Research Report
Statistical Research Unit
Department of Economics
Göteborg University
Sweden

On curve estimation
under order restrictions

Kjell Pettersson

Uppsats för licentiatexamen i statistik
Göteborg
November 2007

Research Report
2007:15
ISSN 0349-8034

Mailing address:	Fax	Phone	Home Page:
Statistical Research Unit	Nat: 031-786 12 74	Nat: 031-786 00 00	http://www.statistics.gu.se/
P.O. Box 640 SE 405 30 Göteborg Sweden	Int: +46 31 786 12 74	Int: +46 31 786 00 00	

On curve estimation under order restrictions

Kjell Pettersson
Statistical Research Unit
Department of Economics
Göteborg University

Robust regression is of interest in many problems where assumptions of a parametric function may be inadequate. In this thesis, we study regression problems where the assumptions concern only whether the curve is increasing or decreasing. Examples in economics and public health are given. In a forthcoming paper, the estimation methods presented here will be the basis for likelihood based surveillance systems for detecting changes in monotonicity. Maximum likelihood estimators are thus derived. Distributions belonging to the regular exponential family, for example the normal and Poisson distributions, are considered. The approach is semiparametric, since the regression function is nonparametric and the family of distributions is parametric.

In Paper I, "Unimodal Regression in the Two-parameter Exponential Family with Constant or Known Dispersion Parameter", we suggest and study methods based on the restriction that the curve has a peak (or, equivalently, a trough). This is of interest for example in turning point detection. Properties of the method are described and examples are given.

The starting point for Paper II, "Semiparametric Estimation of Outbreak Regression", was the situation at the outbreak of a disease. A regression may be constant before the outbreak. At the onset, there is an increase. We construct a maximum likelihood estimator for a regression which is constant at first but then starts to increase at an unknown time. The consistency of the estimator is proved. The method is applied to Swedish influenza data and some of its properties are demonstrated by a simulation study.

Paper 1

Unimodal regression in the two-parameter exponential family with constant or known dispersion parameter

KJELL PETTERSSON

Statistical Research Unit, Department of Economics, Göteborg University,
SE 405 30 Göteborg, Sweden
e-mail: kjell.pettersson@statistics.gu.se

SUMMARY.

In this paper we discuss statistical methods for curve-estimation under the assumption of unimodality for variables with distributions belonging to the two-parameter exponential family with known or constant dispersion parameter. We suggest a non-parametric method based on monotonicity properties. The method is applied to Swedish data on laboratory verified diagnoses of influenza and data on inflation from an episode of hyperinflation in Bulgaria.

KEY WORDS: Non-parametric; Order restrictions; Two-parameter exponential family; Known dispersion parameter; Poisson distribution

1 INTRODUCTION

One of the central subjects of statistics is the estimation of curves. There exists a vast literature on the subject. Examples of methods are regression analysis and time series analysis. Often one has some knowledge in advance of the studied phenomenon that may be used in the analysis of the data. Such knowledge may be that the shape of the curve is known. In, for example, the study of the evolution in time of influenza during a season it is known that the number of reported cases per week of influenza-like illness first tends to increase and after reaching a peak tends to decrease. In such applications, it is reasonable to assume that the curve has a unimodal shape. Existing theory may motivate the use of some parametric formulation of the model. In the absence of such knowledge of the functional form of the curve one may use methods with fewer assumptions.

Some smoothing method may be considered when no information of the shape of the curve is available. One may for example calculate a simple moving average with all non-zero weights equal. Since the weights may be regarded as a discrete log-concave function unimodality will be preserved as pointed out by Frisén [1]. Anderson and Bock [2] found, however, that the location of the maximum is generally not preserved.

Example of another kind of method, which may be considered when no information about the shape of the curve is available, is to use smoothing splines. One procedure is described by Silverman [3]. In order to produce a good fit to data and to get a smooth curve he minimizes the following quantity with respect to g:

$$\sum\left(x(t)-g(t)\right)^2+\alpha\int g''(u)^2du\,,$$

where $x(t)$ is the observed value at time t $(t=1,2,...n)$, $\int g''(u)^2du$ is a roughness penalty and α is a smoothing parameter. This method does not preserve unimodality since the weights are in general not log-concave [1].

Information about the shape of the curve can be of different kinds. Sometimes it is known that the curve is concave. Hildreth's [4] method of concave regression may then be used. This method gives consistent estimates of the curve [5]. Holm and Frisén [6] propose a method for estimating concave or convex and increasing or decreasing functions. Dahlbom [7] modified their algorithm and extended the method to estimate sigmoid and unimodal concave functions. In her paper, there is an extensive analysis of the properties of the estimators for different curve forms. The assumption of concavity is unrealistic in some applications. In, for example, a study of influenza it is shown by [8] that in the up-phase an exponential function seems to describe the number of laboratory diagnosed cases rather good. To the down-phase, an exponentially decreasing function can be fitted. Such a mixture of exponential functions is not concave. We do not consider methods for estimating unimodal functions under restrictions of concavity in the present paper.

Gill and Baron [9] consider a method for estimating a continuous change of the canonical parameter of an exponential family from a constant level to a linear function. By using a non-linear transformation of the time-scale the results can be generalized to a non-linear continuous change of the parameter. They give conditions for consistent estimators of the change-point. They consider parameters with known behaviour after the change-point.

As was seen above, there exist different methods to estimate regressions with order restrictions. However, here we concentrate on regression where the only restriction on shape is that of unimodality.

Davies and Kovac [10] describe methods for nonparametric regression controlling the number of local extremes. The methods considered are the run-method and the taut-string multiresolution method. In the run method there is a restriction on the maximum run-length

of the sign of the residuals. The taut-string method was first proposed by [11] and extended to nonparametric regression by [12]. In the integrated process, one constructs upper and lower limits. A taut string is a function within those limits with the shortest length. The derivative of the taut string is the estimator of the curve. The estimates at the local extremes may be adjusted to get better results. The two methods have been used to estimate a unimodal curve.

In the estimation of a monotone function regression splines may be used as described by Ramsay [13]. The idea is to define a knot sequence which partitions the interval into subintervals. In each subinterval a non-negative linear combination of a small number of monotone splines are fitted, This type of method has been used by Meyer [14] to estimate a unimodal density with known mode. To each side of the mode she fits monotone regression splines under the restriction of continuity at the mode.

None of the methods, for unimodal regression mentioned above, however, give maximum likelihood estimators. Such estimators were constructed for the normal distribution with known variances in [1] but here we aim at maximum likelihood estimation for a wider class of members of the exponential family. These maximum likelihood-estimates are needed in surveillance, see e.g. [15].

In a study of influenza, [8] the Poisson distribution and the normal distribution were used to describe the distribution at given time points. The Poisson distribution belongs to the exponential family and has one parameter. The one-parameter exponential family may be regarded as a special case of the two-parameter exponential family with known or constant dispersion parameter. The normal distribution with constant variance is in the class of the exponential family with constant dispersion parameter.

These are the motivations of this paper, in which we study unimodal regression for variables with distributions belonging to the two-parameter exponential family with constant or known dispersion parameter. This kind of estimator is of interest for example in some economical problems and for outbreaks of infectious diseases as will be further discussed in Section 5. Andersson, Bock, Frisén and Pettersson have analyzed outbreaks of influenza in order to construct of methods for online detection of onsets and peaks of influenza [15-20]

The outline of the paper is the following: In Section 2 the model is described. In Section 3 we give the estimator. Some properties of the estimator are given in Section 4. Some applications of the method are described in Section 5. Concluding remarks are given in Section 6.

2 THE MODEL

2.1 The family of distributions

We observe a random process, $X(t)$, at n discrete values of the ordering variable t which will here be called time. The process may be defined as well in discrete time as in continuous time. In both cases, inferences are only for the behaviour of the process at the observed time points. We further restrict the attention to processes for which $X(t)$ has a distribution belonging to the two-parameter exponential family with constant or known dispersion parameter. The one-parameter exponential family may be regarded as a special case with a known dispersion parameter. We also assume that $X(t)$ and $X(u)$ are independent for $t \neq u$ and that $t,u \in (1,2,...,n)$. In applications, the assumption of independency may often be a realistic since we observe the process at discrete time points.

We write a probability function belonging to the exponential family in the canonical form as in [21]

$$f\left(x(t);\theta(t),\phi(t)\right)=\exp\left\{\frac{\left(x(t)\theta(t)-b\left(\theta(t)\right)\right)}{a\left(\phi(t)\right)}-c\left(x(t);\phi(t)\right)\right\} \tag{1}$$

where $\theta(t)\in\Theta(t)$ is constant for each t and $a(\cdot)$, $b(\cdot)$ and $c(.)$ are some functions. $\phi(t)$ is the dispersion parameter, which is regarded as a nuisance parameter. $a\left(\phi(t)\right)>0$ is of the form $a\left(\phi(t)\right)=\frac{\phi(t)}{\omega(t)}$ where $\omega(t)>0$ are known weights for all t. We also assume that the dispersion parameter $\phi(t)$ either is known or if unknown $\phi(t)=\phi$ for all t. f can be either the p.d.f. for a continuous random variable, e.g. the exponential distribution or the probability function for a discrete random variable such as the Poisson distribution.

It is also assumed that the family is regular as defined by Brown [22], i.e. that if $\Xi(t)=\left\{\theta(t):\int_{-\infty}^{\infty}\exp\left[\frac{\left(x(t)\theta(t)\right)}{a\left(\phi(t)\right)}-c\left(x(t);\phi(t)\right)\right]dx(t)<\infty\right\}$ then the parameter space $\Theta(t)$ is defined as $\Theta(t)=\text{int}\left(\Xi(t)\right)$. The parameter space shall thus be an open set. If f is the probability function for a discrete random variable, then the integral should be replaced by a sum.

