Operators on Inhomogeneous Time Series

Gilles Zumbach and Ulrich Müller

Olsen & Associates
Research Institute for Applied Economics
Seefeldstrasse 233, 8008 Zürich, Switzerland.

e-mail: gilles@olsen.ch
phone: +41-1/386 48 24
Fax: +41-1/422 22 82

Keywords: inhomogeneous time series; operators; convolution; exponential moving average; derivative; volatilities.

Document id: GOZ.1998-11-01

February 1, 2000

Abstract

We present a toolbox to compute and extract information from inhomogeneous (i. e. unequally spaced) time series. The toolbox contains a large set of operators, mapping from the space of inhomogeneous time series to itself. These operators are computationally efficient (time and memory-wise) and suitable for stochastic processes. This makes them attractive for processing high-frequency data in finance and other fields. Using a basic set of operators, we easily construct more powerful combined operators which cover a wide set of typical applications.

The operators are classified in macroscopic operators (that have a limit value when the sampling frequency goes to infinity) and microscopic operators (that strongly depend on the actual sampling). For inhomogeneous data, macroscopic operators are more robust and more important. Examples of macroscopic operators are (exponential) moving averages, differentials, derivatives, moving volatilities, etc . . . .

Contents

1 Introduction **3**

2 Basic concepts **5**

2.1 Notations and mathematical preliminaries 5

2.2 Macroscopic and microscopic operators 6

3 Convolution operators: General considerations **7**

3.1 Linear operators 7

3.2 Range and width 8

3.3 Convolution of kernels 8

3.4 Build-up time interval 9

3.5 Homogeneous operators 11

3.6 Robustness 11

4 The menagerie of convolution operators **12**

4.1 Exponential moving average EMA$[\tau]$ 12

4.2 The iterated EMA$[\tau, n]$ 12

4.3 Moving average MA$[\tau, n]$ 13

4.4 Moving norm, variance and standard deviation 15

4.5 Differential $\Delta[\tau]$ 16

4.6 Derivative $D[\tau]$ and γ-Derivative $D[\tau, \gamma]$ 19

4.7 Volatility 19

4.8 Standardized time series $\hat{z}$, moving skewness and kurtosis 24

4.9 Moving correlation 24

4.10 Windowed Fourier transform 25

5 Microscopic operators **27**

5.1 Backward shift operator $\mathcal{B}$, or lag operator L 28

5.2 Time translation operator $\mathcal{T}$ 28

5.3 Regular Time Series RTS$[t_0, \delta t]$ 28

5.4 Microscopic return r and the difference operator δ 28

5.5 Microscopic derivative ∂ 29

5.6 Microscopic volatility 29

5.7 Tick frequency f and activity A 29

References **31**

1 Introduction

Nowadays, empirical research in finance is confronted to an ever increasing amount of data, caused in part by increasing computer power and communication speed. Many time series can be obtained at high frequency, often at market tick-by-tick frequency. These time series are inhomogeneous because market ticks arrive at random times. Inhomogeneous time series by themselves are conceptually simple; the difficulty lies in efficiently extracting and computing information from them.

Time series are the common mathematical framework used to represent the world of economics and finance. Among time series, the first important classification can be done according to the spacing of data points in time. Regularly spaced time series are called homogeneous, irregularly spaced series are inhomogeneous. An example of a homogeneous time series is a series of daily data, where the data points are separated by one day (on a business time scale which omits the weekends and holidays).

In most books on time series analysis, the field of time series is restricted to homogeneous time series already in the introduction (Granger and Newbold, 1977; Priestley, 1989; Hamilton, 1994). This restriction induces numerous simplifications, both conceptually and computationally, and was justified before cheap computers and high-frequency time series were available.

The goal of this paper is to revisit the time series framework in order to build a computational toolbox for inhomogeneous time series. In practice, this toolbox is attractive enough to be also applied to homogeneous time series. Given a time series z, such as an asset price, the general point of view is to compute another time series, such as the volatility of the asset, by the application of an operator $\Omega[z]$. We construct a set of basic operators that can be combined to compute more sophisticated quantities, for example different kinds of volatility or correlation. A few stylized properties of these operators are explored, but our main emphasis is to build a sufficient vocabulary of operators well suited to high-frequency data analysis. In this process, we should keep in mind a few important considerations:

- The computations must be efficient. Even if powerful computers are becoming cheaper, typical tick-by-tick data in finance is 100 or even 10'000 times denser than daily data. Clearly, we cannot afford to compute a full convolution for every tick. For these reasons, our basic workhorse is the exponential moving average (EMA) operator which can be computed very efficiently through an iteration formula. A wealth of complex but still efficient operators can be constructed by combining and iterating the basic operators.

- A stochastic behavior is the dominant characteristic of financial processes. For tick-by-tick data, not only the values but also the time points of the series are stochastic. In this random world, point-wise values are of little significance and we are more interested in average values inside intervals. Thus the usual notion of return also has to be changed. With daily data, a daily return is computed as $r_i = p_i - p_{i-1}$, i. e. as a point-wise difference between the price today and the price yesterday. With high-frequency data, a better definition of the daily return may be the difference between the average price of the last few hours and an average price from one day ago. In this way, it is possible to build smooth variables well suited to random processes. The calculus has to be revisited in order to replace point-wise values by averages over some time intervals.

- Analyzing data typically involves a characteristic time range; a return $r[\tau]$, for example, is computed on a given time interval τ. With high-frequency data, this characteristic time interval can vary from few minutes to several weeks. We have taken care of making explicit all these time range dependencies in the formulation of operators.

- We usually want smooth operators with smooth kernels (weighting functions of moving averages). A simple example of a discontinuous operator is an average with a rectangular weighting function, say of range τ. The second discontinuity at now$-\tau$, corresponding to forgetting events, creates unnecessary noise. Instead, we prefer kernels with a smooth decay to zero. Only at $t =$ now, we often prefer a jump in the kernel form. This jump gives a positive weight to the last arrived information and thus a rapid response in real time. For a discontinuous kernel, the weight at $t =$ now is inversely proportional to the range of the operator. Therefore, there is a trade-off between a fast reaction but more noise, and a smooth average behavior but a slow reaction time. Beside this fundamental noise created by the advance of events, it is better to have continuous and smooth operators.

For operators acting on inhomogeneous time series, an important distinction has to be made between 'microscopic' and 'macroscopic' operators. The essential difference is that a microscopic operator depends on the actual (random) sampling, whereas a macroscopic operator extracts average behaviors of its time series argument. A possible technical definition of macroscopic operators is that they have a well-defined limit when the quotes becomes infinitely dense. Practically, if the quotes are sufficently dense inside the range of the operator, we are close enough to this limit. A macroscopic operator is essentially immune to small variations of the individual quotes, including adding or removing quotes, whereas microscopic operators are not. For inhomogeneous time series, macroscopic operators are better behaved and more robust than microscopic operators.

Examples of a microscopic operator are the time difference δt between ticks (e. g. in $\delta t_i = t_i - t_{i-1}$) and the backward shift operator $\mathcal{B}$ defined in Section 5.1. The archetype of a macroscopic operator is the exponential moving average (EMA) that computes a moving average with an exponentially decaying weight of the past. For homogeneous time series, this distinction is unnecessary because the sampling frequency is fixed and there is no reason to take a continuous-time limit, or to formally add or remove ticks. Moreover, because the workhorse of homogeneous time series analysis is the backward shift operator $\mathcal{B}$ (which is microscopic), most of the conventional time series analysis becomes unusable for inhomogeneous time series. This classification between operators is further explored in Section 2.2.

The generalization to inhomogeneous time series introduces a number of technical peculiarities. Because of their time-translation invariance, all macroscopic operators can be represented by convolutions. A convolution is defined as an integral, therefore the series should be defined in continuous time. Actual data is known only at discrete sampling times, so some interpolation needs to be used in order to properly define the convolution integral. The same problem is present when constructing an artificial homogeneous time series from inhomogeneous data. Another technical peculiarity originates from the fact that our macroscopic operators are ultimately composed of iterated moving averages. All such EMA operators have non-compact kernels: the kernels decay exponentially fast, but strictly speaking they are positive. This implies an infinite memory; a build-up must be done over an initialization period before the value of an operator becomes meaningful. All the above points are discussed in detail in Section 3.

High-frequency data in finance has a property which implies another technical difficulty: strong intra-day and intra-week seasonalities due to the daily and weekly pattern of human activities. A powerful deseasonalization technique is needed, such as using a transformed business time scale (Dacorogna et al., 1993). Essentially, this scale is a continuous-time generalization of the familiar daily business time scale (which contains five days per week, Saturdays and Sundays omitted). A continuous business time scale ϑ allows us to map a time interval $\mathrm{d}t$ in physical time to an interval $\mathrm{d}\vartheta$ in business time, where $\mathrm{d}\vartheta/\mathrm{d}t$ is proportional to the expected market activity. All the operators presented in this paper can be evaluated on any business time scale. The only modification consists in replacing physical time intervals by the corresponding business time intervals. As this extension is straightforward, all the formulae are given in physical time and only few remarks on scaled time are made when interesting.

The plan of this article is as follows. Some basic concepts such as the notation and the distinction of macroscopic and microscopic operators are explained in Section 2. The main theoretical considerations are given in Section 3. In Section 4, a set of convenient macroscopic operators, including different moving averages and derivatives is presented. Armed with powerful basic operators, it is then easy to introduce moving volatility, correlation, moving skewness and kurtosis, and to generalize to complex-valued operators. In Section 5, we examine microscopic operators applied to inhomogeneous time series.

This paper has been written so as to be self-contained and to serve as a general reference. Examples are given with data taken from the foreign exchange market. When not specified, the data set is USD/CHF for the week from Sunday, October 26, 1997 to Sunday, November 2. On the time axis, the labels correspond to the day in October, with the points 32 and 33 corresponding to November 1 and 2. This week has been selected because on Tuesday October 28 some Asian stock markets crashed, causing turbulences in many markets around the world, including the FX market. Yet, the relation between a stock market crash originating in Asia and the USD/CHF foreign exchange rate is quite indirect, making this example interesting. The prices of USD/CHF for this example week are plotted in Figure 1. All the figures for this week have been computed using high-frequency data; the results have finally been sampled each hour using a linear interpolation scheme. The computations have been done in physical time, therefore exhibiting the full daily and weekly seasonalities contained in the data.

Finally, we want to emphasize that the techniques presented in this paper are suitable to be applied to a wide range of statistical computations in finance, for example the analysis needed in risk management. An early application can be found in (Pictet et al., 1992); a very different, recent application is the Scale of Market Shocks (Zumbach et al., 1998).

