HJM: A Unified Approach to Dynamic Models for Fixed Income, Credit and Equity Markets

René A. Carmona[1]

Bendheim Center for Finance
Department of Operations Research & Financial Engineering,
Princeton University, Princeton, NJ 08544, USA
email: rcarmona@princeton.edu

Summary. The purpose of this paper is to highlight some of the key elements of the HJM approach as originally introduced in the framework of fixed income market models, to explain how the very same philosophy was implemented in the case of credit portfolio derivatives and to show how it can be extended to and used in the case of equity market models. In each case we show how the HJM approach naturally yields a consistency condition and a no-arbitrage conditions in the spirit of the original work of Heath, Jarrow and Morton. Even though the actual computations and the derivation of the drift condition in the case of equity models seems to be new, the paper is intended as a survey of existing results, and as such, it is mostly pedagogical in nature.

1 Introduction

The motivation for this paper can be found in the desire to understand recent attempts to implement the HJM philosophy in the valuation of options on credit portfolios. Several proposals appeared almost simultaneously in the literature on credit portfolio valuation. They were written independently by N. Bennani [3], J. Sidenius, V. Piterbarg and L. Andersen [26] and P. Shönbucher [41], the latter being most influential in the preparation of the present survey. After a sharp increase in volume and liquidity due to the coming of age of the single tranche synthetic CDOs, markets for these credit portfolios came to a stand still due to the lack of dynamic models needed to price forward starting contracts, options on options, ... So the need for dynamic models prompted these authors to build analogies between the original HJM approach to interest rate derivatives and derivatives on credit portfolio losses. The common starting point of these three papers is the lithany of well documented shortcomings of the market standard for the valuation of Collaterized Debt Obligations (CDOs). The Gaussian copula model on which the standard is intrinsically a *one period* static model which cannot be used to price forward starting contracts. The valuation by expectation of these forward starting contracts require the analysis of a term structure of forward loss probabilities. The HJM modeling of the dynamics of the forward instantaneous interest rates, suggests how to choose dynamic models for the these forward loss probabilities. The three papers mentioned above try to take advantage of this analogy with various degree of generality and success.

The goal of this paper is to review the salient features of the HJM modeling philosophy as they can be applied to three different markets: the fixed income markets originally considered by Heath, Jarrow and Morton, the credit markets and the

* This research was partially supported by NSF DMS-0456195.

equity markets. In each of the three cases considered in this paper, the financial market model is based on a set of financial securities which are assumed to be liquidly traded. A basic assumption is that the price of each such security is observable, and any quantity of the security can be sold or bought at this observed price. These prices are used to encapsulate what the market is telling the modeler, and the thrust of the HJM modeling approach is to postulate dynamical equations for the prices of all these liquid instruments and to check that the multitude of all these equations does not introduce inconsistencies and arbitrage in the market model.

The classical HJM approach is reviewed in Section 3. Our informal presentation does not do justice to the depth of the original contribution [14] of Heath, Jarrow and Morton. It is meant as a light introduction to the modeling philosophy, our main goal being to introduce notation which are used throughout the paper, and to emphasize the crucial steps which will recur in the discussion of the other market models. Section 5 is devoted to the discussion of the recent works [26] of Sidenius, Pitterbarg and Andersen and [41] Schoenbucher on the construction of dynamic models for credit portfolios in the spirit of the HJM approach. These two papers are at the root of our renewed interest in the HJM modeling philosophy. It is while reading them that we realized the impact they could have on the classical equity models. The latter are usually calibrated to market prices by constructing an implied volatility surface, or equivalently a local volatility surface as advocated by Dupire and Derman and Kani in a series of influential works [19][16]. As we explain in Section 6, the construction of these surfaces is only the first step in the construction of a dynamic model. A dynamic version of local volatility modeling was touted by Derman and Kani in a paper [17] mostly known for its discussion of implied tree models. Motivated by the fact that the technical parts of [17] dealing with continuous models are rather informal and lacking mathematical proofs, Carmona and Nadtochiy developed in [7] the program outlined in [17]. On the top of providing a rigorous mathematical derivation of the so-called drift condition, they also provide calibration and Monte Carlo implementation recipes, and they analyze the classical Markovian spot models as well as stochastic volatility models in a generalized HJM framework. We present their results in the last section of this paper.

Acknowledgements. I would like to thank Dario Villani and Kharen Musaelian for introducing me to the intricacies of the credit markets. Their insights were invaluable: what they taught me cannot be found in textbooks !!!

2 General Mathematical Framework

This section is very abstract in nature. Its goal is to set the notation and the stage for the discussion of a common approach to three different markets.

2.1 Mathematical Notation

Throughout this paper we assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\{\mathcal{F}_t\}_{t\geq 0}$ is a right continuous filtration of sub-σ-fields of $\mathcal{F}$, $\mathcal{F}_0$ containing all the null sets of $\mathbb{P}$. Most often, we assume that this filtration is a Brownian filtration in the sense that it is generated by a Wiener process $\{W_t\}_{t\geq 0}$. We allow this Wiener process to be multi-dimensional, and in fact, it can even be infinite dimensional. The facts from infinite dimensional stochastic analysis which are actually needed to prove the results discussed in this paper in the infinite dimensional setting can be found in many books and published articles. Most of them can be derived without using too much functional analysis. For the sake of my personal convenience, I chose to refer the interested reader to the book [9] for definitions and details about those infinite dimensional stochastic analysis results which we rely upon.

In order to compute cash flow *present values*, we use a discount factor which we denote by $\{\beta_t\}_{t\geq 0}$. The latter is a non-negative adapted stochastic processes. Typically we use for β_t the inverse of the bank account B_t which is defined as the solution of the ordinary (possibly random) differential equation:

$$dB_t = r_t B_t\, dt, \qquad B_0 = 1, \tag{1}$$

where the stochastic process $\{r_t\}_{t\geq 0}$ has the interpretation of a short interest rate. In this case we have

$$\beta_t = e^{-\int_0^t r_s\, ds}. \tag{2}$$

Notice that $\{\beta_t\}_{t\geq 0}$ is multiplicative in the sense that

$$\beta_{s+t}(\omega) = \beta_s(\omega)\beta_t(\theta_s\omega), \qquad\qquad \omega \in \Omega,$$

where $\{\theta_t\}_{t\geq 0}$ is a semigroup of shift operators on Ω. For the sake of illustration, we should think of the ω's in Ω as functions of time, in which case $[\theta_s\omega](t) = w(s+t)$.

We shall assume that $\mathbb{P}$ is a pricing measure. This means that the market price at time $t = 0$ of any liquidly traded contingent claim which pays a random amount ξ at time T, say p_0, is given by (notice that the pay-off ξ is implicitly assumed to be a $\mathcal{F}_T$ integrable random variable):

$$p_0 = \mathbb{E}\{\beta_T \xi\}$$

where $\mathbb{E}\{\,\cdot\,\} = \mathbb{E}^{\mathbb{P}}\{\,\cdot\,\}$ denotes the expectation with respect to the probability measure $\mathbb{P}$. In other words, $\mathbb{P}$ is a pricing measure if prices of contingent claims are given by $\mathbb{P}$-expectations of present values of their future cashflows.

If we also assume that the market is free of arbitrage, then the price p_t at time $t < T$ of the same contingent claim is necessarily given by the conditional expectation

$$p_t = \frac{1}{\beta_t}\mathbb{E}\{\beta_T \xi | \mathcal{F}_t\}$$

which shows that $\{\beta_t p_t\}_{t\geq 0}$ is a $\mathbb{P}$-martingale in the filtration $\{\mathcal{F}_t\}_{t\geq 0}$. In other words, if $\mathbb{P}$ is a pricing measure, the discounted prices are $\mathbb{P}$-martingales.

Notice that we do not assume that such a pricing measure is unique. In other words, we allow for incomplete market models in our discussion.

2.2 Liquidly Traded Instruments

We next assume that our economy is driven by a set of liquidly traded instruments whose prices at time t, we denote by P_t^α. We can think of the vector $\mathbf{P}_t = (P_t^\alpha)_{\alpha\in\mathcal{A}}$ of these observable prices as a state vector for our economy. We will not make the completeness assumption that

$$\mathcal{F}_t = \sigma\{\mathbf{P}_s;\ 0 \leq s \leq t\}, \qquad\qquad t \geq 0.$$

These instruments are fundamental for the analysis of the market, and a minimal requirement on a dynamical model of the economy will be that such a model provides prices for forward starting contracts and European call and put options on these basic instruments. In particular, at each time t, we should be able to compute the quantity

$$\mathbb{E}\{\beta_T (P_T^\alpha - K)^+ | \mathcal{F}_t\} \tag{3}$$

for every maturity $T > t$ and strike $K > 0$. Since a measure μ on the half line $\mathbb{R}_+$ is entirely determined by the knowledge of its call transform, i.e. the values of the integrals

$$\int_{\mathbb{R}_+} (x - K)^+ \ \mu(dx),$$

for $K > 0$, the knowledge at time t, of the prices of all the call options completely determines the distributions under the conditional measure $\mathbb{P}_t$, of all the random variables P_T^α for all $T > t$ and all $\alpha \in \mathcal{A}$.

Here, for each $t > 0$, we define the random measure $\mathbb{P}_t$ as the (regular version of the) conditional distribution given $\mathcal{F}_t$ of the discounted version of $\mathbb{P}$. In other words, $\mathbb{P}_t$ is characterized by the requirement that the equality

$$\mathbb{E}\{\beta_{t+T}\Phi\Psi \circ \theta_t\} = \mathbb{E}\{\beta_t \Phi \mathbb{E}^{\mathbb{P}_t}\{\beta_T \Psi\}\}$$

holds for all bounded random variables Φ and Ψ which are $\mathcal{F}_t$ and $\mathcal{F}_T$ measurable respectively.

Remark. Notice that if instead of simply requiring the knowledge of the prices of all the European call options we were to also require the knowledge of the prices of all the path dependent options, then for each $\alpha \in \mathcal{A}$, the entire (joint) distribution under $\mathbb{P}_t$ of $(P_T^\alpha)_{T \geq t}$ would be determined. In the situation of interest to us, only the one-dimensional marginal distributions of $\mathbb{P}_t$ are determined by the prices we can observe.

2.3 Dynamic Market Model

All the information about the market model should be contained in the specification of a pricing measure $\mathbb{P}$. However, as we explained earlier, it seems that a reasonable market model should

- *be consistent with the prices of the liquidly traded instruments quoted on the market*, in other words, the numerical values P_t^α observed on the market should be recovered as conditional expectations under the pricing measure $\mathbb{P}$ of the discounted cashflows of the corresponding instruments;
- *allow for the pricing of forward starting contracts (e.g. European call options on call options* using the identified liquidly traded instruments as underlyers. In other words, it should provide a way to compute the time evolution of the conditional (random) measures $\mathbb{P}_t$, or at least its marginal distributions.

The first bullet point involves simply reproducing the prices of the basic liquid instruments at time $t = 0$. It usually goes under the name of calibration. The restriction of the measure $\mathbb{P}$ to $\mathcal{F}_0$ is typically trivial and the computation of these prices involves only regular expectations with respect to $\mathbb{P}$ which can be computed at time $t = 0$. So this first bullet point does not seem to involve the dynamics of the stochastic evolution of the characteristics of the market model: it looks like a static requirement for a one period model.

On the other hand, the second bullet point involves information about the model (and hence the pricing measure $\mathbb{P}$) of a more dynamic nature. For this reason, if will appear to be preferable to specify this dynamic information about $\mathbb{P}$ by specifying $\{\mathbb{P}_t\}_{t \geq 0}$ as a stochastic process in the space of probability measures on the possible future time evolutions of the vectors $\{\mathbf{P}_{t+s}\}_{s \geq 0}$ of basic instruments. This is the main thrust of the HJM approach to fixed income market models as it was originally introduced by Heath, Jarrow and Morton, and this is the point of view we take to review in the remaining part of this paper, recent developments in modeling credit and equity markets.

3 The Classical HJM Approach

The goal of this section is purely of a pedagogical nature. It is not intended as a rigorous *exposé* of the original work of Heath, Jarrow and Morton: it is merely an informal discussion aimed at a very general audience. In the case of fixed income markets (also called interest rate derivatives markets), the simplest form of interest rate is the spot rate whose value at time t we denote by r_t. As we will emphasize in several instances, any market model needs to provide with the distribution of the stochastic process $\{r_t\}_{t\geq 0}$, even if its role is limited to the introduction of the bank account and the discount factor as in the previous section. Many market models have been based on the specification of the dynamics of this process. For this reason they are called *short rate models*. Despite the limitations which we are about to document, they remain very popular, mostly because of their versatility and the existence of closed form formulae for the prices of many liquidly traded instruments.

There are several sets of liquid interest rate derivatives actively traded and quoted daily. Coupon bearing bonds, caps, floors, swaptions, are some of them. But because most of them can be viewed as portfolios of zero coupon bonds, or European options on zero coupon bonds, and because this section aims at recasting classical material (which can be found in most financial mathematics textbooks) into the framework adopted in the paper, we find convenient to choose, for the set of liquidly traded securities, the ensemble of all the zero coupon non-defaultable bonds.

For the sake of definiteness, we denote by $B(t,T)$ the price at time t of such a zero coupon bond with maturity T. We shall often use the term ”Treasury” (which essentially means that the bond will not default) interchangeably with ”non-defaultable”. The entire face value will be paid at time T by the issuer of the bond to the buyer as long as $T > t$. So at time $t = 0$, all the prices $B(0,T)$ can be observed and the entire curve

$$T \hookrightarrow B_0(T) = B(0,T) \tag{4}$$

is known. So as stated in the first bullet point of Subsection 2.3 above, a first requirement for a model given by a pricing measure $\mathbb{P}$ is to reproduce these prices exactly.

As we are about to see, this innocent looking condition cannot always be satisfied by the short interest rate models which need to be re-calibrated frequently to satisfy, at least approximatively this requirement. Indeed, short interest rate models are endogenous term structure models as the initial term structure of zero coupon bond prices (4) is an output of the model instead of being an input observed in the market place. This last point is one of the main components of the HJM approach.

Since the cash flows of a zero coupon bond reduce to paying its nominal amount (which we conveniently normalize to 1) at time T, the price has to be given by

$$B_0(T) = \mathbb{E}\{\beta_T\} = \mathbb{E}\{e^{-\int_0^T r_s ds}\}, \tag{5}$$

recall that $\beta_0 = 1$. So if the parameters of the pricing measure $\mathbb{P}$ allow for the computation of the expectation in the above right hand side, the value of this expectation will have to coincide with the observed price $B_0(T)$ if we want to satisfy the first bullet point above.

Using Instantaneous Forward Rates Instead. For reasons that will become clear later, if the zero coupon prices $B(t,T)$ are (or assumed to be) smooth in the maturity variable T, it is more convenient to work with the forward rates defined by

$$f(t,T) = -\frac{\partial}{\partial T} \log B(t,T) \tag{6}$$

rather than the bond prices directly. Since the bond prices can be recovered from the forward rates

$$B(t,T) = e^{-\int_t^T f(t,u)\,du} \tag{7}$$

the term structure of interest rates can be given equivalently by the forward curves. In particular, observing all the bond prices $B_0(T)$ at time $t = 0$ is equivalent to observing all the forward rates $f_0(T)$, and the initial forward rate curve

$$T \hookrightarrow f_0(T)$$

can be the object of the calibration efforts (in the case of short rate models) or it can serve as initial condition (in the case of HJM dynamical models).

3.1 Short Rate Models

Since the prices of the basic instruments of the market can be computed as expectations over the short interest rate, recall formula (5), the simplest prescription for a pricing measure $\mathbb{P}$ is to describe the dynamics of the short rate process. Typically, a short rate model assumes that under the pricing measure $\mathbb{P}$, the short interest rate r_t is the solution of a stochastic differential equation of the diffusion form (i.e. Markovian):

$$dr_t = \mu^{(r)}(t,r_t)\,dt + \sigma^{(r)}(t,r_t)\,dW_t \tag{8}$$

where the drift and volatility terms are given by real-valued (deterministic) functions

$$(t,r) \hookrightarrow \mu^{(r)}(t,r) \quad \text{and} \quad (t,r) \hookrightarrow \sigma^{(r)}(t,r)$$

such that existence and uniqueness of a strong solution hold. For the sake of illustration, we consider only one specific example. Indeed, the goal of this section is not to present the theory of short rate models. They are mentioned only as motivation for the introduction of the HJM modeling approach.

We choose the **Vasicek** model because of its simplicity, but for the purpose of the present discussion, a **CIR** model of the square root diffusion could have done as well. In the case of the Vasicek model, the dynamics of the short rate are given by the stochastic differential equation:

$$dr_t = (\alpha - \beta r_t)\,dt \,+\, \sigma\,dW_t. \tag{9}$$

This equation is simple enough (linear) to be solved explicitly. The solution is given by

$$r_t = e^{-\beta t} r_0 + (1 - e^{-\beta t})\frac{\alpha}{\beta} + \int_0^t e^{-\beta(t-s)} \sigma dW_s. \tag{10}$$

$\{r_t\}_{t \ge 0}$ is a Gaussian process whenever r_0 is, and at each time $t > 0$ there is a positive probability that r_t is negative. Despite this troubling feature (not only can an interest rate be negative in this model, but it is almost surely unbounded below!), this model is very popular because of its tractability and because a judicious choice of the parameters can make this probability of negative interest rate quite small. The tractability of the model is due to the fact that the random variable $\int_0^t r_s ds$ is Gaussian with mean and variance which can be explicitly computed from the parameters α, β and σ of the model, and from this fact, one gets an explicit formula for the expectation (5) giving the price of the zero coupon bonds. We get:

$$B_0(T) = e^{a(T) + b(T) r_0} \tag{11}$$

where r_0 is the current value of the short rate, and where the functions $a(T)$ and $b(T)$ are given by:

$$b(T) = -\frac{1}{\beta}\left(1 - e^{-\beta T}\right) \tag{12}$$

and

$$a(T) = \frac{4\alpha\beta - 3\sigma^2}{4\beta^3} + \frac{\sigma^2 - 2\alpha\beta}{2\beta^2}T + \frac{\sigma^2 - \alpha\beta}{\beta^3}e^{-\beta T} - \frac{\sigma^2}{4\beta^3}e^{-2\beta T}. \tag{13}$$

Alternatively, if we use the forward curve instead of the zero coupon bond curve we get:

$$f(t,T) = re^{-\beta(T-t)} + \frac{\alpha}{\beta}\left(1 - e^{-\beta(T-t)}\right) - \frac{\sigma^2}{2\beta^2}\left(1 - e^{-\beta(T-t)}\right)^2. \tag{14}$$

from which we get an expression for the initial forward curve $T \hookrightarrow f_0(T)$ by setting $t = 0$. Notice that such a forward curve converges to the constant $(2\alpha\beta - \sigma^2)/2\beta^2$ when $T \to \infty$. This limit can be given the interpretation of a *long rate* (as opposed to the short rate) when $\sigma^2 < 2\alpha\beta$. In any case, a Vasicek forward curves flattens and becomes horizontal for large maturity T. The graph of a typical example of a forward curve given by the Vasicek model is given in the left pane of Figure 1. We used the parameters $\alpha = 13.06$, $\beta = 2.5$ and $\sigma = 2$ to produce this plot. We clearly see the flattening of the curve on the right part of the plot.

