6	22	46	186	144	221	310	361	478
1	6	0	3	14	68	121	140	172	226	371	346
-1	-2	0	19	27	50	139	200	226	246	285	369
-1	-1	2	11	41	119	196	277	258	349	406	519
-1	-0.5	3	16	52	146	237	332	363	422	493	627
-1	0	6	26	68	172	291	404	434	517	633	767
-1	0.5	9	37	90	230	364	500	534	644	757	952
-1	1	19	60	125	277	463	577	687	785	963	1198
-1	1.5	38	97	178	395	596	801	840	1040	1229	1528
-1	2	68	151	257	486	770	1026	1097	1335	1582	1962
-1	2.5	107	223	355	642	913	1205	1374	1665	1987	2547
-1	3	153	294	473	832	1166	1548	1763	1995	2539	2892

[3]Some data has been interpolated as it was missing or obviously inconsistent with the rest of the smile: Floor at strike 2.5% (Maturity: 7 years), Floor at strike 3% (Maturity: 10 years), Caps at strike 4% (Maturities: 5, 7, 12, 15 years), Caps at strike 4.5% (Maturities: 12, 20 years), Caps at strike 5% (Maturities: 5, 7, 20 years), Caps at strike 6% (Maturities: 20, 30 years). The higher proportion of prices that had to be cleaned compared to the zero-coupon option case (in the previous page) shows that the year-on-year options market can be less liquid than the zero-coupon options market.

Using the result shown in the above lemma, one calculates the zero-coupon strikes on the equivalent Price Index options:

Call:1 / Put:-1	Strike \Mat.	1	2	3	5	7	10	12	15	20	30
1	1	101	102	103	105.1	107.2	110.5	112.7	116.1	122	134.8
1	1.5	101.5	103	104.6	107.7	111	116.1	119.6	125	134.7	156.3
1	2	102	104	106.1	110.4	114.9	121.9	126.8	134.6	148.6	181.1
1	2.5	102.5	105.1	107.7	113.1	118.9	128	134.5	144.8	163.9	209.8
1	3	103	106.1	109.3	115.9	123	134.4	142.6	155.8	180.6	242.7
1	3.5	103.5	107.1	110.9	118.8	127.2	141.1	151.1	167.5	199	280.7
1	4	104	108.2	112.5	121.7	131.6	148	160.1	180.1	219.1	324.3
1	4.5	104.5	109.2	114.1	124.6	136.1	155.3	169.6	193.5	241.2	374.5
-1	-2	98	96	94.1	90.4	86.8	81.7	78.5	73.9	66.8	54.5
-1	-1	99	98	97	95.1	93.2	90.4	88.6	86	81.8	74
-1	-0.5	99.5	99	98.5	97.5	96.6	95.1	94.2	92.8	90.5	86
-1	0	100	100	100	100	100	100	100	100	100	100
-1	0.5	100.5	101	101.5	102.5	103.6	105.1	106.2	107.8	110.5	116.1
-1	1	101	102	103	105.1	107.2	110.5	112.7	116.1	122	134.8
-1	1.5	101.5	103	104.6	107.7	111	116.1	119.6	125	134.7	156.3
-1	2	102	104	106.1	110.4	114.9	121.9	126.8	134.6	148.6	181.1
-1	2.5	102.5	105.1	107.7	113.1	118.9	128	134.5	144.8	163.9	209.8
-1	3	103	106.1	109.3	115.9	123	134.4	142.6	155.8	180.6	242.7

One notes that the lognormal volatility levels implied by the prices of index options still exhibit skew and are generally much lower than the implied volatilities found in the equity or FX volatility surfaces, generally ranging from 10% to 50%. Finally, as the market is not perfectly liquid, the volatility surface is not perfectly smooth, especially for high or low strikes in the longest maturities.

Call:1 / Put:-1	Strike / Mat.	1	2	3	5	7	10	12	15	20	30
1	1	1	1.4	2	3	3.8	4.9	5.4	5.8	6.1	7.1
1	1.5	1	1.4	2	3.2	3.9	4.9	5.3	5.6	5.6	5.7
1	2	1	1.4	2	3.3	4.1	5.2	5.6	6	6.4	7.1
1	2.5	1	1.6	2.2	3.4	4.2	5.2	5.7	6.1	6.3	6.6
1	3	1.1	1.7	2.5	3.7	4.4	5.6	6	6.4	6.9	7.5
1	3.5	1.1	1.9	2.6	3.8	4.7	5.7	6.2	6.8	7.1	7.5
1	4	1.2	2	2.8	4.2	4.8	5.9	6.5	6.9	7.7	8
1	4.5	1.3	2.3	3	4.2	5.3	6.2	6.8	7.4	8	8.6
-1	-2	1.5	2.2	2.6	2.8	3.7	4.6	5	5.6	6.3	7.4
-1	-1	1.2	1.8	2.3	3	3.6	4.5	5	5.6	6.5	7.9
-1	-0.5	1.2	1.7	2.2	2.9	3.7	4.6	5.1	5.6	6.3	8.1
-1	0	1.2	1.6	2.1	2.9	3.7	4.7	5.1	5.6	6.3	7.4
-1	0.5	1.1	1.5	2	3	3.8	4.7	5.1	5.6	6.1	7
-1	1	1	1.4	1.9	3.1	4.1	5.1	5.1	6	6.6	7.4
-1	1.5	1	1.4	1.9	3.2	4	4.9	5.2	5.7	6.1	6.6
-1	2	1	1.7	2	3.5	4.4	5.4	5.8	6.3	6.5	7.2
-1	2.5	1	1.7	2.2	3.5	4.4	5.2	5.6	6.1	6.5	6.8
-1	3	1.1	1.8	2.3	3.8	4.6	5.6	5.9	6.4	7	7.6

Using the result shown in the above lemmas, once calculates the implied volatility from the equivalent Price Index European and forward-starting options:

Call:1 / Put:-1	Strike / Mat.	1	2	3	5	7	10	12	15	20	30
1	1	1	1.9	2.4	2.9	3.2	3.8	3.7	4.4	5.4	7.8
1	1.5	1	1.7	2.3	3.1	3.1	3.3	3	3.6	4.2	6.8
1	2	1	1.7	2.3	3.1	3.1	2.6	3.5	3.3	4.1	5.2
1	2.5	1	1.8	2.2	3.1	3.4	2.5	2.8	3.4	3.2	4.7
1	3	1.1	1.8	2.3	2.6	3.7	2.6	3.1	2.7	3.3	4
1	3.5	1.1	1.8	2.3	2.8	2.9	2.7	2.6	2.9	3.1	3.6
1	4	1.2	2	2.4	3.1	3.2	2.9	2.9	3.2	3	3.4
1	4.5	1.3	2	2.6	3.4	3.4	3	3.1	3.3	3.3	3.2
1	5	1.4	2.1	2.7	3.4	3.5	2.6	3.6	3.5	3.5	3.5
1	6	1.5	2.3	2.9	3.7	3.8	2.4	3.2	3.5	3.7	4.1
-1	-2	1.5	2.3	2.5	2.8	4	3.9	3.6	3.1	3.4	4.2
-1	-1	1.4	2	3	3.4	3.6	3.6	3	3.1	3.2	3.9
-1	-0.5	1.2	1.9	2.8	3.3	3.5	3.5	2.8	3.2	3.1	3.8
-1	0	1.2	1.9	2.6	3.1	3.5	3.4	2.5	3.3	3.4	3.5
-1	0.5	1	1.8	2.5	3.2	3.3	3.3	2.2	3.4	3	3.7
-1	1	1	1.7	2.3	2.8	3.6	3.2	3.4	3.6	3.4	3.7
-1	1.5	1	1.7	2.2	3.1	3.2	3.1	3.4	3.8	3	3.8
-1	2	1	1.6	2.2	2.7	3.8	3.1	3.4	3.8	3.1	3.9
-1	2.5	1	1.6	2.1	2.7	3.5	2.9	2.8	3.9	3.3	4.9
-1	3	1.1	1.2	2.5	2.9	3	3.2	3	4	5	5.7

The above data shows that the forward skew, which is the volatility skew observed in the forward-starting options, is very persistent across long maturities, as it happens in the equity markets. Given the relevance of these aspects, in the following section we explore the modelling approaches that have been used to model the forward skew.

5.4 Issues with skew and forward skew: pricing year-on-year trades

The equity analogy proposed in the previous section is a powerful tool to model skews and forward skews in a macroeconomic model like the one proposed in chapters 3 and 4. Chapter 4 showed a calibration example where only at-the-money options were considered: calibrating to the skew was not doable because a simple diffusive process driven by Gaussian shocks can hardly reproduce the steep skews that observed in the market.

Gatheral [58] shows how the local volatility model, although can perfectly fit the terminal distributions, tends to exhibit very flat forward skews in the longer maturities, which is not consistent with cliquet prices observed in the market. Cliquets are a forward-starting type of trade that is sensitive to forward skew. The main reason for this behaviour is that the total volatility is, loosely speaking, an integral over all possible paths of the local volatility function: because the integral is a smoothing operator, the incremental volatility at a long maturity tends to be small, as it is "summed" with all the volatility from inception. This considerably reduces the likelihood of extreme market moves in long maturities. Gyöngy [62] provides the theoretical underpinning of the qualitative argument exposed above, showing that the local volatility of a diffusion processes can be thought as the expectations of a stochastic volatility, under some conditions. Further disadvantages of local volatility models include numerical parameter instability (as the local volatility is a function of some derivatives, that are not a smoothing operator), numerical instability when building trees and PDE numerical resolution grids, and the slowness due to having to resort to Monte Carlo simulations.

Stochastic volatility models, like the Heston model [67], better capture the forward skew, because the volatility can randomly increase during the whole life of the option: therefore at long maturities it is still possible to have sudden extreme moves, and this increases the value of long-dated low- and high-strikes forward-starting options. These arguments are explained in more detail both in Gatheral [58] and Rebonato [103]. Another distinctive advantage of these models is that European option prices can be recovered by using characteristic function methods, which speeds up the calibration process, as shown for example in Gatheral [58]. However, we feel it would be cumbersome to incorporate a stochastic volatility term in the macroeconomic model proposed in chapter 3: therefore this approach is not viable in our framework.

For completeness, it must be remembered that in recent years hybrid local-stochastic volatility models have been presented and used in the industry: the diffusion coefficient is the product of a local volatility function and a stochastic volatility term. They are based on the result in Gyöngy [62] mentioned above and essentially calibrate both a local volatility and a stochastic volatility model, where the vol-of-vol of

the latter is properly rescaled in order to minimise the calibration error. More details are provided in Tian, Zhu, Klebaner & Hamza [113].

Alternatively, one can add jumps (as it happens in the Merton model) which increases the skew, but only in the very short maturities, as Gatheral [58] shows. The intuition is that jumps can make the probability of sudden moves higher, especially if they happen in first times of the simulation: as time goes by, the effect of jumps is diluted and the skew is driven more by the compensator.

It should be noted that the Merton model with lognormal jumps can be written as an uncertain-parameter diffusion model, where the uncertainty is driven by the number of jumps that have occurred[4]. This feature is more in line with the macroeconomic framework developed in chapter 3, and is our choice to develop this work. To overcome the flattening of the forward skew, one can use time-varying parameters to maintain the forward skew to some desired levels.

Some simulations reported below show that this is actually the case: here we compare the 1 year forward skew, measured as the difference in lognormal implied volatility at strikes 80% of at-the-money and 120% of at-the-money, when one runs a constant-parameters jump-diffusion against the case when one runs a time-varying parameters jump-diffusion:

Maturity	1	2	3	5	7	9	10
constant-par	-1.87	-1.03	-0.83	-0.56	-0.46	-0.36	-0.17
time-varying par.	-1.87	-2.97	-3.95	-4.39	-2.28	-0.49	-0.26

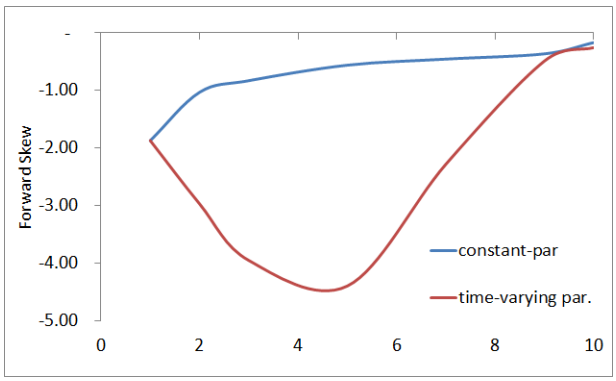


Figure 5.5: Forward skew (implied vol for 120% strike versus the 80% strike for different maturities.

The experiment is run with a jump diffusion with parameters $\sigma = 2.5\%$ (Brownian volatility), $h = 0.25$

[4]The proof of this result is reviewed in the next section.

(jump intensity), $\mu = -2.5\%$ (log-jump average), and $\delta = 5\%$ (log-jump standard deviation). The time varying version is created by multiplying the log-jump average size μ and the log-jump standard deviation δ by the maturity. The aim of this experiment is not to target any specific shape, but to show that a constant-parameter model delivers a shallow forward-skew for long maturities, which can be somewhat overcome by using time-varying parameters.

With these ideas in mind, in the following sections we quickly review the Merton model and propose an extension to the case when its parameters, namely the hazard rate $h(t)$, the log-jump mean $\mu(t)$, the log-jump standard deviation $\delta(t)$, and the Brownian volatility term $\sigma(t)$, are step functions of time.

Finally, a word of caution should be issued regarding market liquidity. We are building a model that can be calibrated both to zero-coupon and year-on-year inflation trades, thanks to the equity analogy. The two markets may be separated as the user of one type of trade may not be the same of the other one. This would mean that selling a year-on-year trade and hedging its volatility exposure with a zero-coupon trade may expose the trader to the basis between these two markets. Therefore this approach can be used to spot illiquidity areas in the market and to create relative-value opportunities: it should not be forgotten that this strategy can expose the trade to profit and loss volatility due to the illiquidity of these markets. The equity analogy helps us again: forward-starting options are normally embedded in cliquet-style structured products that are sold to retail clients with a margin, and are normally hedged by the issuers using European options, that are more liquid. European options can be easily bought or sold (a "two-way market"), while cliquets are sold by banks and held by retail investors (a "one-way market").

5.5 Merton jump-diffusion model (JD)

5.5.1 From diffusions to jump-diffusions

The previous section contained a brief overview of some modelling strategies that have been used in literature and in the industry to address the option market skew and keep the model tractable at the same time: from this point we focus our attention to an extension of the Merton's jump-diffusion model that has yielded promising results when calibrating to inflation options skews, as shown at the end of this chapter.

Three features make this model a good candidate to attack the inflation skew problem: firstly, the model was designed originally for a positive underlying, and the lemmas proved in the previous sections of this chapter have shown how to convert inflation options, both zero-coupon and year-on-year, into options on the price index, which is by construction a positive quantity. Secondly, the Merton model

shows closed forms for both vanilla options and terminal distributions, based respectively on series of Black-Scholes option prices and Gaussian distributions with weights that are Poisson probabilities: this makes the model tractable analytically and speeds up any calibration routine. Finally, because the jump-diffusion is a type of Lévy process, well-known characteristic function techniques and related numerical methods can also be employed to provide an alternative way to calculate options prices and underlying distributions alike.

In the next subsection we briefly review the original Merton model (originally presented in 1976 by Merton [97]) based on Matsuda [93] and highlight the main results used to develop our theory in the rest of the chapter.

5.5.2 The original model

To model the dynamics of the price index $I(t)$, let us define the scalar Brownian motion process $\{W(t)\}_{t\geq 0}$ and the Poisson process $\{N(t)\}_{t\geq 0}$ with constant intensity $h \geq 0$ and deterministic jump size equal to 1: this process is multiplied by a stochastic jump size J. The log of the stochastic jump size is normally distributed with mean μ and variance δ^2, i.e. $\log J \sim \mathcal{N}(\mu, \delta^2)$: the Brownian motion $\{W(t)\}_{t\geq 0}$, the Poisson process $\{N(t)\}_{t\geq 0}$, and its jump size J are independent. Further, the real numbers m and s are used in the below SDE to model the drift and the diffusion respectively:

$$dI(t) = I(t)[(m - hk)dt + sdW(t) + (J - 1)dN(t)]. \tag{5.4}$$

We point out that the drift includes the compensator term hk. Here the real constant k is the average jump size, given by $k = e^{\mu + \frac{\delta^2}{2}} - 1$.

A standard application of Ito's lemma yields the differential for $\log I(t)$:

$$d \log I(t) = \left(m - hk - \frac{s^2}{2}\right) dt + sdW(t) + \log J dN(t). \tag{5.5}$$

This equation is then integrated yielding:

$$\log I(t) = \log I(t_0) + \int_{t_0}^{t} \left(m - hk - \frac{s^2}{2}\right) ds + \int_{t_0}^{t} sdW(s) + \sum_{j=N(t_0)}^{N(t)} \log\ J_j \tag{5.6}$$

which yields the expression for $I(t)$:

$$I(t) = I(t_0) e^{\int_{t_0}^{t} \left(m - hk - \frac{s^2}{2}\right) ds + \int_{t_0}^{t} sdW(s) + \sum_{j=N(t_0)}^{N(t)} \log\ J_j}. \tag{5.7}$$

Here we are assuming that $J_0 = 0$.

To get an expression for the distribution of $I(t)$ (and therefore to price options), Merton's strategy is to condition the value of $I(t)$ to the number of jumps that have occurred between times t_0 and t: because jumps are independent and lognormally distributed, the distribution of $I(t)$ conditional to the event $\{N(t) = j, j \in \mathbb{N}\}$ is known, yielding:

$$\mathbb{P}(x_t) = \sum_{j=0}^{+\infty} \frac{e^{-h(t-t_0)}(h(t-t_0))^j}{j!} \frac{1}{(2\pi(s^2(t-t_0)+j\delta^2))^{\frac{1}{2}}} e^{-\frac{1}{2}\frac{(x(t)-[(m-hk-\frac{s^2}{2})(t-t_0)+j\mu])^2}{(s^2(t-t_0)+j\delta^2)}} \tag{5.8}$$

where we defined $x(t) = \log(I(t)/I(t_0))$ to make the notation lighter. Some useful properties of this distribution follow as standard results for Lévy processes. The characteristic function is:

$$\Phi_{x(t)}(\omega) = e^{(t-t_0)[i\omega(m-hk-\frac{s^2}{2})-\frac{(\sigma\omega)^2}{2}+h(e^{i\omega\mu-\frac{(\delta\omega)^2}{2}}-1)]}. \tag{5.9}$$

Here i is the imaginary unit such that $i^2 = -1$. To lighten notation, we write $x(t) = x$ and $\log(J(t)) = y$.

An alternative way to express the same characteristic function is to write it as a series of characteristic functions of normal distributions, using the Poisson probability weights, means, and variances used in ((5.8)):

$$\Phi_{x(t)}(\omega) = \sum_{j=0}^{+\infty} \frac{e^{-h(t-t_0)}(h(t-t_0))^j}{j!} e^{[i\omega((m-hk-\frac{s^2}{2})(t-t_0)+j\mu)-\frac{(s^2(t-t_0)+j\delta^2)(\omega)^2}{2}]}. \tag{5.10}$$

The Fokker-Plank IPDE for the transition probability $p(x) = \mathbb{P}(x(t) = x|x(t_0) = x_0)$ is:

$$\frac{\partial p(x)}{\partial t} = -m\frac{\partial p(x)}{\partial x} + \frac{s^2}{2}\frac{\partial^2 p(x)}{\partial x^2} + h\int_{-\infty}^{+\infty}[p(x-y)-p(x)]\frac{e^{-\frac{1}{2}\left(\frac{y-\mu}{\delta}\right)^2}}{(2\pi\delta^2)^{\frac{1}{2}}}dy. \tag{5.11}$$

European option prices $V(K, t-t_0)$ with strike K and time to maturity $t-t_0$ are obtained similarly as series of Black-Scholes option prices $V_{BS}(K, t-t_0, n_i, \sigma_i)$ weighted by Poisson probabilities and using n_i as risk-neutral drift and σ_i as volatility:

$$V(K, t-t_0) = \sum_{j=0}^{+\infty} \frac{e^{-h'(t-t_0)}(h'(t-t_0))^j}{j!} V_{BS}(K, t-t_0, n_j, \sigma_j) \tag{5.12}$$

where $n_j = n(t) - hk + \frac{j\log(1+k)}{t-t_0}$, $\sigma_j^2 = s^2 + \frac{j\delta^2}{t-t_0}$, and $h' = h(1+k)$. Consistently with the notation in the previous chapters, $n(t)$ is the short nominal rate.

Finite differences techniques are also used to recover European option prices $V(I(t))$ by solving the

following IPDE:

$$\frac{\partial V(I(t))}{\partial t} + \frac{s^2 I(t)^2}{2}\frac{\partial^2 V(I(t))}{\partial I(t)^2} + (m - hk)I(t)\frac{\partial V(I(t))}{\partial I(t)} + h\mathbb{E}[V(JI(t)) - V(I(t))] = n(t)V(I(t)). \quad (5.13)$$

In the above IPDE we have taken the expectation under the risk-neutral measure: the drift m and the compensator hk are the drift and the compensator in the risk-neutral measure respectively.

Statistical properties like moments of $x(t)$ are easily recovered by differentiating the characteristic functions:

$$\mathbb{E}[x(t)] = m - hk - \frac{s^2}{2} + h\mu$$

$$\mathbb{E}[x(t)^2] = s^2 + h\delta^2 + h\mu^2$$

$$\mathbb{E}[x(t)^3] = h(3\delta^2\mu + \mu^3)$$

$$\mathbb{E}[x(t)^4] = h(3\delta^4 + 6\delta^2\mu^2 + \mu^4).$$

The option market volatility skew is determined by the log-jump average μ and amplified by the intensity h, while the convexity of the volatility smile is driven by the log-jump standard deviation δ and the absolute value of μ. As it happens in many models, the interaction between these three parameters is complex and one recovers similar shapes of the volatility smile using different parameter sets.

5.6 Some comments on the numerical implementation

The series of Poisson probabilities appearing both in the price distribution ((5.8)) and in the forms of option prices ((5.12)) are truncated in the numerical applications that follow by finding the smallest integer value m such that $\sum_{j=0}^{m} \frac{e^{-h(t-t_0)}(h(t-t_0))^j}{j!} > 1 - 10^{-a}$, where the positive integer number a is the wanted accuracy level. Clearly, a higher intensity h or a longer time horizon $t - t_0$ require a higher m given a. In most applications in this chapter we chose $a = 4$, which appears to be a good compromise between numerical accuracy and performance.

When implementing finite differences methods for the Fokker-Plank equation, one needs to compute some probabilities that are not available in closed forms. Further, when running a Monte Carlo simulation to price options one may need to verify ex-post that the distribution is the one expected or what the impact of the time grid is or whether the number of simulations is high enough. Another element to determine is the accuracy of the numerical inversion of the characteristic function. Finally, one may want to better understand what the series truncation impact is in the series forms proposed above for option prices or probabilities.

To perform these checks, we implemented three independent routines to obtain the distribution and cross-check the results:

1. Monte Carlo pricing of undiscounted option prices $V(K_i)$ at many closely spaced strikes K_i (equally spaced with interval $\Delta K = 1$), and subsequent calculation of the implied distribution by the relationship $p(K_i, t) = \frac{\partial^2 V}{\partial K_i^2} \cong \frac{V(K_{i+1})-2V(K_i)+V(K_{i-1})}{(\Delta K)^2}$. Here $K_{i+1} = K_i + \Delta K$ and $K_{i-1} = K_i - \Delta K$.

2. Numerical inversion of the characteristic function, expressed either for the jump-diffusion or for a truncated sum of Poisson-weighted characteristic functions of the normal distribution.

3. Calculation of the truncated sum of normal distributions using the above-mentioned Poisson weights.

The below graph shows the good agreement between the different methods using a parameter set $m = 0$, $s = 0.2$, $h = 1$, $\mu = 0$, and $\delta = 0.05$ over a 10 years time horizon. The initial level $I(t_0)$ is 100. The Monte Carlo simulation is run with 600,000 paths and 50 time steps. A high number of paths is needed to obtain a smooth simulation given that the jump intensity is relatively low. All numerical methods produce very similar results, and the Monte Carlo simulation error is well under control. When truncating the series, we chose the accuracy $a = 4$.

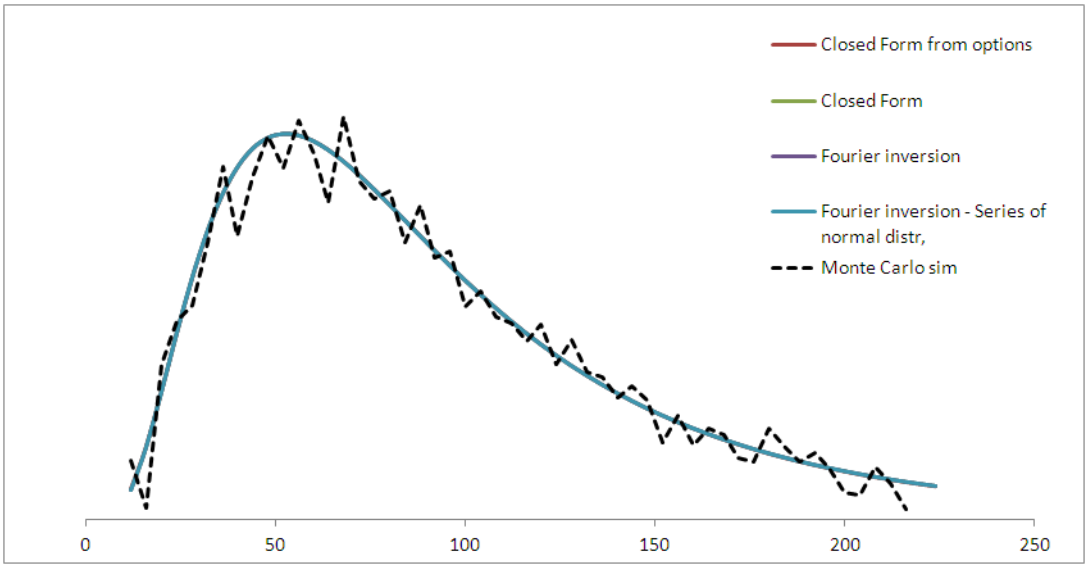


Figure 5.6: Price distribution (Monte Carlo simulation, 600,000 simulations, 50 time steps over 10 years) compared against the same distribution obtained with other numerical methods (Fourier Inversion, second derivative of option prices w.r.t. strike, and closed forms). All methods but the Monte Carlo yield very similar results. The Monte Carlo error appears overall under control.

5.7 Time-varying jump-diffusion models (TV-JD)

To improve calibration, we introduce a time-inhomogeneous version of the model, where $m(t)$, $s(t)$, $h(t) \geq 0$, $\mu(t)$, and $\delta(t)$ are deterministic scalar real bounded functions of time and have the same meaning as the constant variables used in the original Merton model. These functions are step functions: this request improves the analytical tractability while not adding any meaningful practical constraint, given that on the market only a discrete set of maturities are quoted. The SDE to consider is:

$$dI(t) = I(t)[(m(t) - h(t)k(t))dt + s(t)dW(t) + (J-1)dN(t)]. \tag{5.14}$$

Here the function $k(t)$ is the average jump size at time t, given by $k(t) = e^{\mu(t)+\frac{\delta(t)^2}{2}} - 1$, given that the logarithm of the jump size at time t is normally distributed with mean $m(t)$ and variance $\delta(t)^2$. The main disadvantage of this approach is that the series-based distribution (5.8) and option prices formulas (5.12) do not hold any more.

To overcome this problem, we propose a routine to convert for a given maturity T the bounded functions $m(t)$, $s(t)$, $h(t) \geq 0$, $\mu(t)$, and $\delta(t)$ into constant levels $m^*(T)$, $s^*(T)$, $h^*(T) \geq 0$, $\mu^*(T)$, and $\delta^*(T)$ such that the terminal distribution of I_T implied by (5.14) is the same as the one implied by the classic time-homogeneous Merton jump-diffusion below:

$$dI(t) = I(t)[(m^*(T) - h^*(T)k^*(T))dt + s^*(T)dW(t) + (J-1)dN(t)] \tag{5.15}$$

where $k^*(T)$ is the average jump size, given by $e^{\mu^*(T)+\frac{\delta^*(T)^2}{2}} - 1$, given that the logarithm of the jump size at all times $t \in [t_0, T]$ is normally distributed with mean $m^*(T)$ and variance $\delta^*(T)^2$. Here $h^*(T)$ is the intensity of the Poisson process $N(t)$. Clearly this method should only be used with payoffs that are not path-dependent as we are calibrating only to the terminal distribution.

Therefore we calibrate a time-inhomogeneous version of the jump-diffusion model and convert it into its equivalent time-homogeneous version for a given maturity. This approach retains the best of the two models: the improved calibration flexibility of the time-inhomogeneous model coupled with the analytical tractability of the time-homogeneous original model.

Before proposing our numerical algorithm, we notice that moving from the time-inhomogeneous model to the time-homogeneous is straightforward in some special cases:

1. If the intensity h is constant while the log-jump mean and variance are functions of time ($\mu(t)$ and $\delta(t)^2$ respectively), one writes the jump distribution as a sum of normal distributions multiplied by the indicator of each of the M time intervals $[t_{i-1}, t_i)$ when each normal distribution is relevant (we

remind the reader that the time-varying functions are step functions): $f(\log J(t)) = p(\log J(t), t \in [t_0, t_M)) = \sum_{i=1}^{M} f_i(\log J(t), \mu_i, \delta_i^2) I_{\{t \in [t_{i-1}, t_i)\}}$. Here $f_i(\log J(t), \mu_i, \delta_i^2) = \frac{e^{-\frac{1}{2}\left(\frac{(\log J(t)-\mu_i}{\delta_i}\right)^2}}{(2\pi\delta_i^2)^{\frac{1}{2}}}$. Therefore one defines $\mu^*(t_M) = \mathbb{E}[\log J(t) | t \in [t_0, t_M)]$.

One writes $\mathbb{E}[\log J(t) | t \in [t_0, t_M)] = \int_{-\infty}^{+\infty} \log J(t) p(\log J(t) | t \in [t_0, t_M)) d \log J(t)$.

