On PMF-Forecasting for Count Processes

C.H. Weiß, A. Homburg, G. Frahm

Department of Mathematics & Statistics,
Helmut Schmidt University, Hamburg

L.C. Alwan

Sheldon B. Lubar School of Business,
University of Wisconsin-Milwaukee

R. Göb

Department of Statistics,
University of Würzburg

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH
STAT

Outline

- On Models for Count Time Series
- Coherent Forecasting of Count Processes
- DFG-Research Project (Projektnummer 394832307):
  "Coherent Forecasting and Risk Analysis for Count Processes"
- **On PMF-Forecasting for Count Processes**
  (as part of *"Proceedings of ITISE 2021"*)

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

On Models for Count Time Series

Brief Survey

On Models for Count Data

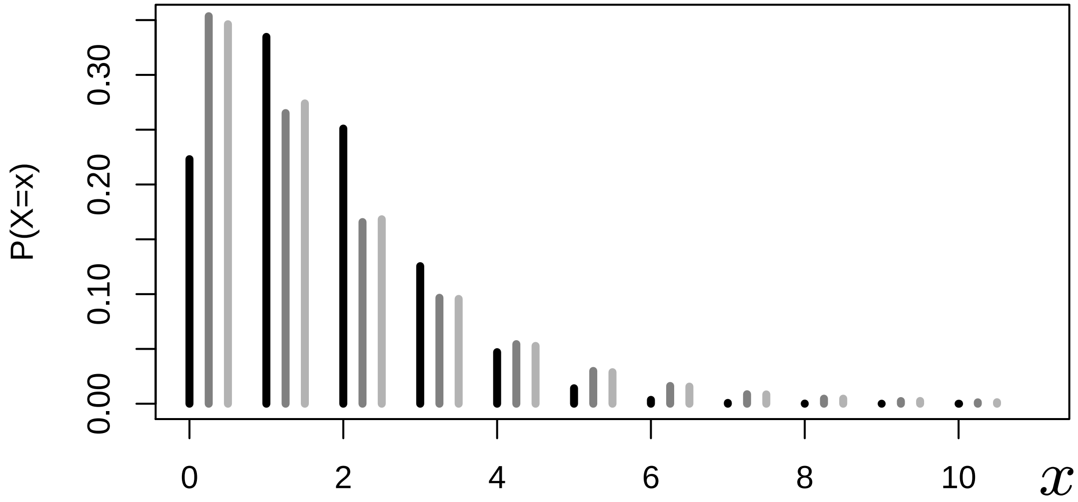
Count data: X is (unbounded) count random variable if X takes only values from $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$.

Example: Random number of e-mails at some day, etc.

Default choice: **Poisson distribution** Poi(λ),

$$P(X = x) \; = \; e^{-\lambda} \cdot \frac{\lambda^x}{x!}.$$

Equidispersion: $E[X] = \lambda = V[X]$.



Equidispersed Poisson (black) vs. **overdispersion** (grey):

$$V[X] > E[X],$$

e. g., negative-binomial distribution.

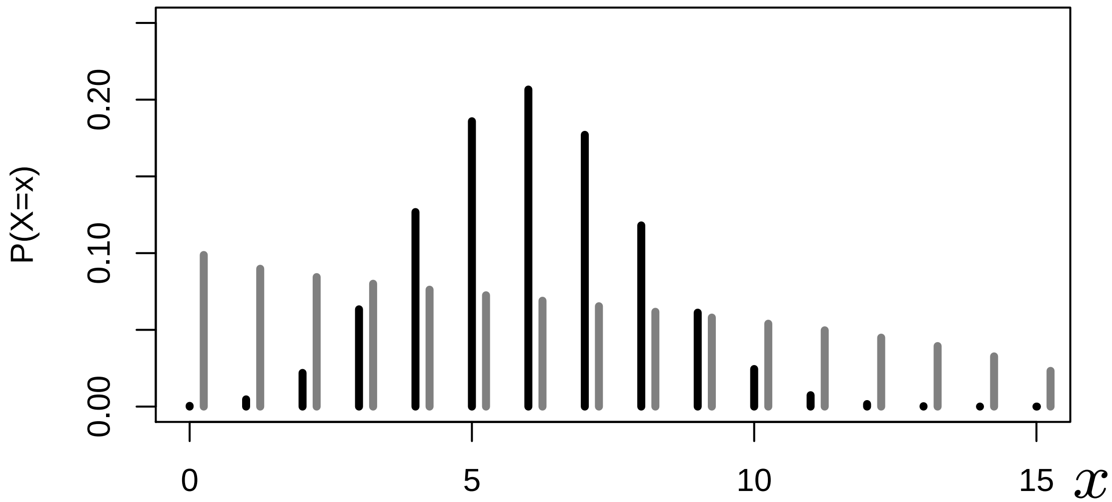
Bounded counts: X with range $\{0,\ldots,n\}$ for given $n \in \mathbb{N}$.

Example: occupied rooms in hotel with n rooms, etc.

Default choice: **Binomial distribution** $\text{Bin}(n,\pi)$,

$$P(X=x) \;=\; \binom{n}{x}\cdot\pi^x\,(1-\pi)^{n-x}.$$

Binomial dispersion: $n\,V[X] = E[X]\,(n-E[X])$.



Binomial (black) vs. **extra-binomial variation** (grey) e. g., beta-binomial distribution.

Time series: chronolog. ordered sequence of observ. $(x_t)_{t\in\mathcal{T}_0}$.

Example: stock prices, water levels, population . . .

If observations are count values, i. e.,

$x_t \in \mathbb{N}_0 = \{0, 1, 2, \ldots\} \Rightarrow$ **count time series**.

Time series of interest in statistics if

values stem from random phenomenon $\to$ **Process**:

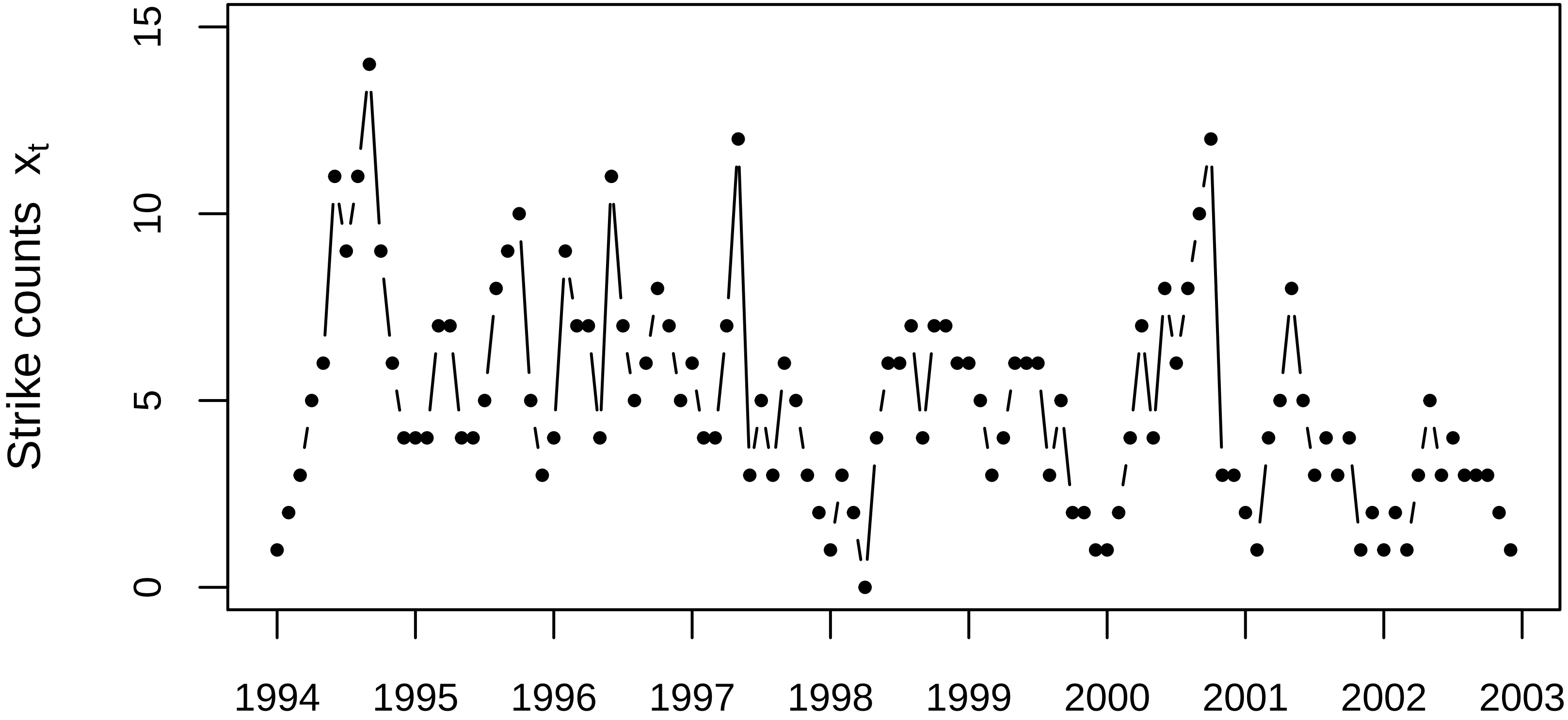
family of random variables $(X_t)_{\mathcal{T}}$, realized at times $t \in \mathcal{T}$.

Time series: realizations $(x_t)_{\mathcal{T}_0}$ from process $(X_t)_{\mathcal{T}}$,

where $\mathcal{T}_0 \subseteq \mathcal{T}$.

