Chapter 6

CONTROLLING TYPE I ERROR RATE IN MONITORING STRUCTURAL CHANGES

Abstract

In this chapter we develop some asymptotically power one partially sequential nonparametric tests for monitoring structural changes. Our test procedures are based on Wilcoxon score. We use the idea of curved stopping boundaries. We derive some exact results and perform simulation studies to provide various properties of the tests. We see that one of the proposed procedures significantly controls the type I error rate. This procedure may be very effective for fluctuation monitoring. We illustrate the procedures by using real life data from the stock market.

6.1 Introduction

The present chapter is motivated from a suggested future research problem in Chu et al. (1996). In Chapter 5 we mention that they used partially sequential sampling scheme in monitoring structural changes. In fact, their work may be looked upon as a two sample testing procedure. A historical data of prefixed size, available at the time of designing fluctuation monitoring, is regarded as the first sample. Thereafter the second sample observations are recorded sequentially. Obviously statisticians do not like to consider more samples than are necessary to detect instability. This again leads to partially sequential stopping procedure as in the previous chapters.

However all such sequential stopping rules, introduced in earlier chapters, terminate with probability one under both the null and alternative hypotheses. Thus the monitoring stops even when there is no fluctuation in the populations. This is really unwarranted in various econometric as well as environmental monitoring. Here a process needs to be monitored ceaselessly. Particularly, when we assume a negligible cost of sampling under no fluctuation, we do not need to stop at all unless there is a signal.

Chu et al. (1996) emphasizes the need for developing partial sequential tests which rarely terminate when no fluctuation is observed. They also discussed the importance of controlling type I error in monitoring structural changes. Thus, we require framing a rule in such a way that a termination will signal instability. Hence we certainly need to construct tests with asymptotically or approximately power one and simultaneously to achieve control over type I error rate.

Research related to controlling type I error rate in statistical inference are getting more and more importance. As the pattern of the alternative is unknown or vague in most cases, deriving optimal tests become complicated. Statisticians are, in these days, avoiding the long practice of using mathematically sound optimal tests. Those tests, mainly in the area of sequential clinical trials, are continually being replaced by some heuristic tests that are more practical and easy to handle. Such heuristic tests, if properly designed, are likely to have greater appeal in testing some econometric or environmental hypothesis as well. One may see Huang et al. (2005) for an illustration of controlling type I error in adjusted O'Brien's test.

In the present chapter we primarily consider the problems of testing H_0 against H_{B1} and against $\tilde{H}_{B1}$. With the sampling technique described in Chapters 4 and 5, such problems may be looked upon as the problems of testing the identity of two continuous univariate distribution functions against one or two sided location alternatives. We develop tests in the context of fluctuation monitoring following the partially sequential sampling scheme. For this purpose, we consider the idea of both the usual ranks as well as the sequential ranks using Wilcoxon score. We perform exact study for detailed comparisons between the two procedures. We also provide an interesting illustration using real life data.

6.1.1 Monitoring Structural Changes

Structural stability is one of the most important criteria in modeling the time series data, because very small changes in population parameters disturb the entire model. Consequently, the model fails and the forecasts lose their accuracy. For different time series model, many econometricians study some powerful tests in this regard. Among them, a few are Andrews (1993), Hawkins (1987) and Ploberger et al. (1989). Unfortunately, those tests are only able to detect an instability signal within the dataset of fixed size obtained a priori. Those tests are popularly known as one shot tests.

In the real world, we observe a steady arrival of new data. Therefore we continuously need to check the validity of the prior models since break can occur at any time point. Here such one-shot tests fail because their repeated applications increase the probability of type I error. Chu et al. (1996) discussed how sequential tests efficiently handle such problem. They introduced the use of partial sequential tests for monitoring fluctuations in linear regression model with unknown slope parameter. Nevertheless, the assumption of linearity does not hold in many practical situations. As a result, the modeling and its monitoring become difficult. A more realistic way would be to assume that, when there is a structural stability, data are generated from some unknown continuous population, say F, and if there is any breakdown, the population distribution switches over to, say G. Chu et al. (1996) considered the situation when sampling costs are negligible under null hypothesis of no structural changes. But they used stopping rule which terminates with probability one both under null and alternative as in Wolfe (1977) or Switzer (1983). They pointed out the limitations of existing partially sequential procedures. For that reason, unlike previous works, we, motivated by Siegmund (1985), consider curved stopping boundaries in partially sequential sampling schemes. We indicate how tests with almost power one can be achieved for practical utility.

6.1.2 Motivating Example: Watching Stock Market Dynamics

In the recent days, Indian stock markets become the center of global attraction. Various foreign institutions and non-resident Indians are investing heavily in Indian markets. As a result, some features of the stock market are rapidly changing. National Stock Exchange (NSE) is one of the major stock exchanges in India. Here we have a stock price index number based on fifty major stocks and is popularly known as NIFTY (Natioal Index based on Fifty major stocks). This is the most important index in National Stock Exchange. It has registered a growth of about 200% during the last eight financial years. In the financial year 2005-06, the growth is really mammoth and is about 30%. It is evident from Figure 6.1.1.

