Chapter 5

The Hull-White Model and multiobjective calibration

This chapter is the first chapter with empirical and numerical applications[1] of this work. We have shown the foundations of interest rate theory and we have derived the general conditions of consistency. However, only little attention has been devoted to the empirical applications of these concepts. In this chapter we address these applications.

5.1 The Hull-White Model

Our test case is the following model, studied by Hull and White [28] (henceforth HW):

$$dr(t) = [\Phi(t) - ar(t)] + \sigma dW(t). \tag{5.1}$$

The Hull-White model improve Ho-Lee model incorporating mean-reversion and providing closed formulas for liquid options like interes rate caps. This model is one of the simplest Gaussian HJM models which preserves the Markov property, allowing very efficient numerical methods for the pricing of any kind of options. On the negative side, it does not capture large humps of the term structure of volatilities (TSV hereafter). The model instantaneous-forward volatility curve $T \to \sigma_f(t,T)$, as we will prove later, is monotonically decreasing and this fact often allows only small humps in the caplet curve. However, as pointed out by Brigo and Mercurio [14, pp. 91–92], in case of decreasing TSIR curves, large

[1] Some of the results presented in this chapter are also reported in [20].

humps can be produce even by this model.

Summarizing, it exhibits a relative good performance when it is chosen as a parsimonious solution for bussiness cycles with monotonically decreasing TSV, as it is shown by [2].

5.1.1 Markovianity of the HW model

The short-rate differential process (2.11) in Proposition 2 is not a Markov process in general. Notice in fact that time t appears in the expression (2.12) for the drift both as extreme of integration and inside the integrand function. However, the HW model as a particular Gaussian HJM have a suitable specification of σ for which the short-rate r is indeed a Markov process. This happens because we can write the model volatility function $\sigma(t,T)$ as a separable specification:

$$\sigma(t,T) = \varsigma(t)\varrho(T) \tag{5.2}$$

with $\varsigma(t)$ and $\varrho(T)$ strictly positive and deterministic functions of time. Under such a separable specification, the short-rate process becomes

$$\begin{aligned} r(t) := F(t,t) &= F(0,t) + \int_0^t \left(\varsigma(u)\varrho(t)\int_u^t \varsigma(u)\varrho(s)\,ds\right)\,du + \int_0^t \varsigma(s)\varrho(t)\,dW(s) \\ &= F(0,t) + \varrho(t)\int_0^t \left(\varsigma^2(u)\int_u^t \varrho(s)\,ds\right)\,du + \varrho(t)\int_0^t \varsigma(s)\,dW(s) \end{aligned} \tag{5.3}$$

Notice that, if we introduce the deterministic function $A(\cdot)$

$$A(t) := F(0,t) + \varrho(t)\int_0^t \left(\varsigma^2(u)\int_u^t \varrho(s)ds\right)du,$$

by taken differentials in (5.3) we can write

$$\begin{aligned} dr(t) &= A'(t) + \varrho'(t)\int_0^t \varsigma(s)dW(s) + \varsigma(t)\varrho(t)\,dW(t) \\ &= \left[A'(t) + \varrho'(t)\frac{r(t)-A(t)}{\varrho(t)}\right]\,dt + \varsigma(t)\varrho(t)\,dW(t) \\ &= [a(t) + b(t)r(t)]\,dt + c(t)\,dW(t) \end{aligned} \tag{5.4}$$

where

$$\begin{aligned} a(t) &:= A'(t) - \frac{\varrho'(t)}{\varrho(t)} A(t), \\ b(t) &:= \frac{\varrho'(t)}{\varrho(t)} \text{ ; and,} \\ c(t) &:= \varsigma(t)\varrho(t) = \sigma(t,t). \end{aligned} \tag{5.5}$$

We can finally derive the HJM forward-rate dynamics that is equivalent to the original short-rate dynamics (5.1). To this end, let us set

$$\sigma(t,T) = \sigma e^{-a(T-t)},$$

where a and σ are real constants, so that

$$\begin{aligned} \varsigma(t) &= \sigma e^{at}, \\ \varrho(T) &= e^{-aT}, \\ A(t) &= F(0,t) + \tfrac{\sigma^2}{2a^2}(1-e^{-at})^2; \end{aligned} \tag{5.6}$$

and after some tedious but trivial algebra we have for the HW model:

$$\begin{aligned} a(t) &:= F_t(0,t) + aF(0,t) + \frac{\sigma^2}{2a}(1-e^{-2at}), \\ b(t) &:= -a \\ c(t) &:= \sigma. \end{aligned} \tag{5.7}$$

The resulting short-rate dynamics is then given by

$$dr(t) = \left[F_t(0,t) + aF(0,t) + \frac{\sigma^2}{2a}(1-e^{-2at}) - ar(t) \right] dt + \sigma\, dW(t), \tag{5.8}$$

which is equivalent to (5.1) when combined with the identity

$$\Phi(t) := F_t(0,t) + aF(0,t) + \frac{\sigma^2}{2a}(1-e^{-2at}).$$

In conclusion, we then have that the HW model is a short-rate markovian model that admits a one-factor HJM formulation. Turning back to the Musiela parametrization, we have that the volatility specification for this model is

$$\widetilde{\sigma}(t,x) := \sigma(t,t+x) = \sigma e^{-ax}, \tag{5.9}$$

which falls into the class of HJM models with deterministic volatility

$$\widetilde{\sigma}(f_t, x) = \widetilde{\sigma}(x).$$

5.2 The Nelson and Siegel family and Invariance

In order to illustrate the theoretical ideas shown in the previous chapter, we now move from abstract theory to the investigation of a number of concrete forward curve families and the Hull-White model. More fundamental results can be found in Björk and Christensen [10] in detail. The leading example is the popular forward curve family introduced by Nelson and Siegel [36], first introduced informally in previous chapter. We analyze the consistency of this family and several variations of it with the Hull-White model. We adapt some of the theoretical results to this Gaussian case study without no further technical discussion for the general case. From now, we again remove the symbol $\widetilde{}$ as in Chap. 4., i.e. we will consider the HJM model under the Musiela parametrization.

