MATH STAT

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

Binomial Models for Count Data Time Series with a Finite Range

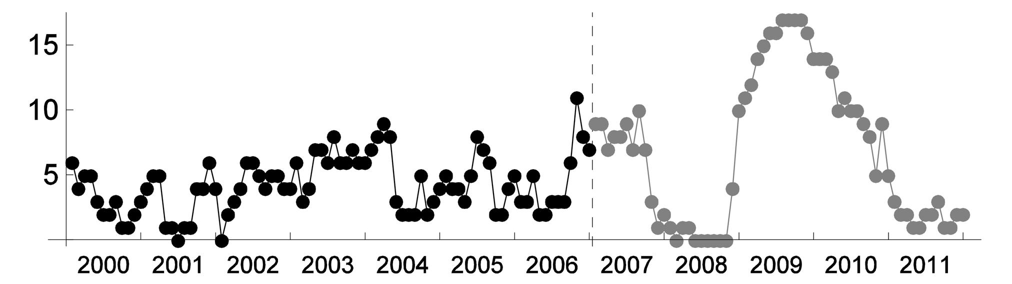
Christian H. Weiß

Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany; **weissc@hsu-hh.de**.

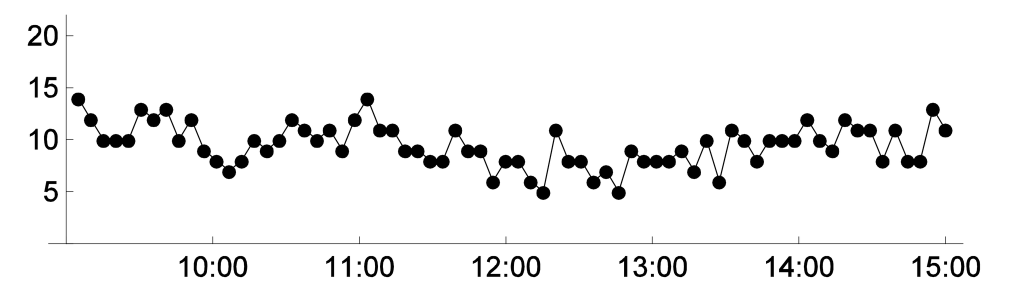
Introduction

The analysis and modelling of time series of counts have become a popular research area during the last decades, but most efforts concentrated on the case of the infinite range $\mathbb{N}_0 = \{0,1,\ldots\}$. Meanwhile, a number of real applications also demonstrated the relevance of models and approaches for time series of counts having a finite range of the form $\{0,\ldots,n\}$ with an $n \in \mathbb{N}$.

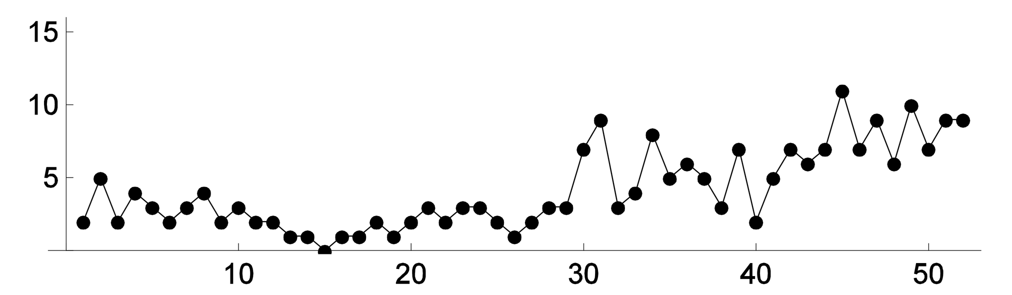
Real-data examples:



Monthly counts of Euro countries (among $n = 17$ member states) having stable prices [WK14].



Number of traded securities companies among $n = 22$ such companies at Korea stock market (5-min, Feb. 8, 2011) [WK13b].



Weekly number of districts in Germany (2011, $n = 38$ districts) with infections by hantavirus [WP14].

McKenzie's Binomial AR(1) Model

This model is based on the *binomial thinning* operation "∘" by [SvH79]: $\alpha \circ X := \sum_{i=1}^{X} Y_i$, where the Y_i are i.i.d. Bernoulli random variables with success probability α, which are also independent of the count data random variable X.

Definition: Let $\pi \in (0;1)$ and $\rho \in (\max\{-\frac{\pi}{1-\pi}, -\frac{1-\pi}{\pi}\};1)$. Define $\beta := \pi(1-\rho)$ and $\alpha := \beta + \rho$. Fix $n \in \mathbb{N}$.

