Consistent Calibration of HJM Models to Cap Implied Volatilities

Flavio Angelini Stefano Herzel

University of Perugia

Abstract

This paper proposes a calibration algorithm that fits multi-factor Gaussian models to the implied volatilities of caps using the respective minimal consistent family to infer the forward rate curve. The algorithm is applied to three forward rate volatility structures and their combination to form two-factor models. The efficiency of the consistent calibration is evaluated through comparisons with non-consistent methods. The selection of the number of factors and of the volatility functions is supported by a Principal Component Analysis. Models are evaluated in terms of in-sample and out-of-sample data fitting as well as stability of parameter estimates. The results are analyzed mainly focusing on the capability of fitting the market implied volatility curve and, in particular, of reproducing its characteristic humped shape.

Keywords: HJM models, forward rates, consistent families, calibration, humped volatility
JEL Classification: E43, C13
Mathematics Subject Classification (1991): 60H30

1 Introduction

The framework proposed by Heath, Jarrow and Morton (1992) for modeling forward rates leaves the user with a wide choice of alternatives, some fundamental, others more subtle. Fundamental is the choice of the number of driving risk factors as well as the shape of the volatility functions, but also important is the selection of the construction criteria for the initial curve and of the model calibration procedure.

The main contribution of this paper is to propose a calibration algorithm that fits multi-factor Gaussian models to the implied volatilities of caps using the respective minimal consistent family to infer the initial forward rate curve. The paper employs one-factor models with from one to three parameters and their combinations to get models with more factors.

The use of models with more than one factor was already proposed in the original paper by Heath et al., motivated by the standard approach of Principal

Component Analysis. Two components are usually considered enough to model the variance of the forward rates, although it depends on how long maturities are included in the sample. Amin and Morton (1994), in one of the first studies on empirical aspects of HJM models, observed that one factor is enough if only short maturities are considered. They compared the performances of some popular one-factor model by measuring the stability of parameter estimates, the fitting capabilities and the ability to exploit perceived mispricing. An important conclusion of their investigation was that "the number of the parameters used in the model has a strong effect on the behavior of the model, in fact stronger than the form of the model used". Driessen et al. (2003) performed an empirical investigation of hedging capabilities of models with up to three factors, with the motivation that more than one factor may be necessary when also swaptions are considered. Their results do not show a clear difference between two and three-factors, as far as prediction on prices is concerned. Brace and Musiela (1994) used a model with two factors so that it would be "capable of generating as many different yield curves as possible". Their motivation was that one is able to hedge just within the class of yield curves that can be "reached" by the model. This idea was later raised again, but in a slightly different and more general way by Björk and Christensen (1999). They pointed out that the set of curves produced depends both on the model and on the initial curve chosen and defined the concept of "consistency" between a family of forward curves and a model, providing appropriate consistency conditions. Their work has been extended by Filipović (2001), who developed a general theory of consistency for HJM models. From a practical point of view, it has to be remarked that only yields for a small set of maturities can be inferred from market data, hence the curve required by the HJM framework is arbitrarily chosen by the user. Using a consistent family is a way of limiting the arbitrariness of such a choice. From another perspective, the consistent calibration algorithm proposed in the present paper may be interpreted as a way to integrate the information extracted from the derivatives into the estimation of the yield curve.

The crucial point of a consistent calibration is that the parameters of the model are determined at the same time as the initial curve. This is different from the usual cross-sectional approach where the initial curve is estimated separately and then used to determine the parameters. For a given model the algorithm identifies the corresponding minimal consistent family and find the optimal fit to the term structure of cap volatilities and of interest rates. To this scope, a general result on consistent families for multi-factor models is proven.

Costs and benefits of the consistent algorithm are evaluated through a comparison with two alternative approaches: the first one entails a non-consistent calibration with a consistent family, and may be considered as a specification test; the second one is the usual cross-sectional calibration adopting the popular parametric family proposed by Nelson and Siegel (1987) to fit the forward rate curve. This extends the analysis of Angelini and Herzel (2002) who considered the case of the extended Vasicek model introduced by Hull and White (1990) and showed that the shape of the initial curve influences the performance of the model and that consistency plays a relevant role.

It is market practice to quote caps in terms of implied volatilities; prices are then recovered from the standard Black formula. The market implied volatility term structure is usually observed to be either humped or decreasing. An important empirical issue is to check whether a model is able to reproduce such features.

A comparative study of the performances of different models on a data set consisting of yields and cap volatilities from the Euro-market from 25-2-2001 to 4-7-2001 is presented. Three forward rate volatility structures, and their combinations, are selected for the analysis and only up to two-factor models are considered: the continuous time version of the model proposed by Ho and Lee (1986), with one parameter, the two-parameter model introduced by Hull and White (1990) and the three-parameter model defined independently by Mercurio and Moraleda (1996) and Ritchken and Chuang (1999). Models are evaluated in terms of in-sample and out-of-sample data fitting as well as stability of parameter estimates. The results are analysed mainly focusing on the capability of fitting the market implied volatility curve and, in particular, of reproducing its characteristic humped shape. As a matter of fact, models' performances differ mostly on periods of humped volatility. The analysis identifies a two-factor model that outperforms the others and that, to our knowledge, was never considered in previous empirical studies.

A Principal Component Analysis (PCA) is applied on the time series of interest rate yields similarly to Bühler et al. (1999) and Driessen et al. (2003), among others. Since the approach proposed here is cross-sectional, the PCA is performed only to get an insight on the number of factors and on the shape of the volatility functions and not as a calibration tool. This analysis shows that the principal loading factors can be well fitted by some of the models selected. Moreover, some of the parameter estimated on a cross-sectional basis are close to the corresponding estimates evaluated through the PCA, suggesting further evidence of the good properties of data description.

The paper is structured as follows: Section 2 gives a brief overview of the HJM framework and presents the models analyzed. Section 3 contains the Principal Component Analysis of the data considered. Section 4 is devoted to the description of the concept of consistent families, together with some general results and the computation of all the families needed in the calibration procedures. In Section 5 the calibration procedure is described in detail. Section 6 is devoted to empirical results, first comparing the consistent calibration algorithm to the non-consistent approaches, then presenting the results of the fitting of the different models. The last section draws the conclusions and gives final comments.

2 The Models

In their seminal paper, Heath, Jarrow and Morton (1992) introduced a class of interest rate models based on the dynamics of instantaneous forward rates. The instantaneous forward rate (hereafter forward rate) contracted at time t and with time to maturity x is defined as

$$f(t,x) = -\frac{\partial \log P(t,t+x)}{\partial x},$$

where $P(t,t+x)$ is the discount factor, i.e. the price at time t of a zero coupon bond with time to maturity x. The evolution of forward rates is modeled, under the martingale measure, by the infinite dimensional diffusion process driven by the d-dimensional Brownian motion W

$$\begin{cases} df(t,x) = \beta(t,x)dt + \sigma(t,x)\cdot dW \\ f(0,x) = f_0(x) \end{cases},$$

where $\sigma(t,x)$ is a d-dimensional process and $\cdot$ denotes scalar product. The initial curve $\{f_0(x),\ x \geq 0\}$ is interpreted as the observed forward rate curve.

The process $\beta(t,x)$ is completely determined by the HJM no-arbitrage condition,

$$\beta(t,x) = \frac{\partial}{\partial x}f(t,x) + \sigma(t,x)\cdot\int_0^x \sigma(t,u)du,$$

as shown in Brace and Musiela (1994). Hence a particular model in this class is specified by the choice of a volatility structure.

This paper is concerned with Gaussian models where $\sigma(t,x) = \sigma(x)$ is a deterministic function depending only on time to maturity, hence $f(t,x)$ is a Gaussian process. The most popular models in this class are the continuous time version of the Ho-Lee model (Ho-Lee (1986)), denoted by HL, and the extension of the Vasicek model (Vasicek (1977)) introduced by Hull and White (1990), denoted by HW. These are one-factor models with volatility structures given respectively by $\sigma_{HL}(x) = \sigma_0$ and $\sigma_{HW}(x) = \sigma_1 e^{-ax}$. More recently, Mercurio and Moraleda (1996) and Ritchen and Chuang (1999), proposed a one-factor model (henceforth MM) with humped volatility of the forward rate, namely $\sigma_{MM}(x) = (\sigma_2 + \gamma x)e^{-bx}$.

Such volatility structures can be combined to form two-factor models. For instance, a two-factor model where the first factor is given by the Ho-Lee structure and the second by the Hull-White structure was introduced by Heath et al. (1992). It will be denoted by HL-HW and analogous notation will be used for the other models considered.

An important feature of Gaussian models is analytical tractability, as they provide closed formula for price of options, since bond prices are log-normal (see for instance Björk (1998), Prop. 19.18). A drawback of these models is that they allow for negative forward rates.

This paper will consider models with at most two factors. Usually, a factor analysis of the yield curve based on historical data, suggests that two components already explain a high percentage of the variation in the curve.

