

Time aggregation problems in Financial Time Series

David Ceballos Hornero and M^a Teresa Sorrosal i Forradellas

Universitat de Barcelona

Department of Actuarial, Financial and Economical Mathematics

Av. Diagonal 690 - 08034 - Barcelona (Spain)

Tel. +34 93 4021951 Fax +34 93 4021953

ceballos@eco.ub.es

sorrosal@eco.ub.es

ABSTRACT: The aim of this communication is to analyse the process of estimation of the adequate time unit in financial series. The problem in time aggregation is the loss of information, that will be acceptable if it is possible to fit a financial model. We present an example of time aggregation problems in Financial Time series over the daily Deutsche Aktienindex (DAX) using three comparative methods: descriptive statistics, Fourier analysis and Artificial Neural Networks.

KEYWORDS: time aggregation, financial time, statistics, Fourier analysis, ART networks.

INTRODUCTION.

Time is always the basic introductory concept in all dynamic construction. With time we show two ideas: an understanding of variability existing in the World (changes) and a quantitative description by comparison with the variability of a reference object (clocks) [Levich, A.P. 1995]. This second idea connects with the measure of time, which is a physical sense of time. But researchers are not always pleased with a physical sense of time because, for the first idea, they understand changes according to the specificity of each Science. They use a different “clock” from a physical process. Also there are authors that analyse an active concept of time [Oleinik, V.P. et al. 2000].

In Finance, time is not only useful to order data, but also to calculate the earning rate of decisions. Although some authors have defended a specific financial time, like [Guitton, H. 1970] and [Ceballos, D. 2001], it is usual to work in Finance with a physical measure of time. In this way, financial data are compiled in time series according to an external and physical scale like astronomical or atomic clocks. Moreover, in financial analysis, an

operating and mathematical time is used to guarantee both robustness and credibility of results and the followed methodology.

A physical time scale not always can be representative of the dynamics of the studied financial series. Then, in order to estimate a significant model it is necessary to change the periodicity of data. In this way, it is possible to show and to explain financial dynamics. This change of time unit has as a result a loss of information and the synthesis of new information.

The subject of time unit or time aggregation problem in financial time series is the content of this communication. We structure the paper in five parts. First of all, we explain time aggregation problems in Finance, and afterwards we present three methods to analyse time series and its results. These methods are Descriptive Statistics, Fourier Analysis and Artificial Neural Network. With them we estimate the form, the periodicity and the clustering of data, respectively. Finally, we comment the results in the conclusion.

TIME AGGREGATION PROBLEMS.

Time aggregation implies a loss of information about the underlying data processes. On contrary, it generates a more smoothing process, which can be reduced to a linear and understanding model. That trade-off is common in economic data. In financial data we observe scaling properties, that is to say, data have a similar behaviour at different scales of time. This property is normally justified by fluctuations and big jumps of financial series. Financial series are not smoothing at any periodicity.

Although it is difficult to estimate a linear model for forecasting in Finance, it is possible to fit models in order to explain and understand financial phenomena. For example, some authors use econometric models based in Box-Jenkins methodology as ARCH, GARH or EGARCH, which work with a variance that is function of time. Other ones work with chaotic models as [Mandelbrot, B.B. 1997] and [Peters, E.E. 1991] do. Finally, it is possible to adjust a probability distribution to the financial data [Masoliver, J. et al. 2000 and Ghasghaei, S. et al. 1996].

In all this possibilities, the researcher works with the initial periodicity of data, normally daily or one of high-frequency. The fact that data are ordered in seconds, minutes, quarters, days, weeks does not imply that the relevant periodicity for financial information is the initial one. It is true that the highest-frequency data contain the others. But knowledge is a synthesis and loss of information is justified by the possibility of prediction or explanation.

From a mathematical perspective, the properties to predict, explain or understand the dynamics of a time series are: stability, repetition or clustering of data.

Stability of data means that parameters and descriptors of the time series are constant in time, which means whether or not data have the same properties at different time intervals and at different periodicity. This is analysed in the next section of descriptive statistics.

We apply Fourier analysis to study repetition of data. Whether data can be calculated from its history or not. This analysis consists in representing data by a sum of sines or cosines.

In clustering, we search through an Artificial Neural Network if we can form some groups in order to explain the dynamics of a time series.

DESCRIPTIVE STATISTICS.

