Ann. Inst. Statist. Math.
Vol. 48, No. 2, 215–228 (1996)

A STATE-SPACE APPROACH TO POLYGONAL LINE REGRESSION

NOBUHISA KASHIWAGI

The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

(Received August 16, 1993; revised October 9, 1995)

Abstract. A non-Gaussian state-space model is proposed to estimate a switching trend from serial data taken at equally spaced intervals. A procedure to detect structural changes in a linear trend is also proposed. The results of a simulation study conducted to check the performance of the detection procedure are shown. A numerical illustration is provided using economic time series data.

Key words and phrases: Akaike information criterion, edge detection, non-Gaussian state-space model, smoothing, structural change, switching trend.

1. Introduction

We consider the problem of fitting polygonal line regression (PLR) to serial data taken at equally spaced intervals. PLR enables us to estimate a switching linear trend. Therefore, PLR is useful for the trend analysis of data predicted to involve structural changes in a linear trend. Structural changes in a linear trend are often discussed in the analysis of econometric time series (see, for example, Tsurumi (1988)). Another application of PLR can be found in one-dimensional edge detection in image analysis. In this paper, we suggest a non-Gaussian state-space approach for PLR.

The most traditional way of estimating a switching linear trend may be to detect structural changes in a linear model. Actually, in the single change case, this way has attracted much attention since Quandt (1958), and has been studied in detail as two-phase regression (see, for example, Broemeling and Tsurumi (1986)). However, when the number of structural changes is unknown, this approach sometimes faces statistical and computational difficulties, which are mainly caused by the fact that the detection of an unknown number of structural changes in a linear model is basically a problem of non-nested model selection, and that the number of alternative models, that is, the number of combinations of possible change points increases exponentially with the size of data. To decrease such difficulties, Kashiwagi (1991) suggested to make inference about structural changes based on the posterior probabilities of possible change points, and proposed an approximation procedure to calculate the posterior probabilities. However, even his procedure becomes infeasible, especially when given data involve many structural changes.

The state-space approach offers an attractive way to avoid the difficulties of the traditional linear model approach. Actually, the recent innovation of the state-space approach has shown that dynamics of structural changes can be modeled explicitly and recursive formulas are available to identify the model parameters and the model itself. For example, Harrison and Stevens (1976) proposed multi-process models associated with dynamic linear models based on the state-space representation to handle model uncertainty including sudden changes in an underlying model, and they gave Kalman-filter-like recursive formulas for forecasting. This work was extended to those of monitoring biomedical time series (Gordon and Smith (1990)) and of tracking multiple targets (Shumway and Stoffer (1991)), for example. However, these works are not concerned with smoothing. In the case of on-line analysis, smoothing is not always necessary. But, in retrospective analysis, it is desirable to consider smoothing. A state-space approach including smoothing which can handle structural changes can be seen in Kitagawa (1987), who discussed non-Gaussian modeling of nonstationary time series and suggested to execute the recursive formulas numerically to ensure the feasibility.

The purpose of this paper is to propose, following Kitagawa (1987), a non-Gaussian state-space model which can achieve the purpose of PLR approximately, even when given data involve many structural changes in a linear trend. To avoid the combinatorial problem of possible change points, we seek to provide a switching linear trend by assuming a special non-Gaussian distribution for the system noise instead of detecting structural changes in a linear model directly. However, as a natural consequence of the state-space approach, the estimated trend is only a predictor that is induced from a posterior density, and therefore, it is not always possible to detect structural changes, especially delicate ones by observing the estimated trend. Accordingly, we also propose a procedure based on the AIC (Akaike (1973)) to detect structural changes statistically.

In Section 2, a non-Gaussian state-space model for PLR is proposed. In Section 3, an estimation procedure is presented, and some remarks for the numerical computation are given. In Section 4, a procedure to detect structural changes in a linear trend is proposed, and the results of a simulation study conducted to check the performance of the proposed procedure are shown. In Section 5, examples of application are provided.

2. Proposed model

In this section, we propose a non-Gaussian state-space model for PLR. Let $y_1, \ldots, y_n$ be a sequence of observations taken at equally spaced intervals. The proposed model is:

$$\begin{aligned} y_t &= \mu_t + v_t \\ \mu_t &= 2\mu_{t-1} - \mu_{t-2} + w_t \\ &\quad t = 1, \ldots, n \end{aligned}$$

where μ_t is a trend component, v_t and w_t are noise components, and it is assumed that $v_t \sim$ i.i.d.N(0,σ_v^2). If a Gaussian distribution is assumed also for w_t, this model becomes a standard one (see, for example, Akaike (1980) and Wecker and Ansley (1983)). Instead of a Gaussian distribution, we assume a non-Gaussian distribution for w_t, which is derived by using the property of the second difference on local linearity such that the following proposition holds only for $k = 2$.

