This paper has been presented at the 1997 Matlab World Conference held in Italy in October 1997

HYPIN: Hyper-Interpolation Analysis

Renato Di Lorenzo
RDL&A s.r.l., Via 5 Maggio 18/2, 16147 Genova, Italy
personal phone +39 330 539032; fax +39 10 3730921
e-mail: rdlea@mbox.vol.it or*: rdlea@bigfoot.com*
http://www.vol.it/dilorenzo

Abstract
Mandelbrot (1963) has been the first to question the simple logic of the Capital Asset Pricing Model (Elton e Gruber 1981). In financial markets, he pointed out, are present complex phenomena as the Noah effect (periodical large shocks) and the Joseph effect (positive trends tend to be followed by positive trends; the same is true for negative trends). The author believes that even more complex underlying patterns may be found. In this paper, which is of a pure experimental nature, and whose aim is at the moment purely descriptive, the likelyhood of the existence of a complex web of straight lines, so popular in technical analysis, is investigated and it is found that it has a credible foundation. A cyclic non-periodic structure of trends comes to the surface in addition. These results, as many others referenced to in the bibliography, question seriously wether even modern statistical theory (Gnedenko and Kolmogorov 1954) may be adequate to describe such complex phenomena.

Acknowledgments
The author is indebted to *The Matlab Inc*. and *Teoresi s.r.l.* for their support.

Introduction
One of the paradigms of the so-called technical analysis is that each real time-series is formed by a sequence of linear trends.
This means for instance that the straight lines drawn in figure 1 by simple visual inspection have a sense.

DEM/USD - 1989

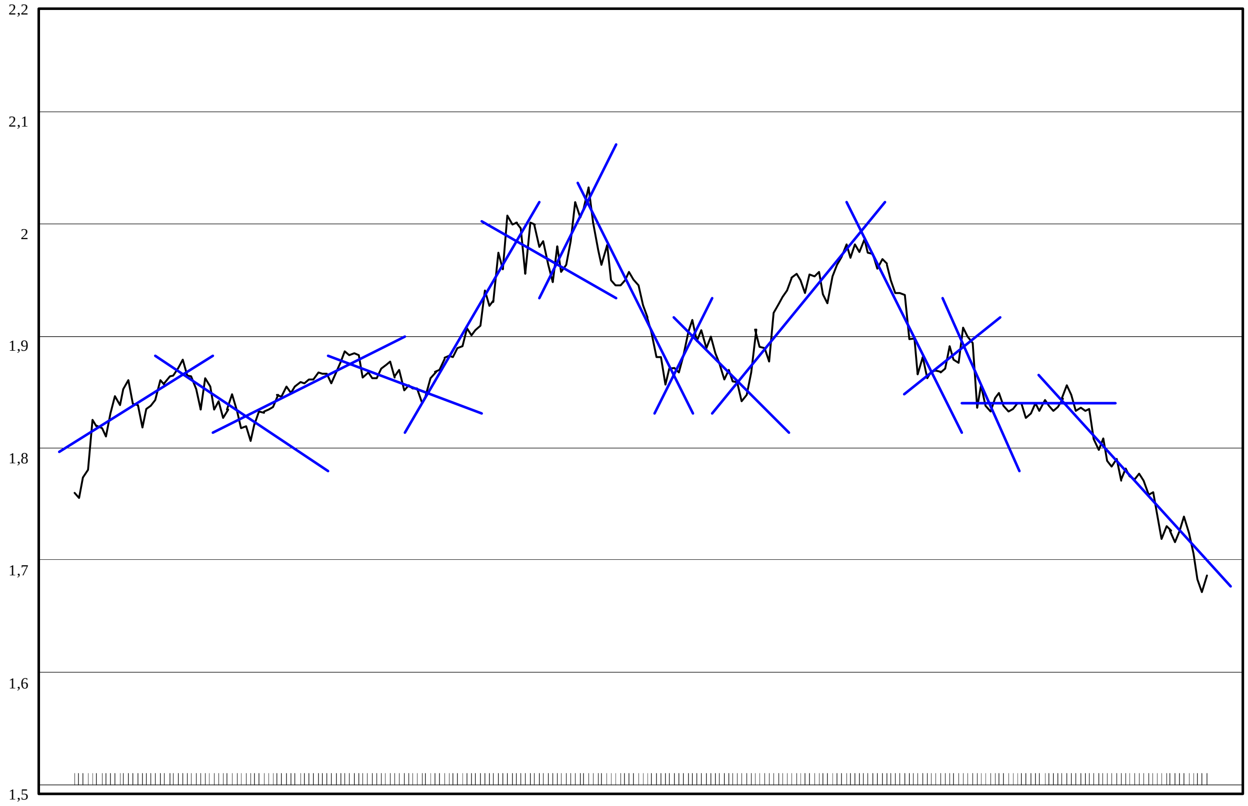


Figure 1   (file 001.cgm)

To understand if such a sense is actually there, the author has devised the method described in what follows, whose proposed name is *HYPIN* an acronim for *Hyper-Interpolation Analysis.*

The Procedure

We will assume we are analyzing a sequence of *n* samples, that are the rercorded prices of a financial asset, precisely the index of the italian Bourse; the results may well apply to other areas of investigation too.

The procedure is as follows.

1. The computer scans all the time series:

$$s_i,\ i=1,2,\ldots,\ n$$

starting from the first sample.

2. For each sample a linear interpolation of the next *l-1* samples:

$$s_i,\ s_{i+1},\ s_{i+2},\ldots,\ s_{i+l-1},\ i=1,2,\ldots,\ n-l+1$$

is made and the major statistical measures (or *statistics* for short) computed:

$seoe_{i,l}$
$c_{i,1}$
$maxc_{i,l}$
$minc_{i,l}$
$range_{i,l}$

respectively:

standard error of estimates
correlation coefficient
maximum estimate of correlation coefficient
minimum estimate of correlation coefficient
range of residuals

(see Appendix 1 for Help Window).
When we talk of *correlation coefficient* we intend to measure how coherent is the behaviour of the Bourse index with the passing of time, i.e. with a straight line.
It is assumed that the outputs of the calculations referred to above are observations of one single underlying statistical process which is not observable as such. It is assumed in addition that each output is independent from all the other ones.
As regards the above mentioned statistics, we will say also that they are *drawns from a population of possible drawns*, and we will assume such population is well-described by the averages of such statistics. Such averages are unobservable as well, but it is assumed that they can be estimated by ordinary statistical methods (see point 4).

3. The average and the standard deviation of the statistical parameters detected in 2. are computed; for instance:

$$Average_l\,(seoe_{i,l}) = (\Sigma_{i=1,n-l+1}\ seoe_{i,l})/\ (n-l+1)$$

4. Then, using the results in 3., the estimate is made concerning the population from which the statistical parameters computed in 2. are drawn; the unobservable averages we are looking for are estimated at a 95% confidence level using standard methods (see Appendix 2 for Help Window).

5. The value of *l* is increased by one unit, and the process is resumed from point 1. again.

The Results

A series of experiments has been performed, as said, on the italian Bourse index. For instance, analyzing 1995, the results shown in figure 2 for the average value of the correlation coefficient (absolute value[1]) have been obtained.

