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

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

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

$$X_t \;=\; \rho \cdot X_{t-1} + \epsilon_t, \qquad \text{where } \rho \in (-1, 1).$$

Innovations ϵ_t i.i.d. with $\mu_\epsilon = \mu\,(1-\rho)$ and $\sigma_\epsilon^2 = \sigma^2\,(1-\rho^2)$, where μ and σ^2 are mean and variance of X_t.

If even Gaussian AR process with normally distrib. innovations, then model fully specified by parameters μ, σ, ρ.

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.

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

and considering symmetry of normal distribution,

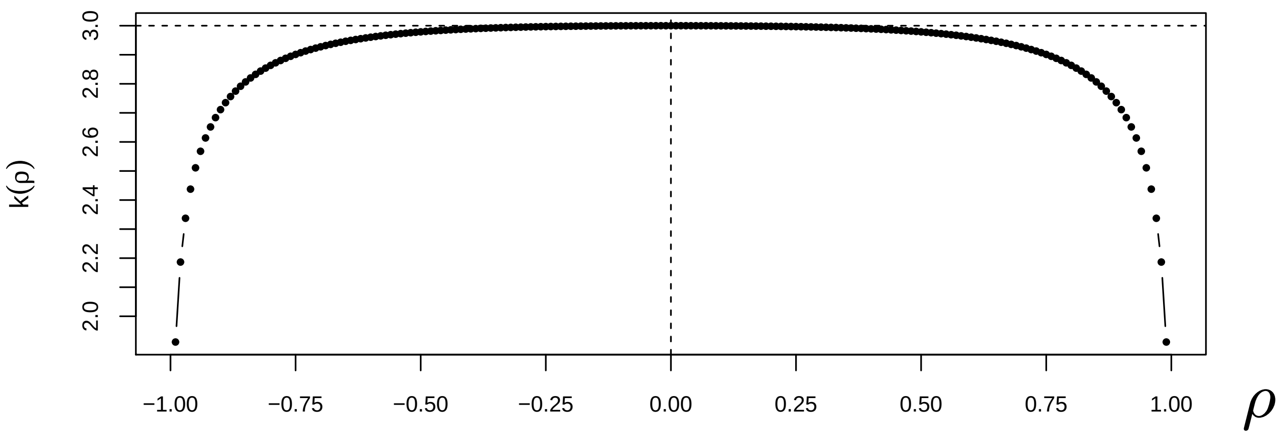
control limits of individuals chart are

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

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

Choice of k
if $\text{ARL}_0 = 370.4$.

i.i.d.: 3-σ-limits,
i.e., $k(0) \approx 3.0$.



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

If true model $P_{\mu,\sigma;\,\rho}$ *not* known,

then estimation from in-control data $x_1, \ldots, x_n$ (Phase I).

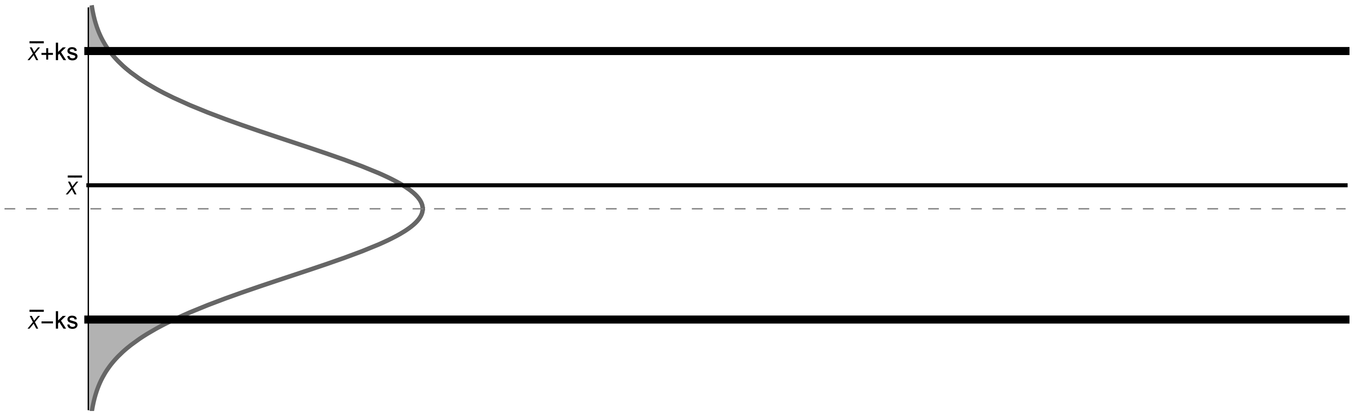
For example, **moment estimates** $\bar{x}$, s and $\hat{\rho} = \hat{\rho}(1)$.