3 Factors and Volatilities

A standard approach to find the driven factors of a set of sources of uncertainty is the Principal Component Analysis. Rebonato (1998), to which we refer for details on the matter, describes the method in the case of interest rate models and gives some results on market data. Rebonato states that three factors usually explain 95-99 % of the entire variability (in terms of variance) of the sample, and that the first one accounts for up to 80-90 %. The first component is fairly constant and it is interpreted as the "average level" or "shift" of the curve, the second component has opposite signs at the opposite ends of the maturity spectrum and is considered as the "twist", the third one has a natural interpretation as the "curvature" and it is called "butterfly". Indeed, this behaviour is typical of any highly correlated set of variables, not only on interest rate data. In particular Rebonato shows a study for the UK market in the years 1989-1992 which confirms, in an optimistic way, the general results stated above, namely 92% of the variance explained by the first factor and 99.1% by the second factor.

The analysis can be used not only to determine the number of factors that explain in a satisfactory way the data, but also to identify the shape of the volatility functions. As already suggested by Heath et al. (1992) this can be pushed even further to calibrate the model to a set of yield curve data. Several approaches have been proposed, see for instance Bühler et al. (1999) and Driessen at al. (2003). Here we impose a parametrical form to the loading factors and estimate the parameters via nonlinear regression.

Although the method adopted here to calibrate a HJM model is based on cross-sectional data, namely prices of derivatives, an analysis on historical data is also performed for different reasons. First to justify, on our data set, the use of models with no more than two factors. Second, to get a first insight on the shape of the volatility functions to be employed for the cross-sectional calibration. Finally, to compare the parameter estimates on historical data with those on cross sections.

The data set, provided by Datastream, consists of a time series of daily discount factors with thirteen different maturities (3,6,9 months and from 1 to 10 years) from 28/5/99 to 4/7/01 for a total of 549 days. Instantaneous forward rates are not observable and it is common to use as a proxy the continuously compounded forward rate from $t + x$ to $t + x + \tau$

$$F(t, t + x, t + x + \tau) = -\frac{\log P(t, t + x + \tau) - \log P(t, t + x)}{\tau}, \quad (1)$$

but this choice is not satisfactory when τ is large, because this approximation of the derivative may be fairly inaccurate. Another possibility is to interpolate the term structure with a smooth curve and then differentiate to recover the instantaneous forward rate. This approach, used also by Bühler et al. (1999) and by Driessen et al. (2003) has the drawback of introducing spurious data.

Instead, we let

$$Y(t, x) = \log P(t, t + x);$$

it is well known (see for instance Björk (1998), Prop. 15.5) that, if the forward rate follows

$$df(t, x) = \beta(t, x)dt + \sigma(t, x) \cdot dW,$$

then

$$dY(t,x) = B(t,x)dt + S(t,x) \cdot dW,$$

where $S(t,x) = -\int_0^x \sigma(t,u)du$. Hence, if $\sigma(t,x)$ depends only on x, so does $S(t,x)$; moreover, we assume that the same holds for $B(t,x)$, so that the stationarity, which is necessary for the PCA, holds. The functions $S(x)$ for the HL, HW and MM models are respectively

$$\begin{aligned} S_{HL}(x) &= -\sigma_0 x, \\ S_{HW}(x) &= \frac{\sigma_1}{a}\left(e^{-ax} - 1\right), \\ S_{MM}(x) &= (A + Bx)e^{-bx} - A, \end{aligned}$$

where $A = \frac{\sigma_2 b + \gamma}{b^2}$ and $B = \frac{\gamma}{b}$.

To perform the component analysis let us set the time lag to one day and compute

$$\Delta Y(t_j, x_i) = Y(t_{j+1}, x_i) - Y(t_j, x_i)$$

for $j = 1, \ldots, 549$ and $i = 1, \ldots, 13$. At this point the procedure is standard: the sample covariance matrix C is computed on the time series of the 13 variables $\Delta Y(t_j, x_i)$ and decomposed as

$$C = V \Lambda V',$$

where V is an orthogonal matrix, V' its transpose and Λ a diagonal matrix. The elements of Λ, i.e. the eigenvalues of C, correspond to the variance of each factor and the columns of V represent the loading factors, i.e. the volatility function of $Y(t,x)$. The first factor explains 91.4% of the total variance, the first two 97.2% and the first three 99%.

The analysis suggests that most of the variance of the process can be modeled by two loading factors, whose shapes are exhibited in Figure 1. The first component is linear, hence it is well fitted by the HL volatility function S_{HL} (full line). The corresponding estimate for σ_0 is 0.0067. As for the second component, S_{MM} fits in a very satisfactory way. The non-linear regression problem has many local solutions. However, most of the solutions provide a good fitting. In Figure 1 we show one of the fittings, with parameters $\sigma_2 = 0.01$, $\gamma = -0.0062$ and $b = 0.32$. These are obtained from the estimates for A, B and b inverting formulas above. Instead, S_{HW} gives a very poor fitting of the second component, as it is clear given that its first derivative has constant sign.

4 Consistent Initial Curves

Every day, financial institutions produce points of the yield curve using quite a standard technique, starting from some input market data as short term interest rates, future prices and swap rates. The main problem is that only rates with few maturities can be bootstrapped directly from the market, typically quite enough for short terms but only with one-year lag for maturities longer than one or two

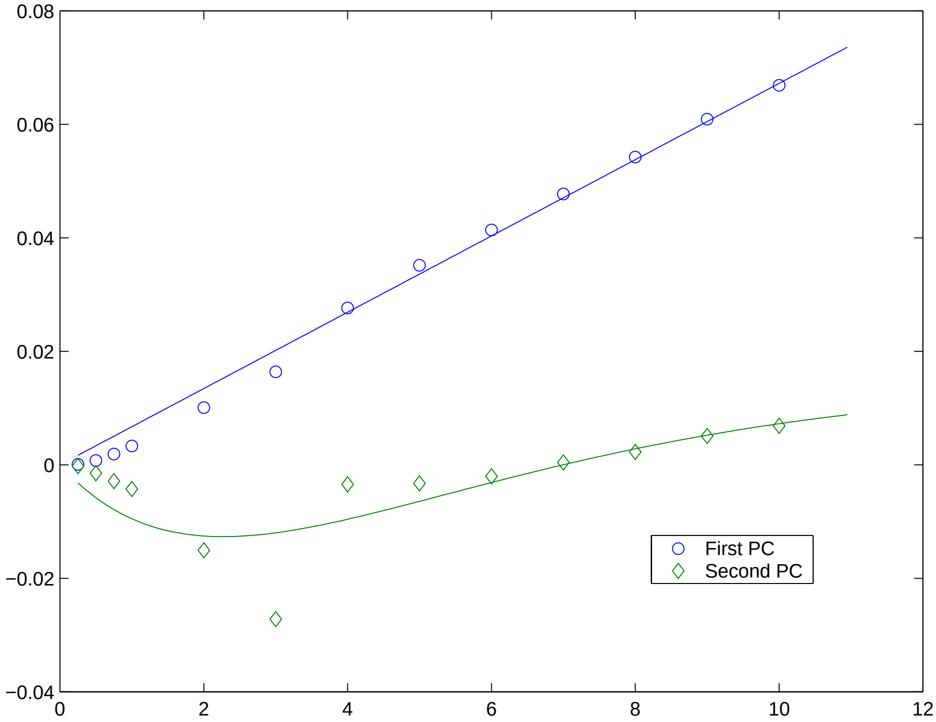


Figure 1: Principal Component Analysis on $Y(t, x) = \log P(t, t + x)$: the first component is fitted with S_{HL} ($\sigma_0 = 0.0067$), the second component with S_{MM} ($\sigma_2 = 0.01$, $\gamma = -0.0062$ and $b = 0.32$).

years. Already dealing with some basic derivative contracts such as Caps, this may be not enough, as rates on a finer time grid may be needed. Therefore, one is forced to "interpolate" between maturities. A very popular solution to the problem is to use linear interpolation. The objection is immediate: the curve would not be differentiable. It may be replied that this is not a serious issue in practice or another interpolating method can be chosen, like cubic splines, or some family of approximating curves, like that proposed by Nelson and Siegel (1987). In the latter case one drawback is that the curve does not pass exactly for all the points.

Up to this point the problem appears merely practical. However, in the case of HJM models, a theoretical issue comes out, which may also have practical implications. The model needs a whole curve to be initialized and, again, one has to fit a family of curves to the observed discrete term structure. Then the model, through its equation, produces future forward curves. It seems natural that the curve used for the interpolation should be reachable by the model.

A coherent solution to this is to use a family of curves such that the equation will generate, starting from a curve of that family, curves remaining within the family. This is the concept of consistency. An HJM model $\mathcal{M}$ and a family $\mathcal{G}$ of curves are consistent if all forward curves produced by $\mathcal{M}$ belong to $\mathcal{G}$. Given a n-dimensional parameterized family $G(z,x)$, where x is the real variable and z belongs to some parameter space $Z \in \Re^n$, $\mathcal{G}$ is defined as

$$\mathcal{G} = \{G(z,\cdot) : \; z \in Z\}.$$

Of course, for a model $\mathcal{M}$, the consistent family is not unique. A consistent family is minimal if it is contained in any other family which is consistent.

