Diagnostic Tests for Binomial AR(1) Processes

Christian H. Weiß

Department of Mathematics & Statistics,
Helmut Schmidt University, Hamburg

Hee-Young Kim
Department of Informational Statistics,
Korea University Sejong

HELMUT SCHMIDT
UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH
STAT

Introduction

Motivation & Outline

Introduction

Popular count-data counterpart to conventional AR(1) model:

INAR(1) model by McKenzie (1985).

Let $(\epsilon_t)_{\mathbb{Z}}$ be i.i.d. with range $\mathbb{N}_0 = \{0, 1, \ldots\}$.

Let "$\alpha\circ$" be **binomial thinning** operator with $\alpha \in (0; 1)$,

i. e., $\alpha \circ X \,|X \sim \text{Bin}(X, \alpha)$ (Steutel & van Harn, 1979).

$(X_t)_{\mathbb{Z}}$ referred to as **INAR(1) process** if

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

together with appropriate independence assumptions.

Problem: Not applicable for bounded range $\{0, \ldots, n\}$.

Introduction

AR(1)-like model for bounded counts with range $\{0, \ldots, n\}$:
binomial AR(1) model by McKenzie (1985).

Idea: Replaces innovation term by additional thinning,

$$X_t \;=\; \alpha \circ X_{t-1} \;+\; \beta \circ (n - X_{t-1}),$$

where $\beta := \pi \cdot (1 - \rho)$ and $\alpha := \beta + \rho$ with
$\pi \in (0;1)$ and $\rho \in \Big(\max\{-\frac{\pi}{1-\pi}, -\frac{1-\pi}{\pi}\} \;;\; 1\Big)$.

Leads to **binomial** marginal distribution, $\mathrm{Bin}(n, \pi)$,
with **AR(1)**-like autocorrelation function, $\rho(k) = \rho^k$.

Introduction

Some further properties:

- Stationary, ergodic and ϕ-mixing Markov chain.
- 1-step transition probabilities being given by

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

- Conditional mean and variance

$$\begin{aligned} E[X_t \mid X_{t-1}] &= \rho\cdot X_{t-1} \; + \; n\beta, \\ V[X_t \mid X_{t-1}] &= \rho(1-\rho)\,(1-2\pi)\cdot X_{t-1} \; + \; n\beta(1-\beta). \end{aligned}$$

Introduction

Real applications of binomial AR(1) model:

- Utilization of computer room with $n = 15$ workstations over time (1-min intervals) (Weiß, 2009a), or
- number of securities companies traded at KOSPI market (5-min intervals; $n = 22$ such companies) (W. & Kim, 2013).

Sometimes, essential **model assumptions** violated, e. g.,

- no binomial marginal but extra-binomial variation, like in **beta-binomial AR(1)** model (Weiß & Kim, 2014),
- or higher-order dependence structure, like in **binomial AR(p)** model (Weiß, 2009b).

Introduction

Aim: Identify situations, where

more sophisticated models than binomial AR(1) required.

Outline: Diagnostic tests concerning . . .

- marginal distribution,
  - overall goodness-of-fit test,
  - test for uncovering extra-binomial variation;
- autocorrelation structure, tests based on the sample (P)ACF.

HELMUT SCHMIDT
UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH
STAT

Goodness-of-Fit Tests
for Binomial AR(1) Model

Marginal Distribution

Goodness-of-Fit w.r.t. Marginal Distribution

Null hypothesis: $X_1, \ldots, X_T$ from binomial AR(1) process.

Abbreviate Bin(n, π)-marginal PMF by $\boldsymbol{p} = (p_0, \ldots, p_n)^\top$.

Let N_i be number of X_t equal to $i = 0, \ldots, n$,

relative frequencies $\hat{\boldsymbol{p}} := \hat{\boldsymbol{p}}(T) = \frac{1}{T}(N_0, \ldots, N_n)^\top$.