If ignoring fact that parameters estimated,

then chart design

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

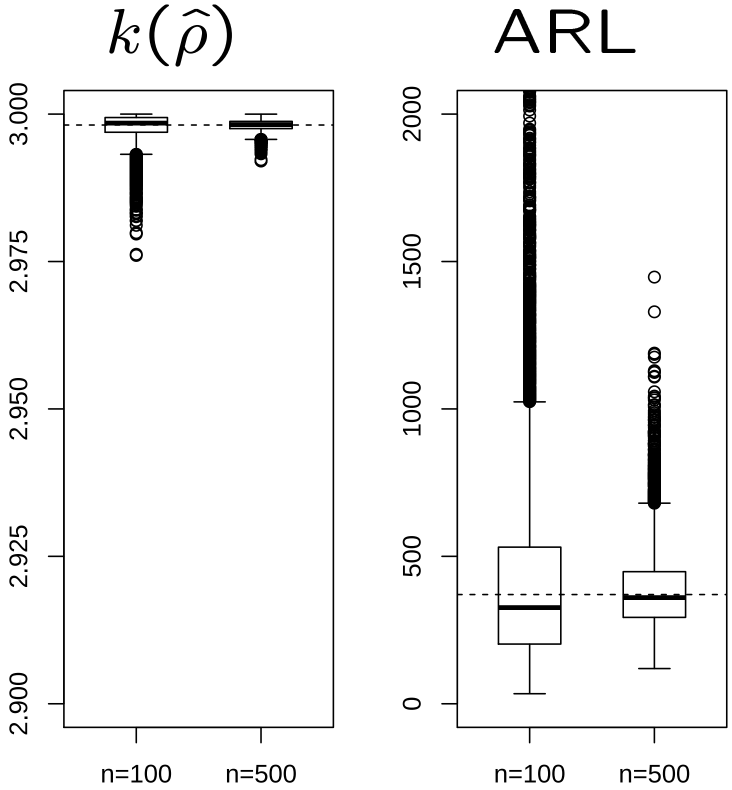
or according to $\bar{x} \ \pm \ k(0) \cdot s$ in i.i.d. case.



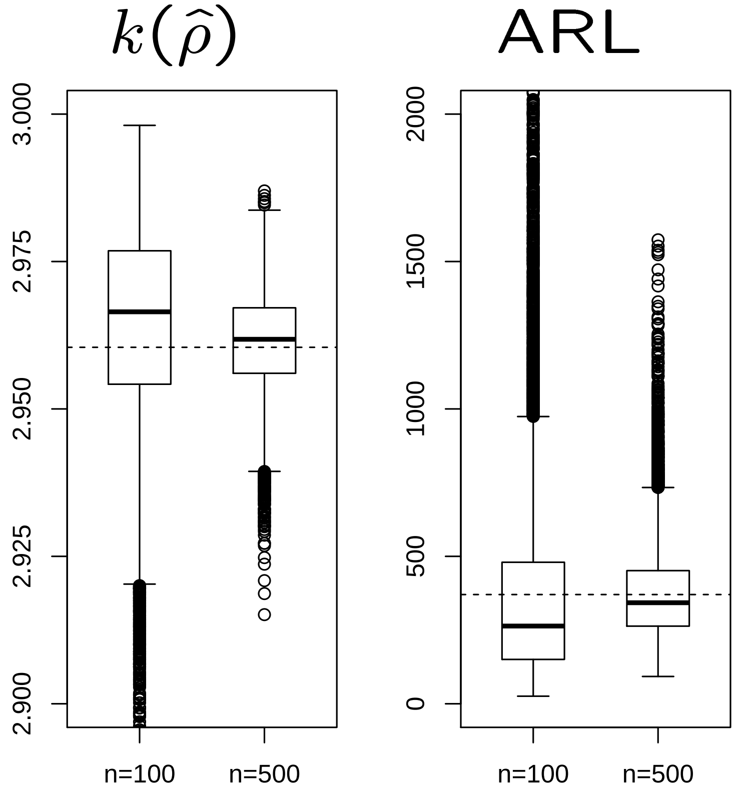
Individuals chart $\bar{x} \ \pm \ k(\hat{\rho}) \cdot s$:

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

(a)



(b)



Actual ARL_0s most often too small!

