



LUDWIG-MAXIMILIANS-UNIVERSITÄT
MÜNCHEN

BACHELORARBEIT

**Probabilistische Vorhersagen
von Zähldaten-Zeitreihen**

Maximilian Hahn

Betreuung:
Professor Dr. Thomas Augustin

Abgabe: 10.02.2021

Inhaltsverzeichnis

1	Einleitung	5
2	Konfidenzintervalle, probabilistische Vorhersagen	7
3	Verfahren	10
3.1	ARIMA	10
3.2	Tscount	11
3.2.1	Poisson-Verteilung	13
3.2.2	Negative Binomialverteilung	13
3.3	Generalized Additive Models for Location, Scale and Shape (GAMLSS)	14
3.3.1	Zero-Inflated Negative Binomialverteilung	19
3.3.2	Zero-Adjusted Negative Binomialverteilung	20
4	Methoden zur Evaluierung der Güte von probabilistischen Vorhersagen	22
4.1	Kalibrierung und Spitze der vorhergesagten Verteilung	22
4.1.1	probabilistische Kalibrierung	23
4.1.2	marginale Kalibrierung	25
4.1.3	Spitze der vorhergesagten Verteilung	25
4.2	Rootogramme	26
4.3	Pinball-Loss	26
4.4	Mittlere quadratische Abweichung	27
5	Beispielhafte Anwendung der Methoden zur Evaluierung der Güte von probabilistischen Vorhersagen	28
5.1	simulierte Zeitreihe	28
5.2	Modelle	29
5.3	Ergebnis	29

6	Anwendung der Verfahren auf reale Daten	33
6.1	Daten	33
6.2	Modellfindung	33
6.3	Ergebnisse	35
6.3.1	Zeitreihe 1	36
6.3.2	Zeitreihe 2	43
6.3.3	Zeitreihe 3	49
6.3.4	Zeitreihe 4	56
7	Fazit	64
	Literatur	68

Abbildungsverzeichnis

2.1	Konfidenzintervall und probabilistische Vorhersage	9
3.1	Wahrscheinlichkeitsfunktionen für unterschiedliche Parametrisierungen der Verteilungen	16
5.1	Simulation PIT Histogramme	30
5.2	Simulation Marginaler-Kalibrierungs-Plot	31
5.3	Simulation Sharpnes Plot	31
5.4	Simulation Rootogramm	32
6.1	Histogramm und Autokorellation für die vier Zeitreihen	34
6.2	Zeitreihe 1 PIT Histogramme	39
6.3	Zeitreihe 1 marginaler Kalibrierungs Plot	40
6.4	Zeitreihe 1 Sharpnes Plot	40
6.5	Zeitreihe 1 Rootogramm	41
6.6	Werte Zeitreihe 1 und Quantile der vorhergesagten Verteilung	42
6.7	Zeitreihe 2 PIT Histogramme	45
6.8	Zeitreihe 2 Marginaler-Kalibrierungs-Plot	46
6.9	Zeitreihe 2 Sharpnes Plot	46
6.10	Zeitreihe 2 Rootogramm	47
6.11	Werte Zeitreihe 2 und Quantile der vorhergesagten Verteilung	48
6.12	Zeitreihe 3 PIT Histogramme	51
6.13	Zeitreihe 3 Marginaler-Kalibrierungs-Plot	52
6.14	Zeitreihe 3 Sharpnes Plot	53
6.15	Zeitreihe 3 Rootogramm	54
6.16	Werte Zeitreihe 3 und Quantile der vorhergesagten Verteilung	55
6.17	Zeitreihe 4 PIT Histogramme	58
6.18	Zeitreihe 4 Marginaler-Kalibrierungs-Plot	59
6.19	Zeitreihe 4 Sharpnes Plot	60
6.20	Zeitreihe 4 Rootogramm	61
6.21	Werte Zeitreihe 4 und Quantile der vorhergesagten Verteilung	62

Tabellenverzeichnis

3.1	Verteilungsannahmen der verschiedenen Modelle	15
6.1	Zeitreihe 1, Pinball-Loss und Mittlere quadratische Abweichung	42
6.2	Zeitreihe 2, Pinball-Loss und Mittlere quadratische Abweichung	48
6.3	Zeitreihe 3, Pinball-Loss und Mittlere quadratische Abweichung	56
6.4	Zeitreihe 4, Pinball-Loss und Mittlere quadratische Abweichung	63

Kapitel 1

Einleitung

Die Vorhersage von Zeitreihen ist ein wichtiger Bestandteil in vielen Anwendungsbereichen, wie der Volkswirtschaftslehre, der Meteorologie und in der Entscheidungsfindung bei diversen betrieblichen Prozessen [9, 29, 11]. Durch die immer größer werdenden Datenmengen in Unternehmen durch Big Data und neue Konzepte wie IoT werden die potenziellen Anwendungsmöglichkeiten für die Zeitreihenvorhersage immer größer [28, 24].

Eine typische Anwendung der Zeitreihenvorhersage in Unternehmen ist die Bedarfsvorhersage [3]. Ziel dieser ist es, den Bedarf an Produkten, für einen gewissen Zeitraum in der Zukunft vorherzusagen. Dabei liegen häufig Zählraten vor, wie beispielsweise die Verkaufszahlen eines Produktes in einem Supermarkt. Klassische Zeitreihenverfahren sind Verfahren für stetige Daten, woraus für Zählratenzeitreihen problematische Modellannahmen bei der Verwendung dieser Verfahren resultieren. Außerdem werden bei der Vorhersage von Zeitreihen klassischerweise Punktvorhersagen betrachtet. In vielen Bereichen, konkret zum Beispiel bei der Bedarfsvorhersage, könnte jedoch eine probabilistische Vorhersage eine bessere Entscheidungsgrundlage sein. Probabilistische Vorhersagen für Zählratenzeitreihen sind in der Literatur bisher jedoch wenig diskutiert. Dies könnte daran liegen, dass die Vorhersage von Verteilungen komplizierter ist als die Punktvorhersage. Eine Punktvorhersage kann, beispielsweise durch die Erwartungstreue des Mittelwertschätzers, auch bei einer falschen Verteilungsannahme immer noch gute Ergebnisse liefern. Bei einer probabilistischen Vorhersage ist die geeignete Verteilungsannahme jedoch essenziell. Eine weitere Schwierigkeit ist die Auswahl des besten probabilistischen Vorhersagemodells. Die Bewertung von Punktvorhersagemodellen ist einfach und gut bekannt. Sie kann zum Beispiel mithilfe der Mittleren Quadratischen Abweichung erfolgen. Zur Bewertung von probabilistischen Vorhersagen gibt es weder einen vergleichbaren offensichtlichen Ansatz, noch gibt es ein allgemein in der Literatur angewendetes Verfahren.

Diese Arbeit soll sich den zwei Schwierigkeiten in der probabilistischen Vorhersage von Zählzeitenreihen widmen. Zum einen werden Verfahren zur Erstellung von probabilistischen Vorhersagen für Zählzeitenreihen vorgestellt, zum anderen werden verschiedene Methoden zur Evaluierung von probabilistischen Vorhersagen präsentiert. Die vorgestellten Verfahren für Zählzeitenreihen sind, das klassische Autoregressive-Integrated-Moving-Aver-Age Zeitreihen Modell (ARIMA) [2], spezielle Zählzeitenreihenmodelle aus dem Tscout[16] R-Paket [19] und das Generalized Additive Model for Location Scale and Shape (GAMLSS) [20], ein Modell für allgemeine Daten, dessen Verwendung zur Erstellung von probabilistischen Vorhersagen für Zählzeitenreihen in dieser Arbeit vorgeschlagen wird. Folgende Methoden zur Evaluierung von probabilistischen Vorhersagen werden vorgestellt. Das Konzept der Kalibrierung und Spitze der vorhergesagten Verteilung [12], Rootgramme [15] und die Pinball-Loss-Funktion.

Die Arbeit ist folgendermaßen strukturiert. Zunächst wird der Begriff der probabilistischen Vorhersage definiert und eine Abgrenzung von probabilistischen Vorhersagen und Konfidenzintervallen gegeben. Anschließend werden die Zeitreihenmodelle ARIMA, Tscout und GAMLSS vorgestellt. Das nächste Kapitel behandelt die Verfahren, die zur Evaluierung der probabilistischen Vorhersagen verwendet werden. Diese Verfahren werden dann im Anschluss auf simulierte Daten angewendet, um ihre Funktionsweise und Interpretation besser zu erklären. Anschließend folgt die Anwendung der Modelle auf reale Zählzeitenreihen sowie die Bewertung der probabilistischen Vorhersagen aus diesen Modellen, mithilfe der vorgestellten Methoden zur Evaluierung.

Kapitel 2

Konfidenzintervalle, probabilistische Vorhersagen

Der Begriff der probabilistischen Vorhersage ist in der Literatur nicht eindeutig definiert. In diesem Kapitel wird definiert, was in dieser Arbeit als probabilistische Vorhersage verstanden wird. Insbesondere soll die probabilistische Vorhersage von Konfidenzintervallen abgegrenzt werden. Die Abgrenzung erfolgt an einem einfachen Zeitreihenmodell, welches die aktuelle Beobachtung mithilfe einer linearen Regression auf den vorherigen Wert unter Normalverteilungsannahme erklärt, das Modell ist auch als AR(1) Modell bekannt. Sei $\{y_t : t = 1, 2, \dots, T\}$ eine Zeitreihe mit Zeitpunkten t . Die Zeitreihe wird nun folgendermaßen modelliert:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t \quad (2.1)$$

Wobei $\epsilon_t \sim N(0, \sigma)$ angenommen wird. Die Koeffizienten β_0, β_1 und die Standardabweichung σ werden aus den Daten als $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}$ geschätzt. Das bedeutet, der bedingte Erwartungswert ist zu jedem Zeitpunkt variabel und hängt von der zurückliegenden Beobachtung ab. Die Standardabweichung hingegen ist über alle Zeitpunkte gleich.

Ein **Konfidenzintervall** spiegelt die Unsicherheit in der Schätzung von β_0 und β_1 wieder. Für das gewählte Beispiel kann das Konfidenzintervall analytisch bestimmt werden. Sei dazu X die Designmatrix. Wobei $X \in \mathbb{R}^{2 \times (n-1)}$ mit dem Einheitsvektor in der ersten Spalte und $y_1, y_2, \dots, y_{(T-1)}$ in der zweiten Spalte. Die Annahme bei diesem Modell ist, dass die y_t unabhängig und identisch normalverteilt sind gegeben der Beobachtung y_{t-1} . Ist die Modellannahme erfüllt, kann ein α -Prozent-Konfidenzintervall für $E(y_t)$ folgendermaßen bestimmt werden: [10, p. 73]:

$$[\hat{\beta}_0 + \hat{\beta}_1 y_{t-1} - t_{T-2} \left(\frac{1-\alpha}{2} \right) \hat{\sigma} \left((1, y_{t-1})(X^T X)^{-1} (1, y_{t-1})^T \right)^{\frac{1}{2}},$$

$$\hat{\beta}_0 + \hat{\beta}_1 y_{t-1} + t_{T-2} \left(\frac{1-\alpha}{2} \right) \hat{\sigma} \left[(1, y_{t-1}) (X^T X)^{-1} (1, y_{t-1})^T \right]^{\frac{1}{2}}$$

Wobei $t_f(p)$ das p Quantil einer T-Verteilung mit f Freiheitsgraden bezeichnet. Alternativ, und für beliebige Verteilungsannahmen, kann ein solches Konfidenzintervall auch durch Bootstrap-Verfahren angenähert werden [7].

In Abbildung 2.1 ist eine simulierte AR(1) Zeitreihe und das sich aus einem geschätzten AR(1) Modell ergebende 90 %-Konfidenzintervall abgebildet.

Eine **probabilistische Vorhersage** gibt die gesamte Verteilung zum Vorhersagezeitpunkt an. In Abbildung 2.1 ist das zentrale 90 % Intervall abgetragen, welches sich aus der probabilistischen Vorhersage ergibt. Dieses ist deutlich breiter als das 90 %-Konfidenzintervall. Das zentrale 90 %-Intervall der probabilistischen Vorhersage ist im Beispiel:

$$[\hat{\beta}_0 - \hat{\beta}_1 y_{t-1} + \Phi(0.9) * \sigma, \hat{\beta}_0 + \hat{\beta}_1 y_{t-1} + \Phi(0.9) * \sigma]$$

Wobei $\Phi(0.9)$ das 90% Quantil der Standardnormalverteilung ist. Allgemein ergibt sich eine probabilistische Vorhersage aus der getroffenen Verteilungsannahme und den Parameterschätzern, die sich für diese Verteilung in Abhängigkeit von Kovariablen ergeben. Probabilistische Vorhersageintervalle, wie in Abbildung 2.1, lassen sich also für jedes parametrische Modell, gegeben der Parameterschätzer, angeben.

Die Beispiele in diesem Kapitel sollen als Erläuterung dienen, was im Rahmen dieser Arbeit als probabilistischen Vorhersage bezeichnet wird. Zur Vereinfachung der Erklärung werden one-step-ahead in-sample-Vorhersagen verwendet. In-sample Vorhersagen sind bei der Evaluierung von Vorhersagemodellen nur beschränkt aussagekräftig, deshalb werden in der späteren Anwendung out-of-sample-Vorhersagen betrachtet.

Im gewählten Beispiel ist die Varianz über alle Zeitpunkte per Annahme gleich, damit hat ein aus der probabilistischen Vorhersage gebildetes Intervall zu jedem Zeitpunkt die gleiche Breite. Die Intervalle verschieben sich mit der Zeit nur in ihrer Lage, welche durch den vorhergesagten Mittelwert bestimmt wird. Flexibler wird die probabilistische Vorhersage beispielsweise, wenn die Varianz zeitabhängig modelliert wird oder andere Verteilungsannahmen getroffen werden. Auf beide Aspekte wird in dieser Arbeit eingegangen.

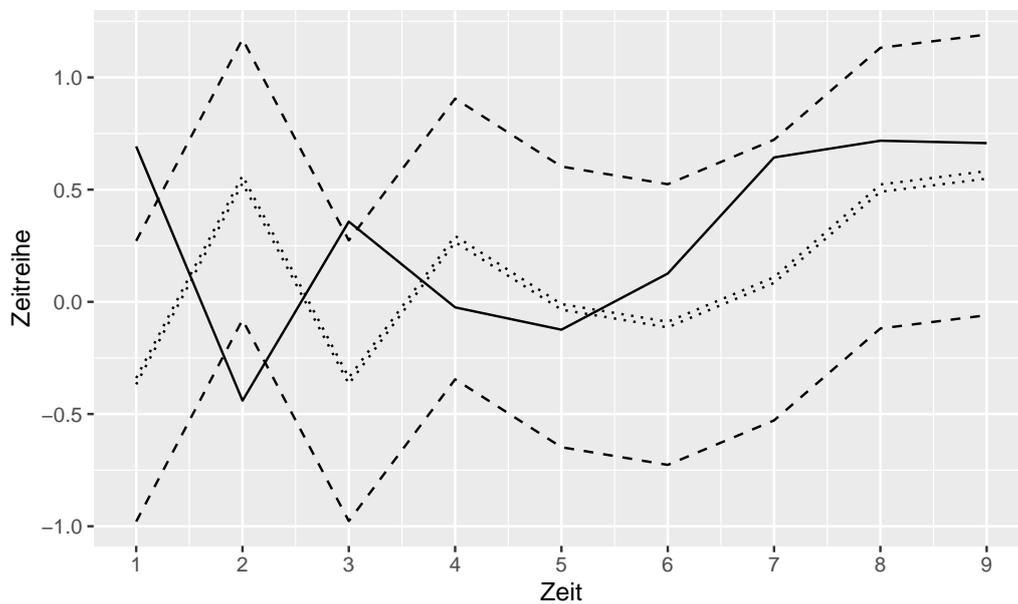


Abbildung 2.1: **Konfidenzintervall und probabilistische Vorhersage:** 90%-Konfidenzintervall (gepunktet) und zentrales 90%-Intervall (gestrichelt) aus der probabilistischen Vorhersage

Kapitel 3

Verfahren

3.1 ARIMA

Autoregressive-Integrated-Moving-Average (ARIMA) Modelle sind traditionell eine der am meisten verwendeten Methoden in der Zeitreihenvorhersage [13]. Die zugrunde liegende Theorie wurde bereits in den 1970er Jahren von Box, Jenkins und Reinsel entwickelt und im Folgenden genau erforscht und beschrieben [1]. Der Fokus in dieser Arbeit liegt auf der Erstellung von probabilistischen Vorhersagen. Im Folgenden wird die grundlegende Theorie der ARIMA-Modelle und die Erstellung von probabilistischen Vorhersagen mit diesen besprochen.

Es soll eine Zeitreihe $Y_t : t \in \mathbb{N}$ betrachtet werden. Wobei der Index t die Zeit darstellt. Das ARIMA-Modell besteht aus drei Teilen. Der autoregressive Teil ist eine lineare Regression auf die zurückliegenden Beobachtungen von Y :

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_t$$
$$\epsilon_t \sim N(0, \sigma)$$

Der Moving-Average Teil ist eine lineare Regression auf zurückliegenden Fehler des Modells:

$$Y_t = b_0 + b_1 \epsilon_{t-1} + b_2 \epsilon_{t-2} + \dots + b_p \epsilon_{t-p} + \epsilon_t$$

In der Praxis hat sich gezeigt, dass viele stochastische Prozesse die Komponenten beider Modelle aufweisen. Es kann zwar gezeigt werden, dass jedes endliche, autoregressive Modell als unendliches Moving-Average-Modell geschrieben werden kann und jedes endliche Moving-Average-Modell als unendliches autoregressives Modell [13]. Dieses Ergebnis ist allerdings von theoretischer Natur, denn in der praktischen Anwendung ist die Schätzung möglichst

weniger Parameter erwünscht. Das Autoregressive-Moving-Average-Modell (ARMA) lässt sich also schreiben als:

$$Y_t = b_0 + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + \dots + b_p\epsilon_{t-p} + \beta_1Y_{t-1} + \beta_2Y_{t-2} + \dots + \beta_pY_{t-p} + \epsilon_t$$

Das ARMA-Modell setzt voraus, dass die Zeitreihe stationär ist. Dies ist in der Praxis oft nicht der Fall. Deshalb wird zum Beispiel durch das Bilden von Differenzen $Y'_t = Y_t - Y'_{t-1}$ eine stationäre Zeitreihe hergestellt [13, Kapitel 8.1]. Das Bilden der Differenzen wird auch als Integrating (I) bezeichnet, damit ergibt sich das ARIMA-Modell:

$$Y'_t = b_0 + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + \dots + b_p\epsilon_{t-p} + \beta_1Y'_{t-1} + \beta_2Y'_{t-2} + \dots + \beta_qY'_{t-q} + \epsilon_t$$

Wegen der Modellannahme $\epsilon_t \sim N(0, \sigma)$ gilt:

$$E(Y'_t) = \mu'_t = b_0 + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + \dots + b_p\epsilon_{t-p} + \beta_1Y'_{t-1} + \beta_2Y'_{t-2} + \dots + \beta_qY'_{t-q}$$

Die Koeffizienten b_i mit $i = 1, 2, \dots, p$, β_j mit $j = 1, 2, \dots, q$ und die Varianz des Fehlerterms ϵ_t für $t = 1, 2, \dots, N$ σ^2 werden zum Beispiel mit einem Maximum-Likelihood Ansatz aus den Daten geschätzt [1, Kapitel 7]. Mit den geschätzten Koeffizienten \hat{b}_i , $\hat{\beta}_j$ ergibt sich der vorhergesagte Erwartungswert zum Zeitpunkt t als:

$$\hat{\mu}'_t = \hat{b}_0 + \hat{b}_1\epsilon_{t-1} + \hat{b}_2\epsilon_{t-2} + \dots + \hat{b}_p\epsilon_{t-p} + \hat{\beta}_1Y'_{t-1} + \hat{\beta}_2Y'_{t-2} + \dots + \hat{\beta}_qY'_{t-q}$$

$$\hat{\mu}_t = \hat{\mu}'_t + Y_{t-1}$$

Mit der geschätzten Varianz $\hat{\sigma}^2$ ergibt sich die probabilistische Vorhersage. Die Dichtefunktion der vorhergesagten Verteilung zum Zeitpunkt t ist:

$$f_t(x) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{(x - \hat{\mu}_t)^2}{2\hat{\sigma}^2}\right)$$

Wobei die Standardabweichung der Normalverteilung $\hat{\sigma}$ zu jedem Zeitpunkt gleich ist und der Erwartungswert $\hat{\mu}_t$ zu jedem Zeitpunkt unterschiedlich ist, abhängig von den zurückliegenden Beobachtungen und den zurückliegenden Fehlern.

Für alle Analysen in dieser Arbeit wird das R-Paket [19] forecast [14] verwendet.

3.2 Tscout

Die klassische Verteilungsannahme für Zähldaten ist die Poisson-Verteilung. Bei der Poisson-Verteilung ist der Erwartungswert gleich der Varianz. Dieser Zusammenhang ist in realen Zähldaten häufig nicht gegeben. Die Varianz ist meist größer als der Erwartungswert, in diesem Fall spricht man

von Überdispersion. Eine Zähldatenverteilung, welche die Überdispersion berücksichtigt, ist die negative Binomialverteilung, hierbei können Erwartungswert und Varianz unterschiedliche Werte annehmen. Dies erlaubt weit flexiblere Modelle, als bei der Verwendung der Poisson-Verteilung. In Abbildung 3.1 sind Dichtefunktionen mit unterschiedlicher Parametrisierung der Poisson-Verteilung und negative Binomialverteilung abgebildet. Für unabhängig, identisch, verteilte Daten sind die Poisson-Verteilung und die negative Binomialverteilung durch klassische "generalized linear models"[18] abgedeckt. Für Zeitreihen ist die Annahme der Unabhängigkeit einzelner Beobachtungen schwierig. In der Literatur gibt es verschieden Ansätze Zähldatenzeitreihen zu modellieren [25]. Ein allgemein anerkannter und verwendeter Ansatz vergleichbar zum ARIMA-Modell bei diskreten Zeitreihen, hat sich bisher nicht etabliert.

Das Tscout- [16] R-Paket ist ein speziell für Zähldatenzeitreihen entwickeltes Paket. Mit Tscout können sowohl Poisson als auch negativ binomial verteilte Modelle geschätzt werden. Dieses Paket soll in dieser Arbeit zur Anwendung kommen.

Mit dem Tscout Paket werden Poisson-Verteilte oder negativ binomial verteilte Modelle geschätzt, bei denen der bedingte Erwartungswert der Verteilung von zurückliegenden Werten der Zeitreihe, als auch von zurückliegenden Werten von sich selbst, abhängt. Für diese Art von Modellen gibt es in der Literatur keine eindeutige Namensgebung. In der Literatur ist das Modell unter anderem als Integer Valued Generalized AutoRegressive Conditional Heteroscedasticity (INGARCH) Modelle zu finden [25, p. 73].