2 Basic concepts

2.1 Notations and mathematical preliminaries

The letter z is used to represent a generic time series. The elements or ticks, (t_i, z_i), of a time series z consist of a time t_i and a scalar value z_i. The generalization to multivariate inhomogeneous time series is fairly straightforward (except for the business time scale aspect!) and will not be discussed. The value $z_i = z(t_i)$ and the time point t_i constitute the i-th element of the time series z. The sequence of sampling (or arrival) times is required to be growing, $t_i > t_{i-1}$. The strict inequality is required in a true univariate time series and is theoretically always true if the information arrives through one channel. In practice,

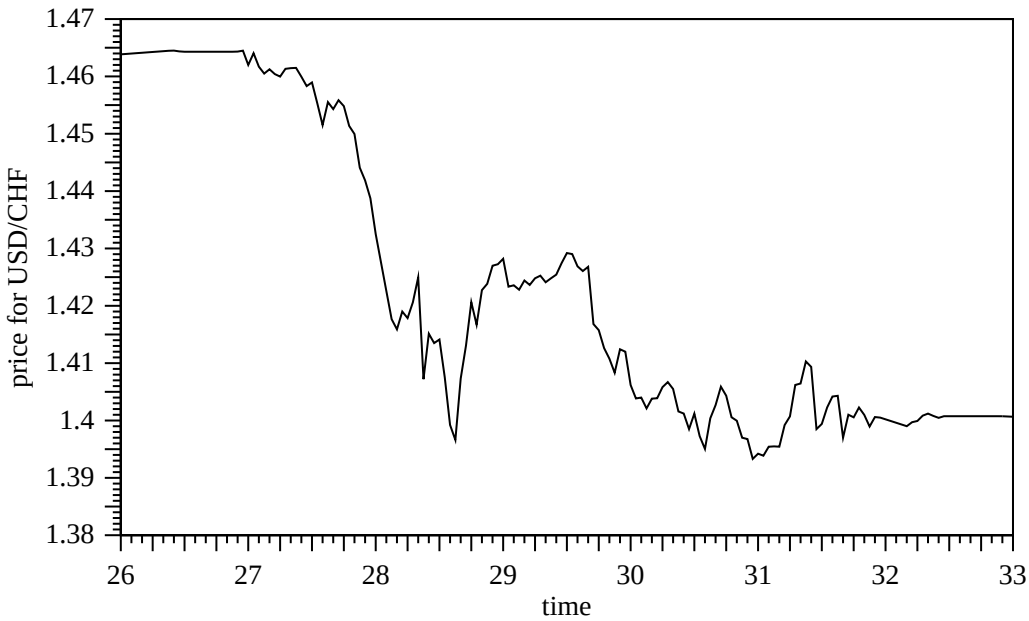


Figure 1: The FX rate USD/CHF for the week from Sunday, October 26, 1997 to Sunday, November 2. The high-frequency data are sampled hourly, using a linear interpolation scheme. From the market quote containing bid and ask prices, the (geometric) middle price is computed as $\sqrt{\text{bid} \cdot \text{ask}}$.

the arrival time is known with a finite precision, say of a second, and two ticks may well have the same arrival time. Yet, for most of the formulae below, the strict monotonicity of the time process is not required. A general time series is inhomogeneous, meaning that the sampling times are irregular. For an homogeneous time series, the sampling times are regularly spaced, $t_i - t_{i-1} = \delta t$. If a time series depends on some parameters θ, these are made explicit between square brackets, $z[\theta]$.

An operator Ω, from the space of time series into itself, is denoted by $\Omega[z]$. The operator may depend on some parameters $\Omega[\theta; z]$. The value of $\Omega[z]$ at time t is $\Omega[z](t)$. For linear operators, a product notation Ωz is also used. The average over a whole time series of length T is denoted by $\mathrm{E}\,[z] := 1/T \int dt\ z(t)$. The probability density function (pdf) of z is noted $p(z)$. A synthetic regular (or homogeneous) time series (RTS), spaced by δt, derived from the irregular time series z, is noted $\mathrm{RTS}[\delta t; z]$. A standardized time series for z is denoted $\hat{z} = (z - \mathrm{E}\,[z])/\sigma[z]$ and $\sigma[z]^2 = \mathrm{E}\left[(z - \mathrm{E}\,[z])^2\right]$.

The letter x is used to represent the logarithmic middle price time series $x = (\ln p_{\text{bid}} + \ln p_{\text{ask}})/2 = \ln(\sqrt{p_{\text{bid}}\, p_{\text{ask}}})$. This quantity is used in the examples and in a few discussions.

2.2 Macroscopic and microscopic operators

A heuristic distinction between microscopic and macroscopic operators has been made in the introduction. This is enough for understanding and applying the essence of this paper.

A mathematically precise distinction of the two operator types would be complex enough to require another paper. This is not our purpose, but the necessary rigorous steps are outlined here:

1. We need to define spaces of time series with different "smoothness" or "continuity" properties. The continuity property of financial data is not simple. Prices move

by small discrete jumps given by the "basis point" granularity. Moreover, for a decentralized market such as foreign exchange, several prices may coexist inside the (known) bid-ask spread. For technical purposes, time series spaces with an increasingly large number of ticks are needed, so that limits can be taken. Given the properties of financial data, this limit can be more complicated than the usual space of continuous Gaussian random paths. In summary, these facts and requirements make the construction of suitable time series spaces quite involved.

2. Once the time series spaces are exactly defined, we can introduce operators for mapping from and to these spaces. Only then we are able to define macroscopic and microscopic operators with mathematical rigor, possibly introducing even finer distinctions between operators.

For inhomogeneous time series with their irregular data sampling, only macroscopic operators are able to extract robust information that characterizes the local behavior of the time series. Robust information means that the result should essentially not depend on small changes of the time series such as adding or removing a few ticks, or altering a few time stamps. Therefore, macroscopic operators are the main subject of this paper. They can be represented by convolutions, which are developed in the next section.

Microscopic operators on inhomogeneous time series depend on the actual sampling and are thus unable to produce robust results. Yet, microscopic operators can also be interesting when the behavior of individual ticks is studied (as opposed to that of local averages). Their discussion is limited to Section 5.

3 Convolution operators: General considerations

3.1 Linear operators

If an operator is linear, time-translation invariant and causal, it can be represented by a convolution with a kernel $\omega(t)$

$$\begin{aligned} \Omega[z](t) &= \int_{-\infty}^{t} dt'\, \omega(t-t')\, z(t') \\ &= \int_{0}^{\infty} dt'\, \omega(t')\, z(t-t') \ . \end{aligned} \tag{1}$$

The kernel $\omega(t)$ is defined only on the positive semi-axis $t \geq 0$, and should decay for t large enough. With this convention for the convolution, the weight given to past events corresponds to the value of the kernel for positive argument. The value of the kernel $\omega(t-t')$ is the weight of events in the past, at a time interval $t-t'$ from t. In this convolution, $z(t')$ is a continuous function of time. Actual time series z are known only at the sampling time t_i and should be interpolated between sampling points. Many interpolation procedures for the value of z(t) between t_{i-1} and t_i can be defined, but three are used in practice:

- previous value $z(t) = z_{i-1}$,
- next value $z(t) = z_i$,
- linear interpolation $z(t) = \alpha(t) z_{i-1} + [1-\alpha(t)] z_i$ with $\alpha(t) = (t_i - t)/(t_i - t_{i-1})$.

The linear interpolation seems preferable as it leads to a continuous interpolated function. Moreover, linear interpolation defines the mean path of a random walk, given the start and end values. Unfortunately, it is non-causal, because in the interval between t_{i-1} and t_i, the value at the end of the interval z_i is used. Only the previous-value interpolation is causal, as only the information known at t_{i-1} is used in the interval between t_{i-1} and t_i. Any interpolation can be used for historical computations, but for the real-time situation, only the causal previous-value interpolation is defined. In practice, the interpolation scheme is almost irrelevant for good macroscopic operators, i. e. if the kernel has a range longer than the typical sampling rate.

The kernel $\omega(t)$ can be extended to $t \in \mathbb{R}$, with $\omega(t) = 0$ for $t < 0$. This is useful for analytical computation, particularly when the order of integral evaluations has to be changed. If the operator Ω is linear and time-translation invariant but non-causal, the same representation can be used except that the kernel may be non-zero on the whole time axis.

We often use two broad families of operators that share general shapes and properties:

- An average operator has a kernel which is non-negative, $\omega(t) \geq 0$, and normalized to unity, $\int dt\, \omega(t) = 1$. This implies that $\Omega[\text{parameters}; \text{Const}] = \text{Const}$.
- Derivative and difference operators have kernels that measure the difference between a value now and a value in the past (with a typical lag of τ). Their kernels have a zero average, $\int dt\, \omega(t) = 0$, such that $\Omega[\text{parameters}; \text{Const}] = 0$.

The integral 1 can also be evaluated in scaled time. In this case, the kernel is no more invariant with respect to physical time translation (i.e. it depends on t and t') but it is invariant with respect to translation in business time. If the operator is an average or a derivative, the normalization property is preserved in scaled time.

3.2 Range and width

The n-th moment of a causal kernel ω is defined as

$$\langle t^n \rangle_\omega := \int_0^\infty dt\, \omega(t)\, t^n \ . \tag{2}$$

The range r and the width w of an operator Ω are defined respectively by the following relations

$$\begin{aligned} r[\Omega] &= \langle t \rangle_\omega = \int_0^\infty dt\, \omega(t)\, t \ , \\ w^2[\Omega] &= \langle (t-r)^2 \rangle_\omega = \int_0^\infty dt\, \omega(t)\, (t-r)^2 \ . \end{aligned} \tag{3}$$

For most operators $\Omega[\tau]$ depending on a time range τ, the formula is set up so that $|r[\Omega[\tau]]| = \tau$.

3.3 Convolution of kernels

A common computation is to successively apply two linear operators:

$$\Omega_C[z] = \Omega_2 \circ \Omega_1[z] = \Omega_2\, \Omega_1\, z := \Omega_2[\Omega_1[z]] \ .$$

It is easy to show that the kernel of Ω_C is given by the convolution of the kernels of Ω_1 and Ω_2.

$$\omega_C = \omega_1 \star \omega_2 \qquad \text{or} \qquad \omega_C(t - t') = \int_{-\infty}^{\infty} dt''\, \omega_1(t - t'')\, \omega_2(t'' - t') \tag{4}$$

or, for causal operators,

$$\omega_C(t) = \int_{-t/2}^{t/2} dt'\, \omega_1(\frac{t}{2} - t')\, \omega_2(t' + \frac{t}{2}) \qquad \text{fort} \geq 0 \ , \tag{5}$$

and $\omega_C(t) = 0$ for $t < 0$. Under convolution, range and width obey the following simple laws:

$$\begin{aligned} r_C &= r_1 + r_2 \ , \\ w_C^2 &= w_1^2 + w_2^2 \ , \\ \langle t^2 \rangle_C &= \langle t^2 \rangle_1 + \langle t^2 \rangle_2 + 2 r_1 r_2 \ . \end{aligned} \tag{6}$$

3.4 Build-up time interval

As our basic building blocks are EMA operators, most kernels have an exponential tail for large t. This implies that, when starting the evaluation of an operator at time T, a build-up time interval must be elapsed before the result of the evaluation is 'meaningful', e.g. that the initial conditions at T are sufficiently forgotten. This heuristic statement can be expressed by quantitative definitions. We assume that the process $z(t)$ is known since T and is modeled before as an unknown random walk with no drift. The definition 1 for an operator Ω computed since T needs to be modified in the following way

$$\Omega[T; z](t) = \int_T^t dt'\, \omega(t - t')\, z(t') \ . \tag{7}$$

The 'infinite' build-up corresponds to $\Omega[-\infty; z](t)$. For $-T < 0$, the average build-up error ϵ at $t = 0$ is given by

$$\epsilon^2 \ = \ E[(\Omega[-T; z](0) - \Omega[-\infty; z](0))^2] \ = \ E[\left(\int_{-\infty}^{-T} dt'\, \omega(-t')\, z(t')\right)^2] \tag{8}$$

where the expectation $E[\,]$ is an average on the space of processes z. For a given build-up error ϵ, this equation is the implicit definition of the build-up time interval T. In order to compute the expectation, we need to specify the considered space of random processes. We assume simple random walks with constant volatility σ, namely

$$E[(z(t) - z(t + \delta t))^2] = \sigma \frac{\delta t}{1y} \ . \tag{9}$$

The symbol $1y$ denotes one year, so $\delta t / 1y$ is the length of δt expressed in years. With this choice of units, σ is an annualized volatility, with values roughly from 1% (for bonds) to 50% (for stocks), and a typical value of 10% for foreign exchange. For $t < -T$, $t' < -T$, we have the expectation

$$E[z(t) z(t')] = z(-T)^2 + \sigma \min(\frac{-t - T}{1y}, \frac{-t' - T}{1y}) \ . \tag{10}$$

Having defined the space of processes, a short computation gives

$$\epsilon^2 = z(-T)^2 \left(\int_T^\infty dt\, \Omega(t) \right)^2 + 2\sigma \int_T^\infty dt\, \omega(t) \int_T^t dt'\, \omega(t') \frac{t'-T}{1y} \ . \tag{11}$$