As it is known, descriptive statistics consists in a set of indicators of global data. Global data are not ordered by time, but they are ordered by value. Descriptive statistics is the study of distribution and characteristics of data, without considering the sequence of the time series. It is supposed that data are values of a random variable and then, time order is not so important as in a deterministic function.

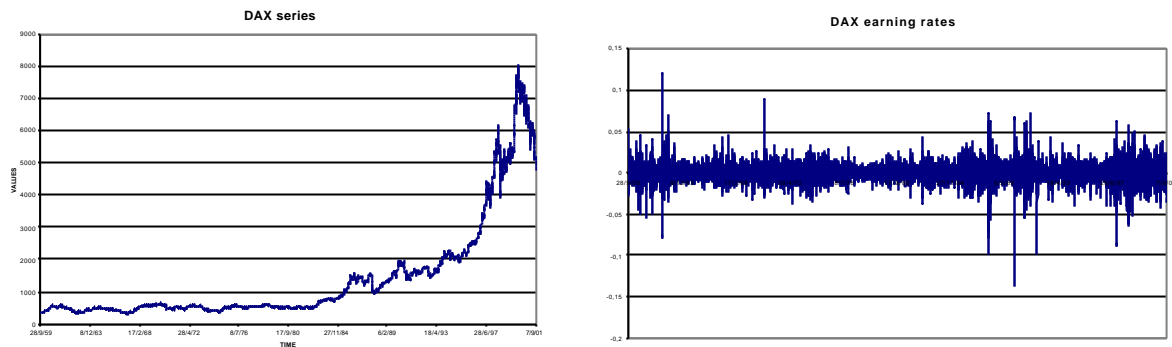
In this case, problem that time poses is double: the analysis time scale and the temporal interval, because according to the analysis period, results can change. Only if time series is stationary, descriptive statistics are constant at different epochs. This means that the estimated model will not be robust if data indicators are not stable in time.

With descriptive statistics we can know the central values of the data (mean and median), its dispersion (variance), its rank (maximum and minimum), its asymmetry, its bell form (kurtosis). These indicators are only significant when they are more or less constant in time, that is to say, they are stationary. This means that the analysis period is indifferent because results do not change. Parameter values and underlying structure are the same.

In Finance, generally, only earning rates are taken into account because of level data are rising in time, what means that one earns always long-dated and in average. Financial series present, commonly, a dependence on the accumulated value, but their increments are independent by its high volatility.

We study earning rates of the Deutsche Aktienindex (DAX) from 1959 to 2001. The original data are daily and we work with time scale from daily to forty years. Earning rates are calculated by the logarithm of the ratio between two consecutive data.

We can observe that this time series is not stationary neither in mean nor variance at any time scale (periodicity) because these indicators change its value in a significant way at different epochs. When time scales are higher than one year, changes in mean and variance are not important for different epochs.



Variance and mean increase in the same proportion as periodicity of data, as theory affirms about random variables. However, in the case of time scales higher than one year, variance is significantly lower than the theoretic value.

In relation to other indicators, the DAX series shows at every time scales and for large time intervals:

- A symmetry in average, which is close to zero.
- A variable kurtosis. From daily to monthly periodicity the observed distributions present leptokurtosis and other time scales are near a normal distribution.
- The rank is stable, except for a periodicity higher than two months.

Periodicity	Mean	Median	Variance	Kurtosis	Asymmetry	Rank
1	0.000248	0.000299	0.000122	8.951662	-0.258514	0.257099
2	0.000496	0.000461	0.000267	5.913127	-0.257873	0.288405
3	0.000745	0.000585	0.000386	4.150960	-0.216604	0.273686
4	0.000994	0.001386	0.000540	4.375836	-0.384512	0.273984
5	0.001242	0.001804	0.000671	4.235108	-0.332458	0.308982
6	0.001490	0.001707	0.000746	2.081986	-0.131807	0.278389
7	0.001739	0.001987	0.000881	4.286116	-0.324431	0.369844
10	0.002484	0.004028	0.001275	1.841813	-0.276839	0.302241
12	0.002981	0.004273	0.001569	1.758929	-0.450347	0.326662
14	0.003478	0.003428	0.001731	2.577454	-0.485333	0.381627
15	0.003726	0.005831	0.002077	4.437850	-0.844077	0.480121
20	0.004968	0.005739	0.002565	1.251348	-0.329417	0.367568
21	0.005216	0.005284	0.003086	9.361563	-1.252645	0.627646
25	0.006210	0.005146	0.003284	1.229633	-0.078726	0.433571
30	0.007452	0.008328	0.004003	1.050498	-0.491243	0.380118
42	0.010433	0.011347	0.005713	2.455912	-0.652296	0.551099
50	0.012420	0.014798	0.007027	2.231231	-0.494415	0.594987

