Process Monitoring for Correlated Gamma Distributed Data Using Generalized Linear Model Based Control Charts

Duangporn Jearkpaporn , Connie M. Borror

Douglas C. Montgomery and George C. Runger

Arizona State University

Tempe, AZ

Outline

- Introduction
  - Problem Statement
  - Review of Model-Based Control Chart
  - Review of Generalized Linear Models
- Recent Researches
- Development of Proposed Method
  - Data Model for Model-Based Control Chart
  - Generalized Likelihood Ratio
  - Monitoring Scheme
  - Simulation Results (ARL Performance)
- Summary and Conclusions

Problem Statement

A need to simultaneously monitor several related process variables.

- All characteristics jointly determine the usefulness of the product.

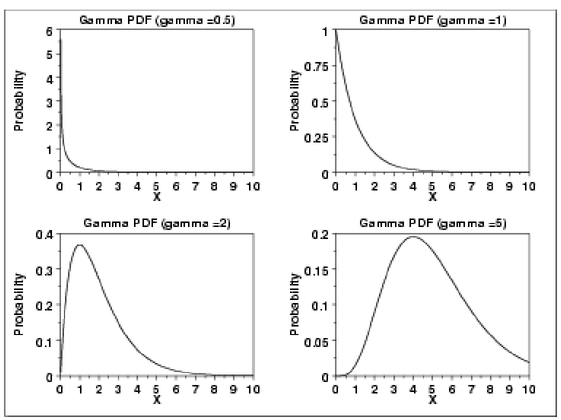
- Mixtures of normally and non-normally distributed variables.

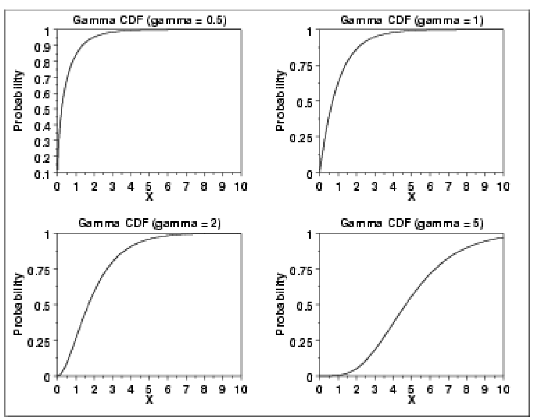
Gamma Distribution

$$f(y;\theta,\kappa)=\frac{1}{\theta^{\kappa}\Gamma(\kappa)}y^{\kappa-1}e^{-y/\theta} \quad , y>0$$

and zero otherwise

- κ is a shape parameter and > 0
- θ is scale parameter and > 0
- $E(y) = \mu = \kappa\theta$ and $Var(y) = \kappa\theta^2$





Model-Based Control Chart

- A model-based control strategy that relates inputs to outputs
  [Mandel (1969), Zhang (1984), Hawkins (1993)]
- Residuals are control statistics
- Model need not be linear
- Benefits
  - ✓ Simple interpretations
  - ✓ Easy to perform.
  - ✓ No correlation over time

Generalized Linear Models (GLM)

- Generalized linear models have three components
  - Random Component
  - Systematic Component, $\eta = \mathbf{X}\beta$
  - Link Function, $g(\bullet)$

  i.e. Ordinary Least Squares (OLS)
    - Normal Distribution
    - Linear combination of inputs, $\mathbf{X}\beta$
    - Identity link, $\mu = g^{-1}(\mathbf{X}\beta) = \mathbf{X}\beta$
- Residual analysis of a GLM is similar to OLS
  - Available from most software packages
  - (Asymptotically) Normally distributed

Recent Research

- Skinner, Montgomery, and Runger's (2001) model-based control chart for counted data Using Generalized Linear Model Based Control Charts.
- Hauck, Runger, and Montgomery (1999) used process knowledge to group variables into one of three charts (model-void, model fixed and cause-selecting).
- Wade and Woodall's (1993) cause-selecting control charts regress input variables on output variables. Various charts of standardized residuals are constructed and used to interpret the signal.

