DEPARTEMENT BEDRIJFSECONOMIE

ESTIMATION OF THE FORWARD RATE CURVE BY MEANS OF METHODS USING HIGHER ORDER DEGREES OF SMOOTHNESS : A PROOF OF CONVERGENCE

by

Sabine LORIMIER

WORKING PAPER
96-227
April 1996

D/1996/2263/1

I thank Prof. Dr. M. Van Wouwe for reviewing the text.

1. Introduction

In recent years a lot of authors have proposed methods to estimate the term structure of interest rates by means of bond prices. The pioneering work was done by McCulloch (1971), who used polynomial splines to estimate the discount function. Because of the exponential nature of the discount function Vasicek and Fong (1982) proposed the usage of exponential splines. Shea (1984, 1985) showed however that spline functions employed to fit the discount function can generate unstable and widely fluctuating forward rate curves, a behaviour that is most unlikely to obtain.
That is why methods immediately fitting the forward rate curve are developed. One of the first articles using this new approach is Delbaen and Lorimier (1990) who define a measure of the degree of smoothness of the forward rate curve based on first differences. In a later work by Lorimier (1995a) the methodology is extended to a continuous time framework and it is proved that the estimated forward rate curve is a polynomial spline and is as smooth as possible under some no arbitrage restrictions. In Adams and Van Deventer (1994) an analogue approach can be found for the continuous time case, but instead of using first derivatives in the measure of degree of smoothness they employ a second derivative.

In this paper a method is presented using mth derivatives and differences to measure the smoothness of a curve in respectively a continuous and discrete time framework.
Furthermore it is show that the solution of the discrete problem converges uniformly to the one of the continuous problem. Finally some examples compare the results obtained with the method where $m = 1$ and $m = 2$.

2. Description of the method in a continuous time framework

Let $P(T_1), \ldots, P(T_N)$ be the prices of N zero-coupon bonds with maturities $T_1 < \cdots < T_N$. The corresponding yields-to-maturity are $Y_1, \ldots, Y_N$ with

$$Y_i = -\frac{\ln P(T_i)}{T_i}, \; i = 1, \ldots, N.$$

To estimate the forward rate curve by means of these bond prices, we first of all require that the following no arbitrage relations are fulfilled:

$$\int_0^{T_k} f(u)\, du = Y_k T_k, \; k = 1, \ldots, N,$$

where $f(u)$, $u \in [0, T]$, is the forward rate curve at time $t = 0$.
Next we will impose a smoothness condition on the forward rate curve, expressing the fact that the forward rates cannot fluctuate much from one period to the next. The measure of degree of smoothness used is based on the mth derivative and is defined as

$$\int_0^T (f^{(m)}(u))^2\, du,$$

where $T = T_N$.
So the forward rate curve is obtained as a solution of the following minimization problem:

$$\min_f \int_0^T (f^{(m)}(u))^2 \, du \tag{2.1}$$

subject to

$$\int_0^{T_k} f(u)\, du = Y_k T_k, \; k = 1, \ldots, N. \tag{2.2}$$

In Lorimier (1995b) it is proved that, if $N \geq m$, the unique solution of problem (2.1)-(2.2) is a $2m$th degree polynomial spline of the form

$$f(t) = \sum_{i=0}^{m-1} f^{(i)}(0) \frac{t^i}{i!} + \sum_{k=1}^{N} a_k h_k(t) \tag{2.3}$$

with

(i) h_k a $2m$th degree polynomial on $[0, T_k]$, C^{2m-1} over $[0, T]$, with
$h_k^{(m)}(u) = ((T_k - u)^+)^m \big/ m!$,
$h_k^{(i)}(0) = T_k^{i+1} \big/ (i+1)!$, $i = 0, \ldots, m-1$,
for $k = 1, \ldots, N$.

(ii) $\{f^{(i)}(0), i = 0, \ldots, m-1\}$ and $\{a_k, k = 1, \ldots, N\}$ solution to

$$\sum_{k=1}^{N} a_k \frac{T_k^{i+1}}{(i+1)!} = 0, \; i = 0, \ldots, m-1, \tag{2.4}$$

$$\sum_{i=0}^{m-1} f^{(i)}(0) \frac{T_k^{i+1}}{(i+1)!} + \sum_{l=1}^{N} a_l < h_l, h_k >_m = Y_k T_k, \; k = 1, \ldots, N, \tag{2.5}$$

where $< .,. >_m$ is the inner product on the Sobolev space $S_m = \{g \mid \text{for } i = 1, \ldots, m-1, g^{(i)} \text{ continuous}; g^{(m)} \in L^2[0, T]\}$ defined as

$$< g_1, g_2 >_m = \sum_{k=0}^{m-1} g_1^{(k)}(0) g_2^{(k)}(0) + \int_0^T g_1^{(m)}(u) g_2^{(m)}(u) \, du.$$

If $N < m$ then the solution is no longer unique. Calculation of the derivatives $f^{(j)}(t)$, $j < m$, shows that in general $f^{(j)}(0)$ and $f^{(j)}(T)$ are not zero. Because of the first property the drawback of the method using first derivatives in the measure of degree of smoothness is no longer valid. But the second property is not quite what we expect. It is namely logical that for a large horizon T the forward rate curve flattens towards the end. That is why the method can be adjusted by imposing the additional constraints

$$f^{(j)}(T) = 0, \; j = 1, \ldots, m-1. \tag{2.6}$$

The unique solution of this new problem (2.1)-(2.2), (2.6) is still a $2m$th spline and is of the form

$$f(t) = \sum_{i=0}^{m-1} f^{(i)}(0)\frac{t^i}{i!} + \sum_{k=1}^{N+m-1} a_k h_k(t) \tag{2.7}$$

with

(i) $h_k, k = 1, \ldots, N$, as in (2.3) and $h_{N+k}, k = 1, \ldots, m-1$, a $(2m-k-1)$th degree polynomial on $[0,T]$ with

$h^{(m)}_{N+k}(u) = (T-u)^{m-k-1} \Big/ (m-k-1)!\,,$

$h^{(i)}_{N+k}(0) = T^{i-k} \Big/ (i-k)!\,,\ i = k, \ldots, m-1,$

$h^{(i)}_{N+k}(0) = 0,\ i = 0, \ldots, k-1.$

(ii) $\{f^{(i)}(0), i = 0, \ldots, m-1\}$ and $\{a_k, k = 1, \ldots, N+m-1\}$ solution to

$$\sum_{k=1}^{N} a_k T_k = 0, \tag{2.8}$$

$$\sum_{k=1}^{N} a_k \frac{T_k^{j+1}}{(j+1)!} + \sum_{k=1}^{j} a_{N+k}\frac{T^{j-k}}{(j-k)!} = 0,\ j = 1, \ldots, m-1, \tag{2.9}$$

$$\sum_{i=0}^{m-1} f^{(i)}(0)\frac{T_k^{i+1}}{(i+1)!} + \sum_{l=1}^{N+m-1} a_l < h_l, h_k >_m = Y_k T_k,\ k = 1, \ldots, N, \tag{2.10}$$

$$\sum_{i=j}^{m-1} f^{(i)}(0)\frac{T^{i-j}}{(i-j)!} + \sum_{l=1}^{N+m-1} a_l < h_l, h_{N+j} >_m = 0,\ j = 1, \ldots, m-1. \tag{2.11}$$

The disadvantage that still exists when using one of the previous methods is that the no arbitrage relations have to be exactly fulfilled. Because the observed prices are rounded prices and often the average of bid and ask rates, small deviations in the no arbitrage relations have to be permitted. Furthermore the market does not always immediately react to small changes in the price, what causes small errors in the no arbitrage relations. To avoid this shortcoming we can adjust the previous methods by solving

$$\min_f \int_0^T (f^{(m)}(u))^2\,du + \sum_{k=1}^{N} \epsilon_k^2$$

subject to

$$\int_0^{T_k} f(u)\,du + \lambda\epsilon_k = Y_k T_k,\ k = 1, \ldots, N$$

with or without the additional constraints (2.6), or equivalently

$$\min_f \int_0^T (f^{(m)}(u))^2\,du + \alpha \sum_{k=1}^{N} \left(Y_k T_k - \int_0^{T_k} f(u)\,du \right)^2, \tag{2.12}$$

with or without (2.6), where $\alpha = 1/\lambda^2$ and where the introduction of the second term permits small deviations in the no arbitrage relations. As in the case where $m = 1$ (see Lorimier (1995a)) this new problem is a multicriterion problem, where there exists a trade-off between the exactness and the smoothness of the fit.

In Lorimier (1995b) it is shown that the solution of problem (2.12) without (2.6) is of the form (2.3) with $\{f^{(i)}(0), i = 0, \ldots, m-1\}$ and $\{a_k, k = 1, \ldots, N\}$ solution to

$$0 = \sum_{k=1}^{N} a_k \frac{T_k^{i+1}}{(i+1)!}, \; i = 0, \ldots, m-1, \tag{2.13}$$

$$a_k = \alpha(Y_k T_k - \sum_{i=0}^{m-1} f^{(i)}(0) \frac{T_k^{i+1}}{(i+1)!} - \sum_{l=1}^{N} a_l < h_l, h_k >_m), \; k = 1, \ldots, N. \tag{2.14}$$

If the additional constraints (2.6) also hold, then the solution is of the form (2.7) with $\{f^{(i)}(0), i = 0, \ldots, m-1\}$ and $\{a_k, k = 1, \ldots, N+m-1\}$ solution to

$$0 = \sum_{k=1}^{N} a_k \frac{T_k^{j+1}}{(j+1)!} + \sum_{k=1}^{j} a_{N+k} \frac{T^{j-k}}{(j-k)!}, \; j = 0, \ldots, m-1, \tag{2.15}$$

$$a_k = \alpha(Y_k T_k - \sum_{i=0}^{m-1} f^{(i)}(0) \frac{T_k^{i+1}}{(i+1)!} - \sum_{l=1}^{N+m-1} a_l < h_l, h_k >_m), \; k = 1, \ldots, N, \tag{2.16}$$

$$0 = \sum_{i=j}^{m-1} f^{(i)}(0) \frac{T^{i-j}}{(i-j)!} + \sum_{k=1}^{N+m-1} a_k < h_k, h_{N+j} >_m, \; j = 1, \ldots, m-1. \tag{2.17}$$

In the next section we will present analogue methods in a discrete time framework.

3. Description of the method in a discrete time framework

Let f_j be the forward rate today for the period $[jh, (j+1)h]$ and let $\overline{T}_s = \lfloor T_s/h \rfloor$, with $\lfloor x \rfloor$ the largest integer smaller than x. Define $\Delta_j^{(m)}, j = 1, \ldots, \overline{T}_N - m$, as the mth difference in the following way:

$$\Delta_j^{(0)} = f_{j-1}, \tag{3.1}$$

$$\Delta_j^{(m)} = \Delta_{j+1}^{(m-1)} - \Delta_j^{(m-1)}. \tag{3.2}$$

The following lemma states relations that exist between $\Delta_j^{(m)}$ and f_j.

Lemma 3.1

1) $\Delta_j^{(m)} = \sum_{k=0}^{m} (-1)^k \binom{m}{k} f_{m+j-k-1}$.

2) $f_{m+j} = \sum_{k=1}^{j+1} \Delta_k^{(m)} \binom{m+j-k}{j-k+1} + \sum_{k=0}^{m-1} (-1)^{m-k+1} \binom{j+m-k-1}{j} \binom{m+j}{k} f_k,$

$$= \sum_{k=1}^{j+1} \Delta_k^{(m)} \binom{m+j-k}{m-1} + \sum_{k=0}^{m-1} h^k \binom{m+j}{k} \frac{\Delta_1^{(k)}}{h^k}.$$

Proof

1) We will proof the statement by means of induction with respect to m.
For $m = 1$ it is easily verified that the equality is true. Using (3.2) and the induction hypothesis for $m-1$ we have that

$$\begin{aligned}
\Delta_j^{(m)} &= \Delta_{j+1}^{(m-1)} - \Delta_j^{(m-1)}, \\
&= \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} f_{m+j-k-1} - \sum_{k=1}^{m} (-1)^{k-1} \binom{m-1}{k-1} f_{m+j-k-1}, \\
&= \sum_{k=1}^{m-1} (-1)^k \left[\binom{m-1}{k} + \binom{m-1}{k-1} \right] f_{m+j-k-1} + f_{m+j-1} - (-1)^{m-1} f_{j-1},
\end{aligned}$$

what proofs the first statement, because

$$\binom{m-1}{k} + \binom{m-1}{k-1} = \binom{m}{k}, \tag{3.3}$$

for all m and k.

