Part I

Theory and Practice of Model Risk Management

1 Understanding Model Risk

1.1 WHAT IS MODEL RISK?

In the last years, during and after the credit crunch, we have often read in the financial press that errors on 'models' and lack of management of 'model risk' were among the main causes of the crisis. A fair amount of attacks regarded mathematical or quantitative models, like the notorious Gaussian copula, that were accused to be wrong and give wrong prices for complex derivative, in particular credit and mortgage-related derivatives. These criticisms to valuation models have been shared also by bank executives and people that are not unexperienced on the reality of financial markets. In spite of this it is not very clear when a model must be considered *wrong*, and as a consequence it is not clear what model risk is.

We can probably all agree that *model risk is the possibility that a financial institution suffers losses due to mistakes in the development and application of valuation models*, but we need to understand which mistakes we are talking about.

In the past, model validation and risk management focused mainly on detecting and avoiding errors in the mathematical passages, the computational techniques and the software implementation that we have to perform to move from model assumptions to the quantification of prices. These sources of errors are an important part of model risk, and it is natural that model risk management devotes a large amount of effort to avoid them. We will devote a share of the second part of this book to related issues. However, they regard that part of model risk that partially overlaps with a narrow definition of operational risk: the risk associated to lack of due diligence in tasks for which it is not very difficult to define what should be the right execution. Is this what model validation is all about? In natural science, the attempt to eliminate this kind of error is not even part of model validation. It is called *model verification*, since it corresponds to verifying that model assumptions are turned correctly into numbers. The name model validation is instead reserved to the activity of assessing if *the assumptions of the model are valid*. Model assumptions, not computational errors, were the focus of the most common criticisms against quantitative models in the crisis, such as 'default correlations were too low'.

The errors that we can make in the assumptions underlying our models are the other crucial part of model risk, probably underestimated in the past practice of model risk management. They are the most relevant errors in terms of impact on the reputation of a financial institution that works with models. A clear example is what happened with rating agencies when the subprime crisis burst. When they were under the harshest attacks, rating agencies tried to shield themselves from the worst criticisms by claiming that the now evident underestimation of the risk of credit derivatives was not due to wrong models, but to mistakes made in the software implementation of the models. Many market operators, that knew the models used by rating agencies, did not believe this justification, and it had no other effect than increasing the perception that wrong models were the real problem. What is interesting to notice is that admitting wrong software appeared to them less devastating for their reputation than admitting wrong models.

Unfortunately, errors in mathematics, software or computational methods are easy to define and relatively easy to detect, although this requires experience and skills, as we will see in the second part of the book. Errors in model assumptions, instead, are very difficult to detect. It is even difficult to define them. How can we, as the result of some analysis, conclude that a model, intended as a set of assumptions, has to be considered wrong? We need to understand when a valuation model must be called wrong in order to answer to our first crucial question, *what is model risk?*

In this section we look for the answer. The first sources we use to clarify this issue are the words of a few legendary quants that in the past have tried to say when models are right or wrong in order to give a definition of model risk. You will see that not even among quants there is consensus about what model risk is. But then, when we apply these approaches to past crises to understand how they could have protected us from the worst model losses, we will see that the different approaches can lead to similar practical prescriptions.

1.1.1 The Value Approach

As early as 1996, before both the LTCM collapse and the credit crunch, the two events that put most critical pressure on the risk involved in using mathematical pricing models, one of living legends of quantitative finance, Emanuel Derman, wrote a paper titled *Model Risk*. This is a natural starting point to define our subject, also because it can be seen as the foundation of one of the two main schools of thought about model risk. The views of the author on the subject are further specified by a later paper written in 2001 that addresses model validation prescriptions, under the title *The Principles and Practice of Verifying Derivatives Prices*.

Derman notices first that the previous years had seen the emergence of an *'astonishingly theoretical approach to valuation of risky products. The reliance on models to handle risk'*, he points out, *'carries its own risk'*. Derman does not give a definition of model risk, but he indicates some crucial questions that a model validator should have in mind:

1. Is the payoff accurately described?
2. Is the software reliable?
3. Has the model been appropriately calibrated to the prices of the simpler, liquid constituents that comprise the derivative?
4. *'Does the model provides a realistic (or at least plausible) description of the factors that affect the derivative's value?'*

Can we deduce a definition of model risk from these points? The first two points are not trivial. When speaking of approximations and numerics in Chapter 6 we will talk of errors to avoid in implementation, and we even devote the entire Chapter 10 to the errors that can be made in the description of a payoff. However, these points do not add to our understanding of Derman's ideas about the nature of the errors we can make in model assumptions.

The third point instead underlines a feature that models must have: the capability to price consistently with the market the simpler instruments related to a derivative, namely to perform the so-called *calibration*. This is an important issue, on which we will focus later on. But not even this point clarifies what model risk is. All banks, now, calibrate their models to liquid market prices. For any asset class or financial product there are many models which are different from each other and yet can all be calibrated very well to the market. Once we have satisfied this calibration constraint, are we sure that model risk has been eliminated, or instead the core of model risk is crucially linked to the fact that we have *different* models allowing for good calibration, so that calibration does not solve our model uncertainty?

A better clarification is given in the fourth point. From this we can deduce a definition of model risk. Once we are sure that we have correctly implemented payoff and software, and our model appears calibrated to the liquid underlying products, we have a residual risk that seems to be the core of model risk:

> *Model risk is the risk that the model is not a realistic/plausible representation of the factors affecting the derivative's value*

This is confirmed when Derman says that for less liquid or more exotic derivatives one must verify the *'reasonableness of the model itself'*. There is more. Derman (1996) gives an account of the things that can go wrong in model development, and he starts from some examples where lack of realism is surely the crucial problem:

> *'You may have not taken into account all the factors that affect valuation ... You may have incorrectly assumed certain stochastic variables can be approximated as deterministic .... You may have assumed incorrect dynamics ... You may have made incorrect assumptions about relationships'.* E. Derman, Model Risk.

So, is Derman saying that we should try to find out what the *true model* is? No, in fact he never uses those somewhat exoteric concepts like the *true model* or *right model*. He states, and it is hard to disagree, that a model is always an *'attempted simplification of a reality'*, and as such there can be no true or perfectly realistic model. But realism and reasonableness, coupled with simplicity, must remain crucial goals of a modeller, and their lack creates model risk.

Is Derman saying that we must look for realism and reasonableness in all aspects of the model? Not either. We must care for those aspect that have a relevant impact, limiting the analysis to '*the factors that affect the derivative's value*'.

This approach to model risk is probably the one shared by most practitioners of finance and beyond, and does not appear too far away from the views expressed more recently by Derman. For example, in the 'Financial Modeler's Manifesto', written with Paul Wilmott, another legend of quant finance, we read among the principles that a modeler should follow *'I will never sacrifice reality for elegance without explaining why I have done so. Nor I will give the people who use my model false comfort about its accuracy'*. We refer to this, and to Derman's recent book 'Models Behaving Badly – Why Confusing Illusion with Reality Can Lead to Disaster, on Wall Street and in Life', whose title is already very telling, for more about Derman's views.

It is clear to everyone that knows finance and does not confuse it with mathematics and not even with physics, that there is not such a thing as the 'true value' of a derivative that the model should be able to compute. However realism and capability to describe the actual behaviour of the relevant risk-factors are crucial principles to judge a model, and more realistic models should be preferred. Somewhat, we can say that the right model and the right value do not exist in practice, but wrong models and wrong values do exist, they can be detected and we should commit ourselves to find models giving values as 'little wrong' as possible, and then manage the residual unavoidable risk. This is the reason why we talk of 'Value approach'.

There are cases where we can all agree that the price given by some models does not correspond to the value of a derivative. Most of these cases are trivial. If we are selling an out-of-the money option on a liquid volatile underlying, the model we use must incorporate some potential future movement of the underlying. We cannot use a deterministic model, assuming no volatility. Otherwise we would be selling the option for nothing, based on an assumption that can be disproved just waiting a bit and seeing the price of the underlying move in the market.

We will see other examples which are less trivial and yet we can easily spot that some assumptions are not realistic. To give an example regarding the infamous credit models, you will see in Chapter 2 the case of default predicted exactly by spreads going to infinity according to standard structural models or in Chapter 3, speaking of Gaussian copula, again a default predicted exactly, and some years in advance, by the default of another company. These assumptions are unrealistic and yet they are hidden in two very common models. When they do not impact in a relevant way the value of a derivative, we can consider them harmless simplifications. When, like in the examples we will analyze, we can show that they impact strongly the value of a derivative, we should raise a warning. At times it is more difficult to say when a relevant feature of a model is realistic or not; in this case we will have to use our judgement, collect as much information as possible and try to make the best possible choice.

You may at first think that everyone must agree with such a reasonable and no-nonsense approach, and with the definition of model risk it implies. It is not like that. A view on Model Risk that starts from completely different foundations is analyzed in the next section.

1.1.2 The Price Approach

If Derman has been one of the fathers of quantitative modelling between the end of the eighties and the nineties, Riccardo Rebonato marked the development of interest rate models – the field where the most dramatic quantitative developments have been done – between the end of the nineties and the subsequent decade. He has been a master in bridging the gap between complex mathematics and market practice. After the turn of the century Rebonato wrote a paper titled *Theory and Practice of Model Risk Management* that presents a view on the subject strongly different, at first sight, from the classic view explained above.

Rebonato (2003) takes the point of view of a financial institution, which is worried not only of the material losses associated to model risk, but even more of the effect that evidence of model risk mismanagement can have on the reputation of a financial institution and its perceived ability to control its business. Under this point of view, this classic definition of model risk and model validation are misplaced. In fact derivatives need to be marked-to-market, as we will see in Section 1.3, and this means that the balance-sheet value of a derivative must come as much as possible from market prices.

If this is the situation, what should the main concern of a model validation procedure be? Should we worry so much that ‘*the model provides a realistic (or at least plausible) description of the factors that affect the derivative’s value*’? Well . . . at least this is not the first concern we must have, since, to use the words of Rebonato, *‘Requiring that a product should be marked to market using a more sophisticated model (ie, a model which makes more realistic assumptions) can be equally misguided if . . . the market has not embraced the “superior” approach.’*

These considerations lead Rebonato to an alternative definition of model risk, that has become so popular that we can consider it the motto of a different approach to model risk, the Price approach:

> *‘Model risk is the risk of occurrence of a significant difference between the mark-to-model value of a complex and/or illiquid instrument, and the* price *at which the same instrument is revealed to have traded in the market’*. Rebonato R., *Theory and Practice of Model Risk Management*

Rebonato (2003) justifies this view pointing out that the real losses that hit an institution’s balance sheet usually do not appear *‘because of a discrepancy between the model value and* *the "true" value of an instrument'*, but through the mark-to-market process, because of a discrepancy between the model value and the market price.

Fair enough. It is hard to disagree with such statements. As long as the market agrees with our model valuation, we do not have large losses due to models. When we evaluate with a model which is the same one used to reach market prices, we do not have model losses arising from mark-to-market thus we have no accounting losses. More interestingly, we can also avoid material losses, because, if the market agrees with our valuation model, we can always sell an asset or extinguish a liability at the price at which we have booked. This is true even if the market model is, to use the words of Rebonato, *'unreasonable, counterintuitive, perhaps even arbitrageable'*.[1]

This has another implication. When the market price can be observed quite frequently, there is little time during which the model price and market price of a derivative can diverge, so that big model risk is unlikely to be generated. If a bank notices a mispricing, this will be controlled by provisions such as stop-losses and will not generate losses so big to worry an institution, although they can worry a single trader. The problem arises with very complex or illiquid products, for which market prices are not observed frequently. Then the model price of a derivative and its market price can diverge a lot, and when eventually the market price gets observed a large and sudden loss needs to be written in the balance-sheet, with effects on a bank which are also reputational.

The different definition of model risk given by Rebonato (2003) requires, at least at first sight, a different approach to model validation. Large losses with reputational damage emerge when a sudden gap opens between market price and model booking. This can happen for three reason:

1. The reason can be that we were using a model different from the market consensus, and when we are forced to compare ourselves with the market – because of a transaction or because the market consensus has become visible – this difference turns into a loss. From this comes the first prescription of the Price approach, given strongly in Rebonato (2003), to gather as much information as possible on the approach currently used by the majority of the market players. This can be done through different channels. We follow Rebonato (2003) and we add some more of our own, which have become more important after Rebonato's paper was written.
   A. Some channels are based on the idea that if we can observe prices from counterparties, then we can perform *reverse-engineering* of these prices, namely we can understand which models were used to generate them. Examples of how this can be performed are in Chapter 2, in Section 4.1 and throughout the book. How can we collect counterparty prices when the market is not liquid?
      - getting as much information as possible about the deals which are struck in the market or other closeout prices such as those for unwindings and novations.
      - analyzing the collateral regulations with counterparties. Collateral is the amount of guarantees (usually cash) exchanged between banks in order to protect the reciprocal

[1] Some could argue that losses may arise, even if we use the same model used by the market, from the fact that we are hedging with an unreasonable model. We discuss similar issues in Chapter 5, where we will see that the above argument has some solid foundations, but also that real hedging strategies do not follow strictly model assumptions, so that it can be difficult to quantify the hedging losses due to unreasonableness of a valuation model. According to Rebonato (2003), in any case, losses incurred because of an 'incorrect' hedging strategy are unlikely to be of such magnitude to have a major impact, and thus should not be the focus of model risk management. More recently, Nawalkha and Rebonato (2011) points out that when a derivative is hedged, losses due to model errors may cancel out, at least in part, between the derivative and the hedge.

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

exposures from counterparty risk. The amount of collateral must be kept equal to the expected discounted exposure, that corresponds approximately to the price of all deals existing between two counterparties. We can observe this frequent repricing from our counterparties, in some cases also specifically for a single deal, to get information on the models they use.

- monitoring broker quotes (that usually do not have the same relevance as prices of closed deals) and consensus pricing systems such as Mark-it Totem. This is a service that collects quotes from market operators on a range of different over-the-counter derivatives, eliminates the quotes that appear not in line with the majority, and then computes an average of the accepted quotations. The market operators whose quotes were accepted get informed about the average. There are derivatives for which this service provides a very relevant indication of market consensus. Today, this is considered an important source of market information.

B. A few channels suggested by Rebonato (2003) regard gathering *market intelligence* by
- attending conferences and other technical events where practitioners present their methodologies for evaluating derivatives.
- asking the salesforce for any information they have about counterparty valuations. Additionally, salespeople can inform us if the prices computed with our models appear particularly competitive in the market (are we underestimating risk?) or are regularly beaten by competitors' prices (are we being too conservative?).
- Rebonato (2003) says finally that 'contacts with members of the trader community at other institutions are invaluable'. We can rephrase it, less formally, as follows: keep in touch with your college mates that work in other banks and make them speak out about the model they use at the third pint of beer at the pub.

2. If, thanks to any of the above channels, we are confident that we are using the same model prevailing in the market and this model is not changing, the only cause for large gaps between our booking and market prices can be the model/operational errors like software bugs or errors in describing the payoff. Therefore these errors must be avoided.
3. The two points above do not appear to help us in the past examples of big market losses. In 1987 there appeared to be a market consensus on the use of something similar to the Black and Scholes formula to price equity derivatives. After the market crash in October 1987 the pricing approach changed dramatically, with a clear appearance of the smile. The market consensus had moved from a lognormal model to some approximation of a model with fat-tails, may it be a stochastic volatility model or a model admitting jumps, and this was a big source of losses. Those that had sold out-of-the-money puts for nothing had to book a loss not only because of the fall of the underlying, but also because the volatility used by market player to evaluate them became much higher than the one used for at-the-money options. Even following the above points 1) and 2) of the Price approach, we would have been completely exposed to such losses. Similar market shifts in the pricing approach to interest rate derivatives characterized the aftermath of the LTCM crisis in 1998. And we have recently experienced the most dramatic event of this type with the subprime crisis and the fall of the Gaussian copula based pricing framework for CDOs. This gives the third way in which we can have a large gap between the way we were pricing and the market price: even if we are using the market consensus model, the market consensus can suddenly change. This issue is taken into account by Rebonato (2003) that, after presenting knowledge of the market approach as the first task of a model risk manager, adds that *'the next important task of the risk manager is to surmise how today's accepted pricing methodology might change in the future.'*

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

There is actually one fourth case that I have to add, but that we can resume in the third one. It is the case when our market intelligence research reveals that there is no model consensus in the market, a case that we analyze in Chapter 2. Also in this case the diligent risk manager will try 'to surmise' which future consensus will emerge. Some other indications on how to behave in this case are given in Chapter 2.

Now the crucial question that a model risk manager surely will ask is: how the hell can we surmise or anticipate how the market model is going to change? Can we devise some patterns in the dramatic changes in model consensus that have led to big model losses? It is unavoidable to start our analysis of this point from the recent, still hurting credit crunch crisis. In the following account I do not minimally try to be exhaustive in describing the reasons and the mechanism of the crisis; with the amount of books and papers written on this that would be indubitably redundant. I will try instead to focus only on the modelling aspect of what happened in 2007, and in doing this I will try to single out what I find are the most relevant elements.

1.1.3 A Quant Story of the Crisis

Let us recall what was the situation before the subprime crisis burst. An efficient market intelligence would have revealed that there existed a consensus, agreed upon at least among the most active market participants, over the pricing of these credit derivatives where the crisis burst first.

Rating agencies and banks used the Gaussian copula, that we resume here and analyze in detail in Chapter 3, for computing the prices of bespoke CDO's. For the few that, during the crisis, were not able to learn what CDOs are, see the next section. We call 'bespoke' those CDO's which are written on a portfolio of credit risks whose correlations are not liquidly traded. The predominant mass of CDO's, including mortgage-backed CDO's, were bespoke. While the Gaussian copula was used by the majority of players, there were differences in the computation of correlations. Rating agencies computed correlations historically while banks had a mixed approach. On one hand they kept an approach consistent with rating agencies since they needed ratings to market their products, on the other hand they often performed mark-to-market of CDO's by a Gaussian copula with a correlation smile given by some mapping approach that will be explained in Section 3.5.

The modelling frameworks used made it almost always possible to extract from a portfolio of defaultable mortgages a good size of senior and mezzanine CDO tranches (explained below) whose risk was evaluated to be quite low, allowing in particular to give high ratings to these tranches. Senior and mezzanine tranches had been at the heart of the expansion of portfolio credit derivatives before the crisis, and then they were the first market where the crisis burst. The optimistic judgement on the riskiness of these products was crucial to fuel the growth of their market. In fact, the underlying mortgages generated a high level of returns, which kept the spread paid by these products much higher than a risk-free return (even 200bp over Libor for AAA securities) in spite of the low risk certificated by high ratings. This correspondence of high returns and low certified risk made the products very attractive.

In the following Section 1.2.1 we explain better how the demand and supply markets for CDOs were composed, which provides an even better understanding as to why rating was a crucial element for the investment choices of funds and also banks. There we tackle also another issue that you may have already heard of: did rating agencies and banks really believe the above picture? The issue is tricky. It may be that the modelling framework we are going to present was so much liked in the market because, by minimizing the risk of CDO's, it matched well the distorted perception of risk of some operators with an artificially short-term investment horizon, like those we will see in Section 1.2.1. More likely, there were surely bona-fide players that truly believed the optimistic picture (I have met some of them), there were some others that were bending models to their purposes, and a large mass of operators that did not have elements to make an informed judgement and followed someone else's advice.

Here this is of limited relevance to us, because what counts was that there was a consensus on the modelling assumptions for valuation. This model consensus was followed by the active operators and as such it protected those using it from model losses, as noticed by the Price approach to model risk, no matter if the model was right or wrong, believed by all players or not. The losses arose when the model consensus fell, and the causes of this fall we are going to study, to understand how the market consensus on a model can suddenly change.

The pre-crisis market model

CDO's are derivatives where one party buys protection from the losses that can arise from a portfolio of investments, for example mortgages, while the other party sells protection on these portfolio losses. What makes them special is that here *the loss is tranched.* What does it mean? If we buy protection on the tranche with *attachment point A* (for example 3% of the total notional) and *detachment B* (for example 6%) we only receive the part of the loss that exceeds A and does not exceed B.