It is also assumed the first and second derivatives of $b(\theta)$ with respect to θ exist and that $\frac{\partial^2 b\left(\theta(t)\right)}{\partial\theta^2}>0$. It is a well-known fact that

$$\mu(t)=E\left(X(t)\right)=\frac{\partial b\left(\theta(t)\right)}{\partial\theta}$$

$$Var\left(X(t)\right)=a\left(\phi(t)\right)\frac{\partial^2 b\left(\theta(t)\right)}{\partial\theta^2}$$

2.2 The regression function

In the present paper, we restrict attention to the case when the expected value of the process first is increasing and after having reached a peak decreases. The methods are easily modified for the case when the expectations first decrease and after a trough increase.

We define unimodality as follows: There exists a t' such that

$$\left.\begin{aligned}&\mu_{\max}=\mu(1)\geq\ldots\geq\mu(n)\ \text{ for } t'=1\\&\mu(1)\leq\ldots\leq\mu(t'\text{-}1)\leq\mu_{\max}\ \text{ and }\ \mu_{\max}\geq\mu(t')\geq\ldots\geq\mu(n)\ \text{ for } t'\in(2,3,\ldots,n)\\&\mu(1)\leq\ldots\leq\mu(n)=\mu_{\max}\ \text{ for } t'=n+1\end{aligned}\right\} \tag{2}$$

where $\mu_{\max}=\max_{1\leq t\leq n}\mu(t)$ and there is at least one strict inequality in (2).

3 THE MAXIMUM LIKELIHOOD ESTIMATOR

For $X(1),X(2),\ldots,X(n)$, independently distributed random variables, $X(t)$, $t\in(1,2,\ldots,n)$, having a distribution belonging to the two-parameter exponential family (1) we assume that

there are $m(t)$ observations on $X(t)$ for each *t*. In Lemma 1, we study the case of a monotone regression. Denote the maximum likelihood estimator of $\boldsymbol{\mu} = \left(\mu(1), \mu(2), ..., \mu(n)\right)'$ subject to $\mu(1) \le \mu(2) \le ... \le \mu(n)$ by $\tilde{\boldsymbol{\mu}} = \left(\tilde{\mu}(1), \tilde{\mu}(2), ..., \tilde{\mu}(n)\right)'$. Then the following may be shown.

Lemma 1:
(a) $\tilde{\boldsymbol{\mu}}$ is given by minimizing

$$\sum_{t=1}^{n} \left(\bar{x}(t) - \mu(t)\right)^2 \frac{m(t)}{\phi(t)} \tag{3}$$

with respect to $\mu(t)$ $(t = 1, 2, ..., n))$ under the restriction of isotonicity for $\mu(t)$ if $\phi(t)$ is known for all *t*.
(b) $\tilde{\boldsymbol{\mu}}$ is given by minimizing

$$\sum_{t=1}^{n} \left(\bar{x}(t) - \mu(t)\right)^2 m(t) \tag{4}$$

with respect to $\mu(t)$, $(t = 1, 2, ..., n))$, under the restriction of isotonicity for $\mu(t)$ if $\phi(1) = \phi(2) = ... = \phi(n)$

The maximum likelihood estimator of $\boldsymbol{\mu}$, subject to $\mu(1) \ge \mu(2) \ge ... \ge \mu(n)$ and the family of distributions of Section 2.1, is obtained by minimizing (3) and (4) respectively under the restriction of antitonicity.

Proof: Silvapulle and Sen [23] consider the following case: Let $x_1(t), ..., x_{m(t)}(t)$ be $m(t)$ independent observations on the random variable $X(t)$ from group t, $(t = 1, ..., n)$.

We want to find the maximum likelihood estimator of $\left(\mu(1), ...\mu(n)\right)$ where $\mu(t) = E\left(X(t)\right)$ under the restriction $A\boldsymbol{\mu} \ge 0$. A is a matrix in which each row is a permutation of $(-1, 1, 0, ..., 0)$ and $\mu = \left(\mu(1), ..., \mu(n)\right)'$. Assume that the distribution of $X(t)$ belongs to the exponential family with parameters $\theta(t)$ and $\phi(t)$. Part (a) of proposition 2.4.3 in [23] states that the maximum likelihood estimator $\tilde{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ under the restriction $A\boldsymbol{\mu} \ge 0$ is the value of $\boldsymbol{\mu}$ at which

$$\sum_{t=1}^{n} \left(\bar{x}(t) - \mu(t)\right)^2 \frac{m(t)}{\phi(t)} \tag{5}$$

reaches its minimum subject to $A\boldsymbol{\mu} \ge 0$ if $\phi(1), \phi(2), ..., \phi(n)$ are known constants. Part (b) of proposition 2.4.3 in [23] states that the maximum likelihood estimator $\tilde{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ under the restriction $A\boldsymbol{\mu} \ge 0$ is the value of $\boldsymbol{\mu}$ at which

$$\sum_{t=1}^{n} \left(\bar{x}(t) - \mu(t)\right)^2 m(t) \tag{6}$$

reaches its minimum subject to $A\boldsymbol{\mu} \ge 0$ if $\phi(1) = \phi(2) = ... = \phi(n) = \phi$, say. In isotonic regression the i:th row of A has -1 in position i, +1 in position $i+1$ and 0 in all other positions. For antitonic regression the i:th row has +1 in position i, -1 in position $i+1$ and 0

in all other positions. Thus the maximum likelihood estimator $\tilde{\boldsymbol{\mu}}$ under the restriction of isotonicity and antitonicity respectively is thus given by minimizing $\sum_{t=1}^{n}\left(\bar{x}(t)-\mu(t)\right)^2\frac{m(t)}{\phi(t)}$ with respect to $\mu(t)$ under the restrictions of isotonicity and antitonicty respectively if $\phi(t)$ is known for all t. For $\phi(1)=\phi(2)=...=\phi(n)=\phi$ $\tilde{\boldsymbol{\mu}}$ is given by minimizing $\sum_{t=1}^{n}\left(\bar{x}(t)-\mu(t)\right)^2 m(t)$ with respect to $\mu(t)$ under the restrictions of isotonicity and antitonicty respectively □

An estimator $\tilde{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ under the restriction of unimodality may be constructed in the following way [1]. Regard the following partitions of the observations $x(1),x(2),...,x(n)$:

$$\left(\{\varnothing\},\{x(1),...,x(n)\}\right),\left(\{x(1)\},\{x(2),...,x(n)\}\right),...,\left(\{x(1),...,x(n)\},\{\varnothing\}\right). \qquad (7)$$

For each of these partitions we fit monotone regressions to each part. To the first part is fitted an isotonic regression and to the second an antitonic regression. Denote the likelihood for the fitted unimodal regression for the i:th partition by L_i $(i=1,2,...n+1)$. $\tilde{\boldsymbol{\mu}}$ is given by the partition with gives $\max_{1\le i\le n+1}(L_i)$.

Theorem 1: $\tilde{\boldsymbol{\mu}}$ defined above is the maximum likelihood estimator of $\boldsymbol{\mu}$ under the restriction of unimodality for the regular exponential distribution with known or constant dispersion parameter as described in Section 2.1.

Proof: First, we regard the maximum likelihood estimator of $\boldsymbol{\mu}$ for a given t' as defined by (2) Assume that $t'=1$. Then $\mu(t)$ is an antitonic function of t. The maximum likelihood estimator of $\boldsymbol{\mu}$ is given by lemma 1. Denote the maximum of the likelihood function in this case by L_1. If $t'=n+1$ then $\mu(t)$ is an isotonic function of t. The maximum likelihood estimator of $\boldsymbol{\mu}$ follows from Lemma 1. Denote the maximum of the likelihood function by L_{n+1}. Now assume that $t'\in(2,3,...,n-1)$. According to the definition of t' in (2) then the curve is first increasing and then decreasing. If $\mu(1)\le...\le\mu(\text{t'-1})\le\mu_{\max}$ and $\mu_{\max}\ge\mu(t')...\ge\mu(n)$ for a given t', we fit an isotonic function according to lemma 1 to $x(1),...,x(t'-1)$. Denote the maximum of the likelihood by $L_{t'}^{I}$. Fit an antitonic function according to lemma 1 to $x(t'),...,x(n)$. Denote the maximum of the likelihood by $L_{t'}^{A}$.Then the maximal likelihood for the curve is $L_{t'}=L_{t'}^{I}\cdot L_{t'}^{A}$. Our maximum likelihood estimator $\tilde{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ under the restriction of unimodality is given by the partition, which maximizes $L_{t'}$ for $t'\in(1,2,...n+1)$ □

The method is illustrated by a numerical example in section 5.1.

The parameter $\phi(t)$ in (1) may be interpreted as a scale parameter and $\omega(t)$ as the number of observations on $X(t)$ for $t\in(1,2,...,n)$. For normally distributed variables, the scale parameter is the variance. When a unimodal regression is fitted to observations on normally distributed variables with varying variances, we correct for the differences in scale by weighting the observations inversely to their variances. If the dispersion parameter is constant

for all observations, we have the same scale for all observations and we therefore do not correct for differences in scale and give all observations a weight, which is proportional to the number of observations for all time points.

In section 2 it was stated that $\mu = \frac{\partial b(\theta)}{\partial \theta}$ and since it was assumed that $\frac{\partial^2 b(\theta)}{\partial \theta^2} > 0$ then μ is a strictly increasing function of θ. This motivates the following corollary.

Corollary: The maximum-likelihood estimator of $\theta(t)$, $\tilde{\theta}(t)$, is given by $b'^{-1}_{\theta}(\tilde{\mu}(t))$, where b'^{-1}_{θ} denotes the inverse of $\frac{\partial b(\theta)}{\partial \theta}$.

