

Multivariate EWMA Control Chart

Summary

The **Multivariate EWMA Control Chart** procedure creates control charts for two or more numeric variables. Examining the variables in a multivariate sense is extremely important when the variables are highly correlated, since joint out-of-control conditions can occur without any individual variable violating its control limits when plotted separately. The EWMA chart is similar to the T-Squared chart, except that the points plotted on the chart are a weighted average of current and past observations.

Sample StatFolio: *mviewma.sgp*

Sample Data:

The file *grit.sgd* contains measurements made on $n = 56$ batches of “grit”, from Holmes and Mergen (1993). The data represent the percentages of large, medium, and small particles in the grit. The table below shows a partial list of the data in that file:

| <i>Large</i> | <i>Medium</i> | <i>Small</i> | <i>Means</i> | <i>Covariances</i> |
|--------------|---------------|--------------|--------------|--------------------|
| 5.4 | 93.6 | 1.0 | 5 | 3.5 |
| 3.2 | 92.6 | 4.2 | 90 | -5.5 |
| 5.2 | 91.7 | 3.1 | | -5.5 |
| 3.5 | 86.9 | 9.6 | | 13.5 |
| 2.9 | 90.4 | 6.7 | | |
| 4.6 | 92.1 | 3.3 | | |
| 4.4 | 91.5 | 4.1 | | |
| 5.0 | 90.3 | 4.7 | | |
| 8.4 | 85.1 | 6.5 | | |
| 4.2 | 89.7 | 6.1 | | |
| 3.8 | 92.5 | 3.7 | | |
| 4.3 | 91.8 | 3.9 | | |
| 3.7 | 91.7 | 4.6 | | |
| 3.8 | 90.3 | 5.9 | | |
| 2.6 | 94.5 | 2.9 | | |

Since the percentages in the first three columns sum to 100% in each row, it is only necessary to create a chart based on the first 2 columns. In addition to the data, the file contains columns with the standard means and covariances for the percentage of large and medium particles, established when monitoring of the process was first begun.

Data Input

The data to be analyzed consist of $p = 2$ or more numeric columns containing the variables of interest, with optional entry of the variable means and variance-covariance matrix.

- **Data:** 2 or more numeric columns containing the n samples, one sample per row.
- **Subgroup numbers or size:** If the data were obtained as individuals, leave this field blank or enter “1”. If the data were collected in subgroups, each of size m , enter the single value m . In such a case, each consecutive m rows in the file will be considered to form a subgroup. If the subgroup sizes are not equal, enter the name of an additional numeric or non-numeric column containing group identifiers. The program will scan this column and place all sequential rows with identical codes into the same group.
- **Standard Means:** For an “initial study” or Phase I analysis where the data will be used to determine the control limits, leave this field blank. For a “control-to-standard” or Phase II analysis, enter the name of a column contain p means.
- **Standard Covariances:** For an “initial study” or Phase I analysis, leave this field blank. For a “control-to-standard” or Phase II analysis, enter the name of a column contain the p^2 variances and covariances. In entering the values in a covariance matrix, enter the values in the first row of the matrix, then the values in the second row, and so forth. Note: if you select *Save Results* after performing a Phase I analysis, the covariance matrix will be saved in this exact format.

- **Labels:** optional labels for each subgroup. The labels will be applied in sequence to the subgroups when plotting the control charts.
- **Select:** subset selection.

Analysis Summary

The *Analysis Summary* shows the number of observations included in the analysis and the location of the control limits on the control charts.

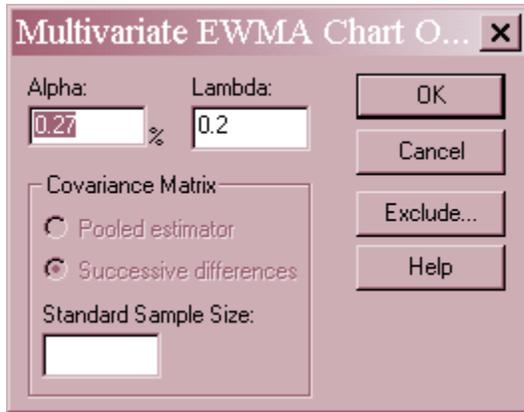
| <u>Multivariate EWMA Charts</u> | | | | |
|--|--------|-----|--------|---------------|
| Data variables: | | | | |
| Large | | | | |
| Medium | | | | |
| Number of observations included = 56 | | | | |
| Number of observations excluded = 0 | | | | |
| Smoothing parameter lambda: 0.2 | | | | |
| Initialization: centerline | | | | |
| Phase 2 - covariance specified based on standard | | | | |
| Chart | Alpha | LCL | UCL | Beyond limits |
| T-Squared | 0.0027 | 0.0 | 11.829 | 15 |

