HEATH–JARROW–MORTON INTEREST RATE DYNAMICS AND
APPROXIMATELY CONSISTENT FORWARD RATE CURVES

CLAUDIA LA CHIOMA AND BENEDETTO PICCOLI

Abstract. We study a finite-dimensional approach to the Heath–Jarrow–Morton model
for interest rate and introduce a notion of approximate consistency for a family of func-
tions in a deterministic and stochastic framework. This amounts to asking the decrease
of the minimum distance in least squares sense. We start from a general linearly pa-
rameterized set of functions and extend the theory to a nonlinear Nelson–Siegel family.
Necessary and sufficient condition to have approximately consistency are given as well
as a criterion of stability for the approximation.

1. Introduction

In this paper we present some results on the finite dimensional approach to Heath–
Jarrow–Morton interest rate modelling. The most used approach for the term structure
is the infinite dimensional HJM model: it is well known that one model is determined
by the choice of a particular form for the volatility structure $\sigma(t, T)$, and the dynamic is
consequently computed after arbitrage theory consideration. In spite of its capability to
capture the evolution of the interest rate, such model is difficult to handle, since it evolves
following an infinite dimensional stochastic equation (one for every considered maturity).

Then it is interesting to investigate if an infinite dimensional forward rate model can be
approximated by a finite dimensional stochastic system. This approach has been motivated
by a requirement from practitioners: whenever we have to calibrate a model to a set of data,
we deal with a finite set of maturities, even if possibly large.

Several contributions were provided in this setting by Björk and alt. [2, 3, 5, 4] and
Filipović and Teichmann [8, 9, 10] with a geometric approach. They characterized
the existence of a local finite dimensional realization giving conditions on the volatility function
and on the initial curve.

In this paper we propose a new approach introducing a notion of approximately consistent
family of forward curve in a deterministic and stochastic sense. Recalling the fact that in
practice we deal with a finite set of maturities, we ask compatibility of a finite dimensional
family only at this finite set of points. More precisely, first, measuring an approximation
error by mean squares, approximate consistency is stated as the decrease in time of the
minimal error with respect to the mean of the rates.

We start from the study of the Nelson–Siegel family as a key example and continue the
analysis with more general linearly parameterized set of functions. By direct computations
we can provide some necessary and sufficient conditions for consistency: Theorem 4.1 and
4.5. An interesting consequence is the possibility of giving explicit conditions for consistency
for the case of polynomial functions, see Corollary 4.2, and for special choice of the volatility

Date: April 17, 2006.

Key words and phrases. interest rate modelling, HJM models, NS families, finite dimensional family,
consistency.
function assuming that the HJM evolution provokes translation movements independent of time or of maturity, see Example 4.1. Some results are then extended in case of fitting of the whole term structure, thus with respect to the error in $L^2$.

Then, in Section 6, we assume that the parameters of the family may vary according to a stochastic evolution. The determination of a stochastic process for parameters minimizing the expected error is a too difficult problem, thus we study the error of the process obtained by least squares for every couple $(t, \omega)$ and consequently define approximate consistency. Then the stochastic error is defined as the expected value of the approximation error for such process and is estimated in terms of the deterministic error.

Then we deal with the four parameters Nelson and Siegel family as first example of not linearly parameterized family. In fact, the last parameter $Z_4$ appears in some exponential function. In this case the existence of parameters realizing the least squares error is not immediately granted. However, for a fixed $Z_4$, the family is linear and we reduce to the previous case. Thus we give conditions ensuring the existence of the minimum, computing the asymptotic behavior of the least squares error as function of the parameter $Z_4$ for $Z_4$ tending to infinity, see Proposition 7.1. These conditions can be stated in terms of some variations of data with respect to maturities.

Finally, we define stability as the persistency of existence of a minimum for least squares error. More precisely, assuming that at initial time one has perfect data fitting, the aim is to estimate the maximal time for which existence of minimum still holds. We provide a formula to estimate this maximal time in the general case.

2. Heath–Jarrow–Morton interest rate models

One of the most important object in financial markets is the interest rate. There are two ways to model its term structure in continuous time: the equilibrium approach and the arbitrage approach. The former starts from a description of the underlying economy, the evolution of exogenous factors and of the preferences of an agent; the latter starts from a choice of the stochastic evolution of one or more interest rates and derives the prices of derivatives under the assumption of no arbitrage opportunities. In this paper we focus our attention on the last approach.

The choice of a volatility for an HJM model can be based on the description of the short rate, as in [6, 7] or [12], attempting to obtain the shape of the entire term structure, or describing the forward rate curve, having fixed the initial data, that is the short rate. A great number of articles dealt with the framework of pure diffusion models, the attention having been focused on the study of the parameters.

The approach introduced by Heath-Jarrow-Morton [11] chooses as state variable the whole yield curve of the term structure, whose dynamic is given by

$$df(t, T) = \mu(t, T)dt + \tilde{\sigma}(t, T) \cdot dW_t;$$

here $t$ is the running time and $T$ is the maturity. To avoid arbitrage opportunities the following condition has to be satisfied:

$$\mu(t, T) = \tilde{\sigma}(t, T) \int_t^T \tilde{\sigma}^T(t, s)ds$$

This equation shows that choosing the volatility parameter is sufficient to get a complete description of the model.

We note that the only important variable in this modelling is the time to maturity, $x = T - t$, therefore it is better to use the Musiela parametrization: call $r(t, x) = f(t, t + x)$,
the evolution equation is
\[ dr(t, x) = \left( \frac{\partial}{\partial x} r(t, x) + \sigma(t, x) \int_0^x \sigma^T(t, y) dy \right) dt + \sigma(t, x) dW_t, \]
where \( \sigma(t, x) = \tilde{\sigma}(t, t + x) \).

3. Nelson–Siegel Family

In this section we consider a three-parameters Nelson–Siegel model to approximate the
interest rate evolution when the data are known only on a finite number of maturities. Thus
the space in which the approximation is studied is \( \mathbb{R}^M \), for some \( M \in \mathbb{N} \).

Let \( (\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \in \mathbb{N}}, \mathbb{P}) \) be a filtered probability space, \( \mathbb{P} \) the risk-neutral measure, and \( W \) a
one-dimensional standard Brownian motion. Assume that for every fixed \( T \) the correspond-
ing term structure of forward rate \( f(t, T) = r(t, x) \) evolves following the HJM model
\[ dr(t, x) = \left( \frac{\partial}{\partial x} r(t, x) + \sigma(t, x) \int_0^x \sigma(t, y) dy \right) dt + \sigma(t, x) dW_t, \quad 0 < t \leq T. \]
We can observe, today, the interest rates \( f_{d, k}(0) \) corresponding to \( M \) different ma-
turities \( x_k \), and consider only the evolutions of the corresponding forward interest rates
\[ dr(t, x_k) = \left( \frac{\partial}{\partial x} r(t, x_k) + \sigma(t, x_k) \int_0^{x_k} \sigma(t, y) dy \right) dt + \sigma(t, x_k) dW_t, \quad k = 1, \ldots, M \]
with initial datum:
\[ r(0, x_k) = d_{0, k} := d_k(0). \]

Remark 3.1. In order to perform the following calculations, we need \( d_k(0) \) for all \( k \in \mathbb{R} \).
These values can be obtained by interpolation from the \( M \) initial benchmarks values.