This distribution is written as: $p(\log J(t) | t \in [t_0, t_M)) = p(\log J(t), t \in [t_0, t_M))/p(t \in [t_0, t_M))$: further one finds the expression $p(\log J(t), t \in [t_0, t_M))/p(t \in [t_0, t_M)) = \sum_{i=1}^{M} f_i(\log J(t), \mu_i, \delta_i^2)[e^{-ht_{i-1}} - e^{-ht_i}]/[e^{-ht_0} - e^{-ht_M}]$.

Therefore the expectation is calculated as $\mathbb{E}[\log J(t) | t \in [t_0, t_M)] = \sum_{i=1}^{M} \mu_i [e^{-ht_{i-1}} - e^{-ht_i}]/[e^{-ht_0} - e^{-ht_M}]$. A similar reasoning can be done for the second moment, to get the variance.

2. If the intensity $h(t)$ is a function of time while the log-jump mean and variance are constant (μ and δ^2 respectively), one introduces the quantities $H(t) = \int_{t_0}^{t} h(s) ds$ and $S(t) = e^{-H(t)}$ respectively. Because the jump size is independent from the jump time, the survival probability is the same whether one uses the time-varying hazard rate $h(t)$ or $h^*(T) = H(T)/T$. Once this quantity has been calculated we move on as in the time-inhomoegenous case.

We consider the most general case, when all functions $m(t)$, $s(t)$, $h(t)$, $\mu(t)$, and $\delta(t)$ are step functions of time. The routine we propose is run for all available maturities and requires the following steps:

1. For the first maturity there is no need to run a calibration, as the time-homogeneous model parameter set is identical to the one of the time-inhomogeneous model, because we are working with step functions.

2. Starting from the second maturity, the Brownian volatility term is calibrated first, by setting $s^*(T)^2 = \frac{1}{T-t_0} \int_{t_0}^{T} s(u)^2 du$. This step is done first because variances can be integrated and the Brownian motion $W(t)$ is independent from the Poisson counting process $N(t)$.

3. The calculation of the jump process parameters $h^*(T)$, $\mu^*(T)$, and $\delta^*(T)$ is the most complex step in the procedure.

   One defines the discontinuous part $D(t) = \sum_{j=N(t_0)}^{N(t)} \log J_j = \int_{t_0}^{t} \log J(s) dN(s)$. Let us introduce a partition of the time interval $[t_0, T]$ in M subintervals $[t_0, t_1], (t_1, t_2], ..., (t_{M-1}, T]$, where the extremes of each interval are the times where the step function jumps (and therefore these are the maturities of the options we are calibrating the model to). We write $t_i = t_{i-1} + \Delta t_i$. We focus on the number of jumps that have happened in each time interval, defined $N(\Delta t_i) = N(t_i) - N(t_{i-1})$: because the Poisson process is a Lévy process, its increments are independent and have the same distribution, which is Poisson with intensity $h_i = h(t_i)$ over the time interval Δt_i. Therefore one writes $\mathbb{P}[N(\Delta t_i) = n] = e^{-h_i \Delta t_i} (h_i \Delta t_i)^n / n!$.

This result is useful to write the probability distribution of $D(t)$ using conditioning to the number of jumps that happen in each time interval:

$$p(d) = \mathbb{P}[D(t) = d] = \mathbb{P}\left[\sum_{j=N(t_0)}^{N(t)} \log J_j = d\right] =$$

$$\sum_{(n_1,...,n_M)\in\mathbb{N}^M} \mathbb{P}\{\sum_{j=N(t_0)}^{N(t)} \log J_j = d|[N(\Delta t_1) = n_1, ..., N(\Delta t_M) = n_M]\mathbb{P}[N(\Delta t_1) = n_1, ..., N(\Delta t_M) = n_M]\}. \tag{5.16}$$

An alternative notation, less concise but maybe more clear, is:

$$\mathbb{P}\left[\sum_{j=N(t_0)}^{N(t)} \log J_j = d\right] = \sum_{n_1=0}^{+\infty} ... \sum_{n_M=0}^{+\infty} \mathbb{P}\{\sum_{j=N(t_0)}^{N(t)} \log J_j = d|[N(\Delta t_1) = n_1, ..., N(\Delta t_M) = n_M]$$
$$\mathbb{P}[N(\Delta t_1) = n_1, ..., N(\Delta t_M) = n_M]\}. \tag{5.17}$$

Because the increments are independent and the jump sizes are lognormal, we write:

$$\mathbb{P}\left[\sum_{j=N(t_0)}^{N(t)} \log J_j = d\right] = \sum_{n_1=0}^{+\infty} ... \sum_{n_M=0}^{+\infty} \mathbb{P}\{\sum_{j=N(t_0)}^{N(t)} \log J_j = d|[N(\Delta t_1) = n_1, ..., N(\Delta t_M) = n_M]\prod_{i=1}^{M} \mathbb{P}[N(\Delta t_i) = n_i]\}$$

$$= \sum_{n_1=0}^{+\infty} ... \sum_{n_M=0}^{+\infty} \mathbb{P}\{\sum_{j=N(t_0)}^{N(t)} \log J_j = d|[N(\Delta t_1) = n_1, ..., N(\Delta t_M) = n_M]\prod_{i=1}^{M} e^{-h_i\Delta t_i}(h_i\Delta t_i)^{n_i}/n_i!\}$$

$$= \sum_{n_1=0}^{+\infty} ... \sum_{n_M=0}^{+\infty} \left\{\frac{1}{(2\pi\sum_{j=0}^{M} n_j\delta_j^2)^{\frac{1}{2}}} e^{-\frac{1}{2}\frac{(d-\sum_{j=0}^{M} n_j\mu_j)^2}{\sum_{j=0}^{M} n_j\delta_j^2}} \prod_{i=1}^{M} e^{-h_i\Delta t_i}(h_i\Delta t_i)^{n_i}/n_i!\}\right\}. \tag{5.18}$$

We managed to write the probability distribution of the discontinuous part of the time-inhomogeneous jump-diffusion in closed form. The distribution of the sum of the logarithm of the jumps is clearly still normal given that each jump is independent and its logarithm is normally distributed by construction.

Because we are doing a bootstrapping, we set $M = 2$, where the first period is the time up to which the model has already been calibrated (i.e. $[t_0, t_{i-1}]$), and the second period is the interval between the previous time and the maturity we are calibrating the model to (i.e. $(t_{i-1}, t_i]$). This

radically simplifies the above formula:

$$p(d) = \sum_{n_{i-1}=0}^{+\infty} \sum_{n_i=0}^{+\infty} \left\{ \frac{e^{-\frac{1}{2}\frac{(d-n_{i-1}\mu^*_{i-1}-n_i\mu_i)^2}{(n_{i-1}\delta^{*\,2}_{i-1}+n_i\delta_i^2))}}}{(2\pi(n_{i-1}\delta^{*\,2}_{i-1}+n_i\delta_i^2))^{\frac{1}{2}}} \frac{e^{-h_i\Delta t_i}(h_i\Delta t_i)^{n_i}}{n_i!} \frac{e^{-h^*_{i-1}(t_{i-1}-t_0)}(h^*_{i-1}(t_{i-1}-t_0))^{n_{i-1}}}{n_{i-1}!} \right\}. \tag{5.19}$$

We should note that for the first interval we use the constant parameters that have been calibrated at the previous bootstrapping step (μ^*_{i-1}, δ^*_{i-1}, and h^*_{i-1}).

In order to truncate the infinite sums above, one applies the technique presented in the previous section: firstly one builds a guess for the intensity $\hat{h}_i = [h^*_{i-1}(t_{i-1}-t_0) + h_i\Delta\ t_i]/(t_i - t_0)$, and then one chooses an accuracy integer level $a > 1$ and finds the smallest integer N^* such that $\sum_{j=0}^{N^*} \frac{e^{-\hat{h}_i(t-t_0)}(\hat{h}_i(t-t_0))^j}{j!} > 1 - 10^{-a}$.

The truncated probability takes the final form:

$$p(d) \cong \sum_{n_{i-1}=0}^{N^*} \sum_{n_i=0}^{N^*} \left\{ \frac{e^{-\frac{1}{2}\frac{(d-n_{i-1}\mu^*_{i-1}-n_i\mu_i)^2}{(n_{i-1}\delta^{*\,2}_{i-1}+n_i\delta_i^2))}}}{(2\pi(n_{i-1}\delta^{*\,2}_{i-1}+n_i\delta_i^2))^{\frac{1}{2}}} \frac{e^{-h_i\Delta t_i}(h_i\Delta t_i)^{n_i}}{n_i!} \frac{e^{-h^*_{i-1}(t_{i-1}-t_0)}(h^*_{i-1}(t_{i-1}-t_0))^{n_{i-1}}}{n_{i-1}!} \right\}. \tag{5.20}$$

Taken a set of candidate constant levels μ^*_i, δ^*_i, and h^*_i, for each integer $n \in \{0, 1, ..., N^*\}$, which represents the total number of jumps in the calibration interval $[t_0, t_i]$, one writes the error term E_n, defined as $E_0(d) = p_0^C(d) - p_0^V(d)$. The term $p_0^C(d)$ refers to expression (5.20) where one uses the candidate values μ^*_i, δ^*_i, and h^*_i for both μ^*_{i-1}, δ^*_{i-1}, and h^*_{i-1} and μ_i, δ_i, and h_i (in practice this is the probability obtained using a constant set of parameters); the term $p_0^V(d)$ refers to expression (5.20) (in practice this is the true probability, calculated using a time-inhomogeneous parameter set).

We also propose an alternative notation for the same probabilities, which is less concise but possibly more clear: $P_n^C(d, n)$ represents the probability density of the discontinuous part valued in d, given that exactly n jumps have occurred assuming that the model parameters are constant (i.e. from the time-homogeneous model we are calibrating). Instead, the probability density of the discontinuous part valued in d, given that that exactly k jumps have occurred in the past (i.e. in the interval $[t_0, t_{i-1})$ and exactly $n-k$ jumps have occurred in the interval $[t_{i-1}, t_i)$, is denoted by $P_n^V(d, [k, n-k])$. Let us also denote here the Gaussian distribution with mean μ and variance δ^2 by $g(\mu, \delta^2)$.

For example, for $n = 0$ we write explicitly: $E_0(d) = p_0^C(d) - p_0^V(d) = P_0^C(d, 0)\mathcal{D}(0) - P_0^V(d, [0])\mathcal{D}(0)$, where $\mathcal{D}(0)$ is the Dirac delta function in 0.

For $n = 1$, we write explicitly $E_1(d) = p_1^C(d) - p_1^V(d) = P_1^C(d, 1)g(\mu^*_i, \delta^{*2}_i) - \{P_1^V(d, [0, 1])g(\mu_i, \delta_i^2)$

$+ P_1^V(d,[1,0])g(\mu_{i-1},\delta_{i-1}^2)\}$.

For $n=2$, we write $E_2(d) = p_2^C(d) - p_2^V(d) = P_2^C(d,2)g(2\mu_i^*, 2\delta_i^{*2}) - \{P_2^V(d,[0,2])g(2\mu_i, 2\delta_i^2) + P_2^V(d,[2,0])g(2\mu_{i-1}, 2\delta_{i-1}^2) + P_2^V(d,[1,1])g(\mu_i^* + \mu_{i-1}^*, \delta_i^2 + \delta_{i-1}^2)\}$.

This explicit calculation is carried out up to N^*, and each term has $n+2$ terms to be added to get $E_n(d)$.

The cost function to minimise is a probability-weighed sum of square error terms:

$$\sum_{n=0}^{N^*}(E_n(d))^2\mathbb{P}[N(t)=n] = \sum_{n=0}^{N^*}(E_n(d))^2\frac{e^{-\hat{h}_i(t_i-t_0)}(\hat{h}_i(t_i-t_0))^n}{n!}.$$

To provide the final explicit form to minimise numerically, the cost function, one remembers that there are also normal distributions to discretise: therefore one calculates the normal distributions only at the $L+1$ points $d_j \in \{d_0 = d_{min}, d_1 = d_{min} + \Delta d, d2 = d_{min} + 2\Delta d, ..., d_{L-1} = d_{min} + (L-1)\Delta d, d_L = d_{min} + L\Delta d = d_{max}\}$. One also adds Gaussian weights to the sum. The final form for the cost function to minimise is:

$$\sum_{j=0}^{L}\frac{e^{-\frac{(d_j)^2}{2}}}{(2\pi)^{\frac{1}{2}}}\sum_{n=0}^{N^*}(E_n(d_j))^2\frac{e^{-\hat{h}_i(t_i-t_0)}(\hat{h}_i(t_i-t_0))^n}{n!}. \tag{5.21}$$

This cost function is the expected square error of replacing the original time-varying (i.e. time-inhomogeneous) model with a new constant-parameters (time-homogeneous) model. The above expression is proposed in approximated closed form and can therefore be minimised by choosing the jump intensity and distribution parameters that fit a given target. We stress the importance of a closed-form expression that lets one reduce the computation times significantly.

4. Finally, known the parameters $h^*(T)$, $\mu^*(T)$, and $\delta^*(T)$, where $T = t_i$, one calculates the compensator $h^*(T)k^*(T)$, where $k^*(T) = e^{\mu^*(T)+\frac{\delta^*(T)^2}{2}} - 1$ and therefore calculates the deterministic part of the process. One imposes that $m^*(T) - h^*(T)k^*(T) = \int_{t_0}^{T}[m(s) - h(s)k(s)]ds$.

To conclude, we show a numerical example where we convert a time-inhomogeneous jump diffusion into its time-homogeneous equivalent, for all maturities. The implementation is done with accuracy level $a = 4$.

The time-inhomogeneous parameter set is:

time	drift	Brownian vol	Jump intensity	Log jump exp	Log jump st dev
1	0	0.1	1	0.01	0.06
2	0	0.11	2	0.02	0.05
3	0	0.12	3	0.03	0.04
4	0	0.13	4	0.04	0.03
5	0	0.14	5	0.05	0.02
6	0	0.15	6	0.06	0.01

The equivalent time-homogeneous parameter set obtained with our algorithm is:

time	drift	Brownian vol	Jump intensity	Log jump exp	Log jump st dev
1	0	0.1	1	0.01	0.06
2	0	0.10512	1.50949	0.01658	0.05385
3	0	0.1103	1.9972	0.02338	0.04745
4	0	0.11554	2.46954	0.03006	0.04406
5	0	0.12083	2.96107	0.03686	0.03911
6	0	0.12616	3.44693	0.04392	0.03626

We show that the distribution of the two processes is extremely close after running a Monte Carlo simulation for maturity 3 years with $60,000$ paths (using a time step of 1 year). The small differences are due both to Monte Carlo error and to the approximations used in the method proposed (truncation of the Poisson infinite sums to N^* terms and quadrature of the normal distribution using $L+1$ intervals). The parameter L is set to 21.

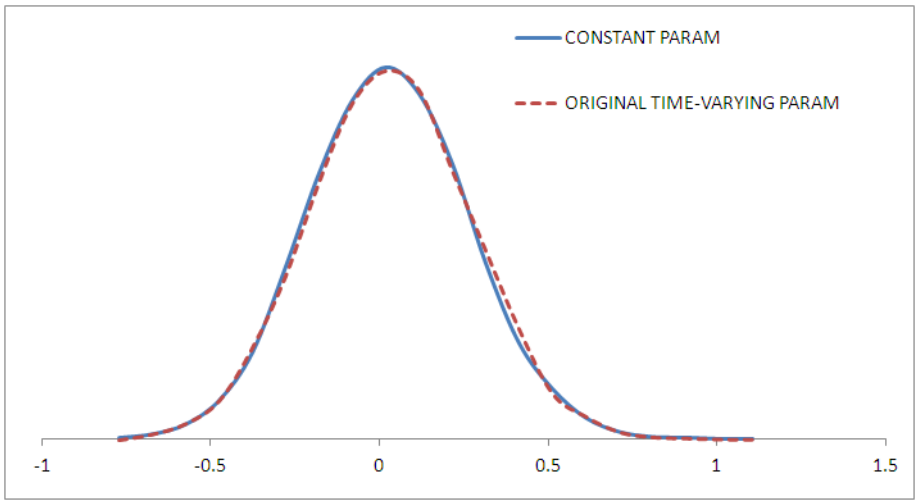


Figure 5.7: Price distribution (Monte Carlo simulation, 60,000 simulations, 3 time steps over 3 years) comparing the original time-inhomogeneous model distribution against the equivalent time-homogeneous model distribution found with the proposed algorithm).

5.8 Uncertain-parameters time-varying jump-diffusion models (UP-TV-JD)

Here we propose an extension of the time-inhomogeneous jump-diffusion model presented in the previous section by using uncertain-parameters techniques: this lets us further improve the calibration capabilities while keeping the model still tractable.

An uncertain-parameters model is a model where some parameters are random variables that become known at inception. We work with a time-inhomogeneous jump-diffusion model where the drift $m_i(t)$, the Brownian diffusion term $s_i(t)$, the jump intensity $h_i(t)$, mean $\mu_i(t)$, and variance $\delta_i^2(t)$ can take a finite number of values depending on the state $\mathcal{P} \in \{1, 2, ..., M\}$, which is a $\mathcal{F}_0$-measurable random variable. Here $\mathcal{F}_t$ is the σ-algebra generated by the state $\mathcal{P}$, the driving Brownian motion $\{W(t)\}_{t\geq 0}$, the Poisson process $\{N(t)\}_{t\geq 0}$, and the jump size process $\{J(t)\}_{t\geq 0}$. The state $\mathcal{P}$ has a known probability distribution with $p(\mathcal{P} = i) = w_i$. We loosely think to a random draw that happens at inception that determines the parameter set driving the time-inhomogeneous jump-diffusion process. The unconditional distribution of the resulting process is a mixture of distributions with weights w_i. More details and full definitions can be found in Brigo, Mercurio & Rapisarda [24] and Castagna [37].

This type of model has better calibration capabilities, at the expenses of increased computational costs (for example, in a Monte Carlo simulation one has to simulate the state $\mathcal{P}$ in addition to the other random variables): the challenge is how to retain the improved flexibility of this approach while compacting the uncertain-parameters model into an equivalent model that has no uncertainty around

the parameter set and for which the dynamics are known.

Brigo [16] starts with a local volatility diffusion scalar SDE in the form: $dX(t) = f(t, X(t))dt + s(t, X(t))dW(t)$. It is assumed that the drift, the diffusion terms, and the initial condition $X(t_0) = x_0$, are such that the above SDE has a unique strong solution whose support is in the interval $(b, +\infty)$: for example, for a geometric Brownian motion with $X(t_0) = x_0 > 0$, we have $b = 0$. Further, one assumes that the unique strong solution of the above SDE has density $p(t, X(t))$ that is absolutely continuous with respect to the Lebesgue measure on the interval $(b, +\infty)$ and satisfies the Fokker-Plank equation.

A mixture condition on the probability distribution $p(t, X(t))$ is imposed, by requesting that $p(t, X(t)) = \sum_i^M w_i p_i(t, X(t))$, where the weights w_i are positive real numbers such that $\sum_{i=1}^M w_i = 1$ and $p_i(t, X(t))$ are some generic probability densities. The objective is, given a density, to specify the SDE: in practice to find the functions $f(t, X(t))$ and $s(t, X(t)))$ such that the density $p(t, X(t))$ can be expressed via the mixture $\sum_i^M w_i p_i(t, X(t))$. Brigo finds a general result by integrating the Fokker-Plank equation twice with respect to the state variable $X(t)$, and then focuses on the specific case where the densities $p_i(t, X(t))$ are the marginal distributions associated to some instrumental diffusions $dX(t) = f_i(t)dt + s_i(t)dW(t)$, with $f_i(t)$ and $s_i(t)$ being deterministic functions of time. The same assumptions made above regarding the existence of a unique solution on the support $(b, +\infty)$, on the absolute continuity of the probability densities $p_i(t, X(t))$, and the related Fokker-Plank equations hold.

The main results are that $f(t, X(t)) = \sum_{i=1}^M f_i(t) w_i^*(t)$ and $s^2(t, X(t)) = \sum_{i=1}^M s_i^2(t) w_i^*(t)$, where $w_i^*(t) = w_i g(m_i(t), v_i(t)) / \sum_{i=1}^M [w_i g(m_i(t), v_i(t))]$, $m_i(t) = \int_{t_0}^t f_i(s)ds$, and $v_i(t) = \int_{t_0}^t s_i^2(s)ds$. In this context $g(m_i(t), v_i(t))$ is a Gaussian distribution with mean $m_i(t)$ and variance $v_i(t)$. The main result of that paper is that an uncertain-parameters model originates a local volatility model via its Markovian projection.

The new contribution in this section is to extend Brigo's result to jump-diffusions, and to suggest some numerical routines to solve the problem in an efficient way. As discussed in the previous section, a time-inhomogeneous jump-diffusion model can generate very steep forward skews even at long maturities, which is an advantage when compared to the local volatility models (and, as shown by Brigo, also to uncertain-parameters simple diffusions). We assume that the initial condition and all coefficients of all SDEs presented from this point are such that the SDEs admit a unique strong solution, whose probability densities satisfy the related Fokker-Plank IPDEs.

We assume that the following jump-diffusion SDE is governing the evolution of $X(t)$:

$$dX(t) = f(t, X(t))dt + s(t, X(t))dW(t) + X(t)(J(t) - 1)dN(t). \tag{5.22}$$

We observe that the model above is the same as the one in (5.14), where one assumes that the intensity

of the Poisson process $\{N(t)\}_{t\geq 0}$ is $h(t, X(t)) \geq 0$, the logarithm of the jump at time t has a normal distribution with mean $\mu(t)$ and variance $\delta^2(t)$[5]: we refer to this distribution as $g(\mu(t), \delta^2(t))$. The Brownian motion $\{W(t)\}_{t\geq 0}$, the Poisson process $\{N(t)\}_{t\geq 0}$, and the jump size $J(t)$ are all independent from each other. As above the distribution of $X(t)$ is referred to as $p(t, X(t))$. Here we assume that the drift $f(t, X(t))$ also contains the compensator $h(t, X(t))k(t) = h(t, X(t))[e^{\mu(t)+\frac{1}{2}\delta^2(t)} - 1]$, which is equivalent to writing $f(t, X(t)) = m(t, X(t)) - h(t)k(t)$, where $m(t, X(t))$ can be thought as the original drift term without the compensator term.

We introduce the M instrumental SDEs

$$dX(t) = X(t)[f_i(t)dt + s_i(t)dW(t) + (J(t) - 1)dN(t)] \tag{5.23}$$

where $i \in \{1, 2, ..., M\}$, the intensity of the Poisson process $\{N(t)\}_{t\geq 0}$ is $h_i(t)$, the logarithm of the jump size at time t has a normal distribution with mean $\mu_i(t)$ and variance $\delta_i^2(t)$: we refer to this distribution as $g_i(\mu_i(t), \delta_i^2(t))$. Under the above SDE the marginal distribution of $X(t)$ is $p_i(t, X(t))$. The same assumptions of independence between the Brownian motion $\{W(t)\}_{t\geq 0}$, the Poisson process $\{N(t)\}_{t\geq 0}$, and the jump size $J(t)$ hold. Here we assume that the drift $f_i(t)$ also contains the compensator $h_i(t)k_i(t) = h_i(t)[e^{\mu_i(t)+\frac{1}{2}\delta_i^2(t)} - 1]$, which is equivalent to writing $f_i(t) = m_i(t) - h_i(t)k_i(t)$, where $m_i(t)$ is the original drift term without the compensator term.

Our task is to find some expressions for the parameters of the SDE (5.22) as a function of the parameters of the M instrumental SDEs such that $p(t, X(t)) = \sum_{i=1}^{M} w_i p_i(t, X(t))$, where the weights w_i are positive real numbers such that $\sum_{i=1}^{M} w_i = 1$ and $p_i(t, X(t))$ are the probability densities implied by the instrumental SDEs. To do this we follow Brigo [16] for the first part, essentially up to the expression for the drift and diffusion terms, and then we make some Ansatz to get to an expression of a numerical error expression that has to be brought to zero in order to find the right parameters for the intensity and the jump distribution.

We start by writing the full expression of the Fokker-Plank IPDE implied by (5.22), where, to make the notation lighter, we write $p(t, X(t)) = p(X(t))$.

We also define the lognormal distributions $\phi_i(J(t), \mu_i(t), \delta_i^2(t))$ and $\phi(J(t), \mu(t), \delta^2(t))$ such that the distributions of the logarithm of the jumps $J(t)$ are respectively $\frac{e^{-\frac{1}{2}\left(\frac{(\log J(t)-\mu_i(t)}{\delta_i(t)}\right)^2}}{(2\pi\delta_i^2(t))^{\frac{1}{2}}}$ and $\frac{e^{-\frac{1}{2}\left(\frac{(\log \ J(t)-\mu(t)}{\delta(t)}\right)^2}}{(2\pi\delta^2(t))^{\frac{1}{2}}}$. With all this notation the Fokker-Plank IPDE reads:

$$\frac{\partial p(X(t))}{\partial t} = -\frac{\partial[f(t, X(t))p(X(t))]}{\partial X(t)} + \frac{1}{2}\frac{\partial^2[(s(t, X(t))^2 p(X(t))]}{\partial X(t)^2} \tag{5.24}$$

[5]The two parameters $\mu(t)$ and $\delta^2(t)$ can be local parameters, i.e. one should write $\mu(t, X(t))$ and $\delta^2(t, X(t))$. This case will be fully discussed later in the chapter. We ignore this to keep the notation light.

$$+h(t,X(t))\int_0^{+\infty}[p(X(t)/J(t))-p(X(t))]\phi(J(t),\mu(t),\delta^2(t))dJ(t).$$

One can think to the integral term above $\int_0^{+\infty}[p(X(t)/J(t))-p(X(t))]\phi(J(t),\mu(t),\delta^2(t))dJ(t)$ as the equivalent of the log-space integral $\int_{-\infty}^{+\infty}[p(x-y)-p(x)]\frac{e^{-\frac{1}{2}\left(\frac{y-\mu}{\delta}\right)^2}}{(2\pi\delta^2)^{\frac{1}{2}}}dy$ that was presented in the Fokker-Plank equation 5.11 on page 149. The former is expressed in space terms (so the jump distribution $\phi(J(t),\mu(t),\delta^2(t))$ is a lognormal distribution), while the latter is expressed in log-space, and therefore one uses the normal distribution for the log-jump. One remembers the assumption $p(t,X(t))=\sum_{i=1}^M w_ip_i(t,X(t))$ and writes the above IPDE as:

$$\frac{\partial[\sum_{i=1}^M w_ip_i(X(t))]}{\partial t}=-\frac{\partial[f(t,X(t))\sum_{i=1}^M w_ip_i(X(t))]}{\partial X(t)}+\frac{1}{2}\frac{\partial^2[s(t,X(t))^2\sum_{i=1}^M w_ip_i(X(t))]}{\partial X(t)^2}+$$
$$+h(t,X(t))\int_0^{+\infty}\sum_{i=1}^M w_i[p_i(X(t)/J(t))-p_i(X(t))]\phi(J(t),\mu(t),\delta^2(t))dJ(t). \tag{5.25}$$

One integrates the above IPDE with respect to $X(t)$:

$$\int_b^u\frac{\partial[\sum_{i=1}^M w_ip_i(X(t))]}{\partial t}dX(t)=\int_b^u\frac{-\partial[f(t,X(t))\sum_{i=1}^M w_ip_i(X(t))]}{\partial X(t)}dX(t)$$
$$+\int_b^u\frac{1}{2}\frac{\partial^2[s(t,X(t))^2\sum_{i=1}^M w_ip_i(X(t))}{\partial X(t)^2}dX(t)$$
$$+\int_b^u h(t,X(t))\int_0^{+\infty}\sum_{i=1}^M w_i[p_i(X(t)/J(t))-p_i(X(t))]\phi(J(t),\mu(t),\delta^2(t))dJ(t)dX(t). \tag{5.26}$$

Making the integration explicit where possible and moving out the sums yields:

$$\int_b^u\frac{\partial[\sum_{i=1}^M w_ip_i(X(t))]}{\partial t}dX(t)=-[\sum_{i=1}^M f(t,U(t))w_ip_i(U(t))]+\frac{1}{2}\frac{\partial[s(t,U(t))^2\sum_{i=1}^M w_ip_i(U(t))]}{\partial U(t)}$$
$$+\int_b^u h(t,X(t))\int_0^{+\infty}\sum_{i=1}^M w_i[p_i(X(t)/J(t))-p_i(X(t))]\phi(J(t),\mu(t),\delta^2(t))dJ(t)dX(t). \tag{5.27}$$

A second integration with respect to $U(t)$ yields:

$$\int_b^y\int_b^u\frac{\partial[\sum_{i=1}^M w_ip_i(X(t))]}{\partial t}dX(t)dU(t)=-\int_b^y f(t,U(t))[\sum_{i=1}^M w_ip_i(U(t))]dU(t)$$
$$+\frac{1}{2}\int_b^y\sum_{i=1}^M\frac{\partial[s(t,U(t))^2w_ip_i(U(t))]}{\partial U(t)}dU(t)$$
$$+\int_b^y\int_b^u h(t,X(t))\int_0^{+\infty}\sum_{i=1}^M w_i[p_i(X(t)/J(t))-p_i(X(t))]\phi(J(t),\mu(t),\delta^2(t))dJ(t)dX(t)dU(t). \tag{5.28}$$

One rearranges the above expression and writes:

$$\int_b^y \int_b^u \sum_{i=1}^M w_i \frac{\partial[p_i(X(t))]}{\partial t} dX(t)dY(t) = -\sum_{i=1}^M \int_b^y f(t,U(t))[w_i p_i(U(t))]dU(t) + \frac{s(t,Y(t))^2}{2} \sum_{i=1}^M [w_i p_i(Y(t))]$$
$$+ \int_b^y \int_b^u h(t,X(t)) \int_0^{+\infty} \sum_{i=1}^M w_i [p_i(X(t)/J(t)) - p_i(X(t))]\phi(J(t),\mu(t),\delta^2(t))dJ(t)dX(t)dU(t). \quad (5.29)$$

We observe that the Fokker-Plank IPDE is also valid for the instrumental processes, yielding the IPDEs:

$$\frac{\partial p_i(X(t))}{\partial t} = -\frac{\partial[f_i(t)X(t)p_i(X(t))]}{\partial X(t)} + \frac{1}{2}\frac{\partial^2[[\boldsymbol{s}_i(t)X(t)]^2 p_i(X(t))]}{\partial X(t)^2} \quad (5.30)$$

$$+h_i(t)\int_0^{+\infty} [p_i(X(t)/J(t)) - p_i(X(t))]\phi_i(J(t),\mu_i(t),\delta_i^2(t))dJ(t).$$