The process $(X_t)_{\mathbb{N}_0}$, defined by the recursion

$$X_t = \alpha \circ X_{t-1} + \beta \circ (n - X_{t-1}) \quad \text{for } t \geq 1,$$

where all thinnings are performed independently of each other and where the thinnings at time t are independent of $(X_s)_{s<t}$, is referred to as a *binomial AR(1) process*. [McK85]

Stochastic Properties

- Stationary, ergodic Markov chain with marginal distribution $\text{Bin}(n,\pi)$, transition probabilities $p_{k|l} := P(X_t = k | X_{t-1} = l)$ given by

$\sum_{m=\max\{0,k+l-n\}}^{\min\{k,l\}} \binom{l}{m}\binom{n-l}{k-m} \alpha^m (1-\alpha)^{l-m} \beta^{k-m} (1-\beta)^{n-l-k+m}$;

- autocorrelation function $\rho(k) = \rho^k$ (AR(1)-like);
- explicit formulae for conditional moments, higher-order joint moments and cumulants, and many more. [W09a,WK13a]

Approaches for Parameter Estimation

A number of approaches have been investigated (asymptotics, finite-sample performance):

- maximum likelihood [WK13b, WP12, WP14];
- conditional least squares [CL10, WP12,WK13a];
- method of moments [WK13b];
- squared differences [WK13a,b].

Diagnostic Tools

The dispersion behavior of a count data random variable X with range $\{0,\ldots,n\}$ is measured by the *binomial index of dispersion*, which is a function of n, mean μ and variance σ^2:

$$I := I(n,\mu,\sigma^2) := \frac{n\,\sigma^2}{\mu(n-\mu)}.$$

$I = 1$ for $X \sim \text{Bin}(n,\pi)$, and a distribution satisfying $I > 1$ is said to exhibit *extra-binomial variation*. The empirical version of I, denoted as $\hat{I}$, is used as a diagnostic tool. For an underlying binomial AR(1) process, it holds [WK14]:

$$\sqrt{T}(\hat{I} - 1) \xrightarrow{D} \text{N}\left(0,\ 2\left(1-\tfrac{1}{n}\right)\tfrac{1+\rho^2}{1-\rho^2}\right).$$

[WK14] utilize this result for testing for extra-binomial variation (→ beta-binomial or density dependent binomial AR(1) model).

To test for the whole marginal distribution, the following result is useful. Let N_i denote the number of X_t equal to i, and let $p_0,\ldots,p_n$ be the marginal $\text{Bin}(n,\pi)$-probabilities. [W09a] showed that for *Pearson's χ^2-statistic* X, we have

$$\text{X} := \sum_{i=0}^{n} \frac{(N_i - T\cdot p_i)^2}{T\cdot p_i} \xrightarrow{D} \sum_{j=1}^{n} \frac{1+\rho^j}{1-\rho^j}\cdot Z_j^2,$$

where $Z_1,\ldots,Z_n$ are i.i.d. N(0,1).

Further Thinning-based Models

Binomial AR(p) Model

Higher-order autoregressions are obtained through a probabilistic mixture approach [W09b]:

$$X_t = \sum_{k=1}^{p} D_{t,k}\cdot\left(\alpha \circ X_{t-k} + \beta \circ (n - X_{t-k})\right)$$

$$\text{with} \quad (D_{t,1},\ldots,D_{t,p}) \sim \text{Mult}(1;\phi_1,\ldots,\phi_k).$$

Together with appropriate independence assumptions, the typical $\text{AR}(p)$-like autocorrelation structure is observed, while the marginals still follow the $\text{Bin}(n,\pi)$-distribution.

Beta-Binomial AR(1) Model

[WK14] introduce an additional dispersion parameter $\phi \in (0;1)$ and assume beta-distributed thinning parameters, i.e., *beta-binomial thinnings*:

$$X_t = \alpha_\phi \circ X_{t-1} + \beta_\phi \circ (n - X_{t-1}) \quad \text{for } t \geq 1.$$

Still $\rho(k) = \rho^k$, but now X_t exhibits extra-binomial variation:

$$I = 1 + \frac{(n-1)\left(1-2\pi(1-\pi)(1-\rho)\right)}{\left(\frac{1}{\phi}-1\right)(1+\rho)+\left(1-2\pi(1-\pi)(1-\rho)\right)} \quad \in (1;n).$$

Density Dependent Binomial AR(1) Models

[WP14] suggest the model parameters (π,ρ) and (α,β) at time t to depend on the process up to that time t through the density X_{t-1}/n. If, e.g.,

ρ is fixed but $\pi_t := a + b\cdot X_{t-1}/n$ with $a, a+b \in (0;1)$,

then

$$I = \frac{1+\rho}{1+\rho-2\left(1-\frac{1}{n}\right)\rho b-\left(1-\frac{1}{n}\right)(1-\rho)b^2},$$

which might be both <1 or >1. This includes the binomial INARCH(1) model (see below) as the boundary case $\rho \to 0$.

