2007

Pade Approximants And One Of Its Applications

Tame-kouontcho Fowe
University of Central Florida

PADE APPROXIMANTS AND ONE OF ITS APPLICATIONS

by

TAME FOWE
B.S. University of Douala, 2002

A thesis submitted in partial fulfillment of the requirement
for the degree of Master of Science
in the Department of Mathematics
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2007

Major Professor: Ram N Mohapatra

ABSTRACT

This thesis is concerned with a brief summary of the theory of Padé approximants and one of its applications to Finance. Proofs of most of the theorems are omitted and many developments could not be mentioned due to the vastness of the field of Padé approximations. We provide reference to research papers and books that contain exhaustive treatment of the subject. This thesis is mainly divided into two parts. In the first part we derive a general expression of the Padé approximants and some of the results that will be related to the work on the second part of the thesis. The Aitken's method for quick convergence of series is highlighted as Padé$[L/1]$. We explore the criteria for convergence of a series approximated by Padé approximants and obtain its relationship to numerical analysis with the help of the Crank-Nicholson method. The second part shows how Padé approximants can be a smooth method to model the term structure of interest rates using stochastic processes and the no arbitrage argument. Padé approximants have been considered by physicists to be appropriate for approximating large classes of functions. This fact is used here to compare Padé approximants with very low indices and two parameters to interest rates variations provided by the Federal Reserve System in the United States.

Dedicated to my parents,
siblings and brother-in-law
for all the help
I received from them

ACKNOWLEDGMENTS

To Professor Ram Mohapatra

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

After the proof of the famous Weirstrass theorem on the approximation of continuous functions defined over a compact set by polynomials, considerable research on existence, uniqueness of best approximation by polynomials and goodness of their approximation were considered. Several operators were also defined to approximate a class of functions by such operators. The most notable of these results being Korovkin's results on approximation by positive linear operators (see [13]). Results on saturation order and saturation class of functions were also obtained. Most of this type of research was motivated by polynomials, because polynomials have the pleasant property of depending linearly on their coefficients. After that mathematicians were interested to study approximation out of a useful family of functions, the members of which do not depend linearly on their parameters. These are the rational functions, that is, functions which are ratio of two polynomials. Existence, characteristics and uniqueness of Padé approximants have been studied in detail (see [13], [29]). Our aim is to consider the Taylor series analogue of rational functions called Padé approximations (see [13], page 173- 179, [1]).

This chapter is divided into two main sections. A short historical mention on Padé approximants is introduced at the outset. Then the theory of Padé approximants is outlined starting with its basic definition. Then we mention its relationship to numerical analysis including Crank-Nicholson Method.

1.1 History

Padé approximations originated from the study of continued fractions dating back to Euclide around 300 BC. Lagrange in 1776 thought about the idea and Henri Eugène Padé (1863-1953), French mathematician presented a systematic study.

In fact, it was Charles Hermite who gave his student Henri Eugène Padé the approximant to study in the 1890's. As a result of that, in 1892, Henri Eugène Padé published an article concerning the approximate representation of a function by rational fractions in the Scientific Transactions of the Ecole Normale Superieure in Paris. He retired in 1834 and in 1908 became the youngest rector appointed in France. After a long and deserving career in academics, he died at the age of 89.

Three-quarters of a century later, the advent of arithmetical computers led scientists to consider various methods of representing functions, especially rapidly converging functions.

1.2 Definition of Padé approximants

Given a power series

$$f(z) = \sum_{i=0}^{\infty} c_i z^i \tag{1.1}$$

where c_i = 0, 1, 2… the Padé approximation to f(z) is a rational function

$$[L/M] = \frac{\sum_{k=0}^{L} a_k z^k}{\sum_{k=0}^{M} b_k z^k} \tag{1.2}$$

which has a Maclaurin series expansion that agrees with (1.1) as many terms as possible. There are $L+1$ independent coefficients in the numerator and M independent coefficients in the denominator, making $L+M+1$ unknown coefficients in all since we took $b_0=1$.We use the notation:

$$\sum_{i=0}^{\infty} c_i z^i = \frac{\sum_{k=0}^{L} a_k z^k}{\sum_{k=0}^{M} b_k z^k} + O(z^{L+M+1}) \tag{1.3}$$

By cross-multiplying, we find that

$$\left(\sum_{k=0}^{M} b_k z^k\right)\left(\sum_{i=0}^{\infty} c_i z^i\right) = \sum_{k=0}^{L} a_k z^k + O(z^{L+M+1}) \tag{1.4}$$

Equating the coefficients of $z^{L+k}, k=1,\dots,M$, we find

$$\begin{cases} b_M c_{L-M+1} + b_{M-1} c_{L-M+2} + \dots + b_0 c_{L+1} = 0; \\ b_M c_{L-M+2} + b_{M-1} c_{L-M+3} + \dots + b_0 c_{L+2} = 0; \\ \quad \vdots \\ \quad b_M c_L \quad + \ b_{M-1} c_{L+1} \quad + \dots + b_0 c_{L+M} = 0 \end{cases} \tag{1.5}$$

If j < 0, we define $c_j = 0$ for consistency. Since $b_0 = 1$, equations (1.5) become a set of M linear equations for the M unknown coefficients of the denominator, which in the matrix notation are given by:

$$\begin{bmatrix} c_{L-M+1} & c_{L-M+2} & c_{L-M+3} & \cdots & c_L \\ c_{L-M+2} & c_{L-M+3} & c_{L-M+4} & \cdots & c_{L+1} \\ c_{L-M+3} & c_{L-M+4} & c_{L-M+5} & \cdots & c_{L+2} \\ \vdots & \vdots & \vdots & & \vdots \\ c_L & c_{L+1} & c_{L+2} & \cdots & c_{L+M-1} \end{bmatrix} \begin{bmatrix} b_M \\ b_{M-1} \\ b_{M-2} \\ \vdots \\ b_1 \end{bmatrix} = - \begin{bmatrix} c_{L+1} \\ c_{L+2} \\ c_{L+3} \\ \vdots \\ c_{L+M} \end{bmatrix}. \tag{1.6}$$

From (1.6) the b_k's, $k=1,2,\dots M$, may be found if the matrix is nonsingular. The numerator coefficients a_i follow from (1.4) by equating the coefficients of $1, z, z^2, \dots, z^L$. We get,

$$\begin{cases} a_0 = c_0, \\ a_1 = c_1 + b_1 c_0, \\ a_2 = c_2 + b_1 c_1 + b_2 c_0, \\ \vdots \\ a_L = c_L + \sum_{i=0}^{\min(L,M)} b_i c_{L-i} \end{cases} \tag{1.7}$$

.

Thus (1.6) and (1.7) normally determine the Padé numerator and denominator if they exist and then are called Padé equations.

The power series converges for all z in the disc $|z| < R$ and diverges for z such that $|z| > R$, given that its circle of convergence is $|z| = R$. On the circle of convergence $|z| = R$, the power series may or may not converge (See Titchmarsh, Theory of Functions). If the power series converges to the same function for $|z| < R$ with $0 < R < \infty$, then a sequence of Padé approximants may converge for $z \in \mathrm{D}$ where D is a domain larger than $|z| < R$. This indeed is relevant thanks to analytic continuation. It assures the existence of an analytic function which coincides with the power series on the disc $|z| < R$ and expands on a bigger set including the disc.