By doing some substitutions of the term $\frac{\partial p_i(X(t))}{\partial t}$ one gets:

$$\int_b^y \int_b^u \sum_{i=1}^M w_i \{-\frac{\partial[f_i(t)X(t)p_i(X(t))]}{\partial X(t)} + \frac{1}{2}\frac{\partial^2[[s_i(t)X(t)]^2 p_i(X(t))]}{\partial X(t)^2}$$
$$+ h_i(t)\int_0^{+\infty} [p_i(X(t)/J(t)) - p_i(X(t))]\phi_i(J(t),\mu_i(t),\delta_i^2(t))dJ(t)\}dX(t)dU(t) =$$
$$-\sum_{i=1}^M \int_b^y f(t,U(t))[w_i p_i(U(t))]dU(t) + \frac{s(t,Y(t))^2}{2}\sum_{i=1}^M [w_i p_i(Y(t))]$$
$$+ \int_b^y \int_b^u h(t,X(t)) \int_0^{+\infty} \sum_{i=1}^M w_i [p_i(X(t)/J(t)) - p_i(X(t))]\phi(J(t),\mu(t),\delta^2(t))dJ(t)dX(t)dU(t). \quad (5.31)$$

By rearranging the above expression in a more compact way one gets:

$$s(t,Y(t))^2 = \frac{2\sum_{i=1}^M w_i}{\sum_{i=1}^M [w_i p_i(Y(t))]} \left\{ \int_b^y \left[-U(t)f_i(t)p_i(U(t)) + f(t,U(t))p_i(U(t))\right] dU(t) \right\}$$
$$+ \frac{2\sum_{i=1}^M w_i}{\sum_{i=1}^M [w_i p_i(Y(t))]} \left\{ \int_b^y \int_b^u \left[\frac{1}{2}\frac{\partial^2[[s_i(t)X(t)]^2 p_i(X(t))]}{\partial X(t)^2}\right] dX(t)dU(t) \right\}$$
$$+ \frac{2}{\sum_{i=1}^M [w_i p_i(Y(t))]} \left\{ \int_b^y \int_b^u \int_0^{+\infty} \sum_{i=1}^M w_i h_i(t) \left[p_i(X(t)/J(t)) - p_i(X(t))\right]\phi_i(J(t),\mu_i(t),\delta_i^2(t))\right] dJ(t)dX(t)dU(t) \right\}$$
$$- \frac{2}{\sum_{i=1}^M [w_i p_i(Y(t))]} \left\{ \int_b^y \int_b^u \int_0^{+\infty} \sum_{i=1}^M w_i h(t,X(t)) \left[p_i(X(t)/J(t)) - p_i(X(t))\right]\phi(J(t),\mu(t),\delta^2(t))\right] dJ(t)dX(t)dU(t) \right\}. \quad (5.32)$$

Performing some further integrations one obtains:

$$s(t,Y(t))^2 = \frac{2\sum_{i=1}^M w_i}{\sum_{i=1}^M [w_i p_i(Y(t))]} \left\{ \int_b^y \left[f(t,U(t)) - U(t) f_i(t)) \right] p_i(U(t)) dU(t) \right\} + \frac{\sum_{i=1}^M w_i [s_i(t) Y(t)]^2 p_i(Y(t)}{\sum_{i=1}^M [w_i p_i(Y(t))]}$$
$$+ \frac{2}{\sum_{i=1}^M [w_i p_i(Y(t))]} \left\{ \int_b^y \int_b^u \int_0^{+\infty} \sum_{i=1}^M w_i h_i(t) \left[p_i(X(t)/J(t)) - p_i(X(t)) \right] \phi_i(J(t), \mu_i(t), \delta_i^2(t)) \right] dJ(t) dX(t) dU(t) \right\}$$
$$- \frac{2}{\sum_{i=1}^M [w_i p_i(Y(t))]} \left\{ \int_b^y \int_b^u \int_0^{+\infty} \sum_{i=1}^M w_i h(t, X(t)) \left[p_i(X(t)/J(t)) - p_i(X(t)) \right] \phi(J(t), \mu(t), \delta^2(t)) \right] dJ(t) dX(t) dU(t) \right\}. \tag{5.33}$$

5.8.1 Solving the IPDE

The main contribution here is to propose two strategies to solve the above IPDE and show a numerical implementation. This implementation does not need numerical inversion of Fourier transforms but simple numerical calculations of integrals: this is possible thanks to the fact that the distribution is known explicitly and can be regarded as a distinctive advantage of this model over more general models based on generic Lévy processes. More details on numerical methods for the IPDEs related to Lévy processes can be found in Cont & Tankov [45].

The first two lines of the above equation are solved, following Brigo [16], by setting:

$$f(t,(X(t)) = \sum_{i=1}^M w_i p_i(t, X(t)) X(t) f_i(t) / \sum_{i=1}^M w_i p_i(t, X(t)) \tag{5.34}$$

$$s(t, X(t))^2 = \sum_{i=1}^M w_i p_i(t, X(t)) s_i^2(t) X^2(t) / \sum_{i=1}^M w_i p_i(t, X(t)). \tag{5.35}$$

Because we incorporated the compensator term $h(t, X(t))k(t)$ in the above condition, we rewrite this expression by using two conditions.

$$\sum_{i=1}^M w_i p_i(t, X(t)) X(t) [m_i(t) - h_i(t) k_i(t)] / \sum_{i=1}^M w_i p_i(t, X(t)) = [m(t, X(t)) - h(t, X(t)) k(t)] \tag{5.36}$$

$$m(t,(X(t)) = \sum_{i=1}^M w_i m_i(t) X(t) / \sum_{i=1}^M w_i p_i(t, X(t)). \tag{5.37}$$

We propose the following Ansatz:

$$\sum_{i=1}^M w_i p_i(t, X(t)) h_i(t) k_i(t) / \sum_{i=1}^M w_i p_i(t, X(t)) = h(t, X(t)) k(t). \tag{5.38}$$

If the above conditions are satisfied, we are left with the final integral equation to be solved:

$$\int_b^y \int_b^u \int_0^{+\infty} \sum_{i=1}^{M} w_i h_i(t) \left[p_i(X(t)/J(t)) - p_i(X(t))\right] \phi_i(J(t), \mu_i(t), \delta_i^2(t))\Big] \, dJ(t) dX(t) dU(t) =$$
$$\int_b^y \int_b^u \int_0^{+\infty} \sum_{i=1}^{M} w_i h(t, X(t)) \left[p_i(X(t)/J(t)) - p_i(X(t))\right] \phi(J(t), \mu(t), \delta^2(t))\Big] \, dJ(t) dX(t) dU(t). \tag{5.39}$$

To solve this last problem, we perform some algebraic manipulations and then propose a zero-finding routine. Because above we have added the constraint $\sum_{i=1}^{M} w_i p_i(t, X(t)) h_i(t) k_i(t) / \sum_{i=1}^{M} w_i p_i(t, X(t)) = h(t, X(t)) k(t)$, known $k(t)$ one gets the local intensity

$$h(t, X(t)) = \sum_{i=1}^{M} w_i p_i(t, X(t)) h_i(t) k_i(t) / [k(t) \sum_{i=1}^{M} w_i p_i(t, X(t))].$$

If we know $k(t) = e^{\mu(t) + \frac{1}{2}\delta(t)^2} - 1$ we can determine the values of both $\mu(t)$ and $\delta(t)$. To move from a two-dimensional zero-finding problem to a more tractable one-dimensional problem, we add the constraint $\delta(t) = r|\mu(t)|$, where r is a positive real constant and where we assume that $\mu(t) \neq 0$.

Thanks to these constraints, the above integral equation is written only in terms of the unknown $\mu(t)$:

$$\int_b^y \int_b^u \int_0^{+\infty} \sum_{i=1}^{M} w_i h_i(t) \left[p_i(X(t)/J(t)) - p_i(X(t))\right] \phi_i(J(t), \mu_i(t), \delta_i^2(t))\Big] \, dJ(t) dX(t) dU(t) =$$
$$\int_b^y \int_b^u \int_0^{+\infty} \sum_{i=1}^{M} w_i h(t, X(t)) \left[p_i(X(t)/J(t)) - p_i(X(t))\right] \phi(J(t), \mu(t), (r|\mu(t)|)^2)\Big] \, dJ(t) dX(t) dU(t). \tag{5.40}$$

At this point a choice has to be made: we can choose a simpler global solution to the problem, which may deliver a good but not exact fit, or a local solution that can achieve exact fit. If one opts for the first option, one finds numerically an intensity $h(t)$ that is not a function of the state level by numerically solving the integral equation above for μ and freezing the probabilities $p_i(X(t))$ for example to a central level $p_i(X(t_0))$: because there are two integrations in the state level dimensions, errors are cancelled and compensated (integration is a smoothing operator).

This means that one finds an intensity that globally allows a good fit but does not guarantee an exact local fit: there is also an obvious computational advantage in having a single local parameter, because when simulating one does not need to interpolate or recalculate the local intensity. To sum up all these points, the integral equation to solve for $\mu(t)$ is finally:

$$\int_b^y \int_b^u \int_0^{+\infty} \sum_{i=1}^M w_i h_i(t) \left[p_i(X(t_0)/J(t)) - p_i(X(t_0))\right]\phi_i(J(t), \mu_i(t), \delta_i^2(t))\right] dJ(t)dX(t)dU(t) =$$

$$\int_b^y \int_b^u \int_0^{+\infty} \sum_{i=1}^M w_i \frac{\sum_{j=1}^M w_j p_j(t, X(t_0))h_j(t)k_j(t)}{\sum_{j=1}^M w_j p_j(t, X(t_0))k(t)} \left[p_i(X(t_0)/J(t)) - p_i(X(t_0))\right]\phi(J(t), \mu(t), (r|\mu(t)|)^2)\right] dJ(t)dX(t)dU(t)$$

where we have plugged in the Ansatz (5.38). Intuitively the above approach is likely to work better when the mixture distributions are not too different from each other, given that one is looking for a single distribution to summarise multiple distributions.

If one needs an exact fit to the distribution, one needs to ignore the two integrations in the state dimension and rewrite the above problem only considering the jump distribution: this avoids the compensations that were happening in the above multiple integrations. This is therefore rewritten with all known parameters being on the left hand side and the expression containing the unknown $\mu(t)$ on the right hand side. One removes the integrations across the space to focus on the jump size:

$$\int_0^{+\infty} \sum_{i=1}^M w_i h_i(t) \left[p_i(X(t)/J(t)) - p_i(X(t))\right]\phi_i(J(t), \mu_i(t), \delta_i^2(t))\right] dJ(t)$$

$$\int_0^{+\infty} \frac{\sum_{i=1}^M w_i h_i p_i(X(t))[e^{\mu_i(t)+\frac{\delta_i^2(t)}{2}} - 1] \sum_{i=1}^M w_i[p_i(X(t)/J(t)) - p_i(X(t))]}{[e^{\mu(t,X(t))+\frac{(r|\mu(t,X(t))|)^2}{2}} - 1] \sum_{i=1}^M w_i p_i(t, X(t))} \phi(J(t), \mu(t, X(t)), (r|\mu(t, X(t))|)^2)]dJ(t). \tag{5.42}$$

Known the mixture parameters $h_i(t)$, $\mu_i(t)$, and $\delta_i(t)$, (and therefore $k_i(t)$ as well) and made an assumption on r, one finds $\mu(t, X(t))$ by discretising the integrals above and by performing a search for zeros, for example with the bisection method. With one gets $\delta(t, X(t)) = r|\mu((t, X(t)), X(t))|$, and then $k(t, X(t)) = e^{\mu(t,X(t))+\frac{1}{2}\delta(t,X(t))^2} - 1$.
The final result reads $h(t, X(t)) = \sum_i^M w_i p_i(t, X(t))h_i(t)k_i(t)/[k(t)\sum_{i=1}^M w_i p_i(t, X(t))]$. In practice, for a level $X(t)$ one finds $\mu(t, X(t))$ such that $\sum_i^M w_i p_i(t, X(t)) = p(t, X(t))$.

We show some numerical examples of this routine implementation that show that it is working in a satisfactory way.

As an example, we assume to be working with a two-dimensional uncertain-parameters model whose jump diffusion equations have parameters $m_1 = 0, s_1 = 0, \mu_1 = 0.025, \delta_1 = 0.017, h_1 = 8$ and $m_1 = 0, s_1 = 0, \mu_1 = 0.035, \delta_1 = 0.02, h_1 = 4$ respectively: for clarity we are dropping the time notation from the parameters. The two models are have weighs $w_1 = 0.8$ and $w_2 = 0.2$, and we are interested to find the local parameters of the jump diffusion for the one year maturity. Before doing any calculation, we

notice that for simplicity the drift and the diffusion parts are both zero, so we do not expect any values for these parameters: this does not affect the generality of our result. We implemented a numerical calculation for the error term (5.41), as a function of the average log-jump size μ. We assume that $r = 0.6582$, based on the ratio of the weighted average of the mixture log-jump standard deviations (0.8×0.01 7+ 0.5×0.02=0.03) on the weighted average of the log-jump averages ($0.8 \times 0.025 + 0.2 \times 0.04$=0.035). Therefore we expect the parameter μ to be very close to 0.03: indeed our zero-finding routine finds $\mu = 0.02700698$. Therefore $\delta = 0.01817064$ and $h = 6.7909844$.

To check these results, we compare the distributions of the original mixture against the distribution of the new jump-diffusion process that summarises the uncertain-parameters model.

The below graph refers to the case where we find a global intensity $h(t, X(t_0))$, which, as explained above, does guarantee a good fit but not an exact one, especially on the tails.

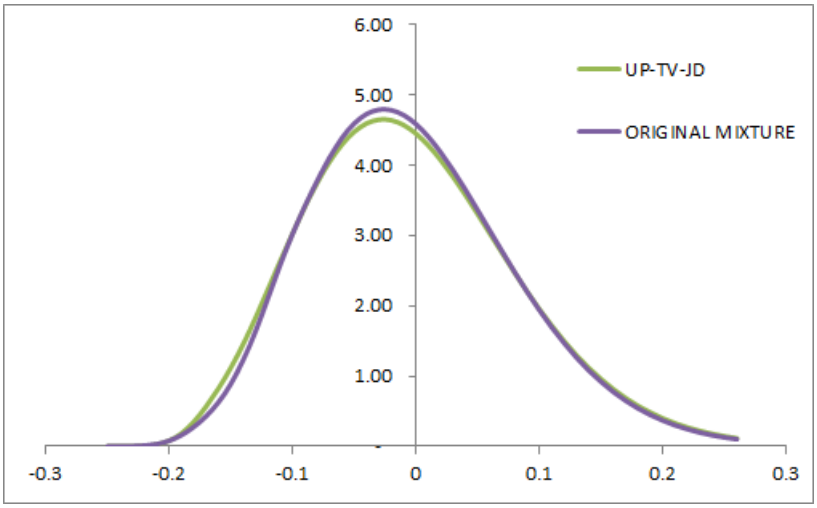


Figure 5.8: Log-Price distribution comparison: the green line is the original mixture, the mauve line is the UP-TV-JD model estimated numerically based on the IPDE analysis shown above.

We calculate to the exact fit via local parameters. The below tab shows what levels of the local intensity parameter $\mu(t, X(t))$ one has to choose to ensure an exact fit between the original mixture distribution and the local-UP-TV-JD model. This has been calculated by numerically solving equation (5.42) as a function of $\mu(t, X(t))$ on a log-price grid that can be made as fine as needed. The error shown is caused by the numerical implementation of the integrations.

Log price	Abs. error	$\mu(t, X(t))$
-0.2	0.00000025	0.02840599
-0.15	0.00000062	0.02862033
-0.1	0.00000001	0.02831014
-0.05	0.00000076	0.02873111
0	0.00000017	0.0284884
0.05	0.00000062	0.02838171
0.1	0.00000052	0.02831064
0.15	0.00000077	0.02826631
0.2	0.00000031	0.0282461

Lastly, we make some remarks regarding the use of the proposed methodology in the wider context of skew modelling: if we have calibrated an uncertain-parameters time-inhomogeneous jump-diffusion model to the lognormal price index volatilities implied by market prices of inflation zero-coupon and/or year-on-year options, this method allows one to compress the uncertain-parameter model into a local-volatility time-inhomogeneous jump-diffusion, where the parameters of the jump distribution $\mu(t)$, and $\delta(t)$ are a function of time and can be a function of the state of the process if one opts for the exact fit via local calibration. One crucially notes that the methodology developed in the previous section (which allows to find the equivalent time-homogeneous jump diffusion given a time-inhomogeneous jump diffusion) is used in the calibration phase as it allows one to price options in closed form.

This more complex model can be used to price path dependent exotic derivatives either via a Monte Carlo simulation or via an IPDE finite differences grid.

5.9 Skew market calibration example

We extend the calibration exercise presented in chapter 4 to capture the prices of high- and low-strikes zero-coupon inflation options. To achieve this result, we calibrate an uncertain-parameters time-varying jump-diffusion model (where the mixture is composed by two separate models), that can be compacted using the IPDE technique presented above. The resulting time-varying jump-diffusion model can then be regarded as an uncertain-parameters versions of the CTCB model.

The calibration is done using the data as of 7th December 2012: the curves and the at-the-money option prices where listed in chapter 4, while the prices of zero-coupon and year-on-year inflation options were listed earlier in this chapter (section 3).

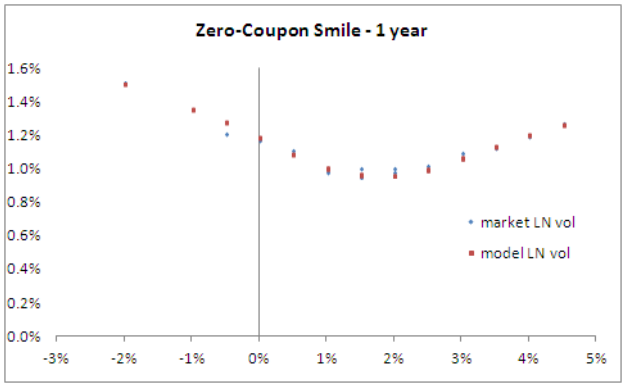
The main finding of this exercise is that the prices of zero-coupon and year-on-year inflation options

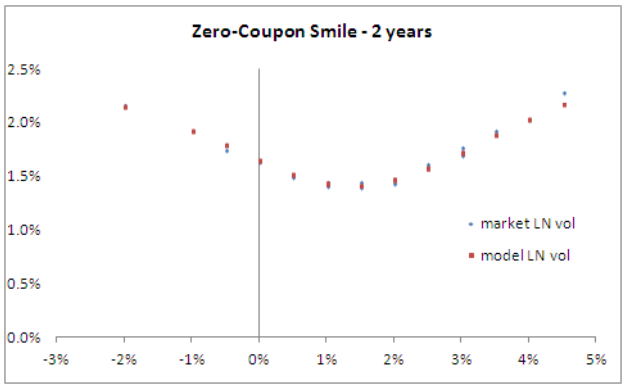
are not consistent with each other, especially for long maturities. We choose to calibrate to the prices to zero-coupon inflation options and use this parameter set to price year-on-year options to check if the prices are consistent with each other, and we find that this is not the case. This shows that these two markets are not perfectly liquid and that an arbitrage can be potentially set up.

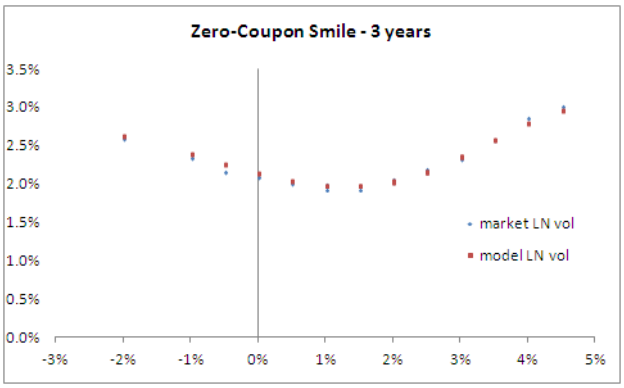
The table below shows the model parameters for the bivariate jump-diffusion mixture that we have fitted to the zero-coupon inflation option prices for the maturities up to three years.

Maturity	σ_1	h_1	μ_1	δ_1	w_1	σ_2	h_2	μ_2	δ_2	w_2
1	0.00411	0.5	0.00621	0.01	0.4	0.00444	0.85	-0.00521	0.01	0.6
2	0.00993	0.5	0.00326	0.03851	0.4	0.01111	0.85	0.00444	0.00181	0.6
3	0.01461	0.5	0.03554	0.05308	0.4	0.01263	0.85	-0.02056	0.00001	0.6

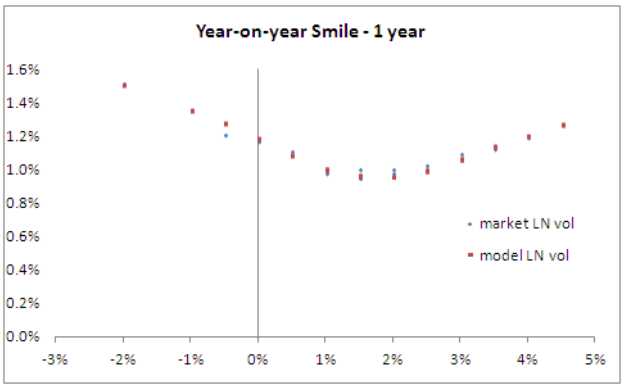
The graphs below show the calibration fit to the lognormal price index volatilities implied by the zero-coupon inflation option prices. The fit is very good. The market volatility for the -0.5% strike are clearly an outlier perhaps due to low liquidity.

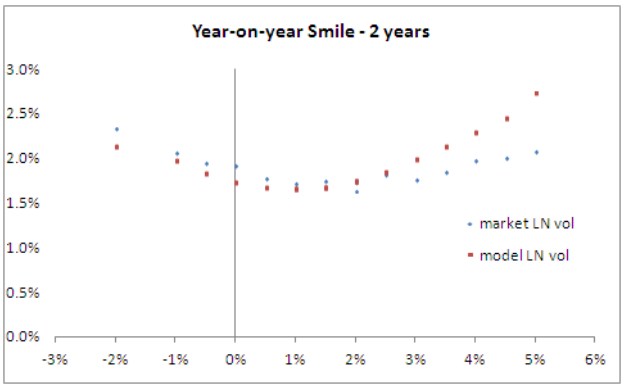


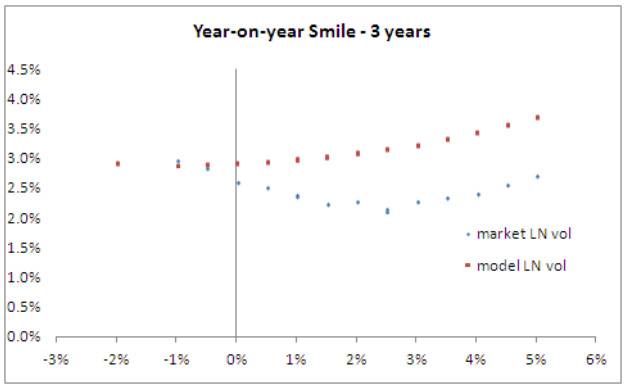




The graphs below show instead the fit to the lognormal price index volatilities implied by the market prices of the year-on-year options obtained with the parameter set calibrated on the zero-coupon options. From these it is clear that the prices of the two trade types are not consistent. In particular, the year-on-year volatilities look cheap and less skewed on the upside compared against the prices of the zero-coupon options.






Year-on-year Smile - 3 years
4.5%
4.0%
3.5%
3.0%
2.5%
2.0%
1.5%
1.0%
0.5%
0.0%
-3%
-2%
-1%
0%
1%
2%
3%
4%
5%
6%
market LN vol
model LN vol

Chapter 6

Counterparty and funding risk aspects of inflation derivatives

In the recent years much attention of financial mathematics researchers and practitioners alike has been devoted to modelling counterparty and funding risks of derivatives. The Lehman crisis and the following credit squeeze have shown that a pure risk-free valuation of derivatives no longer explains the prices observed in the markets: in particular, the likelihood of default of the counterparty and the investor has to be taken into account. Regulatory bodies and accounting principles have followed the market practice, by requiring CVA (Credit Valuation Adjustment) and DVA (Debt Valuation Adjustment) to be accounted for in the balance sheet (FASB, IFRS accounting regulations) and by adding a CVA capital buffer in the Basel III regulations. Regulators and accounting principles are not always consistent: for example the Basel III framework ignores the DVA benefit.

To mitigate counterparty risk, financial institutions post collateral to each other: collateral is typically remunerated at OIS (Overnight Index Swap) rate, which has some funding implications. Meanwhile, funding has become expensive, and therefore any interest on any cash amount needed or obtained to set up a trade and to post collateral has to be taken into account in the trade pricing. It is becoming common market practice to take into account a Funding Valuation Adjustment (FVA), although this has given rise to much debate and there is no broadly accepted way to define it, let alone to calculate it. Even worse, there are some overlaps between FVA and DVA, which makes even the definition of the FVA a matter of debate. Finally, the use of collateral to mitigate credit risk has consequences on the discounting rate that has to be used to price derivatives. It is easy to imagine that all these new features pose significant modelling challenges and divert the attention from the derivative payoff complexity to the general pricing framework that takes into account all these features, namely counterparty and funding

risks (they are generally referred to as XVA).

The model proposed in chapter 3 is both powerful and analytically simple to be adapted and extended to include counterparty and funding risks aspects of inflation and fixed income derivatives: in this chapter we add the credit model for counterparty and investor default risk and extend the model to cope with other asset classes. To do this, we use a Marshall-Olkin model, where only a part of the intensities is directly correlated with the economy: we split the intensities into a systematic and an idiosyncratic part, where the latter is not related to the economy at all. We briefly review the Marshal-Olkin model: more references can be found in Marshall & Olkin [92], Lindskog & McNeil [89], Brigo, Pallavicini & Torresetti [28], Brigo & Choudarkis [19], Mai & Scherer [91], and Brigo, Mai & Scherer [21]. The final result is a general framework where some meaningful macroeconomic dynamics drive both the evolution of the market and the default of market participants. Further, we explore some methods to reduce the number of parameters of the Marshall-Olkin model, which can explode if not managed properly when the number of counterparties is high.

The subject is in continuous evolution and there are still many open questions that have not found a final answer yet: the objective of this chapter is not to take any side in these debates or propose a solution to the open questions (this would be a considerable task): we present some instruments and make some reasonable assumptions to extend the CTCB model presented in chapter 3 to include counterparty and funding aspects of inflation and fixed income derivatives. A Monte Carlo simulation concludes the chapter by showing some results under different market and collateralisation regimes. The simulation results can be checked by using some approximations proposed here for the first time.

6.1 Definitions, choices and fundamental results

In this section we define the main concepts that are needed to understand the theory behind the counterparty and funding adjustments. We also make explicit choices for our pricing framework. In this chapter the notation for the negative part X^- refers to a negative quantity: $X^- = \min(0, X)$. For the indicator $\mathbb{I}$, that takes value one when a certain condition $\mathcal{A}$ is satisfied and zero if not, we use indifferently the notations $\mathbb{I}_{\{\mathcal{A}\}}$ or $\mathbb{I}[\mathcal{A}]$, whenever we feel the notation is easier to read.

CSA mechanics. The Credit Support Annex (CSA) is a legal document that details the rules used to post or receive collateral in a derivatives transaction. In this chapter we define the collateral margin process $\{M_i(t)\}_{t\geq 0}$ as the value of collateral posted (if negative) or received (if positive). This is a function of the value of the derivatives portfolio that a given financial institution $\mathcal{B}$ has with a given counterparty $\mathcal{C}_i$ (netting set): we indicate this value process as $\{V_i(t)\}_{t\geq 0}$. We assume that the financial institution $\mathcal{B}$ has trades with N counterparties: $i \in \{1, ..., N\}$. To make notation lighter, we drop the counterparty

$\mathcal{C}_i$ indicator i whenever we assume to be working with only one counterparty, and write the collateral as $M(t)$.

For example, if there is perfect collateralisation, we always have $V_i(t) = M_i(t)$. Independent amounts, i.e. collateral amounts not directly linked to the value of the derivatives portfolio, can also be posted. Alternatively collateral can be updated only when the exposure $V_i(t) - M_i(t)$ is above or below a certain threshold, and/or can be updated on a fixed set of dates t_j: $M_i(t) = V_i(t_j)$ for $t \in (t_j, t_{j+1}]$. Finally in case of almost perfect collateralisation, the full value of the trades is posted/received one instant later, i.e. $V_i(t^-) = M_i(t)$: this leaves both parties exposed only to jumps and discontinuities. Collateral can be posted either in cash or in securities: in this chapter we assume that only cash is posted, and this cash is remunerated at collateral rate. We also assume that collateral can be always rehypothecated, which means that is re-pledged to another counterparty by the financial institution that has received it in another transaction. The main rules used to post or receive collateral can be found for example in Cesari et al. [38].

CSA and discounting in the post-Lehman environment. Piterbarg [101] and Kenyon & Stamm [82] show that in a fully collateralised transaction, discounting has to be performed using OIS rates[1]: if there is no collateralisation an unsecured rate like Libor should be used. Piterbarg also shows that in case of non perfect collateralisation the Black-Scholes equation for a contingent claim portfolio $V(t, S(t))$ becomes $\mathcal{L}_S V(t) = n'(t)M(t) + n(t)(V(t, S(t)) - M(t))$, where $n'(t)$ is the instantaneous OIS rate, $n(t)$ is the the instantaneous Libor rate, the diffusion operator $\mathcal{L}_S$ is defined as $\mathcal{L}_S = \partial_t + n^S S(t)\partial_S + \frac{1}{2}\sigma^2(S(t), t)S(t)^2\partial_{SS}$, and $n^S(t)$ is the repo rate on funding secured by the asset $S(t)$. We have written here $V(t) = V(t, S(t))$ to stress the derivative price dependence from the underlying level $S(t)$. Piterbarg makes this example for an equity derivative, assuming that there are no dividends and starting from the usual Brownian $\mathbb{P}$-dynamics $dS(t) = S(t)[\mu(S(t), t)dt + \sigma(S(t), t)dW(t)]$: this result is generalised for any payoff of any asset class. For simplicity, in this chapter we assume that there is only one currency and ignore the cases where the collateral and the derivatives portfolio currencies are different.

