Journal of Statistical Sciences, Autumn and Winter, 2022
Vol. 16, No. 2, pp 469-492
DOI: 10.29252/jss.16.2.469

Integer-valued Autoregressive Model Based on Innovations with Discrete Exponential-Weibull Distribution

Noorollahi, M. J., Deiri, E., Baloui Jamkhaneh, E.,
Department of Statistics, Qaemshahr Branch, Islamic Azad University, Qaemshahr, Iran.

Corresponding author: E. Deiri, e.deiri53@qaemiau.ac.ir

Received: 9/1/2022 **Revised:** 5/11/2022 **Accepted and Published Online:** 17/11/2022.

Introduction

In real-life phenomena, we often encounter discrete data sets, such as the number of cancer cells, the number of rainy days, the volume of stocks, etc. Also, statistical inference based on correlated count data is frequently performed in biomedical studies, for example, the count of cancer cells, the number of patients with contagious diseases, and the number of successful trials. So, introducing a convenient model for counting time series is felt nowadays. The literature on integer-valued time series models has been introduced based on the Binomial thinning operator and its generalizations. As pioneers of the Integer-valued Autoregressive (INAR) models, we can cite McKenzie (1986) and Al-Osh and Alzaid (1987). This paper proposes a new flexible discrete Exponential-Weibull Distribution (DEW) and a new INAR(1) model based on the DEW innovations. This new DEW distribution provides several appealing statistical properties in terms of handling all forms of dispersion, skewness, and kurtosis. This subsequently allows for the modeling of count time series observations. In this context, the INAR(1) process is developed with Binomial thinning operators. The efficiency and superiority of the process in fitting counts data of deaths due to COVID-19 disease are compared with other competing models.

Material and Methods

This paper is concentrated on the new discrete distribution and its corresponding INAR(1) model. Several parametric and non-parametric estimation approaches are considered, including the conditional maximum likeli-

hood (parametric), conditional least squares (non-parametric), and Yule-Walker (non-parametric) estimation methods. Several INAR(1) processes are considered to fitting the COVID-19 data sets. The goodness of fit measures consists of the Akaike information criterion (AIC), Bayesian information criterion (BIC), consistent Akaike information criterion (CAIC), and Hanna-Quinn (HQIC).

Results and Discussion
The simulation comparison is conducted concerning the mean square errors (MSE), which shows the superiority of the conditional maximum likelihood estimation method. Due to evaluating the performance of estimators in terms of MSE, 100 iterations are considered for different sample sizes. Based on the goodness of fit measures, it is concluded that the DEW-INAR(1) process is preferred among other INAR(1) processes.

Conclusion
The main focus of the manuscript is to introduce the discrete version of the Exponential-Weibull distribution and the modeling of its corresponding integer-valued autoregressive process. The flexibility and comprehensiveness of the discrete Exponential-Weibull distribution in fitting different types of count data is deduced by examining the statistical characteristics. Also, parameter estimation methods and Monte Carlo simulation studies are presented. According to the simulation results, the conditional maximum likelihood parametric estimation method performs better than non-parametric approaches. Using the data of COVID-19, the efficiency of the new process is confirmed in comparison with the classical INAR(1) models.

Keywords: INAR(1) process, Discrete Exponential-Weibull, Dispersion index, Heavy-tailed.

Mathematics Subject Classification (2010): 62P10, 62M20.

©The Author(s). The Publisher is Iranian Statistical Society.
This is an open access article distributed under the terms and conditions of (CC BY-NC 4.0)

مجله علوم آماری، پاییز و زمستان ۱۴۰۱
جلد ۱۶، شماره ۲، ص ۴۶۹ – ۴۹۲
DOI: 10.29252/jss.16.2.469
مقاله پژوهشی

مدل خودبازگشتی گسسته‌مقدار بر اساس نوفه‌هایی با توزیع نمایی-وایبل گسسته

محمدجواد نورالهی، عین‌الله دیری، عزت‌الله بالوئی جامخانه
گروه آمار، واحد قائمشهر، دانشگاه آزاد اسلامی

نویسنده مسئول : عین‌الله دیری، e.deiri53@Qaemiau.ac.ir

تاریخ دریافت: ۱۴۰۰/۱۰/۱۹ تاریخ بازنگری: ۱۴۰۱/۸/۱۴ تاریخ پذیرش و انتشار: ۱۴۰۱/۸/۲۶

چکیده: در این مقاله، به منظور مدل‌سازی داده‌های سری زمانی گسسته‌مقدار، فرایند خودبازگشتی گسسته‌مقدار جدید بر اساس توزیع نمایی-وایبل گسسته معرفی شده است. نظر به اهمیت توزیع‌های گسسته در مدل‌سازی داده‌های شمارشی، همتای گسسته توزیع نمایی-وایبل معرفی و برخی ویژگی‌های آماری آن از قبیل تابع بقا، نرخ خطر، تابع مولد گشتاور، چولگی و کشیدگی بررسی می‌شود. شاخص‌های پراکندگی فیشر، چولگی و کشیدگی، بیانگر انعطاف‌پذیری و کارایی توزیع نمایی-وایبل گسسته در برازش انواع مختلف داده‌های شمارشی است. توزیع نمایی-وایبل گسسته، برازش داده‌هایی با ویژگی‌های مختلف پراکندگی (کم‌پراکندگی، بیش‌پراکندگی و همسان)، دم راست بلند (چوله به راست) و دم سنگین را پوشش می‌دهد. پارامترهای مدل با استفاده از سه رویکرد ماکسیمم درستنمایی شرطی، کمترین توان‌های دوم شرطی تعمیم‌یافته و یول-واکر برآورد شده است. در پایان، کارایی و برتری فرایند مدنظر در برازش داده‌های تعداد فوت ناشی از بیماری COVID-19 نیز، در مقایسه با سایر مدل‌های رقیب بررسی می‌شود.