Classical **Pearson statistic**

$$\mathrm{I}_1 \;:=\; \sum_{i=0}^{n} \frac{(N_i - Tp_i)^2}{Tp_i} \;=\; T\,\big(\hat{\boldsymbol{p}} - \boldsymbol{p}\big)^\top \mathbf{D}^{-1}\,\big(\hat{\boldsymbol{p}} - \boldsymbol{p}\big) \quad \text{with } \mathbf{D} = \mathrm{diag}(\boldsymbol{p})$$

asymptotically follows quadratic form distribution under H_0,

$$\mathrm{I}_1 \quad \xrightarrow{D} \quad \sum_{j=1}^{n} \frac{1 + \rho^j}{1 - \rho^j} \cdot Z_j^2 \quad \text{for } T \to \infty. \qquad \text{(Weiß, 2009a)}$$

Goodness-of-Fit w.r.t. Marginal Distribution

Problem: Asymptotic approximation for I_1 often bad, because outer categories with low expected frequencies.

Solution: Following Cochran's rule, construct categories

$$\{0,\ldots,k\},\ \{k+1\},\ldots,\ \{n-l-1\},\ \{n-l,\ldots,n\}.$$

Define matrix $\mathbf{A}^{(k,l)} \in \{0,1\}^{(n-l-k+1)\times(n+1)}$ as

$$\mathbf{A}^{(k,l)} = \left(\underbrace{\begin{matrix}1 & \cdots & 1\\ & & \\ & 0 & \end{matrix}}_{k}\middle|\underbrace{\begin{matrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{matrix}}_{n-l-k+1}\middle|\underbrace{\begin{matrix} & 0 & \\ & & \\ 1 & \cdots & 1\end{matrix}}_{l}\right)$$

such that

$$\mathbf{A}^{(k,l)}\,\boldsymbol{y} = \big(y_0+\ldots+y_k,\ y_{k+1},\ \ldots,\ y_{n-l-1},\ y_{n-l}+\ldots+y_n\big)^{\top}.$$

Goodness-of-Fit w.r.t. Marginal Distribution

Theorem: (Kim & Weiß, 2015)

Let $\mathbf{P}^\top = (p_{i|j})$ denote binomial AR(1)'s transition matrix,

define $\boldsymbol{\Sigma} = 2\,(\mathbf{I} - \mathbf{P}^\top + \boldsymbol{p}\,\mathbf{1}^\top)^{-1}\,\mathbf{D} - \mathbf{D} - \boldsymbol{p}\boldsymbol{p}^\top$.

Define $\boldsymbol{\Sigma}^{(k,l)} := \mathbf{A}^{(k,l)}\,\boldsymbol{\Sigma}\,\left[\mathbf{A}^{(k,l)}\right]^\top \in \mathbb{R}^{(n-k-l+1)\times(n-k-l+1)}$

with non-zero eigenvalues $\lambda_1^{(k,l)}, \ldots, \lambda_r^{(k,l)}$.

Then **combined Pearson statistic**

$$\mathrm{I}_1^{(k,l)} = T\left[\mathbf{A}^{(k,l)}\left(\widehat{\boldsymbol{p}} - \boldsymbol{p}\right)\right]^\top \mathrm{diag}(\mathbf{A}^{(k,l)}\,\boldsymbol{p})^{-1}\,\mathbf{A}^{(k,l)}\left(\widehat{\boldsymbol{p}} - \boldsymbol{p}\right)$$

satisfies

$$\mathrm{I}_1^{(k,l)} \quad \xrightarrow{D} \quad \sum_{j=1}^{r} \lambda_j^{(k,l)} \cdot Z_j^2.$$

Goodness-of-Fit w.r.t. Marginal Distribution

Finite-sample performance: (10,000 replications)

