

Analyse und Modellierung diskretwertiger Zeitreihen

Forschung an der Professur für Quantitative Methoden der Wirtschaftswissenschaften

Prof. Dr. Christian H. Weiß

Fächergruppe Mathematik/Statistik, Helmut-Schmidt-Universität, Holstenhofweg 85, 22043 Hamburg.

Diskretwertige Zeitreihen

Unter einer Zeitreihe versteht man eine zeitlich geordnete Folge von Beobachtungen, $x_1, \dots, x_T$, von der man annimmt, dass sie Realisation eines zugrundeliegenden stochastischen Prozesses $(X_t)_{\mathbb{N}}$ ist. Um in der Lage zu sein, zukünftige Werte vorherzusagen, oder um den weiteren Verlauf von $(X_t)_{\mathbb{N}}$ mit dem Ziel zu überwachen, strukturelle Änderungen schnellstmöglich aufzudecken, ist eine Modellierung des Prozesses von zentraler Bedeutung, welche wiederum eine eingehende statistische Analyse der verfügbaren Zeitreihe $x_1, \dots, x_T$ voraussetzt. Spätestens mit dem wegweisenden Buch von [BJ70] ist die Zeitreihenanalyse zu einer etablierten Disziplin in Forschung, Lehre und Anwendung geworden. Allerdings wurde über lange Zeit hinweg eine Zeitreihe automatisch als reellwertig angenommen, so dass sich Fachbücher zur Zeitreihenanalyse bis zum heutigen Tag nahezu ausschließlich mit diesem Fall befassen. In Anwendungen dagegen ist man häufig mit Zeitreihen aus diskretwertigen Beobachtungen konfrontiert, etwa mit Zeitreihen aus Zählraten (also mit Wertebereich $\mathbb{N}_0 = \{0, 1, \dots\}$) oder gar qualitativen Zeitreihen mit nominalem oder ordinalem Wertebereich. In derartigen Fälle können die üblichen Modelle und Verfahren für reellwertige Zeitreihen nicht angewandt werden.

Erst in den vergangenen etwa 25 Jahren hat sich ein stetig wachsendes Interesse in Forschung und Praxis bezüglich diskretwertiger Zeitreihen entwickelt; eine Überblick bietet das Buch [W18]. Dabei wurden zahlreiche Modelle und Ansätze zur statistischen Analyse diverser Typen von diskretwertigen Zeitreihen entwickelt, und dies zum Teil an der Professur für Quantitative Methoden der Wirtschaftswissenschaften der HSU Hamburg. Über derartige Beiträge und zukünftige Forschungsrichtungen soll im Folgenden exemplarisch berichtet werden.

Zählratenzeitreihen

Der in der Praxis am häufigsten anzutreffende Typ von diskretwertiger Zeitreihe sind Zählratenzeitreihen; zwei derartige Beispiele zeigen die Abb. 1 und 2.