During the period April 1, 1998 to March 31, 2000, NSE observed in all 505 trading days. In that phase, Indian economy was in vulnerable state. In the last decade of earlier century, there was lot of changes in central government and their policies. Successive elections and various other internal and external affairs caused that weakness. During those days, we observed 505 data regarding intra-day change in NIFTY. This is the difference between opening and closing value of NIFTY, measured in each day. This intra-day change figures suppose to play a major role in *margin trading, buy today sell tomorrow (BTST) trading and profit brooking.* In a volatile market, this is likely to follow a random pattern rather than trend. A run test on 505 observations shows that the observed number of runs is 248 against the expected number of 253.4515 runs. Thus the observed number of runs is quite close to the expected number of runs. In fact, the observed number of runs indicates that the p-value is 0.6271. Hence, no evidence against randomness is present. So we may conclude that the data, though essentially time sequential in nature, is a random sample from an unknown distribution. Further, the data seems to show heavier tail than normal, as apparent from the normal probability plot (Figure 6.1.2). However the data is more or less symmetric in nature. It can be seen from the density plot (Figure 6.1.3).

From April 1, 2000, the first full budget of a stable government came into effect after a long time. We begin our monitoring at this juncture. Data recorded on or after April 1, 2000 are referred to as second sample observations. We want signal under abnormal volatility pattern or under rapid growth. It is legitimate to think intra-day difference pattern will show some shift in location. We present the analysis of the data and conclusions in subsection 6.5.

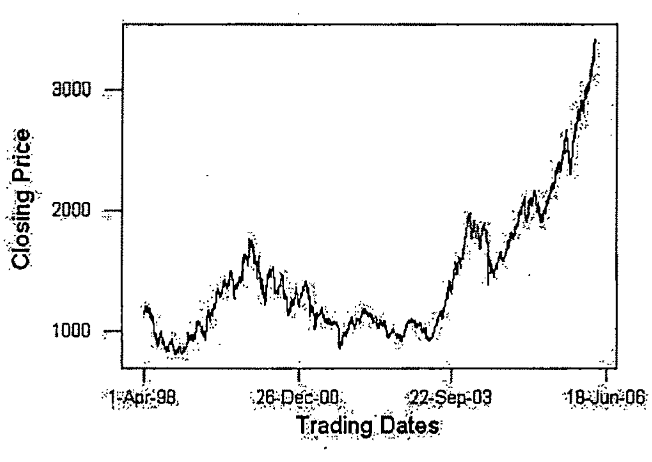


Figure 6.1.1. NIFTY over different Trading Days (Closing Price)

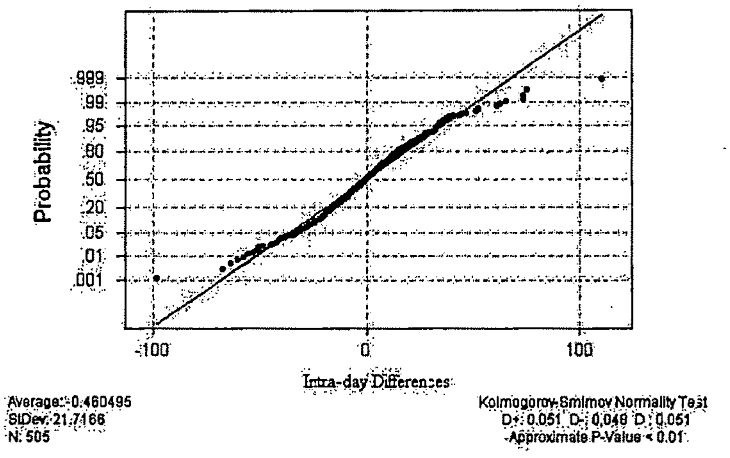


Figure 6.1.2. Normal Probability Plot of the Initial Sample

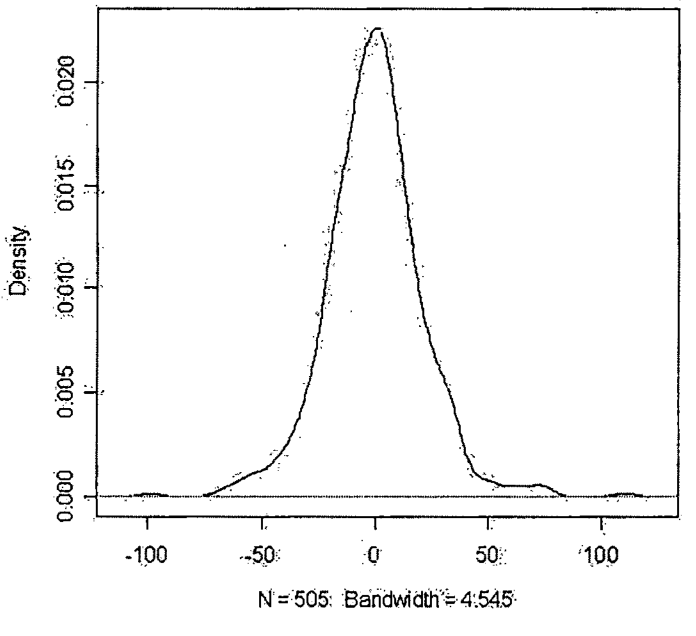


Figure 6.1.3. Nonparametric Density Plot of the Initial Sample

6.1.3 Some other Scopes of Applications

In Monitoring Monsoon Effect on Ground Water Arsenic Level:

Some of the districts of West Bengal in India [See Purkait and Mukherjee(2006)] and Bangladesh are highly arsenic prone zone. Geologists often doubt a change in the level of arsenic contamination in the ground water before and, on or after the monsoon. Arsenic level of some relatively low hit zone increases during or just after the monsoon and the reverse effect for some high hit zone takes place because of water transmission in the monsoon. Here one may consider the pre-monsoon data as the first sample observations of fixed size. Observations, recorded sequentially as soon as monsoon begins, may be referred to as the second sample. The action is required only if there is a significant change in arsenic level. An increase in arsenic level will have an worse impact on the locality while a decrease may affect adjacent areas. Here again we may use a partially sequential stopping rule which leads to a two tailed test with power one.