We use Cramer's rule to calculate the unknowns $b_0, b_1, ..., b_M$ from (1.6).and replace them in the denominator $\sum_{k=0}^{M} b_k z^k$ of (1.2). The final expression of $\sum_{k=0}^{M} b_k z^k$ can written as a determinant we call $Q^{[L/M]}(z)$. The expression of $Q^{[L/M]}(z)$ is given in the following page.

$$Q^{[L/M]}(z) = \begin{vmatrix} c_{L-M+1} & c_{L-M+2} & \cdots & c_L & c_{L+1} \\ c_{L-M+2} & c_{L-M+3} & \cdots & c_{L+1} & c_{L+2} \\ \vdots & \vdots & & \vdots & \vdots \\ c_{L-1} & c_L & \cdots & c_{L+M-2} & c_{L+M-1} \\ c_L & c_{L+1} & \cdots & c_{L+M-1} & c_{L+M} \\ z^M & z^{M-1} & \cdots & z & 1 \end{vmatrix} \tag{1.8}$$

By multiplying the above matrix by $\sum_{i=0}^{\infty} c_i z^i$, this results to

$$Q^{[L/M]}(z)\sum_{i=0}^{\infty} c_i z^i = \begin{vmatrix} c_{L-M+1} & c_{L-M+2} & \cdots & c_L & c_{L+1} \\ c_{L-M+2} & c_{L-M+3} & \cdots & c_{L+1} & c_{L+2} \\ \vdots & \vdots & & \vdots & \vdots \\ c_{L-1} & c_L & \cdots & c_{L+M-2} & c_{L+M-1} \\ c_L & c_{L+1} & \cdots & c_{L+M-1} & c_{L+M} \\ \sum_{i=0}^{\infty} c_i z^{M+i} & \sum_{i=0}^{\infty} c_i z^{M+i-1} & \cdots & \sum_{i=0}^{\infty} c_i z^{i+1} & \sum_{i=0}^{\infty} c_i z^{i} \end{vmatrix}$$

By subtracting z^{L+1} times the first row from the last, z^{L+2} times the second row from the last, etc., up to z^{L+M} times the row right above the last one, we get the last row of the determinant above as given in the last row of the determinant below:

$$P^{[L/M]}(z) = \begin{vmatrix} c_{L-M+1} & c_{L-M+2} & \cdots & c_L & c_{L+1} \\ c_{L-M+2} & c_{L-M+3} & \cdots & c_{L+1} & c_{L+2} \\ \vdots & \vdots & & \vdots & \vdots \\ c_{L-1} & c_L & \cdots & c_{L+M-2} & c_{L+M-1} \\ c_L & c_{L+1} & \cdots & c_{L+M-1} & c_{L+M} \\ \sum_{i=0}^{L-M} c_i z^{M+i} & \sum_{i=0}^{L-M+1} c_i z^{M+i-1} & \cdots & \sum_{i=0}^{L-1} c_i z^{i+1} & \sum_{i=0}^{L} c_i z^{i} \end{vmatrix} \tag{1.9}$$

From the theory above we get the theorems.

Theorem 1.2.1 With the definitions given in (1.8) and (1.9),

$$Q^{[L/M]}(z)\ \sum_{i=0}^{\infty} c_i z^i - P^{[L/M]}(z) = O(z^{L+M+1}) \tag{1.10}$$

The proof of this theorem is available in [Encyclopedia of Mathematics, Padé approximant, Chapter 1, page 6], see ref. [1].

In 1846, Jacobi proved the following theorem

Theorem 1.2.2 [Jacobi, 1846, see ref. [1]] With the definitions (1.8) and (1.9), the [L/M] Padé approximant of $f(z)=\sum_{i=0}^{\infty} c_i z^i$ is given by $[L/M]=\dfrac{P^{[L/M]}(z)}{Q^{[L/M]}(z)}$ given that $Q^{[L/M]}(0)\neq 0$.

If we display the approximants on a table we get what is called the Padé table.

Table 1: Padé table

L/M	0	1	2	…
0	[0/0]	[1/0]	[2/0]	…
1	[0/1]	[1/1]	[2/1]	…
2	[0/2]	[1/2]	[2/2]	…
⋮	⋮	⋮	⋮	⋱

1.3 Convergence Theory

For row sequences on the Padé table, de Montessus's theorem proves convergence for functions meromorphic in a disk. The theorem establishes that the region of bad approximation becomes arbitrarily small. For [L/0] on the first row, they converge inside a circle of convergence whose radius may be 0, finite or infinite.

The second row of [L/1] approximants has convergence governed by the following Beardon theorem. This theorem is stated without proof. For proof see [Encyclopedia of Mathematics, Padé approximant, page 277, ref. [1]].

Theorem 1.3.1 Let $f(z)$ be analytic in $|z| \le R$. Then an infinite subsequence of [L/1] Padé approximants converges to $f(z)$ uniformly in $|z| \le R$.

Theorem 1.3.2 [de Montessus, 1902, see ref. [1]]. Let $f(z)$ be a function which is meromorphic in the disk $|z| \le R.$, with m poles at distinct points $z_1, z_2, ..., z_m$ with

$$0 < |z_1| \le |z_2| \le ... \le |z_m| < R .$$

Let the pole of $f(z)$ at z_k have multiplicity μ_k, and let the total multiplicity be

$\sum_{k=1}^{m} \mu_k = M$ precisely. Then

$$f(z) = \lim_{L \to \infty} [L/M]$$

uniformly on any compact subset of $D_m = \{z, |z| \le R, z \ne z_k, k = 1,2,...,m\}$.

Bear in mind that the understanding of this theorem requires the Cauchy-Binet formula, (see ref [1]), which is useful to calculate the determinant of two matrices. The Cauchy-Binet formula is stated in [Encyclopedia of Mathematics, Padé approximant, chapter 6, page 281].

The proofs of the above theorems and related theory are given in [Encyclopedia of Mathematics, Padé approximant, pages 281-290].

1.4 Properties of the Padé approximant

We now formulate some algebraic properties of Padé approximants related to power series without implying any convergence properties.

Property 1.4.1 (Duality) Let $g(z) = \{f(z)\}^{-1}$ and $f(0) \ne 0$,

then $[L/M]_g(z) = \{[M/L]_f(z)\}^{-1}$. It is assumed that both the Padé approximants exist.

Property 1.4.2 (Homographic invariance under transformations). Let $f(z)=\sum_{i=0}^{\infty}c_i z^i$. We define a linear fractional transformation (which preserves the origin) with argument $w=\frac{az}{1+bz}$, and a function $g(w)=f(z)$. Then

$$[M/M]_g(w)=[M/M]_f(z)$$

provided both the Padé approximants exist.

Property 1.4.3 (Homographic invariance of value transformations)

Let us have $f(z)=\sum_{i=0}^{\infty}c_i z^i$ and we define $g(z)=\frac{a+bf(z)}{c+df(z)}$.

if $c+df(0)\neq 0$, then

$$[M/M]_g(z)=\frac{a+b[M/M]_f(z)}{c+d[M/M]_f(z)}$$

given that $[M/M]_f(z)$ exists.

Property 1.4.4 (Truncation theorem) Let $f(z)=\sum_{i=0}^{\infty}c_i z^i$ and

$$g(z)=\sum_{i=0}^{\infty}g_i z^i=\left\{f(z)-\sum_{i=0}^{k-1}c_i z^i\right\}z^{-k}.$$

Then $\left[L-k/M\right]_g(z)=\left\{[L/M]_f-\sum_{i=0}^{k-1}c_i z^i\right\}z^{-k}$ for $k\geq 1$, $L-k\geq M-1$, given that the Padé approximants exist.

Property 1.4.5 (Unitarity) [Gammel and McDonald] Let $f(z)=\sum_{i=0}^{\infty}c_i z^i$ be unitary, meaning that $f(z)f^*(z)=1$.

If $[M/M]=[M/M]_f(z)$ is a diagonal Padé approximant of $f\left(z\right)$, then

$$[M/M][M/M]^*=1,$$

where the asterisk indicates the complex conjugate.

The proofs of each property mentioned above are quite straightforward and are given in Encyclopedia of Mathematics, Chapter 1, pages 32-37 in ref. [1].

CHAPTER 2 : APPLICATION OF PADE APPROXIMATION TO NUMERICAL ANALYSIS

Some numerical methods have been developed to find the Padé approximants among them are, ε - algorithm, $\eta -$algorithm, the Q.D. Algorithm and the Aitken's Δ^2 method, just to list a few. We shall mention these briefly below.

2.1 Aitken's Δ^2 method as [L/1] Padé approximants

We will describe only the Aitken's Δ^2 method as [L/1] Padé approximants which was implemented in 1926. This method allows an acceleration of a sequence for convergence. Given a sequence of real or complex numbers,

$$\mathbb{S} = \{ S_n , \text{n} = 0, 1, 2 \ldots\},$$

such that $S_n \to S$ as $\text{n} \to \infty$, the problem is to find a new sequence which converges faster to S.