An important point related to consistent families is that they represent finite dimensional realizations of a given model, meaning that the dynamics of the entire forward rate curve can be recovered from the dynamics of the variables in the parameter space Z of the family. For more details on the topic we refer to Björk (2001) or Björk and Landen (2002). In the latter, it is also shown how to construct such realizations. This interpretation has a relevant practical advantage in terms of simulation of the term structure at a future time, since one only needs to simulate the finite dimensional process followed by z. This will substantially improve the efficiency of any Monte-Carlo procedure. The calibration algorithm described in Section 5 may be employed as an initial step to determine the current values of the z's.

The present paper uses this consistent approach. It is therefore necessary to determine consistent families for HJM Gaussian models. The main reference on the topic is the paper by Björk and Christensen (1999) (for more general discussions see Björk (2001) and Filipović (2001)). The main theorem (Theorem 4.1 in Björk and Christensen (1999)), stated in the Gaussian case and without technical details, is reported for convenience of the reader.

Theorem 4.1. *The family $\mathcal{G}$ is consistent with the model $\mathcal{M}$ if and only if*

$$G_x(z,x) + \sigma(x) \cdot \int_0^x \sigma(u)' du \quad \in \quad Im[G_z(z,x)] \tag{2}$$

$$\sigma(x) \quad \in \quad Im[G_z(z,x)] \tag{3}$$

for all $z \in Z$.

$G_x(z,x)$ stands for partial derivative with respect to x and $Im[G_z(z,x)]$ is the tangent space to $\mathcal{G}$.

Condition 3 is interpreted componentwise for σ. Condition 2 is called *the consistent drift condition* and condition 3 is called *the consistent volatility condition.* Note that, in the general case, there is an extra term in the left hand side of the consistent drift condition.

Although stated in quite abstract terms, the theorem is fairly easy to apply in concrete cases. Alternatively, for the one-factor models mentioned in Section 2, Proposition 7.2 and Proposition 7.3 in Björk and Christensen (1999) may be directly applied, without going through any computation, to get the following results:

1. the family

$$G_{HL}(z,x) = z_1 + z_2 x$$

is consistent with the HL model;

2. the family

$$G_{HW}(z,x) = z_1 e^{-ax} + z_2 e^{-2ax}$$

is consistent with the HW model;

3. the family

$$G_{MM}(z,x) = (z_1 + z_2 x)e^{-bx} + (z_3 + z_4 x + z_5 x^2)e^{-2bx}$$

is consistent with the MM model.

It is easy to check that the above families are minimal for the corresponding model.

All the applications and examples contained in Björk and Christensen (1999) refer to one-factor models, apart from the last section where the HL-HW model is taken into consideration. For the purpose of the paper, where consistent families for multi-factor models are needed, the following proposition is useful.

Proposition 4.2. *Let the model $\mathcal{M}$ be given by*

$$\sigma(x) = \begin{pmatrix} \sigma_1(x), & \sigma_2(x), & \ldots, & \sigma_d(x) \end{pmatrix}$$

and let $G_i(z^i, x)$, $i = 1, \ldots, d$, be families of curves consistent respectively with the one-factor model with volatility function $\sigma_i(x)$. Then the family

$$G(z^1, \ldots, z^d, x) = G_1(z^1, x) + \ldots + G_d(z^d, x)$$

is consistent with $\mathcal{M}$.

Proof. From Theorem 4.1, it is enough to check the consistent drift condition and the consistent volatility condition. The tangent space to the family $G(z^1, \ldots, z^d, x)$ is

$$Im[G_{(z^1,\ldots,z^d)}(z^1, \ldots, z^d, x)] = < Im[G_{1z^1}(z^1, x)], \ldots, Im[G_{dz^d}(z^d, x)] >,$$

that is the vector space generated by the tangent spaces to each family. Therefore, since each σ_i belongs to $Im[G_{iz^i}(z^i, x)]$, the volatility condition is satisfied. It is also easy to see that the drift condition is satisfied because

$$\sigma(x) \cdot \int_0^x \sigma(y)' dy := \sum_{i=1}^{d} \sigma_i(x) \int_0^x \sigma_i(y) dy$$

by definition and

$$G_x(z^1, \ldots, z^d, x) = \sum_{i=1}^{d} G_{ix}(z^i, x).$$

This concludes the proof. □

Notice that, via a re-parametrization, it may be possible to reduce the number of parameters of the family $G(z^1, \ldots, z^d, x)$. For instance, for the HL-HL model the family would be $G(z^1, \ldots, z^d, x) = z_1 + z_2 x + z_3 + z_4 x$, which is clearly the same as $z'_1 + z'_2 x$. Of course the HL-HL model is equivalent to the one-factor HL model, but this may also happen in other situations.

Proposition 4.2 and the results for one-factor models provide more examples of families which are consistent with two-factor models:

1. the family

$$G_{HL-HW}(z, x) = z_1 + z_2 x + z_3 e^{-ax} + z_4 e^{-2ax}$$

   is consistent with the HL-HW model;

2. the family

$$G_{HW-HW}(z, x) = z_1 e^{-a_1 x} + z_2 e^{-2a_1 x} + z_3 e^{-a_2 x} + z_4 e^{-2a_2 x}$$

   is consistent with the HW-HW model;

3. the family

$$G_{HL-MM}(z, x) = (z_1 + z_2 x) + (z_3 + z_4 x)e^{-bx} + (z_5 + z_6 x + z_7 x^2)e^{-2bx}$$

   is consistent with the HL-MM model.

These results may also be proven by using Theorem 4.1. Not only does Proposition 4.2 provide a shortcut to the particular results we need but also is quite useful, as it will appear in what follows, to formulate a consistent algorithm for the fitting of a multi-factor model to the term structures of caps and bonds.

Notice that some of the aforementioned families are minimal for one particular model, but could still be consistent with others. This is the case, for instance, of the family $G_{MM}(z, x)$ which is also consistent with the HW model or the case of $G_{HL-MM}(z, x)$ which is also consistent with the HL, HW, MM and HL-HW models. Hence, if the family $G_{HL-MM}(z, x)$ fits well the observed forward curves, but one would like to use a more simple model, with fewer factors or parameters, than the HL-MM model, the HL, HW, MM and HL-HW models still represent consistent choices.

5 The Calibration Algorithm

Each day the input data are: a vector $P(\bar{x})$ of prices of zero coupon bond and a term structure v_T of implied volatilities of at-the-money interest rate caps, where $\bar{x} = (x_1, \ldots, x_n)$ and $T = (T_1, \ldots, T_m)$ are vectors of maturities. A cap is said to be at-the-money when its price equals that of the corresponding floor, i.e. when the strike is the forward swap rate, that is defined in terms of $P(x)$ (Definition 1.5.3 in Brigo and Mercurio (2001) to which we refer for details on caps). The market quotes caps in terms of implied volatilities, recovering the actual prices using (a linear combination of) Black's formula. Let $Cap^{Bl}(T_i, P(x), v_{T_i})$ be the Black's formula for the price at time 0 of an at-the-money T_i-cap, where $P(x)$ is the entire curve of discount factors. Of course the formula should also depend upon other contractual characteristics like reset times, but they are skipped for simplicity.

Let $\sigma(x;\theta)$ be the volatility function of the Gaussian model $\mathcal{M}$, where θ is a vector of parameters. The objective is to find parameters to fit the market implied volatilities. Let $Cap^{\mathcal{M}}(T_i, P(x);\theta)$ be the model formula for the price of an at-the-money T_i-cap. Remind that, for a Gaussian model, this is a closed formula because caps are portfolios of options on bonds. Again, the dependency on other contractual characteristics is dropped.

For every maturity T_i, we want to minimize

$$\log Cap^{\mathcal{M}}(T_i, P(x);\theta) - \log Cap^{Bl}(T_i, P(x), v_{T_i}), \tag{4}$$

which is similar to minimize relative errors. For a given vector of parameters θ, let $v_{T_i}^{\mathcal{M}}(\theta)$ be the model T_i-cap implied volatility, that is the solution $v_{T_i}^{\mathcal{M}}$ of the equation

$$Cap^{Bl}(T_i, P(x), v_{T_i}^{\mathcal{M}}) = Cap^{\mathcal{M}}(T_i, P(x);\theta). \tag{5}$$

Observe that Black's formula for at-the-money options is approximately proportional to the volatility, and this also holds for caps, since an at-the-money cap is a portfolio of nearly at-the-money options; namely

$$Cap^{Bl}(T_i, P(x), v_{T_i}) \approx \alpha v_{T_i}.$$

Hence the function (4) is a good approximation of

$$\log v_{T_i}^{\mathcal{M}}(\theta) - \log v_{T_i}, \tag{6}$$

Therefore, a minimization based on (4) leads to a fitting to implied volatilities. Note that working directly on (6) would be computationally inefficient because Equation (5) should be inverted at each step.

Let us now describe in detail the entire calibration procedure which makes use of consistent families. In a consistent calibration the parameters of the model are determined together with the initial forward rate curve. This is different from the usual fitting of HJM models, where the two steps are separate, but similar to that employed for short rate models fitted to bonds and options.

Let $G(z,x;\theta)$ be a family consistent with $\mathcal{M}$ and $P(z,x;\theta) = e^{-\int_0^x G(z,s;\theta)ds}$ be the corresponding discount function. Notice the dependency of the family from the vector of parameters θ which holds in general, and in all of the models considered (except for the HL model).

