Ein erweitertes Poisson INAR(1)-Modell

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich **Mathematik**

Christian H. Weiß

Fachbereich Mathematik

Technische Universität Darmstadt

Populär bei **reellwertigen** stationären Prozessen:

ARMA(p,q)-Modelle. Sei $(\epsilon_t)_{\mathbb{Z}}$ weißes Rauschen, dann

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

wobei $\alpha_1, \ldots, \alpha_{\mathsf{p}}, \beta_1, \ldots, \beta_{\mathsf{q}} \in \mathbb{R}$ geeignet gewählt.

Beispiel: AR(1)-Modell $X_t \;=\; \alpha \cdot X_{t-1} + \epsilon_t$
mit Autokorrelationsfunktion $\rho_X(k) = \alpha^k$.

Nicht auf Zähldatenprozesse anwendbar: I. A. ist $\alpha \cdot X \notin \mathbb{N}_0$.

Modellierung von Zähldatenzeitreihen

Poisson-INAR(1)-Modell

INAR(1)-Modell:

Ersetze Multiplikation '$\cdot$' in AR(1)-Rekursion

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

durch **binomial thinning** ('binomiale Ausdünnung'):

Hat X Wertebereich $\mathbb{N}_0$, und ist $\alpha \in (0;1)$, so

$$\alpha \circ X \;:=\; \textstyle\sum_{i=1}^{X}\; Y_i,$$

wobei Y_i unabhängige, binäre ZV mit $P(Y_i = 1) = \alpha$,

auch unabhängig von X. (Steutel & van Harn, 1979)

Es gilt: $E[\alpha \circ X] \;=\; E[\alpha \cdot X]$, aber stets $\alpha \circ X \in \mathbb{N}_0$.

Beispiel:

INAR(1)-Modell mit Poisson-verteilten Innovationen:

Sei $\alpha \in (0;1)$, sei $(\epsilon_t)_{\mathbb{N}}$ i.i.d. $Po(\lambda)$ und $X_0 \sim Po(\frac{\lambda}{1-\alpha})$.
Prozess $(X_t)_{\mathbb{N}_0}$, definiert durch Rekursion

$$X_t \;=\; \alpha \circ X_{t-1} \;+\; \epsilon_t, \qquad\qquad t \geq 1,$$

sowie durch geeignete Unabhängigkeitsannahmen,
heißt **Poisson-INAR(1)-Prozess**. (McKenzie, 1985)

Grundlegende Eigenschaften:

- $(X_t)_{\mathbb{N}_0}$ stationäre Markovkette, Randverteilung $Po(\frac{\lambda}{1-\alpha})$,
- $E[X_t] \;=\; V[X_t] \;=\; \frac{\lambda}{1-\alpha}$ (Equidispersion),
- AR(1)-Abhängigkeit: $\rho_X(k) = \alpha^k$,
- $P(k|l) \;:=\; P(X_t = k \mid X_{t-1} = l)$

$$= \sum_{j=0}^{\min(k,l)} \binom{l}{j} \alpha^j (1-\alpha)^{l-j} \cdot e^{-\lambda} \frac{\lambda^{k-j}}{(k-j)!} \;>0.$$

Interpretation von $\alpha \circ X := \sum_{i=1}^{X} Y_i$:

- Population von Größe X zu gewisser Zeit t.
- Später zur Zeit $t+1$: Population geschrumpft, weil manche Individuen verstorben.
- Sterben alle Individuen unabhängig voneinander mit Wahrscheinlichkeit $1-\alpha$

  $\Rightarrow$ *Zahl der Überlebenden* gegeben durch $\alpha \circ X$.

