Weighted Discrete ARMA Models for Categorical Time Series

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Christian H. Weiß, Osama Swidan

Department of Mathematics & Statistics,

Helmut Schmidt University, Hamburg

Funded by DFG – Projektnr. 516522977.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH
STAT

Discrete ARMA Models for Time Series

Introduction

Classical ARMA models popular for *real-valued* t.s.,

but cannot be applied to *discrete-valued* t.s.

Many attempts to define "ARMA-like" models (Weiß, 2018),

where **NDARMA model** (new discrete ARMA) of

Jacobs & Lewis (1983) applicable even to **categorical t.s.**:

Let X_t have categorical range $\mathcal{S} = \{s_0, \ldots, s_d\}$,

let innovations $(\epsilon_s)_{s \leq t}$ be i. i. d. on $\mathcal{S}$.

Let $\boldsymbol{D}_t = (D_{t,-\mathrm{q}}, \ldots, D_{t,0}, \ldots, D_{t,\mathrm{p}})$ be i. i. d. multinomial via

Mult(1; $\phi_{-\mathrm{q}}, \ldots, \phi_0, \ldots, \phi_{\mathrm{p}}$), then

$$X_t \; = \; \sum_{i=1}^{\mathrm{p}} D_{t,i}\, X_{t-i} \; + \; D_{t,0}\, \epsilon_t \; + \; \sum_{j=1}^{\mathrm{q}} D_{t,-j}\, \epsilon_{t-j}.$$

NDARMA model

$$X_t \;=\; \sum_{i=1}^{\mathsf{p}} D_{t,i}\, X_{t-i} \;+\; D_{t,0}\, \epsilon_t \;+\; \sum_{j=1}^{\mathsf{q}} D_{t,-j}\, \epsilon_{t-j},$$
$$\boldsymbol{D}_t \;\sim\; \text{Mult}(1;\ \phi_{-\mathsf{q}}, \ldots, \phi_0, \ldots, \phi_{\mathsf{p}})$$

looks like ARMA at first glance, but **major differences:**

X_t randomly selects outcome of either one of last p observations, $X_{t-1}, \ldots, X_{t-\mathsf{p}}$, or one of last $\mathsf{q}+1$ innovations, $\epsilon_t, \ldots, \epsilon_{t-\mathsf{q}}$.

Pros: serial dependence structure satisfies YW equations, random selection mechanism applicable to any range.

Cons: sample paths with long “runs” and sudden jumps, mainly relevant for nominal t.s.; only positive dependence.

⇒ Omit aforementioned "cons" by

new and flexible extension of NDARMA model!

Outline:

- Concept of "weighting operators", tailor-made for, e. g., ordinal time series or negative dependence.
- Stochastic properties of resulting weighted discrete ARMA (WDARMA) models.
- Parameter estimation, data application to an ordinal time series on air quality in Beijing.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH
STAT

Weighting Operators and WDARMA Models

Definition & Approaches

Notations: $\mathbf{0}_k$ $(\mathbf{1}_k)$ is k-dim. vector of zeros (ones),

$\mathbf{I}_k = \text{diag}(\mathbf{1}_k)$ is $k \times k$-identity matrix,

$\mathbf{E}_k$ $(\mathbf{O}_k)$ is $k \times k$-matrix of ones (zeros),

$\mathbb{S}_k := \{\boldsymbol{u} \in (0;1)^k \mid \mathbf{1}_k^\top \boldsymbol{u} = 1\}$, $\overline{\mathbb{S}}_k := \{\boldsymbol{u} \in [0;1]^k \mid \mathbf{1}_k^\top \boldsymbol{u} = 1\}$

denote open and closed k-part unit simplex, respectively.

If X categorical r.v. with range $\mathcal{S}$, we assume **pmf vector**

$\boldsymbol{p} = (p_0, \ldots, p_d)^\top \in \mathbb{S}_{d+1}$, i. e., each $p_i = P(X = s_i) > 0$.

We associate **weight vector** $\boldsymbol{w}_j$ with each $s_j \in \mathcal{S}$,

but requirement $\boldsymbol{w}_j \in \overline{\mathbb{S}}_{d+1}$ allows for zero entries.

Weight vectors $\Rightarrow$ **weight matrix $\mathbf{W} = (\boldsymbol{w}_0, \ldots, \boldsymbol{w}_d)$**,

which is left-stochastic: $\mathbf{1}_{d+1}^\top \mathbf{W} = \mathbf{1}_{d+1}^\top$.