For a portfolio with 100 mortgages in it, all for the same notional and with zero recovery in case of a default, the situation of a buyer of the above 3%–6% tranche is as follows (notice that the buyer of a tranche is the protection seller). Until the first three defaults in the portfolio, he suffers no losses. At the fourth loss, namely when total losses have just exceeded its 3% attachment point, he loses $\frac{1}{3}$ of the nominal of its tranche. He will lose another third at the fifth loss, and the last third at the sixth loss, when its 6% detachment point is touched. From the seventh default on, he will lose nothing more. He has already lost everything. For him the best situation is when there are either 1 or 2 or 3 defaults, because he loses nothing, and the worst situation is any in which there are 6 or more defaults, because in this case, irrespective of the precise number of defaults, he has lost everything.

Such a tranche was often called 'mezzanine', since it has intermediate position in the capital structure. A tranche $0\% - X\%$, that suffers the first losses, is called an equity tranche for any $X\%$, while a tranche positioned at the opposite end, $X\% - 100\%$, of the capital structure is called a senior tranche. Also tranches that were intermediate but with sufficiently high attachment and detachment were usually called senior.

The expected loss for an investor depends on the correlation assumed among the default events. Let us consider an investor that has sold protection for a nominal of 100, first on the most equity tranche possible, the $0\% - 1\%$, with maturity of 5 years. We suppose that all the mortgages have a $20\% = 0.2$ probability to default within 5 years, and they have a 1, or 100%, default correlation ρ. In the market standard, that will be fully understood (including its tricky and misleading aspects) in Chapter 3, a 100% default correlation means, in this case, that all mortgages will default together. What is the distribution of the loss in 5 years?

$$Loss_{0\%,1\%} = \begin{cases} 0 & \text{with 0.8 probability} \\ 100 & \text{with 0.2 probability} \end{cases},$$

so that the expected loss

$$\mathbb{E}[Loss_{0\%,1\%}|\rho = 1] = 20.$$

If instead we say there is zero default correlation, then the one-hundred default events for the one-hundred mortgages are independent. Now the probability of having zero defaults is $0.8^{100} \approx 0$, so that

$$Loss_{0\%,1\%} = \begin{cases} 0 & \text{with 0 probability} \\ 100 & \text{with 1 probability} \end{cases},$$

leading to

$$\mathbb{E}[Loss_{0\%,1\%} | \rho = 0] = 100.$$

Take instead a protection sale on the most senior tranche, $99\% - 100\%$. Under correlation 100%, the distribution of the loss is

$$Loss_{99\%,100\%} = \begin{cases} 0 & \text{with 0.8 probability} \\ 100 & \text{with 0.2 probability} \end{cases},$$

so that

$$\mathbb{E}[Loss_{99\%,100\%} | \rho = 1] = 20.$$

If instead we say there is zero default correlation, now the probability of having 100 defaults is $0.2^{100} \approx 0$, so

$$Loss_{99\%,100\%} = \begin{cases} 0 & \text{with 1 probability} \\ 100 & \text{with 0 probability} \end{cases},$$

$$\mathbb{E}[Loss_{99\%,100\%} | \rho = 0] = 0.$$

We can notice first that an equity tranche is more risky than a senior tranche. They are the same for 100% correlation, but for all lower levels of correlation the senior tranche is less risky. Then we notice that for equity tranches risk decreases with correlation, while for senior tranches risk is increasing with correlation from almost no risk at 0 correlation up to maximum risk at unit correlation.

Now we give a rough description (improved in Chapter 3) of the market model for these derivatives, trying in particular to explain how this modelling framework allowed regularly to extract from a bunch of mortgages a number of tranches with low risk.

The market model was made up, following the approach of a Gaussian copula, by default probabilities and correlations. The historical approach, favoured by rating agencies, based the correlations on observing past data and extrapolating some conclusions from it. The mapping approach, often used by banks and funds, was based on a modification of today correlations from some reference markets which are very different from and much less risky than the bespoke CDOs to which it was then applied. We will show in 3.5 that this approach, which was supported by mathematical considerations with very little financial justifications, was biased towards underestimating the correlations of subprime CDOs and in general of all CDOs more risky than the reference markets. This bias was not immediate to detect, probably because of the lack of transparency and intuitiveness of the methodology. We have included the unveiling of this model error in Chapter 3 devoted to stress testing of models.

In this section we focus instead on the historical estimation approach, because it was this approach, used by rating agencies, that led to those favourable ratings which were the crucial driver of the market growth. And it was the break-down of this historical approach that then ignited the crisis. The users of this approach took as an input the historical default rates of mortgages, divided into the national rate and the regional rates, which were often rather different from the national one. From these data they could compute the correlation among the default events of the different borrowers. The historical evidence was that subprime borrowers, that are known for being unreliable, defaulted most often for their personal financial problems, with a low dependence on the regional trend of the economy and an even lower one on the national trend. The historical evidence on the default of subprime mortgagers, formally organized as in Remark 1, was the foundation of the tendency to give low correlation to subprime mortgagers, reducing the riskiness of senior tranches in subprime CDO.

In the years preceding the crisis, someone suspected that this model may not be anymore reasonable for the current times. In fact during the first decade of this century the number of subprime mortgages had been increasing, while the losses on them had been low, and this was due to a fact not taken into account by historical data. During the entire decade house prices had been increasing, and the evolution of the financial system had made it easy to perform *equity withdrawals*, which means the mortgager getting cash from an increase in the price of his house, without selling it. The simplest way for a mortgager to get this is to refinance his debt. If I bought a house for $100.000, using a $100.000 mortgage guaranteed by my house, but after one year my house is worth $150.000, I can go to another bank and get a $150.000 mortgage guaranteed by my house. I can use $100.000 to extinguish the previous mortgage and spend the remaining $50.000, including paying regularly the interest on my mortgage. Clearly at the same time I have also increased my total indebtment, increasing in the long or medium run my risk of default.

Why were banks or mortgage companies happy about this? Again, because of the increasing house prices: mortgage lenders that, with a default, became proprietors of a house with a price higher than the value of the mortgage, and easy to sell, can have a gain and not a loss from the default. This led to an expansion of mortgages, that in turn sustained the increase of house prices on which the mortgage expansion itself was based.

It is clear that the picture was changed by the new situation: now the fact of having losses on mortgages depended crucially on the trend of house prices, since as long as the trend is increasing losses are less likely. This should alter also our reasoning on correlation, since the dependence on a common trend creates stronger correlation. If the real reason that made the market function is the one we described above, a generalized decrease in house prices should first of all create problems to refinance the debt for all mortgagers, increasing the probability that they default together, and secondly, after a default, it increases the probability that these defaults generate losses due to lower house prices. Rating agencies knew this somewhat, but this did not change dramatically their correlation assumptions: the large number of AAA ratings remained. This is justified as follows by Brunnermeier (2009):

'Many professional investors' statistical models provided overly optimistic forecasts about structured mortgage products for a couple of reasons: 1) they were based on historically low mortgage default and delinquency rates that arose in a credit environment with tighter credit standards, and *2) past data suggested that housing downturns are primarily regional phenomena—the U.S. had never experienced a nation-wide housing slowdown. The seemingly low cross-regional correlation of house prices generated a perceived diversification benefit that especially boosted the evaluations of AAA-rated tranches.*'

The rating agencies followed again historical observations, and they noticed that, at least in the recent past considered, 'the U.S. had never experienced a nation-wide housing slowdown'.

This is the crucial observation, together with the other 'housing downturns are primarily regional'. House prices had gone down in single states, but then, when looking at the national numbers, the house prices had never decreased during the historical period used for evaluating CDO's. Thanks to this evidence, the correlation was increased only for mortgagers belonging to the same state, but not for mortgagers living in different states. Since the CDO's designed by banks tried to pool together names coming as much as possible from different states, the rating agency models gave low correlation to the names in the pool, making senior tranches deserve a high rating.

Thus for the first approach that rating agencies had used in the past correlation of subprime was low since subprime are based mainly on idiosyncratic risk. For the more up-to-date model, that took into account the link between subprime losses and house prices, the crucial implicit assumption justifying low correlation was in assuming that the national house trend can only be increasing, what Oyama (2010) calls the system of *loans with real estate collateral based on the myth of ever-increasing prices*.[2]

What happened to this myth in 2007? If you want a more detailed technical account of the modelling framework used by agencies, banks and funds to compute correlations, you can read the following remark. Otherwise, you can get directly to the answer in the next section.

Remark 1. ***Technical Remark on Factor Models.*** *Rating agencies were using factor models, where the default time τ of a mortgager happens before the mortgage maturity* T *in case a standardized Gaussian random variable*

$$X \sim N(0, 1)$$

is lower than a threshold H*,*

$$\Pr(\tau \leq T) = \Pr(X \leq H) = \Phi(H),$$

where Φ is the cumulative probability function of a standardized Gaussian distribution, so that once $\Pr(\tau \leq T)$ *has been estimated we can say that default happens before maturity when*

$$X \leq \Phi^{-1}(\Pr(\tau \leq T)).$$

This model lacks in any real dynamics, in the sense that with such a model one can find only silly answers to questions such as: given that the mortgager has survived until a date T_1 in the future, what is the likelihood that he will survive until $T_2 > T_1$? But we will leave this aspect to Chapter 3, when we analyze the liquidity problems that followed the subprime crisis and the difficulties to deal with them using this model. For the time being, we focus on the fact that the variable X *is the one through which these models can capture the dependency, and therefore the correlation, between the default times of different mortgagers. They assume that for the mortgager '*i*' of the state '*a*' we have a factor* X *shaped as follows*

$$X_i = \gamma_{US} Y_{US} + \gamma_a Y_a + \gamma_i Y_i$$

where γ_{US} is the factor which is common to all mortgagers in the US, γ_{a} is a term common to only the mortgagers in state a *and independent of the US factor, and* $\mathrm{Y_i}$ *is an idiosyncratic*

[2] Once, during a workshop on the crisis, when we arrived at this point one guy, anticipating that this low correlation was what in the end turned out wrong, told me: 'You see? The problem were not wrong models, but wrong data!'. I have already expressed what I think in the preface: I find this comment paradoxical, particularly when coming from a modeller. The data in this case were historical data about house prices, and they were actually right. What was wrong was the choice to extrapolate the recent past trend to the future, without introducing in the model the likelihood of an important deviation from it. This was absolutely a modelling choice!

factor that takes into account the probability that mortgager i *defaults independently of the trend of the national or regional economy. The loadings* γ_{US}*,* γ_a *and* γ_i *are the weights of the three different terms. If we believe that the dependency on the national factor* γ_{US} *is the crucial one, we are going to have*

$$\gamma_{US} \geq \gamma_a, \gamma_i,$$

if instead we believe that the mortgagers usually default for their personal reasons, we are going to set

$$\gamma_i \geq \gamma_{US}, \gamma_a.$$

It is logic to take these three factors independently. In fact if there are links between the different states in the US, this will be captured by a higher weight γ_{US}*, while if there is a link between mortgagers in the same state* a*, this will be captured via a higher weight* γ_a*. Notice that if we take the three factors* $\mathrm{Y_{US}}$*,* $\mathrm{Y_a}$ *and* $\mathrm{Y_i}$ *to be all standardized Gaussians* $\mathbb{N}(0, 1)$*, and we set*

$$\gamma_{US}^2 + \gamma_a^2 + \gamma_i^2 = 1$$

we have kept the property that $\mathrm{X_i}$ *is* $\mathbb{N}(0, 1)$*, in fact*

$$\mathbb{E}[X_i] = \gamma_{US}\mathbb{E}[Y_{US}] + \gamma_a\mathbb{E}[Y_a] + \gamma_i\mathbb{E}[Y_i] = 0$$

and

$$Var[X_i] = \gamma_{US}^2 Var[Y_{US}^2] + \gamma_a^2 Var[Y_a^2] + \gamma_i^2 Var[Y_i^2] = 1.$$

The interesting thing for us is that this factor model also determines the correlation between the default risk of different mortgagers. In fact for two mortgagers i *and* j *belonging to the same state* a *we have*

$$\begin{aligned} Corr[X_i X_j] &= \mathbb{E}[X_i X_j] \\ &= \mathbb{E}\left[(\gamma_{US}Y_{US} + \gamma_a Y_a + \gamma_i Y_i)\left(\gamma_{US}Y_{US} + \gamma_a Y_a + \gamma_j Y_j\right)\right] \\ &= \gamma_{US}^2\mathbb{E}[Y_{US}^2] + \gamma_a^2\mathbb{E}[Y_{US}^2] = \gamma_{US}^2 + \gamma_a^2. \end{aligned}$$

while if the two mortgagers belong to two different states a *and* b *we have*

$$\begin{aligned} Corr[X_i X_j] &= \mathbb{E}[X_i X_j] \\ &= \mathbb{E}\left[(\gamma_{US}Y_{US} + \gamma_a Y_a + \gamma_i Y_i)\left(\gamma_{US}Y_{US} + \gamma_b Y_b + \gamma_j Y_j\right)\right] = \gamma_{US}^2. \end{aligned}$$

The historical evidence was that subprime borrowers had a low dependence on the regional trend of the economy and an even lower one on the national trend. Thus γ_{US} *and* γ_a *were low, leading to low default correlation. Then the importance of the trend of house prices became more important: the effect of the national economy on the probability of default of a mortgager was through the possibility that national house prices went down; the effect of the regional economy was through the possibility that house prices in a given state went down. Then there was the residual factor* $\mathrm{Y_i}$ *associated to the classic default risk of a subprime: he loses his job and after his default it is difficult to sell his house, independently of the trend of the housing market. Inspired by the above historical evidence, analysts took* γ_{US} *to be very low, since the national housing trend had always been increasing and could not be a reason for defaults. The* *dominant factor was the state factor γ_a, since state housing trends can turn from increasing to decreasing and a decreasing trend can lead to default. Thus they had a very low correlation for names belonging to different states, and a higher one for names belonging to the same state, getting low correlations for CDOs diversified across states, as most CDOs were. We are back to the crucial question: what happened to the myth of ever-increasing national house prices in 2007?*

The strike of reality

We follow Brunnermeier (2009), an early but very accurate account of the preparation and burst of the crisis, along with some other sources, to describe those seemingly minor events in 2007 that had such a strong impact on the future of finance.

An increase in mortgage subprime defaults was registered as early as February 2007, but it seemed a temporary slow-down with no consequences. However, something else happened in March. From the *Washington Post* on 25 April 2007, we read that sales of homes in March fell 8.4 percent from February, the largest one-month drop since January 1989, when the country was in a recession. Operators tried to play down the relevance of such figures. David Lereah, chief economist for the Realtors group, attributed the downturn partly to bad weather in parts of the country in February that carried over to transactions closed in March.

But there was something more relevant. The median sales price fell to $217,000 in March, from $217,600 in March 2006. It is a very small decrease. But in light of the above analysis it is easy to see just how disquieting it must have appeared to operators in the CDO market. The situation became even worse later, and did not only concern 'houses', but real estate in general. Figure 1.1 illustrates the dramatic reversal of the price trend in the crucial sector of commercial property, that also happened around the summer of 2007, with some early signs in the preceding months.

Many operators in those days seemed to change their minds about the prospects of the market. UBS shut down their hedge fund focused on subprime investments. Moodys put a number of tranches on downgrade review: while not yet a downgrade, it usually anticipates one. Others tried to carry on as normal. Bear Sterns in mid June injected liquidity to save one of their funds investing in subprime mortgages, that was experiencing liquidity troubles. It is interesting to note that Bear Sterns had no contractual obligation to do this, but acted to save its reputation.

From 25 to 26 July 2007 the CDX 5-year maturity index, a good measure of the average credit risk of US senior corporations, jumped by 19% from 57bp to 68bp. Nor was the reaction limited to the US. The i-Traxx 5-year maturity spreads, an indicator of the confidence of market operators on the credit perspectives of the European economy, jumped from 36bp to 44bp, a 22% increase that was by far the largest one day-jump in its history to date. For the financials sub-index, the situation was even more dramatic: the jumps was from 23bp to 33bp, a jump of 43%. From Monday, 22 July to Friday, 27 July, just one working week, the spread of financials almost tripled, from 23bp to 59bp.

It seems market operators had put two and two together. If house prices go down, mortgage equity withdrawals become difficult and defaults in the subprime markets are doomed to rise. This will result in banks and mutual funds becoming proprietors of the houses of the defaulted mortgages. In a context of falling house sales and falling house prices, this will turn into material losses.

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

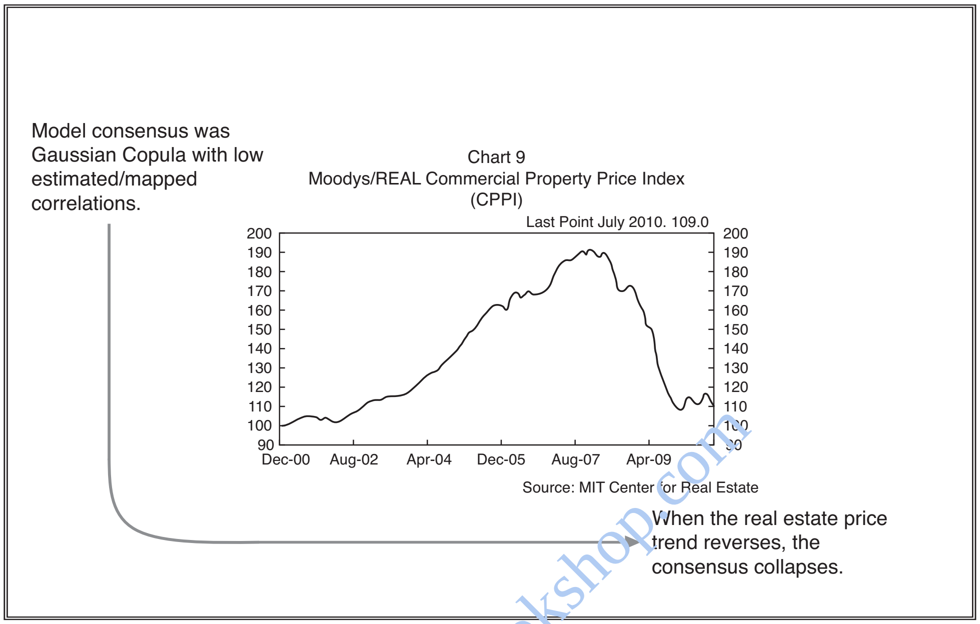

Figure 1.1 The credit crunch is the first example of model consensus collapse that we analyze

If banks can suffer losses from the default of mortgages, all the mortgage-based derivatives that were sold as virtually risk-free AAA assets will soon become junk. There are so many of them around that the whole financial system will suddenly be in big trouble, banks will have to reduce their lending and this will turn into an increased risk of default for all economic players worldwide. The decrease in national house prices shattered the foundations of a splendid, if fragile, edifice: the economic system built in the first decade of the 21st century.

The first wave of this tide hit the CDS and CDO market. On 31 July American Home Mortgage Investment Corporation announced it was unable to meet its obligations, and it defaulted officially on 6 August. Everything that followed has been recounted in hundreds of books, and we will not reprise it here. We will mention the topic again in Section 3.4, where we explain that after the subprime crisis burst and the initial clustered credit losses, these losses generated liquidity stress and a panic that exacerbated the liquidity stress. There we will show why the Gaussian copula market model is particularly unfit too for the management of the risk of clustered losses, an element that certainly did not help anticipate the real risks associated with CDO investments, nor did it help ease the panic and the liquidity crunch once the crisis had burst. But that's another story and one told in Chapter 3.[3]

[3] After this recap, one might wonder: are we simplifying how things actually went, thanks to the benefit of insight? That's for sure. But not that much: the crisis was really triggered by such a simple event as a decrease in real estate prices. There has been a lot of talk of Black Swans in the aftermath of the crisis, but the swan that hit us was a plain white swan, already seen in the not-too-recent past. What we had forgotten is that swans can migrate and stay away for some time . . . but they usually come back. Such delusions are not uncommon in finance. The 'new paradigm' of economy at the end of the 1990s predicted that we were not going to see recessions or crises again, but that also was belied by reality.

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

1.1.4 A Synthetic View on Model Risk

Let us go back to our initial question. What can trigger the market suddenly to abandon a consensus pricing methodology, as happened with the subprime crisis? The analysis of the crisis shows what happened in that case: *an event related to the fundamentals was observed*. There was a decrease in house prices at national level. This reminded market operators that the model used was crucially based on a hypothesis extremely biased towards an aggressive scenario, that of ever-increasing house prices, *that could be macroscopically disproved by evidence*. The solidity of the market could be destroyed by a generalized decrease in house prices, a scenario that had previously been considered impossible. Now this scenario was becoming a reality in the housing market. We can say that the crisis burst when *an element of unrealism of the model was exposed to be more relevant than previously thought*.

