Guaranteed Conditional ARL Performance in the Presence of Autocorrelation

Christian H. Weiß, Detlef Steuer
Department of Mathematics & Statistics, Helmut Schmidt University, Hamburg

Carsten Jentsch
Department of Economics, University of Mannheim

Murat Caner Testik
Department of Industrial Engineering, Hacettepe University
Monitoring of Autocorrelated Processes

Introduction
Let \((X_t)\) be autocorrelated process of variables data.

Two common approaches for monitoring (Testik, 2005):
- Fit time series model and monitor residuals; or
- adjust control limits and monitor original observations \((X_t)\).

Here, we consider second approach.

Presented approach applies to general AR processes, but for ease of presentation, focus on AR(1) process

\[
X_t = \rho \cdot X_{t-1} + \epsilon_t, \quad \text{where } \rho \in (-1, 1).
\]

Innovations \(\epsilon_t\) i.i.d. with \(\mu_\epsilon = \mu (1 - \rho)\) and \(\sigma_\epsilon^2 = \sigma^2 (1 - \rho^2)\), where \(\mu\) and \(\sigma^2\) are mean and variance of \(X_t\).
If even Gaussian AR process with normally distrib. innovations, then model fully specified by parameters $\mu, \sigma, \rho$.

**Abbreviation:** $X_t \sim P_{\mu,\sigma;\rho}$.

W.l.o.g., we assume $\mu = 0$ and $\sigma = 1$ in analyses.

Note: $\rho = 0$ corresponds to i.i.d. case.

In the sequel, focus on individuals control chart, but our method could be extended to other types of control chart as well.
Monitoring of Autocorrelated Processes

If true model $P_{\mu,\sigma;\rho}$ known,

and considering symmetry of normal distribution,

control limits of individuals chart are

$$\mu \pm k \cdot \sigma \quad \text{with} \quad k := k(P_{\mu,\sigma;\rho}).$$

Gaussian case: $k$ only depends on $\rho$ (Schmid, 1995): $k = k(\rho)$.

Choice of $k$

if $\text{ARL}_0 = 370.4$.

i.i.d.: $3-\sigma$-limits,

i.e., $k(0) \approx 3.0$.

Computation using R’s spc package (Knoth, 2016).

Christian H. Weiß — Helmut Schmidt University, Hamburg
If true model $P_{\mu,\sigma;\rho}$ not known, then estimation from in-control data $x_1, \ldots, x_n$ (Phase I). For example, moment estimates $\bar{x}, s$ and $\hat{\rho} = \hat{\rho}(1)$. If ignoring fact that parameters estimated, then chart design

$$\bar{x} \pm k(\hat{\rho}) \cdot s,$$

or according to $\bar{x} \pm k(0) \cdot s$ in i.i.d. case.
Monitoring of Autocorrelated Processes

Individuals chart $\bar{x} \pm k(\hat{\rho}) \cdot s$:

box plots of $k(\hat{\rho})$ and true ARL for (a) $\rho = 0.2$ and (b) $\rho = 0.6$.

Actual ARL$_0$s most often too small!
So need to replace $k(\hat{\rho})$ by corrected value $k(\bar{x}, s \mid P_{\mu,\sigma; \rho})$.
But would require again knowledge about $P_{\mu,\sigma; \rho}$.
Circumvent problem by not intending to reach $\text{ARL}_0$ exactly, only $\text{ARL} \geq \text{ARL}_0$ with a probab. $1 - \alpha$ (Gandy & Kvaløy, 2013): **guaranteed conditional performance** (GCP).
So not majority of ARL values below intended $\text{ARL}_0$ (as before), but only $\alpha \cdot 100\%$ of all cases.

**Crucial question**: how $k$ for GCP without knowing $P_{\mu,\sigma; \rho}$?

**Literature**: only i.i.d. case (Gandy & Kvaløy, 2013).

**Here**: novel GCP approach for time series (especially AR(1)) with bootstrap implementation.

Christian H. Weiβ — Helmut Schmidt University, Hamburg
Guaranteed Conditional Performance under AR Dependence

Bootstrap Approach
Since \( P_{\mu,\sigma;\rho} \) not known, we estimate \( \bar{X}, S \) and \( \hat{\rho} := \hat{\rho}(1) \).

