Research Report
Department of Statistics
Göteborg University
Sweden

On monitoring of environmental and other autoregressive processes

Magnus Pettersson

Research Report 1998:5
ISSN 0349-8034

Mailing address:	Fax	Phone	Home Page:
Department of Statistics	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se
Göteborg University	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
Box 660			
SE 405 30 Göteborg			
Sweden			

On monitoring environmental and other autocorrelated time series

by

Magnus Pettersson

Department of Statistics,
School of Economics and Commercial Law, Göteborg University,
Box 660, SE-405 30 Göteborg, Sweden

Summary

Statistical surveillance is used for monitoring a sequence of data arriving step by step. These techniques have been applied in many places in society and lately the interest and need for rational methods to be used on environmental data have been growing. In many cases, both for environmental time series and time series from other applications, the data is not independent. This is a violation against the requirements for most standard tools that are used in practice and have to be handled in some way.

This licentiat thesis consists of two parts: A case study on fish catches (1) and a study of the properties of some methods used to monitor time series (2).

In the first paper, a case concerning past data from landed catches of six economically interesting fish species in Lake Mälaren in central Sweden is studied. In 1990 the catches of vendace (*Coregonus albula*) suddenly dropped and the question discussed is whether statistical process control methods are useful for monitoring similar data. The data is examined from both univariate and multivariate viewpoints. In the univariate part, the construction of an alarm procedure for a change in the mean in an AR(1) process is briefly discussed, with this application in mind. The main conclusion is that statistical methods could have been useful for this application.

In the second paper, comparisons between two methods often suggested in literature to be used for AR(1) processes are presented. Further, comparisons are made with a direct Shewhart and a likelihood ratio based method. We can conclude that neither of the two main alternatives studied here is uniformly the best choice. The residual method works best for immediate detection.

1. Pettersson, M. (1998). Monitoring a Freshwater Fish Population - Statistical Surveillance of Biodiversity. *Environmetrics.* 9, pp 139-150.

2. Pettersson, M. (1998). Evaluations of some methods for statistical surveillance of an AR(1) process. *Research Report* 1998:4, Department of Statistics, Göteborg University.

ENVIRONMETRICS
Environmetrics, 9, 139–150 (1998)

MONITORING A FRESHWATER FISH POPULATION: STATISTICAL SURVEILLANCE OF BIODIVERSITY

MAGNUS PETTERSSON*
Department of Statistics, Göteborg University, Box 660, SE-40530 Göteborg, Sweden

SUMMARY

Statistical surveillance comprises methods for repeated analysis of stochastic processes, aiming to detect a change in the underlying distribution. Such methods are widely used for industrial, medical, economic and other applications. By applying these general methods to data collected for environmetrical purposes, it might be possible to detect important changes fast and reliably. We exemplify the use of statistical surveillance on a data set of fish catches in Lake Mälaren, Sweden, 1964–93. A model for the 'in control' process of one species, vendace (*Coregonus albula*), is constructed and used for univariate monitoring. Further, we demonstrate the application of Hotelling's T^2 and the Shannon–Wiener index for monitoring biodiversity, where a set of five economically interesting species serve as bioindicators for the lake. © 1998 John Wiley & Sons, Ltd.

KEY WORDS vendace; recursive residuals; Shewhart test; AR process; Fourier series; species correlation matrix; Shannon–Wiener index; Hotelling's T^2; Lake Mälaren; catch data

1. INTRODUCTION

There is a growing interest in studying fundamental changes in the earth's environment which is creating new opportunities for people dealing with environmental data. Often politicians, biologists and others try to find out if changes in our environment have occurred by monitoring one or more variables of ecological interest over time. Topics of interest include global warming, deterioration of water or soil quality, increasing incidence of cancer diseases caused by environmental factors, and changes in biodiversity. The increasing awareness and interest in the status of the environment has given rise to large data collection programmes. However, there is a risk that data are only being collected and stored and are not dealt with in a systematic way.

The ordinary hypothesis testing approach is to divide the data into two disjoint sets: before and after a possible change point at an unknown time. However, these tests cannot be reused directly. Since we are monitoring data to be able to detect a possible change at an unknown time and make repeated analyses, we have to use statistical surveillance instead (Wetherill and Brown 1991).

* Correspondence to: M. Pettersson, Department of Statistics, Göteborg University, Box 660, SE-40530 Göteborg, Sweden. e-mail: Magnus.Pettersson@statistics.gu.se

Contract grant sponsor: Swedish National Council for Research in the Humanities and Social Sciences

CCC 1180–4009/98/020139–12$17·50

Received 10 April 1997
Accepted 19 September 1997

By using these techniques, it might be possible to design procedures for monitoring changes in the environment and to sound an alarm as a change in the system, quickly and accurately.

This paper will give an introduction to the use of statistical surveillance in environmental science. We will study a case from a data set on fish catches in Lake Mälaren in Sweden, where we will be able to evaluate the usefulness of these statistical methods in monitoring the environment. We will study the detection of change in the level of one species, by using univariate monitoring procedures, and extend the model for monitoring the correlation between the species. The emphasis in the paper is on identifying a useful model that can be used to transform the data into a form where standard SPC methods can be applied.

Data from the catches of fish made by professional fishermen around Lake Mälaren in central Sweden have been collected since 1964 by the four regional authorities surrounding the lake. Of the species living in the lake, six have a major economic interest: burbot (*Lota lota*), eel (*Anguilla anguilla*), perch (*Perca fluviatilis*), pike (*Esox lucius*), pike-perch (*Lucioperca lucioperca*) and vendace (*Coregonus albula*). We will use five of them as indicators of the biodiversity in the lake and evaluate the performance of different monitoring procedures. The eel has been excluded since its population is dependent on artificial breeding, and is therefore increasing over time. We will not discuss the relevance of these specific species as bioindicators but instead concentrate on the statistical aspects of the problem.

From 1987, the catch of vendace decreased. At first, this decline was considered part of an assumed 6–8 year cycle of all fish in the lake, but when the expected increase did not occur in 1990 the authorities began searching for a possible cause. As we will see below, period lengths other than that assumed might better fit the data. No statistical analysis to detect departures from the 'in control' pattern has been performed previously. We will study the data material, kindly provided by the Fresh Water Laboratory in Drottningholm, from different viewpoints.

The aim of this paper is to evaluate, retrospectively, how different statistical models and methods may be applied to the current application. Although we are certain now that something happened in 1989 or 1990, we will go back in time and, without using this prior information, see what would have been done with the data available at each time point. As part of the technique in this situation, recursive residuals (Brown *et al.* 1975) will be used.

The analysis described in this paper is based on the landed catches of fish made by professional, mostly part-time, fishermen. Since we lack information about the effort, we will only use the catch data for analysis. Assuming that these figures are correlated with the abundance of each fish species, these catch data will suffice. Official statistics on the number of fishermen and the value of their equipment give reason to believe that the fishing activity has been fairly constant over time.

In Section 2, we will give an overview of statistical surveillance methods. In Sections 3 and 4, a data set from Lake Mälaren is studied using different models for the data to show the impact of model selection. Section 3 focuses on one species, vendace (*Coregonus albula*), while Section 4 discusses application of multivariate methods on five species at the same time. Finally, Section 5 discusses the conclusions and ideas for further study.

2. STATISTICAL SURVEILLANCE

Often data arrive one by one or in groups at discrete time steps. When the system producing the sequence of measurements behaves in some predicted or prescribed way we say that it is 'in control'. We assume that at a stochastic time τ the system leaves that state and goes 'out of control'. The aim of the surveillance procedure is to detect when the system goes 'out of control', under some given performance criteria, e.g. fixed false alarm probability at a certain time. In many cases *ad hoc* methods are constructed or data are viewed by an expert, who decides whether to take action or not. Using methods of statistical surveillance makes the monitoring more accurate since the performance of the methods can be evaluated and different methods can be compared with each other.

Statistical methods for detecting changes in the underlying distribution of a sequence of data have been used in many other applications. Examples from medicine, economics and forensic science can be found in Frisén (1992; 1994), Arnkelsdóttir (1995), Sveréus (1995) and Charnes and Gitlow (1995). Earlier among others Berthoux *et al.* (1978), Kjelle (1987), Settergren Sörensen and la Cour Jansen (1991) and Vaughan and Russell (1983) have applied SPC to environmental data.

We have a process of stochastic variables $X(t)$, $t = 1, 2, \ldots$, which can be univariate or multivariate, i.e. $X(t)$ is a vector of dimension $p \times 1$. Note that $X(t)$ is monitored at *discrete* time points. Further, we define the cumulated process up to time s as $X_s = \{X(t), t = 1, 2, \ldots, s\}$.

At each time step s we will formulate two possible states that we want to distinguish between: $D(s)$ and $C(s)$, that is whether the system is 'in control' or 'out of control' at time s, respectively. Given the data X_s, we will evaluate the evidence of $C(s)$ versus $D(s)$ to a specified level of certainty. Note that even if we use a formulation similar to hypothesis testing this is not the case.

The 'out of control' alternative $C(s)$ will be formulated differently for different applications. In this paper, we describe statistical surveillance in the case of a change in the mean of one species. The choice of critical event can have important effects on the performance of the surveillance procedure used (Sveréus 1995).

2.1. Recursive residuals

The 'in-control' model can also contain parameters with unknown values that have to be estimated. Two natural ways to deal with these unknowns are to estimate them during a 'run-in period' or to update the estimates in each time step by using the cumulated data. The use of recursive residuals (Brown *et al.* 1975) is an example of the latter idea. Instead of monitoring the process $\{X(t)\}$ we use the *residual process*

$$R(t) = X(t) - \hat{\mu}_{t-1}(X(t)), \tag{1}$$

where $\hat{\mu}_{t-1}(X(t))$ denote the expected value of $X(t)$ estimated using X_{t-1}. The new process, $\{R(t)\}$, is monitored with some univariate method. For example, when the mean level is constant, but unknown,

$$\hat{\mu}_{t-1}(X(t)) = \frac{1}{t-1}\sum_{i=1}^{t-1} X(i).$$

Similarly, an ARMA process can be monitored from the forecast errors, i.e. the residuals between the real values and their forecasts. For example, since the forecast errors are i.i.d. with the same distribution as $\varepsilon(t)$ (Wei 1990), an AR(1) process $X(t) = \phi_1 X(t-1) + \varepsilon(t)$ can be monitored using

$$R(t) = X(t) - E(X(t) \mid X_{t-1}) = X(t) - \phi_1 X(t-1),$$

where ϕ_1 have been estimated during 'run-in'.

2.2. Methods

Several methods for detecting the change in distribution have been designed. For univariate problems the first method was the Shewhart chart (Shewhart 1931), followed by CUSUM (Page 1954), EWMA (Roberts 1959) and the likelihood ratio method (Shiryaev 1963; Frisén and de Maré 1991). Bayesian approaches can be found in Zacks (1983). In this paper, we will only use Shewhart tests on the residuals and forecast errors – not because it is the optimal method, but because it is easy to apply and therefore suitable for benchmarking. With the Shewhart method, an alarm is triggered when the last observation exceeds a critical limit, i.e. when $|X(s)| > c$. The limit c in traditional SPC literature is set to $3{\cdot}09\sigma(s)$ or $3\sigma(s)$, where $\sigma(t) = \sqrt{\mathrm{var}(X(t))}$ (Wetherill and Brown 1991).

For multivariate problems, two natural strategies are either to monitor each process separately (an alarm is triggered at the first alarm of an individual process) or to transform the data into a univariate sequence. The likelihood ratio method can equally well be applied for a multivariate sequence as for a univariate one. A survey of methods for detecting changes in more than one variable can be found in Wessman (1996). In this paper we will study the Hotelling's T^2 statistic (Hotelling 1947) and the Shannon–Wiener index (Shannon and Weaver 1949).

3. MODELLING AND MONITORING VENDACE

In this section we will study the data for vendace (*Coregonus albula*) from a univariate point of view. We will suggest different models for the 'in control' state, compare them and discuss their performance on the data set. We will study models where the mean is considered to be constant or periodic. Further, we will use a model where we assume data to be an aperiodic ARMA(p, q) process. Ideally, the 'in control' state should be given by knowledge of the biological process, but in this paper we will have to use data to determine it. Further, we will re-estimate the parameters at each time to show the impact of re-estimation on the surveillance.

For current purpose, we find it sufficient to describe the alternative models by the residual mean squares, RMS. Suppose we estimate l parameters in the model using $\{X(1), \ldots, X(s-1)\}$; we denote the estimated expected value of $X(i)$ by $\hat{\mu}_{l,s-1}(X(i))$ and define

$$\mathrm{RMS}(X_s, s-1) = \frac{1}{s-l}\sum_{i=1}^{s}(X(i) - \hat{\mu}_{l,s-1}(X(i)))^2. \tag{2}$$

Although the present data series only consists of 30 time steps, we see possible periodic patterns. Using a frequency domain approach, we can estimate the periodically varying mean. Assume we have an additive process with independent and known mean and constant variance, i.e. $X(t) = \mu_l(t) + \varepsilon(t)$, where $\varepsilon(t)$ are i.i.d., $N(0, \sigma)$. Then the transformed process $X^c(t)$, defined by $X^c(t) = X(t) - \mu_l(t)$, becomes a white noise process that can be used for surveillance. When $\mu_l(t)$ is unknown we will in the following use estimated values $\hat{\mu}_l(t)$, and when also l is unknown we estimate l first and then estimate $\mu_l(t)$ using $\hat{l}$ and get $\hat{\mu}_{\hat{l}}(t)$.