$n = 10, \pi = 0.1$

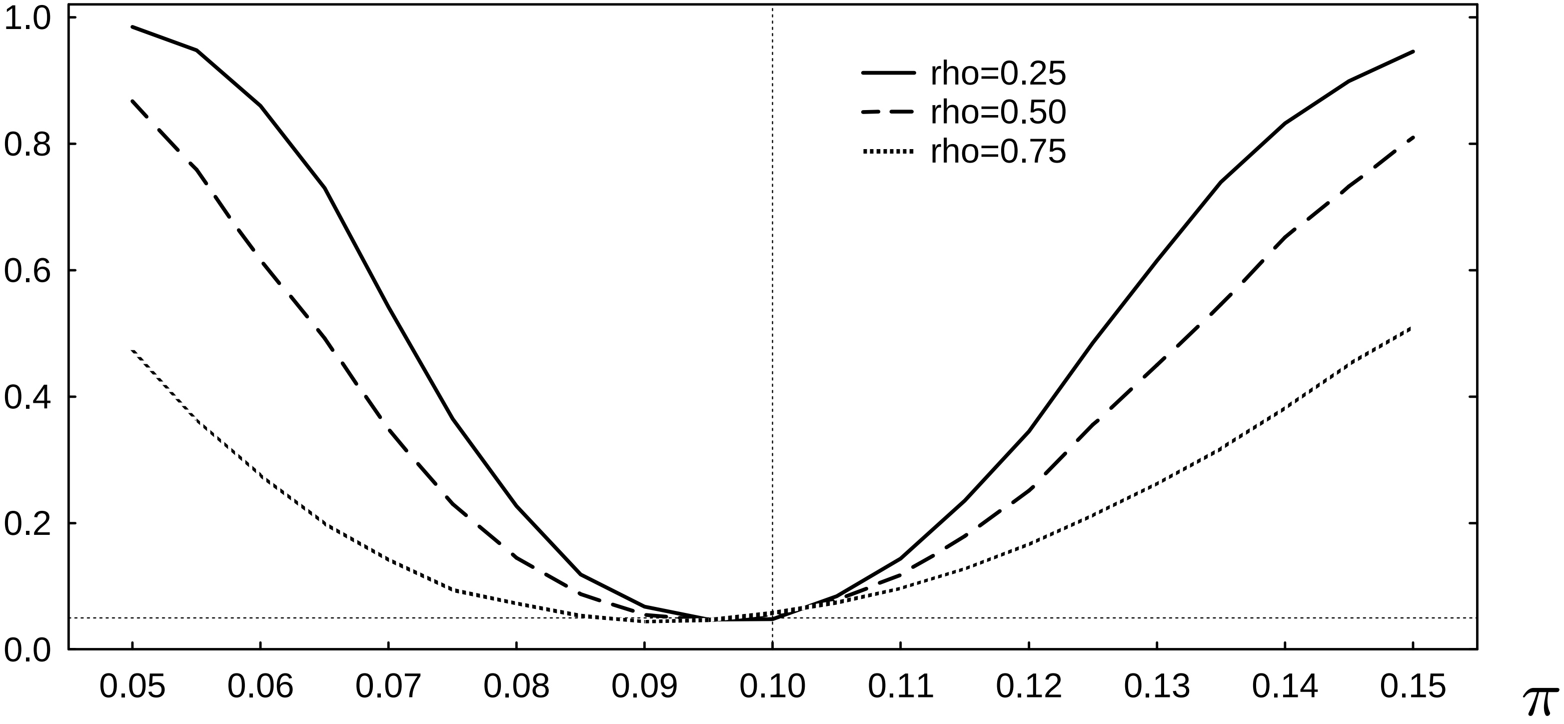
T	k	l	$\rho = -0.10$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$
100	0	6	5.09	5.50	5.97	5.72
250	0	6	4.91	4.98	5.39	5.35
500	0	5	5.64	5.61	5.34	6.17
1000	0	5	5.36	5.42	5.58	5.88

$n = 10, \pi = 0.5$

T	k	l	$\rho = -0.50$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$
100	2	2	5.28	4.70	4.67	5.38
250	1	1	5.28	5.33	5.51	6.15
500	1	1	5.00	4.67	5.17	5.70
1000	0	0	5.79	5.41	6.01	6.29

Goodness-of-Fit w.r.t. Marginal Distribution

Finite-sample performance: (10,000 replications)



Power with regard to changes in π.

Null model: $(n, \pi_0;\ T;\ k, l) = (10, 0.1;\ 100;\ 0, 6)$.

