1 The LMM-SABR Model

So far we have briefly presented the standard (ie, deterministic-volatility) LIBOR market model and the SABR model and we have explained why we like both these approaches. We will show that the SABR model describes surprisingly well how the smile moves in reality when the underlying moves. This feature is at the root of successful delta hedging and these positive results reinforce our belief that the SABR model is a good place to start. What we need is a dynamic extension of the static SABR model. This can be provided by the SABR-LIBOR Market Model.

The LIBOR Market Model was the standard for pricing complex products before smiles appeared in the interest-rate markets. Monotonically-declining smiles can, to some extent, be shoe-horned within the framework of the LIBOR market model. **'Hockey-stick' smiles, however, require more radical surgery.**

There are many ways to obtain non-monotonic smiles in interest rates.The differences among most of these stochastic-volatility models are not great. The precise distributional assumptions for the volatility process are sometimes different; displaced diffusion sometimes replaces the CEV process; the numerical implementation can be more or less cumbersome. However, since volatility is a latent variable, determining what its 'true' distribution should be inhabits a murky area between heroic econometric analysis and an act of faith. As for the 'correct' exponent for the CEV process of the forward rates, empirical analysis seems to suggest that no single exponent works well for all levels of forward rates. For these 'fine details' of the stochastic-volatility models, we therefore believe that practicality and tractability should win the day.

What all the stochastic-volatility models mentioned above *do* share, however, is an inability to fit accurately the market smile surface in a manner which is simple, quick, robust and financially justifiable (ie, without too many time-dependent 'fudge factors'.)

If this is the case, why not take the market-standard SABR model as the blue-print for the required stochastic-volatility extension of the LIBOR market model? We do not know whether the distributional assumptions for the volatility or for the CEV process for the forward rates implied by the SABR model are, statistically speaking, the 'best' possible modelling choices. But, in such an uncertain landscape, the more the dynamic extension of the LIBOR market model 'looks and feels' like the SABR model, the better our hopes will be to recover the smile surface, again, in a manner which is simple, quick, robust and financially justifiable.

This is the route taken by Henry-Labordere (2007) and Rebonato (2007). We believe that the second approach is simpler, both to understand and to use. For this reason, the rest of the book is devoted to showing how to turn a promising idea into a complete pricing framework – by which we mean not just deriving the no-arbitrage 'equations of motions', but also the calibration, the empirical evidence to implement it, and the hedging.

2 The Equations of Motion

We start directly from a formulation that allows for a natural decomposition of the stochastic part of the dynamics into a component that only depends on the volatilities and one that only depends on the correlations.

Therefore, for $i = 1, 2, ..., N$ let the joint dynamics of the N forward rates and their instantaneous volatilities be defined by:

$$df_t^i = \mu_t^i(\{\boldsymbol{f}\}, \{\boldsymbol{s}\}, \boldsymbol{\rho})dt + \left(f_t^i\right)^{\beta_i} s_t^i \sum_{j=1}^{N_F} b_{ij} d\hat{z}_j \tag{1}$$

$$ds_t^i = g(t, T_i)dk_t^i \tag{2}$$

$$\frac{dk_t^i}{k_t^i} = \mu_t^{k_i}(\{\boldsymbol{f}\}, \{\boldsymbol{s}\})dt + h_t^i \sum_{j=1}^{N_V} c_{ij} d\hat{w}_j \tag{3}$$

where N_F and N_V are the number of factors driving the forward-rate and volatility processes, respectively, and with $N_F \leq N$, $N_V \leq N$. The notation $\mu_t^i(\{\boldsymbol{f}\}, \{\boldsymbol{s}\})$ and $\mu_t^{k_i}(\{\boldsymbol{f}\}, \{\boldsymbol{s}\})$ indicates that, in general, the no-arbitrage drifts of the forward rates and of the volatilities depend on all the forward rates and volatilities.

If one sets

$$\sum_{j=1}^{N_F} b_{ij}^2 = 1 \tag{4}$$

$$\sum_{j=1}^{N_V} c_{ij}^2 = 1$$

and

$$\mathbb{E}\left[d\hat{z}_j d\hat{z}_k\right] = \delta_{jk} dt \tag{5}$$

$$\mathbb{E}\left[d\hat{w}_j d\hat{w}_k\right] = \delta_{jk} dt \tag{6}$$

$$\mathbb{E}\left[d\hat{z}_j d\hat{w}_k\right] = x_{jk} dt \tag{7}$$

where δ_{ij} is the Kronecker delta and x_{ij} is defined later, then each forward rate

f_t^i and each factor k_t^i will have a CEV or lognormal instantaneous volatility s_t^i and h_t^i, respectively.

3 The Nature of the Stochasticity Introduced by the Model

A few observations are in order. We have introduced stochasticity in the volatility by implicitly defining a quantity, say, p_t, defined as:

$$p_t = \left(\frac{s_t^i}{g(t, T_i)}\right) \tag{8}$$

This can be regarded as the ratio of the total volatility to its deterministic part (or, as we shall see, to the time-t_0 expectation of the stochastic volatility). If, at time t_0, the functional dependence on expiry, T_i, of the expectations of the volatilities for all the forward rates, f^i, is perfectly captured by the function $g(t, T_i)$, then the initial values of the the processes k_t^i would be identically equal to 1 (or, for that matter, to the same constant):

$$k_0^i = 1, \quad i = 1, 2, ..., N \tag{9}$$

If this is not the case the initial values k_0^i will not all be identical. In this case, we show below how to choose the $g(t, T_i)$ in such a way that these non-constant initial values for the scaling factors are in some sense optimal. For the moment, we simply note that, with a slight abuse of notation, we have used the same symbol for the scaling factors introduced before and the stochastic process in Equation (3). This is not a coincidence. We pointed out there that, if all the $\{k^i\}$ had been chosen to be exactly equal to 1, the evolution of the smile surface would have been exactly self-similar. In the present formulation the factors $\{k^i\}$ have become stochastic processes, and therefore the future smile surface will display stochasticity. This is obviously desirable. Indeed, it is the whole point behind introducing a stochastic-volatility model. **However, in general we still want that today's expectation of the future smile surface should not display a deterministic change – unless, of course, we believe that the smile should undergo a deterministic transformation**. The formulation above ensures that this is the case.

It also says something about future **conditional expectations of the smile surface**. If all the future stochastic values of the terms $\{k^i\}$ evolved to have the same value, this would roughly correspond to a shift upward or downward of future expectation the whole conditional smile surface. This is desirable, because indeed we observe that the dominant mode of deformation of the smile surface is a quasi-parallel shift.

To the extent that the future values $\{k^i\}$ are different, however, the future expectation of the conditional smile surface implies a change in its shape. **If we want to limit the degree of these deformations, we therefore would like to impose a high degree of correlation among the processes** $\{k^i\}$.

As far as the function $g(t, T_i)$ is concerned, we will require in the following that it should be time-homogeneous:

$$g(t, T_i) = g(T_i - t) \tag{10}$$

We will use the functional form discussed before, but nothing hangs on this. The reader could even decide to dispense with time homogeneity. The equations that follow would still remain valid, but, in normal market conditions, the ‘physics’ would be poorer – the more so, the stronger the dependence on calendar time.

4 A Simple Correlation Structure

The correlation, ρ_{ij}, between two forward rates (say, i and j) is given by:

$$\rho_{ij} = \sum_{k=1}^{N_F} b_{ik} b_{jk} \tag{11}$$

The correlation, r_{ij}, between two instantaneous volatilities (say, i and j) is given by:

$$r_{ij} = \sum_{k=1}^{N_V} c_{ik} c_{jk} \tag{12}$$

Define the matrix $\boldsymbol{b} = \left\{ b_{ij},\ i = 1, ..., N,\ j = 1, ..., N_F \right\}$. Then

$$\boldsymbol{\rho} = \boldsymbol{b}\boldsymbol{b}^T \tag{13}$$

where the superscript T denotes transpose. Similarly, define the matrix $\boldsymbol{c} = \left\{c_{ij},\ i = 1, ..., N,\ j = 1, ..., N_V\right\}$. Then

$$\boldsymbol{r} = \boldsymbol{c}\boldsymbol{c}^T \tag{14}$$

For exogenously-assigned matrices $\boldsymbol{\rho}$ and $\boldsymbol{r}$, the exact recovery of all their elements will in general not be possible unless $N_V = N_F = N$. If $N_V = N_F = N$ the matrices $\boldsymbol{b}$ and $\boldsymbol{c}$ are given by the eigenvectors of the matrices $\boldsymbol{\rho}$ and $\mathbf{r}$, respectively.

Consider now the correlation between forward rate i and volatility q, R_{iq}. This is given by:

$$R_{iq} = \sum_{k=1}^{N_F} \sum_{s=1}^{N_V} b_{ik} c_{qs} x_{ks} \tag{15}$$

The quantities $\{b_{ik}\}$ and $\{c_{qs}\}$ are already fixed from the requirement to match the exogenous matrices $\boldsymbol{\rho}$ and $\boldsymbol{r}$. For an exogenous $N \times N$ matrix $\boldsymbol{R}$ we are then left with $N_V \times N_F$ quantities (the x_{sk}) to fix the N^2 quantities $\boldsymbol{R}$. An exact solution will in general not be possible unless $N_V = N_F = N$.

5 A More General Correlation Structure

We present in this section a more symmetric correlation structure that sometimes can make the book-keeping and the implementation simpler and neater. The main advantage is that now all the Brownian increments (ie, the increments that we denoted $\{d\hat{z}_i\}$ and $\left\{d\hat{w}_j\right\}$ above) are *all* assumed to be independent, and are therefore treated on the same footing.

