Diagnosing and Modelling Extra-Binomial Variation for Time-Dependent Counts

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich **Mathematik**

Christian H. Weiß

Department of Mathematics,

Darmstadt University of Technology

Hee-Young Kim

Institute of Economics,

Korea University Seoul

Binomial AR(1) Processes

Definition & Properties

Aim:

Obtain counterpart of AR(1) model $X_t = \alpha \cdot X_{t-1} + \epsilon_t$,

but for process of counts

with finite range $\{0, \ldots, n\}$.

Binomial thinning operator (Steutel & van Harn, 1979):

$$\alpha \circ X := \sum_{i=1}^{X} Y_i, \qquad \text{where } Y_i \text{ are i.i.d. } \mathrm{Bin}(1, \alpha),$$

i. e., $\alpha \circ X \sim \mathrm{Bin}(X, \alpha)$ and has range $\{0, \ldots, X\}$.

($\approx$ number of "survivors" from population of size X)

Fix $n \in \mathbb{N}$.

Parameters $\pi \in (0;1)$, $\quad \rho \in \left(\max\{-\frac{\pi}{1-\pi}, -\frac{1-\pi}{\pi}\}\ ;\ 1\right)$.

Define thinning probabilities $\beta := \pi\,(1-\rho)$ and $\alpha := \beta + \rho$.

Binomial AR(1) process $(X_t)_{\mathbb{N}_0}$ with range $\{0,\ldots,n\}$ defined by the recursion

$$X_{t+1} \;=\; \underbrace{\alpha \circ X_t}_{\text{survivors}} + \underbrace{\beta \circ (n - X_t)}_{\text{newly occupied}} \qquad \text{for } t \geq 0,$$

thinnings performed independently, independent of $(X_s)_{s<t}$.

(McKenzie, 1985)

Well-known properties:

Ergodic Markov chain, transition probabilities

$\mathbb{P}(X_{t+1}=k \mid X_t=l) \; = $

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

uniquely determined stationary distribution: $\text{Bin}(n,\pi)$.

Autocorrelation function: $\rho_X(k)=\rho^k$ for $k\geq 0$.

Regression properties:

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

Weiß & Pollett (2012):

***h*-step regression properties:**

Define $\beta_h = \pi(1-\rho^h)$ and $\alpha_h = \beta_h + \rho^h$ for $h \geq 1$.

Then

$$\mathbb{P}(X_{t+h} = k \mid X_t = l) \; = \; \sum_{m=\max\{0,k+l-n\}}^{\min\{k,l\}} \binom{l}{m} \binom{n-l}{k-m} \alpha_h^m (1-\alpha_h)^{l-m} \; \beta_h^{k-m} (1-\beta_h)^{n-l+m-k},$$

$$E[X_{t+h} \mid X_t] \; = \; \rho^h \cdot X_t \; + \; n\beta_h,$$

$$\mathrm{Var}[X_{t+h} \mid X_t] \; = \; \rho^h(1-\rho^h)(1-2\pi) \cdot X_t \; + \; n\beta_h(1-\beta_h).$$

Further recent results together with H.-Y. Kim:

- closed-form expressions for $\mu(s_1, \ldots, s_{r-1})$ for $r \leq 4$;
  (Weiß & Kim, 2011)

- asymptotics of YW-, CLS-, SD- and ML-estimator,
  finite-sample performance,
  effect of jackknife. (Weiß & Kim, 2011, 2012)

Diagnosing Extra-Binomial Variation

Motivation & Approaches

Motivation: Relation between variance and mean of $\mathsf{B}(n,\pi)$-distribution determined by

$$I_{\mathsf{d}} \;=\; 1, \qquad \text{where } I_{\mathsf{d}} \;:=\; \frac{n\sigma^2}{\mu(n-\mu)}.$$

I_{d} referred to as *binomial index of dispersion*.

If $I_{\mathsf{d}} > 1 \rightarrow$ **extra-binomial variation** (overdispersion).

How to diagnose extra-binomial variation?

How to diagnose extra-binomial variation?

Empirical binomial index of dispersion:

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

How to diagnose extra-binomial variation?