3.1. Modelling

We will first study a model where the mean function is any function with period l, i.e. for some l we have $\mu_l(t) = \mu_l(t+l)$. For a given l we estimate μ_l by using the disjoint time subsets

$\{t, t+l, t+2l, \ldots \leqslant s\}$ and we define $N_{l,s}(t) = \#\{t, t+l, t+2l, \ldots \leqslant s\}$, for $t = 1, \ldots, l$. The maximum likelihood estimate for $\mu_{l,s}(t)$, given l, becomes

$$\hat{\mu}_{l,s}(t) = \frac{1}{N_{l,s}(t)} \sum_{i \in \{t, t+l, t+2l, \ldots \leqslant s\}} X(i), \quad t = 1, \ldots, l-1.$$

The estimated variance σ^2 using X_s for the estimation becomes

$$\hat{\sigma}_s^2 = \frac{1}{s-l-1} \sum_{i=1}^{s} (X(i) - \hat{\mu}_{l,s}(X(i)))^2 = \frac{1}{s-l-1} \mathrm{RSS}, \tag{3}$$

for $s > l+1$, where RSS denotes the residual sum of squares.

Instead of estimating a mean level for each part of the sample, we can fit a more parsimonious Fourier series of order 1 (see for example Tolstov 1962; Churchill and Brown 1987), i.e.

$$\mu_l(t) = \mu + \beta_1 \cos\left(2\pi \frac{t}{l}\right) + \beta_2 \sin\left(2\pi \frac{t}{l}\right).$$

This model needs three parameters to be estimated for the mean (cf. l above) independent of l. Parameters are estimated using linear regression and the variance is estimated analogously with (3).

In the time domain approach to the problem, we identify the process and estimate the parameters using the Box–Jenkins approach (Box and Jenkins 1966; or Wei 1990). As usual we define the ARMA(p, q) model with mean $\mu(t)$ as

$$X(t) = \mu(t) + \varphi_1 X(t-1) + \cdots + \varphi_p X(t-p) - \theta_1 \varepsilon(t-1) - \cdots - \theta_q \varepsilon(t-q) + \varepsilon(t),$$

where $\varepsilon(t)$ are i.i.d. and $N(0, \sigma^2)$. The one-step-ahead forecast errors will be i.i.d. with mean 0 and variance $\sigma^2 = \mathrm{var}(\varepsilon(t))$, and can therefore be used for surveillance by for example the Shewhart method. We find that a suitable model would be the AR(1), having the three parameters shown in Table I.

The sample mean and the Yule–Walker estimate are used for estimating μ and φ_1, respectively. The variance σ^2 is estimated by $\hat{\sigma}^2 = \widehat{\mathrm{var}}(X_t)(1 - \hat{\varphi}_1 \hat{\rho}_1)$. The forecast errors are plotted in Figure 2. Diagnostic checking shows that the fit seems to be accurate enough, although there is an indication of a possible 9 year cycle.

The RMS, defined using the general definition (2), are shown in Figure 1 and Table II. As expected, the i.i.d. have the maximum RMS. Adding only one parameter and using the AR(1) model would yield a notable improvement. Using reasonably low values of l, we get local minima for both Fourier series RMS(l) and periodic mean RMS(l) for $l = 9$ years. But it is obvious from Figure 1 that both the Fourier series model and the periodic mean model are sensitive to the choice of l. For $l \leqslant 8$ and $l \geqslant 11$ we get a better fit with the AR(1) model. There is also a disadvantage with the periodic mean model that we have to estimate $l+1$ parameters.

Table I. Estimates of the AR(1) model

	Parameter		
	Mean	φ_1	SE
Estimated value	158	0·39	31·2

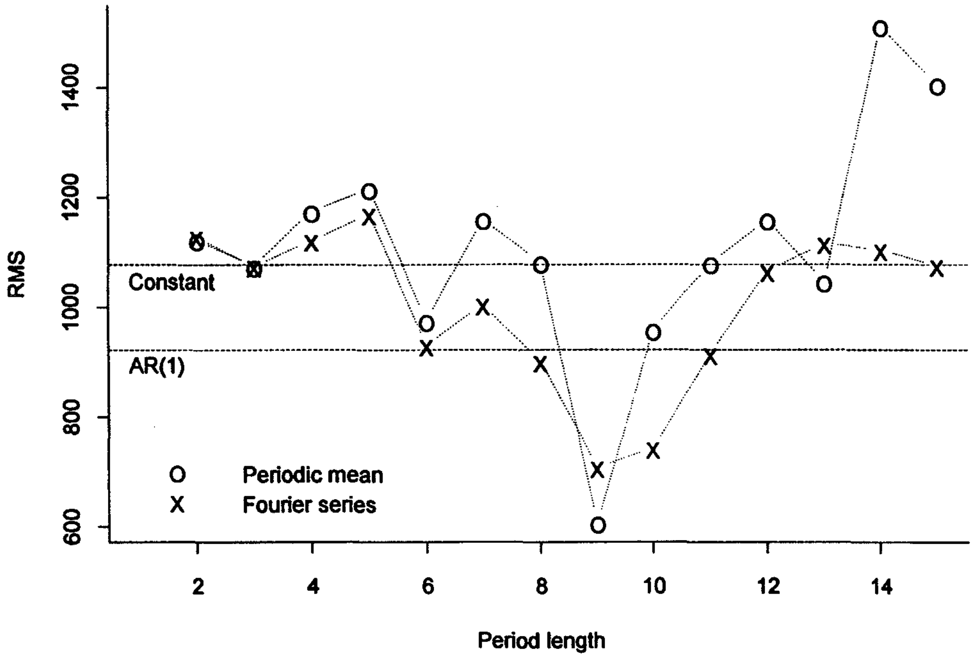


Figure 1. Residual mean square (RMS) for different models of the vendace data: the constant mean model, the periodic mean model, the Fourier series model and the AR(1) model. The constant mean and the AR(1) model are independent of period length

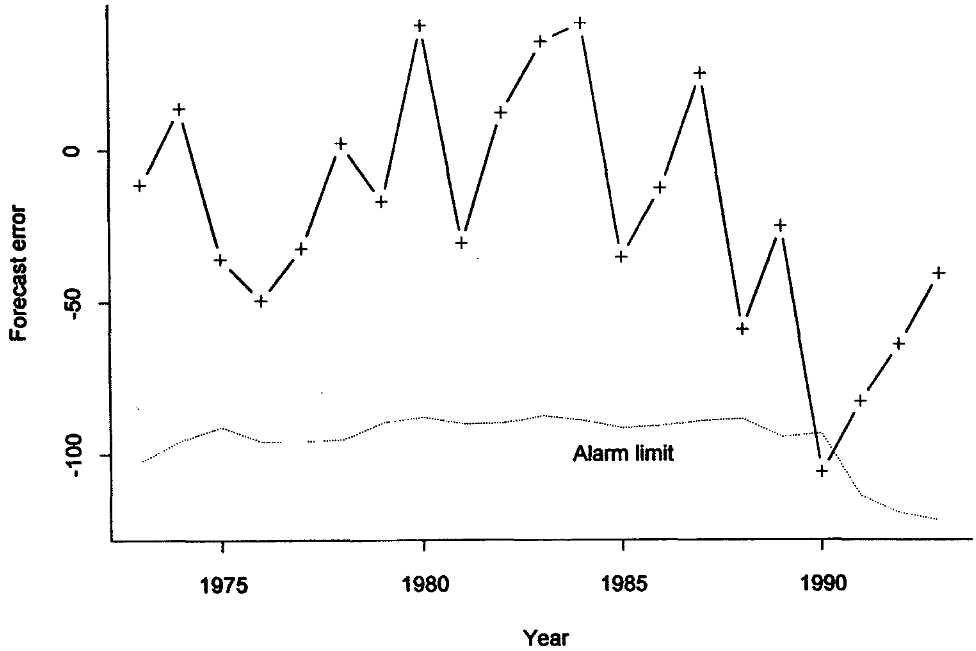


Figure 2. One-step-ahead forecast errors for the catches of vendace. Forecasts are based on the AR(1) model. The alarm limit is the $3\sigma_{t-1}(t)$ Shewhart limit with $\sigma_{t-1}(t)$ estimated from the data X_{t-1}

Table II. Comparison between the RMS for the univariate models

Model	Number of parameters	RMS
i.i.d.	2	1077
AR(1)	3	922
Fourier series ($l = 9$)	4	708
Periodic mean ($l = 9$)	10	607

We conclude that there are many different models that might fit the data, and we therefore need better knowledge about the actual ecological model generating the data to be able to make a choice. If we consider a model with a periodic mean, a period of 9 years seems to be suitable. We also find that the models with a small number of parameters, the AR and the Fourier series models, give a sufficient improvement of the goodness of fit.

3.2. Monitoring

We will now apply the models developed above for monitoring a change in distribution of the vendace population. We will only apply the Shewhart test to the data although other methods might give better performance. The mean and standard deviation are re-estimated at each time step, and the residuals are used for surveillance, using the formulation (1). Analogously, we define $\hat{\sigma}_{s-1}(X(s))$, the standard deviation of the residual at s. The Shewhart test prescribes that an alarm is triggered at time s if $|X(s) - \hat{\mu}_{s-1}(X(s))| > 3\hat{\sigma}_{s-1}(X(s))$. Using the models described above we would get an alarm in 1990. Table 3 shows the standardized deviation from the expected value, given the estimated mean and variance.

Table III. The standardized deviation from the mean in 1990

Model	S (1990)
AR(1)	−3·41
i.i.d.	−3·68
Fourier series ($l = 9$)	−5·71
Periodic mean ($l = 9$)	−5·84

With the surveillance procedures we have used we would not get an alarm earlier than 1990. However, the Shewhart method is not always the optimal method to use, and another choice of method might have detected a change earlier. Although an alarm is triggered in 1990 for all models, the difference between the values in Table III for the considered models show us that depending on which model we choose we will get different detection power.

Since the data material, owing to the long time steps of one year, will still be small for many years, ecological background or other prior information is needed to restrict attention to only a small number of possible models and interesting critical events.

4. MONITORING FIVE SPECIES SIMULTANEOUSLY

In this section we will attempt to compare the performance of different monitoring procedures, based on the information from five monitored species. The aim is to see whether it is possible to detect a change earlier if all species have been monitored simultaneously. None of the species in the current material have a detectable changepoint earlier than 1990. A minimax procedure, which sounds an alarm whenever any of the processes change, would therefore not detect a change earlier than 1990.

One way of combining the information from the multiple sources, and designing a common system for monitoring all at the same time, is by creating an index that can be used for surveillance. Statistical methods for surveillance of multiple processes have been suggested by many authors. We will use the often applied Hotelling's T^2.

4.1. Diversity indices

Several indices of biodiversity with different statistical and demographic properties have been suggested. Overviews can be found in for example Colinvaux (1986), Magurran (1988), Noss (1990) or Pielou (1975). However, no universally accepted index exists. A widely used statistic for biodiversity is the Shannon–Wiener index (Shannon and Weaver 1949) H', originally designed for measuring information content. It is defined as

$$H' = -\Sigma p_i \log(p_i),$$

where p_i is the proportion of species i measured by some suitable unit. For a given N, H' is maximized when $p_i = 1/N$ for $i = 1, \ldots, N$. When we measure diversity, any definition of 'amount' can be used that has a relevant meaning for the studied species. We will use the landed mass of each species.

In Figure 3 standardized values of H' are plotted, based on estimated mean and variance of H' up to one year earlier than the current year, i.e.

$$H''(t) = \frac{H'(t) - \hat{E}_{t-1}(H')}{\hat{\sigma}_{t-1}(H')}.$$

We see that a Shewhart $3\hat{\sigma}$ limit will give no alarm at all.

4.2. Hotelling's T^2

Hotelling's T^2-statistic (Hotelling 1947) is defined as

$$T^2(t) = (X(t) - \mu)^{\mathrm{T}}\Sigma^{-1}(X(t) - \mu),$$

where μ and Σ are the mean vector and covariance matrix, respectively. The T^2-statistic can detect deviations from both mean and variance, but is most sensitive for changes in mean, especially when all the means are changing at the same time and direction.

We assume that the data come from a multinormal distribution

$$X(t) \sim MN_q(\mu(t), \Sigma),$$

where the sequence is i.i.d. The dimension of $X(t)$ and $\mu(t)$ is $1 \times q$ and the dimension of Σ is $q \times q$. The mean for each process, estimated by using X_s, is denoted $\hat{\mu}_s$. Further, we estimate Σ successively over time, $\hat{\Sigma}_s$ estimated using X_s, by

$$\hat{\Sigma}_s = \frac{1}{s-q}\sum_{i=1}^{s}(X(i) - \hat{\mu}_s)^{\mathrm{T}}(X(i) - \hat{\mu}_s) \quad \text{when } s > q.$$

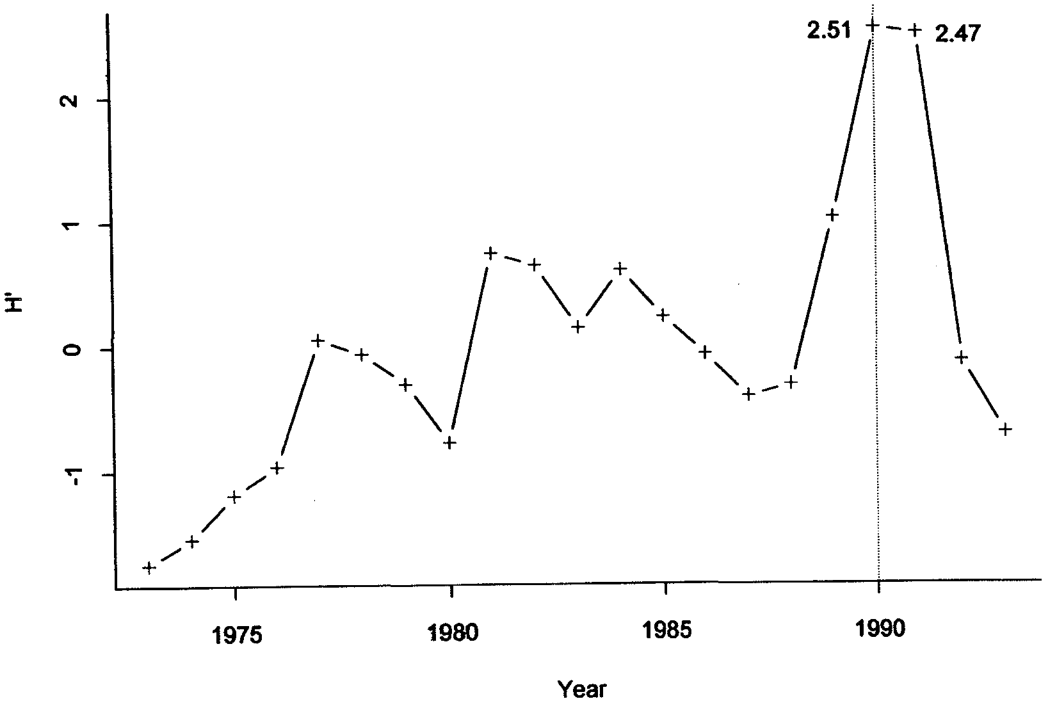


Figure 3. Standardized Shannon–Wiener index. The standardization is made on the sample mean and standard deviation, i.e. H'_{t-1}. A value exceeding 3 would trigger an alarm with the Shewhart method

In order to reduce the number of parameters, we group the correlations. Guided by Figure 4, we define three groups (4) and assume that the correlations are equal within the groups, thereby reducing the number of parameters from 15 to 8.