Data Model for Model-Based Control Chart

- Consider p Gamma variables, y_j
- y_j is a function of k inputs, x_i with the link $g(\bullet)$
- $E(y_j) = \mu = g^{-1}(\beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \ldots + \beta_{jk}x_k)$

$$y_j \sim gamma\left(\kappa_j, \theta_j = \frac{g^{-1}(\beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \ldots + \beta_{jk}x_k)}{\kappa_j}\right)$$

Data Model for Model-Based Control Chart (Con't)

- Assume log link, one x and p y's
- All y's have different means

$$\mu_j = \exp(\beta_{j0} + \beta_{j1}x)$$

- Variance of y is $\mu^2/\kappa = \exp[2\cdot(\beta_{j0} + \beta_{j1}x)]/\kappa$

$$y_j \sim gamma\left[\kappa_j, \theta_j = \frac{\exp(\beta_{j0} + \beta_{j1}x)}{\kappa_j}\right]$$

Generalized Likelihood Ratio

- Likelihood ratio statistics for detecting mean changes for models given above were developed.
- Detecting the mean change is equivalent to testing H_0: $\mu = \mu_0$ vs. H_1: $\mu = \mu_1$. The GLR for testing this is

$$-2\left\{\ln\left(\frac{\Gamma(\kappa_0)}{\Gamma(\hat{\kappa}_1)}\right)+(\hat{\kappa}_1-\kappa_0)\ln y+\ln\left(\frac{\hat{\kappa}_1^{\hat{\kappa}_1}}{\kappa_0^{\kappa_0}}\right)+\ln\left(\frac{\mu_0^{\kappa_0}}{\hat{\mu}_1^{\hat{\kappa}_1}}\right)-\frac{y\hat{\kappa}_1}{\hat{\mu}_1}+\frac{y\kappa_0}{\mu_0}\right\}$$

where $\hat{\mu}_1$ and $\hat{\kappa}_1$ are the MLE of μ and κ under H_1.

- Under H_0 , test statistic ~ $\chi^2(m)$

  where m = # parameter we wish to test.

- Reject H_0 if this statistic > c

Generalized Likelihood Ratio (Con't)

Consider the case where

- A mean shift is due to the scale parameter changes.
- Assume κ is known and equal $\kappa_0 = \kappa_1 = \kappa$
- A univariate case $p = 1$.

The test statistic simplifies to

$$r = sign(y - \hat{\mu})\left\{2\left[y \ln\left(\frac{y}{\hat{\mu}}\right) + \frac{y - \hat{\mu}}{\hat{\mu}}\right]\right\}^{\frac{1}{2}}$$

where $\hat{\mu} = g^{-1}\left(\mathbf{x}'\hat{\beta}\right)$

A likelihood ratio statistic is the ***deviance residual***

Generalized Likelihood Ratio (Con't)

For the shift in p variables, the test statistic is

$$2\sum_{j=1}^{p}\kappa_j\left\{-\ln\left(\frac{y_j}{\hat{\mu}_j}\right)+\left(\frac{y_j-\hat{\mu}_j}{\hat{\mu}_j}\right)\right\}$$

where $\hat{\mu}_j = g^{-1}\left(\mathbf{x}'\hat{\boldsymbol{\beta}}\right)$ the fitted mean of the j^{th} output under H_0 (the mean before the shift).

Monitoring Scheme

Consider a process with p output variables.

- An initial set of n observations is used to fit GLM for each of p variables.
- Construct control limits based on deviance residuals obtained from GLM.
- Observe new data.
- Compute deviance residuals for new observations.
- Plot deviance residuals on p individual Charts.
- A signal on any of the p charts indicates that the variable it monitors is out of control.