Empirical binomial index of dispersion:

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

Derive asymptotics by using

- $(X_t)_{\mathbb{Z}}$ is φ-mixing with geometrically decreasing weights (Billingsley, 1968, p. 167f);
- central limit theorem (Billingsley, 1968, p. 177);
- expressions for $\mu(s_1,\ldots,s_{r-1})$ (Weiß & Kim, 2011)

How to diagnose extra-binomial variation?

Theorem:

If $X_1, \ldots, X_T$ stem from stationary binomial AR(1) model, then

$$\sqrt{T}\,(\hat{I}_{\mathrm{d}} - 1) \ \stackrel{\mathrm{D}}{\to}\ \mathrm{N}\left(0,\ 2(1-\tfrac{1}{n})\,\frac{1+\rho^2}{1-\rho^2}\right).$$

$\Rightarrow$ **Critical value** for test based on $\hat{I}_{\mathrm{d}}$:

$$1 + z_{1-\beta} \cdot \sqrt{\frac{2}{T}\,(1-\tfrac{1}{n})\,\frac{1+\rho^2}{1-\rho^2}},$$

where $z_{1-\beta}$: $(1-\beta)$-quantile of $\mathrm{N}(0,1)$-distribution.

Real-data application:

European Union (EU): changes in consumer prices measured via Harmonised Index of Consumer Prices (HICP).

Price stability in Euro area: annual rates of change in HICP (inflation rates) should be below 2 %.

2009: Euro area with $n = 17$ member states ("EA17").

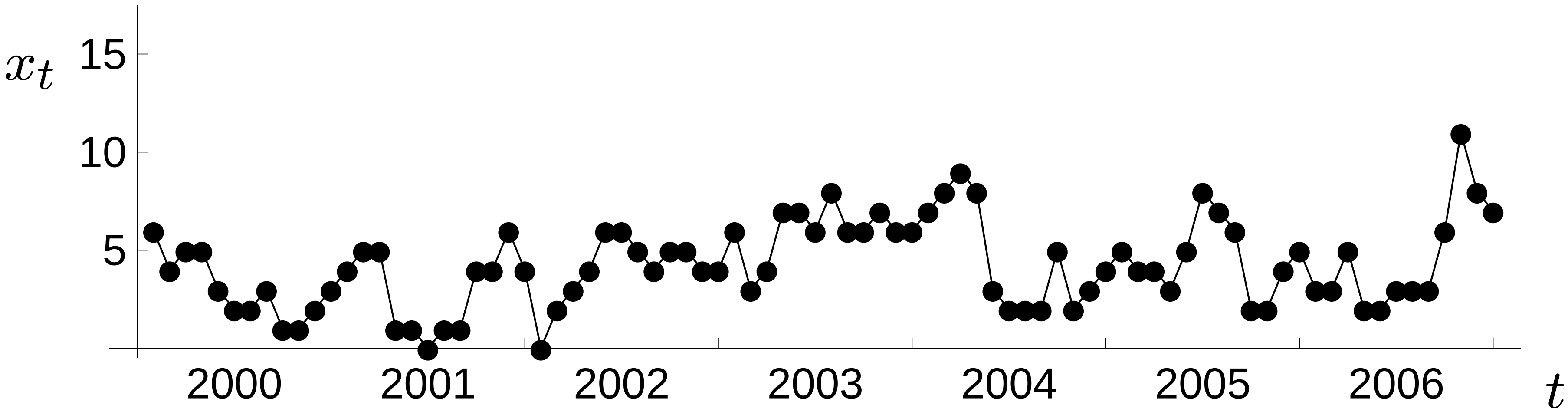
Question: How many EA17 countries have stable prices?

From monthly inflation rates (Eurostat, Jan. 2000 – . . .), we computed corresponding counts x_t for each month t.

Real-data application:

We first restrict to Jan. 2000 to Dec. 2006

$\Rightarrow$ time series $x_1, \ldots, x_T$ of length $T = 84$.