We define $\hat{z}(\theta)$ as

$$\hat{z}(\theta) := \arg\min_z \sum_{i=1}^{n} \left[\log P(\bar{x}) - \log P(z,\bar{x};\theta)\right]^2. \tag{7}$$

Notice that, in all of the cases considered here, the problem is linear in z, hence $\hat{z}(\theta)$ is an explicit function of θ.

At this point an entire zero coupon curve $P(\hat{z}(\theta), x; \theta)$ is determined. The estimates $\hat{\theta}$ are then found by solving the non-linear problem

$$\min_{\theta} \sum_{i=1}^{m} \left[\log Cap^{\mathcal{M}}(T_i, P(\hat{z}(\theta), x; \theta), \theta) - \log Cap^{Bl}(T_i, P(\hat{z}(\theta), x; \theta), v_{T_i})\right]^2 . \tag{8}$$

Remind that the strike for caps is the forward swap rate obtained from $P(\hat{z}(\theta), x; \theta)$ and therefore changes at every step of the optimization algorithm.

The above routines are implemented in a MATLAB toolbox. Given a d-dimensional volatility function, the code identifies, making use of Proposition 4.2, the corresponding minimal consistent family. Then it estimates the parameter vector θ, solving the non-linear problem (8) through the Levenberg-Marquardt algorithm, and the corresponding coefficients $z(\theta)$ of the family.

6 Empirical Results

The data set for the analysis consists of 100 daily observations from 15/2/2001 to 4/7/2001 of discount factors for thirteen maturities (3, 6, 9 months and from 1 to 10 years) and of implied volatilities of at-the-money interest rate caps with maturities 1,2,3,4,5,7,10 years, which are shown in Figure 2 . All the data are provided by Datastream.

The models are mainly evaluated in terms of capability of fitting the market volatility term structure and, in particular, of reproducing its hump when it occurs. Attention is also paid to the errors on the interest rate term structure.

The main measure adopted is the relative volatility error (RVE), namely the average on i of the absolute value of the individual relative errors $(\hat{v}^{\mathcal{M}}_{T_i} - v_{T_i})/v_{T_i}$, where $\hat{v}^{\mathcal{M}}_{T_i}$ is the estimated model T_i-cap implied volatility defined as the solution of

$$Cap^{Bl}(T_i, P(\hat{z}(\hat{\theta}), x; \hat{\theta}), \hat{v}^{\mathcal{M}}_{T_i}) = Cap^{\mathcal{M}}(T_i, P(\hat{z}(\hat{\theta}), x; \hat{\theta}), \hat{\theta}).$$

As discussed at the beginning of Section 5, the residuals of the optimization problem (8) are close (although not exactly equal) to the individual relative volatility errors. This is confirmed by the fact that the mean squared error (MSE) for (8) is in agreement with the RVE, that is lower MSE corresponds to lower RVE. Errors on discount bonds are measured in terms of mean squared error of (7) (MSEZ). One-day and one-week out-of-sample tests are also performed. A n-day out-of-sample forecast is obtained from the parameters estimated n days before and the term structure of discount factor of the current day. The model forecasts are then compared to the realized cap prices, in terms of mean squared errors. Finally, as an indication of the size of parameter estimates and of their stability over the sample, we report their means and coefficients of variation.

There is a general agreement that it is excessive to use models with more than two factors. This is confirmed by the factor analysis in Section 3. Therefore the models considered are those presented in Section 2 (i.e. HL, HW, MM) and

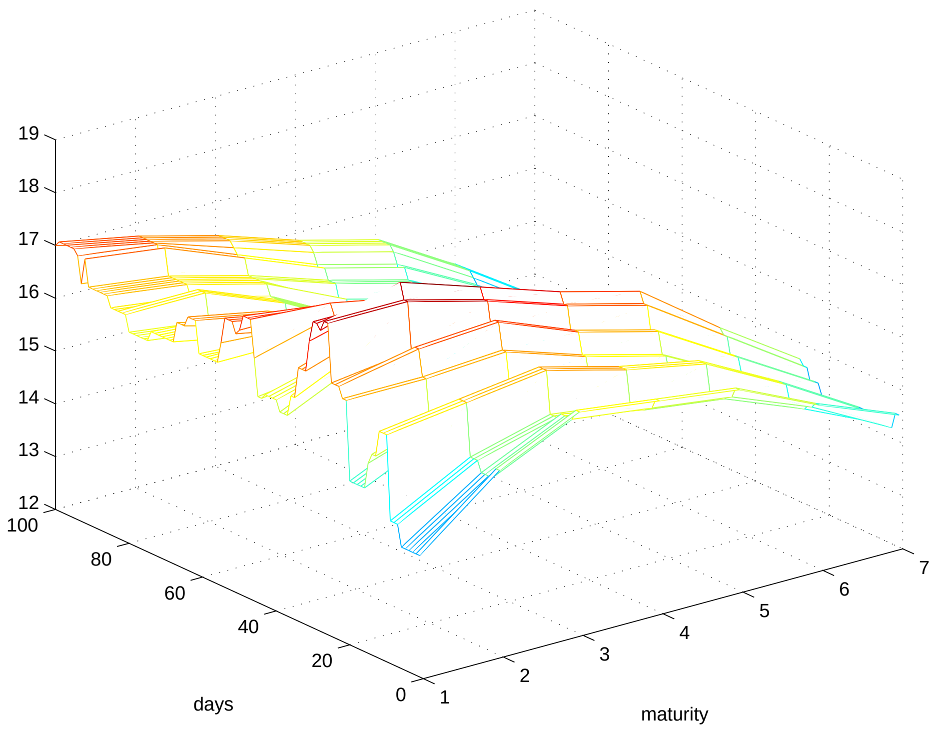


Figure 2: Implied volatilities (in percent) of at-the-money interest rate caps with maturities 1,2,3,4,5,7,10 years from 15/02/2001 to 4/7/2001. Source Datastream.

their combinations, with the exceptions of HL-HL because it is equivalent to HL and of the over-parameterized HW-MM and MM-MM.

Costs and benefits of the consistent approach are analyzed by comparing the results with those obtained by calibrating the models in a non-consistent way. This extends the analysis of Angelini and Herzel (2002) to the data set and the models presented above.

The sample is divided into two subperiods, Period 1 and Period 2. Period 1 runs from the beginning to 19/4/2001 (46 observations) and it is characterized by a humped implied volatility term structure. Period 2 goes from 20/4/2001 to the end (54 observations) and presents a decreasing implied volatility. Moreover two particular days, namely 15/2/2001 (in Period 1) and 4/7/2001 (in Period 2), denoted respectively Day 1 and Day 2, are chosen to present a graphical representation of some of the aspects of the cross-sectional calibration.

6.1 Consistent versus Non-Consistent Calibration

The theoretical justification to the use of a consistent family in the calibration procedure of a HJM model was discussed in Section 4. A matter of coherency should impose the choice of a parametric family consistent with the model adopted. In fact, if one believes in a model one should look for the initial curve among those reachable by the model. The question naturally arises as to what the costs and benefits of the consistent calibration approach are. We attempt to answer that question by comparing the behavior of the consistent method with the usual non-consistent approach in two ways. First by running a specification test for the consistent procedure, then by adopting a popular parametric interpolating family proposed by Nelson and Siegel (1987), which is used by many important central banks for the estimation of the yield curve (see BIS (1999)).

Note that the consistent calibration procedure presents some advantages on the numerical side. The algorithm is so structured that allows the fitting at the same time of bonds and caps, hence involving only one optimization procedure instead of the two necessary for a non-consistent approach. According to our experience this leads to a more robust optimization procedure, less sensible to starting parameters and less likely to incur into local minima.

The specification test entails a "non-consistent calibration with a consistent family". What we mean with this is best illustrated with an example. For the HW model we consider the three-parameter family

$$G(z, x) = z_1 e^{-z_3 x} + z_2 e^{-2z_3 x}$$

and determine the vector of parameters z to minimize the differences of logarithms of bond prices analogously to (7). In this way the optimization problem for bonds is no longer linear and the algorithm may incur into some local minima. To make sure that the algorithm finds a minimum at least as small as that reached by the consistent calibration, we let it start from the optimal point found by the consistent algorithm. Then the parameters of the HW model: $\theta = (\sigma_1, a)$ are estimated from the optimization problem for cap prices analogous to (8). In the consistent calibration, parameter z_3 is constrained to coincide with a . For this reason we measure the deviation from consistency with the difference $\Delta_{HW} = a - z_3$. We proceed analogously for the other models; in particular, for the MM model, we introduce the parameter z_6 and study $\Delta_{MM} = b - z_6$. For every day in the sample, we compute the ratios between consistent and non-consistent MSE and MSEZ. The ratios of MSEZ will always be greater than or equal to one, because, by construction, non-consistent calibrations will always be at least as good, on bonds, as the consistent one. On the other hand, the ratios for caps may turn out to be also smaller than one, because variations in the interest rate term structure may affect the market prices of caps (recovered from the implied volatilities) and this may induce some changes in the objective function.

