Controlling Jumps in Poisson INAR(1) Processes.

Christian H. Weiß

University of Würzburg

Institute of Mathematics

Department of Statistics

For references in this talk, see

Weiß, C.H. (2008b). Controlling jumps in correlated processes of Poisson counts. Manuscript submitted to Applied Stochastic Models in Business and Industry.

Weiß, C.H. (2008a). Serial Dependence and Regression of Poisson INARMA Models. Journal of Statistical Planning and Inference 138(10), 2975-2990.

Weiß, C.H. (2007). Controlling correlated processes of Poisson counts. QREI 23(6), 741-754.

Poisson INAR(1) Processes

Definition & Properties

Definition of INAR(1) process:

Let $(\epsilon_t)_{\mathbb{N}}$ be i.i.d. process with marginal distribution $Po(\mu)$, let $\alpha \in (0;1)$. Let $N_0 \sim Po(\frac{\mu}{1-\alpha})$. If the process $(N_t)_{\mathbb{N}_0}$ satisfies

$$N_t = \alpha \circ N_{t-1} + \epsilon_t, \qquad t \geq 1,$$

plus sufficient independence conditions, then it follows a stationary *Poisson INAR(1) model* with marginal distribution $Po(\frac{\mu}{1-\alpha})$.

McKenzie (1985), Al-Osh & Alzaid (1987, 1988)

Binomial thinning, due to Steutel & van Harn (1979):

N discrete random variable with range $\{0, \ldots, n\}$ or $\mathbb{N}_0$.

Binomial thinning

$$\alpha \circ N \ := \ \sum_{i=1}^{N} X_i,$$

where X_i are independent Bernoulli trials $\sim B(1, \alpha)$.

Guarantees that right-hand side always integer-valued:

$$N_t \ = \ \alpha \circ N_{t-1} \ + \ \epsilon_t.$$

Interpretation of INAR(1) process:

$$\underbrace{N_t}_{\text{Population at time } t} = \underbrace{\alpha \circ N_{t-1}}_{\text{Survivors of time } t-1} + \underbrace{\epsilon_t.}_{\text{Immigration}}$$

Interpretation applies well to many real-world problems:

- N_t: number of users accessing web server, ϵ_t: number of new users, $\alpha \circ N_{t-1}$: number of previous users still active.
- ... and many more, see Weiß (2007).

The INAR(1) model ...

- is of simple structure,
- essential properties known explicitly,
- is easy to fit to data,
- is easy to interpret,
- applies well to real-world problems, ...

In a nutshell: A simple model for autocorrelated counts, which is well-suited for SPC!

Poisson INAR(1) Processes

Dependence and Jumps

Poisson INAR(1): Dependence & Jumps

Basic properties concerning the serial dependence structure of Poisson INAR(1) processes:

- autocorrelation $\rho(k) \;:=\; Corr[N_t, N_{t-k}] \;=\; \alpha^k$,

- $p_{k|l} \;:=\; P(N_t = k \mid N_{t-1} = l)$
$= \;\sum_{j=0}^{\min(k,l)} \binom{l}{j}\alpha^j(1-\alpha)^{l-j} \cdot P(\epsilon_t = k - j)$.

Explicit expression for

bivariate probability generating function (pgf):

$$p_{N_t,N_{t-k}}(z_1, z_2) \;=\; \exp\left(\tfrac{\mu}{1-\alpha}\,(z_1-1)\right)\cdot\exp\left(\tfrac{\mu}{1-\alpha}\,(z_2-1)\right) \cdot\exp\left(\tfrac{\mu}{1-\alpha}\,(z_1-1)(z_2-1)\cdot\alpha^k\right),$$

which is essentially a function of $\rho(k)$.

$\Rightarrow$ Particular type of bivariate Poisson distribution.

Taking partial derivatives, one obtains from $p_{N_t,N_{t-k}}(z_1, z_2)$ that

$$E[N_t \mid N_{t-k} = x] \;=\; \tfrac{\mu}{1-\alpha} \cdot (1-\alpha^k) \;+\; \alpha^k \cdot x,$$

$$V[N_t \mid N_{t-k} = x] \;=\; (1-\alpha^k) \cdot (\tfrac{\mu}{1-\alpha} \;+\; \alpha^k \cdot x),$$

also see Freeland (1998).