	Vendace	Pike	Pike–perch	Perch	Burbot
Vendace	·	*A*	*B*	*A*	*A*
Pike	*A*	·	*B*	*C*	*C*
Pike–perch	*B*	*B*	·	*A*	*B*
Perch	*A*	*C*	*A*	·	*C*
Burbot	*A*	*C*	*B*	*C*	·

(4)

The estimated mean vector has a multinormal distribution, $\hat{\mu}_t \sim MN_q(\mu, \Sigma/t)$, and the estimated variance matrix has a Wishart distribution, $\hat{\Sigma}_t \sim W_q(t-1, \Sigma)$, which is a multivariate extension of the χ^2-distribution (Crowder and Hand 1993). Approximating the distribution of $\tilde{\Sigma}_t$ by $W_q(t-1, \Sigma)$ we get

$$\frac{q}{(t+1)(t-q+1)} T^2 \overset{\text{approx}}{\sim} F_{q,t-q+1}.$$

Using the re-estimated values of μ and Σ, i.e.

$$T^2(t) = (X(t) - \hat{\mu}_{t-1})^{\mathrm{T}} \hat{\Sigma}_{t-1}^{-1} (X(t) - \hat{\mu}_{t-1}),$$

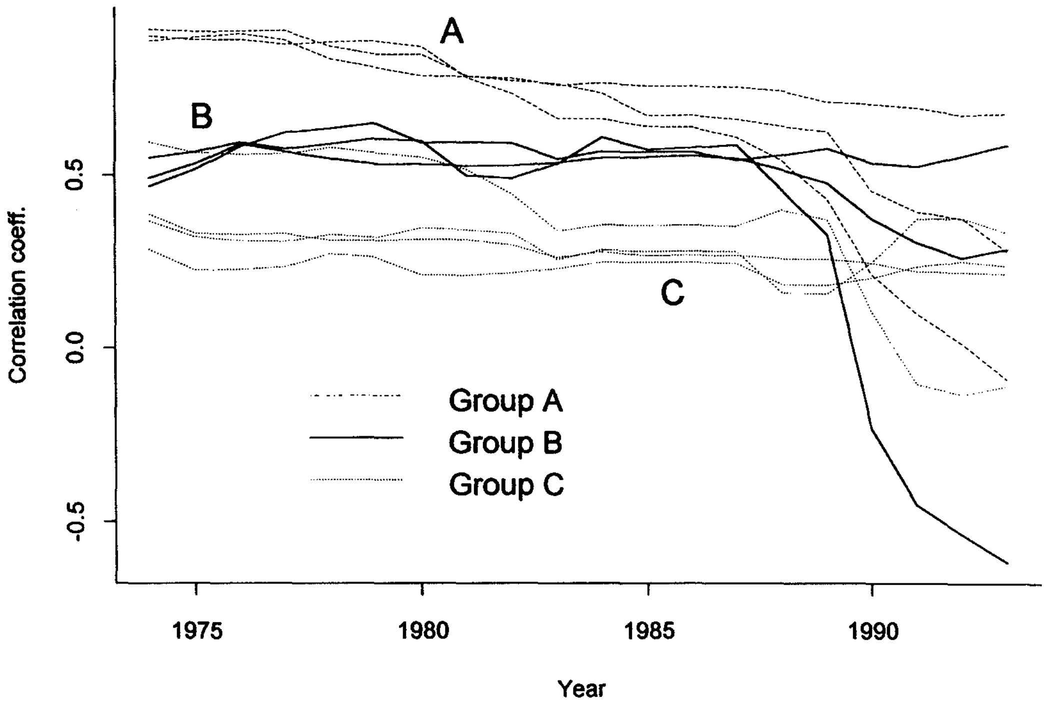


Figure 4. Pairwise correlation coefficients between species, estimated between 1964 and the current year. Correlations have been grouped into three groups (*A*, *B* and *C*) where the correlations are almost equal to each other

we get the sequence of $T^2(t)$-values shown in Figure 5. With the same false alarm rate as for a traditional Shewhart test, we would get an alarm in 1990.

However, by accepting a higher false alarm rate, say $\alpha = 0{\cdot}05$, the alarm limit is crossed already in 1989. Applying the same false alarm rate with the Shannon–Wiener index we would have had an alarm in 1990, but no earlier alarms for the univariate case. It could therefore be possible to detect some changes earlier if we take all species into account simultaneously.

5. DISCUSSION AND CONCLUSIONS

We see from the univariate analysis of the vendace (*Coregonus albula*) data that the choice of model is of great importance. When different cyclic patterns or autocorrelations are present, data have to be modified to take this into account. We find that goodness of fit can be improved by estimating cyclic patterns or autocorrelation. By using either of the four models studied in this paper, the Shewhart procedure would have detected a change by 1990. The weakest reactions come from the AR(1) and i.i.d. models. The periodic models both expected an increased catch by 1990 and, because of that, the difference between the expected and actual catch in 1990 was magnified.

As expected, the species of fish are correlated with each other. Thanks to that, the drop in vendace could either be explained by the other species behaving in the same way or else lead to a decreasing correlation between vendace and other species. Both scenarios are interesting as the ecological causes possibly have to be sought in different places. As with the univariate

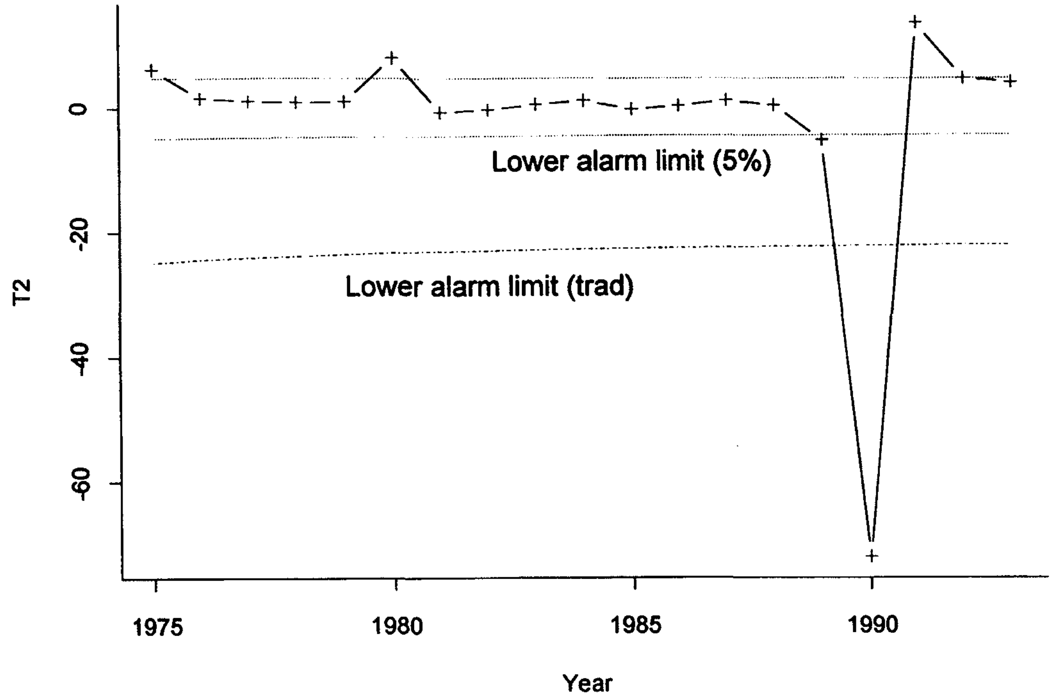


Figure 5. Hotelling's T^2-statistic based on mean and variance re-estimated at each time step. The variance matrix has been replaced by the reduced variance matrix $\tilde{\Sigma}$, where the covariances have been replaced by their arithmetic means within the correlation groups. The critical limits have been estimated using an approximated F-distribution

problem above, the description of the critical event is crucial and is also affected by the 'in control' system.

Neither the Shannon–Wiener index nor the variants of it studied in this paper sound any alarm at all. Hotelling's T^2, however, detects a change already in 1989, one year earlier than the univariate procedures if we accept a higher false alarm rate. With the same false alarm rate as for the Shewhart test, the alarm is called in 1990.

There is great potential and usefulness for statistical surveillance on environmental data. Throughout the world, enormous amounts of data are collected about the environment and stored for analysis. Quality control and statistical process control are used, as was mentioned earlier, in many places in society, and these techniques can also be useful for monitoring environmental data.

ACKNOWLEDGEMENT

This paper is a part of a research project on statistical surveillance at the Göteborg University, which is supported by the Swedish National Council for Research in the Humanities and Social Sciences. I would like to thank Marianne Frisén, the supervisor of the project, for her helpful advice, ideas and encouragement. I am also grateful to Kasra Afsarinejad and the reviewers for valuable comments. Data for the fish example were provided by Olof Filipsson and his colleagues at the Fresh Water Laboratory of the National Board of Fisheries at Drottningholm.

REFERENCES

Arnkelsdóttir, H. (1995). *Surveillance of Rare Events: on Evaluations of the Sets Method*, Research Report 1995:1, Department of Statistics, Göteborg University.
Berthoux, P. M., Hunter, W. G. and Pallesen, L. (1978). 'Monitoring sewage treatment plants: some quality control aspects', *Journal of Quality Technology*, **10**, 139–149.
Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis Forecasting and Control*, Holden-Day, San Francisco.
Brown, R. L., Durbin, J. and Evans, J. M. (1975). 'Techniques for testing the constancy of regression relationships over time', *Journal of the Royal Statistical Society, Series B*, **37**, 49–163.
Charnes, J. M. and Gitlow, H. S. (1995). 'Using control charts to corroborate bribery in Jai Alai', *The American Statistician*, **49**(4), 386–389.
Churchill, R. V. and Brown, J. W. (1987). *Fourier Series and Boundary Value Problems*, McGraw-Hill, New York.
Colinvaux, P. (1986). *Ecology*, Wiley, Chichester.
Crowder, M. J. and Hand, D. J. (1993). *Analysis of Repeated Measures*, Chapman and Hall, London.
Frisén, M. (1992). 'Evaluations of methods for statistical surveillance', *Statistics in Medicine*, **11**, 1489–1502.
Frisén, M. (1994). *Statistical Surveillance of Business Cycles*, Research Report 1994:1, Department of Statistics, Göteborg University.
Frisén, M. and de Maré, J. (1991). 'Optimal surveillance', *Biometrika*, **78**, 271–280.
Hotelling, H. (1947). 'Multivariate quality control: illustrated by the air testing of bombsights', in Eisenharts, C., Hastay, M. V. and Wallis, W. A. (eds), *Techniques of Statistical Analysis*, McGraw-Hill, New York, pp. 111–184.
Kjelle, P.-E. (1987). *Alarm Criteria for the Fixed Gamma Radiation Monitoring Stations*, Research Report 1987:7, National Institute of Radiation Protection.
Magurran, A. E. (1988). *Ecological Diversity and its Measurements*, Princeton University Press, Princeton, NJ.
Noss, R. F. (1990). 'Indicators for monitoring biodiversity: a hierarchical approach', *Conservation Biology*, **4**, 355–364.
Page, E. S. (1954). 'Continuous inspection schemes', *Biometrika*, **41**, 100–114.
Pielou, E. C. (1975). *Ecological Diversity*, Wiley, Chichester.
Roberts, S. W. (1959). 'Control chart tests based on geometric moving averages', *Technometrics*, **1**, 239–250.
Settergren Sørensen, P. and la Cour Jansen, J. (1991). 'Statistical control of hygienic quality of bathing water', *Environmental Monitoring and Assessment*, **17**, 217–226.
Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL.
Shewhart, W. A. (1931). *Economic Control of Quality Control*, Reinhold, Princeton, NJ.
Shiryaev, A. N. (1963). 'On optimum methods for quickest detection problems', *Theory Probability and its Applications*, **8**, 22–46.
Sveréus, A. (1995). *Detection of Gradual Changes: Statistical Methods in Post Marketing Surveillance*, Research Report 1995:2, Department of Statistics, Göteborg University.
Tolstov, G. P. (1962). *Fourier Series*, trans. R. A. Silverman, Dover, New York.
Vaughan, W. J. and Russell, C. S. (1983). 'Monitoring point sources of pollution: answers and more questions from statistical quality control', *American Statistician*, **37**, 476–487.
Wei, W. W.-S. (1990). *Time Series Analysis*, Addison-Wesley, New York.
Wessman, P. (1996). *Multivariate Surveillance*, Research Report 1996:4, Department of Statistics, Göteborg University.
Wetherill, G. B. and Brown, D. W. (1991). *Statistical Process Control: Theory and practice*, Chapman and Hall, London.
Zacks, S. (1983). 'Survey of classical and Bayesian approaches to the changepoint problem: fixed sample and sequential procedures of testing and estimation', *Recent Advances in Statistics*, 245–269.