واژه‌های کلیدی: فرایند INAR(1)، نمایی-وایبل گسسته، شاخص پراکندگی، دم سنگین.
کد موضوع‌بندی ریاضی (۲۰۱۰): 62P10، 62M20.

۱ مقدمه

امروزه با بسیاری از داده‌های گسسته‌مقدار روبرو هستیم که مدل‌سازی این دسته از داده‌ها، برحسب ماهیت گسسته آن‌ها امری ضروری است. در برخی موارد، از مدل‌های تقریبی پیوسته در برازش داده‌های گسسته استفاده می‌شود که

©نویسند(گان). ناشر انجمن آمار ایران است.
این مقاله با دسترسی آزاد تحت شرایط و ضوابط (CC BY-NC 4.0) توزیع شده است.

عملکرد دقیقی را ارائه نمی‌کنند. مدل‌های گسسته در برازش داده‌های شمارشی در علوم مختلفی مانند پزشکی، بیمه، زیست‌شناسی و کشاورزی از اهمیت بسزایی برخوردار است. در سال‌های اخیر، توجه بسیاری به معرفی توزیع‌های گسسته جدید در برازش داده‌های شمارشی، معطوف شده است. مشاهدات شمارشی به صورت اعداد صحیح مثبت که اغلب به صورت متوالی جمع‌آوری می‌گردند، بیان می‌شوند. از سوی دیگر، داده‌های شمارشی به دلیل وجود مشاهدات پرت، اثرات فیزیکی یا پنهان، بیش‌پراکنده و در برخی موارد به صورت همسان یا کم‌پراکنده هستند.

توزیع‌های متعارف دوجمله‌ای، پواسن، هندسی، دوجمله‌ای منفی، به دلیل ماهیت پیچیده و ویژگی‌های پدیده‌های طبیعی مانند چولگی، پراکندگی، نرخ شکست یکنوا یا تک‌مدی، تورم یا عدم تورم، عملکرد مناسبی را نشان نمی‌دهند (المرشدی و همکاران ، ۲۰۲۰؛ الیوا و المرشدی ، ۲۰۲۱). به منظور مدل‌سازی داده‌های شمارشی، بسیاری از توزیع‌های طول عمر پیوسته، به توزیع‌های گسسته تبدیل شده‌اند. در این میان، می‌توان توزیع‌های وایبل گسسته (ناکاگاوا و اساکی ، ۱۹۷۵)، لیندلی گسسته (گومز-دنیز و کالدرین-اوجدا ، ۲۰۱۱)، بور گسسته، پارتو گسسته ،(کریشنا و پوندیر ، ۲۰۰۹) لیندلی گسسته سه پارامتری (الیوا و همکاران ، ۲۰۲۰)، بیلال گسسته (آلتون و همکاران ، ۲۰۲۲) و شبه‌لیندلی گسسته (ایرشاد و همکاران ، ۲۰۲۱) را نام برد.

برای مدل‌سازی مشاهدات شمارشی، فرایندهای خودبازگشتی گسسته‌مقدار مرتبه اول[1] (INAR(1)) مبتنی بر عملگر رقیق‌ساز[2] توسط مکنزی (۱۹۸۶) و الاوش و الزید (۱۹۸۷) معرفی شده است. به منظور برازش بهتر داده‌های به شدت پراکنده یا متورم در صفر، توزیع‌های مختلفی برای نوفه‌ها مانند پواسن‌لیندلی (لیویو و همکاران ، ۲۰۱۸) و بل (هانگ و ژو ، ۲۰۲۱) در نظر گرفته شده است. محمدپور و شیراوژن (۲۰۲۰) فرایند INAR(1) جدیدی را بر اساس عملگر رقیق‌ساز دوجمله‌ای منفی معرفی نمودند که در آن نوفه‌ها به طور متوالی به مقدار فرایند در زمان جاری وابسته است. معرفی عملگر رقیق‌ساز جدید توسط بورگس و همکاران (۲۰۱۷)؛ ویب (۲۰۱۸) و لیو و ژو (۲۰۲۱) مطالعه و بررسی شده است. اخیرا، عملگر رقیق‌ساز دوجمله‌ای منفی وابسته مبتنی بر سری شمارشی هندسی متورم در صفر توسط شما و همکاران (۲۰۲۰) معرفی شده است.

در این مقاله، توزیع نمایی-وایبل گسسته[3] (DEW) با ویژگی‌های برجسته و انعطاف‌پذیری در برازش انواع داده‌های شمارشی معرفی می‌شود. هم‌چنین، فرایند INAR(1) تعمیم‌یافته، مبتنی بر عملگر رقیق‌ساز (شما و همکاران ، ۲۰۲۰)، با نوفه‌های DEW معرفی و کاربرد آن در برازش داده‌های تعداد فوت ناشی از COVID-19 نشان داده می‌شود. در بخش ۲، فرایند INAR(1) با نوفه‌های DEW معرفی شده است. در بخش ۳، پارامترهای فرایند با رویکردهای ماکسیمم درستنمایی شرطی، کمترین توان‌های دوم شرطی تعمیم‌یافته و یول-واکر برآورد شده است. مطالعات شبیه‌سازی مونت‌کارلو و مقایسه روش‌های مختلف برآورد پارامترهای فرایند مدنظر در بخش ۴ ارائه می‌گردد. در بخش ۵، عملکرد فرایند در برازش داده‌های COVID-19 و پیش‌بینی کلاسیک فرایند گزارش می‌شود.

