ANALYSIS OF NON-LINEAR DEPENDENCIES USING THE λ-TEST

Thorsten Poddig, Ingo Möller and Kerstin Petersmeier
University of Bremen – chair of finance
Hochschulring 40, D-28359 Bremen Germany
phone: +49-421-218-7549, fax: +49-421-218-4838
email: imoeller@uni-bremen.de

ABSTRACT: In this paper we propose a modification of Pi's δ-test (Pi and Petersen, 1994), which detects non-linear functional dependencies in sequences of measurements. The λ-test just as the δ-test uses the continuity of the underlying function. Moreover it allows a decision an a significance level and reduces the number of parameters. Thus it simplifies the handling of the test. To illustrate the power of our method, we present a couple of mathematical samples.

KEYWORDS: statistical test, delta-test, dependencies, neural networks, variable selection

INTRODUCTION

Linear methods, like the analysis of correlations and autocorrelations, are most popular to examine financial data sets for functional dependencies. The problem is that it covers only a few kinds of possible dependencies. Even when the data set feeds a neural network, linear methods wouldn't be suitable to preselect the input, because it might overlook the non-linear structures and keeps back helpful information.

To avoid these problems, there are some test stategies published to find out also non-linear dependencies. The most famous among them is the BDS-test, developed by Brock, Dechert and Scheinkman for time series, that we would like to review at the beginning of this article. The presentation is quite a little unusual, but gives in this way a better connection to the following. This gave an impulse to many other test strategies to explore function dependencies. Instead of the BDS-test, which can be helpful to examine univariate time series, the δ-test by Pi and Petersen (Pi and Petersen, 1993) provides an expansion into multivariate analysis. In the last years the δ-test was applied on different financial time series, e.g. currency exchange rates (Pi, 1993) or british yields (Rehkugler et al.,1996). In many cases it was associated with neural network training.

A lack of the δ-test is the need to adjust two parameters in the test procedure and the problem of coming to statistical sifgnificant statements. This paper gives a suggestion to diminish the difficulties with the δ-test and presents an application of the modified method to some mathematical series and to financial data.

First of all it is helpful to distinguish different kinds of dependencies. When a (dependent) variable can be explained completely by other, independent variables, we call this in our nomenclature deterministic. In the analysis of time series the independent variable could consist of earlier values of the same variable. In other words there is a functional dependence between the dependend and the set of the independend variables. If there is a indeterminable part in the system, we denote it as stochastic. This includes noise originates from random influence or from a too minor set of independent variables. Furthermore it could happen that there are no dependencies in the data points and so the system is completely random.

The article is structured as follows: the next section gives a brief review on the BDS-test. An introduction into the methods of the δ-test is described in section 3 and our extensions to the λ-test is in section 4. We applied the λ-test to some mathematical examples and the results are presented in section 5. Section 6 consists of a brief summary.

The BDS-test

Consider a time series $x(i)$ with $1 \leq i \leq M$ and assume, that we try to find out whether $x(i)$ is stochastic generated or whether it is completely random. According to the BDS test, we form m-tuples of m subsequent elements from the time series.

$$\begin{aligned} \mathbf{x} &= \big(x(i),\, x(i-1), \ldots,\, x(i-m+1)\big) \\ &= (x_1,\, x_2, \ldots,\, x_m) \end{aligned} \tag{1}$$

Every m-tuple contents an element with its history and so we call it an m-hitory. We have only $N = M - m + 1$ of the tuples to complete each of them. If the time series is generated by a stochastic process, we assume it in a form

$$x_1 = f(x_2, \ldots, x_m) + r \qquad \text{for all } m\text{-histories} \tag{2}$$

The function $f(x_2, \ldots, x_m)$ might be non-linear and r represents an indeterminable part originates either from insufficient information or from a random noise. If the dependence is deterministic in terms of the independent variables, r vanishs.