Results: Comparing ARL Performance

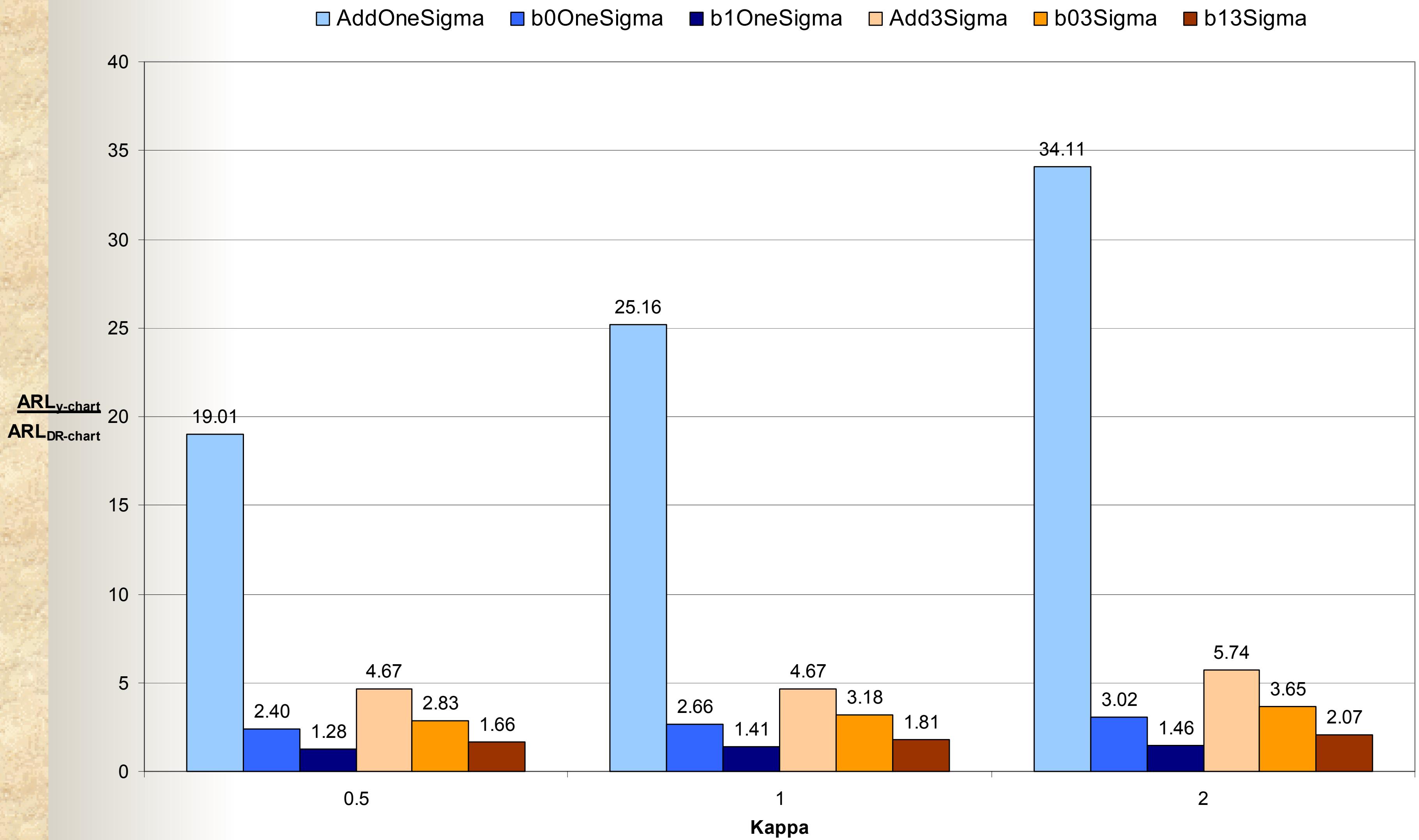
Simulation scenarios:

- Univariate Case.
- Bivariate Case with $\mu_1 = \mu_2$ and $\kappa_1 = \kappa_2$
- Bivariate Case with $\mu_1 \neq \mu_2$ and $\kappa_1 = \kappa_2$