Example 1: Monthly counts of "major strikes" (1994–2002): strikes and lock-outs of $\geq 1\,000$ workers.

Source: U. S. Bureau of Labor Statistics, Weiß (2018).



Example 1: Descriptive statistics.

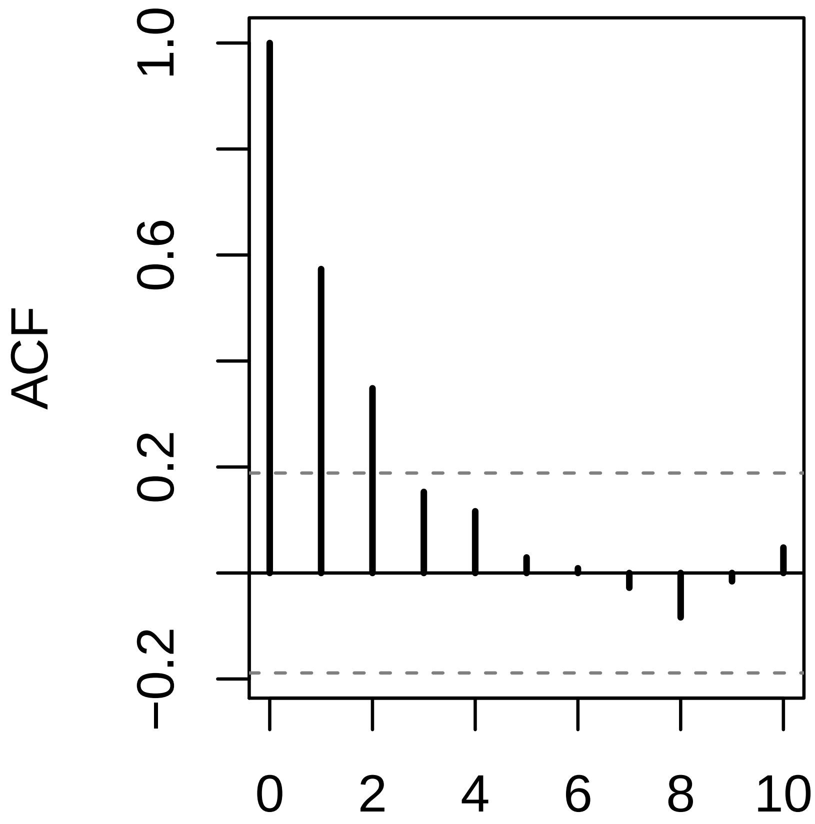
Mean 4.94,

variance 7.85,

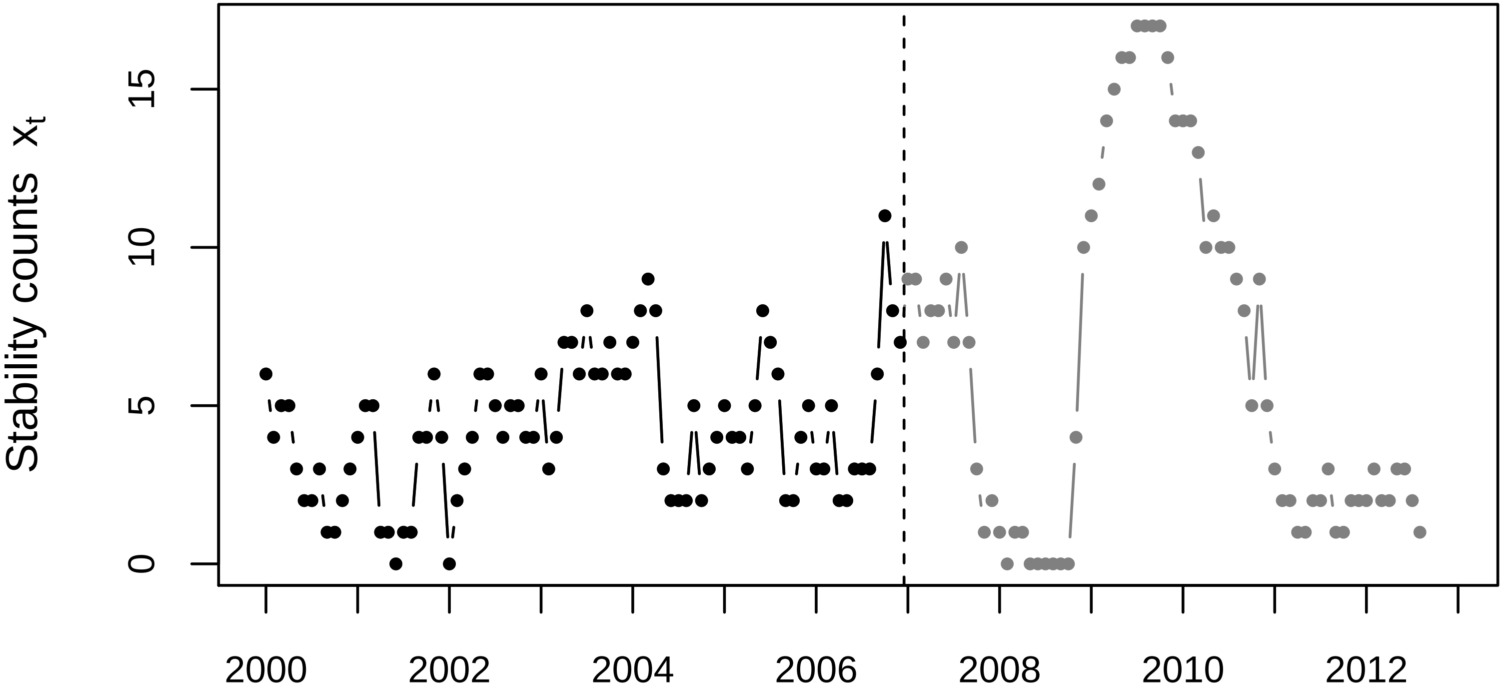
hence overdispersion.

This contradicts
a Poisson model.

Autocorrelation function (ACF):

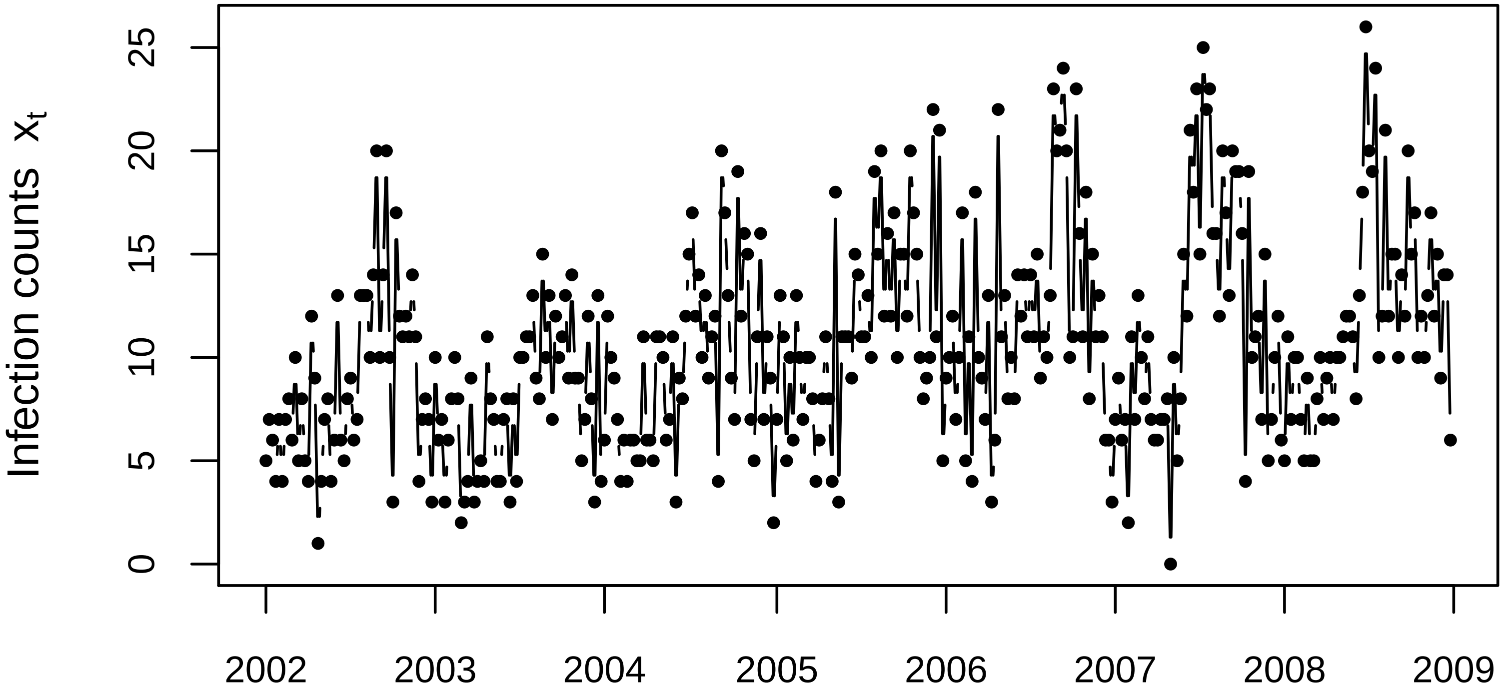


Example 2: Monthly counts of "EA17" countries with stable prices (i.e., inflation $< 2\,\%$), Jan. 2000 to Aug. 2012. So upper bound $n = 17$. Source: Weiß & Kim (2014).



Example 3: Weekly number of new infections with Legionnaires' disease in Germany, 2002–2008.