Es soll erneut eine Zeitreihe $Y_t : t \in \mathbb{N}$ betrachtet werden. Modelliert wird nun der bedingte Erwartungswert der Zeitreihe zum Zeitpunkt t gegeben aller Werte der Zeitreihe und aller bedingter Erwartungswerte der Zeitreihe vor dem Zeitpunkt t . Dabei bezeichnet Y_{t-} alle Werte der Zeitreihe vor dem Zeitpunkt t und λ_{t-} alle Werte des bedingten Erwartungswerts vor dem Zeitpunkt t .

$$E(Y_t | Y_{t-}, \lambda_{t-}) = \lambda_t$$

$$g(\lambda_t) = \beta_0 + \sum_{k=1}^p \beta_k Y_{t-i_k} + \sum_{l=1}^q \alpha_l g(\lambda_{t-j_l}) \quad (3.1)$$

Wobei $g(x)$ eine der bekannten Linkfunktionen Identität, Log oder Logit ist. Die Mengen $\{i_k : k = 1, 2, \dots, p\}$ und $\{j_l : l = 1, 2, \dots, q\}$ sind Mengen aufsteigender natürlicher Zahlen, die es erlauben verschiedene Kombinationen von Lags in das Modell einzubeziehen ($0 < i_1 < i_2 \dots < i_p < \infty$ und $0 < j_1 < j_2 \dots < j_q < \infty$). β_0 ist der Intercept, die Koeffizienten β_k geben den Einfluss zurückliegender Beobachtungen auf $g(\lambda_t)$ an, die Koeffizienten α_l geben den Einfluss der zurückliegenden Erwartungswerte auf $g(\lambda_t)$ an. Die

Koeffizienten β_0, β_k für $k = 1, 2, \dots, p$ und α_l für $l = 1, 2, \dots, q$ sollen aus den Daten geschätzt werden. Im Tscout-Paket werden die Koeffizienten durch die Maximierung der quasi-bedingten Likelihoodfunktion ermittelt [16, Kapitel 3].

Die angegebene Modellspezifikation ist sehr nah an der allgemeinen Spezifikation, die in Tscout verwendet wird. Zusätzlich erlaubt Tscout auch das Einbeziehen externer Kovariablen, wovon in dieser Arbeit kein Gebrauch gemacht wird.

Tscout erlaubt Verteilungsannahmen aus der Poisson-Verteilung und der negativen Binomialverteilung zu treffen. Diese zwei Modelle werden in den folgenden zwei Absätzen besprochen.

3.2.1 Poisson-Verteilung

Eine Poisson-Verteilung ist die klassische Modellannahme für Zählraten. Sei X eine Poisson verteilte Zufallsvariable, dann wird die Wahrscheinlichkeit des Auftretens von x mit folgender Wahrscheinlichkeitsfunktion bestimmt:

$$f(x) = \frac{\lambda^x}{x!} \exp(-\lambda)$$

Der Erwartungswert und die Varianz sind dabei gleich λ . Die Poisson-Verteilung ist also eine Verteilung mit einem Parameter $\lambda = E(X) = Var(X)$. Im Tscout-Paket, unter Verwendung einer Poisson-Verteilung, wird angenommen, dass jede Beobachtung Y_t einer Zeitreihe $\{Y_t : t = 1, 2, \dots, n\}$, aus einer Poisson-Verteilung mit Dichtefunktionen $f(y|\lambda_t)$ stammt. λ_t wird dabei als abhängig von zurückliegenden Werten der Zeitreihe und als abhängig von zurückliegenden Werten von λ selbst wie in Gleichung 3.1 angegeben modelliert. Die verwendeten zurückliegenden Beobachtungen $\{Y_{t-i_k}^i = 1, 2, \dots, p\}$ und die zurückliegenden Erwartungswerte $\{\lambda_{t-i_k}^i = 1, 2, \dots, q\}$ sind entscheidend für die Qualität des Modells. Die verwendete Methode zur Auswahl der Beobachtungen, Erwartungswerte in dieser Arbeit wird in Kapitel 6 besprochen.

3.2.2 Negative Binomialverteilung

Für eine Poisson verteilte Zufallsvariable X gilt $E(X) = Var(X)$. Im Fall realer Daten stellt dies oftmals ein Problem dar, denn die beobachtete Varianz ist hierbei oft größer als der beobachtete Erwartungswert. Dies wird auch als Überdispersion bezeichnet. In diesem Fall ist die Annahme einer negativen Binomialverteilung besser geeignet, denn diese lässt Varianzen zu,

die sich vom Erwartungswert unterscheiden. Mit der Parametrisierung, λ Erwartungswert und ϕ Überdispersionsparameter, ist die Wahrscheinlichkeitsfunktion der negativen Binomialverteilung folgende:

$$f(x) = \frac{\Gamma(\phi + x)}{\Gamma(x + 1)\Gamma(\phi)} \left(\frac{\phi}{\phi + \lambda}\right)^\phi \left(\frac{\lambda}{\phi + \lambda}\right)^x \quad (3.2)$$

Es gilt:

$$E(X) = \lambda$$

$$Var(X) = \lambda + \frac{\lambda^2}{\phi}$$

Für den Überdispersionsparameter gilt $\phi \in (0, \infty)$, je kleiner dieser ist, desto größer ist die Varianz.

Die negative Binomialverteilung ist also eine Verteilung mit zwei Parametern. Der Parameter λ wird abhängig von zurückliegenden Beobachtungen und Erwartungswerten wie in Gleichung 3.1 angegeben modelliert. Der Überdispersionsparameter ϕ wird nicht als abhängig modelliert, sondern wird gegeben der geschätzten Werte für $\lambda_{it} = 1, 2, \dots, T$ konstant über alle Zeitpunkte T geschätzt [16, Kapitel 3]. Im Kontext von Zeitreihen könnte es interessant sein, nicht nur den Erwartungswert einer Verteilung als abhängig von zurückliegenden Beobachtungen zu modellieren, sondern mehrere Parameter der Verteilung. Ein Modell, welches dies ermöglicht, wird im nächsten Kapitel vorgestellt. In Tabelle 3.1 sind die Dichtefunktionen aller in dieser Arbeit verwendeten Verteilungen gegenübergestellt. Dabei wird für jede Verteilung angegeben, welche Parameter jeweils als abhängig von zurückliegenden Werten, oder als fest über alle Zeitpunkte, modelliert werden.

3.3 Generalized Additive Models for Location, Scale and Shape (GAMLSS)

Die bisher verwendeten Verteilungen für Zähldaten, Poisson-Verteilung und negative Binomialverteilung, können bei passender Datenlage bereits gute Ergebnisse erzielen. In realen Zähldatenzeitreihen kann es jedoch auch zu gehäuften Nullwerten kommen. Das bedeutet, die Null tritt häufiger auf, als es bei einer Poisson-Verteilung oder negativen Binomialverteilung der Fall wäre. Die Verwendung der Poisson-Verteilung oder negativen Binomialverteilung ist in diesem Fall problematisch. Je häufiger Nullbeobachtungen auftreten, bei ansonst größeren positiven Beobachtungen, desto ungenauer wird die Modellierung unter der Annahme einer dieser zwei Verteilungen. Sie unterschätzen dann, entweder die Wahrscheinlichkeit mit welcher

Model	Dichte Funktion	variable Parameter	feste Parameter
ARIMA	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ
Tscount Poisson	$f(x) = \frac{\lambda^x}{x!} \exp(-\lambda)$	λ	
Tscount neg. Binomial	$f(x) = \frac{\Gamma(\phi+x)}{\Gamma(x+1)\Gamma(\phi)} \left(\frac{\phi}{\phi+\lambda}\right)^\phi \left(\frac{\lambda}{\phi+\lambda}\right)^x$	λ	ϕ
GAMLSS zero inf. neg. Binomial	$x = 0:$ $f(x) = \nu + (1 - \nu) \frac{\Gamma(\frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})} \left(\frac{1}{1+\sigma\mu}\right)^{\frac{1}{\sigma}}$ $x \geq 1:$ $f(x) = (1 - \nu) \frac{\Gamma(x+\frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})\Gamma(x+1)} \left(\frac{\sigma\mu}{1+\sigma\mu}\right)^x \left(\frac{1}{1+\sigma\mu}\right)^{\frac{1}{\sigma}}$	μ, σ, ν	
GAMLSS zero adj. neg. Binomial	$x = 0:$ $f(x) = \nu$ $x \geq 1:$ $f(x) = c \frac{\Gamma(x+\frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})\Gamma(x+1)} \left(\frac{\sigma\mu}{1+\sigma\mu}\right)^x \left(\frac{1}{1+\sigma\mu}\right)^{\frac{1}{\sigma}}$ mit $c = \frac{(1-\nu)}{1 - \left(\frac{1}{1+\sigma\mu}\right)^{\frac{1}{\sigma}}}$	μ, σ, ν	

Tabelle 3.1: Verteilungsannahmen der verschiedenen Modelle: Dichtefunktion der jeweiligen Verteilungsannahme, variable Parameter sind jene Parameter, die für jeden Zeitpunkt der Zeitreihe variabel sind. Feste Parameter hingegen sind für jeden Zeitpunkt der Zeitreihe gleich.

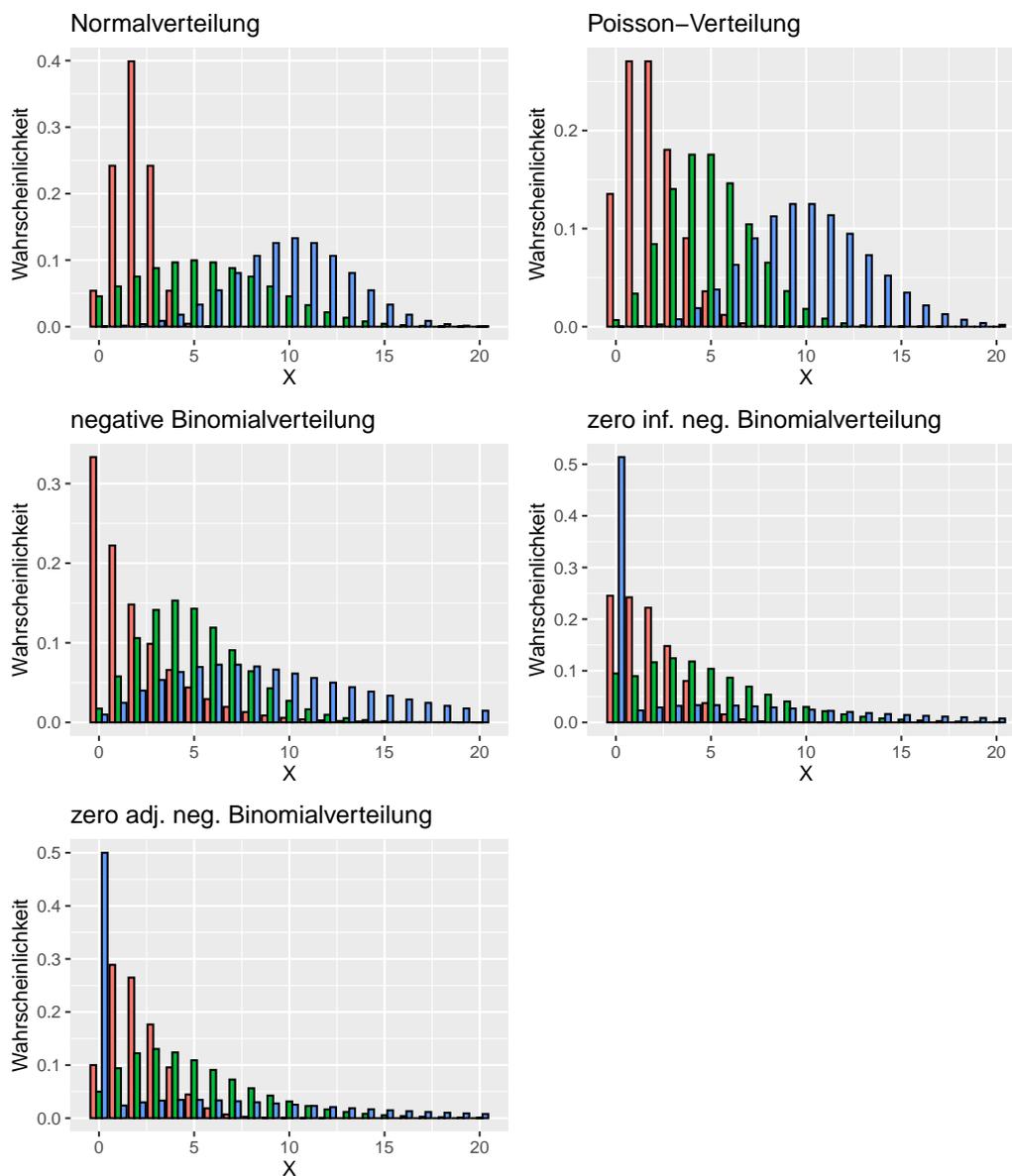


Abbildung 3.1: **Wahrscheinlichkeitsfunktionen für unterschiedliche Parametrisierungen der Verteilungen:** Normalverteilung (μ, σ): rot (2,1), grün (5,4), blau (10, 3); Poisson-Verteilung (λ): rot (2), grün (5), blau (10); neg. Binomialverteilung (λ, ϕ): rot (2,1), grün (5,0.1), blau (10,0.3); zero inf. neg. Binomialverteilung: (μ, σ, ν): rot (2,0.1,0.1), grün (5,0.3,0.05), blau (10,0.5,0.5); zero adj. neg. Binomialverteilung: (μ, σ, ν): rot (2,0.1,0.1), grün (5,0.3,0.05), blau (10,0.5,0.5)

Null auftritt, oder die Wahrscheinlichkeitsverteilung wird insgesamt in Richtung Null gezogen und bildet somit die positiven Werte schlecht ab. Zero-inflated oder zero-adjusted-Verteilungen sind bei solchen Daten mit gehäuft auftretenden Nullwerten eine Alternative. Konkret bieten sich hier die zero-inflated-Poisson-Verteilung, Zero-Inflated-Negativ-Binomialverteilung (ZINBI) und Zero-Adjusted-Negativ-Binomialverteilung (ZANBI) an. Die Anwendung von Modellen mit solchen Verteilungsannahmen auf Zeitreihendaten ist in der Literatur bisher wenig diskutiert. Das R-Paket ZIM [27] [26] ist speziell zur Modellierung von Zähldatenzeitreihen entwickelt worden und erlaubt es, zero-inflated-Poisson und ZINBI-Modelle zu schätzen. Dabei wird im Fall der ZINBI-Modelle der Erwartungswert und die Wahrscheinlichkeit für das Auftreten von Null in Abhängigkeit der Daten modelliert. Der Überdispersionsparameter wird konstant über alle Beobachtungen ermittelt, analog zur negativen Binomialverteilung im Tscout Paket.

In dieser Arbeit wird das GAMLSS-Modell [20] als Ansatz zur Modellierung von Zähldatenzeitreihen vorgeschlagen. Dieser Ansatz bietet zwei Vorteile. Zum einen, können alle Parameter einer Verteilung als Funktion zurückliegender Beobachtungen modelliert werden, dies stellt einen Unterschied zu den bisher in dieser Arbeit betrachteten Verfahren dar, welche ausschließlich den Erwartungswert der Verteilung als Funktion zurückliegender Beobachtungen modellieren. Zum anderen, erlaubt es GAMLSS frei aus einer großen Anzahl an Verteilungsannahmen zu wählen. Konkret sollen ZINBI und ZANBI-Modelle betrachtet werden. Möglich wäre auch ein zero-inflated-Poisson-Modell, wie sich jedoch im Kapitel 5.2 Ergebnisse zeigt, ist das Poisson-Modell nicht gut für die in dieser Arbeit verwendeten Daten geeignet. Daher wird im Folgenden die ZINBI- und ZANBI-Verteilung verwendet.

Die Klasse der GAMLSS-Modelle kann als eine Weiterentwicklung und Kombination von Generalized Linear Model (GLM), Generalized Additive Model (GAM) und Generalized Linear Mixed Model (GLMM) verstanden werden. Es können also Modelle mit linearen Effekten, glatten Effekten und zufälligen Effekten geschätzt werden. Die Weiterentwicklung liegt darin, dass man nicht mehr auf Verteilungsannahmen aus der Exponentialfamilie beschränkt ist, sondern eine Vielzahl an Verteilungen zur Verfügung stehen. Außerdem kann bei diesen Verteilungen jeder Parameter als Funktion der Kovariablen modelliert werden und nicht nur der Erwartungswert, wie bei GLM, GAM und GLMM. Für diese Arbeit genügt die Betrachtung des einfachen linearen Modells. Sei dazu $f(y|\Theta)$ eine Dichtefunktion. Der Parameter $\Theta_t = (\theta_t^1, \theta_t^2, \dots, \theta_t^p)$ parametrisiert dabei die Verteilung. Das bedeutet, es sind theoretisch Verteilungsannahmen mit beliebig vielen Parametern möglich. Konkret sind im GAMLSS- R-Paket Verteilungen mit maximal vier Parametern implementiert [22]. Modelliert wird nun jeder Parameter in Abhängigkeit von Kova-

riablen:

$$g_k(\theta^k) = X_k \beta_k \quad (3.3)$$

Dabei ist $g_k(x)$ eine Linkfunktion. X_k eine $n \times p_k$ dimensionale Designmatrix. Die Abhängigkeit von k bedeutet, dass es für jeden Parameter θ_k der Verteilung eine unterschiedliche Linkfunktion und Designmatrix geben kann. Die Kovariablen können also für jeden Parameter unterschiedlich sein. β_k ist ein Vektor der Länge p_k , der den Einfluss der Kovariablen auf den Parameter θ_k widerspiegelt. Die Parameter β_k sollen aus den Daten geschätzt werden. Die Schätzung der Parameter erfolgt durch die Maximierung der penalisierten Likelihood-Funktion [20]. Die Maximierung erfolgt numerisch und wird entweder durch den Algorithmus von Cole und Green [4] oder den Algorithmus von Rigby und Stasinopoulos [21] erreicht. Eine ausführliche Erklärung der Schätzverfahren und Algorithmen findet sich in „flexible regression and smoothing: using GAMLSS in R“ [23].

X_k ist die Designmatrix und enthält die Kovariablen, von denen Parameter θ_k abhängt. Im Fall einer Zeitreihe bietet es sich an, zurückliegende Beobachtungen als Kovariablen zu verwenden. Dies wird als autoregressives Modell bezeichnet. Um mit GAMLSS eine Zeitreihe zu modellieren, soll in dieser Arbeit ein solches autoregressives Modell verwendet werden. Sei dazu $\{y_t : t = 1, 2, \dots, n\}$ eine Zeitreihe. Jede Beobachtung y_t ist dabei die Realisierung einer Zufallsvariablen mit Dichtefunktion $f(y|\Theta)$. Der Parametervektor $\Theta = (\theta_1, \theta_2, \dots, \theta_K)$ parametrisiert die Verteilung vollständig, wobei K der Anzahl der Parameter der Verteilung entspricht. Das GAMLSS-Modell als autoregressives Modell für eine Zeitreihe ist dann:

$$g_k(\theta_k) = \sum_{j=1}^J \beta_0^k + \beta_j^k * y_{t-l_j} \quad (3.4)$$

Wobei J die Anzahl der verwendeten zurückliegenden Beobachtungen angibt. $L = \{l_j : j = 1, 2, \dots, J\}$ gibt an, wie weit die Beobachtungen jeweils zurückliegen. y_{t-l_j} ist also die Beobachtung, die l_j Zeitpunkte vor der Beobachtung y_t beobachtet wurden. $g_k(x)$ ist wieder eine bekannte Linkfunktion: Identität, Log, Logit. β_0^k ist der Intercept für Parameter θ_k und β_j^k ist der Koeffizient, der angibt, welchen Einfluss die Beobachtung y_{t-l_j} auf die Beobachtung y_t hat.

Die Abhängigkeit aller Parameter θ_t^k $k = 1, 2, \dots, K$ von zurückliegenden Beobachtungen stellt den Unterschied, zu den bisher betrachteten Modellen dar. In diesen Modellen wurde jeweils nur ein Parameter, der bedingte Erwartungswert, als variabel angenommen, die übrigen Parameter wurden als fest über alle Zeitpunkte angenommen.

Für das GAMLSS wird angenommen, dass die einzelnen Beobachtungen unabhängig voneinander sind, gegeben des Parametervektors Θ . Die Parameter sind eine Funktion zurückliegender Beobachtungen, wie in Gleichung 3.4 angegeben. Das impliziert für das autoregressive GAMLSS-Modell, dass die Beobachtungen y_t unabhängig voneinander sind, gegeben einer Menge von zurückliegenden Beobachtungen. Diese Annahme ist durchaus kritisch zu betrachten. Die Wahl der zurückliegenden Beobachtungen, die in ein Modell aufgenommen werden, ist hierbei von großer Bedeutung. In Kapitel 6 wird beschrieben, wie die Auswahl der zurückliegenden Beobachtungen in dieser Arbeit erfolgt.

GAMLSS erlaubt es, eine Verteilungsannahme aus einer Vielzahl von Verteilungen zu treffen. Darunter sind auch eine große Anzahl an Zähldatenverteilungen. In den folgenden zwei Abschnitten werden die Zero-Inflated Negativ Binomialverteilung und die Zero-Adjusted Negativ Binomialverteilung besprochen.

3.3.1 Zero-Inflated Negative Binomialverteilung

Überdispersion und Nullwerte können in der Modellierung von Zähldatenzeitreihen, die Annahme einer klassischen Zähldatenverteilung erschweren. Weist eine Zeitreihe beide Charakteristiken auf, ist die Annahme einer Zero-Inflated Negativ Binomialverteilung (ZINBI) sinnvoll. Die ZINBI ergibt sich als Mischverteilung aus einer Binomialverteilung und einer negativen Binomialverteilung. Die Wahrscheinlichkeitsfunktion einer ZINBI verteilten Zufallsvariablen X ist [22, p. 100],

Für $x = 0$:

$$f(x) = \nu + (1 - \nu) \left(\frac{1}{1 + \sigma\mu} \right)^{\frac{1}{\sigma}} \quad (3.5)$$

Für $x \geq 1$:

$$f(x) = (1 - \nu) \frac{\Gamma(x + \frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})\Gamma(x + 1)} \left(\frac{\sigma\mu}{1 + \sigma\mu} \right)^x \left(\frac{1}{1 + \sigma\mu} \right)^{\frac{1}{\sigma}} \quad (3.6)$$

Bei der ZINBI können Nullwerte auf zwei verschiedene Arten auftreten. Zum einen, wenn die Binomialverteilung der Mischverteilung den Wert Null annimmt, oder wenn die negativ Binomialverteilung der Mischverteilung den Wert Null annimmt. Positive Werte nimmt die ZINBI an, wenn die Binomialverteilung Eins und die negative Binomialverteilung einen Wert ungleich Null annimmt. Erwartungswert und Varianz der ZINBI sind:

$$E(x) = (1 - \nu)\mu$$

$$Var(x) = \mu(1 - \nu) + \mu^2(1 - \nu)(\sigma + \nu)$$

Die ZINBI ist somit eine Verteilung mit drei Parametern. Alle drei Parameter werden im autoregressiven GAMLSS Model in Abhängigkeit der zurückliegenden Werte der Zeitreihe modelliert.

$$g_\mu(\mu_t) = \sum_{j=1}^{J^\mu} \beta_j^\mu * y_{t-l_j^\mu} \quad (3.7)$$

$$g_\sigma(\sigma_t) = \sum_{j=1}^{J^\sigma} \beta_j^\sigma * y_{t-l_j^\sigma} \quad (3.8)$$

$$g_\nu(\nu_t) = \sum_{j=1}^{J^\nu} \beta_j^\nu * y_{t-l_j^\nu} \quad (3.9)$$

Für die Linkfunktion werden im Folgenden die Standardeinstellungen von GAMLSS verwendet, $g_\mu(x) := \log(x)$, $g_\sigma(x) := \log(x)$, $g_\nu(x) := \text{logit}(x)$. Die verwendeten zurückliegenden Beobachtungen, $\{l_j^\mu : j = 1, 2, \dots, J^\mu\}$, $\{l_j^\sigma : j = 1, 2, \dots, J^\sigma\}$ und $\{l_j^\nu : j = 1, 2, \dots, J^\nu\}$ werden in Kapitel 6 für jedes der Modelle angegeben.