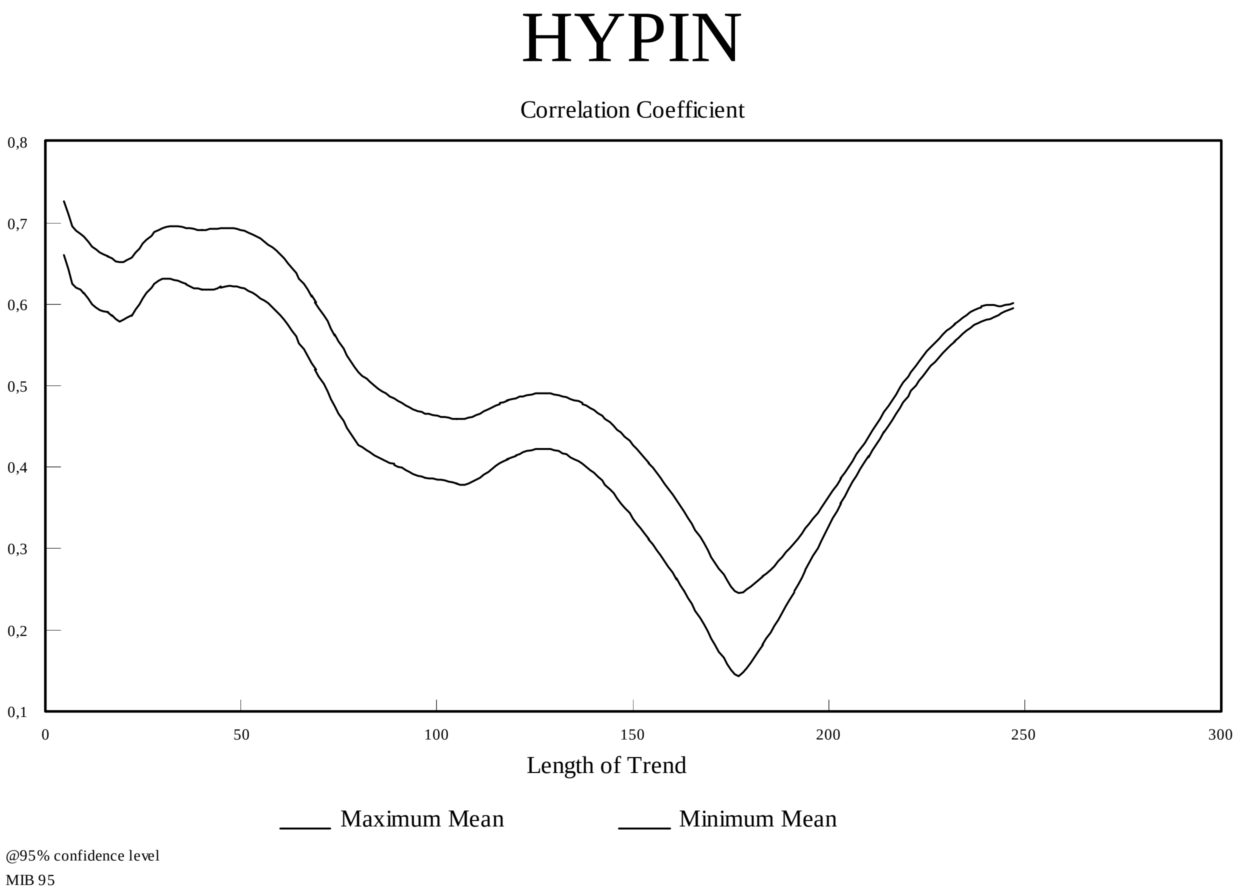


Figure 2 (file mib95.cgm)

For almost any value of *l* the correlation coefficient of the population has not much to say.The most significant value is around 0.7, which is not what is normally looked for to be much excited. However local maxima and minima are present, showing that not all values of *l* are equivalent, i.e. *not all trends were born equal*. The first local maximum is detected at *l=39.*

When the same calculation is done for 1987, the results shown in figure 3 are found.

[1] Here one has to use the absolute value as it is expected that the average of the normal correlation coefficient be almost zero. We will go back to this.

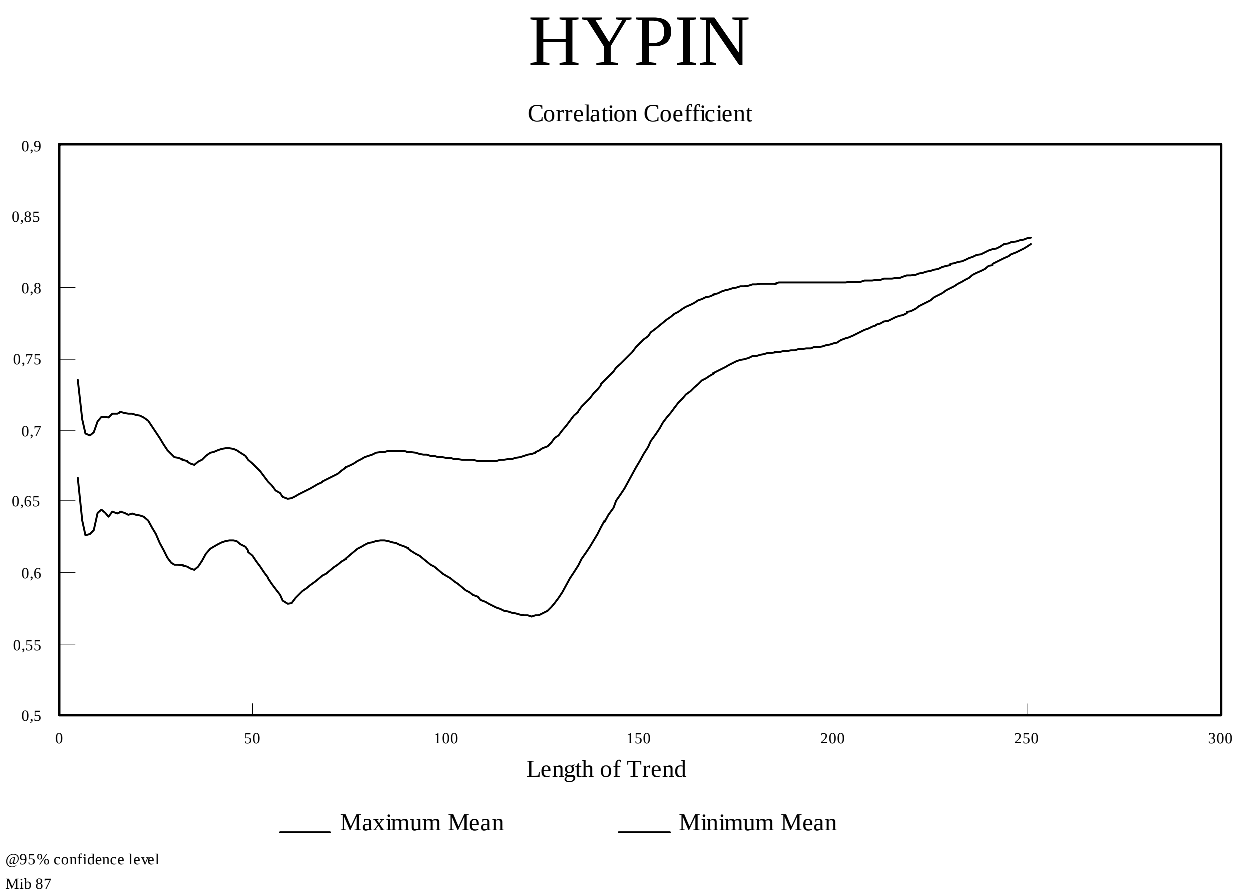


Figure 3 ....(file mibcc-83.cgm)

Here the situation is different: for a larger part of the values of *l* the correlation coefficient is significant. But what is more important is that local maxima and minima are present, but they are not located at the same values of *l* as before.
Thus a first conclusion may be drawn: the process is **not** stationary, at least as regards the value of the random variable *l*.

Let us now look at the data from another point of view.
Suppose you want to examine, during a period of one year, the behaviour of the correlation coefficient once the value of *l* has been fixed.
If we choose *l=11* as an instance, the result shown in figure 4, for year 1995, is obtained.

