Multi-scale description and prediction of financial time series

To cite this article: A P Nawroth et al 2010 New J. Phys. 12 083021

View the article online for updates and enhancements.

Related content
- Capturing rogue waves by multi-point statistics
  A Hadjihosseini, Matthias Wächter, N P Hoffmann et al.
- Towards a stochastic multi-point description of turbulence
  R Stresing and J Peinke
- Towards a stochastic multi-point description of turbulence
  Robert Stresing, Matthias Wächter and Joachim Peinke

Recent citations
- Can log-periodic power law structures arise from random fluctuations?
  Jan Henrik Wosnitza and Jens Leker
- Stochastic analysis of ocean wave states with and without rogue waves
  A Hadjihosseini et al
- Parameter estimation and forecasting for multiplicative log-normal cascades
  Andrés Leövey and Thomas Lux
Multi-scale description and prediction of financial time series

A P Nawroth, R Friedrich and J Peinke

1 Triemlistrasse 49, 8047 Zürich, Switzerland
2 Institute for Theoretical Physics, University of Münster, D-48149 Münster, Germany
3 Institute for Physics, Carl-von-Ossietzky University Oldenburg, D-26111 Oldenburg, Germany
E-mail: a.p.n@web.de, fiddir@uni-muenster.de and peinke@uni-oldenburg.de

New Journal of Physics 12 (2010) 083021 (10pp)
Received 29 January 2010
Published 10 August 2010
Online at http://www.njp.org/
doi:10.1088/1367-2630/12/8/083021

Abstract. A new method is proposed that allows a reconstruction of time series based on higher order multi-scale statistics given by a hierarchical process. This method is able to model financial time series not only on a specific scale but for a range of scales. The method itself is based on the general $n$-scale joint probability density, which can be extracted directly from given data. It is shown how based on this $n$-scale statistics, general $n$-point probabilities can be estimated from which predictions can be achieved. Exemplary results are shown for the German DAX index. The ability to model correctly the behaviour of the original process for different scales simultaneously and in time is demonstrated. As a main result it is shown that this method is able to reproduce the known volatility cluster, although the model contains no explicit time dependence. Thus a new mechanism is shown how, in a stationary multi-scale process, volatility clustering can emerge.

4 Author to whom any correspondence should be addressed.
1. Introduction

A typical feature of complex systems, such as finance, is that these have hierarchical and non-trivial structures on different scales. Due to the hierarchical structure such systems are in general very high dimensional; thus there is the challenge to find appropriate methods to model these structures efficiently. Successful attempts have been made to describe such hierarchical systems as processes in scale rather than time or space itself. Examples are the description of financial data [1]–[3], roughness of surfaces [4, 5], turbulence [6]–[8] and earthquakes [9]. These successful attempts are characterized by the fact that they can correctly provide the joint probability density function $p(y_1, \tau_1; \ldots; y_n, \tau_n)$ of the log returns $y(\tau)$ of the process variable $S$ at different scales $\tau$,

$$y(t, \tau) := \log S(t) - \log S(t - \tau).$$  (1)

In the following the joint statistics of these increment processes on different scales $\tau_i$ are considered, which we refer to as multi-scale description. Thereby the smaller increments are always nested into the larger increments. Note that the method itself is not restricted to this particular increment definition [10]. It should be pointed out that the knowledge of the joint multi-scale probability density function $p(y_1, \tau_1; \ldots; y_n, \tau_n)$ includes also information such as multifractal scaling properties, typically expressed by $\langle y^n(\tau) \rangle \sim \tau^{\xi_n}$. A multifractal scaling property, also often called multiscaling property, is not necessary for our approach presented here. Nevertheless, the concept of these multifractal features is a very strong one and has been shown frequently for different financial data sets. Furthermore, multifractal features are often the basis for setting up powerful models; see e.g. [11]–[13].