The BDS test starts with the null hypothesis, that the elements in a time series are independent and identically distributed. Now the test investigates the probability of two m-histories $\mathbf{x}(i)$, $\mathbf{x}(k)$ differs in a component x_1, $x_2, \ldots,$ x_m at most by an amount of ϵ. To bring it in a formal manner, we note down the difference of two m-histories concerning a component j with a small $\scriptstyle\Delta$

$${\scriptstyle\Delta} x_j(i,k) = |x_j(i) - x_j(k)| \tag{3}$$

The probability for any given pair of m-tuples has a difference of equal or less than ϵ can be denoted as $\mathrm{P}({\scriptstyle\Delta} x_1 < \epsilon)$. To find out, whether a pair of m-tuples differs in *every* component by an amount of ϵ, it is comfortable to determine first of all the maximum absolute difference. If the maximum difference is equal or less than ϵ, this is valid for all components. Therefore we denote with a larger Δ

$$\Delta x_m(i,k) = \max_{j=1,\ldots,m} |x_j(i) - x_j(k)| \tag{4}$$

and yield as the probability for the difference between a pair of m-tuples is equal or less than ϵ in *each* component

$$\mathrm{P}(\Delta x_m < \epsilon) = \mathrm{P}\big(({\scriptstyle\Delta} x_1 < \epsilon) \wedge ({\scriptstyle\Delta} x_2 < \epsilon) \wedge \cdots \wedge ({\scriptstyle\Delta} x_m < \epsilon)\big) \tag{5}$$

If the sequence is independent according to the null hypothesis, then the conjunction at the right hand side of (5) separates into a product of the form

$$\mathrm{P}(\Delta x_m < \epsilon) = \mathrm{P}({\scriptstyle\Delta} x_1 < \epsilon)\mathrm{P}({\scriptstyle\Delta} x_2 < \epsilon) \cdots \mathrm{P}({\scriptstyle\Delta} x_m < \epsilon) \tag{6}$$

With the presumption of an identical distribution, each probability on the right hand side is equal, so we can summarize it to (for an index $m = 1$ we can denote both a small or a large Δ)

$$\mathrm{P}(\Delta x_m < \epsilon) = \mathrm{P}(\Delta x_1 < \epsilon)^m \tag{7}$$

The BDS test exploits this equation by estimating the probabilities from the data set and testing it for equality. The number of all pairs of m-histories, which distance is equal or less than ϵ gives an estimation for the probabiltities [1] , i. e.

$$\hat{C}_m = \hat{\mathrm{P}}(\Delta x_m \leq \epsilon) = \frac{2}{N(N-1)}\, \mathrm{card}(\Delta x_m \leq \epsilon) \tag{8}$$

---

[1] We denote the number of elements in a set fulfilling a certain condition as *card*. Thus $\mathrm{card}(\Delta x \leq \epsilon)$ is the number of all pairs fulfilling the condition $\Delta x \leq \epsilon$. Sometimes we denote in more detail $\mathrm{card}(1 \leq i < k \leq N \mid \Delta x(i,k) \leq \epsilon)$.

A problem is now to recognize real difference and not to trap into random equality. This point were contributed by Brock, Dechert and Scheinkman. They derived a normalization factor[2] σ and demonstrated that in a limit of an infinitesimal time series the following expression is a statistic normally distributed with a mean of zero and a standard deviation of one.

$$z = \frac{\sqrt{N}}{\hat{\sigma}} \left(\hat{C}_m - \hat{C}_1^m \right) \tag{9}$$

So it can be constructed a test-procedure to take a decision on a given significance level with the normal distribution.

But there are also some restrictions using the BDS-test. The test founds on the null hypothesis of an identical and independed distribution. Rejecting the null hypothesis only shows, that one of the assumptions is violated. This could be the independence of the elements in the time series, but another possibility is the not identical distributed random term. Even in financial time series the heteroskedastic behaviour is reported many times. Other problems like trends or unit-roots could happen, too. The rejection of the null hypothesis gives only a hint, that the data hides unknown structures.

Another topic is the size of the data set. There are papers (Ramsey and Yuan, 1989) showing, that the BDS-test needs a huge number of m-tuples to provide valid results. In financial analysis these volume does not exist in most cases. The correct choice of the embedding dimension, i.e. the parameter m, is also a problem, althought there are suitable solutions suggested (Savit and Green, 1991).