Rigid Term Structures for Calibration

As we explained earlier, choosing values for the parameters of the model (α, β and σ in the Vasicek model discussed in this section) in order for the model to reproduce the observed forward curve is what is usually called calibration of the model. Since the Vasicek model depends upon three parameters, three quoted prices, say $B_0(T_1)$, $B_0(T_2)$ and $B_0(T_3)$ for three different maturities T_1, T_2 and T_3 should in principle be enough to determine these parameters. But unfortunately, the curve $T \hookrightarrow B_0(T)$ constructed from formulae (11), (12), and (13) and three parameter values derived from three bond prices does not always look like the curve produced by the market quotes, and most importantly, it changes with the choices of the three maturities T_1, T_2 and T_3. For the sake of illustration, we give in the right pane of Figure 1 the plot of the market zero-coupon forward curve on 3/28/1996, and we super-impose on the same graph the plot of the best least squares fit among the possible forward curves produced by the Vasicek model. This optimal Vasicek forward curve was obtained for the values $\alpha = 13.06$, $\beta = 2.401$ and $\sigma = 1.724$ of the parameters. The fact that a Vasicek forward curves flattens for large maturity makes it impossible to match the typical increase in T found in most practical instances.

A Possible Fix

Several solutions have been proposed to the undesirable rigidity of the initial term structure curves produced by the short rate models. The most popular one is to force some of the coefficients to be time dependent in order for the model to match any market forward curve $T \hookrightarrow f_0(T)$. This is especially simple and useful in the case of the Vasicek model for if the time dependent coefficients are deterministic, the solution process remains Gaussian, and closed form solutions for the values of the forward rates and zero coupon prices can still be derived. To be more specific, formula (10) becomes

$$r_t = e^{-\int_0^t \beta_s ds} r_0 + \int_0^t e^{-\int_s^t \beta_u du}\alpha_s ds + \int_0^t e^{-\int_s^t \beta_u du}\sigma_s dW_s. \tag{15}$$

and since the conditional distribution of the integral $\int_s^t f_u du$ is Gaussian, bond prices

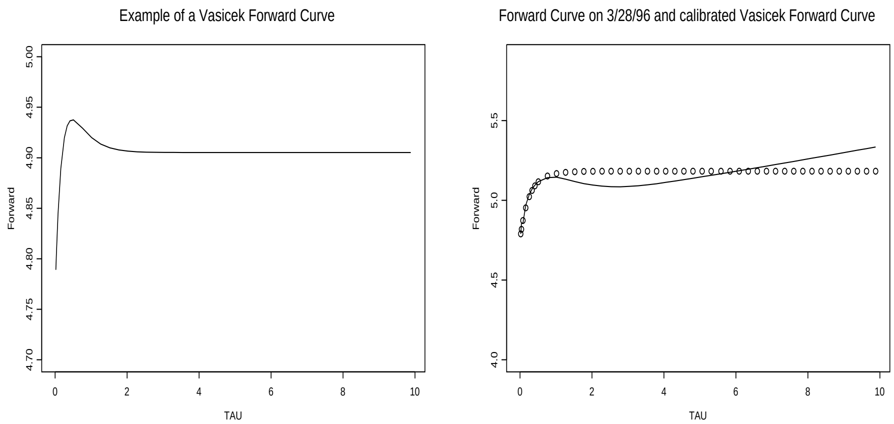

Fig. 1. Typical forward curve produced by the Vasicek model (left) and calibrated Vasicek forward curve (dotted line) to the zero-coupon forward curve on 3/28/1996.

$$B(t,T) = \mathbb{E}\{e^{-\int_t^T r_s ds} | \mathcal{F}_t\}$$

can still be derived from the expression of the Laplace transform of the Gaussian distribution.

This strategy was successfully implemented in the case of the Vasicek model (9) by Hull and White. These two authors proposed to leave the volatility σ and the mean reversion rate β constant, and to replace the parameter α by a deterministic function $t \hookrightarrow \alpha(t)$. In this case, the solution r_t is given by the formula

$$r_t = e^{-\beta t} r_0 + \int_0^t e^{-\beta(t-s)} \alpha_s ds + \sigma \int_0^t e^{-\beta(t-s)} dW_s, \qquad (16)$$

and the forward rate is given by the formula

$$f(t,T) = e^{-\beta(T-t)} r_t + \int_t^T e^{-\beta(T-s)} \alpha_s ds - \frac{\sigma^2}{2\beta^2}[1 - e^{-\beta(T-t)}]^2. \qquad (17)$$

If we replace in this formula t by 0 and $T-t$ by t, we get simple formulae for the initial forward curve and its derivative. From there one easily sees that it is possible to choose the function $t \hookrightarrow \alpha(t)$ to obtain any given (smooth) forward curve. To be specific, if we denote by $\overline{f}_0(T)$ the forward rate observed on the market at time $t=0$ for maturity T, then choosing

$$\alpha_t = \overline{f_0'}(t) + \beta \overline{f_0}(t) - \frac{\sigma^2}{2\beta}(1 - e^{-\beta t})(3e^{-\beta t} - 1)$$

will force the initial forward curve $T \hookrightarrow f_0(T)$ produced by the Vasicek model with this time dependent coefficient $\alpha(t)$ to coincide with the market (observed) forward curve $T \hookrightarrow \overline{f}_0(T)$. The model is now compatible with the current *observed* forward curve, it is *calibrated* to the market.

Model calibration is everyday practice in quantitative finance, and the procedures similar to the Hull-White modification of the Vasicek model are regarded as useful. But despite their popularity with practitioners, these calibration techniques remain problematic for several reasons.

Firstly, this fix is short lived for in general the adequacy of the modified model is limited to a short period. Indeed, the next time we check the forward curve given by the market, it will most likely not agree with the forward curve implied by the model, hence the need to recalibrate and changing the stochastic differential equation as we need to change its coefficients. The relevant question is then, when to recalibrate, and there is no theoretical answer to that question in general. The calibration procedure described above limits the usefulness of the model to a short time period, and de facto, turns a dynamic model into a *one period model*.

But there are other reasons to go beyond the short rate models. Indeed, specifying a short rate model amounts to specifying the (stochastic) dynamics of the whole forward curve by specifying the (stochastic) dynamics of the left-hand point of the curve. Indeed, $r_t = f(t,t)$, and this rigidity is confirmed by the fact that given any two maturities T_1 and T_2 the correlation coefficient between the *random variables* $df(t,T_1)$ and $df(t,T_2)$ is necessarily equal to 1!

Factor Models, Consistency and No-Arbitrage

Short rate models are particular cases of factor models of the terms structure of interest rates. They correspond to the case when the number of factors is one, and the sole factor is the short interest rate itself. More general factor models have been considered, and no-arbitrage conditions in the spirit of the discussion of this section have been derived at various levels of generality. See for example [9] or Proposition 2.2. of [24] for a sample condition.

Notation. We introduce a special notation $\tau = T - t$ for the time-to-maturity of a bond, or yield, or forward, etc. The forward rates (as well as the bond prices) are defined accordingly in terms of this new variable.

$$\tilde{f}_t(\tau) = f(t, t+\tau), \qquad \tau \geq 0. \tag{18}$$

Expressing the forward rates at time t in terms of time-to-maturity τ instead of time-of-maturity T has the advantage of forcing all the forward curves $\tilde{f}_t$ to be defined on the same domain $[0,\infty)$. This convenient notation is often called the Musiela notation.

We concentrate later on the no-arbitrage condition for general HJM models. For the time being, we discuss it in the context of factor models built from parametric families of forward or yield curves. These families are usually introduced in the following way. We start from a function G from $\Theta \times [0,\infty)$ into $[0,\infty)$ where Θ is an open set in $\mathbb{R}^d$ which we interpret as the set of possible values of a vector θ of parameters $\theta_1, \cdots, \theta_d$. In this way, for each $\theta \in \Theta$ the *curve* $G(\theta, \cdot) : \tau \hookrightarrow G(\theta,\tau)$ can be viewed as a possible candidate for the forward curve. For the sake of illustration we give the classical example of the Nelson-Siegel family defined by

$$G(\theta,\tau) = \theta_1 + (\theta_2 + \theta_3\tau)e^{-\theta_4\tau}, \qquad \tau \geq 0. \tag{19}$$

The parameters θ_1 and θ_4 are assumed to be positive. θ_1 represents the asymptotic (long) forward rate, $\theta_1 + \theta_2$ gives the left end point of the curve, namely the short rate, while θ_4 gives an asymptotic rate of decay. The set Θ of parameters is the subset of $\mathbb{R}^4$ determined by $\theta_1 > 0$, $\theta_4 > 0$ and $\theta_1 + \theta_2 \geq 0$ since the short rate should not be negative. The parameter θ_3 is responsible for a hump when $\theta_3 > 0$, or a dip when $\theta_3 < 0$. Other parametric families have been used, the most popular one being the Svensson's family. See [9] and the references therein.

We now introduce factor models from the notion of parametric family formalized above. We assume that we are given a parametric family G as before and we suppose that $\mathbf{\Theta} = \{\Theta_t\}_{t\geq 0}$ is a d-dimensional semi-martingale with values in the parameter space Θ. We then set

$$\tilde{f}_t(\tau) = G(\theta_t, \tau), \qquad t \geq 0,\, \tau > 0.$$

Recall that τ represents the time to maturity. The d components θ_t^j of θ_t are interpreted as economic factors driving the dynamics of the term structure of interest rates. Assuming further that G is twice continuously differentiable in the variables θ^j, we can use Itô's formula and derive the dynamics of $f_t(\tau)$.

As we assume that the measure $\mathbb{P}$ is the measure used by the market to compute prices, in this context, the absence of arbitrage is equivalent to the the fact that all the discounted bond prices $\{B^*(t,T)\}_{t\in[0,T]}$ are local martingales. Recall the discussion of Section 2. Here, the discounted bond price at time t for maturity T is given by

$$B^*(t,T) = \beta_t B(t,T) = e^{-\int_0^t r_s ds} B(t,T) = e^{-\int_0^t f(s,s)ds} e^{-\int_t^T f(t,u)du}. \tag{20}$$

since we are using the inverse of the bank account as discount factor. For each fixed $T > 0$, the process $\{B^*(t,T)\}_{0\leq t\leq T}$ is a local martingale if the drift in its Itô's stochastic differential is 0. Such a condition takes a particularly simple form in the case of a factor model defined as above and when the factors θ_t^j form a d-dimensional Markov diffusion. We refrain from giving the details as we are about to discuss the same condition in a more general setting. The interested reader is referred to [9] p.70 or [25] for details. In the literature on the classical HJM approach to fixed income markets, a pair $(G, \boldsymbol{\Theta})$ satisfying the no-arbitrage condition is said to be *consistent*. Again, see for example [25] and [9]. For the sake of consistency (!!), we use the same terminology in the present situation. The context will make clear whether we mean consistency with a spot model or absence of arbitrage for a factor model.

3.2 The Heath–Jarrow–Morton Approach

In the far reaching paper [14], Heath, Jarrow and Morton proposed to solve the above dilemma by modeling directly the dynamics of the entire term structure of interest rates, in other words, by modeling the dynamics of the forward curve. This seemingly minor change has dramatic consequences: *it kills two birds with one stone* in the sense that both bullet points of Subsection 2.3 are taken care of by this change. Indeed, calibration merely reduces to feeding the initial condition to the dynamical equation (this takes care of the first bullet point), and the time evolution of the conditional probabilities $\mathbb{P}_t$ follows from the same dynamical equation.

In order to be more specific, we consider a pricing measure $\mathbb{P}$, we choose the basic instruments to be at each instant t the discounted bond prices $\{B^*(t,T)\}_{T\geq t}$ defined by equation (20) above, and we assume that for each fixed maturity T, these discounted bond prices form a continuous local martingale for $\mathbb{P}$. The martingale property is our way to guarantee that such a market model is free of arbitrage opportunities. We explain below that enforcing this martingale property in a model leads to a constraint which is known under the name of drift condition. In their original proposal, Heath, Jarrow and Morton suggested to work with the forward rates $\{f(t,T)\}_{t\in[0,T]}$ instead of the actual bond prices. So instead of starting from a dynamical equation of the form

$$dB^*(t,T) = \sum_{i=1}^d \tau^{(i)}(t,T)dW_t^{(i)} \tag{21}$$

for some predictable processes $\{\tau^{(i)}(t,T)\}_{t\in[0,T]}$, they assume that the dynamics of the forward rates $\{f(t,T)\}_{t\in[0,T]}$ are given by stochastic differential equations of the form

$$df(t,T) = \alpha(t,T)dt + \beta(t,T)\cdot dW_t, \tag{22}$$

where the processes $\{\alpha(t,T)\}_{t\in[0,T]}$ and $\{\beta(t,T)\}_{t\in[0,T]}$ are assumed to be predictable with respect to the filtration generated by the Wiener process. Notice that both β and W can be multivariate (i.e. vector valued) in which case the above equation can be understood in developed form as

$$df(t,T) = \alpha(t,T)dt + \sum_{j=1}^{d} \beta^{(j)}(t,T)dW_t^{(j)}. \tag{23}$$

Notice that, as we mentioned in the introduction, the theory allows for $d = \infty$. See for example [8] or the contribution of Ekeland and Taflin to this volume. Note also that elementary stochastic calculus manipulations can be used to derive an equation of the form (23) from a starting point like (21), and conversely, it is easy to go from (21) to (23).

Finally, we note that the dynamics (23) are given by a large number of stochastic equations, one for each maturity T. Equivalently, this can be rewritten as a single equation for a function of T, in other words a semi-martingale given by the solution of a stochastic differential equation with values in a space of functions of T. Still another possibility is to view the forward rate $f(t,T)$ as a random field parameterized by t and T. The reader interested in the interactions between these three points of view is referred to [8].

A Spot Consistency Condition

In most cases of interest, the limit $\lim_{T\searrow t} f(t,T)$ exists almost surely for each fixed t, and as we already mentioned, this limit can be naturally identified with the short interest rate r_t. Such a process $\{r_t\}_{t\geq 0}$ defined as the left hand point of the forward curve is a semi-martingale and its stochastic differential can resemble a stochastic differential equation of the form we used to define short interest rate models, though it turns out that this is generally not the case. This definition of the short rate can also be viewed as a consistency restriction between the specification of the dynamics of the forward curve and the possible prescription of stochastic dynamics for the short rate. It is expressed as:

$$r_t = f(t,t). \tag{24}$$

The Original HJM Drift Condition

The discounted bond prices $B^*(t,T)$ can be written in terms of the instantaneous forward rates $f(t,T)$ as

$$B^*(t,T) = e^{-\int_0^t r_s\,ds} e^{-\int_t^T f(t,u)\,du}$$

and computing their stochastic differentials using the dynamic equation (23), and setting the resulting drift to zero gives another restriction on the coefficients α and β of (23). As explained in most financial mathematics textbooks, this constraint can be written as:

$$\alpha(t,T) = \beta(t,T)\cdot \int_t^T \beta(t,s)ds = \sum_{j=1}^{d} \beta^{(j)}(t,T)\int_t^T \beta^{(j)}(t,s)ds. \tag{25}$$

The above formula shows that the drift is completely determined once the volatilities have been chosen. It was discovered by Heath, Jarrow and Morton [14] and is widely known as the HJM drift condition.

Summary of the Approach

In order to highlight the main components of the HJM modeling philosophy we summarize the preceding discussion in a short list of a few bullet points.

- At any time t, we coded the prices of the liquidly traded instruments (i.e. the zero coupon bonds) by a forward curve.
- We prescribed stochastic dynamics for the elements of the code-book under the pricing measure $\mathbb{P}$.
- We derived a *consistency* condition which holds if the model has to coexist with a short rate model.
- We derived a condition guaranteeing the absence of arbitrage (the discounted prices of all the liquidly traded instruments are local martingales) which took the form of a *drift condition*.

These results are quite satisfactory from the theoretical point of view. However, the business of choosing the number d of factors and the actual volatility processes $\beta^{(j)}(t,T)$ still remains. This issue is especially thorny as the prices of the liquidly traded instruments are supposed to go in the initial condition and not in the choice of the volatility factors. There is no generally accepted solution to this difficult problem. The most popular approach relies on prices of more exotic instruments and the analysis in principal components for the determination of d and the $\beta^{(j)}(t,T)$'s. See for example [9].

4 First Extensions to Equity Markets

Before switching gear and extending the HJM approach to more complex code-books as in the case of credit and equity markets discussed in the following sections, we review two extensions of the HJM approach to the equity markets when the complexity of the code-book is the same as in the classical case described in the previous section where the liquidly traded instruments were coded with a mere one dimensional curve.