$$\Delta^k \mu_t = 0 \Leftrightarrow \text{the consecutive } k+1 \text{ components } \mu_t, \ldots, \mu_{t-k} \text{ are on a line}$$

where Δ is the difference operator. This property suggests that, if we set a constraint of increasing the possibility that the second difference $\mu_t - 2\mu_{t-1} + \mu_{t-2}$ is equal to zero, the three components μ_t, μ_{t-1} and μ_{t-2} will show a tendency to lie on a line. On the other hand, when the three components are not on a line in practice, to achieve the purpose of PLR, it is not always necessary to bring the value of the second difference close to zero, for example, by assuming a Gaussian distribution for w_t as attempted in standard smoothing methods. Accordingly, we assume a mixture distribution having the following density for w_t.

$$p(w_t \mid \alpha_t) = \alpha_t \delta(w_t) + (1 - \alpha_t) f(w_t)$$

where α_t is a mixing proportion, $\delta(w_t)$ is Dirac's delta, and $f(w_t)$ is a density of a certain distribution. In the estimation of trend, it is assumed that $\alpha_t \equiv \alpha$. For $f(w_t)$, most of existing distributions are acceptable, and in this paper, we employ the following symmetric continuous distributions as representative examples.

Gaussian distribution (G)

$$f(w_t \mid \sigma_w^2) = 1/(\sqrt{2\pi}\sigma_w) \exp\{-w_t^2/(2\sigma_w^2)\}.$$

Mixture of Γ distribution (Γ)

$$f(w_t \mid a, b) = |w_t|^{a-1} \exp\{-|w_t|/b\}/\{2b^a \Gamma(a)\}, \qquad a \geq 1, \;\; b > 0.$$

Mixture of uniform distribution (U)

$$f(w_t \mid g, h) = \begin{cases} 1/\{2(h-g)\} & w_t \in [-h, -g] \cup [g, h], \;\; 0 \leq g < h \\ 0 & \text{otherwise.} \end{cases}$$

Corresponding to these distributions, we denote the proposed model by $H_{G,0}$, $H_{\Gamma,0}$ and $H_{U,0}$, respectively. Model $H_{G,0}$ contains a standard smoothing model as a special case, $H_{\Gamma,0}$ involves a prior which can have non-zero peaks, and $H_{U,0}$ corresponds to a kind of flat prior model. The unknown parameters involved in these models, $\boldsymbol{\theta}_G = (\sigma_v^2, \alpha, \sigma_w^2)$, $\boldsymbol{\theta}_\Gamma = (\sigma_v^2, \alpha, a, b)$ and $\boldsymbol{\theta}_U = (\sigma_v^2, \alpha, g, h)$, are estimated using likelihood, and the best fit model among the three is selected by the AIC. These will be discussed in the next section, where the observation and system models in the proposed model are denoted in the density form as $p(y_t \mid \mu_t)$ and $p(\mu_t \mid \mu_{t-1}, \mu_{t-2})$, respectively.

In the state-space approach, the treatment of the initial state, (μ_0, μ_{-1}), sometimes becomes a subject of study. Actually, several authors have discussed this from various viewpoints (for example, Kohn and Ansley (1987) and Kashiwagi and Yanagimoto (1992)). However, we assume a conventional distribution for the initial state as $p(\mu_0, \mu_{-1}) \propto 1$, because an exact treatment of the initial state requires an inordinate amount of computation, and because the main effect of this assumption is limited in many cases to the first part of the series.

3. Estimation procedure

In this section, we present an estimation procedure for the proposed model. The procedure consists of recursive formulas in the state-space approach and of the AIC.

To define the likelihood of the model, we first present the prediction and filtering formulas, which are given, respectively, as

$$p(\mu_t, \mu_{t-1} \mid \boldsymbol{y}_{t-1}) = \int_{R(\mu_{t-2})} p(\mu_t \mid \mu_{t-1}, \mu_{t-2}) p(\mu_{t-1}, \mu_{t-2} \mid \boldsymbol{y}_{t-1}) d\mu_{t-2}$$

$$p(\mu_t, \mu_{t-1} \mid \boldsymbol{y}_t) = p(y_t \mid \mu_t) p(\mu_t, \mu_{t-1} \mid \boldsymbol{y}_{t-1}) / p(y_t \mid \boldsymbol{y}_{t-1})$$

where $\boldsymbol{y}_t = (y_1, \ldots, y_t)'$, $R(\mu)$ is the support of μ and

$$p(y_t \mid \boldsymbol{y}_{t-1}) = \int_{R(\mu_t, \mu_{t-1})} p(y_t \mid \mu_t) p(\mu_t, \mu_{t-1} \mid \boldsymbol{y}_{t-1}) d\mu_t d\mu_{t-1}.$$

The initial condition in these formulas is $p(\mu_0, \mu_{-1} \mid \boldsymbol{y}_0) = p(\mu_0, \mu_{-1})$ and the suffix t runs from 1 to n. The likelihood of the model is defined by

$$L_{D,0}(\boldsymbol{\theta}_D \mid \boldsymbol{y}_n) = \prod_{t=1}^{n} p(y_t \mid \boldsymbol{y}_{t-1})$$

where $D \in (G, \Gamma, U)$. The estimates $\hat{\boldsymbol{\theta}}_D$ of $\boldsymbol{\theta}_D$ can be obtained by maximizing the likelihood. The best fit model among $H_{G,0}$, $H_{\Gamma,0}$ and $H_{U,0}$ can be selected by using the AIC. The AIC's for these models are defined by

$$\begin{aligned} \mathrm{AIC}_{G,0} &= -2 \times \log L_{G,0}(\hat{\boldsymbol{\theta}}_G \mid \boldsymbol{y}_n) + 6 \\ \mathrm{AIC}_{\Gamma,0} &= -2 \times \log L_{\Gamma,0}(\hat{\boldsymbol{\theta}}_\Gamma \mid \boldsymbol{y}_n) + 8 \\ \mathrm{AIC}_{U,0} &= -2 \times \log L_{U,0}(\hat{\boldsymbol{\theta}}_U \mid \boldsymbol{y}_n) + 8. \end{aligned}$$

We may select the model which gives the minimum of AIC. The selected model is denoted by $H_{\hat{D},0}$.