So need to replace $k(\widehat{\rho})$ 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 ARL_0 *exactly*,

only $\mathrm{ARL} \geq \mathrm{ARL}_0$ with a probab. $1-\alpha$ (Gandy & Kvaløy, 2013):

guaranteed conditional performance (GCP).

So not majority of ARL values below intended 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.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH
STAT

Guaranteed Conditional Performance under AR Dependence

Bootstrap Approach

GCP under AR Dependence

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

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

We would need α-quantile $q_{\alpha;\,\rho}$ of deviations

$$k(\bar{X}, S \mid P_{\bar{X},S;\,\widehat{\rho}}) \;-\; k(\bar{X}, S \mid P_{\mu,\sigma;\,\rho}) \;=\; k(\widehat{\rho}) \;-\; k(\bar{X}, S \mid P_{\mu,\sigma;\,\rho}).$$

With appropriate bootstrap scheme (see below),

we compute α-quantile $q^*_{\alpha;\,\widehat{\rho}}$ of

$$k(\bar{X}^*, S^* \mid P_{\bar{X}^*,S^*;\,\widehat{\rho}^*}) - k(\bar{X}^*, S^* \mid P_{\bar{X},S;\,\widehat{\rho}}) \;=\; k(\widehat{\rho}^*) - k(\bar{X}^*, S^* \mid P_{\bar{X},S;\,\widehat{\rho}}),$$

and define corrected limit as

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

$$k^*_{\mathsf{corr};\,\widehat{\rho}} \;:=\; k(\bar{X}, S \mid P_{\bar{X},S;\,\widehat{\rho}}) - q^*_{\alpha;\,\widehat{\rho}} \;=\; k(\widehat{\rho}) - q^*_{\alpha;\,\widehat{\rho}}.$$

This corresponds to **Hall's percentile method**.

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

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

$$\tilde{k}^*_{\mathsf{corr};\,\widehat{\rho}} \;:=\; \tilde{q}^*_{1-\alpha;\,\widehat{\rho}}$$

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

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

We use either nonparametric or parametric AR(1) bootstrap, see, e.g., Kreiss & Paparoditis (2011).

Nonparametric AR(1) bootstrap:

1. Compute the centered observations $Y_t = X_t - \overline{X}$.
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 $\quad X_t^* = Y_t^* + \overline{X}$,
   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., $B = 1\,000$),
   compute quantiles required for corrected limits.

Parametric AR(1) bootstrap: modify step 3 as

3'. Estimate innovations' variance σ_ϵ^2 as

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

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

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

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 $p \in \mathbb{N}$.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Guaranteed Conditional Performance under AR Dependence

Simulation Study

Simulation study: 10 000 replications,

1 000 bootstrap replications, $\alpha = 0.10$, $\text{ARL}_0 = 370.4$.

Proportion of Monte-Carlo replications with ARL below ARL_0:

	Hall's percentile method						Standard percentile method					
	nonparametric, $n =$			parametric, $n =$			nonparametric, $n =$			parametric, $n =$		
ρ	100	500	1000	100	500	1000	100	500	1000	100	500	1000
−0.8	0.112	0.108	0.103	0.110	0.106	0.103	0.102	0.088	0.085	0.100	0.087	0.084
−0.6	0.108	0.103	0.103	0.104	0.103	0.100	0.103	0.095	0.095	0.100	0.096	0.093
−0.4	0.100	0.100	0.102	0.096	0.099	0.100	0.099	0.098	0.100	0.096	0.098	0.099
−0.2	0.092	0.095	0.100	0.088	0.097	0.100	0.092	0.095	0.100	0.088	0.097	0.100
0	0.093	0.095	0.100	0.087	0.093	0.099	0.094	0.096	0.100	0.087	0.093	0.099
0.2	0.097	0.096	0.097	0.094	0.094	0.096	0.096	0.095	0.097	0.094	0.093	0.096
0.4	0.109	0.099	0.100	0.106	0.096	0.098	0.108	0.097	0.097	0.105	0.094	0.096
0.6	0.126	0.101	0.104	0.123	0.101	0.102	0.122	0.094	0.094	0.120	0.094	0.094
0.8	0.160	0.115	0.107	0.159	0.114	0.107	0.150	0.095	0.088	0.147	0.093	0.086