In Monitoring Consumer Satisfaction:

For comparing the degree of consumer satisfaction of certain product of different brands, inference procedures based on sequential sampling plans can be applied with various advantages. For example, mobile phone is growing its popularity and is used in the recent days among the urban middle class people in India. Service providers, in a competitive environment, always like to monitor whether there is any difference in the level of satisfaction among the users. In this process, historical data of a fixed size may be taken as the first sample. Observations are subsequently recorded at the present condition and may be treated as the second sample. We need to raise an alarm whenever the consumer grievance is on significantly higher side. Otherwise, we allow the monitoring to go on. This is an ideal case for one sided partially sequential power one test based on inverse sampling scheme.

Both the examples involve sensitive issues. Therefore, raising a false alarm is always dangerous. Thus we need to control the type I error as much as possible. We organize the rest of the chapter in the following way. Section 6.2 describes the test procedures. The exact distributions of the test statistics are derived in Section 6.3. Some numerical results based on simulations are discussed in Section 6.4. The proposed procedures are illustrated through NIFTY data in Section 6.5. Section 6.6 concludes.

6.2 Statistical Framework and Some Test Procedures

Here we start with the problem of testing H_0 against H_{B1} as in Chapter 4. With same set up as in Sections 4.2 and 4.3, let $\mathcal{U}_k$ and $\mathcal{V}_k$ be, respectively, the numbers of first sample and second sample observations that are not greater than Y_k. Then, for fixed Y_k, $\mathcal{U}_k$ follows binomial distribution with parameters (m, θ_k) and $\mathcal{V}_k$ follows binomial distribution with parameters $(k-1, \theta_k^*)$, independently of each other, where

$$\theta_k = F(Y_k) \quad \text{and} \quad \theta_k^* = G(Y_k) = F(Y_k - \delta).$$

Now, writing $\mathcal{Z}_k = \mathcal{U}_k + \mathcal{V}_k$, and hence

$$\tilde{S}_n = \sum_{k=1}^{n} H_{m+k-1}(Y_k) = \sum_{k=1}^{n} \frac{1}{m+k-1} \mathcal{Z}_k,$$

stopping rule corresponding to (4.3.2) becomes

$$N = \min\left\{n : \tilde{S}_n \geq r/2\right\}.$$

Thus the level condition for the exact N-test reduces to the following inequality

$$P_{H_0}(\tilde{S}_{N_\alpha - 1} \geq r/2) \leq \alpha < P_{H_0}(\tilde{S}_{N_\alpha} \geq r/2). \tag{6.2.1}$$

Given r and α, (6.2.1) yields the cut off point N_α for the exactly distribution free test at level α, and further the exact power of the test is

$$P_\delta(N < N_\alpha) = P_\delta(\tilde{S}_{N_\alpha} \geq r/2).$$

Similar development can be given by using the M-test.

The sequential sampling scheme with curved stopping boundary corresponding to Wolfe (1977) and Orban and Wolfe (1980) for testing H_0 against H_{B1} may be described by the following stopping variable:

$$M_1^c = \min\left\{n : n \geq \eta, \sum_{k=1}^{n}(F_m(Y_k) - \frac{1}{2}) \geq \varphi\sqrt{n}\right\}, \tag{6.2.2}$$

where φ is a suitable constant and $\eta > 4\varphi^2$ is an integer. The partially sequential test then becomes: Stop sampling at $min(M_1^c, m_t)$, and reject H_0 iff $M_1^c < m_t$, where m_t is the maximum number of second sample observations desired to observe during monitoring time t. For example, if we consider daily data and plan to monitor for one year, we set m_t at 365. This is a nonparametric partially sequential version of repeated significance test described in Siegmund (1977,1985). For testing H_0 against $\tilde{H}_{B1}$, an appropriate stopping rule will become:

$$M_2^c = \min\left\{n : n \geq \eta, max(\sum_{k=1}^{n}(F_m(Y_k) - \frac{1}{2}), \sum_{k=1}^{n}(\overline{F}_m(Y_k) - \frac{1}{2})) \geq \varphi\sqrt{n}\right\}, \tag{6.2.3}$$

where, as in other chapters, $\overline{F}_m(Y_k) = 1 - F_m(Y_k)$. (6.2.3) is also equivalent to

$$M_2^c = \min\left\{n : n \geq \eta, |\sum_{k=1}^{n}(F_m(Y_k) - \frac{1}{2})| \geq \varphi\sqrt{n}\right\}. \tag{6.2.4}$$

Consequently, the partially sequential repeated significance test becomes: Stop sampling at $min(M_2^c, m_t,)$and reject H_0 iff $M_2^c < m_t$. Similar development can be made for the sequential rank based procedures. For testing H_0 against H_{B1}, we have the stopping rule

$$N_1^c = \min\left\{n : n \geq \eta, \sum_{k=1}^{n}(H_{m+k-1}(Y_k) - \frac{1}{2}) \geq \varphi\sqrt{n}\right\}, \tag{6.2.5}$$

where φ and η are as before. Now it is easy to see that this stopping rule may be expressed equivalently as:

$$N_1^c = \min\left\{n : n \geq \eta, T_n \geq \frac{n}{2} + \varphi\sqrt{n}\right\}.$$

The partially sequential repeated significance test then becomes: Stop sampling at $min(N_1^c, n_t)$ and reject H_0 iff $N_1^c < n_t$. Interpretation of n_t is similar to that of m_t. Obviously the type I error of the test will be given by $P_{H_0}[N_1^c < n_t]$. When we have both sided alternative $\tilde{H}_{B1}$ we use the stopping rule:

$$N_2^c = \min\left\{n : n \geq \eta, \sum_{k=1}^{n} \mid (H_{m+k-1}(Y_k) - \frac{1}{2}) \mid \geq \varphi\sqrt{n}\right\}$$

with the test criterion as: Stop sampling at $min(N_2^c, n_t)$ and reject H_0 iff $N_2^c < n_t$. This stopping rule may be alternatively expressed as:

$$N_2^c = \min\left\{n : n \geq \eta, \mid T_n - \frac{n}{2} \mid \geq \varphi\sqrt{n}\right\}.$$

Here the type I error is obviously $P_{H_0}[N_2^c < n_t]$. Similar are the developments based on the M_1^c and the M_2^c.