Source: Robert-Koch-Institut, Weiß (2018).



Possible features for real-world count time series:

- typically low counts,
- overdispersion or zero inflation, and
- serial dependence;
- sometimes structural changes, or
- seasonality or trend.

How to model count time series?

Basic approach for **real-valued** processes:

stationary **ARMA(p,q) model**.

Let innovations $(\epsilon_t)_{\mathbb{Z}}$ be white noise, then

$$X_t \;=\; \alpha_1 \cdot X_{t-1} + \ldots + \alpha_p \cdot X_{t-p} + \epsilon_t + \beta_1 \cdot \epsilon_{t-1} + \ldots + \beta_q \cdot \epsilon_{t-q},$$

where $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \in \mathbb{R}$ suitably chosen.

ACF via **Yule–Walker equations**.

Enumerous extensions: SARIMA, ARFIMA, GARCH, . . .

Not applicable to count processes: generally, $\alpha \cdot X \notin \mathbb{N}_0$.

Tailor-made models

required for count time series, e.g.,

- thinning-based models ($\to$ INARMA),
- regression models ($\to$ INGARCH),
- hidden Markov models, ...

In the sequel: look at thinning-based and regression models for count data-generating processes (DGPs).

Conventional ARMA models: “multiplication problem”.

Possible solution: Use appropriate thinning operations.

Most popular:

binomial thinning operator (Steutel & van Harn, 1979):

$\alpha \circ X \ \sim\ \text{Bin}(X, \alpha)$ with $\alpha \in (0; 1)$; has range $\{0, \ldots, X\}$.

In particular, $E[\alpha \circ X] \ =\ E[\alpha \cdot X]$.

($\approx$ number of “survivors” from population of size X)

Let $(\epsilon_t)_{\mathbb{Z}}$ be i. i. d. with range $\mathbb{N}_0 = \{0, 1, \ldots\}$, denote $E[\epsilon_t] = \mu_\epsilon$, $V[\epsilon_t] = \sigma_\epsilon^2$. Let $\alpha \in (0; 1)$.

$(X_t)_{\mathbb{Z}}$ referred to as **INAR(1) process** if

$$\underbrace{X_t}_{\text{Population at time } t} \quad = \quad \underbrace{\alpha \circ X_{t-1}}_{\text{Survivors of time } t-1} \quad + \quad \underbrace{\epsilon_t}_{\text{Immigration}} \quad ,$$

plus appropr. independence assumptions. (McKenzie, 1985)

Mean, dispersion ratio, and ACF given by

$$\mu \;=\; \frac{\mu_\epsilon}{1-\alpha}\,, \qquad I = \frac{\sigma^2}{\mu} \;=\; \frac{\frac{\sigma_\epsilon^2}{\mu_\epsilon} + \alpha}{1+\alpha}, \qquad \text{and} \qquad \rho(k) \;=\; \alpha^k.$$

INAR(1) process:

$$X_t \;=\; \alpha \circ X_{t-1} \;+\; \epsilon_t.$$

INAR(1) process constitutes Markov chain with

transition probabilities $p(x|x_T) = p(X_{T+1} = x | X_T = x_T)$ as

$$p(x|x_T) \;=\; \sum_{s=0}^{\min\{x,x_T\}} \binom{x_T}{s} \alpha^s (1-\alpha)^{x_T - s} \cdot P(\epsilon_t = x - s).$$

Referred to as Poi-, NB-, or ZIP-INAR(1) model, respectively,

if ϵ_t Poisson, negative binomial, or zero-inflated Poisson.

INAR(2) process by Du & Li (1991):

$$X_t \;=\; \alpha_1 \circ X_{t-1} \;+\; \alpha_2 \circ X_{t-2} \;+\; \epsilon_t \quad \text{with } \alpha_1 + \alpha_2 < 1.$$

INAR(2) process constitutes second-order Markov process

with transition probabilities

$$p(x|x_T, x_{T-1}) \;=\; \sum_{j_1=0}^{\min\{x, x_T\}} \sum_{j_2=0}^{\min\{x-x_T, x_{T-1}\}}$$
$$\binom{x_T}{j_1}\, \alpha_1^{j_1}\, (1-\alpha_1)^{x_T-j_1} \cdot \binom{x_{T-1}}{j_2}\, \alpha_2^{j_2}\, (1-\alpha_2)^{x_{T-1}-j_2} \cdot P(\epsilon_t = x - j_1 - j_2).$$

ACF satisfies

$$\rho(1) \;=\; \frac{\alpha_1}{1-\alpha_2}, \qquad \rho(k) \;=\; \alpha_1\, \rho(k-1) + \alpha_2\, \rho(k-2) \text{ for } k \geq 2.$$

BinAR(1) model by McKenzie (1985):

AR(1)-like model for bounded range $\{0,\ldots,n\}$ with some $n \in \mathbb{N}$.

Let $\pi \in (0,1)$ and $\alpha \in \Big(\max\{-\frac{\pi}{1-\pi}, -\frac{1-\pi}{\pi}\}, 1\Big)$ and define $\beta := \pi(1-\alpha)$ and $\gamma := \beta + \alpha$. Then,

$$X_t \;=\; \gamma \circ X_{t-1} \;+\; \beta \circ (n - X_{t-1}) \qquad \text{with } X_0 \sim \text{Bin}(n,\pi).$$

Markov chain with Bin(n,π)-marginal and ACF $\rho(k) = \alpha^k$,

and with transition probabilities $\quad p(x|x_T) \;=$

$$\sum_{m=\max\{0,x+x_T-n\}}^{\min\{x,x_T\}} \binom{x_T}{m}\binom{n-x_T}{x-m}\,\gamma^m(1-\gamma)^{x_T-m}\,\beta^{x-m}(1-\beta)^{n-x_T+m-x}.$$

Regression-type models: solve "multiplication problem" by applying ARMA-like recursion to conditional means

$\rightarrow$ INGARCH models (Ferland et al., 2006).

Poi-INARCH(1) model: $X_t|X_{t-1},\ldots \sim \text{Poi}(\beta + \alpha\, X_{t-1})$

with $\beta > 0$ and $\alpha \in (0,1)$. Mean, variance, and ACF are

$$\mu = \frac{\beta}{1-\alpha}\,, \quad \sigma^2 = \frac{\mu}{1-\alpha^2}, \quad \text{and} \quad \rho(k) = \alpha^k.$$

Markov chain with transition probabilities

$$p(x|x_T) \;=\; \exp(-\beta - \alpha\, x_T)\, \frac{(\beta + \alpha\, x_T)^x}{x!}.$$

Poi-INARCH(2) model:

$$X_t | X_{t-1}, \ldots \ \sim \ \text{Poi}(\beta + \alpha_1\, X_{t-1} + \alpha_2\, X_{t-2})$$

with $\alpha_1 + \alpha_2 < 1$ and ACF like for the INAR(2) model.

Second-order Markov process with transition probabilities

$$\begin{aligned} & p(x | x_T, x_{T-1}) \\ = \ & \exp(-\beta - \alpha_1\, x_T - \alpha_2\, x_{T-1})\, \frac{(\beta + \alpha_1\, x_T + \alpha_2\, x_{T-1})^x}{x!}. \end{aligned}$$

BinARCH(1) model for bounded counts:

$$X_t|X_{t-1},\ldots \sim \mathsf{Bin}\left(n, \beta + \alpha\,\frac{X_{t-1}}{n}\right)$$

with $\beta, \beta + \alpha \in (0, 1)$ and transition probabilities

$$p(x|x_T) = \binom{n}{x} \left(\beta + \alpha\,\tfrac{x_T}{n}\right)^x \left(1 - \beta - \alpha\,\tfrac{x_T}{n}\right)^{n-x}.$$

Log-linear **ll-Poi-AR(1) model** for non-stationary counts,
e. g., with linear trend and harmonic oscillation
(period p, angular frequency $\omega = 2\pi/\text{p}$):

$$X_t | X_{t-1}, \ldots \ \sim \ \text{Poi}(M_t) \text{ with}$$

$$\ln M_t \ = \ \overbrace{\gamma_0 + \gamma_1\, t \ + \ \gamma_2 \cos(\omega t) + \gamma_3 \sin(\omega t)}^{=:\, \ln \mu_t}$$

$$+ \ \alpha_1 \big(\ln (X_{t-1} + 1) - \ln (\mu_{t-1} + 1) \big).$$

Additional dispersion via conditional NB distribution:
ll-NB-AR(1) model assumes $X_t | X_{t-1}, \ldots \ \sim \ \text{NB}\left(1, \frac{n}{M_t + n}\right)$,
where parameter $n > 0$ controls dispersion level.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Coherent Forecasting of Count Processes

Aims & Methods

Let $x_1, \ldots, x_T$ be **count time series**

from underlying **count process** $(X_t)_{t\in\mathbb{Z}=\{\ldots,-1,0,1,\ldots\}}$.

Aim: forecasting of X_{T+h} given the past $x_T, \ldots, x_1$.

Computing, e. g., conditional mean as point forecast (PF) value

does not make sense, because

$E[X_{T+h}|x_T, \ldots, x_1]$ (positive) real number, not count value.

Coherent forecasting: computed forecasts of count process

should be count values themselves.