For complex systems there is a general challenge: to generate surrogates or future paths based on the knowledge of the statistical properties and the recent history of the system. Most often the properties only in scale or only in time are modelled. In order to model the scale properties, one approach is to propose general models and proper parameter sets, see for example [14, 15]. Based on the feature of volatility clustering, which can be measured by the correlation or by return intervals of variables $y(t, \tau)$ and $|y(t, \tau)|$ for the same scale $\tau$ but different times $t$, also aspects of more point statistics have been taken into account; see [11, 12, 16].

Typically, remarkable features such as correlations or the shape of single scale pdfs $p(y, \tau)$ are used for the parameter selection. Some of these approaches provide also a remarkably good description of higher order statistics. In order to model the time properties of complex systems, the dependencies of consecutive points are modelled mostly without considering the form of the
distribution. Known examples are the ARCH [17], GARCH [18], MMAR [11] or MRW [12] just to mention a few; for further models, see e.g. [11, 19]. We propose an alternative, more general approach, which is based on the knowledge of multi-scale properties. In particular, we show how a given time series can be extended into the future in such a way that the scale and time properties are modelled simultaneously. In a previous work [20] the initial idea of this method has been proposed for simulating turbulence data. Here we extend the analysis to another complex system and show more details using a financial data set. In particular, we will discuss problems such as predicting volatility clustering and verifying the quality of the prediction. This issue of the prediction of features of the future time series is an actual topic. Often one tries to predict extreme events by features of ‘patterns’ in the precursor data. As categorized in [21], either the frequency of these patterns is evaluated or the pattern itself that precedes any event in the record is analysed. As an alternative also the statistics of return intervals of selected events can be analysed.

Here we do not merely focus on the occurrence of extreme events, but also we are interested in determining some stochastic features of the process in the future, independent of any threshold consideration. As another important feature, our approach allows us to break down the $n$-point precursor consideration on three-point or, respectively, on two-increment statistics.

2. Method

The aim is to compute the probability of occurrence of the event \{${x}^*, t^*$\}, conditioned on the knowledge of the past points \{${x}_1, t^* - \tau_1; \ldots; {x}_N, t^* - \tau_N$\}, where the following shorthand notation is used:

\[
{x}^* = X(t^*) = \log S(t^*). \tag{2}
\]

This probability can be expressed as

\[
P({x}^*, t^*|{x}_1, t^* - \tau_1; \ldots; {x}_N, t^* - \tau_N). \tag{3}
\]

We now assume that expression (3) has no explicit time dependence. This is a weaker assumption than the stationarity assumption for \(x^*\) itself, due to the fact that certain information, such as current volatility, is already accounted for by the knowledge of \{${x}_1, t^* - \tau_1; \ldots; {x}_N, t^* - \tau_N$\}. Thus our aim will be to determine

\[
P(x^*|{x}_1, \tau_1; \ldots; {x}_N, \tau_N). \tag{4}
\]

The two-point pdf may be introduced in the following form:

\[
p({x}^*; x, \tau) = \langle \delta(x^* - X(t))\delta(x - X(t - \tau)) \rangle
\]

\[
= \int dy \langle \delta(y - (X(t) - X(t - \tau)))\delta(x - X(t - \tau)) \rangle \cdot \delta(y - (x^* - x)), \tag{5}
\]

where the expectation is taken over the ensemble and the time variable \(t\). Statistical independence of the cascade on \(x = X(T)\) may be assumed, because the current price can always be normalized to a reference price, by describing a certain fraction or multiple of
the asset, instead of the asset itself. Therefore the expectation term can be factorized and the following expression is obtained:

\[ p(x^*; x, \tau) = \langle \delta(x^* - X(t)) \delta(x - X(t - \tau)) \rangle \]

\[ = \int dy \langle \delta(y - (X(t) - X(t - \tau))) \rangle \delta(x - (x^* - x)). \quad (6) \]