Evaluation of some methods for statistical surveillance of an autoregressive process

Magnus Pettersson
Department of Statistics
Göteborg University
Box 660
SE-405 30 Göteborg
Sweden
magnus.pettersson@statistics.gu.se

May 5, 1998

Abstract

Statistical surveillance is used for fast and secure detection of a critical event in a monitored process. This paper studies the performance for AR(1) processes.

Two often suggested methods for detection of a shift in the mean, the modified Shewhart and the residual method, are compared and evaluated. Further, comparisons are made with direct Shewhart and a likelihood ratio method.

New evaluation measures, the probability for successful detection and the predictive value, are also applied together with the average run length and run length distributions.

We conclude that neither the modified nor the residual methods is uniformly optimal. The residual method is, however, optimal for immediate detection, but has inferior properties otherwise. For many parameter setups, the modified method will give the better performance.

1. Introduction

Statistical surveillance is used for systematic monitoring of a process with the purpose to detect an unwanted departure from a specified state. Methods for *Statistical Process Control* (SPC) have been widely used for industrial, medical, economical, environmental and many other applications. Several textbooks have been published, for example Box and Luceño (1997), Montgomery (1997) or Wetherill and Brown (1991). Note the difference between hypothesis testing for a change-point on a fix set of data and surveillance: In both cases we do not know if something has happened and when. But statistical surveillance is used for situations where new data arrives at each time step. The procedure is repeated and there is no fixed hypothesis.

One fundamental assumption required by standard methods is that the process is *iid* (*Independent and Identically Distributed*) - a requirement which is often not met in practise. Removing the assumption of independence will affect the performance of the surveillance procedures.

A survey by Alwan and Roberts (1995) of 235 quality control applications, where less than 50% of the studied applications were independent and less than 15% were *iid*, gives a good motivation for studying this problem. Further, Alwan and Roberts (1995) together with Caulcutt (1995) and the discussion following them, testified about the frustration they have met with engineers who tried to apply SPC methods to autocorrelated data since the resulting monitoring system does not have the wanted properties. Stone and Taylor (1995) also pointed out that sometimes not even the ARIMA model is sufficient for the description of the process.

The robustness of CUSUM and EWMA applied directly on the observed process have been discussed by for example Bagshaw and Johnson (1975), Harris and Ross (1991), Johnson and Bagshaw (1974), Montgomery and Mastrangelo (1991), Schmid and Schöne (1997), VanBrackle and Reynolds (1997) and Yashchin (1993).

Among others, two solutions for the non *iid* case have been proposed by several authors: We will call them the *modified Shewhart* method and the *residual* method, respectively. The methods will be described in detail below. The modified Shewhart method have been investigated by Vasilopoulos and Stamboulis (1978), for an AR(2) process. The residual method was suggested for ARIMA-processes by Berthoux *et al.* (1978) and Alwan and Roberts (1988). Since these methods are often suggested and used in practise it is interesting to compare them with each other. Furthermore, we will briefly examplify what will happen if the process parameters are estimated during run-in under an assumed *iid* situation. We will call this method the *direct Shewhart*.

Often comparisons between the methods are limited to average run length. We will extend the evaluation using the predictive value and the probability of successful detection suggested by Frisén (1992). We will in this paper also compare the modified Shewhart and the residual method with examples of the *likelihood ratio method* in order to further examine their properties.

In Section 2 a specification of the situation which is studied is given. In Section 3 the methods compared in this paper are defined in detail. Section 4 contains results on the evaluation measures considered. In Section 5 the results and conclusions are discussed.

2. Specifications

Consider a process that is observed at discrete time steps, $t = 1, 2, \ldots$. The data observed at time t is a continous stochastic variable denoted by $X(t)$. The cumulated data up to time t is denoted by $X_t = \begin{pmatrix} X(1) & \ldots & X(t) \end{pmatrix}$. Consistently, the current value of any variable is denoted by time within parentheses, *eg.* $X(t)$, $\mu(t)$, $\varepsilon(t)$ and $w(t)$, while the cumulated sets are denoted by time in index, *eg.* X_t, μ_t, ε_t and w_t. When the process behaves in the prescribed, wanted or expected way we say that it is "in control". Our general model for the in control part of the process is

$$X(t) = \mu(t) + w(t),$$

where

$$w(t) = \phi \cdot w(t-1) + \varepsilon(t). \tag{2.1}$$

and the correlation $|\phi| < 1$. The variable ε_t is normally distributed white noise with $Var[\varepsilon(t)] = \sigma^2$ and $\varepsilon(t)$ is independent of w_{t-1}. Note that we are defining σ^2 as the variance of the concealed error term, ε. We will in this paper assume that ϕ, μ and σ are known and we can therefore without loss of generality set $\mu(t) = 0$ and $\sigma = 1$.

At an unknown time, τ, the process is disturbed and goes "out of control". We study the case where a shift in μ to a known value, δ, occurs, *i.e.*

$$\mu(t) = \delta \cdot \mathbf{1}_{\{\tau \leq t\}}.$$

Hence the expected value of X is

$$E[X(t)] = \begin{cases} 0 \text{ when } t < \tau \\ \delta \text{ when } t \geq \tau \end{cases}.$$

At each time, s, we want to discriminate between two events, $D(s)$ and $C(s)$, where $D(s) = \{\tau > s\}$ is the event of the process being in control. $C(s) = \{\tau = s\}$ and $C(s) = \{\tau \leq s\}$ will be discussed.

Figure 1 shows an example of an AR(1) process with a shift $\delta = 10 \cdot \sigma$ with $\tau = 40$.

3. Methods for Monitoring an AR(1) Process

When the monitored process is not *iid* but autoregressive the properties of the standard methods are changed. In this paper we will study some methods that are often suggested in the literature for this case: "Direct Shewhart", where the time series structure is not taken into account; "Modified Shewhart", where the limits have been altered to give a specific average run length and "Residual Shewhart", where the forecast errors are used for monitoring. As a benchmark these methods will be compared with the likelihood ratio method. The name "modified Shewhart" was given by Schmid (1995) and exact limits for some processes have been given by Vasilopoulus and Stamboulis (1978). The residual method was suggested by Alwan and Roberts (1988) and Berthoux *et al.* (1978).

We will in this paper restrict attention to the AR(1) process (2.1) with $\phi > 0$.

3.1. Direct Shewhart

If time dependence is not taken into account a user might estimate the mean and the variance during run-in. In the case of an *iid* process, X, the Shewhart procedure, suggested by Shewhart (1931), prescribes that an alarm is called when

$$|X(t)| > k \cdot \sigma,$$

where the constant k is set to give a certain proabability of calling a false alarm. In traditional SPC litterature k is often 3 or 3.09. However, for a stationary AR(1) process the variance of X becomes

$$\sigma_x^2 = Var[X(t)] = \frac{\sigma^2}{1-\phi^2}.$$

Estimating the variance with a very large number of observations and using the same constant k an alarm will be called when

$$|X(t)| > k \cdot \sigma_x = k \cdot \sigma \cdot \frac{1}{\sqrt{1-\phi^2}}. \tag{3.1}$$

Since $(1-\phi^2)^{-1/2} > 1$ these limits will become greater than the limits for an *iid* process with variance σ^2.

3.2. Modified Shewhart

The direct Shewhart will, as we will see in later Sections, have some undesirable properties, *eg.* an ARL_0 (Section 4.1) that is depending on ϕ. A straightforward solution to that problem could be to adjust the control limits of the Shewhart chart to give the wanted ARL_0.

Define $c(\phi)$ as the factor adjusting the limits of the *iid* Shewhart so that an alarm is called when

$$|X(t)| > k \cdot \sigma \cdot c(\phi).$$

Since $\phi > 0 \Longrightarrow Var[X] > \sigma^2$ it follows that $c(\phi) > 1$. In Table 3.1, the adjusting factors have been estimated by computer simulation to yield $ARL_0 = 11$, the limits are also plotted in Figure 3 together with the limits obtained by using the direct Shewhart (3.1) with $ARL_0 = 11$ for $\phi = 0$.

ϕ	**Modified** $c(\phi)$	**Direct** $(1-\phi^2)^{-1/2}$	ARL_0
0.0	1.000	1.000	11.00
0.2	1.014	1.020	11.26
0.4	1.060	1.091	12.17
0.6	1.155	1.250	14.36
0.8	1.363	1.667	20.99

Table 3.1: *Comparison between the adjusting factors of the modifed and direct Shewhart.*

We see that $c(\phi) < (1-\phi^2)^{-1/2}$, *i.e.* the direct Shewhart is having higher alarm limits than the modified. Therefore it follows that that the ARL_0 is higher for the direct than for the modified Shewhart.

3.3. Residual Method

The idea of the residual method is that the current value, $X(s)$, and its expectation given the past value are compared and the difference is used for monitoring. A similar approach is used by the *Food and Drug Administration* (FDA) as a guideline in postmarketing surveillance of adverse effects of drugs, where consequtive quarters are compared (Sveréus, 1995). Also the *National Institute for Radiation Protection* (SSI) uses differences in mean between consequtive 24 hour-period means to detect suddenly increasing background radiation levels (Kjelle, 1987). Other examples of applications of the residual method can be found in Harris and Ross (1991), Montgomery (1997), Notohardjono and Ermer (1986) and Pettersson (1998).

Based on the second last observation, $X(s-1)$, a forecast of $X(s)$ is

$$\widehat{X}(t) = \phi \cdot x(t-1).$$

The residual is here defined as the difference between the observed value and its forecast, *i.e.*

$$R(t) = X(t) - \widehat{X}(t) = X(t) - \phi \cdot X(t-1).$$

When $t < \tau$ the residual $R(t) = \varepsilon(t)$. But generally the residual becomes

$$R(t) = \varepsilon(t) + \Delta(t),$$

where

$$\Delta(t) = E[R(t)] = \begin{cases} 0 & \text{when } t < \tau \\ \delta & \text{when } t = \tau \\ (1-\phi)\,\delta & \text{when } t > \tau \end{cases}. \tag{3.2}$$

For a fixed value of τ

$$Var[R(t)] = Var[\varepsilon(t)] = \sigma^2.$$

and a Shewhart test used for R would call an alarm when

$$|R(t)| > k \cdot \sigma,$$

where k is a constant.

When $\phi > 0$, the expected value will decrease after τ and

$$E[R(t)] < E[R(\tau)], \text{ for } t = \tau+1, \tau+2, \ldots$$

In Figure 1 we see an example of an simulated AR(1) process, with a shift at $t = 40$ of the size 10σ. Figure 2 shows the residuals, *i.e.* forecast errors, of the process in Figure 1, where $E[R(40)] = 10$ and $E[R(t)] = 5$ for $t > 40$. That have earlier been observed by among others Harris and Ross (1991), Ryan (1991), Superville and Adams (1994) and Wardell *et al.* (1994) and for time series analysis by among others Enders (1995), Fox (1972) and Wei (1990).

3.4. Likelihood Ratio Method

It is possible to derive a method which have certain optimality properties. For a fixed false alarm rate and a fixed time, an alarm set based on the *likelihood ratio statistic* (*lr*) have the highest probability of calling an alarm when the process have gone out of control (Frisén and de Maré, 1991). Sequential procedures with minimal expected delay are based on this statistic. This approach will not be studied in detail in this paper, except for some illustrative examples intended to give insight in the properties of the methods studied.

The likelihood ratio statistic, $lr(X_s)$, is defined as

$$lr(X_s) = \frac{f(X_s \mid C(s))}{f(X_s \mid D(s))},$$

where $f(X_s \mid D(s))$ and $f(X_s \mid C(s))$ is the probability density function of X_s under the in- and out-of control states, respectively. Since $X(t)$ given X_{t-1} is normally distributed with

$$E[X(t) \mid X_{t-1}] = \phi \cdot X(t-1) + \Delta(t)$$

and $Var[X(t)] = \sigma^2$ the probability distribution function becomes

$$f_{X(t)|X_{t-1}}(x(t), x(t-1)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \phi \cdot x(t-1) + \Delta(t))^2\right\},$$

where $\Delta(t) = 0$ for $t < \tau$ (3.2). Further, using that

$$f(X_s) = f(X(s) \mid X_{s-1}) \cdot f(X(s-1) \mid X_{s-2}) \cdot \ldots \cdot f(X(1))$$

the lr statistic for $D(s) = \{\tau > s\}$ and $C(s) = \{\tau = k \leq s\}$ reduces to

$$lr(X_s) = \prod_{i=k}^{s} \frac{f(X(i) \mid X_{i-1}, \tau = k)}{f(X(i) \mid X_{i-1}, \tau > s)}.$$

Cancelling constants and using the properties of the exponential function we find that the lr statistic depends on the data only through

$$\begin{aligned}
&\sum_{i=k}^{s} (X(i) - \phi \cdot X(i-1) + \Delta(i))^2 - (X(i) - \phi \cdot X(i-1))^2 \\
&= \sum_{i=k}^{s} [2 \cdot X(i) \cdot \Delta(i) - 2\phi \cdot X(i-1) \cdot \Delta(i)] \\
&= 2\sum_{i=k}^{s} [X(i) - \phi \cdot X(i-1)] \cdot \Delta(i) = 2\sum_{i=k}^{s} R(i) \cdot \Delta(i).
\end{aligned}$$

Now, using the specification (3.2) for Δ we find that the lr statistic depends on the data only through

$$R(k) + (1-\phi) \sum_{i=k+1}^{s} R(i),$$

for $C\{\tau = k\}$ when $k < s$ and $R(s)$ for $C\{\tau = s\}$. Hence the likelihood ratio statistic for immediate detection, $C(s) = \{\tau = s\}$, depends on the data only through $R(s)$. However, for other specifications of $C(s)$ this is no longer the case. The likelihood ratio statistic for $C(s) = \{\tau = k\}$ becomes a function of $R(k), \ldots, R(s)$.