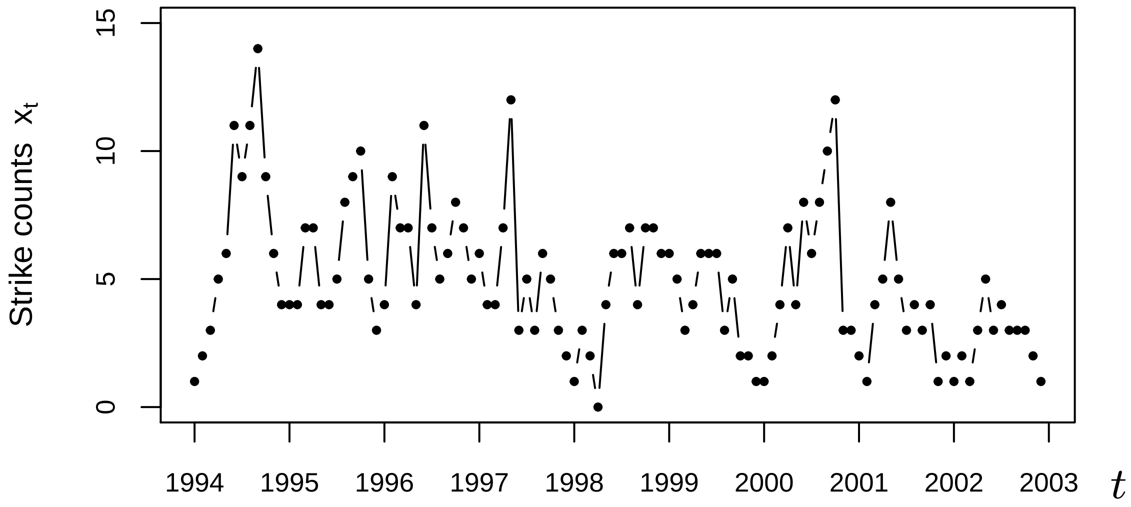
(Freeland & McCabe, 2004; Jung & Tremayne, 2006)

Coherent forecasting achieved by first deriving full h-step-ahead conditional distribution, and then computing

- median (mode) of $X_{T+h}\big|x_T,\ldots,x_1$ as **central point forecast** (PF);
- extreme (upper) quantile of $X_{T+h}\big|x_T,\ldots,x_1$ as **non-central PF**; or
- finite subset of $\mathbb{N}_0$ satisfying coverage requirement as discrete **prediction interval** (PI) for $X_{T+h}\big|x_T,\ldots,x_1$;

see Homburg et al. (2019, 2021) for details.

Recall **Example 1**, counts $x_1, \ldots, x_{108}$ of monthly "major strikes" in U. S. (1994–2002):



Weiß (2018) uses Poi-INARCH(1) model.

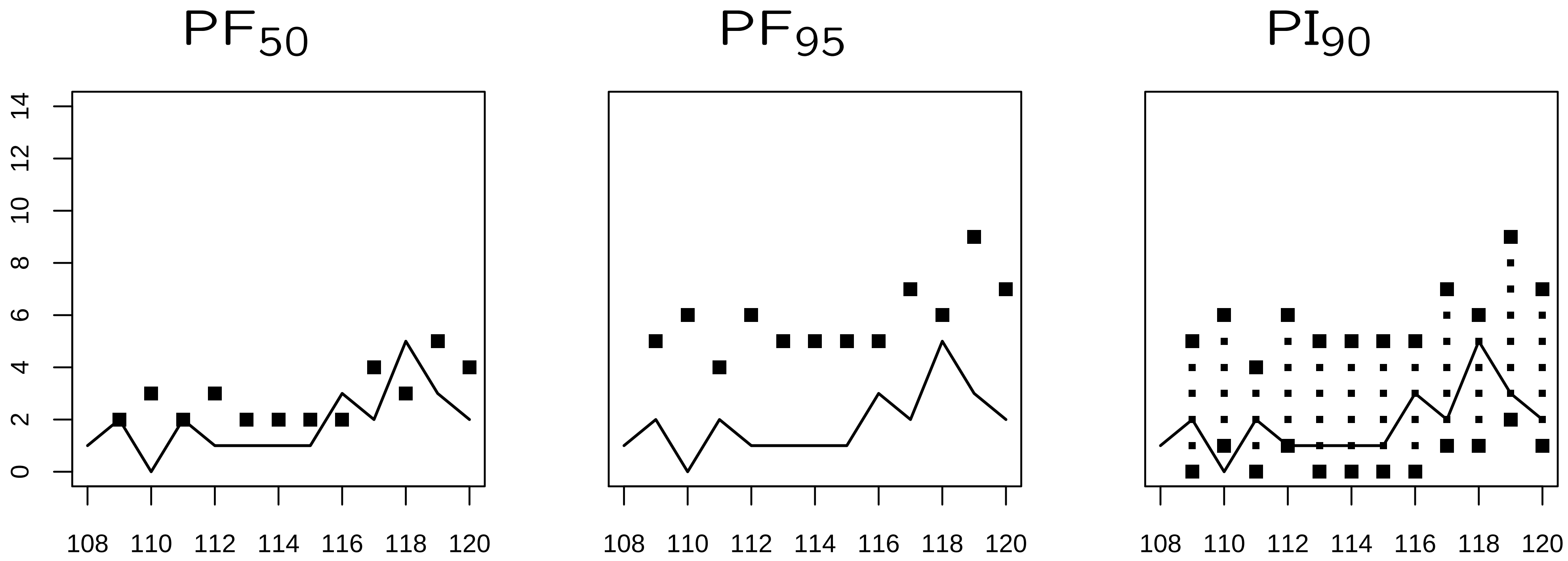
ML estimates (s. e.)

$\widehat{\mu}_{\mathrm{ML}} \approx 4.981\ (0.593)$,

$\widehat{\alpha}_{\mathrm{ML}} \approx 0.636\ (0.081)$.

Out-of-sample forecasting for counts $x_{109}, \ldots, x_{120}$ in 2003, which are $2, 0, 2, 1, 1, 1, 1, 3, 2, 5, 3, 2,$ by successive 1-step-ahead forecasting based on Poi-INARCH(1) model.

Different types of coherent forecasts for 2003 based on fitted Poi-INARCH(1) model (true values shown as line):



PFs rarely agree with actual future observations X_{T+h}
(for real-valued processes, agreement probability even 0,
while truly positive for discrete count processes).

PIs for discrete count processes often true coverage probability
much larger than given coverage requirement
(while in real-valued case, exact match possible).

Thus, several authors (e. g., Boylan & Synteto, 2006; Willemain,
2006; McCabe et al., 2011; Snyder et al., 2012; Kolassa, 2016)
use full predictive probability mass function as forecast value:
PMFFs to judge which outcome with which probability.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

DFG-Research Project:

“Coherent Forecasting and Risk Analysis for Count Processes”

Overview

Coherent Forecasting and Risk Analysis

for Count Processes: (→ project website)

Three-years project, funded by Deutsche Forschungsgemeinschaft (DFG) — Projektnummer 394832307.

PhD candidate: Annika Homburg

Project partners:

- Layth C. Alwan, University of Wisconsin, USA,
- Gabriel Frahm, HSU Hamburg,
- Rainer Göb, University of Würzburg.

Project aims:

Comprehensive performance analysis . . .

- of all different types of coherent forecasts
- for a multitude of models (bounded vs. unbounded counts, stationary vs. non-stationary DGPs, different model orders, special features such as overdispersion or zero inflation, etc.).

For this purpose, appropriate criteria for evaluation and efficient comparison are to be specified.

(. . .)

Project aims:

Although various models for count time series available, practitioners often use Gaussian approximations, e. g., by fitting Gaussian ARIMA models to count time series.

Typical reasons are insufficient communication of count models and ease of implementation of Gaussian ARMA models because of readily available software solutions.

Computing forecasts based on Gaussian approximation (plus appropriate discretization) may cause substantial forecast errors.

(. . .)

Project aims:

Thus, we systematically compare coherent and approximative forecasting: *"Does model-based coherent forecasting lead to notable added value for prediction of count data?"*

Finally,

non-central PFs (extreme quantiles) also for risk analysis.

Risk prediction by quantile forecast and deduced risk measures (e.g., expected shortfall, expectiles, mid-quantiles).

Effect of estimation uncertainty?

Evaluation of the "goodness" of risk prediction?

Project results:

- Homburg et al. (2019): Evaluating Approximate Point Forecasting of Count Processes. *Econometrics* **7**(3), 30. (→ open access)

Abstract: In forecasting count processes, practitioners often ignore the discreteness of counts and compute forecasts based on Gaussian approximations instead. For both central and non-central point forecasts, and for various types of count processes, the performance of such approximate point forecasts is analyzed. (...) We conclude that Gaussian forecast approximations should be avoided.

Project results:

- Homburg et al. (2021a): A Performance Analysis of Prediction Intervals for Count Time Series. *Journal of Forecasting* **40**(4), 603–625. (→ open access)

  Abstract: (. . .) The use of interval forecasts instead of point forecasts allows us to incorporate the apparent forecast uncertainty. When forecasting count time series, one also has to account for the discreteness of the range, which is done by using coherent prediction intervals (PIs) relying on a count model. We provide a comprehensive performance analysis of coherent PIs for diverse types of count processes. (. . .)

Project results:

- Weiß et al. (2021): Efficient Accounting for Estimation Uncertainty in Coherent Forecasting of Count Processes. *Journal of Applied Statistics*, in press. ($\rightarrow$ open access)

  Abstract: (. . .) In practice, forecasting always relies on a fitted model and so the obtained forecast values are affected by estimation uncertainty. (. . .) We propose a computationally efficient resampling scheme that allows to express the uncertainty in common types of coherent forecasts for count processes. (. . .) the obtained ensembles of forecast values can be presented in a visual way that allows for an intuitive interpretation.