For any \( t > 0 \) we define
\[ d_x(t) = d(t, x) := \mathbb{E}[r(t, x)] = d_x(0) + \mathbb{E}\left[ \int_0^t \left( \frac{\partial}{\partial x} r(s, x) + \sigma(s, x) \int_0^x \sigma(s, u) du \right) ds \right]. \]
Using the result in [1, Prop. 18.4] we have
\[ \frac{\partial}{\partial t} d(t, x) = \frac{\partial}{\partial x} d(t, x) + \mathbb{E} \left[ \sigma(t, x) \int_0^x \sigma(s, u) du \right]. \]
Using the characteristic method we get
\[ d(t, x) = d(0, t + x) + \int_0^t \mathbb{E} \left[ \sigma(s, x + t - s) \int_0^{x+t-s} \sigma(s, u) du \right] ds. \]
Therefore, if we assume for simplicity \( x_k = k \), on the set of observed maturities we get
\[ d_k(t) = d_{t+k}(0) + \mathbb{E}\left[ \int_0^t \sigma(s, t + k - s) \int_0^{t+k-s} \sigma(s, u) du ds \right]. \]
Consider a three-parameters Nelson–Siegel family
\[ G(Z(t), k) = Z_1(t) + Z_2(t)e^{-ak} + Z_3(t)ke^{-ak}, \]
where \( Z(t) \in \mathbb{R}^3 \), \( a > 0 \) and \( k \) is the time to maturity, \( k = 1, \ldots, M \).
Our aim is to minimize the distance, in the least squares sense, between the Nelson–Siegel family and the HJM dynamics at any time \( t \). Once \( \{d_k(t)\}_k \) is computed we let \( Z^{\text{det}} \) the point realizing the minimum distance. Thus we define

\[
\Phi^{\text{det}}(t) := \sum_{k=1}^{M} \left[ G(Z^{\text{det}}(t), k) - d_k(t) \right]^2,
\]

which represents a “deterministic” error. The existence and uniqueness of \( Z^{\text{det}} \) is assured by the standard theory, as follows. For instance, at time \( t = 0 \) the error is given by

\[
\Phi^{\text{det}}(0) = \min_{Z \in \mathbb{R}^3} \sum_{k=1}^{M} \left[ G(Z, k) - d_k(0) \right]^2.
\]

Stationary points of such cost are the solutions of

\[
\begin{align*}
2 \sum_{k=1}^{M} \left[ Z_1 + Z_2 e^{-ak} + Z_3 k e^{-ak} - d_k(0) \right] &= 0, \\
2 \sum_{k=1}^{M} \left[ Z_1 + Z_2 e^{-ak} + Z_3 k e^{-ak} - d_k(0) \right] e^{-ak} &= 0, \\
2 \sum_{k=1}^{M} \left[ Z_1 + Z_2 e^{-ak} + Z_3 k e^{-ak} - d_k(0) \right] k e^{-ak} &= 0;
\end{align*}
\]

this is equivalent to

\[
J(a, M) Z = D(0, a, M),
\]

where:

\[
J(a, M) := \begin{pmatrix}
\sum_{k=1}^{M} e^{-ak} & \sum_{k=1}^{M} k e^{-ak} \\
\sum_{k=1}^{M} e^{-2ak} & \sum_{k=1}^{M} k e^{-2ak} \\
\sum_{k=1}^{M} k e^{-ak} & \sum_{k=1}^{M} k e^{-2ak}
\end{pmatrix},
\]

\[
D(0, a, M) := \begin{pmatrix}
\sum_{k=1}^{M} d_k(0) \\
\sum_{k=1}^{M} d_k(0) e^{-ak} \\
\sum_{k=1}^{M} k d_k(0) e^{-ak}
\end{pmatrix}.
\]

We can note that \( \det J(a, M) \neq 0 \) for all \( a > 0 \) and for all \( M \in \mathbb{N} \), therefore the previous system has an unique solution, \( Z^{\text{det}}(0) = (Z^{\text{det}}_1(0), Z^{\text{det}}_2(0), Z^{\text{det}}_3(0))^T \). Notice that the
Hessian matrix coincides with $J(a, M)$: its invariants are all positive, thus it is positive definite, and it follows that $Z^{\text{det}}(0)$ is the unique minimum point.

Arguing as before, for any $0 \leq t \leq T$, $Z^{\text{det}}(t)$ is given by

$$Z^{\text{det}}(t) = J(a, M)^{-1} D(t, a, M).$$

$Z^{\text{det}}$ gives us the best approximation, $t$ being fixed, while the quantity $\Phi^{\text{det}}(t)$ gives the error we incur in substituting the HJM model with the NS family evaluated at $Z^{\text{det}}$. We focus now our attention to the time evolution of the error: $Z^{\text{det}}$ evolves as

$$\dot{Z}^{\text{det}}(t) = < J(a, M)^{-1} \left( \begin{array}{ccc} 1 & e^{-a} & e^{-a} \\ \vdots & \vdots & \vdots \\ 1 & e^{-ak} & ke^{-ak} \\ \vdots & \vdots & \vdots \\ 1 & e^{-akM} & Me^{-akM} \end{array} \right) >,$$

where $< \cdot, \cdot >$ is the standard scalar product.

Once we know the dynamics of the best approximation we get that $\Phi^{\text{det}}(t)$ evolves according to

$$\dot{\Phi}^{\text{det}}(t) = -2 \sum_{k=1}^{M} \left[ G(Z^{\text{det}}(t), k) - d_k(t) \right] \nabla_Z G(Z^{\text{det}}(t), k) \cdot \dot{Z}^{\text{det}}(t) - \dot{d}_k(t),$$

(3.7)

where we have used that $Z^{\text{det}}(t)$ solves $G(Z^{\text{det}}(t), k) = 0$, see (3.6).

We introduce now the notion of approximate consistency.

**Definition 3.1.** We say that the Nelson–Siegel family (3.4) is **approximately consistent** with (3.1) if

$$\dot{\Phi}^{\text{det}}(t) \leq 0 \text{ for all } t \in [0, T].$$

To perform the study of this inequality, we can describe the family (3.4) in the following way. Let $G : \mathbb{R}^3 \rightarrow \mathbb{R}^M$ be the linear mapping such that $Z \mapsto G(Z) = (G(Z, k))_{k=1, \ldots, M}$:

$$G(Z) = GZ = \left( \begin{array}{ccc} 1 & e^{-a} & e^{-a} \\ \vdots & \vdots & \vdots \\ 1 & e^{-ak} & ke^{-ak} \\ \vdots & \vdots & \vdots \\ 1 & e^{-akM} & Me^{-akM} \end{array} \right) Z,$$

and let $d(t) = (d_1(t), \ldots, d_M(t)) \in \mathbb{R}^M$. The values of the family described by $G$ form a linear space, $\mathcal{G}$, whose vector basis is $\{g_1, g_2, g_3\}$,

$$g_1 = 1^T \in \mathbb{R}^M,$$
$$g_2 = (e^{-ak}, \ldots) \in \mathbb{R}^M,$$
$$g_3 = (ke^{-ak}, \ldots) \in \mathbb{R}^M.$$

As before, for any $t$ fixed, we can find $Z^{\text{det}}$ as a solution of

$$2(G \cdot (Z - d(t))) \cdot G = 0.$$

Defining

$$\pi_\mathcal{G} : \mathbb{R}^M \rightarrow \mathbb{R}^M,$$ projection onto $\mathcal{G}$,

$$\pi_{\mathcal{G}^\perp} : \mathbb{R}^M \rightarrow \mathbb{R}^M,$$ projection onto $\mathcal{G}^\perp$,
the previous equation tells us that the projection of the vector \( G \cdot Z - d(t) \) onto the linear space \( \mathcal{G} \) is null, that is that the vector is orthogonal to \( \mathcal{G} \).