4. Results

In this section numerical results comparing the methods are presented. To compare different methods, several evaluation measures have been suggested, see Frisén (1992) and Frisén and Wessman (1998) for overviews. The choise of which measure should be used as guidance has to be decided by using knowledge of the specific application.

We will study an AR(1) process with parameter $0 < \phi < 1$ and without loss of generality we set $\mu = 0$ and $\sigma = 1$. We will use a two-sided Shewhart test, with the limits set to give $ARL_0 = 11$. For many applications this might be too small but it will anyway show the impact of the autocorrelation on the surveillance procedures.

The critical event is a shift in mean from 0 to $\delta \cdot \sigma$ occuring at time τ. To calculate the predictive value and probability of successful detection we need knowledge of the run length given any value of τ, which is an extension from earlier papers on this matter, where only the cases $\tau = 1$ or $\tau = \infty$ have been considered. At calculation of the predictive value, we will a priori assume that τ is geometrically distributed,

$$f_\tau(t) = \nu \cdot (1-\nu)^{t-1}, t = 1, 2, \ldots,$$

where ν is the failure rate or incidence, *i.e.* $\nu = P\{\tau = t \mid \tau \geq t\}$, for $t = 1, 2, \ldots$.

4.1. The Run Length Distribution

The time to the first alarm, that is the run length, t_A, is of special interest. When $t_A < \tau$ the alarm is false and otherwise it is true. The stochastic variable t_A is a stopping time with outcomes in $\{1, 2, \ldots\}$. Figure 4 shows the the probability density function for the run length, f_{t_A}, for the modified and residual method when $\mu(t) \equiv 0$ which is denoted by $\tau = \infty$.

An often used summarizing value is the *Average Run Length* (*ARL*). More specifically, we define

$$ARL_0 = E[t_A \mid \tau = \infty],$$

the average run length when the process is in control. In quality control literature, the ARL_0 is often compared with

$$ARL_1 = E[t_A \mid \tau = 1].$$

For the residual method, the probability of calling a false alarm at a specific time is

$$p_0 = P(|R(t)| > k\sigma) = 2(1 - \Phi(k)),$$

where Φ denotes the cumulative probability density function for the standard normal distribution. The expectation $E[R(t)]$ is depending on the time since the shift (3.2). The probability of calling an alarm for $t \geq \tau$ becomes

$$p_{A0} = P(t_A = \tau) = P(|R(t)| > k \mid t = \tau) = 1 - \Phi(k - \delta) - \Phi(-k - \delta)$$

and

$$p_{A1} = P(t_A = t \mid t > \tau) = 1 - \Phi(k - \delta \cdot (1 - \phi)) - \Phi(-k - \delta \cdot (1 - \phi)).$$

The average run lengths ARL_0 and ARL_1 becomes

$$\begin{aligned} ARL_0 = & \sum_{i=1}^{\infty} i \cdot P(t_A = i \mid \tau = \infty) \\ = & \sum_{i=1}^{\infty} i \cdot p_0 \cdot (1 - p_0)^{i-1} = \frac{1}{p_0} \end{aligned}$$

and

$$\begin{aligned} ARL_1 = & \ 1 \cdot P(t_A = 1 \mid \tau = 1) + E[t_A \mid t_A > 1] \cdot P(t_A > 1 \mid \tau = 1) \\ = & \ p_{A0} + \left(\frac{1}{p_{A1}} + 1\right) \cdot (1 - p_{A0}) \\ = & \ \frac{1 - p_{A0} + p_{A1}}{p_{A1}}. \end{aligned}$$

For the direct Shewhart the ARL_0 depends on ϕ (Figure 5). Therefore it is not directly comparable with the other methods. It will be excluded from analyses with measurements of detection power.

Figure 6 presents the ARL_1 for the residual and modified Shewhart where the values for the latter have been obtained using computer simulations. Comparing them, we find that they both have an ARL_1 that increases with ϕ, but ARL_1 for the residual method is higher than the ARL_1 for the modified method. Using the run lengths would therefore favour the modified Shewhart method. When $\phi \approx 0.6$ there is a substantial difference.

These ARL functions have earlier been described by Schmid (1995), Wardell *et al.* (1994) and Zhang (1997). They found that the modified method has a smaller ARL_1 than the residual method, given a fixed ARL_0. Also Schmid and Schöne (1997) and Superville and Adams (1994) have found the same.

4.2. Probability of Successful Detection

For some applications it is crucial that a change is detected within a certain time, say d time steps. If an alarm is called within d time steps, actions can be taken to prevent the negative effects of the change. A relevant measure for such applications is the *Probability of Successful Detection* (PSD). We define

$$PSD(s,d,\phi) = P\{t_A < s+d \mid t_A \geq s, \tau = s\} = \frac{P\{s \leq t_A < s+d \mid \tau = s\}}{P\{t_A \geq s \mid \tau = s\}}$$

(Frisén, 1992). The PSD is generally a function of the time of the change, τ. The properties of the Shewhart test implies that the PSD for the residual method is constant over time:

$$PSD_{res}(d,\phi) = 1 - (1-p_{A0}) \cdot (1-p_{A1})^{d-1}.$$

Also the PSD for the modified Shewhart is constant over time and have been estimated by computer simulations.

In the special case where $d = 1$, *i.e.* the probability of immediate detection, the residual is better than the modified (Figure 7). When $\phi = 0$ the PSD for both the methods are equal. From Figure 8, where some values of the PSD for $d > 1$ are plotted, we see that the performance of both the residual and modified methods get worse when ϕ grows. Further, we can see that for values $\phi \approx 0.7$ or smaller the modified method will have a higher probability of calling an alarm here. When ϕ is close to one, the PSD becomes higher for the residual method than for the modified depending on that it still has a high probability of calling an alarm at $t = \tau$.

4.3. Predictive Value

As an alarm is called we want to know how certain we can be that a change has occured. A measure for this is the *Predictive Value* (PV), defined as

$$PV(s) = P\{\tau \leq s \mid t_A = s\},$$

(Frisén, 1992). It can be rewritten as the proportion of motivated alarms of all alarms at time s, *i.e.*

$$PV(s) = \frac{P\{t_A = s \wedge \tau \leq s\}}{P\{t_A = s\}} = \frac{PMA(s)}{PMA(s) + PFA(s)},$$

when $P\{t_A = s\} > 0$. When PV is close to 1 the alarm is highly motivated.

We define the *Probability of a False Alarm* (PFA) occuring at time s as

$$PFA(s) = P\{t_A = s \mid \tau > s\} \cdot P\{\tau > s\}. \tag{4.1}$$

The *Probability of a Motivated Alarm* (*PMA*) is not only depending on the time of the alarm, s, but also on the actual time of the change, τ, and the event to be detected. The PMA is calculated by conditioning on τ and using the distribution of τ

$$PMA(s, \delta) = \sum_{t=1}^{s} P\{t_A = s \mid \tau = t\} \cdot P\{\tau = t\}. \tag{4.2}$$

To derive the PV for the residual method we find the PFA using (4.1)

$$PFA(t) = p_0 \cdot (1 - p_0)^{t-1} \cdot (1 - \nu)^t,$$

which is independent of ϕ. Secondly, we use (4.2) to find PMA

$$PMA(s) = \sum_{t=1}^{s} \nu (1 - \nu)^{t-1} \cdot l(s, t, \phi),$$

where $l(s, t, \phi) = P\{t_A = s \mid \tau = t\}$ for $t \leq s$. For the residual method $l(s, t, \phi)$ can be calculated exactly and

$$\begin{aligned} PMA_{res}(s) = \; & \nu \cdot (1 - \nu)^{s-1} \cdot (1 - p_0)^{s-1} \cdot p_A(0) \\ & + \sum_{t=1}^{s-1} \nu \cdot (1 - \nu)^{t-1} \cdot (1 - p_0)^{t-1} (1 - p_A(0))^{s-t-1} \cdot p_A(1). \end{aligned}$$

For the modified Shewhart the l-function, p_0 and p_A have been estimated by computer simulations. In Figure 9 PV for $\phi = 0$, $\phi = 0.2$ and $\phi = 0.9$ of the residual and modified method are plotted. Often it is reasonable to choose an ARL_0 high enough to ensure that the monitoring stops before the $t = ARL_0$ when $\tau = 1$. When $t < ARL_0$

$$\phi' < \phi'' \Rightarrow PV(t, \phi') > PV(t, \phi''),$$

for the cases presented in the figure. Further, when $t = 3, 4, \ldots$ the predictive value for the modified Shewhart is higher than for the residual method. Initially the modified Shewhart is having a very poor PV, which is depending on the high flase alarm probability at $t = 1$ (Figure 4). At $t = 2$ the methods are almost equal, but the modified method is better when $\phi = 0.9$ and the residual when $\phi = 0.2$.

5. Discussion

The direct Shewhart method will have an ARL_0 which is increasing with ϕ. In order to obtain a constant ARL_0 the limits would have to be adjusted and set equivalent with the modified Shewhart. Hence direct Shewhart is not fully comparable with the others.

The likelihood ratio method is optimal in the Neyman-Pearson sense, *i.e.* have the highest probability of calling a true alarm given a specific false alarm probability (Frisén and de Maré, 1991). When the method is optimized to detect a change immediately, *i.e.* when $C(s) = \{\tau = s\}$, the *lr* method and the residual method become equivalent, and in the special case $\phi = 0$ also the modified and direct Shewhart methods become equvivalent with the likelihood ratio. For other specifications of $C(s)$, *eg.* an event occuring at a specified time, $t < \tau$, the *lr* statistic is not a function of $R(s)$ or $X(s)$ only.

Apart from the case $C(s) = \{\tau = s\}$, both the residual and the modified Shewhart are suboptimal. As was seen above, the residual method is not monitoring the *level* of the mean, but instead the *change* in level of the mean. This effect have been observed and discussed by among others Harris and Ross (1991), Ryan (1991), Superville and Adams (1994) and Wardell *et al.* (1992, 1994). Wardell *et al.* (1994) showed that the run length distribution after a shift has occured is almost equal to the in control run length distribution, one time step after the shift.

Clearly, if we can identify the process under study and the requirements we have on the surveillance procedure, it might be possible to construct an optimal *lr* procedure. However, reducing the data with the tranformations presented above will, except in a few special cases, lead to a loss of information and suboptimal procedures. Applying EWMA or CUSUM on the reduced data can not get that information back.

Summarizing the findings about the ARL_1, PSD and PV (for $t < ARL_0$) we see that for most of the cases studied both the modified and the residual methods have worse performance for larger values of ϕ. For $\phi \approx 0.7$ and smaller, the modified method will be better, except for immediate detection. The low $PV(1)$ for the modified method is due to the high probability of a false alarm at $t = 1$. Zhang (1997) pointed out, as a rule of thumb, that the residual method is to prefer when $\delta > 2$ and $\phi > 0.8$.

As have been noted by many authors before, the residual chart does not give a full picture of the process and is only sufficient for some specifications of the considered process. To overcome the disadvantages of the residual and modified methods Adams *et al.* (1994) suggested that both the observed and residual processes should be used simultanously. Alwan (1992) compared the alarms given by either of the observed and residual process. This approach will lead to a multivariate monitoring problem, discussed by *eg.* Jones *et al.* (1970), Kramer and Schmid (1997) and Wessman (1998).

From the results in the earlier Sections we find that one method is not always uniformly better than the other. Neither the residual nor the modified method are optimal except in a few special settings, for example the residual method for the situation of immediate detection of a shift.

Acknowledgment
I would like to thank my supervisor, Marianne Frisén, for lots of encouragment and good advice. Many thanks to all collegaues at the Department of Statistics for providing an inspiring research environment and special thanks to (in alphabetic order) Shakir Hussein, Tommy Johnson, Aila Särkkä and Peter Wessman for comments on an earlier versions of the manuscript.

References

[1] Adams, B. M., W. H. Woodall and C. R. Superville (1994). Discussion. *Technometrics*, 36, pp 19-22.

[2] Alwan, L. C. (1992). Effects of Autocorrelation on Control Chart Performance. *Communications in Statistics - Theory and Methods*, 21(4), 1025-1049.

[3] Alwan, L. C. and H. V. Roberts (1988). Time-Series Modelling for Statistical Process Control. *Journal of Business Economics and Statistics,* 6, 87-95.

[4] Alwan, L. C. and H. V. Roberts (1995). The Problem of Misplaced Control Limits. *Applied Statistics,* 44(3), pp 269-278.

[5] Bagshaw, M. and R. A. Johnson (1975). The Effect of Serial Correlation on the Performance of CUSUM Tests II. *Technometrics*, 17(1), pp 73-80.

[6] Berthoux, P. M., E. G. Hunter and L. Pallesen (1978). Monitoring Sewage Treatment Plants: Some Quality Control Aspects. *Journal of Quality Control*, 10, 139-149.

[7] Box, G. and A Luceño (1997). *Statistical Control by Monitoring and Feedback Adjustment.* John Wiley and Sons.

[8] Caulcutt, R. (1995). The Rights and Wrongs of Control Charts. *Applied Statistics.* 44(3), 279-288.

[9] Fox, A. J. (1972). Outliers in Time Series. *Journal of the Royal Statistical Society, Series B*, 43, pp 350-363.

[10] Enders, W. (1995). *Applied Econometric Time Series.* Wiley.

[11] Frisén, M. (1992). Evaluations of Methods for Statistical Surveillance. *Statistics in Medicine*, 11, 1489-1502.

[12] Frisén, M. and P. Wessman (1998). *Evaluations of likelihood ratio methods for surveillance. Differences and robustness.* Research report 1998:2, Department of Statistics, Göteborg University.