Project results:

- Homburg et al. (2021b): Analysis and Forecasting of Risk in Count Processes. (→ open access) *Journal of Risk and Financial Management* **14**(4), 182.

Abstract: Risk measures are commonly used to prepare for a prospective occurrence of an adverse event. (. . .) It becomes clear that Gaussian approximate risk forecasts substantially distort risk assessment and, thus, should be avoided. In order to account for the apparent estimation uncertainty in risk forecasting, we use bootstrap approaches for count time series. (. . .)

Project results:

- Homburg et al. (2021c): On PMF-Forecasting for Count Processes.

  In *Proceedings of ITISE 2021*, in press.

. . . **this is the topic of the remaining talk!**

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

On PMF-Forecasting for Count Processes

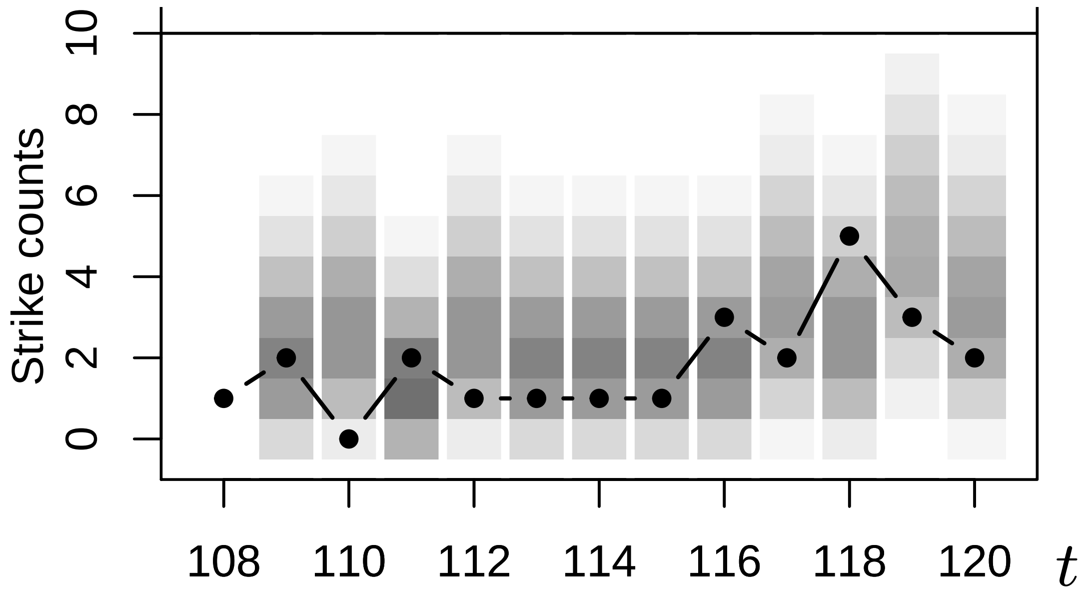
Introduction

PMF-forecasts (PMFFs) provide full predictive PMF as forecast value itself, to judge "plausibility" of future outcomes.

In analogy to Weiß et al. (2021), PMFF at time t plotted as vertical band of gray levels, where intensity proportional to probability (white = zero probability).

Example of strikes counts for year 2003,

PMFFs based on fitted Poi-INARCH(1) model:



Research questions:

- How evaluate performance of PMF forecasting?
- Performance of coherent PMFFs (relying on count model) under estimation uncertainty?
- Performance of approximate PMFFs (reyling on Gaussian ARIMA approximation) under estimation uncertainty?
- Application of PMFFs in practice?

Notations:

Count DGP (X_t) follows model with parameter vector $\boldsymbol{\theta}$.

If parameter vector estimated, we write $\widehat{\boldsymbol{\theta}}$.

If fitting Gaussian ARMA model to $x_1, \ldots, x_T$,

then ARMA parameters $\boldsymbol{\vartheta}$ with estimate $\widehat{\boldsymbol{\vartheta}}$.

True PMFF of X_{T+h} given $x_T, \ldots, x_1$: $\quad \widehat{\boldsymbol{p}}_{T+h}(\boldsymbol{\theta})$,

where xth component $\widehat{p}_{T+h,x}(\boldsymbol{\theta}) = P(X_{T+h} = x \,|\, x_T, \ldots, x_1)$.

Coherent PMFF of X_{T+h} using estimates: $\quad \widehat{\boldsymbol{p}}_{T+h}\big(\widehat{\boldsymbol{\theta}}\big)$.

Approximate PMFF of X_{T+h}: $\quad \widehat{\boldsymbol{p}}_{T+h,\mathrm{a}}\big(\widehat{\boldsymbol{\vartheta}}\big)$.

Equivalently, use cumulative distribution function (CDF), denoted as $\widehat{\boldsymbol{f}}_{T+h}(\boldsymbol{\theta})$, $\widehat{\boldsymbol{f}}_{T+h}(\widehat{\boldsymbol{\theta}})$, and $\widehat{\boldsymbol{f}}_{T+h,\mathrm{a}}(\widehat{\boldsymbol{\vartheta}})$, respectively.

Here, xth component $\widehat{f}_{T+h,x}(\boldsymbol{\theta}) = P(X_{T+h} \leq x \,\big|\, x_T, \ldots, x_1)$.

Approximate PMFF $\widehat{\boldsymbol{p}}_{T+h,\mathrm{a}}(\widehat{\boldsymbol{\vartheta}})$ derived from conditional normal distribution of ARMA approximation, say $\mathrm{N}\big(\widehat{\mu}_{T+h}, \widehat{\sigma}^2_{T+h}\big)$. Let Φ be CDF of $\mathrm{N}(0,1)$, then for $x \in \mathbb{N}_0$,

- **simple normal appr.**: $\quad \widehat{f}_{T+h,\mathrm{a},x}(\widehat{\boldsymbol{\vartheta}}) := \Phi\Big(\dfrac{x-\widehat{\mu}_{T+h}}{\widehat{\sigma}_{T+h}}\Big)$,
- **continuity-corr. n.a.**: $\quad \widehat{f}_{T+h,\mathrm{a},x}(\widehat{\boldsymbol{\vartheta}}) := \Phi\Big(\dfrac{x-\widehat{\mu}_{T+h}+0.5}{\widehat{\sigma}_{T+h}}\Big)$.

PMFF $\widehat{\boldsymbol{p}}_{T+h,\mathrm{a}}(\widehat{\boldsymbol{\vartheta}})$ as discrete differences of $\widehat{\boldsymbol{f}}_{T+h,\mathrm{a}}(\widehat{\boldsymbol{\vartheta}})$.

On PMF-Forecasting for Count Processes

Performance Evaluation

Forecast performance by evaluating

inaccuracy of considered PMFF $\widehat{\boldsymbol{p}}$ w.r.t. true PMFF $\widehat{\boldsymbol{p}}_0$.

Here, $\widehat{\boldsymbol{p}} \in \left\{\widehat{\boldsymbol{p}}_{T+h}\big(\widehat{\boldsymbol{\theta}}\big), \widehat{\boldsymbol{p}}_{T+h,\mathsf{a}}\big(\widehat{\boldsymbol{\vartheta}}\big)\right\}$ and $\widehat{\boldsymbol{p}}_0 = \widehat{\boldsymbol{p}}_{T+h}(\boldsymbol{\theta})$.

Different solutions proposed in literature yet:

- Homburg (2020) distinguishes **local and global criteria**. Global comparison, e. g., by Kullback Leibler divergence, Kolmogorov metric, or Raff's maximum error.
- Most often, types of "squared distance" between $\widehat{\boldsymbol{p}}$ and $\widehat{\boldsymbol{p}}_0$:

(. . .)

Forecast performance evaluation for $\widehat{\boldsymbol{p}}$ vs. $\widehat{\boldsymbol{p}}_0$:

- Most often, types of "squared distance" between $\widehat{\boldsymbol{p}}$ and $\widehat{\boldsymbol{p}}_0$:
  - χ^2-distance by Willemain (2006),
  - unweighted mean squared errors (MSEs) by McCabe et al. (2011), such as $\left\|\widehat{\boldsymbol{p}}-\widehat{\boldsymbol{p}}_0\right\|^2=\sum_{x=0}^{\infty}(\widehat{p}_x-\widehat{p}_{0,x})^2$,
  - quadratic score (qs) or ranked probability score (rps) by Snyder et al. (2012), Kolassa (2016).

