

Towards a Framework for Combining Stochastic and Deterministic Descriptions of Nonstationary Financial Time Series

Ragnar H Lesch
leschrh@aston.ac.uk

David Lowe
d.lowe@aston.ac.uk

Neural Computing Research Group
Aston University, Birmingham B4 7ET, UK
Tel: +44 121 333 4631 Fax: +44 121 333 4586
<http://www.ncrg.aston.ac.uk/>

Abstract

We present in this paper ideas to tackle the problem of analysing and forecasting nonstationary time series within the financial domain. Accepting the stochastic nature of the underlying data generator we assume that the evolution of the generator's parameters is restricted on a deterministic manifold. Therefore we propose methods for determining the characteristics of the time-localised distribution. Starting with the assumption of a static normal distribution we refine this according to the empirical results obtained with the methods and conclude with the indication of a dynamic non-Gaussian behaviour with varying dependency for the time series under consideration.

1 Introduction

Using neural networks to predict the return of a financial asset in the future conditional upon past behaviour has only received limited success so far. In part this is due to either assumptions of deterministic behaviour or problems due to the stochastic nature and long memory effects leading to high input dimensions. Consequently, a vast numbers of data points is then required to estimate reliable statistics about the underlying process. However, requiring a long time history is a problem as financial time series processes are nonstationary.

In this paper we consider an alternative approach to the problem of analysing nonstationary financial time series. We accept the intrinsic stochastic nature of the return generating process by characterising its local distribution via the probability density function, its characteristic function as well as with moments and cumulants.

Consequently, the assumption of a slowly varying structure of the distribution allows us to model the dynamics of the nonstationarity in the time series. This could be addressed by deterministic models which track the drift of the underlying parameters of the distribution and finally forecast those transitions.

We will demonstrate the idea for the proposed methods using the financial time series of the daily Dow Jones Industrial Average (DJIA) stock price index for approximately the last 100 years, consisting of 27611 samples. The closing prices and corresponding returns, defined as the log-difference of two consecutive prices are shown in figure 1. From these figures it can be observed

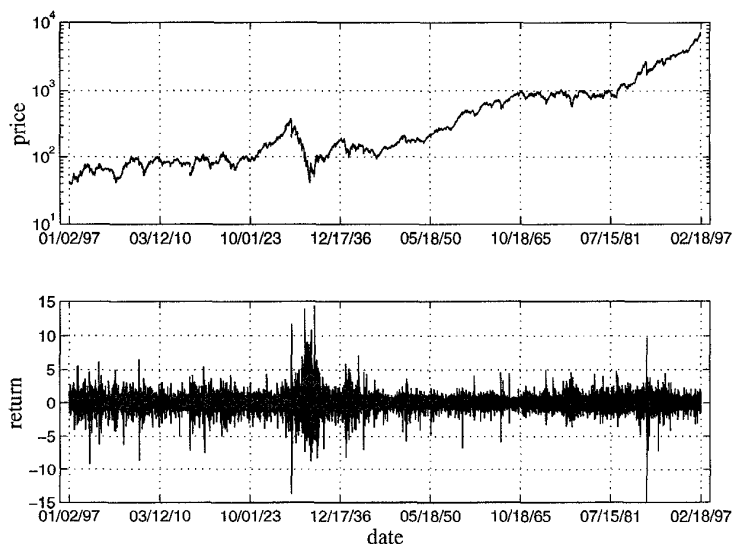


Figure 1: The daily prices of the DJIA stock index (above) and the corresponding returns (below) for the years 1897 until 1997

that the prices follow a nonlinear trend and that extreme returns are clustered indicating a changing variance of the distribution over time.

In the following we describe the possibilities to characterise the stochastic properties of the time series generating process. First, we introduce the approach and results obtained by a direct estimation of the probability density function. Afterwards we outline possibilities to estimate the characteristic

function as an alternative to the p.d.f. In the fourth section a more direct way of obtaining information about the distribution using moments and cumulants will be introduced. Afterwards we suggest a time-dependent mixture model and present evidence for varying distributional properties for the DJIA time series. Finally, a conclusion about the considered time series is given and some extensions for future work will be outlined.

2 Density estimation

Since we want to model financial time series within a probabilistic framework we need to model the probability density function. That will finally enable us to produce better decision criteria than just to give an estimate for the mean of that distribution, very often obtained in a neural network forecasting solution.