Completing the basis of \( \mathcal{G} \) with \( \{g_4^\perp, \ldots, g_M^\perp\} \) we can write
\[
\dot{d}(t) = \sum_{i=1}^{3} \alpha_i(t)g_i + \sum_{i=4}^{M} \beta_i(t)g_i^\perp,
\]
for some \( \alpha, \beta : [0, T] \rightarrow \mathbb{R} \).

This approach suggests us to work in an orthogonal basis. In this perspective we describe a more general problem of which the Nelson–Siegel family is a particular case.

### 4. Finite Dimensional Families

Let \( \mathbf{G} : \mathbb{R}^N \rightarrow \mathbb{R}^M \) be a linear application, \( N \) the dimension of the vector \( Z(t) \) and \( M \) the number of maturities, given by
\[
(G(Z(t), k) = \sum_{i=1}^{N} Z_i(t)g_i(k) = <(Z_1(t), \ldots, Z_n(t))^T, (g_1(k), \ldots, g_n(k))^T>,
\]
where \( \{g_1, \ldots, g_n\} \) is an orthonormal set, \( g_i \in \mathbb{R}^M \). Arguing as in Section 3, if \( N < M \) we can complete the \( \mathcal{G} \) basis with an orthonormal system of vectors obtaining an orthonormal basis for \( \mathbb{R}^M \); let \( \{g_{n+1}^\perp, \ldots, g_M^\perp\} \) be this completion. The definition of \( \Phi^{\text{det}} \) and of approximate consistency are the same as before. In this system we have
\[
\dot{d}(t) = \sum_{i=1}^{N} \alpha_i(t)g_i + \sum_{i=N+1}^{M} \beta_i(t)g_i^\perp,
\]
for some \( \alpha_i, \beta_i : [0, T] \rightarrow \mathbb{R} \).

For any \( 0 \leq t \leq T \) fixed, we are looking for a \( Z^{\text{det}}(t) \) such that
\[
\min_{Z(t) \in \mathbb{R}^N} \sum_{k=1}^{M} |G(Z(t), k) - d_k(t)|^2 = \sum_{k=1}^{M} |G(Z^{\text{det}}(t), k) - d_k(t)|^2,
\]
therefore
\[
\sum_{k=1}^{M} |G(Z^{\text{det}}(t), k) - d_k(t)|\nabla_{Z} G(Z^{\text{det}}(t), k) = 0,
\]
that, by the linearity of \( \mathbf{G} \), is
\[
\sum_{k=1}^{M} |G(Z^{\text{det}}(t), k) - d_k(t)| \begin{pmatrix} g_1(k) \\ \cdots \\ g_n(k) \end{pmatrix} = 0.
\]
This equation has an unique solution given by
\[
Z_j^{\text{det}}(t) = \sum_{k=1}^{M} d_k(t)g_j(k), \text{ for all } j = 1, \ldots, N.
\]
This solution can be easily found using matrix calculus. Looking at (4.3) we see that the vector \( (G(Z^{\text{det}}(t), k) - d_k(t))_{k=1, \ldots, M} \) is orthogonal to the linear space \( \mathcal{G} \). Let us now analyze the evolution of the error (3.5).
Using (4.3), we can rewrite (3.7) as

$$\Phi^{\text{det}}(t) = -2 \sum_{k=1}^{M} \left[ G(Z^{\text{det}}(t), k) - d_k(t) \right] \left( \pi_G(\tilde{d}(t)) k + \pi_{G^+}(\tilde{d}(t)) k \right)$$

$$= -2 \sum_{k=1}^{M} \left[ G(Z^{\text{det}}(t), k) - d_k(t) \right] \pi_{G^+}(\tilde{d}(t)) k.$$ 

**Theorem 4.1.** Assume \( \tilde{d}(t) \in \mathcal{G} \), then \( \Phi^{\text{det}}(t) = 0 \). Hence the family (4.1) is approximately consistent with (3.1).

**Proof.** By assumption, \( \beta_i(t) = 0 \) for all \( i = n + 1, \ldots, M \) in equation (4.2), therefore

$$d_k(t) = \sum_{i=1}^{N} \alpha_i(t) g_i(k), \text{ for all } k = 1, \ldots, M.$$ 

Inserting in (3.7),

$$\Phi^{\text{det}}(t) = -2 \sum_{k=1}^{M} \left[ G(Z^{\text{det}}(t), k) - d_k(t) \right] \pi_{G}(\tilde{d}(t)) k$$

$$= -2 \sum_{i=1}^{N} \alpha_i(t) \sum_{k=1}^{M} \left[ G(Z^{\text{det}}(t), k) - d_k(t) \right] g_i(k) = 0.$$ 


Beside this general result, our aim is to provide a more precise analysis for some specific family of functions and volatility \( \sigma \). Therefore from now on we consider special choices for the volatility \( \sigma \) in order to get approximate consistency. Consider first the following assumption

(\( \mathcal{H} \)) Assume that \( \mathcal{G} \) contains the vector \( g_1 = \frac{1}{\sqrt{M}} (1, \ldots, 1) \) and that \( g_i \) is a vector whose components are polynomial of degree \( i - 1 \), \( i = 1, \ldots, N \).

Under assumption (\( \mathcal{H} \)) we can prove the following:

**Corollary 4.2.** If (\( \mathcal{H} \)) holds, the initial curve verifies \( \frac{\partial}{\partial x} r(0, x) = b \) and the volatility \( \sigma \) is constant, then the finite dimensional family \( \mathcal{G} \) is approximately consistent with the HJM model.

**Proof.** By assumption \( g_1 = \frac{1}{\sqrt{M}} (1, \ldots, 1) \), and \( g_i(k) = \sum_{l=0}^{i-1} a_{i}^l k^l \), polynomials in \( k \) for \( i = 2, \ldots, N \); the coefficients \( a_i^l \) are chosen to satisfy the condition of orthonormal basis and \( a_i^{i-1} \neq 0 \), also \( \sigma(t, k) = \sigma \), then, from equation (3.3) we get

$$d_k(t) = d_{i+k}(0) + \sigma^2 \left( \frac{t^2}{2} + tk \right),$$

therefore

$$\dot{d}_k(t) = \frac{\partial}{\partial x} d_{i+k}(0) + \sigma^2 (t + k);$$

the hypothesis of Theorem 4.1 holds if

$$\dot{d}_k(t) = \sum_{i=1}^{N} \alpha_i(t) g_i(k);$$
therefore, recalling that \( b = \frac{\partial}{\partial x}d(0,x) \mid _{x=t+k} \) for all \( t \) and for all \( k \),

\[
b + \sigma^2(t + k) = \sum_{i=1}^{N} \alpha_i(t)g_i(k)
\]

\[
= \left( \frac{\alpha_1(t)}{\sqrt{M}} + \sum_{i=2}^{N} \alpha_i(t)a_0^i + \sum_{i=2}^{N} \alpha_i(t)a_1^i k + \sum_{i=3}^{N} \alpha_i(t)a_2^i k^2 \right. \\
+ \ldots + \left( \alpha_{N-1}(t)a_{N-2}^i + \alpha_{N-1}(t)a_{N-2}^i k \right) k^{N-2} + \alpha_N(t)a_{N-1}^i k^{N-1};
\]

equating powers we obtain

\[
\begin{align*}
\alpha_N(t)a_{N-1}^i &= 0, \\
\alpha_{N-1}(t)a_{N-2}^i + \alpha_N(t)a_{N-2}^i &= 0, \\
& \ldots \\
\sum_{i=2}^{N} \alpha_i(t)a_1^i &= 0, \\
\sum_{i=2}^{N} \alpha_i(t)a_2^i &= \sigma^2, \\
\left( \frac{\alpha_1(t)}{\sqrt{M}} + \sum_{i=2}^{N} \alpha_i(t)a_0^i \right) &= b + \sigma^2 t.
\end{align*}
\]