Clearly, we are speaking with the benefit of recent hindsight, but the *death of a model* seen in this crisis is a typical pattern of crisis in regard both to quantitative and qualitative models. The losses in the derivatives world in 1987 were driven in part by the appearance of the skew (a decreasing pattern of implied volatilities when plotted against the option strike), that corresponds to abandoning a Gaussian distribution of returns and replacing it with a distribution where there is more likelihood of large downwards movements of the underlying stock prices. This was clearly driven by the fact that such a large downward movement *had just happened in reality*, in the stock market crash of Black Monday, as we can see in Figure 1.2. The model change was effected without any sophistication, but simply by moving implied volatilities to patterns inconsistent with the previous Gaussian model, and it was done very fast.

Even the dot com bubble of the '90s was sustained by a sort of model, mainly qualitative but with some quantitative implications on simple financial indicators, that predicted a *change of*

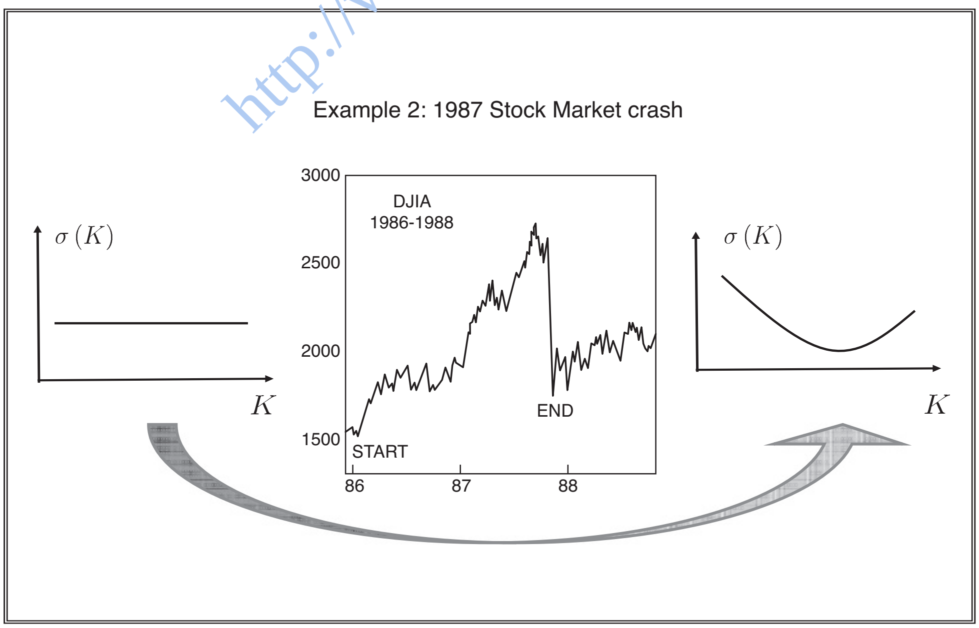

Figure 1.2 Another example of model shift is the 1987 Stock Market crash

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

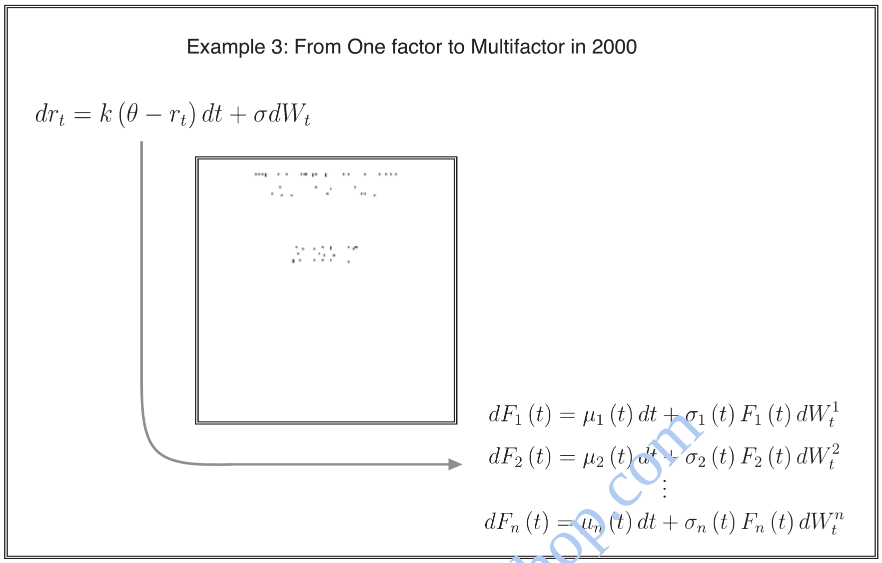

Figure 1.3 An example of a model shift triggered also by a piece of research

paradigm in the economy that should sustain *all* internet companies in obtaining performances never seen before. When the model was disproved by the reality that internet companies had started to default, the bubble burst.

Another example is the hypothesis of deterministic recovery, usually set at 40%, that was used in pricing credit derivatives before the crisis. When the credit crunch and, in particular Lehman's default, showed that recoveries with one single digit were quite likely in a time of crisis, there was a move by many players towards stochastic recovery models.

These conclusions are confirmed by an example given by Rebonato (2003) where the consensus was not changed by a crisis, but *by a new piece of research*. The paper 'How to throw away a billion dollar' by Longstaff, Santa-Clara and Schwartz, around the year 2000, pointed out that if the market term structure was driven by more than one factor, then using a one-factor model exposed banks to losses and prevented them from exploiting opportunities for profit (the issue is explained in more detail in Section 2.8.2 and in Chapter 9). The outcry caused by this piece of research was the final blow that made a majority of banks move to models with a higher number of factors. Since the number of factors driving a realistic representation of the term structure behaviour should certainly be higher than one, this market shift can also be associated with the fact that *an element of unrealism of the model was exposed to be more relevant than previously thought*.

The importance of this factor in the sudden changes of modelling consensus in the previous crises has an interesting consequence. The patterns of model changes show that also for the Price approach the points mentioned by Derman (1996) become important *'You may have not taken into account all the factors that affect valuation . . . You may have incorrectly assumed certain stochastic variables can be approximated as deterministic . . . You may have assumed incorrect dynamics . . . You may have made incorrect assumptions about relationships'*. This means that in the Price approach we also need to understand if a model is sufficiently *realistic*, or at least if it is sufficiently *reasonable* and *robust* to changes in market reality not to expose users to sudden losses as soon as a not particularly unlikely scenario turns out to be true.

This reduces the practical difference between the Price approach and Value approach. The fundamental requirement of the Value approach, that models should give *'a realistic (or at least plausible) description of the factors that affect the derivative's value'*, is also important in the Price approach. Additionally, if an institution is particularly worried about the reputational side of model losses, losses which are revealed to be based on an unrealistic model are particularly difficult to justify. This is true even when the unrealistic model has been used by the majority of players, as shown by the example on rating agencies set out at the beginning of this chapter.

The Price approach makes the fundamental contribution of pointing out an element that appeared to have been overlooked in classic approaches: the importance of understanding the modelling consensus in the market, which we cannot afford to overlook since derivatives are regularly marked to market. However, notice that this could have been done relatively easily for the CDO market before the crisis, by finding out about Gaussian copula, historical correlations and mapping, and yet if the approach had stopped there, assuming that as long as we are consistent with the market consensus we have no model risk, it would still have led to big losses. Avoiding this requires the second step of the approach, surmising how the model consensus can change, and this makes realism relevant also to the Price approach, as the main element whose lack can trigger a sudden change in the model, and as a blueprint that allows us to make an informed guess about which new model may in due course replace the old one.

On the other hand, classic approaches to model validation, focused only on using realistic rather than consensus models, would make the life of a bank rather difficult. For example, if a more realistic model had been used for credit derivatives before the credit crunch, this model would probably have been more pessimistic about the risk of CDOs. The bank using the realistic model would have waited for many years before seeing the predictions of its model come true. In the meantime, the bank would have gone against the market, accumulating mark-to-market losses.

But is it true that the Value approach would have led to such silly behaviour? Only on a superficial interpretation. The Value approach does not underline the importance of consistency with the market as clearly as the Price approach, but it does not overlook it. First of all, the requirement for good calibration is not as trivial as may appear at first sight. Remember that model validation is mainly for complex derivatives that do not have liquid observable price. What Derman says at point 3 of his prescription is that, although the price of the derivative is not observable, we observe the market prices for the simplest derivative that represent the constituent parts of that derivative, and we must be able to calibrate these prices well. For that part of the market that is observable, we must be consistent with the market at least in terms of price fitting. This is not enough to control model risk, as we show with a number of examples in Chapter 2, but it is a basic element of consistency with the market.

There is more about this in Derman (1996). Among the factors that a realistic model should consider is 'market sentiment':

> *'A model may be "correct in principle" but the market may disagree in the short run. This is really another way of saying the model is limited, in the sense that it didn't take other short-term factors into account (including market sentiment) which can influence price.'*

Thus market sentiment is a factor that should influence our choice of the model. This is logical, since it is certainly one of 'the factors that affect the derivative's value'. Additionally, we see that Derman (1996) says it is a short-term factor, that can easily change in the future.

This does not sound very different from the Price approach suggestion of taking into account the market's modelling consensus but being ready for sudden changes in this consensus.

We can now start to draw together a number of points which, according to both approaches, form the backbone of a model validation practice.

- Both approaches require a *price verification* where the mathematics and the implementation of the model are validated, together with the description of the payoff. This is explicit in the Value approach, while in the Price approach is a consequence of the need to remain in line with market prices.
- The Price approach focuses on collecting information about the market consensus, and points out for the first time the importance of this aspect. This is not, however, in contrast to the Value approach, which requires calibration to liquid market prices and takes into account market sentiment.
- Both approaches require consideration the realism and robustness of the model in different financial scenarios. For the Value approach this is the fundamental requirement. For the Price approach it is a consequence of the need to anticipate changes in market consensus.

What should happen when, as in the case we analyze in Chapter 2, we find no market consensus at the second step? In this case we have to guess which model consensus may form in the future, since if a consensus forms we may have to mark the derivative to that consensus. In this case, a model with some features of realism and some simplicity of application is the best choice we can make. It may not predict correctly the model that will emerge, but we have no elements to make any different reasonable guess. When there is no market consensus and we expect it will never form, and we also think we are going to keep the derivative until maturity (albeit this is quite a strong assumption), the focus will be especially on the realism, since what matters in such a case are physical cashflows (a default or a survival, an option ending in or out of the money, a loan prepaid or not).

Remark 2 ***Buy-side vs Sell-side.*** *The analysis of Nawalkha and Rebonato (2011) bears some relation to the trade-off between consistency with the market and realism considered here. One implication of their analysis is that this trade-off can be viewed differently by banks on the sell-side of derivatives and funds or bank treasuries or institutions on the buy-side. On the sell-side we have players who regularly hedge most of their derivatives. In this case, there will be more focus on consistency with the market consensus model – in particular the model consensus for the pricing of the hedging instrument – and on computational efficiency, since these two features allow an efficient hedging. Such efficient hedging will allow in turn some partial cancellation of a model's unrealistic features between derivative and hedges. This corresponds to the requirement for calibration that we will see in our scheme for model validation and risk-management in Section 1.5. On the other hand, as pointed out in Nawalkha and Rebonato (2011), these unrealistic features still generate model risk due to the 'basis' between hedges and derivative, which they say is associated with issues such as correlations or time-dependencies of parameters, at least in interest rate derivatives. We will add to this later in this chapter where we see that some apparently efficient hedges break down when model consensus collapses due usually to its unrealistic features, exposing also the sell-side to large model losses, and in Chapter 5 where we discuss correlation risk in hedging, and in Chapter 9 where we consider time-dependency of parameters.*

Those who do not perform very active hedging, either because they are on the buy-side or simply because, as happens for many complex derivatives, efficient hedging is not possible, will *have an even higher focus on the realism of model features, such as volatility mean-reversion in Nawalkha and Rebonato's example.*

Remark 3 ***Is a realistic model always a sophisticated model?*** *There is one more interesting comment by Rebonato (2003). It regards what happened at the end of the 1998 Russian crisis* 'Anecdotal evidence suggests that, for instance, after the liquidity crunch associated with the Russia crisis, traders reverted to pricing Bermudan swaptions with cruder and more conservative models, such as the Black-Derman-and-Toy, rather than with the indubitably more sophisticated and realistic Libor Market Models'. *Is this a case where there was a shift in the market consensus towards a* less realistic *model? Not quite, for at least two reasons. First, this example does not appear a change in the 'market consensus model'. When the Russian crisis burst only a minority of advanced banks had already adopted the Libor Market Model, while the models used by the majority of the players for Bermudans were still short-rate models like Black-Derman-and-Toy. In fact the Russian default/restructuring was announced on 17 August 1998. As Rebonato (2003) recalls, one crucial driver of the generalized shift towards the Libor Market Model is the aforementioned paper* 'How to throw away a billion dollars' *by Longstaff et al., whose first version appeared in January 1999. This is after the Russian crisis. The same paper pictures the situation as it was at that time:* 'extensive conversations with numerous brokers and dealers indicate that most Wall Street firms use some form of a single-factor Black-Derman-Toy model in valuing their swaption positions and making exercise decisions. For example, the Black-Derman-Toy model is the default valuation model for American-style swaptions in the widely-used Bloomberg system'.

What happened in the aftermath of the crisis was not a reversal of a market consensus about LMM, a consensus that did not exist yet, but rather the retreat of some sophisticated banks that had decided to go 'against the market', adopting early the LMM and betting that the market would soon be forced to follow them. The bet failed and the retreat was towards the low-factor models that had remained the market consensus. The market consensus actually shifted some time later, and it was towards the more realistic LMM, as confirmed by Rebonato (2002a), who also associates this shift with the introduction of advances in Monte Carlo simulation for Bermudans, whose most popular example is the Least Squares Monte Carlo that only became known to most banks around 2000.

It is now time to note that, when the market shifts consensus in a crisis, the move is rarely towards more sophisticated and realistic *models, but rather towards* less sophisticated and more realistic *models! These models underlie assumptions that appear more realistic in the light of recent events, but the implementation needs to be simple: in a tempest one shelters on a rough but solid rock rather than wasting time trying to build a light hypermodern yacht. There are a number of examples: the appearance of the smile in 1987 disrupted the Black and Scholes model with no additional model complication, the abandonment of the CDO market model in 2008 was only carried out by a change in the way of computing the parameters (correlation and recovery).*

It is interesting to recall also the most relevant example related to the aftermath of the Russian crisis. We follow again Rebonato (2002) who notices, 'the Russia default and the near-collapse of LTCM in 1998 brought about unprecedented dislocations in several markets. The interest rate volatility surfaces were not immune from these upheavals'. *The events illustrated dramatically how unrealistic a deterministic volatility model can be, and the smile made a clear appearance in the interest rates market. Players clearly desired models incorporating stochastic volatility, but when did the market really introduce, even simply for quotations, a*

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

stochastic volatility model that became part of market consensus? It was only a few years after the crisis, when a tractable stochastic volatility model was developed, the SABR model introduced by Hagan et al. (2002). The shift happened with little increase in implementation complexity, since the SABR model is just an elaborate way of computing a strike-dependent implied volatility to be put in a Black formula, as we will see in Chapter 6 and Chapter 7.[4] *This is a perfect example of how a model can evolve towards a more realistic model that allows better understanding of the market without excessive computational complications. The most likely model towards which the market will turn is the one which is more realistic without dramatically increasing computational complexity.*

1.2 FOUNDATIONS OF MODELLING AND THE REALITY OF MARKETS

After this introduction into the concept of model risk, we will go into the details of model risk management and provide a number of examples. In order to do so, we now need to review the foundations of the mathematical models used for pricing, and to define precisely the different components of the standard way to compute prices. In the meantime, we will see the relation between the standard pricing procedure and the reality of financial markets.

1.2.1 The Classic Framework

Consider a financial product that pays X_T at time T. The price of this derivative is formally written as

$$\Pi_t = \mathbb{E}^Q \left[D(t,T) X_T \,|\, \mathcal{F}_t \right] \tag{1.1}$$

There are a number of symbols to describe here.

Risk-free discounting

First, $D(t, T)$. This is the discount factor, so that

$$D(t,T)X_T$$

is the discounted payoff. Usually $D(t, T)$ is written mathematically as

$$D(t,T) = e^{-\int_t^T r(s)\,ds} = \frac{B(t)}{B(T)}.$$

The quantity $B(t)$ is the value at t of the *money market account*, the mathematical representation of a bank account. We assume that the amount of money in a bank account grows every instant with continuous compounding at a rate of interest $r(t)$ called *instantaneous spot rate* or *short rate*. Formally, the equation for its evolution is then

$$dB(t) = r(t)B(t)dt \tag{1.2}$$

[4] Where, incidentally, we will also see how, many years later, SABR, by then a consolidated market consensus, was shaken by new market evidence.

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

so that, by solving the very simple *deterministic* equation (1.2), we find out that if we invest 1 at time 0 we have at t

$$B(t) = e^{\int_0^t r(s)\,ds}.$$

We say that the equation (1.2) is deterministic since we do not have, as happens for example with the equation for the value of a stock in the Black and Scholes model, a term like

$$\ldots + \sigma B_t dW_t.$$

We have no dependence on the instantaneous movement of a random process dW_t. Although interest rates can be stochastic in our models, in which case $B(t)$ is not already known at time 0, its movement in the next instant, given by (1.2), is deterministic. We do not expect it to be modified by some sudden arrival of financial news, that would be represented by a dW_t. This is the reason why the money market account is also called the *locally risk-free asset* ('locally' means: over the next instant). It does not have a volatility σ.

This is a relatively realistic representation of a real-world bank account. Real interest is not paid or accrued instantaneously, but in any case quite frequently. And we do not expect such sudden variations in the value of a bank account as we would with a stock price. For simplicity, in the following reasoning suppose that interest rates are deterministic.

We assume that with this bank account we can both invest the money we earn and fund ourselves by withdrawing from the account. We invest and borrow at the same rate $r(t)$, which may be unrealistic for a household, but less so for a trading desk in a bank. Since all the money that we save for some time, or that we borrow for a period, goes into this account, the way its value changes over time tells us what for us is the *time-value* of money. When we know the time-value of money, we can build a discount factor, that expresses the quantity X_T received at T in the equivalent amount of today's money. The discount factor is given by the ratio between the value of the bank account at t, and its value at T, so, as we wrote above

$$D(t,T) = \frac{B(t)}{B(t)} = e^{-\int_t^T r(s)\,ds}.$$

This representation does not take into account the reality of credit risk and liquidity risk that can affect money lent by one bank to another bank, as we will see in Chapter 4, but these two risks can be more or less neglected in times far from crisis, making this representation of discounting sufficiently realistic to be viable for many years. Even in the aftermath of the credit crunch this representation is still valid for derivatives that are collateralized through a CSA agreement with collateral paying overnight interest, that we consider in Chapter 4.

Risk-neutral expectation

Now let us analyze the other unusual terms in the above expression. You are probably already familiar with all of them, but before proceeding to an in-depth analysis of our models and their application, it is best that we agree on the fundamental definitions. The symbol

$$\mathbb{E}^Q[\cdot|\mathcal{F}_t]$$

indicates *expectation*, using the available *information at t*, indicated as $\mathcal{F}_t$, under a *risk-neutral* or *risk-adjusted probability measure Q*.

Usually the formal framework is a triplet called a probability space

$$(\Omega, \mathcal{F}, Q)$$

where Ω is the set of all possible future scenarios or states of the world ω, $\mathcal{F}$ is the set of all possible claims that we can make about the state of the world – called events – and Q is the probability measure used to give a probability to such events. Thus Q is a set of probability distributions for the different objects in the market, a specification of their rates of growth, volatilities, dependencies, possibly probability of jumps etc. Namely, Q is the hero, or the villain, of this book: it is *the model*.

We will talk about *it* at length so now let us concentrate on the other word we have used to define Q. Q is not only a probability measure. We have said it is a *risk-neutral* or *risk-adjusted* probability measure. You may well know a lot about it, but I have read so many misconceptions on the subject, particularly after the subprime crisis, that I think it useful to go over this concept, albeit in a simple way.