On the other hand, we estimate the trend component μ_t by the smoothing density $p(\mu_t, \mu_{t-1} \mid \boldsymbol{y}_n)$, which can be obtained by using the following recursive formula.

$$\begin{aligned} &p(\mu_t, \mu_{t-1} \mid \boldsymbol{y}_n) \\ &\quad = p(\mu_t, \mu_{t-1} \mid \boldsymbol{y}_t) \\ &\quad\quad \cdot \int_{R(\mu_{t+1})} p(\mu_{t+1} \mid \mu_t, \mu_{t-1}) p(\mu_{t+1}, \mu_t \mid \boldsymbol{y}_n) / p(\mu_{t+1}, \mu_t \mid \boldsymbol{y}_t) d\mu_{t+1} \end{aligned}$$

where the suffix t runs from $n-1$ to 1.

The above procedure can be implemented by using standard numerical methods. An example of the implementation is to approximate each function in the formulas by a piecewise linear function with equally spaced knots, and then to use the trapezoidal rule for integration and a variable metric method (see, for example, Jacoby *et al.* (1972)) to maximize the log likelihood. Although the proposed model involves Dirac's delta, this causes no problem in the implementation, because Dirac's delta appears only as an integrand in the recursive formulas. As the interval of definition of μ_t, one may take a sufficiently long finite interval. According to our experience, it is preferable to take the interval so that, at least, it can cover the interval $[\hat{\mu}_t - 3\hat{h}, \hat{\mu}_t + 3\hat{h}]$, where $\hat{\mu}_t$ is the posterior mean of μ_t with respect to the smoothing density and $\hat{h}$ is the estimate of h in $H_{U,0}$. On the other hand, in maximizing the likelihood, we need to expect the existence of local maxima. A practical solution to this problem may be to use various initial values in the iterative maximization method. Finally note that the computer memory size required in smoothing can be reduced by using the split algorithm in Kashiwagi (1993).

4. Detection of structural changes

4.1 *Proposed procedure*

Except when $\hat{\alpha} = 1$, given data are predicted to involve structural changes in a linear trend. However, it is not always possible to detect structural changes by observing the estimated trend. In this section, we propose a procedure for detecting them statistically.

To detect structural changes in a linear trend, we assume the following hypothesis against $H_{\hat{D},0}$.

$$H_{\hat{D},k} : \alpha_t = \begin{cases} 1 & t = k+1 \\ \alpha & \text{otherwise} \end{cases}$$

where $2 \le k \le n-1$. This hypothesis indicates that the density of the system noise w_{k+1} is given by $p(w_{k+1}) = \delta(w_{k+1})$, suggesting that $w_{k+1} = 0$, that is, the consecutive three components whose center is μ_k are always on a line. Therefore, in the comparison of $H_{\hat{D},k}$ with $H_{\hat{D},0}$, if $H_{\hat{D},k}$ is selected, it can be judged that no structural change in a linear trend has occurred at $t = k$. On the contrary, if $H_{\hat{D},0}$ is selected, it can be judged that there is a possibility that a structural change in a linear trend has occurred at $t = k$, because $H_{\hat{D},0}$ suggests that the system noise w_{k+1} is not always equal to zero but is distributed stochastically. Based on these facts, we detect structural changes in a linear trend by selecting the better one for every k among $H_{\hat{D},0}$ and $H_{\hat{D},k}$. To select it, we use the AIC. The AIC for $H_{D,k}$ is given by

$$\begin{aligned}
\mathrm{AIC}_{G,k} &= -2 \times \log L_{G,k}(\tilde{\boldsymbol{\theta}}_G \mid \boldsymbol{y}_n) + 6 \\
\mathrm{AIC}_{\Gamma,k} &= -2 \times \log L_{\Gamma,k}(\tilde{\boldsymbol{\theta}}_\Gamma \mid \boldsymbol{y}_n) + 8 \\
\mathrm{AIC}_{U,k} &= -2 \times \log L_{U,k}(\tilde{\boldsymbol{\theta}}_U \mid \boldsymbol{y}_n) + 8
\end{aligned}$$

Table 1. Frequency that $H_{U,0}$ was selected in 20 trials.