[13] Frisén, M. and J. de Maré (1991). Optimal Surveillance. *Biometrika*, 78(2), pp 271-280.

[14] Harris, T. J. and W. H. Ross (1991). Statistical Process Control Procedures for Correlated Observations. *Canadian Journal of Chemical Engineering*, 69, pp 48-57.

[16] Jones, R. H., D. H. Crowell and L. E. Kapuniai (1970). Change Detection Model for Serially Correlated Multivariate Data. *Biometrics*, 26(2), pp 269-280.

[17] Kjelle, P.-E. (1987). *Alarm Criteria for the Fixed Gamma Radiation Monitoring Stations,* Research Report 1987:7, National Institute of Radiation Protection.

[18] Kramer, H. G. and W. Schmid (1997). EWMA Charts for Multivariate Time Series. *Sequential Analysis.* 16(2), pp 131-154.

[19] Montgomery, D. C. (1997). *Introduction to Statistical Quality Control.* John Wiley and Sons.

[20] Montgomery, D. C. and C. M. Mastrangelo (1991). Some Statistical Process Control Methods for Autocorrelated Data. *Journal of Quality Technology,* 23(3), 179-193.

[21] Notohardjono, B. D. and D. S. Ermer (1986). Time Series Control Charts for Correlated and Contaminated Data. *Journal of Engineering for Industry,* 108, pp 219-226.

[22] Pettersson, M. (1998). Monitoring a Freshwater Fish Population: Statistical Surveillance of Biodiversity. *Environmetrics.* Vol 9, pp 139-150.

[23] Ryan, T. P. (1991). Discussion. *Journal of Quality Technology,* 23(3), pp 200-202.

[24] Schmid, W. (1995). On the run length of a Shewhart chart for correlated data. *Statistical Papers,* Berlin, 36, pp 111-130.

[25] Schmid, W. and A. Schöne (1997). Some Properties of the EWMA Control Chart in the Presence of Autocorrelation. *The Annals of Statistics,* 25(3), pp 1277-1283.

[26] Shewhart, W. A. (1931). *Economic Control of Manufactured Product.* Princeton, NJ.

[27] Stone, R. and M. Taylor (1995). Time series models in statistical process control: considerations of applicability. *The Statistician,* 44(2), pp 227-234.

[28] Superville, C. R. and B. M. Adams (1994). An Evaluation of Forecast-based Quality Control Scemes. *Communications in Statistics - Simulation and Computation,* 23(3), pp 645-661.

[29] Sveréus, A. (1995). *Detection of Gradual Changes. Statistical Methods in Post Marketing Surveillance.* Research report 1995:2, Department of Statistics, Göteborg University.

[30] VanBrackle III, L. N. and M. R. Reynolds Jr (1997). EWMA and CUSUM Control Charts in the Presence of Correlation. *Communications in Statistics - Theory and Methods*, 26(3), 979-1008.

[31] Vasilopoulos, A. V. and A. P. Stamboulis (1978). Modification of Control Chart Limits in the Presence of Data Correlation. *Journal of Quality Technology*, 10(1), 20-30.

[32] Wardell, D. G., H. Moskowitz and R. D. Plante (1992). Control Charts in the Presence of Data Correlation. *Management Science*, 38, 1084-1105.

[33] Wardell, D. G, H. Moskowitz and R. D. Plante (1994). Run-Length Distributions of Special-Cause Control Charts for Correlated Processes. *Technometrics*, 36(1), 3-27.

[34] Wei, W. W-S. (1990). *Time Series Analysis: Univariate and Multivariate Methods.* Addison.

[35] Wessman, P. (1998). Some Principles for Surveillance Adopted for Multivariate Processes with a Common Change Point. *Communications in Statistics - Theory and Methods.* 27(5).

[36] Wetherill, G. B. and D. W. Brown (1991). *Statistical Process Control - Theory and practice.* Chapman and Hall.

[37] Yashchin, E. (1993). Performance of CUSUM Control Schemes for Serially Correlated Observation. *Technometrics*, 35(1), pp 37-52.

[38] Zhang, N. F. (1997). Detection capability of residual control chart for stationary process data. *Journal of Applied Statistics*, 24(4), 475-492.

Legend to Figures

1. Simulated example of an AR(1) process ($\phi = 0.5$) with a shift in mean, $\mu(t) = 10 \cdot \mathbf{1}\{t = 40, 41, \ldots\}$.

2. The residuals, $R(t)$, from the process in Figure 1.

3. The alarm limits for the direct (solid) and modified (dotted) shewhart for different values of ϕ.

4. The *pdf* of the in control run length, $f^0_{t_A}(t)$, for the modified Shewhart (cross) and the residual method (ring) for a process with high autocorrelation ($\phi = 0.9$).

5. The ARL_0 for direct shewhart for different values of ϕ. The limits were set to make $ARL_0 = 11$ for $\phi = 0$.

6. The ARL_1 for the modified Shewhart (dotted) and the residual method (solid) for different values of ϕ, where $ARL_0 = 11$.

7. The $PSD(1)$, *i.e.* the probability of immediate detection after a change, for the modified (dotted) and residual method (solid) for different values of ϕ, where $ARL_0 = 11$.

8. The $PSD(d)$, *i.e.* the probability of detection before d timesteps after the change, for the modified (dotted) and residual (solid) method for different values of ϕ, where $ARL_0 = 11$. The upper and lower pairs have $d = 7$ and $d = 3$, respectively.

9. The $PV(t)$, *i.e.* the predictive value of an alarm at time t, for the modified and residual method, where $\phi = 0.2$ and $\phi = 0.9$. The solid lines for the residual method and dotted for the modified Shewhart. (X) marks for the situation where $\phi = 0.2$ and (o) for $\phi = 0.9$. The dotted line without marks is for $\phi = 0$. The incidence $\upsilon = 0.1$.