Figure 6.1 represents the results of such a test for the HW (left) and the MM (right) models. We chose these two cases because they well represent the two possible outcomes of the test, depending on the fitting capability of the model: a sensible change in the estimates and a very small, virtually negligible, change. For each model we made two plots: the first one represent the ratios of MSE of caps and bonds with respect to the difference Δ (top); the second one is the ratio of the MSE of caps as a function of time (bottom).

For the HW model, the top left panel shows that the values of Δ are often different from zero, meaning that the consistency constraint is indeed binding. Such changes in Δ obviously imply an improvement for bond fitting. On the other hand, they result in a worsening, in most of the cases, of the fitting of the caps. In this case, the use of a non-consistent calibration procedure leads to a change in the objective function for caps that makes the fitting less precise. A similar picture, although less pronounced, comes out in the case of the HL-HW model (the HW-HW model is more delicate because there are two couple of parameters to control and we will not analyze it). The top right panel shows the ratio of the MSE as a function of time. It is evident that the behavior depends on the period. In Period 1, the changes in the MSE are lower, quite often in favor of the consistent approach. All of the cases when the relaxation of the consistency constrain gives a significant improvement of the performance are concentrated at the beginning of Period 2. For the rest of the period, the ratio is generally lower than one, particularly in several days.

As for the MM model, we note that relaxing the consistency constraint does not lead to a significant change because the consistent family is rich enough to provide a good fitting to bond prices and, except for a few days, performances on caps do not improve. Such behavior is similar to that of the HL-MM model. Later we will see that these models offer the best overall performances on the sample.

Similarly, the question naturally arises as to how the fit of bond prices compares with standard methods such as have been proposed by Nelson and Siegel (1987). Also, does the in-sample and out-of-sample cap pricing performance improve or weaken if one restricts the forward rate curve in the proposed way compared to Nelson-Siegel? Further, does the consistency restriction lead to more stable parameter estimates for the parameters in the volatility function?

To answer those questions, we run, for every day in the sample and for each model, the standard non-consistent calibration algorithm adopting the Nelson-Siegel family. For convenience of the reader, we recall that the Nelson-Siegel parametrization of the forward rate curve is given by

$$G_{NS}(z,x) = z_1 + z_2 e^{-z_4 x} + z_3 x e^{-z_4 x}.$$

The results are reported in Tables 1 and 2. For this type of comparison, the tables are to be read along the columns. The fitting capabilities of the two approaches are different, confirming the analysis of Angelini and Herzel (2002). As for the fitting of bond prices, the consistent approach shows remarkable benefits. For most models, the consistent family outperforms the Nelson-Siegel one: even in the cases of the HL-HW and HW-HW models, where the number of parameters of the family is four as in the Nelson-Siegel one, the MSEZ of the consistent family are of an order of magnitude smaller. For the MM and the HL-MM model the difference is from one to two orders of magnitude. Just for the HL and HW model the errors on bond prices of the consistent families are slightly greater, but of the same order, although they only have two parameters.

As for in-sample cap fitting, for most of the models the consistent approach gives a better performance, with a minor exception of the HW-HW case in Period 1, and a more evident one of the HL and HW models. As for the out-of-sample cap fitting, the differences among the two approaches become less evident. In the HW case, the non-consistent approach has the drawback of a higher instability of parameter estimates, especially those of parameter a, as shown by the coefficient of variation of Table 2. These results may be compared to Angelini and Herzel (2002) who analyzed the HW model on a different sample, reporting a similar behavior in terms of parameter stability, but a better performance of the consistent family also in terms of cap fitting. In the more parsimonious HL model, there is not relevant difference in the parameter estimates between the two approaches. In general, the coefficients of variation of the estimates vary with the parameter and the interpolation adopted. To summarize the results of the different approaches one can compute the averages of the coefficients of variation for each model. The consistent estimates turn out to be more stable, according to such ranking, for all of the cases, except for the models MM, HL-HW and HW-HW in Period 2.

The examination and comparison of consistent and non-consistent approaches on the proposed data set let us formulate some conclusions and answers to the above posed questions. Our general remark is that the proposed theoretical coherency in the model mainly does bring some benefits to the calibration procedure. In our opinion, the major cost of the consistent calibration approach is that the initial term structure of interest rates is not fitted exactly, as it is the case of any parsimonious family of forward curves. However, in the case of consistent families, this can be offset by enlarging the family, still preserving consistency with the model. Our analysis shows that, for sufficiently rich families (e.g. with four or more parameters), errors on bond fitting are satisfactory. A second possible objection to the use of a consistent approach can be that the consistency constraint may in fact limit the efficiency of the fitting algorithm, giving less precise results. The specification test performed above shows that this is not the case. The comparison with the standard method of Nelson and Siegel (1987) shows that the in-sample fitting of caps and bonds of the consistent approach are generally better, while out-of-sample performances are similar. Moreover, the consistent approach generally leads to more stable estimates for the parameters of the volatility function.

6.2 Period 1 and Period 2

From now on we will only consider the calibration results obtained with the consistent algorithm. The purpose of this section is to compare the performances of the models on extended periods of time. Table 1 exhibits the sample mean of the daily measures, namely MSE and MSEZ (we read the table along the rows corresponding to the consistent approach). Here it is confirmed that the major differences between models are in Period 1, where the volatility term structure is humped. In such a Period the models containing the MM factor outperform all other models. In particular HL-MM is the best, both in terms of cap and bond fitting, but the one-factor, slightly more parsimonious, MM model has a very good performance. The other models substantially improve their results in Period 2, becoming a viable alternative when the implied volatility is decreasing. For a more precise comparison, that confirms the above discussion, the daily RVE's on the entire sample are plotted in Figure 6.2 for all the models.

Out-of-sample tests are reported in Table 1. On average, in the one-day case, the errors produced by MM and HL-MM are much lower than those produced by other models in Period 1, while the differences are again less evident for Period 2. These differences become less pronounced as the forecasting horizon increases.

The mean and the coefficient of variation of parameter estimates are reported in Table 2. It can be observed that one-factor models are generally more stable than two-factor ones. Notice that the coefficient of variation of the models containing the HW factor (especially the two-factor models HL-HW and HW-HW) are generally higher in Period 1, while those of the models containing the MM factor are only slightly higher in Period 2.

In most of the cases, for any given model, there is not a significant difference on the coefficient of variation between the two periods, a part for a worse behavior in the first period of the models containing the factor HW.

It is interesting to compare such cross-sectional estimates with those obtained from the factor analysis in Section 3, where the two-factor model HL-MM was selected and fitted to the loading factors. The mean on the entire sample of the cross-sectional estimates for the parameter of the first factor of the HL-MM model is 0.0067, surprisingly in agreement with the estimate based on historical data. Moreover, these daily estimates are very much stable around their mean, as indicated by the coefficient of variation. The analysis of the second factor is more problematic and less clear, due to various reasons, which will not be further investigated in this paper.

While in Period 1 only the models containing the MM factor are able to capture the hump of the volatility term structure, in Period 2, satisfactory calibration results are also achieved by more parsimonious models like HW. From this analysis it seems that the use of two-factor models like HL-HW and HW-HW does not substantially improve the fitting to the implied volatility and also produces more unstable parameter estimates. Overall the best performances in terms of volatility errors are given by HL-MM and MM, with MM presenting more stable parameter estimates.

The above results show that the form of the model has an impact on its performance. Compare for instance the MM and HL-HW models (three parameters) or the HL-MM and HW-HW models (four parameters). This is different from what observed by Amin and Morton (1994) on one-factor models, which found that "the number of the parameters used in the model has a strong effect on the behavior of the model, in fact stronger than the form of the model used".