3.3.2 Zero-Adjusted Negative Binomialverteilung

Die Zero-Adjusted Negative Binomialverteilung (ZANBI) adressiert das gleiche Problem wie die ZINBI: Überdispersion und gehäufte Nullwerte. Während bei der ZINBI Nullwerte auf zwei verschiedene Arten auftreten können, nämlich, wenn die Binomialverteilung den Wert Null annimmt oder wenn die negative Binomialverteilung den Wert Null annimmt, so treten bei der ZANBI Nullwerte nur mit einer festen Wahrscheinlichkeit ν auf. Positive Werte treten mit der Wahrscheinlichkeit einer negativen Binomialverteilung auf. Wobei die Wahrscheinlichkeitsfunktion der negativen Binomialverteilung mit dem Faktor c korrigiert werden muss, um auf eine gültige Wahrscheinlichkeitsfunktion zu kommen. Die Wahrscheinlichkeitsfunktion der ZANBI ist [22],

Für $x = 0$:

$$f(x) = \nu \quad (3.10)$$

Für $x \geq 1$:

$$f(x) = c \frac{\Gamma(x + \frac{1}{\sigma})}{\Gamma(\frac{1}{\sigma})\Gamma(x + 1)} \left(\frac{\sigma\mu}{1 + \sigma\mu} \right)^x \left(\frac{1}{1 + \sigma\mu} \right)^{\frac{1}{\sigma}} \quad (3.11)$$

mit

$$c = \frac{(1 - \nu)}{1 - \left(\frac{1}{1 + \sigma\mu} \right)^{\frac{1}{\sigma}}}$$

Die ZANBI ist also auch eine Verteilung mit drei Parametern. Alle drei Parameter werden im autoregressiven GAMLSS-Modell in Abhängigkeit der zurückliegenden Werte der Zeitreihe modelliert.

$$g_\mu(\mu_t) = \sum_{j=1}^{J^\mu} \beta_j^\mu * y_{t-l_j^\mu} \quad (3.12)$$

$$g_\sigma(\sigma_t) = \sum_{j=1}^{J^\sigma} \beta_j^\sigma * y_{t-l_j^\sigma} \quad (3.13)$$

$$g_\nu(\nu_t) = \sum_{j=1}^{J^\nu} \beta_j^\nu * y_{t-l_j^\nu} \quad (3.14)$$

Für die Linkfunktion werden erneut die Standardeinstellungen von GAMLSS verwendet, $g_\mu(x) := \log(x)$, $g_\sigma(x) := \log(x)$, $g_\nu(x) := \text{logit}(x)$. Die verwendeten zurückliegenden Beobachtungen, $\{l_j^\mu : j = 1, 2, \dots, J^\mu\}$, $\{l_j^\sigma : j = 1, 2, \dots, J^\sigma\}$ und $\{l_j^\nu : j = 1, 2, \dots, J^\nu\}$ werden in Kapitel 6 für jedes der Modelle angegeben.

Kapitel 4

Methoden zur Evaluierung der Güte von probabilistischen Vorhersagen

In Kapitel 3, Verfahren, wurden zahlreiche Methoden zur Modellierung von Zählzeitenreihen vorgestellt. Um sich für eines der Modelle zu entscheiden, ist es notwendig die Vorhersagegüte der Modelle zu bewerten und entsprechend dieser Bewertung eine Rangfolge der Modelle zu erstellen. Im Fall von Punktvorhersagen ist dies eine einfache, da etablierte und gut bekannte Aufgabe. Bei Punktvorhersagen ist nur die Abweichung der Vorhersage von der Beobachtung von Interesse. Als Metrik wird üblicherweise die Mittlere quadratische Abweichung (MQA) verwendet. Das Modell, welches die geringsten MQA liefert, wird dabei als bestes Modell eingestuft. Im Fall von probabilistischen Vorhersagen ist die Bewertung von Modellen weniger gut etabliert und verlangt komplexere Methoden als im Fall der Punktvorhersage. In dieser Arbeit sollen drei Methoden zur Bewertung von probabilistischen Vorhersagen verwendet werden. Zum einen wird die probabilistische und marginale Kalibrierung, sowie der Interquantilsabstand verwendet, dem Paradigma „maximizing the sharpness of the predictive distributions subject to calibration“ [12] folgend. Weiter kommen Rootgramme und die Pinball-Loss-Funktion zum Einsatz.

4.1 Kalibrierung und Spitze der vorhergesagten Verteilung

Gneiting et al. formulieren das Paradigma „maximizing the sharpness of the predictive distributions subject to calibration“ [12]. Kalibrierung bezieht sich

dabei auf die Kompatibilität der vorhergesagten Verteilung und der Verteilung des datengenerierenden Prozesses. Gneiting et al. formulieren drei Arten der Kalibrierung: Probabilistische Kalibrierung, Überschreitungskalibrierung und marginale Kalibrierung, um die Kompatibilität zwischen den vorhergesagten Verteilungen und der Verteilung des datengenerierenden Prozesses zu überprüfen. Diese Betrachtungen sind jedoch asymptotisch und somit in der Praxis schwer zu überprüfen. Zur Überprüfung der Kalibrierung im konkreten Fall werden empirische Varianten der probabilistischen und marginalen Kalibrierung eingeführt. Zur Überprüfung der Überschreitungskalibrierung ist keine empirische Methode bekannt. Die empirischen Varianten der probabilistischen und marginalen Kalibrierung werden in den folgenden zwei Abschnitten näher erläutert.

Die Maximierung der Spitze der vorhergesagten Verteilung bedeutet, dass unter den Modellen für die Kalibrierung festgestellt wird, dass das Modell, welches die spitzesten vorhersage Verteilungen liefert, als bestes Modell betrachtet werden soll. Als spitze Verteilung wird dabei eine Verteilung mit möglichst geringer Streuung bzw. möglichst kleinen Quantilsdifferenzen betrachtet. Die Maximierung der Spitze bezieht sich also nur auf die Vorhersagen und lässt die Beobachtungen außen vor. Methoden zur Bewertung der Spitze der Vorhersagen werden im dritten Abschnitt vorgestellt.

4.1.1 probabilistische Kalibrierung

Die Überprüfung der probabilistischen Kalibrierung beruht auf dem Probability Integral Transform (PIT). Sei $y_t : t \in \mathbb{N}$ eine Zeitreihe und $F_t(x)$ die vorhergesagte Verteilungsfunktion zum Zeitpunkt t . Definiere nun $p_t := F_t(y_t)$ als PIT-Wert zum Zeitpunkt t . Eine probabilistische Kalibrierung liegt vor, wenn die p_t gleichverteilt sind auf $[0, 1]$. Die Überprüfung der Gleichverteilung der PIT-Werte wird nicht nur von Gneiting et al. [12] vorgeschlagen, sondern dient auch anderen Autoren, um probabilistische Vorhersagen zu evaluieren [8]. Gneiting et al. [12] schlagen vor, die Gleichverteilung der PIT-Werte mit einem Histogramm zu überprüfen, die Gleichverteilung kann jedoch auch mit formalen Hypothesen überprüft werden [5]. In dieser Arbeit werden im Weiteren die Histogramme der PIT-Werte verwendet, da die Struktur des Histogramms bei Abweichung von der Gleichverteilung Hinweise auf den Grund der Abweichung geben kann. Ein U-förmiges Histogramm weist auf Vorhersageverteilungen mit Überdispersion hin. Ein umgekehrt U-förmiges Histogramm deutet auf Vorhersageverteilungen mit Unterdispersion hin [12].

Die Invertierung der Vorhersageverteilung ist nur im Fall stetiger Zufallsvariablen möglich. Damit beschränkt sich die oben angeführte Berechnung der

PIT-Werte auf den Fall, dass die y_t aus einer stetigen Verteilung stammt. Da sich diese Arbeit der probabilistischen Vorhersage von Zählzeitenreihen widmet, stammen die y_t offensichtlich aus einer diskreten Verteilung. Im Falle des ARIMA-Modells wird jedoch implizit angenommen, dass die Zeitreihe normalverteilt und somit stetig ist. Zur Evaluierung dieser Modelle wird im Weiteren die oben erläuterte Herangehensweise verwendet, um ein PIT-Histogramm zu erstellen.

Für alle anderen Modelle wird eine diskrete Verteilung angenommen. Zur Evaluierung probabilistischer Vorhersagen über diskrete Zufallsvariablen schlagen Czado et al. das "non-randomized PIT" vor [6]. Dabei wird, wie bei diskreten Verteilungen, ein Histogramm erstellt, um auf Gleichverteilung zu überprüfen. Die Berechnung der Histogramme ist allerdings komplizierter als im diskreten Fall.

Sei $F_t(y_t)$ die vorhergesagte Verteilung zum Zeitpunkt t ausgewertet an der Beobachtung zum Zeitpunkt t . Betrachte nun folgende Funktion:

$$\bar{P}(u) = \frac{1}{n} \sum_{i=1}^T P^{(i)}(u)$$

Wobei die Funktion $P^{(t)}(u)$ folgendermaßen definiert ist.

Für den Fall $y_t > 0$:

$$P^{(t)}(u) = \begin{cases} 0 & \text{für } u \leq F_t(y_t - 1) \\ \frac{u - F_t(y_t - 1)}{F_t(y_t) - F_t(y_t - 1)} & \text{für } F_t(y_t - 1) \leq u \leq F_t(y_t) \\ 1 & \text{für } F_t(y_t) \leq u \end{cases}$$

Für den Fall $y_t = 0$:

$$P^{(t)}(u) = \begin{cases} \frac{u}{F_t(0)} & \text{für } u \leq F_t(0) \\ 1 & \text{für } u \geq F_t(0) \end{cases}$$

Ein PIT-Histogramm kann nun folgendermaßen erstellt werden. Wähle eine Anzahl Bins J , meist 10 oder 20. Berechne $p_j = \bar{P}(\frac{j}{J}) - \bar{P}(\frac{j-1}{J})$ als Höhe für Bin $j = 1, 2, \dots, J$.

Das resultierende Histogramm kann genauso interpretiert werden wie im diskreten Fall. Gleichverteilung zeigt probabilistische Kalibrierung an, ein U-förmiges Histogramm weist auf Vorhersageverteilungen mit Überdispersion hin, ein umgekehrt U-förmig Histogramm deutet auf Vorhersageverteilungen mit Unterdispersion hin.

4.1.2 marginale Kalibrierung

Das Konzept der marginalen Kalibrierung beruht auf folgender Idee. Wenn die Beobachtung y_t aus der vorhergesagten Verteilungsfunktion $F_t(x)$ stammt, dann muss die Mischverteilung, die sich über alle $t = 1, 2, \dots, T$ ergibt, kompatibel mit der empirischen Verteilung der Beobachtungen sein. Die Mischverteilung der Vorhersagen ergibt sich aus:

$$\bar{F}_T(x) = \frac{1}{T} \sum_{t=1}^T F_t(x)$$

,die empirische Verteilung der Beobachtungen ist

$$\hat{G}_T(x) = \frac{1}{T} \sum_{t=1}^T I(y_t \leq x)$$

Zur Beurteilung der Übereinstimmung der zwei Verteilungen wird der marginale Kalibrierungsplot verwendet. Dazu wird $\bar{F}_T(x) - \hat{G}_T(x)$ gegen die Beobachtungen geplottet. Eine möglichst geringe Abweichung von Null spricht dabei für die marginale Kalibrierung.

4.1.3 Spitze der vorhergesagten Verteilung

Ziel dieser Arbeit ist es, verschiedene Modelle, hinsichtlich ihrer Eigenschaft probabilistische Vorhersagen zu treffen, zu bewerten, und das am besten geeignete Modell zu identifizieren. Im Paradigma „maximizing the sharpness of the predictive distributions subject to calibration“ wird das Modell, das die spitzeften Vorhersageverteilungen liefert und gleichzeitig die probabilistische und marginale Kalibrierung erfüllt, als das beste Modell erachtet.

Die Bewertung der Spitze einer Verteilung kann auf verschiedene Weisen erfolgen. Die folgende Bewertung orientiert sich an dem Vorschlag von Gneiting et al. [12]. Als Maß für die Spitze werden die Boxplots der zentralen 50%- und 90%-Vorhersageintervalle verwendet. Es sei $t = 1, 2, \dots, T$ die Zeit und $Q_t^p(x)$ die $p\%$ Quantilsfunktion der vorhergesagten Verteilung zum Zeitpunkt t . Betrachte nun die Quantilsdifferenzen:

$$D_{0.5}(y_t) = \frac{1}{T} \sum_{t=1}^T Q_t^{0.75}(x_t) - Q_t^{0.25}(x_t)$$

$$D_{0.9}(y_t) = \frac{1}{T} \sum_{t=1}^T Q_t^{0.95}(x_t) - Q_t^{0.05}(x_t)$$

Die Quantilsdifferenzen $D_{0.5}(y_t)$ und $D_{0.9}(y_t)$ für $t = 1, 2, \dots, T$ können nun in Boxplots visualisiert werden. Das Modell mit den geringsten Quantilsdifferenzen ist das Modell mit den spitzesten Vorhersageverteilungen und wird, Kalibrierung vorausgesetzt, als bestes Modell betrachtet. Die Visualisierung in Boxplots bietet den Vorteil, dass ersichtlich ist, ob die vorhergesagten Verteilungen zu allen Zeitpunkten etwa gleich breit sind, oder ob die Breite der vorhergesagten Verteilungen schwankt.

4.2 Rootogramme

Kleiber et al. schlagen sogenannte Rootogramme zur Evaluierung von probabilistischen Zählmodellen vor [15]. Die Idee hinter den Rootogrammen ist ähnlich, wie die hinter den marginalen Kalibrierungsplots, jedoch wird eine andere Visualisierung verwendet. Im Gegensatz zu den marginalen Kalibrierungsplots, bei denen die erwartete und die beobachtete Verteilungsfunktion betrachtet werden, werden bei den Rootogrammen die erwartete und die beobachtete Dichte betrachtet.

Sei $\{y_t : t \in \mathbb{N}\}$ eine Zeitreihe und $f_t(x)$ die Dichtefunktion der vorhergesagten Verteilung. Betrachte nun für alle Ausprägungen der Zeitreihe j , die erwartete Dichte:

$$exp_j := \sum_{t=1}^T f_t(j)$$

die beobachtete Dichte:

$$obs_j := \sum_{t=1}^T I(y_t = j)$$

Die erwartete Dichte und die beobachtete Dichte werden gemeinsam gegen die Menge an natürlichen Zahlen, die die Zeitreihe annimmt (j), abgetragen. Wobei die beobachteten Werte als Balkendiagramm dargestellt werden und die erwarteten Werte mit einer Linie verbunden werden. Ein Beispiel für ein Rootogramm findet sich im nächsten Kapitel. Kleiber et al. [15] präsentieren drei leicht unterschiedliche Arten von Rootogrammen, in dieser Arbeit werden jedoch nur die sogenannten „stehenden Rootogramme“ verwendet.

4.3 Pinball-Loss

Die Pinball-Loss-Funktion ist ein von der oben vorgestellten Methode unabhängiges, nicht parametrisches Verfahren um die probabilistischen Vorhersagen verschiedener Modelle zu vergleichen. Die Pinball-Loss-Funktion für ein

Quantile p , die Beobachtung x_t und das p Quantil der Vorhersageverteilung $Q_t^p(x_t)$ zum Zeitpunkt t ergibt sich als:

$$L_p(x_t, Q_t^p(x_t)) = \begin{cases} (x_t - Q_t^p(x_t))p & \text{für } x_t \geq Q_t^p(x_t) \\ (Q_t^p(x_t) - x_t)(1 - p) & \text{für } x_t \leq Q_t^p(x_t) \end{cases}$$

Für eine Zeitreihe bietet es sich an, die durchschnittliche Pinball-Loss-Funktion für ein Quantil über alle Beobachtungen zu betrachten. Um möglichst die ganze Verteilung abzudecken, können verschiedenen Quantile betrachtet werden. Im Folgenden werden die 0.05, 0.10, 0.25, 0.5, 0.75, 0.90, 0.95 Quantile betrachtet. Der Pinball-Loss Score wird hier folgendermaßen definiert: sei $P = \{0.05, 0.10, 0.25, 0.5, 0.75, 0.90, 0.95\}$

$$S_{pin} = \frac{1}{|P|} \sum_{p \in P} \frac{1}{T} \sum_{t=1}^T L_p(x_t, Q_t^p(x_t))$$

4.4 Mittlere quadratische Abweichung

Diese Arbeit konzentriert sich auf probabilistische Vorhersagemodelle. Der Vollständigkeit wegen soll für alle Vorhersagemodelle auch die Qualität der Punktvorhersagen bewertet werden. Dazu bietet sich die Mittlere Quadratische Abweichung (MQA) an, die zur Evaluierung von numerischen Vorhersagen üblicherweise zur Anwendung kommt.

Sei x_t mit $t = 1, 2, \dots, T$ eine Zeitreihe, $F_t(x)$ die Vorhersageverteilung zum Zeitpunkt t und $Y \sim F_t$. Die MQA wird berechnet als:

$$\frac{1}{T} \sum_{t=1}^T (E(Y) - x_t)^2$$

Kapitel 5

Beispielhafte Anwendung der Methoden zur Evaluierung der Güte von probabilistischen Vorhersagen

Dieses Kapitel soll dazu dienen, die im vorigen Kapitel vorgestellten Methoden zur Evaluierung probabilistischer Vorhersagen, insbesondere der probabilistischen und marginalen Kalibrierung so wie der Rootogramme an einem Beispiel zu illustrieren.

Dazu wurde eine Zeitreihe bekannter Struktur simuliert. Für diese Zeitreihe wurde dann auf fünf verschiedene Weisen probabilistische Vorhersagen erstellt, um die Wirkung verschiedener Vorhersagen auf die Methoden zur Evaluierung zu veranschaulichen.

5.1 simulierte Zeitreihe

Zur Simulation der Zeitreihe wurde auf das R-Paket `Tscount` [16] zurückgegriffen. Es wurde eine negativ binomialverteilte Zeitreihe erstellt. Der Parameter μ der Verteilung wurde als abhängig von der letzten Beobachtung und abhängig vom Wert des Parameters μ zu den Zeitpunkten t_1 und t_2 erzeugt. Dabei wurde die Identitätsfunktion als Linkfunktion verwendet. Der Überdispersionsparameter wurde auf konstant 5 gesetzt. Die Parametrisierung der negativen Binomialverteilung entspricht der üblichen in R verwendeten Parametrisierung.

5.2 Modelle

Die simulierte Zeitreihe wurde auf fünf verschiedene Weisen probabilistisch vorhergesagt:

- **Perfektes Modell:** Die probabilistische Vorhersage entspricht dem Modell, aus dem die simulierten Daten gezogen wurden. Dieses Modell dient hier nur zur Referenz, es ist in der tatsächlichen Anwendung natürlich unbekannt.
- **Optimal geschätztes Modell:** Die Parameter der negativ Binomialverteilung wurden aus den simulierten Daten geschätzt.
- **Modell mit Unterdispersion:** Die simulierte Zeitreihe wurde als Poisson verteilt angenommen und der Parameter der Poisson-Verteilung wurden für jeden Zeitpunkt aus den Daten geschätzt. Das führt dazu, dass die vorhergesagten Verteilungen kleinere Varianzen aufweisen, als die Verteilungen aus der die simulierten Daten gezogen wurden.
- **Modell mit Überdispersion:** Die simulierte Zeitreihe wurde als negativ binomialverteilt angenommen. Der μ Parameter wurde aus den Daten geschätzt, der size-Parameter wurde konstant auf 1 gesetzt. Das führt dazu, dass die vorhergesagten Verteilungen größere Varianzen aufweisen, als die Verteilung aus der die simulierten Daten gezogen wurden.
- **Modell Bias:** Die simulierte Zeitreihe wurde als negativ binomialverteilt angenommen. Der aus den Daten geschätzte Parameter μ wurde konstant um drei erhöht, der size-Parameter wurde konstant auf fünf gesetzt. Dies führt zu einer insgesamt verzerrten probabilistischen Vorhersage, bei der weder die Erwartungswerte noch die Varianzen der Verteilung der simulierten Daten entsprechen.