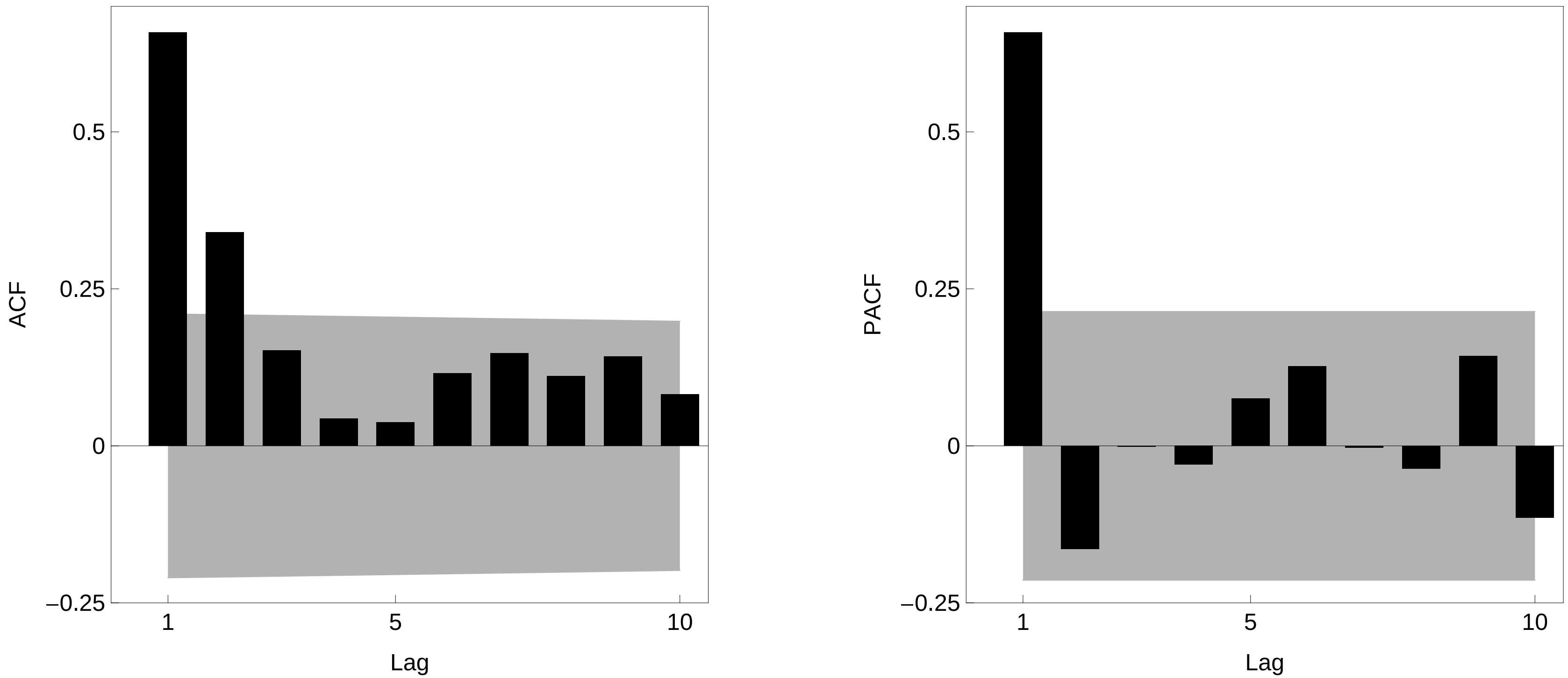
Data look stationary but serially dependent ($\widehat{\rho}(1) \approx 0.658$),

$\bar{x}/n \approx 0.251$, i. e.:

about 25 % of the EA17-countries show stable prizes.

Real-data application:

ACF and PACF of the price stability counts:



AR(1)-like autocorrelation structure,
so binomial AR(1) model appropriate?

Real-data application:

Value of $\widehat{I}_{\mathrm{d}}$ about $1.521 > 1$,

approximate critical value about 1.392,
(significance level 5 %,
computed by plugging-in $\widehat{\rho} \approx 0.658$)

so **significant** extra-binomial variation!

Plausible: EA17 countries economically heterogeneous,
so perhaps no unique π for price stability.

So how to model the price stability counts?