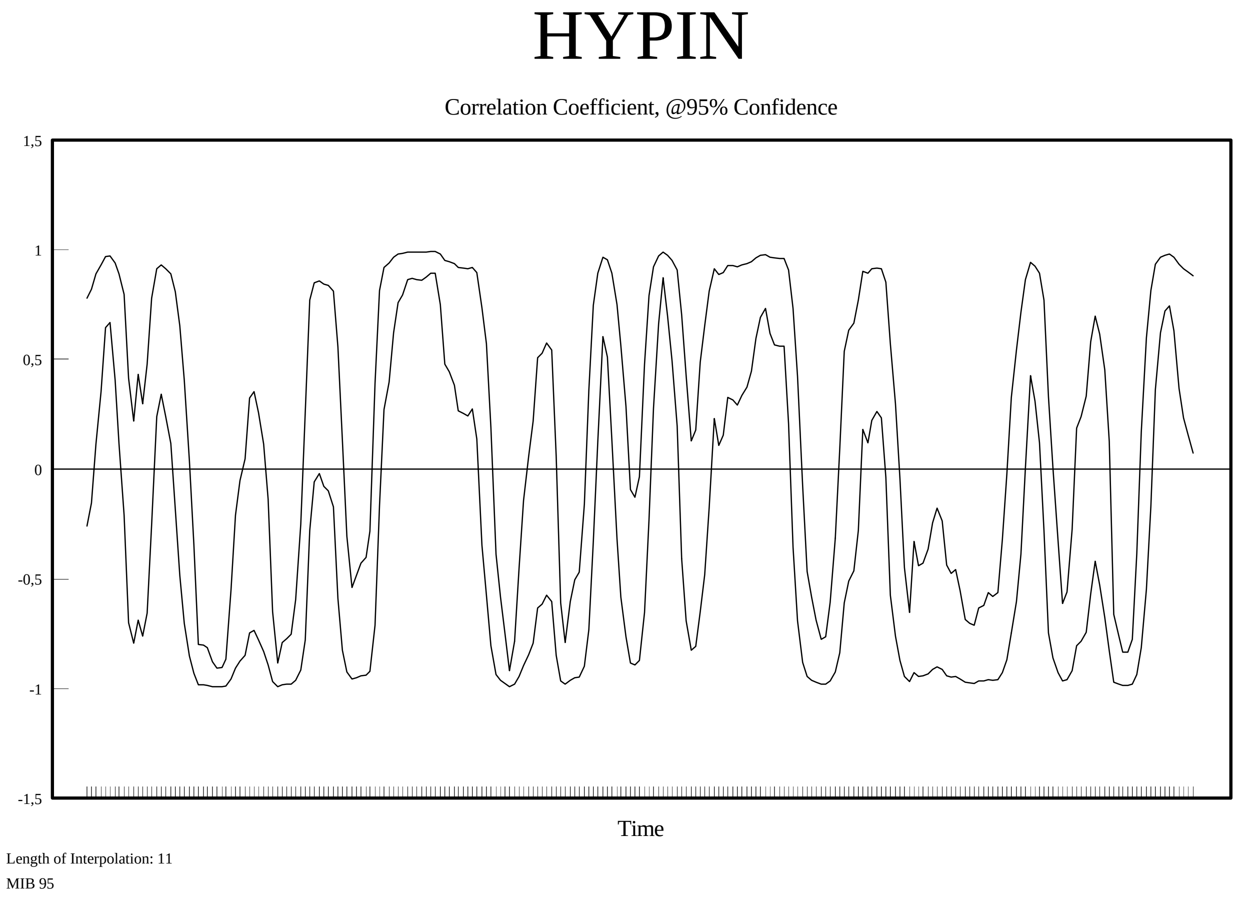


Figure 4 (file min95-11.cgm)

The same graph for 1985 has the aspect shown in figure 5.

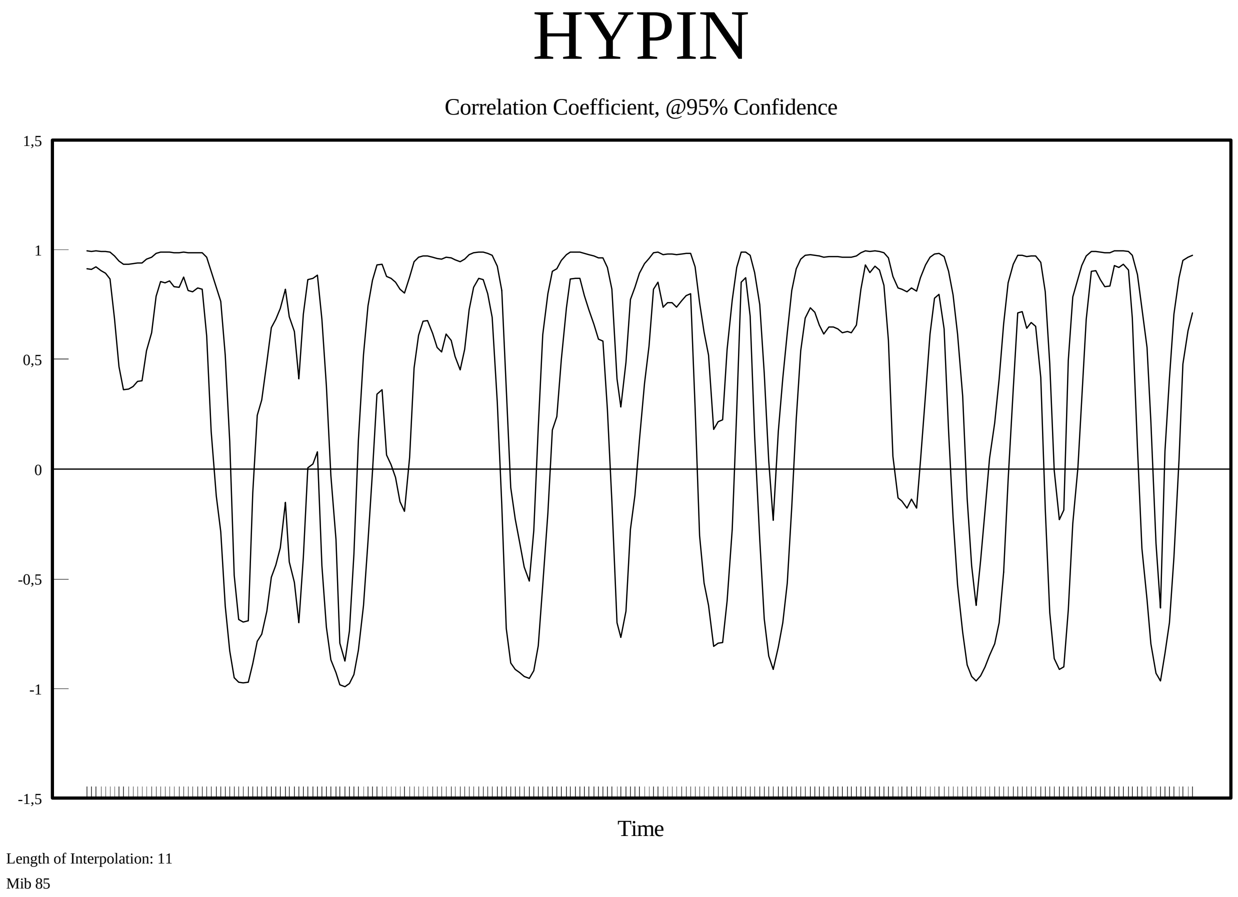


Figure 5 (file min85-11.cgm)

In these graphs the maximum and minimum values of the correlation coefficent are estimated using standard statistical methods (see appendix 1 for the Help Window).
By visual inspection alone, one can hardly detect differences between the two graphs.
What then provoked such differences? There is no other answer than the following: also in 1995 there **were** days when a credible 11 days linear trend started (values of the correlation coefficient well near ±1), but they were too few and too sparse to justify an estimate of the population significantly asserting that the 11-days-long linear trends were there.
HYPIN has posted another result: it is able to detect differences that by visual inspection alone (remember that technical analysis proceeds almost exclusively by visual inspection) are indistinguishable one another.
But the more interesting aspect that has come out is the following: during one year period, be it 1985 or 1995, the estimate of the correlation coefficient seems to have a cyclic non-periodic structure. The name non-periodic is due as, even by visual inspection alone, a fixed period for the cycles seems unlikely to be found.
The cyclic behaviour of the correlation coefficient gives us the key to understand why we used a graph of the **absolute** values of the correlation coefficient to detect its behaviour against *l:* as the true correlation coefficient (with its sign) is alternatively positive and negative, its average would have been near zero.

A statistc has been computed about the length of such cycles from 1985 to 1995, for a total of 11 years. Each year the average duration of cycles has been computed and then, for each year, the maximum and minimum average values estimated with a level of confidence of 95% (see Appendix 1 for the Help Window). Here are the results:

Maximum Length	Minimum Length	Year
21,7147	9,337928	85
30,50319	12,83015	86
32,39027	18,22512	87
24,02687	10,19536	88
22,47015	11,21406	89
25,20181	12,13152	90
5,25964	9,048051	91
21,0434	10,28993	92
20,93916	11,27136	93
23,02822	13,81389	94
19,68344	10,31656	95

It may be noticed that, if the process does not seem stationary as regards the significance of the existence of the linear web, on the contrary, once a length *l* is chosen, year after year the cycles *frequency* does not seem to differ much; in this sense the process shows signs of stationarity. These values again are drawns from an unobservable population of cycle durations, and we again adopt the point of view that such population is well-described by average values. We then can estimate such values by guessing a t-Student distribution (see Appendix 2 for Help Window) for this series of data, and may compute the confidence limits at 95% of the mean of the population from which the samples were drawn using the following standard formula (Spiegel 1975):

$$Max=\mu+t_{0,975}*s/(n-1)^{0,5}$$
$$min=\mu-t_{0,975}*s/(n-1)^{0,5}$$

where μ is the sample mean, s is the sample standard deviation and $t_{0,975}$ is 2,20 when the degrees of freedom are 11; hence:

Average		Standard Deviation		Number of Samples
Max	Min	Max	Min	
23,29644	11,69763	4,575948	2,480969	11

and:

	Maximum	Minimum
Maximum value of	26,47993	13,42364
Minimum value of	20,11295	9,971617

So we may conclude that there is a strong likelyhood that on the italian Bourse, at least as far as the index is concerned, sufficiently credible linear trends 11 days long are present; such trends are alternatively upward or downward directed, and they last from a minimum of 10÷13 days to a maximum of 20÷26 days, twice as much. Note that the 11 days length is almost centered in the interval of likelyhood of the minimum duration.