Risk neutral measures. If one introduces credit risk, collateralisation, and OIS discounting, in the model there is a wider choice of pricing measures. We use as numeraire the OIS bank account $B'(t) = B'(0)e^{\int_0^t n'(s)ds}$ and introduce the OIS risk-neutral measure $\mathbb{Q}'$. Alternatively one can use a defaultable bond $P^i(t, T)$ issued by the counterparty $\mathcal{C}_i$ and introduce the i-credit risky T-forward measure $\mathbb{Q}_i^T$: to do this one has to assume that the bond price, in case of default, can never hit zero (i.e. has strictly positive recovery rate—recovery rates are fully defined below). Finally one can use a first-to-default

[1] The use of OIS rates, that refer to unsecured overnight interbank lending, is strictly speaking theoretically wrong given that these transactions have some residual credit risk left. However it is market practice to use such rate as the best proxy for the risk-free rate. OIS swaps are quoted for all maturities: in such swaps there are fixed payments against the average OIS rate in a given period. These fixed payments are seen as the market-implied path for the OIS rate.

bond $P^{\mathcal{C}_i \wedge \mathcal{B}}(t,T)$ that pays 1 unit of currency at maturity only if both the financial institution $\mathcal{B}$ and the counterparty $\mathcal{C}_i$ survive before the bond maturity: this introduces the first-to-default T-forward measure $\mathbb{Q}^T_{\mathcal{C}_i \wedge \mathcal{B}}$ (again one needs to assume that the numeraire bond has strictly positive recovery).

Positive exposures. Exposure of the financial institution $\mathcal{B}$ with respect to counterparty $\mathcal{C}_i$ is defined as $E_i(t) = V_i(t) - M_i(t)$: to keep notation light we can assume where possible that the financial institution has exposure only towards one counterparty and write $E(t) = V(t) - M(t)$. Positive exposure is defined as $E_i^+(t) = \max[0, E_i(t)]$.

Negative exposures. Negative exposure of the financial institution $\mathcal{B}$ with respect to counterparty $\mathcal{C}_i$ is defined as $E_i^-(t) = \min[0, E_i(t)]$.

Default times. The default times of the financial institution $\mathcal{B}$ and for the counterparty $\mathcal{C}_i$ are denoted τ_B and τ_{C_i} respectively. Again to make notation simpler, if we are working with only one counterparty we drop the counterparty index i and write τ_C. From a probabilistic point of view, the default time is a stopping time.

Recovery rates. In case of default of the financial institution $\mathcal{B}$ and of the counterparty $\mathcal{C}_i$, one assumes that the creditors only recover a fraction R_B and R_{C_i} respectively. Both numbers are between 0 and 1. In this article we make the simplifying assumption that all recovery rates are deterministic. One needs to be careful on zero recovery rates as they do not let one move to the risky T-forward measure, as explained below. This is not a major constraint as we do not use this measure in this chapter.

Loss given default. Assuming that the counterparty $\mathcal{C}_i$ has defaulted at time τ_{C_i}, the loss given default for the financial institution $\mathcal{B}$ is defined as $LGD_i = (1 - R_{C_i})E_i^+(\tau_{C_i})$: this concept is used below to introduce the CVA and is by construction a positive number. One assumes that the financial institution $\mathcal{B}$ has a gain if it defaults when it owes money to counterparty $\mathcal{C}_i$, in which case it makes a positive gain of $LGD_i^B = -(1 - R_B)E_i^-(\tau_B)$. This concept is used below to introduce the DVA.

Unilateral CVA. The unilateral CVA with respect to counterparty $\mathcal{C}_i$ is defined as the risk-neutral expectation of the loss given default for this counterparty: there is no mention to the potential default of the financial institution $\mathcal{B}$ in this definition.

In formulas we have $CVA_i^U(t_0) = (1-R_{C_i})\mathbb{E}^{\mathbb{Q}}_{t_0}[e^{-\int_{t_0}^{\tau_{C_i}\wedge T_i} n(s)ds} E_i^+(\tau_{C_i})]$ which we rewrite as $CVA_i^U(t_0) = \mathbb{E}^{\mathbb{Q}}_{t_0}[e^{-\int_{t_0}^{\tau_{C_i}\wedge T_i} n(s)ds} LGD_i]$: we stress that using this definition the CVA is a positive quantity, even if is a balance sheet liability. Here the time T_i is the maximum maturity across all trades that the financial institution $\mathcal{B}$ has with counterparty $\mathcal{C}_i$. Strictly speaking one does not need to work under the risk-neutral measure: considering the generic measure $\mathbb{N}$ defined by the numeraire asset $N(t)$ one writes: $CVA_i^U(t_0) = (1 - R_{C_i})\mathbb{E}^{\mathbb{N}}_{t_0}[N(t)/N(\tau_{C_i})E_i^+(\tau_{C_i})]$. This said, in the rest of this chapter we work with the risk-neutral measure $\mathbb{Q}$ for the CVA-DVA calculations (and therefore discounting using Libor), to stress the fact that the CVA-DVA amounts are unsecured, even if there is any CSA with the counterparty on

the trades that are part of the netting set. Later in this chapter we will discuss XVA discounting issues more in detail.

We make the dependence on the default time τ_{C_i} explicit and rewrite the CVA expression as: $CVA_i^U(t_0) = (1 - R_{C_i})\mathbb{E}_{t_0}^{\mathbb{Q}}[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n(s)ds} E_i^+(t)\mathbb{I}[\tau_{C_i} \in [t, t + dt)]]$.

Unilateral DVA. The unilateral DVA with respect to counterparty $\mathcal{C}_i$ is defined as the risk-neutral expectation of the loss given default of the investor $\mathcal{B}$, considering only the netting set of positions with counterparty $\mathcal{C}_i$, whose default is ignored. In formulas $DVA_i^U(t_0) = -(1-R_B)\mathbb{E}_{t_0}^{\mathbb{Q}}[e^{-\int_{t_0}^{\tau_B \wedge T_i} n(s)ds} E_i^-(\tau_B)]$ which we rewrite as $DVA_i^U(t_0) = \mathbb{E}_{t_0}^{\mathbb{Q}}[e^{-\int_{t_0}^{\tau_B \wedge T_i} n(s)ds} LGD_i^B]$: we stress that using this definition the DVA is a positive quantity, being a balance sheet asset. As for the CVA, one does not need to work under the risk-neutral measure: considering the generic measure $\mathbb{N}$ defined by the numeraire asset $N(t)$, one writes: $DVA_i^U(t_0) = -(1 - R_B)\mathbb{E}_{t_0}^{\mathbb{N}}[N(t)/N(\tau_B)E_i^-(\tau_B)]$: the same comment made on Libor discounting for CVA applies to DVA.

We make the dependence on the default time τ_B explicit and rewrite the DVA expression as: $DVA_i^U(t_0) = -(1 - R_B)\mathbb{E}_{t_0}^{\mathbb{Q}}[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n(s)ds} E_i^-(t)\mathbb{I}[\tau_B \in [t, t + dt)]]$ where the time T_i is the maximum maturity across all trades that the financial institution $\mathcal{B}$ has with counterparty $\mathcal{C}_i$.

Bilateral (or First-to-default (FTD)) CVA. The bilateral (or first-to-default) CVA is the CVA calculated only over the paths where the counterparty $\mathcal{C}_i$ defaults before the financial institution $\mathcal{B}$.

In formulas: $CVA_i^B(t_0) = (1 - R_{C_i})\mathbb{E}_{t_0}^{\mathbb{Q}}[e^{-\int_{t_0}^{\tau_{C_i} \wedge \tau_B \wedge T_i} n(s)ds} E_i^+(\tau_{C_i})\mathbb{I}_{\{\tau_{C_i} < \tau_B, \tau_{C_i} < T_i\}}]$.

We make the dependence on the default time τ_{C_i} explicit and rewrite the CVA expression as: $CVA_i^B(t_0) = (1 - R_{C_i})\mathbb{E}_{t_0}^{\mathbb{Q}}[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n(s)ds} E_i^+(t)\mathbb{I}[\tau_{C_i} \in [t, t + dt), \tau_{C_i} < \tau_B]]$ where the time T_i is the maximum maturity across all trades that the financial institution $\mathcal{B}$ has with counterparty $\mathcal{C}_i$. The same comments made above on the use of different pricing measures apply. Brigo & Capponi [18] have originally introduced the distinction between bilateral and unilateral valuation adjustments.

Bilateral (FTD) DVA. The bilateral (or first-to-default) DVA is the DVA calculated only over the paths where the financial institution $\mathcal{B}$ defaults before the counterparty $\mathcal{C}_i$.

In formulas: $DVA_i^B(t_0) = -(1 - R_B)\mathbb{E}_{t_0}^{\mathbb{Q}}[e^{-\int_{t_0}^{\tau_{C_i} \wedge \tau_B \wedge T_i} n(s)ds} E_i^-(\tau_B)\mathbb{I}_{\{\tau_B < \tau_{C_i}, \tau_{C_i} < T_i\}}]$.

We make the dependence on the default time τ_B explicit and rewrite the DVA expression as: $DVA_i^B(t_0) = -(1 - R_B)\mathbb{E}_t^{\mathbb{Q}}[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n(s)ds} E_i^-(t)\mathbb{I}[\tau_B \in [t, t + dt), \tau_B < \tau_{C_i}]]$ where the time T_i is the maximum maturity across all trades that the financial institution $\mathcal{B}$ has with counterparty $\mathcal{C}_i$. The same comments made above on the use of different pricing measures apply.

Filtrations. In this chapter we follow the usual convention to use a separate filtration for the market risk factors information (including credit spreads) and for the credit events respectively. In particular we define the sub-filtration $\{\mathcal{F}_t\}_{t \geq 0}$ as the filtration modelling all market information except defaults. We define the default sub-filtration $\{\mathcal{H}_t\}_{t \geq 0}$ as the filtration modelling defaults: $\{\mathcal{H}_t\}_{t \geq 0} = \sigma([\tau_B \leq$

$u] \vee [\tau_{C_1} \leq u] \vee .... \vee [\tau_{C_N} \leq u], u \leq t)$. For the calculations that involve also the default times one uses the σ-algebra $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t$.

CVA/DVA and collateralisation. Perfect collateralisation brings both the CVA and DVA to zero, as there is no counterparty exposure during the life of the trades. In any other cases collateralisation reduces the exposures but not completely, leaving some CVA/DVA. Clearly no collateralisation is the case where the CVA/DVA exposure is maximum: this usually happens when a bank is facing a small or medium enterprise that can not set up a CSA.

Credit-risky versus credit risk-free present value. Given the definitions of CVA and DVA (either unilateral or bilateral), it has become market practice to refer to the value of a derivatives portfolio that the financial institution $\mathcal{B}$ has with counterparty $\mathcal{C}_i$ in terms of risk-free or risk-adjusted value. The risk-free value $V_i^*(t)$ is the value of the derivatives of the netting set of counterparty $\mathcal{C}_i$ calculated assuming that there is no credit risk for both the financial institution $\mathcal{B}$ and counterparty $\mathcal{C}_i$: in practice this means an OIS-discounted payoff (i.e. one removes the counterparty risk by posting collateral), assuming that the collateral value follows instantly the trade value. In order to take into account the credit risk of the financial institution $\mathcal{B}$ and counterparty $\mathcal{C}_i$, one introduces the credit-risky value defined as $V_i(t) = V_i^*(t) - CVA_i^X(t) + DVA_i^X(t)$, for $X \in \{U, B\}$. We note that this is the value of the derivatives portfolio that has to be used in the balance sheets according to FASB and IFRS (even if they don't specify whether one should use unilateral or bilateral valuation adjustments). Bilateral valuation adjustments are mentioned in the Basel II papers. In this chapter, in the following sections we opt for bilateral CVA and DVA for the adjustment, obtaining finally: $V_i(t) = V_i^*(t) - CVA_i^B(t) + DVA_i^B(t)$. We should stress that such definition, although is very common in the industry, is not completely correct from a theoretical perspective if some assumptions are made: Brigo, Perini & Pallavicini [27] stress that separating a CVA and DVA term is not possible under some assumptions.

Risky versus riskless close-out. An essential assumption that has to be made to calculate the valuation adjustments is whether the loss given default of either party includes the CVA and DVA values (replacement close-out) or not (riskless close-out). We refer to Brigo, Morini & Pallavicini [26] (in particular chapter 14) for the definitions.

In this chapter we assume riskless close-out, which simplifies the problem considerably. If one opts for replacement close-out, the problem becomes recursive and backward SDEs techniques are needed to tackle it.

Wrong-way risk. Wrong-way (or alternatively right-way) risk is an expression referring to the effect that the co-movement of the counterparty (or the financial institution) credit risk and the risk-free value of the derivatives portfolio may have on the valuation adjustment. For example, if one assumes that the credit spreads of a counterparty is positively correlated with the inflation levels (for example, the

counterparty is more likely to default in a hyperinflation crisis), a bank buying an at-the-money zero-coupon inflation swap to a counterparty without CSA would be exposed to wrong-way risk. As inflation goes up, the trade becomes an asset for the bank (i.e. it has positive PV), and the counterparty becomes more likely to default. The exposure goes up and the default probability goes up at the same time, making the CVA more severe.

From a modelling perspective, we introduce a correlation term between the increment of the stochastic processes of the credit intensity and the other market variables, and/or use a copula between default times and the market variables. Wrong-way risk usually is a second order effect, unless one is dealing with pathological cases: because of this, it is market practice to calculate CVA/DVA of some products by assuming either no co-movement or no stochasticity in the credit intensities (therefore ignoring wrong-way risk). As we show in the next section, this allows one to develop intuition and to propose some option-based CVA/DVA approximated pricing formulas.

Credit correlation and CVA/DVA. In a bilateral CVA/DVA setting, or when a credit-contingent payoff is part of the netting set, the credit-credit correlation becomes an essential ingredient for the model. In these cases the use of a copula (as opposed to some correlated credit spread/intensity processes) introduces a stronger credit dependence, as suggested also in Brigo, Morini & Pallavicini [26]. It should be stressed that in the case of maximal dependence, simultaneous default is guaranteed only when the default probabilities of two names are the same: this is a well-known limitation of copula functions.

CVA and DVA carry and monetisation. The CVA and DVA are accounting adjustments that are marked-to-market, and therefore give profit and loss volatility to the financial institution. The CVA of a specific trade starts at a negative value at inception and is zero at maturity, as time passes and the default probability becomes smaller. Clearly any change in the mark-to-market of the derivative or in the market-implied probability of default has an impact on the CVA: if these two effects are taken out, the CVA decreases (i.e. becomes less negative) with time, i.e. has a positive carry (theta). If one hedges the CVA credit component, i.e. buys protection with a CDS on the counterparty, the CDS premium paid gives negative carry: clearly this is an idealised situation as the notional of the CDS protection varies with the mark-to-market of the derivative. Charging the CVA means charging the expected cost of buying protection on this trade. The opposite can be said for the DVA. The DVA starts at a positive level and decays to zero at maturity, and therefore has negative carry. Hedging the DVA means selling protection on itself (if possible) or on a very correlated entity (or basket of entities) and therefore receiving the protection premium (positive carry).

The FVA debate: some possible definitions. In the very recent years a new valuation adjustment has been introduced by some important financial institutions: even if they are not obliged to account for it by accounting rules, some banks calculate a funding valuation adjustment (FVA) to take into account

funding (and other) costs of the positions. We stress that there is no commonly agreed definition for it.

We start from a first definition of FVA assuming that there is no credit risk, and later add further elements of complexity to reach a final definition. To start, the FVA could be defined as the expected total funding cost or benefit of a given CSA, discounted by a generic discount rate $n^D(t)$: later in this section we will show that the rate used to discount FVA is actually irrelevant as it disappears in the calculations (this is shown both by Brigo, Pallavicini & Perini [27] and by Elouerkhaoui [53]). We stress that there is funding benefit or cost even if there is no CSA: for example, in the case of a partially collateralized trade with positive mark-to-market, the bank has to fund the cash needed to set it up. In a perfectly collateralised trade there will be an FVA component only if the collateral rate $n^c(t)$ is different from the risk-free rate: in this chapter, as we are assuming that the collateral is cash, we have this effect only if the effective collateral rate is different from the OIS rate, which can be due to liquidity spreads $l^c(t) = n^c(t) - n'(t)$. This component is sometimes referred to as LVA (Liquidity valuation adjustment) or COLVA (Collateral valuation adjustment) in Burgard & Kjaer [30]. We include this component in our FVA definition.

If the collateralisation is not perfect, funding the amount $V_i(t) - M_i(t)$ with counterparty $\mathcal{C}_i$ has a cost (the funding cost $n^B(t)$ of the institution $\mathcal{B}$, that must finance the position minus the collateral to be posted to the counterparty $\mathcal{C}_i$), and a smaller benefit (the OIS rate $n'(t)$ that is paid by the counterparty $\mathcal{C}_i$ on the collateral posted to the institution $\mathcal{B}$). The situation is reversed when collateral is received (funding benefit). Clearly if there is no credit risk there is no reason to include close-outs in the definition. Before giving the definition we define the generic risk-neutral measure $\mathbb{Q}^D$, that is defined by the bank account numeraire that grows with the generic rate $n^D(t)$. Given the above reasoning we sketch a first definition of FVA without credit risk ($FVA^{NCR}_{i,\mathcal{B}}(t)$):

$$FVA^{NCR}_{i,\mathcal{B}}(t_0) = \mathbb{E}^{\mathbb{Q}^D}_{t_0}\left[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n^D(s)ds}\{[n^c(t) - n'(t)]M_i(t) + [n^B(t) - n'(t)](V_i(t) - M_i(t))\}dt\right]. \tag{6.1}$$

At this stage we make these remarks:

1. The funding rate $n^B(t)$ is the short funding rate implied by the traded bonds of the institution $\mathcal{B}$, assuming they exist for all maturities. The financial meaning of this concept is that the institution $\mathcal{B}$ can fund itself receiving either overnight deposits or collateral in a derivatives transaction under CSA.

2. One can write the funding rate $n^B(t)$ as a funding spread $f^B(t)$ over the risk-free OIS rate $n'(t)$, therefore setting: $n^B(t) = n'(t) + f^B(t)$. If one is using intensity-based credit models (fully defined below) and assumes that they are constant, that there is no bond-CDS spread, that interest rates

are independent from credit spreads, and that there is zero recovery rate, then the short-term funding spread $f^B(t)$ is the same as the default intensity $h^B(t)$: in the following section we assume that this is not the case and continue using the funding spread in the FVA definition. We also use the collateral spread $l^c(t) = n^c(t) - n'(t)$ to achieve a lighter notation.

$$FVA_{i,\mathcal{B}}^{NCR}(t_0) = \mathbb{E}_{t_0}^{\mathbb{Q}^D}\left[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n^D(s)ds}\{l^c(t)M_i(t) + f^B(t)(V_i(t) - M_i(t))\}dt\right]. \tag{6.2}$$

3. Funding in an uncollateralised derivatives transaction is regarded as short-term funding (in practice, overnight), given that the sign of the collateral can shift from positive to negative and vice-versa. This funding is under no circumstances to be regarded as term (stable) funding. Clearly there is an open risk of refinancing, in case the funding rate $n^B(t)$ spikes: this is captured when one takes the expectation over the future possible paths. The way this risk is allocated between the different desks of a bank is a matter of debate: the client desk can be charged for term funding, and the treasury desk can be left with the refinancing risk.

4. The above definition of FVA is symmetric and can be summed across different netting sets. Symmetry in this context means that the formula is the same for both short and long cash/collateral positions.

5. In order to achieve even greater generality, we assume that there are different funding and collateral spreads when the institution $\mathcal{B}$ is either receiving cash, or when it is paying or posting it: these spreads are defined as $f_+^B(t)$, $l_+^c(t)$ and $f_-^B(t)$, $l_-^c(t)$ respectively. By taking into account these features the above definition of FVA becomes:

$$\mathbb{E}_{t_0}^{\mathbb{Q}^D}\left[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n^D(s)ds}\{l_+^c(t)[M_i(t)]^+ + l_-^c(t)[M_i(t)]^- + f_+^B(t)[V_i(t) - M_i(t)]^+ + f_-^B(t)[V_i(t) - M_i(t)]^-\}dt\right].$$

The first and second terms represent the LVA. The third term in the integral above ($f_+^B(t)[V_i(t) - M_i(t)]^+$) is the funding benefit: if the financial institution is overall long cash, it uses it to finance itself, thus saving its financing spread over OIS. If instead the financial institution is overall short cash (i.e. $V_i(t) - M_i(t) < 0$), the funding cost is $f_-^B(t)[V_i(t) - M_i(t)]^-$, which is the fourth term of the integral above.

Whenever the funding rates become asymmetric, BSDE techniques are required to perform the calculations, the intuition being that the problem becomes path-dependent (i.e. there are some paths where the collateral account is negative, and one needs to integrate over the simulated funding rate, and vice-versa). The use of BSDE techniques in counterparty and funding risk is explained in Guyon &

Labordère [61] and in Crépey & Bielecki [47]. This is not current industry practice given the computational burden of such techniques: however, Brigo, Liu, Pallavicini & Sloth [20] show that the difference between a simplified and a full calculation can be material when the two spreads $f_+^B(t)$ and $f_-^B(t)$ diverge significantly.

We derive an FVA definition with credit risk from the above formula, by requiring that both institutions $\mathcal{B}$ and $\mathcal{C}_i$ are not defaulted before time t:

$$FVA_{i,\mathcal{B}}(t_0) = \mathbb{E}_{t_0}^{\mathbb{Q}^D}\left[\int_{t_0}^{T_i} e^{-\int_{t_0}^{\tau_B\wedge\tau_{C_i}\wedge T_i} n^D(s)ds}\{l_s^c(t)[M_i(t)]^s + f_s^B(t)[V_i(t)-M_i(t)]]^s\}\mathbb{I}_{\{\tau_B\wedge\tau_{C_i}>t\}}dt\right] \tag{6.3}$$

with $s \in \{+,-\}$, to make the notation lighter. For simplicity, again we do not account for close-out terms in the FVA definition. By making the dependency on the default times explicit, we rewrite the above formula as:

$$\mathbb{E}_{t_0}^{\mathbb{Q}^D}\left[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n^D(s)ds}[l_s^c(t)[M_i(t)]^s + f_s^B(t)[V_i(t)-M_i(t)]]^s]\mathbb{I}[\tau_B\wedge\tau_{C_i}>t]dt\right]. \tag{6.4}$$

The FVA, as discussed above, can be split into three components, that we list here for convenience. Apart from the LVA, defined as $\mathbb{E}_{t_0}^{\mathbb{Q}^D}\left[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n^D(s)ds}[l^c(t)[M_i(t)]]\mathbb{I}[\tau_B\wedge\tau_{C_i}>t]dt\right]$ that has already been mentioned above, many authors split the funding component into a funding benefit adjustment or FBA, defined as $\mathbb{E}_{t_0}^{\mathbb{Q}^D}\left[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n^D(s)ds} f_-^B(t)[V_i(t)-M_i(t)]]^-]\mathbb{I}[\tau_B\wedge\tau_{C_i}>t]dt\right]$, and a funding cost adjustment or FCA, defined as $\mathbb{E}_{t_0}^{\mathbb{Q}^D}\left[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n^D(s)ds} f_+^B(t)[V_i(t)-M_i(t)]]^+]\mathbb{I}[\tau_B\wedge\tau_{C_i}>t]dt\right]$.

FVA as a portfolio metric. Because the FVA now includes absolute values of collateral and exposure levels, and because these are managed centrally by the treasury (i.e. they do not depend on a specific netting set), one can not sum the FVA terms defined so far (one could have done so for the CVA and DVA terms). Therefore one needs to define $T = \max\{T_i,\ i = 1, ..., N\}$ and the overall FVA as:

$$FVA_{\mathcal{B}}(t_0) = \mathbb{E}_{t_0}^{\mathbb{Q}^D}\left[\int_{t_0}^{T} e^{-\int_{t_0}^{t} n^D(s)ds}\{l_s^c(t)[\sum_{i=1}^{N}\mathbb{I}_i(t)(M_i(t))]^s + [f_s^B(t)\sum_{i=1}^{N}\mathbb{I}_i(t)(V_i(t)-M_i(t)]^s\}dt\right]. \tag{6.5}$$

where we have used $\mathbb{I}_i(t) = \mathbb{I}_{\{\tau_B\wedge\tau_{C_i}>t\}}$ to make notation lighter.

The invariance principle for funding. Elouerkhaoui [53] shows that the choice of the discount rate in the FVA definition is irrelevant. This is proven by showing that the price adjusted for FVA is a martingale under two different generic discounting rates. The same result is also found in Brigo, Pallavicini & Perini [27], where different measures and discount rates used in the derivations are simple computational tools and do not carry any specific financial meaning. We stress again that, while the

discount rate is irrelevant for the FVA, for CVA and DVA one should be discounting using Libor, given that these two adjustments are unsecured: as far as discounting is concerned we agree with Elouerkhaoui [53], who suggests to discount CVA and DVA cashflows using unsecured rates. [2]

The FVA debate. The definition and even the inclusion in the price of the FVA have been subject to extensive debate, that we summarise below.

Piterbarg [101] shows that the present value of a collateralised trade can be written as the expected value of the payoff, discounted at the collateral rate, plus a flow term that measures the funding benefit or cost of the difference between the cashflows and the collateral:

$$V(t) = \mathbb{E}_{t_0}^{\mathbb{Q}'}\left[e^{-\int_{t_0}^{t} n^c(s)ds}V(T) + \int_{t_0}^{T} e^{-\int_{t_0}^{t} n^c(s)ds}(n^B(s) - n^c(s))(V(t) - M(t))dt\right].$$

The integral term is *in nuce* the FBA + FCA term, in absence of credit risk. There is no LVA term mentioned in this paper, as the risk free rate here is the same as the cash collateral rate. We have converted the original notation of this paper to make it consistent with our analysis. This result is also useful to show that, in case of perfect collateralisation (i.e. $V(t) = M(t)$) and no collateral spread, the impact of collateral on the valuation of a derivative is entirely captured by the OIS-discounting, which means that there is no FVA term needed.

Hull & White ([75] and [76]) debate with Laughton & Vaisbrot [85] whether or not one should be considering the FVA at all in the price of a derivative. The main argument used by Hull & White is that no FVA is needed in a Black/Scholes set-up, given that "the risk-neutral valuation [...] gives the correct economic valuation for a derivative, taking into account all its market risks". Further, they recognise some overlap with the DVA, and exclude the FVA to prevent any double counting. However, the sole fact that systemic financial institutions like J.P. Morgan have recently started adding an FVA term to their balance sheet seems to go against the Hull & White approach. Laughton & Vaisbrot's arguments are based on the fact that markets are in practice incomplete, that bank funding is exogenous (i.e. "the bank borrows at the rate it can", and that "market-makers give no value to their expected profit or loss upon their own default").

Brigo, Pallavicini & Perini [27] build a general framework to calculate CVA and DVA. When adding funding costs, they show that it is difficult to introduce a purely additive FVA term, given that the equation has a recursive form. They take a different view from Hull & White, by finding that FVA and

[2] For completeness it should be added here that there is no universally accepted market practice regarding how to discount the CVA and the DVA. A counter argument could be that, in absence of bond-CDS basis, the price of the bond is given by the market, and that one could imply two different credit spreads if one is assuming Libor or OIS discounting. Therefore one could discount the CVA and the DVA using OIS instead. The OIS rate would be added to the credit spread in the risky discount factor: following this reasoning, discounting at Libor plus credit spread would clearly be redundant and would require the use of an *ad-hoc* spread. If one chooses this second option the rest of this work is not made inconsistent, one should simply use a different discount rates in the definitions.

DVA are not the same quantity and challenge Hull & White's statement that the FVA should not be considered at all. Their definition of the FVA term is, at least in our view, the most complete found in literature, including collateral costs/benefits, close-outs effects, and hedging. In order to derive their results, they model also the cashflows that happen within the institution, in particular between a trading desk and the treasury. A numerical application can be found in Brigo, Liu, Pallavicini & Sloth [20].

Burgard & Kjaer [30] propose a PDE-based framework that includes asymmetric funding costs, exogenous close-outs, hedging costs, the bank credit risk hedging (by trading the counterparty and its own bonds, assuming that they are liquid enough to do so): they calculate both counterparty and funding valuation adjustment consistently. The idea is that the credit risk of a derivatives transaction is hedged by buying a certain notional of bonds (both of the counterparty and the financial institution), that are assumed to be liquid enough. The main issue with their approach is computational, given that PDE numerical methods are ill-suited for very high dimensionalities. In a later paper [32], the authors explore some different replication strategies available to financial institutions to hedge the close-out upon their own default, that result in different levels of protection for the bondholders (i.e. some strategy ensures a higher protection given that the close-out is always positive) at the expenses of the shareholders (that, before the default happens, see smaller profits if that the close-out is overhedged).

Albanese & Iabichino [1] propose a bank-level FVA definition that ensures no overlap with the DVA. Firstly, they include in their FVA definition only the collateral costs/benefits, and secondly they stress the view that OTC books are an unstable source of funding, and that therefore the FVA should be discounted with the short OIS rate. Their proposal is to calculate an asymmetric FVA term by taking into account the sum of all initial margins and only the collateral funding costs (i.e. to exclude the benefits, which is to take into account only the scenarios where the bank posts collateral and therefore must finance it only receiving OIS). They do not include close-out in the FVA definition. As a result, the DVA is orthogonal to the FVA, which is a useful result, but at the cost that one has to ignore the bond-CDS basis. We may have a payable trade that gives some DVA, but its funding benefit must be calculated using bond spreads.

Elouerkhaoui [53] proposes a general framework that is probably closer to Brigo, Pallavicini & Perini [27], given that he models the relationship between the trading desk and the treasury, stresses that the FVA can be discounted using any rate, and that takes into account the close-outs. Because he uses riskless close-out and symmetric funding rates, he manages to solve the problem in two steps, i.e. to calculate the FVA for the credit-riskless PV, and then uses this PV to perform the CVA and DVA calculations under different close-out assumptions.

