Is It Possible to Reconcile the Caplet and Swaption Markets? Evidence from the U.S. Dollar Market

Article *in* The Journal of Derivatives · October 2011
DOI: 10.3905/jod.2011.2011.1.015

2 authors:


Riccardo Rebonato
University of Oxford


Andrey Pogudin
King's College London

Is It Possible to Reconcile the Caplet and Swaption Markets? *Evidence from the U.S.-Dollar Market*

Riccardo Rebonato and Andrey Pogudin

Riccardo Rebonato is the head of Front Office Risk and Analytics at the Royal Bank of Scotland in London, UK, and a visiting lecturer at the Oxford Centre for Industrial & Applied Mathematics (OCIAM) at Oxford University in Oxford, UK.
riccardo.rebonato@rbs.com

Andrey Pogudin is a senior quantitative analyst in the Quantitative Strategy Group at the Royal Bank of Scotland in London, UK.
andrey.pogudin@rbs.com

We look at the prices of U.S.-dollar caplets and swaptions and ask whether the two markets (which ultimately share the same set of underliers) are internally consistent, given a flexible and parsimonious class of no-arbitrage pricing models. We find that it is difficult to reconcile the observed market prices of all swaptions with plausible assumptions about correlation and volatilities. We find instead that in all market conditions (normal and excited), the swaption market trades extremely close to the no-arbitrage boundaries implied by the chosen class of no-arbitrage models.

We also identify under what conditions the prices of swaptions can be replicated using early-stopping caplets, and we show that when these conditions are not met, the swaption prices will always be below this boundary.

Empirically, we then find that the market prices of swaptions trade very close to this boundary in both normal and excited periods. When the no-arbitrage boundary is exceeded, reversion to the boundary is swift.

In this article, we ask whether the U.S.-Dollar swaption[1] and caplet markets are internally consistent. To explain what we mean by internally consistent, we first note that, for a given currency, the same quantities, that is, forward rates, constitute the common underliers for both markets.[2] By asking whether the caplet and swaption markets are internally consistent, we therefore ask whether the same pricing model *with the same calibration* can simultaneously account for the observed market prices of swaptions and caplets. This fundamental question affects relative valuation, hedging, and risk management, and was already raised as early as 1997 by Jamshidian [1997] in the article that introduced the LIBOR market model (LMM) for interest-pricing pricing.[3]

The precise results will clearly depend on the chosen option model. We explain in the following the reason for using the LMM-SABR model (Rebonato [2007], Rebonato and White [2010], Rebonato, McKay and White [2009]) for our analysis. We can already say, however, that we are looking for almost model-independent indications as to whether the two markets may be congruent. As the LMM-SABR model provides a set of no-arbitrage conditions within which different specific models can be naturally nested, it constitutes a good starting point for our analysis.

Since the two markets are connected by the same underliers (forward rates), the question naturally arises as to how, in an efficient market, swaptions and caplets could fail to display coherence. Copious literature in the area of potential causes for market inefficiency (see, e.g., Shleifer [2000]) points to the limits to and the riskiness of (pseudo)-arbitrage. We therefore discuss in the concluding section to what extent a marked-to-market player could bring about the reconciliation of the two markets if they were indeed found to trade out of line with each other.

THE LMM-SABR PRICING FRAMEWORK

Both the LMM and its offspring, the LMM-SABR, are usually referred to as "models." However, they are not truly models but sets of no-arbitrage conditions. These no-arbitrage conditions have been derived for the LMM-SABR approach under a variety of choices of numeraire by, among others, Henry-Labordere [2006, 2007], and Rebonato, McKay and White [2009]. A number of "models" are then nested as special cases depending on the specification of the volatility, correlation, and (for the LMM-SABR) volatility-of-volatility functions.

The specific LMM-SABR model presented in Rebonato [2007], Rebonato and White [2010], Rebonato, McKay and White [2009] is constructed to recover with great accuracy and with closely similar distributional assumptions the prices produced by the SABR model (see, e.g., Hagan et al. [2002]) for European options. As explained below, the SABR model is a stochastic-volatility extension of the constant elasticity of variance (CEV) model.[4] As the latter has become the pricing standard for caplets and swaptions (the two markets we want to "explain"), this is in itself a very desirable feature. In addition, the LMM-SABR model has been shown to have surprisingly robust econometric foundations (Rebonato, McKay and White [2009]) and a very attractive cross-hedging performance (Rebonato, Pogudin and White [2008])—where cross hedging refers to the ability to vega hedge swaptions of different maturities once the model has been calibrated to caplets.

The most glaring limitation of the LMM (-SABR) models is the absence of jumps. However, in the empirical part of our work, we find the greatest lack of congruence between the caplet and swaption markets in the long-dated expiries. As jumps tend to affect short-dated options much more strongly, it is most unlikely that the absence of jumps can be the cause for these discrepancies. The LMM-SABR framework in general and the Rebonato [2007] specification, in particular, are therefore reasonable starting points for the investigation we conduct in this article.

For the sake of brevity, we report only those features of the SABR and LMM-SABR model necessary to understand the analysis in the article, and the reader is referred to the references above for a fuller treatment. In the SABR model, the process for the forward (swap) rate, f_t^T, is of the stochastic-volatility CEV type:

$$df_t^T = (f_t^T)^{\beta^T} \sigma_t^T dz_t \tag{1}$$

$$\frac{d\sigma_t^T}{\sigma_t^T} = \nu^T dw_t \tag{2}$$

$$E[dz_t dw_t] = \rho^T dt \tag{3}$$

The superscript T indicates that the parameters of the SABR model (β^T, ρ^T, and υ^T), as quoted in the market, depend on the expiry of each forward rate. On any given day, no single set of parameters describes either the whole European swaption *or* caplet market. The situation is reminiscent of the Black implied volatilities that were quoted in the 1990s and also displayed a dependence on the expiry of the forward rate (the so-called "term-structure of volatilities"). This expiry dependence suggests that the Black (and now the SABR) models are reduced-form models that produce the correct market prices for *European* options, without having to specify a more complex latent structure that would only affect multi-forward options (such as swaptions).

In the LMM and in the LMM-SABR framework one then has to make assumptions about this latent structure. An infinity of specifications are consistent with a given set of European swaption (or caplet) prices. Each assumption will specify a particular "model." Financial justification and statistical analysis must therefore be invoked to choose a desirable specification for the volatility and correlation functions. The link between these unobservable inputs and market observables is *via* the smile surface, because any choice for the volatility function will uniquely determine the current and future smile surface. The future smile is, of course, unknown. However, historical observation of smile surfaces shows that their shapes remain remarkably stable as a function of the residual time to maturity of the forward rates (see Exhibit 18). The challenge for a dynamic model is therefore to explain the European prices for all strikes and expiries obtained by the reduced-form SABR parametrization in a way that is both parsimonious, financially justifiable, and reflective of the observed regularities in the shape of the smile surface: future smiles should "look like" the smiles that have been observed in the past. We shall show that some very simple and prima facie choices for the volatility functions fail to pass this last test.

The LMM-SABR model satisfies these requirements. It is characterized by the following constitutive equations:[5]

$$df_t^i = (f_t^i)^{\beta_i} s_t^{T_i} dz_t^i \tag{4}$$

$$s_t^{T_i} = k_t^{T_i} g_t^{T_i} \tag{5}$$

$$\frac{dk_t^{T_i}}{k_t^{T_i}} = \mu_k^i dt + h_t^{T_i} dw_t^i \tag{6}$$

$$E[dz_t^i dz_t^j] = \rho_{ij} dt \tag{7}$$

$$E[dw_t^i dw_t^j] = r_{ij} dt \tag{8}$$

$$E[dw_t^i dz_t^j] = R_{ij} dt \tag{9}$$

The function $s_t^{T_i}$ represents the instantaneous volatility of the forward rate of expiry T_i, and is made up of the product of a deterministic backbone, the function $g()$, and a stochastic multiplicative scaling factor, $k_t^{T_i}$, that randomly changes the level of volatility through time. The intuition behind Equations (5) and (6) is simple: In the limit of the volatility of volatility going to zero, the function $g(t, T)$ describes the instantaneous volatility at time t of a forward rate of expiry T. In the stochastic volatility setting of the LMM-SABR model, the volatility of the forward rate is made stochastic by multiplying the deterministic backbone by a geometric Brownian process with initial value (close to) and volatility $h(t, T)$. In this sense, the function $h(t, T)$ plays the role of the time t volatility of volatility of a forward rate of expiry T.

We note that Rebonato and White [2010] have shown that very accurate analytic estimates of caplet and swaption prices can be obtained from the model parameters without having to resort to a Monte Carlo simulation. This feature greatly simplifies the task presented in this article.

THE $g()$ AND $h()$ FUNCTIONS

To ensure time-homogeneity (that guarantees self-similarity of the future smile surface), we impose that the deterministic volatility backbone, the "volatility of volatility," should only depend on the difference between the current time and time to maturity: $g_t^T = g(T-t)$ and $h_t^T = h(T-t)$. The assumption that underlies this choice is that the only feature that differentiates two forward rates is their residual time to expiry. Fitting market prices of caplets (both with deterministic or stochastic volatility) clearly indicates that the instantaneous volatility of a forward should not be constant throughout its life, with a maximum instantaneous volatility encountered, on average, when it is 6 to 18 months away from expiry. This accords well both with a financial account of how the actions of monetary authorities are likely to affect the volatility of forward rates (see, e.g., Rebonato [2004], Rebonato [2002], and Rebonato, McKay, and White [2009]) and with direct econometric studies of the volatility of forward rates as a function of their residual time to expiry. There is a rich literature supporting humped volatility structures (see, e.g., Amin and Morton [1994], Goncalves and Issler [1996], Ritchken and Chuang [1999], Brigo and Mercurio [2001], Rebonato, McKay, and White [2009] among others). The time-homogeneity[6] assumption is also common in the literature on interest-rate volatility (see, e.g., Rebonato [2004], Amin and Morton [1994], Ritchken and Chuang [1999], Mercurio and Moraleda [2000]).[7] We therefore choose for the forward-rate volatility function the following humped and time-homogeneous functional form:

$$g(T_i, t) = g(T_i - t) = g(\tau_i) = [a + b(\tau_i)]\exp(c\,\tau_i) + d$$

A typical shape for the $g()$ function is shown in Exhibit 1, and Exhibit 2 displays a typical shape of the $g()$ function obtained from fitting its parameters (a, b, c, and d) to caplet prices as described in Rebonato, McKay, and White [2009].

EXHIBIT 1

A Typical Shape for the Time-Homogeneous Deterministic Backbone, $g()$, of the Stochastic Volatility, s

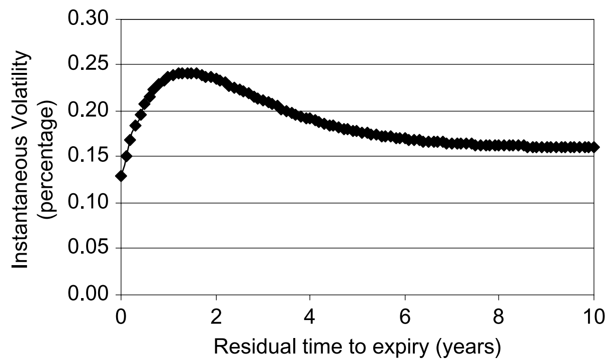
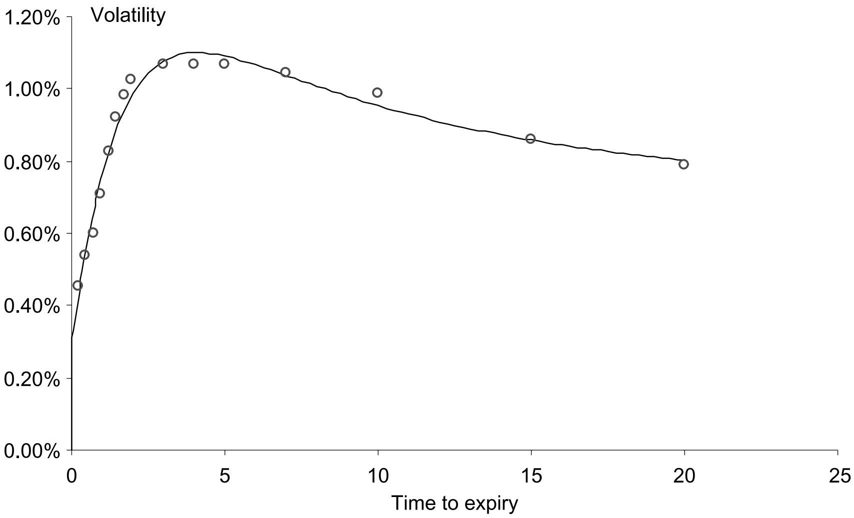


EXHIBIT 2

Calibration of $g()$ Function to Caplet Prices, July 5, 2010, U.S.-Dollar Caplet Market, RMS, $\beta = 0$ (normal case, 1% means 100 bps)



Note the goodness of the fit and the natural recovery of a humped shape from market data. As this shape has a profound effect on the relative pricing of caplets and swaptions, it is discussed in some detail in the next section.

Finally, we note that on any given day, the market implied volatilities of caplets of all expiries cannot in general be exactly recovered using the same parameters a, b, c, and d of the $g()$ function. Perfect pricing is ensured by selecting slightly different values, $k_0^{T_i}$, for the initial values of the process $k_t^{T_i}$ in Equation (6).