Another phenomenon tackled by HYPIN is the following.
The maximum and minimum values of the correlation coefficient (absolute value) for an interpolation length of 83 days, in 1987, is reported in figure 6, and a characteristic behaviour shows up: in practice there have been, during 1987, long periods when each and every day a credible trend 83 days long started; in other (short) periods the correlation coefficient dropped at levels that usually, in normal practice, are rejected.
On the first local maximum M_1 of the correlation coefficient the situation is the one depicted in figure 7. In correspondence with the first (and worst) local minimum m of the correlation coefficient, the situation is the one depicted in figure 8. In correspondence with the second local maximum M_2 we have the situation shown in figure 9.

HYPIN

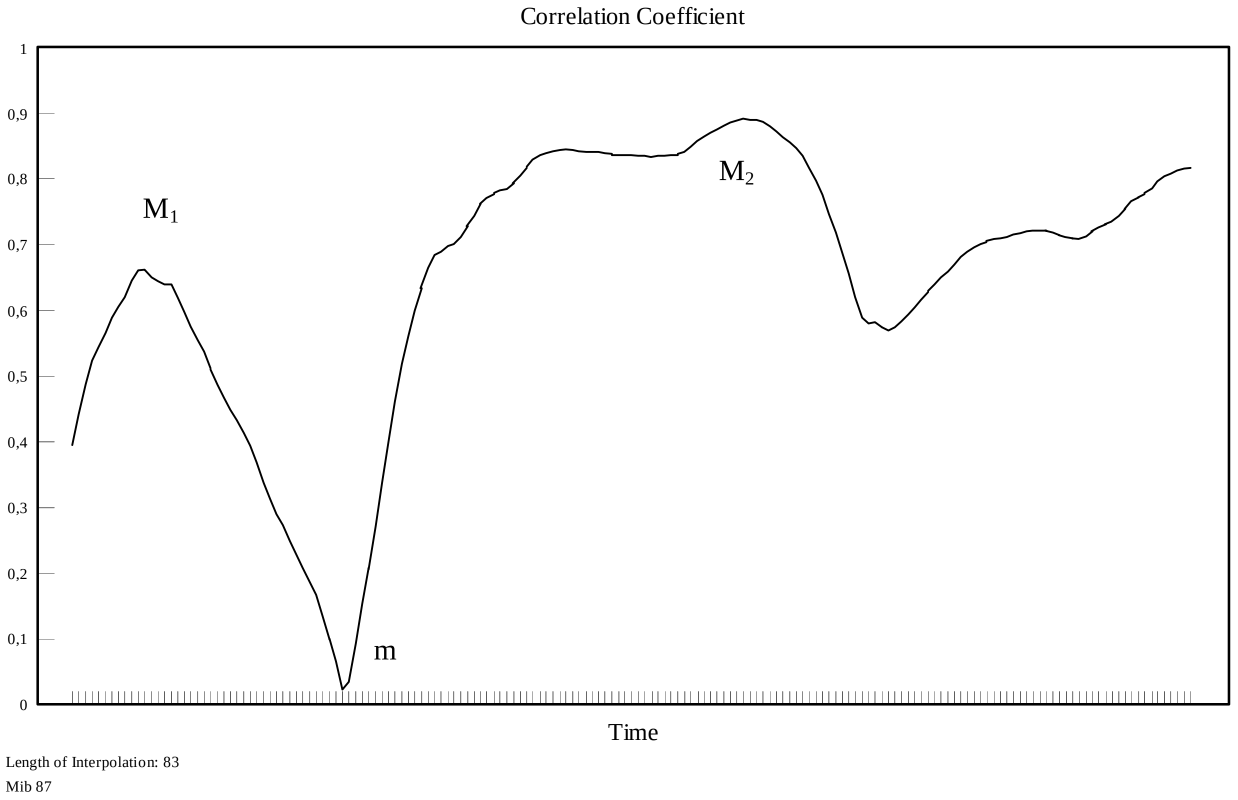
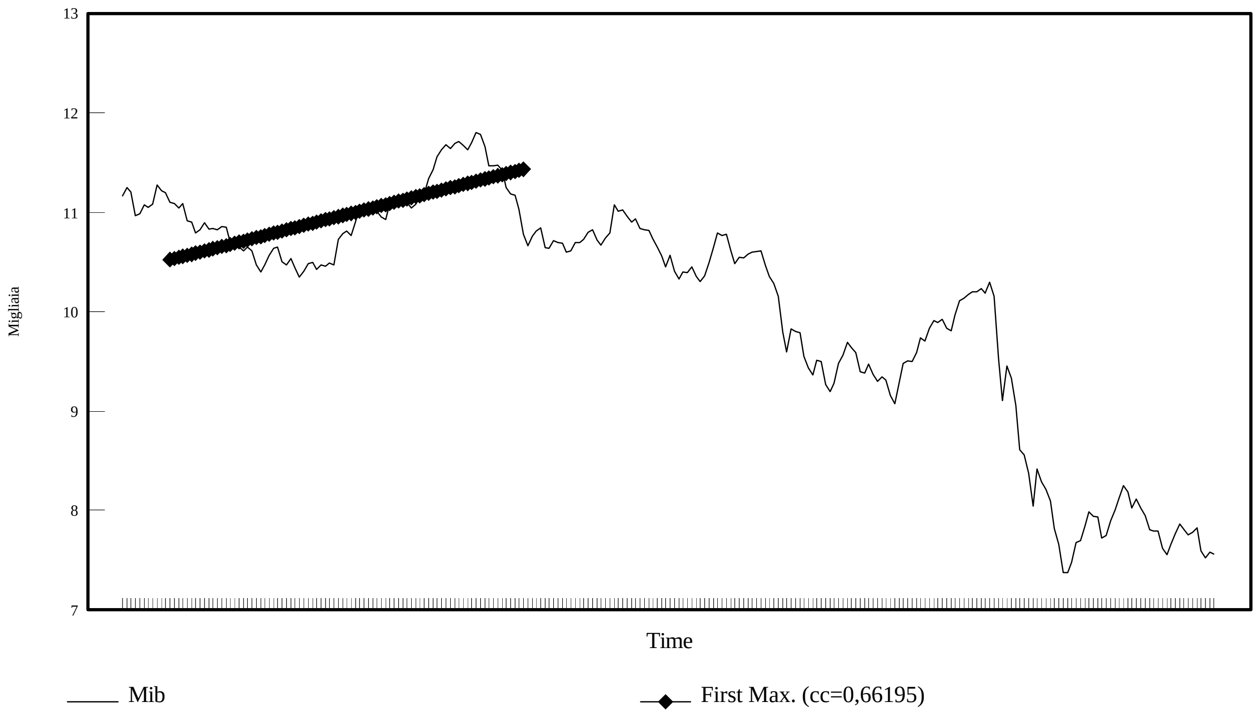


Figure 6 (file mibcc.cgm)