2) The first equality will be proved by induction with respect to m. For $m = 1$ the statement is true. To proof the equality for m, we use that

$$\begin{aligned}
f_{m+j} &= f_{(m-1)+(j+1)}, \\
&= \sum_{k=1}^{j+2} \binom{m+j-k}{j-k+2} \Delta_k^{(m-1)} + \sum_{k=0}^{m-2} (-1)^{m-k} \binom{j+m-k-1}{j+1} \binom{m+j}{k} f_k, \\
&= \sum_{k=1}^{j+2} \binom{m+j-k}{j-k+2} \Delta_k^{(m-1)} + \sum_{k=0}^{m-2} (-1)^{m-k+1} \binom{m+j-k-1}{j} \binom{m+j}{k} f_k \\
&\quad - \sum_{k=0}^{m-2} (-1)^{m-k+1} \binom{m+j-k}{j+1} \binom{m+j}{k} f_k, \\
&= \sum_{k=1}^{j+2} \binom{m+j-k}{j-k+2} \Delta_k^{(m-1)} + \sum_{k=0}^{m-1} (-1)^{m-k+1} \binom{m+j-k-1}{j} \binom{m+j}{k} f_k \\
&\quad - \sum_{k=0}^{m-1} (-1)^{m-k+1} \binom{m+j-k}{j+1} \binom{m+j}{k} f_k,
\end{aligned}$$

where the second equality follows from the induction hypothesis and the third one from (3.3).

Furthermore we have because of (3.3) and (3.2) that

$$\sum_{k=1}^{j+2} \binom{m+j-k}{j-k+2} \Delta_k^{(m-1)}$$

$$= \sum_{k=1}^{j+1} \Delta_{k+1}^{(m-1)} \binom{m+j-k-1}{j-k+1} + \Delta_1^{(m-1)} \binom{m+j-1}{j+1},$$

$$= \sum_{k=1}^{j+1} \Delta_{k+1}^{(m-1)} \binom{m+j-k}{j-k+1} - \sum_{k=2}^{j+1} \Delta_k^{(m-1)} \binom{m+j-k}{j-k+1} + \Delta_1^{(m-1)} \binom{m+j-1}{j+1},$$

$$= \sum_{k=1}^{j+1} \Delta_k^{(m)} \binom{m+j-k}{j-k+1} + \Delta_1^{(m-1)} \left(\binom{m+j-1}{j} + \binom{m+j-1}{j+1}\right),$$

$$= \sum_{k=1}^{j+1} \Delta_k^{(m)} \binom{m+j-k}{j-k+1} + \Delta_1^{(m-1)} \binom{m+j}{j+1},$$

$$= \sum_{k=1}^{j+1} \Delta_k^{(m)} \binom{m+j-k}{j-k+1} + \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} \binom{m+j}{j+1} f_{m-k-1},$$

$$= \sum_{k=1}^{j+1} \Delta_k^{(m)} \binom{m+j-k}{j-k+1} + \sum_{k=0}^{m-1} (-1)^{m-k+1} \binom{m+j}{k} \binom{m+j-k}{j+1} f_k.$$

The last equality follows from

$$\binom{m-1}{m-k-1} \binom{m+j}{j+1} = \binom{m+j}{k} \binom{m+j-k}{j+1}.$$

To prove the last statement in 2) we have to show that

$$\sum_{k=0}^{m-1} (-1)^{m-k+1} \binom{m+j-k-1}{j} \binom{m+j}{k} f_k = \sum_{k=0}^{m-1} \binom{m+j}{k} \Delta_1^{(k)}.$$

This can be proved by means of induction with respect to m. Namely, for $m = 1$ the equality is valid. Because of the relation in 1) and the induction hypothesis, the right-hand side equals

$$\sum_{k=0}^{m-2} \binom{m+j}{k} \Delta_1^{(k)} + \binom{m+j}{m-1} \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} f_{m-k-1}$$

$$= \sum_{k=0}^{m-2} (-1)^{m-k} \binom{m+j-k-1}{j+1} \binom{m+j}{k} f_k + \binom{m+j}{m-1} \sum_{k=0}^{m-1} (-1)^{m-k-1}$$

$$\binom{m-1}{k} f_k,$$

$$= \sum_{k=0}^{m-2} (-1)^{m-k-1} \left[\binom{m+j}{m-1} \binom{m-1}{k} - \binom{m+j-k-1}{j+1} \binom{m+j}{k}\right] f_k.$$

Because

$$\binom{m+j}{m-1}\binom{m-1}{k}=\binom{m+j-k}{j+1}\binom{m+j}{k}$$

and (3.3) the last statement is proved.

QED.

To find the forward rate curve by means of N zero-coupon bond prices we again require that the following no arbitrage relations are fulfilled:

$$\sum_{j=0}^{\overline{T}_k-1} f_j h = Y_k T_k,\ k=1,\dots,N.$$

Using lemma 3.1 this is equivalent to

$$h\sum_{i=0}^{m-1} f_i + h\sum_{i=0}^{\overline{T}_k-m-1}\left[\sum_{j=1}^{i+1}\Delta_j^{(m)}\binom{m+i-j}{i+1-j}+\sum_{j=0}^{m-1}(-1)^{m-j+1}\binom{i+m-j-1}{i}\right.$$
$$\left.\binom{m+i}{j}f_j\right]=Y_kT_k$$

or because of lemma A.1 (see appendix),

$$h\sum_{i=1}^{\overline{T}_k-m}\binom{\overline{T}_k-i}{m}\Delta_i^{(m)}+h\sum_{j=0}^{m-1}\left[1+(-1)^{m-j+1}\sum_{i=0}^{\overline{T}_k-m-1}\binom{i+m-j-1}{i}\binom{m+i}{j}\right]f_j$$
$$=Y_kT_k,$$

for all $k=1,\dots,N$.

Because of lemma A.4 and lemma A.6 the term in f_j, $j=0,\dots,m-1$, can be rewritten as

$$h\sum_{j=0}^{m-1}\sum_{i=j}^{m-1}(-1)^{i-j}\binom{\overline{T}_k}{i+1}\binom{i}{j}f_j$$
$$=h\sum_{i=0}^{m-1}\binom{\overline{T}_k}{i+1}\sum_{l=0}^{i}(-1)^l\binom{i}{i-l}f_{i-l},$$
$$=h\sum_{i=0}^{m-1}\binom{\overline{T}_k}{i+1}\Delta_1^{(i)},$$

where the last equality follows from lemma 3.1. So the constraints transform into

$$h\sum_{i=1}^{\overline{T}_N-m}\binom{\overline{T}_k-i}{m}^+\Delta_i^{(m)}+h\sum_{i=0}^{m-1}\binom{\overline{T}_k}{i+1}\Delta_1^{(i)}=Y_kT_k,\ k=1,\dots,N \tag{3.4}$$

with

$$\binom{\overline{T}_k-i}{m}^+=\frac{(\overline{T}_k-i)^+\dots(\overline{T}_k-i-m+1)^+}{m!}.$$

Therefore the problem to be solved to obtain the forwards f_j, $j = 0, \ldots, \overline{T}_N - 1$, is

$$\min_{\substack{f_0,\ldots,f_{m-1} \\ \Delta_1^{(m)},\ldots,\Delta_{\overline{T}_N-m}^{(m)}}} \sum_{j=1}^{\overline{T}_N-m} \left(\Delta_j^{(m)}\right)^2 \tag{3.5}$$

subject to (3.4).
The forwards f_j, $j = m, \ldots, \overline{T}_N - 1$ can then be found by means of lemma 3.1. The unique discrete forward rate curve is then given by

$$f^h(t) = f_k = \sum_{l=0}^{m-1} h^l \binom{k}{l} \frac{\Delta_1^{(l)}}{h^l} + \sum_{l=1}^{k-m+1} \binom{k-l}{m-1} \Delta_l^{(m)} \tag{3.6}$$

if $t \in (kh, (k+1)h)$ $(k = 0, \ldots, \overline{T}_N - 1)$, with $\binom{k}{l} = 0$ if $k < l$, and where $\Delta_1^{(l)}$, $l = 0, \ldots, m-1$, as function of $f_0, \ldots, f_{m-1}$, and $\Delta_l^{(m)}$, $l = 1, \ldots, \overline{T}_N - m$, are the least squares solution of problem (3.4)-(3.5).

As in the continuous case we can add the additional constraints

$$\Delta_{\overline{T}_N-j}^{(j)} = 0, \ j = 1, \ldots, m-1,$$

to obtain a flat curve towards T_N. Using the fact that because of lemma 3.1,

$$\begin{aligned}
\Delta_{\overline{T}_N-j}^{(j)} &= \sum_{k=0}^{j} (-1)^k \binom{j}{k} f_{\overline{T}_N-k-1}, \\
&= \sum_{k=0}^{j} (-1)^k \binom{j}{k} \left[\sum_{l=1}^{\overline{T}_N-m-k} \Delta_l^{(m)} \binom{\overline{T}_N-k-1-l}{m-1} + \sum_{l=0}^{m-1} \binom{\overline{T}_N-k-1}{l} \Delta_1^{(l)} \right], \\
&= \sum_{l=1}^{\overline{T}_N-m-j} \sum_{k=0}^{j} (-1)^k \binom{j}{k} \binom{\overline{T}_N-k-l-1}{m-1} \Delta_l^{(m)} + \sum_{l=0}^{j-1} \sum_{k=0}^{l} (-1)^k \binom{j}{k} \\
&\quad \binom{m+l-k-1}{m-1} \Delta_{\overline{T}_N-m-l}^{(m)} + \sum_{l=0}^{m-1} \sum_{k=0}^{j} (-1)^k \binom{j}{k} \binom{\overline{T}_N-k-1}{l} \Delta_1^{(l)}, \\
&= \sum_{l=1}^{\overline{T}_N-m-j} \binom{\overline{T}_N-l-1-j}{m-j-1} \Delta_l^{(m)} + \sum_{l=\overline{T}_N-m-j+1}^{\overline{T}_N-m} \binom{\overline{T}_N-l-1-j}{\overline{T}_N-m-l} \Delta_l^{(m)} \\
&\quad + \sum_{l=j}^{m-1} \binom{\overline{T}_N-j-1}{l-j} \Delta_1^{(l)},
\end{aligned}$$

where the last equality follows from lemma A.2 and lemma A.3 in the appendix, we can rewrite the additional constraints as

$$\sum_{l=1}^{\overline{T}_N-m} \binom{\overline{T}_N-l-j-1}{m-j-1} \Delta_l^{(m)} + \sum_{l=j}^{m-1} \binom{\overline{T}_N-j-1}{l-j} \Delta_1^{(l)} = 0, \ j = 1, \ldots, m-1.$$

By performing the following row computations on the previous system of equalities: replace equality j by

$$\frac{1}{h^j}\sum_{p=j}^{m-1}\binom{p}{j} \text{ equation } p,$$

and using lemma A.7, the additional constraints transform into

$$\sum_{l=1}^{\overline{T}_N-m}\frac{1}{h^j}\binom{\overline{T}_N-l}{m-j-1}\Delta_l^{(m)}+\sum_{l=j}^{m-1}\frac{1}{h^j}\binom{\overline{T}_N}{l-j}\Delta_1^{(l)}=0,\ j=1,\ldots,m-1. \tag{3.7}$$

Thus to find the forward rate curve we now solve problem (3.5) subject to (3.4) and (3.7).

When small errors in the no arbitrage relations are allowed, we obtain the following problem

$$\min_{\substack{f_0,\ldots,f_{m-1}\\ \Delta_1^{(m)},\ldots,\Delta_{\overline{T}_N-m}^{(m)}}}\sum_{j=1}^{\overline{T}_N-m}\left(\Delta_j^{(m)}\right)^2+\sum_{k=1}^{N}\epsilon_k^2 \tag{3.8}$$

subject to

$$h\sum_{i=1}^{\overline{T}_N-m}\binom{\overline{T}_k-i}{m}^{+}\Delta_i^{(m)}+h\sum_{i=0}^{m-1}\binom{\overline{T}_k}{i+1}\Delta_1^{(i)}+\lambda\epsilon_k=Y_kT_k,\ k=1,\ldots,N \tag{3.9}$$

with or without the additional constraints (3.7), or equivalently

$$\min_{\substack{f_0,\ldots,f_{m-1}\\ \Delta_1^{(m)},\ldots,\Delta_{\overline{T}_N-m}^{(m)}}}\sum_{j=1}^{\overline{T}_N-m}\left(\Delta_j^{(m)}\right)^2+\alpha\sum_{k=1}^{N}\left(Y_kT_k-h\sum_{j=1}^{\overline{T}_N-m}\binom{\overline{T}_k-j}{m}^{+}\Delta_j^{(m)}\right.$$
$$\left.-h\sum_{j=0}^{m-1}\binom{\overline{T}_k}{j+1}\Delta_1^{(j)}\right)^2$$

with or without (3.7).