Similarly, in order to fit exactly the SABR volatility of volatility for all expiries exactly, constant, expiry-specific factors, $\xi_0^{T_i}$, pre-multiply the $h()$ function, so that the dynamics for the stochastic drivers, $k_t^{T_i}$, becomes:

$$\frac{dk_t^{T_i}}{k_t^{T_i}} = \mu_k^i\, dt + \xi_0^{T_i} h_t^{T_i}\, dw_t^i$$

Needless to say, doing so breaks exact time homogeneity. However, the corrections are in general very small, as can be appreciated by looking at Exhibit 2. Although a formal Bayesian estimation was not employed in the fitting, the adjustment factors $\{k_0^{T_i}\}$ and $\{\xi_0^{T_i}\}$ can be interpreted as the corrections most compatible with the prior of a strictly time-homogeneous evolution of the smile surface, given the posited dynamics and the constraint of exact recovery of the market prices. As we show below, these adjustment factors also provide a transparent quantification of the quality of the fits, which conveys more information than the usual chi-squared statistic.

GAINING INTUITION

The problem of reconciling the prices of caplets and swaptions is not trivial because of the imperfect correlation between the forward rates, $\{f_i\}$, that make up a given swap rate, SR, and the time-dependence of the instantaneous volatility of the forward rates. It is therefore important to gain an intuitive understanding of how these two quantities affect the volatility of swap rates. To this effect, let's first define the *terminal* correlation from time t_0 to time T between forward rates j and k, by $\hat{\rho}_{jk}(T)$:

$$\hat{\rho}_{jk}(T) = \frac{\int_{t_0}^{T} \rho_{jk}(t)\sigma_j(t)\sigma_k(t)dt}{\sqrt{\int_{t_0}^{T} \sigma_j^2(t)dt \int_{t_0}^{T} \sigma_k^2(t)dt}}$$

where $\sigma_j(t)$ is the instantaneous volatility at time t of forward rate j and $\rho_{jk}(t)$ is the instantaneous correlation at time t between forward rates j and k. Clearly, the terminal correlation will in general be lower than one even in the presence of perfect instantaneous correlation as long as the instantaneous volatility functions are not constants. We show below that terminal correlation is directly linked to the swap-rate volatility.

To see how this comes about, we begin by observing that a swap rate, SR_i, can always be expressed as a linear combination of forward rates:

$$SR_i = \sum_{k=1,n_i} w_{ik} f_k$$

where w_{ik} are suitable "weights," defined for example in Rebonato [2004], and n_i signifies the number of forward rates in the i^{th} swap rate, SR_i. Next, consider for simplicity the case of a lognormal process for the swap rate (this assumption is purely made for ease of exposition). Under lognormality, the Black formula

(Jamshidian [1997]) provides the pricing for a European swaption. The volatility input for the Black formula is given by the "implied" (in the lognormal case, root-mean-squared) volatility for the i^{th} swap rate:

$$[\sigma_{Black}^{SR_i}]^2 T_{exp} = \int_0^{T_{exp}} \sigma_{inst}^{SR_i}(u)^2\,du \tag{10}$$

where T_{exp} is the expiry time of the European swaption in question, $\sigma_{inst}^{SR_i}(t)$ is the instantaneous volatility of the associated swap rate at time t, and $\sigma_{Black}^{SR_i}$ is the market-implied volatility for the swaption associated with swap rate SR_i.

If one makes a joint log-normal assumption for all the forward rates and for the swap rate (Rebonato [1999]),[8] a straightforward application of Ito's lemma then links the instantaneous volatility of a given swap rate with the instantaneous volatilities, $\{\sigma_i(t)\}$, of all the underlying forward rates:

$$\sigma_{inst}^{SR_i}(t)^2 = \sum_{j,k=1,n_i} \zeta_j^i(t)\zeta_k^i(t)\rho_{jk}(t)\sigma_j(t)\sigma_k(t) \tag{11}$$

with

$$\zeta_j^i(t) = \frac{w_{ij}(t) + \sum_{k=1,n_i} f_k(t)\frac{\partial w_{ik}}{\partial f_j}}{\sum_{m=1,n_i} w_{im}(t) f_m(t)} \tag{12}$$

and

$$w_{ij}(t) = \frac{P_t^{T_j}}{\sum_{j=1,n_i} P_t^{T_j}\tau_j} \tag{13}$$

where $P_t^{T_j}$ is the time-t price of a discount bond expiring at time T_j and τ_j is the tenor of the j^{th} forward rate.

Rebonato and Jaeckel [2003] explain why it is a good approximation to assume that in the expression above the new weights $\{\zeta_j^i(t)\}$ and the forward rates can be treated as deterministic with a value equal to their realization today, that is:

$$\sigma_{inst}^{SR_i}(t)^2 \simeq \sum_{j,k=1,n_i} \zeta_j^i(t_0)\zeta_k^i(t_0)\rho_{jk}(t)\sigma_j(t)\sigma_k(t) \tag{14}$$

Then the implied (root-mean-squared) swap-rate volatility, $\sigma_{Black}^{SR_i}$, to use in the Black formula to price the ith European swaption will be obtained as:

$$[\sigma_{Black}^{SR_i}]^2 T = \int_0^{T_{\exp}} \sigma_{inst}^{SR_i}(t)^2\,dt$$
$$= \sum_{j,k=1,n_i} \zeta_j(t_0)\zeta_k(t_0)\int_0^{T_{\exp}} \rho_{jk}(t)\sigma_j(t)\sigma_k(t)\,dt$$

This clearly shows the importance of the covariance elements Cov_{jk}:

$$Cov_{jk} = \int_0^{T_{\exp}} \rho_{jk}(t)\sigma_j(t)\sigma_k(t)\,dt \tag{15}$$

It is clear that the value of the covariance element will depend on both the correlation between the two forward rates and the time dependence of the volatility functions. Indeed, Equation (15) shows the link with the terminal correlation defined above and explains how the time dependence of the volatility functions affects the pricing by altering the terminal correlation. However, the time dependence of the instantaneous volatilities affects the pricing of swaptions in two distinct ways, as explained below.

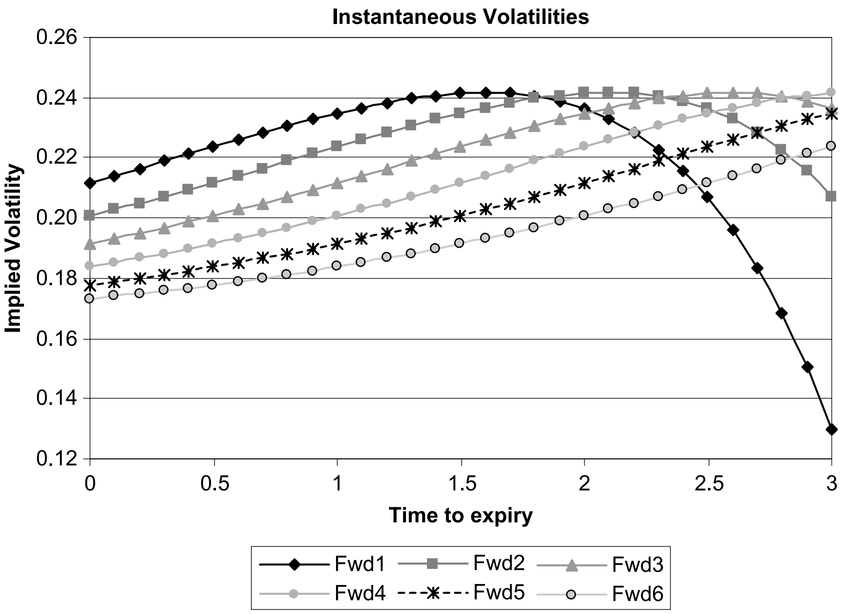
The first effect is shown in Exhibit 3, which depicts the time dependence of the same (time-homogeneous) volatility function as that in Exhibit 1 during the life of a 3Y × 3Y semi-annual European swaption.[9] The expiry of the swaption coincides with the expiry of the first forward rate only. However, by option expiry, all the other forward rates will still be "alive" and will have traversed different portions of their lives. Therefore the covariance elements between the various forward rates (that entail integrals out to time T_{exp}) will, in general, depend on the shape of the volatility function during the time to expiry (consider, for instance, $Cov_{1,5}$).

It is clear by a simple generalization of Schwartz's inequality that for any two portions of the instantaneous volatility functions with given root-mean-squared values over the life of the option, the covariance element (15) will attain its highest possible value when the two functions are constant *over the integration period*.[10] Therefore, within the limits of the approximations and assumptions above, the root-mean-squared volatility of the swaption will almost always be below the value, $\sigma_{Black,\max}^{SR_i}$, given by

$$[\sigma_{Black,\max}^{SR_i}]^2 = \sum_{j,k=1,n_i} \zeta_j^i(t_0)\zeta_k^i(t_0)\hat{\sigma}_j^0\hat{\sigma}_k^0 \tag{16}$$

where the quantities $\hat{\sigma}_j^0$ and $\hat{\sigma}_k^0$ are the root-mean-squared volatilities obtained when the instantaneous volatilities are constant over the integration period.[11]

EXHIBIT 3
The Case of a 3Y x 3Y Swaption



"alive" by swaption expiry. Consider now the case of a short-expiry (3Y) into a long-tail (7Y) swaption. Exhibit 4 shows two possible instantaneous volatilities with the same root-mean-squared volatility over the full life of the last forward rate in the swap rate, that is the 10-year forward rate.

During the 3 years of the life of the swaption, however, the 10-year forward rate will experience a very different volatility in the two cases (around 8% in one case, and around 12% in the other), despite the fact that both functions are compatible with the market price of the 10-year caplet. This shows that the prices of a swaption obtained with two time-dependent volatility functions in Exhibit 4 (that price the caplet identically) can be very different.

So, two instantaneous-volatility-related effects (the "Schwartz-inequality" effect for the covariance element, and the way a non-constant volatility is apportioned over the life of the option)[12] can affect the swaption price. Of course, the impact on the resulting swaption prices of these two effects will depend on the precise interplay between the time to swaption

In general, for fixed root-mean-squared caplet volatilities, the terminal correlation, and hence the covariance elements, and hence the swaption prices will therefore depend on the precise time dependence of the instantaneous volatilities. This is the first way time dependent volatilities affect swaption prices.

To understand the second effect alluded to above, consider (again in a Black world) a caplet with implied (root-mean-squared volatility), $\hat{\sigma}_j(T_j)$, given by

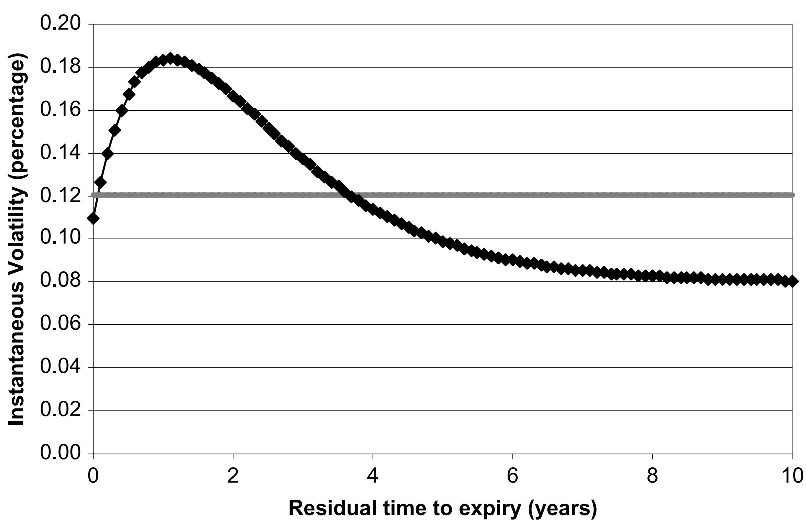
$$\hat{\sigma}_j(T_j) = \sqrt{\frac{1}{T_j}\int_0^{T_j} \sigma(u,T_j)^2\, du} \qquad (17)$$

Note that if the instantaneous volatility of the jth forward rate were constant, $\sigma(t,T_j) = \sigma_j^0$, then the root-mean-squared volatility would be independent of the upper integration limit in Equation (17):

$$\hat{\sigma}_j(\tau) = \sqrt{\frac{1}{\tau}\int_0^{\tau} (\sigma_j^0)^2\, du} = \hat{\sigma}_j(T_j), \quad \forall\tau \;\; 0 \le \tau \le T_j \qquad (18)$$

Now, for a given swaption all but one of the underlying forward rates will still be

EXHIBIT 4
Two Possible Instantaneous Volatilities with Exactly the Same Root-Mean-Squared Value Out to Caplet (but not swaption!) Expiry



expiry and the length of the underlying swap rate and the parameters (mainly d and c) of the $g()$ function.

It is for these combined reasons that commenting on the congruence of the caplet and swaption markets will in general depend on analyzing both the instantaneous volatilities of, and the correlation among, the underlying forward rates. In particular, without analyzing in detail the time dependence of the parametrized function $g()$, it is impossible to say a priori whether instantaneous correlation or time dependence of volatilities will have a bigger effect on the volatility of a swap rate.

The results above are obvious in the Black (lognormal) world discussed in this section. The intuition (although not the details) remain, valid in the more complex case of the LMM-SABR model.

THE DATA

Our empirical investigation was carried out using swaption and caplet market prices for at-the-money and 25-delta (call and put) strikes for every trading day from Jun 24, 2005–Jul 5, 2010. They therefore encompass both the benign market conditions that prevailed until summer 2007, the turbulent period that peaked around September 2008, and the slow return to market normality that followed.