4.1 Realized Variance and Variance Swaps

The goal of this first subsection is to illustrate the HJM framework based on the stochastic dynamics of a family of curves with the example of a class of instruments traded on equity desks. It appears that, when dealing with equity models, both in this section and in Section 6, discounting does not play any significant role except for complicating the nature of the formulae. so without any loss of generality, we assume in both sections that the short interest rate is zero and hence that the bank account B_t and the discount factor β_t are identically equal to 1.

Variance swaps on a stock or an index promise the payment of the realized variance of the log-returns of the underlier to the holder of the swap. They are popular ways for investors to gain pure exposure to variance, or to hedge volatility products. Their prices are given by the expectation of this realized variance up to maturity. Assuming that they can be observed, instead of working from a model of the *spot variance* itself, we follow the approach proposed in [5] by Buehler who chooses to work directly with the dynamics of the entire implied variance swap curve, very much in the spirit of the HJM approach to the term structure of interest rates reviewed in the previous section.

To be specific, we define the annualized variance of a stock or index $S=\{S_t\}_{t\geq 0}$ over a period of n consecutive trading days $0=t_0<t_1<\cdots<t_n=T$ by

$$\hat{V}_n = \frac{252}{n}\sum_{i=1}^{n}\big(\log\frac{S_{t_i}}{S_{t_{i-1}}}\big)^2$$

where 252 represents the number of trading days in one year. We follow the market practice of not subtracting the mean of the daily log-returns. For that reason, $\hat{V}_n$ is not exactly a variance. In any case, a (mean zero) variance swap with maturity T and strike K is a contract which pays $\hat{V}_n - K$ at time T. Since the strike K appears merely as an additive factor, the following analysis will be done by assuming, without any loss of generality, that $K = 0$.

If we assume that the dates $t_0, t_1, \cdots, t_n$ form a partition of the fixed time interval $[0, T]$, and that the mesh of this partition (i.e. the number $\sup_{i=1,\cdots,n}|t_i - t_{i-1}|$) goes to 0, the realized variance $\hat{V}_n$ converges towards the quadratic variation $\langle \log S\rangle_T$ of the logarithm of the underlier. So for the purpose of the mathematical analysis of these instruments, we assume that a variance swap with maturity T pays the realized quadratic variation $\langle \log S\rangle_T$, and we denote by $V_t(T)$ the price at time t of such an instrument.

In this subsection, we assume that there exists a liquid market of variance swaps on the underlier S. This assumption may be far-fetched for most stocks, but it is quite realistic for the major stock indexes. In particular, at time $t = 0$, the prices $V_0(T)$ of variance swaps for all maturities $T \geq 0$ can be observed. At each time t, we use $\{V_t(T);\, T \geq t\}$ for the set of prices of the liquid instruments on which we base our financial market model, and we define a dynamic market model by specifying the stochastic time evolution of this set of prices.

If as before we assume that the market chose a pricing measure $\mathbb{P}$, and if the underlier spot price S satisfies

$$dS_t = S_t\sigma_t dB_t, \qquad t \geq 0,$$

for some Wiener process $\{B_t\}_{t\geq 0}$ and an adapted process $\{\sigma_t\}_{t\geq 0}$, then since we assume that $V_t(T)$ is the price of a liquidly traded instrument, requiring absence of arbitrage implies that:

$$V_t(T) = \mathbb{E}\{\langle \log S\rangle_T|\mathcal{F}_t\} = \mathbb{E}\big\{\int_0^T \sigma_s^2 ds\big|\mathcal{F}_t\big\}.$$

Throughout this paper, we assume that interest rate is 0 (and $\beta_t \equiv 1$) whenever we discuss equity markets.

As in the case of the HJM approach to the term structure of interest rates, we assume that for each fixed t, $V_t(T)$ is a smooth function of the maturity T, and we define the forward variance $v_t(T)$ as its derivative with respect to maturity. In analogy with the term structure of interest rated was coded by the instantaneous forward curve, we capture the term structure of realized variance by the forward variance curve $\tilde{v}_t$ defined by:

$$\tau \hookrightarrow \tilde{v}_t(\tau) = v_t(t+\tau), \qquad \tau \geq 0$$

where by definition $v_t(T) = \partial_T V_t(T)$. Notice that

$$\tilde{v}_t(0) = v_t(t) = \sigma_t, \qquad \mathbb{P} - a.s.$$

gives a simple form of the HJM spot consistency condition (24).

Notice also that, with the above notation, for each fixed T we must have:

$$v_t(T) = \partial_T V_t(T) = \mathbb{E}\{\sigma_T^2|\mathcal{F}_t\}, \qquad 0 \leq t \leq T,$$

which shows that for each fixed T, the process $\{v_t(T)\}_{0\le t\le T}$ is a martingale. Consequently, modeling its dynamics can be done by specifying that it has a semi-martingale decomposition of the form

$$dv_t(T) = \alpha_t(T)dt + \beta_t(T)\,dW_t$$

with $\alpha_t(T) \equiv 0$. So in this particular case, the HJM drift condition takes a trivial form.

The reader interested in factor models for the forward variance $v_t(T)$ and their consistency with no-arbitrage, as well as pricing and hedging of variance swaps in this setting is referred to [5].

Remark. The HJM framework also has bee applied to the commodity markets where most of the trading is done via forward contracts. So like in the case of the fixed income models reviewed in the previous section, the commodity forward markets can be characterized by the set of liquidly traded instruments formed by the forward contracts with a specific set of maturities. So the term structure of forward contracts is captured by a code-book of curves (functions of the date of maturity of the contracts), but since these forward contracts are traded, they must be martingales under the pricing measure chosen by the market and as in the case of the variance swaps markets, the HJM drift condition guaranteeing no-arbitrage is trivial. The drift condition is non-trivial only when the code-book is formed of non-traded instruments.

4.2 European Call Maturity Term Structure

The discussion of this subsection is motivated by the work [39] of Schönbucher on the term structure of implied volatility for a fixed strike K. Schönbucher's results were recently generalized in [42] by Schweizer and Wissel to the case of a fixed convex pay-off function h, when the hockey-stick function $h(x) = (x - K)^+$ is replaced by a general non-negative convex function. We review the main results of this more general version which includes for example power options whose pay-offs are given by the function $h(x) = x^\gamma$ for some $\gamma \ge 1$.

This analysis is made on the stochastic basis of a d-dimensional Wiener process $\mathbf{W} = \{W_t\}_{t\ge}$ on which the dynamics of the underlying stock price are given by an equation of the form

$$dS_t = S_t[\mu_t dt + \sigma_t dW_t^1] \tag{26}$$

where W_t^1 denotes the first component of $W_t \in \mathbb{R}^d$ and where $\{\mu_t\}_{t\ge}$ and $\{\sigma_t\}_{t\ge 0}$ are adapted stochastic processes to be specified.

In this application, as explained earlier, we assume that the market of liquidly traded instruments is formed by the contingent claims with maturity $T > 0$ and pay-off $h(S_T)$ where h is a single non-negative convex function fixed once for all. We denote by $C_t(T)$ the price of such a claim and by $\Sigma_t(T)$ the corresponding implied volatility. Under the assumption of zero interest rate,

$$C_t(T) = \mathbb{E}\{h(S_T)|\mathcal{F}_t\}$$

and $\Sigma_t(T)$ is the unique number σ which recovers the price $C_t(T)$ from the Black-Scholes formula, i.e. the solution of

$$B_h(t, S_t, T, \sigma) = C_t(T)$$

where $B_h(t, S_t, T, \sigma) = \mathbb{E}\{h(S_T)|\mathcal{F}_t\}$ and the expectation is over a geometric Brownian motion with drift $\mu_t \equiv r$ and volatility $\sigma_t \equiv \sigma$. The existence and the uniqueness of such a $\Sigma_t(T)$ are well known in the classical case of the hockey-stick pay-off function $h(x) = (x - K)^+$. We review these facts later in Section 6. In the present situation of a general convex pay-off function h, we need to use a simple no-arbitrage argument which shows that the prices of the call options to satisfy almost surely:

$$C_t(T_1) \le C_t(T_2) \qquad \text{whenever } t \le T_1 < T_2. \tag{27}$$

Indeed, if this inequality is violated with positive probability, it is possible to set up a costless portfolio at time t which can be re-balanced at time T_1 to provide risk-less profit at time T_2, with positive probability. See [42] Proposition 2.1 for details. Moreover,

$$h(S_t) \le C_t(T) \le h(0+) + S_t h'(\infty) \qquad \text{for all } t \le T,$$

the inequalities being strict if h is not affine and the spot process $\{S_t\}_{t\ge0}$ satisfies a mild non-monotonicity condition. These properties guarantee the existence and the uniqueness of the implied volatility in the general case.

The purpose of this subsection is to analyze dynamic models for which the prices of the liquidly traded instruments are coded by their respective implied volatility. In other words, for each time $t > 0$, we want to use the one-to-one correspondence

$$\{C_t(T);\, T \ge t\} \leftrightarrow \{\Sigma_t(T);\, T \ge t\}$$

as a code-book for these prices. We recast the current set-up in the HJM framework described in the previous section by having the implied variance $\Sigma_t(T)^2$ play the same role as the yield to maturity of a discount bond. So in full analogy with the original HJM approach, we replace the implied volatility code book by the code-book of *forward implied variances* $X(t,T)$ defined by

$$X(t,T) = \frac{\partial}{\partial T}\big((T-t)\Sigma_t(T)^2\big) \tag{28}$$

so that we have the familiar expression

$$\Sigma_t(T)^2 = \frac{1}{T-t}\int_t^T X(t,u)du. \tag{29}$$

A dynamic model for our equity market is then determined by prescribing for each maturity T, the dynamics of $X(t,T)$ in the form

$$dX(t,T) = \alpha(t,T)dt + \beta(t,T)dW_t, \qquad 0 \le t \le T. \tag{30}$$

In the previous section, we emphasized the simplifications provided by a switch to a notation system based on the time to maturity $\tau = T - t$. With this in mind we set $\tilde{C}_t(\tau) = C_t(t+\tau)$ and $\tilde{X}_t(\tau) = X(t, t+\tau)$ and the dynamic model is defined for each fixed $\tau > 0$ by:

$$d\tilde{X}_t(\tau) = \tilde{\alpha}_t(\tau)dt + \tilde{\beta}_t(\tau)dW_t, \qquad t \ge 0. \tag{31}$$

The fact that $\mathbb{P}$ is a pricing measure (by which we mean that the underlying spot process $\{S_t\}_{t\ge0}$ and price processes $\{C_t(T)\}_{0\le t\le T}$ of the the liquid instruments are both local martingales is essentially equivalent to a *spot consistency* condition

$$\sigma_t = \tilde{X}(t,0), \qquad \mathbb{P}a.s. \tag{32}$$

for all $t > 0$ (in full analogy with the classical HJM case), together with a *drift condition*

$$\tilde\alpha_t(\tau) = -\frac{\partial^2_{\tau\tau}\tilde C_t(\tau)}{\partial_\tau \tilde C_t(\tau)}\tilde\beta_t(\tau)\int_0^\tau \tilde\beta_t(u)du - \frac{1}{2}\partial_\tau \frac{\partial^2_{\tau\tau}\tilde C_t(\tau)}{\partial_\tau \tilde C_t(\tau)}\tilde X_t(\tau)\Big|\int_0^\tau \tilde\beta_t(u)du\Big|^2$$
$$- S_t\frac{\partial^2_{S\tau}\tilde C_t(\tau)}{\partial_\tau \tilde C_t(\tau)}\sigma_t\tilde\beta^1_t(\tau) - S_t\partial_\tau\Big(\frac{\partial^2_{S\tau}\tilde C_t(\tau)}{\partial_\tau \tilde C_t(\tau)}\Big)\tilde X_t(\tau)\sigma_t\int_0^\tau \tilde\beta^1_t(u)du.$$

Remarks. 1. We stated above that the spot consistency and the drift conditions are *essentially* equivalent to the absence of arbitrage because on the top of some natural technical assumptions, the proof also requires the smoothness of the pay-off function h. See [42] for details.
2. Making explicit the deep and profound relationship between the spot volatility σ_t and the implied volatility Σ_t was done in the more general setting of the full implied volatility surface $(T, K) \hookrightarrow \Sigma_t(T, K)$ by Durrleman in [20]. Despite the fact that his goals were different, most of the computations involved in the derivations of the results of [42] stated above can be found in one form or another in Durrleman's proofs.
3. The results reviewed in this subsection should also be linked to a recent work of Jacod and Protter who study in [27] the problem of the completion of a market by adding derivative instruments. Indeed, in so doing. they derive conditions very similar to the spot consistency and the drift conditions reviewed aove. As an added bonus, and if the equations were not technical enough, Jacod and Protter work in the more general set-up of semi-martingale dynamics with jumps.
4. The complexity of the drift condition (33) and the technicalities involved in its derivation are the main reason why dynamic models for the entire implied volatility surface have not been studied pursed. This is in fact the reason why Schönbucher in [39] and Schweizer and Wissel in [42] limit themselves to dynamic models for a cross section of the implied volatility surface. This complexity is also at the root of the point of view taken by Carmona and Nadtochiy in [7] where they give up on the implied volatility code-book and work with dynamic models based on the local volatility code-book instead. We review the main elements of this approach in Section 6.

5 The HJM Approach for Credit Markets

We now explain how the above modeling philosophy can be used in the case of credit markets. For the sake of simplicity, we assume that the time evolution of the discounting factor is independent of all the default processes underlying the credit derivatives we consider in this section. So for all practical purposes, we can assume that $\{r_t\}_{t\geq 0}$ and $\{B(t,T)\}_{0\leq t\leq T}$ are deterministic. The market of Collaterized Debt Obligations (CDOs for short), and especially the market for single tranches synthetic CDOs saw a tremendous growth in the last five years, and because of their increased liquidity, they became a favorite testbed for quantitative research for the credit markets. As they were the main motivation for the works [26] and [41] which we draw from in this section, for the sake of completeness, we review some of their basic characteristics. In this section, we concentrate on the analysis of these instruments and when we say *credit markets* we mean the markets they span. They provide an appropriate set-up in which we test the HJM approach advocated in this paper. The reader interested in a broader perspective on the credit markets is referred to the textbooks of Schönbucher [40], Duffie and Singleton [18] or Lando [29].

5.1 Single Tranche Synthetic CDO Market Data

We review rapidly the major properties of Single Tranche Synthetic CDOs, often abbreviated as STSCDOs. Not only this will serve as motivation for the following developments, but it will also help us set the notation. Even though these instruments are best understood as derivatives on a portfolio of Credit Default Swaps (known as CDSs), for the sake of time and space, we introduce them independently.

Two parties are involved in any single STSCDO transaction: a counterparty seeking protection against defaults of all or part of a set of firms, and a counterparty selling this protection. To be more specific, a CDO swap with maturity T and notional N, on a tranche with attachment point ℓ_1 and detachment point ℓ_2, is a contract over the period $[0, T]$ whereby the protection seller will compensate the protection buyer for all the default losses in the interval $[\ell_1 N, \ell_2 N]$, in exchange for regular coupon payments computed at a fixed rate (the so-called spread) on a loss depreciating notional. We shall give a formal definition below.

But first, for the sake of illustration, we reproduce the following tables giving bid and ask quotes on the 4th and 5th series of the CDX-IG tranches on December 19, 2005. A hand-picked board of professionals selected a pool of firms as a representative snapshot of an homogeneous slice of the market (here IG stands for Investment Grade, but there exist indexes based on pools of high volatility firms, etc.), and portfolios of credit losses are used to construct an index and tranches which are traded on the market. These indexes are maintained and updated from one series to the next. Each series typically comprise $I = 125$ firms

IG4	0 - 3%	3 - 7%	7 - 10%	10 - 15%	15 - 30%
5-year	38 1/4 - 39 1/4	106 - 112	26 - 32	11 - 16	6 - 7 1/2
7-year	51 3/8 - 52 1/8	244 - 254	47 - 54	26 - 32	8 1/2 - 11
10-year	57 1/2 - 59 1/8	598 - 617	118 - 126	58 - 66	16 - 22

IG5	0 - 3%	3 - 7%	7 - 10%	10 - 15%	15 - 30%
5-year	41 1/4 - 42 1/4	107 1/4 - 112	26 - 29	11 - 14	6 1/2 - 9 1/2
7-year	54 3/4 - 55 5/8	290 - 300	45 - 51	27 - 31	7 - 10
10-year	61 3/4 - 62 3/4	685 - 705	118 - 124	61 - 66	17 - 21

The interpretation of these figures is the following. The quote for the equity tranche (0 - 3%) is the upfront payment (as a percentage of the notional) that is paid *in addition* to the minimal of 500 basis points per year. Quotes for all other tranches are in basis points per year.

We explain the meaning of these quotes by explaining in detail the cash flows associated with one of these tranches. For the sake of definiteness we choose the super-senior tranche with attachment points 15 and 30% on the 5yr CDX-IG index series 4. Let us assume that this tranche traded for 7 basis points. In this case, the protection buyer is to pay 0.07% of the notional per year (in quarterly coupon payments made in arrear). In return, she will be compensated for any losses on the portfolio during the five years that are between 15% and 30% of the principal. The losses are computed from the portfolio underlying the index at the original time of the trade.

The quotes for all the other tranches are defined similarly except for the equity tranche for which the buyer of protection pays an upfront fee and a spread of 500 basis points per year. The published quotes give the bid and ask for the upfront fee expressed as a percent of the notional. The percent of the notional that the protection buyer of the equity-tranche has to pay on December 19, 2005 was between 38.25% and 39.25% for five-year protection.

The index is also quoted to indicate the cost of buying full protection against all $I = 125$ names.