Beta-Binomial AR(1) Processes

Motivation & Properties

Motivation: Relation between variance and mean of B(n, π)-distribution determined by

$$I_{\mathrm{d}} \;=\; 1, \qquad \text{where } I_{\mathrm{d}} \;:=\; \frac{n\sigma^2}{\mu(n-\mu)}.$$

I_{d} referred to as *binomial index of dispersion*.

If $I_{\mathrm{d}} > 1 \rightarrow$ **extra-binomial variation** (overdispersion).

How to diagnose extra-binomial variation? ✓

Popular approach for extra-binomial variation:
beta-binomial distribution.

How to adapt this approach to time-dependent counts?

How to adapt beta-binomial approach to dependent counts?

Let X have range $\mathbb{N}_0$.

Let α_ϕ be random variable independent of X,

which follows distribution $\text{BETA}\big(\frac{1-\phi}{\phi}\cdot\alpha,\ \frac{1-\phi}{\phi}\cdot(1-\alpha)\big)$,

where $\alpha,\phi \in (0;1)$.

$\alpha_\phi \circ X$ obtained from X by **beta-binomial thinning**

if operator '$\circ$' is binomial thinning operator,

performed independently of X and α_ϕ.

How to adapt beta-binomial approach to dependent counts?

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

Beta-binomial AR(1) process $(X_t)_{\mathbb{N}_0}$ with range $\{0,\ldots,n\}$:

$$X_{t+1} \;=\; \alpha_\phi \circ X_t \;+\; \beta_\phi \circ (n - X_t) \qquad \text{for } t \geq 0,$$

α_ϕ, β_ϕ and thinnings performed independently,
and independent of $(X_s)_{s<t}$.

We derived the following properties:

Primitive and hence ergodic Markov chain, with

$$\mathbb{P}(X_t = k \mid X_{t-1} = l) \;=\; \sum_{m=\max\{0,k+l-n\}}^{\min\{k,l\}} \binom{l}{m} \binom{n-l}{k-m}$$
$$\cdot \frac{B\Big(m+\frac{1-\phi}{\phi}\cdot\alpha,\; l-m+\frac{1-\phi}{\phi}\cdot(1-\alpha)\Big)}{B\Big(\frac{1-\phi}{\phi}\cdot\alpha,\; \frac{1-\phi}{\phi}\cdot(1-\alpha)\Big)}\; \frac{B\Big(k-m+\frac{1-\phi}{\phi}\cdot\beta,\; n-l-k+m+\frac{1-\phi}{\phi}\cdot(1-\beta)\Big)}{B\Big(\frac{1-\phi}{\phi}\cdot\beta,\; \frac{1-\phi}{\phi}\cdot(1-\beta)\Big)}.$$

Conditional variance not linear in X_{t-1} anymore:

$$\mathbb{E}[X_t \mid X_{t-1}] \;=\; \rho\cdot X_{t-1} \;+\; n\beta,$$

$$\begin{aligned}\mathrm{Var}[X_t \mid X_{t-1}] \;=\; & \phi\cdot\Big(\alpha(1-\alpha)+\beta(1-\beta)\Big)\cdot X_{t-1}^2 \\ & +\; n\beta(1-\beta)\cdot\Big(1+\phi(n-1)\Big) \\ & +\; X_{t-1}\cdot\Big(\rho(1-\rho)(1-2\pi)\cdot(1-\phi) \;-\; 2n\beta(1-\beta)\cdot\phi\Big).\end{aligned}$$

We derived the following properties:

If $(X_t)_{\mathbb{N}_0}$ stationary, then

$$\mu \;=\; n\pi, \qquad \rho_X(k) = \rho^k,$$

$$\sigma^2 \;=\; n\pi(1-\pi)\cdot\frac{(1-\phi)(1+\rho)+n\phi\cdot\big(1-2\pi(1-\pi)(1-\rho)\big)}{(1-\phi)(1+\rho)+\phi\cdot\big(1-2\pi(1-\pi)(1-\rho)\big)}.$$