QUALITY OF THE CALIBRATION TO MARKET CAPLET PRICES

The LMM-SABR model was first calibrated to caplet prices for each trading day in the dataset above using two different values of the exponent β: $\beta = 0$ and $\beta = 0.5$. The results of the calibration are reported by showing the initial values, $k_0^{T_i}$, of the process $k_t^{T_i}$, (Exhibits 5 and 6) and the constants $\xi_0^{T_i}$ (Exhibits 7 and 8) for the two values of the exponent β. This shows both the goodness of the time-homogeneity assumption and the aptness of the chosen functional form. (Recall that once the adjustment factors $\{k_0^{T_i}\}$ are applied, the market SABR caplet prices are recovered exactly by construction.)

We note that the vast majority of the $\{k_0^{T_i}\}$ and $\{\xi_0^{T_i}\}$ quantities are very close to 1. The only two forward rates for which there is an appreciable difference from 1 are the first and, to a lesser extent, the second IMM (futures) contracts. As a result, on most trading days, the market prices are approximately compatible with the time-homogeneous assumption that underpins our analysis. As mentioned, the largest discrepancies are encountered for the shortest-expiry caplets. This is not surprising because in periods of market excitations, the time-homogeneity assumptions are not satisfied, and this affects most strongly the shortest-expiry forward rates, that is the forward rates whose expiries fall during the short-lived period of market excitation.

The magnitude of the adjustment factors $\{\xi_0^{T_i}\}$ required to make the volatility of volatility match the shape of the market smile for all expiries is greater than that for the factors $\{k_0^{T_i}\}$. However, for most days, the deviations from 1 (the value that indicates perfect time homogeneity) are less than 20%.

We therefore consider the calibration successfully carried out, and with the parameters obtained as described above, we finally are in a position to begin to explore the congruence of the caplet and swaption markets.

RESULTS I: TIME-HOMOGENEOUS *g*() AND *h*() FUNCTIONS

To calculate the model swaption prices, we use the approximate but accurate formulae reported in Rebonato and White [2010], which allow for the calculation of the prices of swaptions implied by a set of forward-rate parameters for the LMM-SABR model without having to use a Monte Carlo simulation. In this approach, the swaption price is produced by obtaining the SABR parameters for the swap rate process

$$dSW_t = \left(SW_t\right)^{\beta}\Sigma_t dz_t \qquad (19)$$

$$\frac{d\Sigma_t}{\Sigma_t} = V dw_t$$

given the parameters of the forward-rate LMM-SABR process. The LMM-SABR forward-rate parameters were chosen to match for each trading day the observed SABR market prices under the time-homogeneity desideratum, as described in the previous section (we show in Appendix the accuracy of these approximations.)

As the derivation of the expressions of the swaption SABR parameters, Σ_0 and V, is somewhat lengthy, we simply report the formulae employed for ease of reference:

$$\Sigma_0 = \sqrt{\frac{1}{T}\sum_{i,j}\left(\rho_{ij}W_i^0 W_j^0 k_0^i k_0^j \int_0^T g_t^i g_t^j dt\right)} \qquad (20)$$

EXHIBIT 5
The Initial Value, $k_0^{T_i}$, of the Process $k_t^{T_i}$ for the Case of $\beta = 0$, Selected Caplets

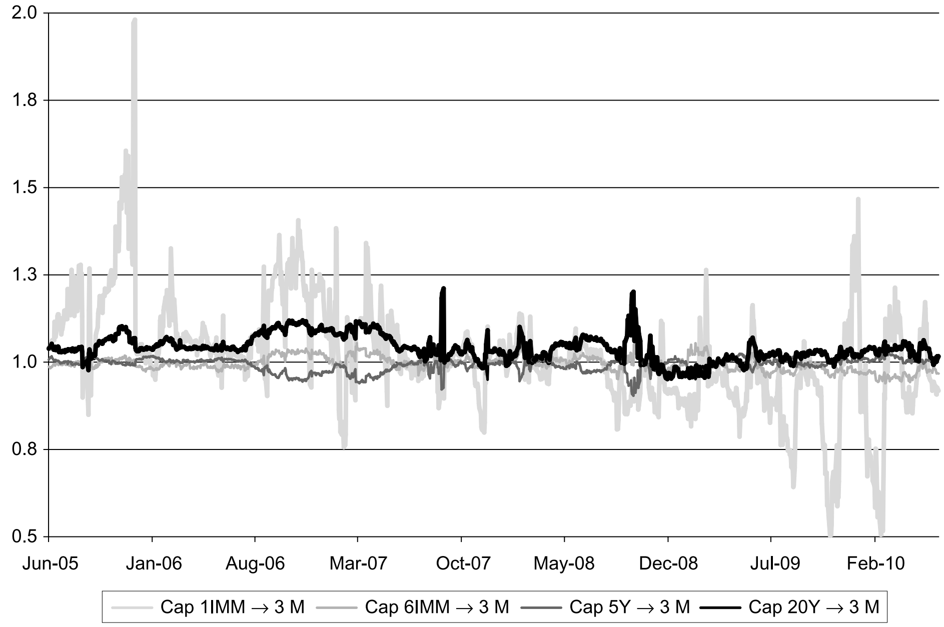


EXHIBIT 6
The Initial Value, $k_0^{T_i}$, of the Process $k_t^{T_i}$ for the Case of $\beta = 0.5$, Selected Caplets

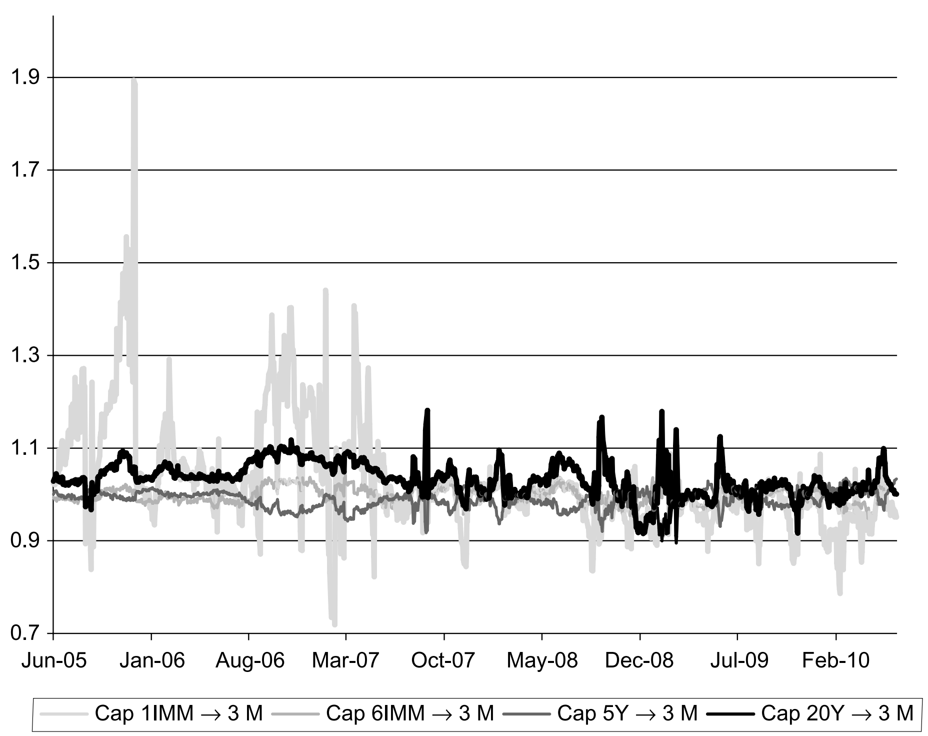


EXHIBIT 7

The Factors $\xi_0^{T_i}$, for the Case of $\beta = 0$

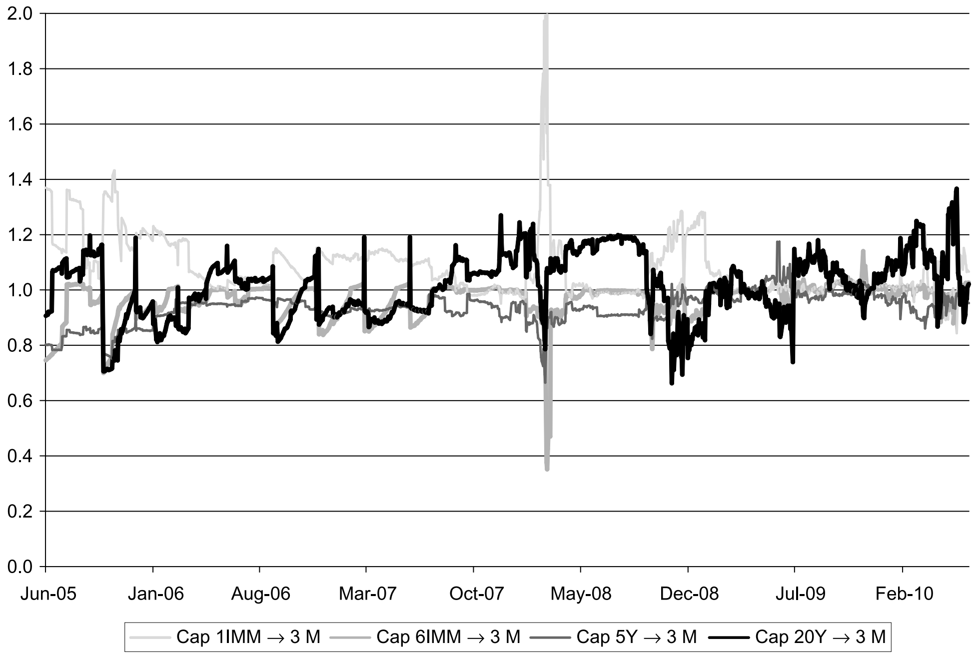
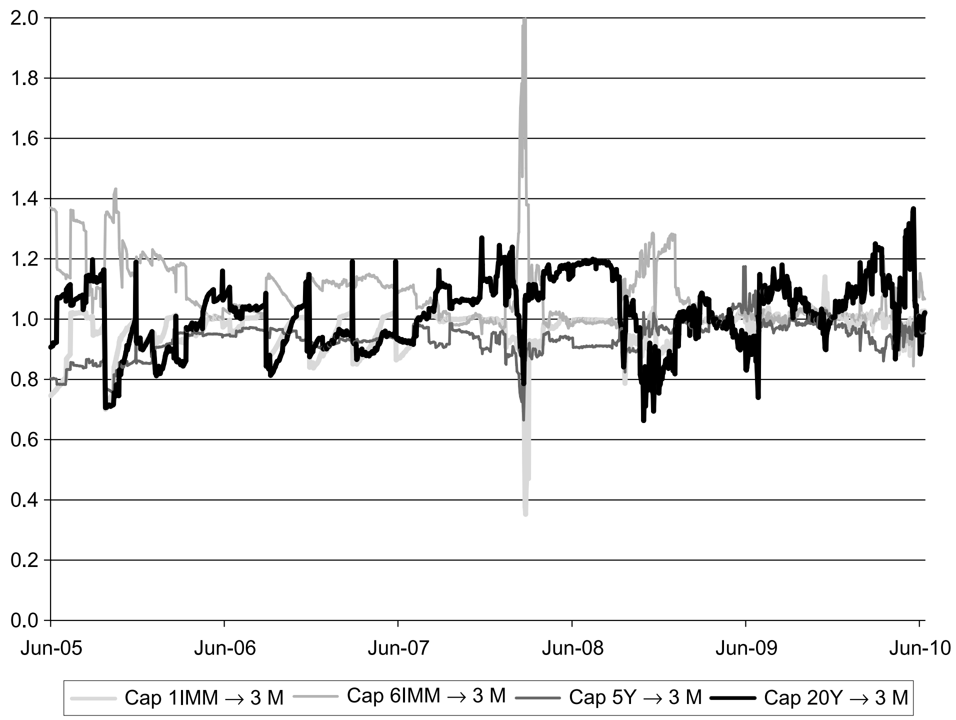


EXHIBIT 8

The Factors $\xi_0^{T_i}$, for the Case of $\beta = 0.5$



$$V = \frac{1}{\Sigma_0 T}\sqrt{2\sum_{i,j}\left(\rho_{ij} r_{ij} W_i^0 W_j^0 k_0^i k_0^j \int_0^T g_t^i g_t^j \hat{h}_{ij}(t)^2 t dt\right)} \quad (21)$$

with

$$\hat{h}_{ij}(t) = \sqrt{\frac{1}{t}\int_0^t h^i(s) h^j(s) ds} \quad (22)$$

and

$$W_k^t = w_k \frac{(f_t^k)^\beta}{(SR_t)^\beta} \quad (23)$$

and refer to Rebonato and White [2010] for a justification. The quantities ρ_{ij} and r_{ij} denote the correlations among the forward rates and their volatilities, respectively. We stress that for reasons explained in the following, all these correlations were set to 1 in our study, unless otherwise stated. As we shall see, doing so makes the results we obtain even stronger.[13]

We have examined in our study a number of combinations of expiries and underlying swap lengths ("tails"). Without great loss of information, we can limit the presentation of the results to the case of short and long expiries into short and long tails (where, in both cases, "long" and "short" indicate 10 and 2 years, respectively). We begin by looking at the values of the at-the-money (ATM) implied volatilities.

In the case of short tails, we find that the agreement is very good throughout the five years in our data, including the exceptionally turbulent period of September 2008, both for short (Exhibits 9 and 10) and long expiries (Exhibits 13 and 14). The results obtained with the exponent $\beta = 0$ perform somewhat better than $\beta = 0.5$.

In the case of long tails, the agreement is also very good until the beginning of the recent financial crisis (summer 2007) but then becomes very poor in September 2008 (see Exhibits 11, 12, 15, and 16). After this date, for long tails, the model values remain consistently below the market prices. Note that we obtained the model values with a correlation of 1 both for ρ and for r. As we discussed below, a more realistic correlation structure would have lowered the model prices even more. We therefore consider our results as correlation-boundary cases.