This system has the solution:

\[
\begin{cases}
\alpha_i(t) = 0, & \text{for } i = 3, \ldots, N, \\
\alpha_2(t) = \frac{\sigma^2}{a_1^2}, \\
\alpha_1(t) = \sqrt{M} \left( b + \sigma^2 t - \sigma^2 \frac{a_0^2}{a_1^2} \right).
\end{cases}
\]

With this choice, we have

\[
\dot{\Phi}^{\text{det}}(t) = -2 \sum_{k=1}^{N} \left[ G(Z^{\text{det}}(t), k) - d_k(t) \right] \dot{d}_k(t)
\]

\[
= -2 \sum_{k=1}^{N} \left[ G(Z^{\text{det}}(t), k) - d_k(t) \right] \left[ \alpha_1(t)g_1(k) + \alpha_2(t)g_2(t) \right]
\]

\[
= -2\alpha_1(t) \sum_{k=1}^{N} \left[ G(Z^{\text{det}}(t), k) - d_k(t) \right] g_1(k) \\
-2\alpha_2(t) \sum_{k=1}^{N} \left[ G(Z^{\text{det}}(t), k) - d_k(t) \right] g_2(k) = 0,
\]

where we have used the first two equation in (4.3).

**Remark 4.3.** From the HJM form of the yield curve, it is easy to see that pure translation movements (plus white noise) of the yield curve are considered only when the drift coefficient is a linear function of the time; analogously we require

\[
\frac{\partial}{\partial t} d_k(t) = \text{const};
\]
therefore the model presents the same translation movement for all the maturities if

\[(4.4) \quad \frac{\partial}{\partial t} d_0(t + k) + E \left[ \sigma(t, k) \int_0^k \sigma(t, u) \, du \right. \]
\[+ \int_0^t \left( \frac{\partial}{\partial t} \sigma(s, t + k - s) \int_0^{t+k-s} \sigma(s, u) \, du + \sigma(s, t + k - s)^2 \right) \, ds \] \quad = \quad \text{const.} \]

Remark 4.4. We can express the movement of the yield curve generalizing in the following way:

\[ \frac{\partial}{\partial t} d_k(t) = \phi(t, k); \]

using the calculation from the previous remark we get

1. \( \sigma(t, x) = \text{const.} \) The movements are expressed by

\[ \phi(t, k) = \partial_t d_0(t + k) + \sigma^2(t + k). \]

2. \( \sigma(t, k) = ak. \) Inserting in \((4.4)\) and performing all the calculation we get

\[ \frac{a^2 k^3}{6} = \text{const, for all } k, \]

from which we can conclude \( a = 0. \)

Example 4.1. Let us consider different choices for the volatility \( \sigma: \)

1. \( \sigma(t, t + k) = \sigma > 0. \) Inserting in equation \((4.4)\) we get

\[ \partial_t d_{t+k}(0) + \sigma^2(k + t) = \text{const,} \]

from which we get

\[ d_0(t + k) = -\sigma^2 \int_0^{t+k} (s + k) \, ds. \]

This implies that from a constant volatility we cannot replicate pure translation movements, because the previous equation cannot hold for every \( k. \)

2. \( \sigma(t, k) = ak. \) Inserting in \((4.4)\) and performing all the calculation we get

\[ \frac{a^2 k^3}{6} = \text{const,} \] for all \( k, \)

Theorem 4.5. Let \( \{ \beta_i(t) \}_{i=N+1, \ldots, M} \) be as in \((4.2)\). Then \( \dot{\Phi}^{\text{det}}(t) \leq 0 \) if and only if

\[ \sum_{i=N+1}^M \beta_i(t) \gamma_i(t) \geq 0, \]

where \( \{ \gamma_i(t) \}_{i=N+1, \ldots, M} \) are the coordinates of \( G(Z^{\text{det}}) - d \) in \( G \).
Since \((G(Z^{\text{det}}(t), k) - d_k(t))_{k=1,\ldots,M} \in \mathbb{G}^+\), we have
\[G(Z^{\text{det}}(t), k) - d_k(t) = \sum_{j=N+1}^{M} \gamma_j(t) g_j^+(k), \quad \text{for all } k = 1, \ldots, M,\]
for some \(\gamma : \mathbb{R}^+ \rightarrow \mathbb{R}\); inserting in the previous inequality we get
\[-\sum_{k=1}^{M} \left[ G(Z^*(t), k) - d_k(t) \right] \pi_{G^+} (d(t))_k = - \sum_{i=N+1}^{M} \beta_i(t) \sum_{j=N+1}^{M} \gamma_j(t) \sum_{k=1}^{M} g_j^+(k) g_i^+(k) \leq 0.\]
The system \(\{g_i(k), g_j^+(k)\}_{i=1,\ldots,N; j=N+1,\ldots,M}\) forms an orthonormal basis for \(\mathbb{R}^M\) therefore
\[\sum_{k=1}^{M} g_j^+(k) g_i^+(k) = \delta_{ij}.\] This leads to the thesis. \(\square\)

5. Error in \(L^2\)

In this section we extend the previous analysis to the whole term structure, namely we study a continuous set of maturity; consider the finite dimensional linear family
\[(5.1) \quad G(Z(t), \tau) = \sum_{i=1}^{M} Z_i(t) g_i(\tau), \quad \tau \in [0, T],\]
where \(\{g_1, \ldots, g_M\}\) is an orthonormal system of functions, \(g_i \in L^2(0, T; \mathbb{R})\). Given a volatility structure \(\sigma(t, T)\) for the yield curve and following the procedure of Section 3, we define
\[d_\tau(t) := d_{\tau+\tau}(0) + E \left[ \int_0^T \sigma(s, t + \tau - s) \int_0^{t+\tau-s} \sigma^T(s, u) du ds \right].\]
We analyze the distance in the \(L^2\) norm, then for any \(t \geq 0\) we define \(Z^{\text{det}}(t) \in \mathbb{R}^N\) as the solution of
\[(5.2) \quad \min_{Z(t) \in \mathbb{R}^N} \int_0^T \left[ G(Z(t), \tau) - d_\tau(t) \right]^2 d\tau = \int_0^T \left[ G(Z^{\text{det}}(t), \tau) - d_\tau(t) \right]^2 d\tau,\]
while the distance is defined as
\[\Phi^{\text{det}}(t) := \int_0^T \left[ G(Z^{\text{det}}(t), \tau) - d_\tau(t) \right]^2 d\tau;\]
from (5.2) we get
\[\nabla_Z \int_0^T \left[ G(Z^{\text{det}}(t), \tau) - d_\tau(t) \right]^2 d\tau = 0,\]
then
\[2 \int_0^T [G(Z^{\text{det}}(t), \tau) - d_\tau(t)] \nabla G(Z^{\text{det}}(t), \tau) = 0, \quad \Leftrightarrow \]
\[2 \int_0^T [G(Z^{\text{det}}(t), \tau) - d_\tau(t)] g_i(\tau) d\tau = 0, \quad \text{for all } i = 1, \ldots, N.\]
By orthonormality of the system \(\{g_i(\tau)\}_{i=1,\ldots,N}\), we have
\[Z^{\text{det}}(t) = \int_0^T d_\tau(t) g_i(\tau) d\tau, \quad \text{for all } i = 1, \ldots, N;\]
its dynamic is
\[\dot{Z}^{\text{det}}(t) = \int_0^T d_\tau(t) g_i(\tau) d\tau.\]
We can follow the same approach as in Section 3 and 4, with the same definition of consistent families. Complete the system \( \{g_1, \ldots, g_n\} \) in order to obtain an Hilbert basis of \( L^2([0, T]) \) and let \( \{g_i\}_{i=N+1}^{\infty} \) be the basis of \( G^\perp \). We can write

\[
\hat{d}(t) = \sum_{i=1}^{N} \alpha_i(t) g_i + \sum_{i=N+1}^{\infty} \beta_i(t) g_i^1,
\]

for some \( \alpha_i, \beta_i : \mathbb{R}^+ \to \mathbb{R} \). Moreover there exist \( \{\gamma_j(t)\}_{j=N+1}^{\infty} \) such that

\[
G(Z^{det}(t), \tau) - d_r(t) = \sum_{j=N+1}^{\infty} \gamma_j(t) g_j^2(\tau), \quad \text{for } \tau \in [0, T].
\]

