Monitoring of stationary and non-stationary processes: on - line detection of the structural shifts

E.A.Grebenuk[1]

[1] Institute of Control Sciences,
Profsoyuznaya, Moscow 117997,
Tel: 495-334-7640, Russia
(e-mail: lngrebenuk@rambler.ru)

Abstract. This paper is devoted to the problem of detection of the structural shifts in non-stationary processes. The procedures of the change detection and diagnosing the type of the change for difference-stationary processes integrated of the first-order is offered.

Keywords. structural changes, model estimation, non- stationary processes, statistics, change detection

1 Introduction

A great number of time series emerging while observing financial, geophysical, or ecological processes are non stationary. The paper is devoted to the study of a class of non stationary random processes, usually referred to as difference-stationary processes or integrated processes of the order d, I(d).[1]. Both the analysis and the forecast of the process behavior are available if the chosen model is adequate to the process. If the process goes through changes and the model does not describe the process adequately any more, it is necessary to detect the changes as soon as possible. In this case, one may introduce the corrections to the initial model and thus improve the quality of the forecast of the process behavior. For example, the processes of the changes with time in the pollutant material concentrations that emerge in the problems of the ecological monitoring can be described by autoregressive models. These processes can become non stationary in some time intervals, especially if the general ecological situation deteriorates. The further deterioration of the ecological situation may be accompanied by increasing the drift of the process or by the increasing the its stochastic trend. Meanwhile the improvement and stabilization of the ecological situation would return the process to the stationary state. A timely detection of the changes in the process and the corrections made to the model would make it possible to increase the precision in the estimates of the probabilities that the process would turn into a non stable one as well as to improve the detections of the transformations themselves. The following types of changes are investigated:

- The changes of the drift, the changes of the stochastic trend and the changes in the type of a process from a non stationary to a stationary process, in case a non stationary process are detected.
- The changes in the type of process from stationary to a non stationary one, in case a stationary process are detected.

The statistics of the process observations have been designed and its properties were analyzed. It is shown that investigated types of changes are the designed statistics changes: its mean and variance. These changes may be detected by means of the on-line detection algorithms, such as cumulative sum (CUSUM) charts [2-3], exponentially weighted moving average (EWMA) charts, etc. In report is offered procedure of the change detected and diagnosing the type of the change for difference-stationary processes integrated of the first-order.

2 Problem statement

Let the initial non- stationary process $y_1, y_2, \dots, y_t$, is integrated of the first order and described by model:

$$y_t = \mu + \sum_{i=1}^{k} \alpha_i y_{t-i} + \varepsilon_t, \quad (0)$$

where y_{t-i} - observation for the process in time $t-i$, ε_t - a sequence of independently and identically distributed random variable with normal distribution $N(0, \sigma_o^2)$, where coefficients $\alpha_i, i = 1, \ldots, \alpha_k$ in condition $\sum_{i=1}^{k} \alpha_i = 1$, μ-constant, defining the drift value. Consider the following changes of the process (1).

1) The drift or coefficient change at time $t = t_p + 1$ without type changes of the process (it remain non- stationary):

$$y_t = \begin{cases} \mu + \sum_{i=1}^{k} \alpha_i y_{t-i} + \varepsilon_t, & \text{if } t \le t_p \\ \mu_1 + \sum_{i=1}^{k} \alpha_i' y_{t-i} + \varepsilon_t, & \text{if } t > t_p \end{cases}, \quad (2)$$

where $\sum_{i=1}^{k} \alpha_i = \sum_{i=1}^{k} \alpha_i' = 1$, $\mu_1 \ne \mu$ constant.

2) The stochastic trend change at time $t = t_p + 1$:

$$y_t = \begin{cases} \mu + \sum_{i=1}^{k} \alpha_i y_{t-i} + \varepsilon_t, & \text{if } t \le t_p \\ \mu + \sum_{i=1}^{k} \alpha_i y_{t-i} + \eta_t, & \text{if } t > t_p \end{cases}, \quad (0)$$

where $E\varepsilon_t = E\eta_t = 0,\ Cov(\varepsilon_t, \eta_{t+j}) = 0,\ j = 0,1,2,\ldots, Cov(\eta_t, \eta_{t+k}) = 0,\ k = 1,2,\ldots,$

$\text{var}(\varepsilon_t) = \sigma_0^2,\ \text{var}(\eta_t) = \sigma_1^2 \ne \sigma_0^2$.