The first term is the 'error at initialization', corresponding to the decay of the initial value $\Omega[-T](-T) = 0$ in the definition 7. A better initialization is $\Omega[-T](-T) = z(-T) \int_0^\infty \omega(t)$, corresponding to a modified definition for $\Omega[T](t)$:

$$\Omega[T; z](t) = z(-T) \int_{-\infty}^T dt'\, \omega(t-t') + \int_T^t dt'\, \omega(t-t')\, z(t') \ . \tag{12}$$

Another interpretation for the above formula is that z is approximated by its most probable value $z(-T)$ for $t < T$. With this better definition for Ω, the error reduces to

$$\epsilon^2 = 2\sigma \int_T^\infty dt\, \omega(t) \int_T^t dt'\, \omega(t') \frac{t'-T}{1y} \ . \tag{13}$$

For a given kernel ω, volatility σ and error ϵ, Eq. 13 is an equation for T. Most of the kernels introduced in the next section have the scaling form $\omega(\tau, t) = \tilde{\omega}(t/\tau)/\tau$. In this case, the equation for $\tilde{T} = T/\tau$ reduces to

$$\epsilon^2 = 2\sigma \frac{\tau}{1y} \int_{\tilde{T}}^\infty dt\, \tilde{\omega}(t) \int_{\tilde{T}}^t dt'\, \tilde{\omega}(t')\, (t' - \tilde{T}) \ . \tag{14}$$

Since this equation cannot be solved for general operators, the build-up interval should be computed numerically. This equation can be solved analytically for the simple EMA kernel, and gives the solution for the build-up time

$$\frac{T}{\tau} = -\ln \epsilon + \frac{1}{2} \ln \left(\frac{\sigma}{2} \frac{\tau}{1y} \right) \ . \tag{15}$$

As expected, the build-up time interval is large for a small error tolerance and for processes with high volatility.

For operators more complicated than the simple EMA, eq. 14 is in general not solvable analytically. A simple rule of thumb can be given: the fatter the tail of the kernel, the longer the required build-up. A simple measure for the tail can be constructed from the first two moments of the kernel as defined by eq. 2. The aspect ratio $AR[\Omega]$ is defined as

$$AR[\Omega] = \langle t^2 \rangle_\omega^{1/2} / \langle t \rangle_\omega . \tag{16}$$

Both $\langle t \rangle$ and $\sqrt{\langle t^2 \rangle}$ measure the extension of the kernel and are usually proportional to τ; thus the aspect ratio is independent of τ and dependent only on the shape of the kernel, in particular its tail property. Typical values of this aspect ratio are $2/\sqrt{3}$ for a rectangular kernel and $\sqrt{2}$ for a simple EMA. A low aspect ratio means that the kernel of the operator has a short tail and therefore a short build-up time interval in terms of τ. This is a good rule for non-negative causal kernels; the aspect ratio is less useful for choosing the build-up interval of causal kernels with more complicated, partially negative shapes.

3.5 Homogeneous operators

There are many more ways to build non-linear operators; an example is given in Section 4.8 for the (moving) correlation. In practice, most non-linear operators are homogeneous of degree p, namely $\Omega[ax] = |a|^p\,\Omega[x]$ (here the word 'homogeneous' is used in a sense different from that in in the term 'homogeneous time series'). Translation-invariant homogeneous operators of degree pq take the simple form of a convolution

$$\Omega[z](t) = \left[\int_{-\infty}^{t} dt'\,\omega(t-t')\,|z(t')|^p\right]^q \tag{17}$$

for some exponents p and q. An example is the moving norm (see Section 4.4) with ω corresponding to an average and $q = 1/p$.

3.6 Robustness

Data errors (outliers) should be filtered prior to any computation. Outlier filtering is difficult and sometimes arbitrary for high-frequency data in finance; this data is stochastic with a fat-tailed distribution of price changes (Pictet et al., 1998). Sometimes it is desirable to build robust estimators to reduce the impact of outliers and the choice of the filtering algorithm. The problem is acute mainly when working with returns, for example when estimating a volatility, because the difference operator needed to compute the return r from the price x is sensitive to outliers. The following modified operator achieves robustness by giving a higher weight to the center of the distribution of returns r than to the tails:

$$\Omega[f;r] = f^{-1}\,\{\,\Omega[f(r)]\,\} \tag{18}$$

where f is an odd, monotonic function over $\mathbb{R}$. Possible mapping functions $f(x)$ are

$$\mathrm{sign}(x)|x|^\gamma \;=\; x\,|x|^{\gamma-1}\;, \tag{19}$$

$$\mathrm{sign}(x) \qquad \text{(this corresponds to } \gamma \to 0 \text{ in the above formula)}, \tag{20}$$

$$\tanh(x/x_0)\;. \tag{21}$$

Robust operator mapping functions defined by eq. 19 have an exponent $0 \le \gamma < 1$. In some special applications, operators with $\gamma > 1$, emphasizing the tail of the distribution, may also be used. In the context of volatility estimates, the usual L^2 volatility operator based on squared returns can be made more robust by using the mapping function $f = \mathrm{sign}(x)\sqrt{|x|}$ (the signed square root); the resulting volatility is then based on absolute returns as in eq. 39. More generally, the signed power $f(x) = \mathrm{sign}(x)|x|^p$ transforms an L^2 volatility into an L^{2p} volatility. This simple power law transformation is often used and therefore included in the definition of the moving norm, moving variance or volatility operators, eq. 32. Yet, some more general transformations can also be used.

4 The menagerie of convolution operators

4.1 Exponential moving average $\mathrm{EMA}[\tau]$

The basic exponential moving average (EMA) is a simple average operator, with an exponentially decaying kernel

$$\mathrm{ema}(t) = \frac{e^{-t/\tau}}{\tau} \ . \tag{22}$$

This EMA operator is our foundation stone, because its computation is very efficient and other more complex operators can be built with it, such as MAs, differentials, derivatives and volatilities. The numerical evaluation is efficient because of the exponential form of the kernel, which leads to a simple iterative formula first proposed by (Müller, 1991):

$$\begin{aligned} \mathrm{EMA}[\tau; z](t_n) &= \mu\ \mathrm{EMA}[\tau; z](t_{n-1}) + (\nu - \mu)\, z_{n-1} + (1-\nu)\, z_n \ \text{, with} \\ \alpha &= \frac{t_n - t_{n-1}}{\tau} \\ \mu &= e^{-\alpha} \end{aligned} \tag{23}$$

and ν depends on the chosen interpolation scheme,

$$\nu = \begin{cases} 1 & \text{previous point} \\ (1-\mu)/\alpha & \text{linear interpolation} \\ \mu & \text{next point} \end{cases} \tag{24}$$

Thanks to this iterative formula, the convolution is never computed in practice; only few multiplications and additions have to be done for each tick. In Section 4.10, the EMA operator is extended to the case of complex kernels.

4.2 The iterated $\mathrm{EMA}[\tau, n]$

The basic EMA operator can be iterated to provide a family of iterated exponential moving average operators $\mathrm{EMA}[\tau, n]$. Practitioners of "Technical Analysis" have applied simple and (occasionally) iterated EMA operators to homogeneous time series since a long time. Iterated EMA operators for inhomogeneous time series were first explored by (Müller, 1991) and are systematically developed and discussed here. A simple recursive definition is

$$\mathrm{EMA}[\tau, n; z] = \mathrm{EMA}[\tau; \mathrm{EMA}[\tau, n-1; z]] \tag{25}$$

with $\mathrm{EMA}[\tau, 1; z] = \mathrm{EMA}[\tau; z]$. This definition can be efficiently evaluated by using the iterative formula 23 for all its basic EMAs. There is one subtle point related to the choice of the interpolation scheme 24. The EMA of z necessarily has an interpolation scheme different from that used for z. The correct form of $\mathrm{EMA}[\tau; z]$ between two points is no longer a straight line but a non-linear (exponential) curve. (Theoretically, it is straightforward to derive the corresponding exact interpolation formula). When using one of the interpolation schemes of eq. 24 after the first iteration, we are making a small error. Yet, if the kernel is wide as compared to $t_n - t_{n-1}$, this error is indeed very small. As a suitable approximation, we recommend using linear interpolation in the second and all

further EMA iterations, even if the first iteration was based on the next-point interpolation. The only exception occurs if z_n is not yet known; then we need a causal operator based on the previous-point interpolation.

The kernel of $\mathrm{EMA}[\tau, n]$ is

$$\mathrm{ema}[\tau, n](t) = \frac{1}{(n-1)!} \left(\frac{t}{\tau}\right)^{n-1} \frac{e^{-t/\tau}}{\tau} \ . \tag{26}$$

This family of functions is related to Laguerre polynomials which are orthogonal with respect to the measure e^{-t} (for $\tau = 1$). Through an expansion in Laguerre polynomials, any kernel can be expressed as a sum of iterated EMA kernels. Therefore, the convolution with an arbitrary kernel can be evaluated by iterated exponential moving averages. Yet, the convergence of this expansion may be slow, namely high-order iterated EMAs may be necessary, possibly with very large coefficients. This typically happens if one tries to construct operators that have a decay other (faster) than exponential. Therefore, in practice, we construct operators 'empirically' from a few low-order EMAs, in a way to minimize the build-up time. The set of operators provided by this paper covers a wide range of computations needed in finance.

The range, width and aspect ratio of the iterated EMA are

$$\begin{aligned} r &= n\tau \ , \\ \langle t^2 \rangle &= n(n+1)\tau^2 \ , \\ w^2 &= n\tau^2 \ , \\ AR &= \sqrt{(n+1)/n} \ . \end{aligned} \tag{27}$$

The iterated $\mathrm{EMA}[\tau, n]$ operators with large n have a shorter, more compact kernel and require a shorter build-up time interval than a simple EMA of the same range $n\tau$. This is indicated by the aspect ratio AR which decreases towards 1 for large n. Each basic EMA operator that is part of the iterated EMA has a range τ which is much shorter than the range $n\tau$ of the full kernel. Even if the tail of the kernel is still exponential, it decays faster due to the small basic EMA range τ.

In order to further extend our computational toolbox, we build another type of compact kernel by combining iterated EMAs, as shown in the next section. As the iterated EMAs, these combined iterated EMAs have a shorter build-up time interval than a simple EMA of the same range.