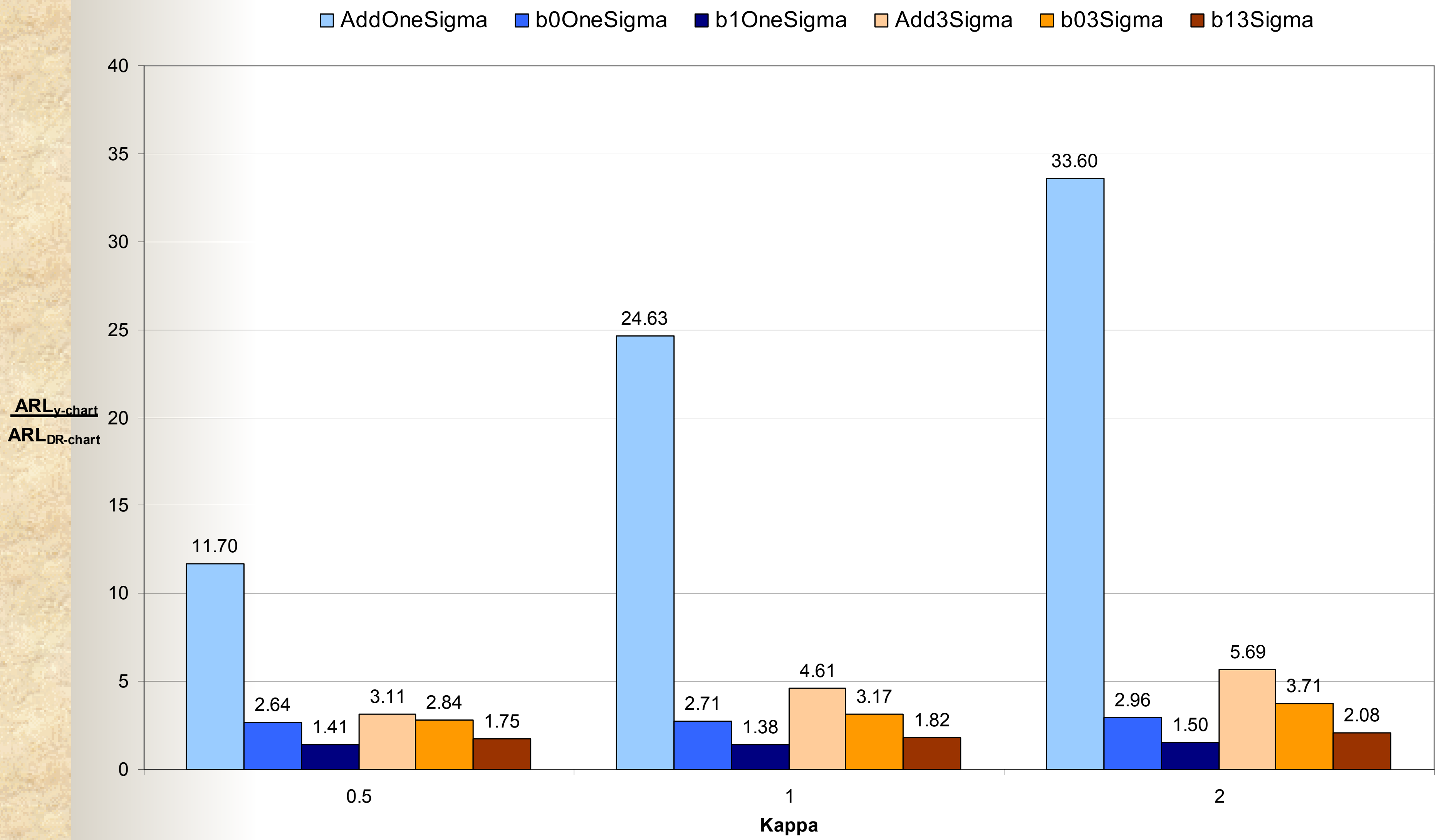
Two types of mean shift are introduced in each scenario:

- Additive shift $\mu = g^{-1}(\beta_0 + \beta_1 x) + \delta$
- Multiplicative Shift
  - ➔ $\mu = g^{-1}(\beta_0 + \beta_1 x + \delta)$
  - ➔ $\mu = g^{-1}(\beta_0 + (\beta_1 + \delta) x)$