60	0.014904	0.013162	0.008004	0.735315	-0.234852	0.559993
100	0.024839	0.025969	0.015026	0.724270	-0.130645	0.695272
250	0.062099	0.068624	0.037420	-0.554224	-0.209546	0.763829
500	0.124197	0.070598	0.036265	-0.283407	0.332712	0.763948
1050	0.260814	0.193035	0.085611	1.447863	1.211871	0.955008
2625	0.652035	0.764614	0.380154	-2.225163	-0.632583	1.322226

In our opinion DAX series is not random because of all these time scale-dependent changes and stability. Therefore, we reject the adjust of a modified gaussian distribution with fat tails. Also, it can be observed that the dimension of DAX series is fractal, for example, through R/S analysis [Peters, E.E. 1991]. A fractal dimension means that time series has a fluctuating evolution but it is not totally random. It has a hidden deterministic dynamics [Ceballos, D. 2001] and therefore it is chaotic, deterministic but with random appearance.

We propose an extension of the statistical analysis with a specific financial time. Financial time is the time of Stock Markets, a short-term time of speculation and fluctuation. This specific time measures jumps in prices, which are common in financial series. Financial time allows us to incorporate an order in data because between two consecutive jumps of the same type, there are other data. In this way, we follow the definition of financial time that is explained in [Ceballos, D. 2001]. We classify data according to its value in earning rates lower than -2%, between -2% and -0.5%, between losses of 0.5% and earnings of 0.5%, earnings between 0.5% and 2% and earning rates higher than 2%.

With this classification of data we observe that the adjust to the independent term is better because adjusted R^2 (regression coefficient) has values between 0.60 (yearly) and 0.82 (daily) according to the time scale (lower than one year) and time interval (higher than two years), while with only one category of jump the adjusted R^2 was not higher than 0.1.

The most significant category is earning rates higher than 2%. This shows data change around five levels, which are the corresponding means of each category. The fact that results are similar until a time scale of one year can be explained by the stability of the rank.

Analysing data from financial time, we observe that the categories more significant are earning rates between -2% and -0.5% and between 0.5% and 2%. These two categories are stable and homogeneous until a periodicity of one month. They show a small variance and a stable average distance between two consecutive data of three – five periods. Categories of earning rates lower than -2% and higher than 2% show a stable average temporal distance near thirty days. These results are only valid for a periodicity lower than one month.

FOURIER ANALYSIS.

Fourier analysis is based in the result that every continuous function can be expressed as a sum of sines or cosines. If that sum has a small quantity of elements, then the period of these trigonometric functions represents the repetitive cycles that the initial function contains. The function would be multiperiodic.

For estimating the algebraic approach Fourier Transform is applied. Fourier Transform connects time (t) and frequency (w).

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cdot \cos(2 \cdot \pi \cdot k \cdot t) + b_k \cdot \sin(2 \cdot \pi \cdot k \cdot t)) \quad \text{Fourier series}$$

$$F(w) = \frac{1}{2 \cdot \pi} \int_{-\infty}^{+\infty} f(t) e^{-i \cdot w \cdot t} \cdot dt = \int_{-\infty}^{\infty} \cos(2 \cdot \pi \cdot w \cdot t) \cdot f(t) \cdot dt \quad \text{Fourier Transform}$$

When one works with observed data and not with an algebraic function, one uses numerical approach of the Fast Fourier Transform (FFT). This approach is a fast algorithm to calculate the Fourier Transform of the time series. It is an efficient algorithm for a number of data that is an integer two-power and with an homogeneous periodicity.

In this case, time is relevant because order of data is fundamental in the definition of cycles, and then, as in financial time there are two important variables: time and value. In the next table we present the obtained results. We have estimated the FFT for the earning rates of DAX at different time scales, using the programme ORIGIN 5.0.