Why on earth should we consider a risk-neutral probability measure? The intuition is dramatically simple. In the above expression we have written that a price is an expectation. This should set alarm bells ringing because if the expectation was taken under the *real-world probability measure* P – the probability measure that we consider when we express our usual probabilistic views, such as 'there is 1 in 6 probability that a roll of a dice gives you a 6' – this way of computing prices would be completely wrong. In fact, it would mean evaluating a security based only on its expected rate of return, and this leaves out something relevant. What?

Say, for example that X_T is the return promised by an internet start-up at the end of the '90s. We may have been told that according to analysts the expected return is 50%. The market expected very high returns from such companies at the time of the dot com bubble of the late '90s and in some Google-style cases they were right. But this 50% expected return may come out of two equally likely scenarios; in one scenario the return will be 125%, in the other it will be −25% (better in any case than most realized returns on internet companies other than Google). The expectation is

$$\frac{1}{2}125\% - \frac{1}{2}25\% = 50\%,$$

as in a fixed-return investment paying you 50% guaranteed, but the two investments are not the same.

In the first case there is a *risk*, and you should take it into account, because, unless you are a very special investor, you will prefer a 50% return guaranteed to the internet company. A normal investor in the financial market is averse to risk, and demands a premium for every risk it takes. Standard expectations do not take risk into account, since for a standard expectation the internet company and the fixed investment are exactly the same. In order to avoid this we consider a special expectation, the risk-adjusted or risk-neutral one.

There may be different ways of taking risk into account, for example we could use a higher discount rate $r^{Risk} > r$ for risky products, or we could compute an expectation and then subtract some term that expresses risk. Quants preferred instead to keep the simplicity of an expectation with no modified discounting and no additional terms, but they understood that they had to correct somehow the real world distribution P to take into account risk; in particular they understood that, due to risk aversion, they had to *reduce* expected rates of return compared to those in the real world. This is the meaning of setting up a risk neutral probability measure. Far from being a way to 'neglect risk', as I once read in the financial press, risk

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

neutral valuation is a way of taking risk fully into account, to *reduce* the value of securities, and certainly not to inflate it.

In the standard simplified approach, risk is usually measured by the *volatility* of an asset. How much should we reduce rates of return to take volatility into account such that an expectation is correctly reduced by the consideration of risk? First answer this question: what is the only security whose distribution is not affected by this change of probability to take risk into account?

The bank account $B(t)$ of (1.2), because it has no volatility and thus no risk. If you try to modify its rate of return to take into account its own risk, you realize that there is nothing to modify, because its risk is zero. It evolves at the spot rate $r(t)$ under the real world probability measure and also under the risk-adjusted probability measure.

Now, if we want this fact to tell us something about the general features of the risk-neutral distributions, we have to add two subtle assumptions, that we will discuss below.

No-arbitrage pricing

The first assumption is *absence of arbitrage opportunities*. The intuition is that an arbitrage is a free lunch, an investment strategy that costs nothing, has no possibility of generating losses, and a positive possibility of generating gains. To give a simple example, we have an arbitrage opportunity when two *tradable* payoffs which are exactly the same have two different prices. This inconsistency can always be transformed into a free lunch: we buy it where it costs less and immediately sell it where it costs more. If the market is arbitrage-free, this must be impossible. In the simplified market representation of standard quantitative finance, where assets can be described only based on their expected rate of return and their risk, when two *tradable* investments have the same risk and different returns we have an arbitrage. If the market is arbitrage-free, this too must be impossible.

Under a risk-neutral probability measure, returns have already been diminished to neutralize risk, so all assets can be treated as if they have zero risk. But if the risk is zero for all tradable assets and the market is arbitrage-free, then the expected rate of return must be the same for all tradable assets. We know that one of them, the locally riskless bank account, has an expected rate of return equal to $r(t)$. The other assets must have the same expected rate of return, therefore they also must have a rate equal to $r(t)$.

Above we have used two assumptions: the first assumption has been stressed explicitly, and it is absence of arbitrage. The second assumption was more covert: for the absence of arbitrage to have an effect on the securities, we have required that the securities be *tradable*. If they are not tradable, they cannot be used to build an arbitrage opportunity, therefore they are not forced to be consistent with the locally riskless asset. We analyze arbitrage first, and will return later to the other assumption of tradability, that with a term more common among practitioners that we may call liquidity while academics often call it completeness.

Is assumption of absence of arbitrage opportunities a reasonable assumption? Lack of arbitrage opportunity should be expected if all operators in the market were rational, well informed and committed to exploit all opportunities for profit. Experience tells us that many investors are not rational in this way, use very partial information and are guided by emotions. As a consequence, standard mathematical finance has been based on less heroic assumptions: the existence of a reduced number of rational and well informed professional players who are committed to exploiting every opportunity for profit, and can act as efficient arbitrageurs eliminating arbitrage opportunities. When an arbitrage opportunity arises, they will take advantage of it by buying assets that are underpriced and selling assets that are overpriced, eliminating underpricing and overpricing.

This assumption sounds more reasonable than the previous one. First, on intuitive grounds, anyone who claims that arbitrage opportunities are abundant in the market should always be asked if he is fabulously rich. If he is not, it is not clear why he knows of so many free lunches and yet, rather than exploiting them, goes around passing preposterous judgements about market functioning. But there is something more. Notice that, if the principles of risk-neutral valuation had been taken into account by market operators in the years before the crisis, such things as subprime credit derivatives with AAA rating and paying 200bps over treasury bonds would have appeared at least awkward, or even suspicious. In fact, for such assets the rating was taken by many market operators as the only measure of risk. AAA rating means negligible risk; in an arbitrage-free market an almost zero risk should be associated with a return similar to the one of treasury bonds, not 200bps over that. Thus, far from being a reason behind the existence of such derivatives, as someone claimed in the financial press, the risk-neutral pricing approach tells us that, if the market is arbitrage-free, there must have been something wrong with such derivatives, in particular the risk of these products could not really have been negligible.

But then, is the existence of such derivatives a proof that arbitrage opportunities do exist in real markets? On the one hand, the answer can be a resounding *no*. In fact, time has shown even too well that the risk was non-negligible and the return was risky, so a spread over the risk-free one was fully justified! *In the long term*, what appeared to be a blatant arbitrage opportunity for protection sellers, receiving a high fee in return for no risk, as was indicated by the ratings, was completely belied by the explosion of the subprime crisis. The problem was with the ratings, not with the market extra-return.

But the devil stays in this detail. *In the long term* . . . the apparent arbitrage opportunity detectable before the crisis has been revealed to be no such thing. But in the *short/medium term*, many of the valuations made by the market appeared very far from a rational reality of fundamentals for excess of *optimism*. And later, someone has argued that immediately after the crisis burst the quotes were far from the reality of fundamentals for excess of *pessimism*.

In the end, the no-arbitrage principle seems to have been confirmed over the long term, and banks appear to believe in it by continuously trading in derivatives and hedging them with the underlyings: the practice of hedging, and even the name ‘derivative’ securities, would be meaningless lacking no-arbitrage, which is in the end just the principle of consistency between different securities. Some market evidence that strikingly supports the idea of consistency between the prices of different securities is given later in the book, like the replications in Chapter 4, or the cross-asset calibrations of Chapter 11. But the idea that the no-arbitrage mechanism works perfectly and instantaneously, always and in all markets, as claimed by dogmatic theorists, appears just a practical simplification. It would require a perfection in the market and in the working of professional operators that does not appear realistic, so we always have to treat it with a generous pinch of salt.

We may add that in some very liquid markets, like the interest rate derivatives market that is split into only a few macro markets (dollar, euro, yen etc) with very strong pressure from a very large number of professional traders on the very same products, traders really believe in fundamental market efficiency. This is confirmed by the importance given in the last few years to model developments only intended to make advanced models fully arbitrage-free, at times thanks to what may appear as minor details (the no-arbitrage drifts of BGM, the issue of convexity adjustments, that we will see in Chapter 6). On the other hand in the credit

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

market, which is more opaque and split among hundreds of issues and thousands of portfolios, longer lasting arbitrages appear much more likely and mathematical details to enforce precise no-arbitrage are much less important than more fundamental considerations.

Bubbles and cheaters

There is something more to say here. A crucial element of market reality that is not sufficiently considered in the analysis of market behaviour is that, even if operators are perfectly rational, informed and committed, in some situations, usually generated by external distortions, they have incentives to act in a way that, for an external observer, may not appear in line with the reality of fundamentals. Here are some simplified but quantitatively precise examples of how this can happen; all the examples have a connection to events observed in the recent credit crunch.

Example 4 ***Examples of rational cheat pricers.*** *After the crisis there was a wave of criticisms against the practice of linking the remuneration of market operators only to short-term gains rather than to long-term risks. Leaving aside any moral considerations, we show below that this can have a mathematical effect on the way prices are computed. Let us consider a protection seller in some credit derivative, who has a loss at some default time τ if this happens before the maturity T of the product. The expected value of their loss is*

$$ELoss(0) = \mathbb{E}\left[e^{-r\tau}1_{\{\tau<T\}}\right]LGD$$

where LGD is the Loss Given Default. From now on, you can consider $\mathbb{E}$ as indicating an expectation under a risk-adjusted pricing measure, unless differently specified. We model τ as the time of the first jump of a Poisson process (see Appendix and Chapter 2), getting

$$ELoss(0) = \int_0^T e^{-r\tau}\lambda e^{-\lambda\tau}d\tau LGD = \lambda\int_0^T e^{-(r+\lambda)\tau}d\tau LGD$$
$$= \frac{\lambda}{r+\lambda}\left(1-e^{-(r+\lambda)T}\right)LGD.$$

Now suppose that the person who has to take the decision about selling protection or not is a bank's salesperson or a fund-manager who will get, if the deal is closed, a fantastic bonus at the end of the year sufficient to retire for the rest of his life. This is the perfect 'short-termist'. Why should he care about those cashflows, maybe negative, that will happen after the end of the year? He has no incentive to do so, thus his real maturity will not be the product maturity T, say 5 years, but his personal horizon αT, say 1 year with $\alpha = 0.2$. For him

$$ELoss^{Short}(0) = \frac{\lambda}{r+\lambda}\left(1-e^{-(r+\lambda)\alpha T}\right)LGD.$$

His expected loss will definitely be lower compared to other market players that will not get such a bonus, and this will make him more aggressive. It's even funnier if we take $r = 0$,

$$ELoss(0) = \left(1-e^{-\lambda T}\right)LGD, \qquad ELoss^{Short}(0) = \left(1-e^{-\lambda\boxed{\alpha T}}\right)LGD$$

For the other market players who do not know he is a 'short-termist', this will appear a model disagreement, since they may not understand that the disagreement is between T and αT, but *may think that the disagreement is between λ and λα,*

$$ELoss^{Short}(0) = \left(1 - e^{-\boxed{\lambda\alpha}\, T}\right) LGD.$$

By the way, the decisions of our short-termist may be underpinned by a model with default probability λα that actually fits correctly the risk appetite of this investor by misrepresenting the default probability. This can easily distort the market.

The presence of many such investors can easily distort a market even when the other market players have understood they are dealing with short-termists. In fact, they may decide to exploit the existence of such a bubble by selling to a cheater just before αT. The expected loss becomes:

$$\begin{aligned} ELoss^{Opportunist}(0) &= \mathbb{E}\left[1_{\{\tau<\alpha T\}}\right] LGD + \mathbb{E}\left[1_{\{\tau>\alpha T\}} ELoss^{Short}(\alpha T)\right] \\ &= \mathbb{E}\left[1_{\{\tau<\alpha T\}} LGD\right] = \left(1 - e^{-\lambda \boxed{\alpha T}}\right) LGD. \end{aligned}$$

So that the non-short-termist price would also coincide with the short-termist's one until the bubble bursts.

Another such 'price cheater' may be the manager of a bank who has invested so much on an asset class X *that he knows the bank's default or survival will depend crucially on the future price* X_T *of this asset class (suppose for simplicity that* T *is the maturity of its investments). He does not care if the bank defaults, because in any case he will not be affected by any subsequent loss, and it may also be that the bank is bailed out by the government. How does he regard an additional simple investment on* X*, a forward agreement with payoff $X_T - K$? The value of this investment for an investor can be written as*

$$\Pi(0) = \mathbb{E}\left[e^{-rT}(X_T - K)\right]$$

but, if the investor does not care what happens after its own default time τ^{Inv} it should possibly be written as

$$\Pi^{Survival}(0) = \mathbb{E}\left[e^{-rT} 1_{\{\tau^{Inv}>T\}}(X_T - K)\right].$$

Our initial investor strongly exposed to X *may know that his bank's default corresponds to the case where this investment has a negative return, $X_T < K$. This means*

$$\left\{\tau^{Inv} > T\right\} = \{X_T > K\}$$

and the value of the payoff becomes

$$\begin{aligned} \Pi^{Survival}(0) &= \mathbb{E}\left[e^{-rT} 1_{\{X_T>K\}}(X_T - K)\right] \\ &= \mathbb{E}\left[e^{-rT}(X_T - K)^+\right]. \end{aligned}$$

A forward contract has been transformed into an option contract, that under any model yields

$$\Pi^{Survival}(0) \geq \Pi(0).$$

The price-cheater is eager to pay a much higher price for this contract, distorting the market. The other players will see a price that, incidentally, is consistent with the hypothesis that the prices of X cannot be below K at the maturity T. This fits pretty well with the case of the ever-increasing house prices that many of the most aggressive and active operators appeared to assume in their pricing of mortgage-backed derivatives. Perhaps they really believed in this

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

aggressive hypothesis, but it may also be that they were aware to have such an exposure to this market that a decrease in the prices of the ultimate underlying real estates, would have triggered their own default. A normal investor may be confused by this pricing, until the hypothesis of ever-increasing house prices is revealed to be totally unrealistic. Going beyond the quite sterile debate about market efficiency, this is an issue that has a lot to do with model risk.

Back to the crisis. Banks, funds and ratings

There are other examples that we can take, less simplified and even more relevant to the crisis. In the market of CDOs and similar credit derivatives at the heart of the crisis, demand and supply were often driven by different forces. The demand was represented by those willing to sell protection on CDO tranches (buyers of risk), the supply was made up of those willing to structure CDOs and buy protection on their tranches (sellers of risk). Among the risk-buyers we have institutional investors such as pension funds. The decisions about the investment of such funds are not made by the households that put their money in the fund. They are made by the fund manager.

This can be a perfect example of an operator that does not always have clear incentives to act in a way which in the end is rational for the people who put their money in the fund. The manager has one obvious goal: maximizing the return from the investment. Clearly the investors also have the goal of keeping the risk minimized, but often this goal is pursued by setting very precise restraints on the manager's mandate. For example, some funds can only invest in securities with very high rating. The rating is a measure of the probability of default of a security. High rating meant an extremely low probability of default. For securities like bonds or tranches bought by institutions, such as funds, that often keep them on banking books, namely do not perform mark-to-market, the default probability is the crucial measure of risk, so this appeared an efficient way of controlling risk.

However, the rating is based on a limited set of information, strongly biased towards the use of historical data, and just gives qualitative information. More importantly, the rating agency is not the entity that suffers the loss in case of default. The agency may have negative impacts if its estimations are proven wrong, but it may also have incentives to make over-optimistic estimations, for example because it is paid by the issuers themselves.

Even without assuming unfair behaviour on the part of rating agencies that had behaved quite fairly for decades in giving ratings to *issuers*, we have to recognize that the approach to ratings for *issuances* had some weaknesses, which is interesting to mention in a book on model risk. Ratings were based on one model. The issuers could propose an issuance, see the rating and, if that rating did not satisfy the issuer, they could change at almost no cost some features of the issuance, in an attempt to exploit all the unavoidable simplifications that were in the rating model (like increasing the maturity to take advantage of some mean-reversion features, see Barucci and Morini (2009)). The rating model was in fact, in the end, very well known to issuers. Thus many derivatives were *overfit* to the details of the model behaviour, making the rating hardly robust in the face of serious discrepancies between model and reality. This was not possible when the ratings were given to issuer companies. A company cannot be changed overnight to adapt to a model, a derivative can.

But let us go back to our fund manager. Once the rating constraints are satisfied, he can say to investors that risk is under control and he will just concentrate on finding the highest returns. So if ratings are biased, his investment behaviour will be biased, even if everyone is rational.

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

In the crisis this was probably one important factor inflating the demand of CDOs. Banks could easily match this demand by securitizing their high return and high risk exposures at a cost that was probably cheaper than it would have been without the rating-sustained demand. We can add that the high ratings made selling protection on senior and mezzanine tranches attractive not only for funds but also for banks themselves, because of the capital charges dictated by the Basel Accords, that were much lower for securities with high rating. It may even be convenient for a bank to take its own mortgages, securitize them, and then buy them back.

It was difficult for non-distorted traders to bring this situation back to equilibrium. Betting against the market exposes one to mark-to-market losses in the short to medium term. A bank trader will find it difficult to make them acceptable to the risk managers monitoring the VAR of their books. A fund trader usually receives the money to invest from investors who may withdraw their support in the face of mark-to-market losses. This risk makes the trades against the mispricing less appealing, as explained in Rebonato (2003).

We have already seen that these bubbles usually burst, like in the subprime crisis example. But this is not the consequence of a dogmatic assumption of perfectly efficient markets with perfectly rational and undistorted operators who know the 'true values'. It happens when the fog really clears, and the models justifying the bubble are exposed as unrealistic in an incontrovertible way, as we have seen in the analysis of the kind of events that, in market practice, trigger model evolution in the most dramatic and relevant cases. But in the meantime distortions can influence the market for non-negligible periods.

A synthetic view of model risk that takes into account both the reality of fundamentals and the deviations from it that we can see in the market price, is confirmed to be a reasonable approach. Moreover, the different forces we have seen so far, some driving models to change to incorporate evidence in the reality, others drifting them away from our perception of the reality of the fundamentals, are confirmation that we crucially need a dynamic management of model risk. Consider also that in many real world cases we cannot distinguish, just by observing market quotes and using our no-arbitrage models, if a market is 'distorted/arbitrageable' or 'realistic'. We devote Chapter 11 to this topic. When, however, we believe there is a divergence between what we think is a realistic model and the model the market is using, only the tools of model risk management – for example the model reserves and the limits to exposures subject to model risk that we will mention later – allow us to take into account the market model and at the same time manage the risk that its unrealism is exposed, driving a painful shift in market consensus. These tools are what we need for dealing with cases where we have *uncertainty* about the right model to use, which is the topic of the next section.

1.2.2 Uncertainty and Illiquidity

The first section, starting from the analysis of Derman and Rebonato, has tried to better understand model risk by analyzing which features, such as realism or consistency with market consensus, a model should have. But even when the goal is clearer, a lot of uncertainty can remain about which model to use. In fact it can be by no means obvious which one is most realistic among different representations of reality. There can be no consensus in the market, where there may also be the presence of distorted operators. One more reason for uncertainty that we will see later on in this section, relates to the different risk aversions of the different operators. Thus we need to define more precisely what *model uncertainty* is in order to understand how to deal with that.

Formalizing Uncertainty

Uncertainty is what finance is all about. Investment decisions are choices between different uncertain returns. Pricing is computing the properly adjusted expectation of an uncertain future payoff. As a consequence, our first thought may be that model uncertainty is just an element that adds uncertainty to our payoff, so that we may deal with model uncertainty as we normally do with the uncertain quantities that appear in a payoff . . . we can take expectation of all that is uncertain and this is the price, with no uncertainty left.

However, if we think a bit more deeply, we realize that this is not the only way, and probably it is not the best way, to deal with model uncertainty. Being uncertain about which model we should use is very different from using a model that assumes that a payoff is uncertain, since it is stochastic according to a specific law.

A very instructive example of this difference can be built starting from a class of models that had a moment of popularity shortly after 2000. They are probably the simplest extension of Black and Scholes.