Fig 1

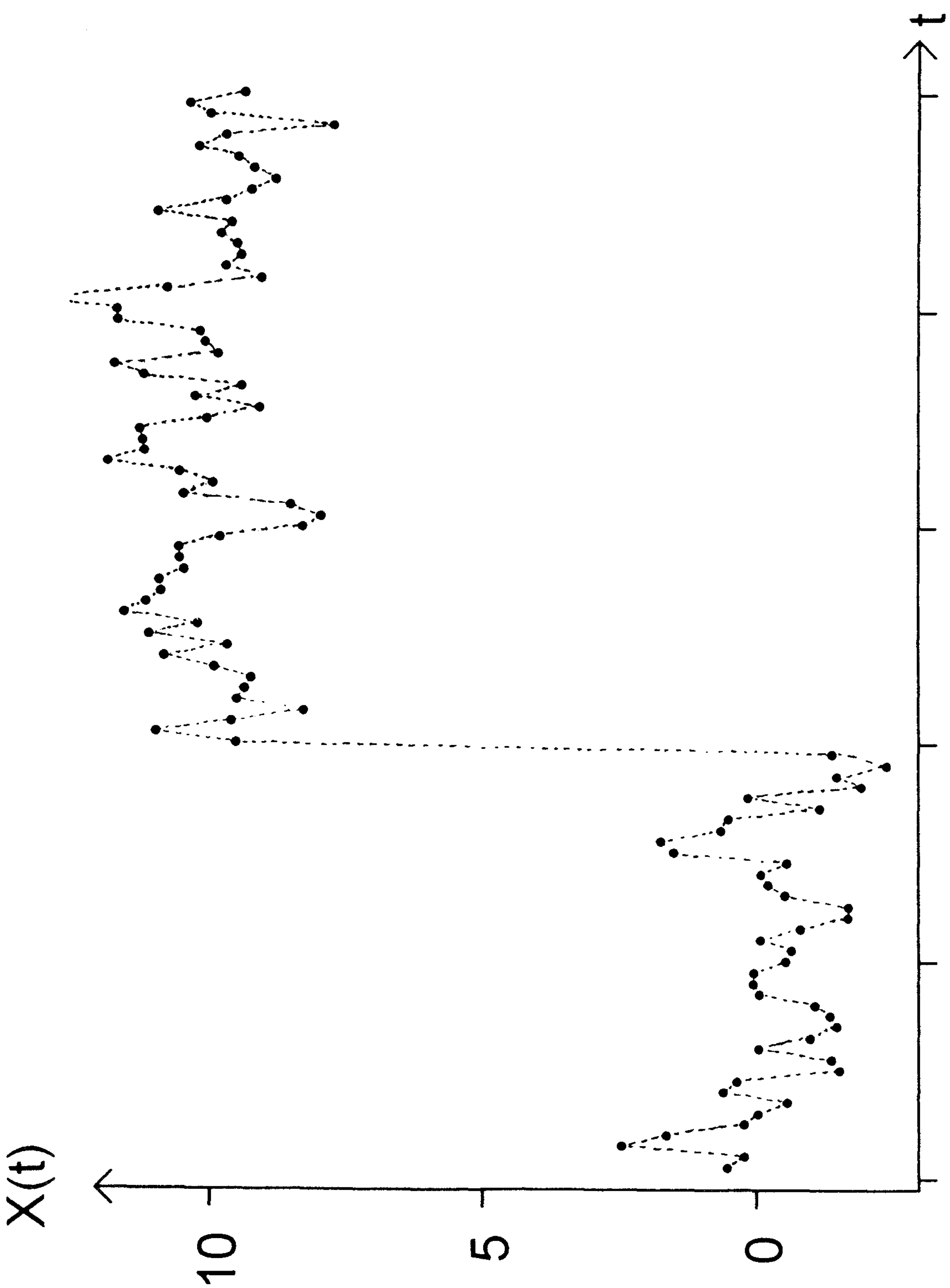


Fig 2

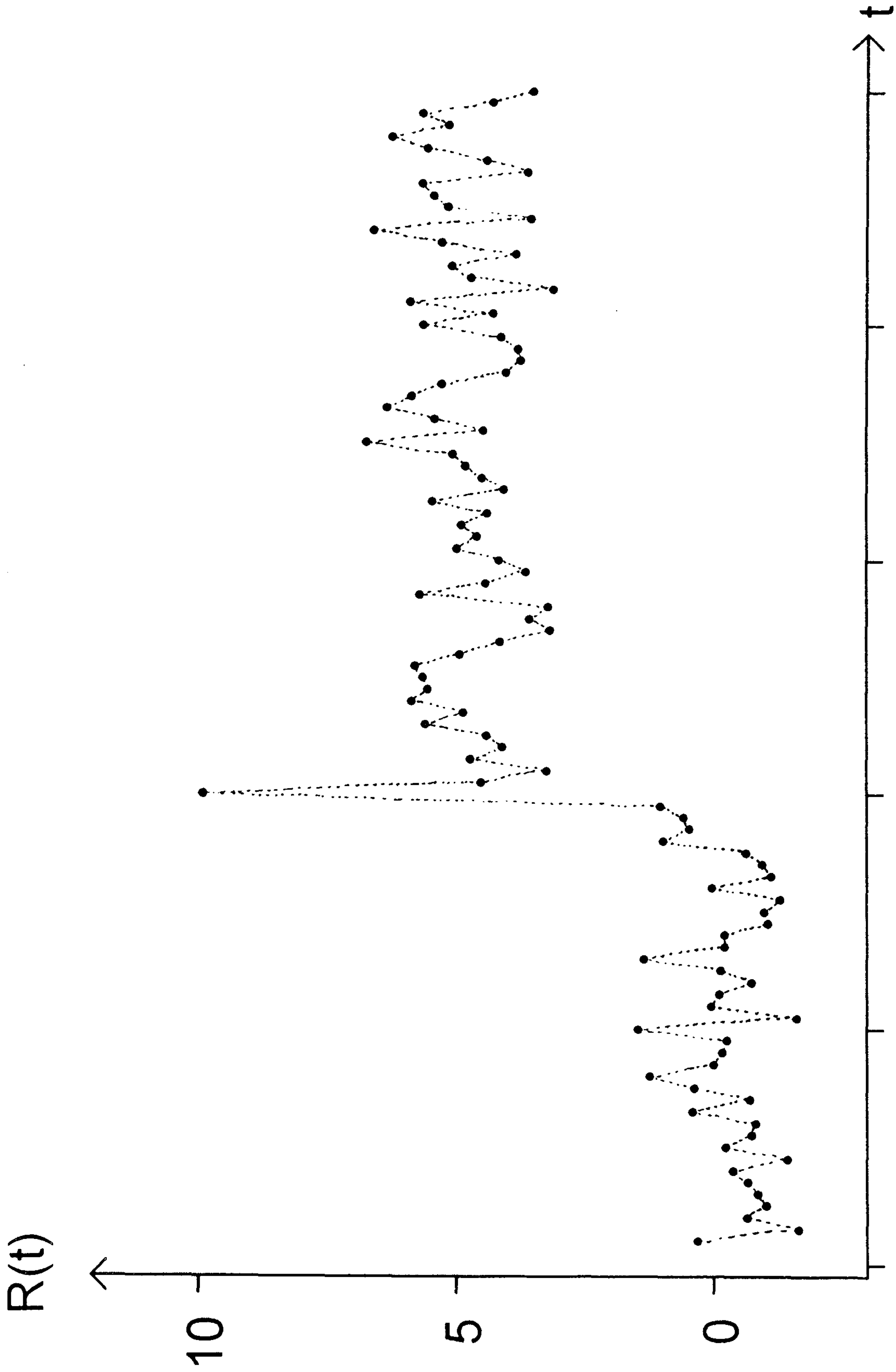


Fig 3

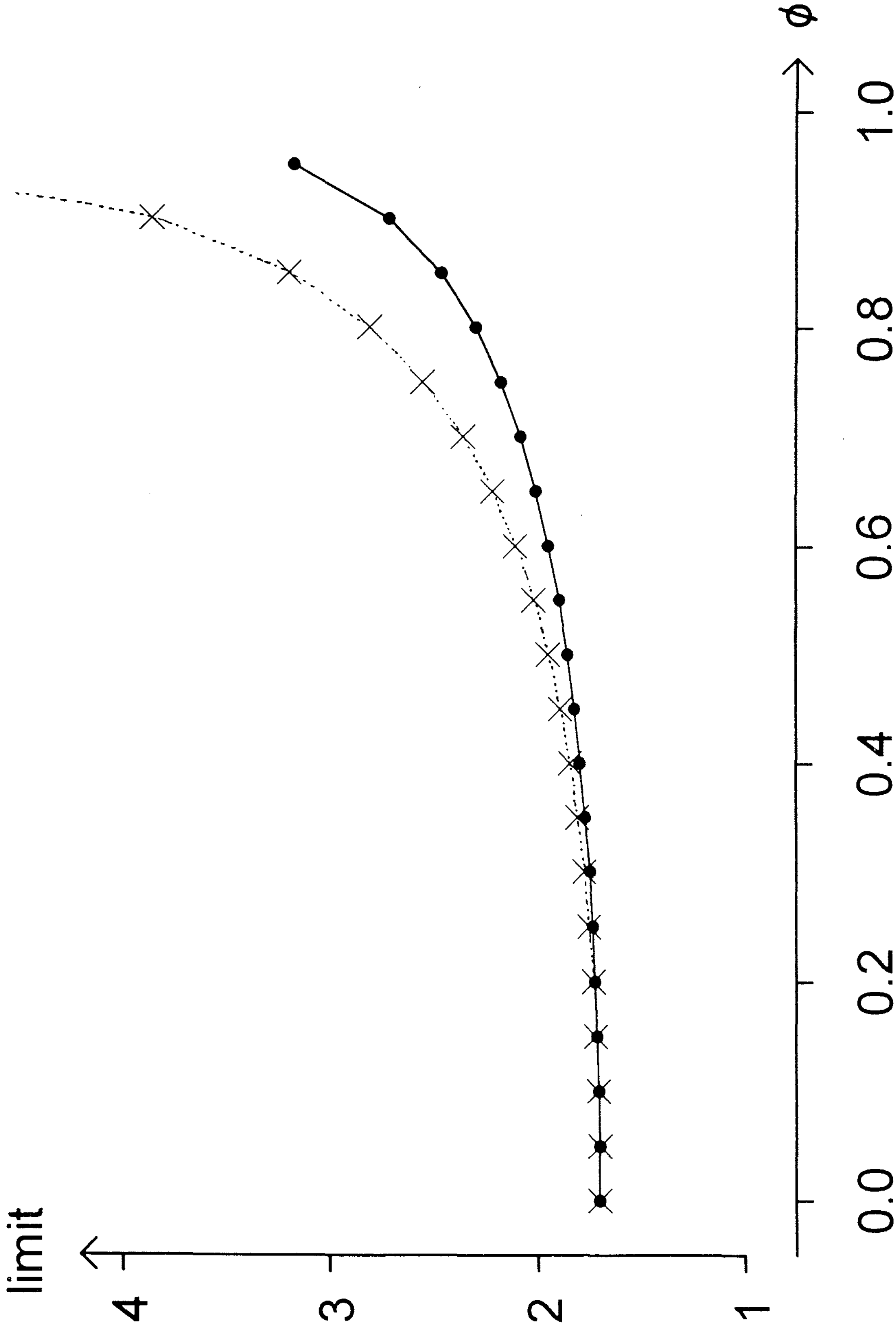


Fig 4

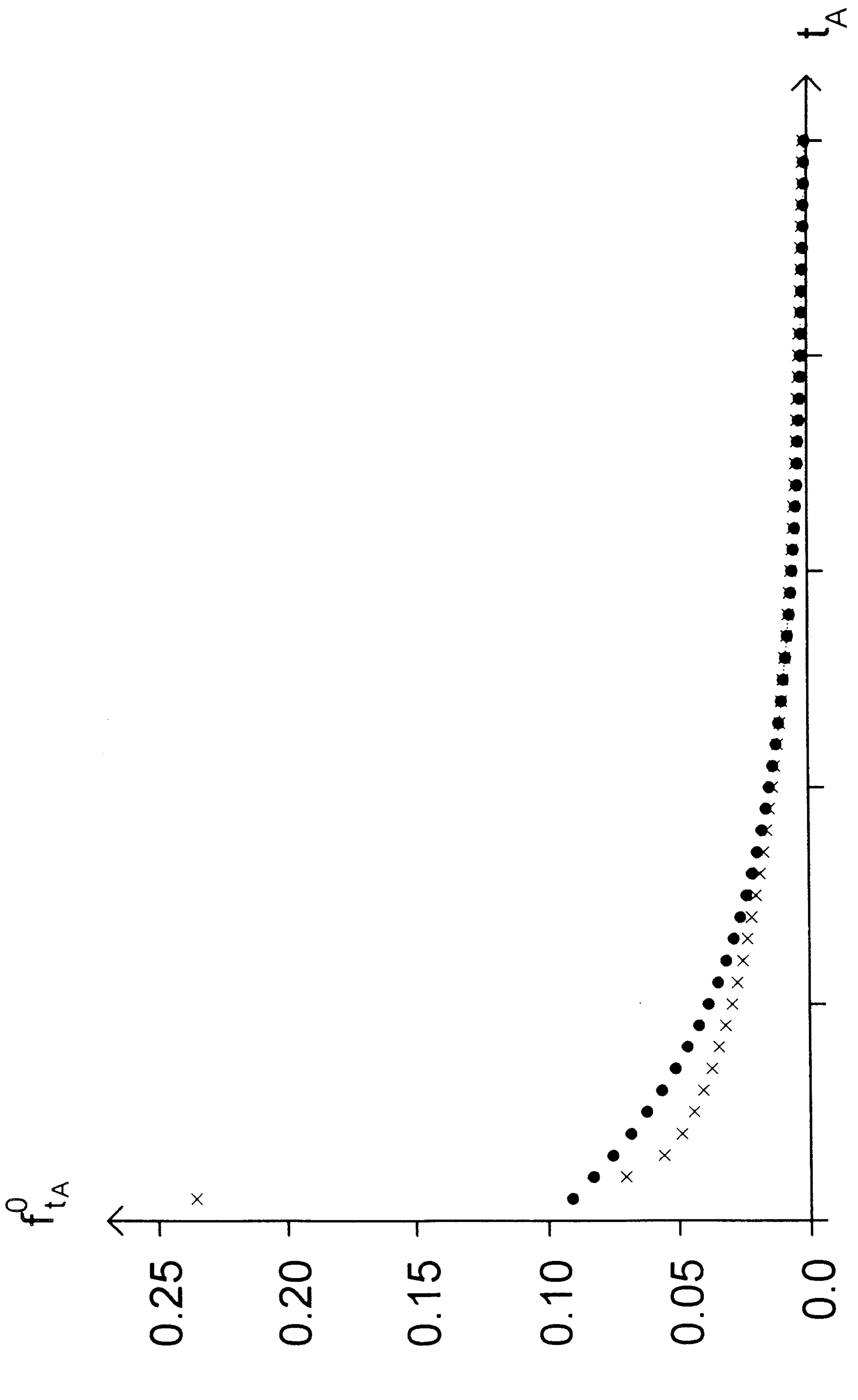
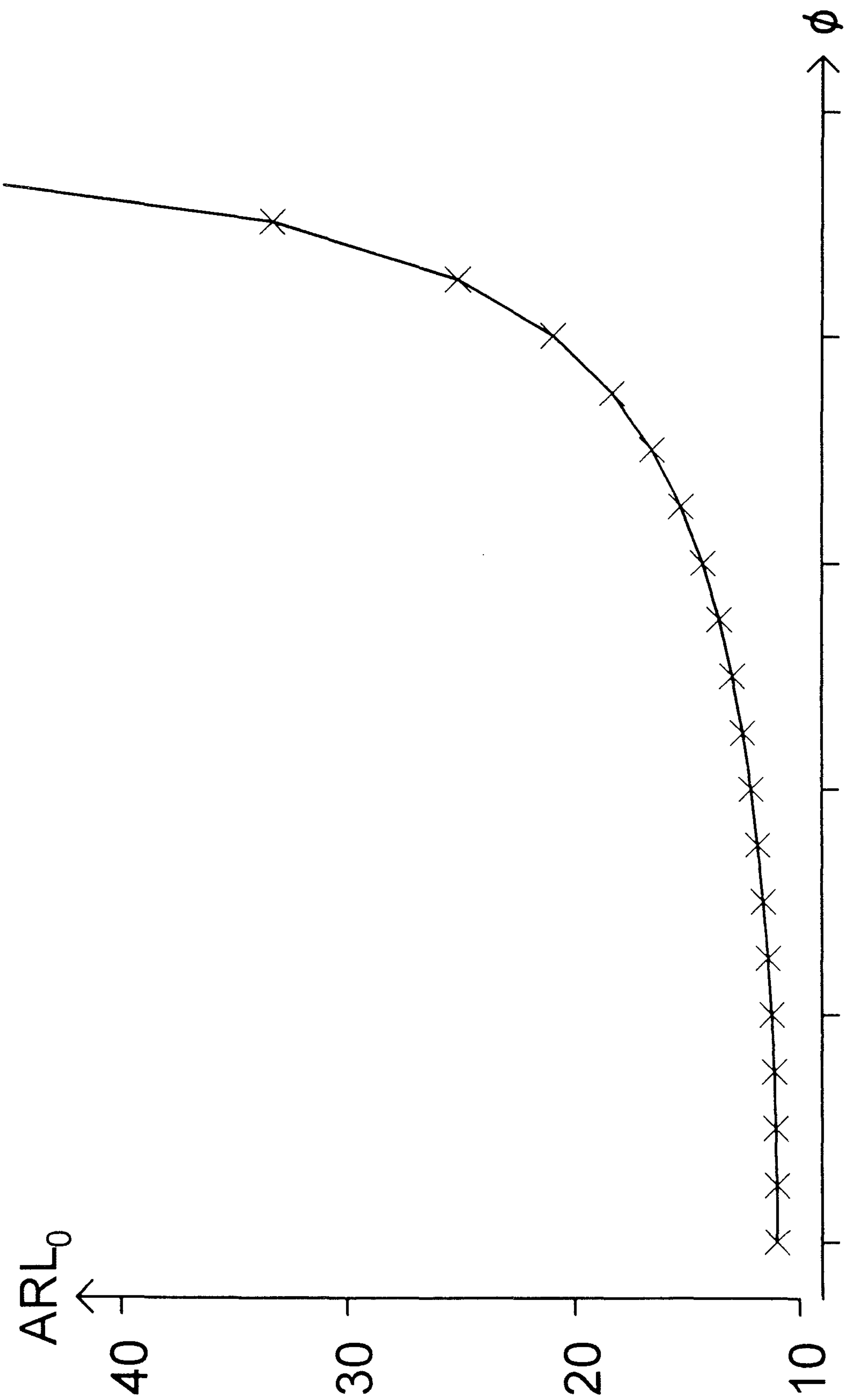