5.3 Ergebnis

Wie zu erwarten, ist das optimal geschätzte Modell das beste Modell. Es ist kalibriert, sowohl bei der probabilistischen Kalibrierung, als auch bei der marginalen Kalibrierung ist es kaum vom perfekten Modell zu unterscheiden. Da für kein anderes Modell Kalibrierung festgestellt wurde, liefert der Sharpnes-Plot hier keine zusätzliche Information zur Modellwahl. Das Rootogramm stellt die gleiche Information in etwas anderer Weise dar. Das optimal geschätzte Modell hat die beste Anpassung zwischen beobachteter Dichte

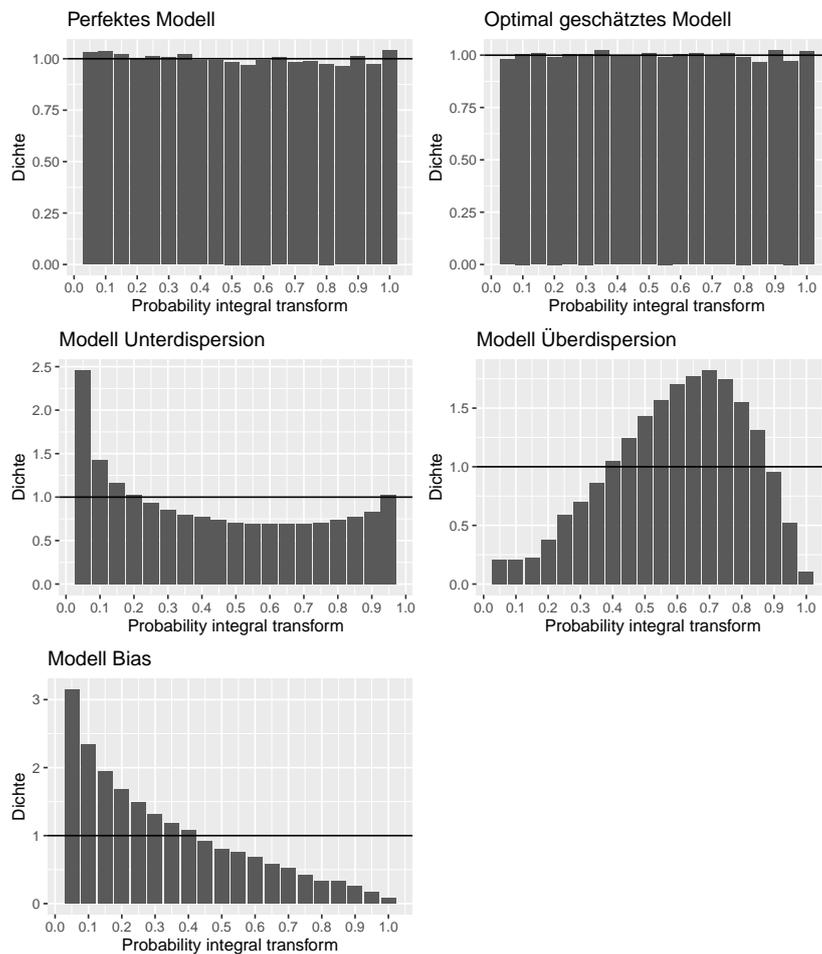


Abbildung 5.1: Simulation PIT Histogramme

und geschätzter Dichte. Somit lässt es sich klar als das beste Modell identifizieren. Für die anderen Modelle lassen sich deren Eigenschaften gut an deren Rootogramm ablesen. Beim Modell Unterdispersion ist die geschätzte Dichte spitzer als die beobachtete Dichte, beim Modell Überdispersion ist die geschätzte Dichte breiter als die beobachtete Dichte, beim Modell Bias stimmen geschätzte Dichte und beobachtete Dichte weder im Erwartungswert noch in der Varianz.

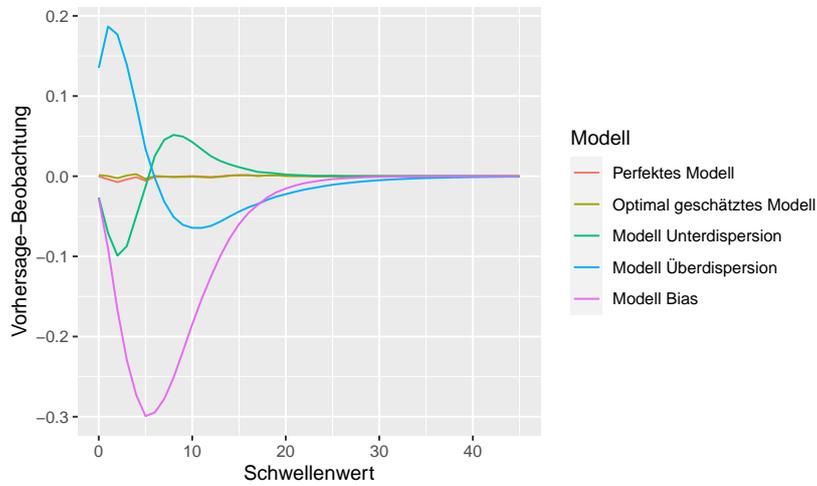


Abbildung 5.2: Simulation Marginaler-Kalibrierungs-Plot

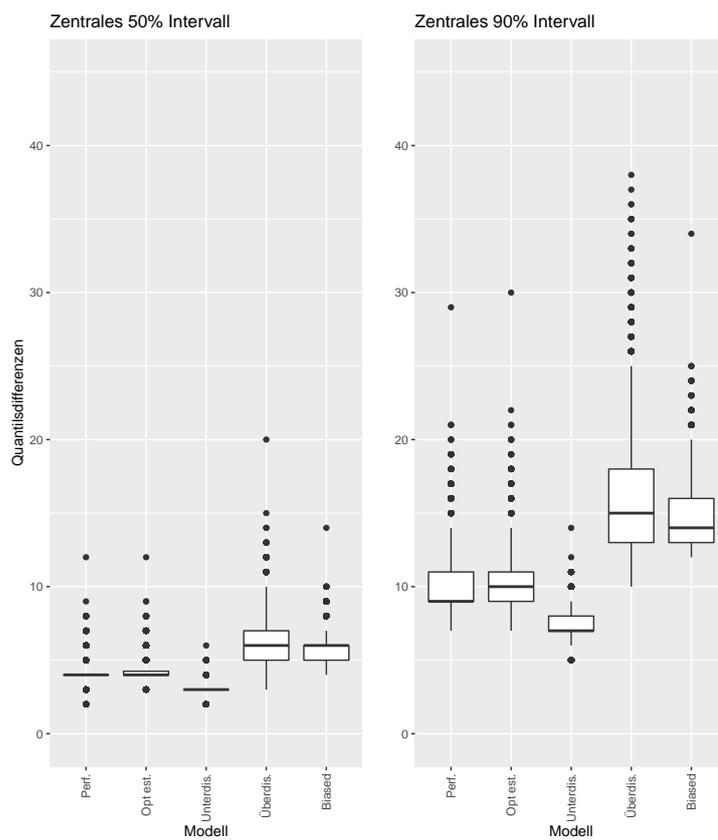


Abbildung 5.3: Simulation Sharpnes Plot

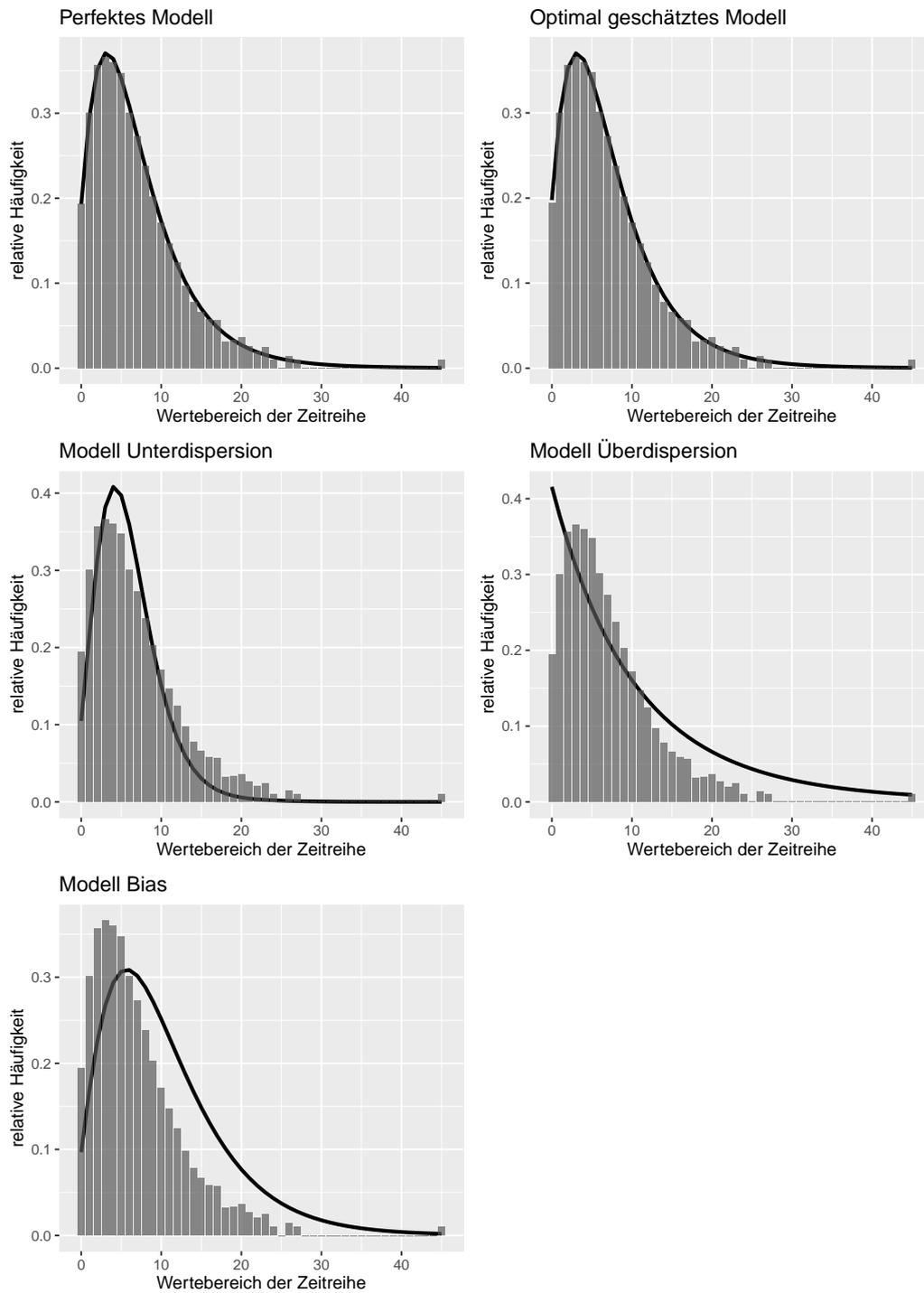


Abbildung 5.4: Simulation Rootogramm

Kapitel 6

Anwendung der Verfahren auf reale Daten

In diesem Kapitel sollen die vorgestellten Methoden zur Erstellung von probabilistischen Vorhersagen auf reale Zählzeitenreihen angewendet werden und die Qualität der Vorhersagen anhand der vorgestellten Methoden zur Evaluierung von probabilistischen Vorhersagen bewertet werden.

6.1 Daten

Es werden Zeitreihen aus der „M5 Accuracy competition“ verwendet [17]. Der M5-Datensatz enthält die Verkäufe für 30490 Produkte, die in verschiedenen Supermärkten an 1919 aufeinanderfolgenden Tagen verkauft wurden. Aus diesen 30490 Zeitreihen wurden 4 Zeitreihen ausgewählt, auf welche dann die oben vorgestellten Zeitreihen-Methoden und die Methoden zur Evaluierung der probabilistischen Vorhersagen angewendet wurden. Die vier Zeitreihen werden im Folgenden als Zeitreihe 1-4 bezeichnet. Dabei entspricht jede Zeitreihe der Anzahl der verkauften Stücke eines bestimmten Produkts in einem Supermarkt an 1919 aufeinanderfolgenden Tagen. In Abbildung 6.1 sind für jede der vier Zeitreihen Histogramme für die gesamten 1919 Tage und partielle Korrelationsplots abgebildet.

6.2 Modellfindung

Wie in Abbildung 6.1 zu sehen ist, weisen alle Zeitreihen große Autokorrelation über sehr viele zurückliegende Beobachtungen auf. Dies liegt wohl auch daran, dass die Zeitreihen nicht stationär sind. Aufgrund dessen ist es nicht offensichtlich, wie viele zurückliegende Beobachtungen in ein Modell

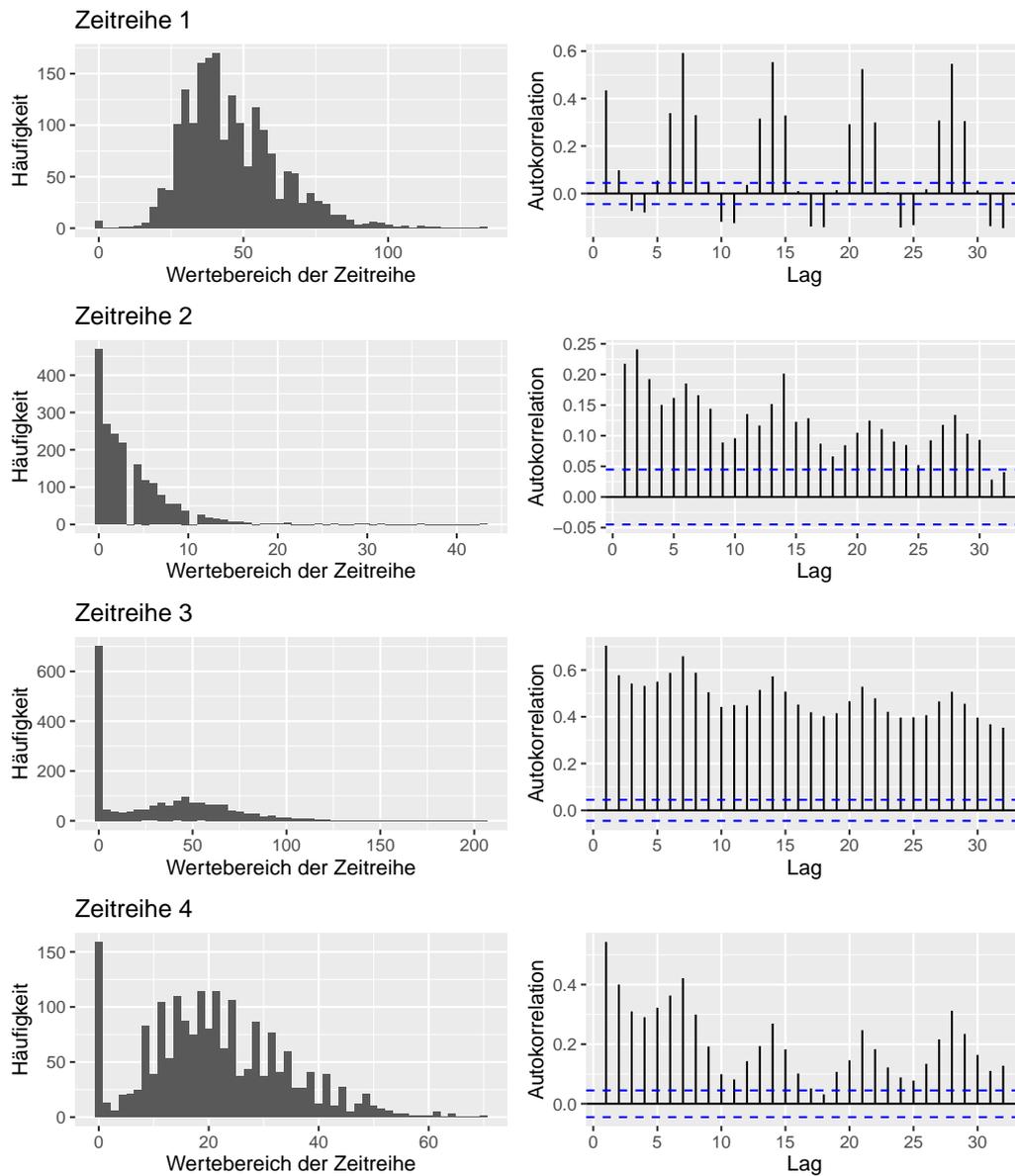


Abbildung 6.1: Histogramm und Autokorrelation für die vier Zeitreihen

aufgenommen werden sollen. Einerseits, um konsistent über alle Modelle vorzugehen, andererseits, um die Modellfindung innerhalb einer Verteilung nicht zu kompliziert zu gestalten, wird wie folgt verfahren. Für alle Modelle werden maximal 12 zurückliegende Beobachtungen verwendet, dies entspricht in diesem Fall einem zweiwöchigen Zeitraum. Zunächst wird ein Modell mit allen 12 zurückliegenden Beobachtungen geschätzt, anschließend wird je ein Modell mit sukzessiv weniger Beobachtungen geschätzt, wobei dabei jeweils die am weitesten entfernte Beobachtung weggelassen wird. Das beste Modell, und somit die beste Menge an zurückliegenden Beobachtungen, wird dann mithilfe des Akaike-Informationskriterium bestimmt. Dieses Vorgehen wird für die Tscout- und GAMLSS-Modelle verwendet. Für ARIMA-Modelle steht mit der Funktion „auto.arima“ aus dem forecast [14] R-Paket eine effiziente und etablierte Methode zur Verfügung, um die optimalen Lags in einem Modell zu ermitteln. Die ARIMA-Modelle in dieser Arbeit werden dementsprechend auch mithilfe der „auto.arima“ Funktion geschätzt.

6.3 Ergebnisse

Die in dieser Arbeit vorgestellten Modelle sollen anhand ihrer Vorhersagegüte bewertet werden. Dazu ist es notwendig, die Modelle auf Daten zu evaluieren, welche nicht zum Schätzen der Modelle verwendet wurden. In klassischen prediktiven Klassifikations- und Regressionsmodellen wird diese Problematik durch die Verwendung einer Kreuzvalidierung gelöst. Für Zeitreihen gestaltet sich die Kreuzvalidierung jedoch schwieriger, da die Vorhersagen abhängig sind von zurückliegenden Fehlern (ARIMA) oder von zurückliegenden Vorhersagen (Tscout). Für das autoregressive GAMLSS besteht keine solche Abhängigkeit und es könnte eine gewöhnliche Kreuzvalidierung verwendet werden. Eine einfache Alternative zur Kreuzvalidierung, die auch für Zeitreihendaten funktioniert, ist, die Daten in einen Trainings- und einen Testdatensatz zu teilen und dann die Modelle auf den Trainingsdaten zu schätzen und auf den Testdaten zu validieren. Für die hier verwendeten Zeitreihen ergeben sich damit allerdings zwei Schwierigkeiten. Einerseits sind die verwendeten Verfahren zur Evaluierung teilweise asymptotisch begründet und bedürfen damit einer gewissen Menge an Testdaten um zuverlässig interpretierbar zu sein. Andererseits sollen in der Schätzung ebenfalls möglichst viele Beobachtungen verwendet werden. Um beiden Anforderungen gerecht zu werden, wird deshalb eine Zeitreihen-Kreuzvalidierung "Time Series Cross-Validation" [13, Kapitel 3.4] zur Anwendung kommen. Das Vorgehen hierzu gestaltet sich wie folgt: Die Zeitreihe wird auch hier in Trainings- und Testbeobachtungen aufgeteilt. Wobei als Testbeobachtung hier nur die letzte Beobachtung

der Zeitreihe verwendet wird und als Trainingsbeobachtungen alle Übrigen. Anschließend wird die letzte Beobachtung aus der Zeitreihe entfernt und die Aufteilung in Trainings- und Testbeobachtungen wird wiederholt. Insgesamt wird dieser Prozess 900-mal wiederholt. Dadurch ergeben sich 900 Verteilungsvorhersagen für jede Zeitreihe und jedes Modell, die dann mithilfe der Methoden zur Evaluierung bewertet werden. Die Ergebnisse der Evaluierung für alle vier Zeitreihen finden sich in den folgenden Abschnitten.

Die optimale Menge der zurückliegenden Beobachtungen (Lags) wird, wie in Kapitel 6.1 beschrieben, bestimmt. Für die Tscout und die GAMLSS Modelle werden die Lags einmal für alle 900 Kreuzvalidierungsdurchgänge (CV-Durchgänge) bestimmt. Für die ARIMA Modelle wird das Modell in jedem der 900 CV-Durchgänge mit der „auto.arima“ Funktion bestimmt. Somit ergeben sich für die ARIMA Modelle für jeden CV-Durchgang unterschiedliche Lags. Bei den Tscout und GAMLSS Modellen werden die Lags nur einmal zu Beginn der Kreuzvalidierung bestimmt und dann für alle Durchgänge einheitlich verwendet. Für die Tscout und GAMLSS Modelle konnten die Lags nicht für jeden CV-Durchgang flexibel bestimmt werden, da die Ermittlung der optimalen Lags für diese Modelle sehr rechenintensiv ist.

6.3.1 Zeitreihe 1

Für Zeitreihe 1 sind im ARIMA-Modell, für die unterschiedlichen CV-Durchgänge, entweder drei AR-Terme und keine MA-Terme, oder fünf AR-Terme und 1 MA-Term enthalten. Das Tscout Modell, Poisson Modell und Negativ Binomial Modell (NBI-Modell) enthält jeweils 10 zurückliegende Beobachtungen und 10 zurückliegende Erwartungswerte. GAMLSS-, ZINBI- und ZANBI-Modelle enthalten 10 zurückliegende Beobachtungen.

- **Probabilistische-Kalibrierung:** Wie in Abbildung 6.2 gut zu erkennen ist, weisen die vorhergesagten Verteilungen des Poisson-Modells Unterdispersion auf. Sowohl ZINBI- als auch ZANBI-Modelle weichen bei der Zeitreihe 1 stark von der Normalverteilung ab. Warum das so ist, wird in Abbildung 6.6 deutlich. Das ARIMA- und das NBI-Modell können am ehesten als gleichverteilt betrachtet werden. Somit wird für diese Modelle von probabilistischer Kalibrierung ausgegangen.
- **Marginale-Kalibrierung:** In Abbildung 6.3 sind die Differenzen aus vorhergesagter und beobachteter Verteilungsfunktion für alle fünf Modelle für verschiedene Schwellenwerte der Zeitreihe abgetragen. ZINBI- und ZANBI-Modelle weisen große Differenzen zwischen vorhergesagter

und beobachteter Verteilungsfunktion auf. Das ARIMA-Modell weist größere Differenzen auf, als das NBI-Modell. Das Poisson-Modell liegt zwischen ARIMA- und NBI-Modell. Für das NBI-Modell wird somit von marginaler Kalibrierung ausgegangen.

- **Spitze:** In Abbildung 6.4 sind die Quantilsdifferenzen der vorhergesagten Verteilungen der fünf Modelle als Boxplots abgetragen. Hier soll die spitzeste Verteilung unter den Modellen bei denen Kalibrierung vorliegt ermittelt werden. Für das NBI-Modell liegt probabilistische und marginale Kalibrierung vor, für das ARIMA-Modell nur probabilistische Kalibrierung. Die Quantilsdifferenzen für das NBI-Modell sind deutlich kleiner als die des ARIMA-Modells. Somit ist das NBI-Modell, unter Verwendung der Kalibrierung und Spitze, das beste Modell.
- **Rootogramme:** In Abbildung 6.5 sind die Rootogramme für die fünf Modelle abgetragen. Das Balkendiagramm nähert dabei die Dichte der Mischverteilung der Daten an. Die Kurve gibt die Dichte der Mischverteilung der vorhergesagten Verteilungen an. Für das ARIMA-Modell passen sowohl das Histogramm, als auch die Dichte offensichtlich nicht zusammen. Für das Poisson-Modell liegt Unterdispersion vor, kleine und große Werte werden vom Modell unterschätzt. ZINBI- und ZANBI-Modell liefern identische Vorhersagen, bei denen jeweils die Wahrscheinlichkeit für Nullwerte deutlich überschätzt wird. Das NBI-Modell liefert die beste Anpassung an die Daten. Nur die Nullwerte werden vom Modell nicht erfasst.
- **Zeitreihe und Quantilvorhersagen:** In Abbildung 6.6 ist der Verlauf der Zeitreihe für einen zufällig ausgewählten Zeitpunkt für 30 aufeinander folgende Tage abgebildet. Für jedes der fünf Modelle sind zusätzlich die 10 %-, 50 %- und 90 %- Quantile der vorhergesagten Verteilungen abgebildet. Auffällig ist, dass ZINBI- und ZANBI-Modelle im ausgewählten Zeitraum schlecht funktionieren. Die Wahrscheinlichkeit für Nullwerte in der Mischverteilung nimmt häufig so große Werte an, dass die ganze Verteilung nur aus dem Wert Null besteht. Dies ist auch der Grund, warum das PIT Histogramm der zwei Modelle für den Wert eins so stark von der Gleichverteilung abweicht. Wenn die vorhergesagte Verteilung Null ist, dann wird die Verteilungsfunktion an jedem Punkt größer und gleich Null den Wert Eins annehmen.