Included in the table are:

- **Smoothing parameter lambda:** the value of the EWMA parameter λ , specified on the *Analysis Options* dialog box. The default value of λ is determined from the settings on the *Control Charts* tab of the *Preferences* dialog box, accessible from the *Edit* menu.
- **Phase:** If Phase 1, the method for estimating the covariance matrix is displayed. If Phase 2, the assumptions about the input covariance matrix are shown.
- **Chart:** the type of chart. For individuals data, only a T-Squared chart is created. For grouped data, a generalized variance chart is included if the subgroup size exceeds the number of variables.
- **Alpha:** the false alarm probability of the chart, specified using *Analysis Options*. For standard 3-sigma control charts, $\alpha = 0.0027$.
- **LCL:** the lower control limit.
- **UCL:** the upper control limit.
- **Beyond limits:** the number of points on the control chart that are beyond the control limits.

In the example, the process generated 15 out-of-control signals on the EWMA T-Squared chart.

Analysis Options



- **Alpha:** the false alarm probability for positioning the control limits. For the equivalent of a standard 3-sigma control chart, set $\alpha = 0.27\%$.
- **Covariance Matrix:** details about the covariance matrix. There are 4 possibilities:
 1. If the *Standard Covariances* field was left blank on the data input dialog box and the data are in subgroups, no entry is necessary since the covariances will be estimated from the pooled within-group variability.
 2. If the *Standard Covariances* field was left blank on the data input dialog box and the data are individuals, select *Pooled estimator* to estimate the covariance between variables j and k using the usual estimator

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad (1)$$

Select *Successive differences* to estimate the covariance using

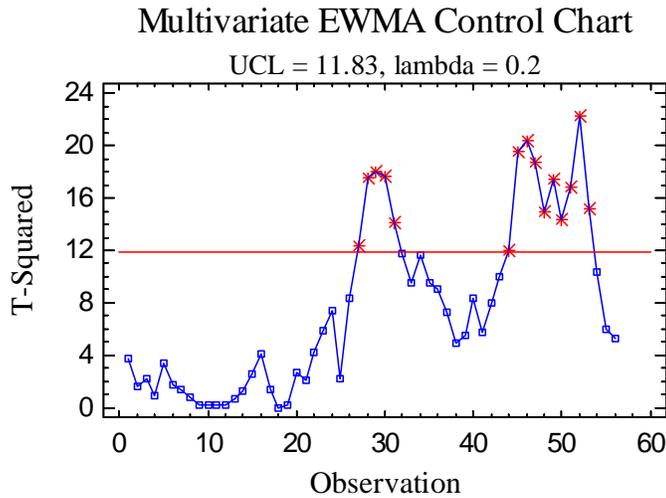
$$s_{jk} = \frac{1}{2(n-1)} \sum_{i=1}^n (x_{ij} - x_{i-1,j})(x_{ik} - x_{i-1,k}) \quad (2)$$

The second estimator is a more “local” estimator, in the sense that it captures only short-term variability, in a similar manner to the way in which a moving range is used to estimate variability for a standard individuals chart.

3. If an entry was made in the *Standard Covariances* field and the estimates provided were obtained from a previous sample, enter the size of that previous sample (if grouped, the number of subgroups) in the *Standard Sample Size* field.
 4. If an entry was made in the *Standard Covariances* field and the covariances are assumed to be known, leave the *Standard Sample Size* field empty.
- **Lambda:** a value for the EWMA parameter $0 < \lambda < 1$. The value of λ controls the amount of weight given to the past history of the process. The smaller the value, the more weight given to older observations or subgroups. This also impacts the average run length of the chart.