Consider the space $\mathcal{H}_\gamma$ introduced in Sect. 4.3.2

Corollary 3 *Consider as given the mapping*

$$G : \mathcal{Z} \to \mathcal{H}_\gamma$$

where the parameter space $\mathcal{Z}$ is an open connected subset of R^d, $\mathcal{H}_\gamma$ a Hilbert space and the forward curve manifold $\mathcal{G} \subseteq H_\gamma$ is defined as $\mathcal{G} = \mathrm{Im}(G)$. The family $\mathcal{G}$ is consistent with the one-factor model $\mathcal{M}$ with deterministic volatility function $\sigma(\cdot)$, if and only if

$$G_x(z,x) + \sigma(x)\int_0^x \sigma(s)ds \in Im\left[G_z(z,x)\right], \tag{5.10}$$

$$\sigma(x) \in Im\left[G_z(z,x)\right], \tag{5.11}$$

for all $z \in Z$.

Proof. See Theorem 4 in Sect. 4.3.2 for the particular case of deterministic volatility $\sigma(f_t, x) = \sigma(x)$. □

The statements (5.10) and (5.11) are the particular CDC and CVC for the deterministic volatility case. These are easy to apply in concrete cases as shown Björk and Christensen [10] or De Rossi [17], among others.

5.2.1 The NS family

The NS forward curve manifold $\mathcal{G}$ is parametrized by $z \in \mathcal{Z} = \mathbb{R}^4$. Recall that the curve shape $G(z, x)$ is given by the expression

$$G(z, x) = z_1 + z_2 e^{z_4 x} + z_3 x e^{-z_4 x}. \tag{5.12}$$

For $z_4 = 0$ the *Frechet* derivatives $G_z(z, x)$ and $G_x(z, x)$ are easily obtained as

$$\begin{array}{rcl} G_z(z, x) & = & [1 \quad e^{-z_4 x} \quad x e^{-z_4 x}]^T, \\ G_x(z, x) & = & (z_3 - z_2 z_4 - z_3 z_4 x) e^{-z_4 x}. \end{array} \tag{5.13}$$

Henceforth, we write $\mathcal{Z}_{NS} = \{[z_1 \ z_2 \ z_3 \ z_4]^T \ : z_4 \neq 0\}$ for the NS parameter space and $\mathcal{G}_{NS} = G(\mathcal{Z}_{NS})$ for the associated manifold. For the HW model characterized by the volatility function (5.9), the consistency conditions of Theorem 4 become

$$\left\{ \begin{array}{rcl} G_x(z, x) + \frac{\sigma^2}{a} [e^{-ax} - e^{-2ax}] & \in & Im[G_z(z, x)], \\ \sigma e^{-ax} & \in & Im[G_z(z, x)]. \end{array} \right. \tag{5.14}$$

To investigate whether $\mathcal{G}_{NS}$ is invariant under HW dynamics, we consider now simplest consistency condition, the CVC: let us consider for constants α_i, such $\forall x \geq 0$ we have

$$\sigma e^{-ax} = \alpha_1 + \alpha_2 e^{-z_4 x} + \alpha_3 x e^{-z_4 x} - \alpha_4 (z_2 + z_3 x) x e^{-z_4 x}. \tag{5.15}$$

One can easily see that is is possible iff $z_4 = a$. So as a first hint, let us fix $z_4 = a$ in the parametrization and introduce the restricted NS family:

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

Now the CVC is verified, while in the CDC we look for β_i such that $\forall x \geq 0$ we have:

$$(z_3 - a z_2 - a z_3 x) e^{-ax} + \frac{\sigma^2}{a} \left(e^{-ax} - e^{-2ax} \right) = \beta_1 + \beta_2 e^{-ax} + \beta_3 x e^{-ax}. \tag{5.16}$$

This equation can never be verified, due to the extra exponential e^{-2ax}, so we have proved the following.

Proposition 8 (Nelson-Siegel and Hull-White.) *The Hull-White model is*

inconsistent with the Nelson-Siegel family.

Let us include this extra exponential in the parametrization, thus, we finally introduce the augmented NS family:

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

Now, both CDC and CVC are verified for this family.

Proposition 9 (Augmented Nelson-Siegel and Hull-White.) *The augmented Nelson-Siegel family is consistent with the Hull-White model.*

Proof. The Frechet derivatives are in this case

$$\begin{aligned} \frac{\partial G_{ANS}}{\partial z}(z,x) &= [1 \quad e^{-ax} \quad xe^{-ax} \quad e^{-2ax}]^T, \\ \frac{\partial G_{ANS}}{\partial x}(z,x) &= [z_3 - a\,(z_2 + z_3 x)]\, e^{-ax} - 2az_4 e^{-2ax}. \end{aligned} \tag{5.17}$$

Now the set $Im[\partial_z G_{ANS}]$ is "large" enough to trivially satisfy CVC and CDC due to the extra component e^{-2ax}. The derivation of the balancing equations analogous to (5.15) and (5.16) are left to the meticulous reader. □

5.3 The Minimal Consistent family and Realizations of Gaussian Models

It should be also noted that $\sigma(x)$ is a one dimension *quasi-exponential* function (QE for short), because is of the form

$$f(x) = \sum_i e^{\lambda_i x} + \sum_i e^{\alpha_i x}[p_i(x)\cos(\omega_i x) + q_i(x)\sin(\omega_i x)], \tag{5.18}$$

with $\lambda_i, \alpha_i, \omega_i$ being real numbers and p_i, q_i are real polynomials.