A model with random parameters

Following Brigo et al. (2004), in such models one assumes that, under the pricing measure, the underlying of an option looks like a standard lognormal diffusion

$$dS(t) = r(t)S(t)dt + \sigma^I(t)S(t)dW(t)$$

but has in fact a crucial difference: here the volatility is uncertain, in the sense that it is a random variable, independent of $dW(t)$, with distribution

$$\sigma^I(t) = \begin{cases} \sigma^1(t) \text{ with prob } p_1 \\ \sigma^2(t) \text{ with prob } p_2 \end{cases} . \tag{1.3}$$

Brigo et al. (2004) say that the intuition underlying this model is that 'the asset volatility is unknown and one assumes different scenarios for it'. In our examples we consider only two scenarios so we have two possible deterministic volatility functions. The greatest advantage of this model is the pricing formula for a call option with strike K and maturity T. We have as usual that the option price is

$$\Pi(K, 0, T) = P(0, T)\,\mathbb{E}\left[(S(T) - K)^+\right],$$

where $P(t, T)$ is the t price of a risk-free bond with maturity T, with

$$P(t, T) = D(t, T)$$

when interest rates are deterministic as they are in this example, and

$$P(t, T) = \mathbb{E}_t\left[D(t, T)\right]$$

when they are not ($\mathbb{E}_t[\cdot]$ is a shorthand for $\mathbb{E}[\cdot|\mathcal{F}_t]$). We can use the law of iterated expectations (see Appendix) and write

$$\Pi(K, 0, T) = P(0, T)\,\mathbb{E}\left[\mathbb{E}\left[(S(T) - K)^+ | \sigma^I\right]\right].$$

Notice that the external expectation regards a random variable that can take only discrete values, so the expectation is given by a weighted average of these values, where the weights are the probabilities of the different volatility scenarios:

$$\Pi(K, 0, T) = P(0, T) \sum_{i=1}^{2} p_i \mathbb{E}\left[(S(t) - K)^+ | \sigma^I(t) = \sigma^i(t) \right].$$

Now look at the inner expectation. Taking into account that $\sigma^I(t)$ is independent of $dW(t)$, we have that, conditional to $\sigma^I(t) = \sigma^i(t)$, $S(t)$ is just a geometric brownian motion with volatility $\sigma^i(t)$. Therefore the option price is just the average of two Black and Scholes prices.

$$p_1 B\&S\left(S(0), K, rT, \int_0^T \left[\sigma^1(t)\right]^2 dt \right) + p_2 B\&S\left(S(0), K, rT, \int_0^T \left[\sigma^2(t)\right]^2 dt \right). \quad (1.4)$$

This pricing formula for a call option is as simple as Black and Scholes itself, but it is consistent with the presence of some form of smile in the market.

If instead we need to compute a price with Monte Carlo, that we describe in Section 6.2.3, we must simulate different paths. For simulating up to T *we must know the values of all model quantities for all t* with $0 < t \leq T$. This is a fundamental feature for calling something *a model*, but at times we revisit it only when we have to perform Monte Carlo simulation. And here an issue arises with the above model. *When do we get to know* $\sigma^i(t)$*?* When is this variable observed? It seems that we need to observe it very soon, since we need to know it for all $t > 0$. This is exactly what Brigo et al. (2004) do: 'The volatility uncertainty applies to an infinitesimal initial time interval with length ε at the end of which the future volatility value is drawn'.

This specification should help us understand better the nature of the above model. The intuition about it of Brigo et al. (2004) is that we are uncertain about the volatility, in fact the model is called 'uncertain volatility model'. But is this a real representation of *model uncertainty*? If the answer were yes, we might conclude that the correct approach to model uncertainty is taking an average of the prices under each model, as done in (1.4).

Unfortunately, here there is no model uncertainty because the model is only one. In a simulation we first – at an infinitesimal time ε – draw σ^I, then once we know that on this scenario $\sigma^I = \sigma^i$, with $i = 1$ or $i = 2$, we use this volatility function until maturity. This is *one specific model* with random volatility, it is not a representation of model uncertainty.

Uncertainty on model parameters

The situation most similar to the one above but involving model uncertainty would be a situation where we know for certain that the underlying is lognormal but we have model uncertainty about the true value of the asset volatility. In this case:

1. at ε we will *not* know the right volatility. There is no precise moment when the volatility will be revealed. In models with uncertain parameters there is essentially a jump of the volatility in ε, regulated by the distribution of (1.3), in model uncertainty there is only our ignorance.
2. the probabilities p_1 and p_2 are part of no model, they are only a representation of our ignorance. We have no way of estimating them historically, for example, and fixing them can be extremely difficult.
3. other market players may know the real value of σ, or have in any case a different degree of uncertainty about it and therefore they give different probabilities to the scenarios. A market

player with more information than us may actually not even include σ^2 in his pricing since he knows the right volatility is σ^1, while a player with less information than us may also have included a value σ^3 to which we have implicitly given 0 probability.

This alters the picture. How are we to deal with this? Cont (2006) analyzes model uncertainty and says that in reality market operators work with a multiplicity of models, or triplets

$$(\Omega, \mathcal{F}, Q_i),$$
$$Q_i = Q_1, Q_2, \ldots Q_N$$

and calls the uncertainty on the right probability measure – the right model – a kind of Knightian uncertainty, from Knight (1921), who first described that type of uncertainty *where we are unable to give a distribution to the future events*, as opposed to cases of uncertainty *where we do not know the future outcome but we know its probability distribution*. The latter is what we usually call 'risk'.[5] Hoeting et al. (1999) suggests using in a Bayesian framework, an approach similar to the averaging of formula (1.4) also for model uncertainty. But this makes no sense in our context, since it makes uncertainty *on* the model indistinguishable from uncertainty *in* the model.

The distinction between the two forms of uncertainty is crucially based on knowledge vs ignorance of the probability distribution. Model risk is a case where we have uncertainty on the distributions. This case is completely different from the typical examples of risk to be found in finance books, which are almost always examples where there is no uncertainty about the distribution. In my opinion the best such example is a bet on a fair roulette. The outcome of the roulette is uncertain, but the probabilities are known precisely.[6]

The irony is that, while most books about risk focus on how to deal with roulette-type problems, the majority of problems in financial markets belong to the other category, where we do not know the distribution, and model risk is a central problem.

Understanding that finance belongs to the second category has some dramatic consequences. In my opinion the most crucial difference, particularly relevant in trading, between the models with random parameters (that still belong to the roulette case) and uncertainty on model parameters (model risk) is that in the first case the probabilities p_1 and p_2 *are assumed to be the same for all players*, while in the second case, even if we can guess them, *the probabilities will be different for different players*, and some players will have higher variance (more uncertainty) while others may have minimum variance because they know better what the best model is. This clearly shows that the approach that works reasonably in the first category, averaging across scenarios, may be dramatically inappropriate in the second situation.

Cont (2006) notices that the typical approach of banks when faced with model uncertainty is not to average across models but rather to adopt a worst case approach. Suppose that we were a bank with model uncertainty about σ^1 or σ^2, where we take $\sigma^2 > \sigma^1$. There is only one way to make sure that we will not suffer losses from any model error mispricing: we must use σ^1 to price when we are option buyers, and σ^2 when we are option sellers. This would

[5] So, according to Rama Cont, this book should be called 'Understanding and Managing Model Knightian Uncertainty' rather than 'Understanding and Managing Model Risk'. I agree, but there was not enough room on the cover.

[6] By the way, notice a curiosity. Roulette is a good example of standard risk, with no model uncertainty, for an ordinary casino player. For a sophisticated cheater who uses sophisticated laser scanners and computers to predict the roulette outcome, like those that won more than £1 million at the Ritz Hotel in London in 2004, then model uncertainty can be relevant. When we speak of roulette, we do not refer to this quite special case.

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

imply for a call option:

$$P^{ask} = B\&S\left(S(0), K, rT, \left(\sigma^2\right)^2 T\right), \; P^{bid} = B\&S\left(S(0), K, rT, \left(\sigma^1\right)^2 T\right).$$

This is the old 'conservative approach' so popular with many banks. If the above bid-ask is too large, the bank will not be very competitive in this option market and will probably remain quite out of it, focusing on markets where they do not have such uncertainty. Gibdoa and Schmeider (1989) explain that one can even formalize this conservative or worst case approach, showing that it corresponds to maximization of utility when we are faced with total ignorance on the probabilities. A relevant reference is in Avellaneda et al., (1995) and Avellaneda and Paras (1996) that, in an analysis where the uncertainty is centred on the volatility, similarly to this example, take a worst-case approach.

Following this line, Cont (2006) proposes two measures of model uncertainty: one is essentially the conservative one, where we use the model that maximizes the price when we have to sell the derivative, and the model that minimizes the price when we have to buy it, so that model uncertainty is quantified as

$$\max_{Q_{i=1,\ldots,N}} \text{Price} - \min_{Q_{i=1,\ldots,N}} \text{Price},$$

while in the second approach the investor should weight more or less models depending on their higher or lower capability to price liquid market instruments.

The first approach appears oversimplified. The second one is more elaborate, but probably it does not add much in practice because model uncertainty is usually among models that all have the same fitting capability, since in current market practice models with bad fit would have been excluded from the start, as we will see in Chapter 2.

Another interesting consideration in Cont (2006), that in my opinion remains the best paper on the formalization of model uncertainty, is in respect of the links between model uncertainty and liquidity, and model uncertainty and static hedging. According to Cont, there is no model uncertainty when:

1. the market is liquid (model uncertainty is always lower than bid-ask spread!)
2. we can set up a model-independent static hedge

These two points appear correct for a theoretical definition of model uncertainty, but in Section 1.3.4 we will see some examples that diminish their relevance when we move from theoretical model uncertainty to practical model risk management. We provide a case where the disruption of a modelling standard exploded all of a sudden exactly in a liquid market, the interbank interest rates market. This broke down the standard practices for discounting cashflows, scrambled the hedging strategies for the liquid products traded in the interbank market, and also soon made the market illiquid. In reality we should never be complacent about model risk and model losses even if today the market appears liquid.

We will also see that what was considered a 'model-independent static hedge' suddenly broke down when the market consensus on the modelling framework changed. This example shows that in practice we must be very careful to define a hedge as model-independent and static, since there can be crucial modelling assumptions behind such hedges that are usually hidden but can be exposed in times of model shifts.

The above section also sought to clarify the strong link between model risk and illiquid markets, namely markets traded at low frequency with high bid-ask spreads. Model uncertainty is an important element that drives the existence of large bid-ask spreads. The two things also tend to shrink together: when a market become more liquid, bid-ask spreads naturally reduce, and at the same time we get to know more of what a market thinks of a price, reducing part of model uncertainty. One final consideration regards a topic that, if you ever took any Master's courses in finance, may have appeared rather theoretical, but can have dramatically practical consequence, as we shall see in abstract in the following section and then with a very practical example in Chapter 2. We are speaking of *market completeness*.

Illiquid markets, ignorance and completeness

It was pointed out that there is another hidden assumption in the standard pricing framework where securities are priced as expected values of their payoffs after the drifts of the underlyings have been reduced to r in order to take risk into account. The assumption is that all the underlyings must be traded. If they were not traded, there would not be any arbitrage trades to bring the drift in their dynamics to r after the risk is taken into account. The assumption of an arbitrage-free market does not justify the above pricing framework if it is not coupled with the possibility for arbitrageurs to trade the underlyings to eliminate arbitrage opportunities.

A market where there are non-traded underlyings is called incomplete. Another way to define it is as a market where not all derivatives can be perfectly hedged. The link between lack of completeness and price uncertainty or model risk is quite strong: when a derivative can be perfectly hedged, it means it can be perfectly replicated with liquid products, namely I can 'build up' the derivative based on simpler, visible products, which suggests that a unique visible price should exist in an arbitrage-free market. The idea of perfect replicability is strongly linked to the two fundamental assumptions of Black and Scholes: continuous price processes in continuous time, that allow me to hedge out the infinitesimal changes in price in an infinitesimal time. Unfortunately these are features that we lose easily when moving to real markets. In the following we investigate more thoroughly the link between the theoretical framework of incomplete markets and the reality of model risk.

Even if with non-traded underlyings the basic pricing framework of mathematical finance does not work, when pricing we should in any case consider that the drift μ we observe in the real world must be reduced to take risk into account. We will use our own assessment of the *price of risk*, mainly of how many units of return we have to take away from μ to take all risk into account. The typical parameterization given to this implicit operation carried out by a market investor is based on the concept of a linear price of risk, a quantity γ such that, if an asset has a volatility σ, we must reduce its drift by

$$\sigma_t \gamma_t .$$

For the above stock price the dynamics to use in the risk-adjusted pricing measure would be

$$dS(t) = (\mu_t - \sigma_t \gamma_t)\, S(t)dt + \sigma_t S(t)dW(t).$$

When the asset is traded the working of the arbitrageurs will act so that the risk adjusted return is the same as the return of the riskless asset, namely

$$\mu_t - \sigma_t \gamma_t = r, \tag{1.5}$$

which has the important consequence that the price of risk

$$\gamma_t = \frac{\mu_t - r}{\sigma_t} \tag{1.6}$$

is fixed by the market that determines μ_t, and we can speak of the *market price of risk*. In particular, a market price of risk will emerge for any source of risk like $dW(t)$, so we should more explicitly indicate (1.6) by γ_t^W.

But when the underlying is not traded, we do not have (1.5) and the market does not tell us what is the agreement on γ_t^W. We can use our personal market price of risk, depending on our risk aversion, without creating any arbitrage opportunity in the market, as long as we always use the same γ_t^W every time we consider the risk of $W(t)$.

So, even if all market players knew exactly the parameter μ_t of the real-world probability distribution, we would not know which drift the other players will use in pricing, since the pricing drift has to be $\mu_t - \sigma_t \gamma_t$ and γ_t can differ for different market players. This is not so different from the model uncertainty case we have seen just above, so this element also adds to our final model uncertainty.

When a bank structures a complex derivative, it usually prefers to make it dependent on a source of risk which is not liquidly traded. The availability of a liquidly traded market for such underlying risks, from which anyone could easily extract the price of the derivative, would make the derivative less attractive for banks, eliminating many opportunities for profit. A bank that considers itself more sophisticated than most counterparties finds it more attractive if there is no full consensus on how to price the crucial risk on which a derivative depends. In the end the majority of complex derivatives in the reality of financial markets, if not all of them, have underlyings that also depend on risks which are not liquidly traded.

Now we have a better understanding of how in this context uncertainty on the price comes from at least three distinct yet interconnected elements:

1. If the market for the underlying risk is illiquid, we do not see prices giving information on the risk neutral dynamics and it is more difficult to estimate the parameters of its real-world dynamics.[7]
2. If the market for the underlying risk is illiquid, we are also in the context of incomplete markets, and therefore the price can be affected by risk preferences and the uncertainty on them.
3. Where there is so little consensus, it is more likely that distorted prices like those seen in 1.2.1 emerge and prosper for some time, adding to the total uncertainty.

So, in the reality of complex derivatives we usually know exactly neither the parameters under the real-world probability measure nor the market price of risk, as in the main example in Chapter 2. In the above simplified example, this would mean that we do not know neither

$$\mu_t, \gamma_t, \sigma_t.$$

and our uncertainty on the parameters under the real-world probability measure compounds with the uncertainty on the right price of risk to use.

The theory of incomplete markets gives an organic theoretical explanation of the second form of uncertainty. It explains that in an incomplete market the price for each player depends on its own risk aversion. For the first type of uncertainty, linked to uncertainty on the real-world underlying dynamics, we do not have such an established and organic explanation, but the

[7] Notice that if a risk factor is not traded but is anyway observable, one may in any case consider using historical estimations. This is unlikely to solve our uncertainty on models and model parameters. For example, observing the history of the underlying we can estimate historically the parameters but if they are time-dependent this is irrelevant. There are even cases, like credit, where the real underlying is the default indicator $1_{\{\tau<T\}}$ – but this quantity is zero until default and observing its history tells us absolutely nothing about the parameters to model it.

above analysis based on Cont (2003), with its focus on a worst-case approach, suggests that even in this situation risk aversion plays a role in determining how different models will be used by different players. In market practice different operators ponder models considering their realism and the market opinions on them, but in the end they are influenced in model choice by their risk appetite and their side in the deal. This is one analogy between model uncertainty and the choice of the appropriate risk adjustment. A second analogy is that there is uncertainty also in the choice of the risk adjustment. An institution has to estimate its own risk aversion, and at the same time find out about the risk aversion of its competitors, since it may have to deal with them in the future.

In the end the two forms of uncertainty do not appear very different to deal with in practice. And since when a bank designs a pricing model it usually does not distinguish between the real-world dynamics and price of risk that expresses risk aversion, these two forms of risk will be indistinguishable in the market. But it is important that a model risk manager does not forget the presence of these two components.

Remark 5 ***Incompleteness in advanced (realistic) models.*** *In the above example the interaction between risk preferences and uncertainty on real world dynamics only took into account the drift parameters, while the volatility parameter is not affected. Volatility in Black and Scholes can be affected by model uncertainty, but not by the change of 'pricing measure' that we have to make in order to take risk preferences into account. This is what Girsanov's Theorem says and is also a consequence of the intuition we gave about risk-neutral pricing: we modify drifts to take volatility into account, not the other way around.*

In practice this interplay can be more relevant. In fact, the separation between preference-sensitive parameters and preference-independent parameters exists also in other models, but with very different implications. For example, in the Heston model the drift of the stochastic volatility depends on the risk-preferences if the market is not complete for this volatility risk factor. Thus the risk preferences enter also the diffusion coefficient of the stock dynamics through the stochastic volatility. And in a model with jumps the parameters related to the frequency and size of the jumps are preference-sensitive. For example, if we consider a jump model where the jump can have an infinite number of possible sizes, we cannot think of any practical market which is complete, because this would require an infinity of traded instruments (each possible jumps size works as a risk-factor and therefore we would need an infinity of liquid markets to have the market price of risk for all of them). See Andersen and Andreasen (2001) and Cont and Tankov (2004) on jump models.

We know that the Black and Scholes model bears only a faint relation to reality. When we move towards more realistic models, we have to add features like stochastic volatility and jumps. Thus we move towards models where market incompleteness is very likely, and where very many parameters are affected by it becoming preference-dependent. And even when we are forced by practical constraints to use simplified versions of these models, they can at best be seen as approximations, reasonable projections, one hopes, of realistic models on the space of simpler and more tractable models. In this case their parameters, like the 'volatility' of the Black and Scholes model, summarize in some way the stochastic volatilities and jumps of the real world, and it is arguable whether to consider them preference-independent. All the elements of our uncertainty are likely to compound there too.

We conclude that in the practice of markets the problem of market incompleteness and the problem of model uncertainty interact dramatically. In reality we will have an uncertainty on the pricing model – which is under the risk-adjusted measure Q – that compounds our uncertainty on the parameters of the 'real' P dynamics and our uncertainty on how to quantify the risk aversion.

1.3 ACCOUNTING FOR MODELLERS

Usually modellers do not care much about accountancy matters. However, there are some accountancy issues that strongly affect our job. In fact, models are not only used to give a value to a derivative the day the derivative is bought or sold to a counterparty. Models are used along the life of a derivative, from inception to maturity, to update its value for accountancy purposes. It is in this accountancy process, and not in real money, that most profits and losses emerge for banks. The famous crises of the past, from the LTCM debacle to the credit crunch, were not generated by actual losses incurred by banks as money outflows, but by losses that emerged from the daily updates of the accountancy values of derivatives. For most financial crises the feared money outflows were anticipated by the balance-sheet accountancy losses, and these balance-sheet losses were enough to force funds like LTCM or banks like Lehman into liquidation, generating plunges in confidence and liquidity shortages. Thus, the part they have in determining accountancy values is one crucial role of models, if not the most crucial. Although we do not need to be experts in accounting details, the fundamental prescriptions and first principles given by national or international accountancy boards on how a derivative should be evaluated for accountancy purposes must not be ignored.

1.3.1 Fair Value

The principles to be followed for the valuation of financial assets and liabilities are given by international organizations such as the International Accounting Standards Board (IASB) or the Financial Accounting Standards Board (FASB). The IASB is an international body which since 1973 has tried to standardize accountancy rules at global level. In the past it published the International Accounting Standards (IAS) and more recently the International Financial Reporting Standards (IFRS). The IASB meets in London and its recommendations have a strong relevance, for example in the countries of the European Union, although they are reformulated by European and national bodies before becoming binding rules. In almost every country there are autonomous local bodies; the most important example is the above-mentioned FASB, the body designated by the SEC (Securities and Exchange Commission) to formulate generally accepted accounting principles (GAAP) valid for the US. When an accountant or auditor speaks to a quant and, after being hushed by talks of martingales and *cadlag* processes, tries to have his own moment of superiority by mentioning dozens of strange acronyms, he probably refers to the main sources of accountancy principles.