5.2 First Mathematical Model

As the largest volume of transactions involve derivative contracts written on synthetic portfolios identified and maintained by the Dow Jones (CDX in the US and iTraxx in Europe), we restrict ourselves to a fixed credit portfolio of I firms, and we denote by τ_i the time of default of firm i. In practice, one is limited to a finite horizon T^* and one only observes $\tau_i \wedge T^*$. Motivated by single tranche synthetic CDOs, we mostly consider instruments with maturities 3, 5, 7 or 10 years, so T^* can safely be assumed to be 10.

We denote by $\{L(t)\}_{t\geq 0}$ the cumulative portfolio loss (appropriately normalized) up to and including time t. We denote by $N(t)$ the nominal of the portfolio at time t so that $N(0)$ denotes the initial nominal. Note that $N(t)$ is a non-increasing function of time and that

$$L(t) = 1 - \frac{N(t)}{N(0)}$$

is a non-decreasing function of time which satisfies

$$L(0) = 0, \quad \text{and} \quad L(t) \leq 1.$$

Since the purpose of the present paper is mostly pedagogical, we make several assumptions with the mere intent to avoid unnecessary technicalities and simplify the notation.

Motivated by the example of the Dow Jones indexes, and especially by the actively traded Investment Grade (IG for short) North America index, we assume that the portfolio is symmetric in the sense that the credit exposure due the possible default of any single firm does not change with the firm in question. So typically, we restrict ourselves to firms included in the CDX and ITraxx indexes published by Dow Jones, and when we discuss CDOs, we consider only single tranche synthetic CDOs on these indexes. So not only do we assume that the individual firm nominal amounts are the same, but we also assume that the recovery rates in case of default are also deterministic, and the same for all the firms. So ignoring an irrelevant scaling factor, for the sake of definiteness we assume that

$$L(t) = \frac{1}{I}\sum_{i=1}^{I} D_i(t)$$

where $D_i(t)$ is the default indicator of firm i defined as:

$$D_i(t) = \mathbf{1}_{\{\tau_i \leq t\}}$$

τ_i being the stopping time giving the time of default of firm i. See for example the discussions in [18] and [29] for what kind of event can trigger or constitute default. Defined in this way, $L(t)$ represents the relative number of defaults prior to and including t, given the fact that there was no default at time $t = 0$. $\{L(t);\, t \geq 0\}$ is a stochastic process with non-negative piecewise constant and non-decreasing sample paths with values in the finite set $\mathcal{I} = \{0, 1/I, 2/I, \cdots, (I-1)/I, 1\}$.

CDO Mechanics and Liquidly Traded Instruments

Even though this is not exactly the case (as the membership in these portfolios is reviewed on a regular basis), we shall assume that the set of firms included in the portfolio is fixed and does not change over the life of the derivatives we consider. Moreover, for the sake of simplicity, we shall assume that the discounting factor β_T

is deterministic (or independent of the default times) and hence, can be taken out of the expectations.

The prices of the basic instruments playing the role of the prices of the zero coupon bonds, are the tranche and index spreads. To be more specific, we shall assume that for each $i = 0, 1, 2, 3, 4$ the spread $s_i(T)$ is observable for each maturity $T = 1, 3, 5, 7, 10$. By convention, we assume that s_0 is the spread on the index, s_1 is the spread on the equity tranche, s_2 the index on the lower mezzanine trance, etc. In order to explain how each spread is computed, we introduce the tenor structure

$$T_1 < T_2 < \cdots < T_n$$

of the days on which the coupon payments are to take place, and we continue the analysis of the tranche with attachment point ℓ_1 and detachment point ℓ_2 introduced earlier. Recall that we now assume that the portfolio nominal has been scaled down to 1.

Let us denote by s the rate of the coupon payments, and let us first evaluate the protection payments received by the protection buyer. Recall that, each time a loss L occurs, we assume that the part RL of the loss is recovered independently of the existence of the protection contract.

For notational convenience, for each time t, we define the quantity $L(t, \ell_1, \ell_2)$ by

$$L(t, \ell_1, \ell_2) = (L(t) - \ell_1)^+ - (L(t) - \ell_2)^+.$$

It gives at time t, the cumulative losses in the tranche. Indeed, it is equal to 0 if there were not enough losses to affect the tranche (i.e. if $L(t) < \ell_1$), it is equal to the tranche nominal $\ell_2 - \ell_1$ if the tranche was wiped out by losses (i.e. if $L(t) > \ell_2$), and it gives the lost part of the tranche nominal (i.e. $L(t) - \ell_1$) in the remaining cases (i.e. when $\ell_1 \leq L(t) \leq \ell_2$). So the expected present value (at time $t = 0$) of all the default losses recovered under the protection contract is

$$PL = (1 - R) \sum_{i=1}^{n} \beta_{T_i} [\mathbb{E}\{L(T_i, \ell_1, \ell_2)\} - \mathbb{E}\{L(T_{i-1}, \ell_1, \ell_2)\}] \tag{33}$$

We now consider the cashflows to the protection seller. At each coupon payment date T_i, she should receive the interest accumulated over the period $[T_{i-1}, T_i]$ computed on the remaining tranche nominal $(\ell_2 - \ell_1) - L(T_i, \ell_1, \ell_2)$. So the expected present value (at time $t = 0$) of all these coupon payments is

$$IL = s \sum_{i=1}^{n} \beta_{T_i} (T_i - T_{i-1}) \mathbb{E}\{(\ell_2 - \ell_1) - L(T_i, \ell_1, \ell_2)\} \tag{34}$$

The (fair) spread of the tranche at time $t = 0$ is the break even value of s making the expected present values of the two legs (34) and (33) equal to each other. Hence, the spread is given by the formula:

$$s = (1 - R) \frac{\sum_{i=1}^{n} \beta_{T_i} \mathbb{E}\{L(T_i, \ell_1, \ell_2) - L(T_{i-1}, \ell_1, \ell_2)\}}{\sum_{i=1}^{n} \beta_{T_i} (T_i - T_{i-1}) \mathbb{E}\{(\ell_2 - \ell_1) - L(T_i, \ell_1, \ell_2)\}} \tag{35}$$

Since we want to work with as many tranches as possible at once, we give up our notation $\ell_1 < \ell_2$ for the attachment/detachment points limiting the tranche, and for the sake of convenience, we shall from now on use the notation $0 = K_0 < K_1 < \cdots < K_k = 1$ for the end points of the tranche intervals.

Our goal is to extract the values of all the expectations

$$C_{i,j} = \mathbb{E}\{(L(T_i) - K_j)^+\}, \qquad i = 1, 2, \cdots, n, \;\; j = 1, \cdots, k, \tag{36}$$

from the values at time $t = 0$ of the spreads for all the available maturities (typically $\tau_1 = 1$, $\tau_2 = 3$, $\tau_3 = 5$, $\tau_4 = 7$, and $\tau_5 = 10$ years) on the index and all the liquidly traded tranches. This problem is not well posed as there are many more expectations than spread quotes. We use a simple form of regularization method to extract a set of expectations from the set of observable spreads. This is in stark contrast with the situation encountered next section when we discuss dynamic models for the equity markets. There, the expectations are directly observable.

The simplest regularization method leads to the least squares estimation We estimate them by solving the least squares minimization problem

$$\mathbf{C} = [C_{i,j}]_{i,j} = \arg\inf_{\mathbf{C}} \tag{37}$$

$$\sum_{j,k} w_{j,k} \Big| s_j(\tau_k) - R' \frac{\sum_{T_i \le \tau_k} \beta_{T_i} [C_{i,j} - C_{i,j-1} - C_{i-1,j} + C_{i-1,j-1}]}{\sum_{T_i \le \tau_k} \beta_{T_i} (T_i - T_{i-1})(K_j - K_{j-1} - C_{i,j} + C_{i,j-1})} \Big|^2,$$

where $R' = 1 - R$, and where for each maturity τ_k and tranche label j, the weights $w_{j,k}$ are chosen to be increasing in liquidity and decreasing in the size of the bid-ask spread. Unfortunately, this naive idea is unrealistic because of the large discrepancy between the number of reliable observations $s_j(\tau_k)$ and the number of desired C_{ij}. Even Levenberg-Marquard algorithms cannot provide a stable solution. The only known fixes are based on hand-waiving arguments and their reliability is questionable. See nevertheless [35] or [36]. Despite all that, it is common to assume that the numbers

$$C_{i,j} = \int (x - K_j)\, d\mu_{T_i}(x), \qquad i = 1, 2, \cdots, n, \;\; j = 1, \cdots, k, \tag{38}$$

are known! Here we use the notation μ_T for the distribution of the cumulative loss $L(T)$. As we already mentioned, for any measure μ, the *call transform* C_μ defined by

$$K \hookrightarrow C_\mu(K) = \int (x - K)^+ \, d\mu(x), \tag{39}$$

completely determines the measure μ. In general, a measure μ cannot be completely recovered from the mere knowledge of $C_\mu(K)$ for a small number of values of K, unless extra information on μ is available, e.g. μ is a finite sum of point masses.

As we will see in the next section, there are many ways to extrapolate these values of $C_{i,j}$ in between the attachment points to obtain for each T_i a convex function of the continuous variable K which coincides with the values derived above for all the $K = K_j$. We postpone the discussion of this point to the review of what is known in the case of equity options in Section 6 below.

So it is commonly assumed that at time $t = 0$, one knows the values of all the expectations $\mathbb{E}\{(L(T_i) - K)^+\}$ for all $K > 0$ which is equivalent to the full knowledge of the marginal distributions of the cumulative loss $L(T_i)$ at all the coupon payment times T_i under the distribution $\mathbb{P}$. This is the common starting point of the two papers on dynamic credit portfolio models which we review in this section.

Loss Distribution Dynamics

Having a hold of the marginal distributions of $L(T_i)$ is enough to price many instruments consistently with the spreads quoted on the market on day $t = 0$. However, this may not be enough for *forward starting contracts*. Let us consider for example the case of a tranche swaption, i.e. an option to enter a tranche swap contract (with maturity T and attachment/detachment points $\ell_1 < \ell_2$) at a later time $0 < T_0 < T$ at a spread level $\mathbf{s}$ fixed today at time $t = 0$. The value today of such an option is given by

$$\beta_{T_0}\mathbb{E}\{PL(T_0) - IL(T_0)\}$$

Here, the protection leg $PL(T_0)$ is the random variable equal to the value of PL computed from formula (33) provided we replace in formula (33) the expectations $\mathbb{E}\{\,\cdot\,\}$ by conditional expectations $\mathbb{E}_{T_0} = \mathbb{E}\{\,\cdot\,|\mathcal{F}_{T_0}\}$ with respect to the sigma-field $\mathcal{F}_{T_0}$ of the information which will be available at time T_0. Similarly, the investment leg $IL(T_0)$ is the random variable equal to the value of IL computed from formula (34) for the spread $s = \mathbf{s}$ and the conditional expectation $\mathbb{E}_{T_0} = \mathbb{E}\{\,\cdot\,|\mathcal{F}_{T_0}\}$ instead of the plain expectation with respect to $\mathbb{P}$.

So attempting to price forward starting contracts requires for each future time $t > 0$, to go through the calibration procedure described earlier at time $t = 0$ for the probability structure given by the (unconditional) pricing measure $\mathbb{P}$, using all the information available at time t by replacing $\mathbb{P}$ by its conditional version $\mathbb{P}_t = \mathbb{P}\{\,\cdot\,|\mathcal{F}_t\}$.

The above discussion justifies the introduction of the following notation which will be needed to describe dynamical models. For each $t \le T$, we denote by $P_t(T,\,\cdot\,)$ the distribution of the cumulative loss $L(T)$ conditioned by $\mathcal{F}_t$. In other words,

$$P_t(T,x) = \mathbb{P}\{L(T) \le x|\mathcal{F}_t\}, \qquad x \in [0,1].$$

Since $L(T)$ takes only finitely many values, the $I+1$ values $x = i/I$ for $i = 0,1,\cdots,I$ to be specific, we can talk about its density. We shall use a lower case to denote this density

$$p_t(T,x) = \mathbb{P}\{L(T) = x|\mathcal{F}_t\}, \qquad x \in [0,1].$$

These distributions will be called the forward loss distributions.

5.3 Two Different Approaches

It is important at this stage to emphasize the main difference between the approach of [26] and the one of [41], as this main difference lies in the choice of the filtration $\{\mathcal{F}_t\}_t$. In [41], the filtration $\{\mathcal{F}_t\}_t$ is the full filtration containing all the information available at time t, including both the economic factors and the default information. In these conditions, even after conditioning by $\mathcal{F}_t$, the above marginal probabilities of the loss distribution are discrete and can take only finitely many values between $L(t)$ and 1, typically the numbers $L(t)$, $L(t)+1/I$, $\cdots$, 1. However, in [26] the filtration used for conditioning in the definition of the forward loss distributions is a smaller filtration, say $\{\mathcal{M}_t\}_t$, which at each time t contains only information on economic factors and not necessarily on the actual default times. Intuitively, if one thinks of an intensity model for the time of default τ_i, the knowledge of $\mathcal{M}_t$ will determine the intensity $\lambda_i(t)$ at time t, but no information on the exponential random variable needed to compute the probability of arrival of the i-th default. This lack of information on the default arrival forces an integration with respect to the exponential random variable in order to compute the forward loss probabilities as defined by conditioning with respect to $\mathcal{M}_t$, and this integration justifies the assumption that the forward probabilities as defined above are smooth functions of the variable x. The densities $p_t(T,x)$ then appear to play the same role as the instantaneous forward rates in the classical HJM theory as they are derivatives of the forward rates given by the loss cumulative distribution functions. We come back to this approach at the end of this subsection.

In any case, for each fixed t, and for each $T \geq t$, we denote by $\mu_{t,T}$ the distribution of $L(T)$ under the conditional probability $\mathbb{P}_t$, and as usual, we denote by $\tau = T - t$ the time to maturity. Since the sample paths of the process $\{L(t+\tau)\}_{\tau \geq 0}$ are non-decreasing $\mathbb{P}_t$ – almost surely, for each fixed t, the measures $\{\tilde{\mu}_{t,t+\tau}\}_{\tau \geq 0}$ are non-decreasing in the balayage order in the sense that for every convex function ϕ, it holds

$$\int \phi(x) d\tilde{\mu}_{t,t+\tau_1}(dx) \leq \int \phi(x) d\tilde{\mu}_{t,t+\tau_2}(dx)$$

whenever $\tau_1 \leq \tau_2$. A classical result of Kellerer [28] implies the existence of a Markov process $\{Y_\tau\}_{\tau \geq 0}$ with the marginal distributions $\{\tilde{\mu}_{t,t+\tau}\}_{\tau \geq 0}$. Notice that this process depends upon t, but for the sake of notation we shall not emphasize this fact.

Schönbucher's Approach

In the case of the full filtration $\{\mathcal{F}_t\}_t$, the Markov process $\{Y_\tau\}_\tau$ has finite state space. Hence its distribution is entirely captured by its infinitesimal generator. The latter is a family of $(I+1) \times (I+1)$ Q-matrices indexed by $\tau \geq 0$ as the Markov process is not necessarily time homogeneous. Notice that we use the finite set $\{0, 1/I, 2/I, \cdots, (I-1)/I, 1\}$ as common state space for all these Markov processes instead of limiting the state space to the smaller set $\{L(t), L(t)+1/I, \cdots, 1\}$ which depends upon the realization of the random loss $L(t)$. Our choice is justified by the need to define dynamic equations which are more easily stated if all the Markov processes have the same state space.

We denote by $\{A_t(\tau); \tau \geq 0\}$ the infinitesimal generator of the Markov process $\{Y_\tau\}_{\tau \geq 0}$, and we denote by $\{a_t(\tau, x, y)\}_{x,y \in \{0,1/I,2/I,\cdots,(I-1)/I,1\}}$ the entries of the Q-matrix $A_t(\tau)$. We shall use this family of Q-matrices as a code-book for the information contained in the forward stochastic model as given by $\mathbb{P}_t$ once calibrated to the observable quotes at time t.

A classical fact from the theory of finite state Markov processes says that for each $\tau > 0$, the off-diagonal entries $a_t(\tau, x, y)$ are non-negative for $x \neq y$ as they have the interpretation of rate of jump from state x to state y. Because of this interpretation, as the sample paths of $L(t+\tau)$ are non-decreasing, the rates $a_t(\tau, x, y)$ should be zero whenever $y < x$, which implies that the matrices $A_t(\tau)$ are upper diagonal. Notice that the last row is identically zero since the state 1 (corresponding to the default of all the firms in the portfolio) is absorbing. Finally, the fact that the matrices $A(\tau)$ form the infinitesimal generator of a Markov process also imply that

$$a_t(\tau, x, x) = -\sum_{y \neq x} a_t(\tau, x, y), \qquad x \in \{0, 1/I, 2/I, \cdots, (I-1)/I, 1\},$$

which shows that the only entries that matter in the characterization of the code-book are the entries in each row, to the right of the diagonal.