Now the case of the conditional pdf of higher order will be considered. We start at

\[ p(x^*; x_1, \tau_1; \ldots; x_N, \tau_N) \]

\[ = \int dy_1 \ldots \int dy_N \langle \delta(x_N - X(t - \tau_N)) \delta(y_1 - (X(t) - X(t - \tau_1))) \rangle \delta(x^* - x_N) \delta(x^* - x_1) \ldots \delta(x^* - x) \]

\[ = \int dy_1 \ldots \int dy_N \langle \delta(x_N - X(t - \tau_N)) \delta(y_1 - (X(t) - X(t - \tau_1))) \rangle \delta(x^* - x_N) \delta(x^* - x_1) \ldots \delta(y_N - (x^* - x_N)) \]

\[ = \int dy_1 \ldots \int dy_N \delta(y_1 - (x^* - x_1)) \ldots \delta(y_N - (x^* - x_N)) \cdot p(x_N) \cdot p(y_1, \tau_1; \ldots; y_N, \tau_N). \quad (7) \]

Thereby again the above-mentioned independence of the cascade on \( x = X(T) \) was used. Therefore it can be written as

\[ p(x^*; x_1, \tau_1; \ldots; x_N, \tau_N) = p(x^*; x_1, \tau_1; \ldots; x_N, \tau_N). \quad (8) \]

Using the same argument for

\[ p(x_1; x_2, \tau_2 - \tau_1; \ldots; x_N, \tau_N - \tau_1) \]

and combining both equations yields an expression for the conditional probability density

\[ p(x^*|x_1, \tau_1; \ldots; x_N, \tau_N) = \frac{p(x^* - x_1, \tau_1; \ldots; x^* - x_N, \tau_N)}{p(x_1 - x_2, \tau_2 - \tau_1; \ldots; x_1 - x_N, \tau_N - \tau_1)}. \quad (10) \]

Because of the involved scales the dimension of the joint probability density is very high. Therefore it is in general very difficult to compute it from empirical time series. However, the description and computation can be highly simplified if Markov properties can be assumed. This is the case if

\[ p(y_i, \tau_i|y_{i+1}, \tau_{i+1}; \ldots; y_n, \tau_n) = p(y_i, \tau_i|y_{i+1}, \tau_{i+1}) \quad (11) \]

is true for all \( i \) and \( n > i \). Without loss of generality we take \( \tau_i < \tau_{i+1} \). It should be noted that the Markov property can be tested for a given data set [2, 7], where also evidence can be found for the validity of the Markov property. In this case the joint probability density can be substantially simplified:

\[ p(y_1, \tau_1; \ldots; y_n, \tau_n) = p(y_1, \tau_1|y_2, \tau_2) \cdot \cdots \cdot p(y_{n-1}, \tau_{n-1}|y_n, \tau_n) \cdot p(y_n, \tau_n). \quad (12) \]

Most interestingly, it has also been shown that these fundamental conditioned pdfs \( p(y_i, \tau_i|y_j, \tau_j) \) and their \( \tau \) dependence can be expressed by the Fokker–Planck or Kolmogorov equation

\[ -\partial_t p(y, \tau | y_i, \tau_i) = \left( -\partial_y D^{(1)}(y, \tau) + \partial_y \partial_y D^{(2)}(y, \tau) \right) p(y, \tau | y_i, \tau_i), \quad (13) \]
where $D^{(1)}$ denotes the drift term and $D^{(2)}$ the diffusion term. Both terms can be estimated directly from the data [2, 7], [22]–[24]. In the following we restrict the discussion to Markov processes in scale and to right-justified increments, i.e. the smaller increment is nested into the larger one and has the right end point in common according to equation (1). In the case of Markov properties, equation (10) simplifies to

$$p(x^*|x_1, \tau_1; \ldots; x_N, \tau_N) = \frac{p(x^* - x_N, \tau_N)}{p(x_1 - x_N, \tau_N - \tau_1)} \frac{p(x^* - x_1, \tau_1)}{p(x_1 - x_2, \tau_2) p(x_2 - x_3, \tau_3) \ldots p(x_N - x_N, \tau_N)}$$

Using different values for $x^*$ it is possible to construct a discrete approximation of $p(x^*|x_1, \tau_1; \ldots; x_N, \tau_N)$ using equation (14) and the conditional probabilities of first order, which can be estimated directly from the data.