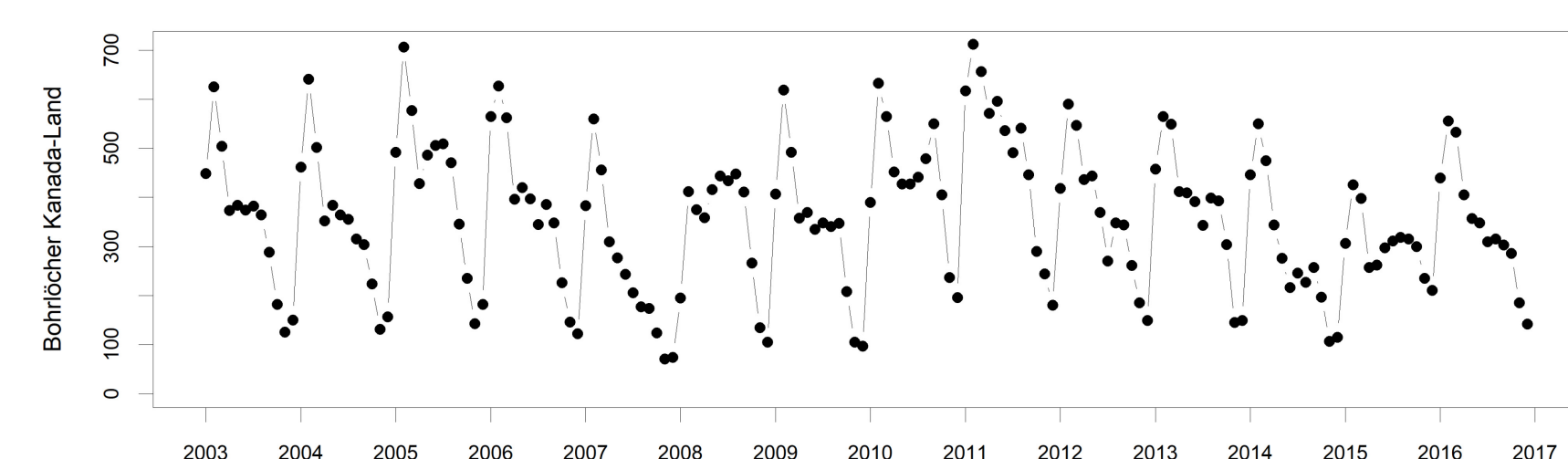


Abb. 1 Monatliche Zahl aktiver Bohrlöcher in Kanada-Land, 2003–2016. Quelle: Baker Hughes.

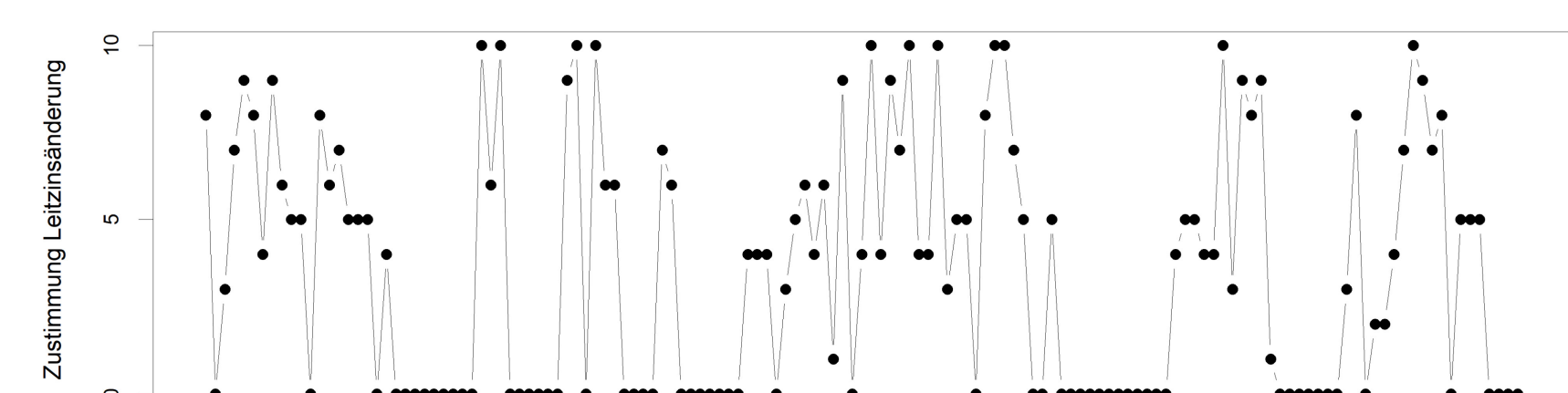


Abb. 2 Anzahl der $n = 10$ Mitglieder der Rates für Geldpolitik der Nationalbank Polens, die bei Sitzungen in 2002–2013 für eine Änderung des Leitzinses stimmten. [MWKS17].

Zählratenzeitreihen können sehr unterschiedliche Charakteristika aufweisen. Manche von ihnen (z.B. jene aus Abb. 1) verfügen über den unbegrenzten Wertebereich $\mathbb{N}_0$, andere über einen beschränkten Träger der Form $\{0, \dots, n\}$ (mit $n = 10$ in Abb. 2). Manche verhalten sich stationär (vgl. Abb. 2), andere weisen systematische Änderungen wie etwa Saisonalität auf (vgl. Abb. 1). Die Randverteilung ist zudem oft durch Eigenschaften wie Überdispersion oder Nullinflation geprägt, Letzteres zeigt Abb. 2.

Aktuelle Forschungsbeiträge:

- Eine Reihe von Arbeiten befassen sich mit der Modellierung und Analyse von Zählratenzeitreihen mit beschränktem Wertebereich. Beispielsweise werden in [MWKS17] vier verschiedene Modelle vorgeschlagen und analysiert, welche unterschiedliche Muster von Nullinflation erklären können. [WP14] entwickeln zustandsabhängige Modelle, [SWSP14] Modelle für bivariate Zählratenzeitreihen, und [MSWSP16] Modelle mit einem Schwellenwertmechanismus.