Notice that the transition probabilities

$$p_t(\tau_1, \tau_2, x, y) = \mathbb{P}_t\{L(t+\tau_2) = y \,|L(t+\tau_1) = x\}$$

contain the same information as the infinitesimal generator matrices $\{A_t(\tau); \tau \geq 0\}$ as the two sets of matrices are related by the forward Kolmogorov equations which read:

$$\frac{\partial}{\partial \tau_2} p_t(\tau_1, \tau_2, x, y) = \sum_{k=0}^{I} p_t(\tau_1, \tau_2, x, k/I) a_t(\tau_1, k/I, y) \tag{40}$$

with initial condition

$$p_t(\tau_1, \tau_2, x, y)|_{\tau_2=\tau_1} = \mathbf{1}_{\{x=y\}} \,.$$

Using now the fact that we restrict ourselves to upper triangular Q-matrices, and the fact that the diagonal element of each row is the negative of the sum of the other elements of the row, we see that:

$$\frac{\partial}{\partial \tau_2} p_t(\tau_1, \tau_2, x, y) = \sum_{k=xI}^{yI-1} p_t(\tau_1, \tau_2, x, k/I) a_t(\tau_1, k/I, y) - p_t(\tau_1, \tau_2, x, y) \sum_{k=yI+1}^{I} a_t(\tau_1, y, k/I). \tag{41}$$

Finally it is easy to see that, once in the form (41), these equations can be solved inductively for the transition probabilities. One gets:

$$p_t(\tau_1, \tau_2, x, y) = \begin{cases} 0 & \text{if } y < x \\ \exp[\int_{\tau_1}^{\tau_2} a_t(\tau_1, s, x, x) ds] & \text{if } y = x \\ \sum_{k=xI}^{(y-1)I} \int_{\tau_1}^{\tau_2} p_t(\tau_1, s, x, k/I) \exp[\int_s^{\tau_2} a_t(\tau_1, s, y, y) ds] & \text{if } y > x \end{cases} \tag{42}$$

Notice that for each fixed $t > 0$, the connection between the forward loss distributions $\mathbb{P}_t\{L(t+\tau) = \,\cdot\,\} = p_t(\tau, \cdot)$ and the transition probabilities $p_t(\tau_1, \tau_2, \cdot, \cdot)$ of the Markov process $\{Y_\tau\}_{\tau \geq 0}$ is given by the relation

$$p_t(\tau, x) = p_t(0, \tau, L(t), x), \qquad x \in \{0, 1/I, 2/I, \cdots, (I-1)/I\}$$

since the marginal distributions of the Markov process $\{Y_\tau\}_{\tau \geq 0}$ are $\{\tilde{\mu}_{t,t+\tau}\}_{\tau \geq 0}$.

In order to avoid obscuring the main ideas by technicalities, we shall assume that the occurrence of more than one default at a time is impossible. This implies that the cumulative portfolio loss process $\{L(t+\tau)\}_{\tau \geq 0}$ can only jump by the amount $1/I$, and consequently the Q-matrix $A_t(\tau)$ are bi-diagonal in the sense that the only non-zero terms are the diagonal entries $a_t(\tau, x, x)$ and their neighbors $a_t(\tau, x, x+1/I)$ as long as $x < 1$. So under this assumption, the code-book reduces to a set of exactly I functions of time (to maturity) namely

$$\{a_t(\tau, x);\, \tau \geq 0\}_{x \in \{0, 1/I, 2/I, \cdots, (I-1)/I\}} \tag{43}$$

where we used the notation $a_t(\tau, x) = -a_t(\tau, x, x) = a_t(\tau, x, x+1/I)$.

It is explained in [41] that this assumption can be restrictive at times, and work-arounds are proposed to develop the same theory without this assumption. However, for the sake of simplicity, we restrict ourselves to models without simultaneous defaults in order to streamline the presentation of this survey.

HJM Dynamics

As explained in the previous section, the crux of the HJM approach to dynamic modeling is the choice of the dynamics of a code-book for the market data in the form of a set of Itô's stochastic differential equations, and the use of observable market data to feed these dynamic equations with an initial condition. Only then should the modeler worry about the consistency of such a model with a stochastic model for the portfolio loss process, and about the existence of possible arbitrages in the model specified in this way. Recall the list in the summary at the end of Subsection 3.2.

These last two issues are considered in the following two subsections. For the time being, we define the dynamics of the code-book by assuming that the forward default rates satisfy the following system of I stochastic differential equations

$$d\,a_t(\tau, x) = \alpha_t(\tau, x)dt \,+\, \beta_t(\tau, x)dW_t \tag{44}$$

where x varies in $x \in \{0, 1/I, 2/I, \cdots, (I-1)/I, 1\}$ and where for each $\tau \geq 0$ and x, $\{\alpha_t(\tau, x)\}_t$ and $\{\beta_t(\tau, x)\}_t$ are adapted processes with values in $\mathbb{R}$ and $\mathbb{R}^d$ respectively.

A Spot Consistency Condition

Consistency holds if the dynamics given by equation (44) can co-exist with a top down model where the time evolution of the system is derived from the dynamics of the cumulative loss process $\{L(t)\}_t$ specified first. The following result gives a necessary condition for this to hold.

Proposition 1. *Let us assume that the process* $\mathbf{L} = \{L(t)\}_{t\geq 0}$ *of cumulative portfolio losses admits transition rates which only jump by* 1. *Then when viewed as a point process,* $\mathbf{L}$ *has an intensity* $\{\lambda_L(t)\}_{t\geq 0}$ *given almost surely by the formula:*

$$\lambda_L(t) = a_t(0, L(t)), \qquad t \geq 0. \tag{45}$$

The consistency condition (45) is a direct consequence of Aven's theorem [2] and our implicit smoothness assumption on the forward default rates. We reproduce the proof given in [41]. If we fix $t > 0$ and $\epsilon > 0$ we have:

$$\begin{aligned}
&\frac{1}{\epsilon}\mathbb{E}_t\{L(t+\epsilon) - L(t)\} \\
&\quad = \frac{1}{\epsilon}\sum_{n=L(t)}^{I}(\frac{n}{I} - L(t))p_t(\epsilon, n) \\
&\quad = \sum_{n=L(t)+1}^{I}(\frac{n}{I} - L(t))\frac{1}{\epsilon}[p_t(0, \epsilon, L(t), n) \\
&\quad = \sum_{n=L(t)+1}^{I}(\frac{n}{I} - L(t))\frac{1}{\epsilon}[p_t(0, 0, L(t), n) + \epsilon\partial_\tau p_t(0, 0, L(t), n) + o(\epsilon)] \\
&\quad = \sum_{n=L(t)+1}^{I}(\frac{n}{I} - L(t))[-a_t(0, n)p_t(0, 0, L(t), n) \\
&\qquad\qquad + a_t(0, n-1)p_t(0, 0, L(t), n-1) + O(1)] \\
&\quad = a_t(0, L(t)) + O(1)
\end{aligned}$$

where we used Kolmogorov's equation. ∎

Remark. The result of Proposition 1 shows that the jump times of the process $\mathbf{L}$ (i.e. the default times of the portfolio components) are totally inaccessible. So even though such an assumption was never stated explicitly, we are actually working in the framework of reduced form models (i.e. intensity based models) as opposed to structural models for which the time of default are typically announced by increasing sequences of stopping times. See also the discussion of the SPA approach below.

The HJM Drift Condition

Let us assume that for each $\tau > 0$ and $x \in \{0, 1/I, 2/I, \cdots, (I-1)/I\}$ the stochastic processes $\{a_t(\tau, x)\}_t$ is a non-negative semi-martingales with a decomposition of the form (44).

For each fixed $t > 0$, we can view $a_t(\tau, \cdot)$ as the negative of the diagonal elements of a bidiagonal Q-matrix and by solving the forward Kolmogorov equations as before, we can derive expressions for the transition probabilities $p_t(\tau_1, \tau_2, \cdot, \cdot)$ of a Markov process whose marginal distributions $p_t(\tau, x) = p_t(0, \tau, L(t), x)$ we would like to coincide with the forward loss distributions $\mathbb{P}\{L(t+\tau) = x | \mathcal{F}_t\}$. Notice that in this case, if we fix $T = t + \tau$ and vary t in $[0, T]$, the latter are martingales by construction since they are conditional expectations of a fixed random variable.

If we start from prescription (44) the explicit formulae (42) for the transition probabilities can be used to prove that $\{p_t(0, T, x, y)\}_{0 \le t \le T}$ is a semi-martingale for each default levels $x < y$, and we can compute its bounded variation and quadratic variation parts. Substituting $L(t)$ for x, one can show that $p_t(0, \tau, L(t), x)$ also is a semi-martingale, and one can derive its bounded variation and quadratic variation parts in terms of the drift $\alpha_t(\tau, x)$ and volatility $\beta_t(\tau, x)$ of $a_t(\tau, x)$. Now recall that, if the forward default rates $a_t(\tau, x)$ come from an *underlying* loss process $L(t)$, then as we already explained,

$$p_t(0, \tau, L(t), x) = p_t(t + \tau, x) = \mathbb{P}\{L(t+\tau) = x | \mathcal{F}_t\}$$

is necessarily a martingale. Stating that its bounded variation part vanishes leads to the following conclusion.

Proposition 2. *If for each $\tau > 0$ and $x \in \{0, 1/I, 2/I, \cdots, (I-1)/I\}$ the stochastic processes $\{a_t(\tau, x)\}_t$ is a non-negative semi-martingale satisfying (44), then the forward loss distributions $\{p_t(T, x)\}_{0 \le t \le T}$ are martingales if and only if*

$$p_t(0, T-t, L(t), x)\alpha_t(T, x) = -\beta_t(T, x) v_t(0, T-t, L(t), x), \qquad (46)$$

for $x \in \{0, 1/I, 2/I, \cdots, (I-1)/I\}$ where $v_t(\tau_1, \tau_2, x, y)$ is the volatility of the semi-martingale decomposition of the transition probability $p_t(\tau_1, \tau_2, x, y)$ as given by the solution of the forward Kolmogorov's equation.

We refer the interested reader to [41] for the details of the derivation. Condition (46) is called the HJM drift condition because of its striking similarity with the original HJM drift condition (25). However, a crucial difference needs to be emphasized. While the classical drift condition (25) gives explicitly the drift $\alpha_t(T)$ of the codebook in terms of its volatility $\beta_t(T)$, the above drift condition merely states a relation between drift and volatility of the code $a_t(T, x)$. Indeed, the term $v_t(0, T-t, L(t), x)$ which appears in the right hand side of (46) is a function of the code, and hence of its bounded variation part. In other words, the drift term $\alpha_t(T, x)$ is present in the right hand side of (46) which is only an implicit equation for $\alpha_t(T, x)$. We shall encounter the same problem in our discussion of the HJM approach to equity market models in the next section. However, the situation is easier here. Indeed, because of the finite nature of the state space of the loss process, and because of our assumption of the upper-diagonal nature of the Q-matrices and the fact that their last rows are identically zero, these implicit equations can be solved exactly after finitely many iterations. We refer the interested reader to the details provided in [41].

Volatility Structure Calibration

One of the goals of this review is to emphasize how an HJM modeling approach resolves the calibration issue by encapsulating the market prices of the liquidly traded instruments in the initial condition of the dynamic model, and how the resulting dynamic specifications can be restricted to the volatility term as the drift can be determined from the volatility and the observed market prices. Because everything rides on the particular choice of a volatility structure $\beta_t(\tau, x)$ for the forward default rates, the actual volatility specification is of crucial importance. Unfortunately, it still remains a *touchy business* as there is no clear algorithm providing such a volatility structure, even if it is easy to understand the practical consequences of $\beta_t(\tau, x) \equiv 0$, or $\beta_t(\tau, x)$ having a constant sign, or being very large of $x \approx L(t)$ and small otherwise, ... etc. Unfortunately, this difficulty cannot be resolved without further information about the desired market model, whether this information comes from from prices of exotic derivatives or more qualitative properties that the model should reproduce. We refer to our discussion of the same issue in Section 3 in the case of the fixed income markets.

The SPA Approach

If we use the smaller market filtration $\{\mathcal{M}_t\}_t$ to condition the time evolutions of the forward loss distributions, then the

The idea of the SPA approach is to treat the values of the forward loss cumulative distribution functions $P_t(T, x)$ as a family of zero coupon bond prices parameterized by the loss level x, in which case it is natural to introduce the equivalent of instantaneous forward rates by defining

$$f_t(T, x) = -\frac{\partial}{\partial T} \log P_t(T, x) = -\frac{\frac{\partial}{\partial T} P_t(T, x)}{P_t(T, x)} \tag{47}$$

and to construct a dynamic portfolio loss model by specifying a set of stochastic differential equations for these forward loss rates in the form

$$d\, f_t(T, x) = \alpha_t(T, x)dt \,+\, \beta_t(T, x)dW_t \tag{48}$$

in full analogy with the HJM prescription (22) used in the fixed income markets. Even though the notation of this approach follow more closely the notation of the classical HJM approach reviewed in Section 3, it is not as natural as the more involved approached based on Markov process codes discussed above. Indeed, the latter will generalize in a straightforward manner to the case of the equity markets discussed in the next section. Moreover, the former cannot be used without introducing an extra layer of technical derivations involving Markov loss processes, obscuring their original claims of simplicity. The interested reader is referred to [26] for details.

6 The HJM Approach to Equity Markets

This section is devoted to the derivation of arbitrage free dynamic stochastic models for the equity markets. We try to incorporate standard features of these markets, and in so doing, we put ourselves in a situation amenable to the HJM philosophy highlighted in the previous sections. This approach to dynamic equity models was originally advocated by Derman and Kani in [17]. The present discussion is based on the recent work of Carmona and Nadotchyi [7] where explicit formulae, rigorous proofs and numerical examples are given.

So as in the cases of fixed income and credit market models reviewed in Section 3, we first identify a set of instruments liquidly traded to which the model needs to be calibrated. The goal of our modeling effort is to characterize important properties (such as for example absence of arbitrage) of the pricing measure $\mathbb{P}$ used by the market by studying the dynamics of these liquidly traded instruments instead of the dynamics of the instruments underlying them. In this way, calibration is taken care of by merely using observed prices as initial conditions for the dynamical equations. As before, the dynamics of the prices of the basic instruments are given by an infinite dimensional stochastic differential equation, or equivalently by a random field.

6.1 Description of the Market

As always, we consider an economy with a perfect frictionless market without bid-ask spreads, with short sales of call and put options allowed in arbitrary sizes, without taxes, etc. In such an idealized market model, it is natural to choose for the set of liquidly traded securities, the ensemble of all the European call options written on underlying instruments spanning the market. For the sake of simplicity, we assume that one single underlyer (e.g. a *stock*) spans the market under consideration. Choosing more underlyers would force the price process to be multivariate and make the notation unnecessarily complicated without changing much to the nature of the results.

Let us denote by $\{S_t\}_{t\geq 0}$ the price process underlying the derivative instruments forming the market. As stated above, for the sake of simplicity we assume that the market comprises only derivatives written on a single underlying instrument, in other words, we assume that S_t is univariate. Again, for the sake of simplicity, we assume that the discount factor is one, i.e. $\beta_t \equiv 1$, or equivalently that the short interest rate is zero, i.e. $r_t \equiv 0$, and that the underlying risky asset does not pay dividends. These assumptions greatly simplify the notation without affecting the generality of our derivations.

We assume that in our idealized market, European call options of all strikes and maturities are liquidly traded, and that their prices are observable. We denote by $C_t(T, K)$ the market price at time t of a European call option of strike K and maturity $T > t$. We assume that today, i.e. on day $t = 0$, all the prices $C_0(T, K)$ are observable. According to the philosophy adopted in this paper, at any given time t, instead of working directly with the price S_t of the underlying asset, we concentrate on the set of call prices $\{C_t(T, K)\}_{T,K}$ as our fundamental market data. This is partly justified by the well documented fact that many observed option price movements cannot be attributed to changes in S_t, and partly by the fact that many exotic (path dependent) options are hedged (replicated) with portfolios of plain (vanilla) call options.

Remarks. 1. It is well known that in order to avoid arbitrage (at least against static strategies) the observed call prices $C_0(T, K)$ should be increasing in T, non-increasing and convex in K, that they should converge to 0 as $K \to \infty$ and that they should recover the underlying price S_0 for zero strike when $K \to 0$. We shall implicitly assume that the observed surface of initial call surface satisfies these properties.
2. **A More Realistic Set-Up.** In the description of our idealized market, we assumed that European call options of all strikes and all maturities were liquidly traded. This assumption is very convenient, though highly unrealistic. Indeed, the knowledge of all the prices $C_t(T, K)$ determine all the marginal distributions of the underlying instruments under the pricing measure $\mathbb{P}$. This information is not available in real life. In practice, the best one can hope for is, for a finite set of discrete maturities $T_1 < T_2 < \cdots < T_n$, one has quotes for the prices of a finite set of call options. In other words, for each $i = 1, 2, \cdots, n$ one has the prices of calls $C_t(T_i, K_{ij})$ for a finite set $K_{i1} < K_{i2} < \cdots < K_{in_i}$.

This more realistic form of the set-up has seen a recent renewal of interest starting with the work of Laurent and Leisen [30]. Our interest in this problematics was triggered by the recent technical reports by Cousot [13] and Buehler [6] who use Kellerer [28] theorem in the same spirit as the present discussion, and by the recent work of Davis and Hobson [15] which relies instead on the Sherman-Stein-Blackwell theorem [43, 44, 4].

We refer the interested reader to [15] and to the references therein.

From now on, we denote by $\tau = T - t$ the time to maturity of the option and we denote by $\tilde{C}_t(\tau, K)$ the price $C_t(T, K)$. In other words

$$\tilde{C}_t(\tau, K) = C_t(t + \tau, K), \qquad \tau > 0.\ K > 0.$$

We assume that the market *prices by expectation* in the sense that the prices of the liquid instruments are given by expectations of the present values of their cashflows with respect to a probability measure. So saying that $\mathbb{P}$ is a pricing measure used by the market implies that for each time $t \geq 0$ we have

$$\tilde{C}_t(\tau, K) = \mathbb{E}\{(S_{t+\tau} - K)^+ | \mathcal{F}_t\} = \mathbb{E}^{\mathbb{P}_t}\{(S_{t+\tau} - K)^+\}.$$

where we denote by $\mathbb{P}_t$ a regular version of the conditional probability of $\mathbb{P}$ with respect to $\mathcal{F}_t$. We denote by $\tilde{\mu}_{t,t+\tau}$ the distribution of $S_{t+\tau}$ for the conditional distribution $\mathbb{P}_t$. It is an $\mathcal{F}_t$- measurable random measure. With this notation

$$\tilde{C}_t(\tau, K) = \int_0^\infty (x - K)^+ \, d\tilde{\mu}_{t,t+\tau}(dx)$$

and for each fixed $\tau > 0$, the knowledge of all the prices $\tilde{C}_t(\tau, K)$ completely determines the distribution $\tilde{\mu}_{t,t+\tau}$ on $[0, \infty)$.

Remarks. 1. Notice that we do not assume uniqueness of the pricing measure $\mathbb{P}$. In other words, our analysis holds in the case of incomplete models as well as complete models.