In what follows the FASB prescriptions will be taken as the main source when describing the accountancy principles driving the valuation of derivatives. In fact, IASB prescriptions are general and not directly binding in any country, thus for our purposes it is better to look at the rules that apply in one specific country, since they have a direct validity. The US is the most relevant possible example, also due to the feedback that FASB decisions have also on IASB (see below for what happened between September and October 2008). In any case, since 2002 the FASB has undertaken the convergence project with the IASB to keep their prescriptions consistent.

According to both IASB and FASB, derivatives and all financial assets or liabilities which are not held for an investment until maturity (banking book products) but can be bought and sold (trading book products) *should be evaluated following the principle of fair value*.

'Fair valuation' is similar to what traders call *mark-to-market*, and actually is not a principle imposed by accountants on financial market operators but, on the contrary, is an approach that developed in financial markets and then spread to the accountancy practices of the rest of the economy. Mark-to-market as an accounting device first developed among futures traders in the Chicago Mercantile Exchange, prior to 1980. Then it was adopted by big banks and is now an important principle for all corporations.

The IAS 39 document, first issued in 1998 and regularly revised, gives the IASB definition of fair value as

> *'the amount for which an asset could be exchanged, or a liability settled, between knowledgeable, willing parties in an arm's length transaction'*(IAS 39)

while the FAS 157 document issued in September 2006 by the FASB states that

> *'Fair value is the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction'* (FAS 157)

Let us point out a few relevant and non-trivial implications of the above definitions, in particular the one in FAS 157.

1. The idea underlying fair value is that the market is the ultimate reference to give a value to a derivative. If a banks internal model indicates that a derivative has a value different from the one that prevails in the market, the market should win when it comes to accountancy. This principle, however, must be balanced with the considerations below at point 4.
2. Fair value is not a mid-price. In the case of an asset it must be the sell price, with the implication that a radical application of this approach in a market with a relevant bid-ask price may lead, one instant after an asset has been bought, to the buyer marking a loss as wide as the entire difference between ask price and bid price. In the case of a liability, we must consider the price required to extinguish it. Fair value is an *exit value*.
3. It seems that everything that would enter the exit price must be considered. For example, if the product is not collateralized, we should consider the default risk of our counterparty (*credit value adjustment, CVA*) and possibly our default risk (*debt value adjustment, DVA*). The introduction of counterparty specific elements, however, is not dealt with in the same way in all countries. It is a non-trivial issue in any case, since default risk of the counterparty makes equal payoffs have different fair values, while the introduction of our default risk can, according to some market operators and some regulators, induce a form of moral hazard. We investigate this issue in Section 4.2.
4. The expression *'arm's length transaction'* is a conventional legal phrase indicating that the two counterparties to the transaction must be on an equal footing, without one having a contractual power much stronger than the other one. It must not be, for example, a transaction between a parent company and one of its subsidiaries. A similar concept is expressed by FAS 157, which says that fair value must refer to the price in an *'orderly transaction'*. This is easy to understand for financial markets: we must not consider distressed conditions, for example when the seller is forced to sell due to a funding liquidity shortage. The transaction must happen under normal market conditions.

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

When looking at the first point above, the accountancy view appears more in line with the Price approach to model risk, which puts discrepancies between market price and model price as the first cause of model-driven losses, rather than the difference between true or at least realistic value and model price. But some questions remain unanswered. What should we do when a market does not exist? What if such market is illiquid? Such considerations have been addressed since the issuance of the above documents, in particular through the definition of a three-level hierarchy, which we describe in Section 1.3.3 following FAS 157.

One question remains open in spite of this, and it is a crucial one: what should we do if we have reason to think the market prices *are wrong* ? This question was fully addressed by boards of accountants only after the credit crunch, and the answer they gave is so relevant to an understanding of the current interpretation of fair value that we address it first.

1.3.2 The Liquidity Bubble and the Accountancy Boards

The issue that market prices can be wrong and irrational, and therefore can in some cases not be a good foundation for giving a balance-sheet value to derivatives, was raised very often after the subprime crisis burst, when products like CDOs, that had previously been treated as almost default-free, started to be evaluated in the market as if all the underlying debt were on the verge of default, even when a rational analysis suggested that this was not the case. This process peaked after Lehman's default, when many banks complained that the market was dramatically undervaluing assets and overestimating liabilities, they were creating the risk that banks following market prices for fair valuation could also be led to default even if not affected by big material losses.

Accountancy boards and regulators answered complaints on 30 September 2008, when the SEC and the FASB issued a joint clarification in which they pointed out that fair value definition speaks of market prices in *'orderly transactions'*, as we noted in point 4. From this it follows that when a market is affected by illiquidity and panic there cannot be orderly transactions, thus market prices may not constitute fair value – and thus can be corrected by personal judgement.

These statements by the accountancy board may appear surprising, since they seem to say that we should account at market prices only when market prices are deemed to be right. When the markets are instead deemed to be wrong, we should account at what we think is the correct price. What is a correct price, in the view of accountants and regulators? In their joint clarification, called 'Sec Office of the chief accountants and FASB Staff Clarification on Fair Value Accounting', the SEC and FASB state that reliable prices need to come out of *reasonable risk-adjusted, liquidity-adjusted and credit-adjusted 'expected cash flows'*.

Only a fortnight later the international board, the IASB, issued a similar paper, in which they claim that a firm should *not* 'automatically conclude that any observed price is determinative of fair value', and that if necessary it can make 'adjustments to the transactions price that are necessary to measure the fair value'. Even the Accounting Standard Board of Japan published a paper on 5 December along the same lines.

These clarifications say that fair prices should be descriptive of expected future cashflows, and therefore they need to coincide with market prices only when market prices can be broadly believed to be already such realistic descriptions. This sounds closer to the Value approach to model risk than to the Price approach, while at first sight the latter appeared closer to accounting principles. This debate in the field of accountancy rules really mirrors the debate we have already seen from a quant's perspective in Section 1.1. And, similarly to the conclusions reached there, in the current view of the boards market consistency is still treated as the first benchmark to take into account in valuation, but they clarify that this can and must be mitigated by considerations that take account of the reasonableness of market prices.

1.3.3 Level 1, 2, 3 . . . go?

There is another similarity between the accountancy prescriptions and the ideas about model risk that we have seen in Sections 1.1 and 1.2. It is the principle that there is a link between model risk and a shortage of liquid market information. There exists in fact a three-level hierarchy in accounting principles that classifies the prices based on the inputs that enter the valuation procedure: prices based only on objective inputs, coming from liquid markets considered surely appropriate for the product to price, rank 1; while prices based on inputs computed making use of a plainly high amount of personal judgement rank 3. We present this hierarchy, then provide a few examples to show why it must be taken with a pinch of salt from the point of view of model risk management.

1. Level 1 (Liquid specific quotes). The valuation of a derivative falls into this category when it is based on the price coming from a *liquid* market where products *identical* to the derivative to price are traded. When such a market is available, we must give our derivative the price that we read in this market at the valuation moment. We have one single input taken directly from a liquid market, and no transformation of the input is required
2. Level 2 (Comparable quotes or Illiquid quotes). These fair value prices come from either:
   A. a *liquid* market where products *similar* (but not identical) to the derivative to evaluate are traded. Here we apply judgement in the application of the input to the derivative to value.
   B. a market where products *identical* to the derivative to evaluate are traded but the market is rather *illiquid*. Here there is judgement in assuming that I could readily sell the derivative at that price.

   In some cases Level 2 prices require an adjustment to be applied to the derivative under evaluation. This often requires interpolations or even extrapolations (see Chapter 7 for the risk involved in these operations, and how to minimize it) that are often carried out via models and can involve personal judgement.
3. Level 3 (Model with non-quoted parameters). These fair value prices come from the use of a valuation technique that requires inputs which involve a crucial amount of personal judgement of the institution computing the price.

The above description is only a summary of the details on how to classify prices into the three categories given by accountancy boards. For example, the FASB explains that Level 2 prices can also make use of valuation techniques, but differently from Level 3 the inputs of these valuation techniques must be available – at least in illiquid markets or for comparables. The valuation techniques used should always be consistent with the definition of fair value, *assumptions that an unrelated party would use* (thus pricing involves making assumptions about someone else's assumptions).

The general wisdom, also contained in some official documents, is that Level 3 prices are 'marked-to-model' and therefore should involve a lot of model risk, while Level 2 prices involve a more limited amount of model risk, and there is no model risk at all in respect to Level 1 prices, which only involve 'market risk'. This formal scheme does not always coincide with the actual distribution of model risk on different deals.

Consider, for example, the distinction between Level 2 and Level 3 prices. They may both involve the use of models, but Level 3 prices are considered less reliable because they involve models that require some input which is not readily available in a market, but must be chosen by the user's judgement or by procedures such as historical estimations. Such a distinction risks fuelling distorted choices about models. If a realistic model recognizes the existence of two different actual risk factors, and allows for a relation to exist between them, it may require a correlation input to be estimated historically. In this case the product will certainly be classified as Level 3, which usually means that the trader will have some penalization compared to a Level 2 price, in the form, for example, of a recognition of his profit that will be diluted in time. This can make sense, but the problem is that another less realistic model may not recognize that two distinct risks exist, treating them as a single one (always perfect correlation), or may recognize the two separate risks but may assume by construction that they are independent (always zero correlations). This model will not require a correlation to be fixed exogenously, thus the prices generated by this model can be recognized as Level 2 and involve less penalization for the trader! The trader will have a clear incentive to choose this second model, which actually involves much more model risk.

In Chapter 8 we show examples of models that make mistakes about correlations which are misinterpreted as perfect dependencies or independencies, and we will see that such models can be dangerous. Another interesting case are the mapping methods described in Chapter 3. Such methods were used before the crisis to evaluate illiquid products based on the quotes of scarcely related liquid products. The relation between liquid and illiquid products was based, as you will see, on very abstract and rigid mathematical relations that created a high model risk. Their rigidity and abstraction did not involve any exogenous parameter choice, which instead would probably have been beneficial to adapt them to the different market situations. But in a shortsighted interpretation of the above accountancy principles, their main flaw may have played in their favour, making them Level 2 rather than Level 3.

Now consider Level 1 prices. Since these come from a market where products are traded identical to the derivative under analysis, they should involve no model risk at all. Here too, the classification can be more formal than substantial, and does not recognize correctly the presence of model risk in a broad sense. In fact, in deciding that a price is Level 1, there is a clear amount of personal judgement, it is the *judgement required to decide that two assets are 'identical'*. One such decision is often based on some assumptions that may be so hidden and rooted in market practice that they are no longer recognizable as essentially 'modelling' assumptions . . . until they become unrealistic and are rejected by market consensus, at times in the blink of an eye. To illustrate how strongly model assumptions can enter into the evaluation of even simple vanilla derivatives, we show some hidden assumptions that enter the valuation of the simplest plain vanilla swap. By the way, these assumptions were generally agreed upon for decades, but were discarded in a day, and we know which day: 9 August 2007.

1.3.4 The Hidden Model Assumptions in 'vanilla' Derivatives

Suppose we have available a liquid interbank market for a few maturities. In this market we can observe the quotes for loans from bank to bank. Banks will agree on the rate at which such loans must happen: this rate has been called Libor for many years. For the maturity T, the market prevailing rate at time t is $L(t, T)$. Which means that if the notional of the loan is

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

Replicating a FRA

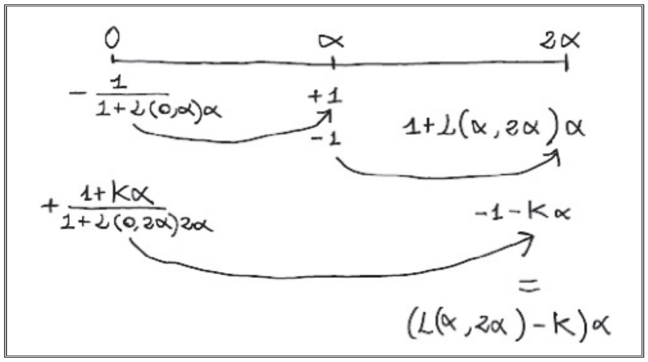

Figure 1.4 FRA standard replication scheme

$1, the lender gives 1 at t, and the borrower must give back

$$1 + L(t, T)(T - t)$$

at T. But also, if the loan has a notional of $\frac{1}{1+L(t,T)(T-t)}$, the lender gives $\frac{1}{1+L(t,T)(T-t)}$ at t, and the borrower must give back 1 at T. Thus we see that this market, at least for the maturity T, gives directly the spot rate and also the value of a contract that pays 1 at T.

Suppose the rates that we see quoted in the market at t include $L(t, \alpha)$ and $L(t, 2\alpha)$. Now we are a bank that has to price, at time $t = 0$, a forward rate agreement with expiry (or fixing) at α and payment at 2α, a product which is the simplest one-period, fixed for floating swap where at 2α one party has a payment indexed to a fixed rate and the counterparty instead pays a floating rate that fixed at α. If the agreed fixed rate is K, the payoff for the payer of the fixed rate is simply at 2α

$$(L(\alpha, 2\alpha) - K)\,\alpha. \tag{1.7}$$

If we are the payer of the fixed rate, we can trivially price it with the following replication, summarized in Figure 1.4. Lend $\frac{1}{1+L(0,\alpha)\alpha}$ to your counterparty for maturity α. You will receive 1 at α. At α lend this 1 for a time α, namely from α to 2α, and you will receive at 2α

$$1 + L(\alpha, 2\alpha)\alpha.$$

This contains part of the payoff of the FRA, $L(\alpha, 2\alpha)\alpha$, but also 1, that does not appear in the FRA payoff (1.7), and on the other hand it does not contain K. In fact at time 0 one must also borrow $\frac{1+K\alpha}{1+L(0,2\alpha)\alpha}$ from the counterparty until 2α. We will pay at 2α

$$-\frac{1 + K\alpha}{1 + L(0, 2\alpha)\,\alpha}\,(1 + L(0, 2\alpha)\,\alpha) = -1 - K\alpha.$$

Put together with what we receive and get

$$L(\alpha, 2\alpha)\alpha - K\alpha,$$

which is exactly the same payoff from the FRA given in (1.7). Thus the price of the FRA must be equal, in an arbitrage-free market, to the price of the replication strategy:

$$\frac{1}{1+L(0,\alpha)\alpha}-\frac{1+K\alpha}{1+L(0,2\alpha)\alpha}.$$

Is this pricing Level 1, 2 or 3? Well, we are working on liquid markets, we are not using any comparable, but only things that ought to be the same in any arbitrage-free market, and we are using no evaluation model. Thus it looks like Level 1.

One may say: no, wait a second. You are using one modelling assumption: you are assuming absence of arbitrage. This is true, in fact all replication arguments like those above work only if there is no arbitrage, since it is absence of arbitrage which dictates that things which have the same payoff ought to have the same price. But the whole Level 1,2,3 classification is based on the assumption that absence of arbitrage does hold. Otherwise the definition of Level 1 would not make sense. An option on the Microsoft stock with maturity 1 year is Level 1 because we can see in the market the price of other options on the Microsoft stock with same strike and same maturity, and use this price to evaluate my option. But this strategy yields a reliable price only if there is absence of arbitrage, otherwise there would be no reason for two options with the same payoff to have the same price. Thus absence of arbitrage is not an additional assumption, but is rather a foundation of the Level 1,2,3 classification, and my pricing of the FRA above involves no other assumption and no illiquid product so it ought to be Level 1.

Usually an FRA is not Level 1 but Level 2 because the evaluation goes through a bootstrap of the term structure that always involves at least interpolation between maturities, but here we assumed rates and thus discounts for exactly the maturities of the FRA. The FRA appears identical to the portfolio of loans; the payoffs are the same.

And yet, in contrast to usual market wisdom, this price comes from some strong model assumptions, and as such it is subject to a fair amount of model risk since the underlying assumptions, like all model assumptions, can be rejected by the market in the future. This happened in this case at the beginning of the credit crunch, precisely on 9 August 2007. We see in Figure 1.5 that the correspondence between FRA prices and the price of their replication strategy held with surprising precision for many years, before breaking down at the beginning of the subprime crisis.

We will analyse this issue in Section 4.1, since this is of particular interest to us as an example of how a market can change its modelling paradigm. But here we anticipate one point: where are the hidden model assumptions in FRA pricing?

First of all, it is clear that the FRA above is treated as default-free. The first thing we have to consider, therefore, is what happens when we introduce risk of default. We may think that this breaks down the above pricing because a default event spoils the identity between the payoff of the FRA and the replication strategy. We will see in Section 4.1 that this is not necessarily the case because the payoff and the replication strategy may be affected by risk of default in the same way. What default risk really breaks in the replication, even for collateralized FRAs, is that with this additional risk *we have no more guarantee that our counterparty in the replication has the same risk of default which is expressed by the market rate* $L(\alpha, 2\alpha)$ *that appears in the payoff*.

It is in this detail, and not so much in the possibility of a default event, that the devil hides. As we see in Figure 1.6, the presence of default risk means that my counterparty C will have a fair lending rate $L^C(\alpha, 2\alpha)$ which is not necessarily the same as the rate $L(\alpha, 2\alpha)$ that appears in

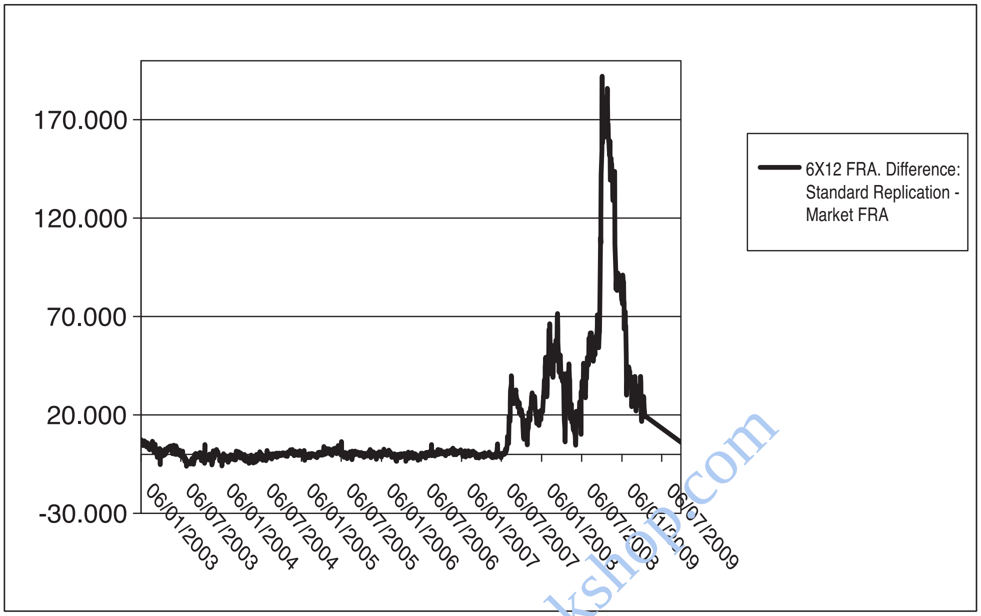

Figure 1.5 Standard replication − Market FRA (6m fixing, 12m payment)

the FRA payoff, potentially breaking down the above replication, that required this condition to have a payoff really equal to (1.7).

And suddenly this reveals the extreme imprecision of the above payoff description: what does it mean that the FRA will pay a floating rate $L(\alpha, 2\alpha)$ at 2α? What is that rate precisely and how will it be computed at 2α? Here we find out the truth. As long as the market is an idealized market, first and foremost free of risk of default, all interest rates are the same since they simply express the market consensus about the time value of money. But when we introduce further elements of realism, we are forced to analyze precisely which rates we are

Replicating a FRA after the credit crunch

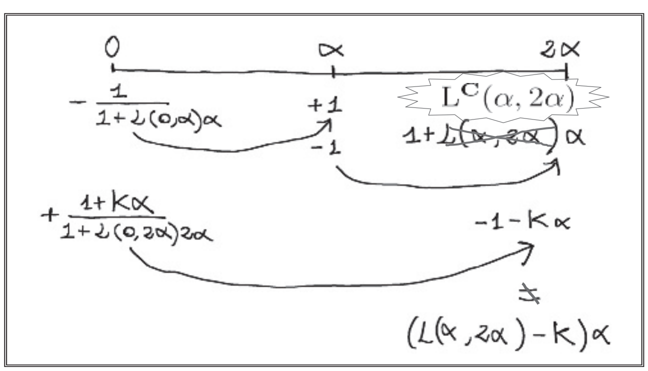

Figure 1.6 FRA replication scheme after the crisis

talking about. And we discover that the FRA is not at all as simple a derivative as we thought, but is more akin to a basket derivative. In fact the rate $L(\alpha, 2\alpha)$ is usually a Libor rate that comes out of a rather elaborate process.