We define $\Delta S_n = S_{n+1} - S_n, \quad \Delta^2 S_n = \Delta(\Delta S_n) = S_{n+2} - 2S_{n+1} - S_n,$

and the sequence $\{T_n, n = 0,1,2,...\}$ where $T_n = S_n - (\Delta S_n)^2 / \Delta^2 S_n$ (2.1)

converges to S. $\lim_{n\to\infty} \frac{T_n - S}{S_n - S} = 0$ which translates to $\{T_n\}_{n\geq 0}$ converges faster to S than $\mathbb{S}$. The process is valid for certain type of convergent sequences like geometric convergent sequences. The connection with Padé approximation resides in defining the series of partial sums S_n which converges to S.

For this purpose, let $\begin{array}{l} c_{n+1} = \Delta S_n = S_{n+1} - S_n, n = 0,1,2,... \\ c_0 = S_0. \end{array}$

Then we form the power series $f(z) = \sum_{i=0}^{\infty} c_i z^i$.

From this, we evaluate $f(1) = S$, by determining $[L/1]_f(1), L = 0,1,2,....$ This is found by using the second row of the Padé table (ref. table 1.1, page 5) then we find the limit as $L \to \infty$.

From (1.8) and (1.10), we get

$$[L/1]_f(1) = P^{[L/1]}/Q^{[L/1]} = \begin{vmatrix} c_L & c_{L+1} \\ \sum_{i=0}^{L-1} c_i & \sum_{i=0}^{L} c_i \end{vmatrix} \div \begin{vmatrix} c_L & c_{L+1} \\ 1 & 1 \end{vmatrix}$$

$$= \frac{(S_L - S_{L-1})S_L - (S_{L+1} - S_L)S_{L-1}}{(S_L - S_{L-1}) - (S_{L+1} - S_L)}$$

$$= \frac{S_{L-1}(S_{L+!} - 2S_L + S_{L-1}) - (S_L - S_{L-1})^2}{S_{L+1} - 2S_L + S_{L-1}}$$

$$= S_{L-1} - \frac{(\Delta S_{L-1})^2}{\Delta^2 S_{L-1}}$$

Since $[L/1]$ has the same expression as (2.1) we conclude that Aitken's method is equivalent to $[L/1]$ Padé approximants.

Theorem 1.2.1 [Henrici, 1964]. Let $S_{n+1} = f(S_n)$ define a convergence real sequence with limit S, let f(x) be twice differentiable at S, and let $f'(S) \neq 1$. Then, with the definition T_n given in (2.1), we have

$$T_n - S = O((S_n - S)^2).$$

2.1.1 Example (using Aitken's method)

Consider the series, $\sum_{i=0}^{\infty} c_i = \frac{1}{2} + \frac{1}{3} - \frac{5}{6} + \frac{1}{4} + \frac{1}{5} - \frac{9}{5} + ...$, where the c_i's are expressed as

$$c_{3m-3} = \frac{1}{2m}, c_{3m-2} = \frac{1}{2m+1}, c_{3m-1} = -\frac{4m+1}{4m^2+2m} \quad \text{for} \quad m = 1,2,3,...$$

We define $S_n = \sum_{i=0}^{n} c_i$ and T_n given by (2.1). Then the following results hold:

(i) $S_{3m-1} = 0$ for $m = 1,2,3,...,$

(ii) $S_n \to 0$ as $n \to \infty$,

(iii) $T_{3m} \to 0, T_{3m-1} \to 1,$ and $T_{3m-2} \to 0$ as $m \to \infty$.

We refer to [Encyclopedia of Mathematics, Padé approximant, Chapter 3, page 67] for more details on numerical methods and Padé approximants.

Now we need to focus on a primordial point when working with sequences. Our attention is on convergence of sequences of Padé approximants to complex functions.

2.2 Relation to numerical analysis: Crank-Nicholson and related methods for diffusion equations

We are interested in finding a continuous function $u(x,t)$ which satisfies the diffusion equation

$$\frac{1}{k}\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0, \qquad t > 0, \tag{5.1a}$$

subject to the boundary condition :

$$u(x,0) = f(x), \quad -\infty < x < \infty \,. \tag{5.1b}$$

Since the solution of this initial-pure problem is known to be

$$u(x,t) = \frac{1}{\sqrt{4\pi kt}} \int_{-\infty}^{\infty} \exp\{\frac{-(\xi - x)^2}{4kt}\} f(\xi) d\xi$$

it is then reduced to numerical integration for all times $t > 0$.

Sometimes, mathematicians encounter equations that are slightly different from (5.1a) with most of the time different boundary values. In those cases, when the Green's-function method is inconclusive, other methods can be used to derive the solution.

To that extent, let us study the partial differential equation for the neutron density $u(x,t)$:

$$\frac{1}{k}\frac{\partial u}{\partial t} - \nabla^2 u = s(x,t)u. \tag{5.2}$$

Considering heat transfer along a finite bar, with the ends maintained at given temperatures and with various heat sources; the temperatures $u(x,t)$ follows the equation

$$\frac{1}{k}\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = s(x,t), \qquad 0 < x < 1, \tag{5.3a}$$

with the boundary conditions at the ends of the bar given by

$$\begin{aligned} u(0,t) &= T_0(t), \\ u(L,1) &= T_L(t), \end{aligned} \tag{5.3b}$$

and initial temperature $u(x,0) = f(x)$.

To obtain numerical solutions of equations such as (5.2) and (5.3) and other types, we substitute derivatives with differences and initially discretize the x-variable in (5.3).

We work with N interior points in $0 < x < L$, leading to the mesh $\Delta x = L/(N+1)$ and the points

$$x_i = i\Delta x, \qquad i = 1, 2, ..., N.$$

We define the approximation scheme (method of lines) by

$$u(x_i, t) \to U_j(t)$$

$$\left.\frac{\partial^2 u}{\partial x^2}\right|_{x=x_i} \to \frac{U_{i+1}(t) - 2U_i(t) + U_{i-1}(t)}{(\Delta x)^2}$$

We solve the 'space-discretized' equations

$$\frac{1}{k}\frac{\partial U_i}{\partial t}-\frac{U_{i+1}(t)-2U_i(t)+U_{i-1}(t)}{(\Delta x)^2}=s(x_i,t) \tag{5.4}$$

for the functions $U_1(t),U_2(t),...,U_N(t)$ given that $U_0(t)=T_0(t)$ and $U_{N+1}(t)=T_L(t)$.As a matter of fact, the system is convergent meaning that $\Delta x\to 0$ as $N\to\infty$.

Equation (5.4) can be rewritten in the matrix form:

$$\frac{\partial U_i}{\partial t}=\sum_{j=1}^{N}A_{ij}U_j(t)+S_i(t), \qquad i=1,2,...,N\,, \tag{5.5}$$

where $S_i(t)=kS(x_i,t)+k\{\delta_{i1}U_0(t)+\delta_{iN}U_{N+1}(t)\}/(\Delta x)^2$.

Equation (5.5) characterizes diffusion equations after space discretization and the method of solution is obtained by time discretization. We use a sequence of time points $t_0=0,t_1,t_2,t_3,...$ and let $\Delta t_k=t_{k+1}-t_k$. The approximation boils down to

$$U_i(t_k)\to U_i^{(k)},k=1,2,...,$$
$$\left.\frac{\partial U_i}{\partial t}\right|_{t=t_k}\to\frac{U_i^{(k+1)}-U_i^{(k)}}{\Delta t_k}.$$

Then (5.5) is replaced by the equation

$$U_i^{(k+1)}=U_i^{(k)}=\Delta t_k\{\sum_{j=1}^{N}A_{ij}U_j^{(k)}+S_i(t_k)\}, \tag{5.6}$$

By setting all boundary and source terms equal to zero, (5.6), is temporarily replaced by

$$U^{(k+1)}=(I+\Delta tA)U^{(k)}, \tag{5.7}$$

where $U^{(k)}=(U_1^{(k)},U_2^{(k)},...,U_N^{(k)})$ is a vector of values at time t_k.

Under the same consideration (no source terms), (5.5) becomes

$$\frac{\partial U_i}{\partial t}=\sum_{j=1}^{N}A_{ij}U_j(t), \tag{5.8}$$

where A_{ij} is a known constant matrix with negative eigenvalues. Equation (5.8) has the solution

$$U^{(k+1)} = \exp(A\Delta t)U^{(k)}, \tag{5.9}$$

Our aim is to calculate the matrix $\exp(A\Delta t)$.