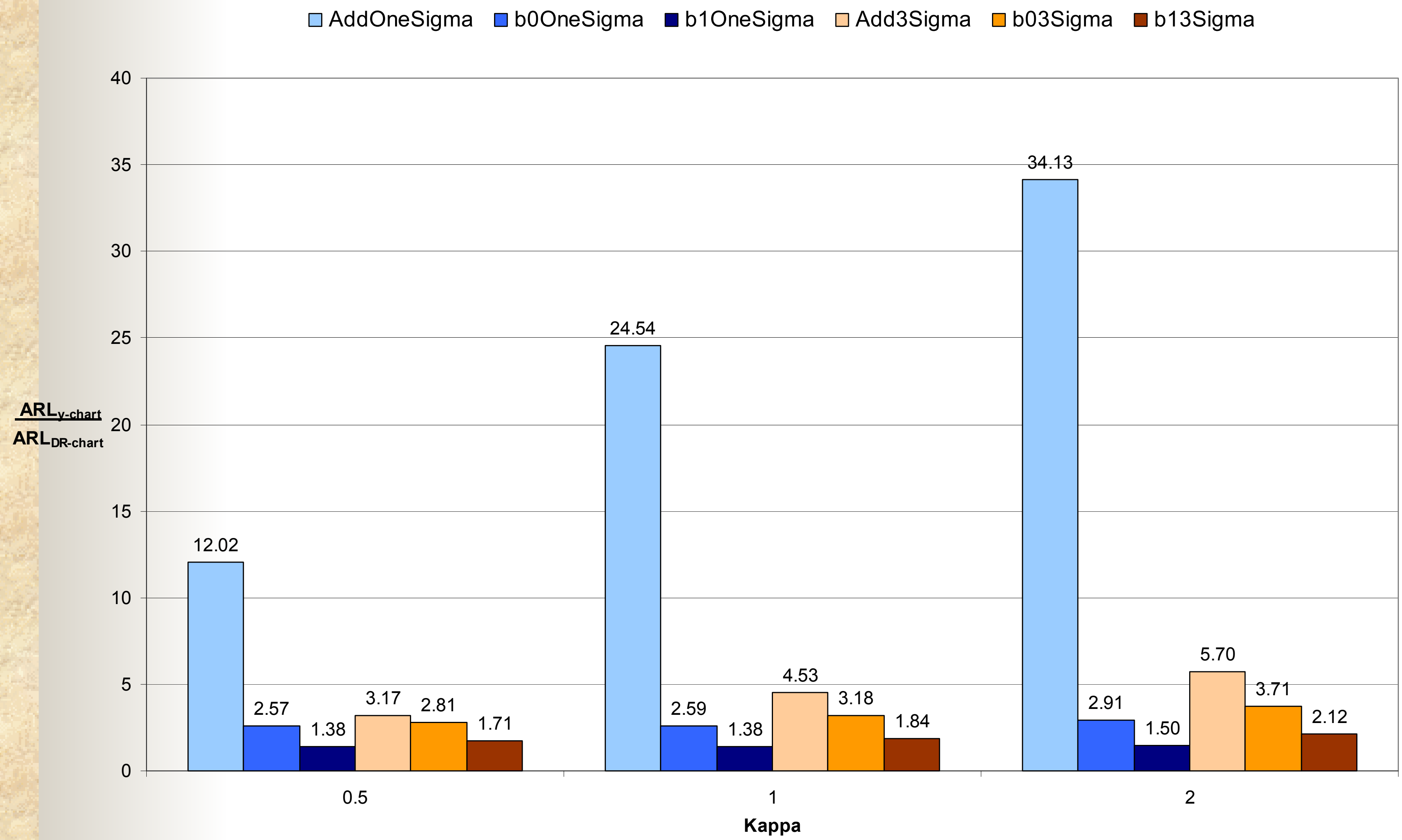
Univariate Case