If $f(x)$ is a q-dimensional QE function, then it admits the following matrix representation

$$f(x) = ce^{Ax}B, \tag{5.19}$$

where A is a $(n \times n)$-matrix, B is a $(n \times q)$-matrix and c is a n–dimensional row

vector, see Björk [5, Lemma 2.1, p. 13]. Thus, $\sigma(x)$ can be written as

$$\begin{aligned} \sigma(x) &= ce^{Ax}b, \text{ where} \\ c &= 1, \\ A &= -a, \\ b &= \sigma. \end{aligned} \tag{5.20}$$

We can write the forward rate equation (4.2) following Björk [5, Proposition 2.1, pp. 8–9]

$$dq_t(x) = \mathbf{F}q_t(x)\,dt + \sigma(x)\,dW_t, \quad q_0(x) = 0 \tag{5.21}$$

$$f_t(x) = q_t(x) + \delta_t(x), \tag{5.22}$$

here $\mathbf{F}$ is a linear operator that is defined by

$$\mathbf{F} = \frac{\partial}{\partial x},$$

and $\delta_t(x)$ is the deterministic process given by

$$\delta_t(x) = f^o(x+t) + \int_0^t \Sigma(x+t-s)\,ds,$$

with

$$\Sigma(x) = \sigma(x)\int_0^x \sigma(s)\,ds.$$

Moreover, $q_t(x)$ has the concrete *finite dimensional* realization

$$dZ_t = -aZ_t\,dt + \sigma\,dW_t, \quad Z_0 = 0, \tag{5.23}$$

$$q_t(x) = e^{-ax}Z_t, \tag{5.24}$$

as a particular result from [8, Definition 2.1, p. 7] with the fundamental concluding remark derived by Björk for QE deterministic volatilities in [8, Proposition 2.3, p. 13]. The SDE (5.23) is linear in the narrow sense [32], with explicit solution

$$Z_t = \sigma e^{-at}\int_0^t e^{as}\,dW_s, \tag{5.25}$$

Now, with the definition of $S(x) = \int_0^x \sigma(u)du$, it is easy to obtain that

$$\int_0^t \Sigma(t+x-s)\, ds = \frac{1}{2}\left[S^2(t+x) - S^2(x)\right],$$

and, therefore, combining these explicit results with decomposition (5.22) we arrive to the forward rate dynamics

$$f_t(x) = f^o(x+t) + \frac{1}{2}\left[S^2(t+x) - S^2(x)\right] + e^{-ax}Z_t. \tag{5.26}$$

Equation (5.26) may be used for building initial forward rate curves $f^o(x)$ time-consistent with the model.

5.3.1 The Minimal Consistent family

Proposition 10 *The family*

$$G_{MIN}(z,x) = z_1 e^{-ax} + z_2 e^{-2ax}, \tag{5.27}$$

is the minimal dimension consistent family with the model characterized by

$$\sigma(x) = \sigma e^{-ax}.$$

Proof. As we mentionead earlier, there is a way to justify (6.7) focusing on forward rate evolution deduced at (5.26) we describe it next. By the definition of $S(x)$, we have that $S'(x) = \sigma(x)$. Then it is easy to derive that deterministic term $\frac{1}{2}\left[S^2(t+x) - S^2(x)\right]$ is of the form

$$g(t)e^{-ax} + h(t)e^{-2ax}.$$

Thus, the forward rate evolution becomes

$$f_t(x) = f^o(x+t) + \left(g(t) + Z_t\right)e^{-ax} + h(t)e^{-2ax}. \tag{5.28}$$

From (5.28) we see that a family which is invariant under time translation is consistent with the model if and only if it contains the linear space $\{e^{-ax}, e^{-2ax}\}$. □

It should be also noted that the map

$$G(z,x) = G_{MIN}(z,x) + \phi(z,x),$$

where $\phi(\cdot)$, is an arbitrary function, is also consistent with this model.

Finally, we list some concluding remarks about the families analyzed.

Lemma 2 *The following hold for the Hull-White model*

- *The Nelson-Siegel family (henceforth NS)*

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

  is not consistent with the model.

- *The family*

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

  is the lowest dimension family consistent with the model (hereafter MIN).

- *The family*

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

  is the simplest adjustment based on restricted NS family that allows model consistency (hereafter ANS).

5.4 Calibration to Market Data Approaches

To calibrate the model by means of real data, we actually need to determine the vector of parameters $\boldsymbol{p} = [\ \sigma\ a\ ]^T$. In order to estimate the forward rate volatility, the statistical analysis of past data can be a possible approach, but the practitioners usually prefer implied volatility, laying within some derivative market prices, based techniques. This way involves a minimization problem where the loss function can be taken as

$$l_C(\boldsymbol{p}) = \sum_{i=1}^{n} (C_i^o - C_i(\boldsymbol{p}))^2,$$

where $C_i(\boldsymbol{p})$ are the i–th theoretical derivative price and C_i^o is the i–th market price one. As we proved on Sect. 3.2, the model price, at $t = 0$, of the cap with equidistant settlement periods is given by

$$C = \sum_{j=1}^{n} \gamma_j = (1 + \tau K) \left(\sum_{j=1}^{n} \kappa P_{j-1}(0) N(-d_-) - P_j(0) N(-d_+) \right), \tag{5.29}$$

where the $d_\pm$ are given by (3.18). Moreover, recall that $P_j(0)$ is the initial x_j-maturity discount bond price $P(0, x_j)$, κ equals to $(1 + \tau K)^{-1}$ with K denoting the *cap rate*. The volatility function $\vartheta(0, \cdot)$ defined by the expressions (3.19) to (3.20) take the particular form for the HW model:

$$\vartheta(0, x_j) = \frac{\sigma}{a} \left(1 - e^{-a\tau}\right) \sqrt{\frac{1 - e^{-2ax_j}}{2a}}.$$

The equations (3.18) and (5.29), also express the effective influence of *ab initio* discount bond curve estimation on cap pricing.

The calibration procedures can be described formally as follows.