Goodness-of-Fit w.r.t. Marginal Distribution

Common type of violating binomial marginal:

extra-binomial variation, i. e., binomial index of dispersion

$$I_{\mathsf{d}} \ := \ I_{\mathsf{d}}(n,\mu,\sigma^2) \ := \ \frac{n\sigma^2}{\mu(n-\mu)} \quad \in \ (0;\infty)$$

becomes > 1 (binomial distribution: $I_{\mathsf{d}} = 1$).

Specialized **test statistic** for this purpose:

$$\hat{I}_{\mathsf{d}} \ := \ \frac{1}{T}\cdot\sum_{t=1}^{T}\frac{n\,(X_t-\bar{X})^2}{\bar{X}\,(n-\bar{X})} \ = \ \frac{n\,(\frac{1}{T}\sum_{t=1}^{T}X_t^2 \ -\bar{X}^2)}{\bar{X}\,(n-\bar{X})}.$$

Theorem (Weiß & Kim, 2014): Under H_0,

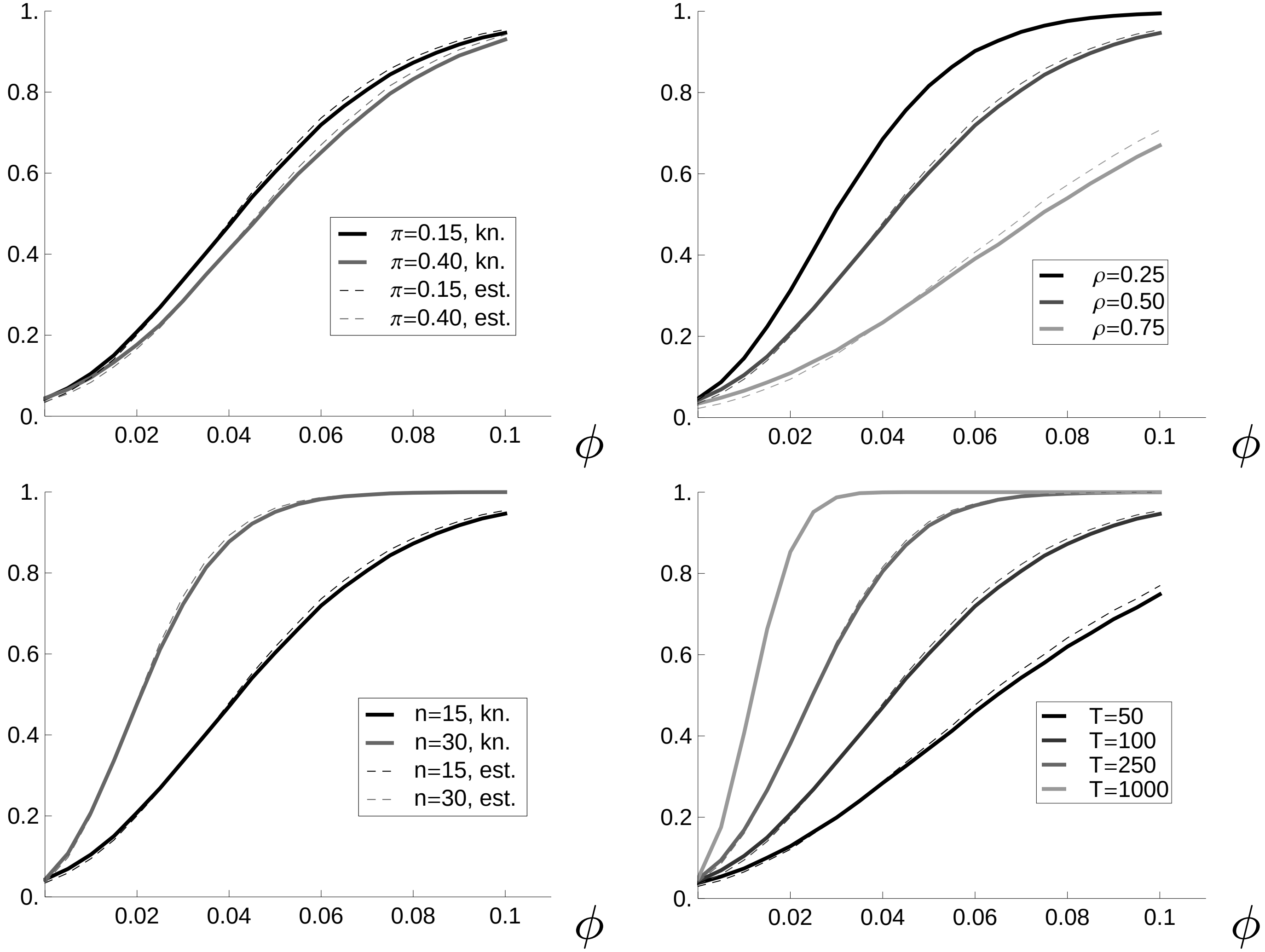
$$\sqrt{T}\,(\hat{I}_{\mathsf{d}}-1) \ \overset{D}{\to} \ N(0,\ \sigma_{\mathsf{d}}^2), \qquad \text{where } \sigma_{\mathsf{d}}^2 \ := \ 2\,(1-\tfrac{1}{n})\,\tfrac{1+\rho^2}{1-\rho^2}.$$