The joint dynamics of the forward rates and their instantaneous volatilities can be now be written as:

$$df_t^i = \mu_t^i dt + \left(f_t^i\right)^{\beta_i} s_t^i \sum_{j=1}^{M} e_{ij} dy_j, \quad i = 1, ..., N \tag{16}$$

$$ds_t^i = g(t, T_i) dk_t^i \tag{17}$$

$$\frac{dk_t^i}{k_t^i} = \mu_t^{k_i} dt + h_t^i \sum_{j=1}^{M} e_{N+i,j} dy_j, \quad i = 1, ..., N \tag{18}$$

for dy, an M-dimensional Brownian increment, with

$$\mathbb{E}\left[dy_j dy_k\right] = \delta_{jk} dt \tag{19}$$

$$M = N_V + N_F \tag{20}$$

$$\sum_{j=1}^{M} e_{ij}^2 = 1 \tag{21}$$

The associated super-correlation matrices, $\boldsymbol{P}$, all fall under the general structure

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{\rho} & \boldsymbol{R} \\ \boldsymbol{R}^T & \boldsymbol{r} \end{bmatrix} \tag{22}$$

with $\boldsymbol{\rho} = [N_F \times N_F]$, $\boldsymbol{r} = [N_V \times N_V]$, $\boldsymbol{R} = [N_F \times N_V]$. The elements of $\boldsymbol{P}$ are given by

$$P_{ik} = \sum_{m=1}^{M} e_{im} e_{km} \tag{23}$$

or, in matrix notation,

$$\boldsymbol{P} = \boldsymbol{e}\boldsymbol{e}^T \tag{24}$$

If $N_V = N_F = N$ then $M = 2N$, the matrix $\boldsymbol{P}$ is of full rank M and the vectors $\boldsymbol{e}_i$ are the eigenvectors of the correlation matrix $\boldsymbol{P}$.

This more general correlation structure has the benefit that we can define the innovations in the forward rate and volatility by

$$dz_t^k \equiv \sum_{j=1}^{M} e_{kj} dy_j, \text{ for } k = 1, ..., N$$

$$dw_t^k \equiv \sum_{j=1}^{M} e_{kj} dy_j, \text{ for } k = N+1, ..., 2N$$

where

$$\mathbb{E}\left[dz_t^j dz_t^k\right] = \rho_{jk} dt \tag{25}$$

$$\mathbb{E}\left[dw_t^j dw_t^k\right] = r_{jk} dt \tag{26}$$

$$\mathbb{E}\left[dz_t^j dw_t^k\right] = R_{jk} dt \tag{27}$$

We can therefore alternatively write the joint dynamics of the forward rates and their volatilities as

$$df_t^i = \mu_i dt + \left(f_t^i\right)^{\beta_i} s_t^i dz_t^i, \quad i = 1, ..., N \tag{28}$$

$$ds_t^i = g(t, T_i) dk_t^i \tag{29}$$

$$\frac{dk_t^i}{k_t^i} = \mu_t^{k_i} dt + h_t^i dw_t^i, \quad i = 1, ..., N \tag{30}$$

Note that all the eigenvalues of the overall matrix $\boldsymbol{P}$ must be positive for $\boldsymbol{P}$ to be a proper correlation matrix. This poses constraint on $\boldsymbol{R}$ (if the sub-matrix $\boldsymbol{R}$ is exogenously assigned). These and related issues are dealt with below.

6 Observations on the Correlation Structure

Specifying an input correlation structure for the LMM-SABR model is more important (and challenging) than for the determinstic-volatility LMM. This is because in order to recover the SABR caplet prices with our dynamic model we will have to impose requirements on the diagonal elements of the sub-matrix $\boldsymbol{R}$, R_{ii}. More precisely, R_{ii} should be exactly equal to the SABR forward-rate/volatility correlation for the ith forward rate, ρ^i_{SABR} :

$$R_{ii} = \rho^i_{SABR} \tag{31}$$

We shall show formally that this is the case in the chapter devoted to the calibration of the model, but it should already be plausible at this stage.

As the N quantities ρ^i_{SABR} are exogenously defined from the market fits, this imposes a serious constraint on the overall matrix $\boldsymbol{P}$ if we want to make sure that it remains a valid correlation matrix (ie, positive definite.) We will show later how this can be accomplished, but we wanted to point out this non-trivial feature at this early stage in the presentation. Specifying the correlation matrix in the traditional LIBOR market model has always been the ‘poor relation’ in the calibration procedure. This is emphatically no longer the case in the extended LMM-SABR model.

Note also that $\boldsymbol{R}$ is not a symmetric sub-matrix (but $\boldsymbol{\rho}$ and $\boldsymbol{r}$ are). Also, speaking of a 'one-volatility factor' model implies that all the elements of $\boldsymbol{r}$ are equal to 1. This choice can pose heavy constraints on the positive definiteness of the matrix $\boldsymbol{P}$ for a sub-matrix $\boldsymbol{R}$ whose diagonal elements have been exogenously assigned with reference to the quantities ρ^i_{SABR}. For instance, suppose that two forward rates are strongly correlated with each other. If we impose that their volatilities should be perfectly correlated with each other, then it would be very difficult to have a correlation between the first forward rate and its volatility very different from the correlation between the second forward rate and it own volatility.

Finally, a comparison between the more symmetric and ‘elegant’ specification of the correlation matrix afforded by Equations (16) to (24) and the simpler specification presented in Equations (1) to (7) is in order.

Treating the forward rates and their volatilities on the same footing as M stochastic variables linked by a complex correlation matrix allows a more compact notation, makes all the M Brownian increments independent and lends itself well to the calibration strategies presented later on. **It has, however, one potential drawback**. If fewer factors than variables are used, it is not easy to specify how the chosen number of factors should be allocated among the factors. For instance, we may want to use six factors, and to ‘use’ four for the forward rates and only two for the volatilities (perhaps on the grounds that the volatilities should be highly correlated anyhow).

However, once we reduce the number of factors, there is no perfect way to tell the model how to carry out this allocation of its resources. The ‘uglier’ formulation of Equations (1) to (7), on the other hand, naturally allows for any partition of the factors that we may like. The price to pay is that the Brownian draws are no longer independent, and the calibration to the correlation matrix becomes more cumbersome. The problem *can* be fixed and we will tackle this topic in Chapter **??**.

7 The Volatility Structure

Let the instantaneous volatilities of the forward rates, $\sigma_i(t, T_i)$, be given by the product of a function, say, $g()$, of the residual time to maturity, $T_i - t$, of the associated forward rate times and forward-rate specific function k_t^i:

$$\sigma_t^i = k_t^i g(T_i - t) \tag{32}$$

For concreteness, the choice for $g(T_i - t)$ that we shall use in the following is the one introduced in Chapter **??**:

$$g(\tau_i) = (a + b\tau_i) \exp(-c\tau_i) + d \tag{33}$$

with $\tau_i = T_i - t$. The quantities k^i follow the process:

$$\frac{dk_t^i}{k_t^i} = \mu_t^{k_i} dt + h_t^T \sum_{j=1}^{M} e_{ij} dy_j^t \tag{34}$$

The function h_t^T has the form

$$h_t^T = h(T - t) \tag{35}$$

and is parametrized by a set of parameters $[l, m, n, ...]$. A possible parametrization is:

$$h_{\tau_i} = \xi_i \left[(\alpha + \beta \tau_i) \exp\left(-\gamma \tau_i\right) + \delta \right] \tag{36}$$

With these choices

$$k_s^i = k_0^i \exp\left[\int_0^s -\frac{1}{2} \xi_i^2 h^2 (T_i - u) du + \xi_i h(T_i - u) \sum_{j=1}^{M} e_{ij} dy_j(u) \right] \tag{37}$$

We have chosen the functions introduced to describe the volatility and the volatility of volatility to be time homogeneous. The terminology is sloppy (and we therefore refine it in the next section), but the concept is essential, at least in conditions of market self-similarity – roughly speaking, in ‘normal’ market conditions. Our preference for time homogeneity translates our belief that, with the same proviso about normality of the market, the only relevant time variable is linked to the residual time to expiry of the various forward rates in the problem. With the important exception of the conditions discussed in the section after the following one, introducing an explicit time or expiry dependence is a ‘fudge’ to be avoided as much as possible. Time-dependent reversion levels, reversion speeds, displacement coefficients or volatilities, etc, make for good fitting but poor physics.

8 The Volatility Structure in Periods of Market Stress

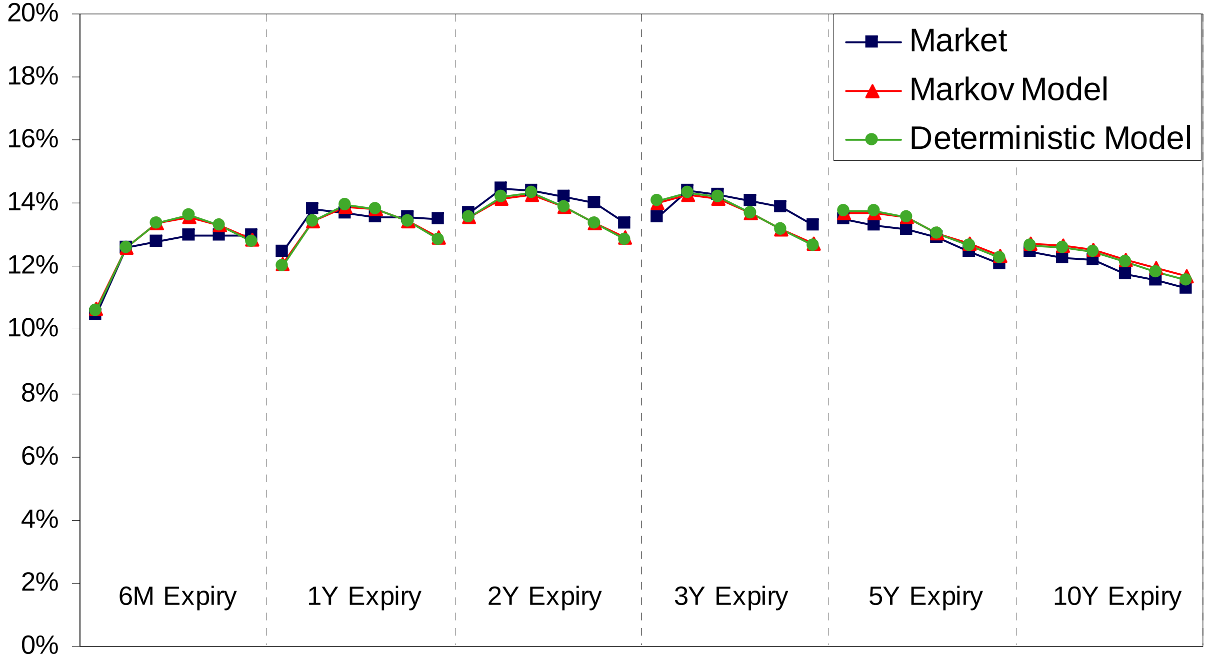
The SABR and LMM-SABR models assume that one (stochastic) volatility regime prevails at all times. In reality a large body of work suggests that markets move, often rapidly, among different states, commonly described as 'normal' and 'excited'. For studies about regime switches in the interest-rate domain which is of greatest interest to us see, eg, Rebonato (2006), White and Rebonato (2007), Rebonato and Gaspari (2006), Rebonato and Chen (2008). (Incidentally, the work by Rebonato and Chen (2008) suggests that *three* states – quiet, normal and excited – may better describe the dynamics of interest rates. We do not pursue this avenue here.)