From $p_{N_t,N_{t-k}}(z_1,z_2)$, one can derive the explicit expression

$$P(N_t = N_{t-k} \pm j) \;=$$

$$\exp\left(-2\,\tfrac{\mu}{1-\alpha}\,(1-\alpha^k)\right)\cdot I_j\!\left(2\,\tfrac{\mu}{1-\alpha}\,(1-\alpha^k)\right), \qquad j\in\mathbb{N}_0,$$

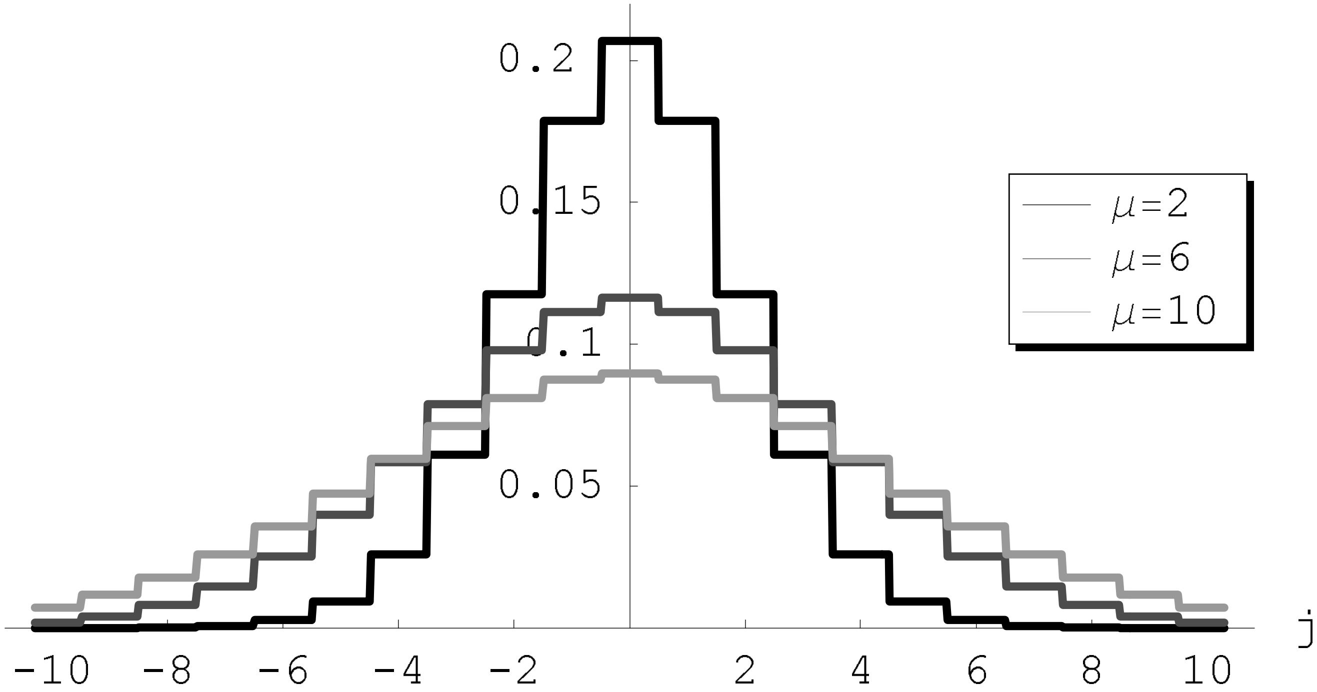
where

$$I_j(z) \;:=\; \sum_{k=0}^{\infty} \frac{(\frac{z}{2})^k\cdot(\frac{z}{2})^{k+j}}{k!\cdot(k+j)!}$$

denotes the **modified Bessel function of the first kind**.

Distribution of jumps $J_t := N_t - N_{t-1}$:

$$P(J_t = \pm j) \;=\; \exp(-2\mu) \cdot I_j(2\mu)\,, \qquad j \in \mathbb{N}_0.$$



(. . .)

⇒ **Important properties** of the distribution of jumps:

Mean and skewness of J_t are equal to 0,

its variance equals 2μ, and

the excess of J_t is given by $\frac{1}{2\mu}$.

Final remark:

Similar results are also derived for higher-order jumps ${J_t}^{(k)} := N_t - N_{t-k}$.

Also INMA(q) models show a similar dependence structure and a similar distribution of jumps.

For further details, see Weiß (2008a).

Controlling Poisson INAR(1) Processes

Control Concepts

Poisson INAR(1) model:

$(N_t)_{\mathbb{N}_0}$ is stationary Poisson INAR(1) process with innovations $(\epsilon_t)_{\mathbb{N}} \sim Po(\mu)$. So $N_t \sim Po(\frac{\mu}{1-\alpha})$.

State of statistical control: $\mu = \mu_0$ and $\alpha = \alpha_0$.

Weiß (2007) proposed the following control charts:

- c-Chart for Poisson INAR(1),
- Residual control chart,
- Conditional control chart,
- Moving average control chart.

Simulation study for ARL performance.

Disadvantages of the charts proposed by Weiß (2007):

- charts cannot be applied universally,
- often insensitive towards a change only in the serial dependence structure, i. e., where mean unaffected,
- exact ARLs are extremely difficult to obtain.

Therefore, . . .

The Combined Jumps Chart

Control Concept

The standard c-**chart**:

Observed counts N_t plotted on chart with control region

$$\mathcal{C}_c(l,u) := \{l,\ldots,u\}, \qquad l,u \in \mathbb{N}_0, \qquad 0 \leq l < \mu_{N,0} < u.$$

Process considered as being in control unless $N_t \notin \mathcal{C}_c(l,u)$.