Proof: Since $\frac{\partial b(\theta(t))}{\partial \theta}$ is strictly increasing in $\theta(t)$ the inverse exists. If the likelihood is maximized for $\mu(t) = \tilde{\mu}(t)$ it is also maximized for the value $\tilde{\theta}(t)$ of $\theta(t)$ for which

$$\tilde{\mu}(t) = \left(\frac{\partial b(\theta(t))}{\partial \theta} \right)_{\theta = \tilde{\theta}} \quad \text{i.e. } \tilde{\theta}(t) = b'^{-1}_{\theta}(\tilde{\mu}(t)). \quad \square$$

4 PROPERTIES OF THE ESTIMATOR

The estimated curve preserves the unimodality since the transformation of the data is log-concave. See Frisén [1] for a proof. In this section we study consistency and bias.

4.1 Consistency.

When the number, *m(t),* of independent observations at each time point, *t*, tends to infinity, we have the following consistency property.

Theorem 2: In the class of distributions given in Section 2.1 $\tilde{\mu}(t)$ is a strongly consistent estimator of $\mu(t)$ for $t = 1, 2, ..., n$ when min *m(t)*$\rightarrow\infty$.

Proof: The theorem follows from the Kolmogorov law of large numbers since it is assumed in Section 2.1 that $\mu(t) = \frac{\partial b(\theta(t))}{\partial \theta}$ exists for $t = 1, 2, ..., n$ $\square$

From this it follows that both the height and the time of the peak will be consistently estimated. Observe that the consistency do not prevail for the case where the number of time points tends to infinity.

If there is only one observation for each time point but the number of time points tends to infinity then $\hat{\mu}_{\max} = \max[\tilde{\mu}(t)]$ is in general an inconsistent estimator of $\mu_{\max} = \max[\mu(t)]$ as is pointed out by Frisén [1] and Dahlbom [24] for a regression with normal distribution and by Woodroofe and Sun [25] for density estimation in the exponential family.

4.2 Bias.

In the case when there is one observation per time point the estimators of the end-points and the maximum points are positively biased, as was pointed out by Dahlbom [24]. She also found that the bias of the estimators of other points at the curve often is negligible. Some results from her simulation experiments of certain curves and the normal distribution will now be reviewed.

We focus our interest to the problem of estimating $\mu_{\max} = \max_{t=1,2,..n}\left[\mu(t)\right]$ when the time point for the maximum is unknown and when there is one observation at each of a fixed number of time points. As an estimator of $\mu_{\max}$ one may use $\hat{\mu}_{\max} = \max_{t=1,2,..n}\left[\tilde{\mu}(t)\right]$. Dahlbom [24] studied, the case when $t_{\max}$ is unique, where $t_{\max} = \arg\max_{t=1,2,...,n} \mu(t)$. One of the models for $\mu(t)$ was a symmetrical and concave second-degree polynomial. The bias of $\hat{\mu}_{\max}$ as an estimator of $\mu_{\max}$ was a decreasing function of the curvature normalised for the standard deviation. When the number of time points in an interval of fixed length increases then the bias tends to increase. Since deviations from symmetry may affect the bias, she also used a third-degree polynomial as a model for $\mu(t)$ with values of the coefficients giving unimodal and concave curves. It was found that moderate deviations from symmetry had small influence on the bias.

The errors in the simulation experiments by Dahlbom were normally distributed. There is no obvious reason to expect much different results for other members of the exponential family. The curves studied by Dahlbom in the simulation experiments are concave. The bias in the estimator of $\mu_{\max}$ for other curve forms and other distributions is not necessarily the same. As mentioned in Section 1 a mixture of exponential functions, one increasing and one decreasing, seems to give a good fit to laboratory diagnosed influenza in Sweden. Such mixtures of exponential functions are not concave.

5 EXAMPLES

5.1 A numerical example.

We have one observation on $X(t)$ at each of the time points $t = 1,2,3,4,5$ where $X(t)$ follows the Poisson distribution $P(\lambda(t))$. The Poisson distribution is in the exponential family of equation (1), with $\theta = \ln(\lambda)$, $b(\theta) = e^{\theta} = \lambda$, $a(\phi) = 1$, $c(x,\phi) = \ln(x!)$ and $\mu(t) = \dfrac{\partial b(\theta(t))}{\partial \theta} = e^{\theta} = \lambda$. We assume that $\mu(t)$ is unimodal as in (2).

We give the calculations for the case there the observed values of X are 1, 3, 1, 5, and 1. In the table below, we give the observed values and the estimates of $\mu(t)$ at different time points and the likelihood for different partitions of the observations. As an example, we get the following likelihood for the partition $\{1,3\},\{1,5,1\}$

$$L = \left(e^{-1}\cdot\frac{1^1}{1!}\right)\cdot\left(e^{-3}\cdot\frac{3^3}{3!}\right)\cdot\left(e^{-3}\cdot\frac{3^1}{1!}\right)\left(e^{-3}\cdot\frac{3^5}{5!}\right)\cdot\left(e^{-1}\cdot\frac{1^1}{1!}\right) = 0.457\cdot 10^{-3}$$

Table 1. The likelihood and the maximum likelihood estimators for each time conditional on each of the possible partitions of the dataset {1, 3, 1, 5, 1}

Partitions	t=1	t=2	t=3	t=4	t=5	Likelihood
$\{\varnothing\},\{1,3,1,5,1\}$	2.5	2.5	2.5	2.5	1	$0.221\cdot10^{-3}$
$\{1\},\{3,1,5,1\}$	1	3	3	3	1	$0.457\cdot10^{-3}$
$\{1,3\},\{1,5,1\}$	1	3	3	3	1	$0.457\cdot10^{-3}$
$\{1,3,1\},\{5,1\}$	1	2	2	5	1	$1.160\cdot10^{-3}$
$\{1,3,1,5\},\{1\}$	1	2	2	5	1	$1.160\cdot10^{-3}$
$\{1,3,1,5,1\},\{\varnothing\}$	1	2	2	3	3	$0.271\cdot10^{-3}$

The conclusion from Table 1 is that the maximum likelihood estimate of the curve is given in the rows for the partitions $\left(\{1,3,1\},\{5,1\}\right)$ and $\left(\{1,3,1,5\},\{1\}\right)$. Thus, the maximum likelihood estimator is $\tilde{\boldsymbol{\mu}} = (1,2,2,5,1)$.

5.2 An economical example.

An economical example where it is of interest to use an inverted U-shaped curve is in the study of inflation before, during and after periods of hyperinflation [26]. See for example monthly data on the inflation for Bulgaria during the time period from 1995 to 2000. After the end of central planning, there was a budget deficit and high monetary growth. The confidence in government was decreasing and the inflation was steadily increasing. During March, the twelve-month inflation was about 2000%. After reforms, Bulgarians became more confident in their currency and inflation decreased. In figure 1, we show the Bulgarian twelve-month inflation in percent during the time period from June 1995 to September 1999. We also show the unimodal regression function fitted under the assumption of normally distributed values with constant variances.

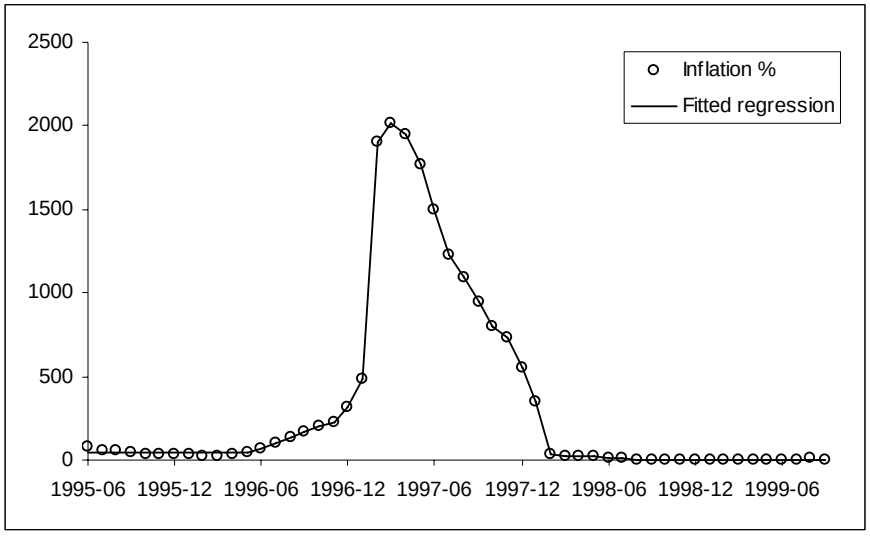

Figure 1. Bulgarian inflation and fitted regression for the years 1996 and 1997

5.3 Application to influenza incidences

In the study of the number of influenza cases during an ordinary season it may be reasonable to assume that the number of cases are initially increasing and that they after having reached a peak are decreasing. It is of interest to study the incidence, since influenza epidemics impose huge costs on society. The Swedish Institute for Infectious Disease control (SMI) publishes information on Swedish influenza incidence. Two types of weekly data are published, namely the number of influenza like illness (ILI) and laboratory diagnosed influenza cases (LDI). Details on the reporting can be found in [27].

The number of reported cases per week of influenza-like illness and the number of laboratory-diagnosed cases per week are counting variables. In some studies, it is assumed that the distribution of the number of cases can be approximated by the normal distribution. However, at the onset of an epidemic there are few cases. Assumption of normality then may assign a non-zero probability that cannot be ignored to a negative number of cases. The assumption of normality of the number of influenza cases has been criticized by Le Strat and Carrat [28]and Rath et al. [29]. Held et al [30] suggest the Poisson distribution for infectious surveillance data and also discuss the negative binomial distribution in cases of over-dispersion relative to the Poisson distribution. Sebastiani et al [31] use a log-normal distribution for ILI data. Andersson et al [8] studied the distributional properties for ILI and LDI in Swedish influenza data from five influenza seasons. They fitted piecewise exponential functions to each season and examined the residuals. The auto-correlations in the residuals were low all seasons for the two variables. Near the peak, there was no evident relation between the squared residuals and the estimated curve. Since there were numerous cases near the peak an assumption of normally distributed values with constant variance may be adequate for that part. In [19], they also studied the onset of an epidemic using ILI-data. They found that the squared residuals depend on the estimated means. Since there was no direct evidence against the Poisson distribution they suggested that this distribution may be used a first approximation.

