Detecting Change in Variance of Unequally Spaced Time Series

Tze-San Lee and N. Hou
Western Illinois University

Key Words: Asbestos exposure data, Brownian motion, variance-change model, likelihood ratio test, Ornstein-Uhlenbeck process.

Abstract. To detect a change in the variance of unequally spaced time series, unobservable errors in the data are modeled by the Ornstein-Uhlenbeck process. By employing the likelihood ratio test, a statistical method is proposed for testing the change-point. The exposure data collected at Lackland Air Force Base from an asbestos-removal team of five workers is used to illustrate the proposed test.

Introduction. A detection of possible changes in the variance of time series has many practical applications. Much work had been done when time series are equally spaced. For example, it has been noticed that the logarithm of the price relatives of daily stock price data is statrionary in mean, but not in variance [Hsu (1977) and Wichern, et al (1976)]. In a study of three work-shifts on the worker's mental performance, Lee (1995) built a variance-change model for the data set. For other related works, see Hsu, et al (1974), Hsu (1979, 1982), Picard (1985), and Tsay (1988). However, no work was ever done on the detection of a change in the variance of unequally spaced time series.

It is the aim of this paper to propose a statistical method for detecting a change in the variance of unequally spaced longitudinal data based upon the likelihood ratio test.

Change-point Hypothesis. Let $y(t_i)$ be the longitudinal data observed at the time points $t_1 < t_2 < \ldots < t_n$ satisfying the following model

$$y(t_i) = \mu + \varepsilon(t_i), i = 1,2,...,n \,, \tag{1}$$

and unobservable errors $\varepsilon(t)$ are modeled by the Ornstein-Uhlenbeck process given as follows:

$$d\varepsilon(t) + \gamma\varepsilon(t)dt = \sigma dW(t), \tag{2}$$

where μ is assumed to be known, $\gamma > 0$, $\sigma > 0$, and W(t) is the Brownian motion or Wiener process (Arnold 1974). Since the proposed test is location invariant, μ is assumed to be zero.

To detect a change in the variance of $y(t_i)$, i = 1, 2, …, n, is equivalent to test the following hypotheses:

H_0: $$Var(y(t_i) \mid y(t_{i-1})) = \frac{\sigma_0^2[1-f_i^2(\gamma_0)]}{2\gamma_0}, i = 2,...,n \,, \tag{3}$$

versus
H_1:

$$Var(y(t_i) \mid y(t_{i-1})) = \begin{cases} [\sigma_1^2(1-f_i^2(\gamma_1))]/2\gamma_1, i = 2,...,k-1 \\ [\sigma_2^2(1-f_i^2(\gamma_1))]/2\gamma_1, i = k,........,n \end{cases} \tag{4}$$

Where γ_0, γ_1, σ_0^2, $\sigma_1^2 \neq \sigma_2^2$ and k are unknown parameters to be estimated, and $f_i(\gamma)$ is given by

$$f_i(\gamma) = \exp(-\gamma\Delta t_i) \,. \tag{5}$$

and

$$\Delta t_i = t_i - t_{i-1} \,. \tag{6}$$

Incidentally, $Var(y(t_1))$ equals $\frac{\sigma_0^2}{2\gamma_0}$ and $\frac{\sigma_1^2}{2\gamma_1}$ in (3) and (4), respectively. Also, the unknown parameter k in (4) is called the change-point for the variance-change model.