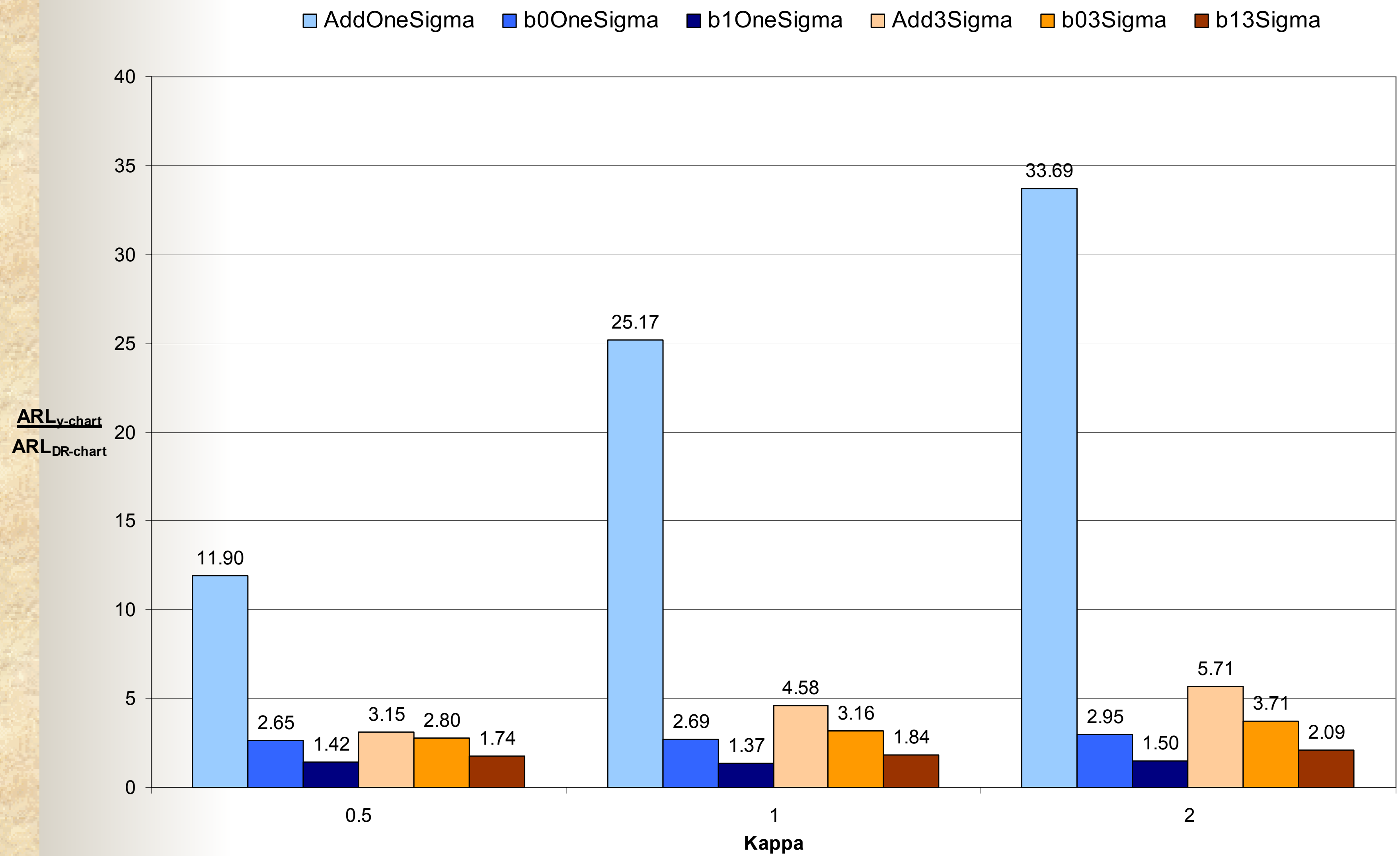
Bivariate Case: Equal Means – y_1 Shifted