For density estimation we consider here three direct methods, an indirect approach will be discussed in the following section. One first direct way is to fit the parameters of an explicit density model such as a Gaussian or Laplacian one, which is known as a parametric method. A second, non-parametric, approach lets the data determine the form of the density where a model is not available. One example for this is a kernel estimator where the final p.d.f. is approximated as the sum of kernels around each data point in the sample [9].

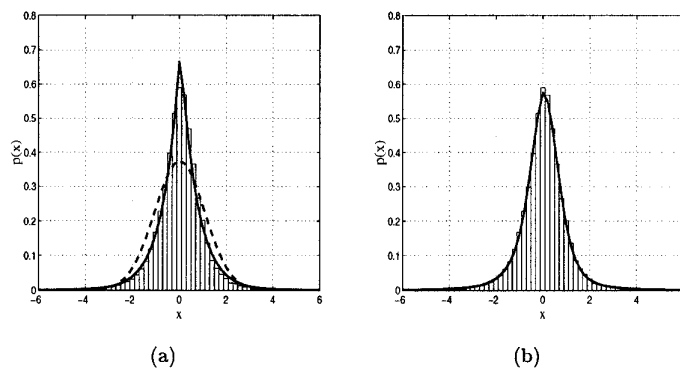


Figure 2: The empirical probability density (solid) for the DJIA returns obtained (a) with a single Laplacian (solid) and a Gaussian (dashed) and (b) with a mixture of one Laplacian and one Gaussian (superimposed with the corresponding histogram respectively)

The third approach combines both methods, taking from the parametric model some prior knowledge about the form of the distribution as a restric-

tion for the empirical density, and from the kernel estimator its flexibility to model densities which are not covered by the chosen prior. Therefore this semiparametric method is also called a mixture model which adjusts its parameters in the way to maximise the likelihood that these parameters have generated the data, e.g. via the EM algorithm [2].

Starting with fitting a single density model to our data we find that a Laplacian distribution is more likely to have created the returns than a Gaussian one. The negative average log-likelihood results are $E_L = 1.344$ for the Laplacian and $E_G = 1.484$ for the Gaussian fit. Figure 2(a) shows the empirical density as a histogram and the Laplacian and Gaussian fit.

In figure 2(b) we contrast a single fit to a mixture of one Laplacian and three Gaussians which gives a negative average log-likelihood $E_M = 1.332$ (a purely Gaussian mixture gives $E_M = 1.333$). In these pictures it can be observed that the empirical density of DJIA returns shows a significant leptocurtic behaviour (higher peak around the mean and fatter tails than the corresponding normal distribution).

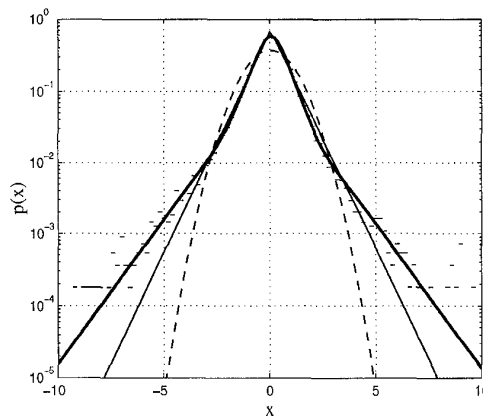


Figure 3: The empirical density estimates (solid) for the DJIA returns with a Laplacian (solid) and a Gaussian (dashed) superimposed with the corresponding histogram

In figure 3 we repeat the plot in figure 2(b) on a log scale in order to show the appropriateness of this configuration since it captures the double-linear scale of the density.

These findings about the shape of the distribution suggest that the time series has not been generated by a static Gaussian. We consider here two possible explanations:

First, the non-Gaussian behaviour could occur as a superposition of local Gaussians separated in time, which means the returns are normally distributed but with varying mean and variance in time. In fact, we will test this

hypothesis in the nonstationary section by imposing a time-related constraint in a Gaussian mixture model.

The second explanation is that indeed, the returns follow a non-Gaussian distribution, like for instance the stable Paretian distribution, a Weibull or a Student t distribution. Nevertheless, even with an alternative non-Gaussian parametric model, we cannot rule out a variation in the parameters of the distribution. Therefore, the stability of the non-Gaussian distribution over time will be tested later as well.

Beside using and hence determining the full density function directly, one might operate in the Fourier space instead and by doing this estimate the characteristic function, an approach we are now going to outline.