This has great relevance for the simultaneous pricing of options of different expiry. To see why this is the case, consider the situation where we are in an excited state. The market ‘knows’ this, but also knows that in a matter of weeks or, at most, months, trading conditions will revert to normal. If this is the case, the market will take this information into account, by pricing short-dated option with a high volatility, medium-dated options with a volatility somewhat elevated with respect to normal times, and long-dated options with a volatility barely unchanged from their typical long-dated values – the more so, the longer-dated the option.

This is when time-homogeneous pricing models, that by and large we like, get into trouble. By construction they assume that calendar time is irrelevant, and that the only variable that matters to determine volatilities, correlations and volatilities of volatilities is the residual time to maturity of the forward rate(s). For time-homogeneous models a one-year option seen as of today will pretty much ‘look’ and behave like a one-year option in five or ten years’ time. But if today we are in an excited state, this means that a time-homogeneous model fitted to short-dated options will propagate the current state of excitation *ad infinitum*. As a consequence the one-year option in ten years’ time will be priced as if the current state of turmoil will still be present, say, ten years from now.

What in normal times is a virtue, in excited periods therefore becomes a serious shortcoming. This 'curse of time homogeneity' is displayed in the figures below, that clearly show that time-homogeneous models become, during excited periods, like a short blanket: either long-dated options are well priced, in which case short-dated ones are underpriced, or *vice versa*. There is no way to cover both the head and the feet of the swaption matrix (or, for other asset classes, of the smile surface) with a time-homogeneous model.

Caption: Fit to market data (US$) on 21-Jul-98 in normal market conditions using a deterministic-volatility time-homogeneous implementation of the LMM. It is apparent that a small number of parameters (five, in this particular case) recover well the at-the-money volatility swaption matrix in its entirety. See White and Rebonato (2008) for further details.



20%
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
Market
Markov Model
Deterministic Model
6M Expiry
1Y Expiry
2Y Expiry
3Y Expiry
5Y Expiry
10Y Expiry

Figure 1:

Caption: Fit to market data (US$) as above carried out four months later. The market is now excited, and the volatility at the short end (6 month into 1 year) is ■22% while the long end (10 years into 1 year) is 10%. The deterministic time-homogeneous volatility model attempts to fit this market by having an instantaneous volatility of the forward rates that is much higher close to maturity than far from it. However this results in significant mispricing of options in common expiry blocks. One could, of course, improve the fits to the short-dated options, but this would systematically misprice the longer-dated ones, and vice versa. See White and Rebonato (2008) for further details.

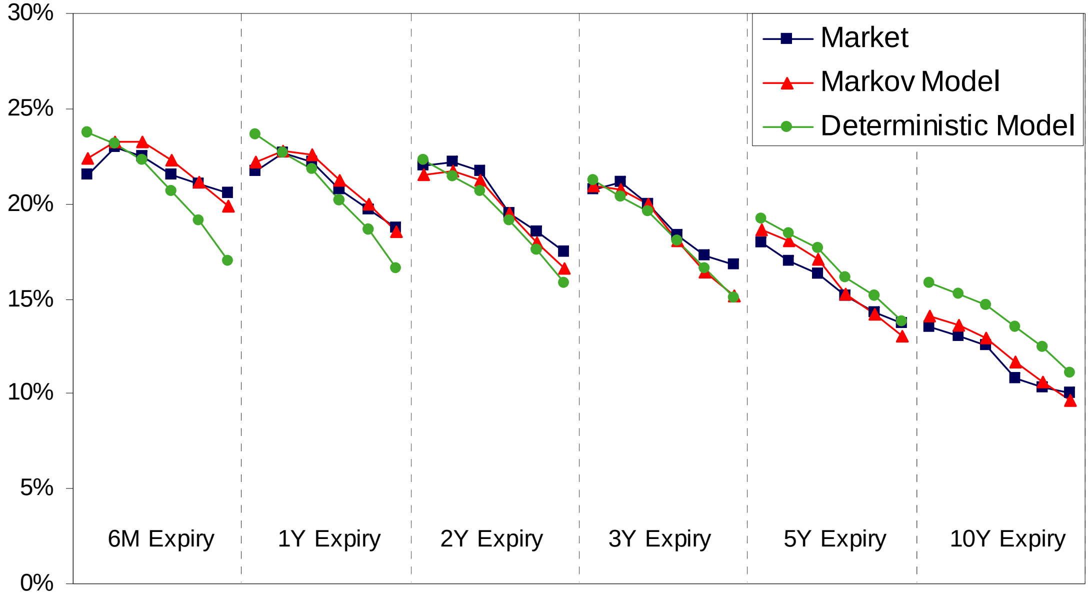

30%
25%
20%
15%
10%
5%
0%
Market
Markov Model
Deterministic Model
6M Expiry
1Y Expiry
2Y Expiry
3Y Expiry
5Y Expiry
10Y Expiry

Figure 2:

Rebonato and White have shown that an effective way to fix this problem in a financially justifiable manner is to posit a two-state Markov chain process for the volatility. This means that at each point in time the volatility can be either in a normal or in an excited state. Given that the volatility is in a given state, it can either migrate to the other state with a transition probability p, or stay in the same state with probability $(1 - p)$. The normal and excited volatility functions can then be chosen to have the humped or exponentially decaying shapes discussed above. Rebonato and Kainth (2004) and White and Rebonato (2008) have shown that even such a conceptually simple model can give a very satisfactory and parsimonious description of the observed prices not only in normal market conditions, but also during periods of market excitation.

The problem is that these Markov-chain models may well be conceptually intuitive, but they are far from simple in terms of getting analytic or numerical approximations. With a lot of acrobatics one can obtain accurate approximations for European swaptions (see White and Rebonato (2008)), making calibration to the swaption matrix feasible, if not simple. When one moves to pricing and, above all, hedging complex options, for which Monte Carlo techniques are needed, the numerical problems become too heavy to deal with for realistic pricing and hedging applications*.

What one needs is a deterministic-volatility approximation to the true two-state Markov chain problem capable of producing *in nuce* the gist of the 'true' results. It must do so without peeking ahead (ie, without knowing when the

*Here is a taste of the problems: suppose, for instance, that I perturb by a small amount a small transition probability. How many Monte Carlo paths do I have to run in order to be able to detect a stable difference with repsect to the underperturbed case?

excited state will occur, or when the market will revert to the normal state). The only piece of information it can avail itself from the 'true' Markov model is the probability of the market being in an excited or normal state today. We show later on the effectiveness of the hedges that this reduced-form approach suggests. In this section we focus on the pricing implications of this reduced-form approach.

Let's assume for the moment that we believe that the probability today of being in an excited state is p. We would recognize this, either using a formal Bayesian estimation technique (see, eg, Rabiner (1989)), or, by fitting the 'proper' two-state Markov chain model described in White and Rebonato (2008). A good sanity check of our estimation of today's state as excited is that the best fit of the LMM-SABR model *to short-expiry options* should yield a monotonically decaying volatility function $g(\cdot)$ (and perhaps $h(\cdot)$ as well). As with the full Markov model we assume that there are two time-homogeneous volatility curves, *normal* and *excited* given by:

$$\sigma_{n,x}(t,T) = \left[a_{n,x} + b_{n,x}(T-t)\right] e^{-c_{n,x}(T-t)} + d_{n,x} \tag{38}$$

To approximate starting in an excited state and decaying to an (absorbing) normal state, we specify the Markov-approximation deterministic volatility as

$$\sigma(t,T;a_n,b_n,c_n,d_n,a_x,b_x,c_x,d_x,\lambda) =$$

$$(1 - pe^{-\lambda t})\sigma(t, T; a_n, b_n, c_n, d_n) + pe^{-\lambda t}\sigma(t, T; a_x, b_x, c_x, d_x) \qquad (39)$$

Note that function (39) is no longer time homogeneous, because of the presence of the unashamedly time-dependent term $e^{-\lambda t}$. This terms does bring in a dependence on calendar time (something we generally argue against), but not as a fitting parameter with no financial justification. Yes, our model says that the future does not look like the present, but, in a situation of transitory market excitation, this is a virtue, not a sin.

The parameter λ is of course related to how long we believe the volatility will remain in the excited state, given that it is in the excited state today. Since we are presenting a 'reduced-form' *Markov* chain model, the permanence in the excited state does not depend on how long the system has already been excited. This may or may not be a shortcoming of the approach, depending on one's views about the market's informational efficiency. It is also clear that the parameter λ is in principle linked to the transition matrix of the 'true' two-state Markov chain. In practice it can be used (within reason!) as a fitting parameter, as discussed below.