Combined Jumps Chart – Concept

The new **jumps chart**:

Observed jumps $J_t := N_t - N_{t-1}$ plotted on chart with control region

$$\mathcal{C}_j(k) := \{-k, \ldots, k\}, \qquad k \in \mathbb{N}.$$

Process considered as being in control unless $J_t \notin \mathcal{C}_j(k)$.

Possible choice of k: $\qquad k \ := \ \left\lfloor 3 \cdot \sqrt{2\mu_0} \right\rfloor.$

Proposition: (Weiß, 2008b)

Let $(N_t)_{\mathbb{N}_0}$ be stationary Poisson INAR(1) process.

Then $(N_t, J_t)_{\mathbb{N}}$ is bivariate Markov chain, range $\mathbb{N}_0 \times \mathbb{Z}$.

Transition probabilities

$$\begin{aligned} p(n,j \mid m,i) \;&:=\; P(N_t = n, J_t = j \mid N_{t-1} = m, J_{t-1} = i) \\ &=\; \delta_{j,n-m} \cdot p_{n|m}. \end{aligned}$$

Marginal probabilities

$$p(n,j) \;:=\; P(N_t = n, J_t = j) \;=\; p_{n,n-j}.$$

Idea: Combine c- and jumps chart, i. e., monitor both N_t and J_t simultaneously.

Advantages:

- c-part sensitive to changes in the mean, jumps-part sensitive to changes in the dependence structure (decreased dependence $\Rightarrow$ larger jumps).
- $(N_t, J_t)_{\mathbb{N}}$ Markov chain $\Rightarrow$ exact ARL computation with approach of Brook & Evans (1972).

Combined Jumps Chart:

Let $l, u, k \in \mathbb{N}_0$ with $l < u$ and $k \leq u - l$.

Observed pairs (N_t, J_t) plotted simultaneously on c-chart with control region $\mathcal{C}_c(l, u)$ and jumps chart with control region $\mathcal{C}_j(k)$.

Process considered as being in control
unless $N_t \notin \mathcal{C}_c(l, u)$ or $J_t \notin \mathcal{C}_j(k)$.

Combined Jumps Chart: ARL computation.

Exact ARL computation through solving appropriate system of linear equations

$$(\mathbf{I} - \mathbf{Q}) \cdot \boldsymbol{\mu} \;=\; \mathbf{1}.$$

Dimension of matrix determined by number of reachable in-control states.

For details and proofs, see Weiß (2008b).

Real-data example:

CJ chart, design

$(l, u, k) = (2, 19, 10)$,

applied to

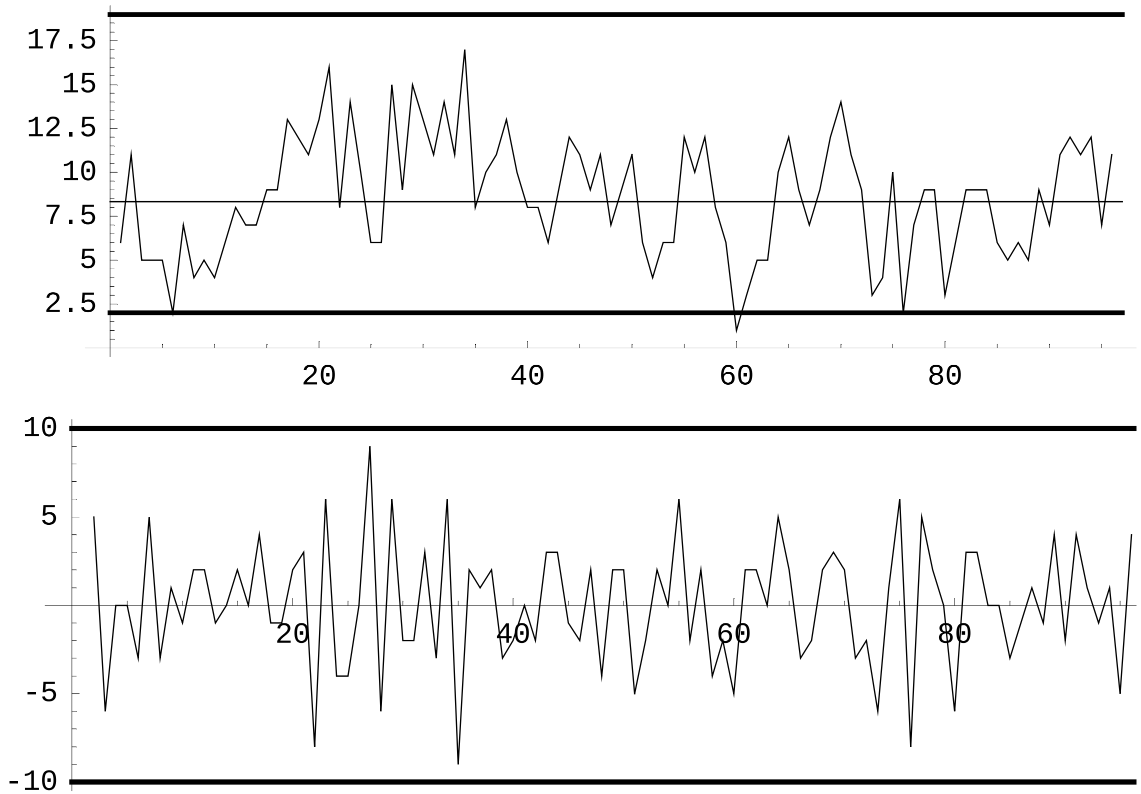
claims count data

(Freeland, 1998).