3 Characteristic function

The characteristic function $\Phi(\omega)$ is defined as the expectation of $e^{j\omega x}$ under the distribution $p(x)$ for $\omega \in \mathbb{R}$ with $j = \sqrt{-1}$, that is $\Phi(\omega) = \int_{-\infty}^{\infty} e^{j\omega x} p(x) dx$ [4; 10]. Once the characteristic function is determined, the p.d.f. can be approximated by the inverse transform $p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega x} \Phi(\omega) d\omega$.

Working with the c.f. instead of with the p.d.f. is justified by the fact that there are some distributions for which the p.d.f. is not defined, whereas the c.f. is always guaranteed to exist, that is for instance the case for the family of stable Paretian distributions [6].

In order to estimate the characteristic function one can use a kernel estimator approximating the density as $p(x) = \frac{1}{N} \sum_{i=1}^N \delta(x - x_i)$ leading to $\Phi(\omega) = \frac{1}{N} \sum_{i=1}^N e^{j\omega x_i}$ [7]. A similar method estimates the c.f. with a neural network architecture where two neural networks are constructed to approximate the real and imaginary part of the c.f. for a discrete grid of supporting points in the ω -space and to interpolate between them [3].

The result for the c.f. of the DJIA returns obtained by the neural network approach is shown in figure 4. Since the time series is real the c.f. needs to be determined only for positive ω . While in figure 4(a) the c.f. $\Phi(\omega)$ is plotted against ω we show in figure 4(b) the log of the characteristic function in order to demonstrate their diverging behaviour in the tails.

It can be observed that the tails of the original distribution are fatter than the corresponding normal distribution $\Phi(\omega) = e^{-\frac{1}{2}\omega^2\sigma^2 + j\omega\mu}$ with the same mean μ and variance σ^2 . Furthermore, while the normal distribution declines quadratically in log-space, the c.f. of the return time series seems to go down only linearly. This makes it difficult to obtain an accurate approximation of the corresponding p.d.f. and especially their tails by the inverse transformation of the characteristic function.

An alternative to these non-parametric methods is suggested in [5] where the p.d.f. was obtained from a parameterised c.f., for instance the stable Paretian distribution, via the Fourier transform. With that it will be possible to apply a maximum-likelihood approach to fit the parameters of the c.f. which account for non-Gaussian behaviour.

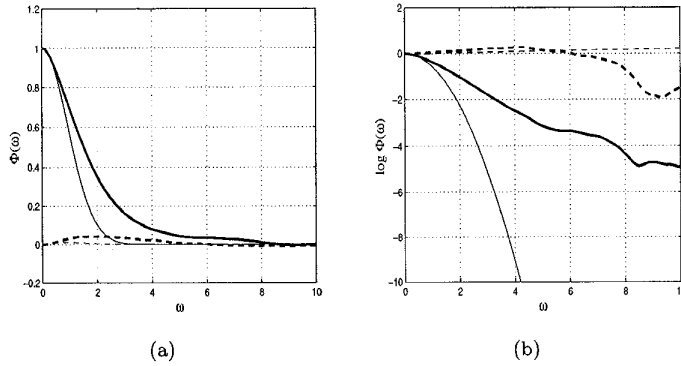


Figure 4: The empirical characteristic function for the DJIA returns (bold) and a fitted Gaussian (thin) with their corresponding real (solid) and imaginary parts (dashed)

4 Moments and Cumulants

Another way to characterise the distribution is to calculate the moments $M_n = \int_{-\infty}^{\infty} x^n p(x) dx$ or cumulants $\kappa_n = \left. \left(\log \sum_{k=1}^{\infty} \frac{s^k}{k!} M_k \right)^{(n)} \right|_{s=0}$ expressed in terms of the moments. They are related to each other via the characteristic function due to $M_n = \left. \frac{d^n \Phi(s)}{ds^n} \right|_{s=0}$ and $\kappa_n = \left. \frac{d^n \log \Phi(s)}{ds^n} \right|_{s=0}$ with $s = jt$. Assuming that the moments and cumulants represent stable characterisations of the underlying stochastic generator, they can then be analysed instead of the full p.d.f.

n	M_n	κ_n
1	0.02	0.02
2	1.14	1.14
3	-0.70	-0.77
4	35.15	31.32

Table 1: Moments and cumulants for the DJIA return time series

In table 1 we put together the moments and cumulants statistics for the (static) marginal distribution of the DJIA returns. Since the skewness reflects the degree of asymmetry of the distribution around the mean, we recognise a slightly stronger negative tail for the return distribution due to its negative skewness. Furthermore, the high value for the fourth moment supports the

findings about the leptocurtic behaviour in a previous section.