Interpretation & Beispiele:

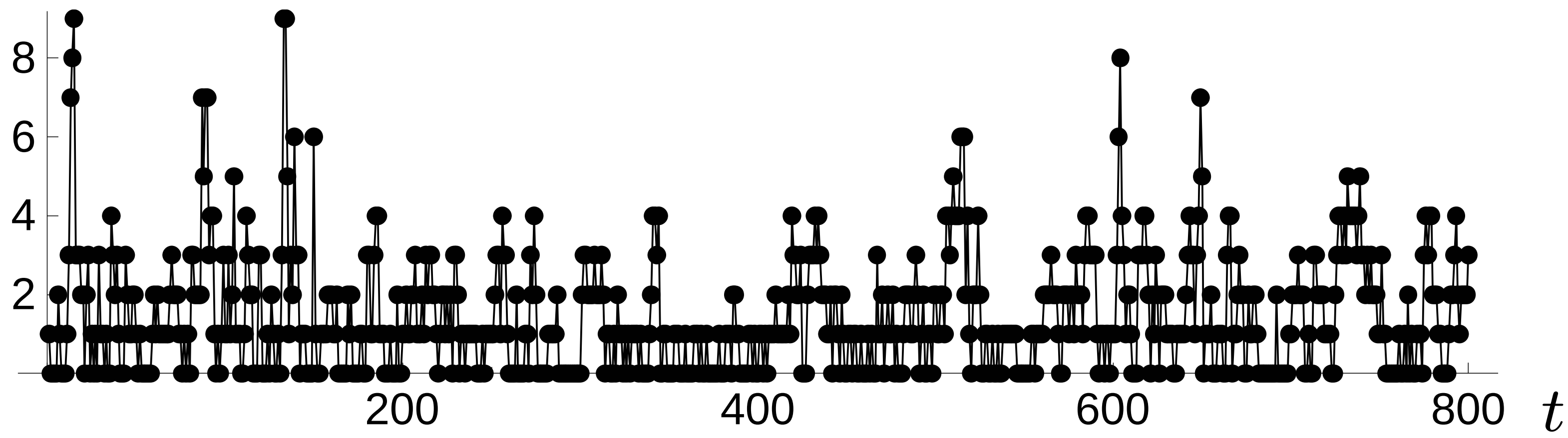
$$\underbrace{X_t}_{\text{Population zur Zeit } t} = \underbrace{\alpha \circ X_{t-1}}_{\text{Überlebende der Zeit } t-1} + \underbrace{\epsilon_t.}_{\text{Immigration}}$$

- X_t: Zahl der Nutzer, die auf Webserver zugreifen. ϵ_t: Zahl neuer Nutzer, $\alpha \circ X_{t-1}$: Zahl früherer Nutzer, welche noch immer aktiv.

- X_t: Zahl der Kunden. ϵ_t: neue Kunden, $X_{t-1} - \alpha \circ X_{t-1}$: Kunden, welche am Ende der vorigen Periode verloren.

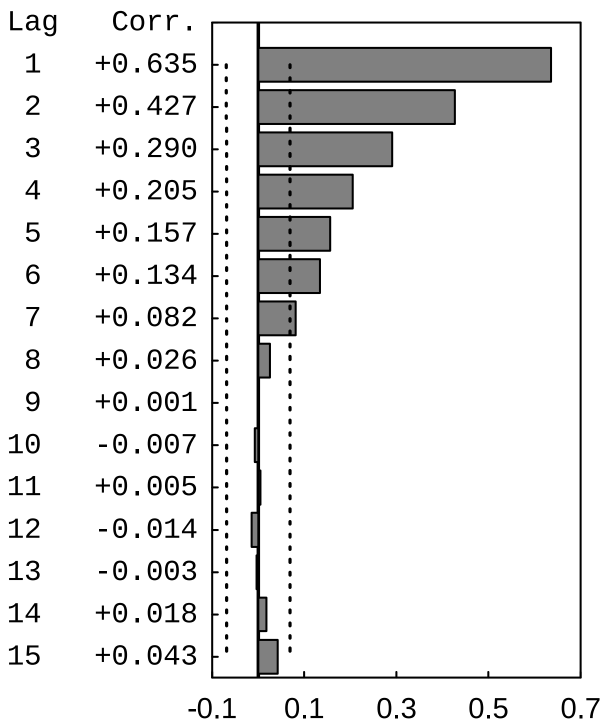
Datenbeispiel aus Jung & Tremayne (2011):