We use $p(z)/q(z)$ to denote the Padé approximant of type $[L/M]$ for $\exp z$, then

$$\exp(A\Delta t) = p(A\Delta t)[q(A\Delta t)]^{-1} + O(\Delta t^{L+M+1}). \tag{5.10}$$

Then (5.9) is approximated by

$$q(A\Delta t)u^{(k+1)} = p(A\Delta t)u^{(k)} \tag{5.11}$$

with local truncation error (based on taking $U^{(k)} = u^{(k)}$)

$$\left|U^{(k+1)} - u^{(k+1)}\right| = O(\Delta t^{L+M+1}).$$

As an example, using the $[1/1]$ Padé approximant, we have from (5.11)

$$(1 - \frac{1}{2}A\Delta t)u^{(k+1)} = (1 + \frac{1}{2}A\Delta t)u^{(k)} \tag{5.14}$$

which is expressed explicitly as

$$-\frac{\mu}{2}u_{l-1}^{(k+1)} + (1+\mu)u_l^{(k+1)} - \frac{\mu}{2}u_{l+1}^{(k+1)} = \frac{\mu}{2}u_{l-1}^{(k)} + (1-\mu)u_l^{(k)} + \frac{\mu}{2}u_{l+1}^{(k)} \tag{5.16}$$

Expression (5.16) has error of $O(\Delta t^3)$.

Equation (5.16) is the familiar Crank-Nicholson method.

For more in-depth theory on this topic, we refer to [Encyclopedia Mathematics, Padé approximants, 2nd Edition, Chapter 10, page 646, see ref. [1]].

CHAPTER 3: APPLICATION OF PADE APPROXIMANTS TO FINANCE

Over the past twenty years several research projects have helped in obtaining new procedures and characterization techniques for studying dynamic relation structures associated with chronological series (see [2 through 12] and [14 through 27]. In the context of time series analysis, several authors have considered rational theory of series in economic modeling and have proposed the use of new techniques. (See [7, 8, and 14]). The main objective of this chapter is to show how some of these techniques can provide appropriate answers to problems arising in he study of time series Economics and interest rates in Finance.

3.1 Application of Padé approximants to the study of interest rates

Many mathematical tools such as the Gaussian model or the Lévy distributions have been used in the field of Finance to model price fluctuations or describe a financial time series. But they turned out to be limited. Therefore mathematicians have been looking for more accurate techniques. The Padé approximants prove powerful instruments in Finance. For that extent, the first part of our study will concentrate on approximations of the discrete data the continuous distributions whereas the second part will show the relevance of the Padé approximants in describing probability densities.

3.1.1 Distribution of interest rates variations

The "term structure" of interest rates refers to the relationship between bonds of different terms. We are interested in analyzing the term structures of samples of N daily interest rates are given

$$I^{[m]}(t), \qquad t = 1, \ldots, N, \tag{1}$$

where $[m]$ in unit year specifies the constant maturity and t refers to chronological opening days. For our study, we will be using the interest rate variations, at lag L days

$$\delta I_L^{[m]}(t) = I_L^{[m]}(t+L) - I_L^{[m]}(t), \qquad t = 1, \ldots, N-L \tag{2}$$

Remarks

1. The study uses a business time rather than a physical time since the opening days t are not always consecutive. Sundays and Holidays are present as well. This difficulty is overlooked in the analysis.
2. For lag L days greater than 1, there is overlapping periods occurring. The distributions defined will consider those specific periods. As an illustration, for $L = 2$, a distribution of the even days and a distribution of the odd ones can be defined, leading to two distributions with non-overlapping two days periods. Consequently, the total number of points is divided by two (or by L for lag L) to get each distribution's number of points.

For a variation $\hat{v}$, a count can be performed to find the number of times $\hat{v}$ occurs in the experimental sample. This number is defined as $N_L^{[m]}(\hat{v})$.

We define the empirical discretized density function

$$\hat{f}_L^{[m]}(\hat{v}) = \frac{N_L^{[m]}(\hat{v})}{N-L}, \tag{3}$$

where $\hat{f}_L^{[m]}(\hat{v})$ refers to the integer values of $\hat{v}$.

We set $\hat{f}_L^{[m]}(\hat{v})$ to be zero outside the interval $[\hat{v}_{\min}, \hat{v}_{\max}]$ and normalized it as

$$\sum_{\hat{v}} \hat{f}_L^{[m]}(\hat{v}) = 1. \tag{4}$$

The densities $f(v)$, continuous functions of the continuous variations v, are normalized to have their integral equal to one. Normalizing the experimental discretized distributions (3) leads to the stepwise integrals

$$\int_{-\infty}^{\infty} \hat{f}_L^{[m]}(\hat{v})\, d\hat{v} = \sum_{\hat{v}} \hat{f}_L^{[m]}(\hat{v}) = 1 \tag{5}$$

Since the mean values of the distributions are zero, the second moments become the variance

$$Variance \equiv \int_{-\infty}^{\infty} v^2 f_L^{[m]}(v)\, dv = \sum_{v} \left(v^2 \hat{f}_L^{[m]}(\hat{v}) \right) \tag{6}$$

The distributions of the variation of interest rates in the continuous variable are expected to be smooth functions. They have a bell shape being maximal for a variation $v = 0$ and decreasing as v tends to infinity on both sides with tails wider than the Gaussian distribution. Moreover they are expected to be symmetrical around $v = 0$ which will make then even functions. This model is supported by the experimental data.

3.1.2 Presentation and statistical description of the data

The data used here are from the American daily spot interest rates for constant maturities equal to one, two, three, four, five, seven, ten and thirty years within February 15, 1977 to August 4, 1997. There are a total of 5108 opening days related to the seven interest rates calculated from bond prices published by the Board of Governors of the Federal Reserve System and available online.

Table 2: Univariate Statistic for the daily changes of the American Spot Interest Rates between February 15,1977 and August 4, 1997

$[m]$		$L=1$	$L=5$	$L=10$	$L=15$	$L=20$	$L=25$	$L=30$
[1]	Mean	0.0000	0.0001	0.0000	0.0001	0.0001	0.0002	0.0004
	Variance	0.0143	0.0860	0.1966	0.3326	0.4836	0.6446	0.8136
	Skewness	-0.1541	-0.8223	-1.0587	-1.0710	-1.1273	-1.1413	-1.1594
	Kurtosis	14.5664	10.7279	10.4445	10.9208	10.9326	10.9088	10.8980
	V min	-1.0800	-2.2700	-3.0600	-4.0700	-5.4500	-6.3100	-6.9100
	V max	1.1000	2.0600	2.4700	3.1400	3.2000	3.7200	4.0100
[2]	Mean	0.0000	-0.0002	-0.0005	-0.0007	-0.0009	-0.0009	-0.0010
	Variance	0.0113	0.0702	0.1594	0.2663	0.3836	0.5109	0.6456
	Skewness	-0.3648	-0.7398	-0.8722	-0.8171	-0.8547	-0.8600	-0.8741
	Kurtosis	12.4522	9.7384	9.1389	9.2421	8.8631	8.5522	8.3537
	V min	-0.8400	-2.0800	-2.7900	-3.3700	-4.4900	-5.2300	-5.9600
	V max	0.8900	1.9900	2.6000	2.8700	3.0900	3.3900	3.7100
[3]	Mean	-0.0001	-0.0004	-0.0010	-0.0015	-0.0019	-0.0022	-0.0025
	Variance	0.0102	0.0622	0.1374	0.2256	0.3225	0.4248	0.5421
	Skewness	-0.1628	-0.4432	-0.5977	-0.5000	-0.5254	-0.5487	-0.5928
	Kurtosis	10.4136	7.7748	7.6880	7.4571	6.7799	6.4504	6.3316
	V min	-0.7900	-1.5700	-2.8300	-3.0500	-3.5800	-4.3200	-5.1200
	V max	0.9200	1.9900	2.6000	3.0400	3.3200	3.6000	3.6500
[5]	Mean	-0.0001	-0.0007	-0.0016	-0.0024	-0.0032	-0.0038	-0.0044
	Variance	0.0091	0.0544	0.1182	0.1901	0.2661	0.3491	0.4409
	Skewness	-0.3064	-0.3938	-0.4482	-0.3617	-0.3448	-0.3109	-0.3025
	Kurtosis	8.7216	6.2382	5.9828	5.9650	5.5396	4.8740	4.4440
	V min	-0.7700	-1.5400	-2.3900	-2.5800	-3.0700	-3.6000	-4.3200
	V max	0.7200	1.6400	2.3600	2.8200	3.0800	3.3800	3.5100
[7]	Mean	-0.0002	-0.0010	-0.0021	-0.0031	-0.0040	-0.0048	-0.0056
	Variance	0.0085	0.0491	0.1044	0.1651	0.2287	0.2981	0.3749
	Skewness	-0.3066	-0.3907	-0.4981	-0.3602	-0.3019	-0.2683	-0.2637
	Kurtosis	8.1780	5.3213	5.4510	5.2149	4.6201	3.9410	3.5034
	V min	-0.7800	-1.3600	-2.4000	-2.5500	-2.9400	-3.1100	-3.6200
	V max	0.7000	1.5300	2.0300	2.6000	2.8000	3.1100	3.2600
[10]	Mean	-0.0002	-0.0012	-0.0025	-0.0037	-0.0049	-0.0060	-0.0070
	Variance	0.0076	0.0440	0.0928	0.1449	0.1984	0.2574	0.3224
	Skewness	-0.2817	-0.5431	-0.6719	-0.4999	-0.3918	-0.3246	-0.3224
	Kurtosis	6.9688	5.4452	5.7715	5.1414	4.1992	3.4414	3.0903
	V min	-0.7500	-1.3600	-2.3500	-2.6000	-2.6200	-2.8400	-3.3700
	V max	0.6500	1.3600	1.8500	2.3600	2.5500	2.8300	2.9700
[30]	Mean	-0.0002	-0.0013	-0.0027	-0.0039	-0.0052	-0.0062	-0.0073
	Variance	0.0062	0.0349	0.0723	0.1114	0.1518	0.1972	0.2475
	Skewness	-0.2245	-0.4463	-0.4792	-0.3334	-0.2077	-0.1437	-0.1078
	Kurtosis	6.8619	4.5314	4.0574	3.7200	3.1661	2.6781	2.3830
	V min	-0.7600	-1.3100	-1.8800	-2.1600	-2.3800	-2.2000	-2.4800
	V max	0.5000	0.9500	1.3800	1.6500	2.0800	2.3700	2.4900

