Specification Tests for Integer-Valued Autoregressive Processes

Sebastian Schweer, University of Heidelberg

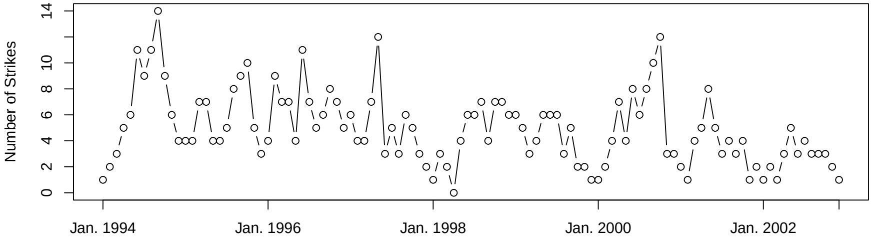
Supported by the *Deutsche Forschungsgemeinschaft*, Grant GRK 1953.

Joint work with Christian Weiß, HSU Hamburg

29.08.2015

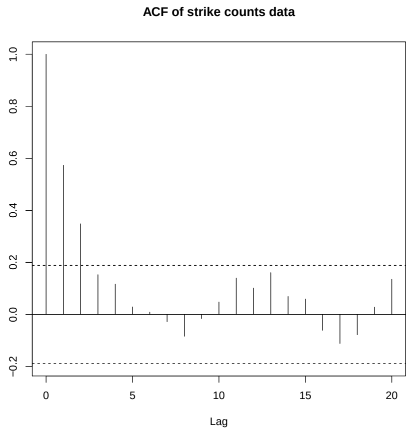
Motivating Example

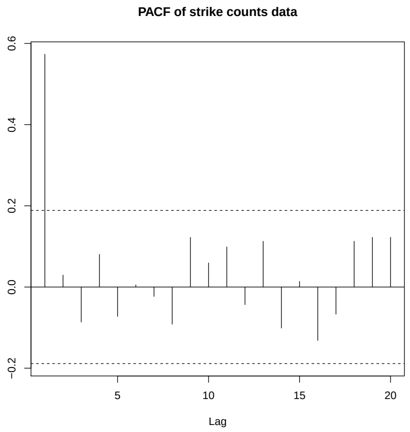
Strike Counts published by the *US Bureau of Labor Statistics*, includes only Strikes with 1000 or more workers being idle:



Motivating Example, ctd.

Strike Counts exhibit Autoregressive Structure:





Count Data Time Series Models

Possible Candidates: Integer-Valued variations of continuous processes.

- INARMA Models
  INAR(1) Process of [Al-Osh and Alzaid, 1987], [McKenzie, 1985],

Count Data Time Series Models

Possible Candidates: Integer-Valued variations of continuous processes.

- INARMA Models
  INAR(1) Process of [Al-Osh and Alzaid, 1987], [McKenzie, 1985],
- INGARCH Models
  INARCH(1) Process of [Ferland et al., 2006]

Overview

1. INAR(1) Processes
   - Definition, Properties
   - Detecting Overdispersion
   - Testing for Time-Reversibility
2. Goodness-of-Fit Testing in General Setting
   - Empirical Probability Generating Function
   - Estimation of the Stationary Distribution
   - Connection to Moment-Based Criteria

INAR(1) Process

Let $(\epsilon_t)_{t\in\mathbb{Z}}$ with range $\mathbb{N}_0$ i.i.d., finite variance, $\alpha \in (0,1)$. If

$$\underbrace{Y_t}_{\text{Population at time } t} = \underbrace{\alpha \circ Y_{t-1}}_{\text{Offspring of Previous Generations}} + \underbrace{\epsilon_t}_{\text{Immigration}} \quad \text{for all } t \in \mathbb{Z},$$

INAR(1) Process

Let $(\epsilon_t)_{t\in\mathbb{Z}}$ with range $\mathbb{N}_0$ i.i.d., finite variance, $\alpha \in (0,1)$. If

$$\underbrace{Y_t}_{\text{Population at time } t} = \underbrace{\alpha \circ Y_{t-1}}_{\text{Offspring of Previous Generations}} + \underbrace{\epsilon_t}_{\text{Immigration}} \quad \text{for all } t \in \mathbb{Z},$$

where

$$\alpha \circ X := \sum_{i=1}^{X} \xi_i,$$

then $(Y_t)_{t\in\mathbb{Z}}$ is an INAR(1) Process.
ξ_i : Bernoulli random variables with $\mathbb{P}(\xi_i = 1) = \alpha$.
All thinning operations are performed independently of each other.