The solution of these discrete methods is always unique, because a strictly convex function is minimized subject to linear constraints, and can be found by standard techniques of least squares optimization.

4. Convergence of the discrete methods

In this section we will show that the discrete forward rate curve $f^h(t)=f_k$, if $t\in(kh,(k+1)h)$, obtained by one of the discrete methods of section 3, converges uniformly to the continuous forward rate curve $f(t)$, solution of the corresponding continuous problem in section 2, if h goes to zero.

To proof this relation we will first of all rewrite the discrete problems in matrix notation and eliminate, by means of the first m constraints, $f_0, \ldots, f_{m-1}$ from the other constraints.

First of all consider problem (3.4)-(3.5). The terms in $f_0, \ldots, f_{m-1}$ are eliminated in the last $N-m$ constraints in (3.4) by replacing the equation s, $s = m+1, \ldots, N$, with

$$\sum_{p=1}^{m} X_{p,s}.\text{equation } p + X_{s,s}.\text{equation } s \tag{4.1}$$

with for $p = 1, \ldots, m$

$$X_{p,s} = (-1)^{p+1} \left(\prod_{l \in \{1,\ldots,p-1,p+1,\ldots,m,s\}} T_l \right) \left(\prod_{\substack{l>r \\ l,r \in \{1,\ldots,p-1,p+1,\ldots,m,s\}}} (T_l - T_r) \right),$$

$$X_{s,s} = (-1)^{m+2} \left(\prod_{l \in \{1,\ldots,m\}} T_l \right) \left(\prod_{\substack{l>r \\ l,r \in \{1,\ldots,m\}}} (T_l - T_r) \right).$$

To see that this elimination works, we will calculate the term in $\Delta_1^{(j)}$, $j = 0, \ldots, m-1$, in equation $s = m+1, \ldots, N$ after applying (4.1). This term equals

$$\sum_{p=1}^{m} h \binom{\frac{T_p}{h}}{j+1} X_{p,s} + h \binom{\frac{T_s}{h}}{j+1} X_{s,s} \tag{4.2}$$

$$\begin{aligned}
&= \sum_{p=1}^{m} (-1)^{p+1} \frac{(T_p - h)\ldots(T_p - jh)}{(j+1)!\,h^j} \prod_{l \in \{1,\ldots,m,s\}} T_l \prod_{\substack{l>i \\ l,i \in \{1,\ldots,p-1,p+1,\ldots,m,s\}}} (T_l - T_i) \\
&\quad + (-1)^m \frac{(T_s - h)\ldots(T_s - jh)}{(j+1)!\,h^j} \prod_{l \in \{1,\ldots,m,s\}} T_l \prod_{\substack{l>i \\ l,i \in \{1,\ldots,m\}}} (T_l - T_i), \\
&= (-1)^m \prod_{l \in \{1,\ldots,m,s\}} T_l \prod_{\substack{l>i \\ l,i \in \{1,\ldots,m,s\}}} (T_l - T_i) \\
&\quad \left[\sum_{p=1}^{m} \frac{g(T_p)}{\prod_{l \in \{1,\ldots,p-1,p+1,\ldots,m,s\}} (T_p - T_l)} + \frac{g(T_s)}{\prod_{l \in \{1,\ldots,m\}} (T_s - T_l)} \right]
\end{aligned} \tag{4.3}$$

with the function $g(t)$ defined as

$$g(t) = \frac{(t-h)\ldots(t-jh)}{(j+1)!\,h^j}.$$

Because the expression (4.3) is the divided difference of the function g with respect to $T_1, \ldots, T_m, T_s$ and because g is a polynomial of degree $j \le m-1$, the expression is zero. Therefore the terms in $\Delta_1^{(j)}$, $j = 0, \ldots, m-1$, disappear when applying (4.1), what proves that the elimination works.

From (4.2) it follows that the coefficients $X_{1,s},\dots,X_{m,s},X_{s,s}$, $s=m+1,\dots,N$, satisfy

$$\sum_{p=1}^{m} T_p^{l+1}X_{p,s}+T_s^{l+1}X_{s,s}=0,\ l=0,\dots,m-1. \tag{4.4}$$

Rewriting problem (3.4)-(3.5), after applying (4.1), in matrix notation yields

$$\min_{x^h} \|x^h\|_2^2 \tag{4.5}$$

subject to

$$A(h)x^h=b \tag{4.6}$$

and with $\Delta_{1,h}^{(0)},\dots,\Delta_{1,h}^{(m-1)}$ solution to

$$h\sum_{j=0}^{m-1}\binom{\overline{T}_k}{j+1}\Delta_{1,h}^{(j)}+h\sum_{j=1}^{\overline{T}_N-m}\binom{\overline{T}_k-j}{m}^{+}x_j^h=Y_kT_k,\ k=1,\dots,m. \tag{4.7}$$

The notations used are

$$\begin{aligned}
x_j^h &= \Delta_j^{(m)},\\
A(h)_{i-m,j} &= \sum_{p=1}^{m} h\binom{\overline{T}_p-j}{m}^{+}X_{p,i}+h\binom{\overline{T}_i-j}{m}^{+}X_{i,i},\\
b_{i-m} &= \sum_{p=1}^{m} Y_pT_pX_{p,i}+Y_iT_iX_{i,i},
\end{aligned}$$

for $i=m+1,\dots,N$ and $j=1,\dots,\overline{T}_N-m$.
The super- or subscript h denotes that a discretization step h is used in the discrete problem.

In the next theorem it is proved that the discrete forward rate curve, as a solution of (4.5)-(4.7) converges uniformly to the continuous forward rate curve (2.3).

Theorem 4.1

Let $(f_0^h,\dots,f_{m-1}^h,\Delta_{1,h}^{(m)},\dots,\Delta_{\overline{T}_N-m,h}^{(m)})$ be the solution of problem (4.5)-(4.7), and let the function $f^h(t)$ defined as $f^h(t)=f_k^h$, for $t\in(kh,(k+1)h)$, $k=0,\dots,\overline{T}_N-1$, be the unique discrete optimal forward rate curve.
Then, if the discretization step h goes to zero, the function $f^h(t)$ converges uniformly to the unique continuous optimal forward rate curve $f(t)$, solution of problem (2.1)-(2.2).

Proof

The solution of (4.5)-(4.6) is given by

$$x^h=A(h)^t(A(h)A(h)^t)^{-1}b\equiv A(h)^tC(h).$$

So we have that for $j = 1, \dots, \overline{T}_N - m$,

$$x_j^h = \sum_{k=1}^{N-m} h \binom{\overline{T}_{m+k} - j}{m}^+ X_{m+k,m+k} C(h)_k + \sum_{p=1}^{m} h \binom{\overline{T}_p - j}{m}^+ \sum_{k=1}^{N-m} X_{p,m+k} C(h)_k,$$

$$= \sum_{k=1}^{N} a_k^h h^m \frac{(T_k - jh)^+ \dots (T_k - (j+m-1)h)^+}{m!} \tag{4.8}$$

with

$$a_p^h = \sum_{k=1}^{N-m} X_{p,m+k} \frac{C(h)_k}{h^{2m-1}}, \; p = 1, \dots, m, \tag{4.9}$$

$$a_{k+m}^h = X_{m+k,m+k} \frac{C(h)_k}{h^{2m-1}}. \tag{4.10}$$

Furthermore, if h goes to zero, then the matrix $h^{2m-1}(A(h)A(h)^t)$ converges to the matrix E with elements $(i, j = 1, \dots, N - m)$

$$\begin{aligned} E_{ij} &= \sum_{p=1}^{m} \sum_{q=1}^{m} X_{p,m+i} X_{q,m+j} \int_0^T \frac{((T_p - u)^+)^m}{m!} \frac{((T_q - u)^+)^m}{m!} du \\ &+ \sum_{p=1}^{m} X_{p,m+i} X_{m+j,m+j} \int_0^T \frac{((T_p - u)^+)^m}{m!} \frac{((T_{m+j} - u)^+)^m}{m!} du \\ &+ \sum_{q=1}^{m} X_{q,m+j} X_{m+i,m+i} \int_0^T \frac{((T_q - u)^+)^m}{m!} \frac{((T_{m+i} - u)^+)^m}{m!} du \\ &+ X_{m+i,m+i} X_{m+j,m+j} \int_0^T \frac{((T_{m+i} - u)^+)^m}{m!} \frac{((T_{m+j} - u)^+)^m}{m!} du, \\ &= X_{m+i,m+i}(X_{m+j,m+j} < h_{m+i}, h_{m+j} >_m + \sum_{q=1}^{m} X_{q,m+j} < h_q, h_{m+i} >_m) \\ &+ \sum_{p=1}^{m} X_{p,m+i}(X_{m+j,m+j} < h_p, h_{m+j} >_m + \sum_{q=1}^{m} X_{q,m+j} < h_p, h_q >_m), \end{aligned}$$

where the last equality follows from (4.4). In the same way as in Lorimier (1995a) it can be shown that the matrix E is strictly positive definite and therefore nonsingular.
So, we have that a_p^h, $p = 1, \dots, m$, converges to

$$a_p \equiv \sum_{k=1}^{N-m} X_{p,m+k} \sum_{j=1}^{N-m} E_{kj}^{-1} b_j \tag{4.11}$$

and that a_{m+k}^h, $k = 1, \dots, N - m$, converges to

$$a_{m+k} \equiv X_{m+k,m+k} \sum_{j=1}^{N-m} E_{kj}^{-1} b_j, \tag{4.12}$$

if h goes to zero.

Remark that because of (4.4) the following relations are valid:

$$\sum_{k=1}^{N} a_k T_k^l = 0,\ l = 1, \ldots, m.$$

From (4.7) it follows that, if h goes to zero, the vector $(\Delta_{1,h}^{(j)}/h^j)_{j=0,\ldots,m-1}$ converges to $(f_0^{(j)})_{j=0,\ldots,m-1}$, solution of

$$\sum_{i=0}^{m-1} f_0^{(i)} \frac{T_k^{i+1}}{(i+1)!} + \sum_{l=1}^{N} a_l < h_l, h_k >_m = Y_k T_k,\ k = 1, \ldots, N.$$

Therefore $(f_0, \ldots, f_0^{(m-1)}, a_1, \ldots, a_N)$ satisfy (2.4)-(2.5), what means that the unique optimal solution of the continuous problem (2.1)-(2.2) is

$$f(t) = \sum_{i=0}^{m-1} f^{(i)}(0)\frac{t^i}{i!} + \sum_{k=1}^{N} a_k h_k(t).$$

To prove the theorem we only have to show that

$$f^h(t) = f_k^h = \sum_{l=0}^{m-1} h^l \binom{k}{l} \frac{\Delta_{1,h}^{(l)}}{h^l} + \sum_{l=1}^{k-m+1} \binom{k-l}{m-1} x_l^h$$

if $t \in (kh, (k+1)h)$, $k = 0, \ldots, \overline{T}_N - 1$, converges uniformly to $f(t)$. Using the form (4.8) this can easily be proved.

QED.

In quite an analogue way the convergence for the problem with the additional constraints can be proved. Namely, first of all the term in $\Delta_1^{(l)}$, $l = 0, \ldots, m-1$, is eliminated from the last $N-m$ constraints in (3.4) as before. To eliminate these terms in the $m-1$ additional constraints (3.7) we again use the first m equations in (3.4) by applying the following row operations: replace equation j, $j = N+1, \ldots, N+m-1$ in (3.7) by

$$\sum_{p=1}^{m} X_{p,j}.\text{equation } p + \sum_{p=j}^{N+m-1} h^{p-j} X_{p,j}.\text{equation } p, \tag{4.13}$$

where $X_{p,j}$, $p = 1, \ldots, m, j, \ldots, N+m-1$, satisfy

$$\sum_{p=1}^{m} h \binom{\overline{T}_p}{l+1} X_{p,j} = 0,\ l = 0, \ldots, j-N-1,$$

$$\sum_{p=1}^{m} h \binom{\overline{T}_p}{l+1} X_{p,j} + \sum_{p=j}^{N+l} h^{N-j} \binom{\overline{T}_N}{l-p+N} X_{p,j} = 0,\ l = j-N, \ldots, m-1.$$

It can be shown that this system of equations is equivalent to

$$\sum_{p=1}^{m} T_p^{l+1} X_{p,N+j} = 0,\ l = 0,\dots,j-1, \tag{4.14}$$

$$\sum_{p=1}^{m} \frac{T_p^{l+1}}{(l+1)!} X_{p,N+j} + \frac{T_N^{l-j}}{(l-j)!} X_{N+j,N+j} = 0,\ l = j,\dots,m-1, \tag{4.15}$$

$$(-1)^{p-j-1} \frac{C_{p-j,p}}{(p+1)\dots(j+2)} X_{N+j,N+j} + X_{N+p,N+j} = 0,\ p = j+1,\dots,m-1, \tag{4.16}$$

for all $j = 1,\dots,m-1$, where, for $q \geq p$,

$$C_{q,p} = \sum_{i_1=1}^{p-q+1} \sum_{i_2=i_1+1}^{p-q+2} \cdots \sum_{i_q=i_{q-1}+1}^{p} i_1.i_2\dots i_q.$$

Therefore the coefficients $X_{1,j},\dots,X_{m,j},X_{j,j},\dots,X_{N+m-1,j}$ can be choosen independent of h.