These results indicate

- that the agreement between model and market prices is always better for the exponent $\beta = 0$ than for $\beta = 0.5$;
- that the agreement is somewhat better for short expiries into short tails than for long expiries into long tails;
- that the agreement tends to break down during turbulent periods for long tails (and especially so for short expiries into long tails).

An explanation for these findings can be offered along the following lines.

EXHIBIT 9
Market and Model Swaption (ATM vol, 2Y x 2Y, $\beta = 0$)

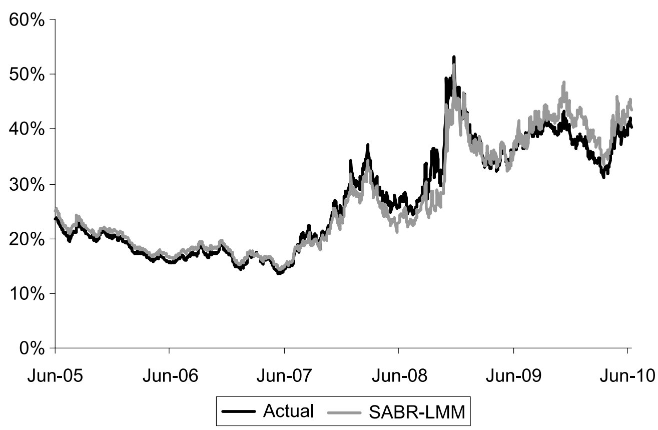
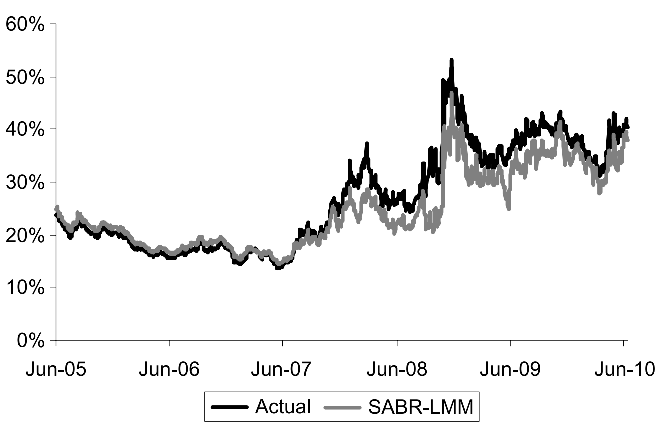


EXHIBIT 10
Market and Model Swaption (ATM vol, 2Y x 2Y, $\beta = 0.5$)



To explain why we obtain better results with $\beta = 0$, we point to work by Rebonato [2003] that clearly shows that for rate levels between approximately 2% and 6%, the dependence of swaption implied volatilities on the swap rate level is better explained by a normal (as opposed to lognormal) behavior. This is also corroborated by the work by Rebonato, de Guillaume, and Pogudin [2010] that explores the dependence of realized (as opposed to implied) volatility as a function of the level of rates. Also, in this work, for the level of swap rates encountered in the present study, a normal behavior (that corresponds to $\beta = 0$) accounts best for empirical data.

EXHIBIT 11

Market and Model Swaption (ATM vol, 2Y x 10Y, $\beta = 0$)

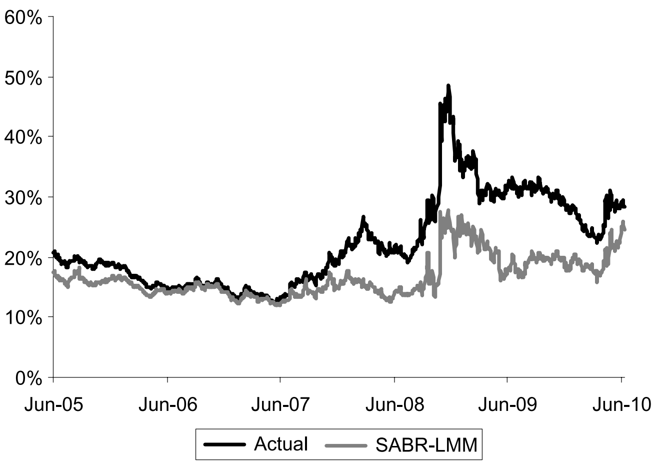


EXHIBIT 12

Market and Model Swaption (ATM vol, 2Y x 10Y, $\beta = 0.5$)

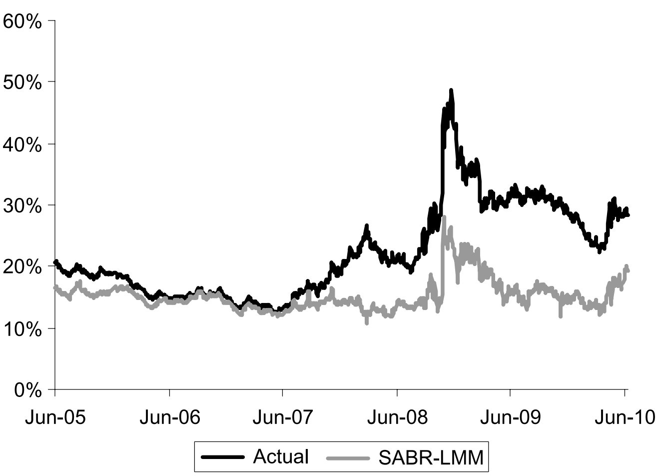


EXHIBIT 13

Market and Model Swaption (ATM vol, 10Y x 2Y, $\beta = 0$)

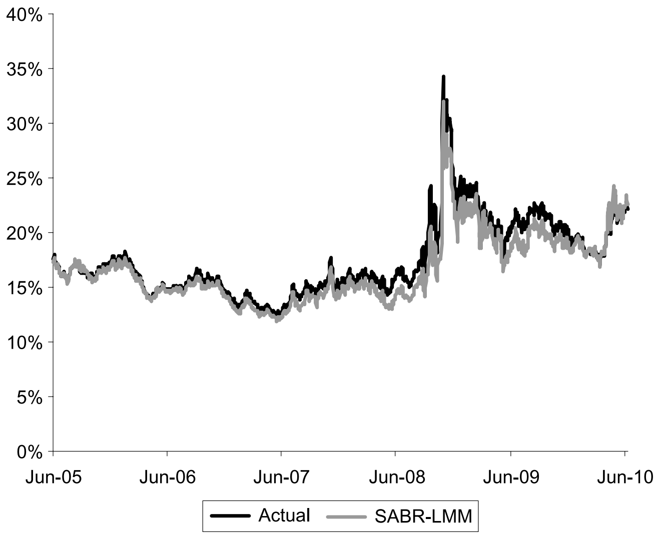
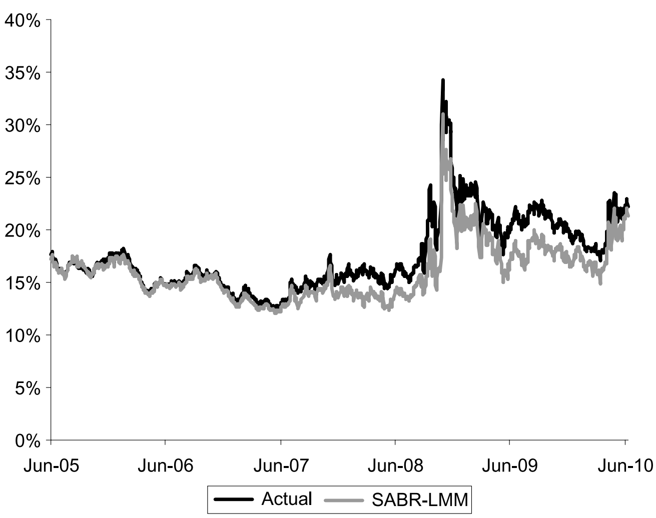


EXHIBIT 14

Market and Model Swaption (ATM vol, 10Y x 2Y, $\beta = 0.5$)



The following remarks are in order.

- Markets evolve, and traders periodically adjust their view on market skews every few months. Rebonato, McKay, and White [2009] show that, over extended periods of time during which the level of rates remains roughly constant, these

EXHIBIT 15

Market and Model Swaption (ATM vol, 10Y x 10Y, $\beta = 0$)

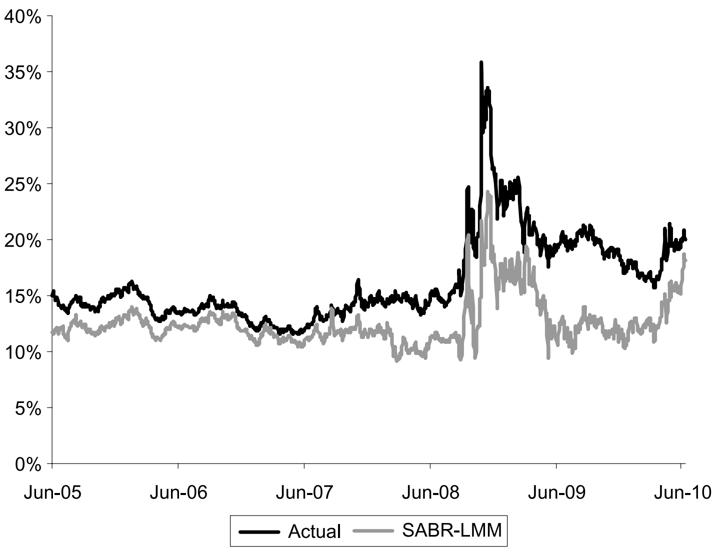
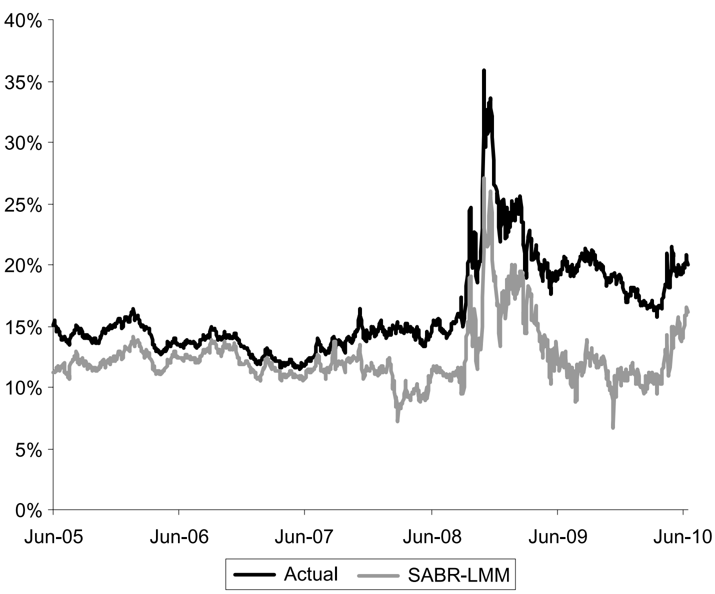


EXHIBIT 16

Market and Model Swaption (ATM vol, 10Y x 10Y, $\beta = 0.5$)



changing views are translated into changes in the correlation coefficient, ρ, between forward rates and volatility (see Equation [3]), rather than the exponent. See Rebonato, de Guillaume, and Pogudin [2010] for a discussion of this point.

- Even in the same period, often the short-expiry swaptions market shows a different skew from the long-expiry swaptions market. Accounting for this feature does not a priori require an expiry-dependent exponent, β, as the effect can be fully accounted for by dynamic structure of the LMM-SABR model (in particular, by the dependence on the residual time to maturity of the volatility-volatility function, $h()$).

In order to understand why during excited periods long-tail swaptions may be underpriced under the time-homogeneous assumption that we have made so far, consider Exhibit 17, which shows two fits to the $g()$ function, one obtained in normal market conditions and one during a period of turmoil.

As one would expect, the instantaneous volatility function fitted during the excited period assumes much higher values for short times to expiries. What is more surprising is that it reaches much *lower* values for long times to expiry, as shown in Exhibit 17. This is as an inescapable consequence of the time-homogeneity assumption. Consider in fact the pricing of a long-tail swaption during a period of market excitation. Assume that the market expects the turbulence to subside (and hence volatilities to decline) after a period of time of weeks or months.[14] The time-homogeneity assumption, however, implies that the future smile surface will remain the same over time. In order to fit at the same time, the observed short-dated (exceptionally high-volatility) option and the long-dated (almost-normal-volatility) options, a time-homogeneous model must therefore imply that all forward rates experience very high volatility when their expiries are short and *extremely low volatility when they have a long time to expiry*. The high instantaneous volatility at the short end is required to price correctly the short-dated and short-tail caplets. In order for long-expiry options to be priced correctly (at a close-to-normal implied volatility level), however, the instantaneous volatility has to decline very sharply for the correct root-mean-squared volatility to obtain.