Für das ARIMA-Modell liegen das 10 %- und 90 %- Quantil relativ weit auseinander. Dementsprechend ist die Zeitreihe überwiegend innerhalb des 10 %-, 90 %- Intervalls. Das Gegenteil dazu bildet das Poisson-

Modell. Hier ist die Verteilung sehr schmal, jedoch liegt die Zeitreihe auch öfter außerhalb des 10 %-, 90 %- Intervalls. Das NBI-Modell scheint die Zeitreihe in den dargestellten 30 Tagen gut abzubilden. Die 10 %-, 90 %- Intervalle sind verhältnismäßig eng, trotzdem liegt die Zeitreihe zu fast allen Zeitpunkten im Intervall.

- **Pinball-Loss / mittlere quadratische Abweichung:** In Tabelle 6.1 sind das Ergebnis der Pinball-Loss-Funktion und die mittleren quadratischen Abweichungen der fünf Modelle eingetragen. Ein geringer Pinball-Loss-Wert zeigt eine gute Anpassung der Vorhersageverteilung an die Zeitreihe. Das NBI-Modell hat den geringsten Pinball-Loss-Wert. Danach folgt das Poisson- und das ARIMA-Modell. Die größten Pinball-Loss-Werte haben das ZINBI- und ZANBI-Modell. Die mittlere, quadratische Abweichung ist für Poisson- und NBI-Modell gleich und die geringste Abweichung aller Modelle. Danach folgt das ARIMA-Modell und anschließend mit Abstand das ZINBI- und ZANBI-Modell.

Für Zeitreihe 1 ist das Ergebnis der Evaluation für alle verwendeten Methoden konsistent. Alle Methoden identifizieren das NBI-Modell als bestes Modell.

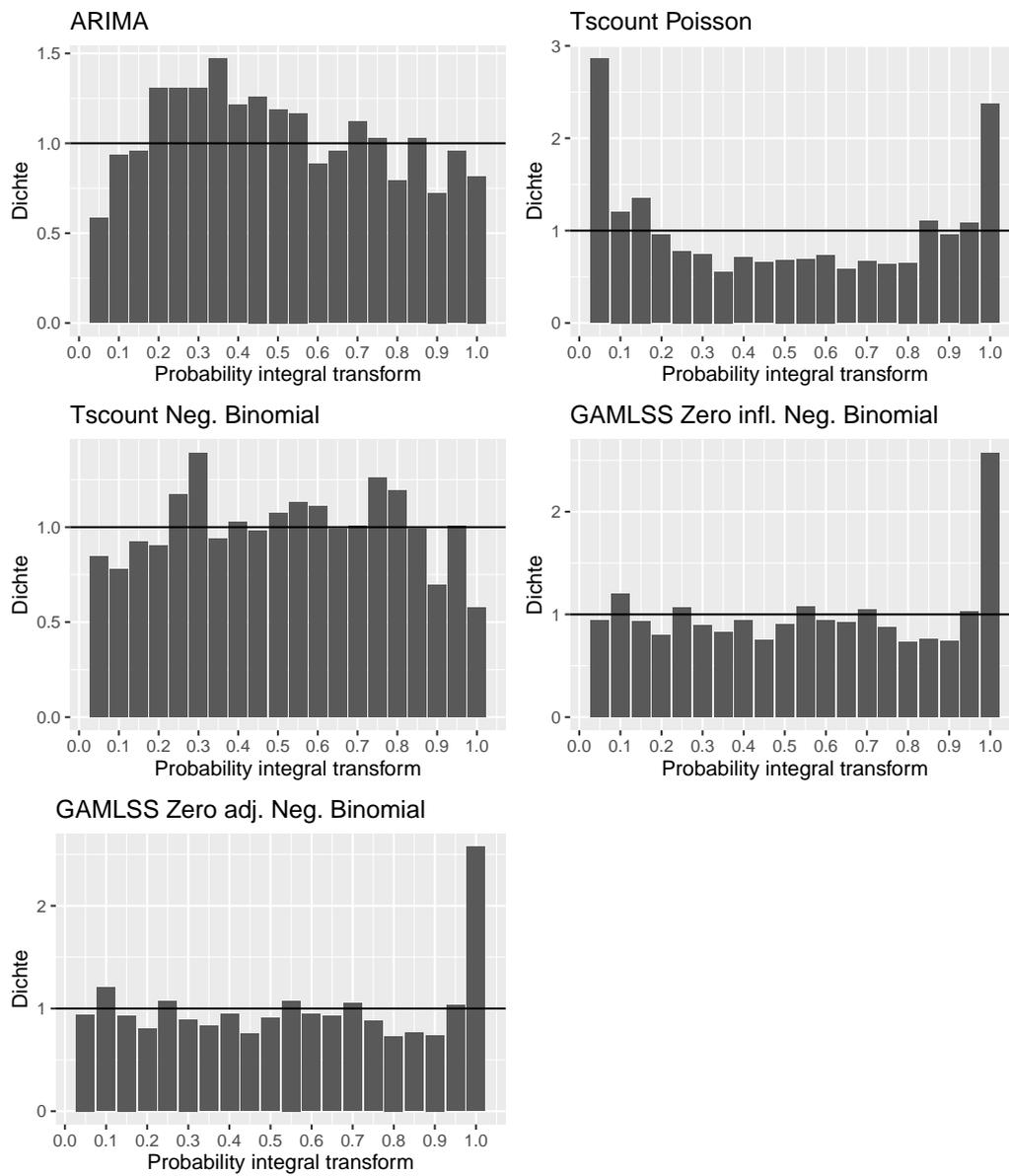


Abbildung 6.2: Zeitreihe 1 PIT Histogramme

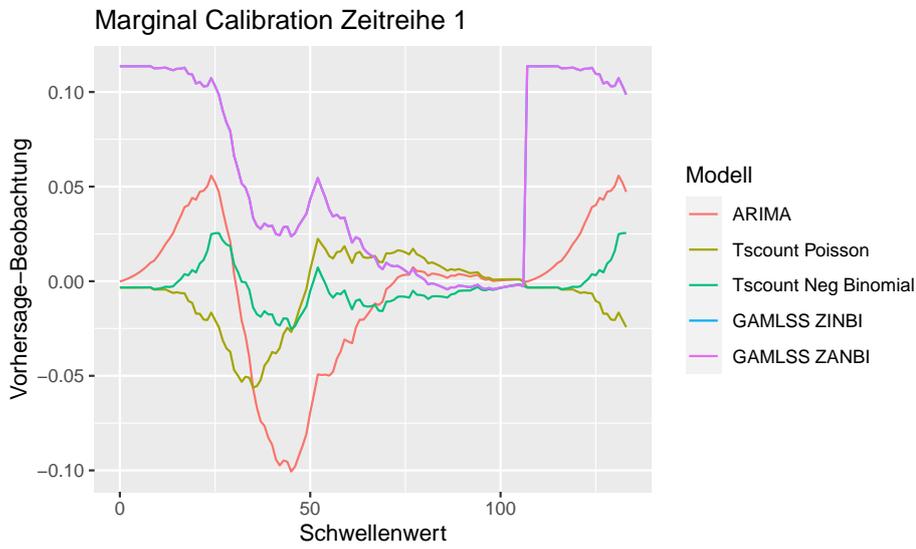


Abbildung 6.3: Zeitreihe 1 marginaler Kalibrierungs Plot

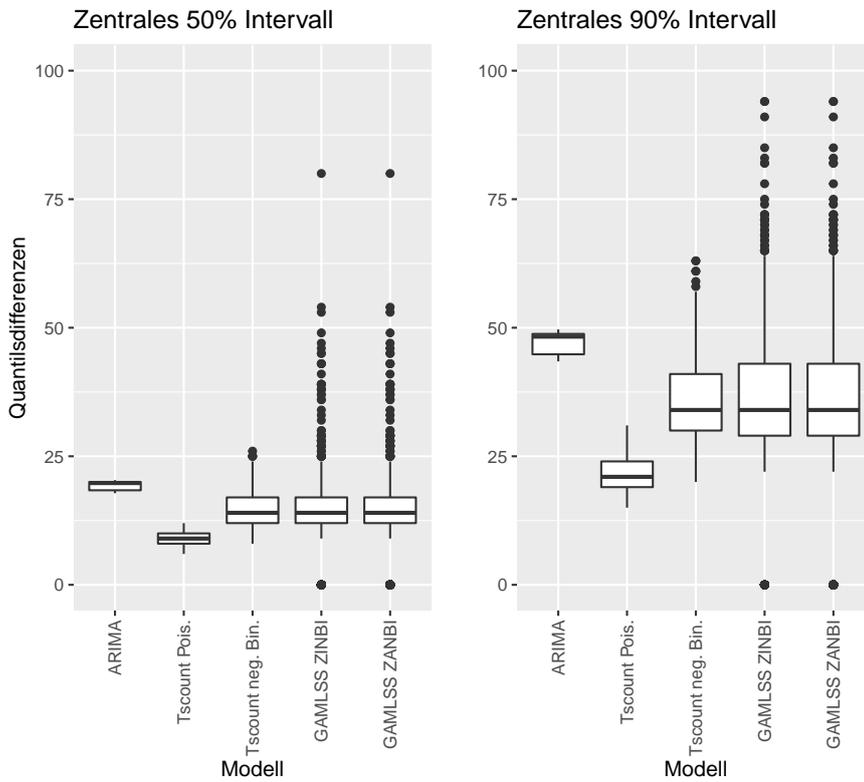


Abbildung 6.4: Zeitreihe 1 Sharpnes Plot

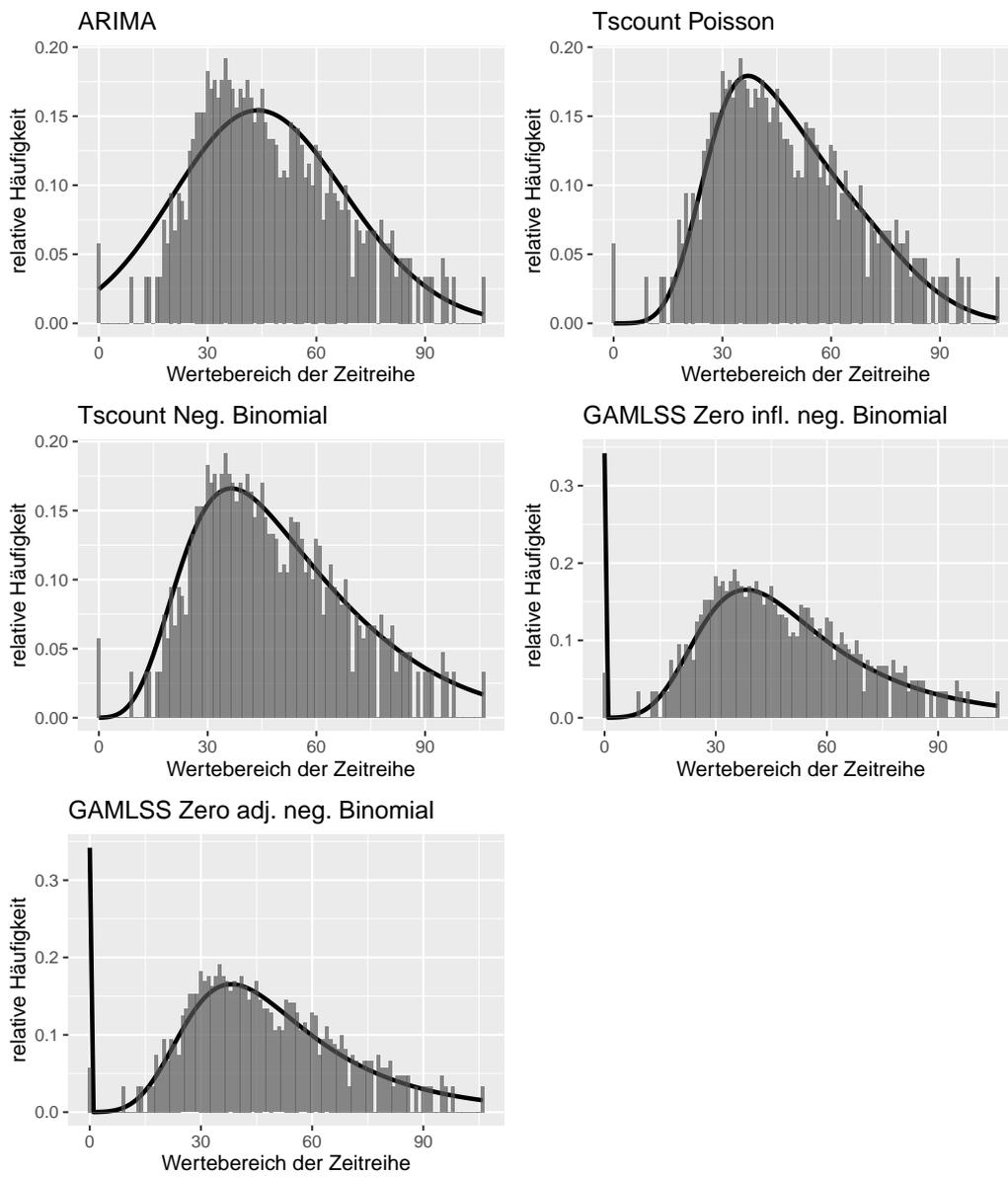


Abbildung 6.5: Zeitreihe 1 Rootogramm

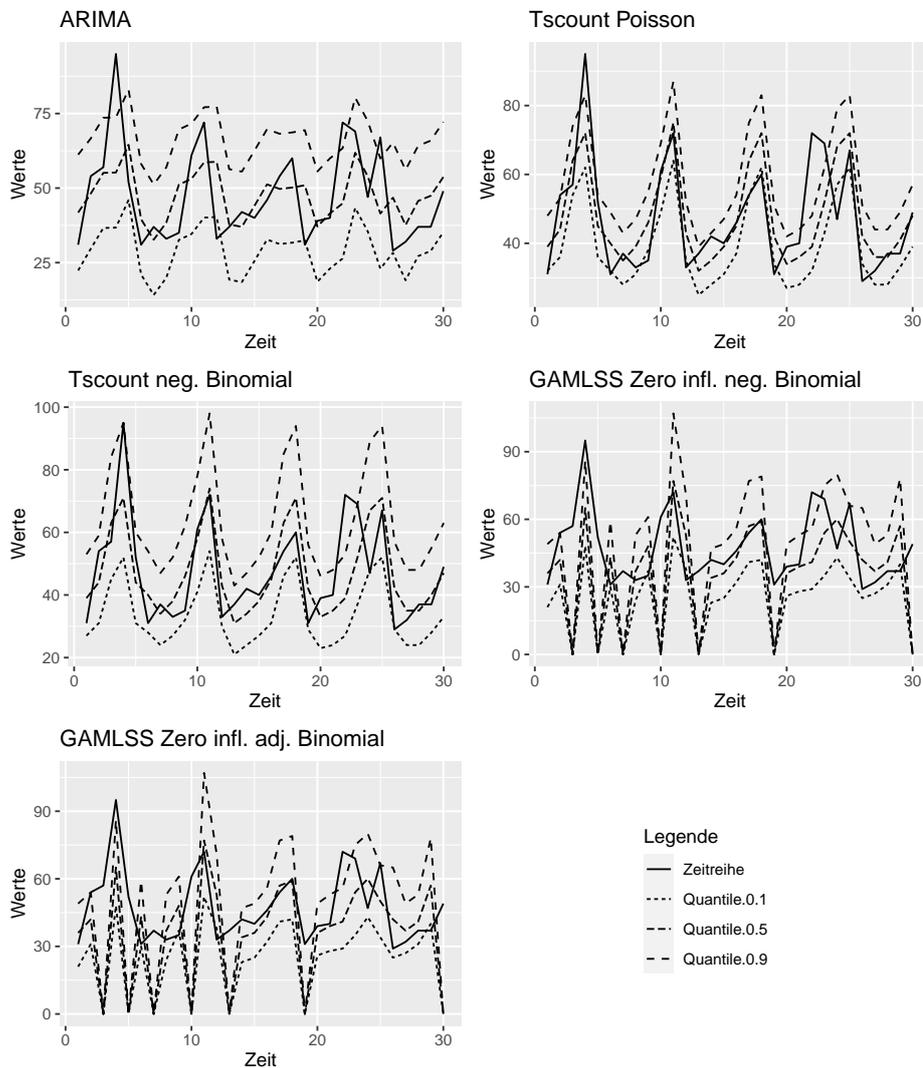


Abbildung 6.6: Werte Zeitreihe 1 und Quantile der vorhergesagten Verteilung

Modell	Pinball-Loss	MQA
ARIMA	2.99	173.94
Tscount Poisson	2.43	108.84
Tscount neg. Binomial	2.31	108.84
GAMLSS zero inf. neg. Binomial	4.58	381.43
GAMLSS zero adj. neg. Binomial	4.58	381.43

Tabelle 6.1: Zeitreihe 1, Pinball-Loss und Mittlere quadratische Abweichung

6.3.2 Zeitreihe 2

Für Zeitreihe 2 sind im ARIMA-Modell für die unterschiedlichen CV-Durchgänge, entweder ein AR-Term und keine MA-Terme, 1 MA-Term und kein AR-Term, 3 MA-Terme und kein AR-Term oder drei AR-Terme und 1 MA-Term enthalten. Das Tscout-Poisson-Modell enthält 10 Lags, das Tscout-NBI-Modell enthält 5 Lags. Das GAMLSS-, ZINBI-Modell enthält 11 Lags und das GAMLSS- ZANBI-Modell enthält 7 Lags.

- **Probabilistische-Kalibrierung:** In Abbildung 6.7 sind die PIT-Histogramme der fünf Modelle für Zeitreihe 2 abgebildet. Das ARIMA- und das Poisson-Modell weichen stark von einer Gleichverteilung ab. Das ZANBI-Modell weist einen großen Ausreißer von der Gleichverteilung bei sehr kleinen Werten auf. Das NBI-Modell ist recht nah an der Gleichverteilung, das ZINBI-Modell erfüllt diese noch etwas besser. Die Annahme der probabilistischen Kalibrierung ist insgesamt für das NBI- und das ZINBI-Modell gerechtfertigt.
- **Marginale-Kalibrierung:** In Abbildung 6.8 ist die marginale Kalibrierung für die Modelle für Zeitreihe 2 abgebildet. Das ARIMA-, das Poisson- und das ZANBI-Modell weisen für Schwellenwerte unter 10 große Differenzen zwischen vorhergesagter und beobachteter Verteilungsfunktion auf. ZINBI- und NBI-Modell haben hier kleinere Abweichungen. Für Schwellenwerte größer 10 ist das ZINBI-Modell besser zu bewerten als das NBI-Modell. Die Annahme der marginalen Kalibrierung ist somit für das NBI- und das ZINBI-Modell gerechtfertigt.
- **Spitze:** Abbildung 6.9 zeigt die Quantilsdifferenzen der Vorhersagen der fünf Modelle. Nur die NBI- und ZINBI-Modelle weisen Kalibrierung auf. Die Quantilsdifferenzen der Vorhersagen des ZINBI-Modells sind minimal kleiner als die des NBI-Modell.
- **Rootogramm:** Abbildung 6.10 zeigt das Rootogramm der fünf Modelle für Zeitreihe 2. Für das ARIMA- und Poisson-Modell weicht die Dichte aus der Mischverteilung der Vorhersagen deutlich von der Dichte der Daten ab. Das ZANBI-Modell unterschätzt die Wahrscheinlichkeit für Nullwerte. Die Anpassung beim NBI- und ZINBI-Modell ist ähnlich.
- **Zeitreihe und Quantilvorhersagen:** Abbildung 6.11 zeigt den Verlauf der Zeitreihe 2 und den Verlauf der Quantile der vorhergesagten Verteilung. Für das ARIMA-Modell zeigt sich hier der Nachteil der Normalverteilung, das 10 % Quantil ist negativ für alle 30 Tage. Zwischen dem NBI- und dem ZINBI-Modell gibt es kaum Unterschiede.

Das 90 % Quantil des NBI-Modells nimmt mal größere, mal kleinere Werte an, als das 90 % Quantil des ZINBI-Modells, welches weniger variabel ist.

- **Pinball-Loss / Mittlere quadratische Abweichung:** Tabelle 6.2 zeigt die Pinball-Loss-Werte und die MQA. Der Pinball-Loss-Wert ist für das ZINBI-Modell am kleinsten, etwas größer ist der Wert für das NBI-Modell. Die Mittlere Quadratische Abweichung (MQA) ist für alle Modelle in etwa gleich.

Für Zeitreihe zwei liefern alle Methoden zur Evaluierung der Modelle konsistente Ergebnisse. Das ZINBI-Modell wird dem NBI-Modell leicht vorgezogen.