Let $\boldsymbol{p}$ be the parameter vector

$$[\sigma \; a]^T$$

for the model under consideration. Assume that we have time series observations of the flat volatilities $\bar{\sigma}_i$ of N at-the-money caps C_i which mature at T_i-times where $i = 1, \ldots, N$. Suppose we are also equipped with the discount bond curve estimation, $P(0, x)$, at time $t = 0$. As we introduce in Sect. 3.1, market participants translate volatility quotes to cash quotes adopting Black model [13]. Thus, according to the Definition 15 in Sect. 3.1., the market price, at $t = 0$, of the cap with regular payment periods is given by

$$C^o = \sum_{j=1}^{n} \gamma_j^o = \tau \sum_{j=1}^{n} P_j(0) \left(L_j(0) N(d_1) - K N(d_2) \right), \tag{5.30}$$

where d_1 and d_2 are given by the identities (3.3) to (3.4).

In addition, recall that according to Remark 2 in Sect. 3.1.1 the cap rates for the ATM plain vanilla caps must fulfill (3.13). By direct inspection, it is clear that this market convention makes the rates K depend on the discount bond

curve estimation $P(0,x)$. Let us denote the market prices of caps by

$$C^o\left(T_i, P(0,x), K_i(P(0,x)), \bar{\sigma}_i\right).$$

Notice that this expression emphasizes explicit and implicit dependence (through ATM *strikes*) on discount bond curve estimation even for market prices. Let

$$C\left(T_i, P(0,x), K_i(P(0,x)), \boldsymbol{p}\right),$$

be the corresponding theoretical price under our particular model.

5.4.1 The Two-Step Traditional Method

Suppose that we are standing at time $t = 0$, the fixed time of calibration under study. For simplicity from now, we remove the dependency on initial time $t = 0$ from discount bond curve $P(0,x) \equiv P(x)$. First, we choose a non-consistent parametrized family of forward rate curves $G(z,x)$.

Let $P(z,x)$ be the zero-coupon bond prices produced by $G(z,x)$.

$$P(z) = [\, P_1(z) \;\ldots\; P_M(z) \,]$$

Let P_k^o be the corresponding discount x_k-bond observations with x_k-times running from $k = 1, \ldots, M$

$$P^o = [\, P_1^o \;\ldots\; P_M^o \,].$$

For each zero-coupon bond denoted with subscript k, the logarithmic pricing error[2] is written as follows

$$\epsilon_k(z) = \log P_k^o - \log P_k(z).$$

Then, we have chosen in this work the sum of squared logarithmic pricing errors, l_P, as the objective loss function to minimize:

$$\min_z l_P(z) = \min_z \| \log P^o - \log P(z) \|_2^2 = \min_z \sum_{k=1}^{M} \epsilon_k^2(z), \tag{5.31}$$

[2]Recall that, for small ϵ_k, it is also the relative pricing error $\frac{P_k^o - P_k(z)}{P_k^o}$.

with

$$\log P_k(z) = -\int_0^{x_k} G(z,u)\,du.$$

Now, via the least squares estimators $\hat{z}$, an entire discount bond curve estimation allows the pricing of caps using the market practice or the HW model. Following a similar scheme for the derivatives fitting as in the preceeding lines, we have

$$\eta_i(\boldsymbol{p}) = \log C_i^o - \log C_i(\boldsymbol{p}),$$

and

$$\min_{\boldsymbol{p}} l_C(\boldsymbol{p}) = \min_{\boldsymbol{p}} \| \log C^o - \log C(\boldsymbol{p})\|_2^2 = \min_{\boldsymbol{p}} \sum_{i=1}^{N} \eta_i^2(\boldsymbol{p}), \tag{5.32}$$

with the vector definitions:

$$C^o = [\, C_1^o \ \ldots \ C_N^o \,] \tag{5.33}$$

$$C(\boldsymbol{p}) = [\, C_1(\boldsymbol{p}) \ \ldots \ C_N(\boldsymbol{p}) \,]. \tag{5.34}$$

Note that here we have dropped the dependencies ($P(x)$, K, T, $\bar{\sigma}$) for simplicity. Moreover, notice that the discount bond curve estimation is external to the model in the sense that there is no need to know first any of the model parameters $\boldsymbol{p}$ for solving non-linear program (5.31).

5.4.2 The Joint Calibration to Cap and Bond Prices

Let us now describe in detail the joint cap-bond calibration procedure which has sense in a consistent family framework. We note that in this situation the parameters of the model are determined together with the initial forward rate curve.

This is different from the traditional fitting of the Hull-White model, where the two steps are separate, as we discussed before. From the expression (6.7), we notice the dependency of the family from the parameter a. Let $G(z,x,a)$ be a family consistent with the HW model (for instance, G_{MIN} and G_{ANS}) and define least-squares estimators, $\hat{z}(a)$

$$\hat{z}(a) = \arg\min_{z} \sum_{k=1}^{M} (\log P_k^o - \log P_k(z,a))^2. \tag{5.35}$$

From the expression

$$\log P_k(z,a) = -\int_0^{x_k} G(z,a,u)\,du = \sum_{j=1}^{n_p} M_{kj}(a) z_j, \tag{5.36}$$

we note that, for consistent families and for a fixed a the problem (5.35) is linear in z-parameters (for the G_{MIN} family $n_p = 2$, and for the G_{ANS} family $n_p = 4$). Thus, $\hat{z}$ is an explicit and continuous function of a.

Strictly speaking, joint calibration must be formalized as a multiobjetive optimization problem (MOO) of the form:

$$\min_{\boldsymbol{p}} \boldsymbol{l}(\boldsymbol{p}) \tag{5.37}$$

where $\boldsymbol{l} = [l_1(\boldsymbol{p})\, l_2(\boldsymbol{p})]^T$ is an objective function vector and $\boldsymbol{p}$ is the *design* vector $[\sigma\, a]^T$. Notice that in this case there are two objectives and two design variables. The *partial loss functions* $l_i(\sigma, a)$ are defined as

$$\begin{aligned} l_1(\boldsymbol{p}) &= \| \log C^o\left[P(\hat{z}(\boldsymbol{p}), \boldsymbol{p})\right] - \log C\left[P(\hat{z}(\boldsymbol{p}), \boldsymbol{p}), \boldsymbol{p}\right] \|_2^2 \\ l_2(\boldsymbol{p}) &= \| \log P^o - M(\boldsymbol{p})\hat{z}(\boldsymbol{p}) \|_2^2 \end{aligned}$$

where

$$\hat{z}(\boldsymbol{p}) = R(a)Q^{-1}(a) \log P^o \tag{5.38}$$

being Q, R the matrices of the reduced QR decomposition of $M(\boldsymbol{p})$ which is defined by the relation (5.36).