It is now easy to see how this has a direct impact on long-tail options—the more so if the expiry is short—during excitation periods. Consider, for instance, the 2Y × 10Y swaption. When the time-homogeneous volatility function is fitted to the caplet market prices for an excited day, during the two years of the option life, most of the underlying forward rates experience in the model the very low instantaneous volatility shown

EXHIBIT 17
Normal and Excited Instantaneous Volatility Functions Fitted on Two Different Days

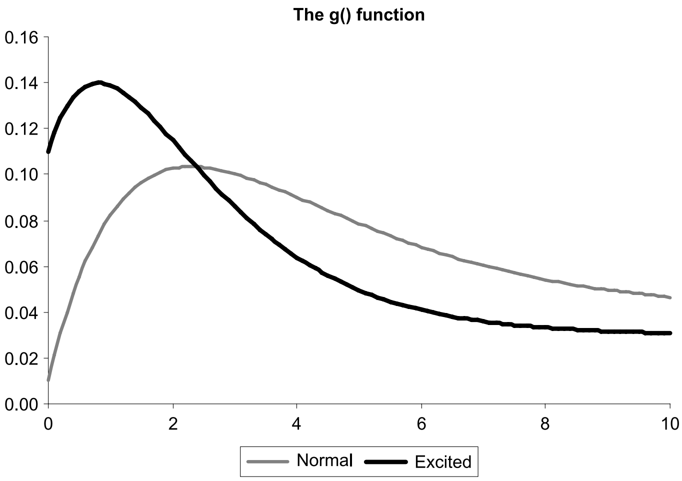
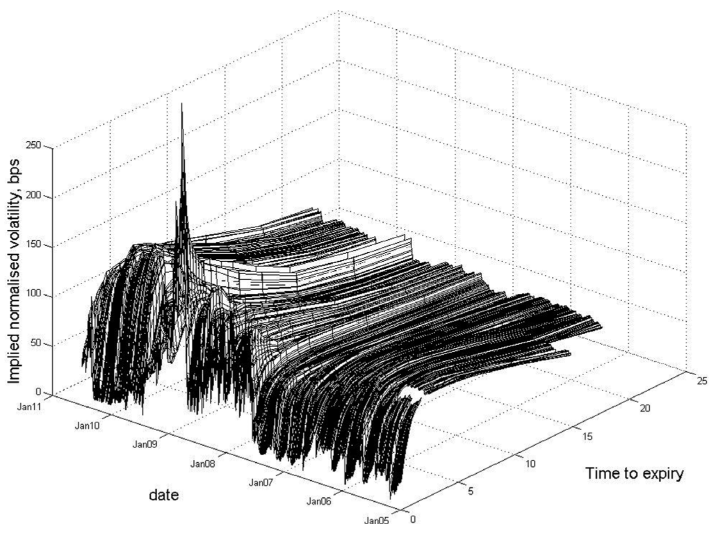


EXHIBIT 18
History of Instantaneous Volatilities Fitted to Caplet Prices over the Historical Period Explored in this Study



in Exhibit 17. This certainly contributes to the mispricing shown in Exhibits 11 and 12. Indeed, note that the discrepancy appears as soon as the market enters in to a (transitory) phase of excitation (summer of 2007) and reaches a maximum in the after-Lehman Brothers period. As markets then slowly return to normal, the discrepancy begins to narrow and almost disappear.

The strength and limitations of the time-homogeneity assumptions are clearly shown in Exhibit 18, which shows the ATM implied volatilities of caplets as a function of expiries for all the trading days in our dataset from June 2005 to July 2010. Most of the time, the time-homogeneity assumption is well justified, as the term structure of ATM volatilities maintains over extended periods a remarkably self-similar shape.[15] However, the exceptional events of the 2007–2009 credit crisis show that the market expectation of where long-expiry volatilities will be relative to the short-expiry level of volatility is not always the same.

RESULTS II: FLAT $g()$ AND $h()$ FUNCTIONS

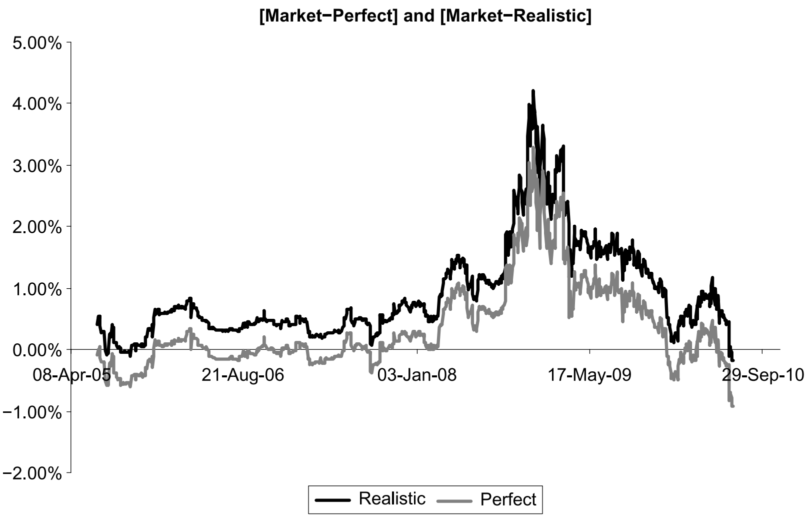
The analysis presented in the previous section indicates that the inadequacy of the time-homogeneous assumption during periods of market excitation can account for some of the systematic underpricing of swaptions implied by the LMM-SABR model calibrated to caplets with $g(t, T) = g(T - t)$. However, it is apparent that *even during normal market conditions*, the model prices for long-dated, long-tail swaptions tend to be lower than the corresponding market prices. Recall that this is true even for prices obtained with perfect instantaneous correlation. Any lower (and econometrically more realistic) value for the instantaneous correlation would produce an even greater discrepancy between model and market prices. See, for instance, the results presented in Exhibit 19 with the flat-volatility caplet calibration (which is, of course, independent of the correlation among forward rates and among volatilities), using either perfect correlation or a "realistic" correlation among forward rates (upper curve) for the 10Y × 10Y case.[16] It is clear that the more realistic imperfect correlation calibration gives rise to an even greater degree of mispricing.

We therefore try to look more deeply at how the market may price swaptions relative to caplets.

Within the LMM-SABR set of no-arbitrage models, the natural way to increase the model prices of swaptions, while still recovering the prices of the market caplets and retaining absence of arbitrage, is to modify the shape of the instantaneous volatility functions in such a way as to keep their root-mean-squared unchanged. Recall from the previous discussion that for any given instantaneous correlation function, the maximum swaption volatility

EXHIBIT 19

Difference between the Market and Model Implied Volatilities for the Perfect Correlation Case, 10Y x 10Y Swaption



will be obtained with constant instantaneous volatilities (i.e., when the $g()$ function assumes the functional form $g(t, T) = d$).[17] In order to see whether the swaption prices can be at all reconciled with the caplet prices, we therefore explore in this section the case of flat $g()$ and $h()$ functions, while retaining the perfect correlation assumption. When these choices are made for the correlation and for the $g()$ and $h()$ functions, the pricing of swaptions can be profitably analyzed with reference to the pricing of early-stopping caplets, defined below.

Early-stopping caplets are the mirror image of forward-starting caplets and swaptions. Consider a swaption with expiry time T_{exp}, strike K, and n underlying forward rates. An early-stopping cap is made up of one "regular-stopping" caplet and $n - 1$ early-stopping caplets. The regular-stopping caplet has the first forward rate in the swap rate as underlier, the same strike as the swaption, and expiries at the expiry time of the swaption (which coincides with the expiry time of the first forward rate). The remaining $n - 1$ early-stopping caplets (all with the same strike) have as underliers the forward rates that expire at times $T_{exp} + i\tau$, i = 1, 2, …, $n - 1$, but with expiry at the expiry time of the swaption: They will expire "early" with respect to a "regular-stopping" caplet, which will still be "alive" by swaption expiry. An early-stopping caplet settles for payment at time $t_{exp_i} = T_{exp} < T_{exp} + i\tau$. At swaption expiry, all the early-stopping caplets will have only intrinsic value (no time value left).

We note in passing that early-stopping caplets are occasionally traded in the market. The main reason for taking regular caplets, rather than early-stopping caplets, as the starting point for the comparison with swaptions in our analysis is their much greater liquidity. We are not aware of early-stopping-caplets quoted beyond a few years. Certainly, they would not cover the forward rates underlying, say, a 10Y × 10Y swaption. We also find their liquidity (bid–offer spread) to be much poorer and virtually non-existent in meaningful size during periods of market turmoil. If early-stopping caplets (and, for that matter, forward starting caplets and/or swaptions) were liquidly traded alongside "regular" caplets and swaptions, Rebonato [2002] shows in detail that either a unique solution or no solution exists for the correlation and volatility within the resolution (three or six months) of the underlying forward rates.

Given the definition above, it is then trivial to prove the following. (The simple results presented in the following can be obtained by successive application of the triangular inequalities presented in Johnson and Nonas [2009]).

Proposition 1. *If the market price of a swaption expiring at time T_{exp} is greater than or equal to the price of a same-strike early-stopping cap made of the underlying early-stopping caplets all with expiry T_{exp}, then there is a model-independent arbitrage profit that can be extracted with a static super-replication strategy. The strategy to realize the arbitrage profit is to sell the swaption and buy the early-stopping cap.*

The proof is straightforward. By the assumption, setting up the strategy either costs nothing or provides a positive cash flow at time t_0. Then note that at expiry time, T_{exp}, there is no time value left either in the swaption or in the early-stopping cap. We therefore only have to look at the intrinsic value. Two situations can arise: either the yield curve is exactly flat, in which case the payoff from the swaption exactly equals the payoff from the early-stopping cap; or it is not flat, in which case the cap can have only the same or more value than the swaption (because the swap rate is a weighted average of forward rates and therefore some caplets can be in the money even if the weighted average is not).

Note that, since only intrinsic values matter, the result is model-independent.

Proposition 2. *If*

- *the instantaneous correlation between any two forward rates is equal to 1;*
- *the g() function is flat: $g(t, T) = d$;*
- *the h() function is flat: $h(t, T) = \delta$;*
- *the yield curve is flat; and*
- *the term structure of volatilities is flat, then any model nested within the set of the LMM-SABR no-arbitrage conditions produces the same price for a swaption and an early-stopping cap.*

The proof is also straightforward: if the yield curve is flat the swap rate is equal to any of the forward rates. Furthermore, with perfect instantaneous correlation, flat $g()$ and $h()$ functions, and flat term structure of volatilities, the volatility of the swap rate is identical to the volatility of the cap. The cap and the swaption must therefore have the same price.

Finally, we have the last result.

Proposition 3. *If any of the conditions in Proposition 2 are not met, the model-independent, arbitrage-free price of a swaption will always be lower than the price of the associated early-stopping cap.*

The result directly follows from Propositions 1 and 2 above.[18]

The simple results above simply show that an early-stopping cap must be worth at least as much as or more than the associated swaption, but they give no indication as to how much more expensive the cap will be when any of the conditions in Proposition 2 fail to be met. In practice, we observe that reasonable variations in the term structure of volatilities produce very small divergences between the prices of the swaption and the early-stopping cap. For similarly strong deviations from a flat yield curve, the discrepancies are found to be greater.

To see the relevance of these findings for swaption pricing, we focus on long-expiry swaptions. We note that for these long expiries, the forward portion of the term structure of rates (the portion that matters for the pricing of a swaption) can often be very flat. When this is true, it is then reasonable to expect that the price of an early-stopping cap will be very close to the price of the associated swaption. This is indeed borne out in practice in Exhibits 20 and 21, which show the prices of a swaption obtained by calibrating the forward rates to caplets with flat volatilities and perfect correlation, for both the 2Y × 2Y and the 10Y × 10Y swaption. Note that for the second series (where indeed we expect the forward yield curve to be flatter), the two curves are, most of the time, virtually indistinguishable. The maximum difference, reached during the extreme post-Lehman Brothers turmoil, remains smaller than one vega.

With these results in mind, we can now turn to the market and model prices of swaptions obtained with perfect correlation and flat $g()$ and the $h()$ functions that price all the caplets correctly. See Exhibits 23 and 24 for the two "corner" cases, that is, for the 2Y × 2Y and 10Y × 10Y European swaptions. Similar results were obtained for other expiries and tails.

The first observation is that we now obtain extremely close agreement between model and market prices of swaptions during normal and excited periods, for both long and short tails and long and short expiries. Recall, however, that we also obtained extremely close agreement between the perfect-correlation, flat-volatility swaption prices and the price of an early-stopping cap

EXHIBIT 20

The Prices of a 2Y x 2Y Perfect-Correlation Swaption with Flat Volatilities and the Price of an Early-Stopping (co-expiry) Cap, also Priced with Flat Volatilities

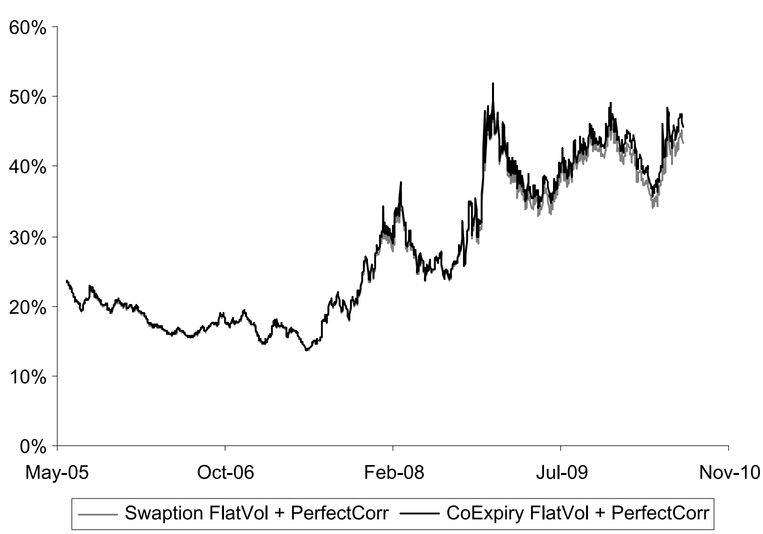
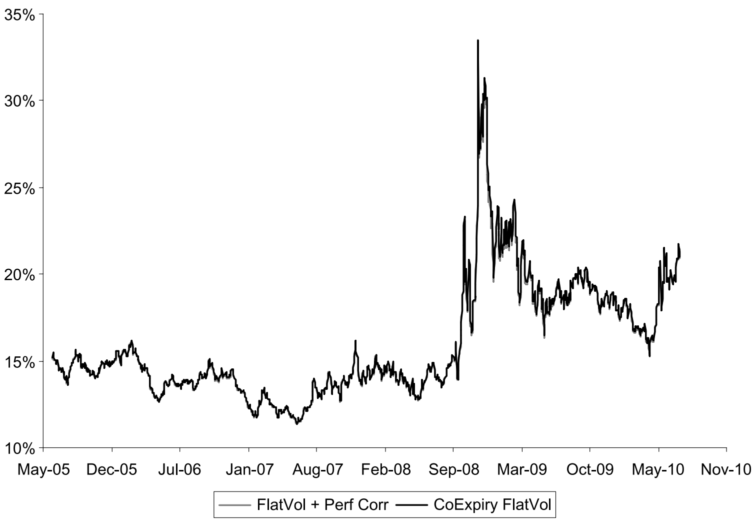


EXHIBIT 21

Same as Exhibit 20 for the 10Y x 10Y Swaptions



(which constitutes a model-independent no-arbitrage boundary). We therefore observe that within the class of the LMM-SABR arbitrage conditions, swaptions trade in the market *very close to the model-independent no-arbitrage value of the associated super-replicating early-stopping cap.*

IMPLICATIONS OF MARKET PRICING

We have provided a convincing story of how the market may be pricing swaptions. We have remarked that this pricing does not expose the trader to model-independent arbitrage. However, pricing swaptions *at* this boundary, as the market appears to be doing, has some unpleasant financial implications. This is most easily understood in terms of the specific parametrization of the LMM-SABR no-arbitrage conditions described above. Indeed, in order to obtain such a close agreement between the market and model prices while pricing caplets correctly, the correction factors, $\left\{k_0^{T_i}\right\}$, must now be very different from 1, indicating that time homogeneity is fully lost, in both normal and excited times.