In-Sample	HL	HW	MM	HL-HW	HW-HW	HL-MM
MSE cons. 1	$6.7\ 10^{-3}$	$4.2\ 10^{-3}$	$2.3\ 10^{-4}$	$4.7\ 10^{-3}$	$3.5\ 10^{-3}$	$1.3\ 10^{-4}$
MSE N-S 1	$5.2\ 10^{-3}$	$3.1\ 10^{-3}$	$5.0\ 10^{-4}$	$4.8\ 10^{-3}$	$3.3\ 10^{-3}$	$1.5\ 10^{-4}$
MSE cons. 2	$2.2\ 10^{-3}$	$7.6\ 10^{-4}$	$3.2\ 10^{-4}$	$3.7\ 10^{-4}$	$3.8\ 10^{-4}$	$2.6\ 10^{-4}$
MSE N-S 2	$2.1\ 10^{-3}$	$5.2\ 10^{-4}$	$4.4\ 10^{-4}$	$5.2\ 10^{-4}$	$4.7\ 10^{-4}$	$2.8\ 10^{-4}$
MSEZ cons. 1	$1.8\ 10^{-5}$	$2.1\ 10^{-5}$	$8.8\ 10^{-7}$	$1.7\ 10^{-6}$	$1.6\ 10^{-6}$	$1.2\ 10^{-7}$
MSEZ N-S 1	$1.5\ 10^{-5}$	$1.5\ 10^{-5}$	$1.5\ 10^{-5}$	$1.5\ 10^{-5}$	$1.5\ 10^{-5}$	$1.5\ 10^{-5}$
MSEZ cons. 2	$2.6\ 10^{-5}$	$2\ 10^{-5}$	$6.1\ 10^{-7}$	$1.5\ 10^{-6}$	$1.5\ 10^{-6}$	$8.0\ 10^{-8}$
MSEZ N-S 2	$1.3\ 10^{-5}$	$1.3\ 10^{-5}$	$1.3\ 10^{-5}$	$1.3\ 10^{-5}$	$1.3\ 10^{-5}$	$1.3\ 10^{-5}$
Out-of-Sample one-day	HL	HW	MM	HL-HW	HW-HW	HL-MM
MSE cons. 1	$7.6\ 10^{-3}$	$5.4\ 10^{-3}$	$1.9\ 10^{-3}$	$5.5\ 10^{-3}$	$4.8\ 10^{-3}$	$1.8\ 10^{-3}$
MSE N-S 1	$6.0\ 10^{-3}$	$4.4\ 10^{-3}$	$2.1\ 10^{-3}$	$5.7\ 10^{-3}$	$4.5\ 10^{-3}$	$1.7\ 10^{-3}$
MSE cons. 2	$3.3\ 10^{-3}$	$2.0\ 10^{-3}$	$1.6\ 10^{-3}$	$1.6\ 10^{-3}$	$1.7\ 10^{-3}$	$1.6\ 10^{-3}$
MSE N-S 2	$3.2\ 10^{-3}$	$1.8\ 10^{-3}$	$1.7\ 10^{-3}$	$1.8\ 10^{-3}$	$1.8\ 10^{-3}$	$1.6\ 10^{-3}$
Out-of-Sample one-week	HL	HW	MM	HL-HW	HW-HW	HL-MM
MSE cons. 1	$1.1\ 10^{-2}$	$1.0\ 10^{-2}$	$7.8\ 10^{-3}$	$9.4\ 10^{-3}$	$9.2\ 10^{-3}$	$7.6\ 10^{-3}$
MSE N-S 1	$9.3\ 10^{-3}$	$9.9\ 10^{-3}$	$8.4\ 10^{-3}$	$9.6\ 10^{-3}$	$9.3\ 10^{-3}$	$8.0\ 10^{-3}$
MSE cons. 2	$5.4\ 10^{-3}$	$4.5\ 10^{-3}$	$4.3\ 10^{-3}$	$4.3\ 10^{-3}$	$4.4\ 10^{-3}$	$4.5\ 10^{-3}$
MSE N-S 2	$5.1\ 10^{-3}$	$4.3\ 10^{-3}$	$4.2\ 10^{-3}$	$4.3\ 10^{-3}$	$4.3\ 10^{-3}$	$4.1\ 10^{-3}$

Table 1: In-sample and out-of-sample for the consistent calibration and the calibration using the Nelson-Siegel family for Period 1 and Period 2 in terms of the mean of mean squared error of the optimization problem (MSE) and mean squared error on zero coupon bond prices (in logarithmic terms) (MSEZ).

	HL	HW	MM	HL-HW	HW-HW	HL-MM
Per. 1 Cons.	σ_0=0.0074 (0.021)	σ_1=0.0073 (0.036) a=-0.0145 (-0.893)	σ_2=0.0067 (0.082) γ=0.0019 (0.435) b=0.154 (0.280)	σ_0=0.0066 (0.322) σ_1=0.0011 (2.012) a=0.1064 (0.770)	$\sigma_{1,1}$=0.0048 (0.586) a_1=0.0416 (2.360) $\sigma_{1,2}$=0.0036 (0.841) a_2=0.0646 (1.196)	σ_0=0.0066 (0.043) σ_2=-0.002 (-0.969) γ=0.0079 (0.257) b=0.5769 (0.073)
Per. 1 N-S	σ_0=0.0074 (0.021)	σ_1=0.0073 (0.041) a=-0.0129 (-1.263)	σ_2=0.0068 (0.091) γ=0.0019 (0.563) b=0.1521 (0.408)	σ_0=0.0059 (0.417) σ_1=0.0017 (1.878) a=0.0128 (2.116)	$\sigma_{1,1}$=0.0065 (0.136) a_1=-0.0108 (-1.230) $\sigma_{1,2}$=0.0026 (0.734) a_2=0.0006 (33.900)	σ_0=0.0067 (0.036) σ_2=-0.0003 (-4.816) γ=0.0060 (0.462) b=0.5706 (0.021)
Per. 2 Cons.	σ_0=0.0073 (0.028)	σ_1=0.0074 (0.034) a=0.0118 (0.647)	σ_2=0.0073 (0.046) γ=0.0006 (0.485) b=0.0795 (0.290)	σ_0=0.0064 (0.182) σ_1=0.0023 (1.111) a=0.2225 (0.893)	$\sigma_{1,1}$=0.0055 (0.158) a_1=0.0031 (6.444) $\sigma_{1,2}$=0.0044 (0.450) a_2=0.1067 (1.950)	σ_0=0.0067 (0.042) σ_2=-0.0014 (-1.366) γ=0.0013 (2.994) b=0.5693 (0.070)
Per. 2 N-S	σ_0=0.0073 (0.029)	σ_1=0.0074 (0.045) a=0.0142 (1.037)	σ_2=0.0074 (0.048) γ=0.0005 (0.423) b=0.0720 (0.174)	σ_0=0.0061 (0.200) σ_1=0.0035 (0.638) a=0.0655 (0.898)	$\sigma_{1,1}$=0.0052 (0.228) a_1=-0.0181 (-5.589) $\sigma_{1,2}$=0.005 (0.286) a_2=0.0316 (0.977)	σ_0=0.0067 (0.039) σ_2=-0.0015 (-1.490) γ=-0.0002 (-17.441) b=0.5698 (0.042)

Table 2: Mean and coefficient of variation (in parentheses) of parameter estimates for the consistent calibration and the calibration using the Nelson-Siegel family for Period 1 and Period 2.

This may be due to several reasons: the calibration procedures differ, mainly on the estimation of the initial curve and on the objective function, the data are of different type and period and Amin and Morton do not consider two-factor models. Driessen at al. (2003) in their empirical comparison of hedging and pricing properties of different multi-factor models considered only Gaussian models with factors of the HW form. They observe that such models offer worse performances than others like the PCA non parametric models and Libor Market Models. Our results suggest that some improvements can be achieved by adopting different volatility functions. Also in this case it has to be emphasized that the data sets and the calibration procedures are different and hence their results are only partially comparable to ours.

6.3 Day 1 and Day 2

Figure 6.3 shows the fitting capabilities of different models in terms of forward rates (left) and implied volatilities (right), for Day 1 (top) and Day 2 (bottom). In order to keep the figures readable, only the results of the HW, HL-HW and HL-MM models are plotted.

The "observed" forward curve, plotted with circles, is computed from the input data $P(\bar{x})$ as the continuously compounded forward rate between available adjacent maturities (1). As for the models, we computed the corresponding forward rates from the model bond prices fitted to the data. For both days, the main qualitative characteristics of the "observed" curves are the throat at short-term maturities (more pronounced in Day 1) and the flattening for long-term maturities. The HL and HW models do not produce the throat nor follow the asymptotic behaviour. The HL-HW and HW-HW models give good results for longer maturities, while having difficulties to fit the throat. The MM and HL-MM models follow quite closely the whole shape.

The market implied volatility term structure is humped for Day 1 and decreasing for Day 2. All of the models produce a good fitting for Day 2, while major differences are evident for Day 1. Only MM (not shown in the figure) and HL-MM are able to recover the humped volatility structure.

Table 3 summarizes the results of the calibration for Day 1 and Day 2. Fitting results do not merely depend on the numbers of parameters. This is evident comparing MM to HL-HW and HL-MM to HW-HW. The difference is clearly made by the capability to reproduce the humped structure. Notice that the performances of the two-factor models HL-HW and HW-HW are not better, if not worse, on implied volatility fitting than those of HL and HW. This is explained by the fact that the objective functions also depend on the fitting of the bond prices for which the two-factor models perform much better. Hence, in this cases, it is not advisable to adopt the two-factor HL-HW and HW-HW models, because they do not provide a real improvement over the corresponding one-factor models, at least in terms of volatility fitting.

Overall, the best performance is given by HL-MM, both on Day 1 and on Day 2, while the best one-factor model is MM.

model	HL	HW	MM	HL-HW	HW-HW	HL-MM
MSE Day 1	$2.1\ 10^{-2}$	$1.2\ 10^{-2}$	$6.4\ 10^{-4}$	$2.1\ 10^{-2}$	$1.3\ 10^{-2}$	$3.1\ 10^{-4}$
MSE Day 2	$1.6\ 10^{-4}$	$1.6\ 10^{-4}$	$1.6\ 10^{-4}$	$7.2\ 10^{-4}$	$5.1\ 10^{-4}$	$7.4\ 10^{-5}$
RVE (%) Day 1	4.6450	3.4810	0.8281	4.7245	3.5635	0.6225
RVE (%) Day 2	0.4525	0.4521	0.4485	0.9781	0.9221	0.3028
MSEZ Day 1	$1.2\ 10^{-5}$	$1.5\ 10^{-5}$	$3.6\ 10^{-7}$	$7.6\ 10^{-7}$	$7.5\ 10^{-7}$	$1.3\ 10^{-7}$
MSEZ Day 2	$3.9\ 10^{-5}$	$3.8\ 10^{-5}$	$4.3\ 10^{-7}$	$5.6\ 10^{-7}$	$7.8\ 10^{-7}$	$7.3\ 10^{-8}$

Table 3: Calibration results for Day 1 and Day 2 in terms of: mean squared error of the optimization problem (MSE), Relative Volatility Error (RVE, in percent) and mean squared error on bond prices (MSEZ).