Rewriting problem (3.5) subject to (3.4), (3.7) in matrix notation after applying the elimination (4.13), yields

$$\min_{x^h} \|x^h\|_2^2 \tag{4.17}$$

subject to

$$A(h)x^h = b \tag{4.18}$$

with x_j^h, $A(h)_{i-m,j}$ and b_{i-m} for $i = m+1,\dots,N$, $j = 1,\dots,\overline{T}_N - m$ as in (4.6) and, with for $i = N+1,\dots,N+m-1$, $j = 1,\dots,\overline{T}_N - m$

$$A(h)_{i-m,j} = \sum_{p=1}^{m} h \binom{\overline{T}_p - j}{m}^{+} X_{p,i} + \sum_{p=i-N}^{m-1} h^{N-i} \binom{\overline{T}_N - j}{m-p-1} X_{N+p,i}$$

$$b_{i-N} = \sum_{p=1}^{m} Y_p T_p X_{p,i}.$$

Furthermore $\Delta_{1,h}^{(0)},\dots,\Delta_{1,h}^{(m-1)}$ are solution to (4.7).
For this new problem the following theorem states that the uniformly convergence still holds.

Theorem 4.2

Let $(f_0^h,\dots,f_{m-1}^h,\Delta_{1,h}^{(m)},\dots,\Delta_{\overline{T}_N-m,h}^{(m)})$ be the solution of problem (4.17)-(4.18), and let the function $f^h(t)$ defined as $f^h(t) = f_k^h$, for $t \in (kh,(k+1)h)$, $k = 0,\dots,\overline{T}_N - 1$, be the unique discrete optimal forward rate curve.
Then, if the discretization step h goes to zero, the function $f^h(t)$ converges uniformly to the unique continuous optimal forward rate curve $f(t)$, solution of problem (2.1)-(2.2), (2.6).

Proof

In the same way as in theorem 4.1 it can be shown that

$$\begin{aligned} x_j^h = & \sum_{k=1}^{N} a_k^h h^m \frac{(T_k - jh)^+ \dots (T_k - (j+m-1)h)^+}{m!} \\ & + \sum_{p=1}^{m-1} a_{N+p}^h h^m \frac{(T_N - jh) \dots (T_N - (j+m-p-2)h)}{(m-p-1)!} \end{aligned}$$

with

$$\begin{aligned} a_p^h &= \sum_{k=1}^{N-m} X_{p,m+k} \frac{C(h)_k}{h^{2m-1}} + \sum_{k=N-m+1}^{N-1} X_{p,m+k} \frac{C(h)_k}{h^{2m-1}}, \ p = 1, \dots, m, \\ a_{m+k}^h &= X_{m+k,m+k} \frac{C(h)_k}{h^{2m-1}}, \ k = 1, \dots, N-m, \\ a_{N+p}^h &= \sum_{l=1}^{p} h^{p-l} X_{N+p,N+l} \frac{C(h)_{N-m+l}}{h^{2m-1}}, p = 1, \dots, m-1. \end{aligned}$$

As in theorem 4.1 the matrix $h^{2m-1}(A(h)A(h)^t)$ converges to the matrix E, with elements $(k = 1, \dots, N-1, \ j = 1, \dots, N-1)$

$$\begin{aligned} E_{kj} = & X_{m+k,m+k}(X_{m+j,m+j} < h_{m+k}, h_{m+j} >_m + \sum_{q=1}^{m} X_{q,m+j} < h_{m+k}, h_q >_m) \\ & + \sum_{p=1}^{m} X_{p,m+k}(X_{m+j,m+j} < h_p, h_{m+j} >_m + \sum_{q=1}^{m} X_{q,m+j} < h_p, h_q >_m), \end{aligned}$$

if h goes to zero. Because the matrix E is strictly positive definite, we have that a_p^h, $p = 1, \dots, m$, converges to

$$a_p \equiv \sum_{k=1}^{N-m} X_{p,m+k}(E^{-1}b)_k + \sum_{k=N-m+1}^{N-1} X_{p,m+k}(E^{-1}b)_k,$$

that a_{m+k}^h, $k = 1, \dots, N-m$, converges to

$$a_{m+k} \equiv X_{m+k,m+k}(E^{-1}b)_k$$

and that a_{N+p}^h, $p = 1, \dots, m-1$, converges to

$$a_{N+p} \equiv X_{N+p,N+p}(E^{-1}b)_{N-m+p}.$$

Using the relations (4.4), (4.14)-(4.15) it can easily be found that

$$\sum_{k=1}^{N} a_k \frac{T_k^{l+1}}{(l+1)!} + \sum_{k=1}^{l} a_{N+k} \frac{T^{l-k}}{(l-k)!} = 0$$

for all $l = 0, \ldots, m-1$.
In an analogue way as in theorem 4.1 it can be shown that $(\Delta_{1,h}^{(i)}/h^i)_{i=0,\ldots,m-1}$ converges to $(f_0^{(i)})_{i=0,\ldots,m-1}$, solution to (2.10), and that $(f_0, \ldots, f_0^{(m-1)}, a_1, \ldots, a_{N+m-1})$ also satisfies (2.11).
Therefore the unique solution of the continuous problem is of the form

$$f(t) = \sum_{i=0}^{m-1} f_0^{(i)} \frac{t^i}{i!} + \sum_{k=1}^{N+m-1} a_k h_k(t)$$

and in the same way as in theorem 4.1 it follows that the discrete forward rate curve converges uniformly to $f(t)$.

QED.

Finally, we will investigate the convergence of the discrete method (3.8)-(3.9), with or without (3.7), a method that allows small errors in the no arbitrage relations.
Let us first assume that the additional constraints (3.7) are not imposed. Then the minimization problem can be rewritten as

$$\min_{x^h, \Delta_1^h} \|x^h\|_2^2 + \alpha h^{2m-1} \|b - M(h)\Delta_1^h - A(h)x^h\|_2^2 \tag{4.19}$$

with, for $j = 1, \ldots, \overline{T}_N - m$, $k = 1, \ldots, N$, $l = 1, \ldots, m$,

$$\begin{aligned} x_j^h &= \Delta_j^{(m)}, \\ b_k &= Y_k T_k, \\ M(h)_{kl} &= h^l \binom{\overline{T}_k}{l}, \\ (\Delta_1^h)_l &= \frac{\Delta_1^{(l-1)}}{h^{l-1}}, \\ A(h)_{kj} &= h \binom{\overline{T}_k - j}{m}^+ . \end{aligned}$$

The superscript h again denotes that a discrete method with discretization step h is used. Remark also that problem (4.19) uses αh^{2m-1} instead of α.

The next lemma presents the form of the optimal solution.

Lemma 4.1

1) The optimal solution of problem (4.19) is solution to

$$x^h - \alpha h^{2m-1} A(h)^t (b - M(h)\Delta_1^h - A(h)x^h) = 0, \tag{4.20}$$

$$M(h)^t (b - M(h)\Delta_1^h - A(h)x^h) = 0. \tag{4.21}$$

2) The optimal solution is given by

$$x^h = h^{2m-1}A(h)^t a^h, \tag{4.22}$$

$$\Delta_1^h = (M(h)^t E(h)^{-1} M(h))^{-1} M(h)^t E(h)^{-1} b \tag{4.23}$$

with

$$E(h) = I_N + \alpha h^{2m-1} A(h) A(h)^t,$$
$$a^h = \alpha E(h)^{-1}(b - M(h)\Delta_1^h).$$

Proof

1) Follows immediately from the first order conditions.

2) From (4.20) it follows that

$$x^h = \alpha h^{2m-1}(I_{\overline{T}_N - m} + \alpha h^{2m-1} A(h)^t A(h))^{-1} A(h)^t (b - M(h)\Delta_1^h). \tag{4.24}$$

Substitution in (4.21) yields

$$M(h)^t (I_N - \alpha h^{2m-1} A(h)(I_{\overline{T}_N - m} + \alpha h^{2m-1} A(h)^t A(h))^{-1} A(h)^t)(b - M(h)\Delta_1^h) = 0$$

or

$$M(h)^t E(h)^{-1}(M(h)\Delta_1^h - b) = 0,$$

what proves (4.21).
Because of the following matrix relation:

$$(I_M + \beta B^t B)^{-1} B^t = B^t (I_N + \beta B B^t)^{-1},$$

for all N-by-M matrices B and constants $\beta > 0$, we have that (4.24) can be rewritten as

$$x^h = \alpha h^{2m-1} A(h)^t (I_N + \alpha h^{2m-1} A(h) A(h)^t)^{-1}(b - M(h)\Delta_1^h),$$

what proves (4.20).

QED.

Using the results in lemma 4.1 the convergence can be proved.

Theorem 4.3

Let $(f_0^h, \ldots, f_{m-1}^h, \Delta_{1;h}^{(m)}, \ldots, \Delta_{\overline{T}_N - m,h}^{(m)})$ be the solution of problem (4.19), and let the function $f^h(t)$ defined as $f^h(t) = f_k^h$, for $t \in (kh, (k+1)h)$, $k = 0, \ldots, \overline{T}_N - 1$, be the unique discrete optimal forward rate curve.
Then, if the discretization step h goes to zero, the function $f^h(t)$ converges uniformly to the unique continuous optimal forward rate curve $f(t)$, solution of problem (2.12).

Proof

First of all remark that the matrix $h^{2m-1}A(h)A(h)^t$ converges to a matrix with elements

$$\int_0^T \frac{((T_k-u)^+)^m}{m!}\frac{((T_l-u)^+)^m}{m!}\,du = H_{kl} - (\tilde{T}_m\tilde{T}_m^t)_{kl},$$

for $k,l = 1,\ldots,N$, where H is the Gram-Schmidt matrix of the functions h_k, $k = 1,\ldots,N$ (defined in section 2) and where

$$(\tilde{T}_m)_{ki} = T_k^i \Big/ i!,\ k = 1,\ldots,N, i = 1,\ldots,m.$$

Therefore the matrix $E(h)$ converges to the matrix $E = I_N + \alpha H - \alpha\tilde{T}_m\tilde{T}_m^t$. Because the matrix $H - \tilde{T}_m\tilde{T}_m^t$ is strictly positive definite, the matrix E is nonsingular.
Furthermore we have that the matrix $M(h)$ converges to the matrix $\tilde{T}_m$, if h goes to zero. Consequently, the vector Δ_1^h converges to

$$F \equiv (\tilde{T}_m^t E^{-1}\tilde{T}_m)^{-1}\tilde{T}_m^t E^{-1} b \tag{4.25}$$

and the vector a^h converges to

$$a \equiv \alpha E^{-1}(b - \tilde{T}_m F). \tag{4.26}$$

Then we have that

$$\tilde{T}_m^t a = \alpha\tilde{T}_m^t E^{-1} b - \alpha(\tilde{T}_m^t E^{-1}\tilde{T}_m)F = 0,$$

because of (4.25) and, from (4.26) it also follows that

$$(I_N + \alpha H)a = \alpha(b - \tilde{T}_m F),$$

what means that (F, a) satisfy the equations (2.13)-(2.14). Thus the unique optimal solution of problem (2.12) is of the form (2.3). In the same way as in theorem 4.1 it can be shown that the optimal discrete forward rate curve converges uniformly to the continuous one, when h goes to zero.

QED.