As for the parameter p, **we will show that it is natural to give it the the interpretation of probability of being today in the excited state**, but we can also simply regard Equation (39) as specifying a more complex, but still perfectly deterministic, form for the volatility function $g(\cdot)$. Figure $\{1\}$ shows what this function may look like for various values of p and $t = 0$. If we pursue the probabilistic interpretation of p, for $p = 0$ we know for certain that we are in a normal state today. With this reduced-form model for $p = 0$ 'nothing happens' and we are back to the LMM-SABR treatment that we have presented so far. **If $p = 1$ we know for certain that we are in an excited state today ($t = 0$) and we price very short-dated options almost exclusively with an excited volatility.**

More interesting is what happens as calendar time increases. See, for instance, Figure {1} for the behaviour of function (39) when $t = \mathbf{1}$. We see that **as time goes by the excited state begins to disappear**, and that **a short-dated option priced in one year's time experiences a different (less excited) volatility than a short-dated option priced today.** How much less excited will depend on the value of λ. This is just what we wanted to achieve.

Before we go any further we should stress that *we should use a function like Equation (39) whenever we know with reasonable certainty that we are in an excited state*. We should do so irrespective of whether we want to follow the 'survival hedging' suggested later on or not. Using function (39) won't buy us very much in normal times (when p is close to zero), but will make a world of difference when p is close to 1.

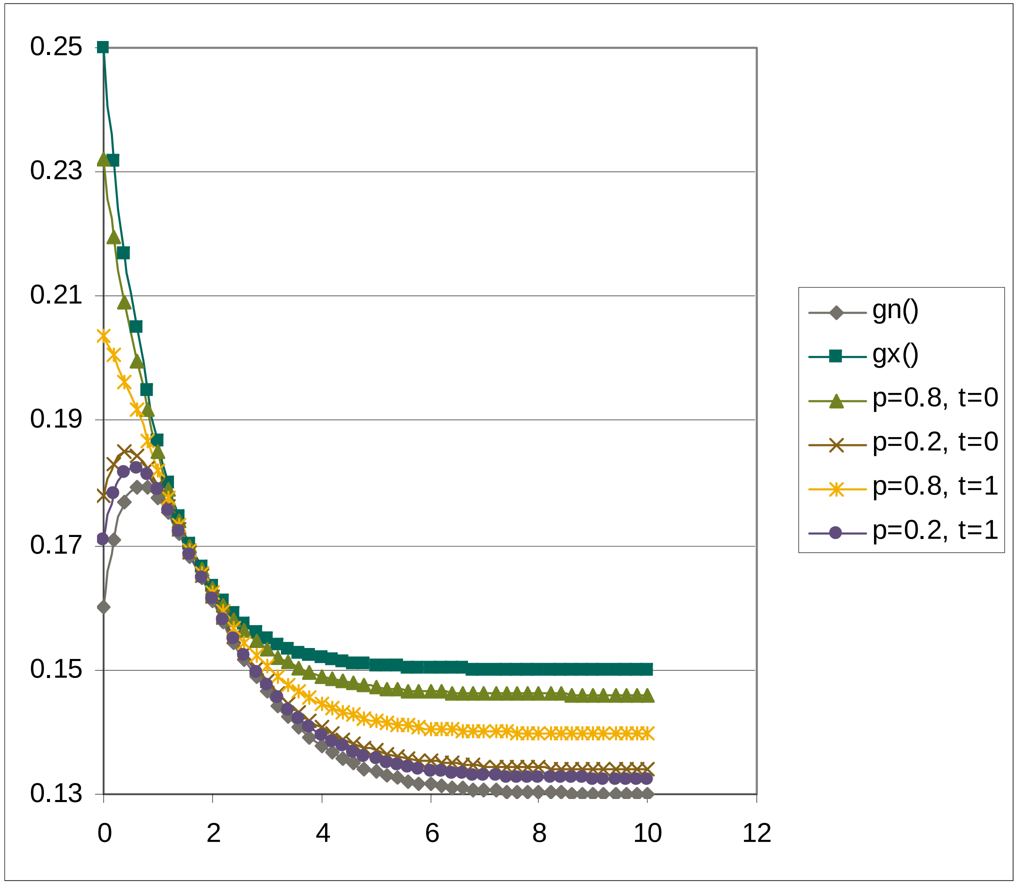

0.25
0.23
0.21
0.19
0.17
0.15
0.13
0
2
4
6
8
10
12
gn()
gx()
p=0.8, t=0
p=0.2, t=0
p=0.8, t=1
p=0.2, t=1

Figure 3:

9 Calculating the No-Arbitrage Drifts

9.1 Preliminaries

We begin by giving some definitions and stating what is for our purposes the fundamental theorem of no-arbitrage. As always in this book, we make no attempt at mathematical rigour and always assume that all technical conditions are satisfied.

Definition 1 *Any traded asset with strictly positive value in every state of the world and that pays no dividends can be used as a numeraire. We shall denote it by* N_t*.*

Definition 2 *For any tradeable assets S_t, the ratio $A_t = \frac{S_t}{N_t}$ is called a relative price. It expresses the value of S in units of N.*

Definition 3 *We denote by $\mathcal{F}_t$ the natural filtration generated by the process S_t.*

With these definitions we can introduce one of the fundamental theorems of no-arbitrage.

Theorem 4 *For any tradeable asset (or portfolio of tradeable assets), S_t, and for any numeraire, N_t, necessary and sufficient condition for the absence of arbitrage is that there should exist a risk-neutral measure, $\mathbb{N}$, associated with the numeraire N_t, such that*

$$\mathbb{E}^{\mathbb{N}}\left[\frac{S_t}{N_t}|\mathcal{F}_s\right] = \frac{S_s}{N_s} \text{ for } t > s \tag{40}$$

In other words, to avoid arbitrage we must find a measure, $\mathbb{N}$, induced by our choice or numeraire, such that the relative price, $A_t = \frac{S_t}{N_t}$, is an $\mathbb{N}$-martingale.

We will not prove this theorem. The sufficient part of the theorem (ie, if the risk-neutral measure exists then there is no arbitrage) is easy to prove. Going the other way (ie, from absence of arbitrage to the existence of a pricing measure) is harder work, and the standard proof is financially not very illuminating. Working in discrete time and with a finite number of assets as usual makes life easier. The typical line of attack makes use of the separating hyperplane theorem and can be found in Duffie (1992), Harrison and Kreps (1979) or Dybvig (1987). Etheridge (2002) provides a step-by-step proof along these lines. Her treatment is about as simple and illuminating as it gets. A more intuitive sketch of the same proof (that does not explicitly mention the separating hyperplane theorem, but makes use of the same intuition) is given by Lengwiler (2004). Finally, Ross (2004) provides a different outline of the same proof using a somewhat less general, but financially more transparent and 'constructive', condition, ie, the maximization of utility of a representative agent. For the reader who wants to understand the standard proof, I recommend to start from Lengwiler (2004) to capture the basic intuition, then to move to Etheridge (2002) for a step-by-step approach and finally to progress to Harrison and Kreps (1979) in order to cross and dot, respectively, all the mathematical ts and is.

In addition to this financial theorem (the ‘physics’ part of the problem), we need some stochastic calculus results, namely, the Radon-Nikodym derivative, the martingale representation theorem and the Cameron-Girsanov theorem. Taken together, these three tools will allow to use the ‘change-of-numeraire technique’, ie, to find in a systematic manner the no-arbitrage drifts under different choices of numeraires. We rely heavily on the approach by Baxter and Rennie (1996) in the following.

Starting with the **martingale representation theorem**, we start from the usual filtered probability space $\{\Omega, \mathbb{P}, \mathcal{F}\}$, and we let

- X be a strictly positive martingale with respect to the filtration generated by a $\mathbb{P}$-Wiener process
- $z(t)$ be a standard Wiener process under $\mathbb{P}$, and
- $q(t)$ a real-valued (not necessarily deterministic!) previsible process, integrable with respect to $z(t)$.

Then, if some technical conditions are satisfied, the following theorem holds:

Theorem 5 *Given X, q and $z(t)$ as above, it is always possible to represent the martingale $X(t)$ in the form*

$$dX_t = X_t q(t) dz(t) \tag{41}$$

Two points are worth making. First, the requirement that X should be a martingale with respect to the filtration generated by a $\mathbb{P}$-Wiener process is essential to the proof, and reflects the fact that the Wiener process should be the only source of randomness in X. Second, despite the notation $(q(t))$, we are *not* saying that the distribution of X_t is log-normal. The 'volatility' $q(t)$ can be relatively horrible, and can certainly depend on X_t itself, and hence be stochastic.

We are now ready to move to the **Cameron-Girsanov's theorem**. We place ourselves again in the usual filtered probability space $\{\Omega, \mathbb{P}, \mathcal{F}\}$, and we let

- $z(t)$ be a standard Brownian motion under a measure $\mathbb{P}$,
- $\mathcal{F}_t$ be the associated Brownian filtration,
- q[†] be a $\mathcal{F}_t$-adapted process which satisfies the Novikov regularity condition $\mathbb{E}^{\mathbb{P}}\left[\exp\left[\int_0^t q(u)^2 du\right]\right] < \infty$, and

[†]With a slight abuse of notation we have used the symbol q both for the 'volatility' of the martingale representation theorem and for process that transforms the $\mathbb{P}$-Brownian motion into z_q. As we shall see, there is a good reason for this choice.

- $z_q(t)$ be defined by

$$z_q(t) = z(t) + qt \tag{42}$$

If these conditions hold, then (see, eg, Mikosh (1998)) the Cameron-Girsanov's theorem actually makes three distinct, but related, statements.