The financial consequences of this loss of time homogeneity can be easily appreciated by looking at Exhibits 26 to 30, which display the future term structure of volatilities implied by the model calibrated to caplets with flat instantaneous volatilities. This calibration clearly implies a radical change of the term structure of volatilities toward shapes that have never been observed in the past, as can be seen by comparing Exhibits 26 to 30 with Exhibit 18, which show a real-world time series for the same quantities.

As long as today's prices are recovered, why should one worry about the realism of the future term structure of volatilities (and, more generally, the future smile surface)? The reason is that, indeed, the future evolution of the smile surface does not affect today's prices, but it will fully determine future prices of plain-vanilla

EXHIBIT 22

The Difference between the Time Series in Exhibit 21

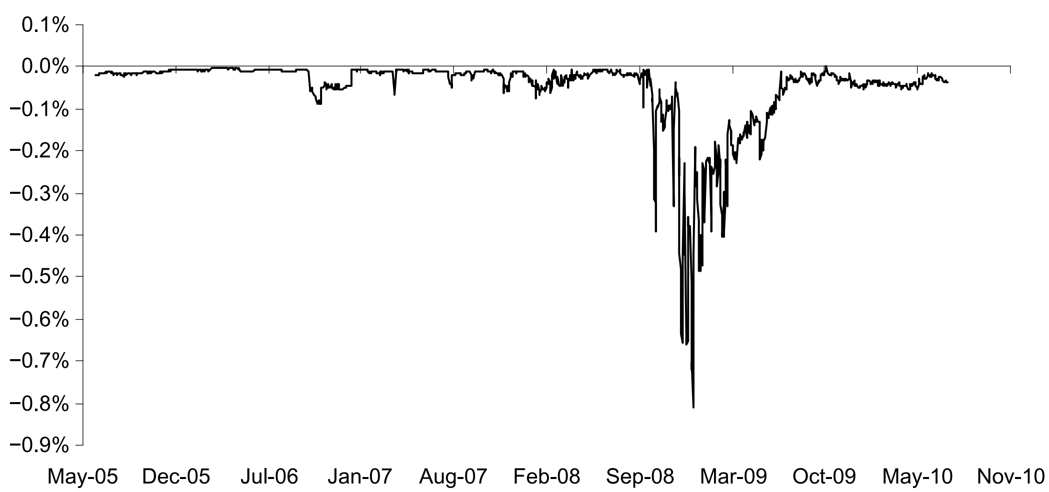


EXHIBIT 23
Market and Model Prices of 2Y x 2Y Swaptions Obtained with Perfect Correlation and Flat Volatilities

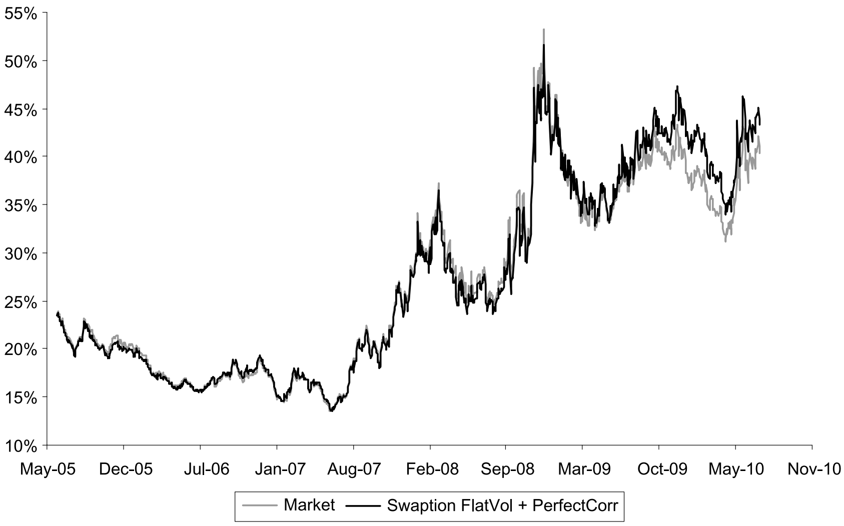
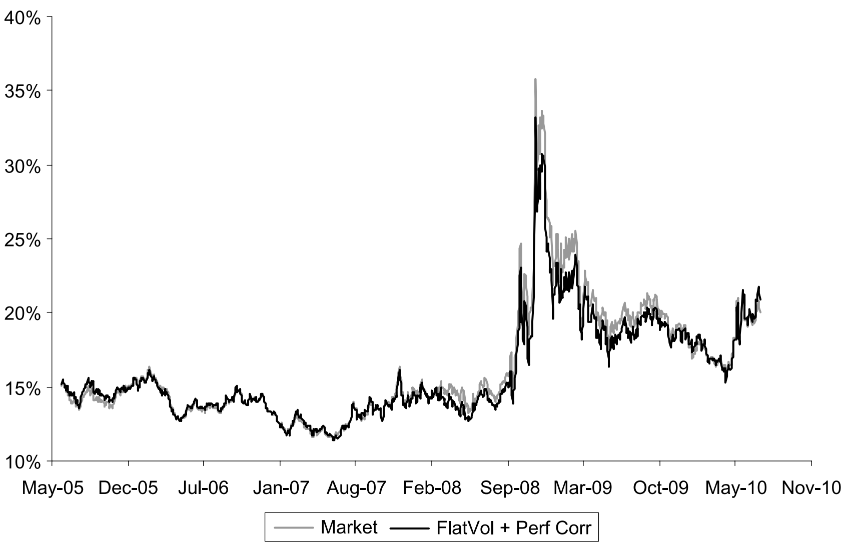


EXHIBIT 24
Same as Exhibit 23 for the 10Y x 10Y Swaptions

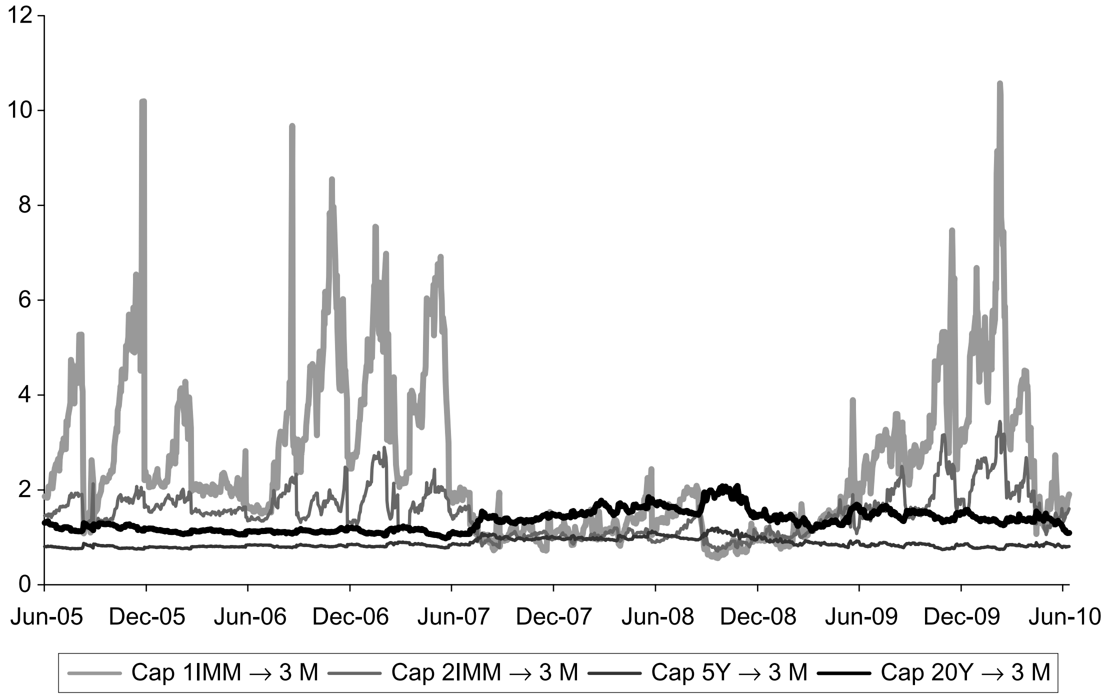


options and therefore affect future vega re-hedging costs.[19] It is therefore important to choose a calibration that will produce future implied volatility surfaces as close as possible to what will be realized in the future, as failure to do so would predict today incorrect future vega re-hedging costs. The effect is not small: a flat $g()$ function implies for the example shown in the Exhibits above that in two years' time, a trader will be able to sell a one-month expiry caplet, which today trades at an implied volatility of 40 basis points, for an implied volatility of approximately 110 basis points. According to the model, the same trader in fifteen years' time will also be able to buy for an 88 basis point implied volatility a caplet that now trades close to 120 basis points.

In sum: our empirical analysis shows that the market prices of swaptions trade close to a boundary value that is attained only if very special modeling conditions (about the instantaneous correlation and the shape of the $g()$ and $h()$ functions) are met. However, these modeling requirements imply an evolution of the term structure of volatilities that is totally at odds with what has been observed over many years (see Exhibit 18), which means that future re-hedging costs may be very different from what a trader is likely to encounter in reality.

Given this disconnect between statistically and financially justifiable prices on the one hand and market prices on the other, why don't swaptions trade even higher? We have pointed out that the flat-volatility swaption prices are often very close to the prices for an early-stopping cap. This occurs when the forward part of the yield curve is close to flat, as is often the case for long-expiry swaptions. We have also shown that in this case a model-independent static super-replicating strategy can, in theory, be put in place using early-stopping caplets. These are far from being liquid instruments, and swaption traders do not regularly use them in their hedging. However, traders are likely to keep these very easy-to-obtain boundary prices in the back of their minds as a "sanity check," beyond which swaptions become "too expensive." There is a market precedent for this: it is known, in fact, that in the case of other static replication strategies that *could* be put in place in practice (such as the replication using plain-vanilla calls

EXHIBIT 25
The Adjustment Factors $\left\{k_0^{T_i}\right\}$ Obtained with Flat Volatilities



*Note: This Exhibit should be compared with Exhibit 5, which shows the same adjustment factors obtained with a time-homogeneous **g**() function. Note that with flat volatilities, the adjustment factors* $\left\{k_0^{T_i}\right\}$ *are now one order of magnitude larger.*