HYPIN

Correlation Coefficient

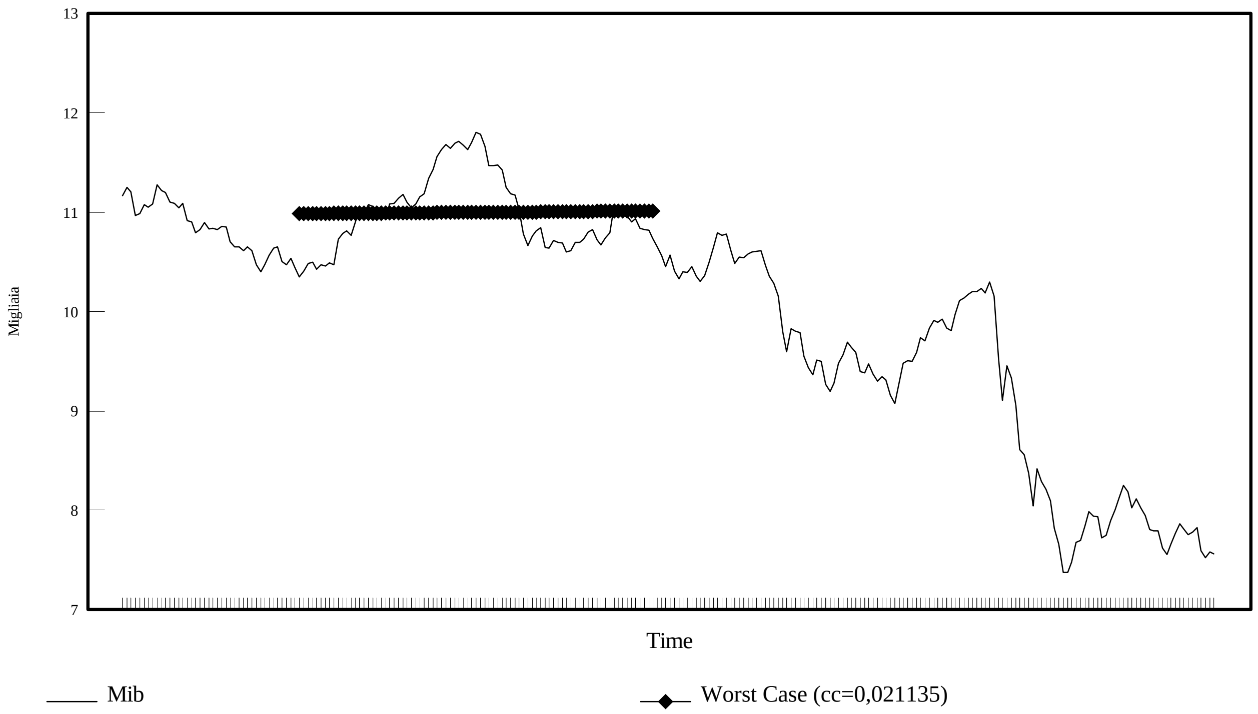


Length of Interpolation: 83
Mib 87

Figure 7 (file mib-b1.cgm)

HYPIN

Correlation Coefficient



Length of Interpolation: 83
Mib 87

Figure 8 (file mib-w.cgm)

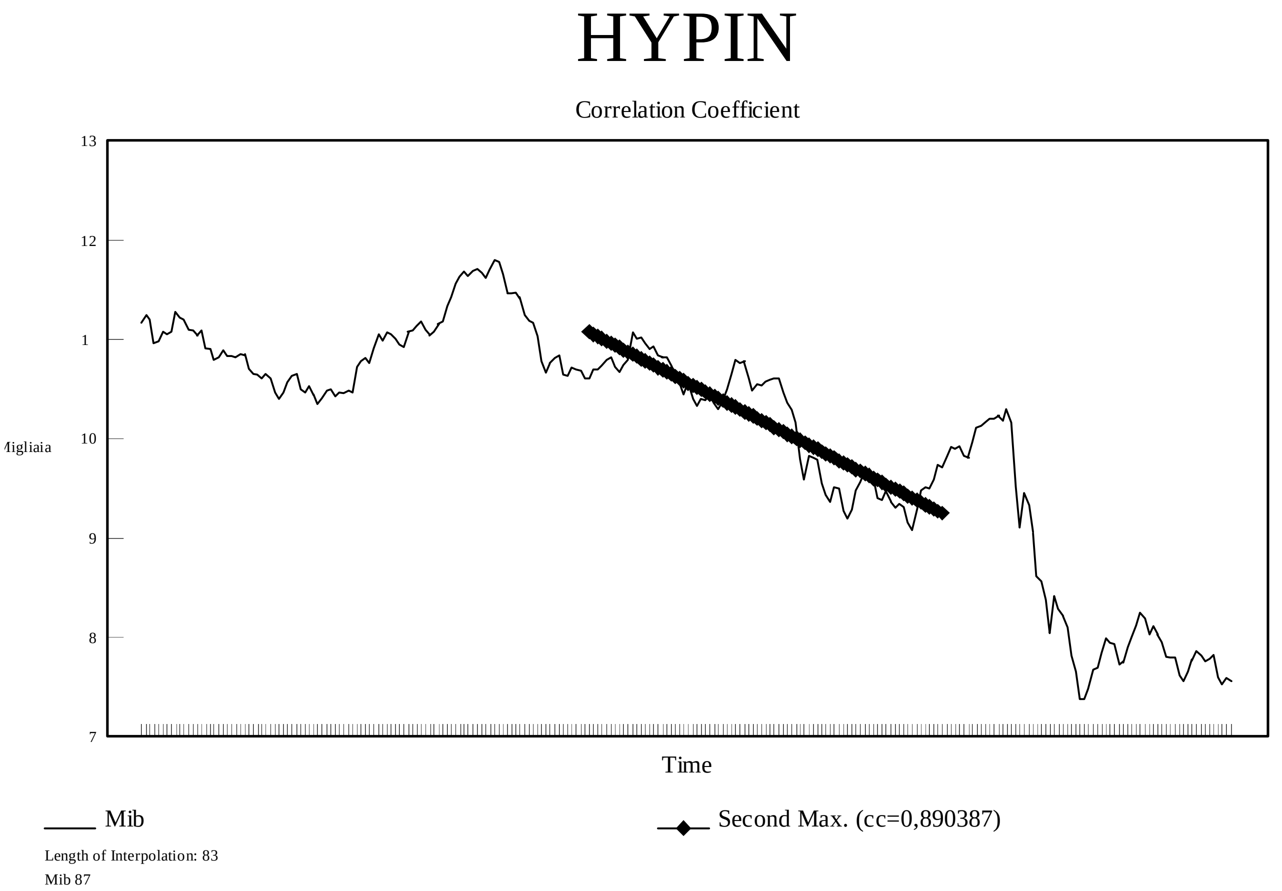


Figure 9 (file mib-b2.cgm)

Well, it is apparent: by visual inspection alone the (much) marked differences between correlation coefficients **cannot** be detected, unless one uses too much phantasy indeed; no really sensible difference can be decided between the three situations just looking at the pictures.

Within the set of the values (range) of the time series - in this case the italian Bourse index - there is a subset formed by all the points that are part of a linear interpolation of length *l* that has a correlation cofficient higher than a certain threshold *t* . We will call this subset a *(t,l)-coherent subrange.* In figure 10 is shown the (*90%, 11 days)-coherent subrange* regarding 1987.

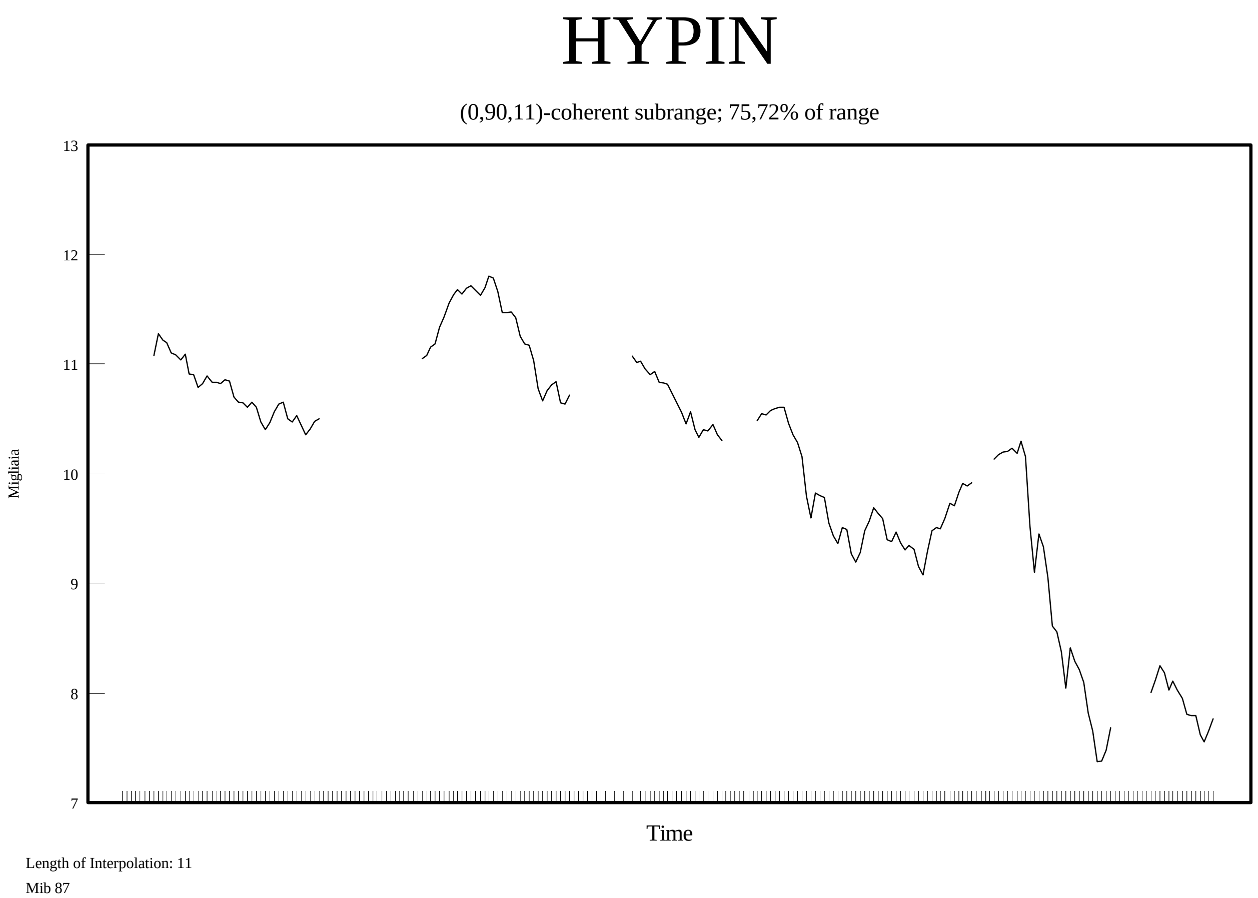


Figure 10 (file coh-87.cgm)