  Note that *increase in expected score* leads to

$$
\begin{aligned}
E\Big[s_{\mathrm{qs}}\big(\widehat{\boldsymbol{p}},X\big)-s_{\mathrm{qs}}\big(\widehat{\boldsymbol{p}}_0,X\big)\,\Big|\,X\sim\widehat{\boldsymbol{p}}_0\Big] \;&=\; \left\|\widehat{\boldsymbol{p}}-\widehat{\boldsymbol{p}}_0\right\|^2,\\
E\Big[s_{\mathrm{rps}}\big(\widehat{\boldsymbol{f}},X\big)-s_{\mathrm{rps}}\big(\widehat{\boldsymbol{f}}_0,X\big)\,\Big|\,X\sim\widehat{\boldsymbol{p}}_0\Big] \;&=\; \left\|\widehat{\boldsymbol{f}}-\widehat{\boldsymbol{f}}_0\right\|^2.
\end{aligned}
$$

Forecast performance evaluation for $\widehat{\boldsymbol{p}}$ vs. $\widehat{\boldsymbol{p}}_0$:

- Most inaccuracy measures related to

$$\left\|\widehat{\boldsymbol{p}} - \widehat{\boldsymbol{p}}_0\right\|^2 \qquad \text{or} \qquad \left\|\widehat{\boldsymbol{f}} - \widehat{\boldsymbol{f}}_0\right\|^2.$$

- Boylan & Synteto (2006): global inaccuracy maybe misleading, because may not exclude poor performance in tails.
- Finally, like in McCabe et al. (2011), we use both global MSEs and two local MSEs (w.r.t. $\boldsymbol{p}$ or $\boldsymbol{f}$):
  - lower-25% MSE $\sum_{x=0}^{\infty}(\widehat{p}_x - \widehat{p}_{0,x})^2\, \mathbb{1}(\widehat{f}_{0,x} \leq 0.25)$,
  - upper-10% MSE $\sum_{x=0}^{\infty}(\widehat{p}_x - \widehat{p}_{0,x})^2\, \mathbb{1}(\widehat{f}_{0,x} \geq 0.90)$.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH
STAT

On PMF-Forecasting for Count Processes

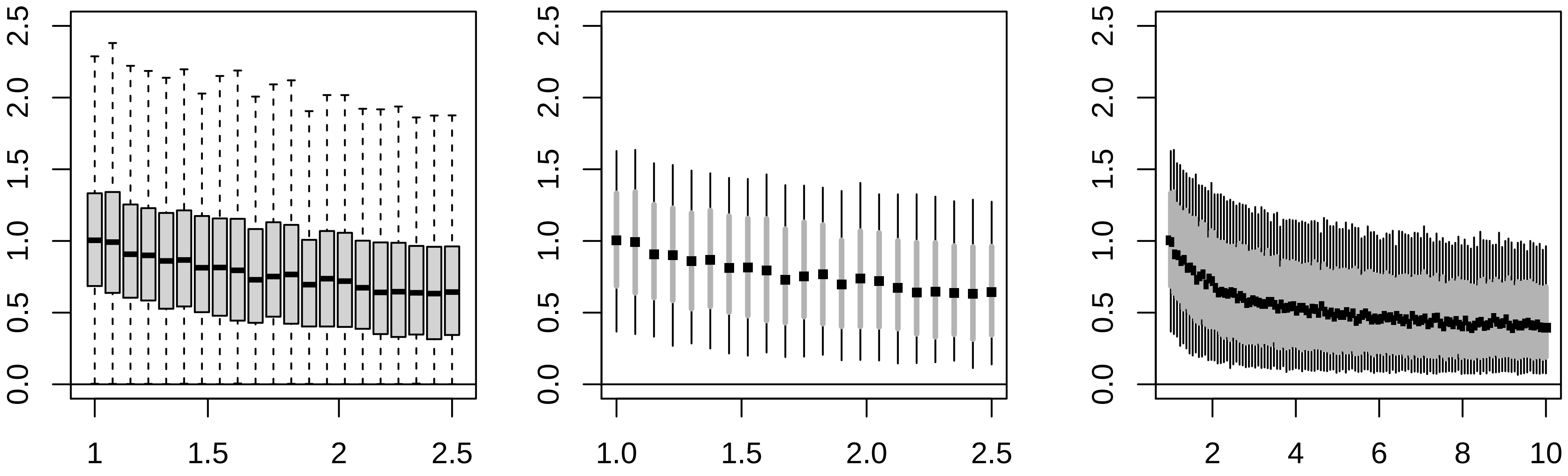
Simulation Study

Comprehensive simulation study ($> 17,000$ scenarios), with 1,000 replications per scenario.

Main visual tool for evaluation: “**lean boxplot**”,

black dot for median, thick grey line connecting quartiles, thin black line connecting 10%- and 90%-quantiles.

ordinary boxplots — lean boxplots — many lean boxplots



Lean boxplots of MSEs $\|\widehat{\boldsymbol{p}}-\widehat{\boldsymbol{p}}_0\|^2$, $\|\widehat{\boldsymbol{f}}-\widehat{\boldsymbol{f}}_0\|^2$ of coherent PMFFs, or MSE differences $\|\widehat{\boldsymbol{p}}_{\mathrm{a}}-\widehat{\boldsymbol{p}}_0\|^2 - \|\widehat{\boldsymbol{p}}-\widehat{\boldsymbol{p}}_0\|^2$, $\|\widehat{\boldsymbol{f}}_{\mathrm{a}}-\widehat{\boldsymbol{f}}_0\|^2 - \|\widehat{\boldsymbol{f}}-\widehat{\boldsymbol{f}}_0\|^2$ between approximate and coherent PMFFs.

Full results as supplementary files ($\rightarrow$ project website).

General findings:

- Conclusions do not differ between PMF-based MSEs and CDF-based MSEs. Thus, focus on PMF-based MSE values.
- Simple normal approximation by far worse than continuity-corrected one, MSEs increased by factor 5–10. Thus, focus on approximate PMFFs with continuity correction.

Performance of Coherent Forecasting:

- Coherent PMFFs generally close to true PMFFs, with decreasing MSEs (of all types) for increasing mean μ and sample size T.
- Increases of dependence α lead to increased MSEs.
- Lower-tail MSE usually larger than upper-tail MSE.

(. . .)

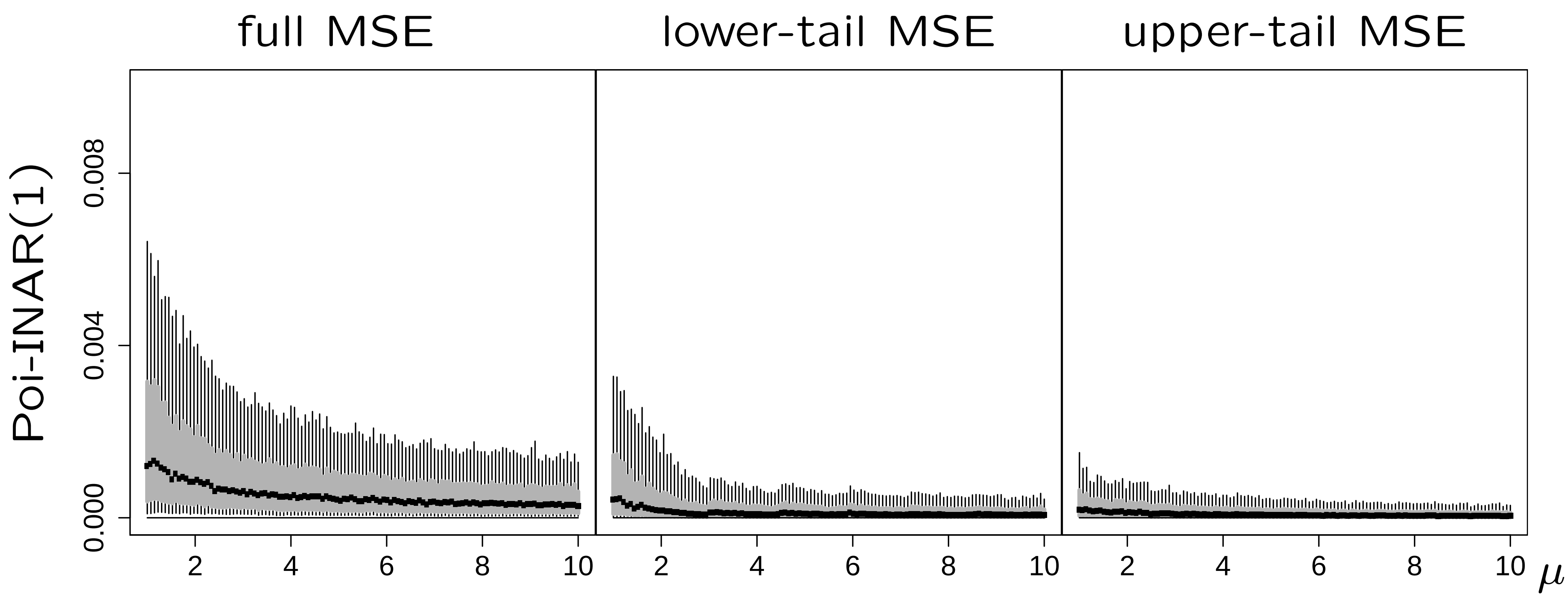


Illustration: Coherent PMFFs for Poi-INAR(1) DGP, different types of MSE with $\alpha = 0.55$, $T = 250$, and $h = 1$, plotted against mean μ.