4.3 Moving average $\mathrm{MA}[\tau, n]$

A very convenient moving average is provided by

$$\mathrm{MA}[\tau, n] = \frac{1}{n} \sum_{k=1}^{n} \mathrm{EMA}[\tau', k] \qquad \mathrm{with} \tau' = \frac{2\tau}{\mathrm{n}+1} \ . \tag{28}$$

The parameter τ' is chosen so that the range of $\mathrm{MA}[\tau, n]$ is $r = \tau$, independently of n. This provides a family of more rectangular-shaped kernels, with the relative weight of the distant past controlled by n. Kernels for different values of n and $\tau = 1$ are shown in Figure 2. Their analytical form is given by

$$\mathrm{ma}[\tau, n](t) = \frac{n+1}{n} \frac{e^{-t/\tau'}}{2\tau} \sum_{k=0}^{n-1} \frac{1}{k!} \left(\frac{t}{\tau'}\right)^k \ . \tag{29}$$

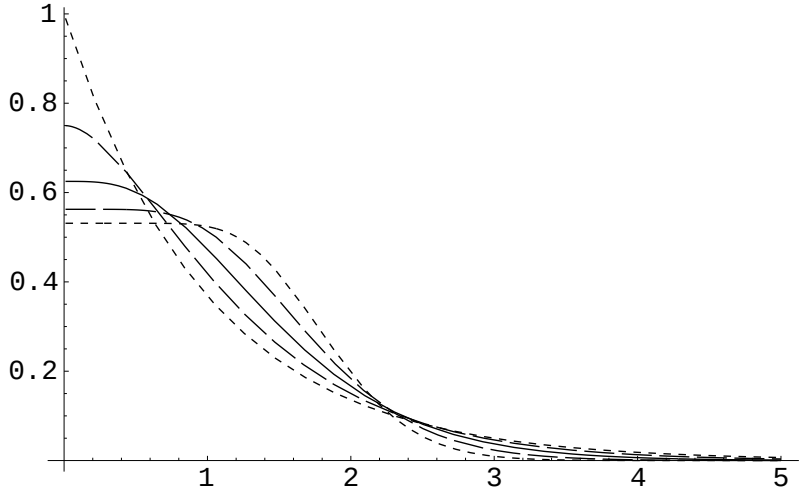


Figure 2: ma$[\tau, n](t)$ for $n = 1, 2, 4, 8$ and 16, for $\tau = 1$

For $n = \infty$, the sum corresponds to the Taylor expansion of $\exp(t/\tau')$, which cancels the term $\exp(-t/\tau')$, making the kernel constant. For finite n, when t/τ' is small enough, the finite sum will be a very good approximation of $\exp(t/\tau')$. Small enough means that the largest term in the sum is of order one: $(t/\tau')^n/n! \sim 1$. For large n, the condition $(t/\tau')^n/n! \sim 1$ corresponds to $t \sim 2\tau$ (using Stirling's approximation $n! \sim n^n$). Therefore, for $t \ll 2\tau$, the series approximates well the Taylor expansion of an exponential

$$\sum_{k=0}^{n-1} \frac{1}{k!} \left(\frac{t}{\tau'}\right)^k \to e^{t/\tau'} \ ,$$

$$\text{ma} \to \frac{n+1}{n} \frac{1}{2\tau} \ .$$

This explains the constant behavior of the kernel for $t \ll 2\tau$. For $t > 2\tau$ large, the exponential always dominates and the kernel decays to zero. Therefore, for large n, this operator tends to a rectangular moving average for which $AR = 2/\sqrt{3}$. For n values of $n \sim 5$ and higher, the kernel is rectangular-like more than EMA-like; this can as seen in Figure 2. These rectangular-like kernels are preferred to the rectangular kernel itself because they fade smoothly rather than abruptly. Abrupt "forgetting" of past events leads to superfluous noise in the operator results.

The decay of MA kernels is shown in Figure 3. The aspect ratio of the MA operator is

$$AR = \sqrt{\frac{4(n+2)}{3(n+1)}} \ . \tag{30}$$

Clearly, the larger n, the shorter the build-up.

This family of operators can be extended by 'peeling off' some EMAs with small k:

$$\text{MA}[\tau, n_{\text{inf}}, n_{\text{sup}}] = \frac{1}{n_{\text{sup}} - n_{\text{inf}} + 1} \sum_{k=n_{\text{inf}}}^{n_{\text{sup}}} \text{EMA}[\tau', k] \qquad \text{with} \quad \tau' = \frac{2\tau}{\text{n}_{\text{sup}} + \text{n}_{\text{inf}}}$$

and with $1 \leq n_{\text{inf}} \leq n_{\text{sup}}$. By choosing such a modified MA with $n_{\text{inf}} > 1$, we can generate a lagged operator with a kernel whose rectangular-like form starts after a lag rather than

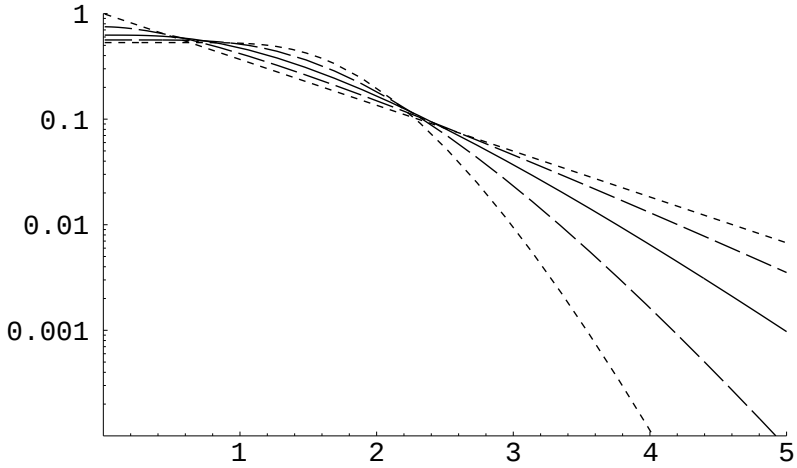


Figure 3: $\mathsf{ma}[\tau, n](t)$ for $n = 1, 2, 4, 8$ and 16, for $\tau = 1$, on a logarithmic scale.

immediately. At the same time, the kernel loses its abrupt behavior at $t = 0$ and becomes fully continuous, thus reducing noise in the results even further. These properties are useful in some applications, but the time delay implied by the lag makes such kernels less attractive for real-time applications.

Almost everywhere, a moving average operator can be used instead of a sample average. The sample average of $z(t)$ is defined by

$$\mathrm{E}\,[z] = \frac{1}{t_e - t_s} \int_{t_s}^{t_e} dt'\, z(t') \tag{31}$$

where the dependency on start-time t_s and end-time t_e is implicit on the left-hand side. This dependency can be made explicit, for example with the notation $\mathrm{E}\,[t_e - t_s; z](t_e)$, thus demonstrating the parallelism between the sample average and a moving average $\mathrm{MA}[2\tau; z](t)$. The conceptual difference is that when using a sample average, t_s and t_e are fixed, and the sample average is a number (the sample average is a functional from the space of time series to $\mathbb{R}$), whereas the MA operator produces another time series. Keeping this difference in mind, we can replace the sample average $\mathrm{E}\,[\cdot]$ by a moving average $\mathrm{MA}[\cdot]$. For example, we can construct a standardized time series $\hat{z}$ (as defined in Section 2.1), a moving skewness, or a moving correlation (see the various definitions below). Yet, be aware that sample averages and MAs can behave differently, for example $\mathrm{E}\left[(z - \mathrm{E}\,[z])^2\right] = \mathrm{E}\left[z^2\right] - \mathrm{E}\,[z]^2$, whereas $\mathrm{MA}[(z - \mathrm{MA}[z])^2] \neq \mathrm{MA}[z^2] - \mathrm{MA}[z]^2$.

4.4 Moving norm, variance and standard deviation

With the efficient moving average operator, we can define the moving norm, moving variance and moving standard deviation operators

$$\begin{aligned} \mathrm{MNorm}[\tau, p; z] &= \mathrm{MA}[\tau; |z|^p]^{1/p} \,, \\ \mathrm{MVar}[\tau, p; z] &= \mathrm{MA}[\tau; |z - \mathrm{MA}[\tau; z]|^p] \,, \\ \mathrm{MSD}[\tau, p; z] &= \mathrm{MA}[\tau; |z - \mathrm{MA}[\tau; z]|^p]^{1/p} \,. \end{aligned} \tag{32}$$

The norm and standard deviation are homogeneous of degree one with respect to z. The p-moment is related to the norm by $\mu_p = \mathrm{MA}[|z|^p] = \mathrm{MNorm}[z]^p$. Usually, $p = 2$ is taken.

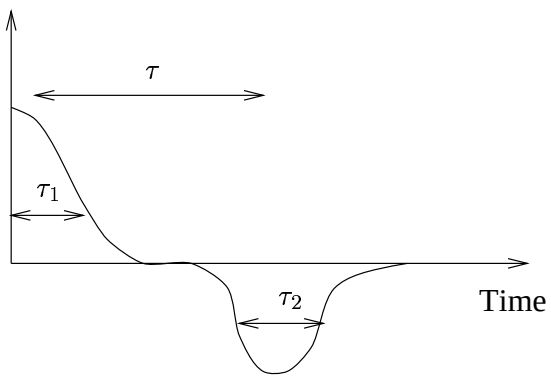


Figure 4: A schematic differential kernel.

Lower values for p provide a more robust estimate (see Section 3.6), and $p = 1$ is another common choice. Yet, even lower values can be used, for example $p = 1/2$.

In the formulae for MVar and MSD, there are two MA operators with the same range τ and the same kernel. This choice is in line with common practice: empirical means and variances are computed for the same sample. Yet, other choices can be interesting, for example the sample mean can be estimated with a longer time range.

4.5 Differential $\Delta[\tau]$

As argued in the introduction, a low-noise differential operator suitable to stochastic processes should compute an 'average differential', namely the difference between an average around time 'now' over a time interval τ_1 and an average around time 'now $- \tau$' on a time interval τ_2. The kernel may look like in Figure 4.

Kernels of a similar kind are used for wavelet transforms. This analogy also applies to other kernels form and is further discussed in Section 4.10.

Usually, τ, τ_1 and τ_2 are related and only the τ parameter appears, with $\tau_1 \sim \tau_2 \sim \tau/2$. The normalization for Δ is chosen so that $\Delta[\tau; c] = 0$ for a constant function $c = c(t) =$ constant, and $\Delta[\tau; t] = \tau$. Note that our point of view is different from that used in continuous-time stochastic analysis. In continuous time, the limit $\tau \to 0$ is taken, leading to the Ito derivative with its subtleties. In our case, we keep the range τ finite in order to be able to analyze a process at different time scales, i. e. for different orders of magnitudes of τ. Moreover, for financial data, the limit $\tau \to 0$ cannot be taken because a process is known only on a discrete set of time points (and probably does not exist in continuous time).

The following operator can be selected as a suitable differential operator:

$$\Delta[\tau] = \gamma \left(\mathrm{EMA}[\alpha\tau, 1] + \mathrm{EMA}[\alpha\tau, 2] - 2 \ \mathrm{EMA}[\alpha\beta\tau, 4] \right) \tag{33}$$

with $\gamma = 1.22208$, $\beta = 0.65$ and $\alpha^{-1} = \gamma(8\beta - 3)$. This operator has a well-behaving kernel that is plotted in Figure 5. The value of γ is fixed so that the integral of the kernel from the origin to the first zero is one. The value of α is fixed by the normalization condition and the value of β is chosen in order to get a short tail. The tail can be seen in Figure 6. This shows that after $t = 3.25\tau$, the kernel is smaller than 10^{-3}, which translates into a small required build-up time of about 4τ.

In finance, the main purpose of a Δ operator is computing returns of a time series of a

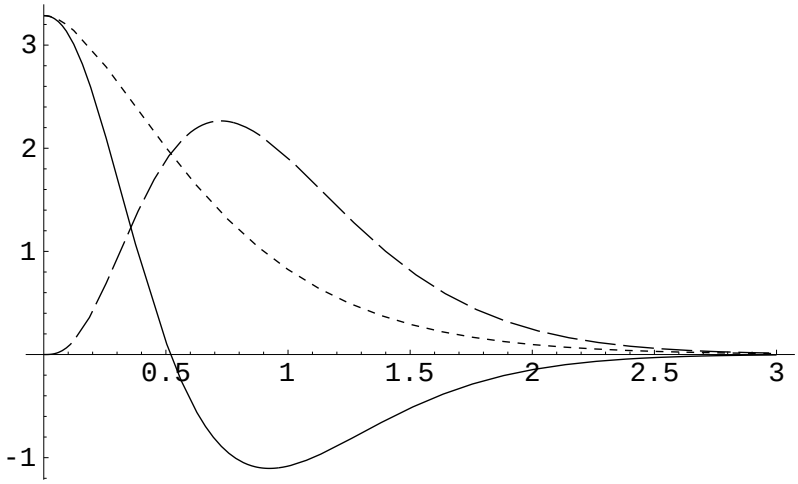


Figure 5: Kernel of the Differential operator (full line) for $\tau = 1$. The dotted curve corresponds to the first two terms of the operator $\gamma(\mathrm{EMA}[\alpha\tau, 1] + \mathrm{EMA}[\alpha\tau, 2])$, the dashed curve to the last term $2\gamma\ \mathrm{EMA}[\alpha\beta\tau, 4]$.