In particular,

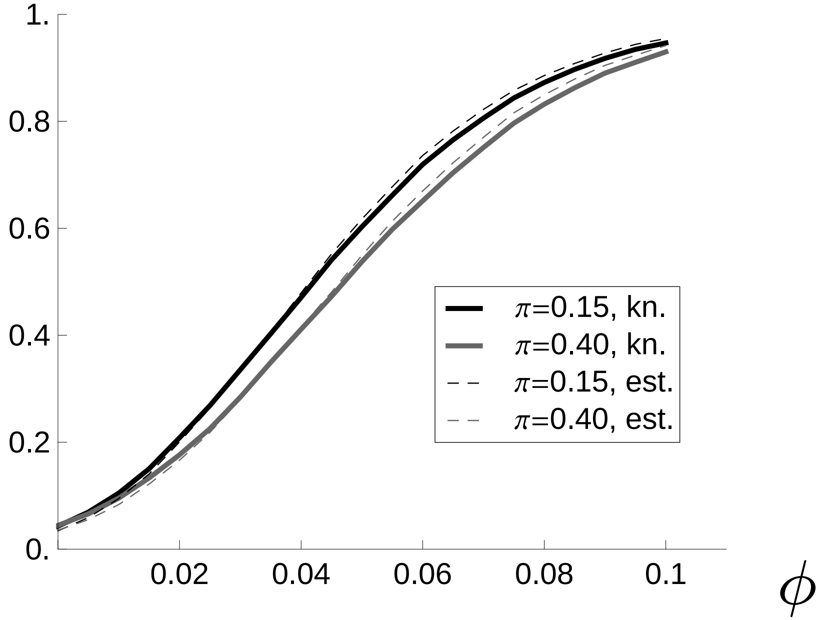
$$I_{\mathrm{d}} \;=\; 1\;+\;\frac{(n-1)\cdot\big(1-2\pi(1-\pi)(1-\rho)\big)}{(\frac{1}{\phi}-1)(1+\rho)+\big(1-2\pi(1-\pi)(1-\rho)\big)}.$$

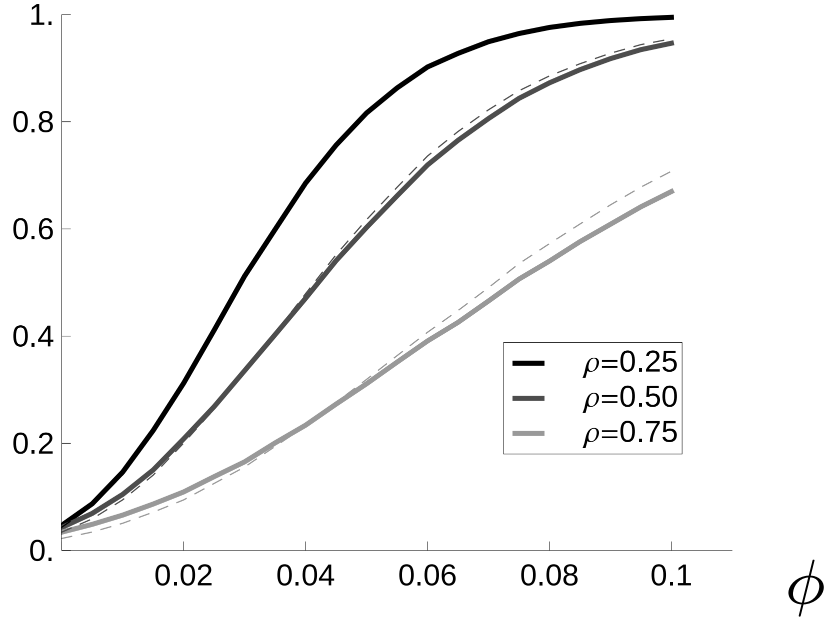
I_{d} strictly increasing in $\phi\in(0;1)$,

I_{d} takes values in $(1;n)$, i. e., extra-binomial variation.

Application 1:

Power analysis for test based on $\widehat{I}_{\mathrm{d}}$:





Power decreases as π, ρ increase (but increases with n, T), little affected by plugging-in $\widehat{\rho}$.

Application 2: Modelling price stability counts.