As can be seen, more than 75% of the time, during 1987, the italian Bourse index was in a highly credible linear trend 11 days long.
A tentative interpretation, which sounds reasonable, is that do **not** belong to the subrange points immediately following the conclusion of a *secondary trend* in the sense of technical analysis.

Now, the problem is: is the 11 days length the only credible length? The answer is **no**.
In figure 11 it is shown that with *l*=5, more than 85% of the range lies in the coherent subrange. In a sense, this is what one might have expected, as points that belong to an 11-days linear trend should likely belong also to a 5-days linear trend.

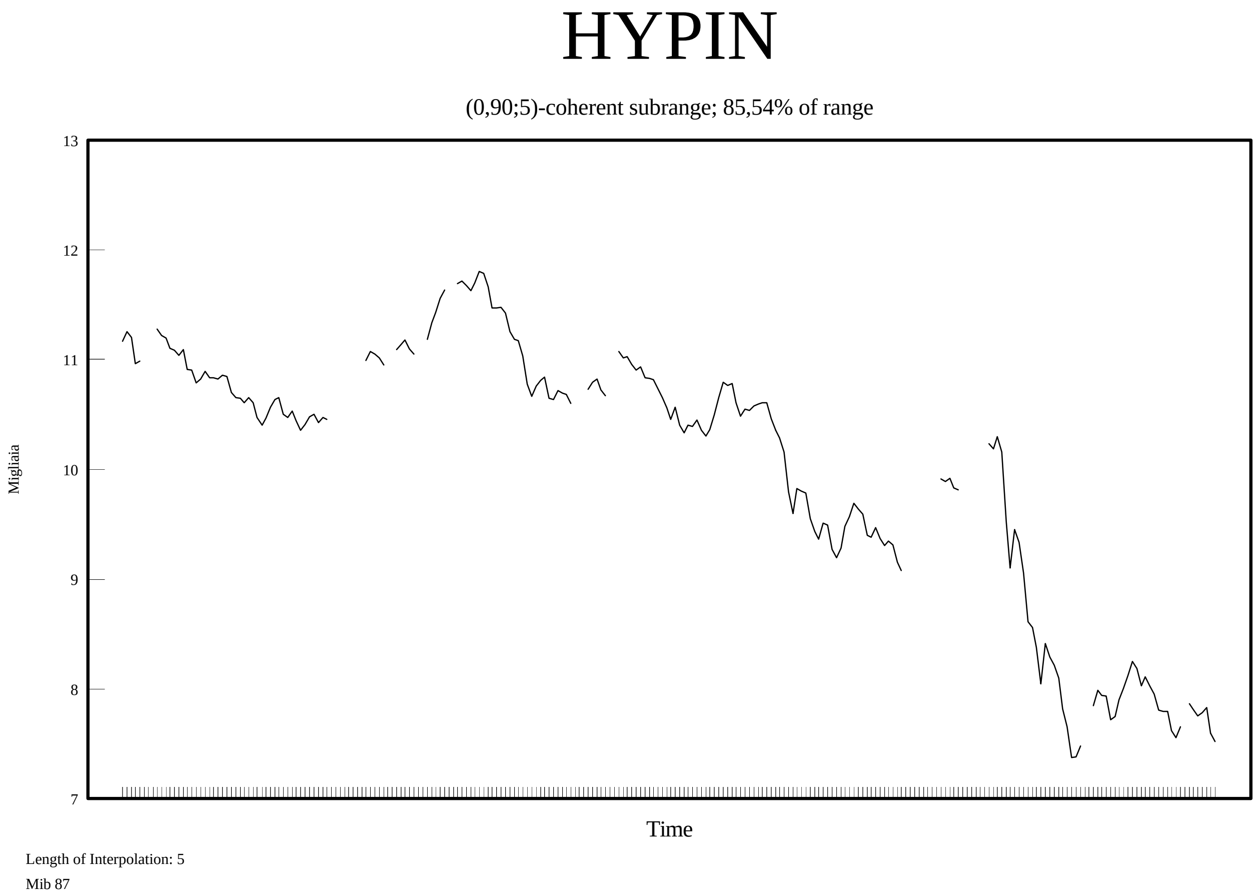


Figure 11 (file coh-87-2.cgm)

This phenomenon, if analyzed on a longer time span, namely 12 years from 1985 to 1996 included, show a lower reading (figure 12).

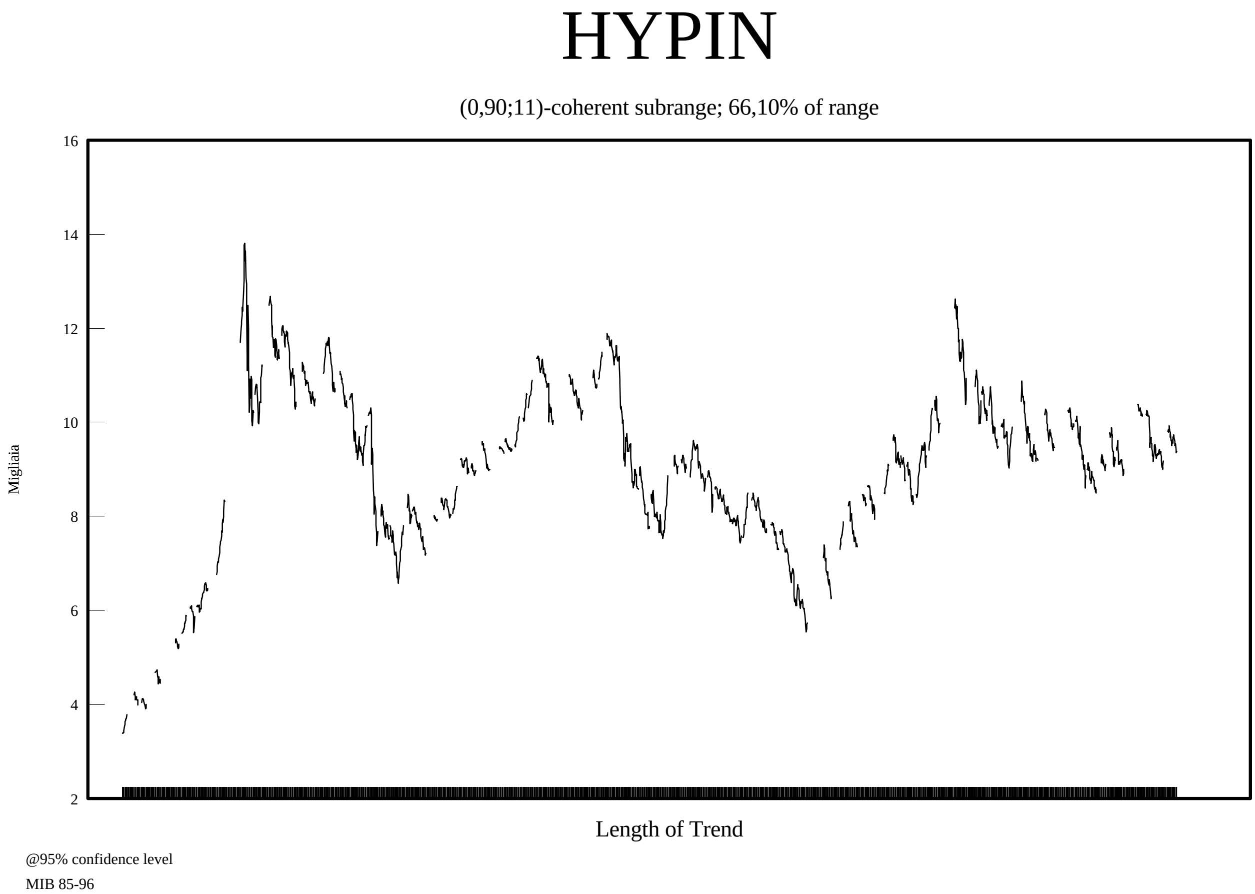


Figure 12 (file coh-t-11.cgm)

If such analysis is carried forward, and the same 12 years span is analyzed for different values of *l* , the picture shown in figure 13 is displaied, where a quite *jagged* function of *l* comes out, showing local maxima and minima, as well as an absolute maximum at about 440 days (two trading years).