Note that it is highly probable that these objectives would both be conflicting, in general, and no single $\hat{\boldsymbol{p}} = (\hat{\sigma}, \hat{a})$ would generally minimize simultaneously the pair of objective functions l_i. Tipically here, there is no single, global solution, and often, it is necessary to determine a *set* of points that all fit a predetermined definition for an optimum. Thus, the predominant concept in defining is that of Pareto optimality [37].

One of the most common and basic approach for multiobjective optimization requires to build a weighted sum of the objectives (see for instance [18, 33]). The result is the following scalarized utility function, which is minimized:

$$\tilde{l} = \omega_1 l_1 + \omega_2 l_2. \tag{5.39}$$

When an appropriate set of solutions is obtained by the single-objective optimization of $\tilde{l}$, the solutions can approximate a Pareto front. The weighted-sum method parametrically changes the weights among objective functions l_1 and l_2 to obtain this Pareto front. If the two weights are positive then minimizing the utility provides a sufficient condition for Pareto optimality, which means the minimum of (5.39) is always Pareto optimal [33, Sect. 4.1.1, pp. 41–42]. Thus, consistent calibration carried out with consistent families involves the entire Pareto optimal set, in contrast to the unique solution appearing in the two-step scalar problem.

At this point, note that the program used by Angelini and Herzel [1, 2] in their works, uses a different goal attainment

$$\min_{\boldsymbol{p}} l_1(\boldsymbol{p}) \tag{5.40}$$

where $l_1(\boldsymbol{p})$, and $\hat{z}(\boldsymbol{p})$ are defined trough the identities (5.32) and (5.35). As a consequence, the program used by these authors is a degenerate case of (5.39) with ω_1 fixed equal to 1 and ω_2 to 0, so it just allows to obtain one point of the implied trade-off front, which would be potentially only a weak Pareto optimum[3] (WPO) not a standard Pareto optimum [33, Sect. 4.1.1, p. 42].

5.5 Empirical Results

In this context the main goal is to analyze the impact that an alternative interpolation scheme has on the fitting capabilities of the model. To this end, we use as a measure, the daily (on average) relative pricing errors, hereafter RPE_C:

$$RPE_C = \frac{1}{N}\sum_{i=1}^{N}\frac{|C_i^o - C_i(\hat{\boldsymbol{p}})|}{C_i^o}$$

The same kind of measure is used for the zero-coupon bond prices and we denote it with RPE_B:

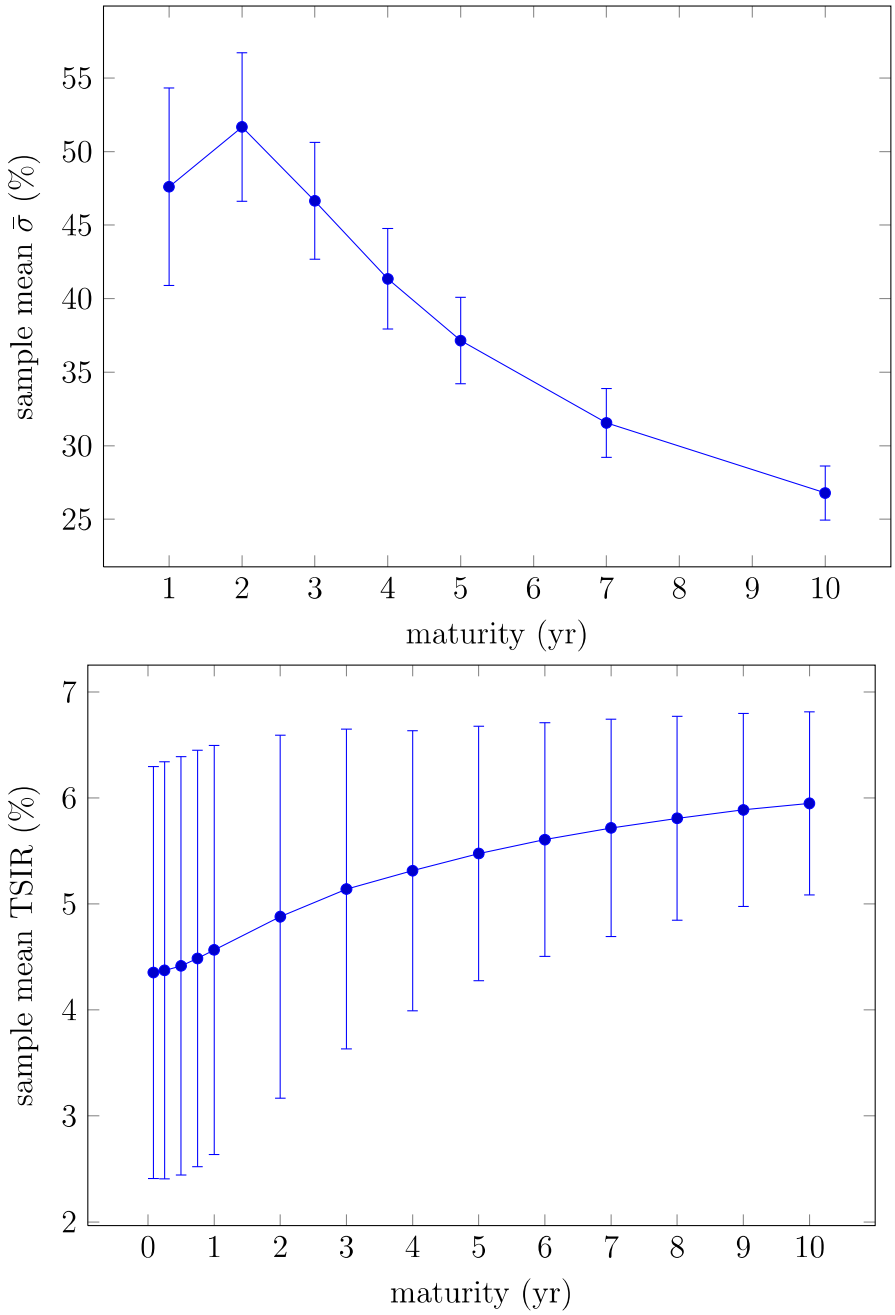
$$RPE_B = \frac{1}{M}\sum_{k=1}^{M}\frac{|P_k^o - P_k(\hat{z}(\hat{\boldsymbol{p}}), \hat{\boldsymbol{p}})|}{P_k^o}$$