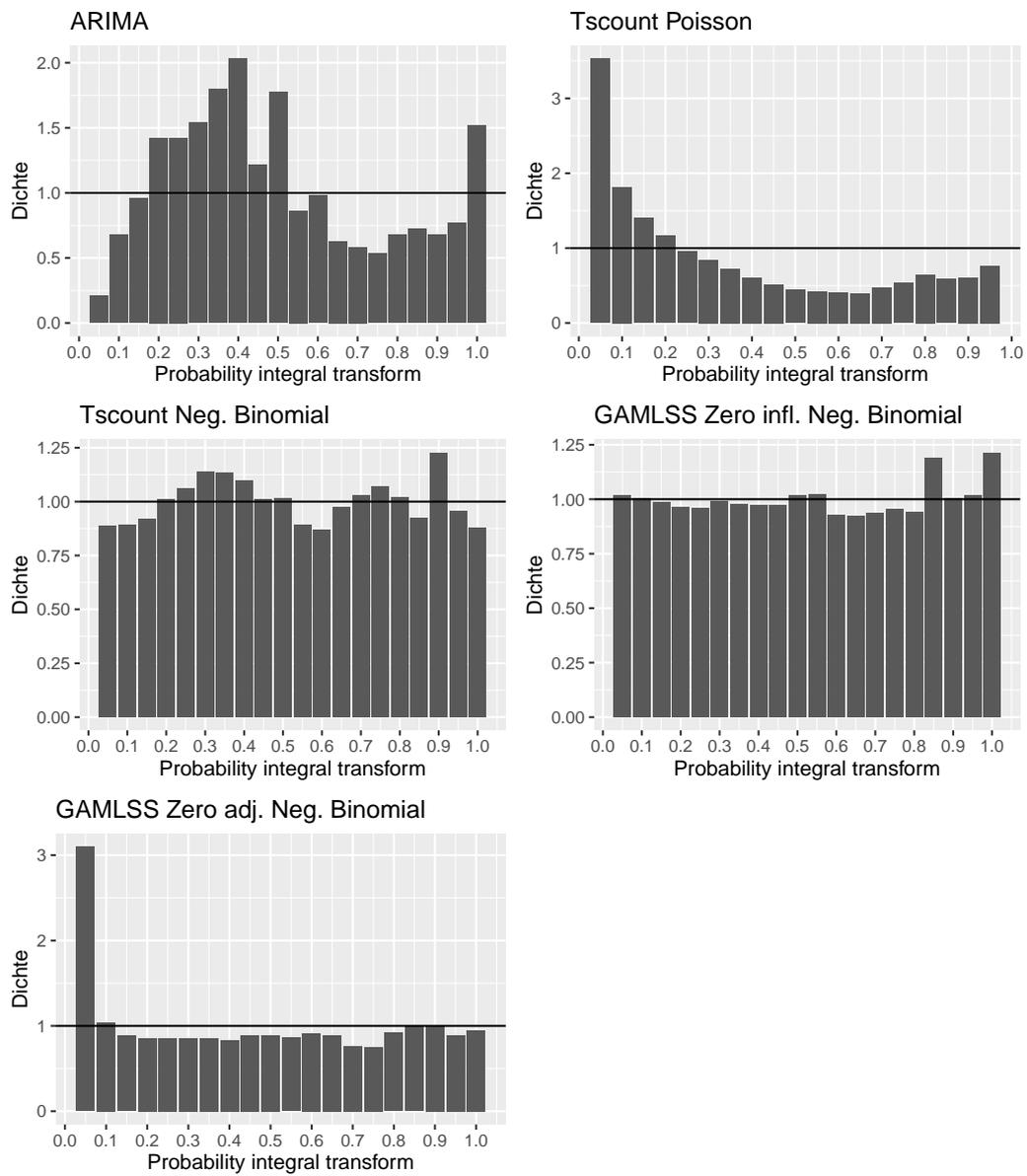


Abbildung 6.7: Zeitreihe 2 PIT Histogramme

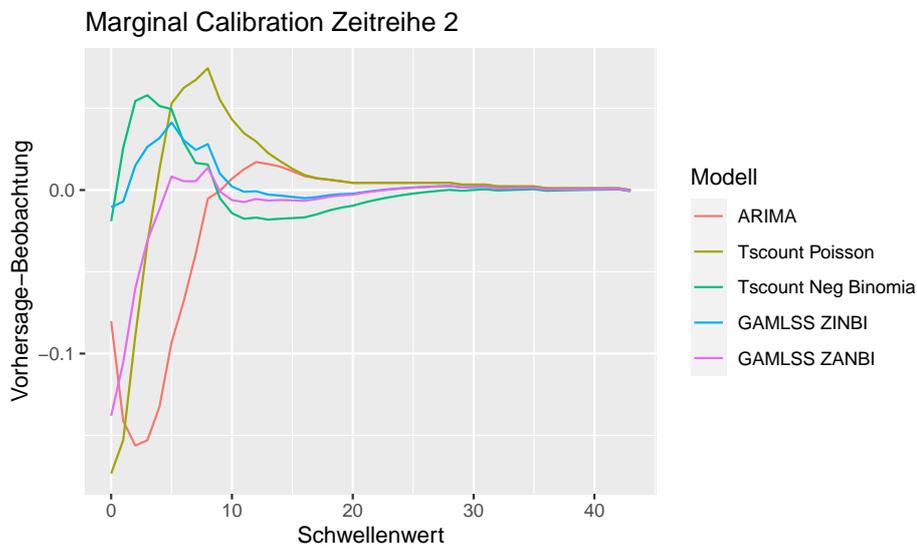


Abbildung 6.8: Zeitreihe 2 Marginaler-Kalibrierungs-Plot

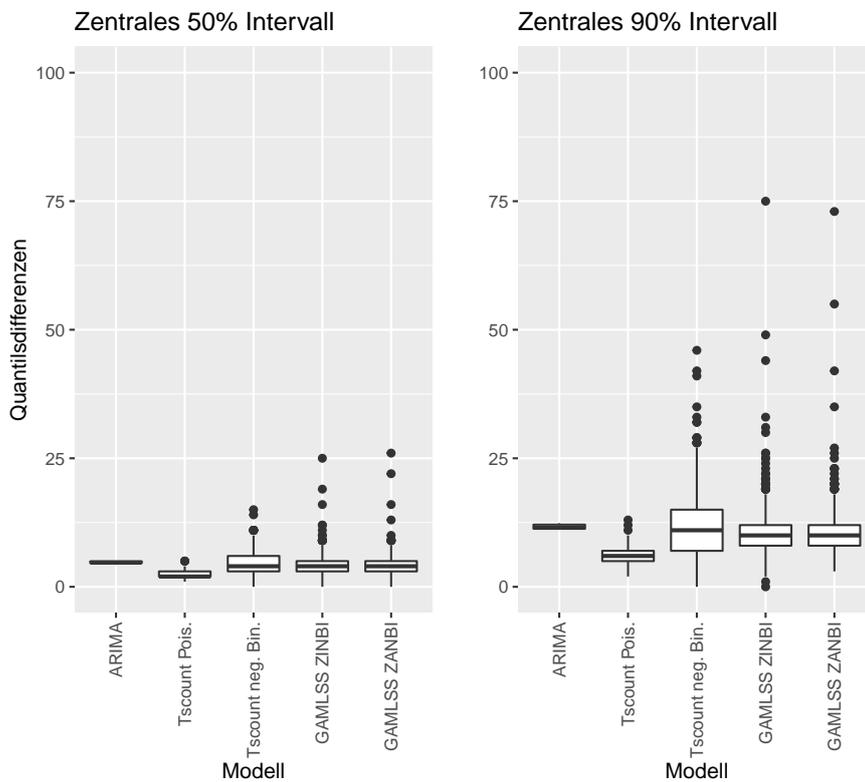


Abbildung 6.9: Zeitreihe 2 Sharpnes Plot

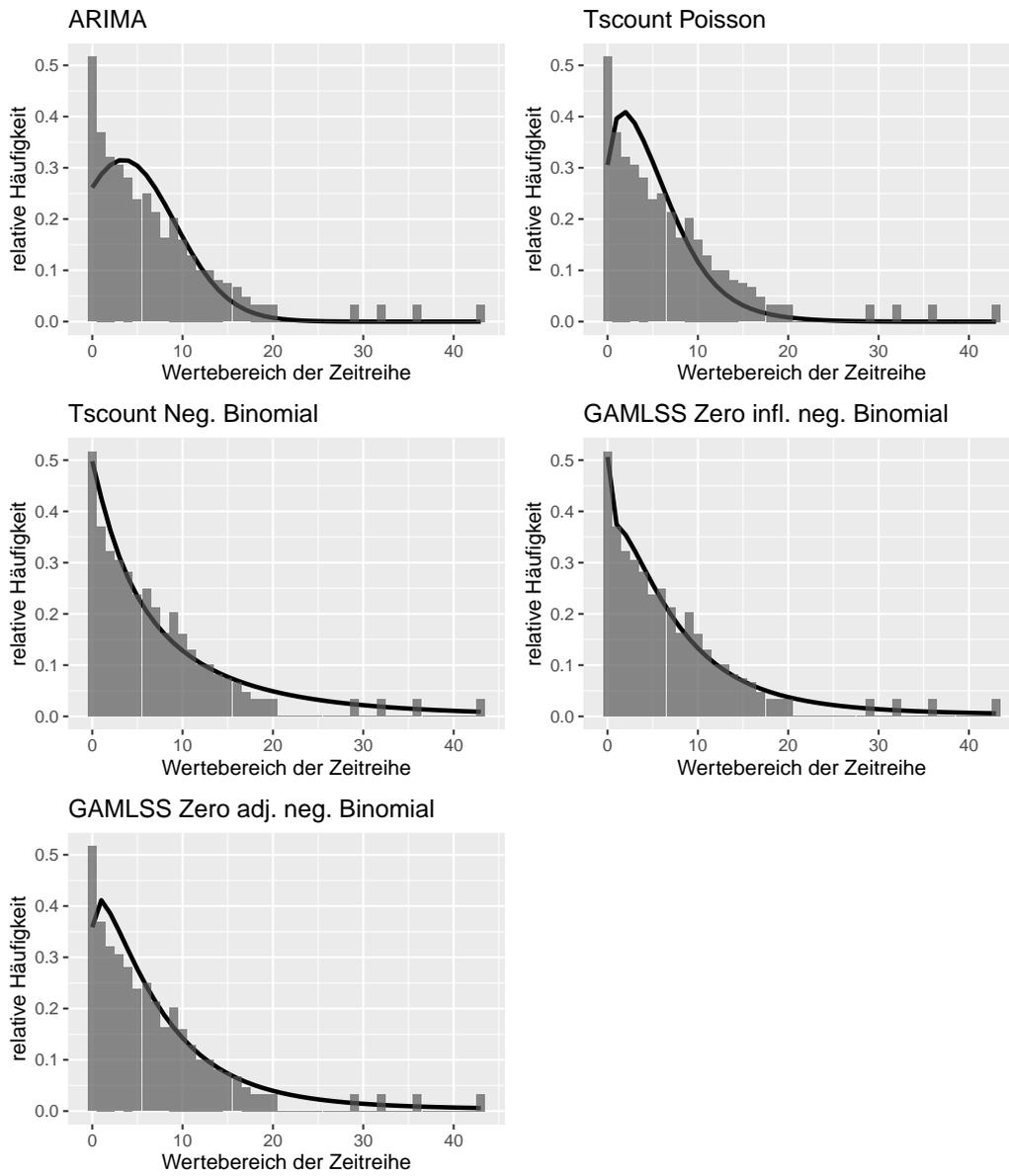


Abbildung 6.10: Zeitreihe 2 Rootogramm

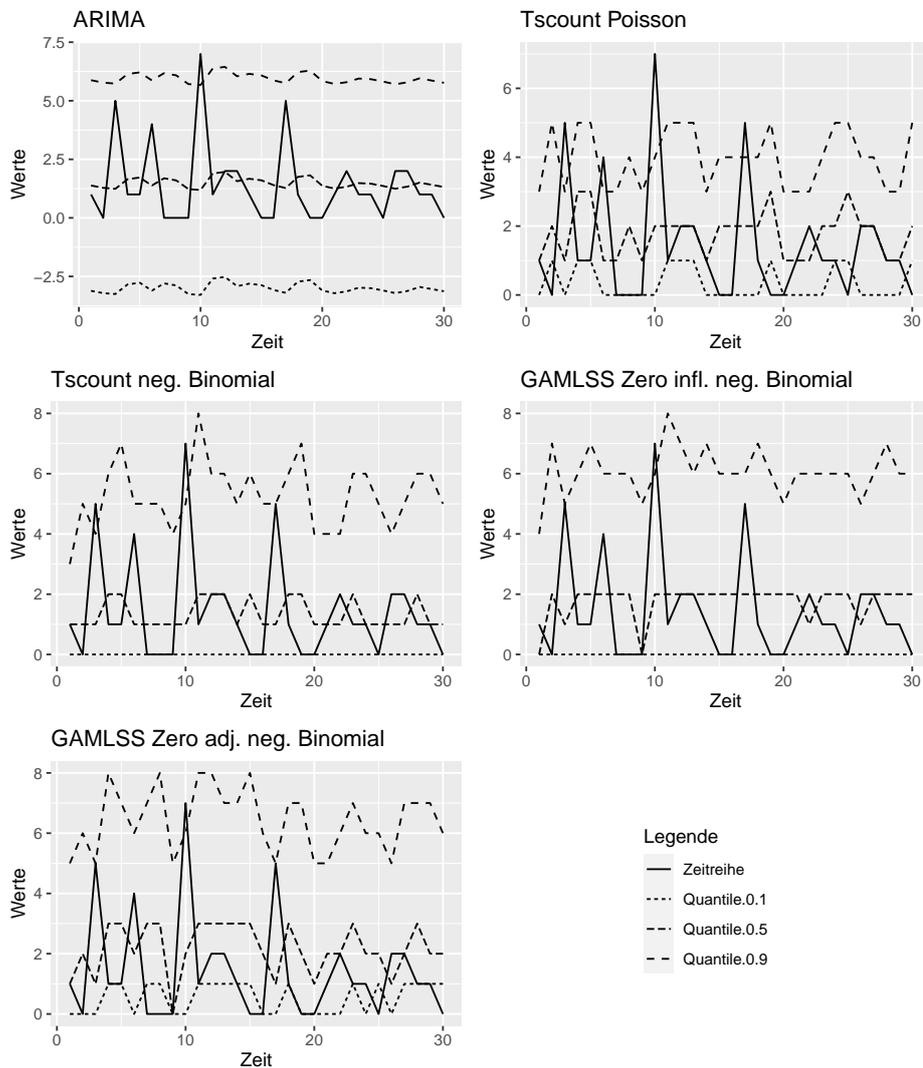


Abbildung 6.11: Werte Zeitreihe 2 und Quantile der vorhergesagten Verteilung

Modell	Pinball-Loss	MQA
ARIMA	0.83	16.03
Tscount Poisson	0.84	15.54
Tscount neg. Binomial	0.77	15.76
GAMLSS zero inf. neg. Binomial	0.76	16.30
GAMLSS zero adj. neg. Binomial	0.78	16.30

Tabelle 6.2: Zeitreihe 2, Pinball-Loss und Mittlere quadratische Abweichung

6.3.3 Zeitreihe 3

Für Zeitreihe 3 enthält das ARIMA-Modell, für die unterschiedlichen CV-Durchgänge, zwischen 1 und 5 MA-Terme und zwischen 0 und 5 AR-Terme. Das Tscout- Poisson-Modell enthält 11 Lags, das Tscout-, NBI-Modell enthält 3 Lags. GAMLSS-, ZINBI- und ZANBI-Modelle enthalten 11 Lags.

- **Probabilistische-Kalibrierung:** Abbildung 6.12 zeigt die PIT-Histogramme für Zeitreihe 3. Poisson- und NBI-Modelle weisen große Abweichungen von einer Gleichverteilung auf. ARIMA-, ZINBI- und ZANBI-Modelle können am ehesten als probabilistisch kalibriert betrachtet werden. Zufriedenstellend ist die Gleichverteilung jedoch in keinem der Modelle.
- **Marginale-Kalibrierung:** In Abbildung 6.13 sind marginale Kalibrierungs-Plots für Zeitreihe 3 abgebildet. Die drei Modelle mit den geringsten Abweichungen von der Null sind das ARIMA-, das ZINBI- und das ZANBI-Modell. ZINBI- und ZANBI-Modelle sind dabei praktisch identisch. Die anderen Modelle weisen abschnittsweise deutlich größere Differenzen auf.
- **Spitze:** In Abbildung 6.14 sind die Quantilsdifferenzen für die verschiedenen Modelle für Zeitreihe 3 abgebildet. Das ARIMA-Modell, das ZINBI- und das ZANBI-Modell sind die Modelle, für die am ehesten von Kalibrierung ausgegangen werden kann. Das ARIMA-Modell weist hier leicht kleinere Quantilsdifferenzen auf, als die zwei anderen Modelle.
- **Rootogramm:** In Abbildung 6.15 sind Rootogramme der fünf Modelle für Zeitreihe 3 abgebildet. Die Dichte aus den Daten und die Dichte der Mischverteilung der Vorhersagen für das Poisson- und NBI-Modell stimmen nicht überein. Die Übereinstimmung ist für das ARIMA-, das ZANBI- und für das ZINBI-Modell besser, jedoch unterschätzt das ARIMA-Modell die Null, und das ZINBI- und ZANBI-Modell überschätzen die Null leicht.
- **Zeitreihe und Quantilvorhersagen:** Abbildung 6.16 zeigt die Zeitreihe und Quantile der vorhergesagten Verteilungen. Offensichtlich nimmt die Zeitreihe die überwiegende Zeit den Wert Null an und in seltenen Fällen recht große Werte. Wie auch schon in den vorherigen Abbildungen festgestellt wurde, haben alle fünf Modelle große Probleme mit dieser Datenlage. Die Phasen, in denen die Zeitreihe Null ist, werden von der ZINBI- und ZANBI-Verteilung gut abgebildet. Die Zeitpunkte

in denen die Zeitreihe positive Werte annimmt, werden allerdings von keinem Modell gut erfasst.

- **Pinball-Loss/ Mittlere quadratische Abweichung:** In Tabelle 6.3 finden sich die Pinball-Loss-Werte und die mittleren quadratischen Abweichungen für die fünf Modelle. ARIMA, ZANBI und ZINBI haben den geringsten Pinball-Loss-Wert. Die mittlere quadratische Abweichung ist für das ARIMA-Modell die geringste, gefolgt von dem ZINBI- und ZANBI-Modell.

Zeitreihe 3 kann insgesamt schlecht durch die Modelle abgebildet werden. Das zeigt sich einheitlich über alle Methoden. Allerdings ist festzustellen, dass es nur anhand der Kalibrierung und Spitze bzw. anhand der Rootogramme schwerfällt zu beurteilen, wie schwierig die Daten für die Modelle sind. Die schlechte Anpassung der Modelle an die Daten wird am leichtesten durch den Vergleich der Zeitreihe und der Quantile der vorhergesagten Verteilung deutlich.