**Theorem 5.1.** Assume that \( \hat{d}_r(t) \in \mathcal{G} \), then

\[
\hat{d}^{det}(t) = 0, \quad \text{for all } t \in [0, T].
\]

Hence the family (5.1) is approximately consistent with (3.1).

**Proof.** See proof of Theorem 4.1.

**Theorem 5.2.** Let \( \{\beta_i(t)\}_{i=N+1}^{\infty} \) as in (5.3). \( \mathcal{G} \) is an approximately consistent family with the given HJM model if and only if

\[
\sum_{i=N+1}^{\infty} \beta_i(t) \gamma_i(t) \geq 0.
\]

**Proof.** See proof of Theorem 4.5

6. **Stochastic Evolution: Finite Dimensional Case**

Until now we have focused our attention on a deterministic approximation of the HJM evolution (3.1)–(3.2) with Musiela parametrization. Now we consider a more general approach to the problem: we give a stochastic evolution for the parameters of the finite-dimensional family in order to get approximate consistency in a stochastic sense.

Consider again the finite dimensional family (4.1) and a volatility structure for the model, \( \sigma(t, T) \). The evolution is given by (3.1)–(3.2) whose solution \( r_k(t) \) is:

\[
r_k(t) := d_k(0) + \int_0^t \left( \frac{\partial}{\partial x} r(s, \tau) + \sigma(s, \tau) \int_0^\tau \sigma^T(s, u) du \right) ds + \int_0^t \sigma(s, s + k) dW_s.
\]

The target is to find a stochastic process \( Z^*(t, \omega) \in L^\infty(0, T; (L^2(\Omega, \mathcal{F}, P))^N) \), i.e. progressively measurable with bounded values in \( L^2 \), that minimizes

\[
\int_0^T E \left[ \sum_{k=1}^{M} G(Z(t, \omega), k) - r_k(t)^2 \right] dt.
\]

This problem is quite difficult to handle thus we simplify our approach in the following way: fix \( (t, \omega) \in [0, T] \times \Omega \) and minimize over all \( Z \in \mathbb{R}^N \), then proceeding as in in Section 4, we determine the process

\[
Z^*_j(t) = \sum_{k=1}^{M} r_k(t) g_j(k), \quad \text{for } j = 1, \ldots, N.
\]

From this equation we get the dynamic of \( Z^*_j(t) \): for \( j = 1, \ldots, N \)

\[
dZ^*_j(t) = \sum_{k=1}^{M} dr_k(t) g_j(k).
\]
which gives the projection of the dynamic onto the linear space $G$.

The error we incur in with this approximation is the distance between the HJM model and the finite dimensional family in the least squares sense

$$\Phi^*(t,\omega) := \sum_{k=1}^{M} \left[ G(Z^*(t), k) - r_k(t) \right]^2.$$ 

**Definition 6.1.** We say that the family (4.1) is approximately consistent in the stochastic sense if

$$\mathbb{E}[d\Phi^*(t, \omega)] \leq 0, \text{ for all } t \in [0, T].$$

To perform the study of this inequality, we analyze the Ito dynamic of $\Phi^*(t, \omega)$. After some calculation

$$d\Phi^*(t, \omega) = -2 \sum_{k=1}^{M} \left[ G(Z^*(t), k) - r_k(t) \right] dr_k(t)$$

$$- \sum_{k=1}^{M} \sum_{h=1}^{M} \sigma(t, k)\sigma^T(t, k) \sum_{i=1}^{N} g_i(h)g_i(k)dt$$

$$-2 \sum_{k=1}^{M} \sum_{i=1}^{N} g_i(k) \sum_{h=1}^{M} \sigma(t, h)\sigma^T(k, h)dt + \sum_{k=1}^{M} \sigma(t, k)\sigma^T(t, k)dt.$$ 

Inserting the dynamic for $dr_k(t)$ we get

$$d\Phi^*(t, \omega) = -2 \sum_{k=1}^{M} \left[ G(Z^*(t), k) - r_k(t) \right] \left[ \frac{\partial}{\partial x} r(t, x) + \sigma(t, k) \int_{0}^{k} \sigma^T(t, u)du \right]dt$$

$$+ \sum_{k=1}^{M} \left\{ \sum_{h=1}^{M} \sigma(t, k)\sigma^T(t, k) \sum_{i=1}^{N} g_i(k)g_i(h) \right\}dt$$

$$-2 \sum_{i=1}^{N} g_i(k) \sum_{h=1}^{M} \sigma(t, h)\sigma^T(k, h) + \sigma(t, k)\sigma^T(t, k)\right\}dt$$

$$-2 \sum_{k=1}^{M} \left[ G(Z^*(t), k) - r_k(t) \right] \sigma(t, k)dW_t.$$ 

Assume that $\sigma$ is a deterministic function, then

$$\mathbb{E}[d\Phi^*(t, \omega)] = d\Phi^{det}(t) + \sum_{k=1}^{M} \left\{ \sum_{h=1}^{M} \sigma(t, h)\sigma^T(t, k) \sum_{i=1}^{N} g_i(k)g_i(h) \right\}dt$$

$$-2 \sum_{i=1}^{N} g_i(k) \sum_{h=1}^{M} \sigma(t, h)\sigma^T(t, k) + \sigma(t, k)\sigma^T(t, k)\right\}dt$$

**6.1. Stochastic Error vs Deterministic Error.** In the study of the approximate consistency of the finite dimensional family (4.1) a central role is played by the choice of the parameter $Z$: it could be a deterministic or a stochastic process, as proposed in the previous sections. Here we compare the two choices, showing that a deterministic evolution produce a lower error than the stochastic one in the least squares sense.
Define the **stochastic error** as the operator

\[
\Phi^{stoch} : L^\infty(0, T; (L^2(\Omega, \mathbb{P}))^n) \rightarrow L^\infty(0, T; \mathbb{R})
\]

\[
Z \mapsto E \left[ \sum_{k=1}^{M} \left( G(Z(t), k) - r_k(t) \right)^2 \right].
\]

Applying Jensen inequality, using the linearity of \( G \), for all \( Z \in (L^2(\Omega, \mathbb{P}))^n \) we have

\[
E \left[ \sum_{k=1}^{M} \left( G(Z(t), k) - r_k(t) \right)^2 \right] = \sum_{k=1}^{M} \int_{\Omega} \left[ G(Z, k) - r_k(t) \right]^2 d\mathbb{P}
\]

\[
\geq \sum_{k=1}^{M} \int_{\Omega} \left[ G(Z, k) - r_k(t) \right] d\mathbb{P}^2 = \sum_{k=1}^{M} \left[ E[G(Z, k)] - d_k(t) \right]^2
\]

\[
= \sum_{k=1}^{M} \left[ G(E[Z], k) - d_k(t) \right]^2,
\]

therefore

\[
\min_{Z \in (L^2(\Omega, \mathbb{P}))^n} E \left[ \sum_{k=1}^{M} \left( G(Z, k) - r_k(t) \right)^2 \right] \geq \min_{Z \in \mathbb{R}^n} \sum_{k=1}^{M} \left[ G(Z, k) - d_k(t) \right]^2.
\]

It follows that:

\[
\Phi^{det}(t) \leq \Phi^{stoch}(Z^*)(t) \leq \Phi^{stoch}(Z^{det})(t).
\]

That means that we can always have a bound of the error only by making use of the deterministic approximation \( Z^{det} \).