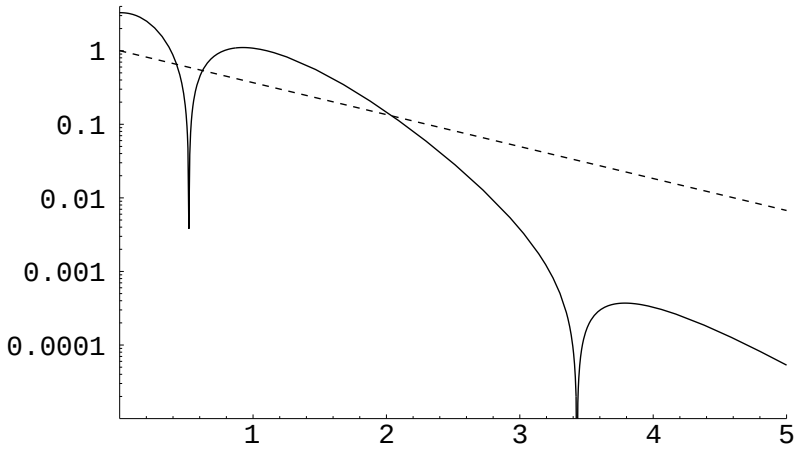


Figure 6: The absolute value of the kernel of the Differential operator (full line), in a logarithmic scale. The dotted line shows a simple EMA with range τ, demonstrating the much faster decay of the differential kernel.

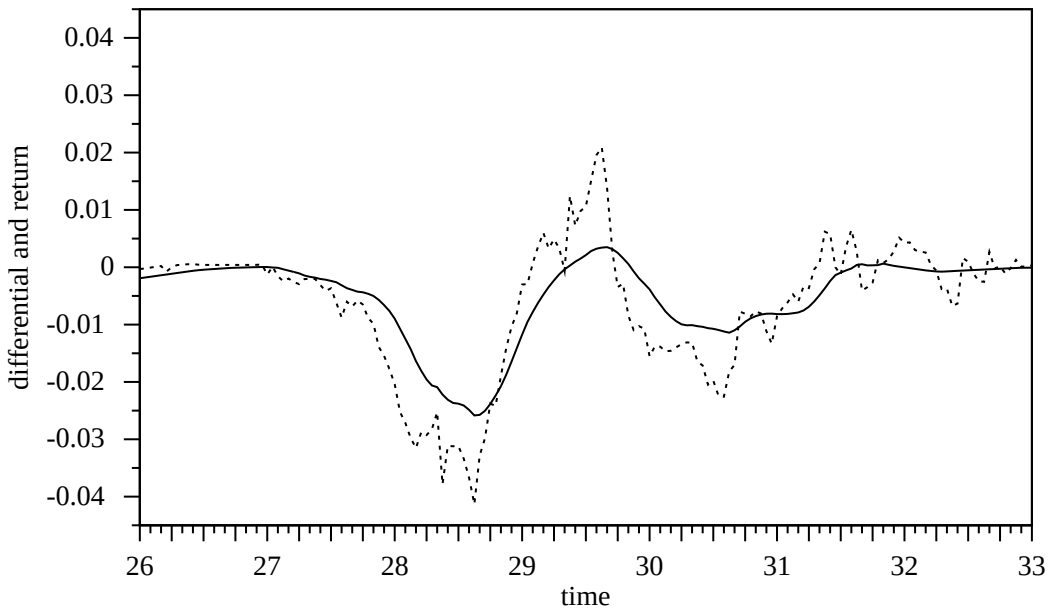


Figure 7: A comparison between the differential computed using the formula (33) with $\tau =$ 24h (full line) and the point-wise return $x(t) - x(t - 24\text{h})$ (dotted line). The time lag of approximately 4h between the curves is essentially due to the extent of both the positive part of the kernel ($0 < t < 0.5$) and the tail of the negative part ($t > 1.5$).

(logarithmic) prices x with a given time interval τ. Returns are normally defined as changes of x over τ; we prefer the alternative return definition $r[\tau] = \Delta[\tau; x]$. This computation requires the evaluation of 6 EMAs and is therefore efficient, time-wise and memory-wise. An example using our 'standard week' is plotted in Figure 7, demonstrating the low noise level of the differential.

The conventionally computed return $r[\tau](t) = x(t) - x(t - \tau)$ is very inefficient to evaluate for inhomogeneous time series. The computation of $x(t-\tau)$ requires many old t_i, x_i values to be kept in memory, and the t_i interval bracketing the time $t - \tau$ has to be searched for. Moreover, the number of ticks to be kept in memory is not bounded. This return definition corresponds to a differential operator kernel made of two δ functions (or to the limit $\tau_1, \tau_2 \to 0$ of the kernel in Figure 4). The quantity $x(t)-x(t-\tau)$ can be quite noisy, so a further EMA might be taken to smooth it. In this case, the resulting effective differential operator kernel has two discontinuities, at 0 and at τ, and decays exponentially, i. e. much slower than the kernel of $\Delta[\tau; x]$. Thus it is cleaner and more efficient to compute returns with the Δ operator of eq. 33.

Another quantity commonly used in finance is $x - \mathrm{EMA}[\tau; x]$, often called a momentum or an oscillator. This is also a differential with the kernel $\delta(t) - \exp(-t/\tau)/\tau$, with a δ function at $t = 0$. A simple drawing shows that the kernel of eq. 33 produces a much less noisy differential. Other appropriate kernels can be designed, depending on the application. In general, there is a trade-off between the averaging property of the kernel and a short response to shocks of the original time series.

4.6 Derivative $D[\tau]$ and γ-Derivative $D[\tau, \gamma]$

The derivative operator

$$D[\tau] = \frac{\Delta[\tau]}{\tau} \tag{34}$$

behaves exactly as the differential operator, except for the normalization $D[\tau; t] = 1$. This derivative can be iterated in order to construct higher order derivatives:

$$D^2[\tau] = D[\tau; D[\tau]] \ . \tag{35}$$

The range of the second-order derivative operator is $2\,\tau$. More generally, the n-th order derivative operator D^n, constructed by iterating the derivative operator n times, has a range $n\tau$. As defined, the derivative operator has the dimension of an inverse time. It is easier to work with dimensionless operators and this is done by measuring τ in some units. One year provides a convenient unit, corresponding to an annualized return when $D[\tau]x$ is computed. The choice of unit is denoted by $\tau/1y$, meaning that τ is measured in year, yet other units could be taken as well.

For a random diffusion process, a more meaningful normalization for the derivative is to take $D[\tau] = \Delta[\tau]/\sqrt{\tau/1y}$. For a space of processes as in Section 3.4, such that eq. 9 holds, the basic scaling behavior with τ is eliminated, namely $\mathrm{E}\left[(D[\tau]z)^2\right] = \sigma^2$. More generally, we can define a γ-derivative as

$$D[\tau, \gamma] = \frac{\Delta[\tau]}{(\tau/1y)^\gamma} \ . \tag{36}$$

In particular

$$\begin{array}{rl} \gamma = 0 & \text{differential,} \\ \gamma = 0.5 & \text{stochastic diffusion process,} \\ \gamma = 1 & \text{the usual derivative.} \end{array} \tag{37}$$

An empirical probability density function for the derivative is displayed in Figure 8. We clearly see that the main part of the scaling with τ is removed when using the γ-derivative with $\gamma = 0.5$.

4.7 Volatility

Volatility is a measure widely used for random processes, quantifying the size and intensity of movements, namely the 'width' of the probability distribution $P(\Delta z)$ of the process increment Δz, where Δ is a difference operator yet to be chosen. Often the volatility of market prices is computed, but volatility is a general operator that can be applied to any time series. There are many ways to turn this idea into a definition, and there is no unique, universally accepted definition of volatility in finance. The most common computation is the volatility of daily prices, Volatility$[x]$, evaluated for a regular time series in business time, with a point-wise price difference $r_i = \Delta x_i = x(t_i) - x(t_i - \tau')$ and $\tau' = 1$ day. The time horizon τ' of the return is one parameter of the volatility; a second parameter is the

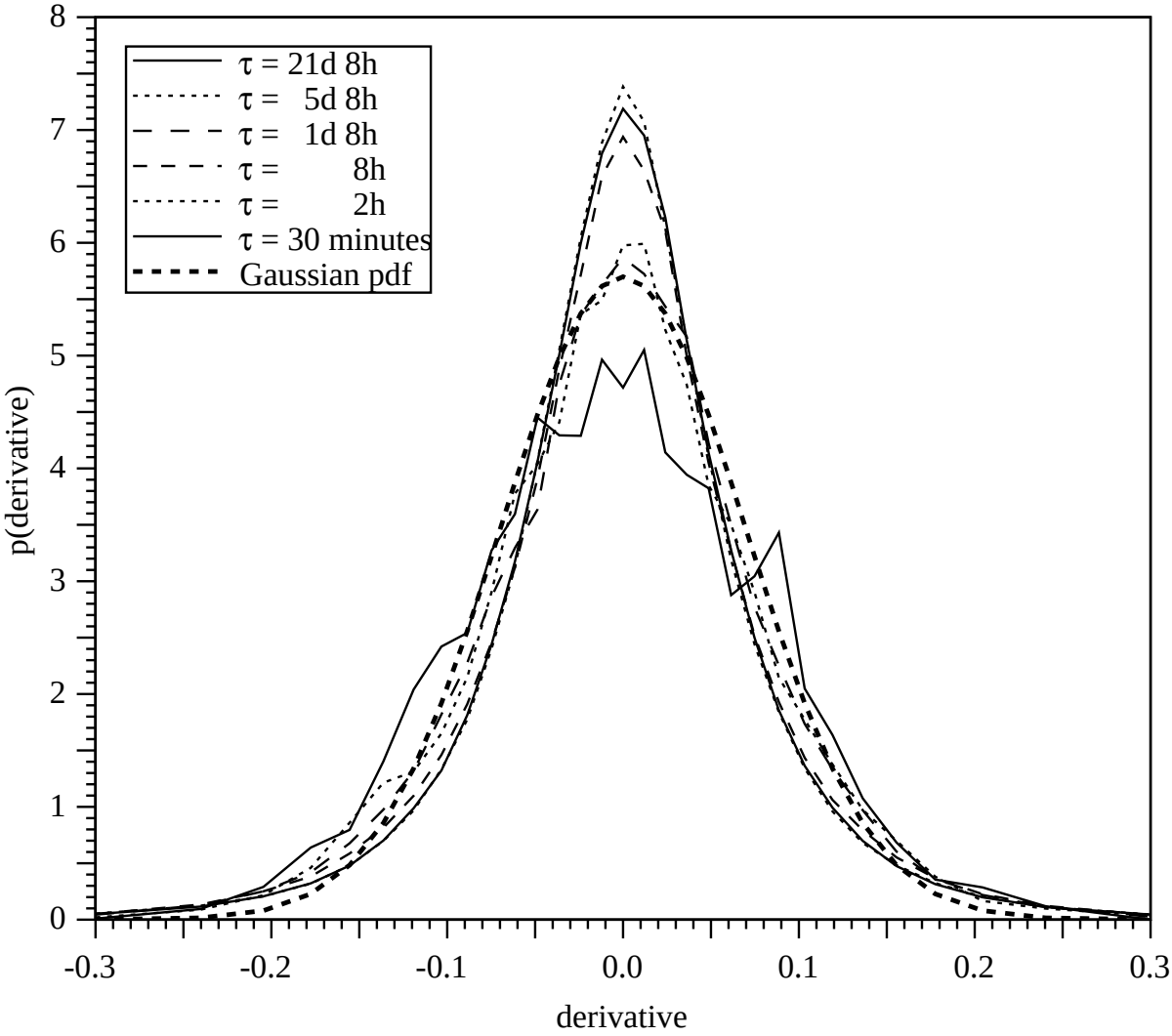


Figure 8: The annualized derivative $D[\tau, \gamma = 0.5; x]$ for USD/CHF from 1 Jan 1988 to 1 Nov 1998. The shorter time intervals τ correspond to the most leptocurtic curves. In order to discard the daily and weekly seasonality, the computations are done on the business ϑ-time scale according to (Dacorogna et al., 1993). The data was sampled every 2 hours in ϑ-time to construct the curves. The Gaussian pdf, added for comparison, has a standard deviation of $\sigma = 0.07$, similar to that of the other curves.