2. **Notation Convention.** In order to help with the readability of the paper, we use a notation without a tilde or a hat for all the quantities expressed in terms of the variables T and K. But we shall add a tilde for all the quantities expressed in terms of the variables τ and K, and a hat when the strike is given in terms of the variable $x = \log K$.

6.2 Implied Volatility Code-Book

In the classical Black-Scholes theory, the dynamics of the underlying asset are given by the stochastic differential equation

$$dS_t = S_t \sigma dW_t, \qquad S_0 = s_0$$

for some univariate Wiener process $\{W_t\}_t$ and some positive constant σ. In this case, the price $\tilde{C}_t(\tau, K)$ of a call option is given by the Black-Scholes formula

$$BS(S, \tau, \sigma, K) = S_t \Phi(d_1) - K\Phi(d_2) \tag{49}$$

with

$$d_1 = \frac{-\log M_t + \tau\sigma^2/2}{\sigma\sqrt{\tau}}, \qquad d_2 = \frac{-\log M_t - \tau\sigma^2/2}{\sigma\sqrt{\tau}}$$

where $M_t = K/S_t$ is the moneyness of the option. We use the notation Φ for the cumulative distribution of the standard normal distribution, i.e.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy, \qquad x \in \mathbb{R}.$$

The Black-Scholes price is an increasing function of the parameter σ when all the other parameters are held fixed. As a consequence, for every real number C (think of such a number as a quoted price for a call option with time to maturity τ and strike K) in the interval between $(S_t - K)^+$ and S_t there exists a unique number σ for which $\tilde{C}_t(\tau, K) = C$. This unique value of σ given by inverting Black-Scholes formula (49) is known as the implied volatility and we shall denote it by $\tilde{\Sigma}_t(\tau, K)$. This quantity is extremely important as it is used by most if not all market participants as the *currency* in which the option prices are quoted. This practice should not be construed as an endorsement of Black-Scholes model. In order to emphasize this fact, I cannot resist the temptation to characterize the implied volatility by the following statement borrowed from Rebonato's book [37]:

the wrong number to put in the wrong formula to get the right price.

For each time $t > 0$, the one-to-one correspondence

$$\{\tilde{C}_t(\tau, K);\, \tau > 0, K > 0\} \leftrightarrow \{\tilde{\Sigma}_t(\tau, K);\, \tau > 0, K > 0\}$$

offers a code-book translating without any loss all the information given by the call prices in terms of implied volatilities, and which we call the Black-Scholes or implied volatility code-book. While Black-Scholes theory predicts a flat profile for the implied volatility surface, one has plenty empirical evidence of the contrary. We refer the reader interested in the empirical properties of the implied volatility surface to the thorough discussion in Rebonato's book [37] and to the references therein. The mathematical analysis of this surface is based on a subtle mixture of empirical facts and arbitrage theories, and it is rather technical in nature. The literature on the subject is vast and it cannot be done justice in a few references. Choosing a few samples for their relevance to the present discussion, we invite the interested reader to consult [10],[20],[31], [33],[21] and the references therein to get a better sense of these technicalities.

Valuation and risk management of complex option positions require models for the time evolution of implied volatility surfaces. [34] and [12] are examples of attempts to go beyond static models, but despite the fact that they consider only a cross section of the surface (say for K fixed), the works of Schönbucher [39] and Schweizer and Wissel [42] are more in the spirit of the HJM approach which we advocate in this section.

At any given time t, absence of (static) arbitrage imposes conditions on the surface of call option prices. As we already mentioned, the surface $\{\tilde{C}_t(\tau, K)\}_{\tau,K}$ should be increasing in τ, non-increasing and convex in K, it should converge to 0 as $K \to \infty$ and recover the underlying price S_0 for zero strike when $K \to 0$. Because of the one-to-one correspondence between call prices and implied volatilities, these conditions can be expressed in terms of properties of the implied volatility surface $\{\tilde{\Sigma}_t(\tau, K)\}_{\tau,K}$ at time t. However inverting Black-Scholes formula (49) is not simple and these conditions become unnecessarily technical. This is one of the reasons why we search for another way to capture the information in the surface of call option prices.

6.3 Choosing another Option Code-Book

As in the standard framework of the Black-Scholes theory, we start from the dynamics of the underlying asset and we try to identify a code-book for the traded instruments in such a way that the dynamics of the codes could be easily manipulated and most importantly, could be used as a starting point to define the dynamics of the market. Since we assume that the filtration is Brownian, without any loss of generality we can assume

$$dS_t = S_t \sigma_t \, dW_t^1, \qquad S_0 = s_0$$

for some adapted non-negative process $\{\sigma_t\}_{t\ge 0}$. If $t > 0$ is fixed, for any τ_1 and τ_2 such that $0 < \tau_1 < \tau_2$, and for any convex function ϕ on $[0,\infty)$ we have

$$\begin{aligned}\int_0^\infty \phi(x)\tilde\mu_{t,t+\tau_1}(dx) &= \mathbb{E}^{\mathbb{P}_t}\{\phi(S_{t+\tau_1})\}\\ &= \mathbb{E}^{\mathbb{P}_t}\{\phi(\mathbb{E}^{\mathbb{P}_t}\{S_{t+\tau_2}|\mathcal{F}_{t+\tau_1}\})\}\\ &\le \mathbb{E}^{\mathbb{P}_t}\{\mathbb{E}^{\mathbb{P}_t}\{\phi(S_{t+\tau_2})|\mathcal{F}_{t+\tau_1}\})\}\\ &= \mathbb{E}^{\mathbb{P}_t}\{\phi(S_{t+\tau_2})\}\\ &= \int_0^\infty \phi(x)\tilde\mu_{t,t+\tau_2}(dx)\end{aligned}$$

from which we see that for any given $t > 0$, the probability measures $\{\tilde\mu_{t,t+\tau}\}_{\tau>0}$ are non-decreasing in the balayage order. This implies the existence of a Markov martingale $\{Y_\tau\}_{\tau\ge 0}$ with marginal distributions $\{\tilde\mu_{t,t+\tau}\}_{\tau>0}$. Since the knowledge of all the call prices $\{\tilde C(\tau,K)\}_{\tau>0,K>0}$ is equivalent to the knowledge of all the distributions $\{\tilde\mu_{t,t+\tau}\}_{\tau>0}$, the Markov martingale $\{Y_\tau\}_{\tau\ge 0}$ is a way to encapsulate the information given by the market at time t by providing the call prices. Obviously, the process $\{Y_\tau\}_{\tau\ge 0}$ contains more information than the mere marginal distributions $\{\tilde\mu_{t,t+\tau}\}_{\tau>0}$ determined by the call option prices. This process can be used to price contracts with path dependent *exotic* pay-offs whose values are not uniquely determined by the *state price densities* of the marginal distributions. The procedure which we just outlined captures perfectly the philosophy and the practice of the market participants: include all the information about the liquidly traded instruments in a model that reproduces all of these prices, and use such a model to price exotic derivatives which cannot be synthesized from the liquid instruments available for trade. As such a model is not uniquely determined by the market prices, there is a lot of freedom in choosing it, and many factors enter the final decision: parsimony, common sense, versatility, basic principles (e.g. maximum entropy, minimum least squares, ...) but in any case, once the choice is made, the only thing left is *hope for the best*.

Notice that, if the process $\{Y_\tau\}_{\tau\ge 0}$ is realized on a Wiener space, then the martingale representation theorem in Brownian filtrations gives that Y_τ can be written as

$$Y_\tau = Y_0 + \int_0^\tau Y_s \tilde a(s)\, dB_s$$

and that, because of the Markov property, the predictable process $\{\tilde a(s)\}_{s\ge 0}$ can be chosen to be of the form $\tilde a(s,\omega) = \tilde a_t(s, Y_s(\omega))$ for some function $(s,y) \hookrightarrow \tilde a_t(s,y)$ of $(s,y) \in [0,\infty)\times[0,\infty)$ and whose graph can be viewed as a surface over the quadrant $[0,\infty)\times[0,\infty)$. Notice that this surface changes with t in an $\mathcal{F}_t$-measurable way. At each time t, we can choose this surface $\{\tilde a_t(\tau,K)\}_{\tau>0,K>0}$ as an alternative code-book for the information contained in the options prices $\{\tilde C(\tau,K)\}_{\tau>0,K>0}$. This code-book is different from the Black-Scholes implied volatility code-book $\{\tilde\Sigma_t(\tau,K)\}_{\tau>0,K>0}$ given by the implied volatilities of the European call options in question. The deterministic version of the surface $\{\tilde a_t(\tau,K)\}_{\tau>0,K>0}$ was introduced in a static framework (i.e. for $t=0$) simultaneously by Dupire [19] and Derman and Kani [16] though with a different definition, as an alternative to the implied volatility surface. The surface $\{\tilde{a}_t(\tau, K)\}_{\tau>0,K>0}$ has been called the local volatility surface for reasons which will become clear later in the paper. From our point of view, the main reason to work with the local volatility surface instead of the implied volatility surface is the ease with which one can check the presence or absence of static arbitrage. Indeed, as we shall see below, the four conditions (increasing in τ, increasing and convex in K, plus the two boundary conditions) guaranteeing the absence of static arbitrage become merely positivity of the numbers $\tilde{a}_t(\tau, K)$. Replacing difficult conditions to check by such a simple one becomes extremely convenient when we deal with dynamic models. The interested reader is invited to consult [32] for a thorough discussion of the connections between local and implied volatility in the static framework (i.e. at time $t = 0$) of stochastic volatility models.

A dynamic version of local volatility modeling was later touted by Derman and Kani in a paper [17] mostly known for its discussion of implied tree models. Motivated by the fact that the technical parts of [17] dealing with continuous models are rather informal and lacking mathematical proofs, Carmona and Natodchy actually develop the program outlined in [17]. While providing a rigorous mathematical derivation of the so-called drift condition, they also discuss concrete examples and provide calibration and Monte Carlo implementation recipes.

We now derive the property of the local volatility surface which got us interested in its dynamics. Notice that the following derivation is done when the time $t > 0$ and the past up to and including time t are fixed. We give details in the case where for example, we assume that the above marginal distributions $\tilde{\mu}_{t,t+\tau}$ have for each $\tau > 0$ a positive density $\tilde{g}_t(\tau, x)$ (continuous as a function of $x > 0$), which once x is held fixed, are continuously differentiable in the variable τ. Then we can conclude that for each t there exists function $(\tau, K) \hookrightarrow \tilde{a}_t(\tau, K)$ such that the process

$$dY_\tau = Y_\tau \tilde{a}_t(\tau, Y_\tau) d\tilde{B}_\tau, \quad \tau > 0 \tag{50}$$

with initial condition

$$Y_0 = S_t$$

is well-defined and has marginal distributions $\tilde{\mu}_{t,t+\tau}$.

We first recall the Breeden-Litzenberger argument which is specific to the *hockey-stick* pay-off function of the European call options. Since the option price with strike K and time to maturity τ is given by

$$\tilde{C}_t(\tau, K) = \int_0^\infty (x - K)^+ \tilde{g}_t(\tau, x) dx$$

we can differentiate both sides twice with respect to K and get:

$$\partial^2_{KK} \tilde{C}_t(\tau, K) = \tilde{g}_t(\tau, K). \tag{51}$$

Next we apply Itô-Tanaka's formula to (50) and the function $f(y) = (y - K)^+$ (see for example [38]). Note that this function f is convex. It is infinitely differentiable everywhere except at $y = K$ where it has a left and a right derivatives. Obviously $f'(y) = 0$ if $y < K$ and $f'(y) = 1$ if $y > K$. Moreover, the second derivative $f''(y)$ in the sense of distributions is the Dirac point mass at K (also called the *delta function* at K). We get:

$$(Y_\tau - K)^+ = (Y_0 - K)^+ + \int_0^\tau \mathbf{1}_{[K,\infty)}(Y_s) dY_s + \frac{1}{2} L^K_\tau$$

where for each $a \in \mathbb{R}$, $\{L^a_t\}_{t\geq 0}$ is the local time of the semi-martingale $\{Y_t\}_{t\geq 0}$ at a. Using the fact that Y is a martingale satisfying $d\langle Y, Y\rangle_s = Y_s^2 \tilde{a}_t(s, Y_s)^2 ds$, by definition of the local time it holds:

$$
\begin{aligned}
L_\tau^K &= \lim_{\epsilon \searrow 0} \frac{1}{2\epsilon} \int_0^\tau \nVdash_{(K-\epsilon, K+\epsilon)}(Y_s) d\langle Y, Y \rangle_s \\
&= \lim_{\epsilon \searrow 0} \frac{1}{2\epsilon} \int_0^\tau \nVdash_{(K-\epsilon, K+\epsilon)}(Y_s) Y_s^2 \tilde{a}_t(s, Y_s)^2 ds,
\end{aligned}
$$

and taking $\mathbb{E}_t$ - expectations on both sides we get:

$$
\begin{aligned}
\tilde{C}_t(\tau, K) &= (S_t - K)^+ + \frac{1}{2} \lim_{\epsilon \searrow 0} \frac{1}{2\epsilon} \int_0^\tau \int^{\mathbb{R}} \nVdash_{(K-\epsilon, K+\epsilon)}(y) y^2 \tilde{a}_t(s, y)^2 g_t(s, y) dy ds \\
&= (S_t - K)^+ + \frac{1}{2} \int_0^\tau K^2 \tilde{a}_t(s, K)^2 g_t(s, K)\, ds.
\end{aligned}
$$

where $g_t(s, y)$ is the density of Y_s for $\mathbb{P}_t$ which is assumed to be continuous in y which justifies taking the limit as $\epsilon \searrow 0$. Finally, taking derivatives with respect to τ on both sides we get:

$$
\partial_\tau \tilde{C}_t(\tau, K) = \frac{1}{2} K^2 \tilde{a}_t(\tau, K)^2 \tilde{g}_t(\tau, K). \tag{52}
$$

Equating the expressions of the density $\tilde{g}_t(\tau, K)$ obtained in (51) and (52) we get the following expression for the local volatility:

$$
\boxed{\tilde{a}_t(\tau, K)^2 = \frac{2\partial_\tau \tilde{C}_t(\tau, K)}{K^2 \partial^2_{KK} \tilde{C}_t(\tau, K)}.} \tag{53}
$$

Equation (53) determines the local volatility surface $\{\tilde{a}_t(\tau, K)\}_{\tau, K}$ from the values of the call prices $\{\tilde{C}_t(\tau, K)\}_{\tau, K}$. Conversely, if we were to start from a prescription giving the local volatility surface $\{\tilde{a}_t(\tau, K)\}_{\tau, K}$, we would derive the set of call option prices $\{\tilde{C}_t(\tau, K)\}_{\tau, K}$ by solving the partial differential equation (PDE for short)

$$
\begin{aligned}
&\partial_\tau \tilde{C}(\tau, K) = \frac{1}{2} K^2 \tilde{a}^2(\tau, K) \partial^2_{KK} \tilde{C}(\tau, K), \qquad \tau > 0,\ K > 0 \\
&\tilde{C}(0, K) = (S_t - K)^+
\end{aligned} \tag{54}
$$

which is sometimes called the Dupire's PDE because it was first advocated by Bruno Dupire in his groundbreaking work [19] on the volatility smile. We call the one-to-one correspondence given by (53) and (54) the local volatility code-book. The correspondence

$$
\boxed{\{\tilde{C}_t(\tau, K);\, \tau > 0, K > 0\} \quad \rightleftharpoons \quad \{\tilde{a}_t(\tau, K);\, \tau > 0, K > 0\}}
$$

defining our code book is analog, though different from the correspondence given by the Black-Scholes code-book. Indeed, to compute the code from the option prices, we need to compute the right hand side of (53) instead of evaluating the Black-Scholes formula, while in order to recover the option prices from the code we solve the partial differential equation (54) instead of inverting the Black-Scholes formula.

Remark: Statistical Estimation. Recalling the discussion of the remark on a "More Realistic Model", on most every day t, the available data are in the form of a finite set of prices $C_t(T_i, K_{i,j})$ (or possibly of implied volatilities $C_t(T_i, K_{i,j})$) on an irregular grid in the (T, K)-plane. The challenge is to construct a smooth surface $\{C_t(T, K)\}_{T \geq t, K > 0}$ or $\{\Sigma_t(T, K)\}_{T \geq t, K > 0}$ through the observations over the finite grid. This problem is discussed with great care in the small book [23] by Flenger, while the book [1] addresses the same problems in a less statistical and more computational spirit. The interested reader is referred to the review [11] written by Carter and Fouque of Fengler's book for an independent perspective on its content. The word of practitioners and academics is divided into two camps neatly delineated by irreconcilable differences. The first camp argues that, in order to rule out any static arbitrage, the price surface $\{C_t(T,K)\}_{T\geq t, K>0}$ needs to go through all the observed market prices $C_t(T_i, K_{i,j})$. The second camp does make this strict requirement, claiming that because these prices are not quoted at the same time of the day (i.e. for different values of t), portfolios leading to arbitrage can in principle be constructed mathematically, but they cannot be implemented in practice because of the lack of simultaneity of the quotes, preventing the *wannabe* arbitrager to set up the arbitrage portfolio identified by the mathematical theory. Both arguments are reasonable and quite convincing, and we will not try to take side on this difficult issue.

We can now hint at our implementation of the HJM philosophy in the case of equity markets: as usual, instead of choosing the dynamics of the underlyer S_t and then deriving a set of equations for the prices of the liquidly traded instruments (the European call option prices in our case), we model directly the dynamics of the prices of the liquidly traded instruments by choosing the dynamics of a specific code-book, and in the present situation, we choose the local volatility code-book.

Another reason for choosing the local volatility code-book over the implied volatility code-book is the fact that the four conditions needed to rule out static arbitrage take a very simple form in the case of the local volatility. Indeed, it is enough to make sure that $\tilde{a}_t(\tau,K)$ is positive to guarantee that $\tilde{C}_t(\tau,K)$ is increasing in T, increasing and convex in K and satisfies the two boundary conditions already discussed. This advantage is priceless when it comes to defining stochastic dynamics.