A more detailed analysis of the error produced in the stochastic approximation could be performed as follows. Given \( \varepsilon > 0 \) a tolerance parameter, we want to find a \( Z \in L^\infty(0, T; (L^2(\Omega, \mathbb{P}))^n) \) to minimize

\[
(6.1) \quad \mathbb{P}\left( \left\{ \sum_{k=1}^{M} \left[ G(Z, k) - r_k(t) \right]^2 > \varepsilon \right\} \right)
\]

Using Chebyshev inequality:

\[
\mathbb{P}\left( \left\{ \sum_{k=1}^{M} \left[ G(Z, k) - r_k(t) \right]^2 > \varepsilon \right\} \right) \leq \frac{1}{\varepsilon} E \left[ \sum_{k=1}^{M} \left[ G(Z, k) - r_k(t) \right]^2 \right];
\]

minimizing over \( Z \in L^\infty(0, T; (L^2(\Omega, \mathbb{P}))^n) \) we obtain

\[
\mathbb{P}\left( \left\{ \sum_{k=1}^{M} \left[ G(Z^*, k) - r_k(t) \right]^2 > \varepsilon \right\} \right) \leq \frac{1}{\varepsilon} \Phi^{stoch}(Z^*)(t).
\]

Therefore we get

\[
\mathbb{P}\left( \left\{ \sum_{k=1}^{M} \left[ G(Z^*, k) - r_k(t) \right]^2 > \varepsilon \right\} \right) \leq \frac{1}{\varepsilon} \Phi^{stoch}(Z^*)(t) \leq \frac{1}{\varepsilon} \Phi^{stoch}(Z^{det})(t).
\]

Thus we can estimate (6.1) again from the stochastic error corresponding \( Z^{det} \).
7. Four parameters Nelson–Siegel family.

In this section we present an extension of the Nelson–Siegel model studied in Section 3. We propose here a four parameter Nelson–Siegel model, assuming that the parameter \( \alpha \) in (3.4) is no more a constant but it is a function of time

\[
G(Z(t), k) = Z_1(t) + Z_2(t)e^{-Z_4(t)k} + Z_3(t)ke^{-Z_4(t)k};
\]

here \( Z(t) = (Z_1(t), Z_2(t), Z_3(t), Z_4(t)) \in \mathbb{R}^4 \) and \( k \) is the time to maturity, \( k = 1, \ldots, M \).

The existence of a \( Z(t) \) realizing a minimum of the error (3.5) is not straightforward since \( G(Z(t), k) \) is a nonlinear family. However, for every fixed \( Z_4 \), \( G \) is a linear parameterized family. Proceeding as in Section 3, a minimum of the error (3.5) is obtained as

\[
Z^{det}(Z_4) = < J(Z_4, M)^{-1} \begin{pmatrix}
1 & e^{-Z_4} & e^{-Z_4} \\
\vdots & \vdots & \vdots \\
1 & e^{-Z_4k} & ke^{-Z_4k} \\
\vdots & \vdots & \vdots \\
1 & e^{-Z_4t} & Me^{-Z_4t}
\end{pmatrix}, \begin{pmatrix}d_1 \\ \vdots \\ d_k \\ \vdots \\ d_M\end{pmatrix} >.
\]

(7.1)

We note that the matrix \( J(Z_4, M) \) is invertible if \( Z_4 > 0 \). Define

\[
\phi_d(Z_4) = \Phi^{det}((Z_1^{det}, Z_2^{det}, Z_3^{det})(Z_4))
\]

the error as function of \( Z_4 \) in approximating the data \( d = \{d_k\}_{k=1}^M \). We prove that, under suitable conditions, there exists a minimum for the function \( \phi_d \) and that the error is bounded by the following quantity

\[
\min_{Z_1} \sum_{k=3}^M (Z_1 - d_k)^2;
\]

this means that a good approximation for the interest rate evolution can be obtained fitting the \( Z_2 \) and \( Z_3 \) parameters on the first two data, and the \( Z_1 \) parameter on the average of the remaining.

**Proposition 7.1.** If there exists \( Z_4 \) such that

\[
\phi(\bar{Z}_4) \leq \lim_{Z_4 \to \infty} \phi_d(Z_4),
\]

then there exists a point of minimum for \( \Phi^{det} \). If \( M \geq 3 \) then

\[
\lim_{Z_4 \to \infty} \phi_d(Z_4) = \min_{Z_1} \sum_{k=3}^M (Z_1 - d_k)^2.
\]

The first statement is obvious, while an intuitive proof of the second can be given as follows.

The functions \( Z_1, Z_2e^{-Z_4k} \) and \( Z_3ke^{-Z_4k} \) form a basis for the four parameter Nelson–Siegel family. Then, it is possible to fit every maturity, even the shorter one (instantaneous interest rate). In fact, consider the two functions of the time to maturity:

\[
f_1(k) = Z_2e^{-Z_4k},
\]

\[
f_2(k) = Z_3ke^{-Z_4k}.
\]
The function $f_1$ attains its maximum for $k = 0$, $f_1(0) = Z_2$, while the function $f_2$ attains its maximum for $k = \frac{1}{Z_4}$.

$$f_2\left(\frac{1}{Z_4}\right) = \frac{Z_3}{Z_4} e^{-1} Z_4 \rightarrow \infty \quad \begin{cases} +\infty & Z_3 > 0, \\ -\infty & Z_3 < 0. \end{cases}$$

Therefore, with a proper choice of the parameters, we can fit every maturity. For example, if we want to fit the first two data, we have to solve the following system

$$\begin{cases}
Z_1 + Z_2 e^{-Z_4} + Z_3 e^{-Z_4} = d_1, \\
Z_1 + Z_2 e^{-2Z_4} + 2Z_3 e^{-2Z_4} = d_2,
\end{cases}$$

obtaining a solution $(Z_2, Z_3)(Z_1, Z_4)$:

$$\begin{cases}
Z_2(Z_1, Z_4) = [2d_1 - 2Z_1 - d_2 e^{Z_4} + Z_1 e^{Z_4}] e^{Z_4}, \\
Z_3(Z_1, Z_4) = [d_2 - Z_1 - d_1 e^{-Z_4} + Z_1 e^{-Z_4}] e^{2Z_4}.
\end{cases}$$

On the other side, looking at the shape of the functions $f_1$ and $f_2$ in Figure 1, we see that the resulting term structure is increasing for the short terms, present an inversion and then decreases to a long term mean value. Thus if we try to fit with these functions some data that are not the first two, the error we incur in tends to infinity with $Z_4$ that tends to infinity.

Define now the \textit{Instantaneous Variance} of the data $d$ as

$$\text{Var}_{k \geq 3}(d) := \left( \sum_{k \geq 3} \left[ \frac{\sum_{j \geq 3} d_j}{M - 2} - d_k \right] \right)^{\frac{1}{2}};$$

then the conclusion of Proposition 7.1 is valid if the following assumption holds:

$$\text{Var}_{k \geq 3}(d(t)) > c, \quad \text{for any } t \in [0, T],$$

where $c$ is a constant sufficiently large.