As $\boldsymbol{w}_j$ columns of $\mathbf{W}$, denote entries as w_{ij}: $\boldsymbol{w}_j = (w_{0j}, \ldots, w_{dj})^\top$.

Finally, random **weighting operator** $\mathcal{W}(\cdot)$, where

$\mathcal{W}(s_j)$ generates categorical value from $\mathcal{S}$ according to $\boldsymbol{w}_j$,

i. e., $P\big(\mathcal{W}(s_j) = s_i\big) = w_{ij}$. **Short-hand notation:** $\mathcal{W}(s_j) \sim \boldsymbol{w}_j$.

If applied to categorical r.v. $X \sim \boldsymbol{p}$, we assume conditionally:

$\mathcal{W}(X) \,|\, \{X = s_j\} \ \sim \boldsymbol{w}_j$. Thus, by conditioning,

$$P\big(\mathcal{W}(X) = s_i\big) \ = \ \sum_{j=0}^{d} w_{ij}\, p_j. \quad (= i\text{th entry of product } \mathbf{W}\,\boldsymbol{p})$$

Requirement $\boldsymbol{w}_j \in \overline{\mathbb{S}}_{d+1}$ implies that

each column of **W** has at least one truly positive entry:

(W1) $\forall\, j \in \{0,\ldots,d\}$, there exists $i \in \{0,\ldots,d\}$ such that $w_{ij} > 0$.

But: To ensure that each state in $\mathcal{S}$ reachable after applying $\mathcal{W}(\cdot)$, analogous property for rows of **W** necessary.

From now on, we also assume:

(W2) $\forall\, i \in \{0,\ldots,d\}$, there exists $j \in \{0,\ldots,d\}$ such that $w_{ij} > 0$.

Consequence: If $X \sim \boldsymbol{p} \in \mathbb{S}_{d+1}$,

then also $\mathcal{W}(X)$'s pmf vector contained in $\mathbb{S}_{d+1}$.

Examples for nominal r.v.:

Identity weighting W $= \mathbf{I}_{d+1}$ preserves given category,

i. e., $\mathcal{W}(s_j) = s_j$ with probability one.

Used to embed NDARMA models into novel WDARMA class.

Reverse weighting W $= d^{-1}\,(\mathbf{E}_{d+1} - \mathbf{I}_{d+1})$:

given category *not* preserved but randomly "flipped"

into other state, also see McGee & Harris (2012).

Used for generating negative dependence in nominal t.s.,

in analogy to Jentsch & Reichmann (2019).

Examples for ordinal r.v.:

Triangular weighting:

$$\boldsymbol{w}_0 = (\tfrac{2}{3}, \tfrac{1}{3}, 0, \ldots)^\top, \ldots, \boldsymbol{w}_j = (\ldots, \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4}, \ldots)^\top, \ldots, \boldsymbol{w}_d = (\ldots, 0, \tfrac{1}{3}, \tfrac{2}{3})^\top.$$

Gives weight to current category and immediate neighbors, thus accounts for natural ordering among states.

Zero inflation, i. e., inflation of lowest state s_0:

$\boldsymbol{w}_j = (1-\omega)\, \boldsymbol{e}_j + \omega\, \boldsymbol{e}_0$, where $\boldsymbol{e}_j$ is jth unit vector.

Thus, s_0 preserved, larger s_j change to s_0 with prob. ω.

Easily generalized to inflate other or multiple states, also see Liu et al. (2022a).

Weighting Operators and WDARMA Models

Inspired by GDARMA model of Gouveia et al. (2018) and Möller & Weiß (2020) (which is defined for *quantitative* t.s.), we propose novel **WDARMA**$(\mathbf{p},\mathbf{q})$ model

$$X_t \;=\; \sum_{k=1}^{\mathrm{p}} D_{t,k}\, \mathcal{W}_{t,k}(X_{t-k}) \;+\; D_{t,0}\, \epsilon_t \;+\; \sum_{l=1}^{\mathrm{q}} D_{t,-l}\, \mathcal{W}_{t,-l}(\epsilon_{t-l}),$$