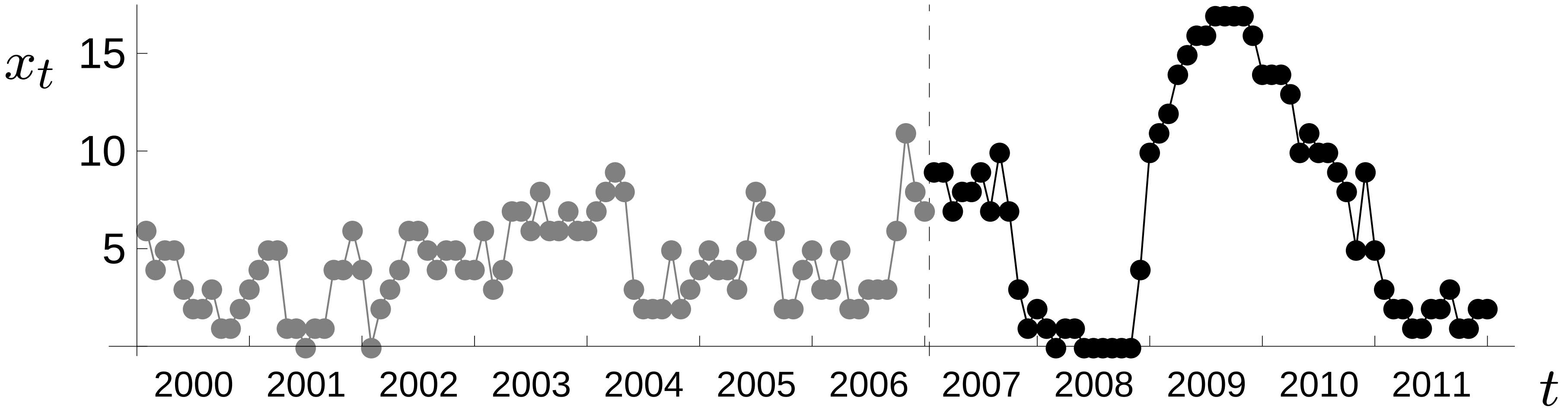
$\hat{\pi}_{\mathrm{ML}} \approx 0.255$, $\hat{\rho}_{\mathrm{ML}} \approx 0.621$, $\hat{\phi}_{\mathrm{ML}} \approx 0.037$.

ML estimates for α and β are 0.718 and 0.097, i. e.:

Country having stable prices in month t also stable prices in month $t+1$ with probability $\approx 72\,\%$,
while prob. for newly having stable prices $< 10\,\%$.

Price stability counts stationary ($\approx$ beta-binomial AR(1)) between Jan. 2000 and Dec. 2006,
but what happened afterwards?

Application 2: Modelling price stability counts.



$\geq$ Sept. 2007: surging oil prices cause HICP inflation.

Then upcoming sub-prime crisis,

point of culmination in Sept. 2008 ("Lehman")

$\Rightarrow$ inflation decreases up to even negative values.

Then Euro area lapsed into severe sovereign debt crisis.

Ideas for future research:

- Diagnostic tests for binomial AR(1) processes, e. g., goodness-of-fit, model order.
- Over- and underdispersion through density-dependent binomial AR(1) models.
- Outliers in binomial AR(1) processes.
- . . .

Thank You

for Your Interest!

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich **Mathematik**

Christian H. Weiß

Department of Mathematics

Darmstadt University of Technology

weiss@mathematik.tu-darmstadt.de

Literature

Billingsley (1968): *Convergence of probab. measures*. 1st edition, Wiley.

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

Steutel & van Harn (1979): *Discrete analogues of self-decomposability and stability*. Ann. Prob. 7(5), 893-899.

Weiß & Kim (2011): *Binomial AR(1) processes: moments, cumulants, and estimation*. Statistics, to appear.

Weiß & Kim (2012): *Parameter estimation for binomial AR(1) models with applications in finance and industry*. Statistical Papers, to appear.

Weiß & Pollett (2012): *Chain binomial models and binomial autoregressive processes*. Biometrics 68(3), 815-824.

Data source: Eurostat,

`http://appsso.eurostat.ec.europa.eu/nui/show.do?wai=true&dataset=prc_hicp_manr`

Background information: Annual Reports of the ECB,

`http://www.ecb.int/pub/annual/html/index.en.html`