Goodness-of-Fit w.r.t. Marginal Distribution

Finite-sample performance: (100,000 replications)

Empirical power for alternative beta-bin. AR(1) with dispersion parameter ϕ

(W. & Kim, 2014)



Goodness-of-Fit Tests for Binomial AR(1) Model

Autocorrelation Structure

Goodness-of-Fit w.r.t. Autocorrelation Structure

Null hypothesis: $X_1, \ldots, X_T$ from binomial AR(1) process.

We consider tests based on sample (P)ACF.

Note that binomial AR(1) is neither Gaussian nor linear,
so usual **Bartlett's formula** for sample ACF does **not hold!**

Our results derived by applying Romano & Thombs (1996).

Notations:

$$\boldsymbol{\gamma} := \big(\gamma(0), \ldots, \gamma(K)\big)^{\top}, \quad \hat{\boldsymbol{\gamma}} := \big(\hat{\gamma}(0), \ldots, \hat{\gamma}(K)\big)^{\top},$$

$$\boldsymbol{\rho} := \big(\rho(1), \ldots, \rho(K)\big)^{\top}, \quad \hat{\boldsymbol{\rho}} := \big(\hat{\rho}(1), \ldots, \hat{\rho}(K)\big)^{\top},$$

$$\boldsymbol{\rho}_{\mathrm{p}} := \big(\rho_{\mathrm{p}}(2), \ldots, \rho_{\mathrm{p}}(K)\big)^{\top}, \quad \hat{\boldsymbol{\rho}}_{\mathrm{p}} := \big(\hat{\rho}_{\mathrm{p}}(2), \ldots, \hat{\rho}_{\mathrm{p}}(K)\big)^{\top}.$$

Goodness-of-Fit w.r.t. Autocorrelation Structure

Theorem: (Kim & Weiß, 2015)

We have

$$\sqrt{T}\left(\hat{\gamma}-\gamma\right) \overset{D}{\to} N(\mathbf{0},\ \mathbf{\Sigma}'),$$

where $(1+i, 1+j)$-th element of $\mathbf{\Sigma}'$ for $0 \leq i \leq j \leq K$ given by

$$\begin{aligned}\sigma'_{1+i,1+j} \;=\;& n\pi(1-\pi)\Big((1-2\pi)^2(j-i+\tfrac{1+\rho}{1-\rho})\cdot\rho^j\\ &+n\pi(1-\pi)(j-i+\tfrac{1+\rho^2}{1-\rho^2})\cdot\rho^{j-i}\\ &+(n-2)\pi(1-\pi)(j+i+\tfrac{1+\rho^2}{1-\rho^2})\cdot\rho^{j+i}\Big).\end{aligned}$$

Goodness-of-Fit w.r.t. Autocorrelation Structure

Theorem: (Kim & Weiß, 2015)