Note that, in developing repeated significance tests, we use no prefixed r as in the existing partially sequential stopping rules. Therefore, in the proposed tests, level conditions for the existing tests, as in equation (2.6.1), will not work. Here the level will depend on m, φ and m_t or n_t as the case may be. However experimenter can only choose φ. Other two factors are usually given. We later show that a higher value of φ ensures lower level. But, as far as exact tests are concerned, such higher values of φ diminish the rate of convergence of power to unity. Again, as m_t (or n_t) increases, type I error increases. If m_t (or n_t) is taken as ∞, which is common in several monitoring problem, power of the tests will almost surely be one . However, a question of type I error rate will arise as it increases with t. We prefer the test where the rate is slow.

6.3 Some Exact Results

In previous chapters, exact behavior of the stopping variable N is presented using simulation studies only. In the present context we introduce a different algorithm, following the line of Fligner and Wolfe (1976), for exact treatment of the stopping variable N or that of N_1^c or N_2^c. For this, we first consider the conditional distribution of Z_k given Y_k. This is given by

$$\begin{aligned} & P(\mathcal{Z}_k = z|Y_k) \\ = \; & P(\mathcal{U}_k + \mathcal{V}_k = z|Y_k) \\ = \; & (1-\theta_k)^m (1-\theta_k^*)^{k-1} \sum_{\ell=\max(0,z-k+1)}^{\min(z,m)} \binom{m}{\ell}\binom{k-1}{z-\ell}\left(\frac{\theta_k}{1-\theta_k}\right)^z \left(\frac{\theta_k^*}{1-\theta_k^*}\right)^{z-\ell}, \end{aligned}$$

where $z = 0, 1, ..., m+k-1$. As $\theta_k = \theta_k^*$ under H_0, the conditional distribution of $\mathcal{Z}_k$, given Y_k, reduces to binomial with parameters $(m+k-1, \theta_k)$ under H_0. Further the unconditional distribution of $\mathcal{Z}_k$ is given by

$$\begin{aligned} & P(\mathcal{Z}_k = z) = E\{P(\mathcal{Z}_k = z|Y_k)\} \\ = & \int_{-\infty}^{\infty} \left\{ (1-\theta_k)^m (1-\theta_k^*)^{k-1} \times \sum_{\ell=\max(0,z-k+1)}^{\min(z,m)} \binom{m}{\ell}\binom{k-1}{z-\ell}\left(\frac{\theta_k}{1-\theta_k}\right)^z \left(\frac{\theta_k^*}{1-\theta_k^*}\right)^{z-\ell} \right\} \\ & dG(y_k), \end{aligned}$$

which, under H_0, reduces to

$$\begin{aligned} P(\mathcal{Z}_k = z) &= \int_{-\infty}^{\infty} \binom{m+k-1}{z} \theta_k^t (1-\theta_k)^{m+k-z-1} dG(y_k) \\ &= \binom{m+k-1}{z} B(z+1, m+k-z) = \frac{1}{m+k} \end{aligned}$$

for all $z = 0, 1, ..., m+k-1$. That means, under H_0, $\mathcal{Z}_k$ follows the discrete uniform distribution over $(0, m+k-1)$. It is also seen that, given $(Y_1, \cdots, Y_n)$, $(\mathcal{Z}_1, \cdots, \mathcal{Z}_n)$ has multinomial distribution, and hence, under H_0, the unconditional distribution of

$(\mathcal{Z}_1, \cdots, \mathcal{Z}_n)$ reduces to

$$P(\mathcal{Z}_1 = z_1, \mathcal{Z}_2 = z_2, \cdots, \mathcal{Z}_n = z_n) = \begin{cases} \frac{1}{(m+1)(m+2)\cdots(m+n)}, & z_i = 0, 1, \cdots, m+i-1; \\ 0, & \text{otherwise,} \end{cases}$$

$i = 1, 2, ..., n$. This implies, under H_0,

$$P(\tilde{S}_n = \tilde{s}_n) = \frac{\#\text{ independent choices of } (z_1, \cdots, z_n) \text{ such that } \sum_{k=1}^{n} \frac{z_k}{m+k-1} = \tilde{s}_n}{(m+1)(m+2)\cdots(m+n)},$$

where # means *number of*. Hence we can find

$$P_{H_0}(\tilde{S}_n \geq r/2) = \sum_{\tilde{s}_n \geq r/2} P_{H_0}(\tilde{S}_n = \tilde{s}_n)$$

after calculating the probabilities at various $\tilde{s}_n \geq r/2$.

Using the above in (6.2.1), we can compute the cut-off point for the N-test. However, under any alternative, the exact distribution of $\tilde{S}_n$ is complicated in nature and can only be evaluated for very small m and r when the probability density function (p.d.f.) of Y is simple and known.