Here we restrict our attention to the LDI data. The LDI-cases are reported weekly from five virus laboratories and between 15 and 20 microbiological laboratories. See [32] and [33] for details. The LDI cases are mostly patients in need of hospital care.

In figure 2, we show the number of cases and the unimodal regression for the number of laboratory diagnosed cases under the assumption of Poisson distributed values for the season 2005/2006

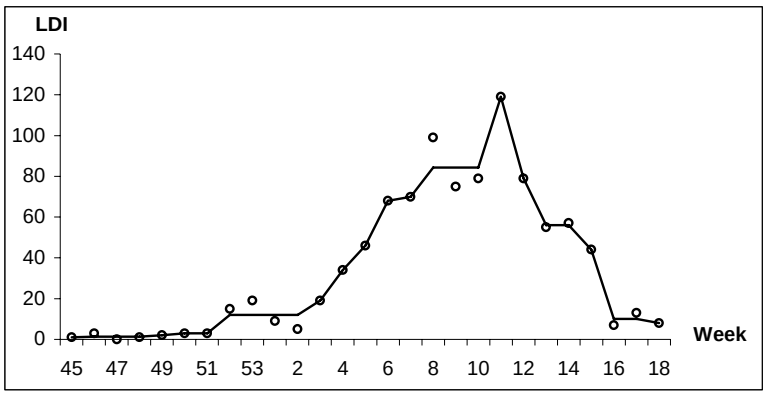

Figure 2. The number of LDI cases and the fitted regression during the season 2005/2006.

6 CONCLUDING REMARKS

We have given examples of situations where it is reasonable to assume that the evolution in time of a process may be described by a unimodal curve and when it can be assumed that the distribution of the observed process may be described by a distribution belonging to the one-parameter exponential distribution.

In unimodal regression with distributions belonging to the exponential family with varying dispersion parameter, the observations are weighted inversely proportional to that parameter if there is one observation per time point. For distributions belonging to a one-parameter exponential family or two-parameter exponential family with constant dispersion parameter, all observations shall have the same weight for one observation per time point. An example of a two-parameter distribution in the exponential family with constant dispersion parameter is the normal distribution with constant variance. An example of a distribution belonging to the one-parameter exponential family is the Poisson distribution. In this paper, we restrict attention to estimation of the unimodal regression curve under the assumptions given above. The results are also important in the construction of surveillance procedures in order to monitor different processes in time. An example is in timely detection of influenza peaks where the Poisson distribution is useful. In some financial time series, the assumption of a normal distribution with constant variance may be used as in the example of Bulgarian hyperinflation. The monitoring of such time series may be of interest to among others actors on the foreign exchange markets.

ACKNOWLEDGEMENTS

The author is grateful to Professor Marianne Frisén and Associate professor Eva Andersson for supervision of this work. Associate professor Dick Durevall has given valuable advice concerning the economical example. Research assistant Linus Schiöler has expertly given technical assistance. The research was supported by the Swedish Emergency Management Agency (grant 0622/204) and the Bank of Sweden Tercentenary Foundation (grant J2003-0558).

REFERENCES

[1] Frisén, M., 1986, Unimodal regression. *The Statistician,* **35,** 479-485.

[2] Andersson, E. & Bock, D. (2001) On seasonal filters and monotonicity. Department of Statistics, Göteborg University.

[3] Silverman, B.W., 1985, Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society B,* **47,** 1-52.

[4] Hildreth, C., 1954, Point estimation of ordinates of concave functions. *Journal of the American Statistical Association,* **49,** 598-619.

[5] Hanson, D.L. & Pledger, G., 1976, Consistency in Concave Regression. *The Annals of Statistics,* **4,** 1038-1050.

[6] Holm, S. & Frisén, M. (1985) Nonparametric regression with simple curve characteristics. Department of Statistics, Göteborg University.

[7] Dahlbom, U. (1994) Estimation of regression functions with certain monotonicity and concavity/convexity restrictions. Göteborg, Ph.D. Thesis. Department of Statistics, Göteborg University.
[8] Andersson, E., Bock, D. & Frisén, M., 2007, Modeling influenza incidence for the purpose of on-line monitoring. *Statistical Methods in Medical Research***,** (in press).
[9] Gill, R. & Baron, M., 2004, Consistent estimation in generalized broken-line regression. *Journal of Statistical Planning and Inference,* **126,** 460.
[10] Davies, P.L. & Kovac, A., 2003, Local Extremes, Runs, Strings and Multiresolution. *The Annals of Statistics,* **29,** 1-65.
[11] Barlow, R.E., Bartholomew, D.J., Bremer, J.M. & Brunk, H.D., 1972, *Statistical inference under order restrictions,* (London: Wiley).
[12] Mammen, E. & Van Der Geer, S., 1997, Locally Adaptive Regression Splines. *The Annals of Statistics,* **26,** 387-413.
[13] Ramsay, J.O., 1988, Monotone regression Splines in Action. *Statistical Science,* **3,** 425-461.
[14] Meyer, M.C. (2007) Algorithms for the Estimation of a Decreasing or Unimodal Density using Shape-Restricted Regression Splines. *International Statistical Institute Conference 2007.* Lisboa.
[15] Bock, D., Andersson, E. & Frisén, M., 2007, Statistical surveillance of epidemics: Peak detection of influenza in Sweden. *Biometrical Journal***,** (in press).
[16] Andersson, E., 2004, Monitoring system for detecting starts and declines of influenza epidemics. *Morbidity and Mortality Weekly Report,* **53,** 229-229.
[17] Frisén, M. & Andersson, E. (2007) On-line detection of outbreaks. Manuscript.
[18] Frisén, M., Andersson, E. & Pettersson, K. (2007) Estimation of outbreak regression. Submitted manuscript.
[19] Frisén, M., Andersson, E. & Schiöler, L. (2007) A non-parametric system for on-line outbreak detection of epidemics. Manuscript.
[20] Bock, D. & Pettersson, K. (2006) Exploratory analysis of spatial aspects on the Swedish influenza data. *Smittskyddsinstitutets rapportserie.* Stockholm, Report from the Swedish Institute for Infectious Disease Control.
[21] Mccullagh, P. & Nelder, J.A., 1989, *Generalized Linear Models,* (London: Chapman and Hall).
[22] Brown, L.D., 1986, *Fundamentals of Statistical Exponential Families,* (Hayward, Calif . IMS).
[23] Silvapulle, M. & Sen, P.K., 2005, *Constrained statistical inference. Inequality, order and shape restriction,* (Wiley).
[24] Dahlbom, U. (1986) Some Properties of Estimates of Unimodal Regression. Licentiate thesis. Department of Statistics, Göteborg University.
[25] Woodroofe, M. & Sun, J.Y., 1993, A Penalized Maximum-Likelihood Estimate of f(0+) When f Is Non-Increasing. *Statistica Sinica,* **3,** 501-515.
[26] Burda, M. & Wyplosz, C., 2005, *Macroeconomics. A European Text,* (New York: Oxford University Press).
[27] Ganestam, F., Lundborg, C.S., Grabowska, K., Cars, O. & Linde, A., 2003, Weekly antibiotic prescribing and influenza activity in Sweden: a study throughout five influenza seasons. *Scandinavian Journal of Infectious Diseases,* **35,** 836-842.
[28] Le Strat, Y. & Carrat, F., 1999, Monitoring epidemiologic surveillance data using hidden Markov models. *Statistics in Medicine,* **18,** 3463-3478.
[29] Rath, T.M., Carreras, M. & Sebastiani, P., 2003, (Ed.) *Automated Detection of Influenza Epidemics with Hidden Markov Models,* (Berlin, Germany Springer-Verlag).

[30] Held, L., Höhle, M. & Hofmann, M., 2005, A statistical framework for the analysis of multivariate infectious disease surveillance data. *Statistical Modelling,* **5,** 187-199.
[31] Sebastiani, P., Mandl, K.D., Szolovits, P., Kohane, I.S. & Ramoni, M.F., 2006, A Bayesian dynamic model for influenza surveillance. *Statistics in Medicine,* **25,** 1803-1816.
[32] Linde, A., Brytting, M., Petersson, P., Mittelholzer, C., Penttinen, P. & Ekdahl, K. (2002) Annual Report September 2001 - August 2002: The National Influenza Reference Center. Swedish Institute for Infectious Disease Control.
[33] Linde, A., Brytting, M., Johansson, M., Wiman, Å., Högberg, L. & Ekdahl, K. (2004) Annual Report September 2003 - August 2004: The National Influenza Reference Center. Stockholm, Swedish Institute for Infectious Disease Control.

Paper 2

Semiparametric estimation of outbreak regression

M. FRISÉN, E. ANDERSSON AND K. PETTERSSON
Statistical Research Unit, Department of Economics, Göteborg University
P.O: Box 640, SE 40530 Göteborg, Sweden

Fax: +46 31 7861274

Marianne.Frisen@Statistics.gu.se +46 31 786 1255

Eva.Andersson@Statistics.gu.se +46 31 786 1264

Kjell.Pettersson@Statistics.gu.se +46 31 786 1284

SUMMARY.

A regression may be constant for small values of the independent variable (for example time), but then a monotonic increase starts. Such an "outbreak" regression is of interest for example in the study of the outbreak of an epidemic disease. We give the least square estimators for this outbreak regression without assumption of a parametric regression function. It is shown that the least squares estimators are also the maximum likelihood estimators for distributions in the regular exponential family such as the Gaussian or Poisson distribution. The consistency of the estimators is proved. The approach is thus semiparametric. The method is applied to Swedish data on influenza, and the properties are demonstrated by a simulation study.