### 3. Results for the DAX index

In order to show the application of the above-described method, it will be applied for the DAX index for the years 1994–2003. The financial data sets were provided by the Karlsruher Kapitalmarkt Datenbank (KKMDB) [25]. The time scales used for the reconstruction are chosen as $\tau_n = 5 \times 2^n$ min with $n = 0, 1, \ldots, 6$. The sample size is quite small and therefore the initial estimate of the probability density in the tails on large scales is not very good. For certain large log returns, there is no realization in the initial sample. Therefore the corresponding probability is estimated as zero. But in order to estimate the joint probability function, which is necessary for the extrapolation, see equation (12), all conditional probabilities are multiplied. Therefore the joint probability becomes zero for these cases. In order to estimate the probability for these events, the conditional probability density functions are fitted with splines. For more details, refer to [26, 27]. In areas with only a few or no data points a kernel density estimation procedure is applied. These smoothed pdfs are now used for the extrapolation.

To give an idea of how this method works we have shown in figure 1 three different sets of data points $\{x_1; \ldots; x_N\}$ together with the constructed probability $P(x^*, t^*|x_1, t^* - \tau_1; \ldots; x_N, t^* - \tau_N)$, which is shown on the right side. In addition the unconditioned probability $P(x^* - x_1, \tau_1)$ is shown. From these three examples, it can be seen clearly how the statistics is sensitively changing with the choice of the initial points $\{x_1; \ldots; x_N\}$, and even bimodal probabilities are obtained for the next step. The main point of our method is that the actually updated conditional pdfs $P(x^*, t^*|x_1, t^* - \tau_1; \ldots; x_N, t^* - \tau_N)$ change from situation to situation, delivering a dynamics that seems not to be stationary, but the underlying scale-dependent pdfs used for the construction are purely stationary. Based on the knowledge of $P(x^*, t^*|x_1, t^* - \tau_1; \ldots; x_N, t^* - \tau_N)$ in a statistical correct sense, a new data point $x^*$ can be drawn. Shifting the above-mentioned procedure by one step, the same procedure can be used to predict the next data point, and so on. Thus we are able to generate new artificial data series.

These synthetic time series can be tested with respect to the proper $n$-scale joint probabilities $p(y_1, \tau_1; \ldots; y_n, \tau_n)$; therefore we use the method to estimate the underlying Fokker–Planck equation. In figure 2 the drift and diffusion terms are shown, which we reconstructed from the original DAX data and the reconstructed data. Based on this result we claim that both data sets underly the same hierarchical stochastic equations and thus are
Figure 1. In the left parts three different time series selected from DAX data are displayed, one with a low volatility (a), one with a higher volatility (b) and one with a kind of bimodal dynamics (c). Based on these initial conditions, the algorithm of multi-scale reconstruction is applied to forecast the pdf for the next point $x^*$ of the time series at a distance of 5 min. The pdfs are displayed in the right part of the figure. All time series were rescaled in such a way that the last point of the time series, $x_1$ has the value zero. In addition, $P(x^* - x_1, \tau_1)$, the unconditioned probability of $x^*$ values, are shown on the right (dotted line).

Stochastically equal in the sense that both data sets have the same $n$-scale statistics. The extrapolated time series were also used to calculate the return distribution $p(y, \tau)$ at the different scales. The results are shown in figure 3. As can be seen the agreement of the pdfs is quite good. Although the algorithm delivers already quite good results, the number and magnitude of scales that should be used is still an open question. Possible solutions may be the analysis of the importance of scales by a wavelet analysis or the HARCH model [28].

New Journal of Physics 12 (2010) 083021 (http://www.njp.org/)
Figure 2. The drift $D^{(1)}(y, \tau)$ (a) and diffusion $D^{(2)}$ (b) coefficients are shown. The considered time scale is $\tau = 9600$ s. The results for the original DAX time series are denoted by black dots and those for the reconstructed time series by white dots. For the results of the reconstructed time series, 15 000 small time series with a length of 57 600 s were produced.