We will give full details of the Libor mechanism in Section 4.1, but we see here some crucial elements of how Libor is determined. The level of Libor is provided by a special average of contributions from a panel including a peculiar set of banks. Only in a market where the risk of default by banks is negligible can we safely assume that all banks have the same fair borrowing rate, and that this is always also the fair borrowing rate for us and for our counterparty in the replication, justifying the replication itself. In fact, in a risk-free market the borrowing rates only express the time-value of money on which the market agrees.

If risk of default is not negligible, we can consider the replication at least approximately valid only if all banks in the panel and also my counterparty have very similar default risk and *this will be expressed at* α *by the rate*

$$L(\alpha, 2\alpha) = L^C(\alpha, 2\alpha).$$

But notice that the latter assumption in italics involves something more than having our counterparty similar to the Libor counterparties today. We are also saying that this will last until α, but this clearly means we are assuming something more. When we allow for credit risk, all rates can be seen as a risk-free rate expressing the time-value of money *plus a credit spread*,

$$L(\alpha, 2\alpha) = L_{rf}(\alpha, 2\alpha) + Spread^{Libor}(\alpha, 2\alpha),$$
$$L^C(\alpha, 2\alpha) = L_{rf}(\alpha, 2\alpha) + Spread^C(\alpha, 2\alpha).$$

What is the volatility of these spreads? What is their correlation? If the volatility is assumed to be zero, then it is true that similar spreads today imply similar spreads at α. The replication can still be approximately satisfactory. But if the volatility is not nil, then we have to worry about the correlation. Is the correlation between $Spread^{Libor}$ and $Spread^C$ equal to 1? Then again we can still use the above replication, but notice that this implies a unit correlation of C with all 16 banks in the Libor panel (even more for Euribor). If *any* of this assumptions does not hold in a relevant way, the above price obtained by the replication can be absolutely wrong. Thus the above price, which by the way all pre-crisis books on interest rates describe as *The price* of a FRA, actually hangs on either:

1. All banks active in the market have a negligible risk of default.
2. If this risk of default is not negligible, it is the same for all banks and its volatility is negligible.
3. If this volatility is not negligible, in any case all banks involved have their risk of default perfectly correlated.

Seeing them written down, we realize that these are not at all light assumptions, they look pretty strong. In spite of this they were approximately valid for the 10 years preceding the credit crunch, so much so that many market players and quants even forgot that they were used implicitly. But in the summer of 2007, they all fell down, from first to last, like dominos, because traders did not believe them to be valid anymore, and this changed market patterns.

The traders were right: in fact, about 12 months later we saw credit spreads sky-rocketing, some banks being hit harder by the credit crisis than others, and in the end we even saw a big, Libor panel bank like Lehman default. As soon as traders adopted the new view, for the first time FRA prices went out of line with (1.7). This shows that even with apparently 'Level 1' or, at worst, 'Level 2' prices there are important modelling assumptions, and additionally they can change quite suddenly so that there is also model risk in such Level 1 or Level 2 prices. In Section 4.1 we will analyze the situation in detail, and we will provide technical tools for dealing with it. We will dig deeper showing, for example, that the replication above implies that not only the counterparty but also we who make the replication must have the same risk of default as Libor. More interestingly, we will see that even for a collateralized FRA, as most FRAs in the market are, the above standard replication failed at the credit crunch, even if there is no risk of default in the FRA. In fact, risk of default is in any case included in the Libor rate L. Finally, we show a replication alternative to the one above, one based on more general assumptions and approximately consistent also with the market prices of FRA after the credit crunch.

So far, what matters to us in this example is to show that model risk can actually arise in liquid markets. If the market remains liquid it is formally exposed only to 'market risk'. But this is only a formal distinction. There are variations in the patterns of a liquid market that can be expressed as 'market risk' because we observe them as variation of quoted prices, but can only be understood and anticipated as changes in the *model* consensus by the market. The loss can be lower when this shift happens in a liquid market, because in liquid markets a bank is more likely to have both long and short positions.

However, we must also consider hedging risk, an aspect often overlooked, or treated in a simplified way, in accountancy principles. The change of the model consensus in a liquid market can break down the fundamental hedging strategies. The change in interest rate modelling assumptions triggered by the subprime crisis emerged as market risk for quoted FRAs; but since it was associated with a change in the model consensus it broke down the hedging strategies of FRAs and swaps, both liquid and illiquid, that were based on the model consensus itself.

We see something similar in one of the examples in Chapter 11: even if one were convinced that there is no model risk in the pricing of liquid caps and swaptions, there is certainly model risk if you hedge cap exposure with swaptions, or the other way around, based on the fact that in standard terms structure models they depend on the same risk factors. This risk became painfully evident as early as the Russian crisis of 1998. Even when we find it convenient to talk about 'market risk' when looking at losses in a single FRA or a single Cap, if we want to explain losses like those of 1998 or 2007 *at a portfolio level* we cannot overlook the fact that at those points in time a modelling paradigm broke down.

Moreover, the change in interest rate modelling assumptions of August 2007 also disrupted the pricing of all swaps that were not liquid, and the foundations of the construction of the discounting curve for derivatives of any asset class. All these hinged on the model assumptions whose change was revealed by changes in liquid market price. Liquid markets are often what we need to observe in order to understand the changes in the pricing of illiquid products. Fundamental model assumptions are certainly more stable in liquid markets, but they can change suddenly even there. An investor will find it very difficult to go against the model consensus of a liquid market, so in a strict sense here we have no 'model uncertainty' because if we want to participate in this market we can only do so at the market's consensus. But also in this case, if the model consensus appear unrealistic or if we detect a possibility that it will change in the near future, tools of model risk management such as limits to the exposures like those given at 2.7.3 can prove useful.

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

Even what happens in liquid markets is strongly related to model risk. Not to mention the fact that with the subprime crisis the interbank market and the FRA market suddenly became illiquid. The fact that the liquidity of a liquid market can dry up confirms what we said in Section 1.2.2: liquid and illiquid markets are affected by model risk in different ways, but setting too sharp a divide between them in terms of model risk management can be dangerous.

1.4 WHAT REGULATORS SAID AFTER THE CRISIS

If this book had been written before the financial crisis of 2007–2009, I am not sure that it would–have included a section on what regulators have to say. Regulators have strongly influenced the world of finance in the past few years, but their role has not always been positive. We have seen in Sections 1.1.3 and 1.2.1 how the importance given to ratings was among the factors that encouraged lax practice in relation to risk management. Some regulators seemed to think that model validation was a formal procedure, and that model risk was minimized if banks were forced to use the same simple, standardized models. The main effects of this approach were: to fuel the so-called regulatory arbitrage, where banks exploit the weaknesses and the rigidities of regulations; to favour procyclicality, making all banks herd together during bubbles and crises; and, most dangerously, to prompt many institutions to commit to following the formal rules given by regulators and then to use this formal compliance as a justification for being particularly lax in terms of substantial model risk management. However, after the crisis regulators issued many documents and uttered numerous words analyzing the mistakes made by banks and those made by the regulators themselves. I think that we should pay attention to what they said for at least two reasons:

1. The guidelines they give about model risk management, in terms of banks' governance, of organization of a risk management unit, and of principles to follow in model validation should become, along the lines of Basel's Pillar II, the benchmark for determining the amount of capital to be held by banks against model risk. Banks that stray too far from these principles will be required to keep more capital.
2. The approach they adopt post-crisis is different from the formal and schematic approaches suggested before the crisis. There is more attention to the substance of model risk, and more awareness that regulators can lay down principles for model risk management but cannot give recipes that work for all banks in all market situations. It seems more evident that model risk management is not an issue of legal compliance but rather a strategic issue that requires a commitment by management tailored specifically to the characteristics of a financial institution and of the markets where the institution operates. Many suggestions are reasonable and should be considered by modellers.

The most important document issued by regulators after the subprime crisis with a strong relevance for model risk is the Basel Committee's 'Supervisory guidance for assessing banks' financial instrument fair value practices', published on 6 February 2009. This is by far the most frequently quoted document by managers and experts at conferences on model risk and validation, and in my opinion it is the best summary of up-to-date regulator views on modelling, as expressed in documents issued by the SEC, FSA, Central Banks and other institutions. The paper 'Supervisory Guidance on Model Risk Management', published by the FED while I was completing this book deserves special mention. Its definition of models is particularly interesting: 'a model is composed of three parts'. They are 'a component which

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

delivers assumptions and data to the model' together with 'a component which transforms inputs into estimates' and, finally, 'a component which translates the estimates into useful business information'. This is a very reasonable assessment to make, as we pointed out in the preface. There are many more similarities between the framework it designs and the approach followed in this book.

Going back to the 'Supervisory guidance for assessing banks' financial instrument fair value practices', we see that it gives 10 principles divided into two groups, 'Valuation Governance and Control' and 'Risk management and Reporting for Valuation'. For the purposes of a technical book such as this, with a focus on model risk management, I have not found it useful to follow the structure of the document. I have skipped those principles I found mainly formal or of little relevance to model risk management and validation, while from the remaining principles I have selected three groups of recommendations which I consider a good summary of the most relevant and innovative elements.

The first group relates mainly to the process of model validation and model management, with specific focus on the role of an institution's management in the process, the use that must be made of the conclusions of the process, and the general principles that should inspire it.

The second group is more specific and relates to how model validation must be adapted to the market and to the specific product under analysis.

The third group of recommendations is really a list of operative steps that should not be omitted in a model validation process.

The second and third groups of recommendations have a real impact on the daily work of a practitioner, and they contain suggestions which are reasonable and in line with the analysis and examples given in this book.

1.4.1 Basel New Principles: The Management Process

1. *'Management must understand the basis of any valuation techniques used by outside parties . . . to determine the appropriateness of the techniques used, the underlying assumptions.' 'Reporting to senior management and to the board should be on a regular basis in an appropriately aggregated and understandable form'*
2. *'Bank valuation methodologies are expected not to place undue reliance on a single information source (eg external ratings) especially when valuing complex or illiquid products'*
3. *'Articulating the bank's tolerance for exposures subject to valuation uncertainty . . . Analysis should be commensurate with the importance of the specific exposure for the overall solvency of the institution'*
4. *'Fair value measurements may involve a significant amount of judgement'*

After the crisis, when sharp criticism was levelled at the senior management of financial institutions, many senior people in banks claimed to be unaware of the valuation techniques used by the institution they managed, and often laid the blame on technical experts. But it was the senior management of financial institutions who decided the strategies and the risk appetite that eventually led to the crisis, and it was they who received the large part of the revenues from it. It was senior management who knew the amount of exposures built towards the different sources of risk.

The excuse that management did not know or understand the valuation techniques used is unacceptable. Either the senior management, who take the investment decisions, claims to understand at least the fundamental practical aspects of the valuations carried out by the technical experts, or the technical experts should be the ones who take investment decisions and get the revenues from them! If things go on as they did before the crisis, the technical experts will decide nothing and eventually will be held to blame for all losses, while the senior management will get all the revenues from risky investment strategies but will not be held responsible when these strategies lead the bank to default or to be bailed out by government.

The first principle above lays the foundations for senior management to be held responsible also for investment decisions taken using valuation models. They are required to understand *the basis* of the valuation techniques used, and to judge on the underlying assumptions. There is no need for them to understand the technical details, but they must take responsibility for the most relevant underlying assumptions. Only they have sufficient information on the bank's business to judge which assumptions are acceptable and which ones should instead be considered too risky. This should mean that the discharge of responsibilities onto models and modellers carried out by some executives during the crisis will no longer be acceptable.

On the other hand, this also requires a commitment on the part of the technical experts. This is the subject of the second sentence in point 1. The choice of a valuation model cannot fall entirely to a quant (who may be completely unaware of market dynamics and the bank's exposures) downloading from the web a popular paper (perhaps one that simply focuses on the details of the derivation of a closed-form formula) and implementing it, with just some generic hints from a trader or salesman and with the underpinning of a purely formal model validation.

The choice of a valuation model must be based on an analysis about its appropriateness and its implications, and this analysis must be reported to senior management, in an *aggregated and understandable form*. The communication must necessarily be aggregate, because senior management get easily lost in the detail and even more importantly, it must be understandable. Many quants, traders and other technically strong practitioners will find this aspect one of the most difficult. How to make things which are usually expressed in a technical language comprehensible for senior management? A way must be found, if technical people do not want to be held responsible again for the misuse of models. And it is not impossible. This book is devoted also to help quants with this. The first point is that we must be sure to have fully understood the model ourselves, not only in its technical details but also in its fundamental assumptions and financial implications, with a practical focus. At times this point is overlooked by quants, who concentrate on mathematical and implementation aspects, therefore we cover this aspect thoroughly in these pages. We hope this will help also communication between those who understand models and those who take decisions based on them. Without this we have the worst model risk.

The document stresses that management must understand the basis of a valuation technique even when it is provided by an external party, such as a rating agency. This is a recognition that excessive reliance on ratings, partly fuelled by regulations, was one of the causes of the crisis. Ratings are explicitly mentioned in the second point, when it is recommended not to use just one source of information.

The third point is simply common sense but is often lost in model validation. Model risk is not linear with respect to exposure. The last crisis was not generated by a bank using a wrong model, in the sense of a model that was liable to be suddenly discarded for lack of realism. This is exactly what the market model for CDOs was, but banks use many models which are even less realistic. It is unavoidable that some models used are like that, but this does not necessarily generate a crisis. The problem before the credit crunch was that an enormous exposure was built on a source of risk that was assessed with a model subject to strong valuation uncertainty and high likelihood of model revision. It was the unfortunate marriage between large exposure and valuation uncertainty that put the world on the verge of disaster.

The third point recalls the steps to be taken to avoid this scenario. A bank must introduce a link between valuation uncertainty in assessing a risk and the size of the exposure to that risk. The higher the valuation uncertainty, the lower should be the exposure. This can be done in different ways, for example by the institution of 'model lines', similar to the credit lines already used by banks, which are discussed in the next section. All effective tools of model risk management, such as model reserves, have in the end this purpose. The goal of model reserves is not to stop a salesperson or a trader from closing a deal, but to reduce the seeming convenience of the deal when we fear that this convenience can be due to misvaluation. This should prevent exposure to this model risk from growing too much and too fast. When standard reserves are not enough, we can think of reserves linked to coefficients that grow when the bank's total exposure grows. It is in the end the very old principle of diversification applied to model risk. Clearly, this requires a firm-wide approach to model risk management, and confirms that those who conduct validation must work closely with those who have a global view of a bank's business.

The second part of the point serves as a reminder that the effort in model validation must be commensurate with the exposure: when a small exposure is started towards a new risk, a limited effort for model validation is acceptable. A bank cannot expend big resources for a small exposure; and additionally banks and traders learn by trial and error, a new model needs to be tested for a while in order to really know its risks. When the exposure starts growing, a previous model validation must not automatically be considered valid: a surplus of effort can be spent on the model used, an effort that was not economically meaningful in the past but is crucial in the face of an increased exposure. This seems redundant, common sense. However, in practice it can happen that risk managers focus a lot on the validation of a new model for a very small deal. Then, when the model is validated, the model is forgotten even if at some future date the sheer survival of the bank hangs on the validity of that model.

The last point, 4, may also sound like trivial common sense to many practitioners. But it is a minor revolution to see it underlined so clearly in so many official documents. The objectivity of valuation, the idea that there must be one single right model that all banks should use, haunted many regulators. It would be good if they were to abandon this dangerous delusion, and the same applies to those managers or salespeople who expect quants to 'tell them the right model'. . . and to the quants who think they have the answer because they have just downloaded a popular paper. Model choice is really a choice, and at times personal judgement is the most important ingredient of all. Recognizing this is a good first step for a better understanding of model risk.

1.4.2 Basel New Principles: The Model, The Market and The Product

1. *'Uncertainty is specific to the instrument and to the point in time the valuation is effected'*
2. *'A bank is expected to consider all relevant market information likely to have a material effect on an instrument's fair value when selecting the appropriate inputs'*
3. *'However, observable inputs or transactions may not be relevant. In such cases, the observable data should be considered, but may not be determinative'*

These three points do not relate to the process of model risk management, but they set precise boundaries on its scope. The first point recalls that uncertainty is specific to one instrument.

We cannot automatically extend to a new instrument the validation that was performed to allow a different application. It makes no sense to think of a model separate from a specific application; the validation cannot state that a model is valid for all computations that can be performed with it.

This is a consequence of the old principle, recognized also by Derman, that ‘all models are wrong, but some are useful’, as the statistician George Box said in the 20th century. A model is a tool that can be useful to solve a problem, but also inapt when solving a related but different problem. Models always make mistakes, and a serious validation is mainly the job of judging if these mistakes are negligible for the product at hand. This makes the scope of validation unavoidably narrow. We will see often in the next chapters, and in particular when discussing stress testing in Chapter 3, that a slight change in the payoff can change, in a dramatic and surprising way, the behaviour of a model. This is clear to most practitioners, but not to everyone. We can still hear old-style risk managers make comments such as ‘we have validated the model for interest rate derivatives’, as if the set of interest rate derivatives was so limited in its potential expansions that it is easy to find a model appropriate for all of them. The only model with such a property would be the ‘true model of interest rate dynamics’, but unfortunately such a model does not exist. At times it may be acceptable to gain efficiency by validating models for ‘categories of products’ rather than single products, as some banks do. But we must be careful in extending the scope of a validation in terms of products. The main risk is the temptation to use this approach in a ‘Lego’ way. The fact that a model has been considered valid for *spread* options as well as *American* options, for example, does not imply it is appropriate for *American spread* options.

The second part of this point reprises that uncertainty also changes with time, so that a validation performed years ago, or in a more recent but different market situation, is not necessarily still valid today. We have already touched on this point, which implies that rules for the revisions of validation are necessary.

The last two points state that the information to use in valuation is all and only the information relevant to the instrument under analysis. All market information must be collected, but irrelevant market information should be excluded. The document explains that they refer mainly to the cases considered by accountants, such as distressed sale. One element to add here is that even liquid and not distressed market information can be irrelevant. There was, particularly before the crisis, a tendency to overextend the calibration set of models, under the illusion that the more market information included, the better. In principle this is not wrong, but it is useless when market information is not relevant to the risk we want to assess. We will see an example of this kind in the next chapter which describes gap risk and option calibration. The risk of this overcalibration approach is that it creates a false sense of confidence. Apparently it increases the robustness of the model, but often it only makes the model less stable and hides the presence of remaining uncertainty on model features that are calibrated to no liquid market but may be crucial for the complex bespoke derivative under evaluation.

1.4.3 Basel New Principles: Operative Recommendations

1. *‘Validation includes evaluations of the model’s theoretical soundness and mathematical integrity and the appropriateness of model assumptions, including consistency with market’*
2. *‘Bank processes should emphasize the importance of assessing fair value using a diversity of approaches and having in place a range of mechanisms to cross-check valuations’*

3. *'A bank is expected to test and review the performance of its valuation models under possible stress conditions, so that it understands the limitations of the models' 'assess the impact of variations in model parameters on fair value, including under stress conditions'*
4. *'Policies should also identify specific triggers (eg indications of deterioration in model performance or quality) that will cause the review'*

These four points give a precise indication of a few crucial steps that should not be omitted in model validation. They include the appropriateness of model assumptions. The considerations contained in the document make this appropriateness of model assumptions very akin to the model realism at the centre of the Value approach. Further, the consistency with the market, central to the Price approach, is explicitly mentioned as an element to be included. It also recalls the need for model verifications, that includes control of the mathematical correctness.

The second point recalls the need for a diversity of approaches when assessing fair values. This seems in striking contrast to some previous regulatory approaches that emphasized the use of simplified models to be taken as the unique reference, also to guarantee homogeneity across banks. This fuelled regulatory arbitrage and pro-cyclicality. Now not only is this sort of homogeneity not required across banks, but it is not needed even internally. This makes sense: only the comparison of really different models gives us a clear view of our model uncertainty, and outlines the advantages and the disadvantages of each single model. We will focus on this in the next chapter.