The first alternative					The second alternative				
	λ					λ			
k	0.125	0.25	0.5	1	k	0.125	0.25	0.5	1
2	0	4	4	9	2	0	0	6	2
3	0	1	1	3	3	0	0	1	3
4	0	0	0	2	4	0	0	2	2
5	0	0	0	0	5	0	0	0	2
6	0	0	0	0	6	0	0	2	1
7	0	0	0	0	7	0	0	1	0
8	0	0	0	0	8	0	0	2	1
9	0	0	0	1	9	0	0	1	1
10	0	0	0	1	10	0	0	1	0
11	0	0	1	1	11	0	0	1	1
12	0	0	1	2	12	0	0	2	2
13	0	1	0	6	13	0	2	5	2
14	0	0	7	13	14	0	1	7	1
15	20	19	18	17	15	20	18	13	3
16	0	1	6	16	16	20	17	9	2
17	0	1	2	6	17	0	1	4	1
18	0	0	1	5	18	0	1	6	3
19	0	0	0	2	19	0	1	5	2
20	0	1	0	1	20	0	0	1	1
21	0	1	0	1	21	0	0	1	1
22	0	1	0	0	22	0	0	1	1
23	0	0	0	0	23	0	0	2	1
24	0	0	0	0	24	0	0	1	0
25	0	0	0	0	25	0	0	1	0
26	0	2	0	2	26	0	0	0	0
27	0	2	0	3	27	0	0	0	2
28	0	0	3	4	28	0	0	1	2
					29	0	0	5	2

where $L_{D,k}(\tilde{\boldsymbol{\theta}}_D \mid \boldsymbol{y}_n)$ is the maximum likelihood for $H_{D,k}$.

The procedure proposed above is a method of testing linearity of consecutive three components individually, but not a method of detecting an unknown number of structural changes in a linear trend simultaneously. However, the simultaneous detection of an unknown number of structural changes causes the combinatorial problem of possible change points. Since the proposed model can follow various types of change of a trend, we think that the proposed procedure is practical enough.

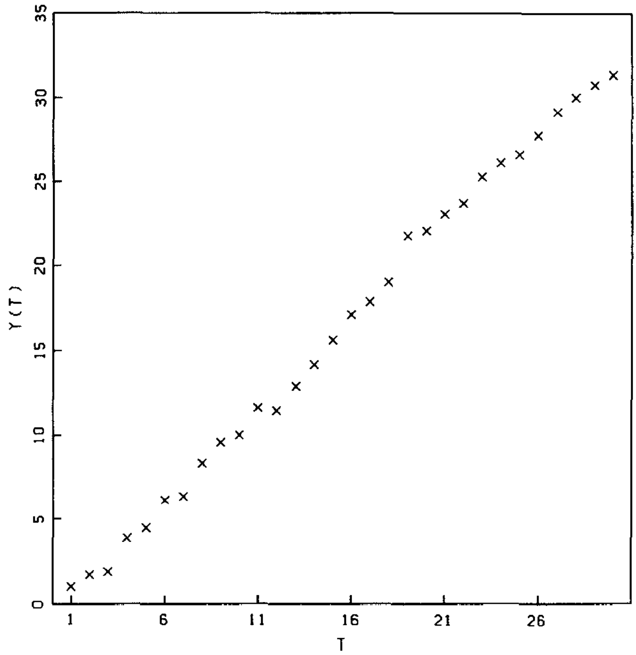

Fig. 1. A sample of the data generated using the second alternative with $\lambda = 0.5$.

4.2 *Simulation*

To check the performance of the procedure proposed in the previous section, we conducted a simulation study. The alternatives we used for generating data are:

$$x_t = \begin{cases} t & t = 1, \dots, 14 \\ 15 & t = 15, \dots, 29 \end{cases} \qquad \text{and} \qquad x_t = \begin{cases} t & t = 1, \dots, 15 \\ t+2 & t = 16, \dots, 30. \end{cases}$$

The first alternative involves a change in slope just at $t = 15$, and the second involves a shift in level between $t = 15$ and 16. We generated data by $y_t = x_t + e_t$, where e_t is a Gaussian random number with mean 0 and variance λ^2. Then we applied the proposed procedure to the data assuming $\hat{D} = U$, because $H_{U,0}$ was always selected as the best fit one among $H_{G,0}$, $H_{\Gamma,0}$ and $H_{U,0}$ in the preliminary analysis. The values assumed for λ are 0.125, 0.25, 0.5 and 1, and the number of trials is 20 for each value of λ.

The results are summarized in Table 1, which shows the frequency that $H_{U,0}$ was selected at each k for each value of λ in each alternative. As seen in this table, when $\lambda = 0.125$, the procedure perfectly detected the structural changes. However, with the increase of λ, the procedure became unable to detect them correctly, and a kind of edge effect began to appear near the change and end points. Such tendency was remarkable in the second alternative. Especially, in the second alternative with $\lambda = 1$, the procedure showed a tendency to select a simple straight line. However, these results may be reasonable, because it is usual that signal disappears with the increase of noise. We show a sample of the data