Let us now consider problem (3.8)-(3.9) with the additional constraints (3.7), that can be rewritten as

$$M_2(h)x^h + M_1(h)y^h = 0$$

with

$$x_k^h = \Delta_k^{(m)},$$

$$(M_1(h))_{jl} = \begin{cases} h^l \binom{\overline{T}_N}{l-j} & l = j,\ldots,m-1, \\ 0 & l = 1,\ldots,j-1, \end{cases}$$

$$(M_2(h))_{jk} = \binom{\overline{T}_N - k}{m-j-1},$$

$$y_j^h = \Delta_1^{(j)}/h^j,$$

for $k = 1, \ldots, \overline{T}_N - m$, $j = 1, \ldots, m-1$. This means that

$$y^h = -M_1(h)^{-1} M_2(h) x^h. \tag{4.27}$$

Elimination of y^h in (3.9) by means of (4.27) yields the following problem in matrix notation, when αh^{2m-1} instead of α is used:

$$\min_{x^h, f_0^h} \|x^h\|_2^2 + \alpha h^{2m-1} \|b - f_0^h t + R(h) M_1(h)^{-1} M_2(h) x^h - A(h) x^h\|_2^2 \tag{4.28}$$

subject to (4.27), where $A(h)$ is defined as in (4.19), $t^t = (T_1, \ldots, T_N)$, $f_0^h = \Delta_1^{(0)}$ and where

$$R(h)_{kl} = h^{l+1} \binom{\overline{T}_k}{l+1}, \quad k = 1, \ldots, N, l = 1, \ldots, m-1.$$

The solution is given by:

Lemma 4.2

The optimal solution of (4.28) is given by

$$x^h = \alpha h^{2m-1} B(h)^t E(h)^{-1} (b - f_0^h t), \tag{4.29}$$

$$f_0^h = \frac{t^t E(h)^{-1} b}{t^t E(h)^{-1} t} \tag{4.30}$$

with

$$B(h) = A(h) - R(h) M_1(h)^{-1} M_2(h),$$
$$E(h) = I_N + \alpha h^{2m-1} B(h) B(h)^t.$$

Proof

The first order conditions for the minimization problem (4.28) are

$$x^h + \alpha h^{2m-1} (R(h) M_1(h)^{-1} M_2(h) - A(h))^t (b - f_0^h t + R(h) M_1(h)^{-1} M_2(h) x^h - A(h) x^h) = 0, \tag{4.31}$$

$$t^t (b - f_0^h t + R(h) M_1(h)^{-1} M_2(h) x^h - A(h) x^h) = 0. \tag{4.32}$$

From the first equality it follows that

$$x^h = \alpha h^{2m-1} (I_{\overline{T}_N - m} + \alpha h^{2m-1} B(h)^t B(h))^{-1} B(h)^t (b - f_0^h t) \tag{4.33}$$

Substitution of (4.33) into (4.32) results in

$$t^t (I_N - \alpha h^{2m-1} B(h) (I_{\overline{T}_N - m} + \alpha h^{2m-1} B(h)^t B(h))^{-1} B(h)^t)(b - f_0^h t) = 0$$

Because

$$I_N - \alpha h^{2m-1} B(h) (I_{\overline{T}_N - m} + \alpha h^{2m-1} B(h)^t B(h))^{-1} B(h)^t = (I_N + \alpha h^{2m-1} B(h) B(h)^t)^{-1}$$

relation (4.30) follows.
In the same way as in lemma 4.1, relation (4.33) also yields (4.29).

QED.

The next theorem shows that the solution still converges to the one of the continuous counterpart.

Theorem 4.4

Let $(f_0^h, \ldots, f_{m-1}^h, \Delta_{1,h}^{(m)}, \ldots, \Delta_{\overline{T}_N-m,h}^{(m)})$ be the solution of problem (4.27)-(4.28), and let the function $f^h(t)$ defined as $f^h(t) = f_k^h$, for $t \in (kh, (k+1)h)$, $k = 0, \ldots, \overline{T}_N - 1$, be the unique discrete optimal forward rate curve.
Then, if the discretization step h goes to zero, the function $f^h(t)$ converges uniformly to the unique continuous optimal forward rate curve $f(t)$, solution of problem (2.12), (2.6).

Proof

First of all remark that from (4.29) it follows that

$$x^h = h^{2m-1} B(h)^t a^h$$

with $a^h = \alpha E(h)^{-1}(b - f_0^h t)$.
Furthermore we have that the matrix $R(h)$ converges to the matrix $\hat{T}_m$ with $(\hat{T}_m)_{kl} = T_k^{l+1}/(l+1)!$, $k = 1, \ldots, N$, $l = 1, \ldots, m-1$, and that the matrix $d(h)$ with $(d(h))_{jl} = M_1(h)_{jl}/h^j$ converges to the matrix $\hat{S}_m$ with $(\hat{S}_m)_{jl} = T_N^{l-j}/(l-j)!$ for $l \geq j$ and zero otherwise, if h goes to zero. So we have that

$$h^u((M_1(h)^t)^{-1} R(h)^t)_{ul} = ((d(h)^t)^{-1} R(h)^t)_{ul},$$

$u = 1, \ldots, m-1$, $l = 1, \ldots, N$, converges to $((\hat{S}_m^t)^{-1} \hat{T}_m^t)_{ul}$.
Therefore the matrix $h^{2m-1} A(h) M_2(h)^t (M_1(h)^t)^{-1} R(h)^t$ converges to a matrix with elements

$$\sum_{u=1}^{m-1} ((\hat{S}_m^t)^{-1} \hat{T}_m^t)_{ul} (< h_k, h_{N+u} >_m - \sum_{i=u}^{m-1} \frac{T_k^{i+1}}{(i+1)!} \frac{T_N^{i-u}}{(i-u)!})$$
$$= (V (\hat{S}_m^t)^{-1} \hat{T}_m^t)_{kl} - (\hat{T}_m \hat{T}_m^t)_{kl},$$

for $k, l = 1, \ldots, N$, with $V_{ku} =< h_k, h_{N+u} >_m$.
In the same way it follows that the matrix $h^{2m-1} R(h) M_1(h)^{-1} M_2(h) A(h)^t$ converges to a matrix with elements

$$(\hat{T}_m \hat{S}_m^{-1} V^t)_{kl} - (\hat{T}_m \hat{T}_m^t)_{kl}, \; k, l = 1, \ldots, N,$$

and that the matrix $h^{2m-1} R(h) M_1(h)^{-1} M_2(h) M_2(h)^t (M_1(h)^t)^{-1} R(h)^t$ converges to a matrix with elements

$$(\hat{T}_m \hat{S}_m^{-1} \hat{H} (\hat{S}_m^t)^{-1} \hat{T}_m^t)_{kl} - (\hat{T}_m \hat{T}_m^t)_{kl}, \; k, l = 1, \ldots, N,$$

where $\hat{H}_{uv} =< h_{N+u}, h_{N+v} >_m$. This proves that the matrix $h^{2m-1} B(h) B(h)^t$ converges to the matrix

$$H - tt^t - V (\hat{S}_m^t)^{-1} \hat{T}_m^t - \hat{T}_m \hat{S}_m^{-1} V^t + \hat{T}_m \hat{S}_m^{-1} \hat{H} (\hat{S}_m^t)^{-1} \hat{T}_m^t,$$

because $\tilde{T}_m\tilde{T}_m^t - \hat{T}_m\hat{T}_m^t = tt^t$. So, we have that the matrix $E(h)$ converges to the matrix $E = I_N + \alpha H - \alpha tt^t - \alpha V(\hat{S}_m^t)^{-1}\hat{T}_m^t - \alpha\hat{T}_m\hat{S}_m^{-1}V^t + \alpha\hat{T}_m\hat{S}_m^{-1}\hat{H}(\hat{S}_m^t)^{-1}\hat{T}_m^t$.
As in the previous theorems it can be proved that the matrix E is nonsingular. Consequently, the value f_0^h converges to

$$f_0 \equiv \frac{t^tE^{-1}b}{t^tE^{-1}t} \tag{4.34}$$

and the vector a^h converges to

$$a \equiv \alpha E^{-1}(b - f_0t), \tag{4.35}$$

when h goes to zero.
Define the vector $\tilde{a}^t = (a_{N+1}, \ldots, a_{N+m-1})$ as

$$\tilde{a} \equiv -(\hat{S}_m^t)^{-1}\hat{T}_m^t a. \tag{4.36}$$

Then we have that

$$\hat{S}_m^t\tilde{a} + \hat{T}_m^t a = 0 \tag{4.37}$$

and that

$$t^t a = 0, \tag{4.38}$$

what means that $(a, \tilde{a})$ satisfy (2.15). Furthermore, from (4.27) we have that

$$\begin{aligned} y^h &= -h^{2m-1}M_1(h)^{-1}M_2(h)B(h)^t a^h, \\ &= -h^{2m-1}M_1(h)^{-1}M_2(h)A(h)^t a^h + h^{2m-1}M_1(h)^{-1}M_2(h)M_2(h)^t(M_1(h)^t)^{-1}R(h)^t a^h. \end{aligned}$$

In quite an analogue way as above, we find that the vector y^h converges to

$$G \equiv -\hat{S}_m^{-1}V^t a + \hat{S}_m^{-1}\hat{H}(\hat{S}_m^t)^{-1}\hat{T}_m^t a, \tag{4.39}$$

with $G^t = (f_0^{(1)}, \ldots, f_0^{(m-1)})$.
From (4.35) together with (4.36), (4.38) and (4.39) it follows that

$$a = \alpha(b - f_0t - \hat{T}_mG - Ha - V\tilde{a}).$$

This means that relation (2.16) is fulfilled for $(f_0, G, a, \tilde{a})$. Finally it is easy to prove that

$$\hat{S}_mG + Va + \hat{H}\tilde{a} = 0,$$

what shows that also relation (2.17) holds.
Thus, the unique optimal solution of the continuous problem (2.12), (2.6) is

$$f(t) = \sum_{i=0}^{m-1} f_0^{(i)}\frac{t^i}{i!} + \sum_{k=1}^{N+m-1} a_kh_k(t).$$

Some calculations then again show that the discrete forward rate curve $f^h(t)$ converges uniformly to the function $f(t)$.

QED.

5. Comparison between methods using first and second differences

In this section the performance of the methods using first and second differences in the measure of degree of smoothness is compared in a discrete time framework with discretization step $h = 1$ day. Therefore we will simulate the data needed to estimate the term structure by means of the CIR-model (see Cox, Ingersoll, Ross (1985)). This means that the yield-to-maturity Y_i of a zero-coupon bond with maturity T_i is calculated as follows:

$$Y_i = \frac{r_0 B(T_i) - \ln(A(T_i))}{T_i} \tag{5.1}$$

with

$$\begin{aligned} A(T_i) &= \left(\frac{2\gamma \exp((\kappa+\gamma)T_i/2)}{(\gamma+\kappa)(\exp(\gamma T_i)-1)+2\gamma} \right)^{2\kappa\theta/\sigma^2}, \\ B(T_i) &= \frac{2(\exp(\gamma T_i)-1)}{(\gamma+\kappa)(\exp(\gamma T_i)-1)+2\gamma}, \\ \gamma &= \sqrt{\kappa^2+2\sigma^2}, \end{aligned}$$

where κ, θ, σ are the parameters of the model and r_0 the current interest rate.
To obtain the corresponding forward rate f_i the following formula can be used:

$$f_i = r_0 \frac{4\gamma^2 \exp(\gamma T_i)}{((\kappa+\gamma)(\exp(\gamma T_i)-1)+2\gamma)^2} + \kappa\theta B(T_i). \tag{5.2}$$

Figures 5.1 and 5.2 present the results obtained with method (3.4)-(3.5) with or without the additional constraints (3.7) for $m = 1$ and $m = 2$, when the data consists of the yields-to-maturity of zero-coupon bonds maturing after 1, 3, 5, 7, 9, 11 months, simulated with (5.1) for $(\kappa, \theta, \sigma, r_0) = (0.75, 0.1, 0.15, 0.0995)$. The simulation is performed on May 8th, 1991.
It is clear that the forward rate curve and the yield curve obtained for $m = 2$ do not flatten before the smallest observable maturity; however for $m = 1$ the flattening effect is clearly visible. This behaviour was to be expected because in section 2 it was shown that $f'(0) \neq 0$ in general, when $m = 2$, while in Lorimier (1995a) it followed that for $m = 1$ the first derivative $f'(0)$ is zero. Finally, we notice that the forward rate curve obtained by means of the method with $m = 1$ and $m = 2$,with (3.7), flattens towards the end, what is caused by the zero derivative $f'(T)$ (see section 2). That this phenomenon creates a curve not completely coinciding with the exact one is of course due to the fact that the time horizon of 11 months is to short to notice the flattening effect in the exact curve. Consequently, the method with $m = 2$ but without (3.7) gives the best results for this

situation. However when T is much larger the method with the additional constraint (3.7) will yield better results than the one without, because now the flattening effect will appear in the exact curve.

To test the methods allowing small errors in the no arbitrage relations, we have to use data that are slightly perturbated. That is why we will use yields-to-maturity calculated as

$$\tilde{Y}_i = Y_i \pm \epsilon, \tag{5.3}$$

where Y_i are the exact yields obtained with (5.1) and where ϵ is a small error. To decide whether the yield is augmented or diminished with ϵ, we assume that the probability for an increase and for a decrease are equal. So for each maturity we generate a discrete uniform random variate X, with value 1 if a plus sign and value 2 if a minus sign has to be used. This can be done by means of the following relation:

$$X = trunc(1 + 2U)$$

with U a uniform $[0, 1]$ random variate, and $trunc(x)$ the truncation of the real value x (see Devroye (1986)).
The relative error on the price of the corresponding zero-coupon bond $P(T_i)$ is then

$$\frac{|\tilde{P}(T_i) - P(T_i)|}{P(T_i)} = |e^{\pm\epsilon T_i} - 1|.$$

Of course in practice the difference between the quoted price and the exact (unknown) price of the bond cannot be larger than, let say, 0.01%. This means, for example for a bond maturing after 3 months with an exact price of 98.65 (calculated with the CIR-model with $\kappa = 0.75$, $\theta = 0.1$, $\sigma = 0.105$ and $r_0 = 0.05$), that the fluctuation $\epsilon \in [-0.0004, 0.0004]$ and for a bond maturing after 1 year with exact price 93.73 (calculated with the same CIR-model) we have that $\epsilon \in [-0.0001, 0.0001]$.