The FVA debate: our (temporary) definition. Given the complexity of the FVA that we outlined in this section, we find it useful to explicitly state the assumptions used to define the FVA in the rest

of this chapter. The differences between the more theoretically consistent approaches (Brigo, Pallavicini & Perini [27], Elouerkhaoui [53], Burgard & Kjaer [30]) and a theoretically less consistent but more practical approach (Albanese & Iabichino [1]) are symptomatic of the complexity of the task. It is also important to stress that Albanese & Iabichino are more interested to the computational aspects of the FVA, rather than to define it in a complete way. As stated above, the objective of this section is not to take any side in these debates but to show an application of the CTCB model to counterparty risk and funding, and therefore we keep the complexity down to the minimum acceptable level: we strike a balance between completeness and complexity similar to the one in the presentation given by Brigo et al. [17] (again, we refer in particular to pages 91-92). Funding spreads may be asymmetric, however the OIS rate paid on collateral may be subject to much smaller bid-ask spreads, which can be ignored.

The FVA is defined by using asymmetric funding rates, by assuming a bond-CDS basis, using OIS discounting and by taking into account positive and negative funding contributions. As discussed above, close-out are not included in the FVA, following Albanese & Iabichino [1]. In formulas this reads:

$$FVA_{\mathcal{B}}(t_0) = \mathbb{E}_{t_0}^{\mathbb{Q}^D}\left[\int_{t_0}^{T} e^{-\int_{t_0}^{t} n^D(s)ds}\{l_s^c(t)[\sum_{i=1}^{N}\mathbb{I}_i(t)(M_i(t))]^s + [f_s^B(t)\sum_{i=1}^{N}\mathbb{I}_i(t)(V_i(t)-M_i(t)]^s\}dt\right]. \tag{6.6}$$

As a consequence, the FVA is part of the general valuation adjustment but is to be considered as an internal transfer pricing tool, given that this has some overlap with the DVA.

If one wanted to define an FVA-type adjustment that can be added to the DVA without any overlap or double-counting, the most widely-used approach in the industry is the so-called marginal method. This method may not be fully accurate but in our view represents a good compromise between calculation simplicity and modelling power. We define a marginal FVA adjustment as:

$$MFVA_{\mathcal{B}}(t_0) = -FVA_{\mathcal{B}}(t_0) - \sum_{i=1}^{N} DVA_i^B(t_0). \tag{6.7}$$

We should remember that in our original definition the DVA is a positive quantity, and therefore subtracting it from the FVA means removing the overlap due to negative exposures. By using this definition, we define a bank-level valuation adjustment defined as:

$$VA_{\mathcal{B}}(t_0) = -\sum_{i=1}^{N} CVA_i^B(t_0) + \sum_{i=1}^{N} DVA_i^B(t_0) + MFVA_{\mathcal{B}}(t_0). \tag{6.8}$$

The main advantages of the marginal definition of the FVA are the following:

1. This definition is general and is not dependent on the collateralisation scheme adopted. Therefore it can be used to measure counterparty and funding effects across the whole book. For example,

a perfectly collateralised interest rates swap has no CVA nor DVA: the LVA term is driven by the collateral margining process, starting at 0 in t_0 (the premium is paid and posted as collateral, so no cashflow is exchanged at inception). An uncollateralised swap has CVA, DVA, and the FVA is driven by the bond-CDS basis, and the funding effects of the premium paid or received at inception. Cases with different collateralisation schemes fall in the middle of this spectrum. In the case where one sells an option without CSA, there is no DVA, there will be CVA, and the FVA is driven by the premium that one has received. Finally, in the case where the bank sells an option, there is no CVA, the DVA is driven by its default probability, and the FVA is driven by the CDS-bond basis.

2. This definition of marginal FVA copes with the potential use of different discount rates: the FVA can be discounted using any rate (as discussed above), while the DVA could be discounted using Libor (as discussed before, this idea is still matter of some debate). Using a marginal definition of the FVA ensures that any double counting is avoided.

3. This definition naturally takes into account the bond-CDS basis.

4. This definition naturally takes into account asymmetric funding rates.

5. This definition results in very simple computations that ensure that there is no overlap between DVA and FVA.

6. The bank-wide FVA term can be split into some counterparty contributions either by using Euler-type weights or other weights defined on some other quantity. This can be useful for trading desks to charge clients and for the treasury to charge desks.

7. The FVA methodology proposed here can be regarded as extremely conservative, which can be problematic in some cases. It implicitly assumes that the financial institution can not monetize its FVA exposure by selling protection on itself or, more realistically, on correlated names. In particular, for a funding intensive trade, the trade is charged the full funding cost which could be recovered (maybe partially) if the financial institution can sell protection on itself (or alternatively create a long position in a very correlated name: this would reduce the FVA mark-to-market volatility). This means that the positive exposure of an uncollateralised trade is charged twice, once to account for the counterparty default (CVA) and once more to account for the financial institution funding costs (FVA). We stress that this topic is still matter of debate in the industry, also in the light of capital requirements. For further discussion on this topic, see Cornalba [46].

Because the aim of this chapter is not to solve these theoretical problems, but to present an application of the CTCB model under credit risk, we do not regard these assumptions as essential, but as a way to

define the FVA quantity. If one prefers to use a different definition of the FVA, this is possible without invalidating the CTCB model extension that we propose in the next sections.

6.2 Some results with no wrong-way risk

Following Brigo & Mercurio [22] and Brigo, Morini & Pallavicini [26] we propose some closed form approximations for the unilateral CVA, DVA, and FVA of a standalone zero-coupon inflation swap under the assumption that there is no wrong-way risk: we assume that the price index process $\{I(t)\}_{t\geq 0}$ is independent from the default of the counterparty and from the funding rate of the financial institution. These approximations are based on time bucketing. The results for inflation derivatives are proposed for the first time here, to the best of our knowledge. These approximations are extremely helpful to develop intuition and test more complex models, even if they are not fully consistent with each other as they are based on different assumptions. The convergence of these approximations to the true levels have been studied in Sarais [105], where we show that approximately 20-30 time steps are enough to discretize the CVA for a 1 year FX forward.

We assume that the trade has maturity T, which we discretize in n subintervals $[t_{i-1}, t_i)$ with $i = 1, ..., n$: clearly $t_n = T$ and t_0 is the current time. The probability of default of the counterparty or the financial institution in the i-th time interval, given that the default has not happened in the past (up the $i-1$-th time interval), are defined as p_i^C and p_i^B respectively. Given that we are going to use market-implied probabilities of default, we assume that they are deterministic. In formulas, $p_i^C = \mathbb{Q}[\tau_C \in (t_{i-1}, t_i] | \tau_C > t_{i-1}]$ and $p_i^B = \mathbb{Q}[\tau_B \in (t_{i-1}, t_i] | \tau_B > t_{i-1}]$.

Proposition 1 *The unilateral standalone CVA calculated at time t_0 for an uncollateralised long zero-coupon inflation swap (i.e. the swap holder receives the price index performance) with maturity T and strike K is given by*

$$CVA^U(t_0) = (1 - R_{\mathcal{C}}) \sum_{i=1}^{n} p_i^C ZCO(Call, K, t_0, t_i)$$

where $ZCO(Call, K, t_0, t_i)$ is the value of a zero-coupon inflation option call, with strike K and maturity t_i. priced at time t_0. For the DVA, one substitutes p_i^C with p_i^B, $(1 - R_{\mathcal{C}})$ with $(1 - R_{\mathcal{B}})$, and calls with puts. The parameter n here is the number of time intervals used for the time bucketing.

To see this, we follow the same logic as Brigo & Mercurio [22]. One defines the unilateral standalone CVA as $\mathbb{E}_{t_0}^{\mathbb{Q}}[V^+(\tau_C)DF(t_0, \tau_C)](1 - R_{\mathcal{C}})$. By bucketing time and assuming no wrong-way risk, the CVA approximation becomes $\sum_{i=1}^{n} p_i^C \mathbb{E}_{t_0}^{\mathbb{Q}}[V^+(t_i)DF(t_0, t_i)(1 - R_{\mathcal{C}})]$. Specifically for the inflation zero-coupon swap, one finds immediately that $\mathbb{E}_{t_0}^{\mathbb{Q}}[V^+(t_i)DF(t_0, t_i)(1 - R_{\mathcal{C}})] = (1 - R_{\mathcal{C}})\mathbb{E}_{t_0}^{\mathbb{Q}}[[I(t_i)/I(t_0) - (1 + K)^{t_i - t_0}]^+ DF(t_0, t_i)]$. The last term is the PV of a zero-coupon call on the price index.

As a side comment we also note that the FCA and FBA terms can be approximated in a very similar way, obtaining

$$FCA^U(t_0) = -\sum_{i=1}^{n} f_-^B(t_i) ZCO(Call, K, t_0, t_i) S(t_i)$$

and

$$FBA^U(t_0) = \sum_{i=1}^{n} f_+^B(t_i) ZCO(Put, K, t_0, t_i) S(t_i)$$

where $S(t_i)$ is the joint survival probability at time t_i. In formulas, $S(t_i) = \mathbb{Q}[\tau_C \wedge \tau_B > t_i]$. Again we assume that survival probabilities are taken from the market, and therefore can be treated as deterministic in this context. In case of a short inflation swap, one substitutes calls with puts in the above formulae.

For the LVA approximation we make different assumptions. For the short-term forward funding spreads $l_+^c(t)$ and $l_-^c(t)$ one assumes that they are deterministic and therefore independent from inflation and discount rates. Further one assumes that both parties can not default. Finally, there are only collateral margin payments, and no other cashflows.

Proposition 2 *The unilateral standalone LVA calculated at time t_0 for a fully collateralised zero-coupon inflation swap with maturity T and strike K is approximately given by*

$$LVA(t_0) = \sum_{i=1}^{n} [l_+^c(t_i) ZCO(Call, K, t_0, t_i) - l_-^c(t_i) ZCO(Put, K, t_0, t_i)](t_{i+1} - t_i)$$

where $ZCO(Call, K, t_0, t_i)$ ($ZCO(Put, K, t_0, t_i)$) is the value of a zero-coupon inflation option call (put), with strike K and maturity t_i, priced at time t_0. The parameter n here is the number of time intervals used for the time discretisation.

Assuming that neither party is defaultable, the standalone LVA of a perfectly collateralised trade is defined as:

$$\mathbb{E}_{t_0}^{\mathbb{Q}^D}\left[\int_{t_0}^{T} e^{-\int_{t_0}^{t} n^D(s)ds} [l_+^c(t)[M(t)]^+ + l_-^c(t)[M(t)]^-] dt\right].$$

Perfect collateralisation lets one write $[M(t)]^+ = [V(t)]^+$, and the same for the negative part. Time bucketing yields $\mathbb{E}_{t_0}^{\mathbb{Q}^D}\left[\sum_{i=1}^{n} e^{-\int_{t_0}^{t_i} n^D(s)ds} \{[l_+^c(t_i)[V(t_i)]^+ + l_-^c(t_i)[V(t_i)]]^-\}(t_i - t_{i-1})\right]$. It is straightforward to write this expression as $\left[\sum_{i=1}^{n} \{[l_+^c(t_i)\mathbb{E}_{t_0}^{\mathbb{Q}^D}[DF(t_0, t_i)V(t_i)^+] + l_-^c(t_i)\mathbb{E}_{t_0}^{\mathbb{Q}^D}[DF(t_0, t_i)V(t_i)^-]]\}(t_i - t_{i-1})\right]$. The discount factor here is using the generic discount rate $n^D(t)$. The terms $\mathbb{E}_{t_0}^{\mathbb{Q}^D}[DF(t_0, t_i)V(t_i)]^+$ and $\mathbb{E}_{t_0}^{\mathbb{Q}^D}[DF(t_0, t_i)V(t_i)]^-$ represent the prices of zero-coupon inflation options priced at time t_0 (calls and puts respectively). In particular we note that $[V(t_i)]^- = \min[0, I(t_i)/I(t_0) - (1+K)^{t_i - t_0}] = -\max[0, -I(t_i)/I(t_0) + (1+K)^{t_i - t_0}]$: this this a short price index put position. i.e. a zero-coupon inflation floor, as shown in chapter 5.

We observe that if the collateral spread is symmetrical (i.e. $l_+^c(t_i) = l_-^c(t_i)$ $i = 1, ..., n$) and the price

of an at-the-money call is equal to the one of an at-the-money put (i.e. there is no inflation volatility skew and discount rates are zero), the LVA of a fully collateralised at-the-money zero-coupon inflations swap is zero, which matches the intuition that, on average, during the life of the trade the funding benefits will offset the funding costs. As above, in case of a short inflation swap, one substitutes calls with puts in the above formulae.

At this stage we recall that in chapter 3 some closed form results for the PV of zero-coupon (and year-on-year) options were found in the CTCB model: these formulae depend on some macroeconomic model parameters and they can be plugged in the above approximated formulae to provide the unilateral standalone CVA, DVA, LVA, and FVA.

6.3 Credit modelling overview

In this section we describe the main credit frameworks used to model counterparty risk, and explain the choices we make to model it in the next sections. This section is by no means a complete treatment of this topic. A complete review of these models can be found in chapter 3 of Brigo, Morini & Pallavicini [26].

6.3.1 Structural versus intensity-based credit models

Credit modelling for CVA and credit derivatives in general is carried out either using structural or intensity-based models.

Structural models follow the intuition of Merton [97] and Black & Cox [13] seminal papers: the default of a given name happens when a state variable $D(t)$ touches for the first time a negative value. The variable $D(t)$ is defined as the difference between the assets and the liabilities of the firm and is modelled as the difference of two exponential Brownian motions. The volatility of the asset and the liabilities are usually calibrated to the CDS-implied probability of default. Some recent extensions of this approach include the AT1P model, reviewed for example in [25]: here the value of the firm follows some diffusive dynamics, and the default happens as soon as $D(t)$ touches for the first time a barrier level $B(t)$; these barrier levels are calibrated again to CDS quotes. The default probability becomes the price of a barrier option, which we price in closed forms assuming constant volatility of the firm value. Another possible extension is the SBVT model, also presented in [25], essentially an uncertain parameters version of the AT1P model. The main criticism to the basic structural approach is that the short-term default probability is zero, given that their dynamics are driven by diffusive processes: these models can not explain inverted credit term structures for example. [3] Intensity-based models

[3] The SBVT model, presented in Brigo & Morini [25], can deal with this problems by using stochastic barriers.

(also known as "reduced form models") are a somewhat simpler model of the credit event: the default is modelled as the first jump of a Poisson process. Therefore, once we know the credit intensity function $h^X(t)$, where $X \in \{\mathcal{B}, \mathcal{C}_1, ..., \mathcal{C}_N\}$, we infer all default and survival probabilities. We further assume that the jumps after the first one do not have any further impact, given that we impose that a name can not default more than once: this approximation is implicit in Lindskog & McNeil [89] and is fully addressed in Brigo, Pallavicini & Torresetti [28].

These models are popular also thanks to the approximation that is derived under constant credit spreads, independency between interest rates and spreads, continuous premium leg, and deterministic recovery, whereby $h^X(t) \cong k^X(t)/(1 - R_X)$. Here $k^X(t)$ is the instantaneous CDS spread of the name X, that can be interpolated from market CDS quotes.

Our choice for this chapter is to use an intensity-based model. The reason is a purely computational one: because CVA, DVA, and FVA are expectations, and because one writes in closed forms the relevant probabilities, the valuation adjustments calculations can be performed without simulating the defaults, at least for non-credit derivatives. This saves significant amounts of time and computation power. Let us recall for example the original definition of unilateral CVA:

$$CVA_i^U(t_0) = (1 - R_{C_i})\mathbb{E}_{t_0}^{\mathbb{Q}}[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n(s)ds} E_i^+(t)\mathbb{I}[\tau_{C_i} \in [t, t + dt)]].$$

By applying the tower law one writes for $t \geq t_0$:

$$\mathbb{E}_{t_0}^{\mathbb{Q}}[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n(s)ds} E_i^+(t)\mathbb{I}[\tau_{C_i} \in [t, t + dt)]] = \mathbb{E}_{t_0}^{\mathbb{Q}}[\mathbb{E}_t^{\mathbb{Q}}[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n(s)ds} E_i^+(t)\mathbb{I}[\tau_{C_i} \in [t, t + dt)]]].$$

By applying the expectation properties, this is equivalent to:

$$\mathbb{E}_{t_0}^{\mathbb{Q}}[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n(s)ds} E_i^+(t)\mathbb{E}_t^{\mathbb{Q}}[\mathbb{I}[\tau_{C_i} \in [t, t + dt)]]] = \mathbb{E}_{t_0}^{\mathbb{Q}}[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n(s)ds} E_i^+(t)\mathbb{Q}[\tau_{C_i} \in [t, t + dt)]].$$

The first jump probability $\mathbb{Q}[\tau_{C_i} \in [t, t+dt)]$ under a Poisson process is written as $h^{C_i}(t)e^{-\int_{t_0}^{t} h^{C_i}(s)ds}dt$, which is an exponential default time. This result can also be obtained in a Marshall-Olkin model, as we will see in the next sections. Therefore we obtain:

$$CVA_i^U(t_0) = (1 - R_{C_i})\mathbb{E}_{t_0}^{\mathbb{Q}}[\int_{t_0}^{T_i} e^{-\int_{t_0}^{t} n(s)ds} E_i^+(t)h^{C_i}(t)e^{-\int_{t_0}^{t} h^{C_i}(s)ds}dt]. \tag{6.9}$$

From the above expression it is clear that we calculate the CVA without having to simulate the default times, but by integration after simulating interest rates, the payoff, and the intensities. The above result is generalised to the DVA and FVA, and to both unilateral and bilateral counterparty valuation

adjustments. It should be stressed that this statement would not hold for credit loss percentiles.

6.3.2 Idiosyncratic versus systematic credit risk

Structural models have the advantage to capture the economic forces behind the assets and liabilities of the firms, and therefore may seem more appropriate to model credit risk in the CTCB framework, which is itself a structural model for inflation. However, given the two main disadvantages of structural credit modelling reviewed above (i.e. they do not allow strictly positive short-term credit spreads and they need to simulate default events for the XVA computation), we choose to use a slightly modified version of the credit intensity model, modified to take into account an idiosyncratic and a systematic component.

Given a name $X \in \{\mathcal{B}, \mathcal{C}_1, ..., \mathcal{C}_N\}$, defined by an intensity term structure $h^X(t)$, by a recovery rate R_X, ones defines $h^X(t) = h_I^X(t) + h_S^X(t)$. Here the positive independent processes $\{h_I^X(t)\}_{t\geq 0}$ and $\{h_S^X(t)\}_{t\geq 0}$ represent the intensities of two independent Poisson processes, respectively the idiosyncratic default component $JTD_I^X(t)$ and the systematic default component $JTD_S^X(t)$: the default event $JTD^X(t) = JTD_I^X(t) + JTD_S^X(t)$ is the first jump of the process defined as the sum of these two processes (because the two processes are independent, the intensity of the sum of the processes is the sum of their intensities). Further, we add some randomness to the intensities: therefore $\{h_I^X(t)\}_{t\geq 0}$ and $\{h_S^X(t)\}_{t\geq 0}$ have to be modelled as positive independent processes, under the constraint that $\mathbb{E}_{t_0}^{\mathbb{Q}}[h^X(t)] = \mathbb{E}_{t_0}^{\mathbb{Q}}[h_I^X(t)] + \mathbb{E}_{t_0}^{\mathbb{Q}}[h_S^X(t)]$, $\forall t \geqslant t_0$ and $\forall X \in \{\mathcal{B}, \mathcal{C}_1, ..., \mathcal{C}_N\}$. All idiosyncratic processes $h_I^X(t)$ are independent from each other and from any other state variable in the model. As we will show in the following sections, all idiosyncratic processes $h_S^X(t)$ are dependent on the economy, which can be seen as a common factor.

6.3.3 Marshall-Olkin models

Brigo, Morini & Pallavicini [26] note that merely correlating the Brownian shocks of the (systematic) default intensities may not yield enough dependence across the default events (we have independently verified this claim when working on a separate project): therefore the most advanced counterparty risk models also include some credit copulas. Credit copulas have a direct impact on bilateral CVA/DVA (given the dependency between the default of the counterparty and the financial institution), on credit derivatives CVA/DVA (the CVA/DVA of a collateralised CDS, when collateralisation is almost perfect (i.e. $M(t) = V(t^-)$), depends crucially by how correlated is the default of the counterparty and the reference entity), and on FVA (the joint survival event is integrated across time in the adjustment computation). A general review of copulas can be found in Embrechts, Lindskog & McNeil [54].

Given that we use intensity-based models and that Gaussian copulas have shown not to generate

enough default dependency (especially for a high number of credit names), we model the credit dependence via a Marshall-Olkin model, that implies a Marshall-Olkin survival copula under an appropriate parametrisation: further, the Marshall-Olkin copula can generate simultaneous defaults, and therefore is a good option to model credit contagion. Li & Pellerey [88] provide a complete description of the Marshall-Olkin copula.

A general literature review of this subject includes the following papers. Marshall & Olkin [92] introduce a survival distribution by modelling the fatal shocks as independent Poisson processes, including the joint shocks. The Marshall-Olkin survival copula is defined as the copula generated by the Marshall-Olkin joint survival distribution. Lindskog & McNeil [89] propose a Common Poisson Shock (CPS) model based on m shocks — modeled as independent Poisson processes — that are linked to the n loss processes by some Bernoulli events. They show that this model implies fatal shocks that are independent Poisson processes: this implies that the survival copula implied by this model is a Marshall-Olkin copula. The dependency structure between the n loss processes is a function of the dependency structure of the Bernoulli events. Brigo, Pallavicini & Torresetti [28] build the GPCL model and calibrate it to the market prices of CDOs. The GPCL model is essentially a CPS model adjusted and reparametrised to prevent names to default more than once (this is an undesirable feature of the CPS model). Brigo & Chourdakis [19] define the self-chaining copulas as the copulas that have a lack of memory property, i.e. that can be iterated N times over N subsequent time intervals of size T/N and give the same joint survival probability as a single iteration of the copula over one interval of size T. This can be regarded as a sort of self-similarity. They show that the the Marshall-Olkin and the Gumbel copula satisfy this property. Brigo, Mai & Scherer [21] show that the Marshall-Olkin multivariate exponential distribution can be characterized in terms of Markovianity of survival indicators. They also review the most important properties of this multivariate exponential distribution, namely that it satisfies the lack-of-memory property, and that it is stable under marginalization, i.e. its lower dimensional margin satisfy the lack-of-memory property as well. Further the Marshall-Olkin copula allows two or more names to default at the same time. Mai & Scherer [91] show that a conditionally independent and identically distributed model (CIID), with exponential variables and Lévy subordinator implies a survival probability function that can be reconciled with the Marshall-Olkin survival function (the Laplace exponent of the Lévy subordinator appears in the survival function). This result is helpful to reduce the number of model parameters to use the model in practical applications. They try and assess the properties of different subordinators. The Lévy subordinator can be interpreted as a common factor: the lack-of-memory of the Marshall-Olkin distribution is the result of the lack-of-memory of the increments of the subordinator (being a Lévy process, it has independent and stationary increments). It should be stressed that the simplest model proposed (a one factor model) can not include all possible Marshall-Olkin laws, but is

still general enough to be helpful in applications.

We propose a brief review of the Marshall-Olkin model to adapt it to our purposes. If one is modelling the default of N names plus the financial institution (so overall $N+1$ possible defaults), he has N sets of positive multiple-default parameters, each modelling the joint default of j names, where $j \in \{2, ..., N+1\}$. Each parameter set has $C_{N+1,j} = \begin{pmatrix} N+1 \\ j \end{pmatrix} = \frac{(N+1)!}{(N+1-j)!(j)!}$ terms, each of which is denoted as $h^{(j)}_{\{A(k,N+1,j)\}}(t)$, where $A(k, N+1, j)$ is the k-th combination of j names defaulting out of $N+1$ ($k = 1, ..., C_{N+1,j}$). If we include also the single name parameters, we work with $2^{N+1} - 1$ intensities.

To make the notation lighter, let us assume we want to model the default of the financial institution $\mathcal{B}$ and N counterparties $\mathcal{C}_1, ..., \mathcal{C}_N$. There are $N+1$ single names intensities, and we define $\Phi^{(1)} = \{\mathcal{B}, \mathcal{C}_1, ..., \mathcal{C}_N\}$. For two names defaulting together we have $C_{N+1,2}$ terms in the set $\Phi^{(2)}$, and so on up to the single event when all $N+1$ names default together: clearly $C_{N+1,N+1} = 1$. In general, $\Phi^{(i)} = \bigcup_{k=1}^{C_{N+1,i}} A(k, N+1, i)$: for example, if $N = 2$, $\Phi^{(1)} = \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{B}\}$, $\Phi^{(2)} = \{(\mathcal{C}_1, \mathcal{C}_2), (\mathcal{C}_1, \mathcal{B}), (\mathcal{C}_2, \mathcal{B})\}$, and $\Phi^{(3)} = \{(\mathcal{C}_1, \mathcal{C}_2, \mathcal{B})\}$. Finally the full set of intensities, including single names and all possible combinations, is defined as $\Phi = \bigcup_{i=1}^{N+1} \Phi^{(i)}$. To calculate the cardinality $\mathfrak{C}$ of this set, one has $\mathfrak{C}(\Phi) = \sum_{i=1}^{N+1} \mathfrak{C}(\Phi^{(i)}) = \sum_{i=1}^{N+1} C_{N+1,i}$.

To see this machinery at work, a simple example may help, following Perini [100]. Let us assume $N = 1$: there exist only one counterparty and the financial institution, so we are modelling 2 defaults in total. If there is no dependence the defaults of only name 1 and only name 2 (name 2 is the financial institution here) are modelled as the first jump of the Poisson processes $\{JTD^1(t)\}_{t\geq 0}$ and $\{JTD^2(t)\}_{t\geq 0}$ respectively, with intensities $h^1(t)$ and $h^2(t)$. Given their independence, their joint survival probability at time t is $S(t,t) = e^{-\int_{t_0}^{t}[h^1(s)+h^2(s)]ds}$.

To introduce some dependency, let us introduce the third Poisson process $\{JTD^{1,2}(t)\}_{t\geq 0}$, with intensity $h^{1,2}(t)$ and independent from $\{JTD^1(t)\}_{t\geq 0}$ and $\{JTD^2(t)\}_{t\geq 0}$: the first jump of the process $\{JTD^{1,2}(t)\}_{t\geq 0}$ represents the event of both names defaulting simultaneously. Let us redefine the default event of names 1 and 2 as the first jump of the newly defined processes $\{JTD^1_*(t)\}_{t\geq 0} = \{JTD^1(t)\}_{t\geq 0} + \{JTD^{1,2}(t)\}_{t\geq 0}$ and $\{JTD^2_*(t)\}_{t\geq 0} = \{JTD^2(t)\}_{t\geq 0} + \{JTD^{1,2}(t)\}_{t\geq 0}$ respectively.
Their joint survival probability at times t_1 and t_2 is $S(t_1, t_2) = e^{-\int_{t_0}^{t_1} h^1(s)ds - \int_{t_0}^{t_2} h^2(s)ds - \int_{t_0}^{\max\{t_1,t_2\}} h^{1,2}(s)ds}$.
If one defines $h^1_*(t) = h^1(t) + h^{1,2}(t)$, $h^2_*(t) = h^2(t) + h^{1,2}(t)$, and $h^{1,2}_*(t) = -h^{1,2}(t)$, we rewrite the above survival probability as

$$S(t_1, t_2) = e^{-\int_{t_0}^{t_1} h^1_*(s)ds - \int_{t_0}^{t_2} h^2_*(s)ds - \int_{t_0}^{\min\{t_1,t_2\}} h^{1,2}_*(s)ds}$$

which is a more expressive way to present it. In particular, one notices the natural constraint that in each time one has to have positive intensities $h^1(t)$, $h^2(t)$, and $h^{1,2}(t)$: this implies the constraint $0 \leq h^{1,2}(t) \leq \min\{h^1_*(t), h^2_*(t)\}$. There is no dependency when $h^{1,2}(t) = 0$: maximal dependency is achieved when $h^{1,2}(t) = \min\{h^1_*(t), h^2_*(t)\}$, that means that both names survive if and only if the riskier survives. This is equivalent to implying that either $h^1(t)$ or $h^2(t)$ is zero, i.e. the independent default intensity for the safer name is switched off, and it defaults only if the riskier defaults at the same time.

One finally notes that the survival probability above implies the Marshall-Olkin copula $C(u,v) = uv \min\{u^{-a_1(t)}, u^{-a_2(t)}\}$, where the parameters are defined as follows: $u = e^{-\int_{t_0}^{t_1} h^1_*(s)ds}$, $v = e^{-\int_{t_0}^{t_2} h^2_*(s)ds}$, $a_1(t) = h^{1,2}(t)/h^1_*(t)$, and $a_2(t) = h^{1,2}(t)/h^2_*(t)$. Clearly, if one imposes $h^1_*(t)$, $h^2_*(t)$ and $h^{1,2}(t)$ one is not free to specify any arbitrary level for $h^1(t)$ and $h^2(t)$, given that these are intensities and have to be positive. Therefore, a Marshall-Olkin copula with exponential marginals does not imply a Marshall-Olkin multivariate survival function, which in turn allows to simulate simultaneous defaults.

The extension to higher dimensionality is straightforward, with the *caveat* that the number of joint-default processes is much higher, to count all possible default combinations. For example, for $N = 2$ and counting the financial institution as the third name, one has to consider the processes $h^{1,2}(t)$, $h^{1,3}(t)$, $h^{2,3}(t)$, and $h^{1,2,3}(t)$ under the constraints $h^1_*(t) = h^1(t) + h^{1,2}(t) + h^{1,3}(t) + h^{1,2,3}(t) \geq 0$, $h^2_*(t) = h^2(t) + h^{1,2}(t) + h^{2,3}(t) + h^{1,2,3}(t) \geq 0$, and $h^3_*(t) = h^3(t) + h^{1,3}(t) + h^{2,3}(t) + h^{1,2,3}(t) \geq 0$.

To be consistent with the modelling of idiosyncratic and systematic single-name default intensity proposed above, we apply the same split used for single-name default intensities to the joint default intensity processes. For the k-th default combination of j names $A(k,N,j)$, we define $h^{(j)}_{\{A(k,N,j)\}}(t) = h^{(j)}_{S,\{A(k,N,j)\}}(t) + h^{(j)}_{I,\{A(k,N,j)\}}(t)$.