We perform such analysis focusing on US market. The real date consists of 282 daily observations, between 2/09/2002 and 30/09/2003. The data set is composed of US discount factors for fourteen maturities (1, 3, 6, 9 months and from 1 to 10

[3] Pareto optimal points are WPO, but WPO are not Pareto optimal.

years) and of implied volatilities of at-the-money interest rate caps with maturities 1,2,3,4,5,7,10 years. This database is provided by Thomson Reuters Datastream.

Figure 5.1: Average of the US market TSIR and TSV with 99% confidence levels.



As it have been explored before, the daily joint calibration of caps and bonds with consistent families must be properly carried out as a MOO problem. In doing so, we choose an *a posteriori articulation* of preferences, that is, we delay

the selection of the solution from the palette of solutions after the weighted-method runs. In response to this articulation of preferences, the decision-maker imposes preferences directly on a set of the potential solution points which depicts the Pareto front. In such a context, weights are typically chosen such that

$$\sum_{i=1}^{Q} \omega_i = 1 \tag{5.41}$$

with $\boldsymbol{\omega} \geq \mathbf{0}$ leading to a convex combination of objectives and this choice can be more helpful than unrestricted weights so as not to repeat any weighting vectors in terms of its relative values [33, Sect. 5.3, pp. 73–74]. According to this, the weighted-sum method was run for every date in sample with the fixed weight vector $\boldsymbol{\omega}$ satisfying (5.41). In doing so, we assume the same ten uniformly spread values

$$\omega_1 = \frac{1}{10} j \qquad j = 1, 2, \ldots, 10 \tag{5.42}$$

and $\omega_2 = 1 - \omega_1$ as the second vector weight component for all trading dates.

We use the following transformation scheme of the objectives, which is often called *scaling* [38]:

$$l_i^{\tau} = \frac{l_i(\boldsymbol{p})}{s_i} \tag{5.43}$$

$$\text{with} \qquad \frac{l_1(\boldsymbol{p}_0)}{s_1} \approx \frac{l_2(\boldsymbol{p}_0)}{s_2} \tag{5.44}$$

where s_i are scalar coefficients, and $\boldsymbol{p}_0$ is a feasible starting point. This approach ensures the objective functions have similar orders of magnitude. Thus, the way to solve our joint calibration problem is to use the weighted-sum method, which is finally stated as:

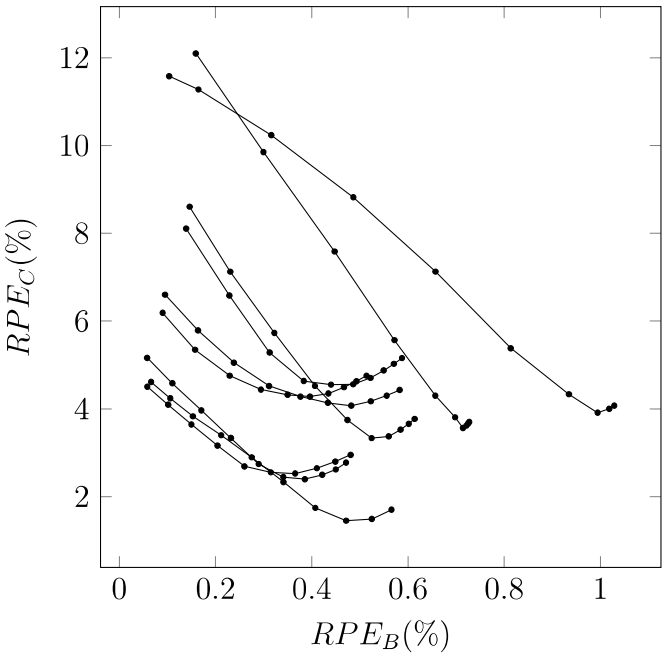
$$\min_{\boldsymbol{p}} \left(\omega_1 \frac{l_1(\boldsymbol{p})}{s_1} + (1 - \omega_1) \frac{l_2(\boldsymbol{p})}{s_2} \right) \tag{5.45}$$

with ω_1 discrete values given by (5.42).

Figure 5.2 shows the in-sample fitting results performed by the MIN family for some dates of the sample under analysis. First of all, notice that efficient frontiers with regular shapes appear nicely revealing the intrinsic multi-objective nature of the consistent calibration. Moreover, note that it can be found different topologies for this frontiers depending on the date. All the days the objectives are conflicting, and beyond a certain point of the Pareto front the better we fit

the discount bonds the worse we calibrate the caps portfolio. However, note that, moving on to the Pareto curve, we can achieve better results for both components of the vector objective without a trade-off until the above-mentioned Pareto point is reached. In other words, the MOO calibration may provide a *better* set of results for the calibration of caps that would produced by single optimizing the scalar objective $l_1(\boldsymbol{p})$.

Figure 5.2: Some daily calibration results for the minimal consistent family.



The tables on Figure 5.3 show, as a numerical example, two different Pareto sets restricting ourselves to the MIN family, the family with the lowest dimensionality. If we look on both tables, it must be noted that for a fixed trading date the best cap fit results may occur with $\omega_1 \neq 1$, even if the objectives are competing.