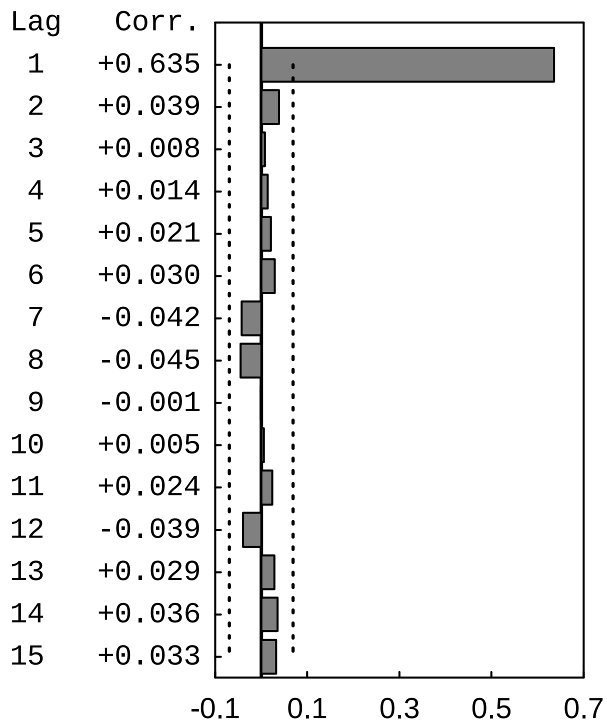
Zeitreihe (Länge 800) zur
Anzahl sog. **iceberg orders** (vgl. Frey & Sandas, 2009)
pro 20 Min. bei 32 Handelstagen im 1. Quartal 2004 bzgl.
Aktien der Deutschen Telekom (XETRA, Deutsche Börse).



Datenbeispiel (iceberg orders):

Autokorrelations- und partielle Autokorrelationsfunktion:





$\Rightarrow$ Klare AR(1)-Abhängigkeit!

Datenbeispiel (iceberg orders):

Mittlere Anzahl: ca. 1.406,
ML-Schätzwerte: $\hat{\lambda}_{\text{ML}} \approx 0.638$ (0.038), $\hat{\alpha}_{\text{ML}} \approx 0.547$ (0.023).

Denkbare Interpretation:
„Reproduktionswahrscheinlichkeit“ von 55 %
(Händler zeigt nächsten Teil des Auftrags),
mittlere „Innovationsrate“ von 0.64 (neue iceberg orders).

Aber:
Empirische Varianz ca. 2.181, d. h.
etwa 55 % **Überdispersion**.

Modellierung von Zähldatenzeitreihen

INARCH(1)-Modell

Definition:

Sei $(X_t)_{\mathbb{Z}}$ Prozess mit Wertebereich $\mathbb{N}_0 = \{0, 1, \ldots\}$,
sei $b > 0$ und $0 < a < 1$.
$(X_t)_{\mathbb{Z}}$ folgt dem **INARCH(1)-Modell**
wenn X_t, bedingt auf $X_{t-1}, X_{t-2}, \ldots$,
Poisson-verteilt ist mit zug. Erwartungswert
$E[X_t \mid X_{t-1}, X_{t-2}, \ldots] = b + a \cdot X_{t-1}$,
d. h.

$$X_t \ \sim \ Po(b + a \cdot X_{t-1}).$$

(Heinen, 2003)

Grundlegende Eigenschaften:

- $(X_t)_{\mathbb{Z}}$ stationäre, ergodische Markovkette,

  (Zhu & Wang, 2009)

- Übergangswahrscheinlichkeiten

$$P(k|l) \;=\; \exp\left(-b - a \cdot l\right) \cdot \tfrac{(b+a\cdot l)^k}{k!} \;\; > 0,$$

- AR(1)-Abhängigkeit: $\rho_X(k) = a^k$,

- alle Momente existieren (Ferland, 2006).