Finally, in case of N names plus the financial institution, the total default intensities are defined as $h^i_*(t) = h^i(t) + \sum_{j=2}^{N+1} \sum_{k=1}^{C_{N+1,j}} h^{(j)}_{A(k,N+1,j)}(t)\mathbb{I}\{i \in A(k,N+1,j)\}$ and split into their idiosyncratic and systematic components as $h^i_{I,*}(t) = h^i_I(t) + \sum_{j=2}^{N+1} \sum_{k=1}^{C_{N+1,j}} h^{(j)}_{I,\{A(k,N+1,j)\}}(t)\mathbb{I}\{i \in A(k,N+1,j)\}$ and $h^i_{S,*}(t) = h^i_S(t) + \sum_{j=2}^{N+1} \sum_{k=1}^{C_{N+1,j}} h^{(j)}_{S,\{A(k,N+1,j)\}}(t)\mathbb{I}\{i \in A(k,N+1,j)\}$. For clarity of notation, $\mathbb{I}\{i \in A(k,N+1,j)\}$ equals one only when the i-th name belongs to the combination *A(k,N+1,j)*, otherwise is zero.

Before moving to other topics we stress that the model that we have introduced is a generalisation of the Marshall-Olkin model. As in a Marshall-Olkin model we split the default intensities of the names in components that model the joint and standalone default: for example, for two names we will have $h^1_*(t) = h^{(1)}_1 + h^{(2)}_{1,2}$ and $h^2_*(t) = h^{(1)}_2 + h^{(2)}_{1,2}$. Our generalisation consists of further splitting all these terms into a systematic and an idiosyncratic component, writing $h^1_*(t) = h^{(1)}_{I,1} + h^{(1)}_{S,1} + h^{(2)}_{I,\{1,2\}} + h^{(2)}_{S,\{1,2\}}$ and $h^2_*(t) = h^{(1)}_{I,2} + h^{(1)}_{S,2} + h^{(2)}_{I,\{1,2\}} + h^{(2)}_{S,\{1,2\}}$. We assume that the idiosyncratic terms $h^{(1)}_{I,1}$,$h^{(1)}_{I,2}$, and $h^{(2)}_{I,\{1,2\}}$ are all independent from each other, in the spirit of the original Marshall-Olkin model, while the systematic

terms $h^{(1)}_{I,1}$,$h^{(1)}_{I,2}$, and $h^{(2)}_{I,\{1,2\}}$ all depend somehow on the macro-economy. This framework allows more flexibility to model different effects: the terms $h^{(1)}_{I,1}$ and $h^{(1)}_{I,2}$ are the classic idiosyncratic single-name defaults, modelling situations that are name-specific (fraud, ...). The terms $h^{(1)}_{S,1}$ and $h^{(1)}_{S,2}$ are used to model standalone defaults that are somehow related to the macro-economy, which may affect only some industries. The term $h^{(2)}_{I,\{1,2\}}$ models default clusters that are independent from the macro-economy, as the ones caused by industry-related issues that do not spill over to the macro-economy. The term $h^{(2)}_{S,\{1,2\}}$ models default clusters that are dependent on the macro-economy. If one sets the terms $h^{(1)}_{S,1}$, $h^{(1)}_{S,2}$ and $h^{(1)}_{I,\{1,2\}}$ to be always zero, one recovers a model that is in the spirit of the original Marshall-Olkin model, where the idiosyncratic effects only affect one name at a time, while the systematic term affects all names at the same time. In this case we would have $h^1_*(t) = h^{(1)}_{I,1} + h^{(2)}_{S,\{1,2\}}$ and $h^2_*(t) = h^{(1)}_{I,2} + h^{(2)}_{S,\{1,2\}}$.

6.3.4 CDS-bond basis

We model the CDS-bond basis as a deterministic curve. The short-term yield of a bond issued by company X, $X \in \{\mathcal{B}, \mathcal{C}_1, ..., \mathcal{C}_N\}$, at time t is defined $n^X(t)$. This was introduced before as the funding rate for the financial institution $\mathcal{B}$: we define it as a deterministic function ϕ of the default intensity $h^X(t)$ multiplied by $1 - R_X$ plus a deterministic spread $l^X(t)$: $n^X(t) = (1 - R_X)\phi(h^X(t)) + l^X(t)$. The function ϕ represents the bootstrapping of the default intensities from the market CDS quotes: in case of constant spreads, independency between risk-free interest rates and spreads, and deterministic recovery, we would have $\phi(h^X(t)) = h^X(t)$. The CDS-bond basis can be either positive or negative, with the constraint that the bond yield $n^X(t)$ is positive if $h^X(t)$ is positive. In practice, this basis is driven by supply and demand of bond and CDS protection on a given name. As noted above, the market for funding can be asymmetric, and a financial institution can borrow at a different rate $n^X_-(t)$ compared to the rate it can lend money at $(n^X_+(t))$: we defined $f^X_+(t) = n^X_+(t) - n'(t)$, and therefore $f^X_+(t) = (1 - R_X)\phi(h^X(t)) + l^X_+(t) - n'(t)$. The same can be done for $f^X_-(t) = (1 - R_X)\phi(h^X(t)) + l^X_-(t) - n'(t)$: in this final case we are assuming that the funding spread asymmetry is driven by an asymmetry in the CDS-bond spreads.

6.4 Multicurve modelling overview

We model interest rates bases by assuming that only the OIS short rate $n'(t)$ is stochastic, and all other rates curves are defined using deterministic spread curves above it. In particular we are considering a single currency economy, where interest rates swaps are quoted on a main reference tenor $\varkappa$, and where the other tenors are $\kappa_1, ..., \kappa_M$. For the avoidance of doubt, in this context the tenor is the frequency of the interest rates swap payments. Therefore these two types of bases are defined, all in terms of short rates:

1. Libor-OIS basis $\beta^{L-OIS}(t)$: this basis, added to the instantaneous short OIS rate, gives an instantaneous unsecured rate $n(t) = n'(t) + \beta^{L-OIS}(t)$ such that the bond prices that one obtains from it, once plugged into the collateralised (i.e. OIS-discounted) swap price formula

$$S^{\varkappa}(t_0, T_\alpha, T_\beta) = \frac{\sum_{i=\alpha+1}^{i=\beta} L^{\varkappa}(t_0, T_{i-1}, T_i) P'(t_0, T_i)}{\sum_{i=\alpha+1}^{i=\beta} P'(t_0, T_i)}$$

returns the market quoted forward swap rates $S^{\varkappa}(t_0, T_\alpha, T_\beta)$ paying at dates $T_{\alpha+1}, ..., T_\beta$ the Libor resetting at dates $T_\alpha, ..., T_{\beta-1}$ against fixed rate $S^{\varkappa}(t_0, T_\alpha, T_\beta)$ on the main tenor $\varkappa = T_{i+1} - T_i$. Above we defined the forward Libor rates observed at time t_0, between times T_i and T_{i+1}, in the tenor $\varkappa$ as $L^{\varkappa}(t_0, T_i, T_{i+1}) = (P(t_0, T_i)/P(t_0, T_{i+1}) - 1)/(T_{i+1} - T_i)$, $P'(t_0, T) = \mathbb{E}^{\mathbb{Q}}_{t_0}[e^{-\int_{t_0}^{T} n'(s)ds}]$, and $P(t_0, T) = \mathbb{E}^{\mathbb{Q}}_{t_0}[e^{-\int_{t_0}^{T} n(s)ds}]$. The bond $P'(t_0, T)$ is clearly not a traded asset.

2. Libor-Libor tenor bases $\beta^{\varkappa-\kappa_i}(t)$, with $i \in \{1, ..., M\}$ are defined in the same way as above. If one defines the κ_i-tenor short rate $n_i(t) = n(t) + \beta^{\varkappa-\kappa_i}(t) = n'(t) + \beta^{L-OIS}(t) + \beta^{\varkappa-\kappa_i}(t)$, and plugs this in the swap pricing formula, one recovers the market quotes for forward swap rates $S^{\varkappa_i}(t_0, T_\alpha, T_\beta)$ on the tenor $\varkappa_i$.

A more complete analysis of this topic can be found for example in Pallavicini & Tarenghi [99], Kenyon & Stamm [82], and Henrard [66].

6.5 Credit modelling in the macroeconomic framework in a multicurve setting: the CR-MC-CTCB model

In this section we extend the model presented in chapter 3 to include a more realistic description of the economy. The idea is to include counterparty risk, funding risk, and a multicurve framework. This is done by including the extensions proposed in the previous sections in a fully consistent way. The benefits are manifold:

1. The model is automatically extended to cover XVA aspects of derivatives transactions without having to adjust the risk-free mark-to-market of the trades using a separate model. This happens when the dynamics used in the XVA model are different from the dynamics used in the counterparty-risk-free model (this is the case in most of the main financial institutions, and the impact may be significant).

2. Macroeconomic dynamics-based modelling helps explain the impact of the economic cycle on defaults. Further, one can test or introduce the assumption that a tighter monetary policy leads to a higher default rate throughout the economy.

3. The extension of this framework to other asset classes is trivial, and therefore the CTCB model can become the general XVA framework of the bank.

4. The extension we propose does not bring about significant computational burden, which is a major advantage in counterparty risk, where computational issues are very common given the high number of simulations required.

5. Because the original macroeconomic dynamics are defined in the physical measure $\mathbb{P}$, and then any pricing is done by calibrating the market prices of risk (as we see in chapter 3 and 4), we use realised model volatilities and zero market prices of risk in order to run simulations in the physical measure. This may be used to calculate various risk metrics such as PFE, VaR, Replacement Risk analysis, and Stress tests. This is a major advantage of the model proposed, given the increasing number of regulatory requests. In practice we use the same model both for pricing-hedging and for risk-related purposes, only changing its parameter set.

6.5.1 Assumptions

To achieve these goals, we make the following assumptions on the economy:

1. All assumptions made in chapter 3 (section 3.1) hold with the exception of 3.1.3.1: we now assume that there is credit risk in the economy. All randomness is driven by a K-dimensional Brownian motion $\{W(t)\}_{t\geq 0}$ and by the default times $\tau_{\mathcal{B}}, \tau_{\mathcal{C}_1}, ..., \tau_{\mathcal{C}_N}$. [4] In particular, one uses the first $D \leq K$ components to model randomness of the economic variables $X(t)$, $I(t)$, $M(t)$, $m_X(t)$, $m_I(t)$,[5] and the systematic default intensities $h_S^X(t)$, with $X \in \Phi$ [6]: this was the n dimensionality used in chapters 3 and 4, while in this chapter N is the number of counterparties. The remaining last $K-D$ components are only used to model the dynamics of the idiosyncratic default intensities $h_I^X(t)$, with $X \in \Phi$ (therefore these processes are independent from both the economic variables and from the systematic default intensities). The parametrisation is such that the idiosyncratic terms are also independent from each other.

2. In terms of filtrations, we refer to the comments made in section 6.1. In particular the intensity processes $\{h_I^X(t)\}_{t\geq 0}$ and $\{h_S^X(t)\}_{t\geq 0}$, with $X \in \Phi$ generate information in $\{\mathcal{F}_t\}_{t\geq 0}$. The default times $\tau_{\mathcal{B}}, \tau_{\mathcal{C}_1}, ..., \tau_{\mathcal{C}_N}$ generate information in $\{\mathcal{H}_t\}_{t\geq 0}$.

[4] In practice, given what we discussed in 6.3.1, we will not need to simulate the default times to calculate the XVAs.

[5] The notation here is the same as of chapter 3: $X(t)$ represents the real GDP, $I(t)$ the price index, $M(t)$ the money supply, $m_X(t)$ the GDP relative growth expectations, and $m_I(t)$ the price index relative growth (inflation) expectation.

[6] It is clear from the context that X is a name index in the set of the default combinations, while $X(t)$ is the real GDP process.

3. All assumptions made in chapter 3 regarding the short rate $n(t)$ are now made for the short OIS rate $n'(t)$. The risk-free rate is proxied with the OIS rate: this approximation has been discussed earlier in this chapter. The practical consequence of this statement is that the CTCB model gives the explicit Hull-White dynamics for the short OIS rate $n'(t)$.

4. We assume OIS discounting. Therefore any reference made in chapter 3 to the zero-coupon bonds $P(t,T)$ are made for the bonds $P'(t,T)$ defined above. Given the deterministic way we defined the Libor-OIS bases, these two bonds only differ for a deterministic positive multiplicative factor. This bond may not exist in practice: here it is used as a theoretical device.

5. As a consequence, whenever in chapter 3 the pricing measure was the risk-neutral $\mathbb{Q}$, using as numeraire the bank account $B(t)$, we now assume that that role is played by the OIS risk-neutral measure $\mathbb{Q}'$ using as numeraire the bank account $B'(t) = B(t_0)e^{\int_{t_0}^{t} n'(s)ds}$, even if such account may not exist in reality (we are assuming that all posted collateral is cash). Similarly, the T-forward measure $\mathbb{Q}^{T'}$ is now defined using the bond $P'(t,T)$ as numeraire.

6. The bond-CDS basis $l(t)$ is deterministic and symmetric, as specified in section 6.3.4.

7. We work with an interest rates multicurve setting, which is now an assumption given the above considerations: this is modelled via the assumptions made in section 6.4. In particular we assume that all bases on the OIS rate $n'(t)$ are deterministic.

8. Credit risk is modelled via the assumptions made in section 6.3: we have $N+1$ credit names in the economy including the financial institution $\mathcal{B}$. This introduces a constraint on the dimensionality of the driving Brownian motion $\{W(t)\}_{t\geq 0}$: $(N+1)+Y = K-D$, where Y represents the number of joint processes of the Marshall-Olkin model. In particular we have:

   (a) For each $X \in \Phi$, a mean-reverting process for the logarithm of the systematic hazard rate $h_S^X(t)$ is driven by the SDE $d\log h_S^X(t) = -m_S^X(t)(\log h_S^X - l_S^X(t))dt + v_S^X(t)\cdot dW^{\mathbb{Q}'}(t)$. Here $\{m_S^X(t)\}_{t\geq 0}$ and $\{l_S^X(t)\}_{t\geq 0}$ are deterministic scalar processes.[7] The process $\{m_S^X(t)\}_{t\geq 0}$ is positive to ensure mean reversion. The process $\{v_S^X\}_{t\geq 0}$ is a deterministic K-dimensional

[7] This model is also known as the "Exponential Vasicek" model with time varying coefficients and is fully reviewed in chapter 3 of Brigo & Mercurio [22]. The main advantage of this model is that mean reversion and strict positivity of the default intensities are guaranteed. However, as it happens for the Dothan model the quantity $\mathbb{E}[e^{\int_0^{\Delta t} h_S^X(u)du}] = \infty$, for an arbitrarily small Δt. This quantity can be regarded as a zero-risk free rate risky bank account. This problem, albeit of theoretical interest, does not have serious practical consequences given that this model is normally implemented in discrete time grids or, in this case, in a Monte Carlo simulation. A Hull-White model would imply negative intensities that have to be manually floored to zero in a simulation, which adds computational burden without clear advantage. In alternative, a CIR model could have been used, however our preference was to use a log-normal model given that the simulation of normal random variables is computationally lighter than Chi-square variables: in fact, one can choose not to simulate the intensity dynamics but, given that their distribution is known in advance, to simulate the intensity levels drawing them from their distribution (this would clearly work only for non-path dependent derivatives where the XVA calculation can be carried out via forward Monte Carlo methods, given that the path simulation is essential otherwise). The important point is that a different choice of the intensity dynamics does not make the XVA model proposed here invalid.

process, with zero level in the last $K-D$ components (that are used to model the volatility of the idiosyncratic parts). The positivity of this intensity is ensured by the use of a Gaussian process for its logarithm.

(b) For each $X \in \Phi$, a mean-reverting process for the logarithm of the idiosyncratic hazard rate $h_I^X(t)$ is driven by the SDE $d\log h_I^X(t) = -m_I^X(t)(\log h_I^X - l_I^X(t))dt + v_I^X(t) \cdot dW^{\mathbb{Q}'}(t)$. Here $\{m_I^X(t)\}_{t\geq 0}$ and $\{l_I^X(t)\}_{t\geq 0}$ are deterministic scalar processes. The process $\{m_I^X(t)\}_{t\geq 0}$ is positive to ensure mean reversion. The process $\{v_I^X\}_{t\geq 0}$ is a deterministic K-dimensional process, and is always zero for its first D components, and then has only one non-zero components, such that all the idiosyncratic hazard rate process are independent from each other: this ensures that this default probability component is idiosyncratic. Again, the positivity of this intensity is ensured by the use of a Gaussian process for its logarithm.

(c) The dynamic model parameters of the processes $\{h_I^X(t)\}_{t\geq 0}$ and $\{h_S^X(t)\}_{t\geq 0}$, $X \in \Phi$, are chosen in the calibration to ensure that the following properties hold:

i. The calibration to the market curve, i.e. $h_*^X(t) = h_{I,*}^X(t) + h_{S,*}^X(t)$, $X \in \Phi^{(1)}$, is such that $(1-R_X)\phi(h_*^X(t))$ reproduces the CDS market-implied default survival probabilities. In practice, for each $X \in \Phi^{(1)}$, $\mathbb{E}_{t_0}^{\mathbb{Q}'}[H_*^X(t)] = \mathbb{E}_{t_0}^{\mathbb{Q}'}[H_{I,*}^X(t)]\mathbb{E}_{t_0}^{\mathbb{Q}'}[H_{S,*}^X(t)]$, where we define $H_{Y,*}^X(t) = e^{-\int_{t_0}^t h_{Y,*}^X(s)ds}$ (with $Y \in \{I,S\}$). We remind that the idiosyncratic intensities are independent from the systematic intensities.

ii. The choice of the parameters, for each $X \in \Phi^{(1)}$ and given the collateralised CDS market implied instantaneous probability of default $\mathbb{E}_{t_0}^{\mathbb{Q}'}[H_*^X(t)]$, is done under the obvious constraints

$$\mathbb{E}_{t_0}^{\mathbb{Q}'}[H_*^X(t)] = \mathbb{E}_{t_0}^{\mathbb{Q}'}[H^X(t)\prod_{j=2}^{N+1}\prod_{k=1}^{C_{N+1,j}} H_{A(k,N+1,j)}^{(j)}(t)\mathbb{I}\{X \in A(k,N+1,j)\}] \geq 0.$$

Each term in the products above is a positive term. By doing the substitutions one

obtains:

$$\mathbb{E}_{t_0}^{\mathbb{Q}'}[H_*^X(t)] = \mathbb{E}_{t_0}^{\mathbb{Q}'}[H^X(t)\prod_{j=2}^{N+1}\prod_{k=1}^{C_{N+1,j}} H_{A(k,N+1,j)}^{(j)}(t)\mathbb{I}\{X \in A(k,N+1,j)\}] =$$

$$\mathbb{E}_{t_0}^{\mathbb{Q}'}[\prod_{j=1}^{N+1}\prod_{k=1}^{C_{N+1,j}} H_{A(k,N+1,j)}^{(j)}(t)\mathbb{I}\{X \in A(k,N+1,j)\}] =$$

$$= \mathbb{E}_{t_0}^{\mathbb{Q}'}[\prod_{j=1}^{N+1}\prod_{k=1}^{C_{N+1,j}} (H_{I,\{A(k,N+1,j)\}}^{(j)}(t)H_{S,\{A(k,N+1,j)\}}^{(j)}(t))\mathbb{I}\{X \in A(k,N+1,j)\}] =$$

$$= [\prod_{j=1}^{N+1}\prod_{k=1}^{C_{N+1,j}} \mathbb{E}_{t_0}^{\mathbb{Q}'}(H_{I,\{A(k,N+1,j)\}}^{(j)}(t))\mathbb{I}\{X \in A(k,N+1,j)\}]\times$$

$$\mathbb{E}_{t_0}^{\mathbb{Q}'}[\prod_{j=1}^{N+1}\prod_{k=1}^{C_{N+1,j}} (H_{S,\{A(k,N+1,j)\}}^{(j)}(t))\mathbb{I}\{X \in A(k,N+1,j)\}].$$

The last calculation is possible because the idiosyncratic intensities are independent from each other, which does not hold for the systematic ones, that are all dependent on the economy.

The choice of the parameters has to be carried out numerically. From a numerical implementation perspective, the terms $\mathbb{E}_{t_0}^{\mathbb{Q}'}(H_{I,\{A(k,N+1,j)\}}^{(j)}(t))\mathbb{I}\{X \in A(k,N+1,j)\}]$ are easily calculated on a standalone basis by using a tree implementation of the intensity dynamics. The term $\mathbb{E}_{t_0}^{\mathbb{Q}'}[\prod_{j=1}^{N+1}\prod_{k=1}^{C_{N+1,j}}(H_{S,\{A(k,N+1,j)\}}^{(j)}(t))\mathbb{I}\{X \in A(k,N+1,j)\}]$ requires more attention as it can only be calculated by simulating all intensities at the same time, which can be cumbersome when these terms are many. A reasonable compromise solution would be to zero-out all systematic terms except the one modelling the default of all names simultaneously. This would significantly reduce the computation budget. Another solution would be to zero-out almost all systematic intensities, but this requires calculating the expectation using Monte Carlo methods on a vector of intensities: this method may be viable if only a small number of intensities is not zeroed-out.

Further interesting attempts to reduce the model parameters number are proposed in Brigo, Mai & Scherer [21], and in Sun, Mendoza-Arriaga & Linetsky [111].

6.5.2 Separable calibration

The proposed calibration strategy is implemented in different steps. As it happened in chapter 4, the calibration of the extended model is still separable, which is a major computational advantage.

1. The model parameters and model volatilities for the variables $X(t)$, $I(t)$, $M(t)$, $m_X(t)$, and $m_I(t)$ are calibrated exactly in the same way proposed in chapter 4, with the *caveat* that all model

volatilities in components $K - D + 1, ..., K$ must be zero to ensure that the idiosyncratic credit component remains independent from the economy.

2. Following the choices made in chapter 4, the calibration is done only to ATM caps/floors and ATM zero-coupon inflation option. This choice is made consistently with the industry practice, where XVA models are not calibrated to volatility skews to save computation resources. Further, at least for an at-the-money swap, the vega exposure of the CVA/DVA term is not significant, given that the short vega of the CVA almost completely tends to offset the long vega from the DVA exposure.

3. As there is no liquid market for credit default swaptions, the intensity volatility functions $v_I^X(t)$ and $v_S^X(t)$ are exogenously determined (maybe using realised volatilities plus some spread, and some proxies). The only constraint is again that the first D components of the idiosyncratic intensity volatilities $v_I^X(t)$ and the last $K - D$ components of the systematic intensity volatilities $v_S^X(t)$ must be zero. The calibration of the different components of the systematic intensity volatilities allows the user to express a view on the correlation between the economy and the systematic credit risk, which can be used to model wrong-way risk. If the market of credit index default swaptions was deemed liquid enough, one could take the implied volatilities from these traded options and rescale them by the ratio of the realised index volatility and the realised single name CDS volatility.

4. The CDS quotes and the exogenous intensity volatilities defined in the previous point are then used to calibrate the intensity dynamics to reproduce the market-implied survival probabilities.

6.6 Monte Carlo simulations

In this simulation we use the calibrated parameters found for 7th December 2012 in the calibration performed in chapter 4. This means that in the current notation $D = 3$ (in chapter 3 we assumed that the dimensionality of the Brownian motion driving the economy was 3) and that we already know the following set of model parameters: $a_I(t)$, $b_I(t)$, $s_I(t)$, $a_X(t)$, $b_X(t)$, $s_X(t)$, and $\lambda(t)$. We extend the volatilities parameters $b_I(t)$, $s_I(t)$, $b_X(t)$, and $s_X(t)$ by adding another 3 components, set to zero to avoid any correlations with the idiosyncratic intensities. We work with wrong-way risk.
Overall, we model the credit intensity of one counterparty and the financial institution, and therefore need another 3 dimensions for the idiosyncratic intensity terms (including the joint Marshall-Olkin term). This means $Y = 1$ and $K = 6$. It should be stressed that the credit market data used for this simulation do not refer to any specific counterparty or financial institution, but are purely fictitious and are used only to provide a numerical example. Log-intensity volatilities have been calibrated to average historic data as per market practice, given that they are hardly observable on an implied basis.

6.6.1 Credit parameters choice

We assume for simplicity that $h_*^B(t_0) = 0.015$, $h_*^C(t_0) = 0.03$: to model the default correlation we propose the split $h^B(t_0) = 0.01$, $h^C(t_0) = 0.025$, and $h_{B,C}^{(2)}(t_0) = 0.005$. Idiosyncraticity is modelled by further splitting these intensities in $h_I^B(t_0) = 0.0005$, $h_I^C(t_0) = 0.0005$, and $h_{I,\{B,C\}}^{(2)}(t_0) = 0.0025$, and $h_S^B(t_0) = 0.0095$, $h_S^C(t_0) = 0.0245$, and $h_{S,\{B,C\}}^{(2)}(t_0) = 0.0025$. We remind the reader that the intensity dynamics are defined in logarithmic terms to ensure that the intensities are always positive.

The collateral basis is assumed to be constant at $l_+^c(t) = 0.001$ and $l_-^c(t) = 0.001$: we are not assuming asymmetric funding costs for the financial institution, which, coupled with the choice of riskless close-out, lets one calculate the XVAs using forward Monte Carlo methods. The long term levels are supposed to be the same as of the initial levels stated above for simplicity, this means $m_J^X(t) = \log h_J^X(t_0)$ for $t \geq t_0$, for $X \in \{\mathcal{B}, \mathcal{C}, (\mathcal{B}, \mathcal{C})\}$ and $Y \in \{I, S\}$. Log-intensities volatilities are assumed to be constant over time and to take the values below: they are compacted in the above matrix $V(t)$:

$$V(t) = \begin{bmatrix} v_I^B(t) \\ v_I^C(t) \\ v_I^{(B,C)}(t) \\ v_S^B(t) \\ v_S^C(t) \\ v_S^{(B,C)}(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.05 \\ -0.1 & -0.1 & -0.1 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0.1 & 0 & 0 & 0 \\ 0.05 & 0.05 & 0.05 & 0 & 0 & 0 \end{bmatrix}.$$

Some comments can be made on how the above matrix influences the modelling of wrong-way risk:

1. The variability due to idiosyncratic moves is of the same order of magnitude of the systematic terms.

2. The calibration results in chapter 4 show an inflation volatility function $s_I(t)$ that is always positive. Therefore, negative values for $v_S^B(t)$ imply that the likelihood of the financial institution defaulting decreases with inflation.

3. The calibration results in chapter 4 show an inflation volatility function $s_I(t)$ that is always positive. Therefore, positive values for $v_S^C(t)$ imply that the likelihood of the counterparty defaulting increases with inflation.

4. For the same reasons, positive terms $v_S^{(B,C)}(t)$ imply that both names are more likely to default in a high inflation environment.

5. The upper-right 3×3 matrix is diagonal, as requested to ensure that the idiosyncratic terms are independent from each other. The upper-left 3×3 matrix contains all zero terms, to ensure that

the idiosyncratic terms are independent from the economic variables. The lower-right 3×3 matrix is only composed of zeros to ensure that the systematic intensity shocks are independent from the idiosyncratic intensity shocks.

Mean reversion speed of log-intensities are supposed to be constant at 1, i.e. $m_J^X(t) = 1$, for $t \geq t_0$, for $X \in \{\mathcal{B}, \mathcal{C}, (\mathcal{B}, \mathcal{C})\}$ and $Y \in \{I, S\}$.

We assume that both recovery rates are 40% (i.e. $R_B = 0.4$ and $R_C = 0.4$) and that the symmetric bond-CDS basis for both names is flat at 30 basis points (i.e. $l^B(t) = 0.003$ and $l^C(t) = 0.003$).

6.6.2 Monte Carlo: some results

We compute the CVA, DVA and LVA for a 10 years-maturity inflation zero-coupon swap, for 100 million EUR notional: this is a standard trade that is done both with corporates without CSA and between financial institutions with a CSA. The position is such that we receive inflation (i.e. the price index performance) and pay a fixed amount of cash $(1 + K(t_0, T))^{T-t_0}$ at maturity T.

We run 2,000 simulations using a 1 week time step: the small size of Monte Carlo error gives us confidence that this parametrisation is good. The same uniform random numbers have been used in the different XVA scenarios to reduce the Monte Carlo error and facilitate the comparison of the results. We have also rerun the simulation using different seeds, and the results were stable.

As specified above, given the assumptions we made, in this case it is possible to calculate the XVAs using a forward Monte Carlo simulation, without having to resort to computationally expensive American Monte Carlo techniques to solve BSDEs. When shocking model parameters in the scenarios shown below, for simplicity in all cases this is done using parallel shocks across the whole curve.

As a first test, we calculate the LVA for different levels of the symmetric collateral spread $l^c(t)$ assuming perfect collateralisation: given that we are assuming symmetric collateral rates, we expect that the LVA is small compared to the other adjustments. This is indeed the case, as shown in the table below: the LVA ranges from -0.036 EUR million when the symmetric spread is decreased to 10 bps, and increases only to -0.057 EUR million when the symmetric spread is increased to a very high level of 200 bps. The basecase CVA is -0.491 EUR million for comparison.

$l^c(t)$	30 bps - basecase	10 bps	200 bps
LVA (EUR million - full collat.)	-0.040	-0.036	-0.057
CVA (EUR million - uncollat.)	-0.491	-0.491	-0.491

It is extremely useful that this set up lets us calculate the LVA from the same calculation process that is run for CVA and DVA, thus saving time and computation resources.

In the four examples that follow we assume there is no CSA, and produce the CVA and DVA, both expressed in EUR million and in basis points running (i.e. CVA and DVA divided by maturity and notional, multiplied by 10,000. This metric is widely used in the industry to compare cost and profitability of different types of trades). CVA and DVA Monte Carlo error is also shown in EUR million.

Finally the experiments shown here are related only to a single unhedged inflation swap: we are not taking into account hedging costs in these experiments, as for example is done in Brigo, Liu, Pallavicini & Sloth [20]. We now consider the CVA and DVA of an uncollateralised position.

1. In this example we move the strike K from ATM (1,95%) -100 bps to ATM + 200 bps. As expected, the DVA (which is a positive adjustment) increases with the strike as the PV of the trade becomes more negative (we pay a higher amount of cash $(1+K(t_0,T))^{T-t_0}$ at maturity T). Qualitatively the DVA behaves as a long inflation floor as a function of the strike (the DVA moves from 0.06 to 2.08 EUR million). The opposite can be said for the CVA, that is a negative adjustment that becomes more negative in a low strike trade, that is more valuable and carries a more significant exposure (the CVA moves from -1.68 to -0.02 EUR million). These observations regarding the qualitative behaviour of the CVA and the DVA are possible thanks to the approximated closed forms for the valuation adjustments that have been presented in section 6.2: their usefulness is clear at this stage. We have just introduced a new model to price CVA and DVA, and we can immediately check that the qualitative behaviour of these adjustments is consistent with what predicted by the approximated closed forms.