Fig 5



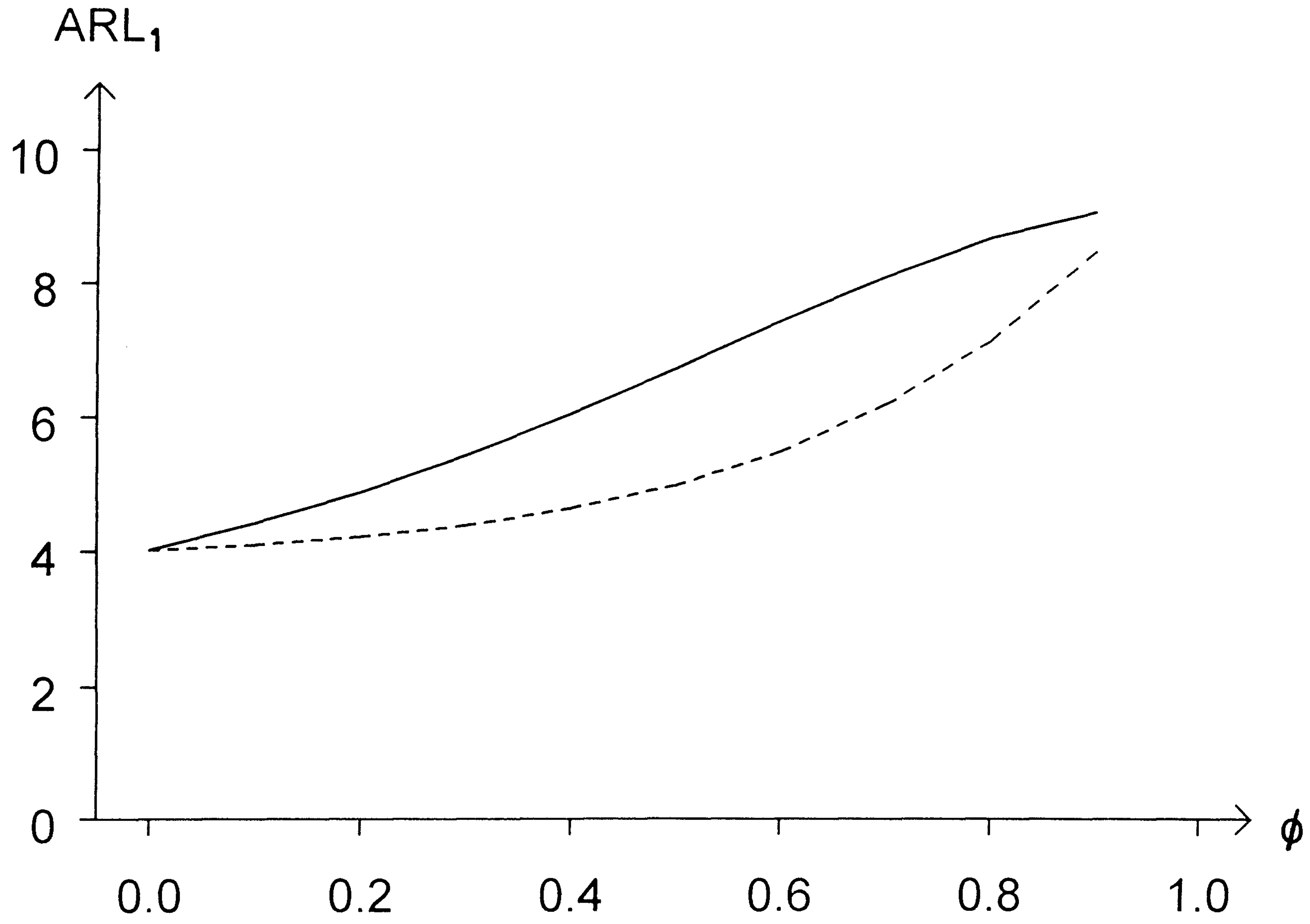


Fig 6

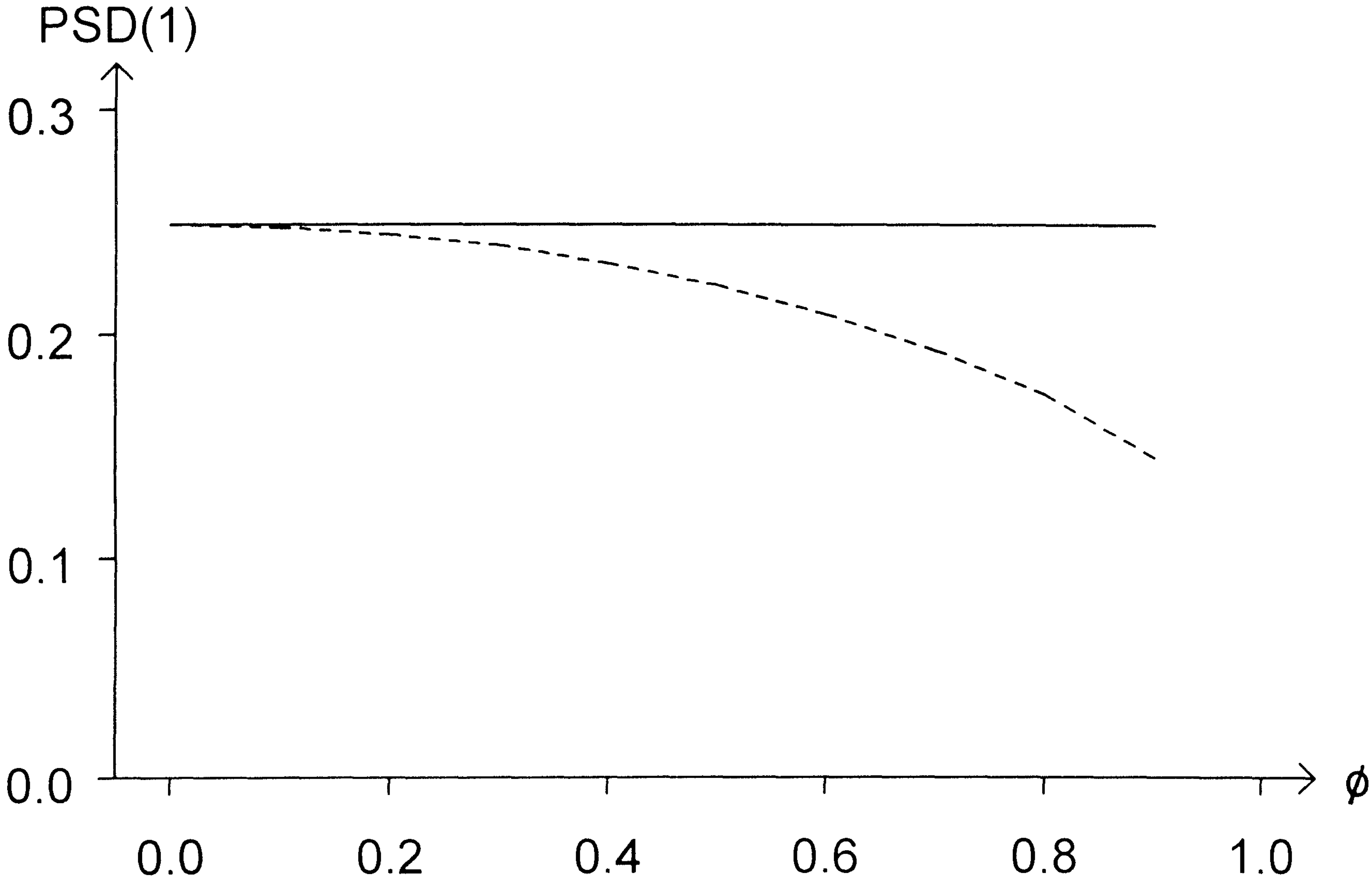


Fig 7

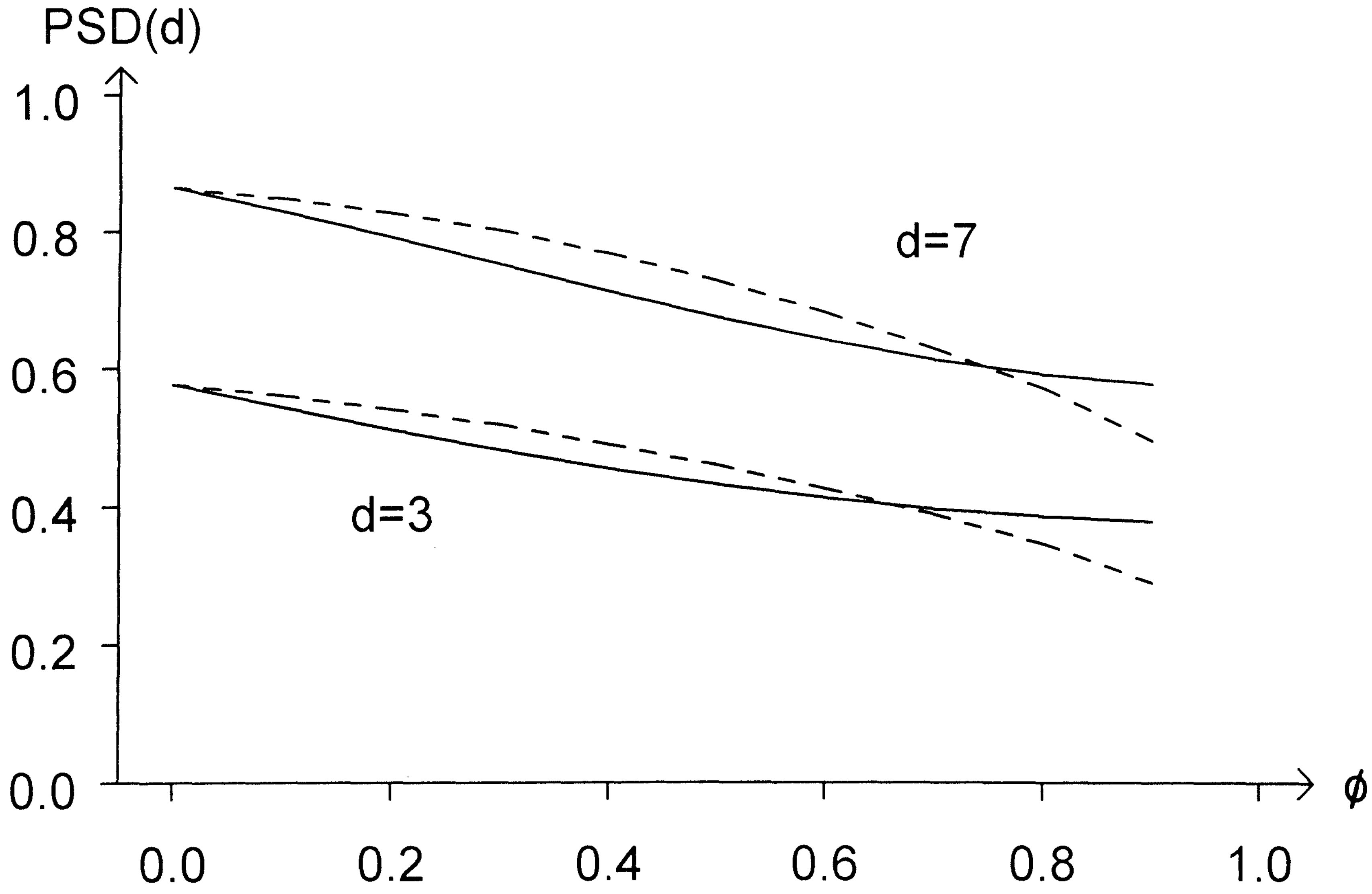


Fig 8

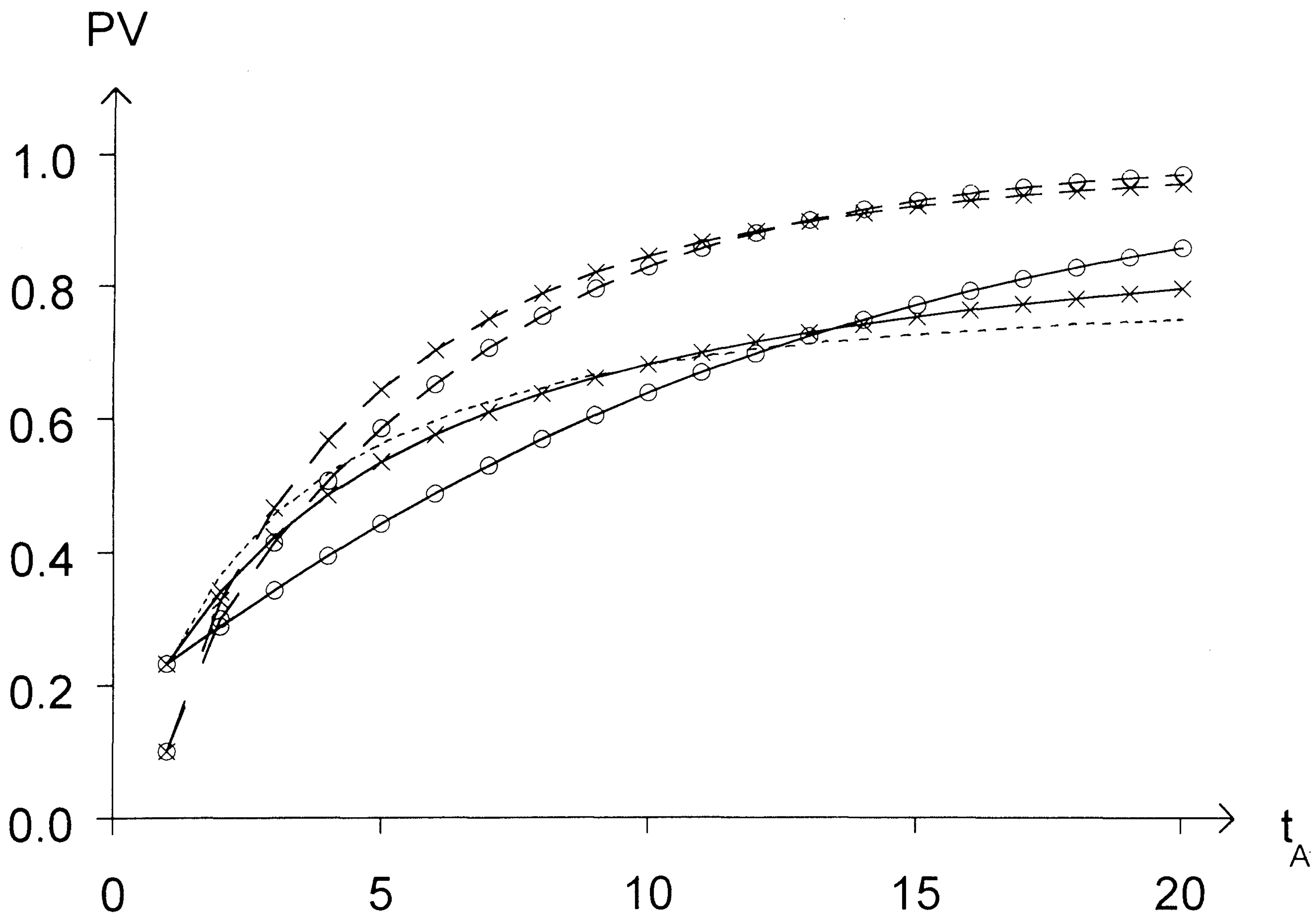


Fig 9

Research Report

1996:1	Ekman, A	Sequential analysis of simple hypotheses when using play-the-winner allocation.
1996:2	Wessman, P	Some principles for surveillance adopted for multivariate processes with a common change point.
1996:3	Frisén, M. & Wessman, P	Evaluations of likelihood ratio methods for surveillance.
1996:4	Wessman, P.	Evaluation of univariate surveillance procedures for some multivariate problems.
1996:5	Särkkä, A.	Outlying observations and their influence on maximum pseudo-likelihood estimates of Gibbs point processes.
1997:1	Ekman, A.	Sequential probability ratio tests when using randomized play-the-winner allocation.
1997:2	Pettersson, M.	Monitoring a Freshwater Fishpopulation: Statistical Surveillance of Biodiversity.
1997:3	Jonsson, R.	Screening-related prevalence and incidence for non-recurrent diseases.
1997:4	Johnsson, T.	A stepwise regression procedure applied to an international production study - a multiple inference solution.
1998:1	Särkkä, A. & Högmander, H.:	Multiple spatial point patterns with hierarchical interactions.
1998:2	Frisén, M. & Wessman, P.:	Evaluations of likelihood ratio methods for surveillance. Differences and robustness.
1998:3	Järpe, E.:	Surveillance of spatial patterns. Change of the interaction in the Ising model.