Now extend **GCP approach** by Gandy & Kvaløy (2013):

We would need \( \alpha \)-quantile \( q_{\alpha;\rho} \) of deviations

\[
k(\bar{X}, S \mid P\bar{X},S;\hat{\rho}) - k(\bar{X}, S \mid P\mu,\sigma;\rho) = k(\hat{\rho}) - k(\bar{X}, S \mid P\mu,\sigma;\rho).
\]

With appropriate bootstrap scheme (see below),

we compute \( \alpha \)-quantile \( q^*_{\alpha;\hat{\rho}} \) of

\[
k(\bar{X}^*, S^* \mid P\bar{X}^*,S^*;\hat{\rho}^*) - k(\bar{X}^*, S^* \mid P\bar{X},S;\hat{\rho}) = k(\hat{\rho}^*) - k(\bar{X}^*, S^* \mid P\bar{X},S;\hat{\rho}),
\]

and define corrected limit as

\[
k^*_{\text{corr};\hat{\rho}} := k(\bar{X}, S \mid P\bar{X},S;\hat{\rho}) - q^*_{\alpha;\hat{\rho}} = k(\hat{\rho}) - q^*_{\alpha;\hat{\rho}}.
\]
GCP under AR Dependence

\[ k_{\text{corr};\hat{\rho}}^* := k(\bar{X}, S \mid P_{\bar{X},S;\hat{\rho}}) - q^*_{\alpha;\hat{\rho}} = k(\hat{\rho}) - q^*_{\alpha;\hat{\rho}}. \]

This corresponds to Hall’s percentile method.

Requires computation of two \( k \)-values per bootstrap sample, \( k(\hat{\rho}^*) \) and \( k(\bar{X}^*, S^* \mid P_{\bar{X},S;\hat{\rho}}) \).

In contrast to i.i.d. case, Hall’s method differs from more simple standard percentile method, where

\[ \tilde{k}_{\text{corr};\hat{\rho}}^* := \tilde{q}_{1-\alpha;\hat{\rho}} \]

with \( \tilde{q}_{1-\alpha;\hat{\rho}} \) being \( (1 - \alpha) \)-quantile of \( k(\bar{X}^*, S^* \mid P_{\bar{X},S;\hat{\rho}}) \).

Requires computation of only one \( k \), \( k(\bar{X}^*, S^* \mid P_{\bar{X},S;\hat{\rho}}) \).

Christian H. Weiß — Helmut Schmidt University, Hamburg
We use either nonparametric or parametric AR(1) bootstrap, see, e.g., Kreiss & Paparoditis (2011).

**Nonparametric AR(1) bootstrap:**

1. Compute the centered observations \( Y_t = X_t - \bar{X} \).
2. Estimate the autoregressive parameter \( \hat{\rho} := \hat{\rho}(1) \), compute residuals \( \hat{e}_t = Y_t - \hat{\rho}Y_{t-1} \) for \( t = 2, \ldots, n \), and center them, i.e., compute \( \tilde{e}_t = \hat{e}_t - \frac{1}{n-1} \sum_{t=2}^{n} \hat{e}_t \).
3. Generate bootstrap observations \( Y_1^*, \ldots, Y_n^* \) according to
   \[
   Y_t^* = \hat{\rho} Y_{t-1}^* + e_t^*,
   \]
   where \( e_t^* \) drawn from \{\( \tilde{e}_2, \ldots, \tilde{e}_n \}\}, \( Y_0^* \) from prerun.
4. Define bootstrap sample $X_1^*, \ldots, X_n^*$ by $X_t^* = Y_t^* + \bar{X}$, compute $k(\hat{\rho}^*)$ and $k(\bar{X}^*, S^* \mid P_{\bar{X}, S; \hat{\rho}})$.