Theorem 16 *The process* $\zeta(t) = \exp[-qz(t) - \frac{1}{2}q^2 t]$ *is a* $\mathbb{P}$*-martingale.*

2. *The process* $\zeta(t)$ *can be used to define a new probability measure* $\mathbb{Q}$*, equivalent to* $\mathbb{P}$*, via the relationship* $\mathbb{Q}(A) = \int_A \zeta_t(\omega) d\mathbb{P}(\omega)$*.*

3. *Under* $\mathbb{Q}$*, the process* $z_q(t)$ *is a standard Brownian motion.*

See, eg, Neftci (1996, Chapter 14, page 291) and *passim* for a nice sketch of the proof of the first part of the theorem, ie, of the fact that $\zeta(t)$ is a $\mathbb{P}$-martingale.

To get some intuition about the Cameron-Girsanov theorem, consider the expression $\mathbb{Q}(A) = \int_A \zeta(\omega) d\mathbb{P}(\omega)$ in the simpler discrete case, where A is now a sub-set of the sample space Ω containing a finite number of elementary events. To make matters even simpler, let's also consider the single-horizon case, when all the uncertainty will be revealed at a single time, T. In this case the expression $\mathbb{Q}(A) = \int_A \zeta_t(\omega) d\mathbb{P}(\omega)$ becomes

$$\mathbb{Q}(A) = \sum_{\omega_i \in A} \zeta(\omega_i^T) \mathbb{P}(\omega_i^T) \tag{43}$$

The transformation $\zeta(\omega_i^T) = \exp[-qz(\omega_i^T) - \frac{1}{2}q^2 T]$ changes the original probability of each event $\mathbb{P}(\omega_i^T)$ into a new probability $\mathbb{Q}(\omega_i^T)$. Under this new measure $\mathbb{Q}$ the probability of event A (that will be revealed at time T) is given by

$$\mathbb{Q}(A) = \sum_{\omega_i \in A} \mathbb{Q}(\omega_i^T)$$

In this expression $\zeta(\omega_i^T)$ is the realization of the $\mathbb{P}$-Brownian motion when event ω_i^T is realized at time T. The transformation is such that every 'new' probability is greater[‡] than zero and smaller than or equal to one, and the sum, $\sum_{\omega_i \in A}$, over any $A \subseteq \Omega$ is strictly positive and smaller than or equal to one. In other terms, in the discrete, single-step case the Cameron-Girsanov theorem tells us that, as long as the transformation is carried out using the quantity $\zeta_T(\omega_i^T) = \exp[-qz(\omega_i^T) - \frac{1}{2}q^2 T]$ with the properties above, we can rest assured that that the transformation changes a good probability measure into another *bona fide* probability measure.

[‡] $\zeta(\omega_i^T)$ must be strictly positive because we are delaing with *equivalent* probability measures.

After the stretching of the probability space carried out by $\zeta(T)$, what was in the original measure $\mathbb{P}$ a good standard Brownian motion will have been 'disrupted' in such a way that it is no longer a Brownian motion under $\mathbb{Q}$. But the Cameron-Girsanov theorem tells us that it doesn't take much to recreate a standard Brownian motion under $\mathbb{Q}$: just add to the original $\mathbb{P}$-Brownian motion a 'drift' term qt, and we are dealing again with a standard Brownian motion under $\mathbb{Q}$.

Clearly, for an 'atomic' set A containing the single discrete event ω ($A = \omega$), one can write

$$\frac{\mathbb{Q}(\omega)}{\mathbb{P}(\omega)} = \zeta_T(\omega) \tag{44}$$

If instead of dealing with discrete probabilities we are using probability densities, Equation (44) becomes

$$\frac{\mathbb{Q}(\omega)}{\mathbb{P}(\omega)} \Longrightarrow \frac{d\mathbb{Q}(\omega)}{d\mathbb{P}(\omega)} \tag{45}$$

where, on the RHS, we now interpret the integration region A to have shrunk to encompass an infinitesimal neighbourhood of ω.

We have considered so far how probability measures (and Brownian motion defined over them) are transformed by the quantity $\zeta_T(\omega)$. But what about expectations? More precisely, what about expectation of quantities that are fully known if the realization of the Brownian motion is known?

To tackle this question, let's consider again the single-horizon problem and consider the unconditional expectation of a Brownian-motion-related random variable. More precisely, consider the payoff, X_T, of a security or derivative contract. Each realization of X_T (discrete or continuous as it may be), constitutes a random event whose value will be revealed at time T. The only really important thing about this random variable is that its value at time T must be perfectly knowable if one knows the realization at time T of the $\mathbb{P}$-Brownian motion, $z(\omega_i^T)$.

Now, to each realization of X_T we can assign a probability $\mathbb{P}(X_T)$ under the measure $\mathbb{P}$. Thanks to the discussion above, we can write under $\mathbb{Q}$ the expression for the probability, $\mathbb{Q}(A)$, (generated by the process $\zeta(T) = \exp[-qz(T) - \frac{1}{2}q^2T]$) of some set A of realizations of X_T (say, the positive realizations of X_T), as

$$\mathbb{Q}(A) = \int_A \zeta(X_T) d\mathbb{P}(X_T) \tag{46}$$

From this expression we can see that the expectation will be given by

$$E^{\mathbb{Q}}\left[X_T|\mathcal{F}_0\right] = E^{\mathbb{P}}\left[\frac{d\mathbb{Q}_T}{d\mathbb{P}_T}X_T|\mathcal{F}_0\right] \tag{47}$$

where the filtration has been denoted by $\mathcal{F}_0$ because we are dealing with a single-step problem. So, in this simplest of cases, the quantity $\frac{d\mathbb{Q}_T}{d\mathbb{P}_T}$ tells us by how much we have to change the measure $\mathbb{Q}$ so that the (unconditional) expectation taken at time 0 of X_T under $\mathbb{P}$ gives us the same value.

If we are looking at a single horizon (T in this case) the Radon-Nikodym derivative $\zeta(X_T)$ defines a random variable. This would be all we need if we were dealing with a European option. But what about a derivative whose value depends on the realization of the Brownian motion that drives our uncertainty at times $t_1, t_2, ..., t_n = T$? What we would like to do is to turn the random variable, $\zeta_0(X_T)$, into a process, $\zeta_t(X_T)$, ie, roughly speaking, a random variable indexed by time. Let's try to do so by considering the *conditional* expectation taken at time t of the Radon-Nikodym derivative $\frac{d\mathbb{Q}(\omega_T)}{d\mathbb{P}(\omega_T)}$ out to a generic horizon T, with $0 \leq t \leq T$. (We continue to follow here the treatment of Baxter and Rennie (1996))

The process ζ_t we have just constructed defines a strictly positive martingale:

$$\zeta_t = E^{\mathbb{P}}\left[\frac{d\mathbb{Q}(\omega_T)}{d\mathbb{P}(\omega_T)}|\mathcal{F}_t\right] = E^{\mathbb{P}}[\zeta_T|\mathcal{F}_t] \quad \text{for } T \geqq t \tag{48}$$

So, the expectation taken at time t of the payoff X_T for any time $t \leq T$ before its value is known with certainty is given by:

$$E^{\mathbb{Q}}\left[\zeta_t X_T | \mathcal{F}_t\right] = E^{\mathbb{P}}\left[\zeta_T X_T | \mathcal{F}_t\right] \tag{49}$$

$$E^{\mathbb{Q}}\left[X_T | \mathcal{F}_t\right] = \frac{1}{\zeta_t} E^{\mathbb{P}}\left[\zeta_T X_T | \mathcal{F}_t\right] \tag{50}$$

where we have 'taken out' of the expectation sign quantities (ζ_t) known at time t. Note that the Radon-Nikodym derivative process takes out exactly what we already know up to time t of the driving Brownian process that, with its full path to T, determines the value of X_T. We can check that expression (49) coincides with Equation (47) when $t = 0$:

$$E^{\mathbb{Q}}\left[\zeta_0 X_T | \mathcal{F}_0\right] = E^{\mathbb{Q}}\left[X_T | \mathcal{F}_0\right] = E^{\mathbb{P}}\left[\zeta_T X_T | \mathcal{F}_0\right] \tag{51}$$

where the middle term follows because $\zeta_0 = 1$.

We have got as much as we need from the mathematical part of the Cameron-Girsanov theorem. We need to inject some financial information into the problem to proceed further. The result that we want to obtain is that the Radon-Nikodym derivative is given by the ratio of the two numeraires that identify the equivalent measures $\mathbb{P}$ and $\mathbb{Q}$. To do this, we mirror again the treatment in Baxter and Rennie (1996), and suppose that in our economy we have $N+2$ securities, S_t^1, S_t^2,..., S_t^N, B_t and C_t. Let's assume that both B_t and C_t satisfy the conditions (see above) for being possible numeraires. Either choice would implicitly define a measure, $\mathbb{Q}^B$ or $\mathbb{Q}^C$, under which relative prices are martingales (again, see above): so, under $\mathbb{Q}^B$, $\frac{S_t^1}{B_t}$, $\frac{S_t^2}{B_t}$, ..., $\frac{S_t^N}{B_t}$ and $\frac{C_t}{B_t}$ are martingales; under $\mathbb{Q}^c$, $\frac{S_t^1}{C_t}$, $\frac{S_t^2}{C_t}$, ..., $\frac{S_t^N}{C_t}$ and $\frac{B_t}{C_t}$ are martingales.