These results can be easily extended to the distribution of N_1^c replacing r by $(n_t + 2\varphi n_t^{1/2})$. Note that, here we have no concept of so called cut off points. However we can calculate the type I error or the $p-value$ as a function of t. This is $P[\tilde{S}_{n_t} < \frac{n_t}{2} + \varphi n_t^{1/2}]$. For the test based on N_2^c, we have the $p-value$ as $P[\mid \tilde{S}_{n_t} - \frac{n_t}{2} \mid < \varphi n_t^{1/2}]$.

For the Orban-Wolfe procedure, we have

$$S_n = \sum_{k=1}^{n} F_m(Y_k) = \frac{1}{m} \sum_{k=1}^{n} \mathcal{U}_k,$$

where, for fixed Y_k, $\mathcal{U}_k$ follows binomial distribution with parameters (m, θ_k), and thus, unconditionally, $\mathcal{U}_k$ follows discrete uniform distribution over $(0, m)$ under H_0. Hence, by the same argument as before, we have, under H_0,

$$\begin{aligned} P(S_n = \tilde{s}) &= \frac{\#\text{ independent choices of } (u_1, \cdots, u_n) \text{ for which } \sum_{k=1}^{n} u_k = m\tilde{s}}{(m+1)^n} \\ &= \frac{\text{coefficient of } t^{m\tilde{s}} \text{ in } (1 + t + t^2 + \cdots + t^m)^n}{(m+1)^n}. \end{aligned}$$

Using the above, we can find the cut-off point for the M-test. Now, under any alternative, unconditionally,

$$P(\mathcal{U}_k = u) = E\ P(\mathcal{U}_k = u|Y_k) = \int_{-\infty}^{\infty} \binom{m}{u} p_k^u (1-p_k)^{m-u} dG(y_k).$$

Consequently, the power is

$$P_H(S_{M_\alpha} \geq r/2) = \sum_{\tilde{s} \geq r/2} P_H(S_{M_\alpha} = \tilde{s}).$$

Developments for the statistics M_1^c and M_2^c are also straightforward and are omitted to save space.

6.4 Some Monte-Carlo Studies

The distributions of the various stopping variables proposed in Section 6.2 can be exactly obtained using the algorithm presented in Section 6.3. However, in general, distributional forms are not very simple. But these can be empirically obtained through some Monte Carlo experiment. Here we carry out simulation studies to enumerate type I error rates and powers under different possible situations.

At the very outset we consider the usual rank based test. Without loss of generality, we choose a fixed sample of size m from the standard normal population. We present the results for $m = 25, 50, 75$ and 100. We further choose an wide range of values of φ, starting from 0.5, in the stopping rule described by (6.2.4). The second sample observations are then sequentially drawn from the standard normal population. We record at least η second sample observations. Here we choose $\eta = \max(10, 4\varphi^2)$. We note the percentage of false signals of instability in 10000 replications of Monte-Carlo experiment at various stages. These are the simulated values of the probability of type I error and we compute these at $m_t = 100, 250, 500, 750, 1000$. Note that m_t is the maximum number of second sample observations desired to be drawn. Thus, even if we consider a daily data, $m_t = 1000$ will imply that a monitoring is planed for about 2.75 years. This will adequately serve most of the practical purposes. We present our findings in Table 6.4.1.

From Table 6.4.1, we see that, for a given m and φ, type I error sharply increases with the delay in inspection time. When $m = 25$ and $\varphi = 2$, the probability of raising a false alarm is only 0.2% after 100 inspections of second sample but, after 1000 inspections, the probability is as high as 29.4%. Such a rapid growth (see Figure 6.4.1) of percentage of unnecessary instability signal is really a big limitation of the usual rank procedure. One way of controlling such a positive hazard is to choose a larger value of φ. To keep this probability below 5% level even after 1000 inspections, we may take $\varphi = 4$, say, when $m = 25$. For $m = 50, 75$ and 100, we may select $\varphi = 3, 2.5$ and 2 to attain that level condition. Given m, it is easy to obtain the optimal φ for which a prefixed level will be attained for a maximum desirable inspection number of second sample observations. But those points will not have any practical utility, and hence is not presented. Further we note that, as m increases, the type I error decreases for a given φ. As for example, if $\varphi = 2$, and $m_t = 1000$, the type I errors are, respectively, 13.8, 7.2 and 3.7 percent for $m = 50, 75$ and 100, compared to 29.4 percent when $m = 25$. This indicates that the type I error rate can be checked choosing a larger number of first sample observations. But, in practice, first sample is given a priori. So a larger choice of φ seems to be the only way out. However larger value of φ will reduce the capability of early detection of instability in presence of some fluctuations. To verify this, we compute powers corresponding to the various combinations used in Table 6.4.1. Here we also consider the first sample from $N(0, 1)$. The second sample observations are sequentially recorded from $N(\delta, 1)$. To save space only two choices of δ viz. 0.5 and 1.0 are considered. The results are summarized in Table 6.4.2. The situation for $m = 25$ is also described in Figure 6.4.2.