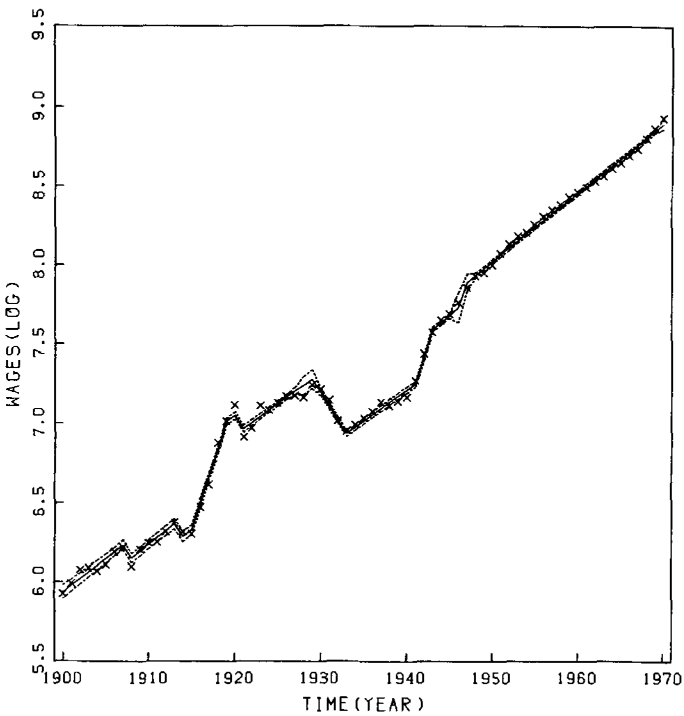

Fig. 2. Logs of the nominal wages in U.S. during 1900–1970 (cross), and the mean (straight line) and a ± 3 sigma interval (dotted line) of the trend component estimated by using $H_{U,0}$.

generated using the second alternative with $\lambda = 0.5$ in Fig. 1. The detection of the structural change may be difficult even by professional eye.

5. Examples

In this section, we show the results of application of the proposed method to economic time series data.

Example 1. The first data set we analyze is the series of the logs of the nominal wages in U.S. during 1900–1970. This data set has been analyzed by Nelson and Plosser (1982), Perron (1989) and Bianchi (1993). Nelson and Plosser (1982) applied unit root models, and Perron (1989) made an intervention analysis. However, their works were not concerned with the detection of an unknown number of structural changes. On the other hand, Bianchi (1993) applied a Bayesian method based on Kashiwagi (1991) to detect an unknown number of shifts in level in the series of the first differences of the data splitting the data into two subsamples. Our purpose is to detect an unknown number of structural changes in a linear trend using all of the data.

The obtained values of $\mathrm{AIC}_{G,0}$, $\mathrm{AIC}_{\Gamma,0}$ and $\mathrm{AIC}_{U,0}$ are -170.67, -169.26 and -171.66, respectively. Therefore, the best fit model is $H_{U,0}$. The maximum

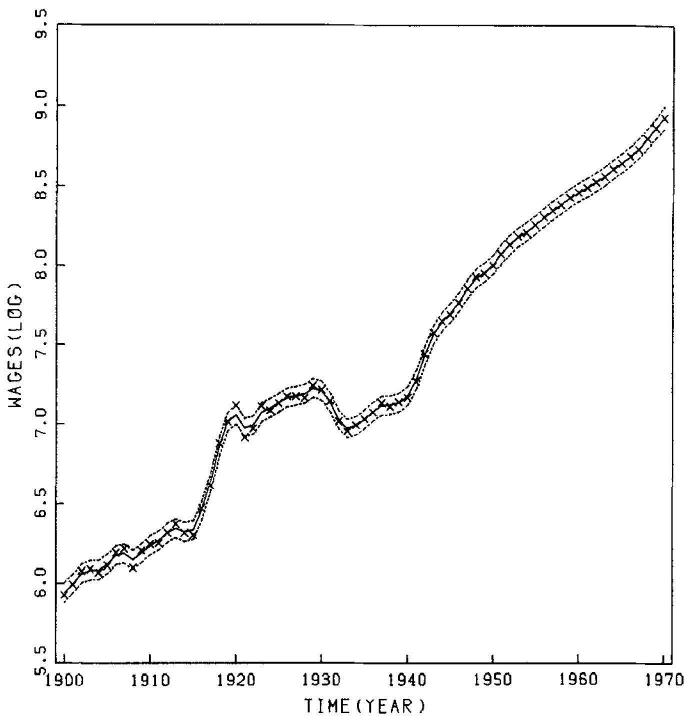

Fig. 3. Logs of the nominal wages in U.S. during 1900–1970 (cross), and the mean (straight line) and a ± 3 sigma interval (dotted line) of the trend component estimated by using the standard smoothing model.

likelihood estimate of $\boldsymbol{\theta}_U$ was obtained as $\hat{\sigma}_v^2 = 0.0304^2$, $\hat{\alpha} = 0.796$, $\hat{g} = 0.126$ and $\hat{h} = 0.128$. Figure 2 shows the mean (straight line) and a ± 3 sigma interval (dotted line) of the trend component estimated by using $H_{U,0}$ as well as the data (cross) plotted against time. The estimated mean changes linearly and turns round clearly at 1907, 1908, 1913, 1914, 1915, 1919, 1920, 1921, 1929, 1933, 1941, 1943, 1946 and 1947. For the comparison, we show the results estimated by using the standard smoothing model, that is, $H_{G,0}$ with $\alpha = 0$ in Fig. 3. The estimated mean changes gradually and many turns are observed other than those mentioned above. Table 2 gives the values of $\mathrm{AIC}_{U,k}$'s. The value of $\mathrm{AIC}_{U,0}$ is less than that of $\mathrm{AIC}_{U,k}$ at 14 time periods. These time periods agree with those detected by observing Fig. 2. It can be considered that structural changes in a linear trend had occurred at these time periods.