Most considerable is the restriction to find dependencies just *within* the time series. It is not possible to check dependencies between different series. However these kind of analysis is in financial forecasting most important. A way to overcome this problem gives the following δ-test (Pi, 93).

THE δ-TEST

Let us consider a sequence of measurements $y(i), 1 \leq i \leq N$ on a dependent variable y. The elements do not need to be a time series. Furthermore we have a set of independent variables $x_1, \ldots, x_m$, so each $y(i)$ belongs to a tuple $x_1(i), \ldots, x_m(i)$. The x_k can be factors of external influence, in a time series it could be earlier observations of the dependent variable y. We assume an even non-linear map given by the expression

$$y = f(\mathbf{x}) + r = f(x_1, \ldots, x_m) + r \tag{10}$$

Pi made the assumption that $f(\mathbf{x})$ is a continuous map for nearly all $\mathbf{x}$. In other words the map $f(\mathbf{x})$ has only a countable number of uncontinuities. To exploit this definiton of continuity it gives the statement, that for all $\epsilon > 0$ and for (nearly) all $\mathbf{x}$ exists a $\delta > 0$ with

$$|f(\mathbf{x}) - f(\mathbf{x}')| \leq \epsilon \quad \text{if} \quad |\mathbf{x} - \mathbf{x}'| \leq \delta \tag{11}$$

Any pair of m-tuples, which are close to each other concerning the y-values, are even close concerning their x-values. With the presence of noise, however, the statement has to be formulated weaker. We drop the strong unequation and proceed to a probabilistic representation. As in the section before we denote the distance between any two data points with regard to the y- and $\mathbf{x}$-values

$$\begin{aligned} \Delta y(i,k) &= |y(i) - y(k)| \\ \Delta x(i,k) &= \max_{j=1,\ldots,m} |x_j(i) - x_j(k)| \end{aligned} \tag{12}$$

---

[2] It is given as

$$\begin{aligned} \hat{K} &= \frac{6}{N(N-1)(N-2)} \operatorname{card}\left(1 \leq i < j < k \leq N \;\middle|\; \Delta x_1(i,j) \leq \epsilon \;\wedge\; \Delta x_1(i,j) \leq \epsilon\right) \\ \hat{\sigma}^2 &= 4\left[\hat{K}^m + 2\sum_{j=1}^{m-1} \hat{K}^{m-j}\hat{C}_1^{2j} + (m-1)^2\hat{C}_1^{2m} - m^2\hat{K}\hat{C}_1^{2m-2}\right] \end{aligned}$$

Now we can construct a conditional probability $\mathrm{P}(\Delta y \leq \epsilon | \Delta \mathbf{x} \leq \delta)$, which is the probability to find a pair of points with an $\mathbf{x}$-distance equal or less than ϵ from the pairs with an y-distance equal or less than δ. In the presence of continuity one would expect a dependency on δ, because the mean distance within the $\mathbf{x}$-values increases with the mean distance in y. If continuity isn't discussed, the probability should be constant or the y-distances are *independent* from the x-distances. In this case we can split the implication into

$$\mathrm{P}(\Delta y \leq \epsilon | \Delta \mathbf{x} \leq \delta) = \frac{\mathrm{P}(\Delta y \leq \epsilon)\, \mathrm{P}(\Delta \mathbf{x} \leq \delta)}{\mathrm{P}(\Delta y \leq \delta)} = \mathrm{P}(\Delta \leq \epsilon) \tag{13}$$

The probability to find a pair of m-tuples within an ϵ-surrounding do not depend on the parameter δ. Again the propabilities are estimated by the frequency of the $N(N-1)/2$ pairs, which obey the certain condition. Thus

$$\hat{\mathrm{P}}(\Delta y \leq \epsilon) = \frac{2}{N(N-1)}\, \mathrm{card}(\Delta y \leq \epsilon) \tag{14}$$

and

$$\hat{\mathrm{P}}(\Delta y \leq \epsilon | \Delta \mathbf{x} \leq \delta) = \frac{\mathrm{card}(\Delta y \leq \epsilon\ \wedge\ \Delta \mathbf{x} \leq \delta)}{\mathrm{card}(\Delta y \leq \delta)} \tag{15}$$