We have

$$\sqrt{T}\left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\right) \overset{D}{\to} N(\mathbf{0},\ \boldsymbol{\Sigma}''),$$

where (i,j)-th element of $\boldsymbol{\Sigma}''$ for $1 \leq i \leq j \leq K$ given by

$$\begin{aligned} \sigma''_{i,j} \ &= \ \rho^{j-i} \cdot \Big(\big(j + \tfrac{1+\rho^2}{1-\rho^2}\big) \cdot (1-\rho^{2i}) - i \cdot (1+\rho^{2i})\Big) \\ &\quad + \tfrac{(1-2\pi)^2}{n\pi(1-\pi)} \cdot \rho^{j} \cdot \Big(\big(j + \tfrac{1+\rho}{1-\rho}\big) \cdot (1-\rho^{i}) - i \cdot (1+\rho^{i})\Big). \end{aligned}$$

So variances

$$\sigma''_{i,i} \ = \ \underbrace{\frac{1+\rho^2}{1-\rho^2} \cdot (1-\rho^{2i}) - 2i\rho^{2i}}_{\text{Bartlett}} + \tfrac{(1-2\pi)^2}{n\pi(1-\pi)} \cdot \rho^{i} \cdot \big(\tfrac{1+\rho}{1-\rho}(1-\rho^{i}) - 2i\rho^{i}\big).$$

Goodness-of-Fit w.r.t. Autocorrelation Structure

Theorem: (Kim & Weiß, 2015)

We have

$$\sqrt{T}\,\hat{\boldsymbol{\rho}}_{\mathrm{p}} \overset{D}{\to} N(\mathbf{0},\ \boldsymbol{\Sigma}'''),$$

where

$$\boldsymbol{\Sigma}''' = \mathrm{diag}\Big(1+\frac{(1-2\pi)^2}{n\pi(1-\pi)}\cdot\frac{\rho^2}{(1+\rho)^2},\ldots,1+\frac{(1-2\pi)^2}{n\pi(1-\pi)}\cdot\frac{\rho^K}{(1+\rho)^2}\Big).$$

So again sample PACFS asymptotically independent.

But difference to Bartlett:

$$1+\frac{(1-2\pi)^2}{n\pi(1-\pi)}\cdot\frac{\rho^i}{(1+\rho)^2} \qquad \text{instead of just} \quad 1.$$

Goodness-of-Fit w.r.t. Autocorrelation Structure

Finite-sample performance: (10,000 replications)

Size and power for Box-type statistics,

$$Q_{\mathsf{BP}} = T\sum_{i=2}^{K}\frac{\widehat{\rho}_{\mathsf{p}}^{2}(i)}{(\sigma_{i,i}''')^{2}},\qquad Q_{\mathsf{BL}} = T(T+2)\sum_{i=2}^{K}(T-i)^{-1}\frac{\widehat{\rho}_{\mathsf{p}}^{2}(i)}{(\sigma_{i,i}''')^{2}},$$