Example 2. The second data set is the series of the logs of the wholesale price index in U.S. during 1890–1970. For this data set, the values of AIC's were obtained as $\mathrm{AIC}_{G,0} = -160.85$, $\mathrm{AIC}_{\Gamma,0} = -160.29$ and $\mathrm{AIC}_{U,0} = -152.64$. Therefore, the best fit model is $H_{G,0}$. The maximum likelihood estimate of $\boldsymbol{\theta}_G$ was obtained as $\hat{\sigma}_v^2 = 0.0302^2$, $\hat{\alpha} = 0.792$ and $\hat{\sigma}_w^2 = 0.204^2$. Figure 4 shows the results estimated by using $H_{G,0}$. The estimated mean clearly turns round at 1896, 1897, 1915, 1920, 1921, 1929, 1932, 1937, 1939, 1945 and 1948. In addition to these,

Table 2. The values of $\mathrm{AIC}_{U,k}$'s in the nominal wages data.

k	(year)	$\mathrm{AIC}_{U,k}$	$\mathrm{AIC}_{U,k}-\mathrm{AIC}_{U,0}$
2	(1901)	−172.12	−0.459
3	(1902)	−172.12	−0.459
4	(1903)	−172.11	−0.444
5	(1904)	−172.11	−0.446
6	(1905)	−172.12	−0.461
7	(1906)	−172.11	−0.451
8	(1907)	−166.67	4.995
9	(1908)	−166.67	4.990
10	(1909)	−172.12	−0.461
11	(1910)	−172.12	−0.461
12	(1911)	−172.12	−0.461
13	(1912)	−172.12	−0.460
14	(1913)	−166.04	5.625
15	(1914)	−166.05	5.609
16	(1915)	−162.11	9.552
17	(1916)	−172.11	−0.448
18	(1917)	−172.11	−0.449
19	(1918)	−172.11	−0.449
20	(1919)	−152.13	19.536
21	(1920)	−164.66	7.002
22	(1921)	−164.66	7.000
23	(1922)	−172.12	−0.461
24	(1923)	−172.12	−0.461
25	(1924)	−172.12	−0.461
26	(1925)	−172.12	−0.461
27	(1926)	−172.12	−0.453
28	(1927)	−172.06	−0.398
29	(1928)	−172.06	−0.400
30	(1929)	−168.38	3.281
31	(1930)	−172.06	−0.396
32	(1931)	−172.12	−0.461
33	(1932)	−172.12	−0.461
34	(1933)	−158.78	12.888
35	(1934)	−172.12	−0.461
36	(1935)	−172.12	−0.461
37	(1936)	−172.12	−0.461
38	(1937)	−172.12	−0.460
39	(1938)	−172.12	−0.459
40	(1939)	−172.12	−0.460
41	(1940)	−172.12	−0.460
42	(1941)	−168.62	3.044
43	(1942)	−172.12	−0.460
44	(1943)	−166.78	4.883
45	(1944)	−172.12	−0.456
46	(1945)	−171.98	−0.313
47	(1946)	−164.83	6.833
48	(1947)	−167.03	4.628
49	(1948)	−172.08	−0.418
50	(1949)	−172.12	−0.461
51	(1950)	−172.12	−0.461
52	(1951)	−172.12	−0.461
53	(1952)	−172.12	−0.461
54	(1953)	−172.12	−0.461
55	(1954)	−172.12	−0.461
56	(1955)	−172.12	−0.461
57	(1956)	−172.12	−0.461
58	(1957)	−172.12	−0.461
59	(1958)	−172.12	−0.461
60	(1959)	−172.12	−0.461
61	(1960)	−172.12	−0.461
62	(1961)	−172.12	−0.461
63	(1962)	−172.12	−0.461
64	(1963)	−172.12	−0.461
65	(1964)	−172.12	−0.461
66	(1965)	−172.12	−0.461
67	(1966)	−172.12	−0.461
68	(1967)	−172.12	−0.461
69	(1968)	−172.12	−0.461
70	(1969)	−172.11	−0.444

Table 3. The values of $\mathrm{AIC}_{G,k}$'s in the wholesale price index data.