Given these choices the price at time t, V_t, of a payoff X_T at time T is equivalently given by:

$$V_t^B = B_t \mathbb{E}^{\mathbb{Q}^B}\left[\frac{X_T}{B_T}|\mathcal{F}_t\right] \tag{52}$$

or by:

$$V_t^C = C_t \mathbb{E}^{\mathbb{Q}^C}\left[\frac{X_T}{C_T}|\mathcal{F}_t\right] \tag{53}$$

We can now ask: how are these two measures related? We can answer this question thanks to the properties of the Radon-Nikodym process presented above. Namely, we can write:

$$\mathbb{E}^{\mathbb{Q}^C}[X_\tau|\mathcal{F}_t] = \frac{1}{\zeta_t}\mathbb{E}^{\mathbb{Q}^B}[\zeta_\tau X_\tau|\mathcal{F}_t] \tag{54}$$

or

$$\zeta_t\mathbb{E}^{\mathbb{Q}^C}[X_\tau|\mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}^B}[\zeta_\tau X_\tau|\mathcal{F}_t] \tag{55}$$

If the process X_t is a $\mathbb{Q}^C$-martingale, we have

$$\mathbb{E}^{\mathbb{Q}^C}[X_\tau|\mathcal{F}_t] = X_t \tag{56}$$

and therefore

$$\zeta_t X_t = \mathbb{E}^{\mathbb{Q}^B}[\zeta_\tau X_\tau|\mathcal{F}_t] \tag{57}$$

ie, under $\mathbb{Q}^B$, it is the process $\zeta_t X_t$ that is a $\mathbb{Q}^B$-martingale. But we also know that, under $\mathbb{Q}^B$, $\frac{S_t^1}{B_t}$, $\frac{S_t^2}{B_t}$, ..., $\frac{S_t^N}{B_t}$ and $\frac{C_t}{B_t}$ are all martingales (the no-arbitrage condition). So, both $\frac{S_t^1}{B_t}$ and $\zeta_t \frac{S_t^1}{C_t}$ are $\mathbb{Q}^B$-martingales – and the same applies, of course, to the other securities. *But, if this is the case, we have a very simple and important result: the Radon-Nikodym process is just given by the ratio of the numeraires*:

$$\zeta_t = \frac{C_t}{B_t} \tag{58}$$

As we have seen ζ_t is a strictly positive martingale. The martingale representation theorem therefore applies and we have:

$$d\zeta_t = \zeta_t q_t dz_t \tag{59}$$

and

$$\zeta(t) = \exp\left[\int_0^t q_s dz_s - \frac{1}{2} q_s^2 ds\right] \tag{60}$$

Equation (60) immediately tells us that the process starts at $\zeta(0) = 1$. The only thing we still need in order to pin down uniquely the measure transformation is the 'volatility' process q_t. In order to obtain this we need to specify a process for C_t and B_t, use Ito's lemma for the ratio $\zeta_t = \frac{C_t}{B_t}$, look at the coefficient of the stochastic part, and we are done. This is exactly how we will apply below the change-of-numeraire technique to derive the drifts we need.

Given the initial condition ($\zeta(0) = 1$) and the 'volatility' of ζ_t, we know everything there is to know about the ratio ζ_t. Therefore, if we know the drift under one measure, we can immediately find the drift under under measure using Equation (55).

We apply this blueprint in the following.

10 Standard LIBOR and LIBOR in Arrears

For a forward rate f_t^i that expires at time T_i and pays at time $T_i+\tau = T_{i+1}$, the discount bond, $P(t, T_{i+1})$ is a possible numeraire. The associated risk-neutral measure, $\mathbb{T}_i$, is known as the terminal measure. The numeraire is called the 'natural numeraire' (or the 'natural payoff' - see Doust (1995)).

The forward rate spanning T_i to T_{i+1} can be written in terms of discount bonds as:

$$f_t^i \equiv f(t; T_i, T_{i+1}) \equiv \frac{P(t, T_i) - P(t, T_{i+1})}{\tau P(t, T_{i+1})} \tag{61}$$

Under the terminal measure we have

$$\mathbb{E}^{\mathbb{T}_i}\left[f(t; T_i, T_{i+1})|\mathcal{F}_s\right] = \frac{1}{\tau}\mathbb{E}^{\mathbb{T}_i}\left[\frac{P(t, T_i) - P(t, T_{i+1})}{P(t, T_{i+1})}|\mathcal{F}_s\right] \tag{62}$$

If we consider the quantity $P(t, T_i) - P(t, T_{i+1})$ as a (portfolio of) assets, we can write $P(t, T_i) - P(t, T_{i+1}) = S_t^i$. We can then apply the fundamental theorem above, and impose that $\frac{S_t^i}{N_t^i}$, which is the ratio of tradable assets, must be a martingale. But since

$$\frac{S_t^i}{N_t^i} = \frac{P(t, T_i) - P(t, T_{i+1})}{P(t, T_{i+1})} = \tau f_t^i \tag{63}$$

it follows that the forward rate f_t^i itself, under its own terminal measure, must be a martingale

$$\mathbb{E}^{\mathbb{T}_i}\left[f_t^i | \mathcal{F}_s \right] = f_s^i \tag{64}$$

This is easy enough for the natural numeraire. In general, however, we want to calculate the no-arbitrage drifts that arise when the numeraire is *not* the natural numeraire. We present in this section the argument in detail in the LIBOR-in-arrears case, ie, for the special case when the forward rate f_t^i resets and pays *at the same time*, T_i. For this problem, using the Radon-Nikodym weaponry to obtain the drift adjustment is a bit like using a cannon to swat a fly; still, we will pursue not only the simpler, *ad hoc* treatment, but also the more general treatment to serve us as a guide in the more complex settings we are interested in.

In the LIBOR-in-arrears case it is natural to discount the payoffs using the discount bond $P(t, T_i)$. This, however, is not the natural numeraire for the forward rate f_t^i. For the LIBOR-in-arrears case we therefore need the dynamics, and, in particular, the no-arbitrage drifts, of f_t^i under the measure $\mathbb{T}_{i-1}$ induced by the numeraire $P(t, T_i)$. As we said, to obtain these we proceed along two distinct routes: the first we call the 'brute-force' approach, and the second the 'change-of-numeraire' route.

10.1 LIBOR in Arrears: Brute-Force Approach

We first note that, by itself, f_t^i is not a traded asset. However, if multiplied by its own natural numeraire, $P(t, T_{i+1})$, it becomes a portfolio of traded assets:

$$\tau_i f_t^i P(t, T_{i+1}) = P(t, T_i) - P(t, T_{i+1})$$

To this portfolio of traded assets we can apply the fundamental theorem of asset pricing, this time under the measure, $\mathbb{T}_{i-1}$, associated with numeraire $P(t, T_i)$[§]. To apply the theorem we we write the fundamental theorem as

$$\mathbb{E}_0^{\mathbb{T}_{i-1}}\left[\frac{\tau_i f_t^i P(t, T_{i+1})}{P(t, T_i)}\right] = \mathbb{E}_0^{\mathbb{T}_{i-1}}\left[\frac{P(t, T_i) - P(t, T_{i+1})}{P(t, T_i)}\right] = \frac{\tau f_0^i P(0, T_{i+1})}{P(0, T_i)} \tag{65}$$

[§]Why do we call $\mathbb{T}_{i-1}$ the measure associated with numeraire $P(t, T_i)$? Because it is the measure under which the forward rate f_{i-1} is driftless!

and we define

$$X_t = \frac{P(t, T_{i+1})}{P(t, T_i)} = \frac{1}{1 + \tau f_t^i} \tag{66}$$

Note now that both X_t and $f_t^i X_t$ are relative prices (because they are assets, or portfolio of assets, divided by our chosen numeraire). By the fundamental theorem of asset pricing they must therefore be martingales under $\mathbb{T}_{i-1}$:

$$\mathbb{E}_0^{\mathbb{T}_{i-1}} \left[f_t^i X_t \right] = f_0^i X_0 \tag{67}$$

$$\mathbb{E}_0^{\mathbb{T}_{i-1}} \left[X_t \right] = X_0 \tag{68}$$

Under the dynamics specified by Equations (**??**), we can use Ito's lemma to write down the equations of motion for X_t:

$$dX_t = \frac{\partial X_t}{\partial f_t} df_t^i + \frac{1}{2}\frac{\partial^2 X_t}{\partial f_t^2}\left\langle df_t^i, df_t^i\right\rangle \tag{69}$$

$$\begin{aligned} &= -\tau X_t^2\left(\mu^{i,\mathbb{T}_{i-1}}\left(f_t^i, t\right)dt + \sigma^i\left(f_t^i, t\right)k_t^i dz_t^{i,\mathbb{T}_{i-1}}\right) + \tau^2 X_t^3 \sigma^i\left(f_t^i, t\right)^2\left(k_t^i\right)^2 dt \\ &= \tau X_t^2\left(\tau X_t \sigma^i\left(f_t, t\right)^2\left(k_t^i\right)^2 - \mu^{i,\mathbb{T}_{i-1}}\left(f_t^i, t\right)\right)dt - \tau X_t^2 \sigma^i\left(f_t^i, t\right)k_t^i dz_t^{i,\mathbb{T}_{i-1}} \end{aligned} \tag{70}$$

The drift of X_t is

$$\frac{\mathbb{E}^{\mathbb{T}_{i-1}}\left[dX_t\right]}{dt} = \tau X_t^2\left(\tau X_t \sigma^i\left(f_t^i, t\right)^2\left(k_t^i\right)^2 - \mu^{i,\mathbb{T}_{i-1}}\left(f_t^i, t\right)\right) \tag{71}$$

However since X_t is a martingale under $\mathbb{T}_{i-1}$, this drift must be zero. This gives the drift of f_t^i under measure $\mathbb{T}_{i-1}$:

$$\mu^{i,\mathbb{T}_{i-1}}(f_t^i, t) = \tau \left(k_t^i\right)^2 \frac{\sigma^i \left(f_t^i, t\right)^2}{1 + \tau f_t^i} \tag{72}$$

This is the drift correction to be applied to the forward rate, f_t^i, in order to avoid arbitrage under the simplified 'stochastic volatility' model specified by Equations (**??**). This result closely resembles the drift obtained in the well-known deterministic-volatility case, that can be obtained if we set $k_t^i = 1$ and we absorb it in the function $\sigma^i(f_t^i, t)^2$:

$$\mu^{i,\mathbb{T}_{i-1}}(f_t^i, t) = \frac{\tau \sigma^i \left(t\right)^2}{1 + \tau f_t^i} \tag{73}$$

Despite the formal similarity note that, in our stochastic-volatility result, the factor k_t^i is a full stochastic process.