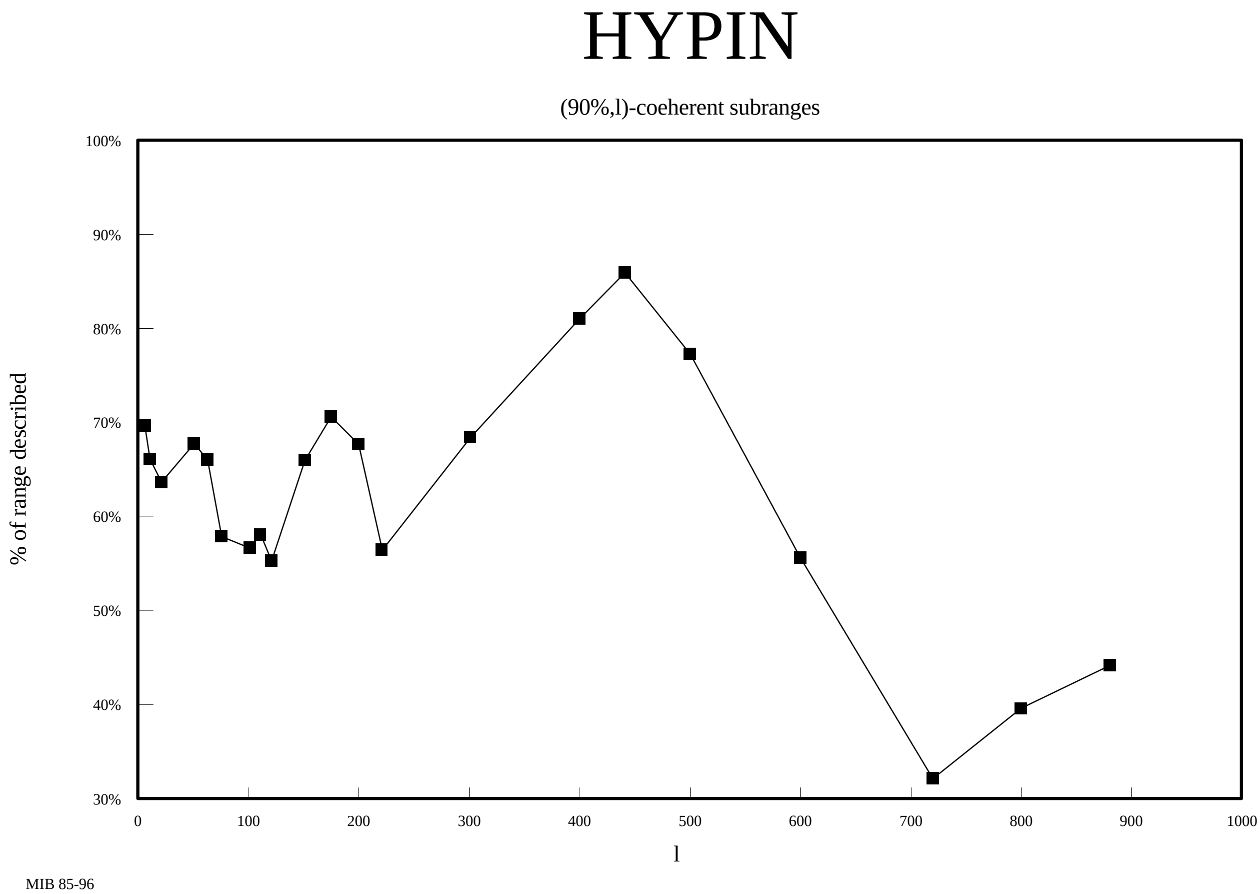


Figure 13 (file coh.cgm)

Conclusions

The last picture and the other findings may be summarized as follow:

- the time series is well described for most of the time by a web of linear trends, but
- not all linear trends were born equal: some describe better the time series than others;
- given a trend length *l*, it appears credible that positive and negative linear trends of such length alternate in a cyclic, non-periodic, but stationary fashion.

Appendix 1 - Help Window

We have said that for each sample a linear interpolation of the *l* samples:

$$s_i,\ s_{i+1},\ s_{i+2},...,\ s_{i+l-1},\ i=1,2,...,\ n-l+1$$

is made using standard least squares methods (Spiegel 1975).
Using the same standard methods the correlation coefficient is computed.
The residuals:

$$d_i,\ d_{i+1},\ d_{i+2},...,\ d_{i+l},\ i=1,2,...,\ n-l+1$$

are the differences between each sample and the corresponding value of the linear interpolation.
The standard error of estimate is the standard deviation of the residuals.
The range of the residuals is the difference between the maximum and the minimum value of the residuals.

The correlation coefficient is itself a random variable, and its value has been computed on a sample drawn from an unobservable population which is assumed nontheless to exist. Thus one important issue is that of estimating the *true* value of it, namely which is likely to be, with a confidence level of, say, 95%, the correlation coefficient of the population from which the sample correlation coefficient was drawn.
To solve this problem (Spiegel 1975) it is used the fact that Fisher's z-transform:

$$z=0,5*\ln((1+r)/(1-r))$$

- where *r* is the sample correlation coefficient - has a probability distribution that is is approximately normal with mean:

$$\mu=0,5*\ln((1+\rho)/(1-\rho))$$

and standard deviation:

$$\sigma=1/(n-3)^{0,5}$$

where ρ is the correlation coefficient of the population and *n* is the size
of the sample under exame.
Then the value of the correlation coefficient of the population, at confidence level of 95%, lies within plus and minus 1,96 standard deviations from the mean.
Thus as *r* and *n* are known, it is possible to compute:

$$z\pm1,96*\sigma=0,5*\ln((1+r)/(1-r))\pm1,96*(1/(n-3)^{0,5})$$

which result in the two values:

$$z_m,\ z_M$$

Then solving for ρ in the two equations:

$$z_m=0,5*\ln((1+\rho)/(1-\rho))$$
$$z_M=0,5*\ln((1+\rho)/(1-\rho))$$

gives the fiduciary limits for the correlation coefficient:

$$\rho_m=(\exp(2*z_m)-1)/(\exp(2*z_m)+1)$$
$$\rho_M=(\exp(2*z_M)-1)/(\exp(2*z_M)+1)$$

Appendix 2 - Help Window

Classical and modern statistics (Spiegel 1975; Gnedenko and Kolmogorov 1954) all deal with the following assumption: *when a random variable is to be handled, it is assumed that there exists an underlying process that generates each drawn of such a random variable; the characteristics of such a process, though in general not observable, may be inferred by the oservation of the samples, i.e. of the actual drawns.* We say also that the the samples that we observe are drawns from a *population* of possible drawns.

As a matter of fact, when we talk about *characteristics* of the underlying process, we talk almost only about tha probability distribution of the above mentioned population, which is called *the limiting probability distribution* exactly because it is inferred by the frequency distributions observed in the samples, in some way through a passage to the limit.

The most important parameter looked for is the average of the population, and it is often said that this is the *true value* of the quantity we are masuring if we are measuring something.

Well, it is shown in classical statistics (Spiegel 1975) that such an average may be inferred using a *t-Student* probability distribution, and that its *true value, - with a 95% confidence*, it is said - may be considered to lie between the following two values:

$$\text{Max}=\mu+t_{0,975}*s/(n-1)^{0,5}$$
$$\text{min}=\mu-t_{0,975}*s/(n-1)^{0,5}$$

where μ is the sample mean, s is the sample standard deviation and $t_{0,975}$ is a number which depends on the number of *degrees of freedom* (which add up to $n-1$) and which is deducted from a table of the *t-Student* distribution. Thus classical statistics allows to inferr from the mean and the average of the sample which the average of the population may be, or at least two limits within which such an average is likely to lie.