The statistics are given for the following maturities [m] = [1], [2], [3], [5], [7], [10] and [30] and the subsets of lags L= 1, 5, 10, 15, 20, 25 and 30. The quantities are expressed in %/year.

We notice an increase of variances and leptokurticity for higher lags. The mean values are approximately zero and therefore the negligible differences of those means from zero can be overlooked. The data can display small asymmetry that will not be reproduced by even distributions.

A short mention on units used is necessary for the understanding of this material. Interest rates have the dimension of the inverse of the time, $[[t]]^{-1}$ and are usually given in $[\%/year]$ with exactly two significant digits. Hence they are expressed in $[b.p/year]$ as an integer number.

- $[b.p/year] = 10^{-4}/year$ is naturally used for interest rates and their variations. It gives a natural binning for the discretized distributions.
- Units $[\%/year]$ and $[year/\%]$ are used in the tables and figure for the different parameters as well as the means and variances
- Passing from one unit to the other is fairly simple using the equalities

$$[b.p/year] = \frac{[\%/year]}{100}$$

$$[year/b.p] = 100 \times [year/\%] = 100 \times century$$

- The value of q_i expressed in units $[year/b.p]^i$ becomes $100^i \times q_i$ in the unit $[year/\%]^i$, while n_i becomes $100^{i+1} \times n_i$. The Padé density $P(v; n_i, q_i)$ in units $[year/b.p]$ for v expressed in $[b.p/year]$ is related to the density $P^{new}(v) = P(v; 100^{i+1} n_i, 100^i q_i)$, where P^{new} is expressed in $[century]$ in terms of v expressed in $[\%/year]$ by

$$P^{new}(v) = P\left(v; 100^{i+1} n_i, 100^i q_i\right)$$

The units and dimensions used are summarized on the table below

Table 3: Units and dimensions

Variable	[[Dimension]]	[Unit] used in body of paper	[Unit] used in the table	Restriction
v	$[[t]]^{-1}$	$[b.p/year]$	$[\%/year]$	
s	$[[t]]^{-1}$	$[b.p/year]$	$[\%/year]$	
f	$[[t]]$	$[year/b.p]$	$[century]$	
P	$[[t]]$	$[year/b.p]$	$[century]$	
n_i	$[[t]]^{1+i}$	$[year/b.p]^{i+1}$	$[century]^{i+1}$	
u_0	$[[t]]^{1/2}$	$[year/b.p]^{1/2}$	$[century]^{1/2}$	
u_i	$[[t]]^{i}$	$[year/b.p]^{i}$	$[century]^{i}$	$for\ i=1,2$
q_i	$[[t]]^{i}$	$[year/b.p]^{i}$	$[century]^{i}$	
d_i	$[[t]]^{i}$	$[year/b.p]^{i}$	$[century]^{i}$	
Mean	$[[t]]^{-1}$	$[b.p/year]$	$[\%/year]$	
Variance	$[[t]]^{-2}$	$[b.p/year]^2$	$[\%/year]^2$	
Skewness	$[[t]]^{0}$			
Kurtosis	$[[t]]^{0}$			
χ^2	$[[t]]^{0}$			

We have recourse to the Hill estimator needed to examine the tail thickness and therefore draw a distribution that characterizes perfectly the tail. The Hill estimator is defined as

$$H_k = \frac{1}{k}\sum_{j=1}^{k} \log \Delta I_j^{[m]} - \log \Delta I_k^{[m]} \tag{7}$$

$\Delta I_j^{[m]}$ denotes the jth largest value of $\left|\delta_1^{[m]}(t)\right|$, $t = 1,...,N-1$. In other words, it is the largest value of the absolute values of the variations of interest rates at the elementary time scale, for a given maturity.

Hill estimator allows us to detect heavy tails and estimate Pareto index $1/H$ which has been consistently approximated as $1/H \approx 3$ .This fact coincides with the decay induced by the choice pf the Padé approximant $[0,4]$. Table 2 gives us outputs of $k = 250$, H, $1+1/H$. $k = 250$ means that the tail contains 5% of the extreme data points.

Table 4: Hill estimators H and the corresponding estimates of the power decrease 1+1/H

[m]	H	1+1/H
[1]	0.428 (0.026)	3.345 (0.144)
[2]	0.407 (0.023)	3.467 (0.141)
[3]	0.407 (0.024)	3.464 (0.144)
[5]	0.372 (0.022)	3.697 (0.157)
[7]	0.337 (0.023)	3.979 (0.207)
[10]	0.324 (0.020)	4.099 (0.203)
[30]	0.329 (0.028)	4.062 (0.2740

Standard errors are given enclosed in parenthesis.

3.2 Use of Padé approximants

We will expose the notation for Padé parameters, their normalization, variance and positivity.

3.2.1 Presentation

The Padé approximant $P^{[M,N]}(v)$ is expressed as a rational fraction of a continuous variable v

$$P^{[M,N]}(v) = \frac{T^M(v)}{B^N(v)} \tag{8}$$

where T and B are polynomials depending on real parameters since the Padé approximants are used to approximate the continuous distributions (5). In this chapter, we adopted a different notation for the Padé approximants compared to the previous chapters. This is due to the use of L as our lag. L has been referring to the degree of the denominator of the Padé approximants in the previous chapters. Therefore to avoid any confusion with decided to denote our Padé approximants with different letters.

To obtain positive probability densities, we consider only values of the Padé approximants leading to that result. The curves given by the Padé approximants should be left-right symmetric to perfectly fit the distributions. Consequently the polynomials T^M and B^N should be even functions meaning that the exponents N and M are even numbers.

To comply with one of the properties of a density function, we should normalize the Padé approximants.

$$\int_{-\infty}^{+\infty} P^{[M,N]}(v)\,dv = 1 \tag{9}$$

We require the variance to be equal to 1.

$$Variance \equiv \int_{-\infty}^{+\infty} P^{[M,N]}(v)\,v^2 dv = 1 \tag{10}$$

Since the Padé approximants considered here are even functions, the mean value will be zero. This agrees with our experimental sample.