Bivariate Binomial AR(1) Model

Defining the thinning operation "⊗" based on the *bivariate binomial distribution of type II* [MO85], [SWSP14] propose a model for bivariate and cross-correlated binomial counts:

$$\boldsymbol{X}_t = (\alpha_1,\alpha_2,\phi_\alpha) \otimes \boldsymbol{X}_{t-1} + (\beta_1,\beta_2,\phi_\beta) \otimes (\boldsymbol{n} - \boldsymbol{X}_{t-1}),$$

where the cross-dependence parameters ϕ_α, ϕ_β might also be negative, i.e., both positive and negative cross-correlation can be generated. The marginals behave like univariate binomial AR(1) models.

Binomial INGARCH Models

These models are motivated by the standard *INGARCH models* [FLO06], which constitute an integer-valued counterpart to the continuous GARCH models.

Binomial INARCH(1) Model

Conditioned on the previous count X_{t-1}, [WK14] assume

$$X_t \sim \text{Bin}(n, a + b\cdot X_{t-1}/n) \quad \text{with } a, a+b \in (0;1),$$

leading to $\rho(k) = b^k$ and extra-binomial variation: $I = \frac{1}{1-\left(1-\frac{1}{n}\right)b^2}$.

Bivariate Binomial INARCH(1) Model

[SWSP14] extended the binomial INARCH(1) model to the bivariate case by assuming

$$\boldsymbol{X}_t \sim \text{BVB}_{\text{II}}\,(n_1,n_2; a_1 + b_1\cdot X_{t-1,1}/n_1, a_2 + b_2\cdot X_{t-1,2}/n_2, \phi).$$

Again, both positive and negative cross-correlation ($\phi \gtrless 0$) can be generated. The marginals behave like univariate binomial INARCH (1) models.

Ongoing Research Activities

- Diagnostic tools concerning the marginal distribution and the autocorrelation structure;
- binomial AR(1) Model with self-exciting threshold;
- higher-order binomial INGARCH models.

References:

[CL10] Cui Y, Lund R (2010) Inference in binomial AR(1) models. *Statist. Probab. Lett.* 80, 1985–1990.

[FLO06] Ferland R, Latour A, Oraichi D (2006) Integer-valued GARCH processes. *J. Time Ser. Anal.* 27, 923–942.

[McK85] McKenzie E (1985) Some simple models for discrete variate time series. *Water Resour. Bull.* 21, 645–650.

[MO85] Marshall AW, Olkin I (1985) A family of bivariate distributions generated by the bivariate Bernoulli distribution. *J. Amer. Statist. Assoc.* 80, 332–338.

[SWSP14] Scotto MG, Weiß CH, Silva ME, Pereira I (2014) Bivariate binomial autoregressive models. *J. Multiv. Anal.* 125, 233–251.

[SvH79] Steutel FW, van Harn K (1979) Discrete analogues of self-decomposability and stability, *Ann. Prob.* 7, 893–899.

[W09a] Weiß CH (2009) Monitoring correlated processes with binomial marginals, *J. Appl. Stat.* 36, 399–414.

[W09b] Weiß CH (2009) A new class of autoregressive models for time series of binomial counts. *Comm. Statist. Theory Methods* 38, 447–460.

[WK13a] Weiß CH, Kim HY (2013) Binomial AR(1) processes: moments, cumulants, and estimation. *Statistics* 47, 494–510.

[WK13b] Weiß CH, Kim HY (2013) Parameter estimation for binomial AR(1) models with applications in finance and industry. *Statist. Papers* 54, 563–590.

[WK14] Weiß CH, Kim HY (2014) Diagnosing and modelling extra-binomial variation for time-dependent counts. *Appl. Stoch. Models Business Industry*, to appear.

[WP12] Weiß CH, Pollett PK (2012) Chain binomial models and binomial autoregressive processes. *Biometrics* 68, 815–824.

[WP14] Weiß CH, Pollett PK (2014) Binomial autoregressive processes with density dependent thinning. *J. Time Ser. Anal.* 35, 115–132.