In a practical application of the δ-test there are still at least two difficulties to solve. Suitable surroundings have to be selected for ϵ and δ and a plot with the probability $\mathrm{P}(\Delta y \leq \epsilon | \Delta \mathbf{x} \leq \delta)$ as a function of δ for different values of ϵ is often used to cover the parameters. The various articles suggested appropriate formula to mean over a parameter range, but a general answer to this question is not given. Also there is no decision rule to reject a null hypothesis on a certain significance level as the BDS-test provides. The applicator himself have to find out whether he accept the hypothesis of dependence by examining the plots. The test values gives only a hint, not more. So we propose a modification of the δ-test.

THE λ-TEST

We retain the principal proceeding of the δ-test and turn the continuity into the presence of dependence. Our null hypothesis assumes no dependence between y and $\mathbf{x}$. Especially the distance between the y-values are independent of the distance in the $\mathbf{x}$-values. We remark that the distribution of y-differences is not presumed, although a normal distribution would simplify the method.

Now generate all combination of pairs and arrange them in order to their distance with regard to the $\mathbf{x}$-difference. Figuratively spoken we have pairs of nearby $\mathbf{x}$-values on one side and pairs of far away $\mathbf{x}$-values on the other one. After that sorting set a parameter λ to a value within the interval $0 < \lambda < 1$ and divide the pairs into two groups. Group G contents the first $n_G = \lambda\, N(N-1)/2$ pairs (the closer one) and group H contents the last $n_H = (1-\lambda)\, N(N-1)/2$ pairs (the more distant one). As a consequence of the null hypothesis this arrangement and dividing must not have any effect of the random y-differences. In particular the two groups should be samples of one and the same distribution. Of course we have in mind, that in presence of a function continuity the group G has likewise closer y-differences and so it would not be seen the same distribution. Here we can apply several statistical tests. The non-parametric Wilcoxon-test for example examines, whether two samples come from the same unknown distribution. Equivalent is the U-test by Mann and Whitney and for that we give a brief summary [3].

Consider a sample G with a size n_G and another sample H with the size n_H as introduced above. A pair of elements (g_i, h_k) and $g_i \in G$ and $h_k \in H$ is called an inversion, if $g_i < h_k$. The statistic of the test is constructed by the number of inversions U, so two samples drawn from the same distribution should not show too far deviation from the expected value $\mathrm{E}U = (n_G\, n_H)/2$. In the limit of large n_G and n_H the expression

$$z = \sqrt{\frac{12}{n_G\, n_H (n_G + n_H + 1)}}\ \left| U - \frac{n_G\, n_H}{2} \right| \tag{16}$$

---

[3] With the assumption of a standard normal distribution it is sufficent to test the equality of means in both samples. That would be provide with the application of an ordinary t-test.

is standard normally distributed and the null hypothesis will be rejected, when $z \geq \Phi^{-1}(\alpha)$ with $\Phi^{-1}(x)$ is the quantile of the standard normal distribution and α the significance level.

The advantages of this modification in the δ-test are the rejection on a significance level. Now it is possible to come to an valid decision (in a statistical sense). Furthermore the problems dealing with the parameter settings are reduced. The question how to choose the remaining parameter λ is still not answered, but we will give a few annotation. Of course λ must not be too close to $\lambda = 0$ or $\lambda = 1$, otherwise the approximation of an normal distribution can not be kept. On the other hand we have to bear in mind the idea of functional continuity. The value should not be to far from $\lambda = 0$. We want to compare the y-distance of data points nearby with the typical y-distance. If we expand the meaning of neighbourhood, we would not get information about the continuity but rather diffuse hints about the data structure.

Some examples

Random numbers

We ensure the capacity of the λ-test by applying it to some examples. First we feed the test with two quasi-independent and identically distributed sequences generated by a random number generator. The dependent variable was uniformly distributed in every case, but for the independent variable we created several series with different distributions: uniform, Gaussian probability and χ^2-probability. Each sequence consists of 500 data points.