For the perturbated data we will use observations with maturities that are no longer equidistant: yields-to-maturity for 7 days, 14 days, 1 month, 2 months, 3 months, 6 months and 1 year are simulated with (5.1), for $(\kappa, \theta, \sigma, r_0) = (0.75, 0.1, 0.105, 0.05)$, on May 8th, 1991, and perturbated with $\epsilon = 5\,10^{-4}$. We employ such a large value of ϵ, because it permits us to better understand how the methods behave when the observations are not exact.
Figures 5.3 and 5.4 present curves obtained with this new data set for the method (3.4)-(3.5) with $m = 2$, with or without (3.7), and with $m = 1$ and the Figures 5.5 and 5.6 give the results for method (3.8)-(3.9). In the latter case the α-value for the discrete method with $m = 1$ is choosen to be equal to 16 and for the discrete method with $m = 2$ to be equal to 16/133225. Because of this special choice of the values of α, we have that the continuous problems with $m = 1$ and $m = 2$, as limits of the discrete problems, use the

same α-value. Namely, in Lorimier (1995b) it is proved that the discrete problem with $m = 1$ and $\alpha = 16$ converges to the continuous problem with $\alpha = 16/h = 5840$. On the other hand from theorem 4.3 and 4.4 it follows that the discrete problem with $m = 2$ (with or without the additional constraint $f'(T) = 0$) and with $\alpha = 16/133225$ converges to the continuous problem with $\alpha = 16/(133225h^3) = 5840$. Thus, in the continuous case the methods with $m = 1$ and $m = 2$ use the same λ-value in the perturbated no arbitrage relations

$$\int_0^{T_k} f(u)\,du + \lambda\epsilon_k = Y_k T_k\,.$$

In Figure 5.3 and 5.4 it can be seen that for small maturities all the methods give practically the same yields and forwards, although for $m = 1$ the flattening effect is slightly visible. Furthermore we see that the estimated curve fluctuates around the exact one. This is of course due to the fact that perturbated data is used in a method which does not allow errors in the no arbitrage relations. It is also clear that the methods with $m = 2$ yield larger fluctuations than the method with $m = 1$, and that the fluctuations are more distinct in the forward rate curve than in the yield curve. The latter is caused by the fact that yields are averaged rates. Finally the forwards obtained with $m = 2$ without the additional constraints (3.7) tend to drift of toward large values when T is large. So the constraints (3.7) are needed to obtain a plausible behaviour of the forward rate curve towards the end.

In Figures 5.5 and 5.6 it is clear that by using method (3.8)-(3.9), which allows small perturbations in the no arbitrage relation, the errors in the observations can be compensated. For small maturities the yields obtained with the methods with $m = 2$ are nearer the exact value than for $m = 1$, because the curve does not flatten at the beginning. The same remark can be made for the forward rates. Although the yields are practically the same for all the methods when the maturity is large, the forwards are better estimated for the method with $m = 2$ without the additional constraint than for the other two methods. Of course, this is to be expected because the horizon T is not very large (only 1 year), what means that the flattening effect is not yet noticable.

Finally Figures 5.7 and 5.8 show results when some real data is used. The curves are now fitted by means of the interbank rates on 7 days, 14 days, 1 month, 2 months, 3 months, 6 months and 1 year on December 3, 1990, as can be found in the files of the National Bank of Belgium.