In-control model:

$\mu_0 = 5.2$

and $\alpha_0 = 0.40$.



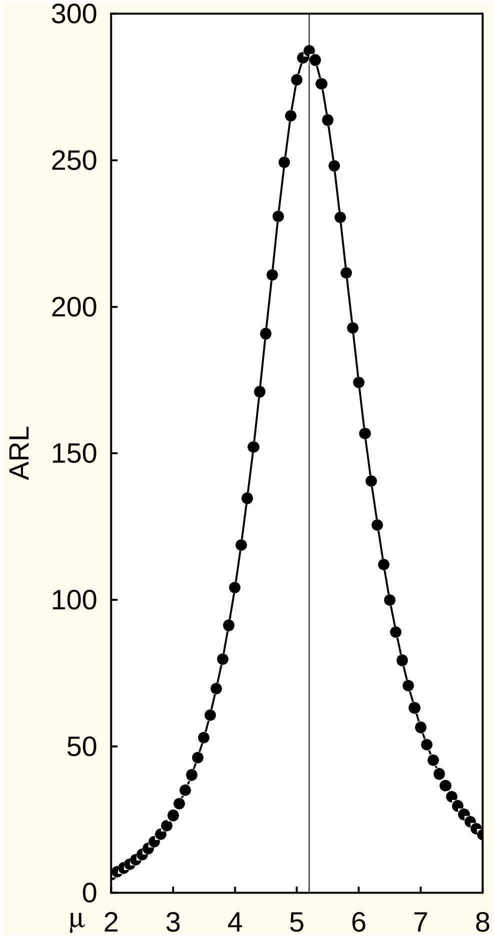
ARL **performance**

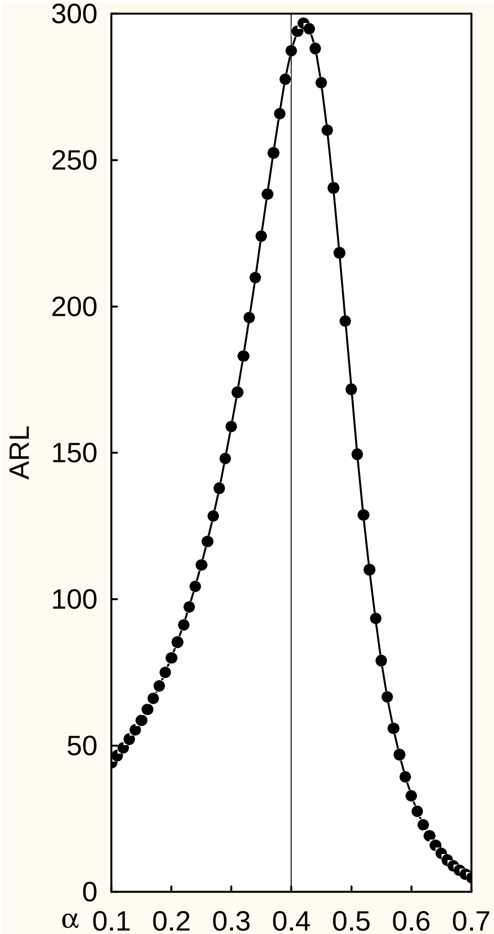
of above

CJ chart

with design

$(l, u, k) = (2, 19, 10)$:





The Combined Jumps Chart

ARL Performance

We study three types of out-of-control situations:

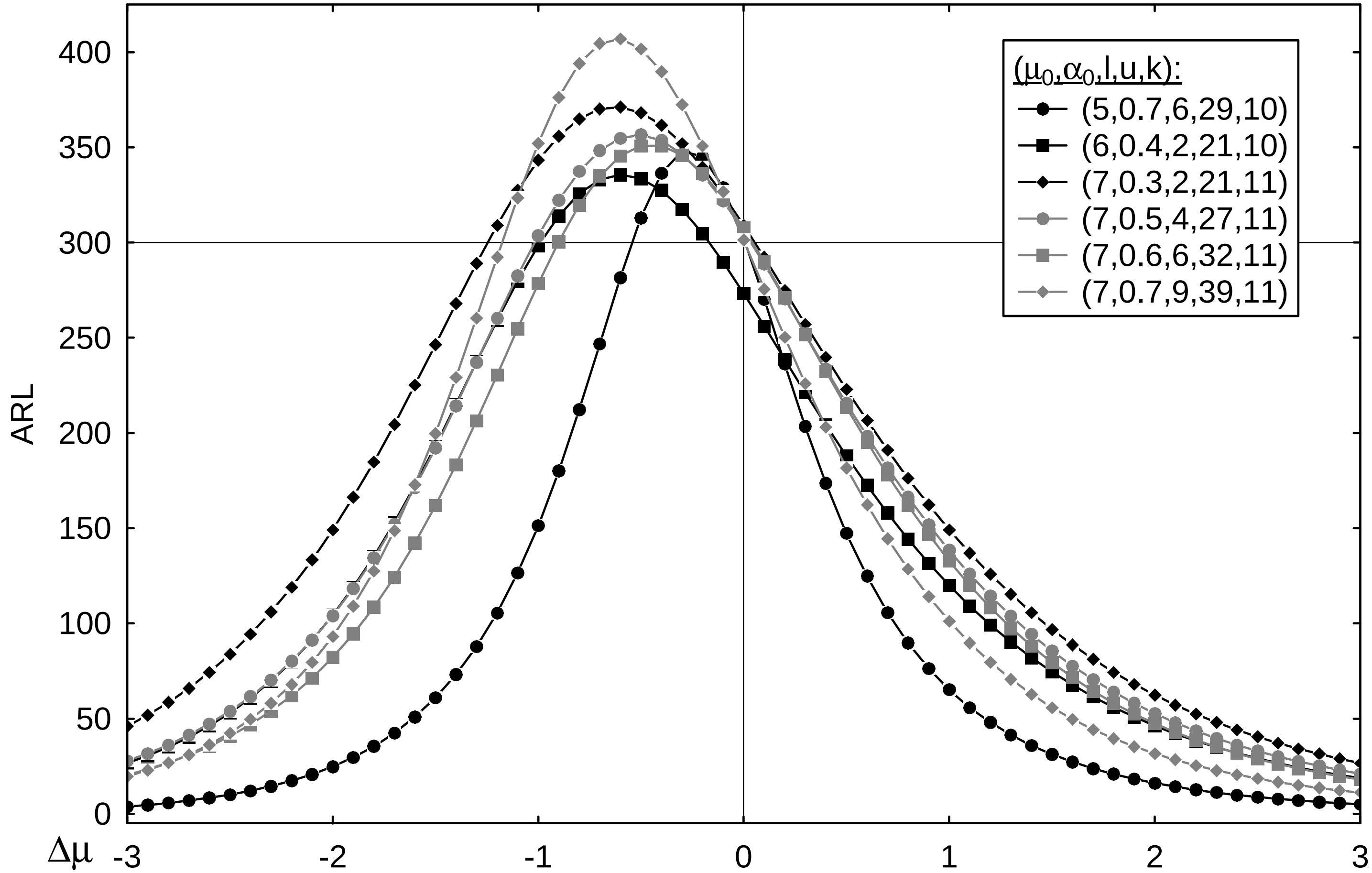
- $\alpha = \alpha_0$ is fixed, but μ varies,

- $\mu = \mu_0$ is fixed, but α varies,

$\Rightarrow$ Marginal process mean $\frac{\mu}{1-\alpha}$ is affected.