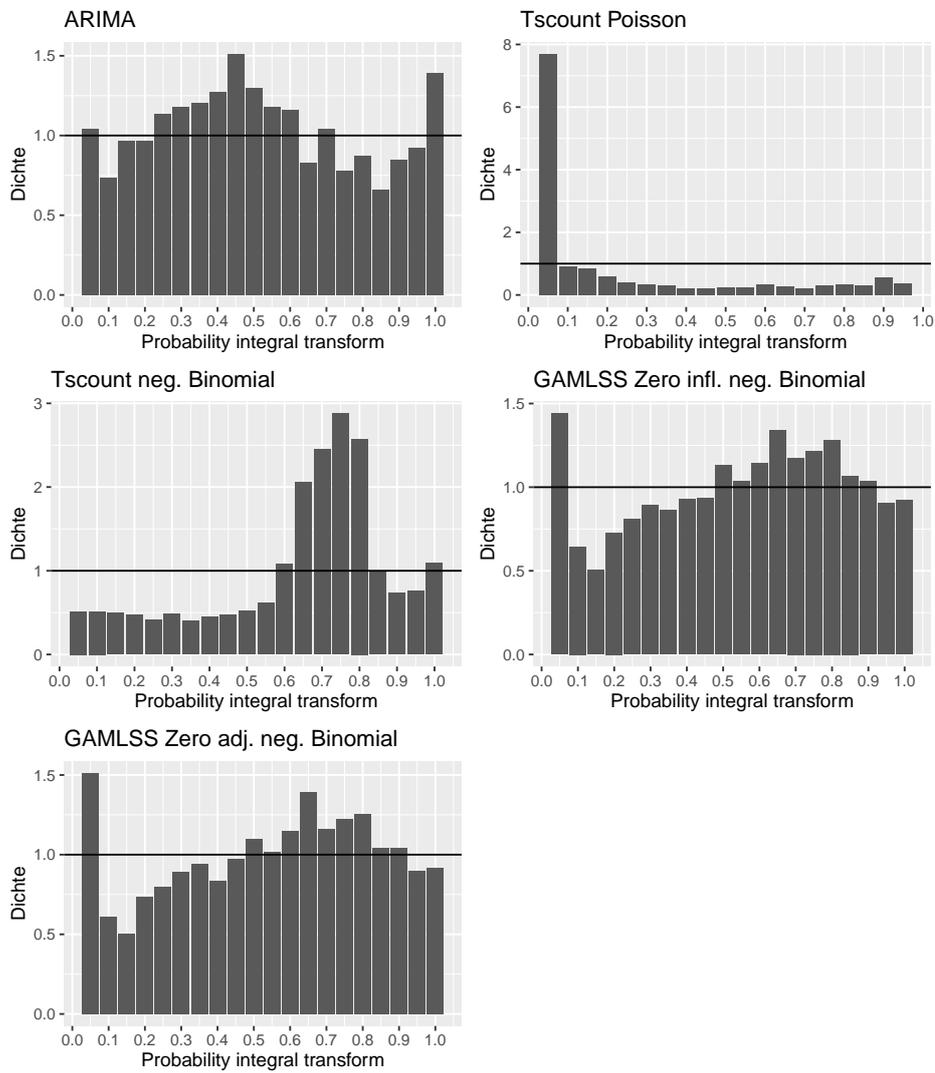


Abbildung 6.12: Zeitreihe 3 PIT Histogramme

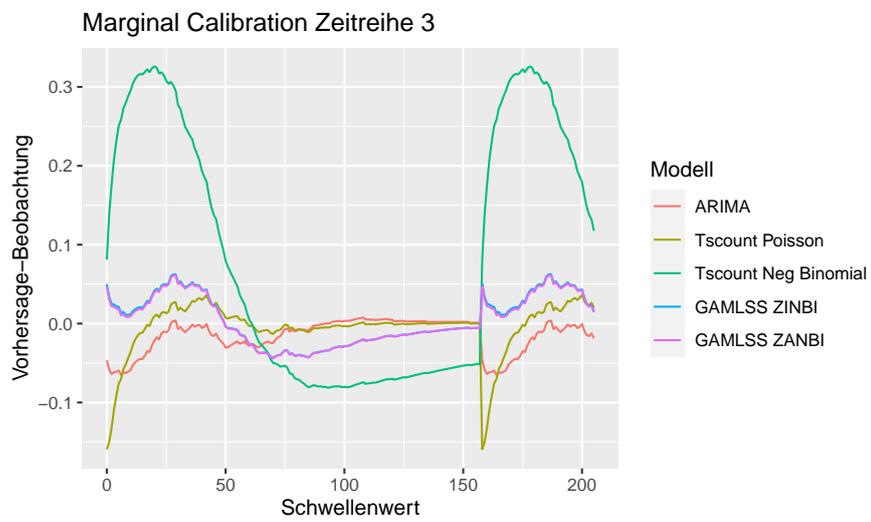


Abbildung 6.13: Zeitreihe 3 Marginaler-Kalibrierungs-Plot

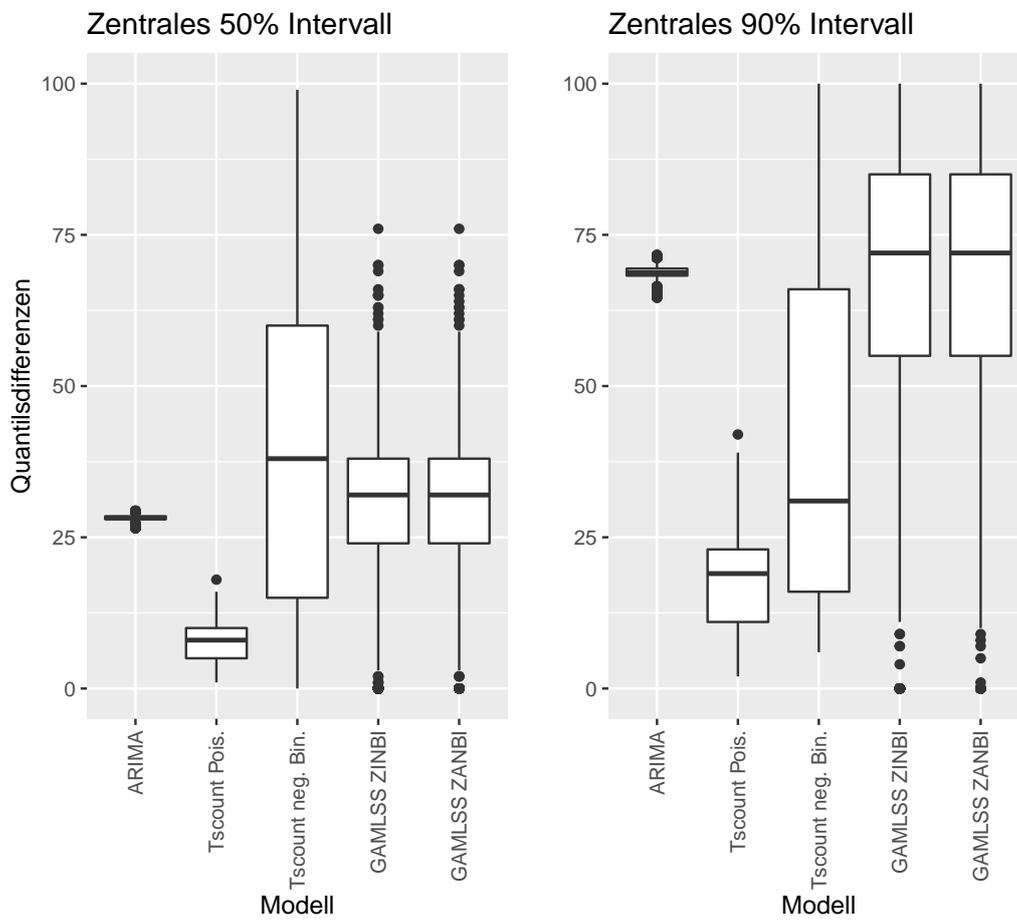


Abbildung 6.14: Zeitreihe 3 Sharpnes Plot

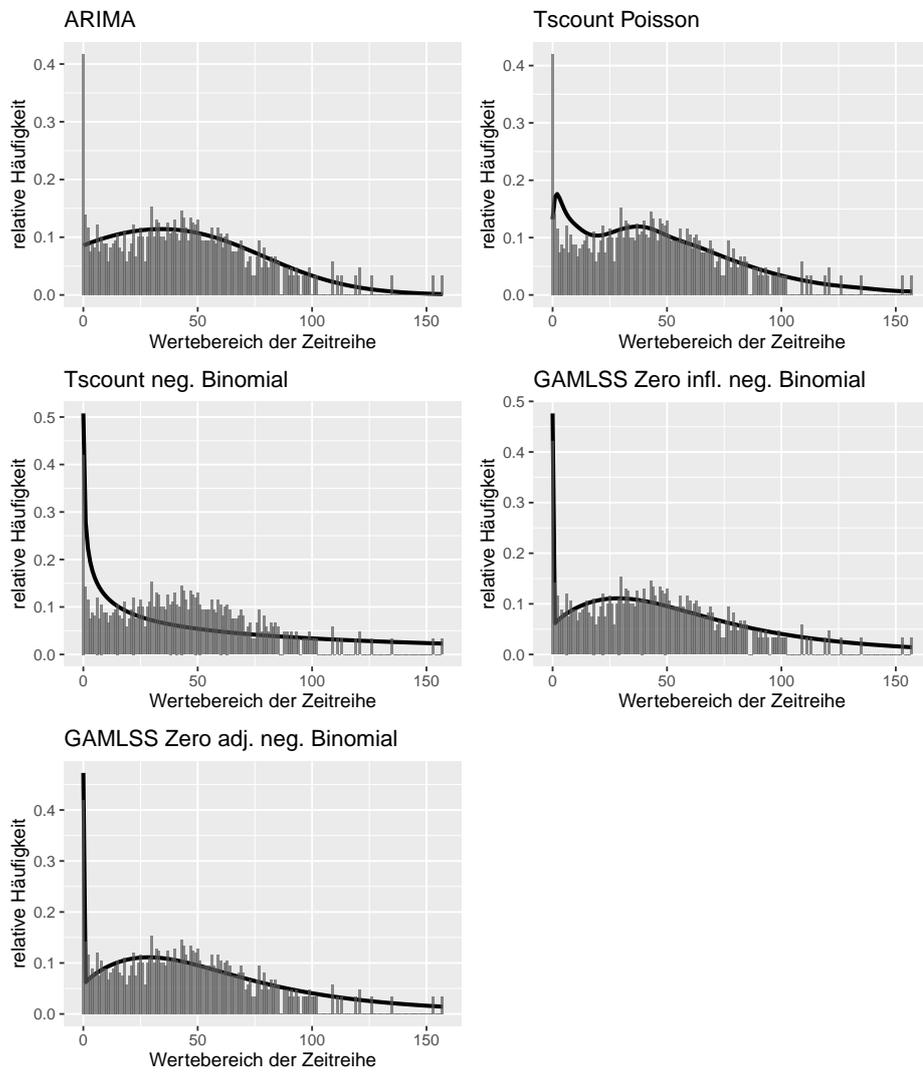


Abbildung 6.15: Zeitreihe 3 Rootogramm

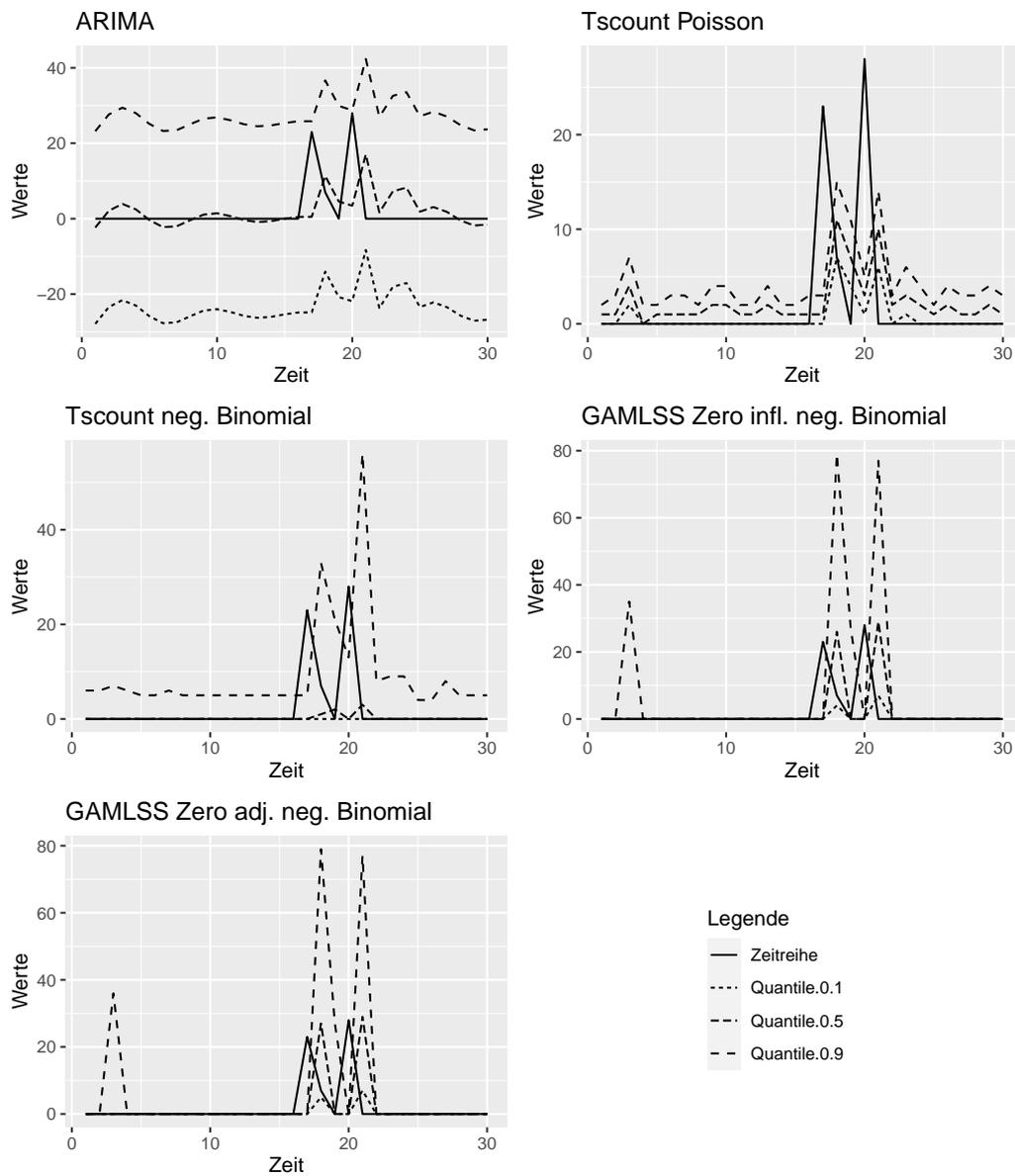


Abbildung 6.16: Werte Zeitreihe 3 und Quantile der vorhergesagten Verteilung

Modell	Pinball-Loss	MQA
ARIMA	5.08	521.00
Tscount Poisson	6.85	583.71
Tscount neg. Binomial	7.06	556.55
GAMLSS zero inf. neg. Binomial	5.08	526.61
GAMLSS zero adj. neg. Binomial	5.08	526.61

Tabelle 6.3: Zeitreihe 3, Pinball-Loss und Mittlere quadratische Abweichung

6.3.4 Zeitreihe 4

Für Zeitreihe 3 enthält das ARIMA Modell, für die unterschiedlichen CV-Durchgänge, zwischen 1 und 5 MA-Terme und zwischen 1 und 2 AR-Terme. Das Tscount-, Poisson-Modell enthält 12 Lags, das Tscount-, NBI-Modell enthält 10 Lags. GAMLSS-, ZINBI- und ZANBI-Modelle enthalten 10 Lags.

- **probabilistische Kalibrierung:** Wie in Abbildung 6.17 zu sehen ist, ist das PIT-Histogramm des Poisson-Modells für Zeitreihe 4 U-förmig. Dies ist ein Indiz für eine Unterdispersion der Vorhersagen, wobei kleine Werte dabei stärker unterschätzt werden als große Werte. Das PIT-Histogramm des NBI-Modells ist umgekehrt U-förmig. Dies deutet auf Überdispersion der vorhergesagten Verteilungen hin. Am ehesten einer Gleichverteilung folgen die PIT-Histogramme des ARIMA-, des ZINBI- und des ZANBI-Modells. Jedoch liegen auch hier größere Abweichungen von der Gleichverteilung vor.
- **marginale Kalibrierung:** Abbildung 6.18 zeigt die marginale Kalibrierung der fünf Modelle für Zeitreihe 4. Alle Modelle weisen größere Differenzen zwischen der vorhergesagten und beobachteten Verteilungsfunktion auf. Am besten sind noch das ARIMA-, das ZANBI- und das ZINBI-Modell zu bewerten. Das ZANBI- und das ZINBI-Modell sind dabei über große Teile identisch.
- **Spitze:** Für das ARIMA-, ZANBI- und ZINBI-Modell wird von Kalibrierung ausgegangen. Das ZINBI- und das ZANBI-Modell haben beide leicht spitzere Verteilungen als das ARIMA-Modell. Somit sind das ZANBI- und ZINBI-Modell bei Betrachtung der Kalibrierung und Spitze die besten Modelle.
- **Rootogramm:** Abbildung 6.20 zeigt die Rootogramme für Zeitreihe 4. Auch hier zeigt sich, dass alle Modelle Probleme haben die Verteilung

der Daten richtig zu erfassen. Die beiden ZINBI- und ZANBI-Modelle treffen die Verteilung noch am besten, allerdings werden hier erneut die Nullwerte deutlich überschätzt. Das ARIMA-Modell unterschätzt diese hingegen deutlich und überschätzt in der Folge die Wahrscheinlichkeit für kleine Werte deutlich. Die restliche Anpassung ist ähnlich wie die des ZINBI-Modells.

- **Zeitreihe und Quantilvorhersagen:** In Abbildung 6.21 ist Zeitreihe 4 und die 10 %-, 50 %-, 90 %- Quantile der vorhergesagten Verteilung im Zeitverlauf abgebildet. Die 10%-, 90%- Quantil-Intervalle sind für das ARIMA-, das ZANBI- und das ZINBI-Modell in den abgebildeten 30 Tagen überraschend ähnlich.
- **Pinball-Loss / Mittlere quadatische Abweichung:** In Tabelle 6.4 sind die Pinball-Loss-Werte und die Mittlere quadatische Abweichung der fünf Modelle eingetragen. Das ZANBI-Modell hat den geringsten Pinball-Loss-Wert und auch die geringste Mittlere quadatische Abweichung.

Zeitreihe 4 kann erneut nicht gut von den Modellen abgebildet werden. Allerdings funktionieren die Modelle hier wesentlich besser, als für Zeitreihe drei. Die Ergebnisse der Methoden zur Evaluierung liefern konsistente Ergebnisse. Das ZANBI- und das ZINBI-Modell werden von allen Methoden als die besten Modelle ausgewählt.