3) The process type change from non-stationary to stationary:

$$y_t = \begin{cases} \mu + \sum_{i=1}^{k} \alpha_i y_{t-i} + \varepsilon_t, & \text{if } t < t_p \\ \mu + \sum_{i=1}^{k} \alpha_i' y_{t-i} + \varepsilon_t, & \text{if } t \ge t_p \end{cases}, \quad (0)$$

where $\sum_{i=1}^{k} \alpha_i = 1,\ \sum_{i=1}^{k} \alpha_i' < 1$.

It is need to detect changes (2)-(4) of the process (1) in time receiving of observations and define the type of these changes.

3 Statistics construction for changes detection and diagnostics

Process $y_1, y_2, \ldots, y_t$, one can consider as the process with regard to its first differences:

$$y_t = \mu + (\alpha_1 + \alpha_2 + \ldots \alpha_k) y_{t-1} + \alpha_2 (y_{t-2} - y_{t-1}) + \ldots + \alpha_k (y_{t-k} - y_{t-1}) + \varepsilon_t = \mu + y_{t-1} +$$
$$(\alpha_2 + \alpha_3 + \ldots + \alpha_k) \Delta y_{t-1} + \ldots \alpha_k \Delta y_{t-k+1} + \varepsilon_t.$$

The equation (1) is transformed to equation:

$$\Delta y_t = \mu + \sum_{i=1}^{k-1} \beta_i \Delta y_{t-i} + \varepsilon_t , \tag{5}$$

where $\Delta y_{i-i} = y_{t-i} - y_{t-i-1}$, $i = 0,...,k$, $\beta_i = \sum_{j=i+1}^{k} \alpha_j$ $, i = 1,\ldots,k-1$, ε_t and μ such as in equation (1). The process (5) is stationary, with expectation value and variance:

$$E(\Delta y_t) = \frac{\mu}{1 - \sum_{i=1}^{k-1} \beta_i}, \quad D(\Delta y_t) = \sigma^2_{\Delta y} .$$

After changes, describing with models (2) - (4), the process models in the first differences in time $t = t_p + 1$ are:

$$\Delta y_t = \mu_1 + \sum_{i=1}^{k-1} \beta_i' \Delta y_{t-i} + \varepsilon_t , , \tag{2a}$$

where $\beta_i' = \sum_{j=i+1}^{k} \alpha_j'$ $, i = 1,\ldots,k-1$ $, \mu_1 \neq \mu$;

$$\Delta y_t = \mu + \sum_{i=1}^{k-1} \beta_i \Delta y_{t-i} + \eta_t , \tag{3a}$$

where the coefficients $\beta_i, i = 1,\ldots,k-1$ are defined in equation (5);

$$\Delta y_t = \mu + \beta_0 y_{t-1} + \sum_{i=1}^{k-1} \beta_i' \Delta y_{t-i} + \varepsilon_t , , \tag{4a}$$

where $\beta_0 = \sum_{j=1}^{k} \alpha_j' - 1$ $\beta_i' = \sum_{j=i+1}^{k} \alpha_j'$ $, i = 1,\ldots,k-1$.

Let $\tilde{\mu}$ and $\tilde{\beta}_i, i = 1,\ldots,k-1$ are the OLS estimates of the parameters μ and $\beta_i, i = 0,\ldots,k-1$ for models (5), calculated by observations $\Delta y_{t-k+1}, \Delta y_{t-k+2}, \ldots$. The statistics

$$ST_t = \Delta y_t - \tilde{\mu} - \tilde{\beta}_0 y_{t-1} - \sum_{j=1}^{k-1} \tilde{\beta}_j \Delta y_{t-j}$$

was constructed and its property changes after process changes was investigate.