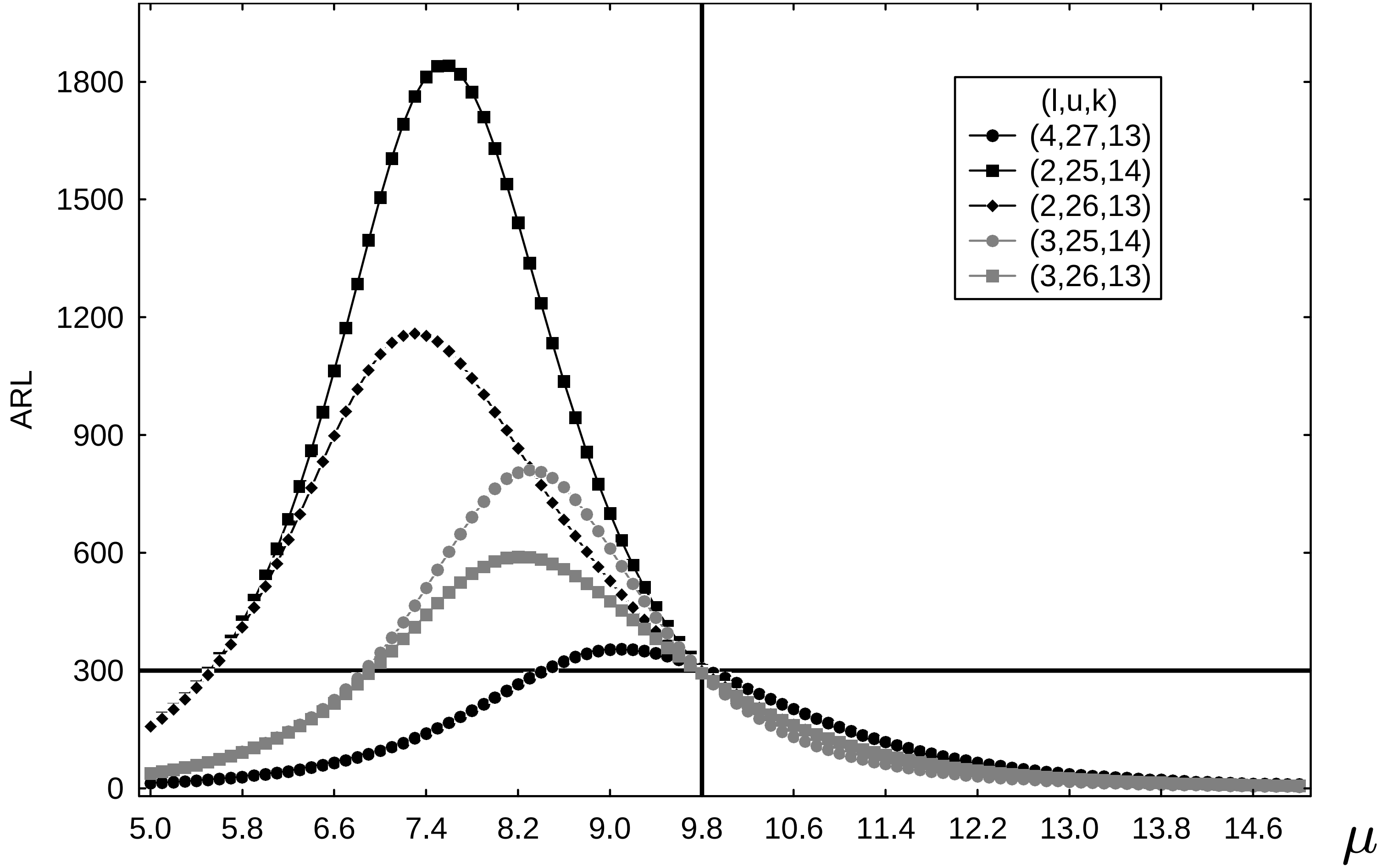
- Marginal process mean $\mu_N = \frac{\mu}{1-\alpha} = \frac{\mu_0}{1-\alpha_0}$ is fixed, but α varies, and therefore also $\mu = \mu_N \cdot (1 - \alpha)$.

Combined Jumps Chart – ARL Performance



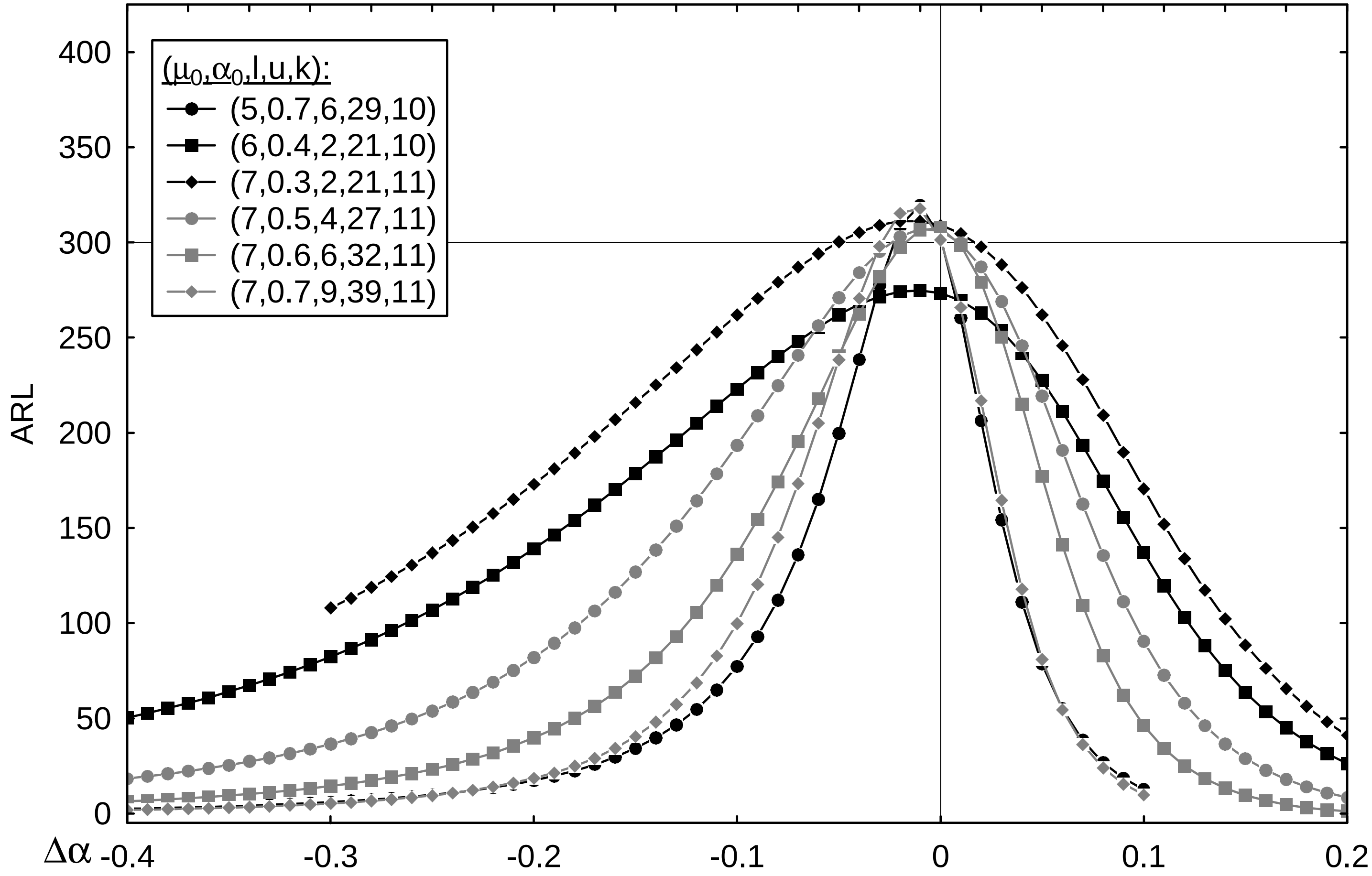
$\alpha = \alpha_0$ fixed, but μ varies: $\Delta\mu := \mu - \mu_0$.