KEYWORDS: Constant Base-line, Monotonic change, Exponential family.

1. INTRODUCTION

One important aim in public health surveillance is to detect disease outbreaks. An outbreak can be characterised as a change from a constant level to a monotonically increasing incidence. This is an important part of surveillance for bioterrorism as well as of surveillance for the detection of new diseases such as the recent SARS and avian flu. Outbreaks are also important in the study of ordinary influenza. For likelihood-based surveillance methods ([1], [2]) maximum likelihood estimates are needed. Such estimators will be given in this article. However, the article will not deal with the sequential issues of on-line detection.

In many applications the "normal" or base-line state can be described by a constant level. At a possibly unknown time, the process changes to a monotonically increasing (or decreasing) regression. It is the case of a monotonically increasing regression following the change point that will be treated here, but the statistical problem is the same for a decreasing regression. This "outbreak" regression is of interest not only at the outbreak of an epidemic disease. We have a similar statistical problem when investigating whether data deviate from a specified econometric model by analysing whether there is a change point after which the residuals are increasing.

Often a parametric regression is used to estimate an outbreak. In many cases, however, for example at the outbreak of influenza, the parameters would vary from case (year) to case. The character of the outbreak also varies from one period to the next, thus making it difficult to use a parametric model without misspecification. In [3] and [4] it is concluded that parametric models are not suitable when the parameters vary much from year to year, as they do for influenza data. The importance of avoiding the effects of estimation errors is also discussed in [5]. Thus, here we suggest a nonparametric approach (with respect to the regression function) utilising only the characteristics of a constant start followed by a monotonic increase.

There are several related nonparametric regression problems. Unimodal or "J-shaped" regression is treated in e.g. [6], [7] and [8]. Concave regression is treated for example in [9]. A broken-line estimation is suggested in [10]. Here the parameter, in a distribution belonging to the exponential family, is constant at first, but at an unknown time there is an onset of a positive constant change. The authors point out that also nonlinear regression can be treated by this approach, after a parametric transformation, and they study conditions for consistent estimation of the change-point. They consider the case where the behaviour of the parameter is known after the change, while this paper requires only that the parameter is monotonically changing with time. In [11] and [12] there are discussions on the use of the extra information by monotonicity restriction in connection with smoothing methods. In [13] there is a discussion about the possibility of using an exploratory graphical method for finding jumps. Smoothing methods are very useful for illustrating the outbreak behaviour, but for some purposes, such as alarm systems and hypothesis testing, maximum likelihood estimates are useful.

The aim of this paper is to derive the least squares and maximum likelihood estimators for outbreak regression under monotonicity restrictions. We study both the case of a known and an unknown change point. The normal distribution and the Poisson distribution are of special interest but other members of the exponential family are also considered. The estimator is semiparametric in the sense that the regression function is

nonparametric while the distributions used for the maximum likelihood estimators are parametric.

In Section 2 the model is specified and notations are given. In Section 3 the least squares estimators are derived. In Section 4 the method is illustrated by an example. Consistency is discussed in Section 5. Maximum likelihood properties are given in Section 6. The properties are demonstrated by a simulation study in Section 7. In Section 8 some concluding remarks are given.

2. MODELS AND SPECIFICATIONS

We observe the process X and at time t we have m(t) observations $x_1(t)$, $x_2(t)$, ..., $x_{m(t)}(t)$, t= 0, 1, … s. Let τ be the time when the monotonic increase starts. Thus τ is the first time for which the regression function is not constant. The change point τ may be known or unknown. The expected value of $X_i(t)$, for τ=j, is denoted by $\mu^{\tau}(t)$. The superscript is suppressed when obvious. At time τ the expected value μ changes from a constant level to an increasing regression:

$$\mu(0)=...=\mu(\tau\text{-}1) < \mu(\tau) \le ... \le \mu(s). \quad (1)$$

The monotonicity restriction contains two parts

$$\mu(0)=...=\mu(\tau\text{-}1) \quad (1a)$$

and

$$\mu(\tau\text{-}1) < \mu(\tau) \le ... \le \mu(s) \quad (1b)$$

We will pay special interest to the situation when $X_i(t)$ is normally distributed and the situation when $X_i(t)$ follows a Poisson distribution, but some results are relevant to all members of the exponential family.

3. LEAST SQUARES ESTIMATION OF AN OUTBREAK REGRESSION

Least squares estimation with monotonicity restrictions was described for example in [14] and [15]. We need optimisation under two restrictions, (1a) and (1b). We will prove that if we first optimise under (1a) and then optimise the resulting series under (1b), we will get estimators with the desired properties. In a situation with more that 1 observation at a specific time (i.e. m(t)>1), the mean is calculated. The suggested estimator, for a specific value τ, is constructed by first considering condition (1a), which is the base for the computation of a provisional series y(t) where

$$Y^{\tau}(t) = \sum_{j=0}^{\tau-1}\sum_{i=1}^{m(t)}\left(X_j(j)\right) / \sum_{t=0}^{\tau-1} m(t) \text{ for } t<\tau \text{ and } Y^{\tau}(t) = \sum_{i=1}^{m(t)} X_i(t)/m(t) \text{ for } t\ge\tau. \quad (2)$$

The next step is to consider condition (1b):

$$\hat{\mu}^{\tau}(t) = g(t\left|Y^{\tau}(1), Y^{\tau}(2), ..., Y^{\tau}(s)\right.), \quad (3)$$

where the function g(t) is the least squares estimator of the provisional series $Y^{\tau}(t)$ under the monotonicity restriction (1b).

The order in which the two conditions (1a and 1b) are used will matter and only this ordering will result in estimators which satisfy the least squares and maximum likelihood conditions under the combined restriction. The estimator can also be seen as a pool-adjacent-violators algorithm (PAVA) [15] as will be demonstrated below.

Theorem 1: For a fixed number of observations s and a fixed time point τ from which $\mu(t)$ is increasing, the least squares estimator under the order restriction (1) is given by $\hat{\mu}^{\tau}(t)$, given in (3).

Proof: Since the ordering of the observations before τ is irrelevant, we can formulate the problem as having τ-1 observations at time τ-1, and the restriction for this new problem is:

$$\mu(\tau\text{-}1) < \mu(\tau) \leq \ldots \leq \mu(s).$$

This problem is an ordinary monotonic regression and the LS estimator is given by PAVA. See for example Section 2.4.1 of [16].■

Theorem 2: When the change point is unknown, the least square estimator of μ(t) is

$$\hat{\mu}(t) = \hat{\mu}^{1}(t) \qquad (4)$$

Proof: All other restrictions are included in the monotonic restriction. Thus, no other joint estimators could have a smaller value of $\sum_{t=0}^{s}\sum_{i=1}^{m(t)}(x_i(t) - \hat{\mu}^{j}(t))^2 = Q(j)$ than Q(1).■

The estimator $\hat{\mu}^{\tau}(t)$ could work as a weighted least squares estimator by using weights, for example $w(t) = 1/\sigma(t)$ where $\sigma^2(t)$ is the variance of each of these observations.

4. CALCULATIONS OF INFLUENZA INCIDENCES

In order to illustrate the computation of the estimator, we give the details for an example with a few observations. This is the number of laboratory-identified cases of influenza in Sweden during the first weeks of the winter 2003/2004.

There are observations x(0), x(1), … , x(7) at time points t=0, 1,... (in this example m(t)≡1). We calculate the estimates for the cases when τ=3 and when τ=6. For τ=3, it is assumed that μ(0)=μ(1)=μ(2) and μ(2)<μ(3)≤μ(4)≤μ(5)≤μ(6)≤μ(7), and for τ=6 it is assumed that μ(0)=μ(1)=μ(2)=μ(3)=μ(4)=μ(5) and μ(5)<μ(6)≤μ(7) respectively. The data (x), the provisional series (y) and the least squares estimators ($\hat{\mu}$) are given in Table 1.

Table 1. The computations for the restrictions $\tau = 3$ and $\tau = 6$

t	x(t)	$y^3(t)$	$\hat{\mu}^3(t)$	$y^6(t)$	$\hat{\mu}^6(t)$
0	0	0	0	1	1
1	0	0	0	1	1
2	0	0	0	1	1
3	2	2	1	1	1
4	0	0	1	1	1
5	4	4	4	1	1
6	23	23	23	23	23
7	38	38	38	38	38

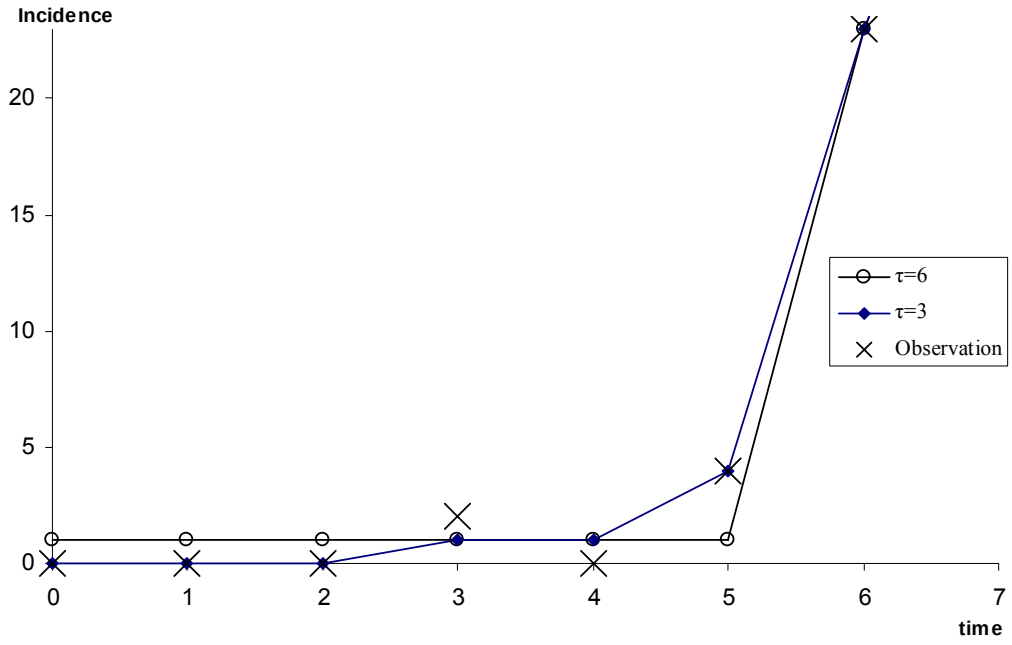

Figure 1: The observed data x and the estimates conditional on the monotonicity restriction $\tau=3$ and $\tau=6$, respectively.