---

[1] First-order Integer-valued Autoregressive Processes
[2] Thinning operator
[3] Exponential-Weibull Distribution

۲ معرفی فرایند INAR(1) تعمیم‌یافته با نوفه‌های نمایی-وایبل گسسته

ابتدا فرایند INAR(1) تعمیم‌یافته جدیدی بر اساس عملگر رقیق‌ساز دوجمله‌ای منفی تعمیم‌یافته (GNB) معرفی و سپس مرور مختصری بر روی عملگر رقیق‌ساز دوجمله‌ای منفی تعمیم‌یافته ارائه می‌شود.

تعریف ۱. (شما و همکاران ، ۲۰۲۰)دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع $\{V_i\}_{i\in\mathbb{N}}$ از توزیع هندسی با پارامتر $\frac{\theta}{1+\theta}$ و متغیر تصادفی برنولی W با پارامتر $\frac{\alpha}{\theta}$، $0 \leq \alpha \leq \theta \leq 1$ را در نظر بگیرید، به طوری‌که V_i و W برای هر $i \in \mathbb{N}$ مستقل هستند. دنباله‌ای از متغیرهای تصادفی وابسته $\{U_i\}_{i\in\mathbb{N}}$ را به صورت $U_i = V_iW$ تعریف کنید. متغیر تصادفی U_i دارای توزیع آمیخته زیر است

$$P(U_i = u) = \begin{cases} 1 - \frac{\alpha}{\theta} & u = 0 \\ (\frac{\alpha}{\theta})\frac{\theta^u}{(1+\theta)^{u+1}} & u = 0, 1, \ldots \end{cases} . \qquad (1)$$

متغیر تصادفی $U_1 + \ldots + U_n$ دارای توزیع آمیخته تباهیده در صفر و دوجمله‌ای منفی با پارامترهای $(n, \frac{\theta}{1+\theta})$ به ترتیب، با نسبت‌های $1 - \frac{\alpha}{\theta}$ و $\frac{\alpha}{\theta}$ است.

تعریف ۲. متغیر تصادفی گسسته‌مقدار نامنفی X و دنباله وابسته از متغیرهای تصادفی $\{U_i\}_{i\in\mathbb{N}}$ مطابق با توزیع (۱) را در نظر بگیرید. عملگر رقیق‌ساز GNB "$*_\theta$"، به صورت $\alpha *_\theta X = \sum_{i=1}^{X} U_{i,t}$، $0 \leq \alpha \leq \theta \leq 1$ تعریف می‌شود.

فرایند INAR(1) تعمیم‌یافته $\{X_t\}$ با استفاده از معادله بازگشتی

$$X_t = \alpha *_\theta X_{t-1} + Z_t, \qquad \alpha < \theta < 1, \qquad t = 1, 2, \ldots, \qquad (2)$$

ارائه می‌شود، که در آن "$*_\theta$" دلالت بر عملگر رقیق‌ساز GNB و $\{Z_t\}$ دنباله‌ای از متغیرهای تصادفی گسسته است. متغیرهای تصادفی $\alpha *_\theta X_{t-1}$ و Z_t به شرط X_{t-1} از یکدیگر مستقل هستند.

۲.۱ توزیع نمایی-وایبل گسسته

فرض کنید $G(x)$ تابع توزیع تجمعی متغیر تصادفی X و $q(t)$ تابع چگالی احتمال چگالی احتمال متغیر تصادفی T باشد، تابع توزیع تجمعی خانواده T-X از توزیع‌ها برای $-\infty \leq a < b \leq -\infty$ و $x \in [a, b]$ به صورت $H(x) = \int_a^{V(G(x))} q(t)dt$ است.

توزیع $q(t)$ را نمایی با پارامتر η و تابع $V(G(x)) = \frac{G(x)}{1-G(x)}$ در نظر بگیرید، به طوری‌که $G(x, \beta, \nu) = 1 - e^{-\frac{x^\beta}{\nu}}$ تابع توزیع وایبل با پارامترهای (β, ν) است. آن‌گاه توزیع نمایی-وایبل (EW) برای $\beta, \eta, \nu > 0$ و

$x \geq 0$ به صورت

$$H(x,\beta,\eta,\nu) = \int_0^{\frac{G(x)}{1-G(x)}} \eta e^{-\eta t} dt = 1 - e^{-\eta(e^{\frac{x^\beta}{\nu}}-1)}, \qquad (3)$$

و تابع چگالی احتمال آن به صورت $h(x,\beta,\eta,\nu) = \frac{\beta\eta}{\nu} x^{\beta-1} e^{\frac{x^\beta}{\nu}} e^{-\eta(e^{\frac{x^\beta}{\nu}}-1)}$ معرفی می‌شود.

فرض کنید T متغیر تصادفی طول عمر با توزیع نمایی با تابع توزیع $Q(\cdot)$ و متغیر تصادفی X بیانگر نسبت بخت است، ریسک این‌که سیستم با طول عمر T، در زمان x از کار بیفتد، برابر با $\frac{G(x)}{1-G(x)}$ است. با استفاده از رابطه (۳) نسبت بخت به صورت $F_Y(y) = P\big(X \leq \frac{G(x)}{1-G(x)}\big) = Q\big(\frac{G(x)}{1-G(x)}\big)$ مدل‌سازی می‌شود (علیزاده و همکاران ، ۲۰۱۷).