Remark. As a last remark, we show that, whenever the underlying is known to satisfy an equation of the form

$$dS_t = S_t\sigma_t\, dW_t \tag{55}$$

for some Wiener process $\{W_t\}_t$ and some adapted non-negative process $\{\sigma_t\}_t$, then at each time t, the local volatility $\tilde{a}_t(\tau,K)$ can be viewed as the current expected variance for time to maturity τ and strike K. More precisely, this means that:

$$\tilde{a}_t(\tau,K)^2 = \mathbb{E}_t\{\sigma^2_{t+\tau}|S_{t+\tau}=K\}. \tag{56}$$

In order to prove this result, it is enough to retrace the steps of the above derivations of (52) and (53) using $S_{t+\tau}$ and its dynamics (55) instead of Y_τ and its own dynamics. This formula is often called Dupire's formula. It is at the origin of the terminology local volatility surface.

6.4 Code-Book Dynamics

We postulate the dynamics of the local volatility surface point by point. For each fixed $T>0$ and $K>0$, we assume that the process $\{a_t(T,K)\}_{0\leq t\leq T}$ is a semi-martingale with decomposition:

$$da_t(T,K) = \alpha_t(T,K)dt + \beta_t(T,K)dW_t, \qquad 0\leq t\leq T. \tag{57}$$

for some d-dimensional Wiener process $\{W_t\}_{t\geq 0}$, and some real valued adapted process $\{\alpha_t(T,K)\}_{0\leq t\leq T}$ and d-dimensional adapted process $\{\beta_t(T,K)\}_{0\leq t\leq T}$ satisfying some mild hypotheses to be specified later. Equivalently, one could specify the dynamics of the local volatility surface parameterized by the time to maturity τ instead of the time of maturity T. In this case, we would assume that

$$d\tilde{a}_t(\tau,K) = \tilde{\alpha}_t(\tau,K)dt + \tilde{\beta}_t(\tau,K)dW_t \tag{58}$$

Using the generalized Îto formula, we see that these two prescriptions are equivalent if and only if

$$\tilde{\alpha}_t(\tau,K) = \alpha_t(t+\tau,K) + \partial_T a(t+\tau,K), \qquad \text{and} \qquad \tilde{\beta}_t(\tau,K) = \beta_t(t+\tau,K). \tag{59}$$

The results of [7] which we review in this section are proven under the following assumption:

Assumption A

For any fixed $\tau \geq 0$ and $K \geq 0$, $\{\tilde{\alpha}_t(\tau,K)\}_{t\geq 0} \in \mathcal{H}^1_{loc}(\mathbb{R})$ and $\{\tilde{\beta}_t(\tau,K)\}_{t\geq 0} \in \mathcal{H}^2_{loc}(\mathbb{R}^d)$. Moreover, we assume that $\mathbb{P}$ - almost surely, for every $t \geq 0$, the functions $(\tau,K) \hookrightarrow \tilde{\alpha}_t(\tau,K)$ and $(\tau,K) \hookrightarrow \tilde{\beta}_t(\tau,K)$ (and hence $(\tau,K) \hookrightarrow \tilde{a}_t(\tau,K)$) are once continuously differentiable in τ and twice continuously differentiable in K.

Also, $\mathbb{P}$ - almost surely
1) for every $t \geq 0$ and all non-negative numbers τ and K

$$|\alpha_t(\tau,K)| + \|\beta_t(\tau,K)\| \leq \lambda_1(\omega,t)$$

$$0 < \lambda_2(\omega,t) \leq \int_0^t \alpha_u(\tau,K)du + \int_0^t \beta_u(\tau,K)dW_u \leq \lambda_3(\omega,t)$$

for some positive adapted processes λ_1, λ_2 and λ_3.

First Technical Results

The first technical result we need to prove before going any further is the fact that for each $\tau > 0$ and $K > 0$, the process $\{\tilde{C}_t(\tau,K)\}_{t\geq 0}$ is a semi-martingale. This result is quite natural. However, its proof is more technical than we would like, and for the purpose of this presentation, we merely outline the major steps of the proof. Complete details can be found in [7].

Under Assumption A, for each fixed $t > 0$, the stochastic differential equation (50) has a unique solution which we denote $\{Y_{t,\tau}\}_{\tau\geq 0}$. Moreover, since $\tilde{a}_t(\tau,K)$ is bounded above and below away from 0, from Feynman-Kac formula and *Theorem 1.3* of [22], the transition density of $\{\log Y_{t,\tau}\}_\tau$ is the fundamental solution of the backward Kolmogorov's equation

$$\partial_\tau u(\tau,x) = \frac{1}{2}\tilde{a}_t^2(\tau,e^x)\partial^2_{xx}u(\tau,x) - \frac{1}{2}\tilde{a}_t^2(\tau,e^x)\partial_x u(\tau,x), \qquad \tau > 0,\ x \in \mathbb{R}$$

From this we conclude that density $g_t(\tau,x)$ of $Y_{t,t+\tau}$ is the fundamental solution of

$$\partial_\tau u(\tau,x) = \frac{1}{2}\cdot K^2\tilde{a}_t^2(\tau,K)\partial^2_{KK}u(\tau,K), \qquad \tau > 0,\ K > 0$$

and the price of the vanilla call option is the solution of barrier problem (54). It is well defined, since after the change of variables

$$\widehat{C}(\tau,x) := \tilde{C}(\tau,\exp x), \qquad \tau \geq 0, x \in \mathbb{R} \tag{60}$$

the option partial differential equation becomes

$$\frac{\partial\widehat{C}(\tau,x)}{\partial\tau}=\frac{1}{2}\tilde{a}_t^2(\tau,e^x)\Delta\widehat{C}(\tau,x)-\frac{1}{2}\tilde{a}_t^2(\tau,e^x)\frac{\partial\widehat{C}(\tau,x)}{\partial x},\qquad T>0,\ x\in\mathbb{R} \quad (61)$$

with initial condition

$$\widehat{C}(0,x)=(S_t-e^x)^+.$$

The proof that option prices are semi-martingales is done in two steps.

1. We first proof the result by replacing the hockey-stick function appearing in the initial condition by a smooth function. In this case, the solution of equation (61) appears as the uniform limit of the results of a finite difference scheme. It is plain to show that any such explicit scheme provides us at each step with a semi-martingale. The convergence being strong enough, one can pass to the limit and prove that the solution of (61) is also a semi-martingale.
2. The general result is obtained by controlling the limit of the solution of (61) when we approximate the hockey-stick initial condition by a smooth regularization.

The details of these arguments are given in [7].

The conclusion of this subsection, and the starting point of the next one are captured by the fact that for each $\tau>0$ and $K>0$ there exist continuous adapted processes $\{\tilde{\mu}_t(\tau,K)\}_{t\geq 0}$ and $\{\tilde{\nu}_t(\tau,K)\}_{t\geq 0}$ such that the following decomposition holds:

$$d\tilde{C}_t(\tau,K)=\tilde{\mu}_t(\tau,K)dt+\tilde{\nu}_t(\tau,K)dB_t. \quad (62)$$

Moreover, the random fields $\{\tilde{\mu}_t(\tau,K)\}_{t,\tau,K}$ and $\{\tilde{\nu}_t(\tau,K)\}_{t,\tau,K}$ satisfy the same assumptions as the random fields $\{\tilde{\alpha}_t(\tau,K)\}_{t,\tau,K}$ and $\{\tilde{\beta}_t(\tau,K)\}_{t,\tau,K}$ appearing in the decomposition of the local volatility $\{\tilde{a}_t(\tau,K)\}_{t,\tau,K}$.

6.5 The HJM Drift Condition

The main goal of this subsection is to derive the following analog of the HJM no-arbitrage analysis.

Theorem 1 (Drift and Consistency Conditions). *The dynamic model of the local volatility surface given by the system of equations*

$$d\tilde{a}_t(\tau,K)=\tilde{\alpha}_t(\tau,K)dt+\tilde{\beta}_t(\tau,K)dW_t,\qquad t\geq 0, \quad (63)$$

with coefficients satisfying assumption A is consistent with a spot price model of the form

$$dS_t=S_t\sigma_t dB_t$$

for some Wiener process $\{B_t\}_t$, and does not allow for arbitrage if and only if the following conditions are satisfied a.s. for all $t>0$:

$$\bullet\tilde{a}_t(0,S_t)=\sigma_t \quad (64)$$

$$\bullet\partial_\tau\tilde{a}_t(\tau,K)\partial^2_{KK}\tilde{C}_t(\tau,K)= \quad (65)$$

$$\left(\tilde{a}_t(\tau,K)\tilde{\alpha}_t(\tau,K)+\frac{\|\beta_t(\tau,K)\|^2}{2}\right)\partial^2_{KK}\tilde{C}_t(\tau,K)+\frac{d}{dt}\langle\tilde{a}_\cdot(\tau,K)^2,\partial^2_{KK}\tilde{C}_\cdot(\tau,K)\rangle_t$$

where we use the notation $\langle\ \cdot\quad\cdot\ \rangle_t$ for the quadratic covariation of two semi-martingales.

Proof:

By construction of the local volatility surface $\{\tilde{a}_t(\tau, K)\}_{\tau>0,K>0}$ we have the equality

$$K^2\tilde{a}_t^2(\tau, K)\partial^2_{KK}\tilde{C}_t(\tau, K) = 2\partial_\tau\tilde{C}_t(\tau, K)$$

which we can rewrite as

$$K^2\tilde{a}_t^2(T-t, K)\partial^2_{KK}\tilde{C}_t(T-t, K) = 2\partial_\tau\tilde{C}_t(T-t, K) \tag{66}$$

for $0 \le t \le T$. Both sides are semi-martingales. We use Îto's rule to compute the differential of $\tilde{a}_t^2(\tau, K)$.

$$\begin{aligned} d\tilde{a}_t^2(\tau, K) &= 2\tilde{a}_t(\tau, K)d\tilde{a}_t(\tau, K) + \|\tilde{\beta}_t(\tau, K)\|^2 dt \\ &= (2\tilde{a}_t(\tau, K)\tilde{\alpha}_t(\tau, K) + \|\tilde{\beta}_t(\tau, K)\|^2)dt + 2\tilde{a}_t(\tau, K)\tilde{\beta}_t(\tau, K)dW_t, \end{aligned}$$

However, we also have

$$\begin{aligned} d\tilde{a}_t^2(T-t, K) = (&-2\tilde{a}_t(T-t, K)\partial_\tau\tilde{a}_t(T-t, K) + 2\tilde{a}_t(T-t, K)\tilde{\alpha}_t(T-t, K) \\ &+ \|\tilde{\beta}_t(T-t, K)\|^2)dt + 2\tilde{a}_t(T-t, K)\tilde{\beta}_t(T-t, K)dW_t, \end{aligned} \tag{67}$$

because the effect of replacing τ by $T-t$ in a stochastic differential is merely an argument substitution (τ by $T-t$) in the local martingale part, while a new term, typically a partial derivative with respect to τ, is also added to the drift or bounded variation part of the differential. Consequently

$$\begin{aligned} &d\left(\tilde{a}_t(T-t, K)^2\partial^2_{KK}\tilde{C}_t(T-t, K)\right) \\ &\quad= \partial^2_{KK}\tilde{C}_t(T-t, K)d\tilde{a}_t^2(T-t, K) + \tilde{a}_t^2(T-t, K)d\left(\partial^2_{KK}\tilde{C}_t(T-t, K)\right) \\ &\qquad + d\langle\tilde{a}_\cdot^2(T-\cdot, K), \partial^2_{KK}\tilde{C}_\cdot(T-\cdot, K))\rangle_t \end{aligned} \tag{68}$$

Since $\tilde{C}_t(\tau, K)$ is a semi-martingale for every fixed $\tau > 0$ and $K > 0$, if we write its decomposition as (recall formula (62))

$$d\tilde{C}_t(\tau, K) = \tilde{\mu}_t(\tau, K)\,dt + \tilde{\nu}_t(\tau, K)dW_t$$

then for each fixed $K > 0$, $\{C_t(T, K) = \tilde{C}_t(T-t, K)\}_{0\le t\le T}$ is also a semi-martingale and its decomposition is given by

$$dC_t(T, K) = d\tilde{C}_t(T-t, K) = [\tilde{\mu}_t(T-t, K) - \partial_\tau\tilde{C}_t(T-t, K)]dt + \tilde{\nu}_t(\tau, K)dW_t. \tag{69}$$

1). Let us first assume absence of arbitrage. As we explained earlier, what we mean by that is the fact that the prices of all the liquidly traded assets are local martingales. In particular, for every fixed $T > 0$ and $K > 0$, the process $\{C_t(T, K) = \tilde{C}_t(T-t, K)\}_{0\le t\le T}$ is a local martingale. On one hand, this implies that $\partial_\tau\tilde{C}_t(T-t, K)$ is a local martingale, and on the other hand that the bounded variation part of the left hand side of equation (66) is equal to 0. Since developing (68) using (67) gives:

$$\begin{aligned} &d\left(\tilde{a}_t(T-t, K)^2\partial^2_{KK}\tilde{C}_t(T-t, K)\right) \\ &\quad= \partial^2_{KK}\tilde{C}_t(T-t, K)d\tilde{a}_t^2(T-t, K) + \tilde{a}_t^2(T-t, K)d\left(\partial^2_{KK}\tilde{C}_t(T-t, K)\right) \\ &\qquad + d\langle\tilde{a}_\cdot^2(T-\cdot, K), \partial^2_{KK}\tilde{C}_\cdot(T-\cdot, K))\rangle_t \\ &\quad= \partial^2_{KK}\tilde{C}_t(T-t, K)\,(2\tilde{a}_t(T-t, K)[\tilde{\alpha}_t(T-t, K) - \partial_\tau\tilde{a}_t(T-t, K)] \\ &\qquad + \|\tilde{\beta}_t(T-t, K)\|^2\Big)\,dt + d\langle\tilde{a}_\cdot^2(T-\cdot, K), \partial^2_{KK}\tilde{C}_\cdot(T-\cdot, K))\rangle_t \\ &\qquad + d(\text{local martingale})_t \end{aligned}$$

and setting the drift component to 0 gives (65).

2). Let us now prove the consistency condition (64). Using the fact that

$$\lim_{\tau\searrow 0}\tilde{C}_t(\tau,K)=(S_t-K)^+ \;\; a.s.$$

one can prove that

$$\lim_{\tau\searrow 0}\int_0^t \tilde{\nu}_u(\tau,K)\cdot dW_u=\int_0^t S_u\sigma_u\mathbf{1}_{\{S_u-K\geq 0\}}dB_u$$

for the uniform convergence in probability. This implies that there a.s. exists a sequence $\{\tau_n\}_{n=1}^{\infty}$ decreasing to 0 for which

$$\lim_{n\to\infty}\int_0^t \tilde{\nu}_u(\tau_n,K)\cdot dW_u=\int_0^t S_u\sigma_u\mathbf{1}_{\{S_u-K\geq 0\}}dB_u$$

which shows that (again because of Tanaka's formula) that

$$\lim_{n\to\infty}\int_0^t \tilde{\mu}_u(\tau_n,K)du=\Lambda_t(K),\quad for\ any\ K>0\ and\ t\in[0,\bar{t}]$$

where $\Lambda_t(K)$ denotes the local time of S_t at K. Since $\tilde{C}_t(T-t,K)$ is a local-martingale in t, we have

$$\tilde{\mu}_u(\tau_n,K)=\partial_\tau\tilde{C}_u(\tau_n,K)=\frac{1}{2}K^2\tilde{a}_u^2(\tau_n,K)\partial^2_{KK}\tilde{C}_u(\tau_n,K)$$

which in turn implies that for any continuous function h with compact support, we have:

$$\int_0^t h(S_u)S_u^2[\sigma_u^2-a_u^2(0,S_u)]du=0$$

from which we can deduce the consistency condition since h is arbitrary.

3) We now consider the converse. As for the proof of the direct part, the details are technical, so we limit our discussion to the main steps, referring the interested reader to [7] for details. If we denote the drift of $\tilde{C}_t(T-t,K)$ by $\tilde{v}_t(T,K)$, smoothness of $\tilde{C}_t(.,.)$, $\tilde{\mu}_t(.,.)$ and $\tilde{\nu}_t(.,.)$ guarantee the required $C^{1,2}$ smoothness of $\tilde{v}_t(.,.)$. Our goal is to show that $\tilde{v}_t(.,.)$ vanishes identically. In order to do so, we first prove that it is the solution of a parabolic partial differential equation, and then we check that the initial condition it satisfies is identically 0. The first step is rather straightforward. By differentiation in the same way as in the first part of the proof, and using the fact that $\tilde{v}_t=\tilde{\mu}_t-\partial_\tau\tilde{C}_t$, we obtain

$$\partial_\tau\tilde{v}_t(\tau,K)=\frac{1}{2}K^2\tilde{a}_t^2(\tau,K)\partial^2_{KK}\tilde{v}_t(\tau,K),\qquad \tau>0,K>0.$$

For the remainder of the proof we show that it is possible almost surely to construct a subsequence $\tau_n\searrow 0$ such that $\tilde{v}_t(\tau_n,\cdot\,)\to 0$ weakly as functions of K. Uniqueness of weak solutions of the above partial differential equation guarantees that we have $\tilde{v}_t(\tau,K)=0$ for all $\tau>0$ and $K>0$. This implies that $\tilde{C}_t(T-t,K)$ is a local martingale in t for any $T>0$ and $K>0$, and since $\tilde{C}_t\leq S_t$ is square integrable we can conclude that $\{\tilde{C}_t(T-t,K)\}_{0\leq t\leq T}$ is a bona fide martingale.