We now prove Proposition 7.1.
Proof. Equation (7.1) can be written as follows:

\[
\begin{align*}
Z_{\text{det}}(Z_4) &= J(Z_4, M)^{-1} \left( \sum_{k=1}^{M} d_k e^{-Z_4 k} \right) \cdot \\
&= J(Z_4) \left( \sum_{k=1}^{M} d_k e^{-Z_4 k} \right) \cdot \\
&= \frac{\text{det} J(Z_4)}{J(Z_4)} \cdot \\
&= \frac{[\text{cof}(J(Z_4))]^T}{\text{det} J(Z_4)}.
\end{align*}
\]

Assume \( M > 3 \), which is typical of real markets; we can write explicitly the expression of \((Z_1^{\text{det}}, Z_2^{\text{det}}, Z_3^{\text{det}})(Z_4)\): by definition

\[
J(Z_4)^{-1} = \frac{[\text{cof}(J(Z_4))]^T}{\text{det} J(Z_4)};
\]

where \( \text{cof} \) indicates the cofactor matrix. Evaluating the previous quantities, we see that there is a finite sum of \( e^{-jZ_4} \) for \( j \geq 1 \), but only some terms contribute in the limit for \( Z_4 \to \infty \). More precisely, those terms corresponding to \( j \leq 6 \). Define

\[
c_{ij} = [\text{cof}(J(Z_4))]_{ij}.
\]

\( J(Z_4) \) is symmetric, then the corresponding matrix \( [\text{cof}(J(Z_4))] \) is symmetric as well, thus we need to compute only the coefficients \( c_{ij} \) for \( 1 \leq i \leq j \leq 3 \). Performing this calculation:

\[
\text{det} J(Z_4) = (M - 2)e^{-6Z_4} + \text{h.o.t.},
\]

where \( \text{h.o.t.} \) stands for higher order terms. Now

\[
\begin{align*}
c_{11} &= e^{-6Z_4} + \text{h.o.t.}, \\
c_{12} &= e^{-4Z_4} - 3e^{-6Z_4} + \text{h.o.t.}, \\
c_{13} &= -e^{-4Z_4} - e^{-5Z_4} - 3e^{-6Z_4}, \\
c_{22} &= (M - 1)e^{-2Z_4} - 4e^{-3Z_4} + (4M - 10)e^{-4Z_4} + \text{h.o.t.}, \\
c_{23} &= -(M - 1)e^{-2Z_4} + 3e^{-3Z_4} - (2M - 6)e^{-4Z_4} + \text{h.o.t.}, \\
c_{33} &= (M - 1)e^{-2Z_4} - 2e^{-3Z_4} + (M - 3)e^{-4Z_4} + \text{h.o.t.}.
\end{align*}
\]

For \( Z_4 \) fixed, the minimal point is given by the following

\[
Z_1(Z_4) = \frac{e^{6Z_4}}{M - 2} \left( \sum_{k=1}^{M} d_k e^{-6Z_4} \sum_{k=1}^{M} (1 - k) d_k e^{-kZ_4} \right) + \text{h.o.t.}
\]

\[
\begin{align*}
&-e^{-5Z_4} \sum_{k=1}^{M} k d_k e^{-kZ_4} + \text{h.o.t.} \\
&= \frac{e^{6Z_4}}{M - 2} \left[ \sum_{k=1}^{M} d_k e^{-6Z_4} - d_2 e^{-6Z_4} - d_1 e^{-6Z_4} \right] + \text{h.o.t.},
\end{align*}
\]

(7.2)
\[ Z_3(Z_4) = \frac{e^{\text{6}Z_4}}{M - 2} \left( e^{-4Z_4} - 2e^{-6Z_4} \right) \sum_{k=1}^{M} d_k \]
\[ + \left( (M - 1)e^{-4Z_4} - 4e^{-3Z_4} + 4(M - 10)e^{-4Z_4} \right) \sum_{k=1}^{M} d_k e^{-kZ_4} \]
\[ + \left( (M - 1)e^{-2Z_4} + 3e^{-3Z_4} - (2M - 6)e^{-4Z_4} \right) \sum_{k=1}^{M} k d_k e^{-kZ_4} \]
\[ + \text{o.h.t.} \]
\[ = \frac{e^{\text{2}Z_4}}{M - 2} \left[ \sum_{k=1}^{M} d_k - (M - 1)d_2 - d_1 \right] \]
\[ + \frac{e^{\text{2}Z_4}}{M - 2} \left[ (2M - 4)d_1 + 2d_2 - 2(M - 1)d_3 \right] + \text{o.h.t.} \]

\[ Z_3(Z_4) = \frac{e^{\text{2}Z_4}}{M - 2} \left[ (e^{-4Z_4} - e^{-5Z_4}) \sum_{k=1}^{M} d_k + (M - 1)e^{-2Z_4} \left( \sum_{k=1}^{M} (k - 1)d_k e^{-kZ_4} \right) \right] \]
\[ + e^{-3Z_4} \left( \sum_{k=1}^{M} (3 - 2k)d_k e^{-kZ_4} \right) \]
\[ + e^{-4Z_4} \left( \sum_{k=1}^{M} (k(M - 3) - (2M - 6))d_k e^{-kZ_4} \right) + \text{o.h.t.} \]
\[ = \frac{e^{\text{2}Z_4}}{M - 2} \left[ \sum_{k=1}^{M} d_k + (M - 1)d_2 + d_1 \right] \]
\[ + \frac{e^{\text{2}Z_4}}{M - 2} \left[ - \sum_{k=1}^{M} d_k + (3 - M)d_1 - d_2 + 2(M - 1)d_3 \right] + \text{o.h.t.} \]
here we have omitted the superscript \( \text{det} \) for simplicity of notation. Now we can evaluate the cost

\[
\begin{align*}
\phi_d(Z_4) &= \sum_{k=1}^{M} \left[ Z_1(Z_4) + Z_2(Z_4)e^{-kZ_4} + Z_3(Z_4)ke^{-kZ_4} - d_k \right]^2 \\
&= \left\{ Z_1(Z_4) + \frac{e^{Z_4}}{M-2} \left[ \sum_{k=1}^{M} d_k - (M-1)d_2 - d_1 \right] \\
&+ \frac{1}{M-2} \left[ (2M-4)d_1 + 2d_2 - 2(M-1)d_3 \right] \\
&+ \frac{e^{Z_4}}{M-2} \left[ - \sum_{k=1}^{M} d_k + (M-1)d_2 + d_1 \right] \\
&+ \frac{1}{M-2} \left[ - \sum_{k=1}^{M} d_k + (3-M)d_1 - d_2 + 2(M-1)d_3 \right] - d_1 \right\}^2 \\
&+ \left\{ Z_1 + \frac{1}{M-2} \left[ \sum_{k=1}^{M} d_k - (M-1)d_2 - d_1 \right] \\
&+ \frac{2}{M-2} \left[ - \sum_{k=1}^{M} d_k + (M-1)d_2 + d_1 \right] - d_2 \right\}^2 \\
&+ \sum_{k>2} |Z_1(Z_4) - d_k|^2 + h.o.t. \\
&= A_1(Z_4) + A_2(Z_4) + \sum_{k>2} |Z_1(Z_4) - d_k|^2 + h.o.t.
\end{align*}
\]

Analyzing these three quantities we can see that

\[
A_1(Z_4) = 0, \quad A_2(Z_4) = 0, \quad \text{for all } Z_4,
\]

and passing to the limit, using (7.2)

\[
\lim_{Z_4 \to \infty} \phi_d(Z_4) = \sum_{k>2} \left[ \frac{\sum_{j \neq 1,2} d_j}{M-2} - d_k \right]^2.
\]

7.1. **Stability of the approximation.** An important point of the proposed technique is the analysis of the stability of the approximation. More precisely, Proposition 7.1 gives a condition for the existence of an optimal choice of parameters: assuming that such condition holds at initial time, we want to estimate the maximal time for which this condition still holds.