Properties of INAR(1) Processes

For an INAR(1) Process $(Y_t)_{t\in\mathbb{Z}}$ with $\mathbb{E}[Y_0] < \infty$,

- the autocorrelation function is $\rho_Y(k) = \alpha^k$,
- $(Y_t)_{t\in\mathbb{Z}}$ is an α-mixing Markov Chain with exponentially decreasing weights.

Properties of INAR(1) Processes

For an INAR(1) Process $(Y_t)_{t\in\mathbb{Z}}$ with $\mathbb{E}[Y_0] < \infty$,

- the autocorrelation function is $\rho_Y(k) = \alpha^k$,
- $(Y_t)_{t\in\mathbb{Z}}$ is an α-mixing Markov Chain with exponentially decreasing weights.

A popular distributional choice: $\epsilon_0 \sim \text{Poi}(\lambda)$. In this case

- $Y_t \sim \text{Poi}(\lambda/(1-\alpha))$,
- $(Y_t)_{t\in\mathbb{Z}}$ is time-reversible,
- the conditional pgf is given by

$$\text{pgf}_{Y_{t+k}|Y_t}(z) = \left(1-\alpha^k+\alpha^k z\right)^{Y_t} \exp\left(\frac{1-\alpha^k}{1-\alpha}\lambda(z-1)\right)$$

Properties of INAR(1) Processes

For an INAR(1) Process $(Y_t)_{t\in\mathbb{Z}}$ with $\mathbb{E}[Y_0] < \infty$,

- the autocorrelation function is $\rho_Y(k) = \alpha^k$,
- $(Y_t)_{t\in\mathbb{Z}}$ is an α-mixing Markov Chain with exponentially decreasing weights.

A popular distributional choice: $\epsilon_0 \sim \mathrm{Poi}(\lambda)$. In this case

- $Y_t \sim \mathrm{Poi}(\lambda/(1-\alpha))$,
- $(Y_t)_{t\in\mathbb{Z}}$ is time-reversible,
- the conditional pgf is given by

$$\mathrm{pgf}_{Y_{t+k}|Y_t}(z) = \left(1-\alpha^k+\alpha^k z\right)^{Y_t} \exp\left(\frac{1-\alpha^k}{1-\alpha}\lambda(z-1)\right)$$

Goal: Test for

$$\mathbf{H}_0 \ : \ (Y_t)_{t\in\mathbb{Z}} \quad \text{is a Poisson INAR(1) Process.}$$

Index of Dispersion

We introduce the Index of Dispersion I_Y

$$I_Y := \frac{\sigma_Y^2}{\mu_Y} \qquad \text{and} \qquad \hat{I}_Y := \frac{S_Y^2}{\bar{Y}}.$$

Index of Dispersion

We introduce the Index of Dispersion I_Y

$$I_Y := \frac{\sigma_Y^2}{\mu_Y} \quad \text{and} \quad \hat{I}_Y := \frac{S_Y^2}{\bar{Y}}.$$

For Poisson INAR(1) Process, $I_Y = 1$, for Strike Counts $\hat{I}_Y \approx 1.602$.

Index of Dispersion

We introduce the Index of Dispersion I_Y

$$I_Y := \frac{\sigma_Y^2}{\mu_Y} \quad \text{and} \quad \hat{I}_Y := \frac{S_Y^2}{\bar{Y}}.$$

For Poisson INAR(1) Process, $I_Y = 1$, for Strike Counts $\hat{I}_Y \approx 1.602$.

Theorem ([S. and Weiß, 2014])

Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process with $\epsilon_t \sim \mathrm{Poi}(\lambda)$, then

$$\sqrt{T} \cdot (\hat{I}_Y - 1) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ 2\,\frac{1+\alpha^2}{1-\alpha^2}\right) \quad \textit{as} \quad T \longrightarrow \infty.$$

Index of Dispersion

We introduce the Index of Dispersion I_Y

$$I_Y := \frac{\sigma_Y^2}{\mu_Y} \quad \text{and} \quad \hat{I}_Y := \frac{S_Y^2}{\bar{Y}}.$$

For Poisson INAR(1) Process, $I_Y = 1$, for Strike Counts $\hat{I}_Y \approx 1.602$.