The same no-arbitrage drift would have been obtained if we had considered the dynamics of $f_t^i X_t$ (which is just as well!). Indeed, this is exactly the line of attack typically used to obtain the no-arbitrage drifts in the deterministic-volatility case. See, eg, Rebonato (2002). However, the approach we have followed does not give us much of a hint as to how we can find the no-arbitrage drift on the volatility, $\eta^i(k_t^i, t)$.To get this drift we must follow a different, more general, line of attack.

10.2 LIBOR in Arrears: The Change-of-Numeraire Approach

Recall the condition

$$\zeta_t X_t = \mathbb{E}^{\mathbb{T}_i}\left[\zeta_\tau X_\tau | \mathcal{F}_t\right] \tag{74}$$

that holds whenever X_t is a martingale under the measure $\mathbb{T}_i$. First we apply this condition to $X_t = f_t^i$ which is a $\mathbb{T}_i$-martingale associated with its natural numeraire $P\left(t, T_{i+1}\right)$. The dynamics of $\zeta_t^i f_t^i$ under the measure we are interested in (the $\mathbb{T}_{i-1}$-measure) are given by:

$$\begin{aligned} d\left(f_t^i \zeta_t^i\right) &= \zeta_t^i df_t^i + f_t^i d\zeta_t^i + df_t^i d\zeta_t^i \\ &= \zeta_t^i \left[\mu^{i,\mathbb{T}_{i-1}}\left(f_t^i, t\right) dt + \sigma^i(f_t^i, t) k_t^i dz_t^{i,\mathbb{T}_{i-1}}\right] + \zeta_t^i f_t^i q_t^i dz_t^{i,\mathbb{T}_{i-1}} + \sigma^i(f_t^i, t) k_t^i \zeta_t^i q_t^i d \\ &= \zeta_t^i \left[\mu^{i,\mathbb{T}_{i-1}}(f_t^i, t) + \sigma^i(f_t^i, t) k_t^i q_t^i\right] dt + \zeta_t^i \left[\sigma^i(f_t^i, t) k_t^i + f_t^i q_t^i\right] dz_t^{i,\mathbb{T}_{i-1}} \end{aligned} \tag{75}$$

where the last term in the second line $\left(\sigma^{i}\left(f_{t}^{i},t\right)k_{t}^{i}\zeta i_{t}q_{t}^{i}dt\right)$ follows because in this particular case we know that there is a perfect functional dependence between the forward rate, f_{t}^{i}, and the Radon-Nikodym derivative, (which is just the ratio of the numeraires):

$$\zeta_{t}^{i}=\frac{P(t,T_{i+1})}{P(t,T_{i})}=\frac{1}{1+\tau f_{t}^{i}} \tag{76}$$

The martingale condition means that we can write:

$$\mu^{i,\mathbb{T}_{i-1}}(f_{t}^{i},t)+\sigma(f_{t}^{i},t)k_{t}^{i}q_{t}^{i}=0\Longrightarrow\mu^{i,\mathbb{T}_{i-1}}(f_{t}^{i},t)=-\sigma(f_{t}^{i},t)k_{t}^{i}q_{t}^{i} \tag{77}$$

We just need the volatility, q_{t}^{i}, of the Radon-Nikodym derivative. This can be obtained from a quick application of Ito's lemma to Equation (76):

$$q_{t}^{i}=-\frac{\tau k_{t}^{i}\sigma^{i}(f_{t}^{i},t)}{1+\tau f_{t}^{i}} \tag{78}$$

and so Equation (77) becomes

$$\mu^{i,\mathbb{T}_{i-1}}(f_t^i,t) = -\sigma^i(f_t^i,t)k_t^i q_t^i = \frac{\left[k_t^i \sigma^i(f_t^i,t)\right]^2 \tau}{1+\tau f_t^i} \tag{79}$$

which is reassuringly identical to the result obtained above using the brute-force approach.

The beauty of this approach is that, following exactly the same blue-print, we can now also obtain at virtually no extra cost the no-arbitrage drift for the volatility. This is how it is done.

10.3 LIBOR in Arrears: The Volatility Drift

We now let $X_t = k_t^i$ be a martingale process under $\mathbb{T}_i$. The dynamics of $k_t^i \zeta_t^i$ under $\mathbb{T}_{i-1}$ are

$$
\begin{aligned}
d(k_t^i \zeta_t^i) &= \zeta_t^i dk_t^i + k_t^i d\zeta_t^i + dk_t^i d\zeta_t^i \\
&= \zeta_t^i \left[\eta(k_t, t)dt + \nu(k_t, t) dw_t^{\mathbb{T}_{i-1}} \right] + k_t \zeta_t q_t dz_t^{\mathbb{T}_{i-1}} + \rho \zeta_t q_t \nu(k_t, t) dt \\
&= \zeta_t \left[\eta(k_t, t) + \rho q_t \nu(k_t, t) \right] dt + k_t \zeta_t q_t dz_t^{i, \mathbb{T}_{i-1}} + \zeta_t \nu(k_t, t) dw_t^{i, \mathbb{T}_{i-1}}
\end{aligned}
\tag{80}
$$

where $dz_t^{i,\mathbb{T}_{i-1}}$ and $dw_t^{i,\mathbb{T}_{i-1}}$ are two standard Brownian motions under $\mathbb{T}_{i-1}$, which shock the forward rate, f_t^i, and the process k_t^i, respectively. But again we know (from the definition of the dynamics) that $k_t^i \zeta_t^i$ is a martingale under

$\mathbb{T}_{i-1}$. It follows that

$$\begin{aligned}\eta^i(k_t^i,t) &= -\rho_i q_t^i \nu^i(k_t^i,t) \\ &= \rho_i \tau k_t^i \frac{\sigma^i(f_t^i,t)\nu^i(k_t^i,t)}{1+\tau f_t^i}\end{aligned} \tag{81}$$

In particular, under the SABR dynamics

$$\begin{aligned} df_t &= \sigma_t \left(f_t\right)^{\beta} dz_t^{\mathbb{T}} \\ d\sigma_t &= \nu \sigma_t dw_t^{\mathbb{T}} \\ \rho dt &= \mathbb{E}\left[dz_t^{\mathbb{T}}, dw_t^{\mathbb{T}}\right] \end{aligned}$$

where df_t and $d\sigma_t$ are driftless under their terminal measure $\mathbb{T}$, the volatility drift under measure $\mathbb{T}_{-1}$ becomes

$$\rho\tau \left(\sigma_t\right)^2 \frac{\left(f_t\right)^{\beta} \nu}{1+\tau f_t} \tag{82}$$

where we have restored the usual SABR notation.

10.4 The Drifts in the General Case of Several Forward Rates

Obtaining the drift corrections for different numeraires is algebraically a bit messier, but conceptually no different. The tool of the Vailliant brackets presented in Rebonato (2002) can help with the algebraic manipulations. So, for forward rates spanning the periods $t_1, t_2, \ldots t_{i,} t_{i+1}, \ldots,$ the general dynamics are

$$\begin{aligned}
df_t^i &= \mu_i(\boldsymbol{f}_t, \boldsymbol{k}_t, t)dt + \sigma_i(\boldsymbol{f}_t, t)k_t^i dz_t^i \\
dk_t^i &= \eta_i(\boldsymbol{f}_t, \boldsymbol{k}_t, t)dt + \nu_i(\boldsymbol{k}_t, t)dw_t^i \\
\left\langle dz_t^i dz_t^j \right\rangle &= \rho_{ij}dt \\
\left\langle dw_t^i dw_t^j \right\rangle &= r_{ij}dt \\
\left\langle dz_t^i dw_t^j \right\rangle &= R_{ij}dt
\end{aligned} \tag{83}$$

where $\boldsymbol{f}_t$ is the vector $\{f_t^1, f_t^2, \ldots\}$ and $\boldsymbol{k}_t = \{k_t^1, k_t^2, \ldots\}$. Using the above procedure it is algebraically tedious, but conceptually straightforward, to show that under the $\mathbb{T}_j$-measure (ie, the measure induced by the numeraire $P(t, T_{j+1})$) the drifts are

$$\mu_i^{\mathbb{T}_j}(\mathbf{f}_t, \mathbf{k}_t, t) = \begin{cases} \sigma_i(\boldsymbol{f}_t, t) k_t^i \sum_{\alpha=j+1}^{i} \frac{\rho_{i\alpha}\sigma_\alpha(\boldsymbol{f}_\alpha, t) k_t^\alpha \tau_\alpha}{1+\tau_\alpha f_t^\alpha} & i > j \\ 0 & i = j \\ -\sigma_i(\boldsymbol{f}_t, t) k_t^i \sum_{\alpha=i+1}^{j} \frac{\rho_{i\alpha}\sigma_\alpha(\boldsymbol{f}_\alpha, t) k_t^\alpha \tau_\alpha}{1+\tau_\alpha f_t^\alpha} & i < j \end{cases} \tag{84}$$

$$\eta_i^{\mathbb{T}_j}(\mathbf{f}_t, \mathbf{k}_t, t) = \begin{cases} \nu_i(\boldsymbol{k}_t, t) \sum_{\alpha=j+1}^{i} \frac{R_{i\alpha}\sigma_\alpha(\boldsymbol{f}_\alpha, t) k_t^\alpha \tau_\alpha}{1+\tau_\alpha f_t^\alpha} & i > j \\ 0 & i = j \\ -\nu_i(\boldsymbol{k}_t, t) \sum_{\alpha=i+1}^{j} \frac{R_{i\alpha}\sigma_\alpha(\boldsymbol{f}_\alpha, t) k_t^\alpha \tau_\alpha}{1+\tau_\alpha f_t^\alpha} & i < j \end{cases} \tag{85}$$

A very important observation is in order. Comparing the no-arbitrage drifts for the forward rates (Equations (84)) and for the volatilities (Equations (85)), *we note that the computationally expensive summation terms have a very similar structure in the two cases*. There are simple and powerful apprimations here.