متغیر نامنفی پیوسته X از توزیع EW با تابع بقای $S(x,\beta,\eta,\nu) = e^{-\eta(e^{\frac{x^\beta}{\nu}}-1)}$ را در نظر بگیرید. به منظور معرفی همتای گسسته EW، متغیر تصادفی X را به فاصله‌های واحد تقسیم‌بندی نموده و متغیر تصادفی گسسته Y را به صورت $Y = [X]$ تعریف نمایید. آن‌گاه تابع جرم احتمال برای $y = 0, 1, 2, \ldots$ به صورت $f_Y(y,\beta,\eta,\nu) = e^{\eta}(e^{-\eta e^{\frac{y^\beta}{\nu}}} - e^{-\eta e^{\frac{(y+1)^\beta}{\nu}}})$ است. به منظور تسهیل در نمایش، با تغییر متغیرهای $e^{-\eta} = \lambda$ و $e^{\frac{1}{\nu}} = \gamma$ داریم

$$f_Y(y,\beta,\lambda,\gamma) = \frac{1}{\lambda}\big(\lambda^{\gamma^{y^\beta}} - \lambda^{\gamma^{(y+1)^\beta}}\big), \quad 0 < \lambda < 1,\ \gamma \geq 1,\ 0 < \beta < \infty. \qquad (4)$$

توابع توزیع تجمعی، بقا و نرخ خطر توزیع DEW به ترتیب عبارتند از

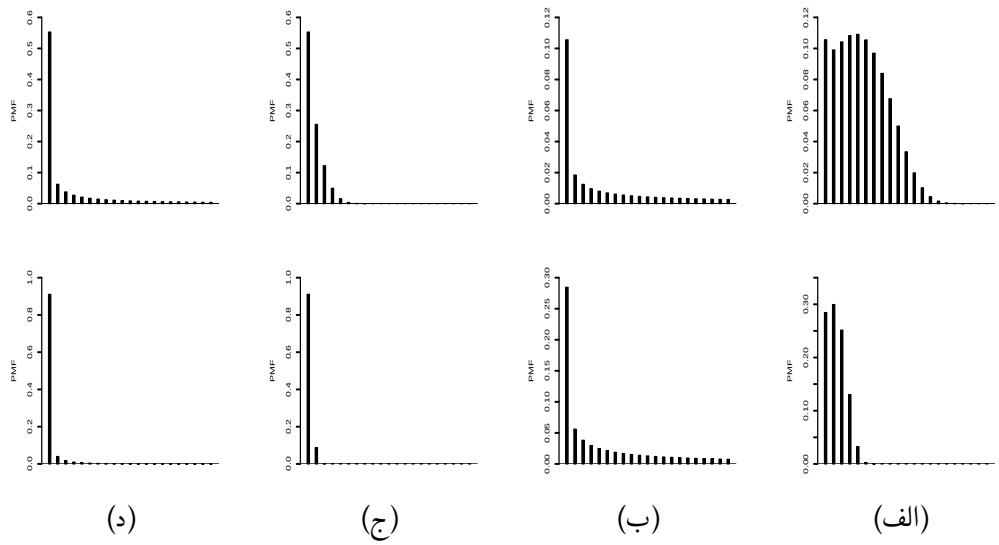
$$\begin{aligned} F(y,\beta,\lambda,\gamma) &= 1 - \lambda^{\gamma^{(y+1)^\beta}-1}, \\ S(y,\beta,\lambda,\gamma) &= \lambda^{\gamma^{(y+1)^\beta}-1}, \\ Hzr(y,\beta,\lambda,\gamma) &= \frac{f(y,\beta,\lambda,\gamma)}{S(y-1,\beta,\lambda,\gamma)} = 1 - \lambda^{(\gamma^{(y+1)^\beta} - \gamma^{y^\beta})}. \end{aligned}$$

نمودارهای توابع جرم احتمال و نرخ خطر توزیع DEW در شکل‌های ۱ و ۲ یکنوایی یا تک‌مدی بودن توابع جرم احتمال و نرخ خطر، توزیع DEW ویژگی‌های منعطفی را نشان می‌دهند. بنابراین در برازش انواع مختلف داده‌های شمارشی کاربرد دارد.

فرع ۱. متغیر تصادفی Z از توزیع DEW با پارامترهای (β,λ,γ) را در نظر بگیرید

الف) تابع مولد گشتاور توزیع DEW به‌صورت $M_Z(t) = \frac{1-e^{-t}}{\lambda}\sum_{z=0}^{\infty} e^{tz}\lambda^{\gamma^{z^\beta}} + e^{-t}$ است.

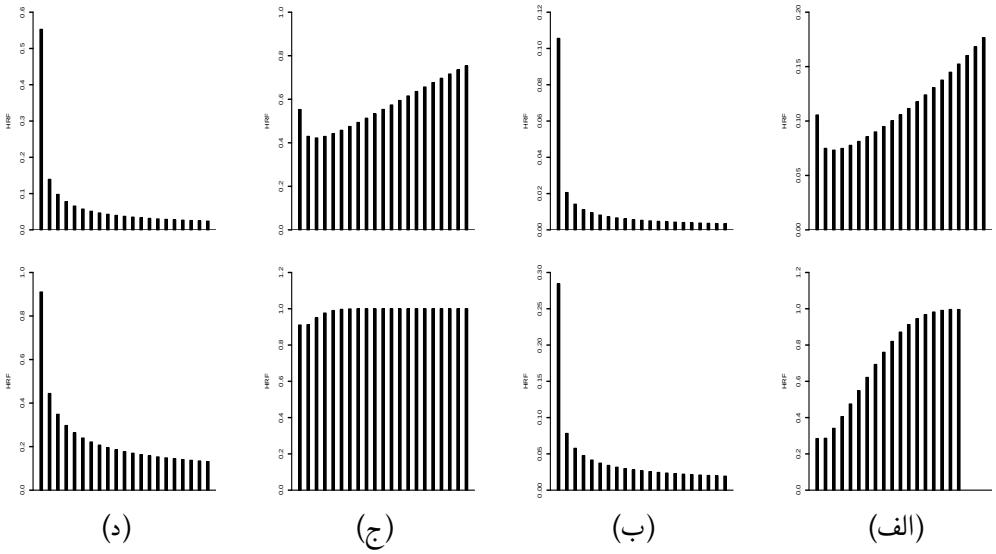
ب) گشتاور غیرمرکزی r-ام به‌صورت $\mathrm{E}(Z^r) = \frac{1}{\lambda}\sum_{z=0}^{\infty}(z^r - (z-1)^r)\lambda^{\gamma^{z^\beta}} + (-1)^r$ است.