- Ein weiterer Schwerpunkt sind die Entwicklung diagnostischer Testverfahren für Zählratenzeitreihen. In [WHP16] werden z.B. Tests zur Aufdeckung von Überdispersion oder Nullinflation in INAR(1)-Prozessen entwickelt, in [WGM17] dagegen Tests zur Erkennung einer Compound-Poisson-Struktur bei INGARCH-Prozessen.
- In aktuell zur Begutachtung eingereichten Arbeiten geht es u.a. um Bootstrap-Verfahren für INAR-Prozesse, um Goodness-of-Fit-Tests bzgl. der Randverteilung von Zählratenzeitreihen, um die Performanz von Informationskriterien bei Zählratenzeitreihen, oder um ein auf binomialer Variation beruhendes ARMA-Modell für den Fall eines beschränkten Wertebereichs.

Aktuelle Forschungsprojekte:

- Dreijähriges Projekt „Kohärente Vorhersage und Risikoanalyse für Zählratenprozesse“, gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 394832307, und initiiert durch vorherige einjährige Förderung durch IFF2016.
- Einjähriges Projekt „Modelldiagnostik für Zählratenzeitreihen“, gefördert durch die IFF2018.

Kategoriale Zeitreihen

Während Zählratenzeitreihen einen quantitativen Wertebereich aufweisen und eine in vielerlei Hinsicht analoge Analyse und Modellierung wie bei klassischen reellwertigen Zeitreihen erlauben, erfordern kategoriale Zeitreihen mit ihrem qualitativen Wertebereich einen von Grund auf eigenständigen Zugang. Zu unterscheiden ist hierbei noch, ob der Wertebereich zumindest eine natürliche Anordnung aufweist (ordinale Zeitreihen) oder ob selbst dies nicht möglich ist (nominale Zeitreihe). Bei ordinalen Zeitreihen sind zumindest Quantile wohldefiniert, und eine grafische Darstellung der Zeitreihe analog zu den Abb. 1 und 2 ist möglich. Bei nominalen Zeitreihen dagegen müssen neue Konzepte zur statistischen und visuellen Analyse entwickelt werden, siehe [W18]. Im Beispiel aus Abb. 3a etwa (der Gesang des Singvogels „Waldtyrann“ ist eine Abfolge von drei möglichen Phrasen, also nominal) ist ein gewöhnlicher Zeitreihenplot nicht möglich, der Rate Evolution Graph erlaubt aber zumindest Einsicht in die Randverteilung. In [W13] wird die asymptotische Verteilung von speziellen Abhängigkeitsmaßen wie Cohen's $\hat{\kappa}(k)$ untersucht, u.a. mit dem Ziel, signifikante Abhängigkeiten aufdecken zu können. Beim Gesang des Waldtyranns (vgl. Abb. 3b) ergibt sich eine periodische Abfolge aus signifikant positiven und negativen Abhängigkeitswerten.