Theorem ([S. and Weiß, 2014])

Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process with $\epsilon_t \sim \mathrm{Poi}(\lambda)$, then

$$\sqrt{T}\cdot(\hat{I}_Y - 1) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ 2\,\frac{1+\alpha^2}{1-\alpha^2}\right) \quad \textit{as} \quad T \longrightarrow \infty.$$

Critical value to the level $\beta = 0.05$ is given by ≈ 1.315.
$\Rightarrow$ Rejection of the Poisson INAR(1) Model.

Testing for Time-Reversibility

The Generalized Autocovariance $\beta_Y(k) = \mathbb{E}\left[Y_t^2 Y_{t-k}\right] - \mathbb{E}\left[Y_t Y_{t-k}^2\right]$ and

$$\hat{\beta}_Y(k) = \frac{1}{T-k} \sum_{t=k+1}^{T} \left(Y_t^2 Y_{t-k} - Y_{t-k}^2 Y_t\right)$$

Testing for Time-Reversibility

The Generalized Autocovariance $\beta_Y(k) = \mathbb{E}\left[Y_t^2 Y_{t-k}\right] - \mathbb{E}\left[Y_t Y_{t-k}^2\right]$ and

$$\hat{\beta}_Y(k) = \frac{1}{T-k} \sum_{t=k+1}^{T} \left(Y_t^2 Y_{t-k} - Y_{t-k}^2 Y_t\right)$$

For Poisson INAR(1) Process, $\beta_Y(1) = 0$, for Strike Counts $\hat{\beta}_Y(1) \approx 2.879$.

Testing for Time-Reversibility

The Generalized Autocovariance $\beta_Y(k) = \mathbb{E}\left[Y_t^2 Y_{t-k}\right] - \mathbb{E}\left[Y_t Y_{t-k}^2\right]$ and

$$\hat{\beta}_Y(k) = \frac{1}{T-k} \sum_{t=k+1}^{T} \left(Y_t^2 Y_{t-k} - Y_{t-k}^2 Y_t\right)$$

For Poisson INAR(1) Process, $\beta_Y(1) = 0$, for Strike Counts $\hat{\beta}_Y(1) \approx 2.879$.

Theorem ([S. and Weiß, 2015])

Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process and let $k \in \mathbb{N}$, then

$$\sqrt{T-k} \cdot \hat{\beta}_Y(k) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_k^2\right).$$

where $\sigma_k^2 = \sum_{t\in\mathbb{Z}} \mathbb{E}[(Y_0^2 Y_{-k} - Y_0 Y_{-k}^2)(Y_t^2 Y_{t-k} - Y_t Y_{t-k}^2)]$.

Testing for Time-Reversibility

The Generalized Autocovariance $\beta_Y(k) = \mathbb{E}\left[Y_t^2 Y_{t-k}\right] - \mathbb{E}\left[Y_t Y_{t-k}^2\right]$ and

$$\hat{\beta}_Y(k) = \frac{1}{T-k} \sum_{t=k+1}^{T} \left(Y_t^2 Y_{t-k} - Y_{t-k}^2 Y_t\right)$$

For Poisson INAR(1) Process, $\beta_Y(1) = 0$, for Strike Counts $\hat{\beta}_Y(1) \approx 2.879$.

Theorem ([S. and Weiß, 2015])

Let $(Y_t)_{t\in\mathbb{Z}}$ be a Poisson INAR(1) process and let $k \in \mathbb{N}$, then

$$\sqrt{T-k} \cdot \hat{\beta}_Y(k) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_k^2\right).$$

where $\sigma_k^2 = \sum_{t\in\mathbb{Z}} \mathbb{E}[(Y_0^2 Y_{-k} - Y_0 Y_{-k}^2)(Y_t^2 Y_{t-k} - Y_t Y_{t-k}^2)]$.

Critical value $\approx 2.386 \Rightarrow$ Rejection of the Poisson INAR(1) Model.