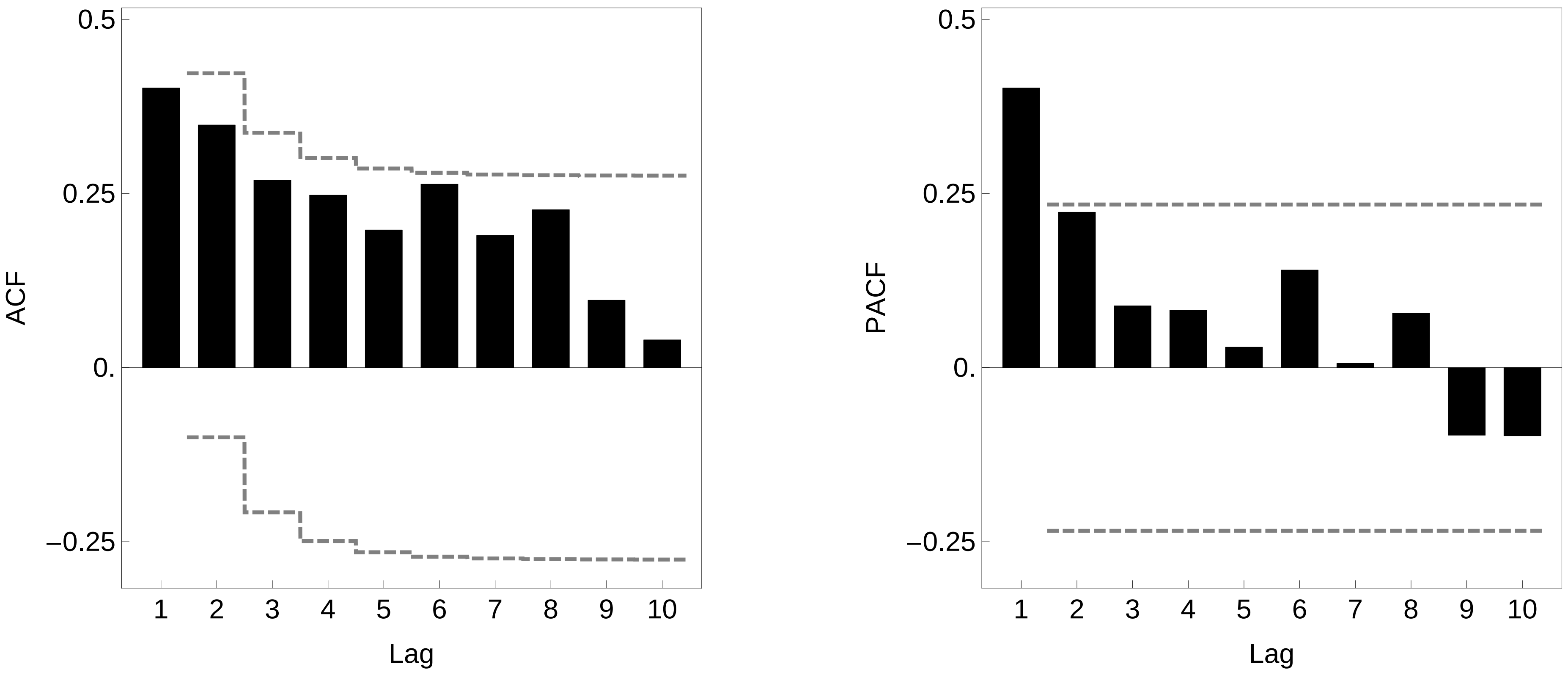
see Kim & Weiß (2015) for details.

Generally, asymptotic theory provides good guide for sample size $T \geq 250$.

Power w.r.t. binomial AR(2), BL usually superior.