The conditions (8), (9) coupled with (7) restrict the possible values of M and N which must satisfy the following inequality

$$N - M \geq 4 \tag{11}$$

for convergence of (9) and (10).

We will avoid too many parameters in the fit, though our experimental sample is quite large. The reason lies in the fact that even Padé approximants with low values for M and N do give satisfactory fits.

We need smoothness of the distribution which can be deterred if numerical minimizations with various parameters are used. In fact it will lead to fake oscillations that belong to no particular known intrinsic structure.

The consideration above helps us to limit ourselves to $N = 8$ and since by (11), $N - M \geq 4$, we take $M = 4$ .Since M and N have to be even, the relevant Padé approximants are [0,4], [0,6], [0,8], [2,6], [2,8], [4,8].

Our objective now is to give an explicit expression of Padé approximants. The parameterization of the numerator is obtained by introducing the complex polynomial of second degree $U^2(v)$

$$U^2(v) = u_0\left(1 + iu_1 v + u_2 v^2\right) \tag{12}$$

Then the numerator of $P^{[4,8]}$ is written as the norm square of $U^2(v)$

$$\begin{aligned} T^4(v) &= \left(U^2(v)\right) \times U^2(v) \\ &= u_0^2\left(1 + \left(u_1^2 + 2u_2\right)v^2 + u_2^2 v^4\right) \\ &\equiv n_0 + n_2 v^2 + n_4 v^4, \end{aligned} \tag{13}$$

where $n_0 = u_0^2, n_2 = u_0^2\left(u_1^2 + 2u_2\right)$ and $n_4 = u_2^2$.

The parameters in the expression above are the $u_i, i = 0,1,2$. The denominator $B^8(v)$ of $P^{[4,8]}$ is calculated using the complex polynomials with four parameters $q_i, i = 1,\ldots,4$

$$Q^4(v) = 1 + iq_1 v + q_2 v^2 + iq_3 v^3 + q_4 v^4 \tag{14}$$

We then get $B^8(v)$ as the norm square of $Q^4(v)$

$$\begin{aligned} B^8(v) &= \bar{Q}^4(v) \times Q^4(v) \\ &= 1 + \left(q_1^2 + 2q_2\right)v^2 + \left(q_2^2 + 2q_4 - 2q_1 q_3\right)v^4 + \left(q_3^2 + 2q_2 q_4\right)v^6 + q_4^2 v^8 \\ &= 1 + d_2 v^2 + d_4 v^4 + d_6 v^6 + d_8 v^8 \end{aligned} \tag{15}$$

We chose the term with degree 0 to be 1 without loss of generality as was previously done for the theory on Padé approximants.

The positivity of the Padé is assured for it is an even function of v being quotient of the norms squares of two complex polynomials.

3.2.2 Normalization, variance and positivity

We can compute analytically the normalization and variance. In fact an analytical continuation switching from real to complex parameter in Q is performed. As a result, the normalization (9) with the numerator and denominator know from (13) and (15) gives us

$$normalization\, Pade[4,8] = \pi \frac{\left(q_2 q_3 - q_1 q_4\right) n_0 - q_3 n_2 + q_1 n_4}{q_1 q_2 q_3 - q_1^2 q_4 - q_3^2} \tag{16}$$

The result gives us an expression with rather three parameters instead of four. We conclude that normalization diminishes the number of parameters from four down to three.

Since $P^{[0,4]}$ is of interest, let us get its explicit expression.

$$normalization\, Pade[0,4] = \pi \frac{n_0}{q_1}, \tag{17}$$

Equating it to 1 for the normalized Padé [0, 4] allows us to derive the value of n_0

$$n_0 = \frac{q_1}{\pi}. \tag{18}$$

The variance expressed in (10) is given by

$$VariancePade[4,8] = \pi \frac{-q_3 q_4 n_0 + q_1 q_4 n_2 + \left(q_3 - q_1 q_2\right) n_4}{q_4 \left(q_1 q_2 q_3 - q_1^2 q_4 - q_3^2\right)} \tag{19}$$

This variance reduces for this particular case to

$$\begin{aligned} Variance\,Pade[0,4] &= -\pi \frac{n_0}{q_1 q_2} \\ &= -\frac{1}{q_2} \end{aligned} \tag{20}$$

The Padé approximants [0,4] is then given by

$$P^{[0,4]}(v) = \frac{q_1}{\pi\left(1 + \left(q_1^2 + 2q_2\right)v^2 + q_2^2 v^4\right)} \tag{21}$$

This expression of Padé approximants [0,4] will be shown to approximate perfectly the normalized experimental distributions $f(v)$ in (3).

3.3 Empirical Padé densities

3.3.1 Determination of parameters q_1 and q_2

The empirical variance is directly computed from the data and thus utilized to determine q_2 thanks to formula (20).

We minimize a merit function to evaluate q_1. A merit function measures the agreement between data and the fitting model for a particular choice of the parameters. The chosen merit function χ^2 is expressed as

$$\chi^2 = \sum_{\hat{v}} \frac{\left(N_{Pade}\left(\hat{v}\right) - N\left(\hat{v}\right)\right)^2}{\sigma\left(\hat{v}\right)^2} \tag{22}$$

The formula (22) features the following characteristics:

- The number of observations from the Padé approximant $N_{Pade}\left(\hat{v}\right) = \left(N - L\right) \times P^{[0,4]}\left(v\right)$ where $P^{[0,4]}\left(v\right)$ is the normalized Padé density given by (21).
- $\hat{v}$ belongs to the interval $\left[\hat{v}_{\min}, \hat{v}_{\max}\right]$. Each extreme value of the interval for $\hat{v}$ corresponds to non-zero value of $N\left(\hat{v}\right)$.
- The errors $\sigma\left(\hat{v}\right)$ in the data can be reasonably assessed considering that interest rates statistically generated induce errors approximately of the order of $\sqrt{N\left(\hat{v}\right)}$.
- The choice $\sigma\left(\hat{v}\right) = 1$, for $\hat{v}$ such that $N\left(\hat{v}\right) = 0$ will make perfect sense since we want to avoid replacing $\sigma\left(\hat{v}\right)$ by 0 in the denominator of formula (22)

3.3.2 Fit with Padé [0, 4]

The hypothesis (at the 5% confidence level) that the data follow the Padé densities is asserted by the goodness of fit as measured by χ^2. The hypothesis fails for $\left[m\right] = 2, L = 1$ because the experimental curve displays an oscillatory behavior around the maximum with three separate peaks. A statistical fluctuation might explain this happening.

Table 5: Padé parameters q_1 and q_2 of our fits for all maturities and subset of the lags

$[m]$		$L=1$	$L=5$	$L=10$	$L=15$	$L=20$	$L=25$	$L=30$
[1]	χ^2	141.4	356.4	576.5	592.6	775.6	738.1	721.4
	q_1	25.21 (0.61)	9.019 (0.276)	6.171 (0.166)	4.825 (0.120)	4.014 (0.097)	4.014 (0.091)	3.051 (0.066)
	q_2	-69.90	-11.63	-5.08	-3.01	-2.07	-1.55	-1.23
[2]	χ^2	368.1	308.6	465.0	558.2	668.5	623.9	595.5
	q_1	24.05 (0.62)	8.218 (0.206)	5.442 (0.149)	4.222 (0.125)	3.427 (0.105)	2.933 (0.088)	2.620 (0.075)
	q_2	-88.51	-14.24	-6.27	-3.75	-2.61	-1.96	-1.55
[3]	χ^2	126.5	293.2	355.6	399.8	562.0	545.6	526.0
	q_1	24.26 (0.61)	7.971 (0.200)	5.167 (0.144)	3.915 (0.087)	3.280 (0.096)	2.823 (0.092)	2.550 (0.082)
	q_2	-97.57	-16.07	-7.28	-4.43	-3.10	-2.33	-1.84
[5]	χ^2	138.0	206.2	292.3	336.9	443.7	503.9	502.4
	q_1	24.28 (0.50)	7.977 (0.213)	5.336 (0.128)	4.135 (0.115)	3.440 (0.100)	2.998 (0.090)	2.761 (0.083)
	q_2	-110.04	-18.39	-8.46	-5.26	-3.76	-2.86	-2.27
[7]	χ^2	136.6	173.5	262.5	335.8	339.2	368.1	520.9
	q_1	23.81 (0.51)	8.036 (0.179)	5.614 (0.124)	4.361 (0.119)	3.547 (0.085)	3.102 (0.072)	2.960 (0.077)
	q_2	-118.07	-20.38	-9.58	-6.06	-4.37	-3.35	-2.67
[10]	χ^2	154.8	179.3	236.9	319.8	377.9	325.4	499.9
	q_1	24.61 (0.52)	8.535 (0.210)	5.832 (0.133)	4.593 (0.133)	3.826 (0.123)	3.286 (0.081)	3.129 (0.089)
	q_2	-131.22	-22.75	-10.78	-6.90	-5.04	-3.88	-3.10
[30]	χ^2	129.3	177.9	251.0	289.9	346.9	317.5	479.7
	q_1	27.85 (0.61)	9.880 (0.228)	6.889 (0.154)	5.484 (0.122)	4.537 (0.126)	3.909 (0.097)	3.645 (0.115)
	q_2	-160.15	-28.68	-13.83	-8.97	-6.59	-5.07	-4.04