4 Statistics properties

On the assumption for model (1) the OLS estimates of the parameters μ and $\beta_i, i = 1,\ldots,k-1$ are the consistent and asymptotically normal:

$$\sqrt{N}(\theta - \tilde{\theta}) \sim N(0, \sigma_\varepsilon^2 D^{-1}(\theta)) ,$$

where N is sample volume for estimation, $\theta^T = (\mu, \beta_1, \ldots, \beta_{k-1})$ and $\tilde{\theta}^T = (\tilde{\mu}, \tilde{\beta}_1, \ldots, \tilde{\beta}_{k-1})$ - vectors of the actual and estimated parameter values, accordingly; $D^{-1}(\theta)$ covariance matrix of the process (5), T" - transposition sign. Let $\Delta\mathbf{Y}^T_{t-k+1} = (1, \Delta y_{t-1}, \ldots, \Delta y_{t-k+1})$. In such designations the model (5), its estimate, $\Delta\tilde{y}$, the statistics ST_t without change are written as:

$$\Delta y_t = \theta^T \Delta\mathbf{Y}_{t-k+1} + \varepsilon_t$$

$$\Delta\tilde{y}_t = \tilde{\boldsymbol{\theta}}^T \Delta\mathbf{Y}_{t-k+1},.$$

$$ST_t = \boldsymbol{\theta}^T \Delta\mathbf{Y}_{t-k+1} - \tilde{\boldsymbol{\theta}}^T \Delta\mathbf{Y}_{t-k+1} + \varepsilon_t .$$

The conditional expectation value and variance of the statistics ST_t are equal, accordingly:

$$E(ST_t \mid \Delta\mathbf{Y}_{t-k+1}) = 0,$$

$$D(ST_t \mid \Delta\mathrm{Y}_{t-k+1}) = \sigma_0^2 \big/ \sqrt{N}\, \Delta\mathrm{Y}^T{}_{t-k+1} D^{-1}(\theta) \Delta\mathrm{Y}^T{}_{t-k+1} + \sigma_0^2 \xrightarrow[\sqrt{N}\to\infty]{} \sigma_0^2 .$$

On the assumption for model (1) the conditional probability density of the statistics ST_t and T_t is the density of normal distribution with zero mean and variance approaching to σ_0^2 with increase of the sample volume. Let ST_t^i are the statistics ST_t after one of changes (2)-(4), $i = 1,2,3$ .In order to estimate properties the statistics ST_t^i , $i = 1,2,3$ after the process changes and define the type of change we calculate these statistics, its means and variances.

1) Let $\boldsymbol{\theta}_M^T = (1, \beta_1^{'}, \ldots, \beta_{k-1}^{'})$ is parameter vector after changes of the drift or coefficients at time $t = t_p + 1$ without changes of the process type. Then, in accordance with (2a), ST_t^1 and its mean and variance are written as

$$ST_t^1 = \theta_M^T \Delta\mathbf{Y}_{t-k+1} - \tilde{\theta}^T \Delta\mathbf{Y}_{t-k+1} + \varepsilon_t = (\theta_M - \theta)^T \Delta\mathbf{Y}_{t-k+1} + (\theta - \tilde{\theta})^T \Delta\mathbf{Y}_{t-k+1} + \varepsilon_t ,$$

$$E(ST_t^1 \mid \Delta\mathbf{Y}_{t-k+1}) = \mu_1 - \tilde{\mu} + \sum_{j=1}^{k-1} (\beta_j^{'} - \beta_j) \Delta y_{t-j} ,$$

$$D(ST_t^1 \mid \Delta\mathrm{Y}_{t-k+1}) = \sigma_0^2 \big/ \sqrt{N}\, \Delta\mathrm{Y}^T{}_{t-k+1} D^{-1}(\theta) \Delta\mathrm{Y}^T{}_{t-k+1} + \sigma_0^2 \xrightarrow[\sqrt{N}\to\infty]{} \sigma_0^2 .$$

When the drift or coefficient are change, the mean of ST_t will be differ from zero, the variance ST_t remains without change.

2) After the stochastic trend change:

$$ST_t^2 = \theta^T \Delta\mathbf{Y}_{t-k+1} - \tilde{\theta}^T \Delta\mathbf{Y}_{t-k+1} + \eta_t = (\theta - \tilde{\theta})^T \Delta\mathbf{Y}_{t-k+1} + \eta_t ,$$

$$E(ST_t^2 \mid \Delta\mathrm{Y}_{t-k+1}) = 0, D(ST_t^2 \mid \Delta\mathrm{Y}_{t-k+1}) = \sigma_0^2 \big/ \sqrt{N}\, \Delta\mathrm{Y}^T{}_{t-k+1} D^{-1}(\theta) \Delta\mathrm{Y}^T{}_{t-k+1} + \sigma_1^2 \xrightarrow[\sqrt{N}\to\infty]{} \sigma_1^2 .$$

When the stochastic trend changes, the variance ST_t changes, the mean ST_t remains equal zero.