Combined Jumps Chart – ARL Performance



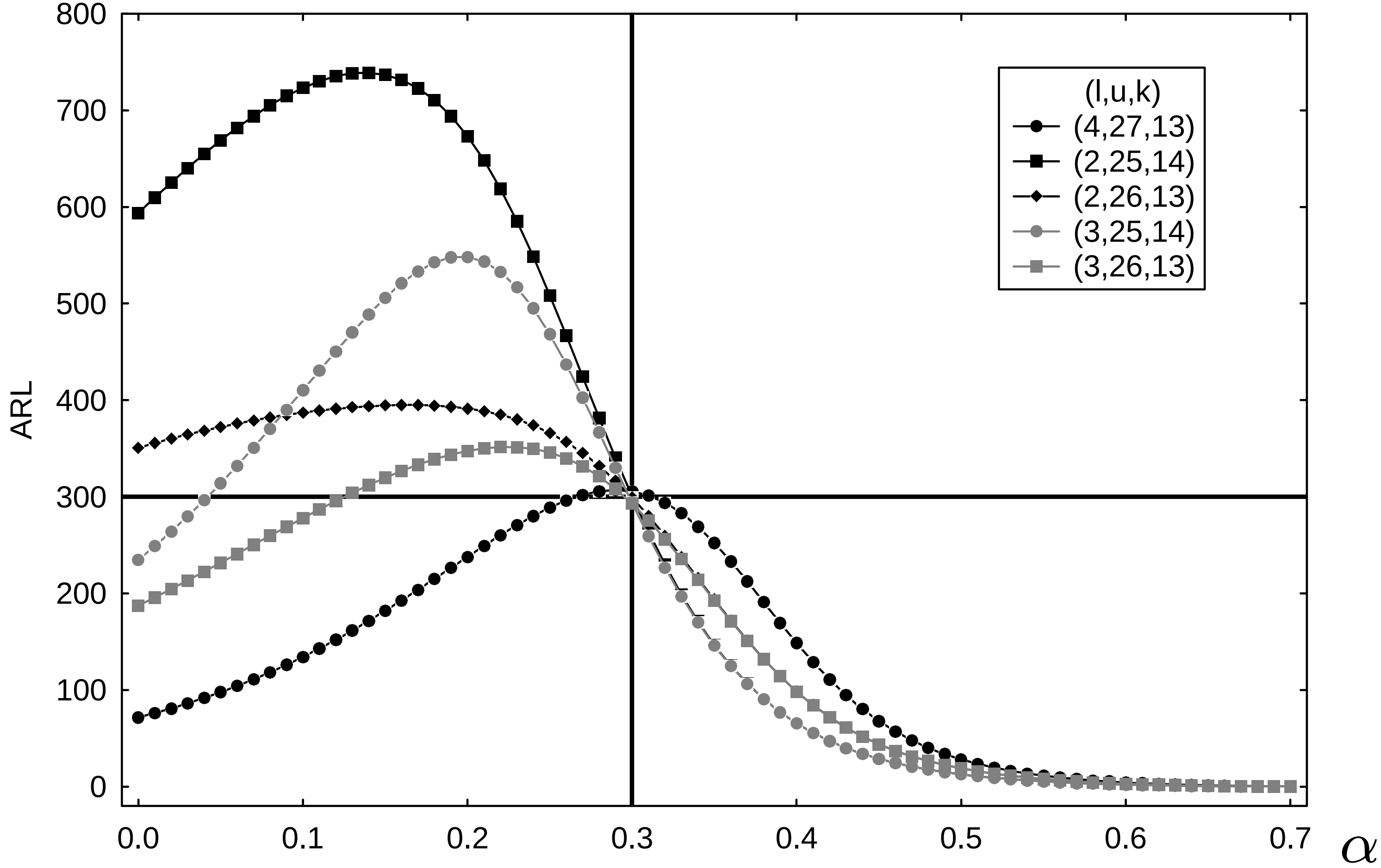
$\alpha = 0.3$ fixed, but μ varies compared to $\mu_0 = 9.8$.