k	(year)	$\mathrm{AIC}_{G,k}$	$\mathrm{AIC}_{G,k}-\mathrm{AIC}_{G,0}$	k	(year)	$\mathrm{AIC}_{G,k}$	$\mathrm{AIC}_{G,k}-\mathrm{AIC}_{G,0}$
2	(1891)	−161.20	−0.348	42	(1931)	−161.16	−0.307
3	(1892)	−161.25	−0.395	43	(1932)	−152.93	7.925
4	(1893)	−161.25	−0.395	44	(1933)	−161.09	−0.235
5	(1894)	−161.23	−0.382	45	(1934)	−161.16	−0.310
6	(1895)	−161.20	−0.343	46	(1935)	−160.86	−0.009
7	(1896)	−160.08	0.770	47	(1936)	−160.64	0.215
8	(1897)	−159.81	1.041	48	(1937)	−159.34	1.510
9	(1898)	−161.10	−0.244	49	(1938)	−160.94	−0.084
10	(1899)	−160.90	−0.051	50	(1939)	−159.77	1.083
11	(1900)	−160.80	0.050	51	(1940)	−160.01	0.837
12	(1901)	−161.05	−0.198	52	(1941)	−161.08	−0.225
13	(1902)	−160.84	0.016	53	(1942)	−160.26	0.596
14	(1903)	−161.03	−0.180	54	(1943)	−160.60	0.257
15	(1904)	−161.10	−0.253	55	(1944)	−161.06	−0.203
16	(1905)	−161.07	−0.220	56	(1945)	−158.44	2.413
17	(1906)	−161.19	−0.338	57	(1946)	−160.69	0.165
18	(1907)	−161.18	−0.330	58	(1947)	−160.19	0.661
19	(1908)	−161.20	−0.348	59	(1948)	−158.73	2.124
20	(1909)	−161.12	−0.266	60	(1949)	−161.16	−0.307
21	(1910)	−160.95	−0.096	61	(1950)	−161.21	−0.353
22	(1911)	−161.17	−0.322	62	(1951)	−161.19	−0.337
23	(1912)	−161.21	−0.354	63	(1952)	−161.25	−0.403
24	(1913)	−161.23	−0.376	64	(1953)	−161.28	−0.430
25	(1914)	−161.22	−0.364	65	(1954)	−161.29	−0.443
26	(1915)	−148.61	12.241	66	(1955)	−161.30	−0.448
27	(1916)	−160.77	0.085	67	(1956)	−161.31	−0.453
28	(1917)	−157.00	3.857	68	(1957)	−161.31	−0.455
29	(1918)	−160.85	0.000	69	(1958)	−161.30	−0.449
30	(1919)	−161.14	−0.292	70	(1959)	−161.30	−0.446
31	(1920)	−134.78	26.074	71	(1960)	−161.29	−0.442
32	(1921)	−139.52	21.334	72	(1961)	−161.29	−0.440
33	(1922)	−161.20	−0.344	73	(1962)	−161.29	−0.436
34	(1923)	−161.23	−0.382	74	(1963)	−161.28	−0.433
35	(1924)	−161.24	−0.384	75	(1964)	−161.28	−0.430
36	(1925)	−161.19	−0.335	76	(1965)	−161.28	−0.428
37	(1926)	−161.21	−0.354	77	(1966)	−161.27	−0.417
38	(1927)	−161.22	−0.369	78	(1967)	−161.24	−0.383
39	(1928)	−161.12	−0.265	79	(1968)	−161.17	−0.317
40	(1929)	−156.45	4.406	80	(1969)	−161.09	−0.243
41	(1930)	−161.05	−0.201				

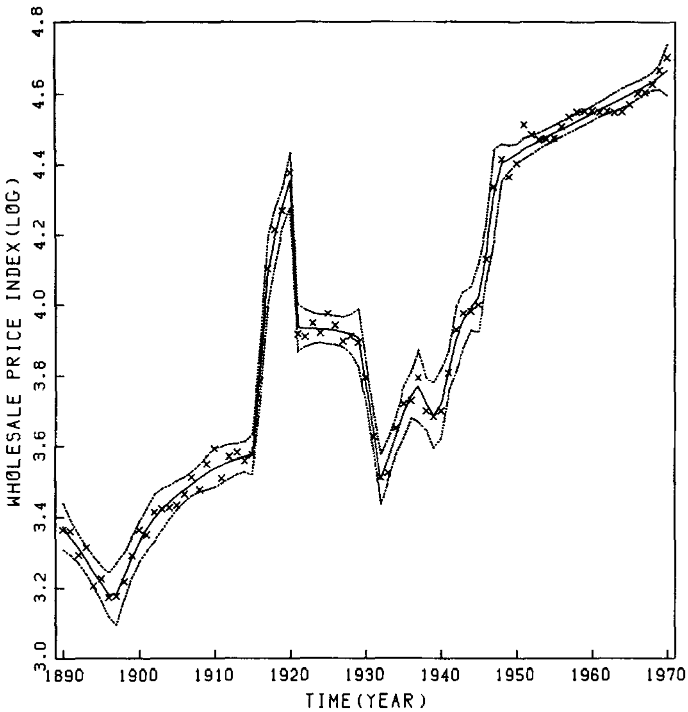

Fig. 4. Logs of the wholesale price index in U.S. during 1890–1970 (cross), and the mean (straight line) and a ± 3 sigma interval (dotted line) of the trend component estimated by using $H_{G,0}$.