We perform the study of this issue, giving an expression to estimate the time interval on which the four parameter Nelson–Siegel approximation remains stable for a general HJM model.

Assume that we can fit exactly the initial data \( \{d_k(0)\}_{k=1}^{M} \), then there exists a \( Z_4^d \) and the corresponding \( (Z_1^d, Z_2^d, Z_3^d)(Z_4^d) \) parameters such that

\[
G(Z_1^d, Z_2^d, Z_3^d, Z_4^d, k) = d_k(0), \quad \text{for all } k = 1, \ldots, M.
\]
Assume also that
\[
0 = \phi_d(Z_i^d) < \lim_{Z_4 \to \infty} \phi_d(Z_4) = \sum_{k \geq 3} \left[ \frac{\sum_{j \geq 3} d_j(0)}{M - 2} - d_k(0) \right]^2.
\]

Consider \( \Delta = \{ \Delta_k \}_{k=1, \ldots, M} \), 0 < \( \Delta_k \ll 1 \), and the corresponding small perturbation of the data \( d^\Delta \):
\[
d^\Delta(t) = \{ d_k(0) + \Delta_k(t) \}_{k=1, \ldots, M}.
\]
If one uses \((Z_1^d, Z_2^d, Z_3^d, Z_4^d)\) to fit \(d^\Delta(t)\), then
\[
\phi_{d^\Delta}(Z_4^d) = \min_{R^3} \sum_{k=1}^M \left[ G(Z_1, Z_2, Z_3, Z_4^d, k) - d_k^2(t) \right]^2
\leq \sum_{k=1}^M \left[ G(Z_1, Z_2, Z_3, Z_4^d, k) - d_k(0) - \Delta_k \right]^2 = \sum_{k=1}^M \Delta_k^2.
\]
To assure the existence of a minimum for the fitting of \(d^\Delta\), it is enough that the perturbation \( \Delta \) satisfies
\[
\phi_{d^\Delta}(Z_4^d) < \sum_{k \geq 3} \left[ \frac{\sum_{j \geq 3} d_j^2(t)}{M - 2} - d_k^2(t) \right]^2 = \lim_{Z_4 \to \infty} \phi_{d^\Delta}(Z_4).
\]
Define
\[
\Delta := \sum_{j \geq 3} \Delta_j = \frac{\sum_{k=3}^M \Delta_k}{M - 2};
\]
performing some calculations we get the inequality
\[
(7.3) \quad \Delta_1^2 + \Delta_2^2 + (M - 2)\Delta^2 - 2 \sum_{k=3}^M \left[ \frac{\sum_{j \geq 3} d_j(0)}{M - 2} - d_k(0) \right] \Delta_k \leq \lim_{Z_4 \to \infty} \phi_d(Z_4).
\]
The interest rate evolves following the HJM model with the Musiela parametrization, therefore the perturbation can be expressed in term of the volatility of the model: comparing with (3.3) we can explicitly write
\[
\Delta_k(t) = d_{k+1}(0) - d_k(0) + \mathbb{E} \left[ \int_0^t \sigma(s, t + k - s) \int_0^{t+k-s} \sigma^T(s, u) du ds \right].
\]
Define
\[
\Sigma = \sup_{s, u \in [0, T]} \sigma(s, u),
\]
and assume
\[
d_k(0) = b \cdot k \quad \text{for all } k,
\]
then
\[
|\Delta_k(t)| \leq |d_{k+1}(0) - d_k(0)| + \mathbb{E} \left[ \int_0^t \sigma(s, t + k - s) \int_0^{t+k-s} \sigma^T(s, u) du ds \right] \leq b \cdot t + \Sigma^2 \left( \frac{t^2}{2} + tk \right).
\]
Using Cauchy-Schwartz (7.3) holds if
\[
\begin{align*}
\frac{\Sigma^4 M}{4} + \frac{\Sigma^4 M^2(M + 1)}{2} t^3 + \left\{ \Sigma^4 M + 2M^2 + 9M^2 - 92M + 116 \right\} + \frac{\Sigma^2 b^2 M(M + 1)}{2} t^2 \\
+ 2 \sum_{k \geq 3} (d)(0) \left( t^2 \text{coef1} + b^2 \text{coef2} \right) - \sum_{k \geq 3} (d)(0) \leq 0,
\end{align*}
\]
\[
\begin{align*}
+ \left[ \Sigma^2 b^2 M^3 + 9M^3 - 92M + 116 \right] + \frac{b^2 M(M^3 + M - 6)}{2} t^2 \\
+ 2 \sum_{k \geq 3} (d)(0) \left( t^2 \text{coef1} + b^2 \text{coef2} \right) - \sum_{k \geq 3} (d)(0) \leq 0.
\end{align*}
\]
where
\[
\text{coeff1} = -2\sum 3M^4 + 6M^3 - 33M^2 - 36M + 108 \\
+ \sum^2 3(1 - 2b^2)M^4 + 2M^3(5 - 6b^2) - M^2(27 - 66b^2) - M(34 - 72b^2) + 48 - 216b^2 \\
+ 2b^2 3M^4 + 6M^3 - 33M^2 - 36M + 108 \\
12(M - 2)
\]
\[
\text{coeff2} = \frac{M^4 - 8M^3 + 23M^2 - 28M + 12}{12(M - 2)}.
\]

Assume now that \( t\sqrt{\text{coeff1}} \geq 0 \) and \( b\sqrt{\text{coeff2}} \geq 0 \), then we have
\[
\begin{align*}
(7.4) & \quad c_0^0 + \sum^4 c_1^1 t^3 + \left\{ \sum^4 c_1^1 + \sum^2 bc_1^1 \right\} t^2 + \left\{ 2\sum^2 bc_1^1 + 2\text{Var}_{k \geq 2}(d)(0)\sqrt{\text{coeff1}} \right\} t \\
+ 5b^2 + b^4 \frac{(M^2 + M - 6)^2}{4(M - 2)} + 2\text{Var}_{k \geq 3}(d)(0)\sqrt{\text{coeff2}} - \text{Var}_{k \geq 3}(d)(0) \leq 0,
\end{align*}
\]

where
\[
\begin{align*}
& c_0^0 = \frac{\sum^4 M}{4}, \quad c_1^1 = \frac{M^4 + 2M^3 + 9M^2 - 92M + 116}{4(M - 2)^2}, \\
& c_1^1 = \frac{M(M + 1)}{2}.
\end{align*}
\]

One can use (7.4) to numerically estimate the maximal time \( t^* \) such that Proposition 7.1 can be applied on \([0, t^*]\).

**References**


**Capitalia, Linea Finanza – Financial Models and Systems – Analytics and Models, Via G. Paisiello, 5, I-00198 Roma, Italy**

E-mail address: claudia.lachioma@capitalia.it

**Istituto per le Applicazioni del Calcolo ”M. Picone”, Consiglio Nazionale delle Ricerche, Via del Policlinico, 137, I-00161 Roma, Italy**

E-mail address: b.piccoli@iac.cnr.it