شکل ۱. تابع جرم احتمال توزیع DEW، سطر اول $\gamma = ۱/۵$ و سطر دوم $\gamma = ۲/۵$، الف- $(\lambda, \beta) = (۰/۸, ۰/۸)$، ب- $(\lambda, \beta) = (۰/۸, ۰/۲)$ ، ج- $(\lambda, \beta) = (۰/۲, ۰/۸)$ ، د- $(\lambda, \beta) = (۰/۲, ۰/۲)$

لازم به ذکر است که مجموع‌های نامتناهی در روابط فوق، به صورت عددی در نرم‌افزار R محاسبه می‌شود. لذا گشتاورهای مرتبه اول تا چهارم توزیع DEW به راحتی قابل دسترسی است. از گشتاورهای مذکور در محاسبه برخی شاخص‌ها از قبیل میانگین، واریانس، چولگی و کشیدگی استفاده می‌شود.

شاخص پراکندگی فیشر (FDI)، به صورت نسبت واریانس به میانگین تعریف می‌شود و بیانگر این است که توزیع مدنظر، برای مجموعه داده‌هایی با بیش‌پراکندگی، کم‌پراکندگی یا پراکندگی یکسان (همسان)، مناسب است. اگر $FDI < ۱$، توزیع کم‌پراکنده، اگر $FDI > ۱$، توزیع بیش‌پراکنده و اگر $FDI = ۱$، توزیع دارای پراکندگی یکسان است. چولگی و کشیدگی نیز، شاخص‌های شکل توزیع هستند که مستقیما توسط گشتاورهای سوم و چهارم محاسبه می‌شوند. مقادیر مثبت چولگی موید دم بلند راست و مقادیر منفی نیز موید دم بلند چپ است. معیار کشیدگی نیز بیانگر پخی یا کشیدگی توزیع (دم‌سنگین یا دم‌سبک) است. معیارهای مذکور برای توزیع DEW و مقادیر مختلف پارامترها در جدول ۱ ارائه شده است. طبق جدول ۱، توزیع DEW، انواع مختلفی از پراکندگی‌ها را شامل می‌شود، زیرا مقادیر FDI، هم شامل اعداد کوچکتر، تقریبا مساوی و بزرگتر از یک است. از سوی دیگر، عدم تقارن توزیع و چولگی به سمت راست (دم راست بلند) و کشیدگی شدید (دم‌سنگین) در توزیع نیز مشاهده می‌گردد. با توجه به جامعیت توزیع DEW در پوشش انواع مختلفی از پراکندگی‌ها و هم‌چنین عدم تقارن و کشیدگی، لذا توزیع DEW در برازش بسیاری از داده‌های گوناگون در زمینه‌های مختلف کاربرد دارد.



شکل ۲. تابع نرخ خطر توزیع توزیع DEW، سطر اول $\gamma = ۱/۵$ و سطر دوم $\gamma = ۲/۵$، الف- $(\lambda, \beta) = (۰/۸, ۰/۸)$، ب- $(\lambda, \beta) = (۰/۸, ۰/۲)$ ، ج- $(\lambda, \beta) = (۰/۲, ۰/۸)$ ، د- $(\lambda, \beta) = (۰/۲, ۰/۲)$

جدول ۱. برخی شاخص‌های آماری توزیع DEW برای مقادیر مختلف پارامترها.

	$\gamma = ۱/۵$					$\gamma = ۲/۵$				
	(λ, β)					(λ, β)				
معیارها	$(۰/۲, ۰/۲)$	$(۰/۲, ۰/۸)$	$(۰/۵, ۰/۵)$	$(۰/۸, ۰/۲)$	$(۰/۸, ۰/۸)$	$(۰/۲, ۰/۲)$	$(۰/۲, ۰/۸)$	$(۰/۵, ۰/۵)$	$(۰/۸, ۰/۲)$	$(۰/۸, ۰/۸)$
میانگین	۲۳/۴۱۶	۰/۷۳۴	۴/۵۷۱	۲۰۶۱/۲	۴/۵۱۸	۰/۰۸۲	۰/۰۴۱	۰/۳۴۹	۱۷/۴۱۲	۰/۹۷۲
واریانس	۸۷۰۲/۱۳۴	۱/۰۲۵	۳۲/۶۱۵	$۹/۲۳۸e^{۶}$	۹/۸۳۶	۰/۳۱۹	۰/۰۳۹	۰/۴۸۴	۱۲۳۴/۹	۰/۸۰۹
FDI	۳۷۱/۶۳۵	۱/۳۹۵	۷/۱۳۵	$۴/۴۸۱e^{۳}$	۲/۱۷۶۸	۳/۸۹۸	۰/۹۶۴	۱/۳۸۶	۷۰/۹۲۴	۰/۸۳۲
چولگی	۱۱/۲۱۸	۱/۵۱۵	۱/۸۱۹	۱/۶۱۷	۰/۴۱۵۵	۱۳/۹۳۷	۴/۷۲۵	۲/۳۲۳	۴/۰۷۷	۰/۵۵۹
کشیدگی	۲۴۴/۴۱۱	۱۲/۶۶۹	۲۱/۱۴۴	۱۵/۱۰۹	۴۱/۳۵۴	۳۴۴/۶۲	۲۳/۶۱۱	۱۳/۴۳۶	۳۹/۳۴۹	۱۳/۹۹۱