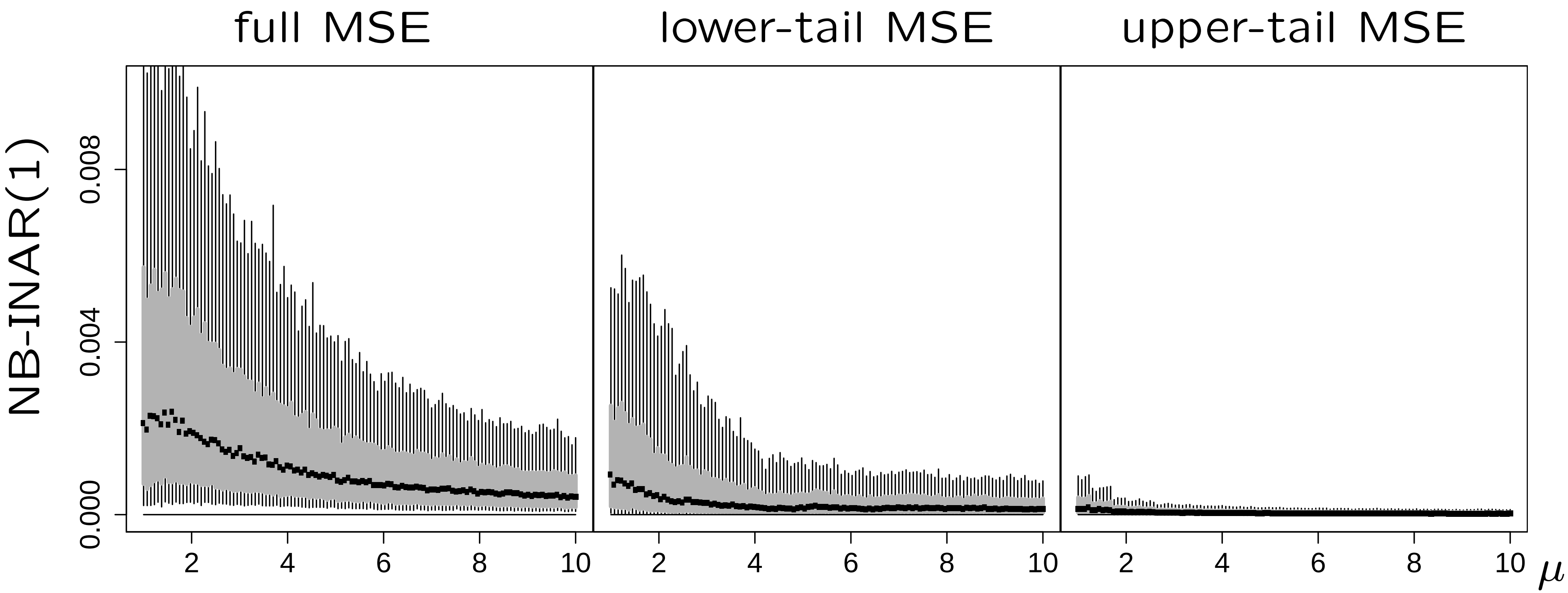


Illustration: NB-INAR(1) DGP with dispersion ratio $I = 2.4$.

Overdispersion causes increasing MSEs, mainly for low μ.

Increase stronger for lower- than for upper-tail MSE.

Performance of Coherent Forecasting:

- (. . .)
- MSE increase even more pronounced for ZIP model, i. e., where overdispersion caused by zero inflation.
- Further increase of model order also further increases MSEs.
- MSEs larger for INAR- than for INARCH-type DGPs.
- Analogous results for bounded counts DGPs, but for $\pi = 0.45$ (nearly symmetric PMF), all types of MSE values get by far smaller.

Performance of Approximate Forecasting:

Now MSE differences $\left\|\widehat{\boldsymbol{p}}_{\mathrm{a}}-\widehat{\boldsymbol{p}}_0\right\|^2-\left\|\widehat{\boldsymbol{p}}-\widehat{\boldsymbol{p}}_0\right\|^2$,

with positive values if approximation causes increased MSEs.

- In large majority, approximate PMFFs clearly larger MSEs. Particularly clear if T increases.
- Discrepancy approximate vs. coherent particularly large for low means, performs rather well only for nearly symmetric bounded counts with $\pi=0.45$.
- Larger discrepancy also for increasing overdispersion or dependence.

(. . .)

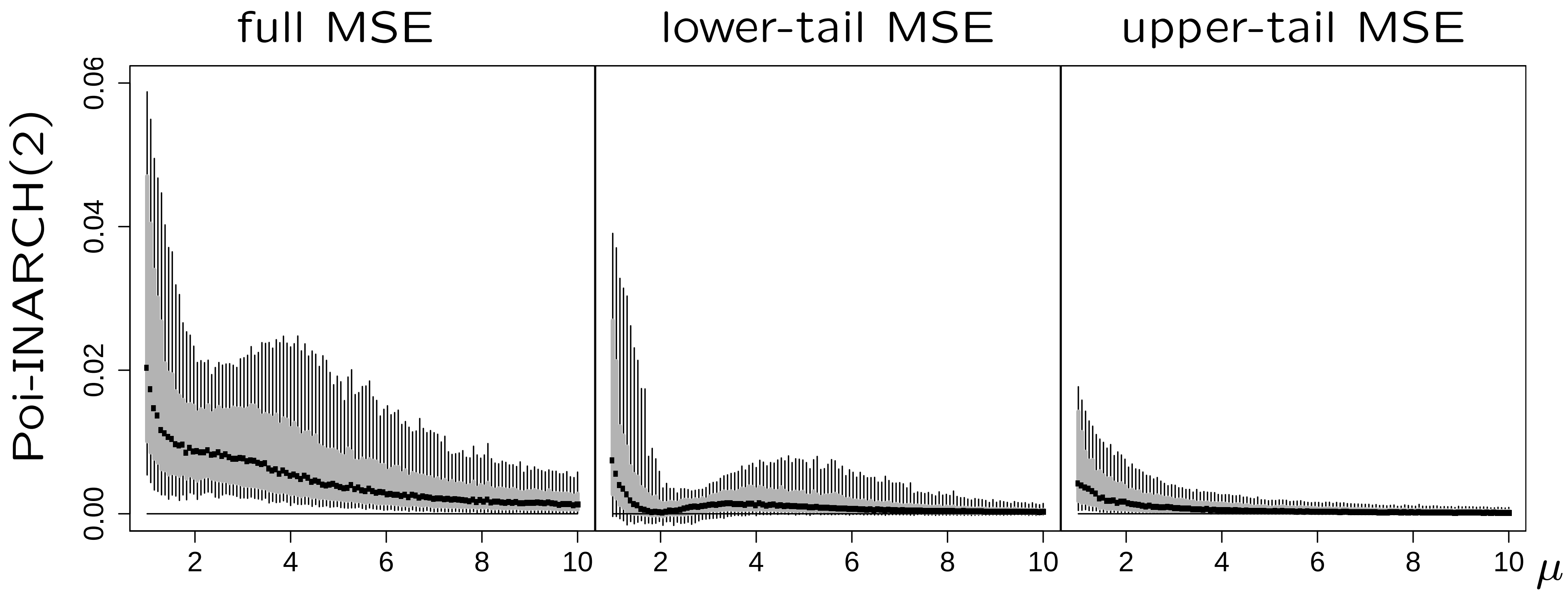


Illustration: MSE differences for Poi-INARCH(2) DGP, with $\alpha = 0.55$, $\alpha_2 = 0.45$, $T = 250$, and $h = 1$.

Approximation especially increases lower-tail MSE.

Conclusions

- For coherent PMFFs, effect of estimation generally low: deteriorations mainly for low sample size T and strong dependence α. More pronounced for low means μ and for overdispersion $I > 1$.

- Strong performance deterioration if approximate PMFFs, even if using continuity correction or increased T. Although approximations tempting w.r.t. implementation benefits, their use in practice **strongly discouraged!**

- **Future (ongoing) research:**
  PMF forecasting under model misspecification.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH
STAT

PMF-Forecasting of Transaction Counts

Application

Real-world data application:

Count time series about transaction numbers per trading day, on structured products from on-market and off-market trading, offered by Cascade-Turnoverdata of Deutsche Börse AG.

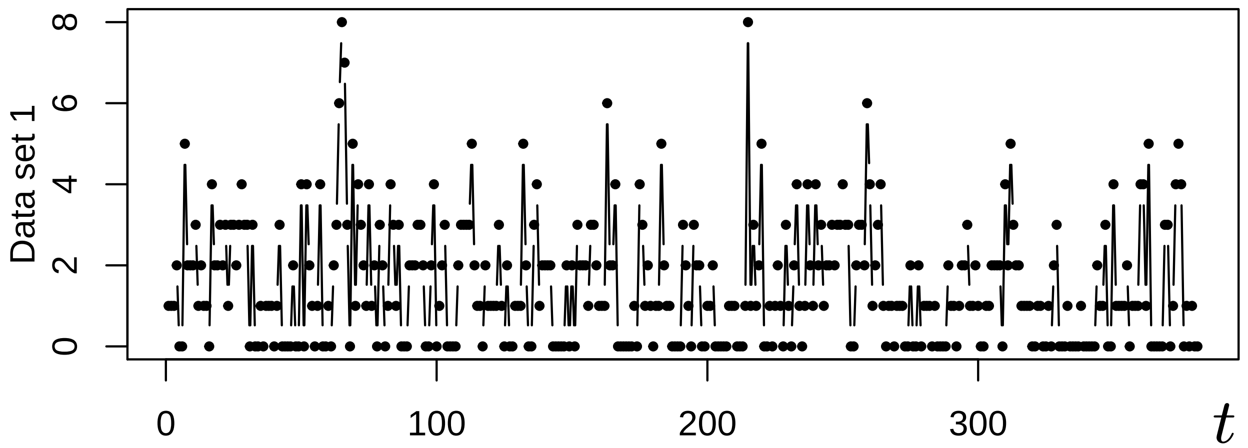
For illustration,
two exemplary time series are considered: $(\ldots)$

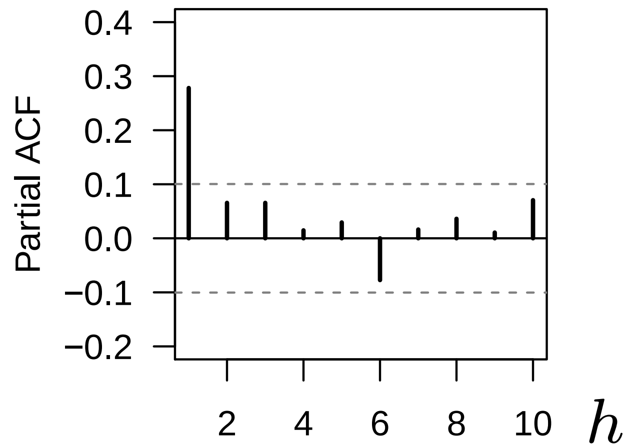
Data set 1:

$T_1 = 381$ counts (Feb. 2017–July 2018) for model fitting,

23 counts from Aug. 2018 left for out-of-sample forecasting.