Weiß (2009):

Rekursionsschema für **Randkumulanten**:

$$\kappa_1 = \frac{b}{1-a}, \qquad \kappa_n = -(1-a^n)^{-1} \cdot \sum_{j=1}^{n-1} s_{n,j} \cdot \kappa_j \quad \text{for } n \geq 2,$$

wobei $s_{n,j}$ die Stirlingzahlen erster Art sind:

$$s_{n,0} = 0 \quad \text{und} \qquad s_{n,n} = 1 \quad \text{für } n \geq 1,$$
$$s_{n+1,j} = s_{n,j-1} - n \cdot s_{n,j} \qquad \text{für } j = 1, \ldots, n \text{ und } n \geq 1.$$

Insbesondere liegt Randverteilung mit Überdispersion vor:

$$\kappa_1 = \frac{b}{1-a} = E[X_t], \quad \kappa_2 = \frac{b}{(1-a)(1-a^2)} = V[X_t].$$

Datenbeispiel (iceberg orders):

ML-Schätzwerte: $\widehat{b}_{\mathsf{ML}} \approx 0.579$ (0.041), $\widehat{a}_{\mathsf{ML}} \approx 0.589$ (0.032).

Interpretation analog Poisson-INAR(1)-Modell.

Empirische Überdispersion: ca. 55 %,

Modellüberdispersion $1/(1-\widehat{a}_{\mathsf{ML}}^2)$: ca. 53 %.

Vergleich der Modelle:

Poisson-INAR(1): AIC $\approx$ 2229.8, BIC $\approx$ 2239.2;

INARCH(1): AIC $\approx$ 2219.7, BIC $\approx$ 2229.1.

Ein erweitertes Poisson-INAR(1) -Modell

„Work in progress“

Motivation:

Wie kann man das Poisson-INAR(1)-Modell mit Rekursion

$$X_t \;=\; \alpha \circ X_{t-1} \;+\; \epsilon_t, \qquad (\epsilon_t)_{\mathbb{N}} \text{ i.i.d. gemäß } Po(\lambda),$$

modifizieren, um Überdispersion zuzulassen?

- Ändere Verteilung der Innovationen (vgl. Weiß (2008));
- ändere thinning und Verteilung der Innovationen (vgl. Weiß (2008));
- Idee: Erlaube seriell abhängige Innovationen!

In Anlehnung an Vorschlag von Buckley & Pollett (2010) (dort bezogen auf infinite-patch-Metapopulationsmodelle) nehmen wir an:

Die Innovation ϵ_t hängt funktional von voriger Beobachtung X_{t-1} ab, d. h. $\epsilon_t \sim Po\big(f(X_{t-1})\big)$, wobei $f:\ \mathbb{N}_0 \to (0;\infty)$.

$\Rightarrow$ $(X_t)_{\mathbb{N}_0}$ ist weiterhin homogene Markovkette, mit

$$\begin{aligned} P(k|l) \ &:= \ P(X_t = k \mid X_{t-1} = l) \\ &= \ \textstyle\sum_{j=0}^{\min(k,l)} \binom{l}{j}\, \alpha^j\, (1-\alpha)^{l-j} \cdot e^{-f(l)}\, \frac{f(l)^{k-j}}{(k-j)!} \ > 0. \end{aligned}$$

Innovationen $(\epsilon_t)_{\mathbb{N}}$ im Allgemeinen nicht mehr i.i.d.!

Analog zu Buckley & Pollett (2010) sei von nun an $f(x) = a \cdot x + b$, mit Achsenabschnitt $b > 0$.