length τ of the moving sample used to compute the 'width'. The most common definition for the width estimator uses an L^2 norm:

$$\text{Volatility}[\tau, \tau'; z] = \left(\frac{1}{n} \sum_{i=0}^{n-1} \left(\delta \ \text{RTS}[\tau'; z] \right)_i^2 \right)^{1/2} \qquad \text{with} \quad \tau = \text{n}\, \tau' \tag{38}$$

where $\text{RTS}[\tau'; z]$ is an artificial regular time series, spaced by τ', constructed from the irregular time series z (see Section 5.3). The operator δ computes the difference between successive values (see Section 5.4). This definition suffers from a couple of drawbacks:

- For inhomogeneous time series, a synthetic regular time series must be created, which involves an interpolation scheme.
- The difference is computed with a point-wise difference. This implies some noise in the case of stochastic data.
- Only some values at regular time points are used. Information from other points of the series, between the regular sampling points, is thrown away. Resulting from this information loss, the estimator is less accurate than it could be.
- It is based on a rectangular weighting kernel, i.e. all points have constant weights of either $1/n$ or 0 as soon as they are excluded from the sample. A continuous kernel with declining weights leads to a better, less disruptive and less noisy behavior.
- By squaring the returns, this definition puts a large weight on large changes of z and therefore increases the impact of outliers and the tails of $P(z)$. Also, as the fourth moment of the probability distribution of the returns might not exist (Müller et al., 1998), the volatility of the volatility might not exist either. In other words, this estimator is not very robust. These are reasons to prefer a volatility defined as an L^1 norm:

$$\text{Volatility}[\tau, \tau'; z] = \frac{1}{N} \sum_{i=0}^{N-1} |\Delta[\text{RTS}[\tau'; z]]_i| \qquad \text{with} \quad \tau = \text{N}\, \tau' \ . \tag{39}$$

There are again many ways to introduce a better definition for inhomogeneous time series. These definitions are variations of the following one:

$$\text{Volatility}[\tau, \tau', p; z] = \text{MNorm}[\tau/2, p; \Delta[\tau'; z]] \tag{40}$$

where the moving norm MNorm is defined by eq. 32 and the differential operator Δ of eq. 33 is used. Let us emphasize that no homogeneous time series is needed, and that this definition can be computed simply and efficiently for high-frequency data because it ultimately involves only EMAs. Note the division by 2 in the MNorm of range $\tau/2$. This is to attain an equivalent of the definition 38 which is parametrized by the total size rather than the range of the (rectangular) kernel.

The variations of this definition mainly include:

- replacing the norm MNorm by a moving standard deviation MSD as defined by eq. 32. This modification subtracts the empirical sample mean from all observations of $\Delta[\tau'; z]$. This is not standard for volatility computations of prices in finance, but might be a better choice for other time series or applications. Empirically, for most data in finance such as FX, the numerical difference between taking MNorm and MSD is very small.

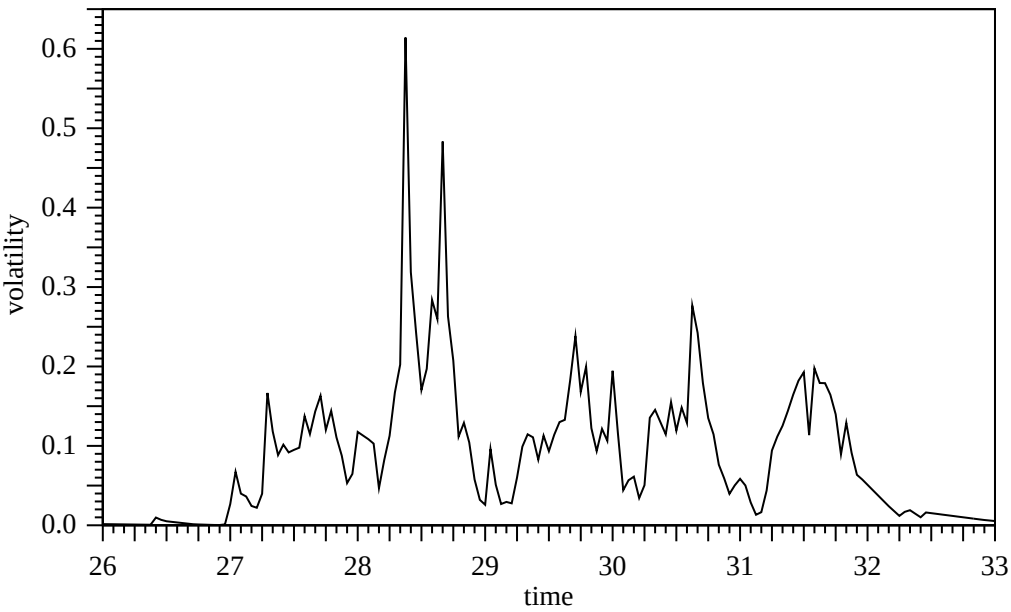


Figure 9: The annualized volatility computed as $\mathrm{MNorm}[\tau/2; D[\tau/32, \gamma = 0.5; x]]$ with $\tau =$ 1h. The norm is computed with $p = 2$ and $n = 8$. The plotted volatility has 5 main maxima correponding to the 5 working days of the example week. The Tuesday maximum is higher than the others, due to the stock crash mentionned in the introduction.

- replacing the differential Δ by a γ-derivative $D[\tau, \gamma]$. The advantage of using the gamma derivative is to remove the leading τ dependence, for example by directly computing the annualized volatility, independently of τ. An example is given by Figure 9.

Let us emphasize that the volatility definition 38 depends on the two time ranges τ and τ' and, to be unambiguous, both time intervals must be given. Yet, for example when talking about a daily volatility, the common language is rather ambiguous because only one time interval is specified. Usually, the emphasis is put on τ'. A daily volatility, for example, measures the average size of daily price changes, i. e. $\tau' = 1$ day. The averaging time range τ is chosen as a multiple of τ', of the order $\tau \geq \tau'$ up to $\tau = 1000\tau'$ or more. Larger multiples lead to lower stochastic errors as they average over larger samples, but they are less local and dampen the possible time variations of the volatility. From empirical studies, one can conclude that good compromises are in the range from $\tau = 16\tau'$ to $\tau = 32\tau'$.

On other occasions, for example in risk management, one is interested in the conditional daily volatility. Given the prices up to today, we want to produce an estimate or forecast for the size of the price move from today to tomorrow, i. e. the volatility within a small sample of only one day. The actual value of this volatility can be measured one day later; it has $\tau = 1$ day by definition. In order to measure this value with an acceptable precision, we may choose a distinctly smaller τ', perhaps $\tau' = 1$ hour. Clearly, when only one time parameter is given, there is no simple convention to remove the ambiguity.

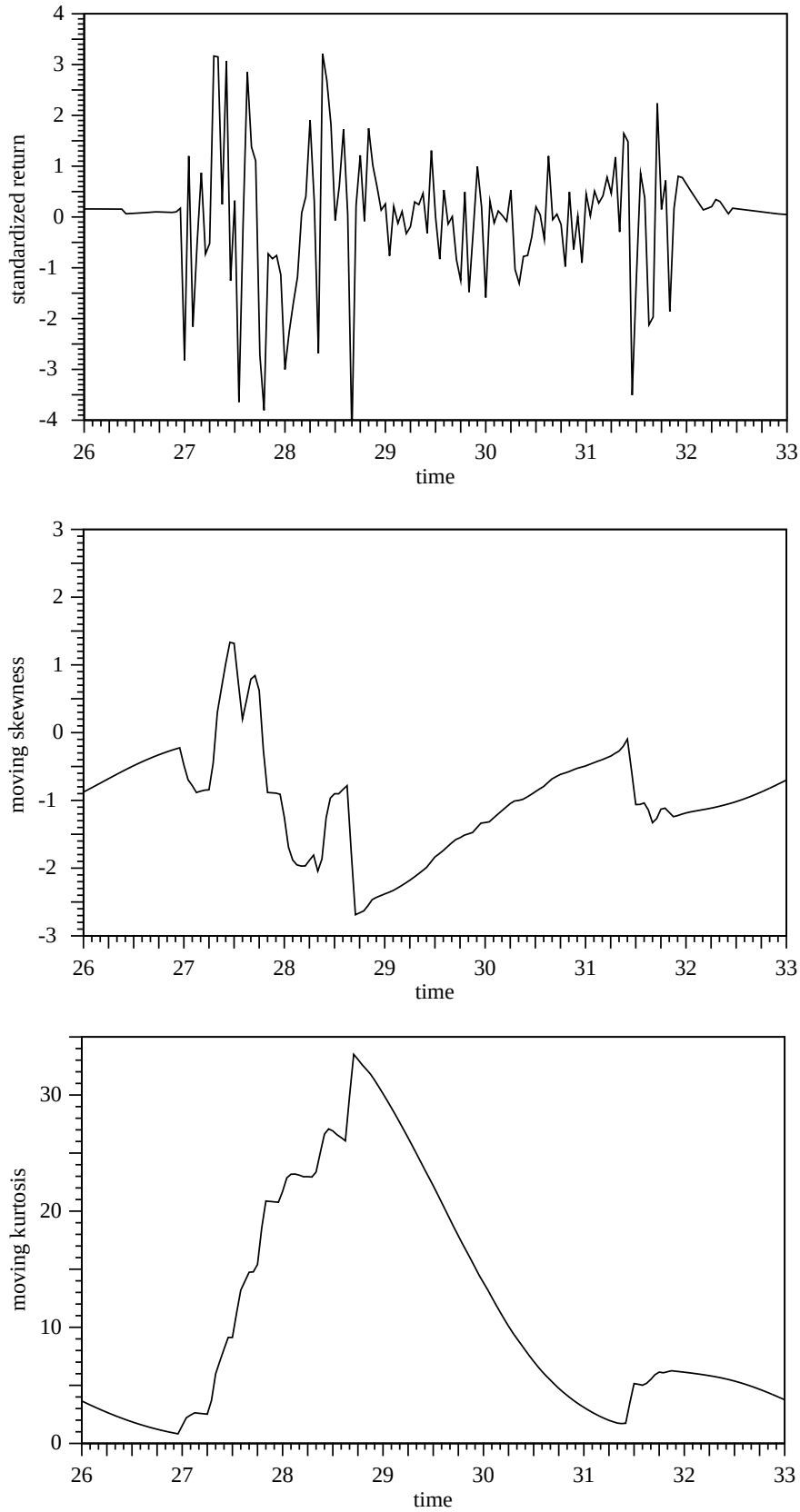


Figure 10: The standardized return, moving skewness and moving kurtosis. The returns are computed as $r = D[\tau = 15 \text{ minutes}; x]$ and standardized with $\tau_1 = \tau_2 = 24$h.

4.8 Standardized time series $\hat{z}$, moving skewness and kurtosis

From a time series z, we can derive a moving standardized time series

$$\hat{z}[\tau] = \frac{z - \mathrm{MA}[\tau; z]}{\mathrm{MSD}[\tau; z]} \ . \tag{41}$$

In finance, z typically stands for a time series of returns rather than prices.

Having defined a standardized time series $\hat{z}[\tau]$, the definitions for the moving skewness and moving kurtosis are straightforward:

$$\begin{aligned} \mathrm{MSkewness}[\tau_1, \tau_2; z] &= \mathrm{MA}[\tau_1; \hat{z}[\tau_2]^3] \ , \\ \mathrm{MKurtosis}[\tau_1, \tau_2; z] &= \mathrm{MA}[\tau_1; \hat{z}[\tau_2]^4] \ . \end{aligned} \tag{42}$$

The three quantities for our sample week are displayed in Figure 10.