Table.6.4.1. Simulation Results on the Type I Error Rate for the Usual Rank Based Test

	m_t / φ	100	250	500	750	1000
m=25	0.5	0.597	0.769	0.852	0.889	0.909
	1.0	0.167	0.358	0.510	0.591	0.642
	2.0	0.002	0.045	0.143	0.227	0.294
	3.0	0.000	0.002	0.024	0.064	0.106
	4.0	0.000	0.000	0.002	0.011	0.030
m=50	0.5	0.596	0.752	0.834	0.870	0.891
	1.0	0.069	0.211	0.356	0.441	0.500
	2.0	0.000	0.005	0.039	0.088	0.138
	3.0	0.000	0.000	0.002	0.010	0.025
m=75	0.5	0.451	0.649	0.764	0.817	0.850
	1.0	0.040	0.131	0.261	0.352	0.419
	2.0	0.000	0.001	0.014	0.041	0.072
	2.5	0.000	0.000	0.002	0.011	0.023
m=100	0.5	0.412	0.611	0.727	0.789	0.822
	1.0	0.029	0.101	0.205	0.287	0.350
	2.0	0.000	0.000	0.005	0.018	0.037

From Table 6.4.2, we see quite expectedly that the power tends to one as m_t or δ increases. However the power sharply decreases as soon as φ increases. Consider the situation for $m = 25$. We see, when $\varphi = 2$ and $\delta = 0.5$, the probabilities of detecting the shift using at most 100, 250, 500, 750 and 1000 second sample observations are, respectively, 17, 60, 82, 89 and 92 percent. But these percentages drop down to 0, 2, 24, 45 and 59 respectively when $\varphi = 4$. So, in case we use daily data, we will be able to detect correctly a shift only in 2% cases even after 8 months (about 250 observations will come in this phase). This is really ridiculous from practical point of view. Now we see how the things can be managed just using the updated stopping rule based on sequential ranks.

Table 6.4.2. Simulation Results on Power for Usual Rank Based Test

	m_t		100		250		500		750		1000	
m	φ	$\delta \rightarrow$	0.5	1.0	0.5	1.0	0.5	1.0	0.5	1.0	0.5	1.0
25	0.5		0.95	1.00	0.98	1.00	0.99	1.00	0.99	1.00	$\simeq 1$	1.00
	1.0		0.77	1.00	0.91	1.00	0.95	1.00	0.97	1.00	0.97	1.00
	2.0		0.17	0.88	0.60	0.99	0.82	1.00	0.89	1.00	0.92	1.00
	3.0		0.00	0.23	0.19	0.92	0.54	0.99	0.70	1.00	0.79	1.00
	4.0		0.00	0.00	0.02	0.56	0.24	0.95	0.45	0.99	0.59	1.00
50	0.5		0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.0		0.82	1.00	0.97	1.00	0.99	1.00	0.99	1.00	$\simeq 1$	1.00
	2.0		0.10	0.92	0.63	1.00	0.88	1.00	0.95	1.00	0.97	1.00
	3.0		0.00	0.16	0.12	0.97	0.55	1.00	0.77	1.00	0.86	1.00
75	0.5		0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.0		0.85	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2.0		0.07	0.95	0.64	1.00	0.92	1.00	0.97	1.00	0.99	1.00
	2.5		0.00	0.62	0.30	1.00	0.79	1.00	0.92	1.00	0.96	1.00
100	0.5		0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.0		0.86	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	2.0		0.06	0.88	0.66	0.99	0.95	1.00	0.99	1.00	$\simeq 1$	1.00

From Table 6.4.3, we see that the type I error rates of the sequential rank based procedure are significantly lower than the corresponding usual rank based procedure for $\varphi = 0.5$ and 1 respectively. For $\varphi = 1$, the sequential rank procedure maintains remarkably low type I error even when the initial sample size is as small as 25 and m_t is as high as 1000. Here we should clearly explain why Table 6.4.3 has fewer rows than Table 6.4.1. The reason is simple. With these specific choices of m and m_t, a higher value of φ, say $\varphi \geq 2$, will produce virtually zero probability for the type I error. Practically we do not need to proceed to that extent. We compare type I error rates of the two procedures for $m = 100$ and 25 taking $\varphi = 1$ in Figures 6.4.3 and 6.4.4 respectively.

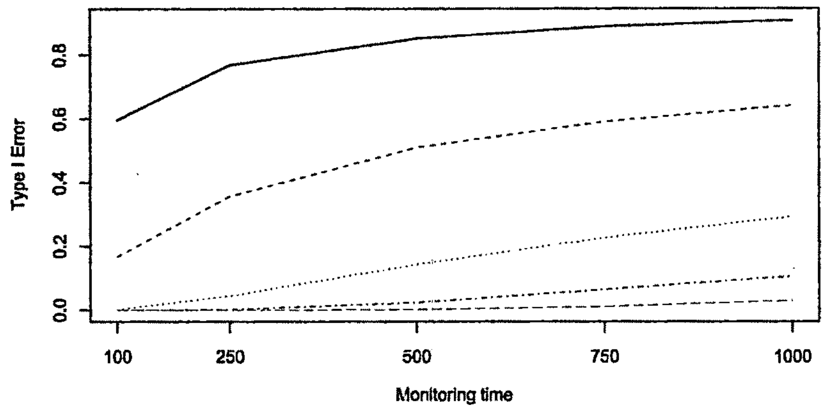


Figure 6.4.1. Type I Error Rate for different φ when $m = 25$ (usual rank test)

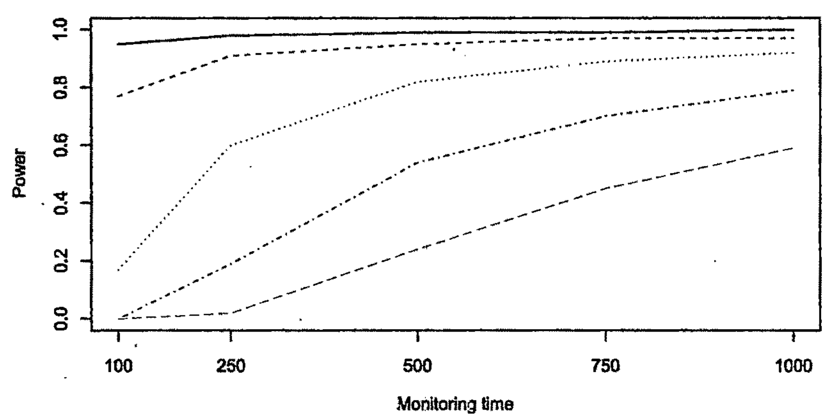


Figure 6.4.2. Power for the different φ when $m = 25$ and $G \sim$ N(0.5,1) (usual rank test)