5. Repeat steps 3 and 4 $B$ times (e.g., $B = 1000$), compute quantiles required for corrected limits.

**Parametric AR(1) bootstrap**: modify step 3 as

3’. Estimate innovations’ variance $\sigma^2_\epsilon$ as

$$\hat{\sigma}^2_\epsilon := \frac{1}{n-1} \sum_{t=2}^{n} \tilde{e}_t^2.$$ 

Generate bootstrap observations $Y_1^*, \ldots, Y_n^*$ via

$$Y_t^* = \hat{\rho} Y_{t-1}^* + e_t^* \text{ with } e_t^* \sim \text{N}(0, \hat{\sigma}^2_\epsilon),$$

where $Y_0^*$ from appropriate normal distribution.
Some remarks:

• For simulation studies as below, ability to efficiently computing ARL, since done for each bootstrap replication. For AR(1), we used R's spc package (Knoth, 2016).

• In practical applications (no Monte Carlo replicates), also simulation-based ARL computation feasible.

• Above autoregressive bootstraps extends easily from order one to general order $p \in \mathbb{N}$. 

Christian H. Weiß — Helmut Schmidt University, Hamburg
Guaranteed Conditional Performance under AR Dependence

Simulation Study
GCP under AR Dependence

**Simulation study**: 10 000 replications, 1 000 bootstrap replications, $\alpha = 0.10$, $\text{ARL}_0 = 370.4$.

Proportion of Monte-Carlo replications with $\text{ARL}$ below $\text{ARL}_0$:

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>100</th>
<th>500</th>
<th>1000</th>
<th>100</th>
<th>500</th>
<th>1000</th>
<th>100</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>−0.8</td>
<td>0.112</td>
<td>0.108</td>
<td>0.103</td>
<td>0.110</td>
<td>0.106</td>
<td>0.103</td>
<td>0.102</td>
<td>0.088</td>
<td>0.085</td>
</tr>
<tr>
<td>−0.6</td>
<td>0.108</td>
<td>0.103</td>
<td>0.103</td>
<td>0.104</td>
<td>0.103</td>
<td>0.100</td>
<td>0.103</td>
<td>0.095</td>
<td>0.095</td>
</tr>
<tr>
<td>−0.4</td>
<td>0.100</td>
<td>0.100</td>
<td>0.102</td>
<td>0.096</td>
<td>0.099</td>
<td>0.100</td>
<td>0.099</td>
<td>0.098</td>
<td>0.100</td>
</tr>
<tr>
<td>−0.2</td>
<td>0.092</td>
<td>0.095</td>
<td>0.100</td>
<td>0.088</td>
<td>0.097</td>
<td>0.100</td>
<td>0.092</td>
<td>0.095</td>
<td>0.100</td>
</tr>
<tr>
<td>0</td>
<td>0.093</td>
<td>0.095</td>
<td>0.100</td>
<td>0.087</td>
<td>0.093</td>
<td>0.099</td>
<td>0.094</td>
<td>0.096</td>
<td>0.100</td>
</tr>
<tr>
<td>0.2</td>
<td>0.097</td>
<td>0.096</td>
<td>0.097</td>
<td>0.094</td>
<td>0.094</td>
<td>0.096</td>
<td>0.096</td>
<td>0.095</td>
<td>0.097</td>
</tr>
<tr>
<td>0.4</td>
<td>0.109</td>
<td>0.099</td>
<td>0.100</td>
<td>0.106</td>
<td>0.096</td>
<td>0.098</td>
<td>0.108</td>
<td>0.097</td>
<td>0.097</td>
</tr>
<tr>
<td>0.6</td>
<td>0.126</td>
<td>0.101</td>
<td>0.104</td>
<td>0.123</td>
<td>0.101</td>
<td>0.102</td>
<td>0.122</td>
<td>0.094</td>
<td>0.094</td>
</tr>
<tr>
<td>0.8</td>
<td>0.160</td>
<td>0.115</td>
<td>0.107</td>
<td>0.159</td>
<td>0.114</td>
<td>0.107</td>
<td>0.150</td>
<td>0.095</td>
<td>0.088</td>
</tr>
</tbody>
</table>

Hall’s method preferable, also nonparametric.