Likelihood Ratio Test. Under H_0, the logarithmic likelihood of $y(t_1), y(t_2), \ldots, y(t_n)$ can be shown to be given by (Robinson 1977)

$$\ln L_0(\gamma_0, \sigma_0) = -\frac{n}{2}\ln\left(\frac{\pi\sigma_0^2}{\gamma_0}\right) - \frac{1}{2}\sum_{i=2}^{n}\ln(1-f_i^2(\gamma_0))$$
$$-\frac{\gamma_0}{\sigma_0^2}\left\{\sum_{i=2}^{n}\frac{S(\gamma_0;\Delta t_i)}{1-f_i^2(\gamma_0)} + [y(t_1)]^2\right\}, \tag{7}$$

where $f_i(\gamma)$ is given by (5), and $S(\gamma; \Delta t_i)$ is given by

$$S(\gamma;\Delta t_i) = [y(t_i) - f_i(\gamma)y(t_{i-1})]^2 \,. \tag{8}$$

By conditioning on that k is given, the logarithmic likelihood of $y(t_1), y(t_2), \ldots, y(t_n)$ under H_1 is given by

$$\ln L_1(\gamma_1, \sigma_1, \sigma_2, k) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln\left(\frac{\sigma_1^2}{2\gamma_1}\right)$$
$$-\frac{1}{2}\sum_{i=2}^{k-1}\ln\left(\frac{\sigma_1^2[1-f_i^2(\gamma_1)]}{2\gamma_1}\right) - \frac{1}{2}\sum_{i=k}^{n}\ln\left(\frac{\sigma_2^2[1-f_i^2(\gamma_1)]}{2\gamma_1}\right)$$
$$-\frac{\gamma_1}{\sigma_1^2}\left\{[y(t_1)]^2 + \sum_{i=2}^{k-1}\frac{S(\gamma_1;\Delta t_i)}{1-f_i^2(\gamma_1)}\right\} - \frac{\gamma_1}{\sigma_2^2}\cdot\sum_{i=k}^{n}\frac{S(\gamma_1;\Delta t_i)}{1-f_i^2(\gamma_1)} \tag{9}$$

Let $\hat{g}_0, \hat{s}_0$, $\hat{s}_1, \hat{s}_2$, and $\hat{g}_1$ given by (A.1-2, 4-6) in the Appendix, denote the maximum likelihood estimator of the parameters γ_0, σ_0, σ_1, σ_2, and γ_1, respectively. Thus, the generalized likelihood ratio test for k under H_0 against H_1 is then given, as follows:

$$-2\ln[\frac{\hat{L}_0}{\hat{L}_1(\hat{k})}] = \max_{2 \le k \le n-1} \{-2\ln[\frac{\hat{L}_0}{\hat{L}_1(k)}]\}, \qquad (10)$$

where $\hat{L}_0 \equiv L_0(\hat{g}_0, \hat{s}_0)$, and $\hat{L}_1(k) \equiv L_1(\hat{g}_1, \hat{s}_1, \hat{s}_2, k)$. As usual, the asymptotic sampling distribution of $\hat{k}$ is approximately a chi-square distribution with one degree of freedom.

Application. Two examples are shown to illustrate the proposed test for detecting the change-point in the variance-change model.

Example 1. The asbestos exposure data collected from a team of five workers is used to illustrate the proposed test. This team was charged to remove the asbestos from a two-story building which was previously used as a clinic at Lackland Air Force Base. The data are given in Table 1. Since the ratio of the largest data to the smallest one for all workers is of the order of magnitude two or higher, we assume that the data follows a lognormal distribution. Further, those data with an asterisk in Table 1 are the overloaded data. Since there is no way to measure the correct level from the overloaded data, a conservative estimate was given by an Air Force bioenvironmental engineer as 5 fibers/cm^3. Compared with other data in Table 1, this observation, 5 fibers/cm^3, can be viewed as an outlier.