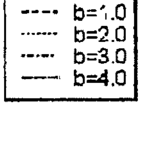


Table 6.4.3. Simulation Results on the Type One Error Rate for the Sequential Rank Based Test

	m_t / φ	100	250	500	750	1000
m=25	0.5	0.427	0.523	0.586	0.622	0.647
	1.0	0.008	0.011	0.013	0.015	0.016
m=50	0.5	0.417	0.518	0.580	0.610	0.632
	1.0	0.007	0.010	0.012	0.013	0.014
m=75	0.5	0.416	0.517	0.577	0.610	0.632
	1.0	0.006	0.009	0.011	0.012	0.013
m=100	0.5	0.413	0.514	0.577	0.610	0.632
	1.0	0.006	0.009	0.011	0.012	0.013

Figures 6.4.3 and 6.4.4 themselves show the benefit of the sequential rank procedure in controlling type I error. Further Table 6.4.4 shows that the power of the test sharply tends to one though little less than the corresponding $M-test$ for a given φ. However, for $m = 100$, if we need to keep the type I error below 5% level at least up to $m_t = 1000$, we take $\varphi = 2$ for the $M - test$, and $\varphi = 1$ for the $N - test$ test. Now we may compare Tables 6.4.2 and 6.4.4. We see that the power of early detection, say, within 100 or 250 second samples, is better in the sequential rank based procedure. In an advanced stage the power of both the tests again become unity. Only in the middle phase the usual rank based test has slightly better power and that too at the expense of heavy type I error rate. Thus we may safely use the sequential rank procedure for all practical purposes. Particularly, when initial sample size is small, the sequential rank procedure is the safest way to check the exploration of type I error.

Table 6.4.4. Simulation Results on Power for Sequential Rank Based Test

	m_t		100		250		500		750		1000	
m	φ	$\delta \rightarrow$	0.5	1.0	0.5	1.0	0.5	1.0	0.5	1.0	0.5	1.0
25	0.5		0.89	1.00	0.94	1.00	0.95	1.00	0.96	1.00	0.96	1.00
	1.0		0.21	0.86	0.27	0.91	0.30	0.93	0.31	0.94	0.32	0.94
50	0.5		0.96	1.00	0.99	1.00	0.99	1.00	0.99	1.00	$\simeq 1$	1.00
	1.0		0.38	0.99	0.55	1.00	0.62	1.00	0.64	1.00	0.65	1.00
75	0.5		0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.0		0.53	1.00	0.74	1.00	0.82	1.00	0.84	1.00	0.86	1.00
100	0.5		0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.0		0.62	1.00	0.85	1.00	0.92	1.00	0.93	1.00	0.94	1.00

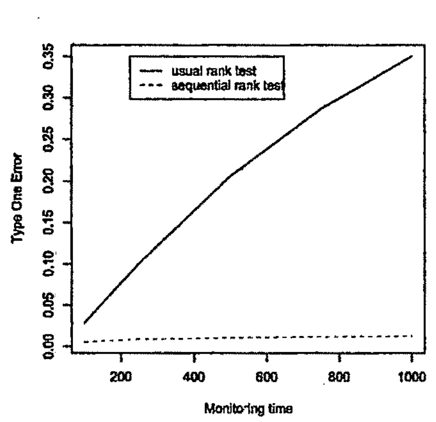


Figure 6.4.3. Comparison of Type I Error Rate for two procedures when $m = 100, \varphi = 1$

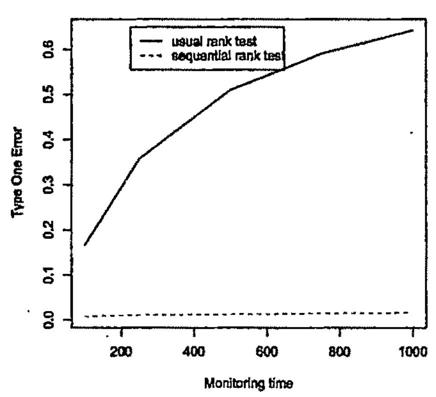


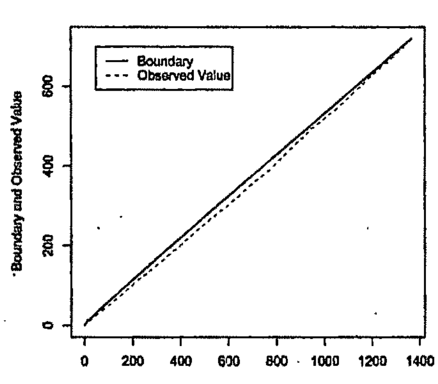
Figure 6.4.4. Comparison of Type I Error Rate for two procedures when $m = 25, \varphi = 1$

6.5 A Data Study

Here we discuss, in detail, the monitoring of the intra-day changes in NIFTY. We describe the data and the problem in section 6.1.3. We observe the value $\aleph_n = \max(S_n, n - S_n)$ for the $M - test$ and $\tilde{\aleph}_n = \max(\tilde{S}_n, n - \tilde{S}_n)$ for the $N - test$, and compare it with $(\frac{n}{2} + \varphi\sqrt{n})$ at each stage. We stop sampling and conclude that a shift in location takes place in the distribution of intra-day change pattern. We may set a time frame till which we will continue monitoring. But, in practice, we do not require any prefix m_t, where the cost of sampling is not serious. This is same as the choice $m_t = \infty$. If the observed value of $\aleph_n$ or $\tilde{\aleph}_n$ exceeds $(\frac{n}{2} + \varphi\sqrt{n})$ at any juncture, we stop sampling to search the reason for the shift.