EWMA Chart

The *EWMA Chart* shows the exponentially weighted value of T^2 for each data value or subgroup:



The EWMA procedure begins by smoothing the observed data vector at time i by

$$EWMA_i = \lambda x_i + (1 - \lambda)EWMA_{i-1} \tag{3}$$

for individuals data and by

$$EWMA_i = \lambda \bar{x}_i + (1 - \lambda)EWMA_{i-1} \tag{4}$$

for grouped data, with $EWMA_0$ set equal to the mean vector μ or \bar{x} . The i -th value of T-squared is then calculated from

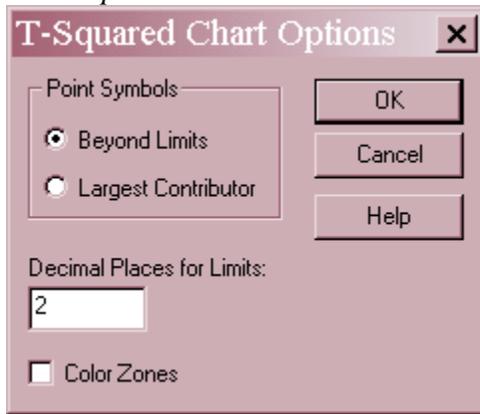
$$T_i^2 = (EWMA_i' - \mu)\Sigma_{Z_i}^{-1}(EWMA_i - \mu) \tag{5}$$

where Σ is the covariance matrix of the input data, and

$$\Sigma_{Z_i} = \frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2i}] \Sigma \tag{6}$$

For the sample data, the chart starts out well below the control limit but then rises during the latter section. 15 out-of-control signals are generated by the chart.

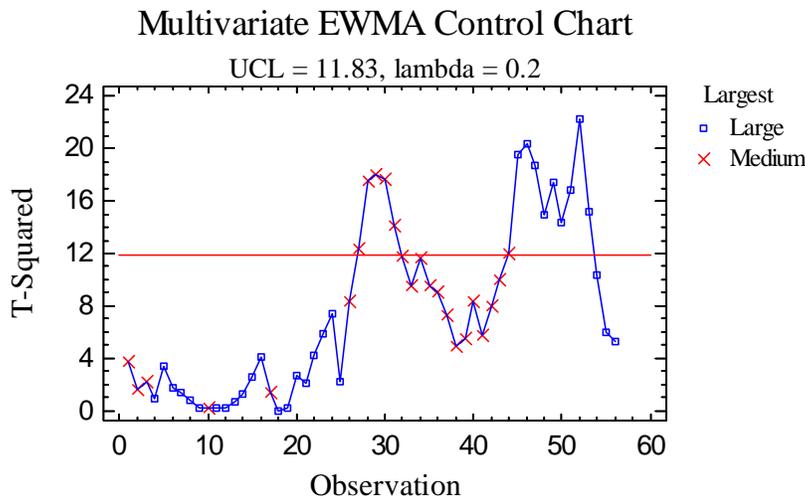
Pane Options



- **Point Symbols:** select *Beyond Limits* to draw special point symbols only for points falling above the control limit. Select *Largest Contributor* to identify the variable that contributes most to each value of T^2 .
- **Decimal Places for Limits:** number of decimal places for displaying the control limit.
- **Color Zones:** check this box to display green and red zones.

Example: Identifying Largest Contributor

If *Largest Contributor* is selected, the chart will take the following form:



Each point on the chart is coded according to the variable that contributes the most to the value of T^2 . In the plot above, the biggest contributor to the first rise appears to be the percentage of *Medium* particles, while the biggest contributor to the second rise appears to be the percentage of *Large* particles.

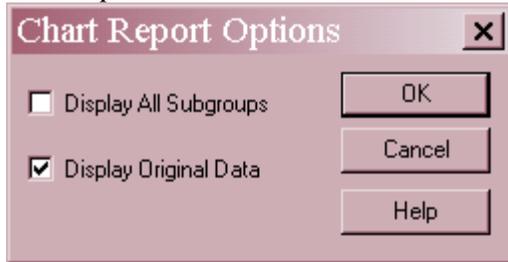
Multivariate Control Chart Report

This pane tabulates the points on the control chart:

| Multivariate Chart Report | | | |
|---------------------------|-----------|---------|---------|
| Observation | T-Squared | Large | Medium |
| 27 | * 12.3143 | 5.35513 | 86.8982 |
| 28 | * 17.5104 | 5.4841 | 86.2185 |
| 29 | * 17.9753 | 5.86728 | 85.6948 |
| 30 | * 17.676 | 6.05382 | 85.5159 |
| 31 | * 14.1056 | 6.10306 | 85.8327 |
| 44 | * 11.9746 | 6.8905 | 85.8034 |
| 45 | * 19.5133 | 7.6924 | 85.0827 |
| 46 | * 20.3941 | 7.79392 | 86.0262 |
| 47 | * 18.7157 | 7.57514 | 86.9009 |
| 48 | * 14.9892 | 7.24011 | 87.5407 |
| 49 | * 17.4784 | 7.53209 | 86.7526 |
| 50 | * 14.3944 | 7.30567 | 87.0021 |
| 51 | * 16.8681 | 7.52454 | 86.5417 |
| 52 | * 22.2222 | 7.93963 | 85.3533 |
| 53 | * 15.1529 | 7.3717 | 86.8827 |

An asterisk indicates any value beyond the control limits. Any points excluded from the analysis using the *Exclude* button are indicated with an X.