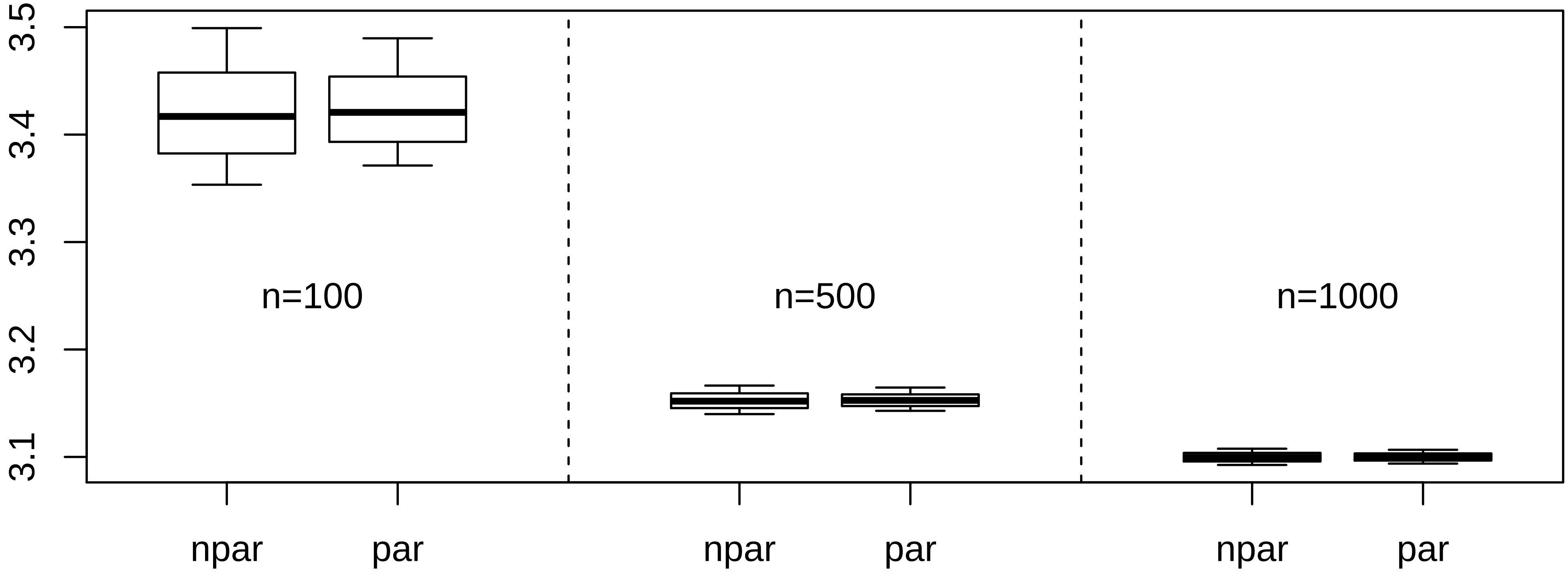
Hall's method preferable, also nonparametric.

Proportion of Monte-Carlo replications with ARL below ARL_0 if using i.i.d. bootstrap approach by Gandy & Kvaløy (2013):