Time series plot and sample PACF of learning sample:





AR(1)-like autocorrelation with overdispersion ($\hat{I} \approx 1.518$)

$\Rightarrow$ fit NB-INAR(1) model.

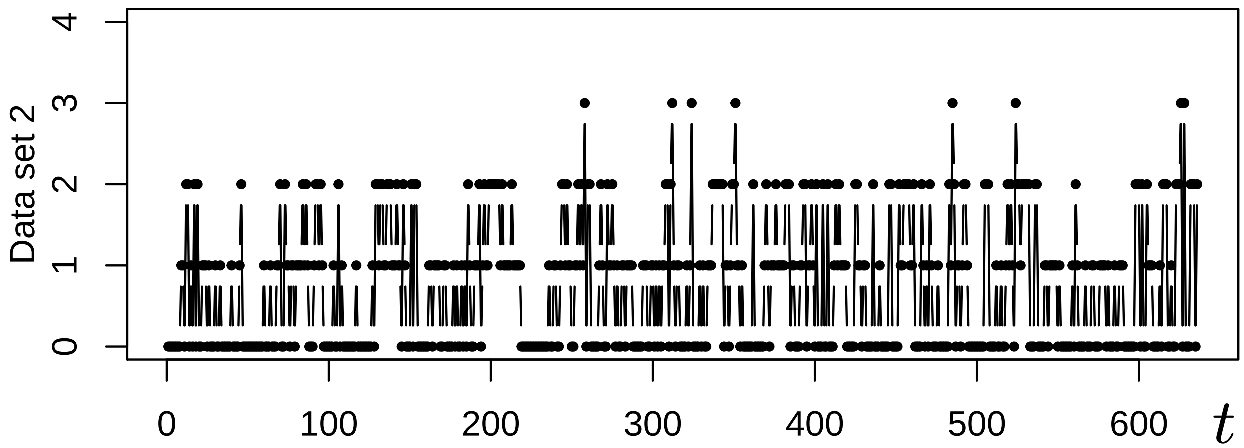
PMF-Forecasting of Transaction Counts

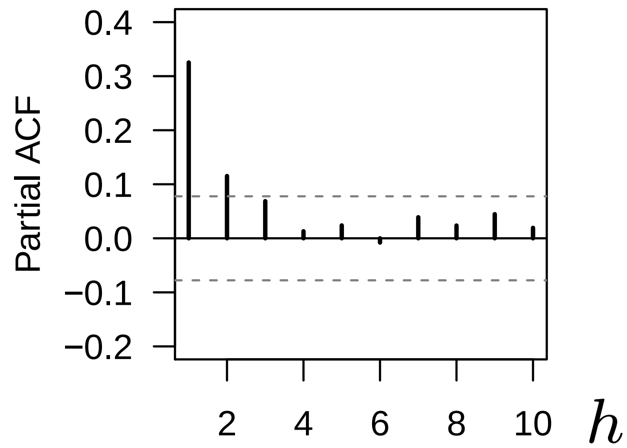
Data set 2 with one additional year of data:

$T_2 = 636$ counts (Feb. 2017–July 2019) for model fitting,

22 counts from Aug. 2019 left for out-of-sample forecasting.

Time series plot and sample PACF of learning sample:

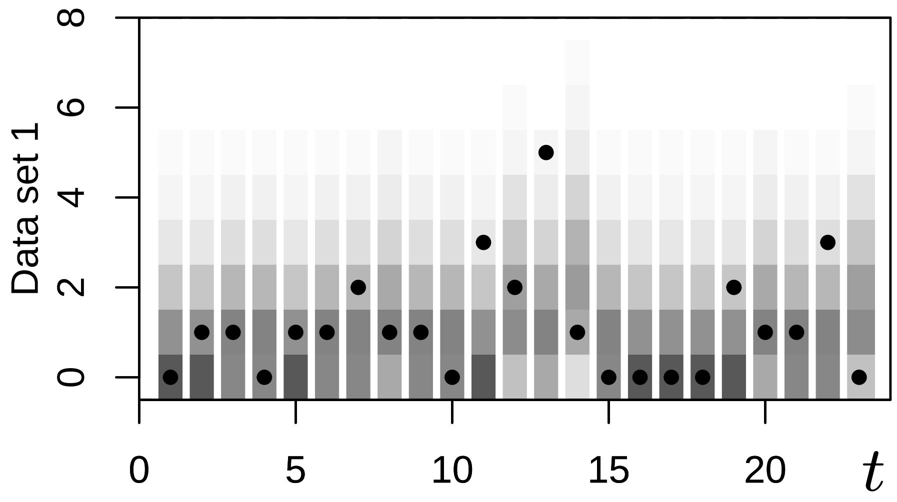


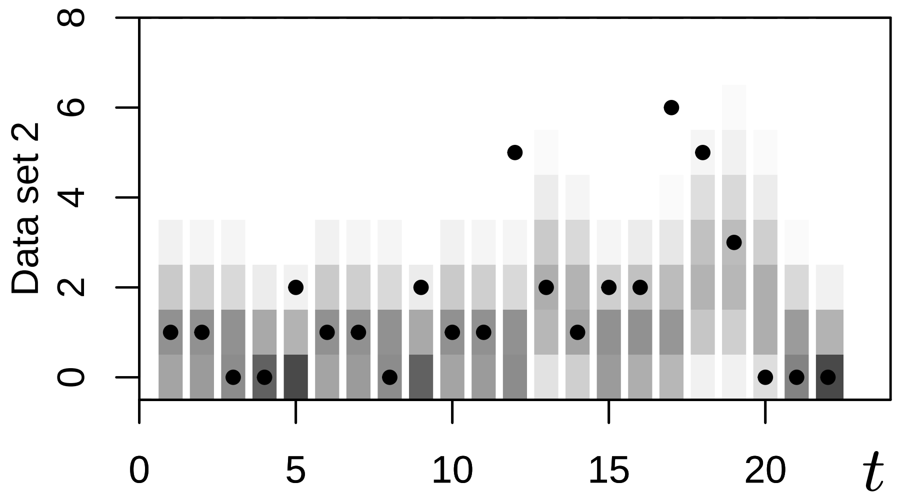


AR(2)-like autocorrelation, nearly equidispersion ($\hat{I} \approx 0.913$)

$\Rightarrow$ fit Poi-INAR(2) model.

PMF forecasting of respective out-of-sample data (black dots) based on fitted models:





PMFFs for data set 1 cover larger area $\rightarrow$ overdispersion.

For data set 2, in turn, some dots in "white area": (...)

Possible application of obtained PMFFs:

integration into **"risk alert" system**.

Achieved transaction counts could be compared to PMFFs to judge their plausibility. Previous figures indicate possibly "unusual order book behavior" for data set 2, namely for counts at $t = 12, 17$.
This could give rise to inform traders on these days.

Real-time risk alerts relevant topic for market infrastructure providers such as Deutsche Börse AG (→ website).

Possible direction for future research:

For risk alerts based on PMFFs,

also take estimation/model uncertainty into account.

PMFFs for data set 2 rely on $T_2 = 636$ observations,

while those for data set 1 only use $T_1 = 381$.

Probably not much estimation effect for data set 2,

but for data set 1 (less data plus overdispersion).

Since risk alerts generated based on tails of PMFF,

careful investigation recommended.

Possible solution: parametric bootstrap approach

in analogy to Weiß et al. (2021).

Thank You

for Your Interest!

Christian H. Weiß

Department of Mathematics & Statistics

Helmut Schmidt University, Hamburg

weissc@hsu-hh.de

Literature

Homburg et al. (2021) On PMF-Forecasting for Count Processes. In *Proceedings of ITISE 2021*, in press.

Boylan & Synteto (2006) Accuracy and accuracy-... *Foresight* **4**, 39–42.

Du & Li (1991) The integer-valued autoregressive ... *JTSA* **12**, 129–142.

Ferland et al. (2006) Integer-valued GARCH proc. *JTSA* **27**, 923–942.

Freeland & McCabe (2004) Forec. discrete valued ... *IJF* **20**, 427–434.

Homburg (2020) Criteria for evaluating ... *CSSC* **49**, 3152–3170.

Homburg et al. (2019) Evaluating approx. point ... *Econometrics* **7**, 30.

Homburg et al. (2021) A performance analysis of predic... *JF* **40**, 603–625.

Jung & Tremayne (2006) Coherent forec. in integer ... *IJF* **22**, 223–238.

Kolassa (2016) Evaluating predictive count data ... *IJF* **32**, 788–803.

(...)

Literature

(. . .)

McCabe et al. (2011) Efficient probabilistic forec. . . *JRSS-B* **73**, 253–272.

McKenzie (1985) Some simple models . . . *Water Res. Bull.* **21**, 645–650.

Snyder et al. (2012) Forecasting the intermittent . . . *IJF* **28**, 485–496.

Steutel & van Harn (1979) Discrete analogues . . . *Ann. Probab.* **7**, 893–899.

Weiß (2018) *An Introduction to Discrete-Valued Time Series*. Wiley.

Weiß et al. (2021) Efficient accounting for . . . *JAS*, in press.

Weiß & Kim (2014) Diagnosing and modeling . . . *ASMBI* **69**, 281–297.

Willemain (2006) Forecast-accuracy metrics . . . *Foresight* **4**, 36–38.

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 394832307.