۲.۲ فرایند INAR(1) تعمیم‌یافته با نوفه‌هایی از توزیع DEW

معادله بازگشتی (۲) را در نظر بگیرید، که در آن $\{Z_t\}$ دنباله‌ای از متغیرهای تصادفی با توزیع DEW و پارامترهای (β, λ, γ) است. فرایند جدید به اختصار DEW-INAR(1) نامیده می‌شود. طبق (۴)، تابع انتقال مرتبه یک به

صورت $P_{\circ j} = P\big(X_t = j|X_{t-۱} = \circ\big) = f_Z(j, \beta, \lambda, \gamma)$ ارائه می‌گردد و برای $i \geq ۱$،

$$\begin{aligned} P_{ij} &= P\big(X_t = j|X_{t-۱} = i\big) = P\big(\alpha *_\theta X_{t-۱} + Z_t = j|X_{t-۱} = i\big) \\ &= (۱ - \frac{\alpha}{\theta}) f_Z(j, \beta, \lambda, \gamma) + \frac{\alpha}{\theta} \sum_{k=\circ}^{j} \binom{i+k-۱}{i-۱} \frac{\theta^k}{(۱+\theta)^{i+k}} f_Z(j-k, \beta, \lambda, \gamma). \end{aligned}$$

در تساوی آخر، با استفاده از تعریف ۲، $\alpha *_\theta X_{t-۱}$ با احتمال $۱ - \frac{\alpha}{\theta}$ برابر با صفر و با احتمال $\frac{\alpha}{\theta}$ دارای توزیع دوجمله‌ای منفی با پارامتر $(i, \frac{\theta}{۱+\theta})$ است.

لم ۱. فرایند $\{X_t\}$ در (۲) ناکاهشی، دوره‌ای و زنجیر مارکف بازگشتی مثبت است، از این‌رو ایستا و ارگودیک است.

برهان: به راحتی مشاهده می‌شود که $\{X_t\}$ یک زنجیر مارکف با $P_{ij} > \circ$ است، بنابراین $\{X_t\}$ ناکاهشی و دوره‌ای است. با استفاده از تکرار (۲)، نتیجه می‌شود

$$X_t \stackrel{d}{=} \alpha *_\theta^{(k)} X_{t-k} + \sum_{i=۱}^{k-۱} \alpha *_\theta^{(i-۱)} Z_{t-i} + Z_t, \tag{۵}$$

که در آن به منظور سهولت نگارش، $\alpha *_\theta^{(i)} = \underbrace{\alpha *_\theta \alpha *_\theta \ldots \alpha *_\theta}_{\text{i مرتبه}}$ در نظر گرفته می‌شود. عبارت اول در سمت راست (۵)، معادل با $o_p(۱)$ است که $\mathrm{E}(\alpha *_\theta^{(k)} X_t) = \alpha^k \mathrm{E}(X_t) \xrightarrow{k\to\infty} \circ$، لذا $X_t \stackrel{d}{=} o_p(۱) + V_k$ و در آن $V_k = \sum_{i=۱}^{k-۱} \alpha *_\theta^{(i-۱)} Z_{t-i} + Z_t$ است. در ادامه، برای هر $\delta > \circ,\ k, l \in \mathbb{N}$ ، عبارت V_k به صورت قریب به یقین همگرا است.

$$\begin{aligned} P\big(\max_{۱\leq q\leq \ell} |V_k - V_{k+q}| > \delta\big) &= P(V_{k+\ell} - V_k > \delta) \\ &= P\big(\sum_{i=۱}^{k+\ell-۱} \alpha *_\theta^{(i-۱)} Z_{t-i} - \sum_{i=۱}^{k-۱} \alpha *_\theta^{(i-۱)} Z_{t-i} > \delta\big) \\ &\leq \frac{۱}{\delta} \mathrm{E}\big[\sum_{i=۱}^{k+\ell-۱} \alpha *_\theta^{(i-۱)} Z_{t-i} - \sum_{i=۱}^{k-۱} \alpha *_\theta^{(i-۱)} Z_{t-i}\big] \\ &= \frac{۱}{\delta}\big[\sum_{i=۱}^{k+\ell-۱} \alpha^i \mu_Z - \sum_{i=۱}^{k-۱} \alpha^i \mu_Z\big] \\ &= \frac{۱}{\delta} \sum_{i=k}^{k+\ell-۱} \alpha^i \mu_Z = \frac{\mu_Z \alpha^k (۱ - \alpha^\ell)}{\delta(۱-\alpha)} \xrightarrow{k\to\infty} \circ, \end{aligned}$$

که در آن $\mu_Z = \mathrm{E}(Z)$ در فرع ۱ تعریف شده است. لذا V_k قریب به یقین همگرا است و $V = \lim_{k\to\infty} V_k$.
با استفاده از (۵) و قضیه اسلاتسکی، می‌توان نتیجه گرفت که X_t در توزیع به V همگرا است. هم‌چنین،

$$\begin{aligned}\lim_{k\to\infty} P_{ij}^{(k)} &= \lim_{k\to\infty} P(X_{t+k}=j|X_t=i) = \lim_{k\to\infty} P(o_p(1)+V_k=j|X_t=i)\\ &= \lim_{k\to\infty} P(V_k=j|X_t=i) = \lim_{k\to\infty} P(V_k=j) = \lim_{k\to\infty} P(X_{t+k}=j) > 0,\end{aligned}$$

لذا $\lim_{k\to\infty} P_{ij}^{(k)} \neq 0$ و $\{X_t\}$ بازگشتی مثبت و ارگودیک است که ایستایی فرایند را نتیجه می‌دهد.