A More General Approach

INAR(1) Process:

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t.$$

A More General Approach

INAR(1) Process:

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t.$$

INARCH(1) Process:

$$Y_t \,|\, Y_{t-1}, Y_{t-2}, \cdots \sim \mathrm{Poi}\left(\alpha \cdot Y_{t-1} + \mu\right).$$

A More General Approach

INAR(1) Process:

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t.$$

INARCH(1) Process:

$$Y_t | Y_{t-1}, Y_{t-2}, \dots \sim \text{Poi}\left(\alpha \cdot Y_{t-1} + \mu\right).$$

More General Framework: $(Y_t)_{t\in\mathbb{Z}}$ ergodic Markov Chain on $\mathbb{N}_0$ with

$$\mathbb{E}[Y_t | Y_{t-1}] = \alpha Y_{t-1} + \lambda$$

for some $\alpha \in (0,1)$, $\lambda > 0$: INCLAR(1) Process [Grunwald et al., 2000].
α-mixing with exponentially decreasing weights.

A More General Approach

INAR(1) Process:

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t.$$

INARCH(1) Process:

$$Y_t | Y_{t-1}, Y_{t-2}, \dots \sim \mathrm{Poi}\left(\alpha \cdot Y_{t-1} + \mu\right).$$

More General Framework: $(Y_t)_{t \in \mathbb{Z}}$ ergodic Markov Chain on $\mathbb{N}_0$ with

$$\mathbb{E}[\, Y_t |\, Y_{t-1}] = \alpha Y_{t-1} + \lambda$$

for some $\alpha \in (0,1)$, $\lambda > 0$: INCLAR(1) Process [Grunwald et al., 2000].
α-mixing with exponentially decreasing weights.

Goal: Test for

$$\mathbf{H}_0 \ : \ f \in \{f_\theta | \theta \in \Theta\} \quad \text{against} \quad \mathbf{H}_1 \ : \ f \notin \{f_\theta | \theta \in \Theta\}.$$

Empirical Probability Generating Function

Define the *(empirical) joint probability generating function*

$$\widehat{\psi}_T(u,v) := \frac{1}{T}\sum_{i=1}^{T} u^{y_i} v^{y_{i+1}} \quad \text{and} \quad \psi(u,v;\theta) := \mathbb{E}_\theta\left[u^{Y_0} v^{Y_1}\right].$$

Test Statistic

$$W_{T,a}(y_1,\ldots,y_{T+1};\theta) := T\int_0^1\int_0^1 \left(\widehat{\psi}_T(u,v) - \psi(u,v;\theta)\right)^2 u^a v^a du dv,$$

$a \geq 0$ a weighting factor. Cp. [Meintanis and Karlis, 2014], [Hudecová et al., 2015].

Asymptotic Distribution of the Statistic

Theorem ([S., 2015])

Let $(Y_t)_{t\in\mathbb{Z}}$ be an INCLAR(1) Process with $\mathbb{E}[|Y_0 Y_1|^{2+\xi}] < \infty$ for some $\xi > 0$ and $\psi(u, v; \theta)$ be twice continuously differentiable, and let the series

$$\sum_{k,l=0}^{\infty} kl \frac{\partial}{\partial\theta}\mathbb{P}_{\theta_0}(Y_0 = k, Y_1 = l) \text{ and } \sum_{k,l=0}^{\infty} \frac{\partial^2}{\partial^2\theta}\mathbb{P}_{\theta'}(Y_0 = k, Y_1 = l)$$

converge, where $\theta' \in \Theta$. Let the estimator $\hat{\theta}$ satisfy standard regularity conditions. Then

$$\sqrt{T}\left(\widehat{\psi}_T(u, v) - \psi(u, v; \hat{\theta}_T)\right) \xrightarrow{\mathcal{D}} \Psi_2,$$

a zero mean Gaussian element in $C[0, 1]^2$ with covariance function $\kappa_2(u_1, v_1; u_2, v_2)$.

Sketch of the Proof

- Finite-Dimensional Distributions: α-mixing with exponentially decreasing weights,
- Tightness: Use Moment Condition by showing that

$$
\begin{aligned}
&\mathbb{E}[|X_T(u_2,0) - X_T(u_1,0)|^2] \leq C_1|u_2 - u_1|^2 \\
&\mathbb{E}[|X_T(0,v_2) - X_T(0,v_1)|^2] \leq C_2|v_2 - v_1|^2 \\
&\mathbb{E}[|X_T(u_2,v_2) - X_T(u_2,v_1) - X_T(u_1,v_2) + X_T(u_1,v_1)|^2] \\
&\leq C_3|u_2 - u_1|^2|v_2 - v_1|^2
\end{aligned}
$$

for $X_T := \sqrt{T}\left(\widehat{\psi}_T(u,v) - \psi(u,v;\hat{\theta}_T)\right)$, see [Lachout, 1988].