5. CONSISTENCY

When the number of observations m(t) increases for each time t we get consistent estimators for the μ vector.

Theorem 3: If the distribution belongs to the exponential family, then $\hat{\mu}^{\tau}(t)$ will give a consistent estimate of $\mu(t)$ which fulfils condition (1).

Proof: Let $m=\min_t\{m(t)\}$. The estimator will use the averages $\overline{X}(t)=\sum_{i=1}^{m(t)} X_i(t)/m(t)$. Since $\overline{X}(t)$ is a strongly consistent estimator of the expected value in the exponential family, so is $\hat{\mu}^{\tau}(t)$, since only averaging and PAVA are used in the transformations of $\overline{x}(t)$. It follows that, with probability one,

$$\lim_{m\to\infty}\max_t |Y^{\tau}(t)-\mu(t)|=0.$$

Thus, with probability one $\hat{\mu}(t)$ satisfies the condition (1) as m goes to infinity. $\hat{\mu}(t)$.■

Unfortunately this consistency does not carry over to the case where there is only one observation for each time but the number of time points increases. For the case when we have a pre-grouping of the time points into classes, the consistency property carries over to the expected values in these time-classes if the number of observations in each time-class increases.

6. MAXIMUM LIKELIHOOD ESTIMATION

For certain distributions the least squares estimators given above are also maximum likelihood estimators. We will consider the regular exponential family with the conditions of the derivatives of the parameters as specified on page 34 of [15] and give special cases of this family.

Theorem 4: The least squares solutions of Theorem 1 and 2 are the maximum likelihood solutions if the values of the dispersion parameter are equal for all times (but possibly unknown).

Proof: This follows from properties of ordinary isotonic regression since the current problem can be expressed in these terms, as demonstrated in the proof of Theorem 1. See for example Section 2.4.2 of [16].■

Theorem 5: The weighted least squares estimator is the maximum likelihood solution for known but possibly different values of the dispersion parameter.

Proof: This follows from properties of ordinary isotonic regression. See for example Section 2.4.2 of [16].■

Corollary 1: The least squares solutions $\hat{\mu}^{\tau}(t)$ and $\hat{\mu}(t)$ in (3) and (4) are the maximum likelihood solutions when the observations at each time follow a normal distribution with equal variances.

Corollary 2: The weighted least squares solutions $\hat{\mu}^{\tau}(t)$ and $\hat{\mu}(t)$ in (3) and (4) are the maximum likelihood solutions when the observations at each time follow a normal distribution with unequal variances, where the variances are known (or their relation to each other is known).

Corollary 3: The (unweighted) least squares solutions $\hat{\mu}^{\tau}(t)$ and $\hat{\mu}(t)$ in (3) and (4) are the maximum likelihood solutions for a Poisson distribution.

This follows from the fact that there is no additional dispersion parameter for the Poisson distribution. One might have expected that weights should be used since the parameter of the Poisson distribution also reflects the variance. However, the only places where the regression differs from the observations are where the estimates by the PAVA are constants. A weighted regression should thus have constant weight.■

The estimated curve (and the corresponding likelihood) may be used for inference such as hypothesis testing or surveillance concerning the start of the influenza season, but such inference will not be treated here.

7. SIMULATION STUDY OF PROPERTIES

We generated data similar to those that can be expected at an influenza outbreak [3] in order to illustrate bias, variance and the influence of the value of τ when the increase starts. In Sweden the monitoring of influenza starts at week 40 each year but the time of the onset varies considerably between years. Thus, also the waiting time until the onset (τ-1 weeks) varies considerably between years, and we investigated several possible scenarios. The reported results are based on at least 1 000 000 replicates. We report results for Poisson and normally distributed variables.

To illustrate the case for a Poisson distribution we generated weekly numbers of laboratory-diagnosed influenza cases (LDI) according to their similarity with the influenza season 2003/2004, which was a "typical" season. The observed process X follows a Poisson distribution with the parameter $\mu(t)$, where

$$\mu(t) = \begin{cases} \mu_0, & t < \tau \\ \exp(\beta_0 + \beta_1 \cdot (t - \tau + 1)), & t \geq \tau \end{cases}$$

where μ_0=1, β_0=-0.26, β_1=0.826.

In Figure 2 the mean and standard deviation (by 2 SD bars) of the estimates of 1 000 000 replicates are given. The cases are generated for different values of τ (τ=4 and τ=8). The estimates were produced with knowledge of the true value. The variation of the estimates is smaller than without the restriction, thus $\mathrm{Var}(\hat{\mu}^{\tau}(t)) < \mathrm{Var}(X(t))$. The effect of the restriction of a constant phase has a major influence on $\mathrm{Var}(\hat{\mu}^{\tau}(t))$ during this phase, and this variance is smaller than the variance for the mean of all the observations during the constant phase. The monotonicity restriction has a small variance-reducing effect when the slope is large in comparison with the variance.

There is a bias, but this is too small to be seen in the scale of Figure 2. Thus the two series (the mean of the estimate and the expected value of the generated data) coincide in Figure 2. The bias ($E[\hat{\mu}] - \mu$) is illustrated in a larger scale in Figure 3.

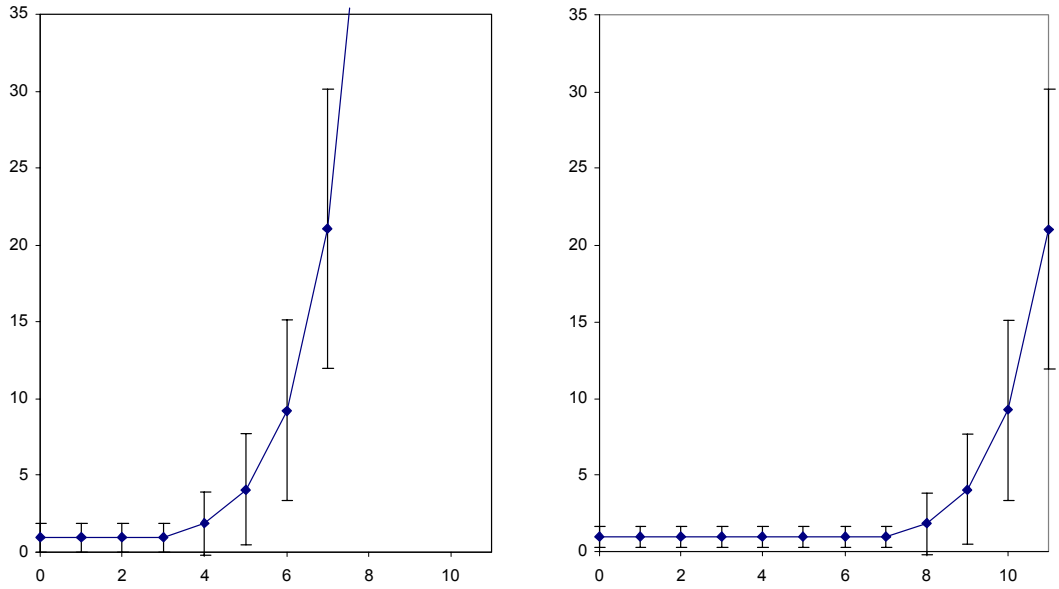

Figure 2. The mean of the estimated values at each time point (dot) and the variation of the estimated values, illustrated by ± 2SD (bars). The true expected value, μ(t), cannot be distinguished from the mean of the estimates, E[$\hat{\mu}$ (t)], in the scale of the figure. The left figure is estimated under the true restriction of τ=4, and the right figure is estimated under the true restriction that τ=8.

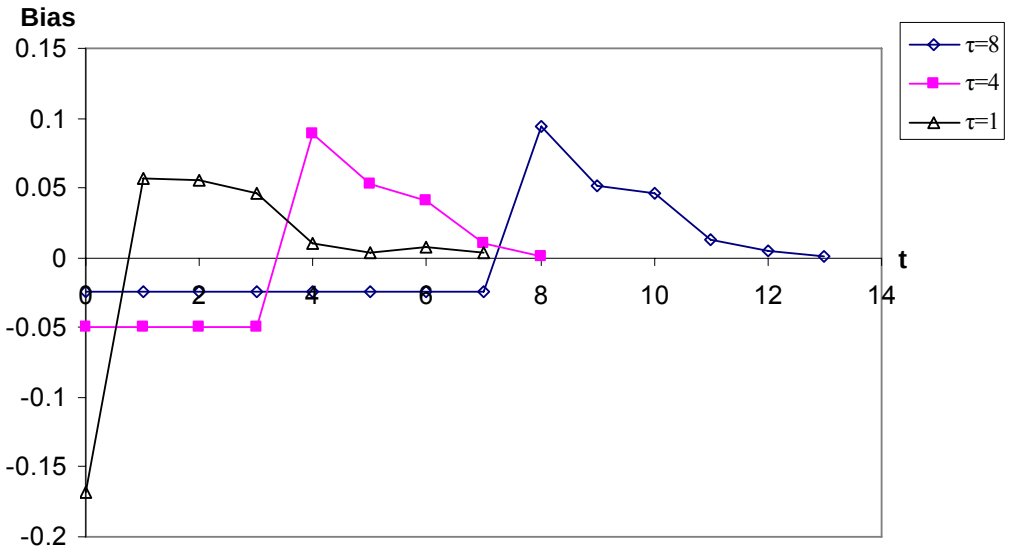

Figure 3. The bias $E[\hat{\mu}]-\mu$ for the situations when data are generated for τ=1 τ=4 and τ=8 and the true value of τ is known in the estimation.