گزاره ۱. فرایند ایستای $\{X_t\}$ ارائه شده در (۲) را در نظر بگیرید. برای $\alpha < \theta < 1$ و $t \geq 1$ داریم
(الف) امید شرطی فرایند به صورت

$$\mathrm{E}\big(X_t|X_{t-k}\big) = \alpha^k X_{t-k} + \frac{1-\alpha^k}{1-\alpha}\mu_Z, \tag{۶}$$

است که در آن μ_Z، امید متغیر تصادفی Z ارائه شده در فرع ۱ است.
(ب) واریانس شرطی فرایند به صورت

$$\mathrm{Var}(X_t|X_{t-1}) = \alpha(\theta-\alpha)X_{t-1}^2 + \alpha(\theta+1)X_{t-1} + \sigma_Z^2, \tag{۷}$$

نشان داده می‌شود و

$$\begin{aligned}\mathrm{Var}\big(X_t|X_{t-k}\big) &= \alpha^k\big(\theta^k-\alpha^k\big)X_{t-k}^2 + \alpha^k\frac{(1+\theta)(1-\theta^k)}{1-\theta}X_{t-k}\\ &+2\mu_Z\alpha^k\big(\frac{1-\theta^k}{1-\theta}-\frac{1-\alpha^k}{1-\alpha}\big)X_{t-k}\\ &+\frac{\alpha\mu_Z(1+\theta)}{1-\theta}\big(\frac{1-\alpha^{k-1}}{1-\alpha}-\frac{\theta(1-(\alpha\theta)^{k-1})}{1-\alpha\theta}\big)\\ &+\mu_Z^2\big(\frac{1-(\alpha\theta)^{k-1}}{1-\alpha\theta}-\frac{1-\alpha^{2k-2}}{1-\alpha^2}\big)\\ &+2\alpha\mu_Z^2\Bigg[\frac{\theta}{1-\theta}\big(\frac{\alpha(1-\alpha^{k-1})}{1-\alpha}-\frac{\alpha\theta(1-(\alpha\theta)^{k-1})}{1-\alpha\theta}\big)\\ &-\frac{\alpha}{1-\alpha}\big(\frac{\alpha(1-\alpha^{k-1})}{1-\alpha}-\frac{\alpha\theta(1-\alpha^{2k-2})}{1-\alpha^2}\big)\Bigg]+\sigma_Z^2\frac{1-(\alpha\theta)^k}{1-\alpha\theta},\end{aligned}$$

که در آن σ_Z^2، واریانس متغیر تصادفی Z ارائه شده در فرع ۱ است.
(ج) تابع خودهمبستگی فرایند به صورت $R(k) = \mathrm{Corr}(X_t, X_{t-k}) = \alpha^k$ نشان داده می‌شود.

برهان: الف) امید شرطی فرایند به صورت

$$\begin{aligned}
\mathrm{E}(X_t|X_{t-1}) &= \alpha X_{t-1} + \mu_Z, \\
\mathrm{E}(X_t|X_{t-2}) &= \mathrm{E}(\mathrm{E}(X_t|X_{t-1})|X_{t-2}) = \alpha^2 X_{t-2} + (1+\alpha)\mu_Z,
\end{aligned}$$

محاسبه می‌شود. با استفاده از استقرا، می‌توان نتیجه گرفت $\mathrm{E}(X_t|X_{t-k}) = \alpha^k X_{t-k} + \frac{(1-\alpha^k)\mu_Z}{1-\alpha}$.
(ب) واریانس شرطی فرایند با استفاده از روش واریانس مکرر به دست می‌آید

$$\begin{aligned}
\mathrm{Var}(X_t|X_{t-1}) &= \mathrm{Var}(\alpha *_\theta X_{t-1}) + \mathrm{Var}(Z_t) \\
&= \alpha(\theta-\alpha)X_{t-1}^2 + \alpha(\theta+1)X_{t-1} + \sigma_Z^2, \\
\mathrm{Var}\big(X_t|X_{t-2}\big) &= \mathrm{E}(\mathrm{Var}(X_t|X_{t-1})|X_{t-2}) + \mathrm{Var}(\mathrm{E}(X_t|X_{t-1})|X_{t-2}) \\
&= \alpha^2(\theta^2-\alpha^2)X_{t-2}^2 + \alpha^2(\theta+1)^2 X_{t-2} + 2\alpha^2(\theta-\alpha)\mu_Z X_{t-2} \\
&\quad +\alpha(\theta+1)\mu_Z + \alpha(\theta-\alpha)\mu_Z^2 + (1+\alpha\theta)\sigma_Z^2, \\
\mathrm{Var}\big(X_t|X_{t-3}\big) &= \alpha^3(\theta^3-\alpha^3)X_{t-3}^2 + \alpha^3(\theta+1)(1+\theta+\theta^2)X_{t-3} \\
&\quad +2\alpha^3\big((\theta-\alpha)+(\theta^2-\alpha^2)\big)\mu_Z X_{t-3} \\
&\quad +\alpha(\theta+1)(1+\alpha(1+\theta))\mu_Z + \big[\alpha(\theta-\alpha)+\alpha^2(\theta^2-\alpha^2)\big]\mu_Z^2 \\
&\quad +2\alpha^2(\theta-\alpha)\mu_Z^2 + (1+\alpha\theta+\alpha^2\theta^2)\sigma_Z^2.
\end{aligned}$$