The parameter q_2 (20) was estimated using directly the variance of the data. The values of q_1 have been obtained by minimizing the χ^2 (22) and their standard errors (given in parenthesis) are obtained by a bootstep method. The goodness of the fit is excellent enough for all cases except for $[m]=2, L=1$ which rejected at 5% level.

3.3.3 Comparison with the Gaussian distribution

We use the comparison of the data with the Gaussian distribution to judge the quality of fit with the Padé approximants. The Gaussian density of the form

$$G(v)=\frac{1}{\sqrt{2\pi\sigma}}e^{-(v-\mu)^2/2\sigma^2} \tag{23}$$

μ is a free parameter and σ is an observed parameter derived from experimental process. Unlike the Padé approximants, the Gaussian densities lead to high χ^2 .Additionally the Kolmogorov-Smirnov test rejected strongly the normal law for all $[m], L$ cases.

We sketched the distribution curves of the Padé, Gaussian and interest rates variations densities for a one year maturity and one day lag. As predicted the Padé distribution curve shares the same properties as our empirical data density curve namely symmetry, positivity, narrow peak and elongated tail. Consequently as the figure 1 below shows, the Padé density curve faithfully mimic the curve of the interest rates variations for a one year maturity and one day lag. Meanwhile, the Gaussian distribution curve ill-fits our data curve since it has a flat top and a short tail. Figure 1 displays the inaccuracy of approximation of our data with a Gaussian distribution whereas showcases the suitability of Padé densities as an appropriate fit.

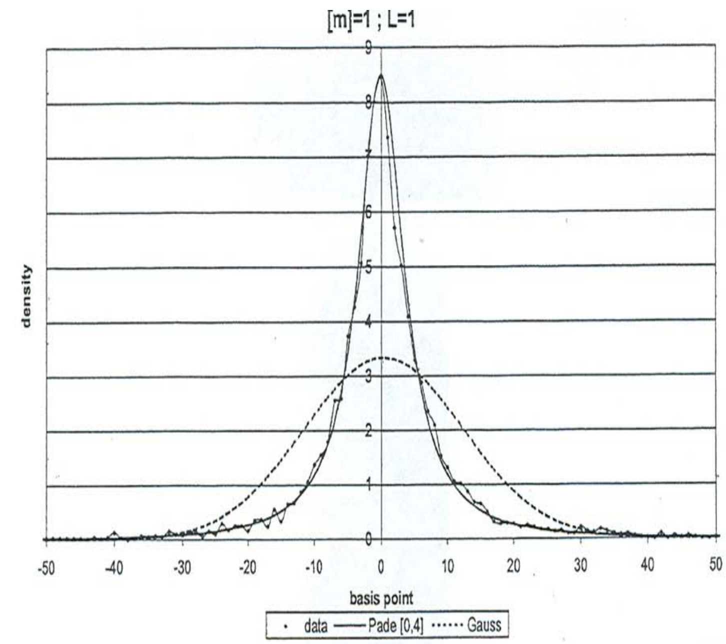

Figure 1: The experimental density of interest rate variations (dots), the Padé density (solid line) and the Gaussian distribution (dashed line) are all plotted in the figure above for the case $[m]=1,\ L=1$. As we notice, the Padé density follows closely the data.

3.3.4 Scaling law and expression of the parameter as functions of the lag

$q_1^{[m]}(L)$ and $q_2^{[m]}(L)$ can be approximated in the log-log plot by linear functions as follows

$$\begin{aligned}\ln\left(q_1^{[m]}(L)\right) &= \lambda_1^{[m]} \ln\left(\frac{L}{L_0}\right) + v_1^{[m]} ,\\ \ln\left(q_2^{[m]}(L)\right) &= \lambda_2^{[m]} \ln\left(\frac{L}{L_0}\right) + v_2^{[m]} .\end{aligned} \tag{24}$$

with L_0 arbitrarily chosen and fixed to 15.

The estimates of $\lambda_i^{[m]}$, $v_i^{[m]}$ and their t-stat are given for all maturities in Table 4 below. The scaling laws relate to λ_2 for the time dependence of the standard deviations

$$\sigma^{[m]}(L) = k^{[m]} L^{E^{[m]}} \tag{25}$$

For the most part in financial time series, standard deviations have this expression with the exponent ordinarily a little larger than 0.5, which should be observed for the Gaussian process. Here, for our case, $E^{[m]} = -\lambda_2^{[m]}/2$. We get absolute values of λ_2 greater than 1 which favors our expectation.

Table 6: Scaling parameters $v_1^{[m]}$ and $v_2^{[m]}$ for the two Padé parameters q_i

	$[m]=[1]$	$[m]=[2]$	$[m]=[3]$	$[m]=[5]$	$[m]=[7]$	$[m]=[10]$	$[m]=[30]$
q_1 $v_1^{[m]}$	1.553 (0.005)	1.416 (0.004)	1.371 (0.003)	1.419 (0.005)	1.462 (0.005)	1.515 (0.004)	1.677 (0.004)
$\lambda_1^{[m]}$	-0.607 (0.006)	-0.644 (0.004)	-0.652 (0.004)	-0.622 (0.006)	-0.606 (0.006)	-0.597 (0.005)	-0.588 (0.005)
R^2 SSR	0,998 0.0186	0.099 0.0105	0.999 0.0086	0.998 0.0181	0.997 0.0197	0.998 0.0152	0.998 0.0138
q_2 $v_2^{[m]}$	1.086 (0.008)	1.308 (0.006)	1.469 (0.006)	1.647 (0.004)	1.788 (0.004)	1.921 (0.003)	2.180 (0.003)
$\lambda_2^{[m]}$	-1.209 (0.009)	-1.200 (0.007)	-1.172 (0.007)	-1.139 (0.005)	-1.109 (0.004)	-1.093 (0.003)	-1.074 (0.004)
R^2 SSR	0.998 0.0510	0.999 0.0329	0.999 0.0290	0.999 0.0137	0.999 0.0118	0.999 0.0059	0.999 0.0074

The standard errors of these scaling parameters are enclosed in parenthesis. The quality of the fit is attested by the values of R^2, the square of the correlation coefficient between explanatory and explained variables, and of the SSR, the sum of the square residuals.

3.3.5 Fits with Padé Approximants with greater indices than $[0,4]$

We further our search of a best fit by investigating the possibility of a better fit to our data than $P^{[0,4]}$. Therefore our next question should be the following: is there a Padé of indices greater than $[0,4]$ which will fit the data more efficiently?

The matter of the fact is $P^{[0,4]}$ pretty much reaches the optimum in terms of data fitting for the following reasons:

1. - First it s primordial to notice that the validity of Padé $[M,N]$ as a fit for our data is tantamount to the existence of a finite variance (10).
2. The Padé $[0,2]$ coincides with the Cauchy distribution.
3. The cases of Padé $[M,N]$ with $M = N$ are not of utility since they cannot be normalized.
4. However Padé with indices higher than $[0,4]$ might produce better fits since they encompass higher number of parameters. This can be possible provided that the condition of finite variance is satisfied.