3) Let $\boldsymbol{\theta}_{MS}^T = (1, \beta_1^{'}, \ldots, \beta_{k-1}^{'})$ is parameter vector for model (5) after changes (4a), in which the non stationary process turn to stationary:

$$\begin{gathered} ST_t^3 = \theta_{MS}^T \Delta\mathbf{Y}_{t-k+1} + \beta_0 y_{t-1} - \tilde{\theta}^T \Delta\mathbf{Y}_{t-k+1} + \varepsilon_t = \\ (\theta_{Ms} - \theta)^T \Delta\mathbf{Y}_{t-k+1} + (\theta - \tilde{\theta})^T \Delta\mathbf{Y}_{t-k+1} + \beta_0 y_{t-1} + \varepsilon_t , \end{gathered}$$

$$E(ST_t^3 \mid \Delta Y_{t-k+1}) = +\beta_0 E(y) + (\theta_{MS} - \theta)^T E(\Delta\mathbf{Y}_{t-k+1}) = \mu,$$

$$D(ST_t^3 \mid \Delta\mathrm{Y}_{t-k+1}) = \sigma_\varepsilon^2 \big/ \sqrt{N}\, \Delta\mathrm{Y}^T{}_{t-k+1} D^{-1}(\theta) \Delta\mathrm{Y}^T{}_{t-k+1} + \beta_0^2 \sigma_y^2 + \sigma_0^2 \xrightarrow[\sqrt{N}\to\infty]{} \beta_0^2 \sigma_y^2 + \sigma_0^2 .$$ When the type of the process changes, the variance ST_t increase, the mean ST_t become equal to drift of initial process. The changes of numerical characteristics of ST_t under the process changes of the

different types described in Table 1.

Table 1

Change type	Mean of ST_t	Variance of ST_t
Without change	0	$\to \sigma_0^2$
1) Drift or coefficients changes without changes of the process type	Different from zero	$\to \sigma_0^2$
2) Stochastic trend change	0	Variance changes, $\to \sigma_0^2$
3)The process type change from non-stationary to stationary	Equal to drift of the initial process μ	Variance changes $\sigma^2 > \sigma_0^2$

5 On-line change detection

Let $\Delta y_{t-k+1}, \Delta y_{t-k+2}, \ldots$ is the training sample for non stationary process without property changes. On training sample we obtain the parameter estimates for model (5):

$$\tilde{\theta}^T = (\tilde{\mu}, \tilde{\beta}_1, \ldots, \tilde{\beta}_{k-1}),\ \tilde{\sigma}_0^2,\ \tilde{D}(\Theta) = \frac{1}{N}\sum \Delta Y_{t-k+1} \Delta Y_{t-k+1}. \quad (6)$$

On observations and estimates (6) we calculate statistics ST_t in each time t. The mean and variance changes of ST_t correspond with the process changes. The change of type 1) detected as change of mean for statistics ST_t under constant variance, the change of type 2) detected as change of variance for statistics ST_t under constant mean. If the third change type was happened, the mean for statistics ST_t change from zero to μ, the variance increase. The sequential algorithms for mean change detection considered in many works, for instance Nikiforov (1994), Siegmund. (1985), Prabhu, S. S., Roberts (1959). Algorithm's to monitor both the mean value of the process and its variance were proposed and investigated Reynolds, M.R., Jr. and Glosh, B.K. (1981) Reynolds, M.R., Jr. and Stoumbos, Z.G (2004) and others.

References

Hamilton J.D. (1994) *Time Series Analysis*. Princeton University Press.. P. 296.

Basseville M., Nikiforov I.V. (1993) *Detection of abrupt changes - theory and application*: New Jersey: Prentice Hall in Information and System Sciences.

Siegmund, D. (1985). *Sequential Analysis. Tests and Confidence Intervals*. Springer–Verlag, New York.

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. in: *Technometrics* 1239-250.

Reynolds, M.R., Jr. and Glosh, B.K. (1981) Designing control charts for means and variances. in *ASQC Quality Congress Transactions, American Society for Quality Control*, San Francisco, 400-407.

Reynolds, M.R., Jr. and Stoumbos, Z.G. (2004b) Control charts and the efficient allocation of sampling resources. Technometrics, 46, 200-214.