In Figure 5.4, we analize more deeply the latter fact this time for both, MIN and ANS, consistent families. We plot the daily distribution of the weight ω_1 which performs the best calibration for caps in sample data. As for the lowest dimensional family, most of the days the weight vector $(\omega_1 = 0.7, \omega_2 = 0.3)$ produces the best cap calibration results and there is a non-negligible number of bussiness dates where other weights than $\omega_1 = 1.0$ produce better goals than it. As for the ANS family, we observe an entirely different distribution, but once

Figure 5.3: Efficient points in the $RPE_B - RPE_C$ space using the method of convex combinations for two different days in sample. The partial objectives, ω_1 and ω_2 are strong conflicting, for the Day 1 (top). In contrast, the latter ones are more cooperative for the Day 2 (bottom).

Day 1

ω_1	ω_2	RPE_B (%)	RPE_C (%)
0.1	0.9	0.1036	11.5795
0.2	0.8	0.1645	11.2761
0.3	0.7	0.3160	10.2359
0.4	0.6	0.4864	8.8214
0.5	0.5	0.6573	7.1267
0.6	0.4	0.8136	5.3813
0.7	0.3	0.9343	4.3329
0.8	**0.2**	**0.9941**	**3.9132**
0.9	**0.1**	**1.0181**	**4.0016**
1.0	**0.0**	**1.0287**	**4.0718**

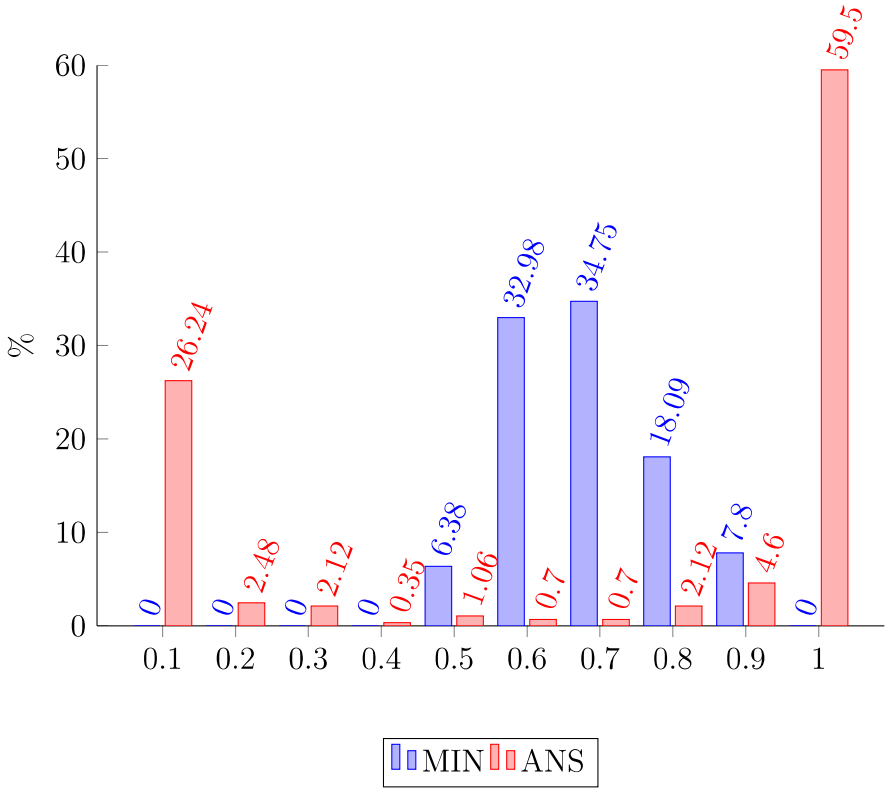
Day 2

ω_1	ω_2	RPE_B (%)	RPE_C (%)
0.1	0.9	0.0902	6.1874
0.2	0.8	0.1574	5.3478
0.3	0.7	0.2295	4.7551
0.4	0.6	0.2944	4.4395
0.5	0.5	0.3499	4.3203
0.6	**0.4**	**0.3962**	**4.2797**
0.7	**0.3**	**0.4345**	**4.3490**
0.8	**0.2**	**0.4670**	**4.4933**
0.9	**0.1**	**0.4930**	**4.6303**
1.0	**0.0**	**0.5138**	**4.7546**

again, we note that the best cap calibration results may be reached whith weights different than $\omega_1 = 1.0$ and even get a large number of them with the smallest weight possible, $\omega_1 = 0.1$.

For the shake of simplicity, from now on we will only consider the calibration results obtained with daily weights choices that produce the best calibration for the caps on every trading date. This *a posteriori* articulation of preferences may be followed by a decision-maker which want to use consistent calibration as a good risk management practice or as an extrapolation tool for marking to