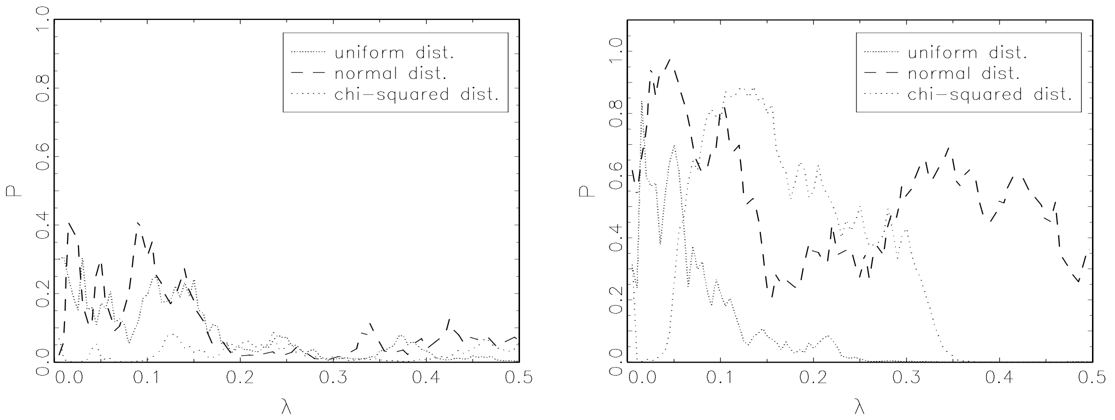

Figure 1: Trying the λ-test by presenting random series. The dependent variable was uniformly distributed on the left hand side and normally distributed on the right hand side, the independent variable was of the form the legend explains. Each series consist of 500 points. The null hypothesis are not rejected in general, all in all we get a fidgety plot.

As in the further text we chose a diagramm, that expresses in our opinion the results at best. We plot the significance level to reject the null hypothesis as a function of λ, i.e. for each sample split we calculated the error probability α to accept the null hypothesis wrongly and turn it to $1 - \alpha$. As remarked in the previous chapter we focus in the discussion on values for λ in a region not too close to zero, because of the asymmetric splitting of both subsamples and not too far from zero, because of the idea of function continuity.

In nearly all cases the λ-test hold the null hypothesis of independence on a 95-percent-level. Even if for some single aspects λ the critical value would be crossed, the complete behaviour were unstable and can not be taken as a valid refuse. Nevertheless we emphasize, that strange distributed variables might lead to an in fact not existing functional dependence. The problem is common for nearly all kind of test procedures, even in a linear context. So a careful analysis of the variable distribution can not be omitted.

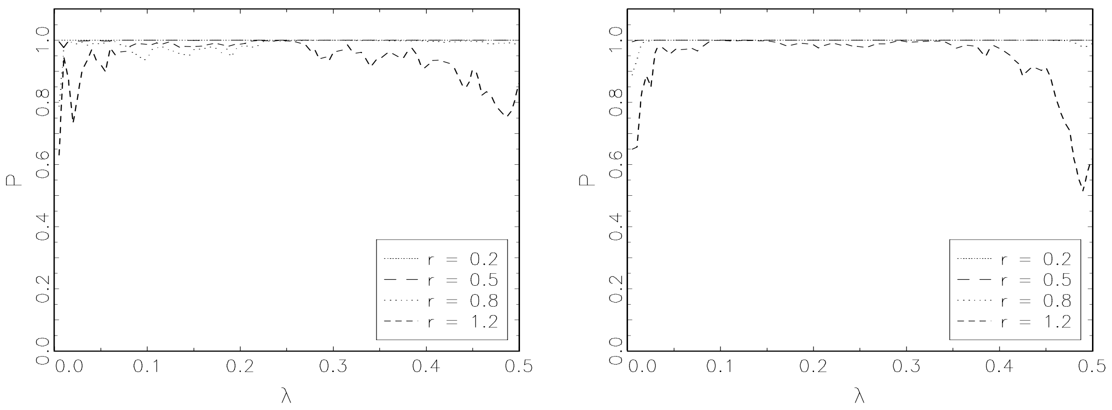

Figure 2: The λ-test on a noisy logistic map is diplay on the left hand side, the Hénon map is on the other side. In both plots it was a sequence of 500 points generated with a start value $x_0 = 0.1$ (and $x_1 = 0.1$). After that we add a normally distributed noise of different strength. As shown in the plot, the test is except for fairly large values of r be able to signal the underlying deterministic process.