Estimating the Stationary Distribution

$$\psi(u, v; \theta) = \mathbb{E}_\theta \left[u^{Y_0} v^{Y_1} \right] = \sum_{k,l=0}^{\infty} u^k v^l \pi_\theta(k) f_\theta(k, l).$$

f_θ is known, the stationary distribution π_θ is unknown in general.
$\Rightarrow$ Consider weak convergence on $c_0 := \{x \in \mathbb{R}^{\mathbb{N}} : \lim_k x_k = 0\}$, with norm $\|x\|_{c_0} := \sup_{k \in \mathbb{N}} |x_k|$.

Estimating the Stationary Distribution

$$\psi(u, v; \theta) = \mathbb{E}_\theta \left[u^{Y_0} v^{Y_1} \right] = \sum_{k,l=0}^{\infty} u^k v^l \pi_\theta(k) f_\theta(k, l).$$

f_θ is known, the stationary distribution π_θ is unknown in general.
$\Rightarrow$ Consider weak convergence on $c_0 := \{x \in \mathbb{R}^{\mathbb{N}} : \lim_k x_k = 0\}$, with norm $\|x\|_{c_0} := \sup_{k \in \mathbb{N}} |x_k|$.

Theorem ([S., 2015])

Let $(Y_t)_{t \in \mathbb{Z}}$ be an INCLAR(1) Process with $\mathbb{E}[|Y_0|^{2+\xi}] < \infty$ for some $\xi > 0$. Then there exists a zero mean Gaussian element Ξ of c_0 with

$$\sqrt{T} \left(\frac{1}{T} \sum_{i=1}^{T} \mathbf{1}_{\{Y_i \leq x\}} - \Pi_\theta(x) \right)_{x \in \mathbb{N}_0} \xrightarrow{\mathcal{D}} \Xi.$$

Effect of Simulating the Stationary Distribution

If Stationary Distribution needs to be Simulated:

1. Generate stationary bootstrap data $(Y_1^{**}, \ldots, Y_S^{**})$ with $f_{\theta'}$.
2. Calculate the statistic $\widehat{W}_{S,T,a}(Y_1, \ldots, Y_{T+1}; \theta')$ by replacing $\pi_{\theta'}(k)$ with the estimator $\frac{1}{S}\sum_{i=1}^{S} \mathbf{1}_{\{Y_i^{**}=k\}}$ for all $k \in \mathbb{N}_0$.

Effect of Simulating the Stationary Distribution

If Stationary Distribution needs to be Simulated:

1. Generate stationary bootstrap data $(Y_1^{**}, \ldots, Y_S^{**})$ with $f_{\theta'}$.
2. Calculate the statistic $\widehat{W}_{S,T,a}(Y_1, \ldots, Y_{T+1}; \theta')$ by replacing $\pi_{\theta'}(k)$ with the estimator $\frac{1}{S}\sum_{i=1}^{S} \mathbf{1}_{\{Y_i^{**}=k\}}$ for all $k \in \mathbb{N}_0$.

Lemma

Let $(Y_t)_{t\in\mathbb{Z}}$ be an INCLAR(1) Process with $\mathbb{E}_\theta[Y_0^{4+\xi}] < \infty$, where $\xi > 0$. Then for large T it holds almost surely that

$$\mathbb{E}\left[\left(W_{T,a}(Y_1, \ldots Y_{T+1}; \theta) - \widehat{W}_{S,T,a}(Y_1, \ldots, Y_{T+1}; \theta)\right)^2\right] = O\left(\frac{T^2}{S}\right).$$

Connection to the Index of Dispersion

For a RV X with pgf G, the moments of X are related to G:
$\mathbb{E}[X] = G'(1^-)$. This corresponds to higher values of a in

$$W_{T,a}(y_1, \ldots, y_{T+1}; \theta) := T \int_0^1 \int_0^1 \left(\widehat{\psi}_T(u,v) - \psi(u,v;\theta) \right)^2 u^a v^a du dv.$$

Connection to the Index of Dispersion

For a RV X with pgf G, the moments of X are related to G: $\mathbb{E}[X] = G'(1^-)$. This corresponds to higher values of a in

$$W_{T,a}(y_1, \ldots, y_{T+1}; \theta) := T \int_0^1 \int_0^1 \left(\widehat{\psi}_T(u,v) - \psi(u,v;\theta) \right)^2 u^a v^a du dv.$$