As could be expected the bias in the constant phase is small since the first step of forming the mean (provisional series) produces an unbiased estimate. In the next step the isotonic regression will produce a too low estimate of the constant phase. The weight of the unbiased estimate is (τ-1)/s, thus the bias will be small for a large value of τ. For the next part of the regression the bias is as expected for an isotonic curve; namely, there is a

negative bias for early time points and a positive bias for late ones. This is illustrated in Figure 4 where a constant value is generated and the estimation is made under the restriction of $\tau=1$.

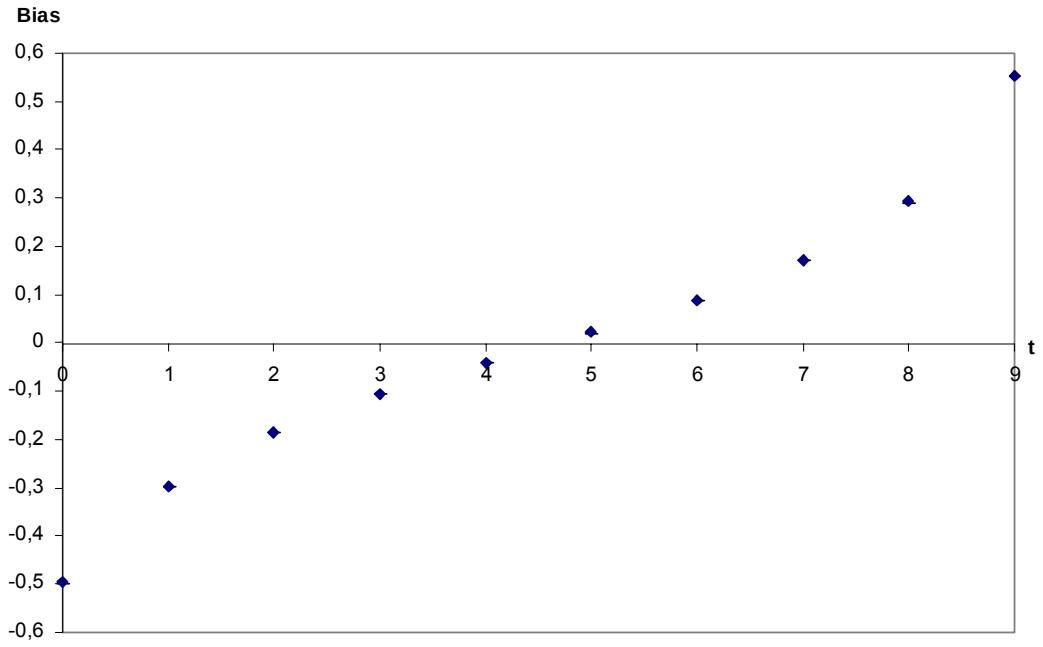

Figure 4. Estimation under the restriction of $\tau=1$ for data generated by a Poisson distribution with a constant mean (i.e. generated under the condition that $\tau=\infty$).

The pattern in Figure 3 will not completely agree with the one in Figure 4 even at the isotonic phase, since we have an exponential increase as soon as the influenza has started. Thus, we will very soon have very little influence of the isotonic regression. The later points will almost always be estimated by the observed values, and the bias will thus decrease to zero.

The effect of misspecification of τ is illustrated in Figure 5. Both curves ($\tau=4$ and $\tau=8$) from which data are generated are the same as those in Figure 2.

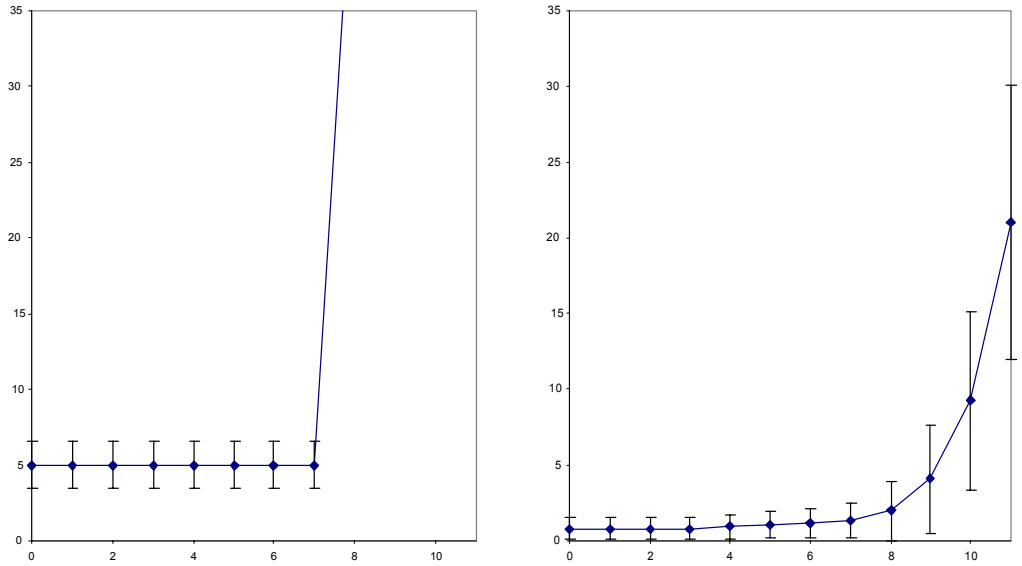

Figure 5. The mean of the estimated values at each time point (dot) and the variation of the estimated values, illustrated by ± 2SD (bars). The effect of error in the assumption of τ is illustrated: in the left figure, the true τ equals 4 but the restriction τ=8 is imposed in the estimation, and in the right figure, the true τ equals 8 but the restriction τ=4 is imposed in the estimation.

In Figure 5 we can see that a restriction of a later change than the true one will give a constant phase at a too high level. In figure b we can see that a restriction of an earlier change than the true one has very little impact. In Figure 6 we illustrate the bias and the standard deviation when no assumption of the value of τ is made but the general maximum likelihood estimator $\hat{\mu}(t)$ is used.

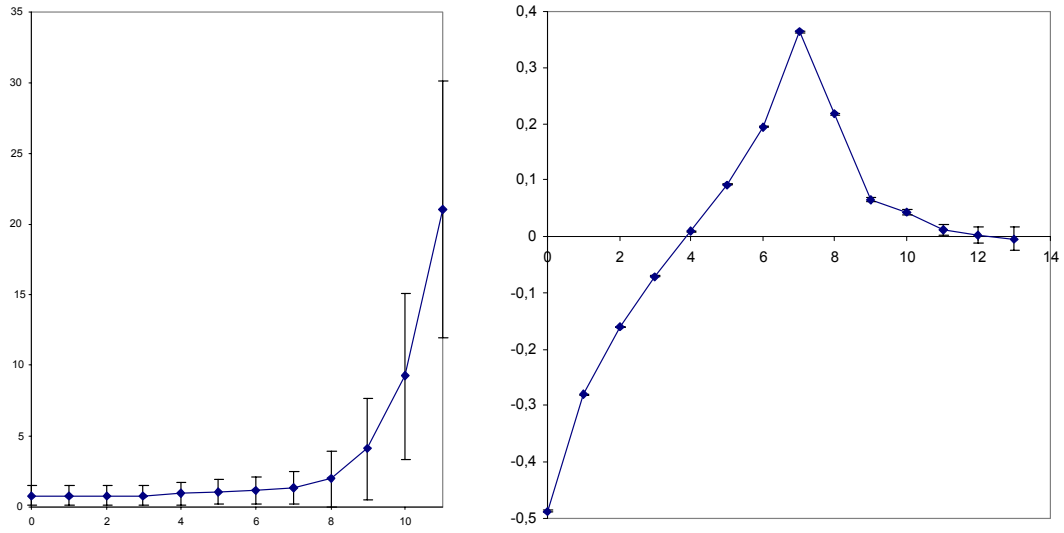

Figure 6. The maximum likelihood estimator, without any information about τ (true τ equals 8). Left: The mean of the estimated values at each time point (dot) and the variation of the estimated values, illustrated by ± 2SD (bars). The true expected value, μ(t), cannot be distinguished from the mean of the estimates, $E[\hat{\mu}(t)]$, in the scale of the figure. Right: The bias at each time point (dot).

In Figure 6 (left) we see that the mean of the estimated curve, even without information on τ, is very close to the real curve in the current scale. Thus, even without knowledge of τ, the estimator produces a reasonable estimate. By comparing the bias in the right panel of Figure 6 with the one in Figure 3 (for τ=8), we can conclude that the knowledge of the value of τ decreases the bias – especially for the constant phase. By comparing the variation of the estimates (±2SD) in Figure 6 (left) with that of Figure 2, we can see that the correct restriction (knowledge about τ) decreases the variation – especially during the constant phase.

For the Poisson distribution, the variance and the expected value are related. In order to examine the effect of variance we generated normally distributed data with constant variance. To illustrate the properties for normal distributions with different variances we generated data with means similar to the number of influenza-like cases (ILI) during the winter 2003/2004. The following model was used for the observed process X:

$X(t) \sim N(\mu(t); \sigma^2)$,

where

$$\mu(t) = \begin{cases} \mu_0, & t < \tau \\ \exp(\beta_0 + \beta_1 \cdot (t - \tau + 1)), & t \geq \tau \end{cases}$$

and μ_0=20, β_0=2.67, β_1=0.68 and different values of the variance σ^2 are used. A normal distribution is a reasonable approximation here since the incidences are rather high. Different scenarios were considered regarding the length of the constant phase.