several turns including gradual changes are observed. However, it seems difficult to detect structural changes exactly only by observing the estimated mean, because some of the turns lie within the range of the sigma interval. For the comparison, we show the results estimated by using the standard smoothing model in Fig. 5. The estimated mean seems to follow the data excessively, and the sigma interval is relatively wide. Table 3 gives the values of $\mathrm{AIC}_{G,k}$'s. The value of $\mathrm{AIC}_{G,0}$ is less than that of $\mathrm{AIC}_{G,k}$ at 22 time periods, that is, at 1896, 1897, 1900, 1902, 1915, 1916, 1917, 1918, 1920, 1921, 1929, 1932, 1936, 1937, 1939, 1940, 1942, 1943, 1945, 1946, 1947 and 1948. If we apply the proposed procedure strictly, these time periods are regarded as change points. However, some of the values of $\mathrm{AIC}_{G,k} - \mathrm{AIC}_{G,0}$ are very close to 0, for example, at 1902, 1918 and 1935. It may be inappropriate to conclude about these time periods hastily. One of causes of such values being obtained may be the assumption that the observation noise is independent. Additionally, in the analysis of econometric time series, it is usual to assume an autocorrelated observation noise. An autocorrelated observation noise is assumable in our framework, though it is necessary to solve some computational difficulties.

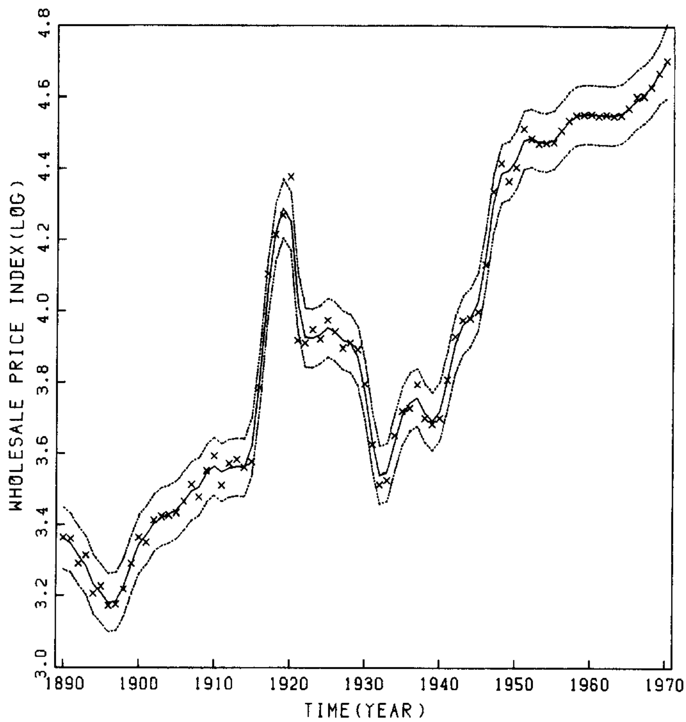

Fig. 5. Logs of the wholesale price index in U.S. during 1890–1970 (cross), and the mean (straight line) and a ±3 sigma interval (dotted line) of the trend component estimated by using the standard smoothing model.

Acknowledgements

The author is grateful to the anonymous referees for their valuable comments.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *Second International Symposium on Information Theory* (eds. B. N. Petrov and F. Csaki), 267–281, Akademiai Kiado, Budapest.

Akaike, H. (1980). Likelihood and the Bayes procedure, *Trabajos de Estadistica*, **31**, 143–166.

Bianchi, M. (1993). Infrequent shocks and signal extraction in economic time series, Ph.D. Thesis, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Broemeling, L. D. and Tsurumi, H. (1986). *Econometrics and Structural Change*, Marcel Dekker, New York.

Gordon, K. and Smith, A. F. M. (1990). Modeling and monitoring biomedical time series, *J. Amer. Statist. Assoc.*, **85**, 328–337.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting (with discussion), *J. Roy. Statist. Soc. Ser. B*, **38**, 205–247.

Jacoby, S. L. S., Kowalik, J. S. and Pizzo, J. T. (1972). *Iterative Methods for Nonlinear Optimization Problems*, Prentice-Hall, New Jersey.

Kashiwagi, N. (1991). Bayesian detection of structural changes, *Ann. Inst. Statist. Math.*, **43**, 77–93.

Kashiwagi, N. (1993). On use of the Kalman filter for spatial smoothing, *Ann. Inst. Statist. Math.*, **45**, 21–34.

Kashiwagi, N. and Yanagimoto, T. (1992). Smoothing serial count data through a state-space model, *Biometrics*, **48**, 1187–1194.

Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series (with discussion), *J. Amer. Statist. Assoc.*, **82**, 1032–1063.

Kohn, R. and Ansley, C. F. (1987). A new algorithm for spline smoothing based on smoothing a stochastic process, *SIAM J. Sci. Statist. Comput.*, **8**, 33–48.

Nelson, C. R. and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series, *Journal of Monetary Economics*, **10**, 139–162.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis, *Econometrica*, **57**, 1361–1401.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes, *J. Amer. Statist. Assoc.*, **53**, 873–880.

Shumway, R. H. and Stoffer, D. S. (1991). Dynamic linear models with switching, *J. Amer. Statist. Assoc.*, **86**, 763–769.

Tsurumi, H. (1988). Survey of Bayesian and non-Bayesian testing of model stability in econometrics, *Bayesian Analysis of Time Series and Dynamic Linear Models* (ed. J. C. Spall), 75–100, Marcel Dekker, New York.

Wecker, E. W. and Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing, *J. Amer. Statist. Assoc.*, **78**, 81–89.