In general, taking into account those higher-order cumulants could overcome the limitations of methods based on normality, like for instance principal component analysis. Furthermore, once extended to the multidimensional case, they can also be used to detect local dependencies in joint distributions, which will be investigated as well in the next section.

5 Nonstationarity

We mentioned earlier that the variance of the DJIA return time series seems to change over time and that this might actually explain the (aggregated) non-Gaussian distribution we found. Therefore we tested the stability of the distribution of the DJIA returns by estimating the first four moments M_n for non-overlapping time windows of 253 days (approximately one year).

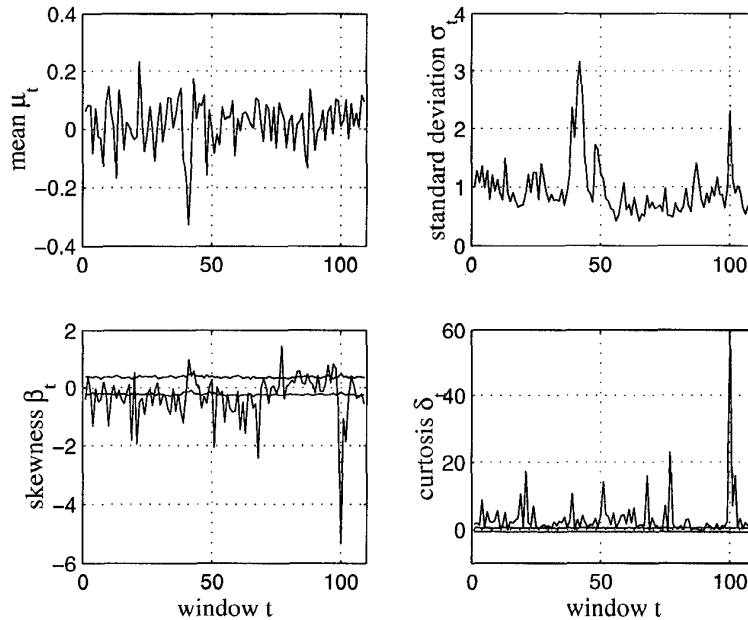


Figure 5: The empirical first four moments for the DJIA returns. Note the confidence intervals for the third and fourth moment reflecting the significance of the estimate

The results in figure 5 show a strong indication for varying moments through time. Unfortunately due to using non-overlapping windows we do not have enough windows to make further analysis in terms of dependencies

of current from previous moments yet, but this will be considered in further studies.

Nevertheless, apart from a potential varying Gaussian distribution we observe also significant non-zero higher order moments indicating a varying non-normal distribution. Thereby the significance, expressed as a 95% confidence interval, is determined by using the surrogate method [11] based on 100 scrambled versions of the original data with the corresponding mean and variance in each time window.

Regarding the skewness we find for the first two-third of the time series negative values while that became less significant in the final part apart from a short drop (late 1980s).

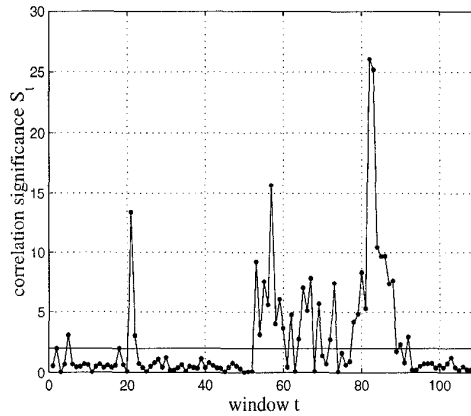


Figure 6: The empirical second order cumulant of the two-dimensional joint distribution of DJIA returns. Note the horizontal line indicating the significance level corresponding to 95% confidence

Another experiment was performed in order to analyse varying dependency in the two-dimensional joint distribution here represented by the second order cumulant. Based on [1] we calculated the significance of this correlation varying in time, again using the surrogate method. The joint distribution was estimated for the time series lagged with itself by one step. In figure 6 we show the variation of the significance of this cumulant, a value higher than 1.96 leads to rejecting the null hypothesis of independence with 95% confidence.