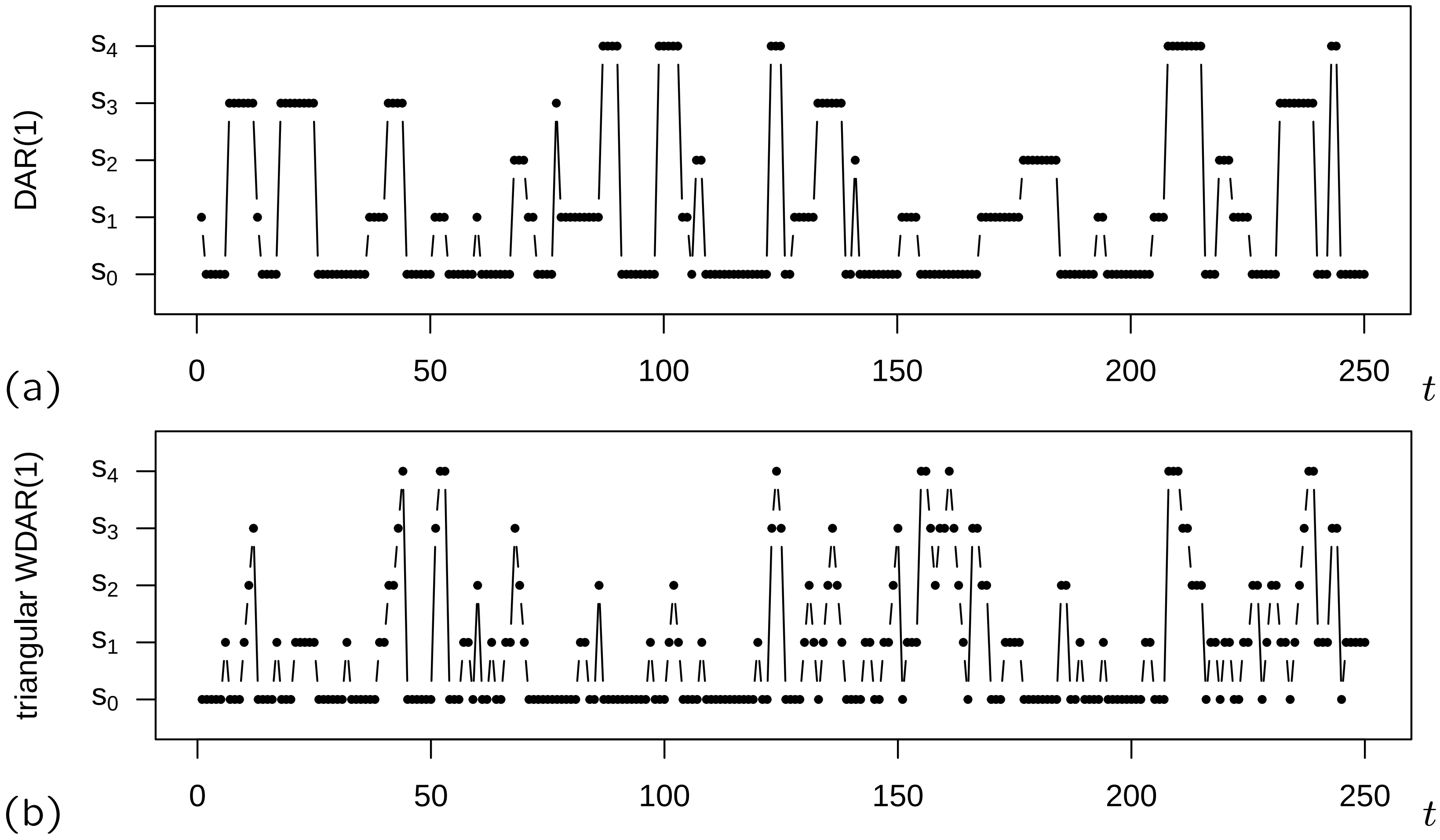
where operators $\mathcal{W}_{t,\cdot}$ executed independently of other r.v.

WDARMA becomes NDARMA if identity weighting.

For simplicity, we assume that all $\mathcal{W}_{t,\cdot}$ have same $\mathbf{W}$.

But existence proof and further stochastic properties extended to different $\mathbf{W}_k$ and $\mathbf{W}_{-l}$ at lags k, l in Weiß & Swidan (2024).

Example: Simulated ordinal WDAR(1) sample paths using (a) identity weighting and (b) triangular weighting.



HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Stationary WDARMA Processes

Stochastic Properties

Theorem: WDARMA(p, q) process $(X_t)_{\mathbb{Z}}$.

Then, $(X_t)_{\mathbb{Z}}$ ergodic and unique stationary solution.

$(X_t)_{\mathbb{Z}}$ is φ-mixing with geometrically decreasing weights $(f_n)_{\mathbb{N}}$,
i. e., there exist $a > 0$ and $0 < \rho < 1$ such that $f_n = a \cdot \rho^n$.

Proof in Weiß & Swidan (2024) uses (W2) and
$(\max\{p, 1\} + \max\{q, 1\})$-dimensional
Markov-chain representation of $(X_t)_{\mathbb{Z}}$,
transition matrix of which shown to be primitive.

Marginal distribution:

Let $\boldsymbol{p} \in \mathbb{S}_{d+1}$ be pmf of stationary marginal dist. of $(X_t)_{\mathbb{Z}}$,

let $\boldsymbol{\pi} \in \mathbb{S}_{d+1}$ be the one of $(\epsilon_t)_{\mathbb{Z}}$.

Denote $\phi^{(\mathsf{p})} := \sum_{k=1}^{\mathsf{p}} \phi_k$ and $\phi^{(\mathsf{q})} := \sum_{l=1}^{\mathsf{q}} \phi_{-l}$.

Proposition: pmf vector satisfies

$$\left(\mathbf{I}_{d+1} - \phi^{(\mathsf{p})}\,\mathbf{W}\right)\boldsymbol{p} \;=\; \left(\phi_0\,\mathbf{I}_{d+1} + \phi^{(\mathsf{q})}\,\mathbf{W}\right)\boldsymbol{\pi}.$$

Bivariate distribution: For pairs (X_t, X_{t-h}) and (X_t, ϵ_{t-h}), let

$$\mathbf{P}(h) \;=\; \big(p_{ij}(h)\big)_{i,j=0,\ldots,d} \quad \text{with } p_{ij}(h) = P(X_t = s_i, X_{t-h} = s_j),$$

$$\mathbf{\Pi}(h) \;=\; \big(\pi_{ij}(h)\big)_{i,j=0,\ldots,d} \quad \text{with } \pi_{ij}(h) = P(X_t = s_i, \epsilon_{t-h} = s_j).$$

Proposition: For stationary WDARMA model, it holds

$$\boldsymbol{\Pi}(h)-\boldsymbol{p}\,\boldsymbol{\pi}^\top = \begin{cases} \sum_{k=1}^{\min\{h,\mathsf{p}\}} \phi_k\,\mathbf{W}\left(\boldsymbol{\Pi}(h-k)-\boldsymbol{p}\,\boldsymbol{\pi}^\top\right) \\ \quad +\ \sum_{l=1}^{\mathsf{q}} \phi_{-l}\,\delta_{lh}\,\mathbf{W}\left(\mathrm{diag}(\boldsymbol{\pi})-\boldsymbol{\pi}\,\boldsymbol{\pi}^\top\right) & \text{if } h>0, \\ \phi_0\left(\mathrm{diag}(\boldsymbol{\pi})-\boldsymbol{\pi}\,\boldsymbol{\pi}^\top\right) & \text{if } h=0, \\ \mathbf{O}_{d+1} & \text{if } h<0. \end{cases}$$

Theorem: For stationary WDARMA model, if $h>0$,

$$\mathbf{P}(h)-\boldsymbol{p}\,\boldsymbol{p}^\top = \sum_{k=1}^{\mathsf{p}} \phi_k\,\mathbf{W}\left(\mathbf{P}(h-k)-\boldsymbol{p}\,\boldsymbol{p}^\top\right)+\sum_{l=h}^{\mathsf{q}} \phi_{-l}\,\mathbf{W}\left(\boldsymbol{\Pi}(l-h)-\boldsymbol{p}\,\boldsymbol{\pi}^\top\right)^\top,$$

while $\mathbf{P}(0)=\mathrm{diag}(\boldsymbol{p})$ and $\mathbf{P}(-h)=\mathbf{P}(h)^\top$.

Stochastic Properties of WDARMA Processes

Above results allow to compute any relevant stochastic properties by solving kind of "Yule–Walker equations".

If $p \geq 2$, solution possible by using Kronecker product and vec-operator, see Weiß & Swidan (2024) for details.