Erweitertes Poisson-INAR(1)-Modell:

Seien $\alpha, a \in (0;1)$ mit $\alpha + a < 1$, sei $b > 0$.
Der Prozess $(X_t)_{\mathbb{N}_0}$ folgt der Rekursion

$$X_t \; = \; \alpha \circ X_{t-1} \; + \; \epsilon_t \qquad \text{für } t \geq 1,$$

wobei, gegeben X_{t-1}, $\alpha \circ X_{t-1}$ und ϵ_t unabhängig sind, auch unabhängig von $X_{t-2}, X_{t-3}, \ldots$, $\epsilon_{t-1}, \epsilon_{t-2}, \ldots$ und $\alpha \circ X_{t-2}, \alpha \circ X_{t-3}, \ldots$,
und ϵ_t verteilt ist gemäß $Po(a \cdot X_{t-1} + b)$.

Interpretation:

$$X_t \;=\; \alpha \circ X_{t-1} \;+\; \epsilon_t, \qquad\qquad \epsilon_t \sim Po(a \cdot X_{t-1} + b).$$

analog zu gewöhnlichem Poisson-INAR(1)-Modell,
aber mit folgendem Zusatzmerkmal:

Immigration wird attraktiver,
wenn aktuelle Population groß ist, d. h.

die grundsätzliche, mittlere Immigrationsrate b
wird erhöht um $a \cdot X_{t-1}$.

Beziehung zu anderen Modellen:

Aus der Definition folgt:

$$X_t \overset{D}{=} Y_{t,1} + \ldots + Y_{t,X_{t-1}} + \nu_t,$$

wobei die $Y_{t,k}$ und ν_s jeweils i.i.d., auch unabhängig voneinander, und wobei $\nu_s \sim Po(b)$ und $Y_{t,k} \sim B(1,\alpha) + Po(a)$.

$\Rightarrow$ $(X_t)_{\mathbb{N}_0}$ verteilt wie „**branching process with immigration**" (Letzteres wegen $b > 0$).

Beziehung zu anderen Modellen: (Forts.)

Zusammen mit der Eigenschaft, dass
$(X_t)_{\mathbb{N}_0}$ **irreduzibel und aperiodisch** ist
(da $(X_t)_{\mathbb{N}_0}$ echt positive Überg.wahrsch. hat),
folgt unter Anwendung eines Satzes aus Heathcote (1966):

Für $\alpha + a < 1$
besitzt $(X_t)_{\mathbb{N}_0}$ eine **stationäre Randverteilung**,
entsprechend ist $(X_t)_{\mathbb{N}_0}$ sogar **ergodisch**.

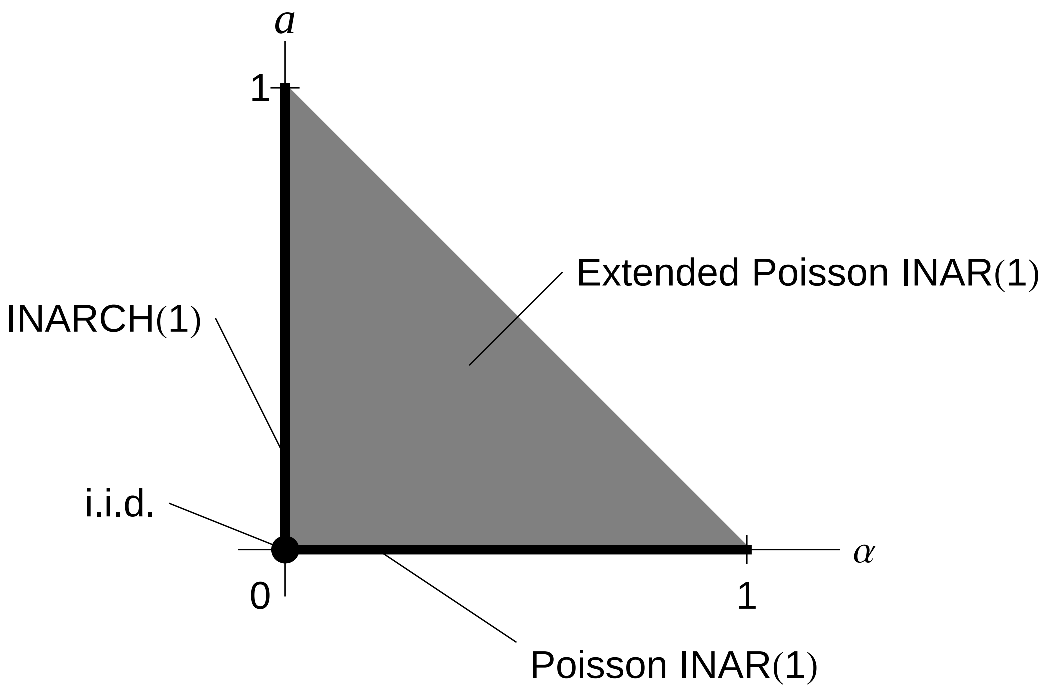
Beziehung zu anderen Modellen: (Forts.)