EXHIBIT 26
Time-0 Term Structure of ATM Volatilities

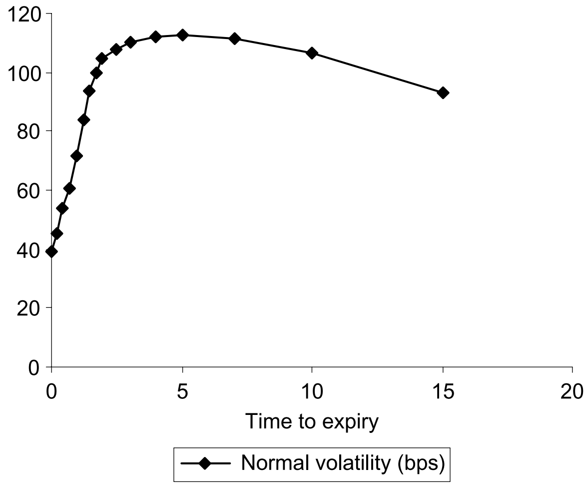


EXHIBIT 27
As in Exhibit 26, in 2 Years

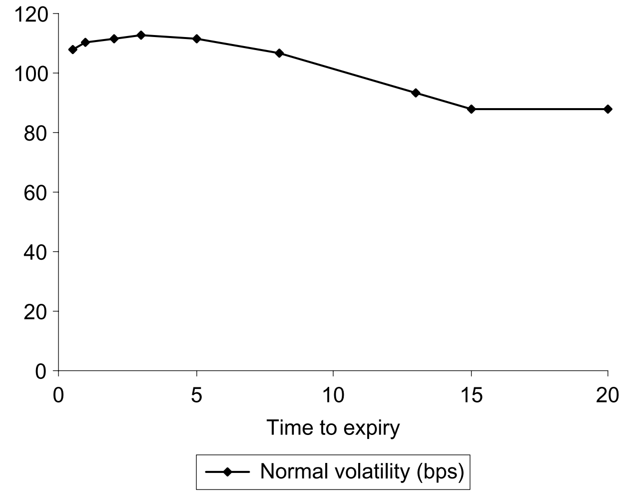


Exhibit 28
As in Exhibit 26, in 5 Years

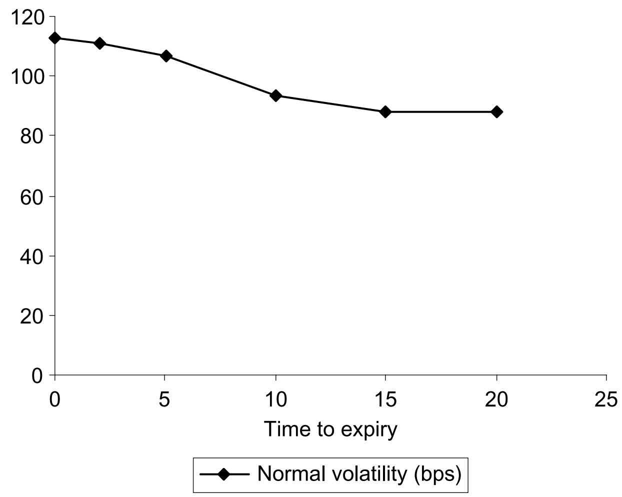


Exhibit 29
As in Exhibit 26, in 15 Years

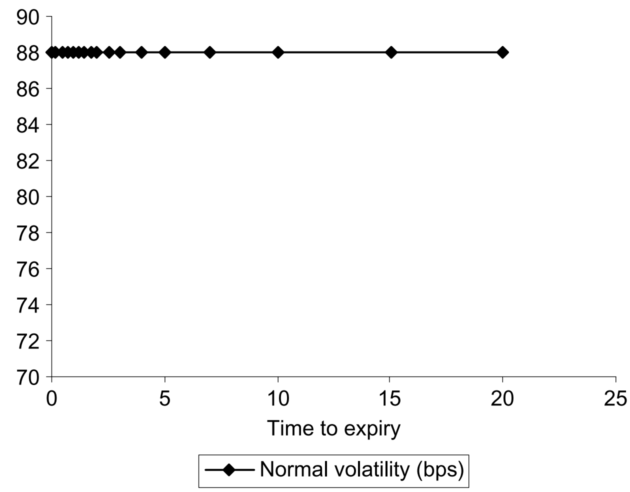
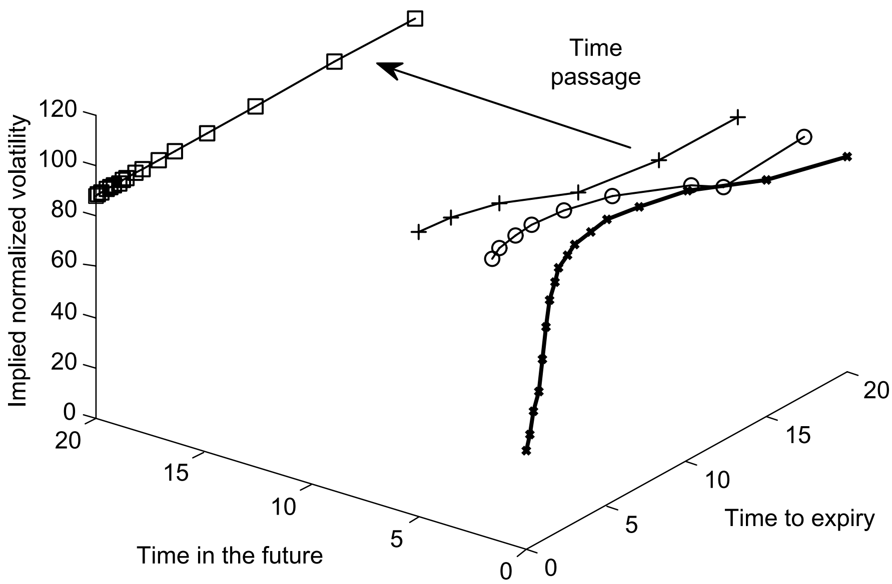


Exhibit 30
Evolution of the Term Structure of ATM Volatilities for the Case of Flat Instantaneous Volatilities



of a double-barrier knock-out option), the strategy is rarely actually deployed in practice; however, the replication construction is extensively used as an alternative way to obtain a price for the option[20] and to ensure that the trader is not exposed to model-independent arbitrage. Indeed, we have noted above that in the case of swaptions, from time to time, even the very special early-stopping-cap no-arbitrage boundary is exceeded, but when this happens, reversion to the boundary tends to be very swift.

To get a feel for the swiftness of the reversion to the model-independent no-arbitrage boundary, contrast Exhibits 10 to 15 and Exhibit 22. When the swaption model is calibrated in a financially plausible manner, reversion to fair pricing between swaptions and caplets takes years (if it occurs at all). Exhibit 22 shows instead

EXHIBIT 31
5Y x 5Y, June 1, 2007, 10K Simulations

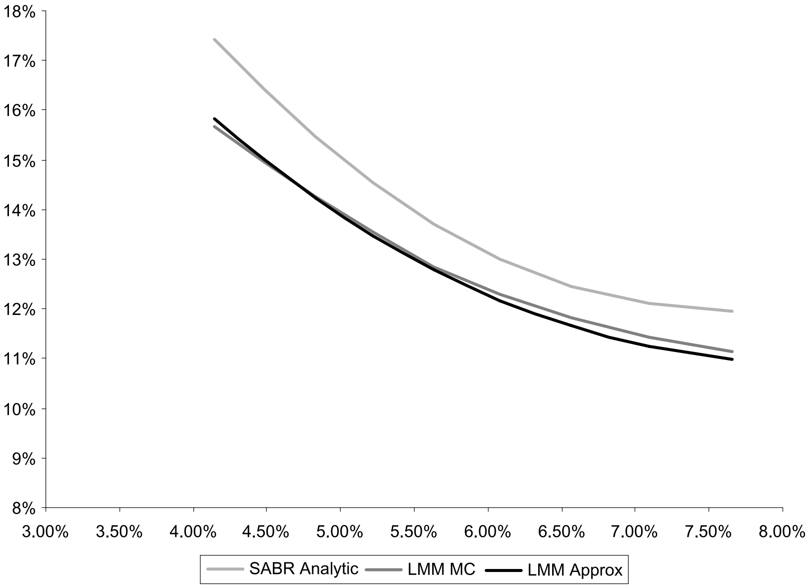
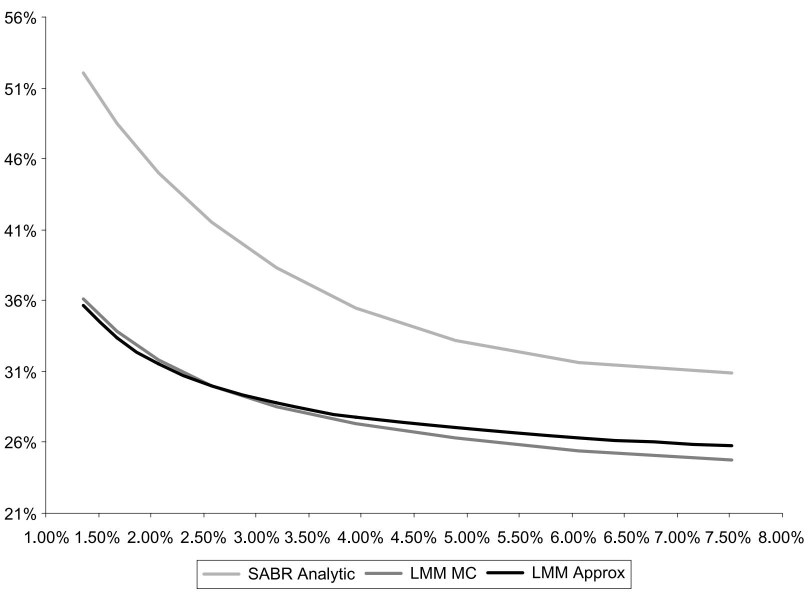


EXHIBIT 32
5Y x 5Y, Dec. 15, 2008, 50K Simulations



that all the way up to early 2008, reversion to the model-independent no-arbitrage boundary typically took days or weeks—and even the divergence from the same boundary that occurred during the Lehman Brothers crisis was reversed in a matter of a few months.

CONCLUSIONS

We have looked at the differences between the market prices and the model prices for U.S.-Dollar European swaptions obtained using the LMM-SABR model. The model was calibrated exactly to caplets by using both realistic humped instantaneous volatility functions and "flat volatilities." We also discussed and investigated the pricing impact of imposing a perfect or imperfect correlation among rates.

We found that during normal market conditions, the agreement between market prices and model prices obtained using the realistic volatilities (and perfect instantaneous correlation) was fair, especially for short expiries. The discrepancy becomes larger, however, during periods of market excitation and for larger expiries. We provided an explanation for this. We observed that introducing a more realistic correlation structure among forward rates made the agreement between market and model prices worse.

Better, and indeed excellent, agreement for all expiries and tails and during both normal and excited periods was obtained using flat volatilities (and, again, perfect instantaneous correlation). We have shown that these model prices are very close to the model-independent no-arbitrage boundary (that we have expressed in terms of early-stopping caps). A first conclusion is therefore that the market appears to trade close to the no-arbitrage boundary.

This conclusion raises some important questions, however. We have also shown in fact that the flat-volatility assumption required to obtain good swaption pricing produces very unrealistic evolutions for the future term structure of ATM volatilities. As the volatility and correlation "markets" are incomplete, this implies that the future re-hedging costs cannot be "locked-in." This raises the question as to why traders appear to endorse these unrealistic choices about instantaneous volatilities and correlations (and hence future term structures of volatilities):

EXHIBIT 33
5Y x 5Y, June 2, 2009, 50K Simulations

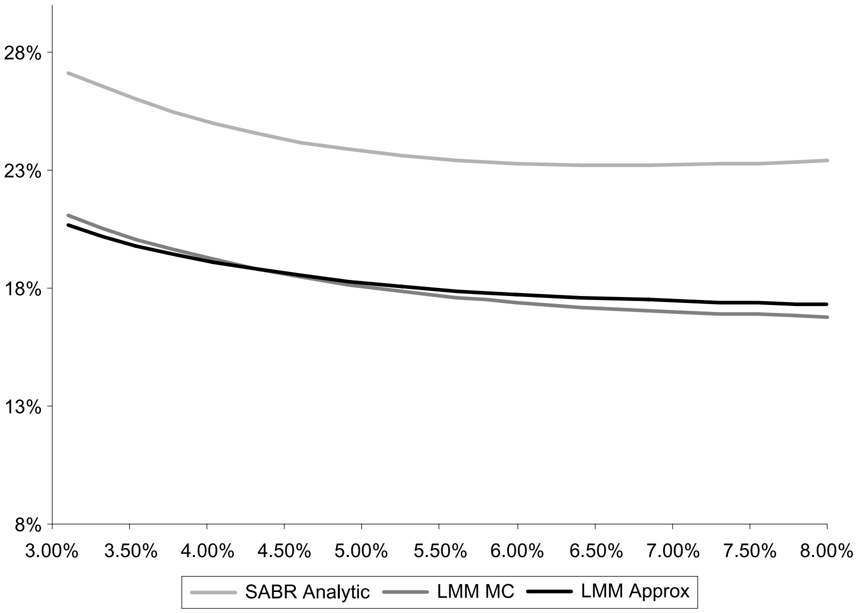
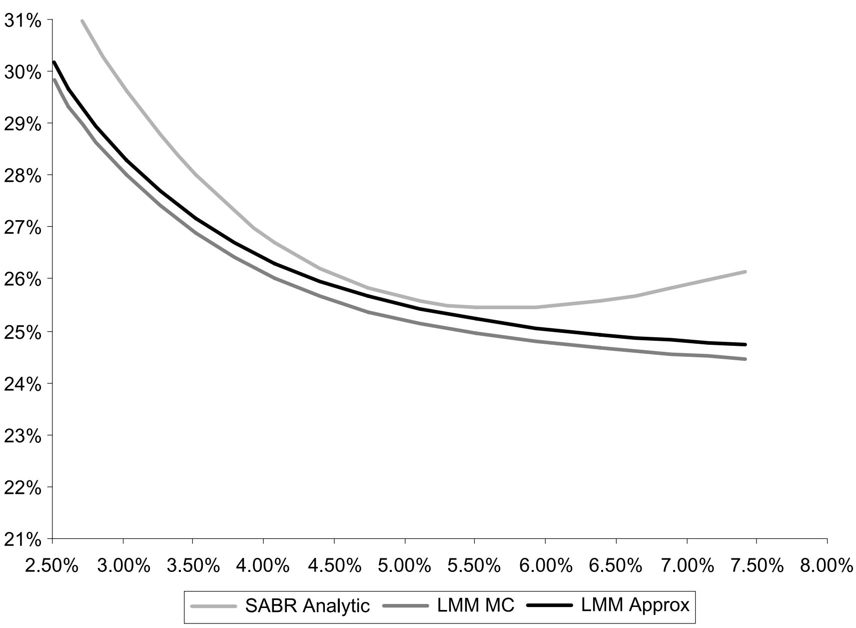


EXHIBIT 34
5Y x 5Y 1-Jul-2010, 50K Simulations



To the extent that these implicit model assumptions about the future smile surface are not borne out in practice, a trader would in fact expose himself to the risk of hedging losses.