10.5 Volatility Drifts in the Swap Measure

The results in this sub-section were first presented, to our knowledge, by Henry-Labordere (2007), but without derivation. We show below how to derive them.

For a set of evenly spaced times T_1, $T_{2,} \ldots T_i, T_{i+1}, \ldots T_N$, with spacing $\tau_i \equiv T_{i+1} - T_i$, we introduce the definition of the following quantities:

- $S^{\alpha\beta}(t)$ for the swap rate for a swap starting at time T_α with final payment at T_β

- $A^{\alpha\beta}(t)$ for the annuity of the swap fixed payments

- $\mathbb{Q}^{\alpha\beta}$ for the measure induced by the choice of the swap annuity $A^{\alpha\beta}(t)$ as numeraire (ie, the measure under which the swap rate $S^{\alpha\beta}(t)$ is a martingale).

With these definitions we have

$$S^{\alpha\beta}(t) = \frac{P(t, T_\alpha) - P(t, T_\beta)}{A^{\alpha\beta}(t)} \tag{86}$$

with

$$A^{\alpha\beta}(t) = \sum_{i=\alpha}^{\beta-1} \tau_i P(t, T_{i+1})$$

The swap rate can be written is terms of the forwards as

$$S^{\alpha\beta}(t) = \sum_{i=\alpha}^{\beta-1} \omega_i(t) f_t^i \tag{87}$$

$$\text{where } \omega_i(t) = \frac{\tau_i P(t,T_{i+1})}{A^{\alpha\beta}(t)}$$

So the weights[¶] themselves are stochastic and depend on all the forwards underlying the swap. However in the swap measure, $\mathbb{Q}^{\alpha\beta}$, the weights are all martingales, ie

$$\mathbb{E}^{\mathbb{Q}^{\alpha\beta}}[\omega_i(t)|\mathcal{F}_s] = \omega_i(s)\ \forall i \text{ and } t \geq s \tag{88}$$

This is because, as Equation (87) shows, the weights are relative prices (ratios of asset prices to the chosen numeraire). Of course the swap rate is also a martingale under the swap measure.

¶From the definition of $A_{\alpha\beta}(t)$ it is clear that $\sum_{i=\alpha}^{\beta-1} W_i(t) = 1$ and that $W_i(t) > 0$, i.e. they have the property of positive weights.

We now specify general stochastic volatility dynamics for the forward rates as

$$\begin{aligned}
df_t^i &= \mu_i(\boldsymbol{f}_t, t)dt + \sigma_i(\boldsymbol{f}_t, t)k_t^i dz_t^i \\
dk_t^i &= \eta_i(\boldsymbol{k}_t, t)dt + \nu_i\left(\boldsymbol{k}_t, t\right)d\hat{z}_t^i \\
\left\langle dz_t^i dz_t^j \right\rangle &= \rho_{ij} dt \\
\left\langle d\hat{z}_t^i d\hat{z}_t^j \right\rangle &= r_{ij} dt \\
\left\langle dz_t^i d\hat{z}_t^j \right\rangle &= R_{ij} dt
\end{aligned} \tag{89}$$

where the notation $\mu_i(\boldsymbol{f}_t, t)$ means the drift on the ith forward is some function of (potentially) all the forward rates and time (as similarly for the other symbols). To lighten the notation we drop the function arguments and time dependence (e.g. f_t^i becomes just f^i). Using Equation (87) we can write down the dynamics of $S^{\alpha\beta}$ as

$$
\begin{aligned}
dS^{\alpha\beta} &= \sum_{i=\alpha}^{\beta-\mathbf{1}} \left\{\omega_i df^i + f^i d\omega_i + df^i d\omega_i\right\} \\
&= \sum_{i=\alpha}^{\beta-\mathbf{1}} \left\{\omega_i \mu_i dt + \omega_i \sigma_i k^i dZ^i + f^i d\omega_i + \sigma_i k^i dZ^i d\omega_i\right\} && (90) \\
&= \sum_{i=\alpha}^{\beta-\mathbf{1}} \left\{\omega_i \mu_i dt + \sigma_i k^i dz^i d\omega_i\right\} + \sum_{i=\alpha}^{\beta-\mathbf{1}} \left\{f^i d\omega_i + \omega_i \sigma_i k^i dz^i\right\} && (91)
\end{aligned}
$$

Note that the second term is purely stochastic (ie, it has no term in dt, since ω_i is a martingale). So,with a slight abuse of notation, we have

$$
\mu_i = -\frac{1}{\omega_i} \frac{\sigma_i k^i \left\langle dz^i d\omega_i \right\rangle}{dt} \tag{92}
$$

and we are left with calculating the stochastic part of $d\omega_i$. We rewrite the weights as

$$\omega_i = \left(\sum_{j=\alpha}^{\beta-1} \frac{\tau_j P\left(t, T_{j+1}\right)}{\tau_i P\left(t, T_{i+1}\right)} \right)^{-1}$$

$$= \left(\sum_{j=\alpha}^{\beta-1} X_{ij} \right)^{-1} \quad \text{where } X_{ij} \equiv \frac{\tau_j P\left(t, T_{j+1}\right)}{\tau_i P\left(t, T_{i+1}\right)} \tag{93}$$

Taking partial derivatives with respect to X_{ik}

$$\frac{\partial \omega_i}{\partial X_{ik}} = -\omega_i^2 \tag{94}$$

$$\frac{\partial_i^2 \omega}{\partial X_{ik} \partial X_{ik'}} = -2\omega_i \frac{\partial \omega_i}{\partial X_{ik'}} = 2\omega_i^3 \tag{95}$$

so

$$d\omega_i = -\omega_i^2 \sum_{j=\alpha}^{\beta-1} dX_{ij} + \omega_i^3 \sum_{j=\alpha}^{\beta-1} \sum_{k=\alpha}^{\beta-1} dX_{ij} dX_{ik} \tag{96}$$

Now let us consider the dynamics of X_{ij}.This can then be written in terms of the forward rates

$$X_{ij}(t) \equiv \begin{cases} \frac{\tau_j}{\tau_i} \prod_{k=i+1}^{j} (1+\tau_k f^k)^{-1} \ i < j \\ \frac{\tau_j}{\tau_i} \prod_{k=j+1}^{i} (1+\tau_k f^k) \ i > j \end{cases} \tag{97}$$

Taking partial derivatives with respect to f_t^k we get

$$\frac{\partial X_{ij}}{\partial f^k} = (2\mathbb{I}_{i>j} - 1) \frac{\tau_k X_{ij}}{1+\tau_k f^k} \tag{98}$$

where $\mathbb{I}_{i>j}$ is the indicator function that $i > j$. The dynamics of X_{ij} are

$$\frac{dX_{ij}}{X_{ij}} = (2\mathbb{I}_{i>j} - 1) \sum_{k=\min(i,j)+1}^{\max(i,j)} \frac{\tau_k}{1+\tau_k f^k} df^k + \mathcal{O}(\left\langle df^k, df^k \right\rangle) \tag{99}$$

Finally the dynamics of $d\omega_i$ are

$$d\omega_i = -\omega_i^2 \sum_{j=\alpha}^{\beta-1} \left\{ (2\mathbb{I}_{i>j} - 1) X_{ij} \sum_{k=\min(i,j)+1}^{\max(i,j)} \frac{\tau_k}{1+\tau_k f^k} \sigma_k k^k dz^k \right\} \tag{100}$$

and the drift on the jth forward rate is

$$\mu_i(\mathbf{f},t) = -\frac{1}{\omega_i}\frac{\sigma_i \kappa^i \left\langle dz^i d\omega_i \right\rangle}{dt} \tag{101}$$

$$= \omega_i \sigma_i \kappa^i \sum_{j=\alpha}^{\beta-1} \left\{ (2\mathbb{I}_{i<j} - 1) X_{ij} \sum_{k=\min(i,j)+1}^{\max(i,j)} \frac{\tau_k}{1+\tau_k f^k} \sigma_k k^i \rho_{ik} \right\} \tag{102}$$

$$= \sigma_i \kappa^i \sum_{j=\alpha}^{\beta-1} \left\{ (2\mathbb{I}_{i>j} - 1) \omega_j \sum_{k=\min(i,j)+1}^{\max(i,j)} \frac{\tau_k}{1+\tau_k f^k} \sigma_k k^k \rho_{ik} \right\} \tag{103}$$

Using exactly the same method we used for the volatility drift under a terminal measure, we have

$$
\begin{aligned}
\gamma &= -\frac{\mu_i}{\sigma_i \kappa^i} \\
&= -\sum_{j=\alpha}^{\beta-1} \left\{ (2\mathbb{I}_{i>j} - 1)\omega_j \sum_{k=\min(i,j)+1}^{\max(i,j)} \frac{\tau_k \sigma_k k^k}{1+\tau_k f^k} \rho_{ik} \right\}
\end{aligned} \tag{104}
$$

But we also know that

$$
\eta_i = -R_{ii}\gamma\nu_i \tag{105}
$$

Therefore

$$
\eta_i = R_{ii}\nu_i \sum_{j=\alpha}^{\beta-1} \left\{ (2\mathbb{I}_{i>j} - 1)\omega_j \sum_{k=\min(i,j)+1}^{\max(i,j)} \frac{\tau_k \sigma_k k^k}{1+\tau_k f^k} \rho_{ik} \right\} \tag{106}
$$

This gives us both the forward rate and volatility drifts in the swap measure. We note in passing that these expressions coincide with the equations provided without derivation in Henry-Labordere (2007). R_{ii} is just the market-given, β-dependent SABR correlation between a swap rate and its own volatility. Any choice of the exponent β that makes the term R_{ii} smaller will also make the volatility drift terms smaller, making life computationally easier (and possibly justifying some approximations of these terms).