The above-presented method for extending time series can be used to simulate many different possible extensions of an existing time series. These simulated paths can be used to analyse the evolution of statistical quantities like standard deviation, quantiles of the distribution or the probability of extreme returns above or below a certain threshold. Therefore predictions can be made about how long low- or high-volatility states, as shown in figure 1, will persist or when extreme events may occur. This would provide a connection to the return interval approach [21].

4. Prediction of volatility

The ansatz presented here allows the prediction of the whole pdf for the next point of the time series as we have shown above. Based on these considerations, the connection between
predicted and realized volatility will be analysed in more detail. It is important to note that, based on one prediction and the corresponding realization, it is not possible to determine the quality of this next step prediction. This is due to the fact that the realization of a next data point of a process with a very broad distribution can be due to its randomness close to the centre, and another realization of a process with a narrow distribution can be far from the centre. Therefore in order to determine the quality of the prediction, the following technique was set up. The data set is divided into two parts; the larger one contains the data of the years 1994–2002, while the smaller one contains the data of the year 2003. The larger data set is used to estimate the unconditional and conditional probability density functions on the right side of equation (14). Then a prediction of volatility for the second smaller data set is attempted. A sequence of points from this second data set is used as an initial condition for equation (14). Now it is possible to estimate the conditional probabilities for the next data point $x^*$ following the sequence of points used as the initial condition. A tuple with the predicted standard deviation and the realized log return is generated. In a second step, these tuples are sorted with respect to the predicted volatility given by the standard deviation of $P(x^*, t^* | x_1, t_1^* - \tau_1; \ldots; x_N, t_N^* - \tau_N)$. Based on 5000 consecutive sorted tuples, the mean predicted volatility $\sigma_{\text{prog}}$ and the mean realized volatility $\sigma_{\text{real}}$ is estimated. In figure 4 the results for the DAX index are presented. The functional dependence between $\sigma_{\text{real}}^2$ and $\sigma_{\text{prog}}^2$ is quite good and linear.

Figure 3. The pdfs obtained from the original time series (1994–2002, solid black line) and from the reconstructed (circles) data set are shown. The considered scales are $\tau = 5 \times 2^n \text{ min}$ with $n = 6, 5, 4, 3, 2, 1, 0$ from top to bottom. The pdfs for different scales are shifted for clarity of presentation. For the reconstruction process smoothing was applied.
Figure 4. The realized variance $\sigma^2_{real}$ with respect to the predicted variance $\sigma^2_{prog}$ for the DAX index is displayed. The scales used for the forecasts are $\tau = 5 \times 2^n$ min with $n = 0, 1, \ldots, 6$. A prediction was attempted for the log return in the next 5 min.

5. Conclusions

A new method has been presented that allows for a multi-scale reconstruction of a time series. More scales are modelled simultaneously, using conditional probabilities of properly chosen increments. A necessary feature of the underlying complex system is that it exhibits Markov properties of the scale-dependent process. This can be tested statistically. In this case all conditional probabilities of arbitrary order can be substituted by conditional probabilities of first order, which can be calculated easily from empirical time series. Using the empirical probability densities a time series can be reconstructed. The reconstructed time series reproduces the probability densities very well. The advantage of this method is the use of the distributions, which are obtained directly from the data. Therefore the modelling of an asymmetric or non-Gaussian distribution is possible. The reconstruction can be applied to any range of scales without a change of the underlying method, as long as Markov properties in scale are present. Furthermore, using this approach the full probability density for the next potential step of a time series is provided. Using an empirical time series as a starting point, this procedure can be applied for prediction purposes. But contrary to many other methods, not only the mean value for the next step is provided, but also the full distribution. Using this full distribution for the log returns, many common risk measures can be calculated. An examples is the standard deviation of the realizations of the process in the next step, as is demonstrated in section 4. But also, more complex risk measures such as VaR (value at risk), EL (expected loss) and ES (expected shortfall) can be easily obtained by integrating the estimated distribution for the next step of the process.