The third point is stress testing, the topic of Chapters 3 and 6. Stress testing has been heavily emphasized by regulators in the last few years, with a double meaning that appears also in the recommendations of this document. On the one hand we should test the performance of models under stress conditions, to understand model limitations. This is what in the following we will call 'model stress testing'; the focus in these tests is on what can 'break' the model. On the other hand, we should stress the parameters to see how fair value changes. In this case the focus is on understanding, using a model, how the value of a derivative or a portfolio can change. This helps us understand which risk factors are crucial for our valuation, and which model parameters are associated with maximum risk. This is what in the following we will call 'portfolio stress testing'.

Stress testing often allows us to understand market conditions that, should they become material, would require a revision of our model validation. This helps set precise quantitative indications on when this should be done. The quantitative triggers that warn us that the model reliability is deteriorating are the focus of point 4. We will see examples of this for the rates market in 2007, for correlation models, and in the analysis of approximations.

1.5 MODEL VALIDATION AND RISK MANAGEMENT: PRACTICAL STEPS

We have seen the different approaches to model risk, the foundations of our modelling framework, the role of fair value and accountancy in modelling, and the latest prescriptions by regulators. From this we can deduce a practical scheme with the steps to follow, first in model validation and then in model risk management. These steps will be the topics of the next chapters. Here we give only a synthetic description of them, since a deeper understanding can be reached only through a detailed description and examples. Bullet points are used to indicate the steps in the process, ★ are indications that relate to more than one step, and ■ are the provisions to be put in place to implement the process.

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

Model Validation and Model Risk Management

1 Model Verification
- Mathematics
- Implementation
- Computational Methods
- Payoff Description

2 Model Validation
- Calibration
- Reasonableness
- Market Intelligence
- Reality Check

★ Model Comparison
- ■ Accept/Reject
- ■ Improvements
- ■ Provisions
- ■ Periodic Revisions

3 Stress Testing
- Stressability
- Computational methods
  - A Payoff features
  - B Market conditions
- Model assumptions
  - A Payoff features
  - B Market conditions
- ■ Improvements
- ■ Provisions
- ■ Triggered Revisions
- ■ Payoff Limitations

★ Model Evolution
- ■ Provisions
- ■ Triggered Revisions

Figure 1.7 A scheme for Model Validation and Model Risk Management

1.5.1 A Scheme for Model Validation

1. **Model Verification (check of the internal consistency)**
The Value approach recommends checking the mathematics and the implementation of a model. The alternative Price approach requires mathematical and numerical correctness in order to get prices similar to those of the other players. The regulators speak of assessing 'mathematical integrity'. In the natural sciences this is called Model Verification: checking that model assumptions are turned correctly into prices and sensitivities. We can distinguish three verifications:
   - **Mathematics**
   Often, a model starts by assuming a dynamics for a financial variable. Then the model developer solves a stochastic differential equation to see which distributions are produced from this dynamics, and from the distributions a number of useful formulas are computed. Examples of this type of model development are short rate models like Vasicek or CIR. In other cases the model starts from a distribution and then derives a consistent dynamics. This is the case for most local volatility models. When a model is used for the first time the passages from dynamics to closed-form formulas or the other way around should be verified. This should be the case for any new model developed by a front office quant, also for any new model that simply appears on the internet – or in a journal. There are errors even in published literature, never be too trusty.
   - **Implementation**
   Verifying a model implementation is not just a boring job where thousands of lines of code must be checked out. Instead, for a quant or a trader, it is an art. There are a number of clever tests that can be performed with a software to ensure that the software is consistent with the underlying modelling assumptions. Although, this book is not

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

about debugging, it is about understanding the consequences of model assumptions and the special cases that reveal the differences between models. Thus everything that follows here can be of some use in verifying the correctness of an implementation.

- **Computational Methods**
  When a price or a sensitivity cannot be computed exactly in closed form, alternative computational methods have to be used. These have to be validated. The most relevant case, from the point of view of model risk, are approximations. They must be tested in a special case scenario when exact methods are available, or against alternative methods like Monte Carlo simulation, which is not exact but can be made very precise at the cost of high computational effort. Moreover, the implementation of Monte Carlo simulation and other numerical methods such as binomial trees must be verified, usually in special cases where exact methods are available. The first verification of approximations and numerical methods is done under normal conditions, namely with the parameters we find in calibrating the model to today's market conditions. The analysis under extreme conditions is part of stress testing (although it can often be practically anticipated here). Both the analysis under normal conditions and the one under stressed conditions are described in Chapter 6.
- **Payoff Description**
  Next we have to verify that the implementation gives a correct description of the payoff. It is recommended explicitly by Derman, and it is a prerequisite for the Rebonato's goal to price derivatives consistency with market consensus (if the payoff was wrongly described, the price would be different even if we are using the consensus model). The errors made in the representation of the payoff are of two sorts. First there are 'the delta-one trader errors': wrong day-count conventions (actual/360 rather than actual/actual...), errors in the start date (it is today+2 days, not today...) and so on. These errors are important, particularly for traders in rather linear payoffs where money is made out of little spreads on large notionals. But for complex derivatives, they are usually not the core risk, and they have little to do with model risk. Second, there are errors in the interpretation of the payoff which are more sneaky. Spotting them requires a minimum experience of legal issues, documentation and market practice, and also the intuition of a modeller. Chapter 10 describes some potential errors in the representation of a payoff.

2. **Model Validation**
   Here we move from a check of internal consistency and correct description of the payoff to the analysis of model assumptions. This is proper Model Validation, and is described in detail in the examples in Chapter 2 where we compare different models applied to the same payoff. Having a plurality of models is crucial here.

   First, all models under analysis must be assessed on the following two points:
   - **Calibration:** Calibrate with acceptable precision the relevant market information. The topic of calibration is covered specifically in Chapter 9, with a focus on the model uncertainty that can remain after calibrations, and on enriching a calibration procedure to reduce this indetermination. The topic of calibration, however, is so crucial that it is covered in the next two chapters, then again in Chapter 5 where we see the links between calibration and hedging, and in Chapter 11 where the goodness of calibration to two markets jointly is used to spot arbitrage opportunities. In Chapter 7.1 we see how a clever calibration can avoid an extrapolation. Here we must also discuss *computational efficiency*. Due to the standard of calibrating models to a large number of derivatives, and of refreshing this calibration as often as possible, efficiency in pricing the calibration

products becomes a prerequisite of a model, even for models that we allow to be time-consuming when pricing non-calibration products. The topic is important for users and should not be completely dismissed by validators: alternative models with very inefficient calibration will be difficult to apply regularly, although they can be used to compute reserves.

- **Reasonableness:** The models must not be plainly unrealistic on elements of market behaviour which are crucial to the value of the derivative under analysis. We know that there exist no perfectly realistic assumptions, all models involve simplifications. However, we must make sure that the model is not senseless for the product at hand. Examples are given in the next chapter, where we will also see that some models do not pass the calibration or reasonableness tests when they are in their standard form, but that they can often be improved so as to pass these preliminary tests.

With the models that have passed the previous two tests, we move to model selection and assessment of model uncertainty through the following analysis.

- **Market Intelligence.** Knowledge of the market consensus on the pricing of a payoff, if there is one, is crucial given todays' practice of marking derivatives to market. A good calibration to simpler derivatives does not always imply that we are using the consensus model for the complex derivative of interest. The tools to use in order to perform market intelligence have already been explained in Section 1.1.2, and here we just revisit them:
  (a) Visible deals, novations, unwindings
  (b) Collateral regulation
  (c) Broker quotes and consensus platforms
  (d) Conferences
  (e) Salesforce information
  (f) Contacts at other institutions

  a, b and c are useful if we are able to perform *reverse engineering* of market quotes in order to understand which models may have produced them. There are a number of examples in the following where we perform reverse engineering. For instance, we explicitly perform reverse engineering of counterparty quotes in 2.6.2, but we also do it to understand how the market consensus changed in Chapter 4. It is an issue so intrinsic to model risk management that it surfaces throughout the book.
- **Reality Check.** We have seen in Section 1.1.4 that realism in the end is important for all model approaches, and even the fair value principle can deviate from consistency with observable prices when they are plainly unrealistic. We must understand which are the crucial risks that affect a payoff, and then use historical analysis, knowledge of the underlying market or plain common sense to understand which models give a reasonable description of these crucial risks (we can neglect those elements of unrealism that do not affect pricing in a crucial way). The issue is touched upon again when speaking of copulas applied to the risk of concentrated losses, when validating mapping against historical scenarios, when discussing the so-called P&L analysis, and also in relation to various other topics.

★ **Model Comparison.** The last two steps can be performed even on a single model, but they make more sense if applied to a family of models. In fact, it is unlikely that the first model analyzed will appear entirely satisfactory from both points of view, thus the task is usually to compare different models in order to understand which one is 'more realistic' or 'more consistent with market consensus'. Derman says explicitly that the use of a battery of models, is crucial, while Rebonato (2003) says 'In order to carry out this task the risk manager will need a variety of properly calibrated valuation models'. In fact, a battery of

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

models is needed to compare the model prices with the market quotes and to understand, via reverse engineering, which model is more consistent with the market quotes. Regulators suggest the use of '*a diversity of approaches*' and '*a range of mechanisms to cross-check valuations*'.

- **Accept/Reject – Improvements – Provisions.** At this point we have to make a first choice about the model to be used for our payoff. We can reject a model, if any serious problems emerged at any of the previous steps, and we can suggest another route, possibly towards one of the models considered at the Model Comparison step. We can suggest improvements in the direction of increasing realism or making the model more consistent with the market standard consensus. We can also accept the model but establish some provisions to control model risk: **Model Reserves**, usually based on some model that emerged in the Model Comparison step, or a **Model Limit** based on exposure if elements emerged in the Reality Check or Market Intelligence step that require caution. See Section 2.7 for more information about provisions.
- **Periodic Revisions.** Specify dates at which the validation must be revised. At this stage we will set periodic revisions, which are scheduled regularly, for example yearly. After one year both the experience in the use of the model and the knowledge of the market where it is applied will have increased, allowing for real improvements. Also market evidence, or the bank's exposure, may have changed.

3. **Stress Testing**

The topic of stress testing, crucial for regulators as we saw in 1.4, is covered partially in Chapter 2, where we see how the use of alternative models can be exploited as stress testing and at length in Chapter 3, where we show different aspects of stress testing and finally in Chapter 6 which covers the testing of computational methods.

- **Stressability.** Verify that it is possible, and relatively easy, within the model, to represent alternative financial scenarios, including extreme scenarios. Lack of this feature should set alarm bells ringing. It means that the model is of no use for understanding the risks of the payoff under stressed conditions, and that we have difficulties in understanding how the model will behave in future market events.
- **Testing computational methods.** Here the analysis done above to validate the non-exact methods is extended to stressed configurations. This point is very important because the precision of non-exact methods such as approximations usually relies on features of the market that are not necessarily permanent, or on specific features of a given payoff such as a short maturity. In part, the same applies to Monte Carlo simulation. The goal is to find out which are the potential conditions that can reduce the precision of our computational methods. There are two types of stress we want to test for:
  A. **Stress payoff features:** maturity, start date (try it forward), strike . . . Stress the features of the counterparty, of the issuer or the reference entity if applicable; stress the number of components if we are speaking of a portfolio derivative.
  B. **Stress market conditions:** design market conditions that you can translate into stressed values for the main parameters: volatilities, correlations, smile-related parameters. Try to understand what happens if you explicitly assume one parameter to be stochastic or, more simply, uncertain (see Chapter 11). Notice one important thing: it is usually better to start from market scenarios and derive parameters from them. This is not always easy. For example, changes in the ATM swaption matrix may correspond for the Libor market model of Chapters 6 and 9 to either correlation moves or volatility moves. In these cases it is more revealing to stress the parameters directly.

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

- **Testing model assumptions** This step involves stress testing to see if the model, which we have judged reasonable at the previous steps, remains such for stressed conditions. The focus is to check that the model does not start giving answers that contradict logic or market experience, and continues to take into account the main risks affecting the payoff. If the model was today judged to be in line with market consensus but problems arise in this part of stress tests, we have to take into account that market consensus is likely to abandon the model under those conditions which make the model appear unreasonable. Here we have the usual two types of stress to perform, but from a different perspective.
  A. **Stress payoff features**
  B. **Stress market conditions**
- ■ **Improvements-Provisions** Usually, a model is not rejected as a consequence of a stress test. When a problem emerges under stressed conditions, the most typical intervention is to improve the model to deal with the problematic condition or to introduce provisions to mitigate this risk. Of particular importance is the following provision:
- ■ **Triggered Revisions** Now we have enough information to specify a different sort of revision, not scheduled regularly but contingent on the occurrence of specific events. When a model has an undesirable behaviour under stressed conditions, we must set thresholds or triggers on calibrated parameters or on market values so that we recognize if we are approaching the critical condition. When this happens a revision of the validation is in order. Many banks include unexpected difficulties in calibration among the events that may trigger a revision of model validation. The revision will probably involve the whole validation process if we have detected a problem in relation to model assumptions; if instead the problem has emerged when stressing computational methods, only the validation of the computational method must be revised.
- ■ **Payoff Limitations** When instead a model exhibits undesirable behaviour for variations of the payoff, or for particular issuers/reference entities, or for a particular range of underlying products, this must be reported so that validation is not extended to these cases, or provisions are set.

★ **Model Evolution**

The Reality Check or the Stress Test may already have revealed that a model, perhaps consistent with today's market consensus, has some problems that future market conditions could render intolerable, triggering a consensus shift. This final point is to make sure that when a model has been chosen and validated we make an explicit analysis of the conditions that could make it inadequate in the future, for us and for the market in general. I consider this step important and devote Chapter 4 to recent examples of changes in market consensus. For example, risk is very high, although it may be hidden under the surface, when there is a market consensus but we have judged this consensus to be based on unrealistic assumptions.

- ■ **Provisions** Such a situation can be managed with provisions like limits to exposure or reserves based on a model more robust than the market approach. One should also take into account the considerations of Remark 3 about the limited sophistication of the models that usually emerge immediately after a crisis, and the possibility that we are in a model bubble as described in Section 1.2.1.
- ■ **Triggered Revisions** Once the conditions that could lead to a shift in the market consensus have been exposed, a quantitative measure of the distance from these conditions must be monitored, and when the distance is reduced, a revision of validation is advisable.

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

1.5.2 Special Points in Model Risk Management

The following points should be given particular attention in model risk management, within the above scheme or separately, because they cover cases where model risk is particularly high, and appears in peculiar ways. Each of these cases is dealt with in a dedicated chapter in the second part of the book.

- **Hedging**
  Valuation models are not used only for valuation at a point in time, but also for dynamic hedging. Hedging is related to model risk management in a double way. First, the assessment of the hedging behaviour of a model, the so-called P&L analysis, can give relevant indications about the behaviour of the model, supplementing the above consideration about the quality of calibration or the realism of the model. However, in order to be useful in model validation, the P&L analysis must be performed under conditions which are often not those under which a trader will really operate, in particular in real hedging a trader will not use the pricing model consistently, but will introduce various adjustments.
  This leads to the other aspect that links hedging to model risk management: the trading strategy implemented using a validated valuation model is not automatically validated, since the model-implied hedging strategy is usually different from the one really implemented, so the latter may deserve a separate validation. Both aspects are covered in Chapter 5.
- **Extrapolation**
  Extrapolations and interpolations are often used in modelling. There are even those who say, as we pointed out in Chapter 7, that modelling in financial markets is only about interpolating and extrapolating a few points that can now be observed in the market. However, in some cases the use of interpolations and in particular of extrapolations, is very explicit. Extrapolation is often used to supplement calibration. When we are using a model, we calibrate it to a narrow area of liquid quotes, and then we use it to create the information corresponding to an artificial market outside this area. This artificial market carries more model risk than the liquid one. At other times, when we are using a quotation system or some parameterization, we even have to fix parameters that we cannot calibrate, and we often do this via extrapolation. Here the risk is even higher. The use of extrapolation can be avoided or minimized, as we show in the examples in Chapter 7.
- **Correlation**
  The modelling of correlation, and more generally the modelling of interdependence among financial variables, should not in principle differ from the modelling of individual financial variables. Instead, it involves more model risk, since correlation is technically more difficult to model than other parameters. Just consider the fact that it is a matrix rather than a single number like volatility. Moreover, correlations are usually observable only historically since there are much fewer quotes giving implied correlations than there are quotes giving implied volatilities. Add to this that correlations tend to be very unstable parameters, so that historical estimation tells us very little about the likely future. This is the reason to single correlation modelling out. We cover it in Chapter 8, treating both the technical difficulties in modelling and risk of making wrong assumptions about interdependency.
- **Arbitrage**
  Arbitrage here is the attempt to secure a profit emerging from the inconsistencies that a model can detect in a market. This is very different from the standard use of models to price financial products. In pricing we start by implicitly assuming that all parts of the market are consistent in the way dictated by our models. Then we calibrate to observable information

and we find a consistent price for a less liquid derivative. The alternative way to use models typical of statistical arbitrage or model arbitrage carries a different model risk, and requires us to manage it differently. The topic is developed, based on two examples, in Chapter 11.

1.5.3 The Importance of Understanding Models

I hope the practical steps we have looked at will help you in the process of model development, model validation, and then in the use of models in practice. But I stressed in the Preface that this book is not about strict rules and bureaucratic checklists, therefore there is an important disclaimer that goes here. It is not the fact of fulfilling the above points, or any other list of points, that can make us effective model developers, validators or users. Checklists are mainly useful for giving some structure to the analysis of a model, and give us hints to look at a model and its application from different perspectives. This is useful since it helps build a deep understanding of how a model works and how it can operate in applications, which is in the end the crucial element allowing good management of model risk. In the previous analysis and in the examples that follow, the crucial part of any achievement in detecting and managing model risk is an understanding of the model behaviour in financial reality, from how the model is built to how it is applied.

The recent credit crunch crisis has further revealed the importance of understanding the models we are using. As for the Gaussian copula, we have seen errors in the way data were used to feed the model, and errors in the way the model was applied, which suggests little understanding of how the model was starting from formal hypothesis to generate numbers, prices and ratings. A deeper stress analysis would have probably helped. The formal correctness of the model, fed with data coming from historical estimations, seemed reassuring to many players. In fact, mathematics and numerics played the negative role of hiding what the model was really doing, namely extrapolating to the future the reality of the previous decade, in a simplified way that minimized the likelihood of a crisis.

In Section 3.3 we will see that reasoning about the nature of correlations in a Gaussian copula suggests a parameterization that introduces the possibility of a systemic crisis, without changing the model but just through an alternative design of correlations. This would have led to a still simplified but more sensible market practice that would have increased the understanding of model implications. In the same chapter we see two more analysis. The first one reveals that the Gaussian copula model chosen for CDOs was badly specified for capturing one of the risks that instead was most relevant at the burst of the crisis, the risk of loss concentration in short time. The second analysis shows that the methods that were used to give parameters to the Gaussian copula for pricing bespoke CDOs, as an alternative to historical analysis, were also biased towards an underestimation of risk for many CDOs. The problem was a poor understanding of the implications of the mathematical model chosen.

Derman (1996) says that *'Having a valuation model doesn't absolve the model user from thinking about the value of a security. Instead, it makes the security value a dependent variable, and requires the user to think about and estimate the values of other independent variables that are easier to grasp and quantify'*. This is the best job a model can do: to improve the understanding of the value of a security, making it easier to reason about it. If the model itself is poorly understood it cannot play this crucial role.

In all parts of this book, understanding the models we use is the primary task. In the next chapter, for example, we will see that only a deep comprehension of the model allows us to

P1: TIX
JWBK527-c01 JWBK527-Morini September 2, 2011 15:10 Printer: Yet to come

understand if it has unrealistic features, and it is also the only way to say, through reverse engineering, if a market quote can come from this model.

My view is that the most dangerous part of model risk is not that the model 'is wrong', but that *the model is not understood*. In Section 1.4 we reported the popular quote 'all models are wrong but some are useful'. Only if we have understood a model deeply will we be able to tell its unavoidable weaknesses from its useful part, and exploit the latter in the best possible way. *Model misunderstanding is the core of model risk*, and it is also that part of model risk that we can really commit to avoid. For the rest, we know that there can be parameters and assumptions in our model that will turn out wrong. If we understand the model properly, however, this will be recognized sooner and our reaction will be more effective.