References

Amerio L., *Analisi Matematica*, Vol. I and II, Di Stefano, Genova, 1960

Arnold C., *PPS Trading System*, Irwin Professional Publishing, Burr Ridge, Illinois, 1995

Baestaens D., Van Den Bergh W. M., Wood D., *Neural Network Solutionsfor Trading in Financial Markets*, Pitman, London, 1994

Bauer R. J., *Genetic Algorithms and Investment Strategies*, Wiley, New York, 1994

Brown R. G., *Smoothing, Forecasting and Prediction of Descrete Time Series*,Prentice-Hall, Englewood Cliffs, 1962

Dacorogna M. M., Gauvreau C. L., Muller U. A., Olsen R. B., Pictet O. V., *Changing Time Scale for Short-term Forecasting in Financial Markets*, Journal of Forecasting, Vol. 15, Iss. No. 3, 1996

Davenport W. B., Root W. L., *An Introduction to the Theory of Random Signals and Noise*, McGraw Hill, New York, 1958

D'haeseleer P., *An Immunological Approach to Change Detection: Theoretical Results*, IEEE Computer Security Foundation Workshop, 10-12 june 1996

Di Lorenzo R. e Sciarretta V.: *Evidenze Statistiche Riguardanti un Nuovo Metodo di Trading sui Mercati Finanziari*, 1996 a, AF-Analisi Finanziaria, n. 24, december 1996; english version available: *Statistical Evidences Concerning a New Method of Trading the Financial Markets.*

Di Lorenzo R.: *A Case Study in Organization Dynamics*, IEEE Transactions on Engineering Mangement, July 1973.

Di Lorenzo R.: *Come Guadagnare in Borsa*, Il Sole 24 Ore, Milano, 1991

Di Lorenzo R.: *Detecting the Differences in the Statistical Structure of the Financial Markets - the PLI and the R/S^{MOM} Algorithms*, 1996 b, to be published

Di Lorenzo R.: *E Chaotic Model of the Financial Markets (CSSP*), 1996 c, submitted to 15th IMACS World Congress, Berlin, 24-29 August 1997

Di Lorenzo R.: *Forecastability and Tradability*, International Conference on Chaos, Fractals & Models '96, University of Pavia, Italy, October 25-27, 1996 d

Di Lorenzo R.: *Guadagnare in Borsa con l'Analisi Tecnica: gli Oscillatori*, Il Sole 24 Ore,Milano, 1994

Di Lorenzo R.: *Guadagnare in Borsa con l'Analisi Tecnica: I Trend*, Il Sole 24 Ore, Milano, 1993

Di Lorenzo R.: *Guadagnare in Borsa con l'Analisi Tecnica: le Candele Giapponesi*, Il Sole 24 Ore,Milano, 1996

Di Lorenzo R.: *Guadagnare in Borsa con l'Analisi Tecnica: le Figure*, Il Sole 24 Ore,Milano, 1995

Di Lorenzo R.: *Guadagnare Investendo all'Estero*, Il Sole 24 Ore, Milano, 1991

Di Lorenzo R.: *Infinite nth Moment Detection: an Un-Decidable Question? An Euristic Discussion*, 1996 e, submitted to ENUMATH 1997, septembre 29th-october 3rd, University of Heidelberg, Germany

Di Lorenzo R.: *Lineamenti di una Teoria Sistemistica della Gestione dell'Impresa*, Elettronica 2, Torino, Italy, 1973

Di Lorenzo R.: *Multiloop Control of the Production Enterprise*, Proceedings of the Fifth Annual Symposium on System Theory, North Carolina State University and Duke University, USA, 1973

Di Lorenzo R.: *Taking into Account Opportunity Cost in the I.R.R. calculations, and the Theory of the Firm*, 1979, working paper

Di Lorenzo R.: *The Application of System Engineering Methods to Economics*, 1972, Internal Note, Fiat, Centro Elettronico Avio, Turin

Di Lorenzo R.: *The ECA Method of Forecasting and its Improvement Via a Genetic Algorithm - The Problem of the Information Content*, 1996 f, to be published

Eldridge R., Bernhardt C., Mulvey I., *Evidence of Chaos in the S&P 500 Cash Index*, in Trippi R. R. edit., Chaos and Nonlinear Dynamics in the Financial Markets, Irwin, Chicago, 1995
Elton E. J., Gruber M. J., *Modern Portfolio Theory and Investment Analysis*, Wiley, New York, 1981
Falconer K., *Fractal Geometry*, Wiley, New York, 1990
Fama E. F., *Mandelbrot and the Stable Paretian Hypothesis*, in The Random Character of Stock Market Prices, Paul H. Cootner ed., M.I.T. Press, 297-307, 1964
Forrester J.W., *Industrial Dynamics*, MIT Press, Cambridge, Mass., 1961
Gnedenko B. V., Kolmogorov A. N., *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Cambridge, Mass., 1954
Granger W. J. and Orr D., *Infinite Variance and Research Startegy in Time Series Analysis*, Journal of the American Statistical Association, June1972, 67, 338
Hertz J., Krogh A., Palmer R. G., *Introduction to the Theory of Neural Computation*, Addison Wesley, Redwood City, 1991
Holland J. H., *Adaptation in Natural and Artificial Systems*, Ann Arbor, MI, 1975
Hurst H. E., *Long-Term Storage Capacity of Reservoirs*, Transactions of the American Society of Civil Engineers, 116, 1951
Lipschutz S., *General Topology*, McGraw-Hill, New York, 1965
Mandelbrot B., Taylor H.M., *On the Distribution of Stock Price Differences*, Operations Research, 15, 1967, 1057-1062
Mandelbrot B., *Statistical Methodology fro Non-Periodic Cycles, from the Covariance to the R/S Analysis*, Annals of Economic and Social Measurement, 1, 1972
Mandelbrot B., *The Variation of Certain Speculative Prices*, Journal of Business of the University of Chicago, 36, 394-411, 1963
Marcuk G. I., *Metody Vycuslitelnoj Matematiki*, Mauka, Moscow, 1984
Millard B. J., *Winning on the Stock Market*, Wiley, Chichester, 1993
Murphy J.J., *Technical Analysys of the Futures Market*, New York Institute of Finance,New York, 1986
Neftci S. N., *A Note on the Use of Local Maxima to Predict Turning Points in Related Series*, Journal of the American Statistica Association, September 1985, 80, 391
Neftci S. N., *Naive Trading Rules in Financial Markets and Wiener-Kolmogorov PredictionTheory: a Study of "Technical Analysis"* Journal of Business, 1991, vol. 64, n.4
Pancini E., *Misure ed Apparecchi di Fisica*, Veschi, Roma, 1965
Peters E. E., *Chaos and Order in the Capital Markets*, Wiley, New York, 1991
Peters E. E., *Fractal Market Analysis*, Wiley, New York, 1994
Pring M. J., *Technical Analysis Explained*, McGraw-Hill, New York, 1980
Savit R., *When Random is not Random: an Introduction to Chaos in Market Prices*, in Trippi R. R. edit., Chaos and Nonlinear Dynamics in the Financial Markets, Irwin, Chicago, 1995
Shannon C. E., Weaver W., *The Mathematical Theory of Communication*, University of Illinois,Urbana, 1963

Spiegel M. R., *Probability and Statistics*, MacGraw Hill, New York, 1975
Stoll R. R., *Set Theory and Logic*, Dover, New York, 1979
Tanizaki H., Mariano R. S., *Prediction, Filtering and Smoothing in Non-Linear and Non-Normal Cases Using the Monte Carlo Integration*, Journal of Applied Econometrics, Volume 9, Number 2, April-June 1994
Trippi R. R. edit., *Chaos and Nonlinear Dynamics in the Financial Markets*, Irwin, Chicago, 1995
Tsay R. S., *Outliers, Level Shifts, and Variance Changes in Time Series*, Journal of Forecasting, 7, 1-20
Vince R., *Portfolio Management Formulas*, Wiley, New York, 1990
Vince R., *The Mathematics of Money Management*, Wiley, New York, 1992
Vince R., *The New Money Management*, Wiley, New York, 1995
Von Neuman J., Morgenstern O., *Theory of Games and Economic Behaviour*, PrincetonUniversity Press, Princeton, 1944
Weiss M. D., *Nonlinear and Chaotic Dynamics: an Economist's Guide*, in Trippi R. R. edit., -iChaos and Nonlinear Dynamics in the Financial Markets, Irwin, Chicago, 1995
Wylie Jr. C. R., *Advanced Engineering Mathematics*, McGraw Hill, New York, 1960
Zemanian A. H., *Distribution Theory and Transform Analysis*, McGraw-Hill, New York, 1965