The reason is probably to be found in the difficulty in carrying out this arbitrage: For instance, even after vega hedging, caplets and swaptions have very different gamma profiles. Rebonato [2004] shows that in order to mitigate these gamma mismatches, rather complex strangle-versus-straddle trades have to be carried out, which bring in a dependence on caplet and swaption smiles. And, of course, even if a trader were right, she would be exposed to the risk that before option expiry, the market may move even more out of line with "fundamentals," thereby causing a negative mark to market of her portfolio and forcing the trader to close her position. The persistence of this arbitrage would therefore be one instance of the limit-to-arbitrage phenomenon discussed in Shleifer [2000]. Finally, carrying out the pseudo-arbitrage required to bring caplets and swaptions in line with each other becomes particularly difficult in periods of market turmoil because of a variety of factors (poorer liquidity, balance-sheet constraints, reduced appetite for relative-value trades, etc.).

So, the data do not show any tendency for the market prices to converge to the prices that a financially appealing calibration (with imperfect correlation and non-flat volatilities) would suggest. On the contrary, market prices of swaptions at times even briefly exceed the price of a super-replicating portfolio of early-starting caplets. On the other hand, when swaption market prices exceed the no-arbitrage boundary discussed above, we have shown that the reversion to the no-arbitrage boundary becomes swift.

Finally, it is important to point out that the data in themselves cannot tell us whether swaptions are "expensive" or caplets, "cheap." For the U.S.-dollar market, however, a case can be made that there exists an excess demand for swaption optionality to hedge the negative convexity experienced by holders of fixed-rate mortgage-backed securities. The agencies (Fannie Mae and Freddie Mac) were well known in the market to be aggressive buyers of swaptions with expiries similar to the expected life of the mortgage-backed securities in their portfolios, and duration to match. After the 2007–2009 crisis, their activity may well have been substantially curtailed, but non-agency holders of mortgage-backed securities still face the same negative convexity problem.

At the same time, discussions with traders indicate that in the U.S.-dollar markets, liability managers are less inclined to buy interest-rate protection on their floating debt than in the case of European issuers. If true, this would reduce the demand for caps.

Note also that the divergence between the "fair model" and market values of swaptions became more pronounced with the onset of the financial crisis. This could be explained by the heightened desire to hedge negative convexity during high-volatility periods, coupled with the reluctance of pseudo-arbitrageurs to commit capital to relative-value "correcting" trades in periods of great turbulence.

All of this, of course, is highly qualitative and anecdotal, but the market intelligence referred to above suggests that, indeed, swaptions could indeed be "expensive" and caplets "cheap" (meaning that there appears to be a net excess demand for swaption optionality not matched by a similar demand for cap protection). It would be interesting to extend this study to the Euro swaption market. A study in this direction is underway.

APPENDIX

ACCURACY OF THE APPROXIMATIONS

As mentioned in the text, the values of the swaptions and caplets were obtained using the approximate analytic formulae in Rebonato and White [2010]. The Exhibits 31 to 34 show the accuracy of these approximations for a reference 5Y × 5Y swaption, for a number of Monte Carlo simulations between 10k and 50k. As found in Rebonato and White [2010], the accuracy did not vary significantly for other tails and expiries.

These Exhibits show the swaption prices obtained using the analytical formulae, a Monte Simulation and the Hagan et al. asymptotic expansion. It is easy to see that the formulae in Rebonato and White [2010] are always more accurate than the Hagan approximation, that in most cases they are accurate within a vega for all strikes, and that they are accurate to a small fraction of a vega for the ATM volatilities for which the bulk of the analysis has been conducted.

ENDNOTES

It is a pleasure to acknowledge the assistance provided by Dr. Jian Chen for numerical calculations and useful discussions, and by Dr. Doust for insightful comments.

[1]Swaptions in the following will always denote European swaptions, unless otherwise stated.

[2]Of course, the underliers of swaptions are, strictly speaking, swap rates. However, these are in turn made up of forward rates, so that these can be looked at as the ultimate underliers. For a given swaption, spanning forward rates co-terminal swap rates, or co-initial swap rates all constitute equivalent frames of reference. For the sake of intuition and for the technical reasons discussed in Rebonato [2004], we use forward rates as our primary variables.

[3]See, for example, Jamshidian [1997], p. 327.

[4]The reader looking for some pointers as to the nature of the model from the acronym would be disappointed: SABR stands for stochastic alpha beta rho, from the symbols used in Hagan et al.'s [2002] original paper. Alpha, in turn, is the volatility for the process of the underlying, beta the exponent of the CEV component of the process, and rho is the correlation between the Brownian increments for the underlying and the volatility.

[5]The equations are expressed in the measure under which the forward rate is a martingale. Expressions for the no-arbitrage drifts of the forward rates and the volatilities are reported in Rebonato, McKay and White [2009]. These drifts are not required in the present study.

[6]The adjective "stationary" is often used instead of "time-homogeneous."

[7]We note in passing that the time-homogeneity (stationarity) assumption is often made for instantaneous (rather than discrete) forward rates, in an HJM [1992] (rather than LMM) framework. Similarly, the humped volatility assumption is generally made for the volatility of instantaneous forward rates. Empirical tests, however (such as those quoted in Amin and Morton [1994], invariably require looking at discrete forward rates (or futures contracts).

[8]As it is well known, the assumption of joint lognormality cannot be correct for all the forward rates and the swap rate at the same time. However, Rebonato [1999] shows that the pricing implications of assuming joint lognormality for forward rates and swap rates are small (at least for at-the-money strikes).

[9]A brief remark on terminology and notation. A swaption expiring in n years' time to enter into a swap with N residual years to maturity is called an $n \times N$ year swap

(pronounced "*n* into *N*"). The length of the underlying swap is often referred to as the "tail."

[10]Note that the integration period does not in general coincide with the expiry of the underlying forward rate.

[11]The root-mean-squared volatility of the swaption will almost always be below the value, $\sigma^{SR_i}_{Black,\max}$, defined above for a short-dated option if the instantaneous volatility function is increasing and the option expiry is short.

[12]We note that the Schwartz inequality will always reduce the value of the covariance element and, hence, the swaption price. However, the apportioning effect could in principle have an indeterminate sign, depending on the specific form of the function $g()$.

[13]The super-correlation matrix that is made up of the forward-rate/forward-rate, volatility/volatility and forward-rate/volatility blocks required to specify the LMM-SABR model must be positive definite. The diagonal elements of the forward-rate/volatility block are constrained to be at the corresponding SABR market values for the corresponding caplets. See Rebonato, McKay, and White [2009] for a thorough discussion. Therefore, the closest correlation matrix must be found that satisfies the exogenous diagonal inputs and is as close as possible to 1 (in a specified norm) for the other blocks. The first of the two procedures in Chapter 7 of Rebonato, McKay, and White [2009] was used in this study for the purpose. The average correlation among forward rates turned out to be 98.8%, and the average correlation among volatilities, 98.0%.

[14]The study by White and Rebonato [2007] on the two-state Markov chain volatility model mentioned above indicates that the half-lives in the excited state typically range from a few weeks to several months.

[15]Time homogeneity of the instantaneous volatility function, of course, guarantees the self-similarity of the future term structure of volatility.

[16]To obtain a realistic correlation among forward rates, we have used the functional form discussed in Rebonato [2004], given by $\rho_{ij} = LongCorr + (1 - LongCorr)\exp[\beta(T_i - T_j)]$. See the discussion in Rebonato [2004] for a justification for this simple choice of correlation. We used $LongCorr = 0.6$ and $\beta = 0.076$ in order to reproduce approximately the empirical results reported in Rebonato [2004].

[17]The statement is true for most swaption expiries and for realistic shapes of the instantaneous volatility function. In theory, if the instantaneous volatility function was upward sloping, a swaption priced with such a volatility for its underlying forward *could* have a higher price than if the volatility was flat. One can see this by reversing the reasoning applied in the discussion of Exhibit 4. This can only occur, however, for very special combinations of expiries, lengths of tails, and shapes of the instantaneous volatility function.

[18]More constructively, the proof also follows from recognizing i) that a cap is a portfolio of options and a swaption is an option on a portfolio of forward rates; ii) that with less-than-perfect instantaneous correlation or with time-dependent volatilities the swap rate volatility will be lower than the cap volatility; and iii) that if the term structure of volatilities is not flat, the swap rate volatility will be lower than in the flat-term-structure-of-volatilities case. The price of the swaption must therefore be lower than the price of the early-stopping cap.

[19]In the absence of actively traded forward-starting European swaptions, the "volatility market" is incomplete, in the sense that it is impossible to "lock in" via traded hedging instruments the prices implied by an arbitrary set of correlation and instantaneous volatility functions (see Sidenius [1999]).

[20]Jaeckel and Rebonato [2000], for instance, show how to price an American option by replication in a way that is consistent with a variety of future possible smiles and compatible with today's price of European calls.

REFERENCES

Amin, K.I., and A. Morton. "Implied Volatility Functions in Arbitrage-Free Term Structure Models." *Journal of Financial Economics*, 35 (1994), pp. 141-180.

Brigo, D., and D. Mercurio. *Interest Rate Models—Theory and Practice*. Springer Verlag, Berlin (2001).

Goncalves, F., and J.V. Issler. "Estimating the Term Structure of Volatility and Fixed-Income Derivatives Pricing." *The Journal of Fixed Income*, Vol. 6, No. 1 (1996), pp. 32-29.

Hagan, P.S., D. Kumar, A.S. Lesniewski, and D.E. Woodward. *Managing Smile Risk Wilmott Magazine*, Autumn 2002.

Heath, D., R.A. Jarrow, and A. Morton. "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claim Valuation." *Econometrica*, 60 (1992), pp. 77-105.

Henry-Labordere, P. "Unifying the BGM and SABR Models: A Short Ride in Hyperbolic Geometry," Working paper, 2006, ssrn.com/abstract=877762.

——. "Combining the SABR and LMM models," *Risk*, October 2007.

Jaeckel, P., and R. Rebonato. "Valuing American Options in the Presence of User-Defined Smiles and Time-Dependent Volatility: Scenario Analysis, Model Stress and Lower-Bound Pricing Application." *Journal of Risk*, Vol. 4, No. 1 (2001), pp. 35-61.

Jamshidian, F. "LIBOR and Swap Market Models and Measures." *Finance and Stochastics*, Vol. 1, No. 4 (1997), pp. 293-330.

Johnson, S., and B. Nonas. "Arbitrage-Free Construction of the Swaption Cube," SSRN Working paper, Available at: SSRN: http://ssrn.com/abstract=1330869.

Mercurio, F., and J.M. Moraleda. "An Analytically Tractable Interest Rate Model with Humped Volatility." *European Journal of Operational Research*, 120 (2000), pp. 205-214.

Piterbarg, V. "A Stochastic Volatility Forward LIBOR Model with Time-Dependent Skew." *Applied Mathematical Finance*, Vol. 12, No. 2 (2003), pp. 147-185.

Rebonato, R. "On the Pricing Implications of the Joint Log-Normality Assumption for the Cap and Swaption Markets." *Journal of Computational Finance*, 3 (1999), pp. 57-76.

——. *Modern Pricing of Interest-Rate Derivatives*. Princeton and Oxford: Princeton, 2002.

——. "Which Process Gives Rise to the Observed Dependence of Swaption Implied Volatilities on the Underlying?" *International Journal of Theoretical and Applied Finance*, Vol. 6, No. 4 (2003), pp. 419-442.

——. *Volatility and Correlation in the Pricing of FX, Equity and Interest-Rate Options*, 2nd ed. Chichester: John Wiley, 2004.

——. "Forward-Rate Volatilities and the Swaption Matrix: Why Neither Time-Homogeneity Nor Time Dependence Will Do," *International Journal of Theoretical and Applied Finance*, Vol. 9, No. 5 (2006), pp. 705-746.

——. "A Time-Homogeneous, SABR-Consistent Extension of the LMM: Calibration and Numerical Results." *Risk*, December 2007.

Rebonato, R., and P. Jaeckel. "Linking Caplet and Swaption Volatilities in a BGM Framework: Approximate Solutions." *Journal of Computational Finance*, Vol. 6, No. 4 (Summer 2003), pp. 41-60.

Rebonato, R., and D. Kainth. "A Two-Regime Stochastic-Volatility Extension of the LIBOR Market Model." *International Journal of Theoretical and Applied Finance*, Vol. 7, No. 5 (2004), pp. 1-21.

Rebonato, R., K. McKay, and R. White. *The SABR/LIBOR Market Model*. Chichester: John Wiley, 2009.

Rebonato, R., A. Pogudin, and R. White. "Delta and Vega Hedging in the SABR and LMM-SABR Models." *Risk*, December 2008, pp. 94-99.

Rebonato, R., and R. White. "Linking Caplets and Swaption Prices in the LMM-SABR Model." *Journal of Computational Finance*, Vol. 13, No. 2 (Winter 2010), pp. 19-46.

Rebonato, R., N. de Guillaume, and A. Pogudin. "Dependence of Magnitude of Rate Moves on Rates Levels," Working paper, RBS, 2010.

Ritchken, P., and I. Chuang. "Interest Rate Option Pricing with Volatility Humps." *Review of Derivatives Research*, 3 (1999), pp. 237-262.

Sidenius, J. "LIBOR Market Models in Practice." *Journal of Computational Finance*, 3 (Spring 1999), pp. 75-99.

Shleifer, A. "Inefficient Markets—An Introduction to Behavioural Finance." *Clarendon Lectures in Economics*, Oxford: Oxford University Press, 2000.

White, R., and R. Rebonato. "A Swaption Volatility Model Using Markov Regime Switching." *Journal of Computational Finance*, Vol. 12, No. 1 (2007), pp. 79-114.