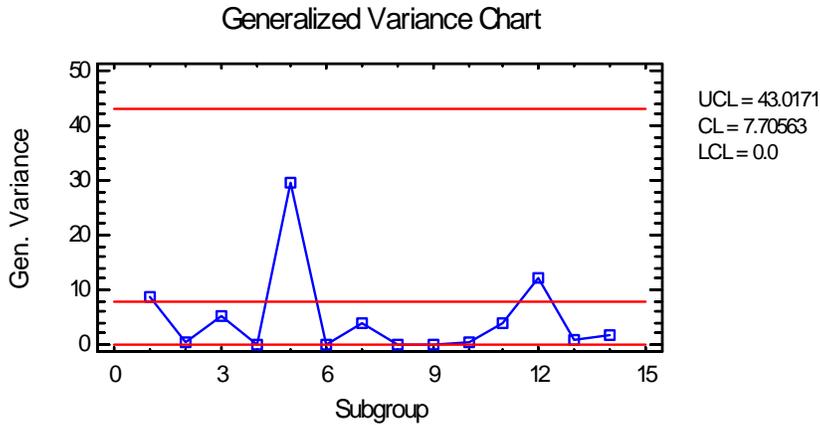
Pane Options



- **Display All Subgroups** – if checked, all observations or subgroups will be tabulated. Otherwise, only points beyond the control limits will be included in the table.
- **Display Original Data** – if checked, the values of each variable will be displayed. Otherwise, only the control chart values will be tabulated.

Generalized Variance Chart

A T-Squared chart is designed to monitor the means of p variables. To monitor the variance, a Generalized Variance Chart by sometimes be displayed:

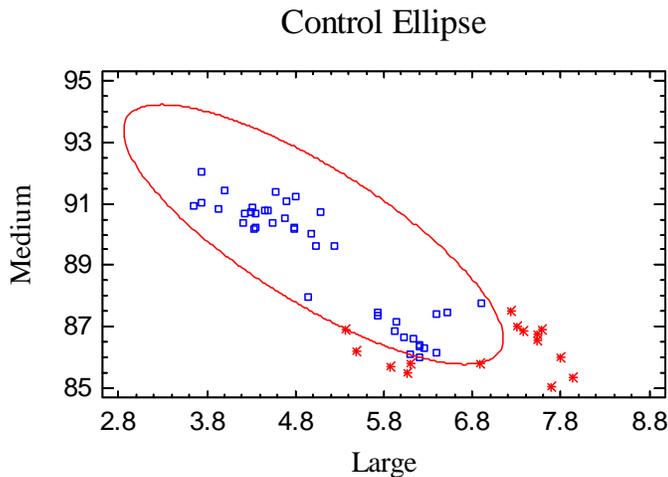


This chart is created only for data arranged in subgroups, and only if the subgroup size is at least $p + 1$. The generalized variance $|S_i|$ for the i -th subgroup is defined as the determinant of its variance-covariance matrix.

The above chart shows the grit data grouped in subgroups of 4 consecutive observations each. Any point beyond the upper control limit would indicate an unusually large amount of variability within that subgroup. In this case, no such points are present.