Figure 5.1

Yield Curve

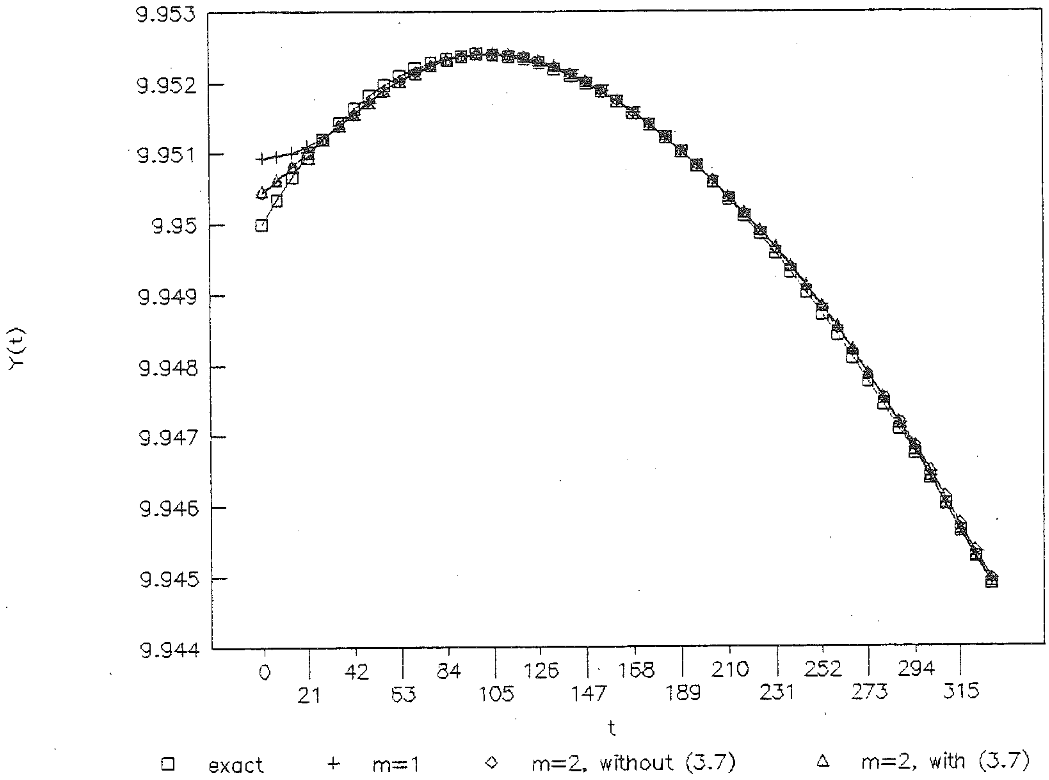


Figure 5.2

Forward Rate Curve

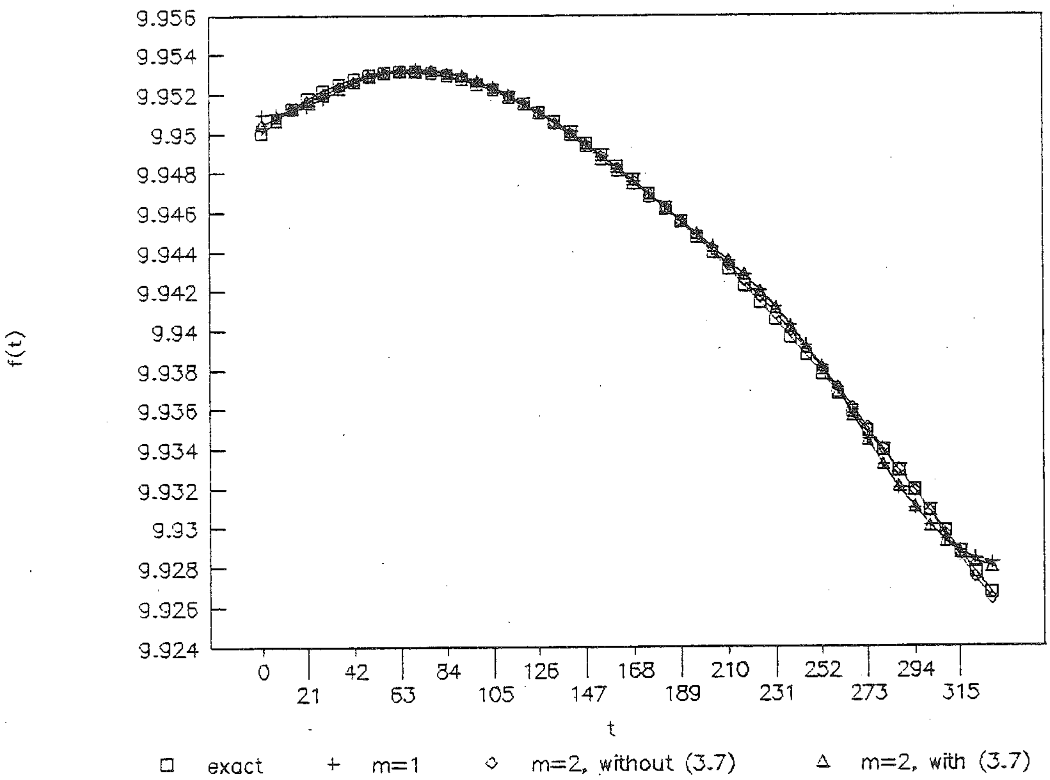


Figure 5.3

Yield Curve

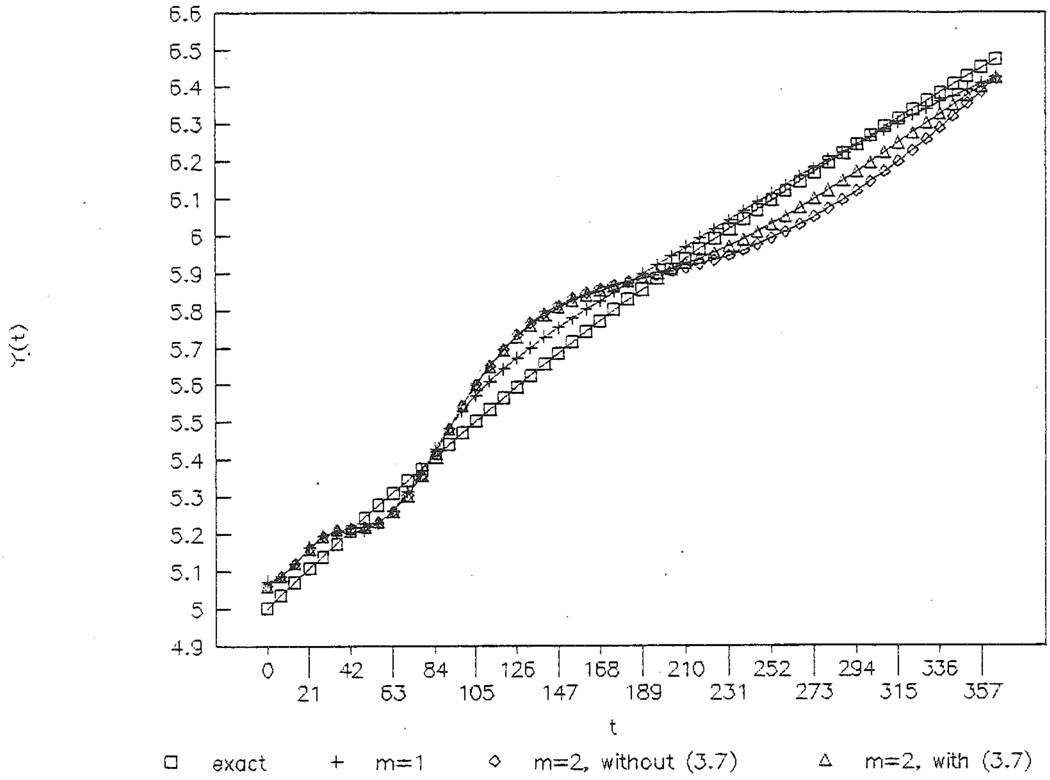


Figure 5.4

Forward Rate Curve

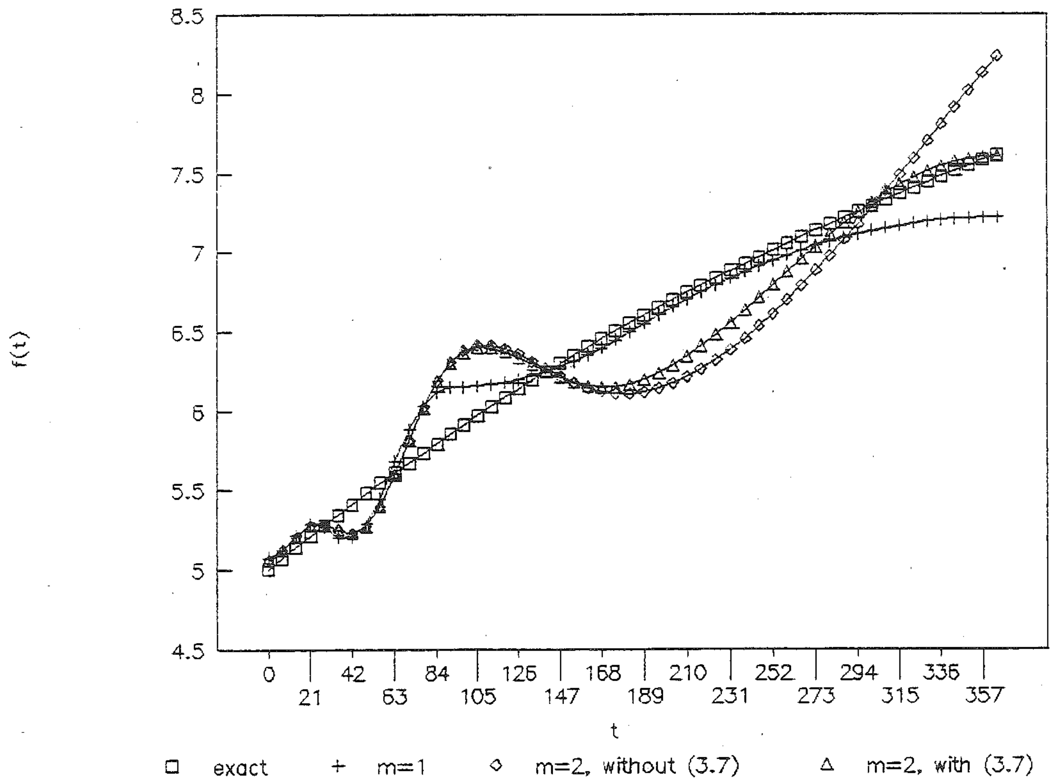


Figure 5.5

Yield Curve

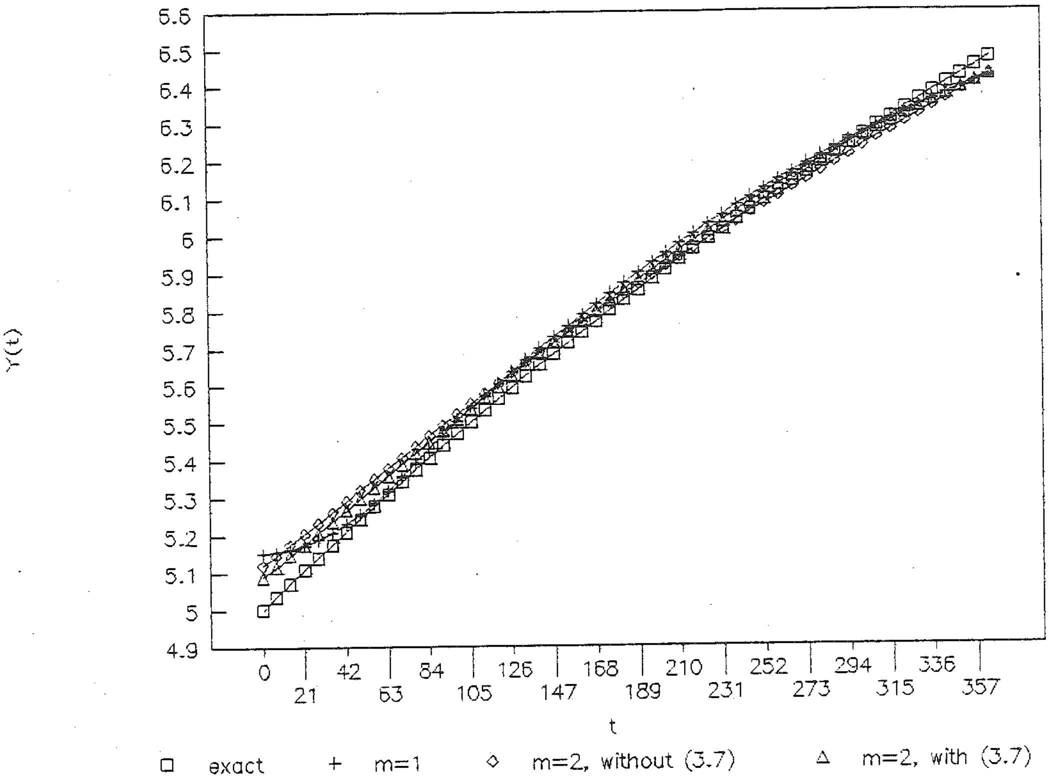


Figure 5.6

Forward Rate Curve

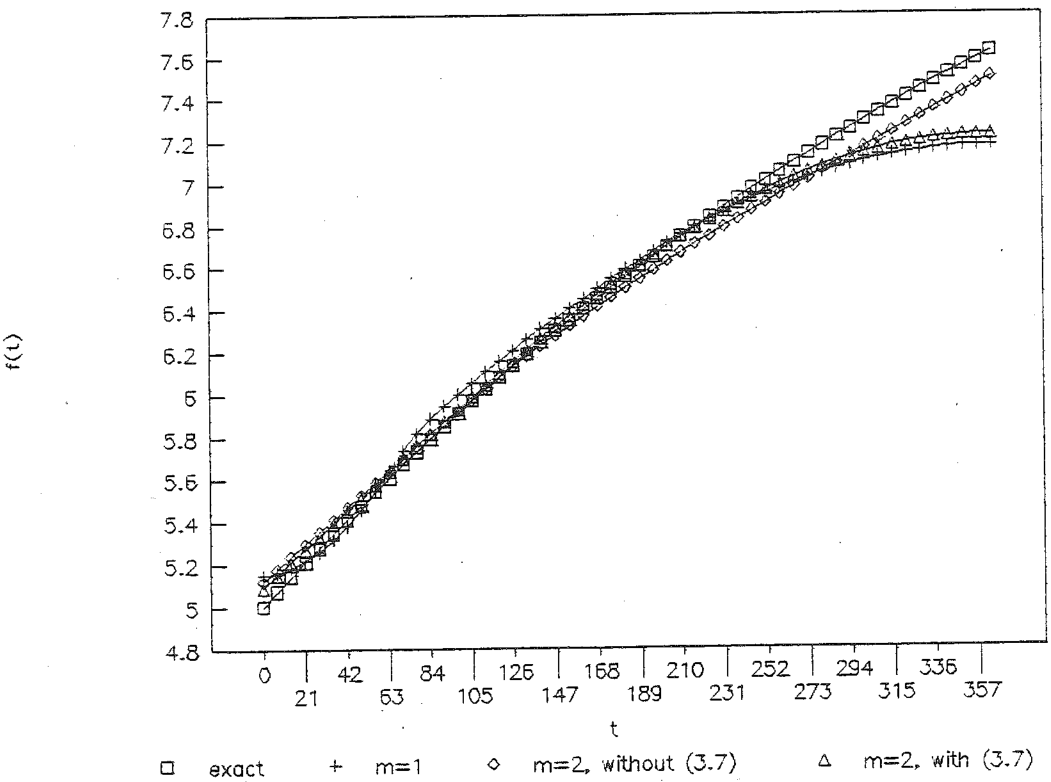


Figure 5.7

Yield Curve

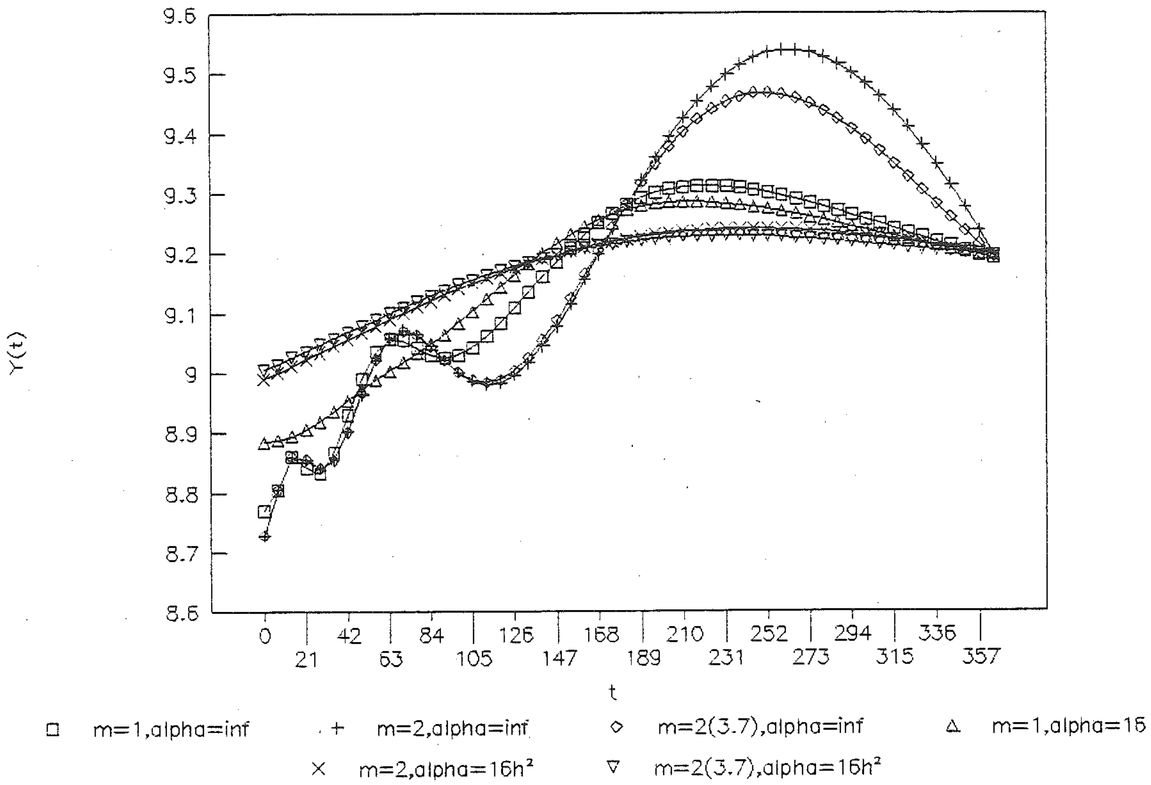
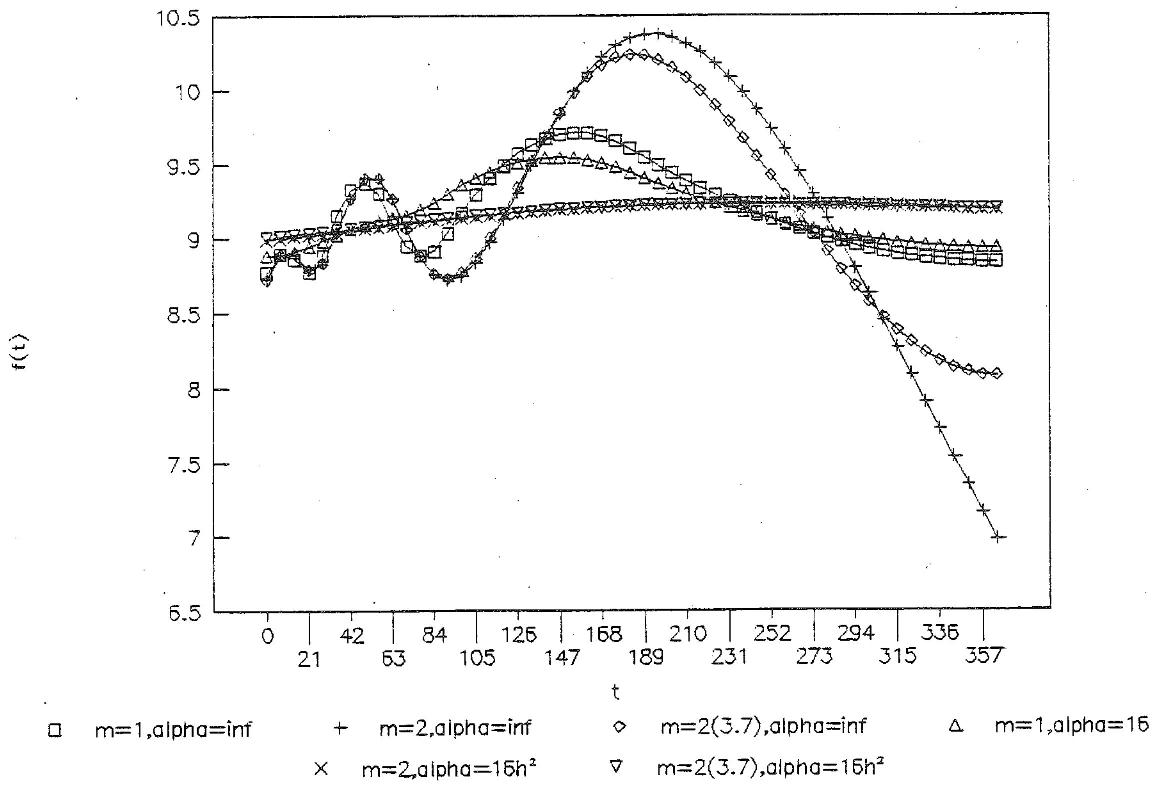


Figure 5.8

Forward Rate Curve



To conclude, we can say that the use of the higher order degree of smoothness gives better results for small maturities, because the forward rate curve no longer flattens. But when the observations are sparse (as can be in practice for long maturities) the higher order methods tend to fluctuate much between the observations, what is an unlikely behaviour. Furthermore, for long maturities the forward rate can drift off, an effect that can be compensated by imposing the additional constraint (3.7). By using method (3.8)-(3.9) smoother curves can be obtained, but which method will give the best results cannot be decided easily.