LOGISTIC MAP AND HÉNON MAP

The next simple system to which to apply the λ-test is the logistic map. It is defined by the equation

$$x_{n+1} = 4x_n(1 - x_n) \tag{17}$$

The map generates a chaotic sequence in the unit interval. We created a series by starting with a seed $x_0 = 0.1$ and add different degrees of normally distributed noise. The standard deviation σ is listed in the legend. As we can see in the plot, our test captures the functional dependencies quite good. The added noise does not trouble the effectiveness of the test. Only for very large noise the results become unclearly.

Another function known in the non-linear dynamics is the Hénon map. It is the simplest example for a two-dimensional map. We fixed the general Hénon map to a specific one defined by the equation

$$x_{n+1} = 1 - 1.4x_n^2 + 0.3x_{n-1} \tag{18}$$

As the seeds for the time series we set $x_0 = 0.1$ and $x_1 = 0.1$. Like we did with the logistic map normally distributed noise were added. The strength given by the standard deviation is specified in the figures. As in the logistic map, the test procedure clears up the dependence.

SUMMARY

The article suggests a modification on the δ-test by H. Pi and C. Petersen. In our approach we retain the idea to exploit the function continuity. The extension consists of splitting all pairs of data points into two groups with different distances in the space of independent variables. After that, we test the null hypothesis, whether both groups were samples of the same distribution with regard to the values of the dependent variable. We can perform the test of identity without a presumption of a specific distribution, in this paper we applied the U-test by Mann and Whitney, but other tests were suitable. So we come to a statistical significant decision. Another advantage is the reduction of the number of adjustable parameters.

We applied the method to two non-linaer maps – the logistic and the Hénon map – and to random variables. We would like to point out, that the analysis of dependencies is even relevant for feedforward network training. The power of artificial neural networks have to be seen in the utilization of non-linear structures, but for the reason of sufficient simplicity a preselection is common strategy. This preselection should take the non-linearity into consideration and the λ-test can be helpful in doing so.

References

Brock, W.A.; Dechert, W.D.; Scheinkman, J.A., 1997, "A test for independence Based on the correlation Dimension", University of Wisconsin, Madison, University of Housten, University of Chicago.

Brock, W.A.; Hsieh, D.A.; LeBaron, B., 1992, "Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Economic Evidence", MIT Press, Cambridge.

Grassberger, P.; Procaccia, I., 1983, "Measuring the Strangeness of Strange Attractors", *Physica D*, **9**, p. 189-208.

Hartung, J., 1998, "Statistik", München.

Kerling, M., 1998, "Moderne Konzepte der Finanzanalyse: Markthypothesen, Renditegenerierungsprozesse und Modellierungswerkzeuge", Bad Soden.

Pi, H., 1993, "Dependency Analysis and Neural Network Modeling of Currency Exchange Rates", *Proc. of the first International Workshop on Neural Networks in the Capital Markets*, London.

Pi, H.; Petersen, C., 1994, "Finding the Embedding Dimension and Variable Dependencies in Time Series", *Neural Computation*, **6**, p. 509-520.

Ramsey, J.B.; Yuan, H., 1989, "Bias and Error Bars in Dimension Calculations and their Evaluation in some Simple Models", *Physical Letters A*, **134**, p. 287-297.

Rehkugler, H.; Kerling, M.; Jandura, D., 1997, "Der Delta-Test – Datenselektion und nichtlineare Finanzanalyse am Beispiel der Prognose der britischen Zinsentwicklung", internal working paper, University of Freiburg.

Savit, R.; Green, M., 1991, "Time Series and Dependent Variables", *Physica D* **50**, p. 95-116.