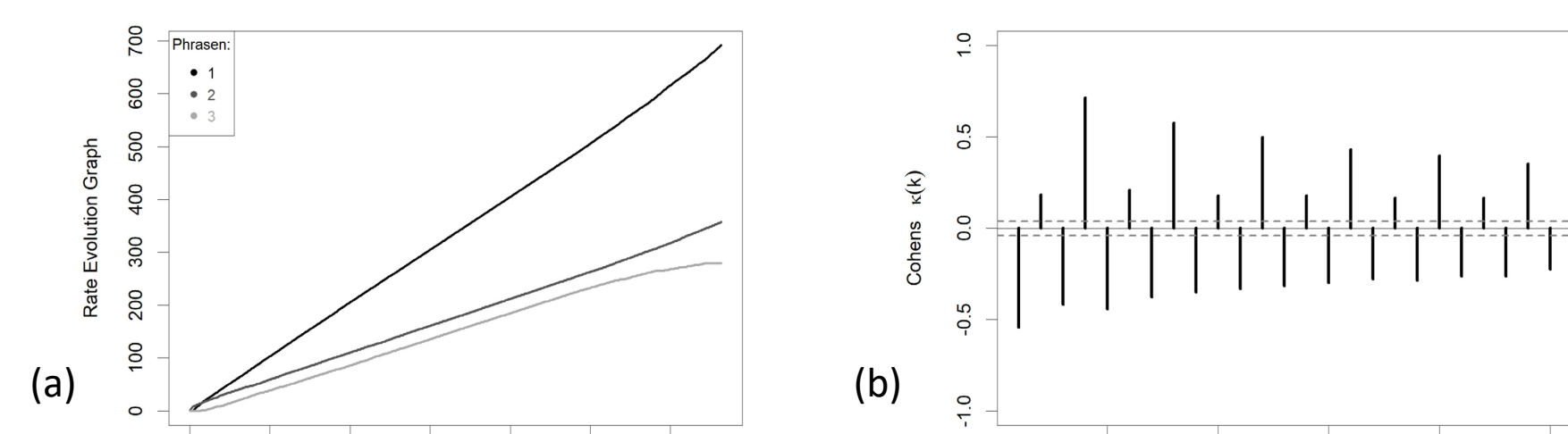


Abb. 3 Gesang des Waldtyranns, Wertebereich aus drei Phrasen: (a) Rate Evolution Graph, (b) Cohen's $\hat{\kappa}(k)$ [W18].

Für die Enzyklopädie *Wiley StatsRef: Statistics Reference Online*, (Balakrishnan et al. (eds.), John Wiley & Sons Ltd) wurden 2017 fünf Beiträge über diskretwertige Zeitreihen verfasst:

- Categorical Time Series Analysis
- Count Time Series Analysis
- Integer-valued Autoregr. Moving-Average (INARMA) Models
- INGARCH and Regression Models for Count Time Series
- Hidden-Markov Models for Count Time Series