---

Christian H. Weiß — Helmut Schmidt University, Hamburg
Proportion of Monte-Carlo replications with ARL below $\text{ARL}_0$ if using i.i.d. bootstrap approach by Gandy & Kvaløy (2013):

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>nonparametric, $n =$</th>
<th>parametric, $n =$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100  500  1000</td>
<td>100  500  1000</td>
</tr>
<tr>
<td>$-0.8$</td>
<td>0.181  0.103  0.059</td>
<td>0.183  0.104  0.060</td>
</tr>
<tr>
<td>$-0.6$</td>
<td>0.138  0.117  0.097</td>
<td>0.142  0.119  0.100</td>
</tr>
<tr>
<td>$-0.4$</td>
<td>0.114  0.109  0.106</td>
<td>0.117  0.113  0.108</td>
</tr>
<tr>
<td>$-0.2$</td>
<td>0.092  0.099  0.101</td>
<td>0.100  0.102  0.103</td>
</tr>
<tr>
<td>$0$</td>
<td>0.093  0.096  0.099</td>
<td>0.100  0.098  0.101</td>
</tr>
<tr>
<td>$0.2$</td>
<td>0.111  0.105  0.104</td>
<td>0.118  0.108  0.107</td>
</tr>
<tr>
<td>$0.4$</td>
<td>0.155  0.127  0.113</td>
<td>0.163  0.132  0.119</td>
</tr>
<tr>
<td>$0.6$</td>
<td>0.232  0.148  0.117</td>
<td>0.238  0.151  0.119</td>
</tr>
<tr>
<td>$0.8$</td>
<td>0.351  0.149  0.084</td>
<td>0.351  0.151  0.085</td>
</tr>
</tbody>
</table>
GCP under AR Dependence

Decile box plots (whiskers end at 10% and 90% quantiles) of corrected limits $k_{\text{corr}}^*; \hat{\rho}$ (Hall’s method) for $\rho = 0.4$:

Christian H. Weiß — Helmut Schmidt University, Hamburg
Decile box plots (whiskers end at 10\% and 90\% quantiles) of corrected limits $k^*_{\text{corr}}; \tilde{\rho}$ (Hall’s method) for nonparametric AR(1) bootstrap with $n = 100$:
Decile box plots (whiskers end at 10% and 90% quantiles) of ARLs (Hall’s method) for nonparametric AR(1) bootstrap with \( n = 100 \):
Decile box plots (whiskers end at 10% and 90% quantiles) of ARLs (Hall’s method) for nonparametric AR(1) bootstrap with $n = 500$:
Decile box plots (whiskers end at 10% and 90% quantiles) of ARLs (Hall’s method) against mean shift $s$, for nonparametric AR(1) bootstrap with $n = 100$: 

Christian H. Weiβ — Helmut Schmidt University, Hamburg
Decile box plots (whiskers end at 10% and 90% quantiles) of ARLs (Hall's method) against mean shift $s$, for nonparametric AR(1) bootstrap with $n = 500$: 

![Box plots](image-url)
In simulations before, concerned with Gaussian AR(1) process. Normality was used for efficient ARL computation.

If innovations’ distribution not known, combine nonparametric bootstrap scheme with simulation-based ARL computation. We used **SA algorithm** (stochastic approximation) by Capizzi & Masarotto (2016) and R package “saControlLimits”.

**Idea:** Use residuals from bootstrap scheme also for ARL simulations within SA algorithm.

To still manage 10 000 Monte-Carlo replicates, we used **warp-speed method** by Giacomini et al. (2013) for simulations.

Christian H. Weiß — Helmut Schmidt University, Hamburg
Decile box plots (whiskers end at 10% and 90% quantiles) of ARLs (Hall’s method) for $\rho = 0.4$ and $n = 1000$:

ARL computation relies on normality (index “N”) or on simulation (index “S”).

Christian H. Weiß — Helmut Schmidt University, Hamburg
Conclusions

- Designing control chart with guaranteed conditional ARL performance in time series context, solutions relying bootstrap schemes for time series.
- Approach exemplified for individuals chart applied to Gaussian AR(1) process, but could be adapted to different charts or data generating process.
- Simulations: Hall’s percentile method leads to reliable chart designs already for moderate sample sizes if autocorrelation not excessively large, and if using appropriate scheme for ARL computation.

Christian H. Weiß — Helmut Schmidt University, Hamburg
Thank You for Your Interest!

Christian H. Weiß
Department of Mathematics & Statistics
Helmut Schmidt University, Hamburg
weissc@hsu-hh.de
Knoth (2016) The spc package (statistical process control), Version 0.5.4.
Testik (2005) Model inadequacy and residuals ... QREI 21, 115–130.