Conclusions

In this paper a method is presented using higher order derivatives or differences to estimate the forward rate curve and the yield curve. The most important results are that the discrete methods give a solution that converges uniformly towards the solution of the continuous method, when the discretization step is small.
From the few examples presented is this paper it is clear that for small maturities the method using higher order differences yields better results than the one using first order differences, because of the flattening effect. On the other hand, the higher order methods can generate curves that fluctuate more between the observed maturities. Therefore it can be usefull to create a method that combines first and higher order differences to obtain curves that do not fluctuate much between the observations and that do not flatten before the first observation.

References

Adams, Kenneth J., and Donald R. Van Deventer, Fitting Yield Curves and Forward Rate Curves with Maximum Smoothness. *Journal of Fixed Income* (June 1994), 52-62.

Cox, John C., Jonathan E. Ingersoll, Jr, and Stephen A. Ross, A Theory of the Term Structure of Interest Rates. *Econometrica*, **Vol 53, No 2** (1985), 385-407.

Delbaen, F., and S. Lorimier, Estimation of the Yield Curve and the Forward Rate Curve starting from a Finite Number of Observations. *Insurance: Mathematics and Economics*, **11** (1992), 259-269.

Devroye, Luc, *Non-Uniform Random Variate Generation.* Springer-Verlag, New York (1986).

Lorimier, Sabine, Interest Rate Term Structure Estimation with Polynomial Splines on the Forward Rate Curve. *Belgian Journal of Operations Research, Statistics and Computational Science*, **Vol 3, No 34** (1995a).

Lorimier, Sabine, *Interest Rate Term Structure Estimation Based on the Optimal Degree of Smoothness of the Forward Rate Curve*. Phd dissertation, University of Antwerp (1995b).

McCulloch, J. Huston, Measuring the Term Structure of Interest Rates. *Journal of Business*, **Vol 34, No 1** (1971), 19-31.

Shea, Gary S., Pitfalls in Smoothing Interest Rate Term Structure Data: Equilibrium Models and Spline Approximations. *Journal of Financial and Quantitative Analysis*, **Vol 19, No 3** (1984), 253-269.

Shea, Gary S., Interest Rate Term Structure Estimation with Exponential Splines: A Note, *Journal of Finance*, **Vol 40, No 1** (1985), 319-325.

Vasicek, Oldrich A., and H. Gifford Fong, Term Structure Modeling Using Exponential Splines, *Journal of Finance*, **Vol 37, No 2** (1982), 339-348.

Appendix

Lemma A.1

For all $n \geq m$ the following equality holds:

$$\sum_{l=0}^{n-m} \binom{m+l-1}{m-1} = \binom{n}{m}.$$

Proof

The statement is proved by means of induction with respect to n. For $n = m$ the relation is true. Furthermore

$$\begin{aligned} \sum_{l=0}^{n-m} \binom{m+l-1}{m-1} &= \binom{n-1}{m} + \binom{n-1}{m-1}, \\ &= \binom{n}{m}. \end{aligned}$$

QED.

Lemma A.2

For all $n \geq j$ we have that

$$\sum_{k=0}^{j} (-1)^k \binom{j}{k} \binom{n-k}{l} = \begin{cases} 0 & l < j, \\ \binom{n-j}{l-j} & l \geq j. \end{cases}$$

Proof

The proof is made by means of induction to j. For $j = 1$ the statement is easily proved. To proof the statement for j we use that

$$\sum_{k=0}^{j} (-1)^k \binom{j}{k} \binom{n-k}{l} = \sum_{k=0}^{j-1} (-1)^k \binom{j-1}{k} \binom{n-k}{l} - \sum_{k=0}^{j-1} (-1)^k \binom{j-1}{k} \binom{n-1-k}{l},$$

what can be obtained by applying (3.3) on the first combinatorial.
If $l < j-1$ the induction hypothesis states that the two sums are zero; if $l = j-1$ the two sums equal one; if $l = j$ we get

$$\sum_{k=0}^{j} (-1)^k \binom{j}{k} \binom{n-k}{l} = \binom{n-j+1}{1} - \binom{n-1-j+1}{1} = 1$$

and, if $l > j$ then

$$\sum_{k=0}^{j} (-1)^k \binom{j}{k} \binom{n-k}{l} = \binom{n-j+1}{l-j+1} - \binom{n-j}{l-j+1} = \binom{n-j}{l-j},$$

what proves the lemma.

QED.

Lemma A.3

For all m, $j = 1, \ldots, m-1$, $s = 0, \ldots, j$ we have that

$$\sum_{k=0}^{s}(-1)^k \binom{j}{k}\binom{m-k+s-1}{m-1} = \binom{m-j+s-1}{s}.$$

Proof

We use induction with respect to j. For $j = 1$ the statement is valid for $s = 0, 1$.
To proof the relation for j, we first of all assume that $s \neq 0, j$. Then because of (3.3) the left-hand side equals

$$\begin{aligned}\sum_{k=0}^{s}(-1)^k \binom{j-1}{k}\binom{m-k+s-1}{m-1} &- \sum_{k=0}^{s-1}(-1)^k \binom{j-1}{k}\binom{m-k+s-1-1}{m-1} \\ &= \binom{m-j+s}{s} - \binom{m-j+s-1}{s-1}, \\ &= \binom{m-j+s-1}{s},\end{aligned}$$

where the second equality follows from the induction hypothesis.
If $s = 0$ then both sides of the equality are one and if $s = j$ then the equality follows from lemma A.2.

QED.

Lemma A.4

For all n, $m = 0, \ldots, n-1$ and $j = 0, \ldots, m-1$ the following equality holds:

$$\sum_{k=0}^{n-m-1}\binom{k+m-j-1}{k}\binom{m+k}{j} = \sum_{k=0}^{j}\binom{n-k-1}{m-k}\binom{m-k-1}{j-k}.$$

Proof

The statement is proved by using induction with respect to j. For $j = 0$ the equality holds because of lemma A.1. Furthermore we have that the left-hand side equals

$$\begin{aligned}\sum_{k=0}^{n-m-1}\binom{k+m-j-1}{k}\binom{m+k-1}{j} &+ \sum_{k=0}^{n-m-1}\binom{k+m-j-1}{k}\binom{m+k-1}{j-1} \\ &= \frac{(m-1)!}{j!\,(m-j-1)!}\sum_{k=0}^{n-1-m}\binom{m+k-1}{k} + \sum_{k=0}^{j-1}\binom{n-k-2}{m-k-1}\binom{m-k-2}{j-k-1}, \\ &= \binom{m-1}{j}\binom{n-1}{m} + \sum_{k=1}^{j}\binom{n-k-1}{m-k}\binom{m-k-1}{j-k},\end{aligned}$$

where the second equality follows from the induction hypothesis and the last one from lemma A.1. This proves the lemma.

QED.

Lemma A.5

For all $n \geq m \geq 1$, $j = 0, \ldots, m-1$,

$$\binom{n}{m}\binom{m-1}{j} - \sum_{k=0}^{j}\binom{n-k-1}{m-k-1}\binom{m-k-2}{j-k} = \sum_{k=0}^{j}\binom{n-k-1}{m-k}\binom{m-k-1}{j-k}.$$

Proof

We use induction with respect to m. For $m = 1$ the left- and right-hand side are both equal to $n-1$. To prove the lemma for m, we first assume that $j \neq 0$. Then the left-hand side equals

$$\begin{aligned}
&\binom{n-1}{m}\binom{m-1}{j} + \binom{n-1}{m-1}\binom{m-1}{j} - \sum_{k=0}^{j}\binom{n-k-1}{m-k-1}\binom{m-k-2}{j-k} \\
&= \binom{n-1}{m}\binom{m-1}{j} + \binom{n-1}{m-1}\binom{m-2}{j} + \binom{n-1}{m-1}\binom{m-2}{j-1} \\
&\quad - \sum_{k=0}^{j}\binom{n-k-1}{m-k-1}\binom{m-k-2}{j-k}, \\
&= \binom{n-1}{m}\binom{m-1}{j} + \binom{n-1}{m-1}\binom{(m-1)-1}{j-1} \\
&\quad - \sum_{k=0}^{j-1}\binom{(n-1)-k-1}{(m-1)-k-1}\binom{(m-1)-k-2}{(j-1)-k}, \\
&= \sum_{k=0}^{j}\binom{n-k-1}{m-k}\binom{m-k-1}{j-k},
\end{aligned}$$

where the last equality follows from the induction hypothesis.
If $j = 0$ then the left-hand side equals

$$\binom{n}{m} - \binom{n-1}{m-1} = \binom{n-1}{m},$$

what equals the right-hand side.

QED.

Lemma A.6

For all m, $n \geq m+1$, $j = 0, \ldots, m-1$ the following relation is valid:

$$1 + (-1)^{m-j+1}\sum_{k=0}^{j}\binom{n-k-1}{m-k}\binom{m-k-1}{j-k} = \sum_{k=j}^{m-1}(-1)^{k-j}\binom{n}{k+1}\binom{k}{j}.$$

Proof

The statement is proved by means of induction with respect to m. For $m = 1$ it is easy to show that both sides of the equality equal n. To prove the relation for m we rewrite the right-hand side by means of the induction hypothesis and by using lemma A.5, what gives, for $j \neq m-1$,

$$\begin{aligned}
&\sum_{k=j}^{m-2}(-1)^{k-j}\binom{n}{k+1}\binom{k}{j}+(-1)^{m-j-1}\binom{n}{m}\binom{m-1}{j} \\
&\quad = 1+(-1)^{m-j}\sum_{k=0}^{j}\binom{n-k-1}{m-k-1}\binom{m-k-2}{j-k}+(-1)^{m-j-1}\binom{n}{m}\binom{m-1}{j}, \\
&\quad = 1+(-1)^{m-j}\left[\binom{n}{m}\binom{m-1}{j}-\sum_{k=0}^{j}\binom{n-k-1}{m-k}\binom{m-k-1}{j-k}\right] \\
&\qquad +(-1)^{m-j-1}\binom{n}{m}\binom{m-1}{j}.
\end{aligned}$$

This is the left-hand side of the relation to be proved. For $j = m-1$ we have, because of lemma A.1, that the right-hand side equals

$$\begin{aligned}
1+\sum_{l=1}^{m}\binom{n-m+l-1}{l} &= \sum_{l=0}^{m}\binom{n-m+l-1}{l}, \\
&= \binom{n}{n-m}, \\
&= \binom{n}{m},
\end{aligned}$$

what equals the right-hand side.

QED.

Lemma A.7

For all n, $l = 0,\ldots,n-1$, $j = 0,\ldots,l$ we have that

$$\sum_{p=j}^{l}\binom{p}{j}\binom{n-p-1}{l-p} = \binom{n}{l-j}.$$

Proof

It is proved by using induction with respect to l. For $l = 0$ both sides are equal to 1. To prove the statement for l, we first assume that $j \neq 0, l$. Then application of (3.3) and the induction hypothesis transforms the left-hand side into

$$\begin{aligned}
&\sum_{p=j}^{l-1}\binom{p}{j}\binom{n-1-p-1}{l-1-p}+\sum_{p=j-1}^{l-1}\binom{p}{j-1}\binom{n-1-p-1}{l-1-p} \\
&\quad=\binom{n-1}{l-1-j}+\binom{n-1}{l-j}, \\
&\quad=\binom{n}{l-j}.
\end{aligned}$$

If $j=0$ then the left-hand side equals

$$\begin{aligned}
&\sum_{p=0}^{l}\binom{n-p-1}{l-p} \\
&\quad=\sum_{k=0}^{l}\binom{n-l+k-1}{k}, \\
&\quad=\binom{n}{n-l}, \\
&\quad=\binom{n}{l-0},
\end{aligned}$$

where the second equality follows from lemma A.1. If $j=l$ then the left- and the right-hand side are equal to 1.

QED.