Control Ellipse

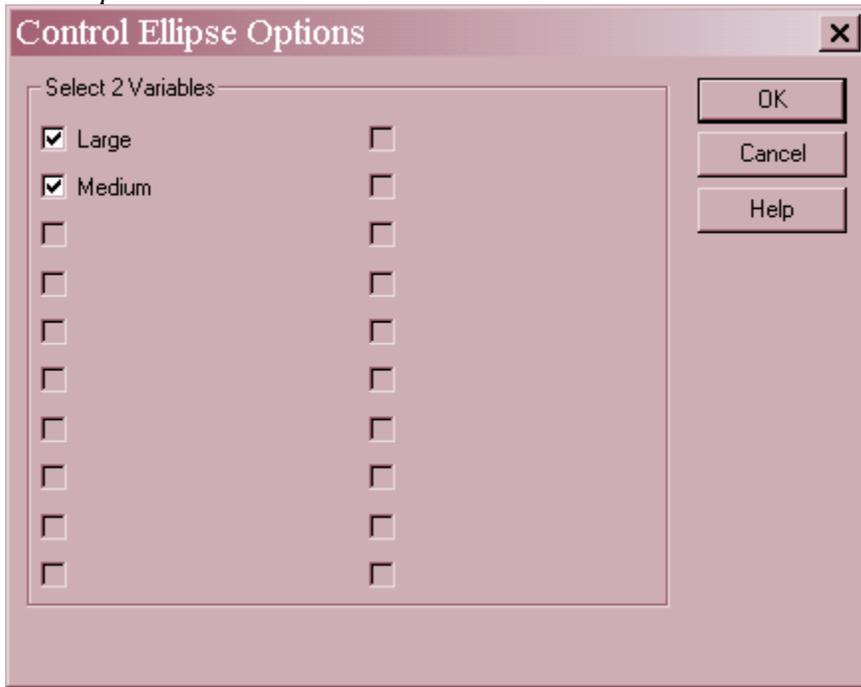
If out-of-control signals are shown on the control chart, it is useful to examine those values in detail. A good chart to use in the case of $p = 2$ variables is the *Control Ellipse*:



The upper control limit on the T-Squared chart corresponds to an elliptical region in the space of any two of the variables, with the other variables held at a fixed value. For $p = 2$, any out-of-control signals will show up as points outside the ellipse.

In the sample data, it will be noticed that some out-of-control singles correspond to a high percentage of *Large* particles while others correspond to a low percentage of *Medium* particles.

Pane Options



- **Select 2 variables:** select any 2 variables to define the control ellipse. The ellipse will be plotted assuming that all other variables are held at their mean levels. Care should be taken in interpreting the plot if $p > 2$, since the true elliptical control region for each point depends on the value of the variables that are not plotted.

EWMA Decomposition

The T^2 statistic can be decomposed into components attributable to each of the variables. One method for measuring the contribution of the j -th variable to an out-of-control T^2 value is by looking at how much smaller T^2 would be if the j -th variable was not included in the analysis. The *T-Squared Decomposition* pane does such a decomposition for each out-of-control signal on the T-Squared chart:

| T-Squared Decomposition | | | |
|-------------------------|---------|----------|------------|
| Observation | | Large | Medium |
| 27 | 12.3143 | 5.90004 | 11.99 |
| 28 | 17.5104 | 7.97741 | 16.9078 |
| 29 | 17.9753 | 5.61899 | 16.0411 |
| 30 | 17.676 | 4.27107 | 14.8204 |
| 31 | 14.1056 | 2.52796 | 10.9768 |
| 44 | 11.9746 | 0.233548 | 2.78432 |
| 45 | 19.5133 | 3.39345 | 0.872939 |
| 46 | 20.3941 | 9.86659 | 0.321611 |
| 47 | 18.7157 | 12.3129 | 1.66376 |
| 48 | 14.9892 | 10.9573 | 2.08558 |
| 49 | 17.4784 | 10.448 | 0.991745 |
| 50 | 14.3944 | 8.40275 | 0.72444 |
| 51 | 16.8681 | 8.8947 | 0.479667 |
| 52 | 22.2222 | 7.8278 | 0.00137625 |
| 53 | 15.1529 | 8.67433 | 0.688634 |

Following Runger, Alt and Montgomery (1996), the table shows

$$d_j = T^2 - T_{(j)}^2 \tag{7}$$

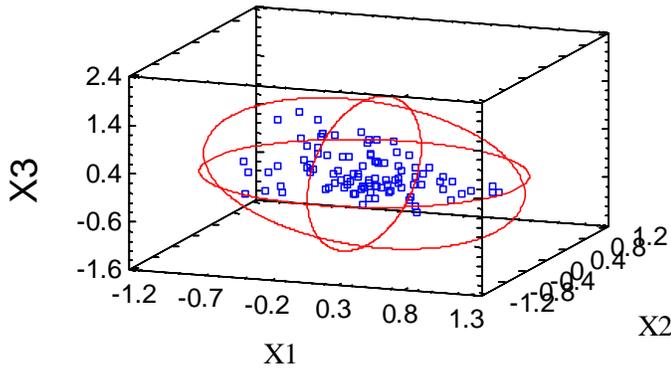
where $T_{(j)}^2$ is the value of the statistic using all variables except the j -th.