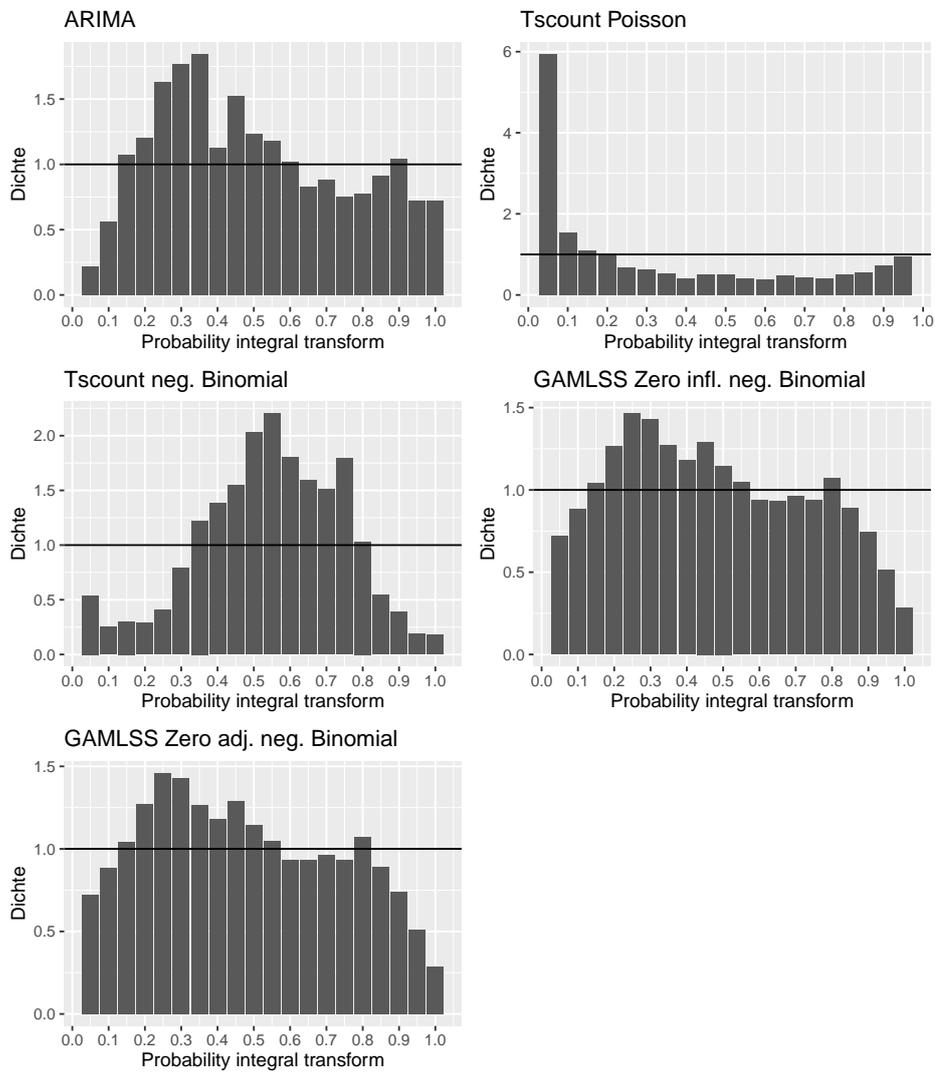


Abbildung 6.17: Zeitreihe 4 PIT Histogramme

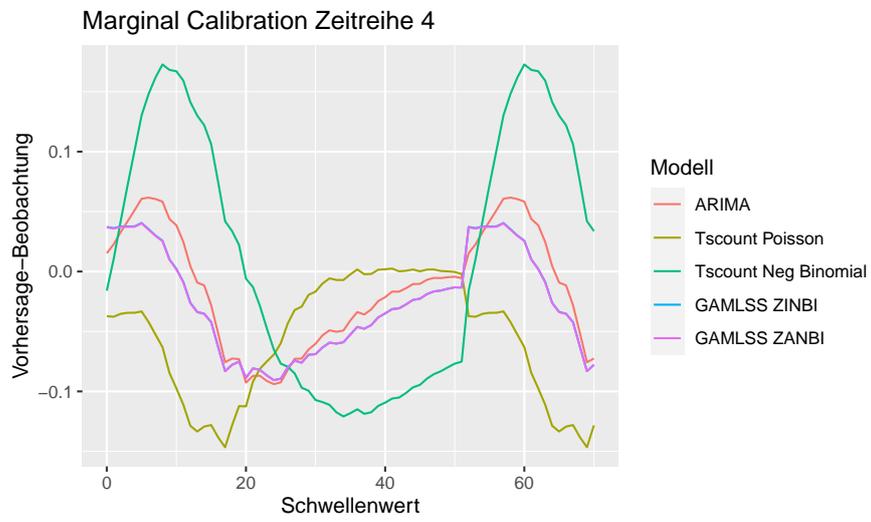


Abbildung 6.18: Zeitreihe 4 Marginaler-Kalibrierungs-Plot

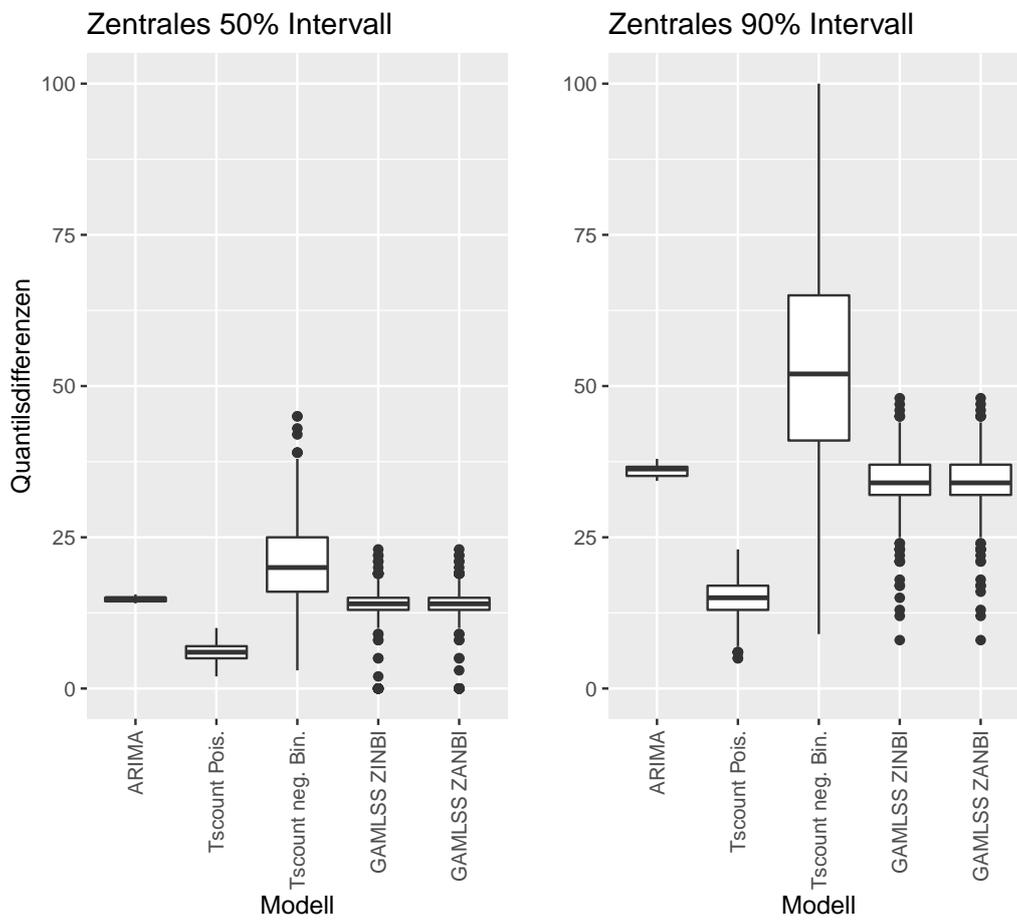


Abbildung 6.19: Zeitreihe 4 Sharpnes Plot

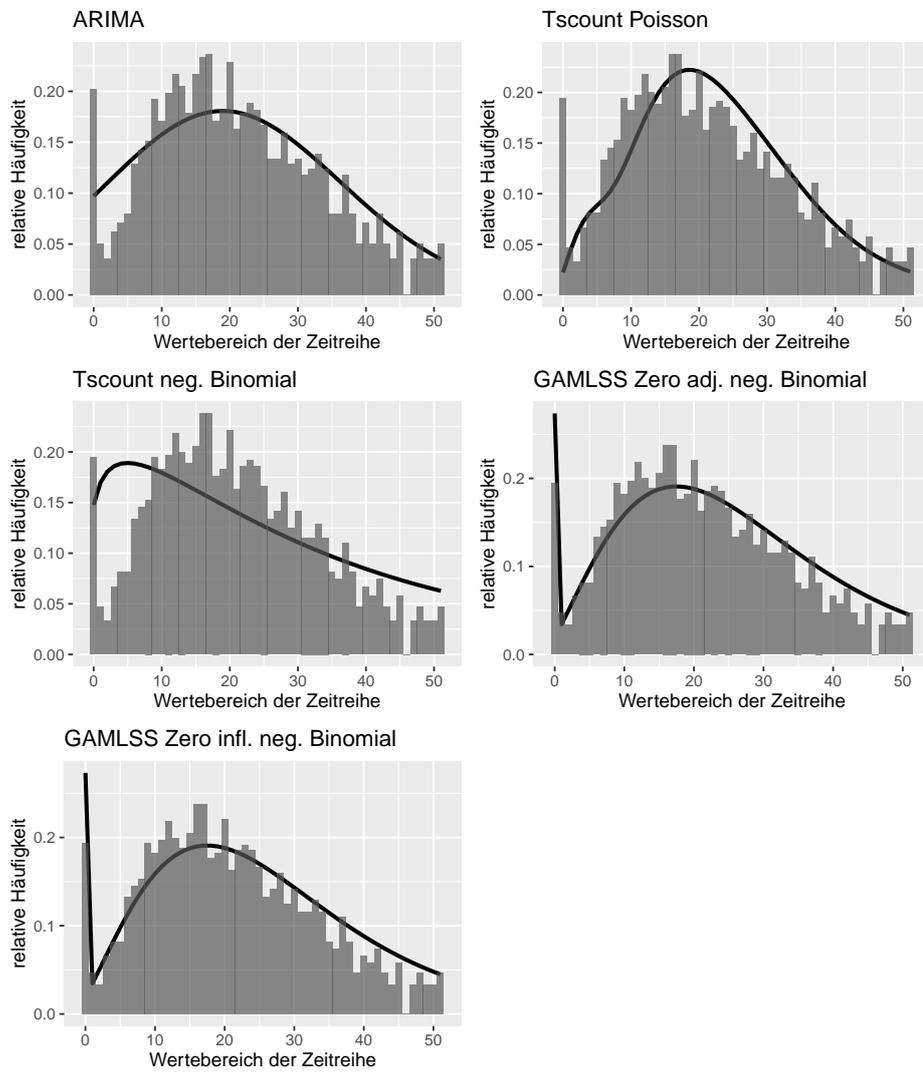


Abbildung 6.20: Zeitreihe 4 Rootogramm

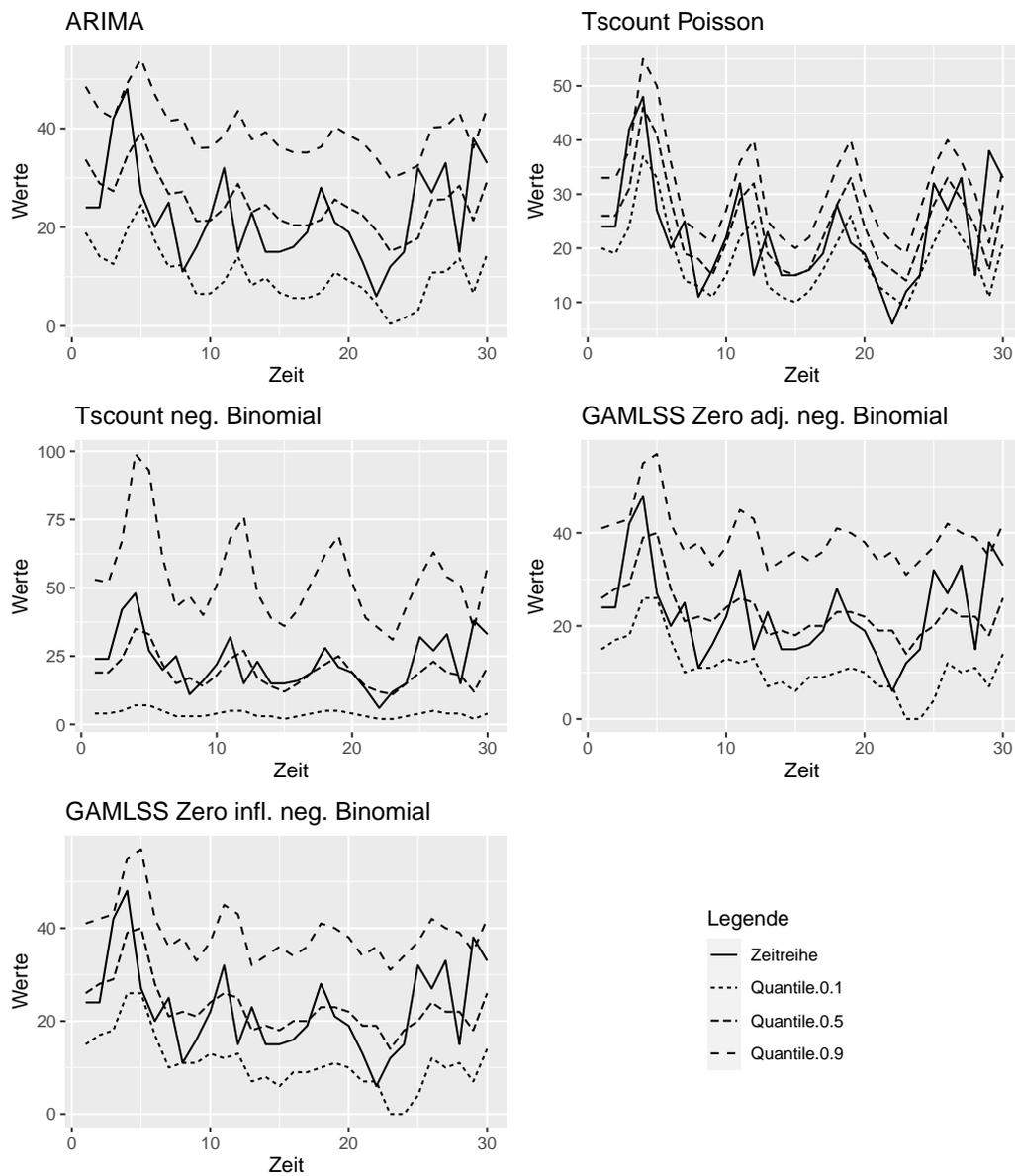


Abbildung 6.21: Werte Zeitreihe 4 und Quantile der vorhergesagten Verteilung

Modell	Pinball-Loss	MQA
ARIMA	2.11	84.13
Tscount Poisson	2.21	75.40
Tscount neg. Binomial	2.34	76.84
GAMLSS zero inf. neg. Binomial	1.97	75.31
GAMLSS zero adj. neg. Binomial	1.96	132.76

Tabelle 6.4: Zeitreihe 4, Pinball-Loss und Mittlere quadratische Abweichung

Kapitel 7

Fazit

In dieser Arbeit wurden zwei Schwerpunkte gelegt. Zum einen auf verschiedenen Modelle zur probabilistischen Vorhersage von Zählzeitenreihen, zum anderen auf mögliche Methoden zur Evaluierung von probabilistischen Vorhersagen. Als Modelle wurden die traditionellen Zeitreihenmethoden ARIMA und das Tscout-Paket zur Modellierung von Zählzeitenreihen mittels Poisson- oder negativer Binomialverteilung vorgestellt. Weiter wurde das allgemeine GAMLSS-Modell als autoregressives Modell auf Zeitreihen angewendet. Das GAMLSS-Modell ist kein spezielles Zeitreihenmodell. Deshalb sollte kritisch hinterfragt werden, ob die Modellannahme für eine gegebene Zeitreihe erfüllt wird. Falls die Modellannahmen hinreichend gut erfüllt werden, bietet das GAMLSS-Modell zwei Vorteile gegenüber dem klassischen Ansatz ARIMA und dem Zählzeitenreihenmodell Tscout. Zum einen kann aus einer Vielzahl an Verteilungen eine Verteilungsannahme getroffen werden, zum anderen können alle Parameter einer Verteilung in Abhängigkeit von zurückliegenden Beobachtungen geschätzt werden. Dadurch kann mit einem GAMLSS-Modell theoretisch eine bessere Anpassung der vorhergesagten Verteilung an die Daten ermöglicht werden. Konkret wurden in dieser Arbeit ein GAMLSS-Modell mit Modellannahme Zero Inflated Negative Binomial (ZINBI) und Zero Adjusted Negative Binomial (ZANBI) verwendet.

Als Methoden zur Evaluierung von probabilistischen Vorhersagen wurden das Konzept der Kalibrierung und Spitze, Rootogramme und die Pinball-Loss-Funktion vorgestellt. Weiter wurden die Mittlere Quadratische Abweichung der Vorhersagen von den Beobachtungen und Grafiken, die den Verlauf der Quantile der vorhergesagten Verteilung dem Verlauf der Zeitreihe gegenüberstellen, verwendet, um eine leichtere Interpretation der Methoden zu ermöglichen.

Um sowohl die Modelle, als auch die Methoden zur Evaluierung zu testen, wurden vier Zeitreihen aus dem Datensatz der „M5 Accuracy Competition“

[17] ausgewählt.

Auf alle vier Zeitreihen wurden die fünf Modelle, ARIMA, Tscount Poisson, Tscount NBI, GAMLSS ZINBI und GAMLSSS ZANBI angewendet. Mittels Zeitreihen-Kreuzvalidierung wurden für jede Zeitreihe und jedes Modell jeweils 900 Tagesbeobachtungen vorhergesagt. Die Methoden zur Evaluierung wurden auf diese 900 Vorhersagen angewendet und die Ergebnisse wurden in Kapitel 6 dargestellt.

Die Ergebnisse in Kapitel 6 zeigen, dass die probabilistische Vorhersage von Zählzeitenzeitreihen eine komplexe Aufgabe ist. Keines der Modelle kann für jede der vier Zeitreihen die besten probabilistischen Vorhersagen erzeugen. Für eine der Zeitreihen liefern alle Modelle nur schlechte Ergebnisse. Insgesamt sprechen die Ergebnisse für das GAMLSS-ZINBI-Modell und das Tscount-NBI-Modell. Das ZINBI-Modell ist flexibel in der Anpassung an die Daten. Sowohl Überdispersion, als auch gehäufte Nullwerte können mit dem ZINBI-Modell gut abgebildet werden. Vorsicht ist allerdings geboten, wenn die Zeitreihe keine gehäuften Nullwerte hat, bzw. die gehäuften Nullwerte selten auftreten. In diesem Fall wird die Wahrscheinlichkeit von Nullwerten vom ZINBI-Modell überschätzt. Falls eine Zeitreihe keine oder wenig gehäuften Nullwerte hat, ist das Tscount-NBI-Modell möglicherweise dem ZINBI-Modell vorzuziehen. Das ZANBI-Modell ist theoretisch genauso gut geeignet, Überdispersion und gehäufte Nullwerte abzubilden wie das ZINBI-Modell, allerdings überschätzt das ZANBI-Modell bei den hier verwendeten Daten die Nullwerte häufiger und stärker als das ZINBI-Modell. Ansonsten schneiden ZINBI- und ZANBI-Modell sehr ähnlich ab. Etwas überraschend sind die Ergebnisse des ARIMA-Modells. Für Zählzeitenzeitreihen mit großen Werten kann das ARIMA-Modell durchaus gute Ergebnisse liefern. Schlecht kann das ARIMA-Modell allerdings gehäuften Nullwerte abbilden. Für Zeitreihen mit kleinen Werten liegt ein Teil der vorhergesagten Verteilung im negativen Bereich, was die Aussagekraft stark einschränkt. Die Ergebnisse liefern wenig Anhaltspunkte für die Verwendung eines Poisson-Modells. Dieses ist für die hier verwendeten Daten zu unflexibel und kann weder Überdispersion noch gehäuften Nullwerte abbilden. Die NBI- bzw. ZINBI-Modelle liefern hier für alle vier Zeitreihen bessere Ergebnisse als ein Poisson-Modell.

Die Ergebnisse in Kapitel 6 zeigen weiterhin, dass die vorgestellten Methoden zur Evaluierung von probabilistischen Vorhersagen leicht anzuwenden sind, falls nur eines der zur Verfügung stehenden Modelle die Daten gut abbildet, lässt sich das passende Modell anhand der Grafiken leicht identifizieren. Falls keines der zur Verfügung stehenden Modelle die Daten herausragend abbildet, wird die Anwendung der Methoden zur Evaluierung von probabilistischen Vorhersagen schwierig. Dann ist es, zum einen schwierig zu entscheiden, bis zu welcher Grenze ein Modell noch eine akzeptable Anpassung an die Da-

ten hat, bzw. wann ein Modell generell als nicht passend betrachtet wird, zum anderen, kann es schwierig sein, die Anpassung verschiedener Modelle zu vergleichen und das beste Modell zu identifizieren oder eine Rangfolge zu erstellen. Als Beispiel dazu soll ein PIT-Histogramm dienen. Es ist nicht klar, ab welcher Verteilung der PIT-Werte nicht mehr von Gleichverteilung ausgegangen werden kann. Außerdem ist es im Vergleich zweier Modelle schwierig zu beurteilen, ob die PIT-Werte des einen Modells besser einer Gleichverteilung folgen, als die PIT-Werte des anderen Modells.

Es scheint ratsam, verschiedene Methoden zur Beurteilung eines Modells zu verwenden. Die Verwendung von nur einer Methode kann möglicherweise einen falschen Eindruck über die Anpassung des Modells an die Daten liefern. Betrachtet man, für Zeitreihe drei zum Beispiel, nur den marginalen Kalibrierungsplot, könnte der Eindruck entstehen, dass das ARIMA-Modell geeignete probabilistische Vorhersagen erzeugt. Betrachtet man zusätzlich die Zeitreihe und die Quantile im Zeitverlauf, ist offensichtlich, dass das ARIMA Modell nicht gut geeignet ist, weil die unteren Quantile im negativen Bereich liegen.

Die Pinball-Loss-Funktion ist eine Alternative zu den grafischen Verfahren. Für die Zeitreihen, bei denen mit Hilfe der grafischen Methoden eine gute Anpassung der vorhergesagten Verteilungen eines Modells festgestellt wurde, liefert die Pinball-Loss-Funktion das gleiche Ergebnis wie die grafischen Verfahren. Das, mittels grafischer Methoden, als bestes Modell ermittelte Modell hat den geringsten Pinball-Loss-Wert. Komplizierter wird die Anwendung der Pinball-Loss-Funktion, falls es schwierig ist mittels grafischer Methoden eindeutig ein passendes Modell zu identifizieren. In diesem Fall kann die Pinball-Loss-Funktion auch zu einem ganz anderen Ergebnis kommen und eine andere Rangfolge der Modelle liefern, als die Reihenfolge die aus den grafischen Methoden abgeleitet wurde.

Ein Kritikpunkt der vorgestellten grafischen Methoden zur Evaluierung von probabilistischen Vorhersagen ist, dass sie sehr aufwendig sind. Für jede Zeitreihe müssen jeweils verschiedene Grafiken betrachtet werden, um zu einem Ergebnis zu kommen. Für die vier Zeitreihen in dieser Arbeit ist dieser Ansatz noch praktikabel. Würde man allerdings auf diese Weise Modelle für alle 30490 im M5-Datensatz enthaltene Zeitreihen ermitteln wollen, wäre der Ansatz sehr aufwendig und wenig praktikabel. In dieser Situation würde sich die Pinball-Loss-Funktion besser eignen, da sie es beispielsweise erlauben würde, das beste Modell, automatisiert anhand des geringsten Pinball-Loss-Werts, zu identifizieren. Mit Hilfe der Pinball-Loss-Funktion ist es allerdings nicht möglich festzustellen, ob die Verteilungsvorhersagen der zur Verfügung stehenden Modelle die Daten überhaupt ausreichend gut abbildet. Die Pinball-Loss-Funktion erlaubt es nur, das Modell mit der besten

Anpassung aller zur Verfügung stehenden Modelle zu ermitteln. Falls keines der Modelle die Daten gut abbildet, bleibt das von der Pinball-Loss-Funktion unentdeckt. Dies ist ein Vorteil der grafischen Methoden, sie erlauben es festzustellen, ob ein Modell sich überhaupt eignet gegebene Daten abzubilden. Zum Abschluss sollen die in dieser Arbeit vorgestellten Modelle kritisch betrachtet werden. Das ARIMA-Modell funktioniert wie erwartet auf Zeitreihen mit kleinen Werten schlecht. Die unteren Quantile der vorhergesagten Verteilung liegen dann im negativen Bereich, was für Zähldaten wenig sinnvoll ist. Für Zeitreihen mit großen Werten und ohne gehäufte Nullwerte funktioniert das ARIMA-Modell ausreichend gut und kann durchaus in Betracht gezogen werden. Poisson- und NBI-Verteilte-Modell aus dem Tscout Paket, haben den Vorteil, dass diese Modelle speziell für Zähldaten entwickelt wurden und somit theoretisch gut fundiert sind. Das Poisson-Modell konnte für die Zeitreihen in dieser Arbeit nicht überzeugen, da es nicht flexibel genug ist, sich den Daten ausreichend anzupassen. Das NBI-Modell funktioniert besser und kann für Zeitreihen ohne gehäufte Nullwerte, oder Zeitreihen, bei denen die gehäuften Nullwerte nur schwach ausgeprägt sind, eine gute Wahl sein. Die ZINBI und ZANBI verteilten GAMLSS-Modelle sind in der Anwendung auf Zeitreihen von der theoretischen Betrachtung her kritisch zu beurteilen, denn es ist fragwürdig, ob alle Modellannahmen im Zeitreihenkontext erfüllt sind. Die empirischen Auswertungen in dieser Arbeit zeigen jedoch keine Probleme bei den GAMLSS Modellen. Im Gegenteil liefert das ZINBI-Modell für die verwendeten Daten die besten Ergebnisse. Dies spricht dafür, dass die Vorteile der Flexibilität der GAMLSS-Modelle die Nachteile, der für Zeitreihen nicht idealen Modellannahmen, überwiegen. Weiter ist zu erwähnen, dass die Möglichkeiten der Modellierung mit GAMLSS keineswegs erschöpft sind. Aus Gründen der Vergleichbarkeit und, um die Modelle nicht zu kompliziert zu gestalten, wurden alle Modelle in Abhängigkeiten von bis zu 12 zurückliegenden Werten der Zeitreihe ermittelt. Hier wäre es beispielsweise möglich, unterschiedliche Abhängigkeiten für jeden Parameter der Verteilung zu verwenden. Weiter könnten externe Kovariablen aufgenommen werden. Die Abhängigkeiten der Zeitreihe von den zurückliegenden Beobachtungen und Kovariablen könnte nicht linear modelliert werden. Außerdem wäre es möglich weitere Verteilungsannahmen zu verwenden, beispielsweise Zähldatenverteilungen mit vier Parametern.

Literaturverzeichnis

- [1] Box, George EP, Jenkins, Gwilym M, Reinsel, Gregory C und Ljung, Greta M: *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [2] Box, George EP und Pierce, David A: *Distribution of residual autocorrelations in autoregressive-integrated moving average time series models*. Journal of the American statistical Association, 65(332):1509–1526, 1970.
- [3] Carbonneau, Real, Laframboise, Kevin und Vahidov, Rustam: *Application of machine learning techniques for supply chain demand forecasting*. European Journal of Operational Research, 184(3):1140–1154, 2008.
- [4] Cole, Timothy J und Green, Pamela J: *Smoothing reference centile curves: the LMS method and penalized likelihood*. Statistics in medicine, 11(10):1305–1319, 1992.
- [5] Corradi, Valentina und Swanson, Norman R: *Predictive density evaluation*. Handbook of economic forecasting, 1:197–284, 2006.
- [6] Czado, Claudia, Gneiting, Tilmann und Held, Leonhard: *Predictive model assessment for count data*. Biometrics, 65(4):1254–1261, 2009.
- [7] DiCiccio, Thomas J und Efron, Bradley: *Bootstrap confidence intervals*. Statistical science, 11(3):189–212, 1996.
- [8] Diebold, Francis X, Gunther, Todd A und Tay, Anthony S: *Evaluating density forecasts with applications to financial risk management*. International Economic Review, 39(4):863, 1998.
- [9] Elliott, Graham und Timmermann, Allan: *Forecasting in economics and finance*. Annual Review of Economics, 8:81–110, 2016.
- [10] Fahrmeir, Ludwig, Kneib, Thomas und Lang, Stefan: *Regression-Modelle, Methoden und Anwendungen*, 2009.

- [11] Gao, Robert, Wang, Lihui, Teti, Roberto, Dornfeld, David, Kumara, Soundar, Mori, Masahiko und Helu, Moneer: *Cloud-enabled prognosis for manufacturing*. CIRP annals, 64(2):749–772, 2015.
- [12] Gneiting, Tilmann, Balabdaoui, Fadoua und Raftery, Adrian E: *Probabilistic forecasts, calibration and sharpness*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69(2):243–268, 2007.
- [13] Hyndman, Rob J und Athanasopoulos, George: *Forecasting: principles and practice*. OTexts, 2018. <https://otexts.com/fpp2/>.
- [14] Hyndman, Rob J und Khandakar, Yeasmin: *Automatic time series forecasting: the forecast package for R*. Journal of Statistical Software, 26(3):1–22, 2008. <https://www.jstatsoft.org/article/view/v027i03>.
- [15] Kleiber, Christian und Zeileis, Achim: *Visualizing count data regressions using rootograms*. The American Statistician, 70(3):296–303, 2016.
- [16] Liboschik, Tobias, Fokianos, Konstantinos und Fried, Roland: *tscount: An R package for analysis of count time series following generalized linear models*. Journal of Statistical Software, 82(5), 2017.
- [17] Makridakis, Spyros, Spiliotis, Evangelos und Assimakopoulos, Vassilis: *The M5 Accuracy competition: Results, findings and conclusions*. Oktober 2020. https://www.researchgate.net/publication/344487258_The_M5_Accuracy_competition_Results_findings_and_conclusions.
- [18] Nelder, John Ashworth und Wedderburn, Robert WM: *Generalized linear models*. Journal of the Royal Statistical Society: Series A (General), 135(3):370–384, 1972.
- [19] R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. <http://www.R-project.org/>.
- [20] Rigby, R. A. und Stasinopoulos, D. M.: *Generalized additive models for location, scale and shape*. Applied Statistics, 54:507–554, 2005.
- [21] Rigby, Robert A und Stasinopoulos, Mikis D: *Mean and dispersion additive models*. In: *Statistical theory and computational aspects of smoothing*, Seiten 215–230. Springer, 1996.

- [22] Rigby, Robert A, Stasinopoulos, Mikis D, Heller, Gillian Z und De Bastiani, Fernanda: *Distributions for modeling location, scale, and shape: Using GAMLSS in R*. CRC press, 2019.
- [23] Stasinopoulos, Mikis D, Rigby, Robert A, Heller, Gillian Z, Voudouris, Vlasios und De Bastiani, Fernanda: *Flexible regression and smoothing: using GAMLSS in R*. CRC Press, 2017.
- [24] Waller, Matthew A und Fawcett, Stanley E: *Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management*. Journal of Business Logistics, 34(2):77–84, 2013.
- [25] Weiß, Christian H: *An introduction to discrete-valued time series*. John Wiley & Sons, 2018.
- [26] Yang, Ming: *Statistical models for count time series with excess zeros*. 2012. <https://ir.uiowa.edu/etd/3019>.
- [27] Yang, Ming, Zamba, Gideon und Cavanaugh, Joseph: *ZIM: Zero-Inflated Models (ZIM) for Count Time Series with Excess Zeros*, 2018. <https://CRAN.R-project.org/package=ZIM>, R package version 1.1.0.
- [28] Yerpude, Samir und Singhal, Tarun Kumar: *Impact of internet of things (IoT) data on demand forecasting*. Indian Journal of Science and Technology, 10(15):1–5, 2017.
- [29] Zhang, Yao, Wang, Jianxue und Wang, Xifan: *Review on probabilistic forecasting of wind power generation*. Renewable and Sustainable Energy Reviews, 32:255–270, 2014.

Eidesstattliche Erklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Datum, Ort

Hahn Maximilian