Finally we focused on segmentation the time series into regimes of constant distribution. Based on the mixture of experts approach [8] we applied a mixture of Gaussians separated in time to the dataset. The model is given by $p_t(x) = \sum_{j=1}^M p(x|j) P_t(j)$ whereby the component densities $p(x|j)$ are time-independent and modelled as Gaussians with adapted mean and variance, while the prior $P_t(j)$ is time-dependent. This restriction is simply applied

by evaluating a normal Gaussian mixture and determine then the priors as a local average of the conditional probabilities $P(j|x_n)$ instead of averaging over the whole dataset.

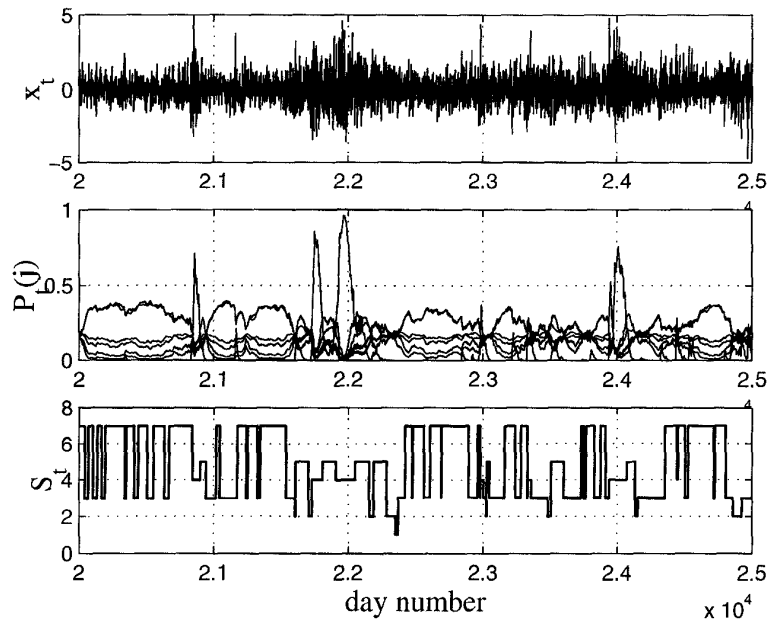


Figure 7: Segmentation of the DJIA return time series (top) with a time-constrained Gaussian mixture model of seven components into regimes of one component generating the data (bottom) based on the time-varying prior (middle)

Figure 7 shows a part of the time series together with the time-varying priors $P_t(j)$ and the corresponding state S_t as the most likely component of the mixture to have generated the data at time t .

There we observe that similar behaviour in the distribution like excessive returns are created by states four and five while the more 'quiet' periods are covered by the states seven and two. It is also worth noting that it seems that two each of the components are highly correlated to each other in terms of the prior which could be interpreted as contributing together to provide a non-Gaussian distribution localised in time.

6 Conclusion

We have presented an overview of techniques which together provide a framework for analysing financial time series using both stochastic and determin-

istic strategies. We have given evidence to support or reject various working hypotheses for how financial time series are generated, and provided a non-Gaussian segmentation driven by the nonstationarity of distributional characteristics.

References

- [1] G. Deco, R. Neuneier, and B. Schürmann. Non-parametric data selection for neural learning in non-stationary time series. *Neural Networks*, 10(3):401–407, 1997.
- [2] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, 39(1):1–38, 1977.
- [3] J. A. Dente and R. V. Mendes. Characteristic functions and process identification by neural networks. *Neural Networks*, 10(8):1465–1471, 1997.
- [4] E. Lukacs. *Characteristic Functions*. Griffin, London, 1970.
- [5] S. Mittnik, T. Doganoglu, and D. Chenyao. Computing the probability density function of the stable paretian distribution. Technical Report 98, Institute of Statistics and Econometrics, Christian Albrechts University at Kiel, Germany, 1996.
- [6] S. Mittnik and S. T. Rachev. Alternative multivariate stable distributions and their applications to financial modeling. In S. Cambanis, editor, *Stable Processes and Related Topics*, pages 107–119. Birkhäuser, Boston, 1991.
- [7] K. Nanbu. Fourier transform method to determine the probability density function from a given set of random samples. *Physical Review E*, 52(6):5832–5838, December 1995.
- [8] K. Pawelzik, J. Kohlmorgen, and K.-R. Müller. Annealed competition of experts for a segmentation and classification of switching dynamics. *Neural Computation*, 8(2):342–358, 1996.
- [9] D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley, New York, 1992.
- [10] A. Stuart and K. Ord. *Distribution Theory*. Kendall’s Advanced Theory of Statistics volume 1. Edward Arnold, sixth edition, 1994.
- [11] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94, 1992.