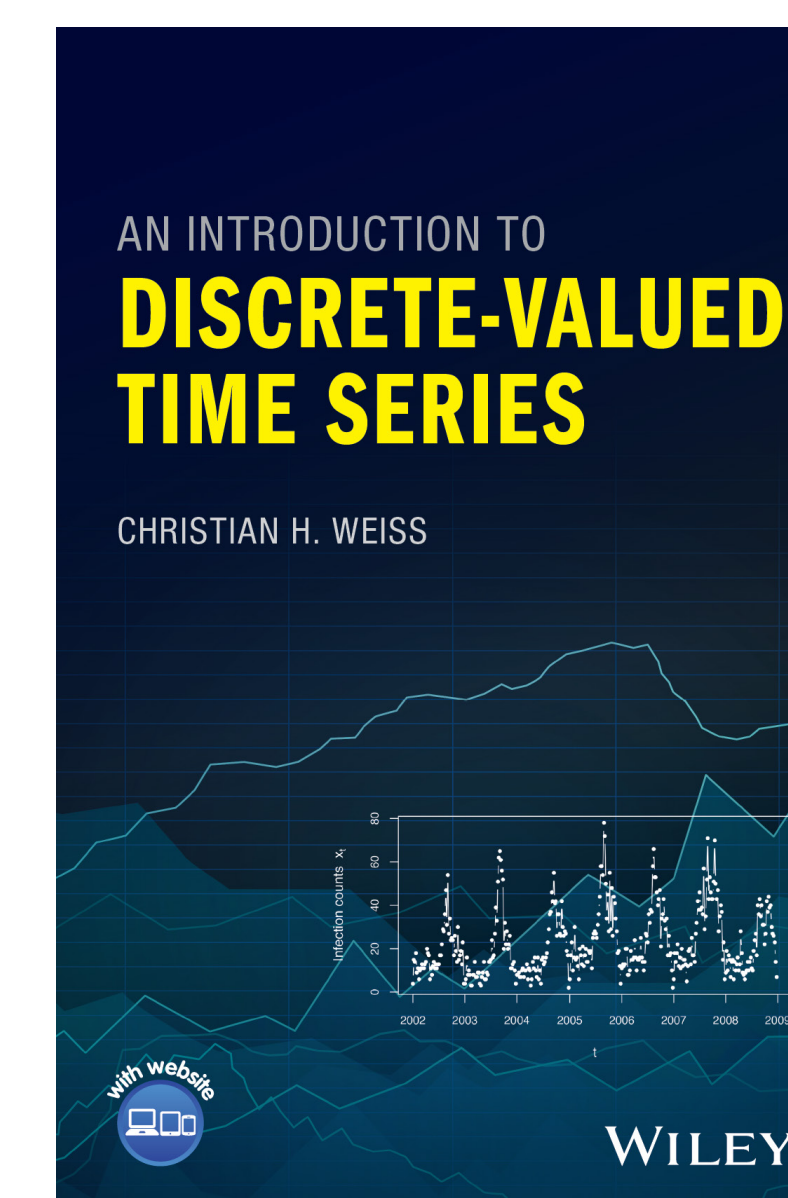
Statistische Prozesskontrolle

Ein weiteres Forschungsgebiet ist die Statistische Prozesskontrolle, bei der ein an sich stationärer Prozess $(X_t)_{\mathbb{N}}$ überwacht werden soll, um strukturelle Änderungen schnellstmöglich aufzudecken. Hierbei werden insbesondere sog. Kontrollkarten eingesetzt, die für den jeweils vorliegenden Prozess („in-control-Modell“) maßgeschneidert entworfen werden.

Solche Kontrollkarten gibt es vornehmlich für reellwertige Prozesse, seit jüngerer Zeit aber auch verstärkt für diskretwertige Prozesse, siehe [W18].

Aktuelle Forschungsbeiträge:

- Ein wichtiger Aspekt ist die Berücksichtigung von Schätzunsicherheit (ausgehend von sog. Phase-I-Daten) beim Design der Kontrollkarte, welches in [WT15,DWTK16] für den Fall autokorrelierter Zeitreihendaten untersucht wird.
- In [W17] werden diverse Kontrollkarten für seriell abhängige kategoriale Prozesse vorgeschlagen und verglichen.
- In aktuell zur Begutachtung eingereichten Arbeiten geht es u.a. um den Einsatz von Risikomaßen zur Bewertung von Kontrollkarten, um einen bootstrap-basierten Ansatz zum Design von Kontrollkarten mit garantierter ARL-Performanz, oder um eine neue kombinierte Shewhart-CUSUM-Karte mit einem Schwellenwertmechanismus.



Ein einführendes Buch zu diskretwertigen Zeitreihen (inkl. Prozesskontrolle) [W18]:

Christian H. Weiß
An Introduction to Discrete-Valued Time Series.

John Wiley & Sons, Inc.,
Chichester, 304 Seiten, 2018.
ISBN 978-1-119-09696-2

<https://www.wiley.com/go/weiss/discrete-valuedtimeseries>

Literatur:

- [BJ70] Box GEP, Jenkins GM (1970) *Time Series Analysis: Forecasting and Control*. 1. Auflage, Holden-Day, San Francisco.
- [DWTK16] Dasdemir E, Weiß CH, Testik MC, Knoth S (2016) Evaluation of Phase I analysis scenarios on Phase II performance ... *Qual. Engg.* 28, 293–304.
- [MWKS17] Möller TA, Weiß CH, Kim HY, Sirchenko A (2017) Modeling zero inflation in count data time series with bounded ... *Meth. Comp. Appl. Prob.*
- [MSWSP16] Möller TA, Silva ME, Weiß CH, Scotto MG, Pereira I (2016) Self-exciting threshold binomial autoregressive processes. *ASTA* 100, 369–400.
- [SWSP14] Scotto MG, Weiß CH, Silva ME, Pereira I (2014) Bivariate binomial autoregressive models. *J. Mult. Anal.* 125, 233–251.
- [W13] Weiß CH (2013) Serial dependence of NDARMA processes. *Comput. Statist. Data Anal.* 68, 213–238.
- [W17] Weiß CH (2017) Control charts for time-dependent categorical processes. In Knoth & Schmid (eds.): *Frontiers in Statistical Quality Control* 12.
- [W18] Weiß CH (2018) *An Introduction to Discrete-Valued Time Series*. John Wiley & Sons, Inc., Chichester.
- [WGM17] Weiß CH, Gonçalves E, Mendes Lopes N. (2017) Testing the compounding structure of the CP-INARCH model. *Metrika* 80, 571–603.
- [WHP16] Weiß CH, Homburg A, Puig P (2016) Testing for zero inflation and overdispersion in INAR(1) models. *Statist. Papers*.
- [WP14] Weiß CH, Pollett PK (2014) Binomial autoregressive processes with density dependent thinning. *J. Time Ser. Anal.* 35, 115–132.
- [WT15] Weiß CH, Testik MC (2015) On the Phase I analysis for monitoring time-dependent count processes. *IIE Transactions* 47, 294–306.

E-Mail:
weissc@hsu-hh.de

URL:
<https://www.hsu-hh.de/mathstat/>