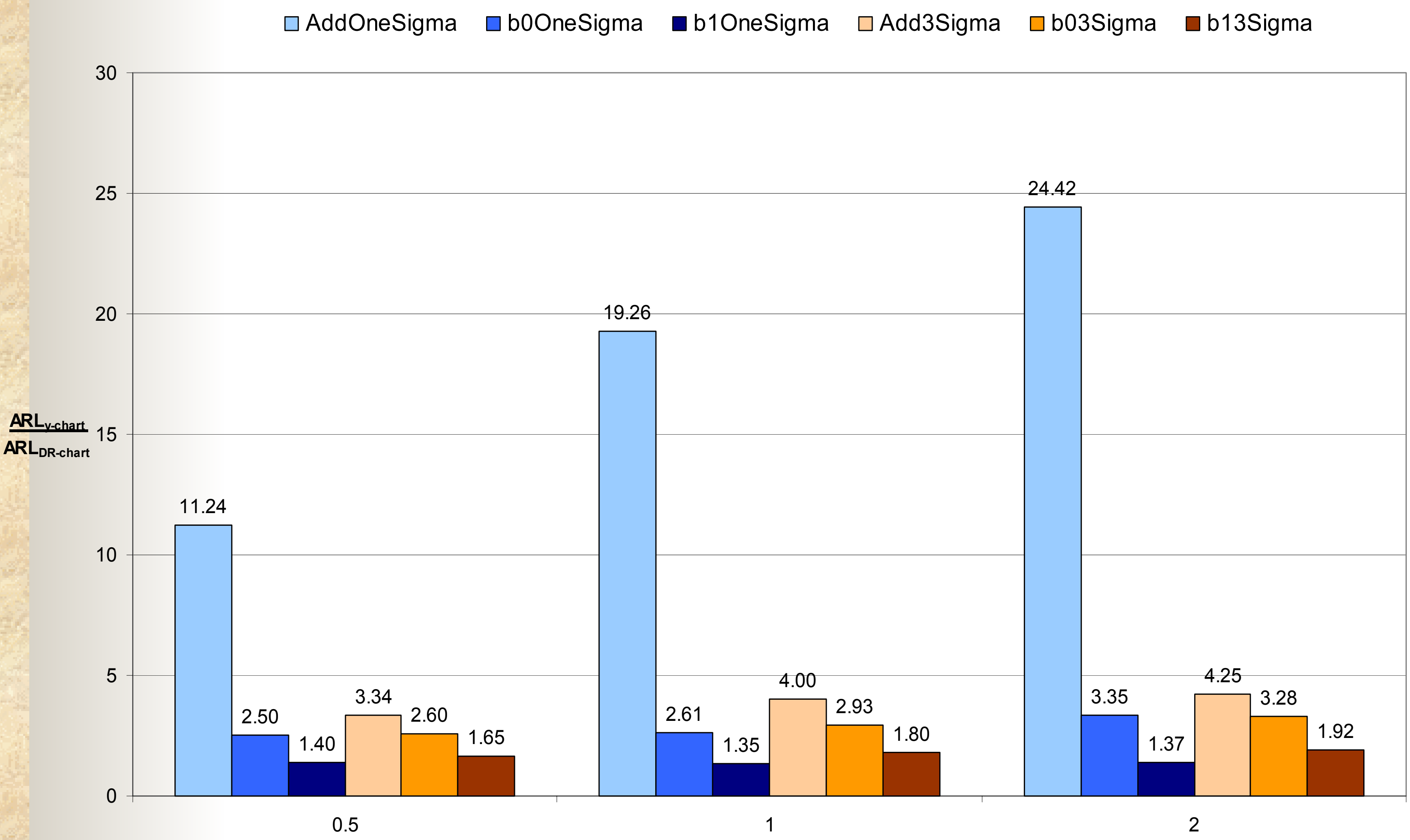
Bivariate Case: Equal Means – y_1 & y_2 Shifted