Our method provides for each historical time series a unique distribution for the realization of the process in the next step. Using a distribution-based risk measure such as VaR, EL or ES,
a direct link between the historical time series and the risk measure is obtained. If a portfolio of assets is considered, then the choice of the portfolio determines the weights for the historical time series and the time series of the considered portfolio can be easily computed. Taking the portfolio’s weights as free variables and the risk measure as an error function, an optimization of portfolios with respect to their risk characteristics is possible. Here it is of advantage that the calculation of the distribution, and hence the distribution-based risk measure, is very fast. Therefore this method can be applied to the cases where historical simulation-based methods are used, e.g. for portfolio optimization.

References

[1] Ghashghaie S, Breymann W, Peinke J, Talkner P and Dodge Y 1996 Nature 381 767
[2] Renner C, Peinke J and Friedrich R 2001 Physica A 298 499
[3] Ausloos M and Ivanova K 2003 Phys. Rev. E 68 046122
[4] Waechter M, Riess F, Kantz H and Peinke J 2003 Europhys. Lett. 64 579
[5] Jafari G, Fazeli S, Ghasemi F, Allaei S V, Tabar M R R, Zad A I and Kavei G 2003 Phys. Rev. Lett. 91 226101
[6] Friedrich R and Peinke J 1997 Phys. Rev. Lett. 78 863
[7] Renner C, Peinke J and Friedrich R 2001 J. Fluid Mech. 433 383
[8] Tutkun M and Mydlarski L 2004 New J. Phys. 6 49
[9] Tabar M R R, Sahimi M, Ghasemi F, Kaviani K, Allamehzadeh M, Peinke J, Mohktari M, Vesaghi M, Niry M D, Bahraminasab A, Tabatabai S, Fayazbakhsh S and Akbari M 2006 Short-Term Prediction of Medium- and Large-Size Earthquakes Based on Markov and Extended Self-Similarity Analysis of Seismic Data (Lecture Notes in Physics vol 705) ed P Bhattacharyya and B K Chakrabarti (Berlin: Springer) pp 281–301
[10] Karth M and Peinke J 2003 Complexity 8 34
[11] Mandelbrot B B, Fisher A and Calvet L 1997 Cowles Foundation Discussion Paper No. 1164
[12] Bacry E, Delour J and Muzy J F 2001 Phys. Rev. E 64 026103
[13] Samanidou F, Zschischang E, Stauffer D and Lux T 2007 Rep. Prog. Phys. 70 407
[14] Breymann W, Ghashghaie S and Talkner P 2000 Int. J. Theor. Appl. Financ. 3 357
[15] Muzy J, Delour J and Bacro E 2000 Eur. Phys. J. B 17 537
[16] Bogachev M, Eichner J and Bunde A 2007 Phys. Rev. Lett. 99 240601
[17] Engle R F 1982 Econnometrica 50 987
[18] Bollerslev T 1986 J. Econometr. 31 307
[19] Dacorogna M M, Gençay R, Müller U, Olsen R B and Pictet O V 2001 An Introduction to High-Frequency Finance (New York: Academic)
[20] Nawroth A P and Peinke J 2006 Phys. Lett. A 360 234
[21] Bogachev M and Bunde A 2009 Phys. Rev. E 80 26131
[22] Nawroth A P, Peinke J, Kleinjans D and Friedrich R 2007 Phys. Rev. E 76 056102
[23] Gottschall J and Peinke J 2008 New J. Phys. 10 083034
[24] Friedrich R, Peinke J and Tabar M R R 2009 Encyclopedia of Complexity and Systems Science ed R A Meyers (Berlin: Springer) pp 3574–97
[25] Lüdecke T 1998 Discussion Paper No. 190 University of Karlsruhe
[26] Green P J and Silverman B W 1993 Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach (London: Chapman and Hall)
[27] Hastie T J and Tibshirani R J 1990 Generalized Additive Models (London: Chapman and Hall)
[28] Müller U A, Dacorogna M M, Davé R D, Olsen R B, Pictet O V and von Weizsäcker J E 1997 J. Empirical Financ. 4 213

New Journal of Physics 12 (2010) 083021 (http://www.njp.org/)