Goodness-of-Fit w.r.t. Autocorrelation Structure

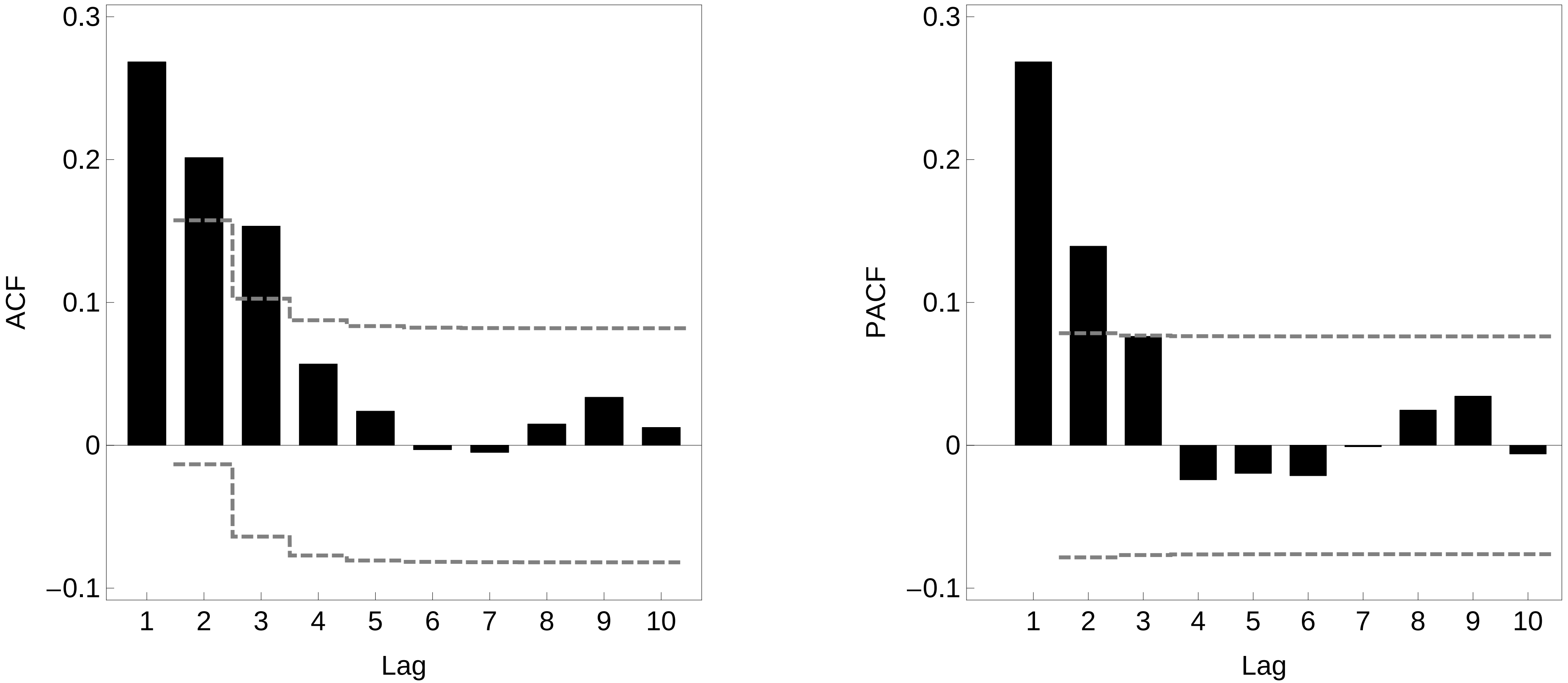
Example of securities counts: (Kim & Weiß, 2015)



No contradiction against null of *first*-order model.

Goodness-of-Fit w.r.t. Autocorrelation Structure

Example of access counts: (Kim & Weiß, 2015)



Contradicts null of *first*-order model, order $p = 2$ or 3.

Conclusions

- Binomial AR(1) model characterized by binomial marginal distribution & AR(1)-like autocorrelation.
- Diagnostic tests concerning marginal distribution (Pearson-type GoF test and dispersion test) and concerning autocorrelation (sample (P)ACF, Box-type test statistics), asymptotic properties and finite-sample performance.
- Future research on particular type of non-binomial behaviour: zero inflation. Development of corresponding models and diagnostic tests.

Thank You

for Your Interest!

Christian H. Weiß

Department of Mathematics & Statistics

Helmut Schmidt University, Hamburg

`weissc@hsu-hh.de`

Literature

Kim & Weiß (2015):
Goodness-of-Fit Tests for Bin. AR(1) Processes. Statistics 49, 291–315.

Weiß & Kim (2014):
Diagn. Mod. Extra-Binomial Variation . . . Appl. Stoch. Mod. 30, 588–608.

McKenzie (1985): *S. simple models for discrete variate t.s.* Water Res. Bull. 21, 645–650.

Romano & Thombs (1996): *Inference for autocorrelations* . . . JASA 91(434), 590–600

Steutel & van Harn (1979): *Discr. analogues of self-decomp.* . . . Ann. Probab. 7, 893–899.

Weiß (2009a): *Monitoring correl. processes w. binom. marginals* J. Appl. Stat. 36, 399–414.

Weiß (2009b): *A new class of autoregr. models* . . . Comm. Stat. Th. Meth. 38, 447–460.

Weiß & Kim (2013): *Parameter estim. for binom. AR(1) models* . . . Stat. Pap. 54, 563–590.