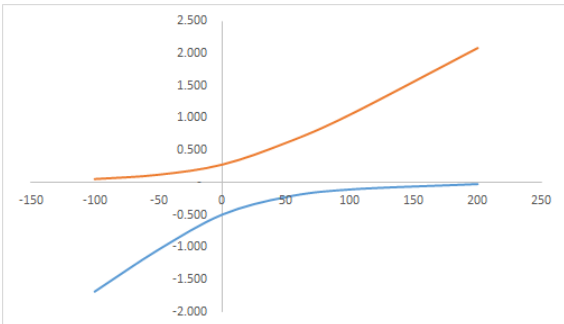


Figure 6.1: CVA (blue) and DVA (red) of an inflation swap (100M EUR, 10Y as a function of its moneyness).

Moneyness	ATM -100	ATM -50	ATM	ATM + 50	ATM + 100	ATM + 200
CVA - uncollat	-1.68	-1.03	-0.49	-0.22	-0.1	-0.02
DVA - uncollat	0.06	0.13	0.28	0.62	1.05	2.08
CVA - MC error	0.02	0.02	0.01	0.01	0.01	0
DVA - MC error	0	0	0.01	0.01	0.01	0.01
CVA - bps running	-16.83	-10.34	-4.91	-2.2	-1.03	-0.19
DVA - bps running	0.58	1.25	2.82	6.19	10.54	20.84

2. In this second example we monitor the DVA behaviour in some scenarios where we shock the volatility $v_S^B(t)$ (and therefore the correlation between the economy and the default intensity of the financial institution $\mathcal{B}$). This shows that the current modelling approach delivers good results when pricing wrong-way risk, given that the DVA is extremely sensitive to the volatility (and therefore to the correlation). The financial intuition behind this findings is that, when inflation is negatively correlated to the credit spread of the financial institution $\mathcal{B}$, a high spread brings about low inflation, that, given the direction of the trade, decreases the PV of the trade and increases the DVA. The DVA is further enhanced by higher credit spreads of the financial institution $\mathcal{B}$. When correlation is positive, this effect disappears as higher spreads are compensated by higher inflation (and therefore higher PV and lower DVA): the DVA moves from 1.2 to 0.31 EUR million.

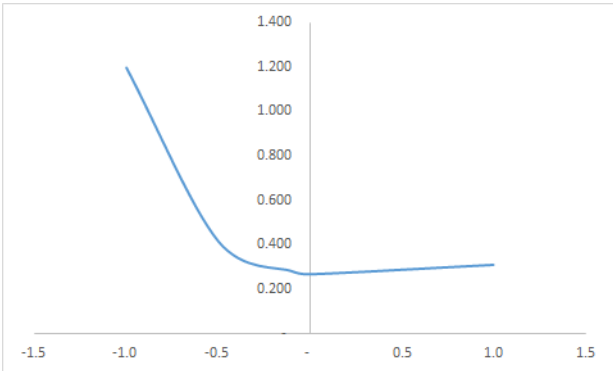


Figure 6.2: DVA of the inflation swap (10 years maturity, 100 EUR million notional) as a function of the financial institution systematic intensity volatility.

$v_S^B(t)$	-1	-0.5	-0.1	0	1
CVA - uncollat	-0.46	-0.49	-0.49	-0.49	-0.39
DVA - uncollat	1.2	0.42	0.28	0.27	0.31
CVA - MC error	0.01	0.01	0.01	0.01	0.01
DVA - MC error	0.05	0.01	0.01	0.01	0.01
CVA - bps running	-4.57	-4.88	-4.91	-4.9	-3.85
DVA - bps running	11.97	4.17	2.82	2.68	3.1

3. The third experiment mirrors the second one: we shock the volatility $v_S^C(t)$ (and therefore the correlation between the economy and the default intensity of the counterparty $\mathcal{C}$). The results mirror the above scenarios: when correlation is higher, higher spreads both increase the CVA and bring about higher inflation: therefore the PV moves up, which further increases the CVA (that moves from -0.41 to -2.18 EUR million). Again this confirms that this modelling of wrong-way risk is powerful for most non pathological situations.

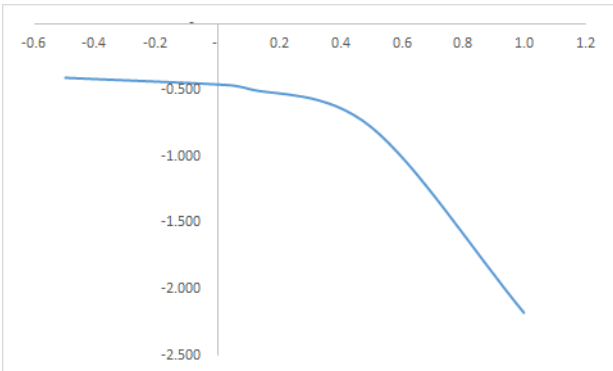


Figure 6.3: CVA of the inflation (10 years maturity, 100 EUR million notional) swap as a function of the counterparty systematic intensity volatility.

$v_S^C(t)$	-0.5	0	0.1	0.5	1
CVA - uncollat	-0.41	-0.46	-0.49	-0.78	-2.18
DVA - uncollat	0.25	0.28	0.28	0.28	0.24
CVA - MC error	0.01	0.01	0.01	0.02	0.12
DVA - MC error	0.01	0.01	0.01	0.01	0.01
CVA - bps running	-4.08	-4.58	-4.91	-7.78	-21.77
DVA - bps running	2.5	2.81	2.82	2.77	2.41

4. To complete our analysis of wrong-way risk, we shock the Marshall-Olkin parameter $h^{(2)}_{I,\{B,C\}}(t)$: because this is an idiosyncratic joint parameter, this has no impact on the economy-credit correlation, but it only affect the correlation between the spread of the financial institution $\mathcal{B}$ and the one of the counterparty $\mathcal{C}$. We propose two cases, one ATM (where CVA and DVA have roughly the same order of magnitude) and a case ATM -100 bps, where CVA is more significant: the latter case has more impact, as the CVA moves from -1.59 to -2.28 EUR million.

   This experiment shows that, as expected, increasing the likelihood of a joint default when the CVA and DVA are balanced (as it happens in the ATM case) has no significant impact on the total adjustment. Instead, when either CVA or DVA is predominant, increasing the joint default likelihood makes the total adjustment higher (in the case below, as the CVA is predominant, the total adjustment becomes more negative as the CVA moves from -15.87 to -22.82 EUR million, while the DVA only moves from 0.52 to 0.92 EUR million.

Moneyness	$h^{(2)}_{I,\{B,C\}}(t)$	0	0.003	0.01	0.02
ATM	CVA - uncollat	-0.47	-0.49	-0.56	-0.64
ATM	DVA - uncollat	0.25	0.28	0.37	0.46
ATM	CVA - MC error	0.01	0.01	0.02	0.02
ATM	DVA - MC error	0.01	0.01	0.01	0.01
ATM	CVA - uncollat - bps running	-4.66	-4.91	-5.61	-6.41
ATM	DVA - uncollat - bps running	2.52	2.82	3.66	4.64
ATM - 100 bps	CVA - uncollat	-1.59	-1.68	-1.95	-2.28
ATM - 100 bps	DVA - uncollat	0.05	0.06	0.07	0.09
ATM - 100 bps	CVA - MC error	0.02	0.02	0.02	0.03
ATM - 100 bps	DVA - MC error	0	0	0	0
ATM - 100 bps	CVA - uncollat - bps running	-15.87	-16.83	-19.54	-22.82
ATM - 100 bps	DVA - uncollat - bps running	0.52	0.58	0.74	0.92

Conclusions

The main contribution of this dissertation is, in our view, to challenge the dichotomy between macroeconomic and pricing models: we provide an enhanced formalisation of a macroeconomic monetary model and use it to price inflation and fixed income derivatives.

The advantages are manifold: the model does not rely on the so-called "Foreign Analogy" and is built on sound economic assumptions, used to model the evolution of the economy, the central bank reaction function, and market liquidity. We assume that the central bank uses the money supply as policy tool, as it has happened in the last years with the so-called "quantitative easing".

The model we propose can be regarded as a structural model for the dynamics of macroeconomic variables, the yield and the inflation swap curves. One of the most striking results is that this macroeconomic model implies a mean-reverting short rate that follows a Hull-White model: this fact is interesting both from a theoretical perspective, as it elegantly links a new model to an established one, and from an economic perspective, as it shows that the mean reversion and its speed are closely linked to the way the central banks implements its monetary policy. Further, from a practical point of view, we know many results for the Hull-White model, which makes option pricing straightforward.

As a consequence, the model lends itself quite naturally to price options on interest rates and on inflation: in many cases (interest rates bond options, caps-floors, swaptions, inflation zero-coupon and year-on-year options) closed form solutions are available thanks to the fact that the terminal distribution of these underlyings is known. In particular we derive in closed form the expression for the year-on-year convexity adjustment, which is a *crux* for any inflation model. Should one need to price exotic and path dependent instruments, the model yields the dynamics as well.

These closed form solutions let us propose a separable calibration strategy that we have successfully implemented and run. The separability of calibration improves its performance significantly, as it is carried out in many simple steps as opposed to a single one-step cumbersome calibration routine. We show some practical applications of this model, in particular for stress tests and for macro-hedging an inflation book using more liquid interest rates derivatives: in both cases the fact that the same model produces inflation dynamics that are consistent with its interest rates dynamics is key.

Building on the analytical tractability of the model we explore how to model inflation options skews: we show some empirical analysis of broker quotes, extend Merton jump-diffusion model to include time-varying and uncertain parameters, and use it to calibrate it to market quotes. It turns out that, even using a very rich parametrisation, the model is not able to fit both zero-coupon and year-on-year skews, thus suggesting that the two markets are not fully liquid and interchangeable.

Finally we remove the initial assumption that there is no credit risk in the economy, and complement the original macroeconomic model with all machinery needed to price credit and funding valuation adjustments. The same model is used to calculate the risk-free PV and its valuation adjustments, which eliminates model risk. Further, in order to test the results, we extend some approximations for valuation adjustments to inflation derivatives, which is a promising direction for further research.

Bibliography

[1] Albanese, C. & Iabichino, S. (2013) *The FVA-DVA Puzzle: Completing the Market with Collateral Trading Strategies*. Available at http://www.albanese.co.uk.

[2] Alexander, R. (2007) *Keeping it Real: Valuation of Inflation-Linked Securities*. MSc Thesis, University of London.

[3] Andersen, A. S. T. (2009) *Inflation Modelling Using Time-Changed Lévy Processes*. Available at SSRN: http://ssrn.com/abstract=1373969.

[4] Andersen, L. & Andreasen, J. (2000) *Jump-Diffusion Processes: Volatility Smile Fitting and Numerical Methods for Option Pricing*. Review of Derivatives Research, 4, 231-262.

[5] Antonov, A., Bianchetti, M. & Mihai, I. (2013) *Funding Value Adjustment for General Financial Instruments: Theory and Practice*. Available at SSRN: http://ssrn.com/abstract=2290987.

[6] Appelbaum, R. (2009) *Lévy Processes and Stochastic Calculus* (2nd edn). Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo, Delhi.

[7] Barclays Capital Research (2005) *Inflation Derivatives, a User's Guide*.

[8] Bayley, R. W. (1994) *Polar Generation of Random Variates with the t-Distribution.* Mathematics of Computation 62(206), 779-781.

[9] Belgrade, N. & Benhamou, E. (2004) *Reconciling Year-on-Year and Zero-Coupon Inflation Swap. A Market Model Approach.* Working Paper.

[10] Benaben, B. et al. (2005) *Inflation-Linked Products*. Risk Books, London, England.

[11] Bertholon, H., Monfort, A. & Pegoraro, F. (2006) *Pricing and Inference with Mixtures of Conditionally Normal Variables.* Available at SSRN: http://ssrn.com/abstract=997975.

[12] Bibby, B. M., Skovgaard, I. M. & Sørensen, M. (2005) *Diffusion-Type Models with Given Marginal Distribution and Autocorrelation Function.* Available at SSRN: www.jstor.org/stable/3318925.

[13] Black, F. & Cox, J.C. (1976) *Valuing Corporate Securities: Some Effects on Bond Indenture Provisions.* Journal of Finance 31(2), 351-67.

[14] Black, F. & Scholes, M. (1973) *The Pricing of Options and Corporate Liabilities.* Journal of Political Economy, 81, 637-654.

[15] Blanchard, O. J. & Khan, C. M. (1980) *The Solution of Linear Difference Models under Rational Expectations.* Econometrica, 48(5), 1305-1311.

[16] Brigo, D. (2002) *The General Mixture-Diffusion SDE and its Relationship with an Uncertain-Volatility Option Model with Volatility-Asset Decorrelation.* Available at http://www.damianobrigo.it.

[17] Brigo, D. (2012) *Counterparty Credit Risk and Collateral Margining.* Available at https://staff.fnwi.uva.nl/p.j.c.spreij/winterschool/14abstracts.html14brigo .

[18] Brigo, D. & Capponi, A. (2008) *Bilateral Counterparty Risk Valuation with Stochastic Dynamical Models and Application to Credit Default Swaps.* Available at http://arxiv.org/abs/0812.3705

[19] Brigo, D. & Chourdakis, K. (2012) *Consistent Single- and Multi-step Sampling of Multivariate Arrival Times: A Characterisation of Self-Chaining Copulas.* Available at SSRN: http://ssrn.com/abstract=2047474.

[20] Brigo, D., Liu, Q. D., Pallavicini, A. & Sloth, D. (2014) *Nonlinear Valuation Under Collateral, Credit Risk and Funding Costs: a Numerical Case Study Extending Black-Scholes.* Available at SSRN: http://ssrn.com/abstract=2430696.

[21] Brigo, D., Mai, J.-F. & Scherer, M. (2014) *Consistent Iterated Simulation of Multi-Variate Default Times: a Markovian Indicators Characterisation.* Available at SSRN: http://ssrn.com/abstract=2431779.

[22] Brigo, D. & Mercurio, F. (2007) *Interest Rates Models - Theory and Practice* (2nd edn). Springer Verlag, Berlin and Heidelberg.

[23] Brigo, D. & Mercurio, F. (2000) *A Mixed-up Smile.* Risk, September, 124-129.

[24] Brigo, D., Mercurio, F. & Rapisarda, F. (2000) *Smile at the Uncertainty.* Risk, May, 97-101.

[25] Brigo, D. & Morini, M. (2005) *CDS Calibration with Tractable Structural Models under Uncertain Credit Quality.* Available at www.damianobrigo.it.

[26] Brigo, D., Morini, M. & Pallavicini, A. (2013) *Counterparty Credit Risk, Collateral and Funding.* John Wiley and Sons, Chicester, England.

[27] Brigo, D., Pallavicini, A. & Perini, D. (2012) *Funding, Collateral and Hedging: Uncovering the Mechanics and the Subtleties of Funding Valuation Adjustments*. Available at SSRN: www.jstor.org/stable/2161528.

[28] Brigo, D., Pallavicini, A. & Torresetti, R. (2007) *Default Correlation, Cluster Dynamics and Single Names: The GPCL Dynamical Loss Model.* Available at SSRN: http://ssrn.com/abstract=956827.

[29] Brody, D., Crosby, J. & Li, H. (2008) *Convexity Adjustments in Inflation-Linked Derivatives*. Asia Risk, November, 52-58.

[30] Burgard, C. & Kjaer, M. (2011) *In the Balance*. Risk, November, 72-75.

[31] Burgard, C. & Kjaer, M. (2011) *Partial Differential Equation Representation of Derivatives with Counterparty Risk and Funding Costs*. Journal of Credit Risk 7(3), 1-19.

[32] Burgard, C. & Kjaer, M. (2013) *Generalised CVA with Funding and Collateral Via Semi-Replication.* Available at SSRN: http://ssrn.com/abstract=2027195.

[33] Björk, T. (1998) *Arbitrage Theory in Continuous Time* (2nd edn). Oxford University Press, Oxford.

[34] Bullard, J. & Mitra, K. (2002) *Learning About Monetary Policy Rules*. Journal of Monetary Economics, 49(6), 1105-1129.

[35] Campbell, J. Y., Shiller, R. H. & Viceira, L. M. (2009) *Understanding Inflation-Linked Bond Markets*. NBER Working Paper Series, 15014.

[36] Cassidy, D. T., Hamp, M. J. & Ouyed, R. (2009) *Pricing European Options with a Log Student's t-Distribution: a Gosset Formula*. Available on arXiv: 0906.4092v1.

[37] Castagna, A. (2010) *FX Options and Smile Risk*. John Wiley and Sons, Chicester, England.

[38] Cesari, G., Aquilina, J., Charpillon, N., Filipovic, Z., Lee, G. & Manda, I. (2010) *Modelling, Pricing, and Hedging Counterparty Risk Exposure: A Technical Guide*. Springer Verlag, Heidelberg.

[39] Cheang, G. H. L. & Chiarella, C. (2011) *A Modern View on Merton's Jump-Diffusion Model*. Research Paper 287. Available at http://www.qfrc.uts.edu.au.

[40] Clarida, R., Gali, J. & Gertler, M. (1999) *The Science of Monetary Policy: A New Keynesian Perspective*. Journal of Economic Perspectives, 37(4), 1661-1707.

[41] Clarida, R., Gali, J. & Gertler, M. (1999) *Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory.* Quarterly Journal of Economics, 115(1), 147-180.

[42] Cochrane, J. H. (2001) *Asset Pricing.* Princeton University Press, Princeton.

[43] Cochrane, J. H. (2014) *Monetary Policy with Interest on Reserves.* Available at http://faculty.chicagobooth.edu /john.cochrane.

[44] Constantinides, G. M. (1992) *A Theory of the Nominal Term Structure of Interest Rates.* The Review of Financial Studies 5 (4), 531-552.

[45] Cont, R. & Tankov, P. (2004) *Financial Modelling with Jump Processes.* Chapman & Hall/ CRC Financial Mathematics Series, Boca Raton, London, New York.

[46] Cornalba, L. (2014) *Funding Value Adjustment and Incomplete Markets.* Available at SSRN: http://ssrn.com/abstract=2498847.

[47] Crépey, S. & Bielecki, T. R. (2014) *Counterparty Risk and Funding: A Tale of Two Puzzles.* Chapman & Hall/ CRC Financial Mathematics Series, Boca Raton, London, New York.

[48] Deacon, M., Derry, A. & Mirfendereski, D. (2004) *Inflation-Indexed Securities.* John Wiley & Sons, Ltd, Chichester, England.

[49] Derman, E. & Kani, I. (1994) *Riding on a Smile.* Risk, 7(2) Feb.1994.

[50] Dodgson, M. & Kainth, D. (2006) *Inflation Linked Derivatives.* Royal Bank of Scotland Group. Available at www.quarchome.org/inflation/InflationLinkedDerivatives20060908.pdf.

[51] Dornbusch, R. & Fischer, S. (1994) *Macroeconomics.* McGraw-Hill Inc, New York.

[52] Duffie, D. (2001) *Dynamic Asset Pricing Theory* (3rd edn). Princeton University Press, Princeton and Oxford.

[53] Elouerkhaoui, Y. (2014) *A Unified Framework for CVA, DVA and FVA Applied to Credit Portfolios.* Available at https://workspace.imperial.ac.uk/mathfin/Public/Seminars.

[54] Embrechts, P., Lindskog, F. & McNeil, A. (2001) *Modelling Dependence with Copulas and Applications to Risk Management.* Available at http://www.math.ethz.ch/ embrecht/ftp/copchapter.pdf.

[55] Fisher, I. (1930) *The Theory of Interest.* The Macmillan Company.

[56] Flashel, P. & Franke, R. (2008) *A Proof of Determinacy in the New-Keynesian Sticky Wages and Prices Model.* Available at http://opus.zbw-kiel.de/volltexte/2008/7367/pdf/EWP-2008-14.pdf.

[57] Fleckenstein, M., Longstaff, F. A. & Lustig, H. (2010) *Why Does the Treasury Issue TIPS? The TIPS-Treasury Bond Puzzle*. Available at http://www.newyorkfed.org/ research/conference/ 2010/cb/Longstaff1.pdf.

[58] Gatheral, J. (2005) *The Volatility Surface: a Practitioner's Guide*. Wiley Finance, Hoboken, New Jersey.

[59] Geman, H., El Karoui, N. & Rochet, J. C. (1995) *Changes of Numeraire, Changes of Probability Measures and Pricing of Options*. Journal of Applied Probability 32, 443-458.

[60] Giacomini, R., Gottschling, A., Haefke, K. & Whie, H. (2008) *Mixtures of t-Distributions for Finance and Forecasting*. Journal of Econometrics 144 (2008), 175-192.

[61] Guyon, J. & Labordère, P. H. (2014) *Nonlinear Option Pricing*. Chapman & Hall/ CRC Financial Mathematics Series, Boca Raton, London, New York.

[62] Gyöngy, I. (1986) *Mimicking the One-Dimensional Marginal Distributions of Processes Having an Ito Differential*. Probability Theory and Related Fields, 71, 501-516.

[63] Hagan, P., Kumar, D., Lesniewski, A. S. & Woodward, D. E. (2002) *Managing Smile Risk*. Wilmott Magazine, July 2002, 84-108.

[64] Harrison, J. M. & Kreps, D. M. (1979) *Martingales and Arbitrage in Multiperiod Securities Markets*. Journal of Economic Theory 120, 381-408.

[65] Harrison, J. M. & Pliska, S. R. (1981) *A Stochastic Calculus Model of Continuous Trading: Complete Markets*. Stochastic Processes and their Applications 11, 313-316.

[66] Henrard, M. (2014) *Interest Rates Modelling in the Multi-Curve Framework*. Palgrave Macmillan, Basingstocke, England.

[67] Heston, S. (1993) *A Closed-Form Solution for Options with Stochastic Volatility, with Application to Bond and Currency Options*. Review of Financial Studies 6, 327-343.

[68] Hurd, M. & Relleen, J. (2006) *New Information from Inflation Swaps and Index-Linked Bonds*. Bank of England Quarterly Bulletin (Spring 2006).

[69] Hughston, L. P. (2004) *International Models for Interest Rates and Foreign Exchange: a General Framework for the Unification of Interest Rate Dynamics and Stochastic Volatility Modelling*. Available at http://www.mth.kcl.ac.uk/finmath/articles.

[70] Hughston, L. P. (1998) *Inflation Derivatives.* Available at http://www.mth.kcl.ac.uk/finmath/articles.

[71] Hughston, L. P. & Macrina, A. (2008) *Information, Inflation, and Interest. Advances in Mathematics of Finance.* Banach Center Publications, Vol. 83. Institute of Mathematics, Polish Academy of Science, Warsaw.

[72] Hughston, L. P. & Macrina, A. (2010) *Pricing Fixed-Income Securities in an Information-Based Framework.* Available on arXiv: 0911.1610v2.

[73] Hughston, L. P. & Rafailidis, A. (2005) *A Chaotic Approach to Interest Rates Modelling.* Finance and Stochastic, Vol. 9 (1), 43-65.

[74] Hull, J. & White, A. (1990) *Pricing Interest-Rate-Derivative Securities.* The Review of Financial Studies, Vol. 3 (4), 573-592.

[75] Hull, J. & White, A. (2012) *The FVA Debate.* Risk, Anniversary edition, July 2012.

[76] Hull, J. & White, A. (2012) *The FVA Debate Continues.* Risk, October 2012.

[77] Jäckel, P. & Bonneton, J. (2009) *Inflation Products and Inflation Models.* EQF 11/036. Available at http://www.pjaeckel.webspace.virginmedia.com/InflationProductsAndInflationModels.pdf .

[78] Jamshidian, F. (1988) *An Exact Bond Option Formula.* Journal of Finance 44, 205-209.

[79] Jarrow, R. A. & Yildirim, Y. (2003) *Pricing Treasury Inflation Protected Securities and Related Derivatives Using an HJM Model.* Journal of Financial and Quantitative Analysis 38(2), 409-430.

[80] Jondeau, E., Poon, S. H. & Rockinger, M. (2007) *Financial Modelling Under Non-Gaussian Distributions.* Springer Verlag London Limited, London, England.

[81] Kenyon, C. (2008) *Inflation is Normal - Revision 1.* Available at SSRN: http://ssrn.com/abstract=1124330.

[82] Kenyon, C. & Stamm, R. (2012) *Discounting, Libor, CVA and Funding.* Palgrave Macmillan, Basingstocke, England.

[83] Kerkoff, J. (2005) *Inflation Derivatives Explained. Markets, Products and Pricing.* Lehman Brothers Fixed Income Quantitative Research. Available at http://www.classiccmp.org/transputer/finengineer.

[84] LaSalle, J. P. (1986) *The Stability and Control of Discrete Processes.* Applied Mathematical Science, Vol 62, Springer, New York.

[85] Laughton, S. & Vaisbrot, A. (2012) *In Defence of FVA: a Response to Hull and White.* Risk, September 2012.

[86] Leippold, M. & Wu, L. (1998) *The Potential Approach to Bond and Currency Pricing.* Available at http://129.3.20.41/ econ-wp/fin/ papers/9903/ 9903004.pdf.

[87] Leung, K. S. & Wu, L. (2008) *Inflation Derivatives: HJM Framework and Market Models.*

[88] Li, X. & Pellerey, F. (2011) *Generalized Marshall-Olkin Distributions, and Related Bivariate Aging Properties.* Journal of Multivariate Analysis 102 (2011) 1399-1409.

[89] Lindskog, F. & McNeil, J. (2003) *Common Poisson Shock Models: Applications to Insurance and Credit Risk Modelling.* ASTIN Bulletin, 33 (2), 209-238.

[90] Lopez de Prado, M. & Foreman, M. D. (2011) *Markowitz Meets Darwin: Portfolio Oversight and Evolutionary Divergence.* Available at SSRN: http://ssrn.com/abstract=1931734.

[91] Mai, J.-F. & Scherer, M. (2009) *A Tractable Multivariate Default Model Based on a Stochastic Time-Change.* International Journal of Theoretical and Applied Finance, Vol 12, No 2, 227-249.

[92] Marshall, A.V. & Olkin, I. (1967) *A Multivariate Exponential Distribution.* Journal of the American Statistical Association, Vol 63, N0 317 (Mar., 1967), 30-44.

[93] Matsuda, K. (2004) *Introduction to Merton Jump Diffusion Model.* Available at http://www.maxmatsuda.com.

[94] McGrath, G. & Windle, R. (2006) *Recent Developments in Sterling Inflation-Linked Markets.* Bank of England Quarterly Bulletin (Q4 2006).

[95] Mercurio, F. (2005) *Pricing Inflation-Indexed Securities.* Quantitative Finance 5(3), 289-302.

[96] Mercurio, F. & Moreni, N. (2006) *Inflation with a Smile.* Risk 19(3), 70-75.

[97] Merton, R. C. (1974) *On the Pricing of Corporate Debt: the Risk Structure of Interest Rates.* The Journal of Finance 29, 449-470.

[98] Merton, R. C. (1976) *Option Pricing When Underlying Stock Returns are Discontinuous.* Journal of Financial Economics 3(1), 125-144.

[99] Pallavicini, A. & Tarenghi, M. (2010) *Interest-Rate Modelling with Multiple Yield Curves.* Available on arXiv: 1006.4767.

[100] Perini, D. (2013) *Marshall-Olkin Copulas.* Manuscript.

[101] Piterbarg, V. (2010) *Funding Beyond Discounting: Collateral Agreements and Derivatives Pricing*. Risk, February, 97-102.

[102] Piterbarg, V. (2012) *Cooking with Collateral*. Risk, August, 58-63.

[103] Rebonato, R. (2004) *Volatility and Correlation. The Perfect Hedger and the Fox* (2nd edn). John Wiley & Sons Ltd, Chicester, England.

[104] Rogers, L. C. G. (1997) *The Potential Approach to the Term Structure of Interest Rates and Foreign Exchange Rates.* Mathematical Finance, 7 157–176.

[105] Sarais, G. (2015) *XVA Hedging: Measuring its Performance and Risks via Discrete-Time Simulations.* Available at SSRN: http://ssrn.com/abstract=2558023.

[106] Schönbucher, P. J. (1997) *Pricing Credit Risk Derivatives*. Working Paper wp-10, London School of Economics, Financial Markets Group, July 1997.

[107] Shaw, W. T. (2006) *Sampling Student's T-Distribution – Use of the Inverse Cumulative Distribution Function.* Journal of Computational Finance (9/4) 37-73.

[108] Shefrin, H. (2008) *A Behavioural Approach to Asset Pricing* (2nd edn). Academic Press, Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sidney, Tokyo.

[109] Shreve, S. (2005) *Stochastic Calculus and Finance* (Vol.1 and 2). Springer Verlag New York Inc, New York, New York.

[110] Sidrauski, M. (1967) *Rational Choice and Patterns of Growth in a Monetary Economy*. American Economic Review, 57(2), 534-544.

[111] Sun, Y., Mendoza-Arriaga, R. & Linetsky, V. (2014) *Marshall-Olkin Distributions, Subordinators, Efficient Simulation, and Applications to Credit Risk*. Available at SSRN: http://ssrn.com/abstract=1702087.

[112] Taylor, J. B. (1993) *Discretion Versus Policy Rules in Practice*. Carnegie-Rochester Conf. Ser. Public Policy, 39, 195-214.

[113] Tian, Y., Zhu, Z., Klebaner, F. & Hamza, K. (2013) *Calibrating and Pricing with Stochastic-Local Volatility Model.* Available at SSRN: http://ssrn.com/abstract=2182411.

[114] Uhlig, H. (1999) *A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily* in Marimon R. & Scotts A. (eds), *Computational Methods for the Study of Dynamic Economies*, Oxford, Oxford University Press, 30-61.

[115] Walsh, C. E. (2003) *Monetary Theory and Policy* (2nd edn). The MIT Press, Cambridge, Massachusetts and London, England.

[116] Woodford, M. (1999) *Optimal Monetary Policy Inertia*. NBER Working Paper N0 7261, July 1999.