Monte Carlo Implementation

We now explain how the drift condition (65) can be used to set up arbitrage-free dynamic models for the local volatility surface. As we already explained, we are not able to use (65) to derive a formula in close form to express the drift surface $\{\tilde{\alpha}_t(\tau, K)\}_{\tau,K}$ as a function of the volatility surface $\{\tilde{\beta}_t(\tau, K)\}_{\tau,K}$. However, it is possible to use a discretized version, in the spirit of the Euler scheme for ordinary stochastic differential equations, to constructively derive Monte Carlo samples of the volatility surface from the mere knowledge of $\{\tilde{\beta}_t(\tau, K)\}_{\tau,K}$.

- Start from a model for $\beta_t(\tau, K)$ (say a stochastic differential equation);
- Get S_0 and $C_0(\tau, K)$ from the market and compute $\partial^2_{KK} C_0$, a_0 and β_0 from its model;
- Loop: for $t = 0, \Delta t, 2\Delta t, \cdots$
  1. Get $\alpha_t(\tau, K)$ from the drift condition (65);
  2. Use Euler to get
     – $a_{t+\Delta t}(\tau, K)$ from the dynamics of the local volatility given by (63);
     – $S_{t+\Delta t}$ from S_t Dynamics;
     – $\beta_{t+\Delta t}$ from its own model;

6.6 Examples.

This last subsection is devoted to the applications of the above approach to two of the most popular spot models.

Markovian Spot Models

Let us first consider the simplest case $\beta \equiv 0$. In this case

$$\tilde{a}_t(\tau, K) = \tilde{a}_0(\tau, K) + \int_0^t \tilde{\alpha}_s(\tau, K) ds$$

and in particular we have

$$\tilde{\alpha}_t(\tau, K) = \frac{d}{dt} \tilde{a}_t(\tau, K).$$

In the present situation, the drift condition (65) reads

$$\partial_\tau \tilde{a}_t(\tau, K) = \tilde{\alpha}_t(\tau, K)$$

and putting the two together we get

$$\partial_\tau \tilde{a}_t(\tau, K) = \frac{d}{dt} \tilde{a}_t(\tau, K)$$

which shows that for fixed K the function $\tilde{a}_t(\tau, K)$, as a function of t and τ, is the solution of a plain (hyperbolic) transport equation whose solution is given by:

$$\tilde{a}_t(\tau, K) = \tilde{a}_0(\tau + t, K)$$

and the consistency condition forces the special form

$$\sigma_t = a_0(t, S_t)$$

of the spot volatility. Hence we proved:

Proposition 3. *The local volatility is a process of bounded variation for each τ and K fixed if and only if it is the deterministic shift of a constant shape and the underlying spot is a Markov process.*

Stochastic Volatility Models

Next we attempt to bridge our analysis of the dynamics of the local volatility with stochastic volatility models widely used in the industry. We start with an explicit form for the dynamics of the stock and spot volatility under a risk-neutral measure, and we derive an *explicit* form for the local volatility surface together with the random fields $\tilde{\alpha}_t(.,.)$, $\tilde{\beta}_t(.,.)$ at each fixed time t.

For the sake of illustration, we consider a simplified version of the SABR model with a stochastic volatility given by a geometric Brownian motion. To be specific we assume that

$$\begin{aligned} dS_t &= S_t \sigma_t dB_t^1 \\ d\sigma_t &= \sigma_t \tilde{\sigma} dB_t^2 \end{aligned}$$

with initial conditions $S_0 = S$ and $\sigma_0 = \sigma$. Here, $\tilde{\sigma} > 0$ is a constant (usually called the vol-vol) and $\{B_t^1\}_{t\geq 0}$ and $\{B_t^2\}_{t\geq 0}$ are standard Wiener processes. If we also assume that these two Wiener processes are independent, by conditioning on $\mathcal{F}^{B^2}$ we can easily obtain a closed form formula for the call prices at time zero:

$$\tilde{C}_0(\tau, K) = \mathbb{E}\left[BS\left(S, \tau, \sqrt{\frac{1}{\tau}\int_0^\tau \sigma_u^2 du}, K\right)\right]$$

where the notation $BS(S, \tau, \sigma, K)$ for the Black-Scholes price of a European call option was introduced in (49). We can then compute the partial derivatives with respect to τ and K passing the derivatives under the expectation and get from (53) the following formula for the local volatility

$$\tilde{a}_0^2(\tau, K) = \frac{S}{K} \cdot \frac{\mathbb{E}\left[\left(2\sigma_\tau^2/\bar{\sigma_\tau} - \bar{\sigma_\tau}\right) e^{-d_1^2/2}\right]}{\mathbb{E}\left[e^{-d_2^2/2}/\bar{\sigma_\tau}\right]} \tag{70}$$

where

$$\bar{\sigma_\tau} = \sqrt{\frac{1}{\tau}\int_0^\tau \sigma_u^2 du} \tag{71}$$

and

$$d_1 = \frac{\log(S/K) + \bar{\sigma_\tau}^2 T/2}{\bar{\sigma_\tau}\sqrt{T}} \quad \text{and} \quad d_2 = d_1 - \bar{\sigma_\tau}\sqrt{\tau}. \tag{72}$$

The independence assumption is often made for the computations to be easier, but it is not necessary. Indeed, similar formula can be obtained if we assume that the two Wiener processes are correlated, say if they satisfy $dB_t^1\, dB_t^2 = \rho\, dt$. In this case, the formula for the price of a call option becomes

$$\tilde{C}_0(\tau, K) = \mathbb{E}\left[BS\left(Se^{\frac{\rho\sigma_0}{\tilde{\sigma}}(\tilde{\sigma}_T - 1) - \sigma_0^2 \frac{\rho^2}{2} T \bar{\sigma}_T^2}, K, \tau, \sqrt{1-\rho^2}\sigma_0 \bar{\sigma}_T\right)\right]$$

where $\bar{\sigma_\tau}$ is defined as above in (71), and $\tilde{\sigma}_t = \sigma_t/\sigma_0$. It now holds

$$\tilde{a}_0^2(\tau, K) = \sigma_0^2\sqrt{1-\rho^2}\frac{\mathbb{E}\left[\sigma_\tau^2/\bar{\sigma_\tau} e^{-d_1^2/2}\right]}{\mathbb{E}\left[e^{-d_2^2/2}/\bar{\sigma_\tau}\right]} \tag{73}$$

where d_1 is now defined by

$$d_1 = \frac{\log(S/K) + \rho\frac{\sigma_0}{\sigma_\tau}(\tilde{\sigma}_\tau - 1) + (0.5 - \rho^2)\sigma_0^2\tilde{\sigma}_\tau^2\tau}{\sqrt{1-\rho^2}\sigma_0\tilde{\sigma}_\tau\sqrt{\tau}}. \tag{74}$$

Example of these local volatility surfaces are given in [7].

6.7 Factor Models and Consistency

In our discussion of the classical HJM approach to the fixed income markets in Section 3, we explained the important role played by the use of factor models based on parametric families of forward curves. Motivated by the computations of the previous subsection, we single out a simple parametric family of two-dimensional surfaces which appear to give a reasonable parametric family for local volatility surfaces. This family is given by the local volatility surfaces of stochastic volatility models where the stochastic volatility is restricted to take only three different values.

Example of a Parametric Family of Local Volatility Surfaces

Parametric families of forward curves have played a crucial role in the major developments in the econometric analysis of interest rate data. Moreover, they were also a major impetus in some of the recent the formulation and the solution of the consistency problem. As far as we know, parametric families of local volatility surfaces have not been introduced and systematically studied, at least with the same intensity, and at least in the academic literature. For the sake of definiteness we introduce a simple example of such a family. For each (multivariate) parameter

$$\Theta = (\sigma, \sigma_1, \sigma_2, p_1, p_2)$$

such as $\sigma > 0$, $\sigma_1 > 0$, $\sigma_2 > 0$, $p_1 > 0$, $p_2 > 0$, and also satisfying $p_1 + p_2 \le 1$, we use formula (53) to define a surface $\tilde{a}_0(\tau, K)$ from a call function $\tilde{C}_0(\tau, K)$ obtained by randomization of the volatility assuming that it takes the values σ_1, σ and σ_2 with probabilities p_1, $1 - p_1 - p_2$ respectively. Consequently,

$$\tilde{a}_0(\tau, K)^2 = \frac{p_1 \partial_\tau C(\sigma_1) + (1 - p_1 - p_2)\partial_\tau C(\sigma) + p_2 \partial_\tau C(\sigma_2)}{p_1 \partial^2_{KK} C(\sigma_1) + (1 - p_1 - p_2)\partial^2_{KK} C(\sigma) + p_2 \partial^2_{KK} C(\sigma_2)} \tag{75}$$

where we use the notation $C(\tilde{\sigma})$ for the Black-Scholes price $\tilde{C}_0(\tau, K)$ if the volatility parameter is $\tilde{\sigma}$. Now, using the following expressions for the partial derivatives of the Black-Scholes price

$$\partial_\tau B(S, K, \tau, \sigma) = \frac{\sqrt{SK}}{\sqrt{2\pi}} \frac{\sigma}{2\sqrt{\tau}} e^{-(\log S/K)^2/2\sigma^2\tau - \tau\sigma^2/8}$$

and

$$K^2 \partial^2_{KK} B(S, K, \tau, \sigma) = \frac{\sqrt{SK}}{\sqrt{2\pi}} \frac{1}{\sigma\sqrt{\tau}} e^{-(\log S/K)^2/2\sigma^2\tau - \tau\sigma^2/8}$$

we get the following formula for the definition of our local volatility parametric family:

$$a^2(\tau, x, \Theta) = \frac{\sum_{i=0}^2 p_i \sigma_i e^{-x^2/(2\tau\sigma_i^2) - \tau\sigma_i^2/8}}{\sum_{i=0}^2 (p_i/\sigma_i) e^{-x^2/(2\tau\sigma_i^2) - \tau\sigma_i^2/8}} \tag{76}$$

where we use the variable x for the log-moneyness $\log(S/K)$ and where we set $p_0 = 1 - p_1 - p_2$ and $\sigma_0 = \sigma$ to simplify the form of the formula. Figure 2 gives an example of such a surface.

This plot clearly hints at one of the major shortcomings of this family: the singular behavior of the surface for short time to maturity, i.e. for $\tau \searrow 0$. Indeed $a^2(\tau, x, \Theta)$ converges toward the maximum of the three $sigma_i$ when $\tau \searrow 0$ and $x \neq 0$, while the same limit is strictly smaller (a weighted average of the σ_i's) when $x = 0$. Possible fixes to this problem include the choice of time dependent volatilities σ_i, and a solution in this spirit is implemented in [7] where a different parametric family is proposed.

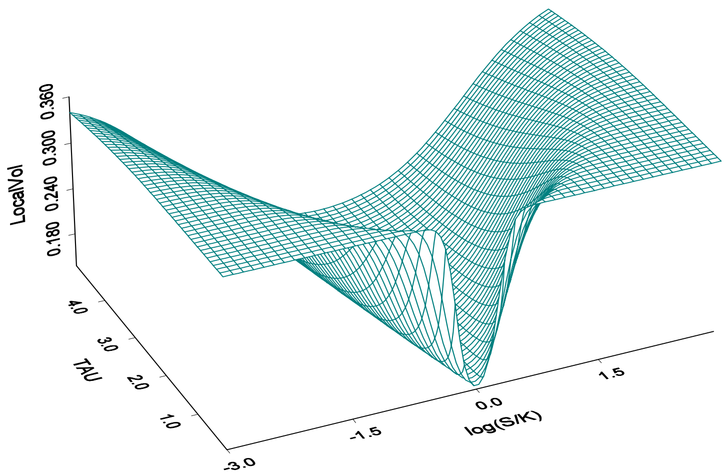

Fig. 2. Parametric local volatility surface from the family described in the text. We used the parameters $\sigma_0 = 0.4$, $\sigma_1 = 0.3$, $\sigma_2 = 0.6$, $p_0 = 0.3$, and $p_1 = 0.5$.

6.8 Local Volatility Factor Models

Studying the consistency of local volatility factor models is a very interesting problem, and as far as we know, such a problem is completely open. As explained in Section 3, factor models are based on the choice of a parametric family as defined in (76) for example. So if we assume that we are given a parametric family G as before and if we suppose that $\boldsymbol{\Theta} = \{\theta_t\}_{t\geq 0}$ is a d-dimensional semi-martingale with values in the parameter space Θ, then consistency of the factor model means that the random field

$$a_t(\tau, K) = G(\theta_t, \tau, K), \qquad t \geq 0,\, \tau > 0,\, K > 0.$$

gives a local volatility model satisfying the no-arbitrage condition.

References

1. Y. Achdou and O. Pironneau, *Computational methods for option pricing*, SIAM.
2. T. Aven, *A theorem for determining the compensator of a counting process*, Scandinavian Journal of Statistics **12(1)** (1985), 69–72.
3. N. Bennani, *The forward loss model: a dynamic term structure approach for the pricing of portfolios of credit derivatives*, Tech. report, November 2005.
4. D. Blackwell, *Equivalent comparisons of experiments*, Annals of Mathematical Statistics **24** (1953), 265–272.
5. H. Buehler, *Consistent variance curve models*, Tech. report, 2007.
6. ______, *Expensive martingales*, Quantitative Finance (2007), (to appear).
7. R. Carmona and S. Nadtochiy, *Hjm dynamics for equity models*, Tech. report, Princeton University, 2007.
8. R. Carmona and M. Tehranchi, *A characterization of hedging portfolios for interest rate contingent claims*, The Annals of Applied Probability **14** (2004), 1267–1294.

9. ______, *Interest rate models: an infinite dimensional stochastic analysis perspective*, Springer Verlag, 2006.
10. P. Carr and D. Madan, *Toward a theory of volatility trading*, vol. Volatility, Risk Publications, pp. 417 – 427.
11. A. Carter and J.P. Fouque, *Review of* SemiParametric Modeling of Implied Volatility *by matthisas fengler*, SIAM Reviews (2007), (to appear).
12. R. Cont and J. da Fonseca, *Dynamics of implied volatility surfaces*, quantitative Finance **2** (2002), 45–60.
13. L. Cousot, *Necessary and sufficient conditions for no-static arbitrage among european calls*, Tech. report, Courant Institute, New York University, Octobber 2004.
14. R. Jarrow D. Heath and A. Morton, *Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation*, Econometrica **60** (1992), 77 – 105.
15. M.H.A. Davis and D.G. Hobson, *The range of traded option prices*, Tech. report, Princeton University, July 2005.
16. E. Derman and I. Kani, *The volatility smile and its implied tree*, Tech. report, Quantitative Research Notes, Goldman Sachs, 1994.
17. ______, *Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure of volatility*, International Journal of Theoretical and Applied Finance **1** (1998), 61–110.
18. D. Duffie and K. Singleton, *Credit risk*, Princeton University Press, 2003.
19. B. Dupire, *Pricing with a smile*, Risk **7** (1994), 32–39.
20. V. Durrleman, *From implied to spot volatility*, Tech. report, Stanford University, April 2005.
21. V. Durrleman and N. El Karoui, *Coupling smiles*, Tech. report, Stanford University, November 2006.
22. E.B Dynkin, *Diffusions, superdiffusions and partial differential equations*, American Mathematical Society, 2002.
23. M.R. Fengler, *Semiparametric modeling of implied volatility*, Lecture Notes in Statistics, Springer Verlag, 2005.
24. D. Filipovic, *Exponential-polynomial families and the term structure of interest rates*, Bernoulli **6(6)** (2000), 1–27.
25. ______, *Consistency problems for heath-jarrow-morton interest rate models*, Lecture Notes in Mathematics, vol. 1760, Springer-Verlag, 2002.
26. V. Piterbarg J. Sidenius and L. Andersen, *A new framework for dynamic credit portfolio loss modelling*, Tech. report, October 2005.
27. J. Jacod and P. Protter, *Risk neutral compatibility with option prices*, Tech. report, Université de Paris VI and Cornell University, April 2006.
28. H. Kellerer, *Markov-komposition und eine anwendung auf martingale*, Mathematische Annalen **198** (1972), 99–122.
29. D. Lando, *Credit risk*, Princeton University Press, 2004.
30. J.P. Laurent and D. Leisen, *Building a consistent pricing model from observed option prices*, World Scientific, 2000.
31. R. Lee, *Implied volatility: Statics, dynamics, and probabilistic interpretation*, International Journal of Theoretical and Applied Finance **4** (2001), 45–89.
32. ______, *Implied and local volatilities under stochastic volatility*, Recent Advances in Applied Probability, Springer Verlag, 2004.
33. ______, *The moment formula for implied volatility at extreme strikes*, Mathematical Finance **14** (2004), 469–480.
34. J. da Fonseca R. Cont and V. Durrleman, *Stochastic models of implied volatility surfaces*, Economic Notes **31** (2002), no. 2.
35. D. Brigo R. Torresetti and A. Pallavicini, *Implied expected tranche loss surface from cdo data*, Tech. report, Banca IMI, March 2007.
36. ______, *Term structure, tranche structure, and loss distributions*, Tech. report, University of Toronto, January 2007.

37. C. Rebonato, *Volatility and correlation: the perfect hedger and the fox.*, 2nd ed., Wiley, 2004.
38. D. Revuz and M. Yor, *Continuous martingales and brownian motion*, 2nd ed., Springer-Verlag, 1990.
39. P. Schönbucher, *A market model for stochastic implied volatility*, Phil. Trans. of the Royal Society, Series A **357** (1999), 2071 – 2092.
40. ______, *Credit derivatives pricing models*, Wiley, 2003.
41. ______, *Portfolio losses and the term structure of loss transitition rates: a new methodology for the pricing of portfolio credit derivatives*, Tech. report, ETH Z December 2005.
42. M. Schweizer and J. Wissel, *Term structures of implied volatilities: Absence of arbitrage and existence results*.
43. S. Sherman, *On a theorem of hardy, littlewood, polya and blackwell*, Proc. National Academy of Sciences USA **37** (1951), 826–831.
44. C. Stein, *Notes on the comparison of experiments*, Tech. report, University of Chicago, 1951.