Periodicity in days	Cycle 1 in days	Cycle 2 in days	Cycle 3 in days	Cycle 4 in days	The highest coefficient
1	909.09	224.2	8.646	3.903	0.00052
2	909.09	224.5	8.646	4.16	0.00043
3	909.09	224.4	8.341		0.00043
4	911.16	233.7	20.90		0.00043
5	910.50	223.8	25.99	16.38	0.00043
6	910.84	224	15.10		0.00043
7	910.30	224	25.77		0.00049
10	900.09	223.8			0.00043
12	901.58	223.9			0.00043
14	903.23				0.00043
15	900.36				0.00043
20	902.53				0.00043
21	903.23				0.00043
25	902.20				0.00043
30	901.98				0.00042
42	903.03				0.00043
50	900.25				0.00041
60	902.66				0.00040

Important cycles are not very significant because its coefficients are small and they do not explain a significant percentage of variability of earning rates. But we observe five stable cycles in time scales between daily and quarterly: about 900 working days or near 3.5 years, 224 working days, near 15 working days or 3 weeks, between 8 and 9 working days and near 4 working days (less than a week). This cycles remain if we introduce a noise in data, which shows that they are not spurious. Also, they remain for different initial data.

Autocorrelation coefficients are small though they rise for the five significant cycles.

Finally, if we consider financial time, dividing data in the five categories the commented cycles disappear. This shows that data are not periodic.

ARTIFICIAL NEURAL NETWORKS.

In order to observe whether among historical DAX values there are hidden temporal patterns or not, we can use (among other possibilities) instruments from connectionist approach. This method is based in a system of information transference between simple elements that are connected within them.

In a first stage, we will choose the type of Artificial Neural Network (ANN) that fits better our problem. Searching equivalent temporal intervals in the studied series can be understood as grouping intervals with a similar statistical behaviour along time. The chosen ANN, then, has to be capable of making groups, that is to say, its purpose has to be clustering, because at the beginning we do not know the number or the kind of resulting categories.

We have decided to use ART (Adaptive Resonance Theory) networks, although other ones like Kohonen maps may be equally useful. We use this kind of ANN because its particular characteristics are appropriate for our example. In this sense, ART networks are designed with an aim: to make groups with a set of patterns or inputs according to some features that the user has established. They undergo unsupervised learning (but supervised learning is also possible), and it is competitive and on line. Contrary to others, an ART network combines the properties of plasticity and stability, and this means that future inputs (future DAX values) can be analysed with the same ANN without losing all the information that it has already stored. This fact is due to a multilayer architecture with feedforward as well as feedback connections between input and output layers.

Finally, the existence of a parameter called vigilance parameter (ρ) that acts as a threshold in one node of the ANN, allows us to influence in the number of resulting groups. If the value of ρ increases, then the homogeneity between elements of the same group also increases. But, in the same way, the number of groups that the ANN creates is bigger. Decreasing the value of the vigilance parameter has the contrary effects, and in the extreme case that $\rho = 0$ all the data is stored in one only prototype.

So, we can sum up how an ART network works in the following way. The first element enters in the system and it becomes the first group and its representative vector. The second input is compared with the prototype of the winning group in the competitive process. If the similarity is sufficient (it is measured by a function of distance and compared with the parameter ρ), the new element is adhered to this group and its prototype is modified in order to collect some characteristics of all the elements that belong to it. If differences are too important, a new group is created. This process continues for the rest of inputs. Because of the variation in prototypes can alter the composition of groups, it will be necessary to introduce again all data until the ANN achieves stability.

Despite the first class of ART networks, designed by S. Grossberg and G. Carpenter, only works with binary inputs, now we dispose of ART2 networks which works with continuous values and we are able to combine them with fuzzy logic also (Fuzzy ART).

In our example, first of all, we have convert the 10.500 values of the daily DAX series from 1959 to 2001 into 33 input vectors. The patterns have been obtained from the DAX return rates at different scales: 1, 2, 3, 4, 5, 6, 7, 10, 12, 14, 15, 20, 21, 25, 30, 42, 50, 60, 100, 250, 500, 1050 and 2625 days, and we have added the series of returns for days of the week and those that we have built using the daily values that exceed 2%, between 0 and 2%, lower than 0%, between -1 and 2% and lower than -1%.

Every series has been simplified to a vector of statistical characteristics that are relevant to explain the behaviour of the whole temporal series. The result is, for every one of this 33 series, a vector of 7 components: mean, median, standard deviation, kurtosis, coefficient of asymmetry, and the value of the autocorrelation function for one and two delays. All components have been normalised to obtain values inside the interval [0, 1].