In particular, exact computation of serial dependence measures

$$\kappa_{\mathsf{nom}}(h) \;=\; \frac{\mathrm{tr}\big(\mathbf{P}(h) - \boldsymbol{p}\,\boldsymbol{p}^\top\big)}{\mathrm{tr}\big(\mathbf{P}(0) - \boldsymbol{p}\,\boldsymbol{p}^\top\big)}, \qquad \kappa_{\mathsf{ord}}(h) \;=\; \frac{\mathrm{tr}\big(\mathbf{F}(h) - \boldsymbol{f}\,\boldsymbol{f}^\top\big)}{\mathrm{tr}\big(\mathbf{F}(0) - \boldsymbol{f}\,\boldsymbol{f}^\top\big)},$$

see Weiß (2020) for background.

Here, $\boldsymbol{f} = (f_0, \ldots, f_{d-1}, f_d)$ with $f_i = P(X \leq s_i)$,
$\mathbf{F}(h) = \big(f_{ij}(h)\big)_{i,j=0,\ldots,d}$ with $f_{ij}(h) = P(X_t \leq s_i, X_{t-h} \leq s_j)$.

Modeling Ordinal Air Quality Time Series

Estimation & Illustration

ML estimation particularly simple in WDAR(p) case, because then, $(X_t)_{\mathbb{Z}}$ is Markov process.

So log-likelihood computes as

$$\ell(\phi_1, \ldots, \phi_\mathsf{p}, \boldsymbol{\pi}) \;=\; \sum_{t=\mathsf{p}+1}^{n} \ln\Big(\sum_{k=1}^{\mathsf{p}} \phi_k\, w_{i_t i_{t-k}} + (1-\phi^{(\mathsf{p})})\, \pi_{i_t}\Big).$$

Estimation by constrained numerical optimization, $\sum_{i=0}^{d} \pi_i = 1$ and $\phi^{(\mathsf{p})} = \sum_{k=1}^{\mathsf{p}} \phi_k < 1$.

Normal asymptotics from Condition 5.1 in Billingsley (1961).

Good finite-sample performance confirmed by simulations, see Weiß & Swidan (2024) for details.

If q $\geq$ 1, still efficient implementation of ML estimation possible by adapting recursive approach of Weiß et al. (2019).

For simplicity, let us focus on **WDARMA**$(1,1)$ **model**.

Define probabilities (for $t=2,\ldots,n$)

$$b_{kl}^{(m)}(t) \;=\; P(\epsilon_t=s_k,\epsilon_{t-1}=s_l,\ X_t,\ldots,X_2 \mid X_1,\epsilon_1=s_m).$$

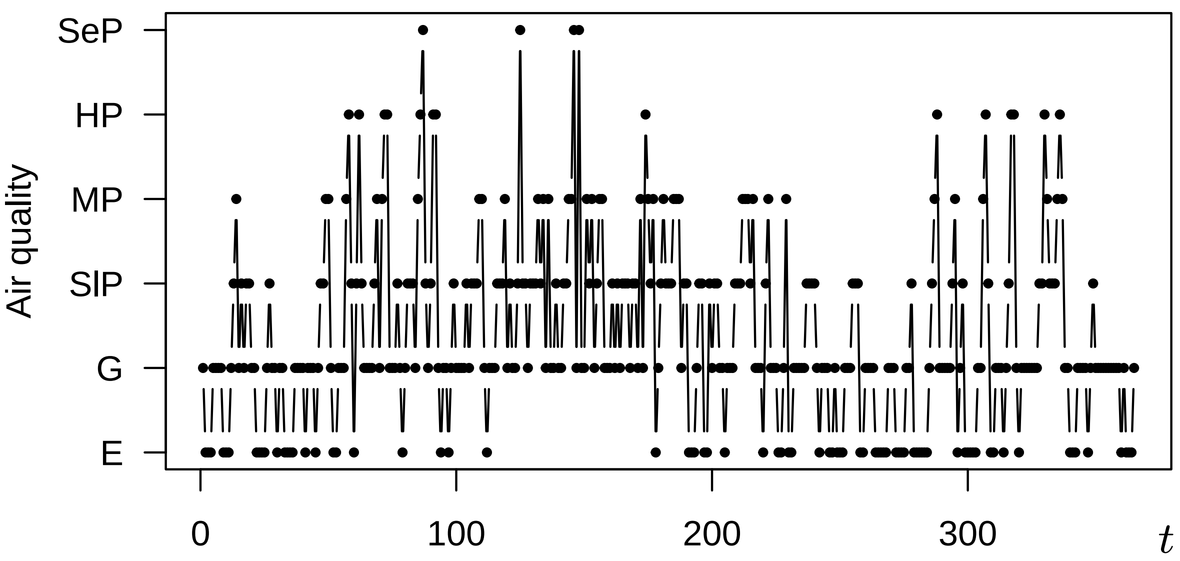
Likelihood follows from double sum $\sum_{k,l=0}^{d} b_{kl}^{(m)}(n)$.