$$X_t \;=\; \alpha \circ X_{t-1} \;+\; \epsilon_t, \qquad \epsilon_t \sim Po(a \cdot X_{t-1} + b).$$

Grenzfall $a \to 0$:
Poisson-INAR(1)-Modell.

Grenzfall $\alpha \to 0$:
INARCH(1)-Modell.

„Brücke“



Stochastische Eigenschaften der Beobachtungen X_t:

$$E[X_t \mid X_{t-1},\ldots] \;=\; (\alpha+a)\cdot X_{t-1}+b,$$
$$V[X_t \mid X_{t-1},\ldots] \;=\; \big(\alpha(1-\alpha)+a\big)\cdot X_{t-1}+b;$$

$$E[X_t^n \mid X_{t-1},\ldots] \;=\; \textstyle\sum_{j=0}^{n} \binom{n}{j}\cdot\mu_{j,B(X_{t-1},\alpha)}\cdot\mu_{n-j,Po(a\cdot X_{t-1}+b)};$$

$$\mu_X \;:=\; E[X_t] \;=\; \tfrac{b}{1-(\alpha+a)},$$
$$\sigma_X^2 \;:=\; V[X_t] \;=\; \mu_X\cdot\tfrac{1-\alpha^2}{1-(\alpha+a)^2} \quad (\rightarrow \text{Überdispersion});$$

$$E[(X_t-\mu_X)^3] \;=\; \mu_X\cdot\tfrac{(1-\alpha^3)(1-\alpha^2+a\alpha+2a^2)+3a\alpha(1-\alpha)}{\big(1-(a+\alpha)^2\big)\big(1-(a+\alpha)^3\big)};$$

$\rho_X(k)=(\alpha+a)^k$ für $k\geq 1$ $\quad(\rightarrow$ AR(1)-Abhängigkeit).

Stochastische Eigenschaften der Innovationen ϵ_t:

$$\epsilon_t \sim Po(a \cdot X_{t-1} + b);$$

$$\mu_\epsilon \;:=\; E[\epsilon_t] \;=\; (1-\alpha)\cdot\mu_X,$$
$$\sigma_\epsilon^2 \;:=\; V[\epsilon_t] \;=\; \tfrac{\sigma_X^2}{1+\alpha}\cdot\big(1-\alpha^2-\alpha a(2-a)\big);$$

$$\rho_\epsilon(k) \;=\; a\cdot(\alpha+a)^{k-1}\cdot\big(1+\alpha a\cdot\tfrac{\sigma_X^2}{\sigma_\epsilon^2}\big) \text{ für } k\geq 1,$$

d. h. seriell abhängige Innovationen.

Parameterschätzung:

Likelihood-Funktion $L(\alpha, a, b) = \prod_{t=1}^{T} P(X_t|X_{t-1})$,
numerische Maximierung des Log-Likelihood
ergibt **ML-Schätzer** $\widehat{\alpha}_{\text{ML}}, \widehat{a}_{\text{ML}}, \widehat{b}_{\text{ML}}$.