The fitting of the data with Padé's ($[4,8]$ and $[2,6]$) features no significant improvement of the criteria. In some cases, the minimization algorithm produces smaller values for the criteria while at the same time forcing the Padé curve to fit small oscillations in the experimental curves. These tiny oscillations are related to the finiteness of the data and thus have no link to the intrinsic structures. A trade-off is made since the minimal parameters are not stable like those fathered by $P^{[0,4]}$ and the minimum curve is sensitive to the starting values.

3.4 Conclusion of the study of Padé approximants as best fit to our data

We exposed and analyzed so far the fits of daily variations of the term structure of interest rates published by the Board of Governors of the Federal Reserve System. The study demonstrated that the Padé Approximants were appropriate fits for our experimental data. More precisely, the values of the parameters given by the Padé Approximants $\left[0,4\right]$ fit smoothly the data as a function of the lag and the maturity. Plus it is in agreement with the asymptotic decrease of the empirical densities and needs not more then two significant parameters to faithfully follow our data. Particularly we stress out the fact that the parameters are well represented in a log-log plot by straight line as function of the lag. This remark relates to the scaling laws reported for other financial time series.

Finally we draw a comparison of Lévy and Padé distributions. The Lévy distributions unlike the Padé distributions have convolution properties. But they both share narrower peaks around the maximum and longer tails in contrast to the Gaussian distribution. Luckily Padé distributions are advantageous compared to Lévy distributions for they are easier to write, tabulate and used for other applications once their best parameters are known.

The theory on application of Padé Approximant as a fit for daily variations of the term structure of interest rates is derived form the paper entitled Phenomenology of the term structure of interest rates with Padé Approximants authored by Jean Nuyts and Isabelle Platten.

CHAPTER 4: CONCLUSION

Padé approximation is utilized in a non-exhaustive list of fields. Its wide use lies on the fact that among its advantages, resides the possibility of deducing from any converging or diverging series of powers, a table of rational approximations of the functions represented by these series. This makes it a powerful mathematical tool. That is the reason why many physicists and applied mathematicians have relied on it and studied it to obtain ever faster computing algorithms. Analytic function theory, difference equations, the theory of moments, continued fractions among others, are fields in Mathematics closely connected with Padé approximations.

ε- algorithm and Q-D algorithms are used in Hydrodynamics, continuum mechanics, quantum mechanics.

Unfortunately, due the vast number of results developed around the Padé approximation, we were able to just cover a tiny portion of that concept. Nevertheless, Mr. Henri Padé probably never expected his work to expand that wide. Many researches are still being carried out in that domain since ever faster convergent algorithms are needed to improve accuracy and hence to contribute to an ever-growing technology.

APPENDIX: ANALYTICITY OF THE DENOMINATOR

An analytical computation of the normalization (16) and the variance (19) was used to determine the denominator of the Padé approximant (14). $Q^n(v)$ has its poles located in the upper-half plane of the complex variable v while its conjugate will have its own in the opposite side for the formula (14) to hold. $Q^n(v)$ is expressed as

$$Q^n(v) = \prod_{j=1}^{n}\left(1 + i\frac{v-s}{v_j}\right)$$

provided that the complex poles are of the form $v_j = s + a_j + ib_{j,}\;\; j = 1,...,4,$ with $b_j > 0.$

Taking $n = 3$ and expanding the expression above while taking into consideration the positivity of b_j, we obtain the following results

$$\begin{aligned} &q_1 > 0, \quad q_2 < 0, \\ &q_3 \le 0, \quad q_3 - q_1 q_2 \ge 0. \end{aligned}$$

For $n = 4$, Q^4 the conditions transform into

$$\begin{aligned} &q_4 \ge 0, \quad q_2 q_3 - q_1 q_4 \ge 0, \\ &q_1 q_2 q_3 - q_1^2 q_4 - q_4^2 \ge 0. \end{aligned}$$

When written in terms of the poles, the analytical form of the normalization and of the variance, lead to positive values.

LIST OF REFERENCES

[1] George, A. Baker, JR and Graves-Morris, P. (1996) Encyclopedia of Mathematics: Padé Approximants 2nd Edition.

[2] Nuyts, J and Platten, I. (2000) Phenomenology of the term structure of interest rates with Padé Approximants. Physica A 299 (2001), pp. 528-546.

[3] G.A. Baker Jr. and P. Graves-Morris, *Padé Approximants*, Encyclopedia of Mathematics and it Applications, Vol. 53, 2nd ed.(Camebridge Univ. Press, Cambridge, 1996).

[4] J.M. Beguin, C. Gourieroux and A. Monfort, Identification of a mixed autoregressive-moving average process: The corner method in: *Time Series,* ed. O.D. Anderson (North-Holland, Amsterdam, 1980) pp.423-436.

[5] A. Berlinet, Estimating the dregrees of an ARMA model, Compstat. Lect. 3 (1984) 61-94.

[6] A. Berlinet, Sequence transformations as statistical tools, appl. Numer. Math. 1 (1985) 531-544.

[7] A. Berlinet and C. Francq, Identification of a univariate ARMA model, Comput. Statist. 9 (1994) 117-133.

[8] A. Berlinet and C. Francq, On the Identifiability of minimal VARMA representations, Statist. Inference Stochastic Processes 1 (1998) 1-15.

[9] G.E.P. Box and G>M. Jenkins, *Time Series Analysis: Forecasting and Control,* revised ed. (Holden Day, San Francisco, CA,1976).

[10] C. Brezinski, *Padé Approximants and General Orthogonal Polynomials* (Birkhäuser, Basel/Boston, 1980).

[11] A. Bultheel, *Laurent Series and Their Padé Approximations* (Birkhäuser, Basel/Boston, 1987).

[12] L. Canina and S. Figlewski, The informational content of implied volatility, Rev. Financial Studies 6(3) (193) 659-681.

[13] E. W. Cheney, *Introduction to Approximation Theory* (McGraw-Hill, New York, 1966)

[14] P. Claverie, D. Szpiro and R. Topol, Identification des modèles à fonctions de transfert: La Méthode Padé Transformée en z, Ann. Econimie Statit. 17 (1990) 145-161.

[15] C. Francq, Identification et minimalité dans les series chronologiques, Thèse, Université des Sciences et Techniques de Languedoc, Montpelier II (1989).

[16] C. Gil and R. Alegria, An application of the Padé approximation to volatility modeling, Internat. Adv. Economic Res. 5(4) (1999) 446-465.

[17] C. González and V. Cano and C. Gil, The epsilon-algorithm for the identification of a transfer-function model: Some applicatons, Numer. Algorithms 9 (1995) 379-395.

[18] H.L. Gray, G.D. Kelley and D.D. Mac Intire, A new approach to ARMA modeling, Common. Statist. Simul. Comput. B 7 (1978) 1-77.

[19] D. Hanssens and L. Liu, Lag specification in rational distributed lag structural models, J. Bus. Econ. Statist. 1 (1983) 316-325.

[20] G.G. Judge, R.C. Hill, W.E. Griffiths and H. Lükepohl, *Introduction to the Theory and Practice of Econometrics,* 2nd ed. (Wiley, New York, 1988).

[21] K. Lii, Transfer function model order and parameter estimation, J. Time Series Anal. 6(3) (1985) 153-169.

[22] L.M. Liu and D.M. Hanssens, Identification of multiple inputs transfer function models, Commun. Statist. Theor. Math. A 11 (3) (1982) 297-314.

[23] H. Lükepohl, *Introduction to Multiple Time Series Analysis,* 2nd ed. (Springer, Berlin, 1993).

[24] H. Lükepohl and D.S. Poskitt, Specification of echelon-form VARMA models, J. Bus. Econ. Statist. 14(1) (1996) 69-79.

[25] B. Mareschal and G. Mélard, The corner method for identifying autoregressive moving average models, Appl. Statist. 37(2) (1988) 301-316.

[26] H. Padé, Sur la représentation approchée d'une fonction par des fractions rationnelles, Ann. Ecol. Norm. Sup. 9 (1982).

[27] C. Pestano and C. González, C., Characterization of the orders in VARMA models, Przeglad Statystyczny XLIII (3/4) (1996) 183-190.

[28] W.C. Pye and T.A. Atchison, An algorithm for the computation of higher order G-transformation, SIAM J. Numer. Anal. 10 (1973) 1-7.

[29] T. J. Rivlin, *An Introduction to Approximation of Functions* (Dover, New York, 1969)