7 Conclusion

This paper proposes a calibration algorithm that fits multi-factor Gaussian models to the implied volatilities of caps using the respective minimal consistent family to infer the forward rate curve. A general result, useful for the computation within the algorithm of consistent families for multi-factor models, has been proven. In a consistent approach the parameters of the model are determined jointly with the initial curve. The objective of the algorithm is to minimize the relative difference of implied volatilities.

Although the algorithm is fairly general, it is only applied to three forward rate volatility structures and their combination to form two-factor models. This choice is also supported by the Principal Component Analysis. Models are evaluated in terms of in-sample and out-of-sample data fitting as well as stability of parameter estimates. The results are analyzed by focusing on the capability of fitting the market implied volatility curve and, in particular, of reproducing its characteristic humped shape. We observe that model performances strongly depend on such a shape. In periods of humped volatility, the Mercurio-Moraleda model and the two-factor HL-MM model outperform all the others, as they are able to capture the hump. Between this two, the best fitting performance is given by the two-factor model, as expected since it has one parameter more, but the one-factor model presents slightly more stable parameters. In periods of decreasing volatilities the differences among models are less evident and more parsimonious models may be adopted.

The empirical findings of the paper also show that models with same number of parameters may behave in a different way depending on the shape of their volatility functions. This is different from what observed by Amin and Morton (1994) on one-factor models, where only the number of parameters seems to matter.

The results obtained by the Principal Component Analysis are in accordance to those of the cross-sectional calibration. In fact, both the procedures identify in the HL-MM model the best fit to the data and, quite remarkable, the parameter estimates of the leading factor are very close to each other.

The choice of consistent families, motivated theoretically by Björk and Christensen (1999) and Filipović (2001), reduces the arbitrariness in the selection of the initial curve. The present paper is one of the first studies of its practical implications and the results are encouraging. A specification test shows that, in most cases, the introduction of the consistency constraint improves the efficiency of the algorithm without worsening the precision of the estimates. Also, the consistent calibration, when compared with a non-consistent method that uses the Nelson-Siegel family to estimate the forward curve, leads to a more precise fitting and comparatively more stable parameter estimates.

References

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

[2] F. Angelini, S. Herzel, Consistent Initial Curves for Interest Rate Models, *Journal of Derivatives*, 9, No. 4 (2002), 8-18.

[3] BIS. Zero-coupon yield curves: Technical documentation. Bank for International Settlements, Basle, March 1999.

[4] T. Björk, *Arbitrage Theory in Continuous Time*, Oxford University Press, 1998.

[5] T. Björk, A Geometric View of Interest Rate Theory, in: *Option Pricing, Interest Rates and Risk Management*, (eds Jouini, E., Cvitanic, J. & Musiela, M) Cambridge University Press, 2001.

[6] T. Björk, B.J. Christensen, Interest Rate Dynamics and Consistent Forward Rate Curves. *Mathematical Finance*, 9, No. 4 (1999), 323-348.

[7] T. Björk, C. Landen, On the construction of finite dimensional realizations for nonlinear forward rate models, *Finance and Stocastics*, 6, No. 3 (2002), 303-331.

[8] Brace, M. Musiela, A Multifactor Gauss Markov Implementation of Heath, Jarrow, and Morton, *Mathematical Finance*, 4, No. 1 (1994), 259-281.

[9] D. Brigo, F. Mercurio, *Interest Rate Models: Theory and Practice*, Springer-Verlag, 2001.

[10] W. Bühler, M. Uhrig-Homburg, U. Walter, T. Weber, An Empirical Comparison of Forward Rate and Spot Rate Models for Valuing Interest Rate Options, *The Journal of Finance*, 54 (1999), 269-305.

[11] J. Driessen, P. Klaassen, B. Melenberg, The Performance of Multi-Factor Term Structure Models for Pricing and Hedging Caps and Swaptions, *Journal of Financial and Quantitative Analysis*, 38, No. 3 (2003), 635-672.

[12] D. Filipović, *Consistency Problems for Heath-Jarrow-Morton Interest Rate Models*, Lecture Notes in Mathematics, 1760, Springer-Verlag, 2001.

[13] D. Heath, R. Jarrow, A. Morton, Bond Pricing and the Term Structure of Interest Rates *Econometrica*, 60, No. 1 (1992), 77-106.

[14] T. Ho, S. Lee, Term Structure Movements and Pricing Interest Rate Contingent Claims. *The Journal of Finance*, 41 (1986), 1011-1029.

[15] J. Hull, A. White, Pricing Interest Rate Derivative Securities. *The Review of Financial Studies*, 3 (1990), 573-592.

[16] J. Hull, A. White, Numerical Procedures for Implementing Term Structure Models I: Single-Factor Models. *The Journal of Derivatives*, Fall (1994), 7-16.

[17] Mercurio F. and J.M. Moraleda, An Analytically Tractable Interest Rate Model with Humped Volatility, discussion paper TI 96-116-2, (1996), Tinbergen Institute, Erasmus University Rotterdam, The Netherlands.

[18] J.M. Moraleda, A. Pelsser, Forward versus Spot Interest Rate Models of the Term Structure: an Empirical Comparison. *The Journal of Derivatives*, Spring (2000), 9-21.

[19] C. Nelson, A. Siegel, Parsimonious Modelling of the Yield Curves. *Journal of Business*, 60 (1987), 473-489.

[20] R. Rebonato, *Interest Rate Option Models*, 2nd edition, Wiley, Chicherster, 1998.

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

[22] O. Vasicek, An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5 (1977), 177-188.

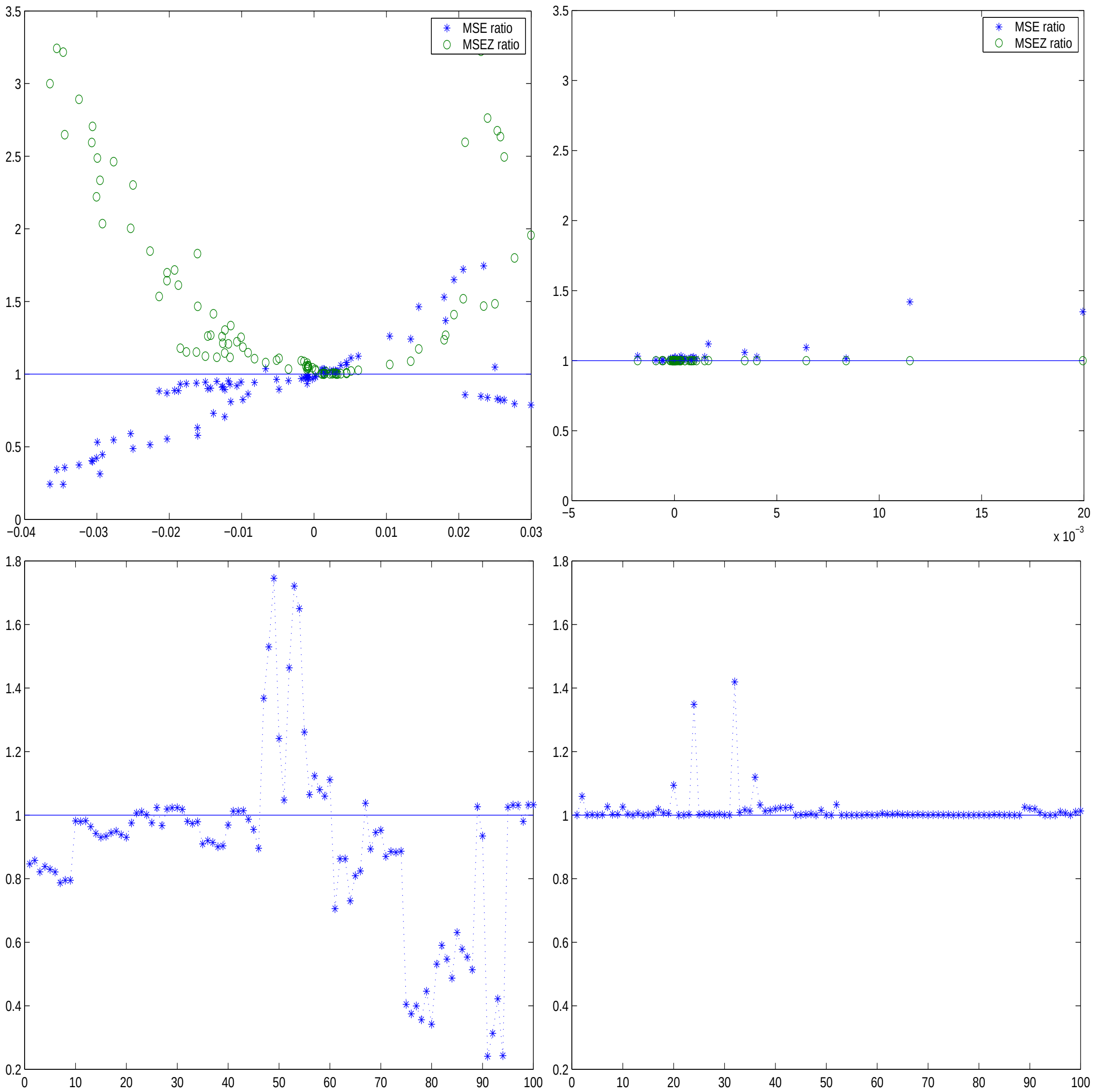
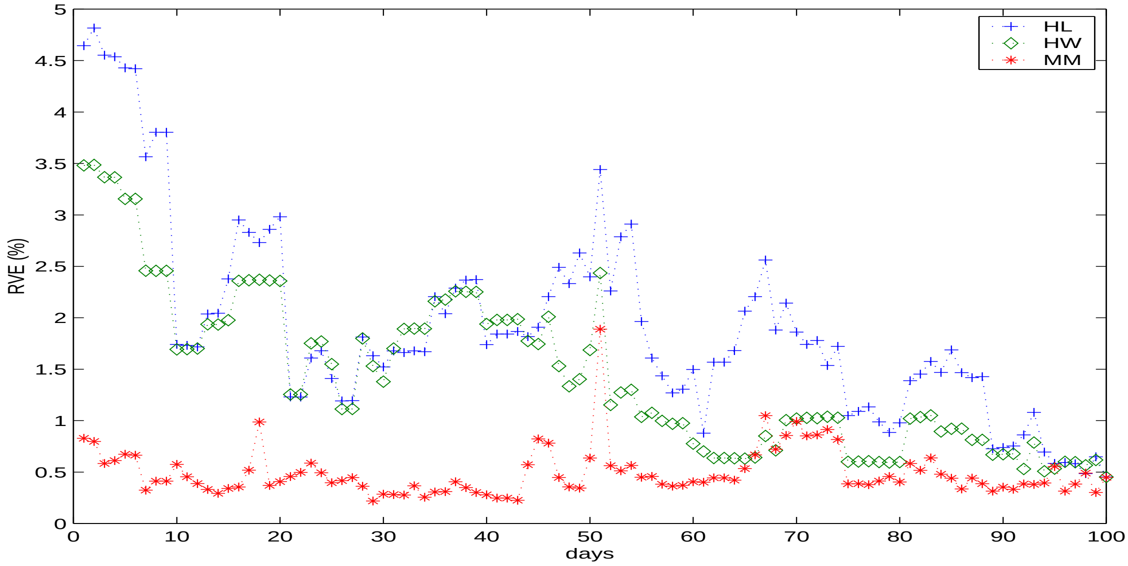