Momentschätzer können basieren auf

$$b = \mu_X \cdot \big(1 - \rho_X(1)\big), \quad \alpha = \sqrt{1 - \frac{\sigma_X^2}{\mu_X} \cdot \big(1 - \rho_X(1)^2\big)},$$
$$a = \rho_X(1) - \alpha.$$

Datenbeispiel (iceberg orders):

ML-Schätzwerte: $\hat{b}_{\mathrm{ML}} \approx 0.567$ (0.040),

$\hat{\alpha}_{\mathrm{ML}} \approx 0.410$ (0.058), $\hat{a}_{\mathrm{ML}} \approx 0.188$ (0.059).

Empirische Überdispersion: ca. 55 %,

Modellüberdispersion $\frac{1-\hat{\alpha}_{\mathrm{ML}}^2}{1-(\hat{\alpha}_{\mathrm{ML}}+\hat{a}_{\mathrm{ML}})^2}$: ca. 29 %.

Vergleich der Modelle:

Poisson-INAR(1): AIC $\approx$ 2229.8, BIC $\approx$ 2239.2;

INARCH(1): AIC $\approx$ 2219.7, BIC $\approx$ 2229.1;

Erw. Poisson-INAR(1): AIC $\approx$ 2212.4, BIC $\approx$ 2226.4.

Datenbeispiel (iceberg orders), Interpretation:

Beobachtete Autokorrelation nicht allein durch „Reproduktionsmechanismus“ bestimmt (Händler zeigt nächsten Teil des Auftrags), sondern auch durch Abhängigkeit der Innovation ϵ_t von voriger iceberg-Zahl X_{t-1} (Steigerung des Basismittels 0.567 um ca. $0.188 \cdot X_{t-1}$).

Liegen viele iceberg orders zur Zeit $t-1$, werden zusätzliche neue iceberg orders angezogen.

(Aufdeckung von iceberg orders: Frey & Sandas (2009).)

Mögliche weitere Schritte:

- Asymptotik von Schätzern,
- Testverfahren zur Abgrenzung der Modelle,
- andere funktionale Abhängigkeit, z. B. Immigration weniger attraktiv bei großer Population;
- „density dependence“ z. B. bei binomialem AR(1)-Modell, eng verwandt zu n-patch-Metapopulationsmodellen.

Vielen Dank

für Ihr Interesse!

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fachbereich
Mathematik

Christian H. Weiß

Fachbereich Mathematik

Technische Universität Darmstadt

Literatur

Buckley & Pollett (2010): *Limit theorems for discrete-time metapopulation models*. Probab. Surveys 7, 53-83.

Ferland et al. (2006): *Integer-valued GARCH processes*. Jour. Time Series Analysis 27(6), 923-942.

Frey & Sandas (2009): *The impact of iceberg orders in limit order books*. CFR working paper 09-06.

Heathcote (1966): *A branching process allowing immigration*. Jour. Roy. Stat. Soc. B 28(1), 213-217.

Heinen (2003): *Modelling time series count data: an autoregressive conditional Poisson model*. CORE Discussion Paper No. 2003-63.

Jung & Tremayne (2011): *Useful models for time series of counts or simply wrong ones?* Adv. in Stat. Analysis 95(1), 59-91.

McKenzie (1985): *Some simple models for discrete variate time series*. Water Resources Bulletin 21(4), 645-650.

Steutel & van Harn (1979): *Discrete analogues of self-decomposability and stability*. Ann. Prob. 7(5), 893-899.

Weiß (2008): *Thinning operations for modelling time series of counts – a survey*. Adv. in Stat. Analysis 92(3), 319-341.

Weiß (2009): *Modelling time series of counts with overdispersion*. Stat. Meth. Appl. 18(4), 507-519.

Zhu & Wang (2009): *Estimation and testing for a Poisson autoregressive model*. Metrika 73(2), 211-230.