After taking the logarithm of the data in Table 1 for each worker, we first removing the sample mean from the data set by using the result of detecting change in the mean of these unequally spaced longitudinal data (Lee 1998). Then, we applied the bisection method (Press, et al 1992) to (A.2) in the Appendix in obtaining $\hat{g}_0$ after substituting (A.1) into (A.2) and then substitute $\hat{g}_0$ back into (A.1) to find $\hat{s}_0$. Similarly, we first substitute (A.4) and (A.5) into (A.6) in the Appendix and then applied the bisection method to (A.6) in obtaining $\hat{g}_1$. After obtaining $\hat{g}_1$, we substitute it into (A.4) and (A.5) to find $\hat{s}_1$ and $\hat{s}_2$, respectively. The estimate $\hat{k}$ for the change-point k is thus obtained from the likelihood ratio test of (10) for all workers. Tables 2(a) and 2(b) are obtained, respectively, with and without including the outliers in the

Table 1. The daily asbestos exposure levels (fibers/cm^3) for five workers.

Date (t_i)	Worker				
	A	B	C	D	E
5-9-91 (1)	.019		.027		.004
5-10-91 (2)	.015				.008
5-13-91 (5)	.009		.006		.006
5-14-91 (6)	Wet[a]	1.102	.02		.194
5-15-91 (7)		1.915	Wet		5.0*
5-21-91 (13)		.071			5.0*
5-24-91 (16)				.03	.053
5-28-91 (20)	.035			.027	5.0*
5-30-91 (22)		.069			.264
6-18-91 (41)	.005		.003	.003	
6-20-91 (43)			.006	.004	.008
6-21-91 (44)			.014	Wet	
7-16-91 (69)				.043	.04
7-17-91 (70)	.025				5.0*
7-18-91 (71)			5.0*		
7-19-91 (72)		.008	Wet		5.0*
7-22-91 (75)			5.0*	.006	5.0*
7-23-91 (76)	5.0*	5.0*	.001		
7-24-91 (77)					.015
7-25-91 (78)			.009	.018	
8-14-91 (98)	.05				
8-15-91 (99)	.031	.081	.094		.121
8-16-91 (100)		.142		.004	
8-19-91 (103)	.054			.059	
8-21-91 (105)	.118	.069	.133	.114	
8-22-91 (106)		.002	.1		
8-23-91 (107)		.111	.002		

[a] Wet means that the filter cassette was wet and torn apart.

calculation. Since the values of $-2\ln[\frac{L_0(\hat{g}_0, \hat{s}_0)}{L_1(\hat{g}_1, \hat{s}_1, \hat{s}_2, \hat{k})}]$ in the last row of Table 2(a) for Worker E is greater than the critical value of $c^2_{.95}(2) = 5.99$ (Hogg and Tanis 1997), the null hypothesis H_0 is rejected for Worker E at the 5% level of significance. Indeed, the variance of the logarithm of the daily asbestos exposure for Worker E has changed from 21.8 (= $\hat{s}_1^2 = 4.67^2$) to 100 (= $\hat{s}_2^2 = 10^2$) when $\hat{k} = 6$, or $t_{\hat{k}} = 13$ (Table 2(a)). After outliers are removed from the data set, the null hypothesis H_0 is then shown to be not rejected for all workers, namely, the variance of the logarithm of the daily asbestos exposure is constant and has not changed during the asbestos removal period because all of their values of $-2\ln[\frac{L_0(\hat{g}_0, \hat{s}_0)}{L_1(\hat{g}_1, \hat{s}_1, \hat{s}_2, \hat{k})}]$ in the last row of Table

2(b) are less than the critical value of $c^2_{.95}(2) = 5.99$. This shows that the likelihood ratio test of (10) is not robust against outliers.

Example 2. This example is the same as Example 1 in Lee (1995). The original data can be found in Table 16.3 of Mansfield (1994), p.659. The data are equally spaced time series. Nevertheless, we apply the proposed test of (10) to detect the change-point in the variance to see whether it works for the equally spaced time series. A change-point $\hat{k} = 20$ was found to be significant with the likelihood ratio being

Table 2. Summary statistics for five workers:

(a) Outliers are included in calculation.