Combined Jumps Chart – ARL Performance



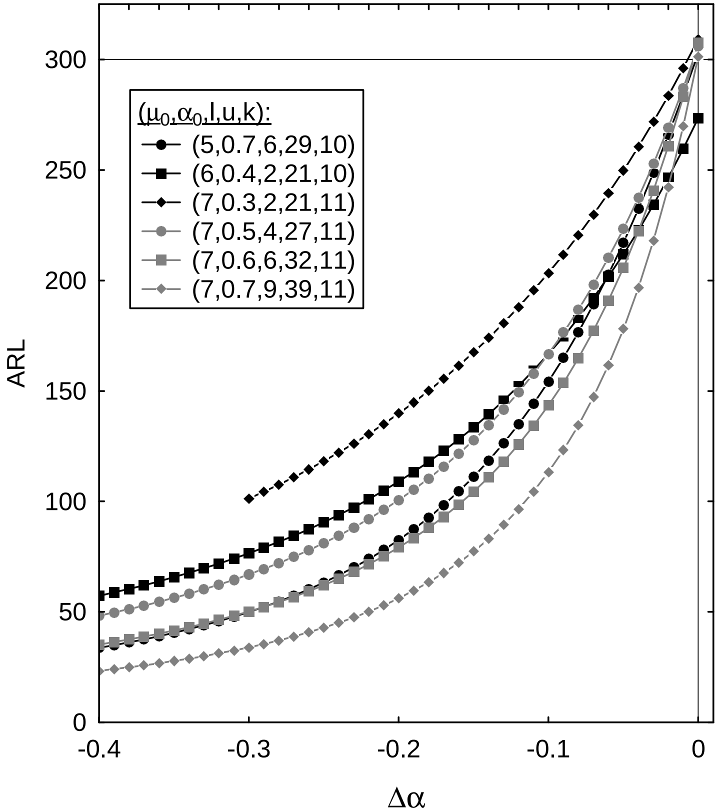
$\mu = \mu_0$ fixed, but α varies: $\Delta\alpha := \alpha - \alpha_0$.

Combined Jumps Chart – ARL Performance

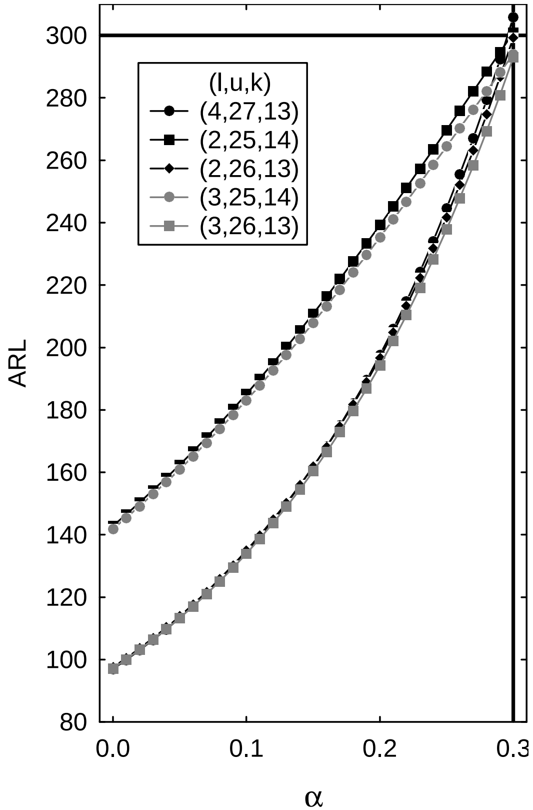


$\mu = 9.8$ fixed, but α varies compared to $\alpha_0 = 0.3$.

Combined Jumps Chart – ARL Performance



(a)



(b)

(a) $\frac{\mu}{1-\alpha} = \frac{\mu_0}{1-\alpha_0}$ fixed, but α varies: $\Delta\alpha := \alpha - \alpha_0$.

(b) $\frac{\mu}{1-\alpha} = 14$ fixed, but α varies compared to $\alpha_0 = 0.3$.

- **INAR(1) model:**
  Simple, easily interpretable model, well-suited for real-world problems from SPC.
  New results concerning serial dependence structure and distribution of jumps.

- **Combined Jumps chart:**
  Exact ARL computation with Markov chain approach, sensitive to various types of out-of-control situations, only three design parameters.
  But design has to be selected carefully!

Thank You
for Your Interest!

Christian H. Weiß

University of Würzburg

Institute of Mathematics

Department of Statistics