Appendix A

DSGE model proofs

In this appendix we provide for convenience the proofs of some significant results that underpin the DSGE model presented in the second chapter, making explicit all approximations. Most of these results are found either in Walsh [115] or in Blanchard & Khan [15].

Proposition 3 *The first order condition* $C_i^{-\sigma} = (1+n_i)\beta\mathbb{E}_i\left[\frac{I_i}{I_{i+1}}C_{i+1}^{-\sigma}\right]$ *is approximately equivalent to the condition* $y_i = \mathbb{E}_i\hat{y}_{i+1} - \frac{1}{\sigma}(\hat{n}_i - \mathbb{E}_i p_{i+1})$ *if one considers the logarithm of the ratio of the economic variables growth rate with respect to their long term equilibrium level.*

One of the first order conditions, that are needed to maximise the economic agent's expected utility stream, is:

$$C_i^{-\sigma} = (1+n_i)\beta\mathbb{E}_i\left[\frac{I_i}{I_{i+1}}C_{i+1}^{-\sigma}\right].$$

Because the economy is closed and there is no public sector nor investment, one writes that the private consumption is equal to the GDP, i.e. $Y_i = C_i$:

$$Y_i^{-\sigma} = (1+n_i)\beta\mathbb{E}_i\left[\frac{I_i}{I_{i+1}}Y_{i+1}^{-\sigma}\right].$$

We remember the definition of the inflation rate: $p_{i+1} = I_{i+1}/I_i - 1$. This implies that $1+\ p_{i+1} = I_{i+1}/I_i$, and therefore $(1+p_{i+1})^{-1} = I_i/I_{i+1}$. One writes:

$$Y_i^{-\sigma} = (1+n_i)\beta\mathbb{E}_i\left[(1+p_{i+1})^{-1}Y_{i+1}^{-\sigma}\right].$$

The following step is to consider the ratio of the economic variables with respect of their equilibrium level. Here we denote the long term equilibrium level with a bar, i.e. the long term equilibrium level

of the variable X is $\bar{X}$. One further assumes that the long term inflation $\bar{p}$ is zero (i.e. prices reach a long term equilibrium) and that the subjective discount factor β is constant over time, i.e. is already in equilibrium: therefore this last variable is ignored. For the output Y_i one considers its growth rate y_i. Finally additive constants are ignored [1]:

$$(y_i/\bar{y})^{-\sigma} = (n_i/\bar{n})\mathbb{E}_i\left[((1+p_{i+1})/(1+\bar{p}))^{-1}(y_{i+1}/\bar{y})^{-\sigma}\right]$$

$$(y_i/\bar{y})^{-\sigma} = (n_i/\bar{n})\mathbb{E}_i\left[((1+p_{i+1})/1)^{-1}(y_{i+1}/\bar{y})^{-\sigma}\right]$$

$$(y_i/\bar{y})^{-\sigma} = (n_i/\bar{n})\mathbb{E}_i\left[(1+p_{i+1})^{-1}(y_{i+1}/\bar{y})^{-\sigma}\right]].$$

One remembers that $\hat{n}_i = n_i/\bar{n} - 1$ and $\hat{y}_i = y_i/\bar{y} - 1$ and writes:

$$(\hat{y}_i+1)^{-\sigma} = (1+\hat{n}_i)\mathbb{E}_i\left[(1+p_{i+1})^{-1}(\hat{y}_{i+1}+1)^{-\sigma}\right].$$

An approximation widely used by macroeconomists is to bring the exponentials – and in general non-linear functions – outside the expectations, which is obviously not exact. This is used in the following derivations:

$$(\hat{y}_i+1)^{-\sigma} \cong (1+\hat{n}_i)\mathbb{E}_i\left[(1+p_{i+1})^{-1}(\hat{y}_{i+1}+1)^{-\sigma}\right].$$

If one considers that $log(x+1) \cong x$ when $x \to 0$, one uses the properties of the logarithm and assumes independence between growth and inflation, clearly making more approximations:

$$-\sigma\log[(\hat{y}_i+1)] = \log[(1+\hat{n}_i)\mathbb{E}_i\left[(1+p_{i+1})\right]^{-1}\mathbb{E}_i[(\hat{y}_{i+1}+1)]^{-\sigma}$$

$$-\sigma\log[(\hat{y}_i+1)] = \log(1+\hat{n}_i) - \log\mathbb{E}_i\left[(1+p_{i+1})\right] - \sigma\log\ \mathbb{E}_i[(\hat{y}_{i+1}+1)]$$

$$-\sigma(\hat{y}_i) = n_i - \mathbb{E}_i\left[p_{i+1}\right] - \sigma\mathbb{E}_i[\hat{y}_{i+1}]$$

$$y_i = \mathbb{E}_i\hat{y}_{i+1} - \frac{1}{\sigma}(\hat{n}_i - \mathbb{E}_i p_{i+1}).$$

Lemma 13 *The two equations*

$$x_i = \mathbb{E}_i x_{i+1} - \frac{1}{\sigma}(\hat{n}_{i+1} - \mathbb{E}_i p_{i+1}) + u_i \tag{A.1}$$

$$p_i = \beta\mathbb{E}_i p_{i+1} + kx_i \tag{A.2}$$

[1] Further details on this approximation procedure can be found in Uhlig [114].

can be compacted in the system

$$\begin{bmatrix} x_i \\ p_i \end{bmatrix} = \frac{1}{\sigma + \delta_x + k\delta_\pi} \left(\begin{bmatrix} \sigma & 1 - \beta\delta_\pi \\ k\sigma & k + \beta(\sigma + \delta_x) \end{bmatrix} \mathbb{E}_i \begin{bmatrix} x_{i+1} \\ p_{i+1} \end{bmatrix} + \begin{bmatrix} 1 \\ k \end{bmatrix} (\sigma u_i - v_i) \right) \tag{A.3}$$

by using the Taylor rule

$$\hat{n}_{i+1} = \delta_\pi p_i + \delta_x x_i + v_i. \tag{A.4}$$

Proof. The proof is done via substitutions.

$$x_i = \mathbb{E}_i x_{i+1} - \frac{1}{\sigma}(\hat{n}_{i+1} - \mathbb{E}_i p_{i+1}) + u_i = \mathbb{E}_i x_{i+1} - \frac{1}{\sigma}((\delta_\pi p_i + \delta_x x_i + v_i) - \mathbb{E}_i p_{i+1}) + u_i$$

$$x_i \left(1 + \frac{\delta_x}{\sigma}\right) = \mathbb{E}_i x_{i+1} - \frac{1}{\sigma}\left(\delta_\pi p_i + v_i - \mathbb{E}_i p_{i+1}\right) + u_i$$

$$x_i \left(1 + \frac{\delta_x}{\sigma}\right) = \mathbb{E}_i x_{i+1} - \frac{1}{\sigma}\left(\delta_\pi \left(\beta \mathbb{E}_i p_{i+1} + k x_i\right) + v_i - \mathbb{E}_i p_{i+1}\right) + u_i$$

$$x_i \left(1 + \frac{\delta_x}{\sigma} + \frac{k\delta_\pi}{\sigma}\right) = \mathbb{E}_i x_{i+1} - \frac{v_i}{\sigma} + \mathbb{E}_i p_{i+1} \frac{1}{\sigma}\left(1 - \delta_\pi \beta\right) + u_i$$

$$x_i = \frac{\sigma}{\sigma + \delta_x + k\delta_\pi} \left[\mathbb{E}_i x_{i+1} + \mathbb{E}_i p_{i+1} \frac{1}{\sigma}\left(1 - \delta_\pi \beta\right) + \left(u_i - \frac{v_i}{\sigma}\right)\right]$$

$$x_i = \frac{1}{\sigma + \delta_x + k\delta_\pi} \left[\sigma \mathbb{E}_i x_{i+1} + \mathbb{E}_i p_{i+1}\left(1 - \delta_\pi \beta\right) + \left(\sigma u_i - v_i\right)\right].$$

With this result in mind we do a substitution in the inflation equation:

$$p_i = \beta \mathbb{E}_i p_{i+1} + k x_i$$

$$p_i = \beta \mathbb{E}_i p_{i+1} + k \left\{ \frac{\sigma}{\sigma + \delta_x + k\delta_\pi} \left[\mathbb{E}_i x_{i+1} + \mathbb{E}_i p_{i+1} \frac{1}{\sigma}\left(1 - \delta_\pi \beta\right) + \left(u_i - \frac{v_i}{\sigma}\right)\right] \right\}$$

$$p_i = \frac{k\sigma}{\sigma + \delta_x + k\delta_\pi} \mathbb{E}_i x_{i+1} + \mathbb{E}_i p_{i+1} \left[\frac{1}{\sigma} \frac{\sigma k}{\sigma + \delta_x + k\delta_\pi}\left(1 - \delta_\pi \beta\right) + \beta\right] + \frac{\sigma k}{\sigma + \delta_x + k\delta_\pi}\left(u_i - \frac{v_i}{\sigma}\right)$$

$$p_i = \frac{1}{\sigma + \delta_x + k\delta_\pi} \left[\mathbb{E}_i x_{i+1} \sigma k + \frac{\sigma}{\sigma} \mathbb{E}_i p_{i+1}\left(k + \beta\left(\sigma + \delta_x\right)\right) + k\sigma\left(u_i - \frac{v_i}{\sigma}\right)\right]$$

$$p_i = \frac{1}{\sigma + \delta_x + k\delta_\pi} \left[\sigma k \mathbb{E}_i x_{i+1} + \mathbb{E}_i p_{i+1}\left(k + \beta\left(\sigma + \delta_x\right)\right)\right] + k\left(\sigma u_i - v_i\right).$$

Lemma 14 *The stability of the above system is ensured if* $k(\delta_\pi - 1) + (1 - \beta)\delta_x > 0$.

Proof. The proof is carried out via an eigenvalue calculation. The stability is ensured if the absolute real part of the eigenvalues is lower than 1. Therefore one calculates the eigenvalues of the matrix:

$$A = \frac{1}{\sigma + \delta_x + k\delta_\pi} \begin{bmatrix} \sigma & 1 - \beta\delta_\pi \\ k\sigma & k + \beta(\sigma + \delta_x) \end{bmatrix}.$$

Because the matrix A is a 2 x 2 matrix, we resort to a result that states the equation for the eigenvalues λ:

$$\lambda^2 - \lambda \ tr(A) + det(A) = 0.$$

One calculates the following:

$$det(A) = \frac{\sigma(k + \beta(\sigma + \delta_x)) - k\sigma(1 - \beta\delta_\pi)}{(\sigma + \delta_x + k\delta_\pi)^2} = \frac{\sigma\beta(\sigma + \delta_x + k\delta_\pi)}{(\sigma + \delta_x + k\delta_\pi)^2} = \frac{\sigma\beta}{\sigma + \delta_x + k\delta_\pi}$$

and

$$tr(A) = \frac{\sigma + k + \beta(\sigma + \delta_x)}{\sigma + \delta_x + k\delta_\pi}$$

and plugs them in the quadratic equation.

We define the two eigenvalues of the matrix A as λ_1 and λ_2 respectively. If both $|\lambda_1| < 1$ and $|\lambda_2| < 1$ hold, we write $|\lambda_1\lambda_2| < 1$. One remembers that $det(A) = |\lambda_1\lambda_2|$, and therefore writes:

$$det(A) = \frac{\sigma\beta}{\sigma + \delta_x + k\delta_\pi} < 1.$$

The implication is that $\sigma(\beta - 1) < \delta_x + k\delta_\pi$, which is always true because $\beta - 1 < 0$ (β is a subjective discount factor, and therefore $\beta \in (0, 1)$) and $\delta_x + k\delta_\pi > 0$ (all these constants are positive).

We check the condition from the request on the $tr(A)$. One remembers that $tr(A) = \lambda_1 + \lambda_2$. If both eigenvalues are smaller than 1 in absolute value, and because the determinant is always smaller than 1, we write:

$$tr(A) = \lambda_1 + \lambda_2 < 1 + det(A).$$

The above result is proved in LaSalle [84]. If one remembers the calculations done above, this condition becomes:

$$\frac{\sigma + k + \beta(\sigma + \delta_x)}{\sigma + \delta_x + k\delta_\pi} < 1 + \frac{\sigma\beta}{\sigma + \delta_x + k\delta_\pi}.$$

By doing the calculations one finally finds the stability condition:

$$k(\delta_\pi - 1) + \delta_x(1 - \beta) > 0.$$

Appendix B

Derivation of a European option pricing formula under a t-distribution

In this appendix we show the derivation of the price of a call option when the underlying follows a Student's t-distribution with n degrees of freedom, location parameter μ and scale parameter σ. The result shown below holds if $n > 1$.

We recall the following definitions of the gamma function and of the analytic expression of the t-distribution:

1. $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$ where $a \in \mathbb{R}$

2. $\Gamma(n) = (n-1)!$ where $n \in \mathbb{N}$

3. $f_t(n, \mu, \sigma) = (\sigma^2 n\pi)^{-\frac{1}{2}} \Gamma(\frac{n+1}{2}) / \Gamma(\frac{n}{2}) (1 + (\frac{t-\mu}{\sigma})^2 \frac{1}{n})^{-(\frac{n+1}{2})}$

4. $F_t(n, \mu, \sigma) = \int_{-\infty}^{t} f_x(n, \mu, \sigma) dx$.

When $\mu = 0$ and $\sigma = 1$ we write in a lighter notation the standardised t-distribution:

1. $f_n(t) = f_t(n, 0, 1)$

2. $F_n(t) = F_t(n, \mu, \sigma)$.

The undiscounted price of a call option at time t with maturity T, underlying S, forward $\bar{S}$ (under

the appropriate pricing measure), and strike K is:

$$\mathbb{E}_t[(S(T)-K)^+] = \int_K^{+\infty} (S(T)-K)(\tau\sigma^2 n\pi)^{-\frac{1}{2}}\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)\left(1+\left(\frac{S(T)-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}\right)^2\frac{1}{n}\right)^{-(\frac{n+1}{2})} dS(T)$$

where we have set the time to maturity $\tau = T - t$.

We start by making a first change of variables by setting $\xi = \frac{S(T)-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}$, which implies $dS(T) = \sigma(\tau)^{\frac{1}{2}} d\xi$ and $S(T) = \xi\sigma(\tau)^{\frac{1}{2}} + \bar{S}$. We also set $k = \frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}$. After this change of variables the above integral becomes:

$$\int_{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}}^{+\infty} (\xi\sigma(\tau)^{\frac{1}{2}} + \bar{S} - K)(\tau\sigma^2 n\pi)^{-\frac{1}{2}}\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)\left(1+(\xi)^2\frac{1}{n}\right)^{-(\frac{n+1}{2})}(\sigma(\tau)^{\frac{1}{2}} d\xi) =$$

$$= \int_{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}}^{+\infty} (\xi\sigma(\tau)^{\frac{1}{2}}) f(\xi) d\xi + (\bar{S} - K)(1 - F(k)).$$

We focus our attention on the first integral, ignoring all multiplicative factors, and calculate the integral:

$$\int_{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}}^{+\infty} \xi\left(1+\frac{\xi^2}{n}\right)^{-\frac{n+1}{2}} d\xi.$$

We propose a new change of variables by setting $z = \left(1+\frac{\xi^2}{n}\right)^{\frac{n+1}{2}}$, which implies $\xi = \pm\left(n\left(z^{\frac{2}{n+1}}-1\right)\right)^{\frac{1}{2}}$ and $d\xi = \pm z^{\frac{1-n}{n+1}}\left(z^{\frac{2}{n+1}}-1\right)^{-\frac{1}{2}}\frac{n}{n+1}dz$. This is shown by doing the following calculations:

$$z = \left(1+\frac{\xi^2}{n}\right)^{\frac{n+1}{2}}$$

$$z^{\frac{2}{n+1}} - 1 = \frac{\xi^2}{n}$$

$$\xi = \pm\left(n\left(z^{\frac{2}{n+1}}-1\right)\right)^{\frac{1}{2}}$$

$$d\xi = \pm d\left(n\left(z^{\frac{2}{n+1}}-1\right)\right)^{\frac{1}{2}} = \pm\frac{n}{2}\left(n\left(z^{\frac{2}{n+1}}-1\right)\right)^{-\frac{1}{2}}\frac{2}{n+1}z^{(\frac{2}{n+1}-1)}dz = \pm\frac{n}{n+1}\left(n\left(z^{\frac{2}{n+1}}-1\right)\right)^{-\frac{1}{2}} z^{\frac{1-n}{n+1}}dz.$$

With this result we can calculate the integral above by applying the change of variables discussed:

$$\int \xi\left(1+\frac{\xi^2}{n}\right)^{-\frac{n+1}{2}} d\xi = \int\left(n\left(z^{\frac{2}{n+1}}-1\right)\right)^{\frac{1}{2}}\frac{1}{z}\frac{n}{n+1}\left(n\left(z^{\frac{2}{n+1}}-1\right)\right)^{-\frac{1}{2}} z^{\frac{1-n}{n+1}}dz = \int\frac{n}{n+1}z^{(\frac{1-n}{n+1}-1)}dz =$$

$$= \int\frac{n}{n+1}z^{(\frac{1-n}{n+1}-1)}dz = \frac{n}{n+1}\frac{1}{\left(\frac{1-n}{n+1}-1+1\right)}z^{(\frac{1-n}{n+1}-1+1)} + c = \frac{n}{1-n}z^{(\frac{1-n}{n+1})} + c.$$

By undoing the last substitution one gets:

$$\frac{n}{1-n} z^{\left(\frac{1-n}{n+1}\right)} = \frac{n}{1-n}\left(\left(1+\frac{\xi^2}{n}\right)^{\frac{n+1}{2}}\right)^{\left(\frac{1-n}{n+1}\right)} =$$

$$= \frac{n}{1-n}\left(1+\frac{\xi^2}{n}\right)^{\frac{1-n}{2}} = \frac{n}{1-n}\left(1+\frac{\xi^2}{n}\right)^{\frac{1-n}{2}+1-1} =$$

$$\frac{n}{1-n}\left(1+\frac{\xi^2}{n}\right)^{\frac{1-n}{2}-1}\left(1+\frac{\xi^2}{n}\right) = \frac{n}{1-n}\left(1+\frac{\xi^2}{n}\right)^{\frac{1-n-2}{2}}\left(1+\frac{\xi^2}{n}\right) =$$

$$= \frac{n}{1-n}\left(1+\frac{\xi^2}{n}\right)^{-\frac{n+1}{2}}\left(1+\frac{\xi^2}{n}\right).$$

With these calculations one finally solves the original integral:

$$\int_{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}}^{+\infty} (\xi\sigma(\tau)^{\frac{1}{2}}) f(\xi) d\xi = \int_{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}}^{+\infty} (\xi\sigma(\tau)^{\frac{1}{2}}) f(\xi) d\xi =$$

$$= \int_{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}}^{+\infty} (\xi\sigma(\tau)^{\frac{1}{2}})(\sigma^2\tau n\pi)^{-\frac{1}{2}}\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)\left(1+\frac{\xi^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} d\xi$$

$$= \int_{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}}^{+\infty} \xi(n\pi)^{-\frac{1}{2}}\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)\left(1+\frac{\xi^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} d\xi =$$

$$= (n\pi)^{-\frac{1}{2}}\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)\int_{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}}^{+\infty} \xi\left(1+\frac{\xi^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} d\xi =$$

$$= (n\pi)^{-\frac{1}{2}}\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)\frac{n}{1-n}\left|\left(1+\frac{\xi^2}{n}\right)^{-\frac{n+1}{2}}\left(1+\frac{\xi^2}{n}\right)\right|_{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}}^{+\infty}$$

$$= (n\pi)^{-\frac{1}{2}}\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)\frac{n}{n-1}\left(1+\frac{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}^2}{n}\right)^{-\frac{n+1}{2}}\left(1+\frac{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}^2}{n}\right)$$

$$= \frac{n}{n-1}\left(1+\frac{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}^2}{n}\right) f_n\left(\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}\right).$$

One finds finally the result:

$$\mathbb{E}_t[(S(T)-K)^+] = \frac{n}{n-1}\left(1+\frac{\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}^2}{n}\right) f_n\left(\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}\right) + (\bar{S}-K)\left(1-F_n\left(\frac{K-\bar{S}}{\sigma(\tau)^{\frac{1}{2}}}\right)\right).$$

One notes that this result converges to the result found for the Normal model when n diverges to $+\infty$.

Appendix C

Option pricing with mixtures

C.1 Introduction

Mixtures of distributions improve calibration to market-observed skews and smiles while keeping the model tractable. The idea of using mixtures to improve model calibration has been explored in literature by many: see for example Brigo & Mercurio [22], Brigo & Mercurio [23], Rebonato [103], and Giacomini, Gottschling, Haefke & White [60].

C.2 Definitions and properties

Definition - A random variable X has a mixture distribution when its density $f(x)$ is expressed as a finite combination of $M \in \mathbb{N}$ densities ("mixture components"), with positive weights summing up to one.

$$f(x) = \sum_{j=1}^{M} f_j(x;\theta_j)p_j = \sum_{j=1}^{M} f_j(x)p_j$$

with

$$p_j \in [0,1]$$

and

$$\sum_{j=1}^{M} p_j = 1.$$

The vector θ_j contains the parameters of the distribution (E.g.: if the distribution is normal, $\theta_j = [\mu_j, \sigma_j]$).

Interpretation - One can think to the mixture as a way to express the distribution of a random variable: this is equivalent to assuming that the distribution $f(x)$ of the random variable X is expressed

as a linear combination of the distributions $f_i(x)$ of the instrumental random variables X_i. Alternatively one can think that the random variable X has unconditional distribution $f(x)$, which can be thought as an expected distribution. There are M states of the world $\{1, 2, ..., M\}$ and that the random variable Y, independent from X, tells what state we are in. Each state of the world has a probability p_j, and X has distribution $f_j(x)$ in the state j.

Moments - Moments of X are calculated as functions of the moments of the components X_i, which we denote as $\mathbb{E}\left[X_i^k\right] = \mathbb{E}^i\left[X^k\right] = \mu_i^k$. Simple integration leads to:

$$\mathbb{E}\left[X^k\right] = \int_\Omega x^k f(x)dx = \int_\Omega x^k \sum_{j=1}^{M} f_j(x)p_j dx = \sum_{j=1}^{M} \int_\Omega x^k f_j(x)p_j dx = \sum_{j=1}^{M} \mathbb{E}\left[X_j^k\right] p_j. \tag{C.1}$$

More generally, using Newton's binomial, we write for the k-th centered moment:

$$\mathbb{E}\left[(X-\mu)^k\right] = \sum_{j=1}^{M} p_j \mathbb{E}^j\left[(X-\mu)^k\right] = \sum_{j=1}^{M} p_j \mathbb{E}^j\left[(X-\mu_j+\mu_j-\mu)^k\right] =$$

$$\sum_{j=1}^{M}\sum_{i=0}^{k} \binom{k}{i} (\mu_j-\mu)^{k-i} p_j \mathbb{E}^j\left[(X_j-\mu_j)^i\right]. \tag{C.2}$$

Properties - We derive the result for the expected value of a function $h(X)$:

$$\mathbb{E}\left[h(X)\right] = \int_\Omega h(x)f(x)dx = \int_\Omega h(x)\sum_{j=1}^{M} f_j(x)p_j dx =$$

$$\sum_{j=1}^{M} \int_\Omega h(x)f_j(x)p_j dx = \sum_{j=1}^{M} \mathbb{E}^j\left[h(X)\right] p_j. \tag{C.3}$$

This property is key to option pricing. Because the present value (PV) of an option is an expected value, if the PV is known in closed form under certain distribution assumptions (E.g., normal) the PV under the mixture of these distributions is just a linear combination of the closed-form PV:

$$PV = \sum_{j=1}^{M} p_j PV_j. \tag{C.4}$$

This property is the main reason of the fortune of mixtures in finance. In many cases (Eg.: Black, Black Normal, Student's t-distribution, ...) the PV of an option is known, but the distribution implied by these models is not good enough to reproduce the skew and smiles traded on the market. Mixtures bridge this gap with a very elegant and intuitive construction.

This property is also used to derive the moment generating function and the characteristic function of

X: they are calculated as the weighed sum of the moment generating functions and of the characteristic functions of the components respectively. The function ν here denotes the logarithm of the Laplace transform.

$$\mathcal{M}^X = \mathbb{E}\left[e^{tX}\right] = \int_\Omega e^{tx} f(x)dx = \int_\Omega e^{tx} \sum_{j=1}^{M} f_j(x)p_j dx =$$

$$\sum_{j=1}^{M} \int_\Omega e^{tx} f_j(x)p_j dx = \sum_{j=1}^{M} \mathbb{E}^j\left[e^{tx}(X)\right] p_j = \sum_{j=1}^{M} \mathcal{M}_j^X p_j = \sum_{j=1}^{M} e^{\nu_j} p_j. \tag{C.5}$$

Positive weights - When calibrating a mixture to the smile, one is typically solving a constrained minimisation problem where he is minimising, for each maturity, the sum across strikes of the square difference between market prices and model prices of vanilla options. The constraint is that the weights p_j have to be positive and add up to one and that the expected value of the underlying respects the forward condition. In order to eliminate the positivity constraints on the weights, Rebonato [103] suggest using some trigonometric functions that can speed up the calibration and eliminate some constraints. Let us assume $M = 2$ and write $p_1 = a^2$ and $p_2 = b^2$: positivity is ensured. To ensure that $p_1 + p_2 = 1$, i.e. $a^2 + b^2 = 1$, we write $a = sin(c)$ and $b = cos(c)$, with $c \in [0, 2\pi)$. Therefore the problem of finding p_1 and p_2, with the constraint $p_1 + p_2 = 1$, boils down to the unconstrained problem of finding the angle c. When $M > 2$ the approach is generalised using polar coordinates in an M-dimensional hypersphere, finding $M - 1$ angles $c_1, c_2, ..., c_{M-1}$ such that

$$p_j = \left(cos(c_j) \prod_{i=1}^{j-1} sin(c_i)\right)^2 \qquad j = 1, 2, ..., M-1$$

$$p_M = \left(\prod_{i=1}^{M-1} sin(c_i)\right)^2.$$

Monte Carlo simulation - Monte Carlo simulations for mixtures are carried out as a two step process. One initially generates the random variable Y and then, based on this variable (that tells in what state of the world we are) generates the actual value of the random variable X.

C.3 Moments of a Gaussian mixture

The previous section contains general results because no distribution for the components is specified. Here we show some basic results that one obtains when the mixture for the random variable u_i has M_u components, each of them has Gaussian distribution with mean $\mu_{i,j}^u$ and variance $(\sigma_{i,j}^u)^2$. Here the time

index i is used again. We introduce a zero-mean constraint for the distribution. The mixture weights are denoted by $p^u_{i,j}$:

$$\begin{cases} \mathbb{E}\left[u_i\right] = \sum_{j=1}^{M} p^u_{i,j}\mu^u_{i,j} \;=\; 0 \\ \mathrm{Var}(u_i) = \mathbb{E}\left[(u_i - \mathbb{E}\left[u_i\right])^2\right] = \mathbb{E}\left[(u_i)^2\right] = \sum_{j=1}^{M} p^u_{i,j}[(\mu^u_{i,j})^2 + (\sigma^u_{i,j})^2] \\ \mathrm{Skew}(u_i) = \mathbb{E}\left[(u_i - \mathbb{E}\left[u_i\right])^3\right] = \mathbb{E}\left[(u_i)^3\right] = \sum_{j=1}^{M} p^u_{i,j}[(\mu^u_{i,j})^3 + 3(\mu^u_{i,j})(\sigma^u_{i,j})^2] \\ \mathrm{Kurt}(u_i) = \mathbb{E}\left[(u_i - \mathbb{E}\left[u_i\right])^4\right] = \mathbb{E}\left[(u_i)^4\right] = \sum_{j=1}^{M} p^u_{i,j}[(\mu^u_{i,j})^4 + 6(\mu^u_{i,j})^2(\sigma^u_{i,j})^2 + 3(\sigma^u_{i,j})^4]. \end{cases} \tag{C.6}$$

The problem of finding the mixture parameters becomes simpler if one uses the result shown in Bertholon, Monfort & Pegoraro [11]. This result lets one calculate the mixture parameters in closed form without relying on numerical algorithms, if one assumes that the distributions are normal and the number of distributions is 2. The authors show that a mixture of two normal random variables can achieve any mean, variance, skew and kurtosis. Furthermore they show the formulas of the mixture parameters, as functions of the target moments.

Their findings can be summarised as follows. Let us assume that the first fourth moments of the random variable X are known, and let us denote them as $(\mu, \sigma^2, \mu^3, \mu^4)$. One standardizes the third and fourth moments by defining $\tilde{\mu}^3 = \mu^3/(\sigma^{3/2})$ and $\tilde{\mu}^4 = \mu^4/(\sigma^4)$. There are two cases:

1. If either $\tilde{\mu}^3 \neq 0$ or if both $\tilde{\mu}^3 = 0$ and $\tilde{\mu}^4 < 3$, one sets:

   (a) a is the unique root $\geqslant 1$ of the polynomial $p(x) = (\tilde{\mu}^3)^2 x^3 + (3 - \tilde{\mu}^4)x^2 - 2$

   (b) $p = \frac{1}{2} - \frac{a^{\frac{3}{2}}\tilde{\mu}^3}{2((a^{\frac{3}{2}}\tilde{\mu}^3)^2+4)^{\frac{1}{2}}}$

   (c) $\mu_1 = \mu + \sigma\left(\frac{1-p}{ap}\right)^{\frac{1}{2}}$

   (d) $\mu_2 = \mu - \sigma\left(\frac{p}{a(1-p)}\right)^{\frac{1}{2}}$

   (e) $\sigma_1 = (\sigma^2(a-1)/a)^{\frac{1}{2}}$

   (f) $\sigma_2 = (\sigma^2(a-1)/a)^{\frac{1}{2}}$.

2. In all other cases, i.e. when both both $\tilde{\mu}^3 = 0$ and $\tilde{\mu}^4 \geqslant 3$ one sets:

   (a) $p = \frac{1}{2} \pm \frac{1}{2}\left(1 - \frac{3}{\tilde{\mu}^4}\right)^{\frac{1}{2}}$

   (b) $\mu_1 = \mu$

   (c) $\mu_2 = \mu$

   (d) $\sigma_1 = (\sigma^2/2p)^{\frac{1}{2}}$

   (e) $\sigma_2 = (\sigma^2/2(1-p))^{\frac{1}{2}}$.