For the current data, *Medium* appears to be the dominant variable for the early out-of-control signals, while *Large* appears to be the dominant variable for the later signals.

3-D Control Chart

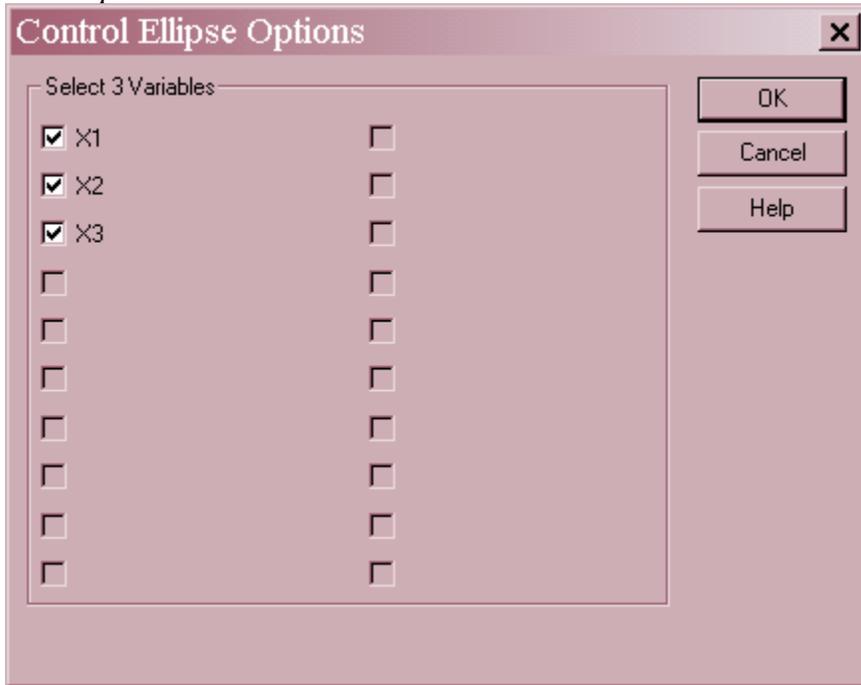
When the data consist of $p = 3$ variables, a 3-D control chart can be very helpful, since the control region is then an ellipsoid.

Control Ellipsoid



The above plot shows the outline of a typical control ellipsoid for 3 variables. Note: to explore this plot, it is very helpful to use the dynamic rotation button on the analysis toolbar.

Pane Options



- Select 3 variables:** select any 3 variables to define the control ellipsoid. The ellipsoid will be plotted assuming that all other variables are held at their mean levels. Care should be taken in interpreting the plot if $p > 3$, since the true control region for each point depends on the value of the variables that are not plotted.

Save Results

The following results can be saved to the datasheet:

1. *T-Squared* – the calculated T^2 values for each observation or subgroup.
2. *Means* – the p variable means.
3. *Covariances* – the p^2 variances and covariances in rowwise order.
4. *Labels* – the labels corresponding to each value of T^2 .
5. *Generalized Variance* - if calculated, the values of $|S|$.

Calculations

T-Squared Control Limit if Covariances are Known

$$UCL = X_{\alpha,p}^2 \quad (8)$$

T-Squared Control Limit if Covariances are Estimated from k Previous Samples

$$UCL = \frac{p(k+1)(k-1)}{k(k-p)} F_{\alpha,p,k-p} \quad (9)$$

T-Squared Control Limit if Covariances are Estimated from Current Data

$$UCL = \frac{(n-1)^2}{n} \text{Beta}_{\alpha,p/2,(n-p-1)/2} \quad (10)$$

Generalized Variance Control Limits

$$UCL = |\Sigma| (b_1 + 3\sqrt{b_2}) \quad (11)$$

$$CL = b_1 |\Sigma| \quad (12)$$

$$LCL = |\Sigma| (b_1 - 3\sqrt{b_2}) \quad (13)$$

where

$$b_1 = \frac{1}{(n-1)^p} \prod_{i=1}^p (n-i) \quad (14)$$

$$b_2 = \frac{1}{(n-1)^{2p}} \prod_{i=1}^p (n-i) \left[\prod_{j=1}^p (n-j+2) - \prod_{j=1}^p (n-j) \right] \quad (15)$$

If Σ is not known, it is replaced by the estimate $|S|/b_1$.