Results are summed up in the following table, where only groups with more than one element appear:

$\rho = 09 \rightarrow 22$ groups	$\rho = 08 \rightarrow 13$ groups	$\rho = 07 \rightarrow 9$ groups
1 day, Mondays	(*) 1, 12, 14, 20, 30 days, Thursdays, between -1 and 2%	1, 12, 14, 15, 20, 25, 30, 42, 60, 100 days, Mondays, Thursdays, between -1 and 2%, between 0 and 2%
2, 3, 4 days, Tuesdays	2, 3, 4, 5 days, Tuesdays	2, 3, 4 days, Tuesdays
6, 10, 15 days	6, 10, 42, 50 days	5 days, Wednesdays
7 days, Fridays	7, 15, 21 days, Fridays	6, 7, 10, 21, 50 days, Fridays
12, 14, 30, 42 days	25, 100 days, between 0 and 2%	500, 1050 days
20, 25 days	Mondays, Wednesdays	Lower than -1%, lower than 0%
	Lower than -1%, lower than 0%	

Taking $\rho = 08$ as an appropriate value for our study, we can expose like relevant features: (i) that the series at one day and the series at 30 days belong to the same category, and (ii) the separation between the behaviour of the series at one day in reference with the series at 2, 3 and 4 days, that are put together in the same group.

The first fact tells us that return series with one month and a half of frequency has a statistical behaviour similar to the daily series. Consequently, the saving for working with the first one and not with the second one is not punished losing an important amount of information. On the other hand, in (ii) it is surprising the different behaviour between the series at any frequency shorter than a week and the daily series.

CONCLUSION.

We have verified that the analysis of a financial series is complex because its variability is very high in time. As a result, instability is at different scales of time. This outcome can be stated as a sign of fractal dimension and chaos.

Time aggregation problem is the trade-off between loss of information when we change periodicity and the best understanding of data evolution thanks to the estimated model. In Finance, it is showed, through the studied example, that the analysis of this problem is a hard work and the “good” periodicity is surely a high-frequency one, that is to say, intraday data. Estimating what is the significant time scale of data is not intuitive, but it can help us to understand the underlying dynamics.

This analysis about daily DAX series from 1959 to 2001 shows that the financial series does not present any scale of time with a stability of its descriptive statistics. Neither significant regularity nor clustering of time scales help us to know the best time unit. But we can conclude that daily periodicity is worse than the other time scales of the week. These other time scales have an homogeneous and more stable behaviour. That means that

periodicity of three or four days is preferable to daily data, although results do not significantly improve .

Moreover, using a specific financial time, which divides values in five categories, improves outcomes, but not in a sufficient way.

A financial explanation of the outcomes lies in the fact that financial behaviour is more random in daily data than in earning rates of three or four days. In this case, forecasts would be more believable.

In conclusion, through there is evidence that in economical data a higher periodicity of data smoothes the time series, with financial data this result can not be generalised.

BIBLIOGRAPHY.

- Ceballos Hornero, D. 2001. "An Approach to Financial Time". 4th Italian-Spanish Conference on Financial Mathematics. Alghero.
- Freeman, J.A., Skapura, D.M. 1993. Redes Neuronales, Algoritmos, Aplicaciones y Técnicas de Programación Addison Wesley. Buenos Aires.
- Ghashghaie, S. et al. 1996. "Turbulence cascades in foreign-exchange markets". *Nature* 381, pp. 767-770.
- Guitton, H. 1970. A la recherche du temps économique. Fayard. Paris.
- Körner, T.W. 1989. Fourier Analysis. Cambridge University Press . Cambridge.
- López Cachero, M. 1993. Fundamentos y métodos de estadística. Pirámide. Madrid.
- Mandelbrot, B.B. 1997. Fractals and Scaling in Finance: Discontinuity, Concentration and Risk. Springer. New York.
- Masoliver, J. et al. 2000. "A dynamical model describing stock market price distribution" *Physica A* 283, pp. 559-567.
- Oleinik, V.P. et al. 2000. "Time, what is it? Dynamical Properties of Time" *Physical Vacuum and Nature* 5, pp. 65-82.
- Peters, E.E. 1991 Chaos and Order in the Capital Markets. John Wiley&Sons. New York.
- Priestley, M. 1981. Spectral analysis and time series. Academic Press. London.
- Levich, A.P. [Editor]. 1995. On the way to understanding the Time Phenomenon: the constructions of time in Natural Science. Part I. World Scientific. New York.
- Weenink, D. 1997. "Category ART: A Variation on Adaptative Resonance Theory Neural Networks", Institute of Phonetic Sciences, Amsterdam, *Proceedings* 21, pp. 117-129.
- <http://www.it.uom.gr/pdp/digital.htm>