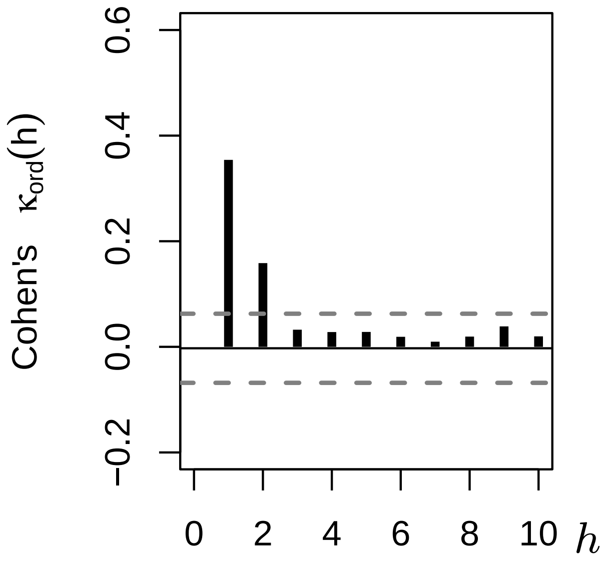
Distribution of (X_1,ϵ_1) from above YW equations.

Recursive scheme:

$$\begin{aligned} b_{kl}^{(m)}(t+1) \;&=\; P(X_{t+1} \mid \epsilon_{t+1}=s_k,\epsilon_t=s_l,\ X_t)\,P(\epsilon_{t+1}=s_k)\sum_{j=0}^{d} b_{lj}^{(m)}(t) \\ b_{kl}^{(m)}(2) \;&=\; \delta_{l,m}\,P(\epsilon_2=s_k)\,P(X_2 \mid X_1,\ \epsilon_2=s_k,\epsilon_1=s_m). \end{aligned}$$

Data example: Daily air quality in Beijing in 2018, taken from Liu et al. (2022b), with $d+1=6$ categories s_0 =“excellent” (E), . . . , s_5 =“severely polluted” (SeP).





AR-like dependence explainable by ordinary DAR model. But ordinal t.s., WDAR with triangular weighting more reasonable.

Information criteria for candidate models:

	DAR(1)	DAR(2)	DAR(3)	WDAR(1)	WDAR(2)	WDAR(3)
AIC	1009.3	1004.1	1006.7	958.9	**955.4**	958.2
BIC	1032.7	1031.4	1037.9	**982.3**	982.7	989.4

Parameter estimates (s. e.) of relevant WDAR models:

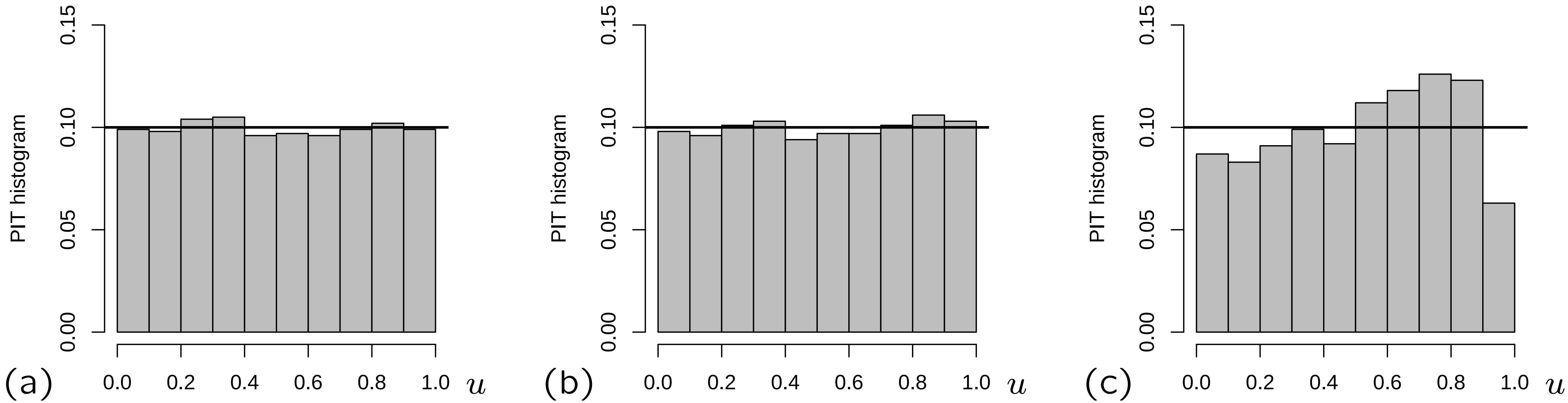
	π_1	π_2	π_3	π_4	π_5	ϕ_1	ϕ_2
WDAR(1)	0.521 (0.049)	0.194 (0.042)	0.081 (0.026)	0.034 (0.015)	0.018 (0.010)	0.529 (0.051)	—
WDAR(2)	0.547 (0.056)	0.184 (0.048)	0.070 (0.028)	0.034 (0.016)	0.014 (0.010)	0.502 (0.054)	0.076 (0.044)

Sample properties & stochastic properties of WDAR fits:

Marginal	f_0	f_1	f_2	f_3	f_4	IOV	skew
Data	*0.211*	*0.625*	*0.847*	*0.951*	*0.989*	*0.471*	*0.449*
WDAR(1)	0.196	0.616	0.833	0.933	0.978	0.494	0.422
WDAR(2)	0.203	0.627	0.841	0.935	0.980	0.488	0.434

Serial	$\kappa_{\text{ord}}(1)$	$\kappa_{\text{ord}}(2)$	$\kappa_{\text{ord}}(3)$	$\kappa_{\text{nom}}(1)$	$\kappa_{\text{nom}}(2)$	$\kappa_{\text{nom}}(3)$
Data	*0.354*	*0.159*	*0.032*	*0.221*	*0.100*	*-0.024*
WDAR(1)	0.361	0.155	0.070	0.208	0.068	0.026
WDAR(2)	0.364	0.200	0.107	0.207	0.095	0.043

PIT histograms of fitted (a) WDAR(1) model, (b) WDAR(2) model, and (c) ZOBPAR model:



ZOBPAR model originally proposed by Liu et al. (2022b).

Altogether, both WDAR model perform nearly equally, more parsimonious WDAR(1) model preferable choice.

Conclusions

- WDARMA models flexibly adapted to negative serial dependencies or ordinal data.
- Stationary, ergodic, and φ-mixing; closed-form marginal and bivariate probabilities.
- Efficient implementation of ML estimation, successfully applied to air quality data.

Future research:

- Approaches for forecasting WDARMA processes.
- Control charts for monitoring WDARMA processes.

Thank You for Your Interest!

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Christian H. Weiß

Department of Mathematics & Statistics

Helmut Schmidt University, Hamburg

weissc@hsu-hh.de

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Projektnummer 516522977.

Literature

Weiß & Swidan (2024) Weighted discrete ARMA models for categorical time series. *Journal of Time Series Analysis*, in press.
(→ open access)

Billingsley (1961) *Statistical Inference for Markov Processes*. Chicago Press.

Gouveia et al. (2018) A full ARMA model ... *SERRA* **32**, 2495–2514.

Jacobs & Lewis (1983) Stationary discrete auto... *JTSA* **4**, 19–36.

Jentsch & Reichmann (2019) Generalized binary ... *Econometrics* **7**, 47.

Liu et al. (2022a) Modeling air quality ... *SERRA* **36**, 2835–2845.

Liu et al. (2022b) Modeling normalcy-dominant ... *JTSA* **43**, 460–478.

McGee & Harris (2012) Coping with ... *J Prob Stat* **2012**, 417393.

Möller & Weiß (2020) Generalized discrete ARMA ... *ASMBI* **36**, 641–659.

Weiß (2018) *An Introduction to Discrete-Valued Time Series*. Wiley.

Weiß (2020) Distance-based analysis ... *JASA* **115**, 1189–1200.

Weiß et al. (2019) INARMA modeling ... *Stats* **2**, 284–320.