Here we do not like to infer anything within the first quarter of monitoring. There are 61 trading days in the first quarter (April-June'00). So we take $\eta = 61$. For our inspection, we like to choose $\varphi = 1.0$. Since we have sufficiently large volume of data, this choice of φ is expected to work satisfactorily. Figure 6.5.1 illustrates the monitoring and boundary crossing for the usual rank procedure. We see that $\aleph_n$ signals an alarm after 1369 working days on Sept 6, 2005. However, at that time, past months growth rate of NIFTY was usual and about 3 to 4.5%. The quarterly growth rate of NIFTY was about 10%. Moreover there was no threat to economy. This is certainly a false alarm caused by the usual rank procedure.

But suppose we use the sequential rank procedure described in Figure 6.5.2. Then



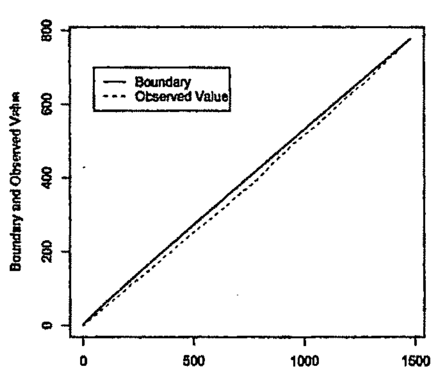


Figure 6.5.1. Monitoring Based on Usual Rank Procedure

Figure 6.5.2. Monitoring Based on Sequential Rank Procedure

the signal comes only on February 21, 2006 which is the trading day number 1480 from the beginning of monitoring. Really, that was a booming phase. At that phase, NIFTY has featured an increasing trend with steep positive slope (see the Figure 6.1.1). Then the quarterly growth rate of NIFTY was around 15% and the monthly growth rate was about $4.8 - 6.6\%$. Just few days before the signal day, NIFTY crosses 3000 mark for the first time in history. Moreover, 28th of February is the budget day in India and everyone was expecting a market friendly budget to come out. This is a nice detection of a genuine shift of intra-day trading pattern.

Some may argue that the usual rank test detected the change five and half months earlier than the sequential rank procedure. The latter only delays the detection and the price is paid for trying to control the type I error rate. But we see that any detection of shift in share price problems leads to either of the two situations. A possible right shift increases the confidence of the intra-day traders while a left shift depresses them to a large extent. In either case, the market may become highly volatile and such high volatility hampers industrial growth. That is why this kind of detection is always sensitive. One must take some considerable time to reach a final decision, and hence we make ourselves very sure before rejecting the null hypothesis. The usual rank procedure always leaves some room for doubt. But the sequential rank procedure is more conservative as it controls the type I error to a great extent and this procedure raises an alarm only when there is sufficient reason.

6.6 Concluding Remarks

Here, unlike all the previous works, some partially sequential sampling schemes are introduced using a curved stopping boundary. In subsection 6.4 we discuss how sequential rank based test effectively controls the type I error rates. However, in that section, various computational results are presented only for two sided tests. But the type I error rates for one sided tests can be easily obtained from Tables 6.4.1 and 6.4.3 for the usual rank based test and the sequential rank based test respectively. For a given combination of m, φ and m_t, the type I error corresponding to each one sided test is about half of the type I error of the corresponding two sided test. This is because of symmetry of the null distributions of the random variables associated with the stopping variables. This can also be easily verified empirically using simulations. The derivations of the asymptotic behaviours of the proposed procedures require various results related to stochastic processes and we leave this for future study. But the powers of such one sided tests, as in type I errors, cannot be obtained from the corresponding powers of the both sided tests. However, one-sided analogues of both the usual rank procedure and the sequential rank procedure are easily developed and the comparison between them are similar.

Finally, it is worth mentioning that for time sequential procedures sometimes shifts occur at unknown time points not at the beginning of the sampling. We see in Chapter 4 that, for linear stopping boundary, sequential rank test can even improve power in such occasion if a shift occurs at a later stage. The same result can also be empirically established in case of curved stopping rules.

In the light of this, we may mention that the two procedures can simultaneously be used in case of an industrial monitoring of the quality of a product. Then any signal raised by the usual rank based test may be compared with the warning limit in statistical quality control. Similarly, an analogy may be drawn between the action limit and the signal of instability by the sequential rank based test. From our illustration in section 6.5, one may also think that the early signal of shift, produced by the usual rank based test, is just an alert. The true shift is detected by the sequential rank based test.

In this connection, it is also possible to provide some asymptotic results by assuming that for each m, there exists a positive integer $\eta = \eta(m)$ such that, as $m \to \infty$, $\eta \to \infty$ but $\frac{\eta}{m} \to \xi_0 \in [0, \infty)$. Then using standard techniques one can easily see that the usual rank procedure can be approximated by a Wiener process. This approximation may be used to calculate the associated boundary crossing probabilities. However the sequential rank procedure can only be approximated by a Gaussian process. Here boundary crossing probabilities are not so straightforward. This is itself another research problem and we omit it at present.

It will be very interesting to combine the two approaches into a single inference procedure by following the usual rank procedure until it sounds an *early alert* and then switching over to the partially sequential rank procedure for the final detection of structural change. In that case we need to develop the properties of the partially sequential procedure conditionally on the previous *early alert* from the usual rank procedure. We leave this problem for future research.