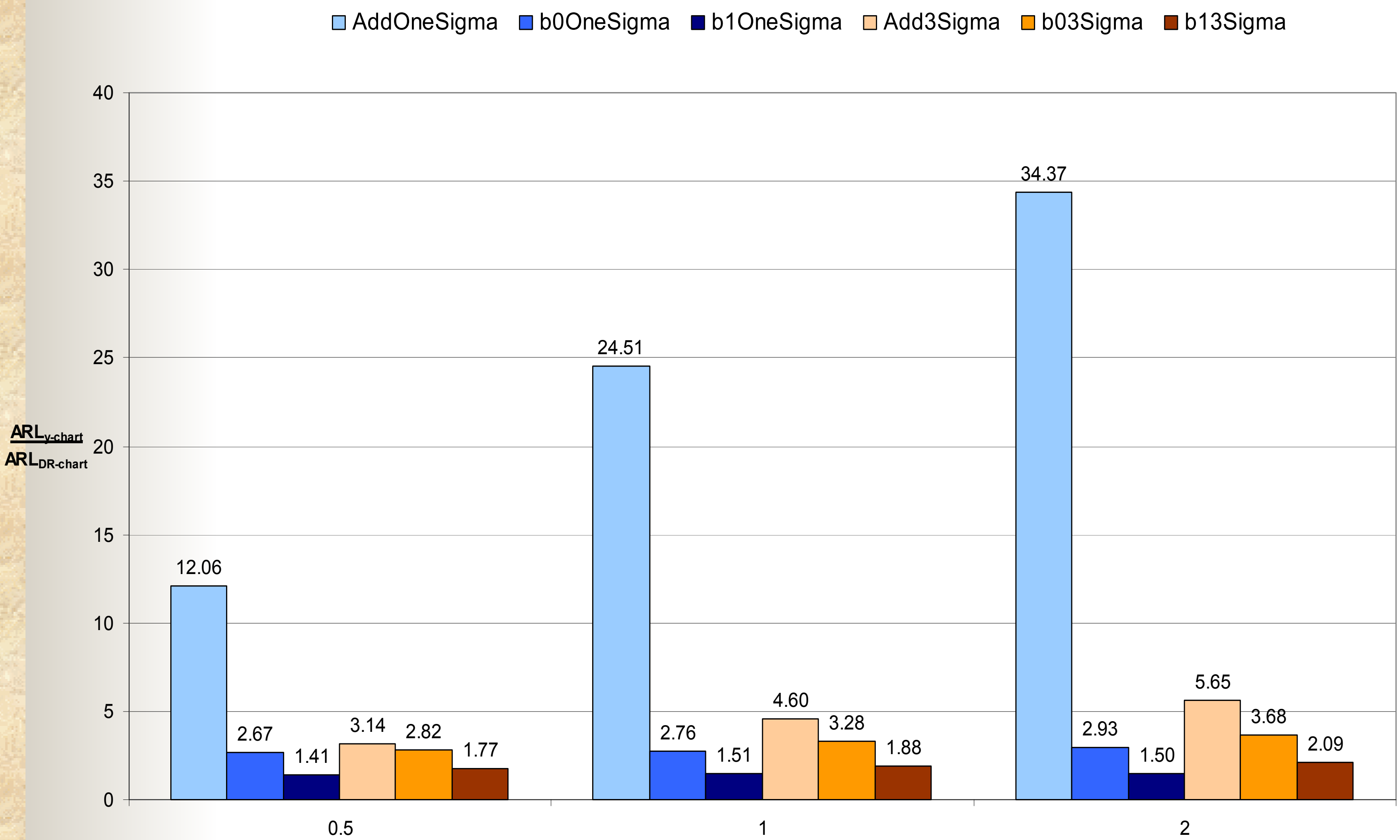
Bivariate Case: Unequal Means – y_1 Shifted



Bivariate Case: Unequal Means – y_2 Shifted



Bivariate Case: Unequal Means – y_1&y_2 Shifted



Summary and Conclusions

This study shows:

- Overall, deviance residuals based control chart outperform the individual Shewhart charts.
- Deviance chart
  - is best to detect the additive shift.
  - detect the shift quicker than individual Shewhart charts when outputs are correlated and non-normal.
- Ability of model-based control chart to handle relationships between the responses.
- Application of the GLM in process monitoring

What Else?

- Apply the same idea to
  - CUSUM for detecting the drift in mean
  - Cascade Processes
- Using other methods in model fitting
  - GEE
  - Robust fitting technique

Question ?

References

- Engelhardt, B. (1992), Introduction to Probability and Mathematical Statistics, 2nd ed., Duxbury Press, Belmont, California, pp.418.
- Hauck D.J., G.C. Runger, and D.C. Montgomery (1999), "Multivariate statistical process monitoring and diagnosis with grouped regression-adjusted variables," *Commun Stat-Simul* 28(2) pp. 309-328.
- Hawkins, D.M., "Regression Adjustment for Variables in Multivariate Quality Control," *JQT*, pp.170-182.
- Mendel, B.J. (1969), "The Regression control Chart," *JQT*, pp.1-9.
- Skinner, K. R., Runger, G. C., and Montgomery, D. C. (2001), "Process Monitoring for Multiple Count Data Using Generalized Linear Model Based control Charts," submitted to IJPR.
- Wade, M.R. and W.H. Woodall (1993), "A review and analysis of cause-selecting control charts," *J. of Quality Technology* 25(3) pp. 161-169.
- Zhang, G.X. (1984), "A New Type of Control Charts and a Theory of Diagnosis with Control Charts", *World Quality congress Transactions.* American society for Quality Control, pp. 175-185.