Theorem

Let $(Y_t)_{t \in \mathbb{Z}}$ be a Poisson INAR(1) process. Then

$$\lim_{a \to \infty} a^6 W_{T,a} = 14 \left(\frac{1}{T} \sum_{i=1}^{T} Y_i \right)^2 \cdot \left(\sqrt{T}(\hat{I}_Y - 1) \right)^2 + O\left(\frac{1}{T} \right)$$

Connection to Generalized Autocovariance

For testing Time-Reversibility, define

$$V_{T,a}(Y_1, \ldots, Y_{T+1}) := T \int_0^1 \int_0^1 \left(\widehat{\psi}_T(u,v) - \widehat{\psi}_T(v,u) \right)^2 u^a v^a du dv.$$

Connection to Generalized Autocovariance

For testing Time-Reversibility, define

$$V_{T,a}(Y_1,\ldots,Y_{T+1}) := T\int_0^1\int_0^1\left(\widehat{\psi}_T(u,v)-\widehat{\psi}_T(v,u)\right)^2 u^a v^a du dv.$$

Theorem

Let $(Y_t)_{t\in\mathbb{Z}}$ be a time-reversible INCLAR(1) process and let all moments of Y_0 exist. Let $\frac{a^4}{T}\to 0$ as $a\to\infty$. Then

$$\lim_{a,T\to\infty} a^8 V_{T,a}(Y_1,\ldots,Y_{T+1}) = \lim_{T\to\infty} 6\left(\sqrt{T}\cdot\hat{\beta}_T(1)\right)^2.$$

Application of the Test

For the Strike Count Data:

Model	a	statistic	critical value	p-value
INARCH(1)	0	0.00492	0.03347	0.512
	2	0.00117	0.00463	0.321
	6	0.00028	0.00076	0.191
	8	0.00002	0.00037	0.689
	10	0.00015	0.00020	0.084
Poi(λ) INAR(1)	0	0.01886	0.01535	0.033
	2	0.00501	0.00282	0.011
	10	0.00021	0.00011	0.009

Questions for Future Research

- Extensions to More Sophisticated Structures, [Fokianos and Neumann, 2013]
- Circumventing Necessity to Estimate Stationary Distribution?
- Criteria for Count Data Time Series Models vs. "Continuous Models"? [Bisaglia et al., 2015]

Bibliography

Al-Osh, M. A. and Alzaid, A. A. (1987).
First-order integer-valued autoregressive INAR(1) processes.
Journal of Time Series Analysis, 8(3):261–275.

Bisaglia, L., Gerolimetto, M., et al. (2015).
Forecasting integer autoregressive processes of order 1: are simple ar competitive?
Economics Bulletin, 35(3):1652–1660.

Ferland, R., Latour, A., and Oraichi, D. (2006).
Integer-valued GARCH process.
Journal of Time Series Analysis, 27(6):923–942.

Fokianos, K. and Neumann, M. H. (2013).
A goodness-of-fit test for Poisson count processes.
Electronic Journal of Statistics, 7:793–819.

Grunwald, G. K., Hyndman, R. J., Tedesco, L., and Tweedie, R. L. (2000).
Theory & Methods: Non-gaussian conditional linear AR(1) models.
Australian & New Zealand Journal of Statistics, 42(4):479–495.

Hudecová, Š., Hušková, M., and Meintanis, S. G. (2015).
Tests for time series of counts based on the probability-generating function.
Statistics, 49(2):316–337.

Bibliography

Lachout, P. (1988).
Billingsley-type tightness criteria for multiparameter stochastic processes.
Kybernetika, 24(5):363–371.

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

Meintanis, S. G. and Karlis, D. (2014).
Validation tests for the innovation distribution in INAR time series models.
Computational Statistics, 29(5):1221–1241.

Schweer, S. (2015).
A goodness-of-fit test for integer-valued autoregressive processes.
Journal of Time Series Analysis, to appear.

Schweer, S. and Weiß, C. H. (2014).
Compound Poisson INAR (1) processes: Stochastic properties and testing for overdispersion.
Computational Statistics and Data Analysis, 77(9):267–284.

Schweer, S. and Weiß, C. H. (2015).
Testing for Poisson arrivals in INAR(1) processes via time-reversibility.
Submitted.