	Worker				
	A	B	C	D	E
n (sample size)	11	11	14	10	16
$\hat{g}_0$	0.14	1.25	7.9	0.23	7.6
$\hat{s}_0$	0.73	3.39	10.0	0.62	8.5
$\hat{g}_1$	0.14	0.37	7.8	0.44	8.0
$\hat{s}_1$	0.39	2.51	5.39	0.91	4.67
$\hat{s}_2$	0.99	6.47	12.3	0.35	10.0
$\hat{k}$	7	10	7	9	6
$-2\ln(\frac{\hat{L}_0}{\hat{L}_1})$	4.60	1.04	3.95	1.56	10.3

(b) Outliers are not included in calculation.

	Worker				
	A	B	C	D	E
n (sample size)	10	10	12	-	10
$\hat{g}_0$	0.12	7.9	2.28	-	4.2
$\hat{s}_0$	0.46	4.99	3.26	-	2.96
$\hat{g}_1$	0.14	7.7	6.0	-	7.3
$\hat{s}_1$	0.22	1.72	2.83	-	1.09
$\hat{s}_2$	0.51	6.21	6.89	-	4.60
$\hat{k}$	3	5	7	-	4
$-2\ln(\frac{\hat{L}_0}{\hat{L}_1})$	1.42	5.68	4.21	-	5.34

$-2\ln[\frac{L_0(\hat{g}_0, \hat{s}_0)}{L_1(\hat{g}_1, \hat{s}_1, \hat{s}_2, \hat{k})}] = 10.8$. This result is the same as the one obtained in Mai (1999), p. 15. This shows that the proposed test can also be applied to for the equally spaced time series.

Concluding Remarks. Based upon the likelihood ratio test, a statistical method for detecting a change in the variance of unequally spaced time series is presented. From an application to the asbestos exposure data of five workers, it seems that the proposed method works very well. Nevertheless, it can been seen from the illustrative example that the proposed test is not robust. This is attributed to the inherent property of the likelihood ratio test. Due to that the sample size is usually small in practical applications, it is desirable that a finite sampling distribution of $\hat{k}$ is to be worked out.

Acknowledgements. The asbestos exposure data of Table 1 were provided by Lt. Mareno-Ferguson while the author was appointed as a Summer Faculty Research Fellow during the summer of 1995 to work in the Occupational and Environmental Medicine Division, Armstrong Laboratory at Brooks AFB, San Antonio, Texas. Maple and SYSTAT were used to facilitate numerical calculation and data analysis, respectively.

Appendix. From the equation of $\frac{\partial \ln L_0}{\partial s_0^2} = 0$, we have

$$\hat{s}_0^2 = \frac{2g_0}{n}[y^2(t_1) + \sum_{i=2}^{n} \frac{S(g_0;\Delta t_i)}{1 - f_i^2(g_0)}], \qquad \text{(A.1)}$$

where $S(g_0; \Delta t_i)$ and $f_i(g)$ are given by (6) and (7), respectively.

From the equation of $\frac{\partial \ln L_0}{\partial g_0} = 0$, $\hat{g}_0$ satisfies the following equation:

$$0 = \frac{n}{2g_0} - \sum_{i=2}^{n} f_0(g_0;\Delta t_i) - \frac{1}{s_0^2}\{y^2(t_1) + \sum_{i=2}^{n}[f_1(g_0;\Delta t_i) y^2(t_i)$$

$$+ f_2(g_0;\Delta t_i) y(t_i) y(t_{i-1}) + f_3(g_0;\Delta t_{i-1}) y^2(t_{i-1})]\} \qquad \text{(A.2)}$$

where f_i, i = 0, 1, 2, 3, are given, respectively, by

$$f_0(g;\Delta t_i) = \frac{\Delta t_i f_i^2(g)}{1 - f_i^2(g)}, \qquad \text{(A.3a)}$$

$$f_1(g;\Delta t_i) = \frac{1 - (1 + 2g\Delta t_i) f_i^2(g)}{[1 - f_i^2(g)]^2}, \qquad \text{(A.3b)}$$

$$f_2(g;\Delta t_i) = \frac{2f_i(g)[(1 + 2g\Delta t_i) f_i^2(g) + g\Delta t_i - 1]}{[1 - f_i^2(g)]^2} \qquad \text{(A.3c)}$$

and

$$f_3(g;\Delta t_i)=\frac{f_i^2(g)[1-2g\Delta t_i-(1+2g\Delta t_i)f_i^2(g)]}{[1-f_i^2(g)]^2}$$