4.9 Moving correlation

Several definitions of a moving correlation can be constructed for inhomogeneous time series. Generalizing from the statistics textbook definition, we can write two simple definitions:

$$\mathrm{MCorrelation}_1[\tau; y, z] = \mathrm{MA}[\ (y - \mathrm{MA}[y])(z - \mathrm{MA}[z])\ ]/(\mathrm{MSD}[y]\ \ \mathrm{MSD}[z]) \ , \tag{43}$$

$$\begin{aligned} \mathrm{MCorrelation}_2[\tau; y, z] &= \mathrm{MA}[\ \frac{(y - \mathrm{MA}[y])\ (z - \mathrm{MA}[z])}{\mathrm{MSD}[y]\ \ \mathrm{MSD}[z]}\ ] \\ &= \mathrm{MA}[\ \hat{y}\ \hat{z}\ ] \ , \end{aligned} \tag{44}$$

where all the MA and MSD operators on the right hand sides are taken with the same decay constant τ. These definitions are not equivalent because the MSD operators in the denominator are time series that do not commute with the MA operators. Yet both definitions have their respective advantages. The first definition obeys the inequality $-1 \leq \mathrm{MCorrelation}_1 \leq 1$. This can be proven by noting that $MA[z^2](t)$ for a given t provides a norm on the space of (finite) time series up to t. It happens because the MA operator has a strictly positive kernel, that acts as a metric on the space of time series. In this space, the triangle inequality holds $\sqrt{MA[(y+z)^2]} \leq \sqrt{MA[y^2]} + \sqrt{MA[z^2]}$, and, by a standard argument, the inequality on the correlation follows. With the second definition (44), the correlation matrix is bilinear for the standardized time series. Therefore, the rotation that diagonalizes the correlation matrix acts linearly in the space of standardized time series. This property is necessary for multivariate analysis, when a principal component decomposition is used.

In risk management, the correlation of two time series of returns, x and y, is usually computed without subtracting the sample means of x and y. This implies a variation of eqs. 43 and 44:

$$\mathrm{MCorrelation}'_1[\tau; y, z] = \mathrm{MA}[\ y\ z\ ]/(\mathrm{MNorm}[y]\ \ \mathrm{MNorm}[z]) \ , \tag{45}$$

$$\mathrm{MCorrelation}'_2[\tau; y, z] = \mathrm{MA}[\ \frac{y\ z}{\mathrm{MNorm}[y]\ \ \mathrm{MNorm}[z]}\ ] \tag{46}$$

where again the same τ is chosen for all MA operators.

In general, any reasonable definition of a moving correlation must obey

$$\lim_{\tau\to\infty} \mathrm{MCorrelation}[\tau; y, z] \to \rho[y, z] \tag{47}$$

where $\rho[y, z]$ is the theoretical correlation of the two stationary processes x and y. Generalizing the definition (44), the requirements for the correlation kernel are to construct a causal, time translation invariant, and a linear operator for $\hat{y}$ and $\hat{z}$. This leads to the most general representation

$$\mathrm{MCorrelation}[\hat{y}, \hat{z}](t) = \int_0^\infty \int_0^\infty dt'\, dt''\, c(t', t'')\, \hat{y}(t - t')\, \hat{z}(t - t'') \ . \tag{48}$$

We also require symmetry between the arguments $\mathrm{MCorrelation}[\hat{z}, \hat{y}] = \mathrm{MCorrelation}[\hat{y}, \hat{z}]$. Moreover, the correlation must be a generalized average, namely $\mathrm{MCorrelation}[\mathrm{Const}, \mathrm{Const}'] = \mathrm{Const}\mathrm{Const}'$, or for the kernel, $\int\int_0^\infty dt'\, dt''\, c(t', t'') = 1$. There is a large choice of possible kernels that obey the above requirements. For example, eq. 44 is equivalent to the kernel $c(t', t'') = \delta(t' - t'')\, \mathrm{ma}(\frac{t'+t''}{2})$, but other choices might be better than this one.

4.10 Windowed Fourier transform

In order to study a time series and its volatility at different time scales, we want to have a tool similar to a wavelet transform, yet adapted to causal signals. The motivation is to reveal structures of price movements related to certain frequencies. Similarly to wavelet transforms, we want a double representation in time and frequency, but we do not require to have an invertible transformation because our aim is to analyze rather than further process the signal. This gives us more flexibility in the choice of the transformations.

A simple causal kernel with such properties is like $\mathrm{ma}[\tau](t)\ \sin(kt/\tau)$. Essentially, the sine part is (locally) analyzing the signal at a frequency k/τ and the MA part is taking a causal window of range τ. As we want a couple of oscillations in the window 2τ, we choose k between $k \sim \pi$ and $k \sim 5\pi$. Larger k values increase the frequency resolution at the cost of the time resolution. To carry out this program efficiently, we compute an EMA with a complex τ; this is equivalent to including a sine and cosine part in the kernel. The nice computational iterative property of the moving average is preserved.

The first step is to study complex iterated EMAs. The kernel of the complex ema is defined as

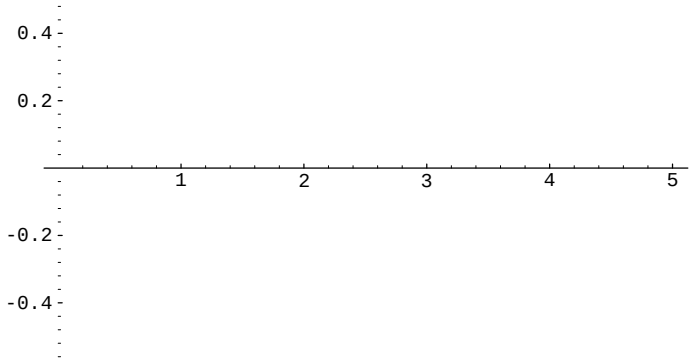
$$\mathrm{ema}[\zeta](t) = \frac{e^{-\zeta t}}{\tau} \qquad \text{where} \quad \zeta = \frac{1}{\tau}(1 + ik) \ . \tag{49}$$

where ζ is complex but τ is again a real number. The choice of the normalization factor $1/\tau$ is somewhat arbitrary (a factor $|\zeta|$ will produce the same normalization for the real case $k = 0$) but leads to a convenient definition of the windowed Fourier kernel below. By using the convolution formula, one can prove iteratively that the kernel of the complex $\mathrm{EMA}[\zeta, n]$ is given by

$$\mathrm{ema}[\zeta, n](t) = \frac{1}{(n-1)!} \left(\frac{t}{\tau}\right)^{n-1} \frac{e^{-\zeta t}}{\tau} \ , \tag{50}$$

which is analogous to eq. 26. The normalization is such that, for a constant function $c(t) = c$,

$$\mathrm{EMA}[\zeta, n; c] = \frac{c}{(1 + ik)^n} \ . \tag{51}$$


0.4
0.2
1
2
3
4
5
-0.2
-0.4

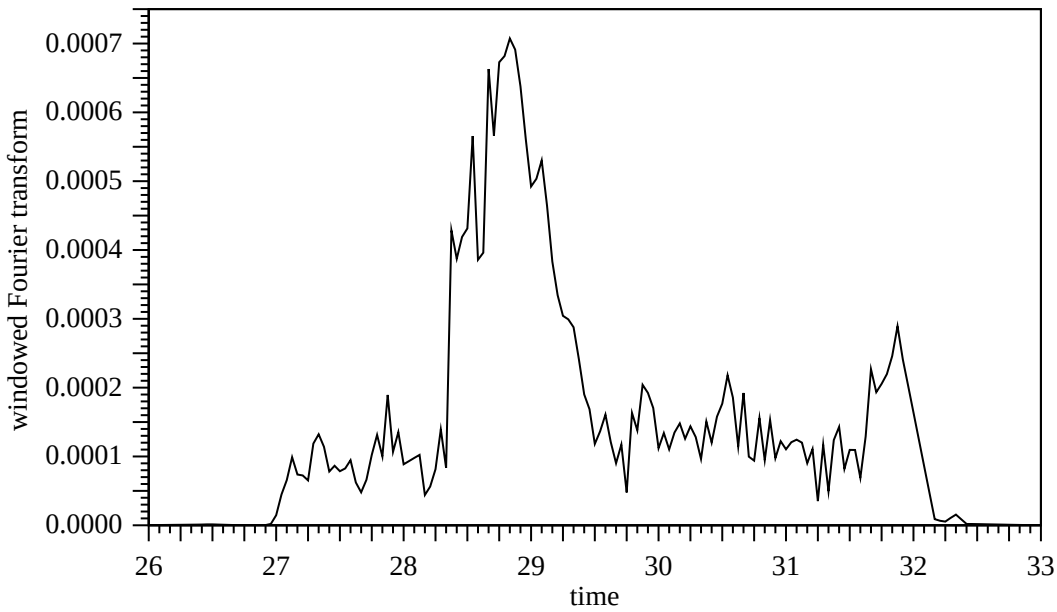


Figure 12: The normed windowed Fourier transform, with $\tau = 1$ hour, $k = 6$ and $n = 8$.

In eq. 54, we are only interested in the amplitude of the measured frequency; by taking the absolute value we have lost information on the phase of the oscillations.

Windowed Fourier transforms can be computed for a set of different τ values to obtain a full spectrum. However, there is an upper limit in the range of computable frequencies. Results are reliable if τ clearly exceeds the average time interval between ticks. For τ values smaller than the average tick interval, results become biased and noisy; this sentence applies not only to windowed Fourier transforms but also to most other time series operators.

Figure 12 shows an example of the normed windowed Fourier transform for the example week. The stock market crash is again nicely spotted as the peak on Tuesday 28. Using our computational tool box of operators, other quantities of interest can be easely derived. For example, we can compute the relative share of a certain frequency in the total volatility. This would mean a volatility correction of the normed windowed Fourier transform. A way to achieve this is to divide NormedWF by a suitable volatility, or to replace z by the standardized time series $\hat{z}$ in eq. 54.

5 Microscopic operators

As discussed in Section 2.2, it is in general better to use macroscopic operators because they are well behaved with respect to the sampling frequency. Yet, some microscopic operators are interesting because they allow to extract tick-related information at the highest possible frequency. An example of such an operator is the microscopic volatility defined below. The computation of the tick frequency requires (by definition) microscopic operators. We also want to extend to inhomogeneous time series the usual operators applied to homogeneous time series, such as the shift operator.

5.1 Backward shift operator $\mathcal{B}$, or lag operator L

The backward shift operator $\mathcal{B}$ shifts the value of the time series by one event backward $\mathcal{B}[z]_i = (t_i, z_{i-1})$, but the time associated to each event is not changed. Some authors use the equivalent lag operator L instead. It shifts the time series values but leaves the time part untouched. The inverse operator $\mathcal{B}^{-1}$ will shift the series forward. It is well defined for regular and irregular time series. Only for a homogeneous time series spaced by δt, this operator is equivalent to a time translation by $-\delta t$ (followed by a shift of the time *and* value series by one event with respect to the irrelevant index i).

5.2 Time translation operator $\mathcal{T}$

The operator $\mathcal{T}$ translates the time series by δt forward $\mathcal{T}[\delta t; z]_i = (t_i + \delta t, z_i)$, namely it shifts the time part but leaves the time series values untouched. Note that for an inhomogeneous time series, this operator defines a series with another set of time points.