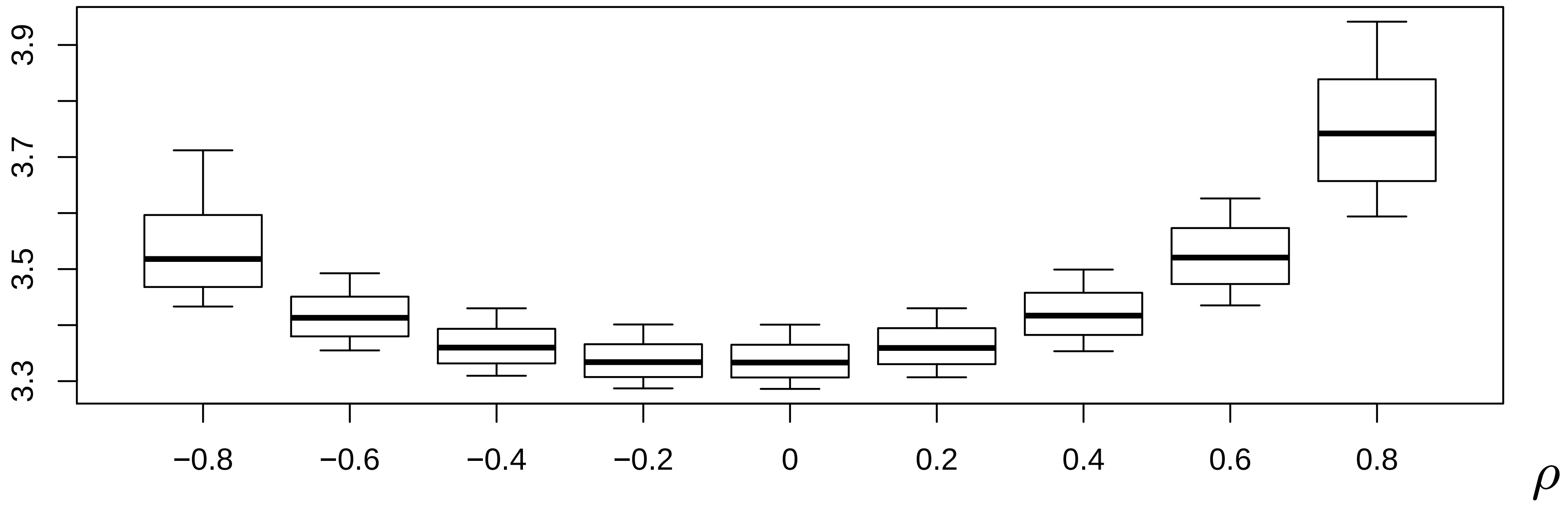
	nonparametric, $n =$			parametric, $n =$		
ρ	100	500	1000	100	500	1000
−0.8	0.181	0.103	0.059	0.183	0.104	0.060
−0.6	0.138	0.117	0.097	0.142	0.119	0.100
−0.4	0.114	0.109	0.106	0.117	0.113	0.108
−0.2	0.092	0.099	0.101	0.100	0.102	0.103
0	0.093	0.096	0.099	0.100	0.098	0.101
0.2	0.111	0.105	0.104	0.118	0.108	0.107
0.4	0.155	0.127	0.113	0.163	0.132	0.119
0.6	0.232	0.148	0.117	0.238	0.151	0.119
0.8	0.351	0.149	0.084	0.351	0.151	0.085

MATH STAT

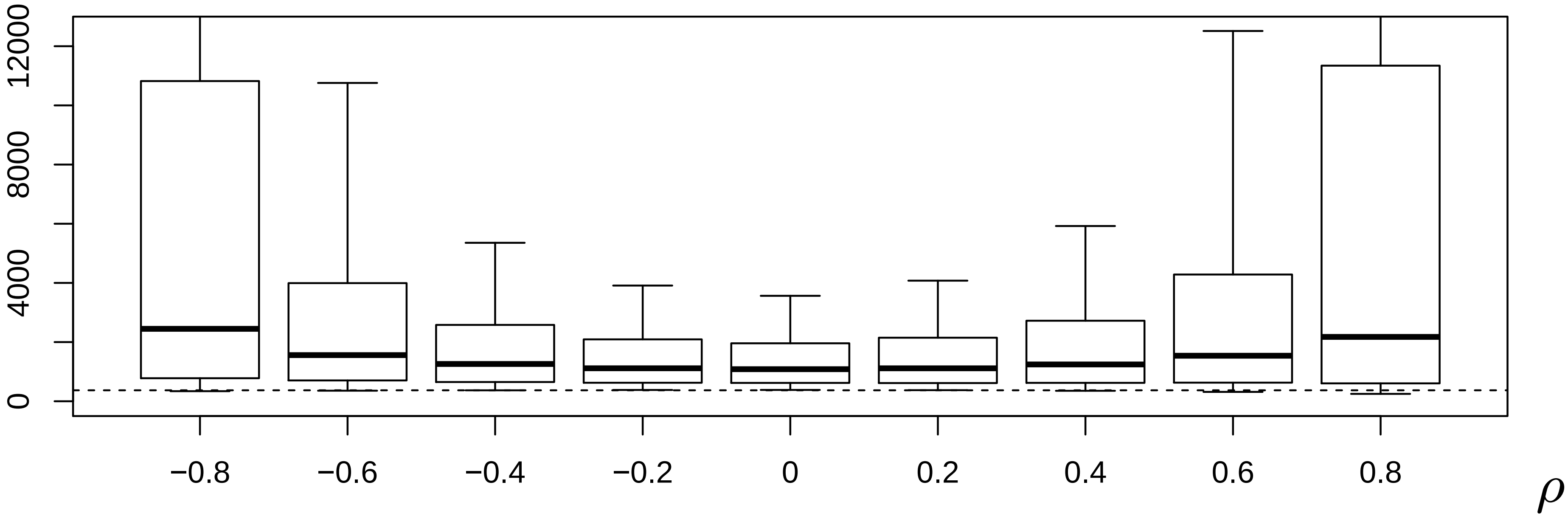
Decile box plots (whiskers end at 10 % and 90 % quantiles) of corrected limits $k^*_{\text{corr};\,\hat{\rho}}$ (Hall's method) for $\rho = 0.4$:



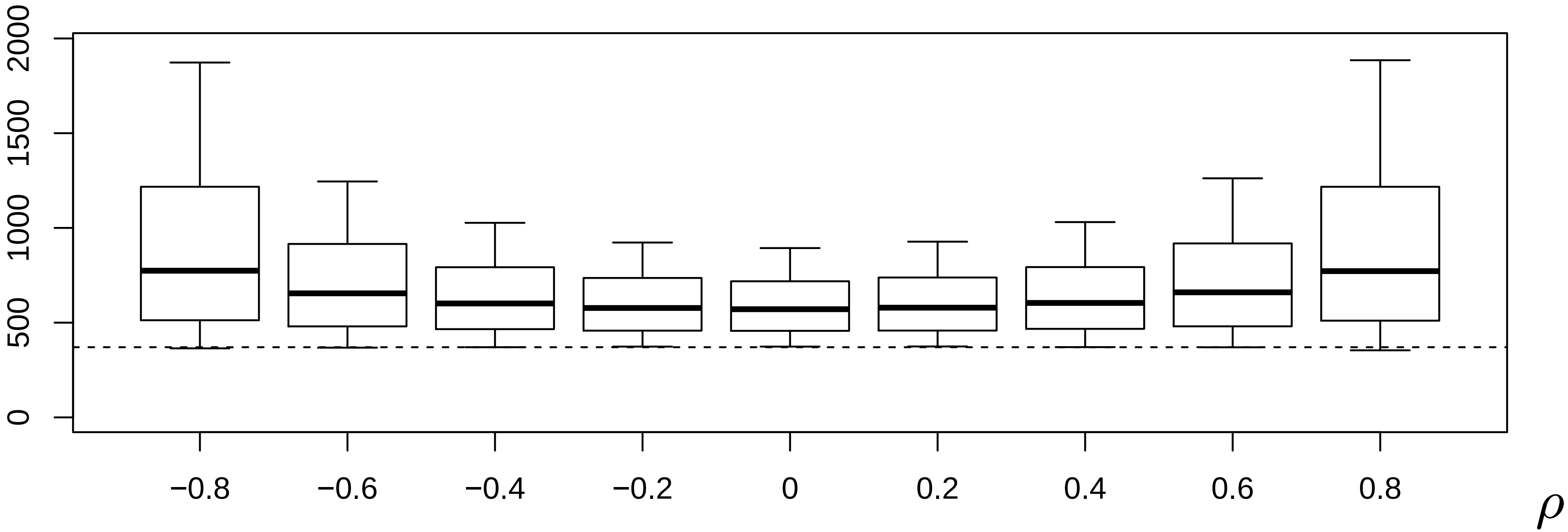
Decile box plots (whiskers end at 10 % and 90 % quantiles) of corrected limits $k^*_{\text{corr};\,\widehat{\rho}}$ (Hall's method) for nonparametric AR(1) bootstrap with $n = 100$:



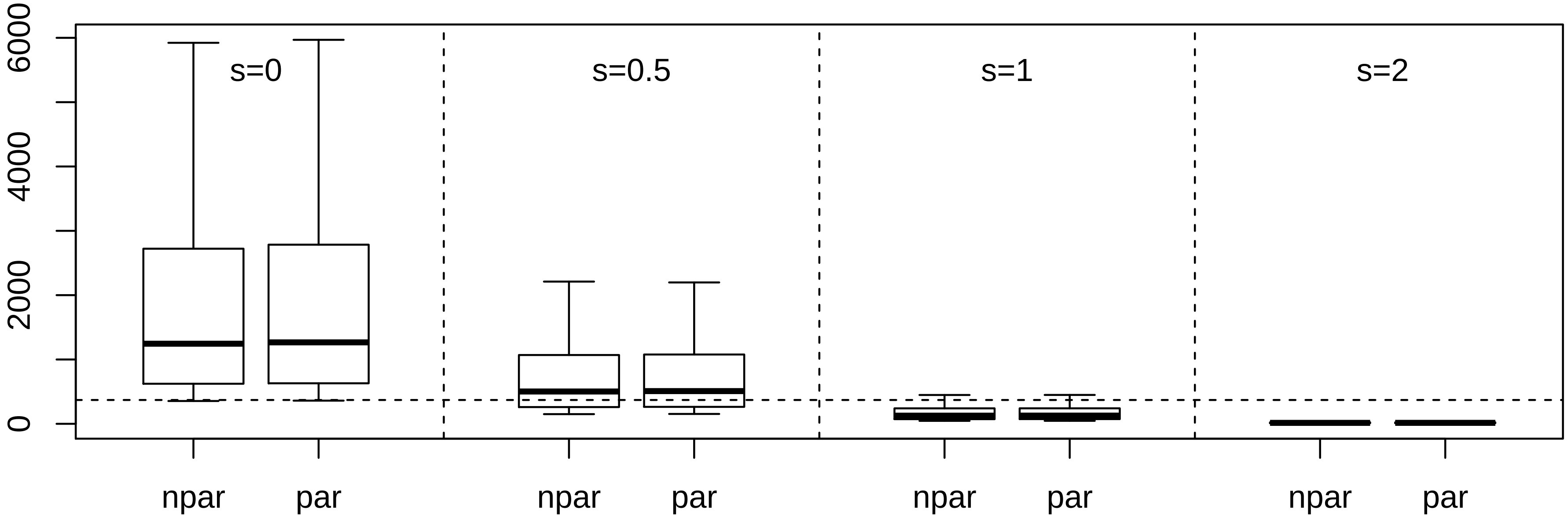
Decile box plots (whiskers end at 10 % and 90 % quantiles) of ARLs (Hall's method) for nonparametric AR(1) bootstrap with $n = 100$:



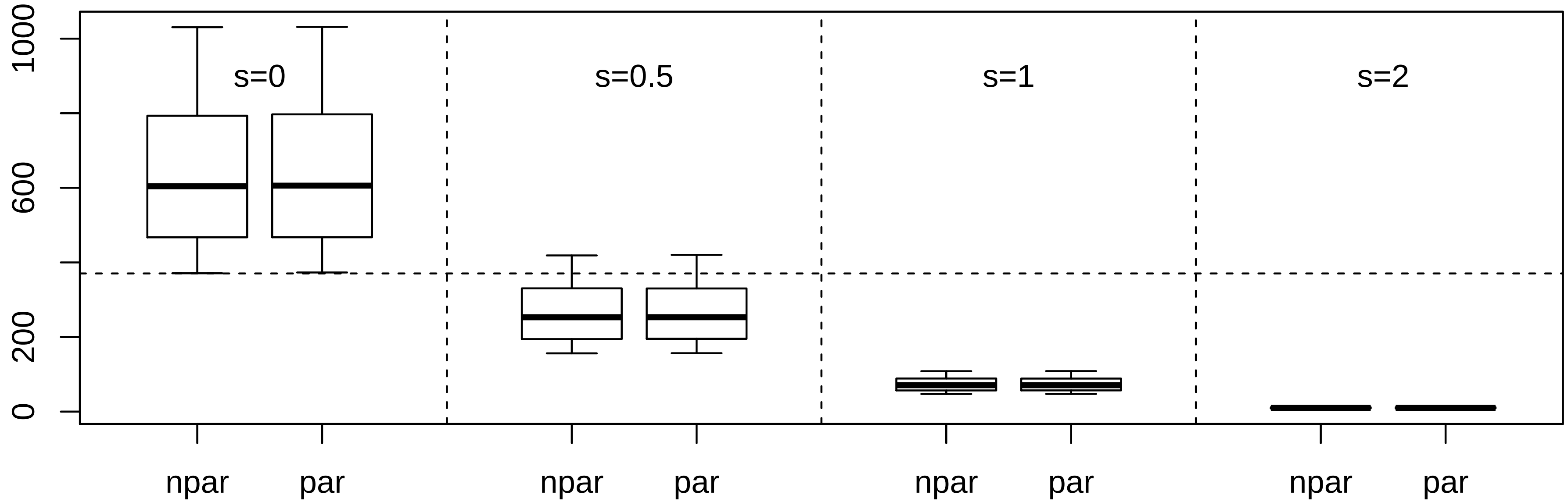
Decile box plots (whiskers end at 10 % and 90 % quantiles) of ARLs (Hall's method) for nonparametric AR(1) bootstrap with $n = 500$:



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 $n = 100$:



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 $n = 500$:



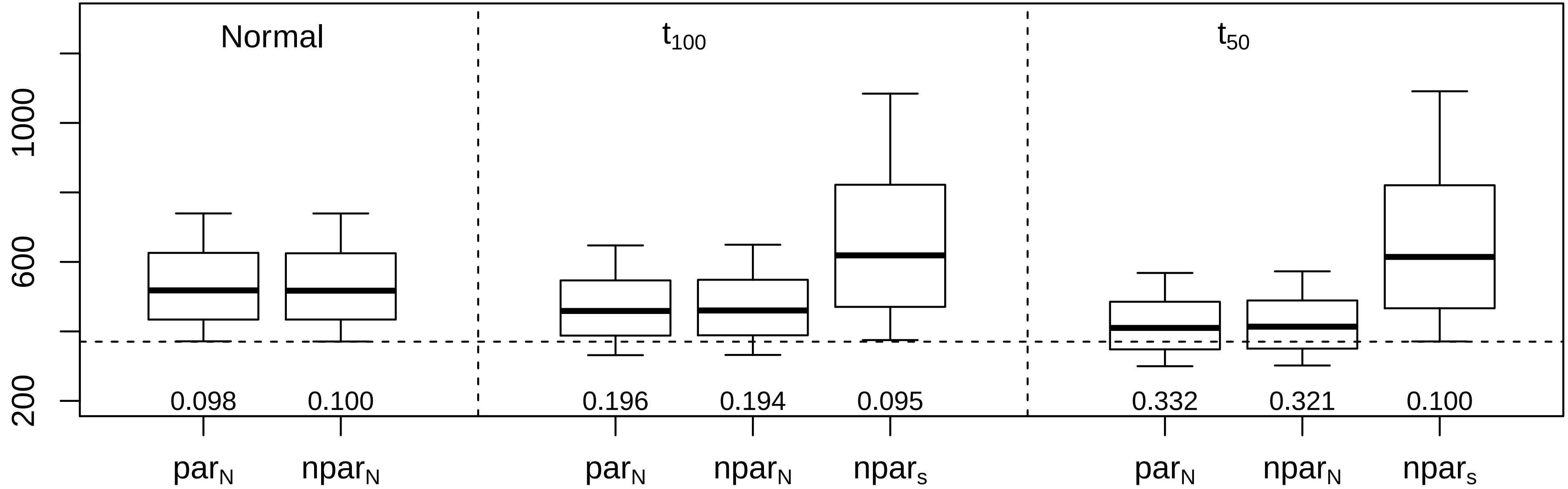
In simulations before, concerned with Gaussian AR(1) process. Normality was used for efficient ARL computation.

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.

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").

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.

Thank You

for Your Interest!

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Christian H. Weiß

Department of Mathematics & Statistics

Helmut Schmidt University, Hamburg

`weissc@hsu-hh.de`

Literature

Capizzi & Masarotto (2016) Efficient control chart ... IIE 48, 57–65.

Gandy & Kvaløy (2013) Guaranteed condit. ... Scand J Stat 40, 647–668.

Giacomini et al. (2013) A warp-speed method ... Econ Theo 29, 567–589.

Kreiss & Pap. (2011) Bootstrap methods ... J Kor Stat Soc 40, 357–378.

Knoth (2016) The spc package (statistical process control), Version 0.5.4.

Schmid (1995) On the run length of a Shewhart ... Stat Pap 36, 111–130.

Testik (2005) Model inadequacy and residuals ... QREI 21, 115–130.