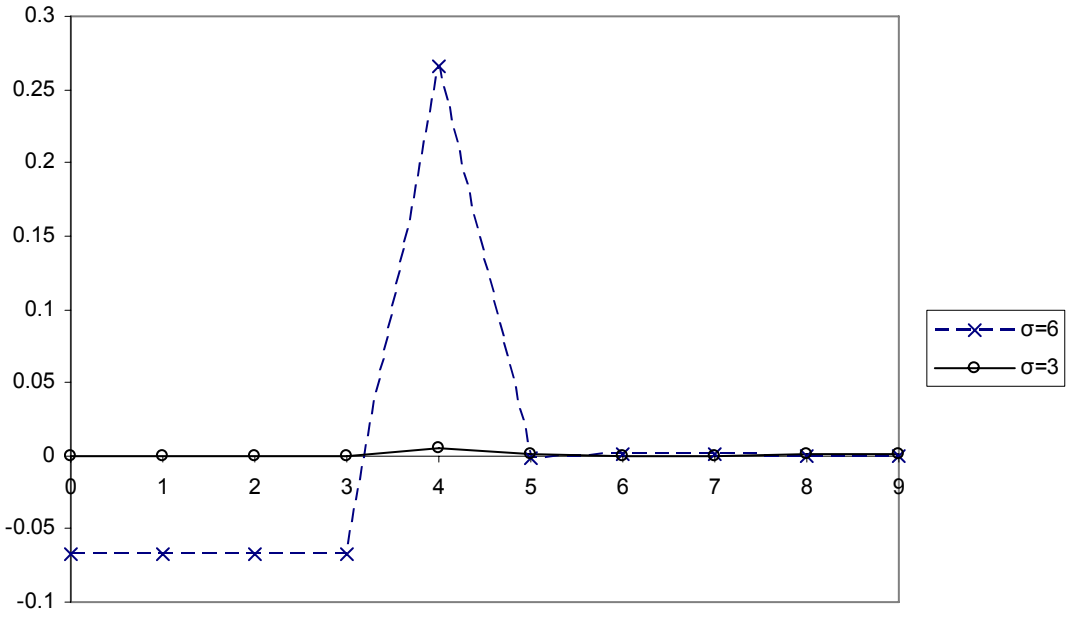

Figure 7. Average bias of the ML-estimator for normally distributed data with standard deviation σ=3 and σ=6 respectively. The data are generated for τ=4 and this is used as a restriction in the estimation.

In Figure 7 we can see that the bias is small for a small variance. This illustrates that the estimator is consistent.

8. CONCLUDING REMARKS

The results presented here on outbreak regression are of importance not only in pure estimation contexts but also for on-line surveillance based on maximum likelihood estimators. The outbreak of a disease can often be characterised as a change from a constant level to a monotonically increasing incidence. Surveillance systems for detecting outbreaks are crucial in surveillance for bioterrorism as well as in surveillance for the detection of new diseases, see [17]. Outbreak detection is also important in the study of ordinary influenza [18]. Surveillance systems based on likelihood ratios have important optimality properties [2]. For nonparametric surveillance as in [4] (nonparametric with respect to the shape of the curve), maximum likelihood estimates are useful as a basis for maximum likelihood ratios. The estimators presented here can be used for likelihood-based surveillance to detect the onset of an increasing incidence. Smoothing methods are useful for the description of the outbreak behaviour but will not give the required maximum likelihood estimators.

Sometimes it is reasonable to believe that the regression is continuous and has continuous derivatives. However, this condition can always be satisfied by some definition of estimates between the discretely observed times. Thus this is no restriction to the estimates. When a smooth curve is needed for illustration, it is possible to fit a smooth curve (such as a spline [19]) to the maximum likelihood estimates.

One may be interested in estimating the time, τ, of the onset of the increasing phase and also the level of the constant phase, $\mu(0)$. The maximum likelihood estimation of the curve by the proposed method will also give maximum likelihood estimates of these parameters. Generally, however, there will not be one unique maximum likelihood estimator of τ. No other value of τ can give a larger value of the likelihood than $\tau=1$, since $\mu(0)= \ldots =\mu(i-1)<\mu(i)\leq \ldots\leq\mu(s)$ is a special case of, or on the limit of, $\mu(0)<\mu(1)\leq\ldots\mu(i)\ldots\leq\mu(s)$. The maximum likelihood estimator of $\mu(0)$ will be unique but biased since the maximum likelihood estimators of τ and $\mu(0)$ are closely related. This problem of bias in the endpoints is shared with other maximum likelihood estimators of ordered statistics such as the usual monotonic regression. In order to get unbiased estimators of τ and $\mu(0)$, more (parametric) structure is needed, for example a certain size of the change. Note that when the maximum likelihood statistic derived here is used for test or surveillance purposes, the bias is not a problem. In such cases there are natural false alarm requirements which give the user the opportunity to state that only important deviations should be detected. This corresponds to the above-mentioned parametric size condition for the estimator but is expressed by probability and does not require any parametric assumption for the curve.

The estimator is consistent (for a large number of observations at each time) but not unbiased. The direction of the bias is that the estimates are too low. However, the bias is very small for the constant phase. For the increasing phase the bias is smaller to start with due to the stabilisation by the constant phase. A long constant phase exaggerates this tendency.

9. REFERENCES

[1] Andersson, E., 2002, Monitoring cyclical processes - a nonparametric approach. *Journal of Applied Statistics,* **29,** 973-990.

[2] Frisén, M., 2003, Statistical surveillance. Optimality and methods. *International Statistical Review,* **71,** 403-434.

[3] Andersson, E., Bock, D. & Frisén, M., 2007, Modeling influenza incidence for the purpose of on-line monitoring. *Statistical Methods in Medical Research,* (in press).

[4] Bock, D., Andersson, E. & Frisén, M., 2007, Statistical surveillance of epidemics: Peak detection of influenza in Sweden. *Biometrical Journal,* (in press).

[5] Ion, R.A. & Klaassen, C.A.J., 2005, Non-parametric Shewhart control charts. *Journal of Nonparametric Statistics,* **17,** 971-988.

[6] Frisén, M., 1986, Unimodal regression. *The Statistician,* **35,** 479-485.

[7] Andersson, L. & Frisén, M., 2002, Verifications of turning points. *Journal of Nonparametric Statistics,* **14,** 623-645.

[8] Andersson, E., Bock, D. & Frisén, M., 2005, Statistical surveillance of cyclical processes. Detection of turning points in business cycles. *Journal of Forecasting,* **24,** 465-490.

[9] Hildreth, C., 1954, Point estimation of ordinates of concave functions. *Journal of the American Statistical Association,* **49,** 598-619.

[10] Gill, R. & Baron, M., 2004, Consistent estimation in generalized broken-line regression. *Journal of Statistical Planning and Inference,* **126,** 460.

[11] Efromovich, S., 2001, Density estimation under random censorship and order restrictions: From asymptotic to small samples. *Journal of the American Statistical Association,* **96,** 667-684.

[12] Frisén, M., 2007, Discussion on "Sequential design and estimation in heteroscedastic nonparametric regression" by Sam Efromovich. *Sequential Analysis,* **26,** 45-47.

[13] Kim, C.S. & Marron, J.S., 2006, SiZer for jump detection. *Journal of Nonparametric Statistics,* **18,** 13 - 20.

[14] Barlow, R.E., Bartholomew, D.J., Bremer, J.M. & Brunk, H.D., 1972, *Statistical inference under order restrictions,* (London: Wiley).

[15] Robertson, T., Wright, F.T. & Dykstra, R.L., 1988, *Order restricted statistical inference,* (Chichester: John Wiley & Sons Ltd).

[16] Silvapulle, M. & Sen, P.K., 2005, *Constrained statistical inference. Inequality, order and shape restriction,* (Wiley).

[17] Sonesson, C. & Bock, D., 2003, A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society A,* **166,** 5-21.

[18] Andersson, E., Kuhlmann-Berenzon, S., Linde, A., Schiöler, L., Rubinova, S. & Frisén, M., 2008, Predictions by early indicators of the time and height of yearly influenza outbreaks in Sweden. *Scandinavian Journal of Public Health***,** to appear.

[19] Silverman, B.W., 1985, Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society B,* **47,** 1-52.

Research Report

2007:1	Andersson, E.:	Effect of dependency in systems for multivariate surveillance.
2007:2	Frisén, M.:	Optimal Sequential Surveillance for Finance, Public Health and other areas.
2007:3	Bock, D.:	Consequences of using the probability of a false alarm as the false alarm measure.
2007:4	Frisén, M.:	Principles for Multivariate Surveillance.
2007:5	Andersson, E., Bock, D. & Frisén, M.:	Modeling influenza incidence for the purpose of on-line monitoring.
2007:6	Bock, D., Andersson, E. & Frisén, M.:	Statistical Surveillance of Epidemics: Peak Detection of Influenza in Sweden.
2007:7	Andersson, E., Kühlmann-Berenzon, S., Linde, A., Schiöler, L., Rubinova, S. & Frisén, M.:	Predictions by early indicators of the time and height of yearly influenza outbreaks in Sweden.
2007:8	Bock, D., Andersson, E. & Frisén, M.:	Similarities and differences between statistical surveillance and certain decision rules in finance.
2007:9	Bock, D.:	Evaluations of likelihood based surveillance of volatility.
2007:10	Bock, D. & Pettersson, K.	Explorative analysis of spatial aspects on the Swedish influenza data.
2007:11	Frisén, M. & Andersson, E.	Semiparametric surveillance of outbreaks.
2007:12	Frisén, M., Andersson, E. & Schiöler, L.	Robust outbreak surveillance of epidemics in Sweden.
2007:13	Frisén, M., Andersson, E. & Pettersson, K.	Semiparametric estimation of outbreak regression.
2007:14	Pettersson, K.	Unimodal regression in the two-parameter exponential family with constant or known dispersion parameter