5.3 Regular Time Series $\mathrm{RTS}[t_0, \delta t]$

From the time series z, irregularly spaced in time, the operator RTS constructs an artificial homogeneous time series at times $t_0 + k\delta\, t$, regularly spaced by δt, rooted at t_0. This involves an interpolation scheme, for example 'previous point' or 'linear interpolation'. Depending on the interpolation scheme, this operator can be causal or not. The regular time series can also be constructed as being regular on a given business time scale rather than in physical time. A simple example is provided by daily data in a business time which omits the weekends and holidays. Such a time series is inhomogeneous in physical time but homogeneous in this particular business time. The RTS operator allows to move from inhomogeneous data to the familiar world of homogeneous time series and to work with the usual tools. For many computations, it is mandatory to have homogeneous data, for example when modeling financial data with ARMA or GARCH processes. Another example is the computation of empirical probability distributions. Such computations are done with a smooth version of the formula

$$p(z) = \frac{1}{T} \int_0^T dt\, \delta(z - z'(t)) \tag{55}$$

with $z'(t)$ the (continuously interpolated) empirical data. In the time integral, a measure can be added, or the integral can be evaluated in business time to account for the seasonalities. The evaluation of the time integral is computationally heavy, and it is much simpler to generate a regular time series and to use the familiar binning procedure in order to obtain a histogram of z. Note also that a moving probability distribution can be defined by replacing the time integral by a MA operator (see the remark at the end of Section 4.3).

5.4 Microscopic return r and the difference operator δ

From the tick-by-tick price time series, the microscopic return for a quote is defined as $r_i = x_i - x_{i-1}$. This return can be attributed to one quote, even if, strictly speaking, it is related to two subsequent quotes. Note that this is a 'microscopic' definition that involves

neither a time scale nor an interpolation scheme. Using the backward shift operator $\mathcal{B}$, the return time series can be defined as

$$r = x - \mathcal{B}[x] = (1 - \mathcal{B})\ x = \delta x \tag{56}$$

where we have defined a microscopic difference operator

$$\delta = (1 - \mathcal{B})\ . \tag{57}$$

where the operator δ should not be confused with the δ function also used in some parts of this paper. The lag n difference operator is defined as $\delta[n] = (1 - \mathcal{B}^n)$.

5.5 Microscopic derivative ∂

The microscopic derivative operator is defined as

$$\partial[\delta t_0]x_i = \frac{x_i - x_{i-1}}{\delta t_0 + t_i - t_{i-1}} = \left.\frac{\delta x}{\delta t_0 + \delta t}\right|_i\ . \tag{58}$$

The constant δt_0 regularizes the expression when $t_i = t_{i-1}$. A reasonable value of δt_0 must be small, the actual choice depends on the application.

Similarly to the microscopic γ-derivative, a microscopic γ-derivative can be defined

$$\partial[\delta t_0, \gamma]x = \frac{\delta x}{(\delta t_0 + \delta t)^\gamma}\ . \tag{59}$$

The best parameters should follow a study yet to be done for the random process of x. The constant δt_0 regularizes the expression when $t_i = t_{i-1}$.

These derivatives are potentially very noisy and can be averaged. In general, the macroscopic derivative D (definition 36) seems more relevant for applications to random processes.

5.6 Microscopic volatility

The microscopic volatility is simply the norm of the microscopic derivative:

$$\text{MicroscopicVolatility}[\tau; z] = \text{MNorm}[\tau/2; \partial z]\ , \tag{60}$$

which also depends on the implicit parameters δt_0 and γ of ∂z. Let us emphasize that this definition does not require a regular time series (and that it is not an MA of the macroscopic definition of volatility!). In a way, this definition uses all the information available on the process z. The constant τ controls the range on which the volatility is computed. The microscopic volatility for our standard example week is displayed in Figure 13.

5.7 Tick frequency f and activity A

The tick frequency $f(t_i)$ counts the number of ticks per time unit. At time t_i, it is defined as (see for example (Guillaume et al., 1994))

$$f[T](t) \quad = \quad \frac{1}{T} N\{t_i \mid t_i \in [t - T, t]\} \tag{61}$$

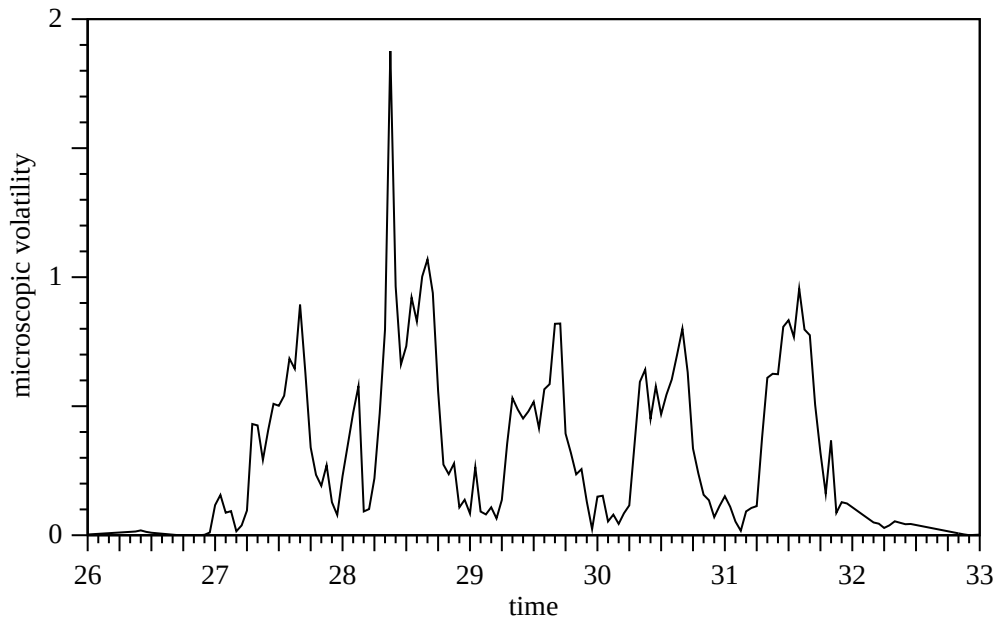


Figure 13: Microscopic volatility, computed with $\gamma = 0.5$, $\delta t_0 = 0.001$ seconds, the time interval expressed in years (annualized), and $\tau = 1$h.

where $N\{t_i\}$ counts the number of elements in a set and where T is the sample time interval during which the counting is computed. The tick frequency has the dimension of an inverse time and is expressed in units such as ticks/minute or ticks/day.

This simple definition has some properties that may not always be desired:

- This formula is computationally cumbersome when computed on a moving sample, especially for large T.
- It is an average over a rectangular window. We often prefer moving averages whose kernel (= weighting function) fades more smoothly in the distant past.
- If no quotes are in the interval spanned by T, this definition will give $f = 0$. A related problem is the unusable limit $T \to 0$ if one wants to measure an instantaneous quote rate.

For these reasons, we prefer the definition 63 below. The tick rate is defined as

$$a[\delta t_0]_i = \frac{1}{\delta t_0 + t_i - t_{i-1}} \ . \tag{62}$$

The tick rate has the same dimension as the tick frequency. This definition has the advantage of being related only to the time interval between two subsequent ticks. Following eq. 62, an activity can be attributed to one tick, analogously to a return that is attributed to the i'th tick by $r_i = x_i - x_{i-1} = (\delta x)_i$.

The activity A is the average tick rate during a time interval τ:

$$A[\tau; z] = \mathrm{MA}[\tau/2; a[z]] \ . \tag{63}$$

In order to avoid the spurious singularity when $t_{i-1} = t_i$, the MA operator has to be evaluated with the next-point interpolation (see formula 24). This makes the computation

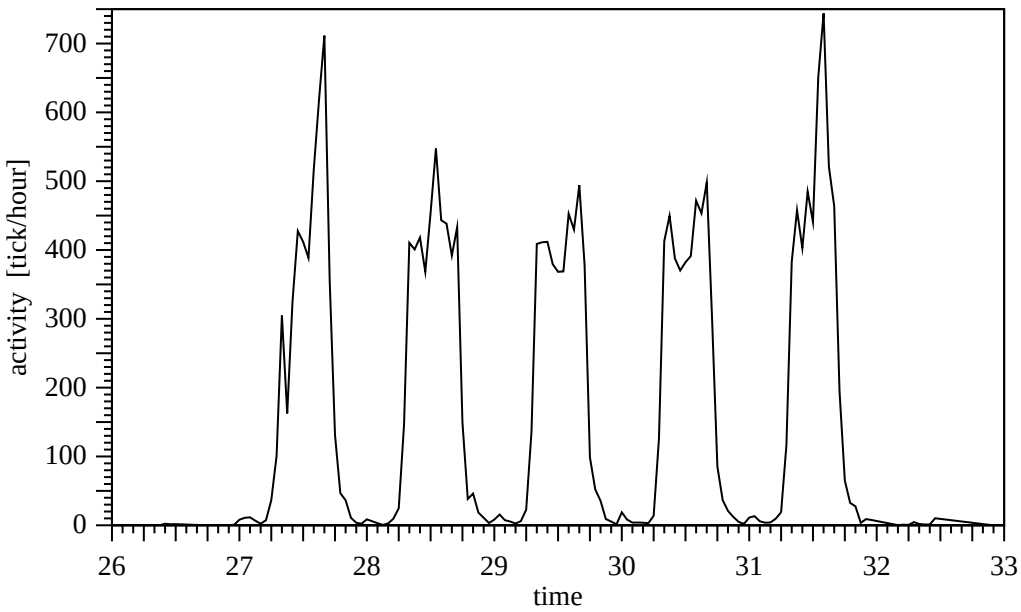


Figure 14: The activity expressed in ticks/hour, computed with $\tau = 1$h. The 5 working days of the example week can be clearly seen.

numerically stable even when an extremely small value of δt_0 in eq. 62 is chosen. Figure 14 shows the behavior of A in our example week.

At first glance, the activity A looks rather different from the tick frequency f. Yet, an interesting feature of this definition is to be equivalent to the tick frequency f when the MA operator is a rectangular moving average with $\tau = T/2$ (this is easy to prove by computing the integral of the piecewise constant function a). However, the activity A has some advantages: it is much simpler to compute on a moving window and the weighting function of the past can be controlled through the choice of the MA kernel.

References

Dacorogna M. M., Müller U. A., Nagler R. J., Olsen R. B., and Pictet O. V., 1993, *A geographical model for the daily and weekly seasonal volatility in the FX market*, Journal of International Money and Finance, **12**(4), 413–438.

Granger C. W. J. and Newbold P., 1977, *Forecasting economic time series*, Academic Press, London.

Guillaume D. M., Dacorogna M. M., Davé R. D., Müller U. A., Olsen R. B., and Pictet O. V., 1994, *From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets*, Internal document DMG.1994-04-06, Olsen & Associates, Seefeldstrasse 233, 8008 Zürich, Switzerland.

Hamilton J. D., 1994, *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Müller U. A., 1991, *Specially weighted moving averages with repeated application of the EMA operator*, Internal document UAM.1991-10-14, Olsen & Associates, Seefeldstrasse 233, 8008 Zürich, Switzerland.

Müller U. A., Dacorogna M. M., and Pictet O. V., 1998, *Heavy tails in high-frequency financial data*, published in the book "A practical guide to heavy tails: Statistical Techniques for Analysing Heavy Tailed Distributions", edited by Robert J. Adler, Raisa E. Feldman and Murad S. Taqqu and published by Birkhäuser, Boston 1998, 55–77.

Pictet O. V., Dacorogna M. M., and Müller U. A., 1998, *Hill, bootstrap and jackknife estimators for heavy tails*, published in "A practical guide to heavy tails:"Statistical Techniques for Analysing Heavy Tailed Distributions", edited by Murad Taqqu and published by Birkhäuser, 283–310.

Pictet O. V., Dacorogna M. M., Müller U. A., Olsen R. B., and Ward J. R., 1992, *Real-time trading models for foreign exchange rates*, Neural Network World, **2**(6), 713–744.

Priestley M. B., 1989, *Non-linear and non-stationary time series analysis*, Academic Press, London.

Zumbach G. O., Dacorogna M. M., Olsen J. L., and Olsen R. B., 1998, *Introducing a scale of market shocks*, Internal document GOZ.1998-10-01, Olsen & Associates, Seefeldstrasse 233, 8008 Zürich, Switzerland.