(A.3d)

Similarly, we have from $\frac{\partial \ln L_1}{\partial s_1^2}=\frac{\partial \ln L_1}{\partial s_2^2}=0$

$$\hat{s}_1^2=\frac{2g_1}{k-1}[y^2(t_1)+\sum_{i=2}^{k-1}\frac{S(g_1;\Delta t_i)}{1-f_i^2(g_1)}] \quad (A.4)$$

$$\hat{s}_2^2=\frac{2g_1}{n-k+1}\sum_{i=k}^{n}\frac{S(g_1;\Delta t_i)}{1-f_i^2(g_1)} \quad (A.5)$$

and $\hat{g}_1$ satisfies the following equation:

$$0=\frac{n}{2g_1}-\sum_{i=2}^{n}f_0(g_1;\Delta t_i)-s_1^{-2}\{y^2(t_1)+$$

$$\sum_{i=2}^{k-1}[f_1(g_1;\Delta t_i)y^2(t_i)+f_2(g_1;\Delta t_i)y(t_i)y(t_{i-1})+$$

$$f_3(g_1;\Delta t_{i-1})y^2(t_{i-1})]\}-s_2^{-2}[\sum_{i=k}^{n}f_1(g_1;\Delta t_i)y^2(t_i)$$

$$+f_2(g_1;\Delta t_i)y(t_i)y(t_{i-1})+f_3(g_1;\Delta t_{i-1})y^2(t_{i-1})].$$

(A.6)

where f_i , i = 0, 1, 2, 3, are given, respectively, by (A3a-d).

References

Arnold, L. (1974). Stochastic Differential Equations: Theory and Applications. John Wiley & Sons, New York.

Hogg, R. V. and Tanis, E. A. (1997). Probability and Satatistical Inference, 5th edition, p. 675, Prentice Hall, Upper Saddle River, New Jersey.

Hsu, D. A., Miller, R. B., and Wichern, D. W. (1974). On the stable Paretian behavior of stock-market prices, J. Amer. Statist. Assoc., 69, 108-113.

Hsu, D. A. (1977). Tests for variance shift at an unknown time point, Appl. Statist., 26, 279-284.

_____ (1979). Detecting shifts of parameter in Gamma sequences with applications to stock price and air traffic flow analysis, J. Amer. Statist. Assoc., 74, 31-40.

_____ (1982). A Bayesian robust detection of shift in the risk structure of stock market returns, J. Amer. Statist. Assoc., 77, 29-39.

Lee, T-S. (1995). A test to detect change in variance. Computing Science and Statistics, 27, 541-544.

_______ (1998). Detecting change in mean of unequally spaced longitudinal data, Computing Science and Statist., 30, 112-115.

Mai, H. C. (1999). Change-Point Models for Time Series, M. S. thesis, Western Illinois University, Macomb, Illinois.

Mansfield, E. (1994). Statistics for Business and Economics, 5th edition, W. W. Norton, New York.

Picard, D. (1985). Testing and estimating change-points in time series, Advances in Appl. Prob., 17, 841-867.

Press, W. H., et al (1992). Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, Cambridge, England.

Robinson, P. M. (1977). Estimation of a time series model from unequally spaced data. Stochastic Proc. & their Appl. 6, 9-24.

Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series, J. Forecasting, 7, 1-20.

Wichern, D.W., Miller, R.B., and Hsu, D.A. (1976). Changes of variance in first-order autoregressive time series models with an application, Appl. Statist., 25, 248-256.