Figure 3: Results of a specification test for the HW (left) and MM (right) models. The top panels represent the ratios of MSE for caps and bonds with respect to the differences $a - z_3$ and $b - z_6$ of the estimates of the specification test, the bottom ones are time series of daily ratios of MSE for caps.



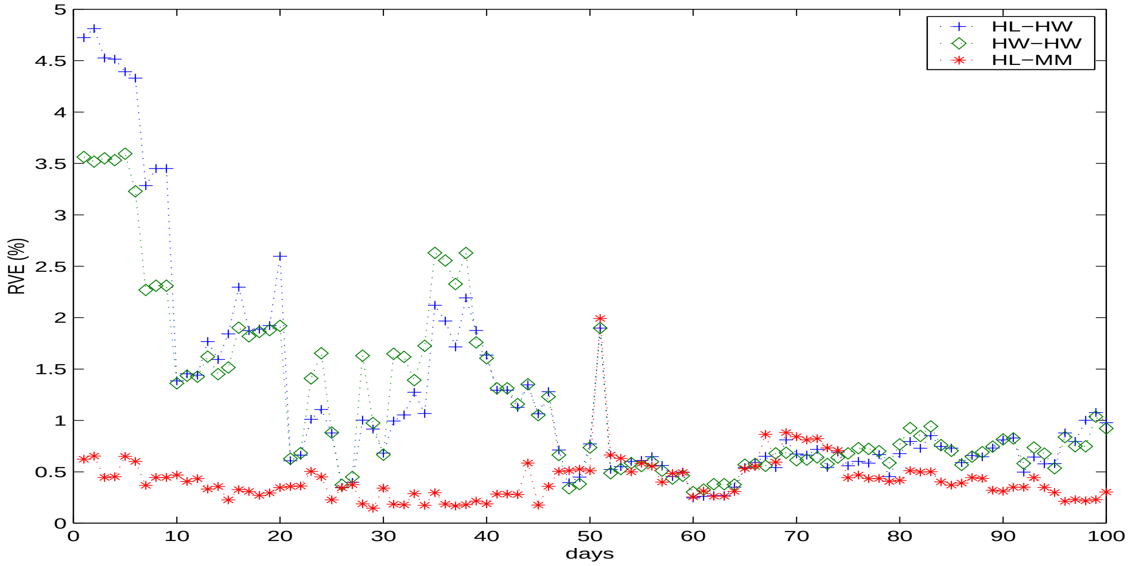


Figure 4: Daily RVE (in percent) for one-factor models (HL, HW, MM, top) and for two factors models (HL-HW, HW-HW, HL-MM, bottom).

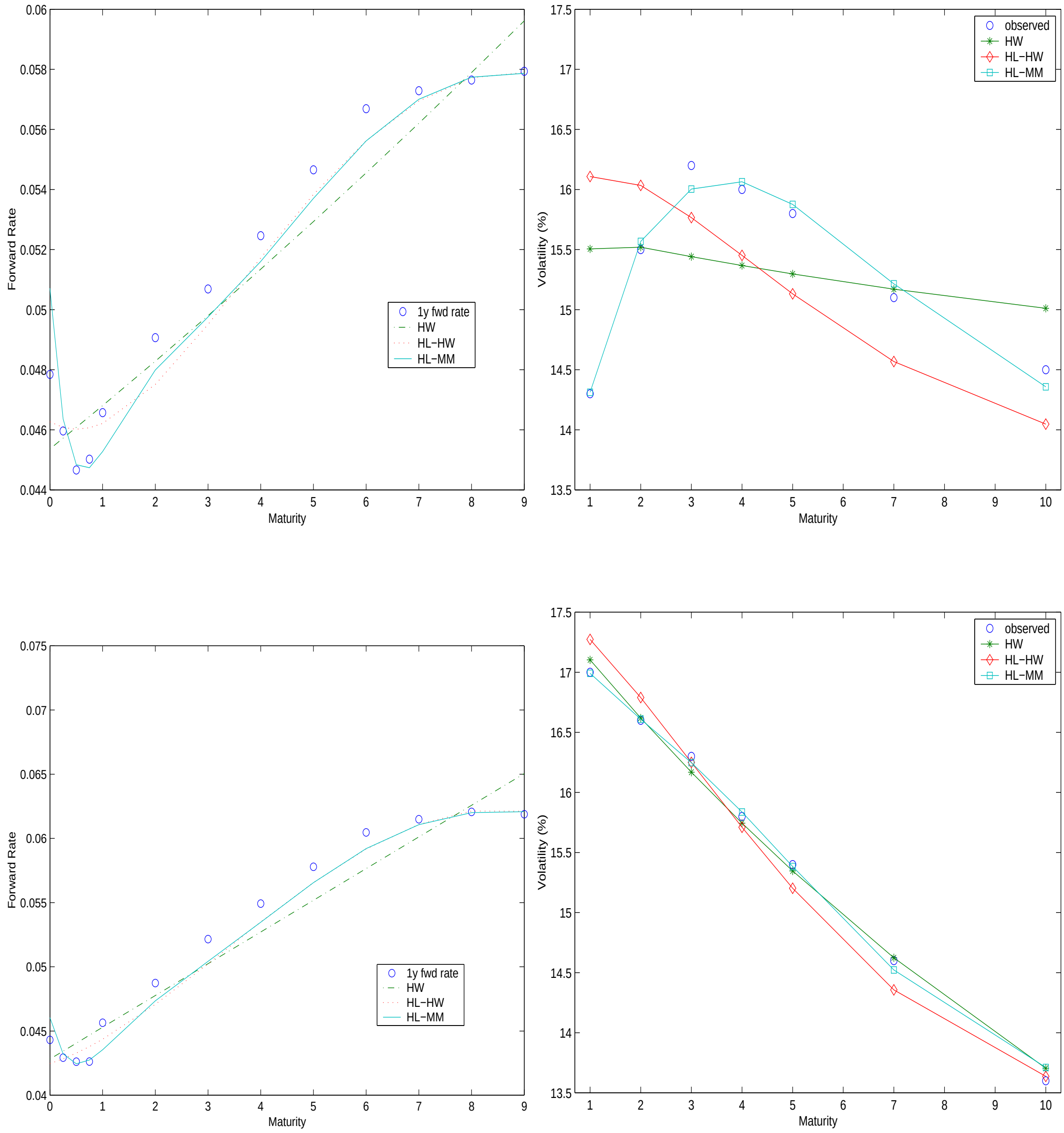


Figure 5: Fitting of the model to the continuously compounded forward rate curve obtained from input bonds (left) and to the at-the-money cap implied volatility term structure (right), for Day 1 (top) and Day 2 (bottom).