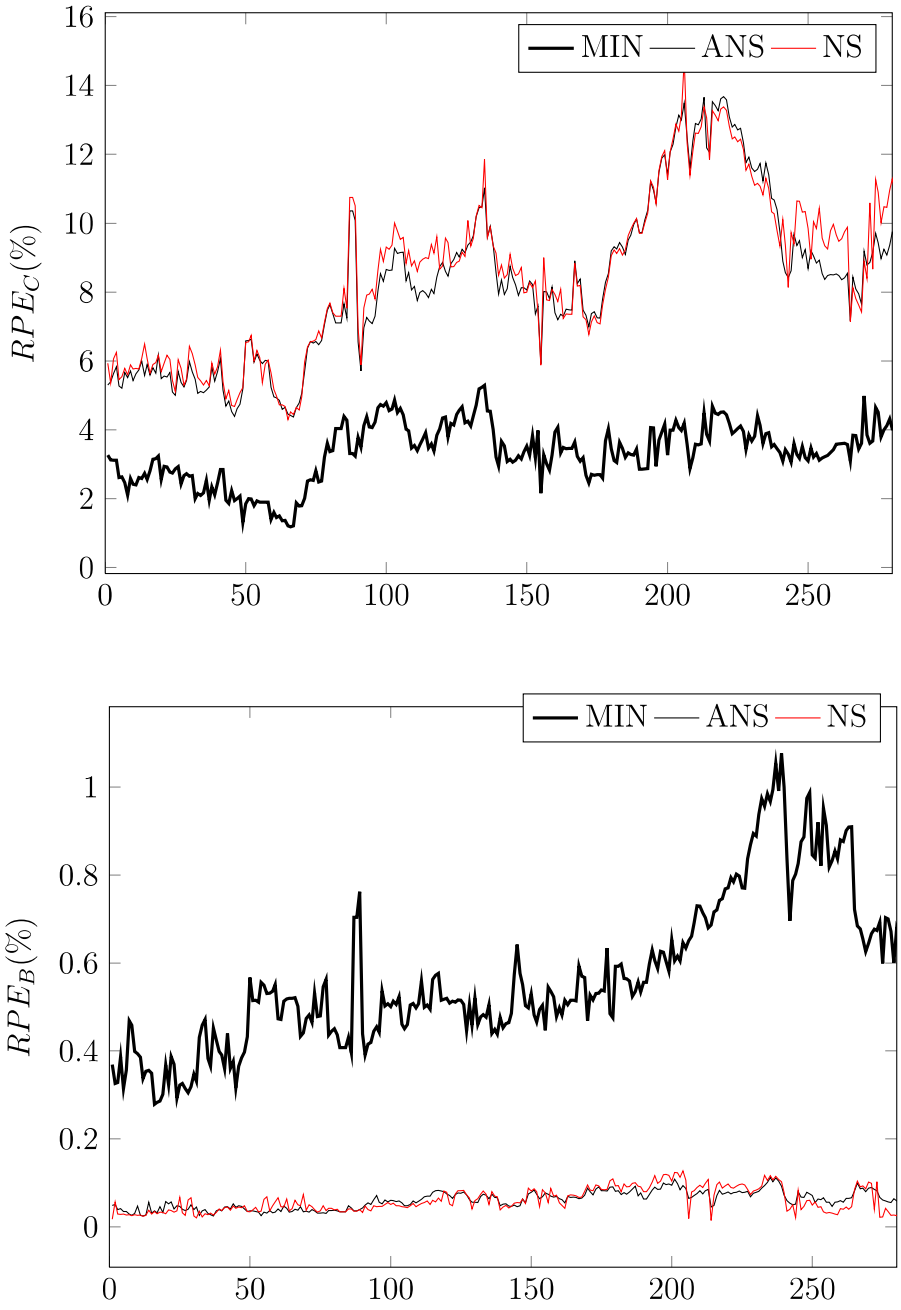
Figure 5.4: Daily empirical distribution of weights with the best RPE_C for both consistent families as produced by the multi-objective calibration.



market less liquid interest-rate derivatives. Following this particular articulation of preferences we choose a single Pareto optimum from the set that estimates the complete Pareto curve. In Figure 5.6, we compare summary statistics of the parameter estimates and the in-sample fit measures reported by NS, MIN and ANS families. In addition, Figure 5.5 shows the comparison of in-sample fitting results in time series.

The two consistent families under study report better RPE results when we restrict the analysis to cap data. For RPE on bonds, only the ANS family outperforms NS in the sample. Recall that this fact is acceptable since the MIN family is a forward curve with less number of parameters than the other ones proposed. Moreover, on caps, note that the MIN family appears to give better results than its consistent counterpart, ANS. Now, this behaviour can be explained because the major of dates considered, market data make the objective functions l_1 and l_2 of this family to strongly conflict as seen in Figure 5.4.

Figure 5.5: Time Series Comparison.



5.6 Concluding Remarks

When calibrating the Hull-White model, a TSIR curve choice to fit a few market data observations is needed. In particular it seems to be natural to use families of curves which do not modify their structure under the future evolution of the model, the so-called consistent families.

Figure 5.6: Summary statistics for the calibration results. In-sample descriptive statistics are carried out using the daily Pareto points with the best derivative fit outcomes.

Summary Statistics

	MIN	ANS	NS
σ	0.0145	0.0127	0.0164
a	-0.0096	0.0243	0.0256
$C_v(\sigma)$	0.1113	0.3118	0.14
$C_v(a)$	-3.2659	2.4268	1.8831
RPE_C (%)	3.3361	8.3011	8.5458
RPE_B (%)	0.5688	0.061	0.0609

In this work, we choose three families of curves (two consistent families and the popular Nelson-Siegel family) and we conclude that this choice have an effective impact on the quality of in-sample fitting for US-market data. Moreover, this paper extends the seminal calibration algorithm proposed in Angelini and Herzel [1].

In a consistent approach the parameters of the model are estimated jointly with the esmation of initial discount bond curve. Therefore, from a rigorous point of view, joint calibration of caps and bonds must be viewed as a multi-objective optimization nonlinear problem. Although the main purpose of the algorithm is to minimize the relative differences of cap prices too, note that the bi-objective extension of the consistent calibration presents more general features. Such extension is structured to allow more numerical outcomes and we observe that it allows to better fit results for both, caps and bonds, than the above mentioned. In particular, it is possible to find better cap calibration outcomes with $\omega_1 \neq 1$, and this is definitively different from what worked Angelini and Herzel [1] on Hull-White model, where only the fixed $\omega_1 = 1$ seems to be considered for all consistent families. The empirical findings of this paper show that, in general, consistent calibration on every date must to be carried out by analyzing the entire shape of the Pareto curve.

In this sense, this work confirms and complements the shown by Angelini and Herzel [1, 2] restricted to a Euro data set. We articulate preferences restricting possible outcomes on every date, by choosing the Pareto points which are responsible of better fit results on caps. Then the minimal consistent family gives the

best performance in terms of caps pricing errors and becomes a good candidate for the calibration of the Hull-White model. The ANS consistent family performs very close to the Nelson-Siegel family, though it seems to be the best solution for estimating the discount bond function. Now, this could be explained in the context of vector optimization. We show empirically the usual competing behaviour followed by the objectives through the sample considered. Then, the minimal parameterized consistent family relax the performance on the estimation of the discount bond curve function, allowing minor relative pricing errors on caps.