با استفاده از اصل استقرا نتیجه می‌شود که

$$\begin{aligned}
\mathrm{Var}(X_t|X_{t-k}) &= \alpha^k\big(\theta^k-\alpha^k\big)X_{t-k}^2 + \alpha^k(1+\theta)\sum_{i=0}^{k-1}\theta^i X_{t-k} \\
&\quad +2\mu_Z\alpha^k\sum_{i=0}^{k-1}\big(\theta^i-\alpha^i\big)X_{t-k} + \mu_Z(1+\theta)\sum_{i=1}^{k-1}\alpha^i\sum_{j=0}^{i-1}\theta^j \\
&\quad +\mu_Z^2\sum_{i=0}^{k-1}\alpha^i(\theta^i-\alpha^i) + 2\alpha\mu_Z^2\sum_{i=1}^{k-1}\alpha^i\sum_{j=1}^{i}(\theta^j-\alpha^j) + \sigma_Z^2\sum_{i=0}^{k-1}(\alpha\theta)^i.
\end{aligned}$$

(ج) اثبات به راحتی به دست می‌آید و لذا حذف شده است.

تذکر ۱. اگر $k \to \infty$، آنگاه $\lim_{k\to\infty} \mathrm{E}\big(X_t \mid X_{t-k}\big) = \dfrac{\mu_Z}{1-\alpha}$ و

$$\lim_{k\to\infty} \mathrm{Var}\big(X_t \mid X_{t-k}\big) = \frac{\alpha\mu_Z(1+\theta)}{(1-\alpha)(1-\alpha\theta)} + \frac{\alpha(\theta-\alpha)\mu_Z^2}{(1-\alpha)^2(1-\alpha\theta)} + \frac{\sigma_Z^2}{1-\alpha\theta},$$

که به ترتیب امید غیرشرطی و واریانس غیرشرطی فرایند DEW-INAR(1) هستند.

۳ برآورد پارامترهای فرایند DEW-INAR(1)

برآوردگرهای کمترین توان‌های دوم شرطی تعمیم‌یافته (MCLS) پارامترهای فرایند DEW-INAR(1)، با حداقل نمودن تابع

$$Q(\delta_1) = \sum_{t=2}^{n} \big(X_t - \mathrm{E}(X_t|X_{t-1})\big)^2 = \sum_{t=2}^{n} \big(X_t - \alpha X_{t-1} - \mu_Z\big)^2,$$

به دست می‌آید، که در آن $\delta_1 = (\alpha, \mu_Z)$ است. برآوردگرهای MCLS به صورت

$$\begin{aligned}
\hat{\alpha}_{MCLS} &= \frac{(n-1)\sum_{t=2}^{n} X_t X_{t-1} - \sum_{t=2}^{n} X_t \sum_{t=2}^{n} X_{t-1}}{(n-1)\sum_{t=2}^{n} X_{t-1}^2 - (\sum_{t=2}^{n} X_{t-1})^2}, \\
\hat{\mu}_{Z,MCLS} &= \frac{\sum_{t=2}^{n} X_t - \hat{\alpha}_{MCLS}\sum_{t=2}^{n} X_{t-1}}{n-1},
\end{aligned}$$

حاصل می‌شود. برآوردهای MCLS پارامترهای (β, λ, γ) با حل عددی $\hat{\mu}_{Z,MCLS}$ به دست می‌آید.

امید شرطی مرتبه اول فرایند تابعی از پارامتر θ نیست، لذا به منظور برآورد پارامتر θ از روش تعمیم‌یافته (کارلسن و تجوستیم ، ۱۹۸۸) به صورت $S(\theta) = \sum_{t=2}^{n} \big(V_t - \mathrm{Var}(X_t|X_{t-1})\big)^2$، استفاده می‌شود، که در آن $V_t = (X_t - \mathrm{E}(X_t|X_{t-1}))^2 = \big(X_t - \hat{\alpha}_{MCLS}X_{t-1} - \hat{\mu}_{Z,MCLS}\big)^2$ و $\mathrm{Var}(X_t|X_{t-1})$ در (۷) معرفی شده است. پارامترها با مقادیر برآوردهای متناظرشان جایگزین شده و به صورت $\mathrm{Var}(X_t|X_{t-1}) = \hat{\alpha}_{MCLS}(\theta - \hat{\alpha}_{MCLS})X_{t-1}^2 + \hat{\alpha}_{MCLS}(\theta+1)X_{t-1} + \hat{\sigma}_{Z,MCLS}^2$ نشان داده می‌شود که σ_Z^2 تابعی از پارامترهای (β, λ, γ) است و توسط $(\hat{\beta}_{MCLS}, \hat{\lambda}_{MCLS}, \hat{\gamma}_{MCLS})$ برآورد می‌شود. از این‌رو، برآوردگر MCLS پارامتر θ به صورت

$$\hat{\theta}_{MCLS} = \frac{\sum_{t=2}^{n}(V_t + \hat{\alpha}_{MCLS}^2 X_{t-1}^2 - \hat{\alpha}_{MCLS}X_{t-1} - \hat{\sigma}_{Z,MCLS}^2)(X_{t-1}^2 + X_{t-1})}{\hat{\alpha}_{MCLS}\sum_{t=2}^{n}(X_{t-1}^2 + X_{t-1})^2},$$

به دست می‌آید با توجه به $\mathrm{E}(X_t) = \frac{\mu_Z}{1-\alpha}$ و $\mathrm{Corr}(X_t, X_{t-1}) = \alpha$ و با استفاده از میانگین نمونه و تابع خودهمبستگی نمونه، برآوردگرهای یول واکر (YW) پارامترهای (α, μ_Z) به صورت

$$\hat{\alpha}_{YW} = \frac{\sum_{t=2}^{n}(X_t - \overline{X})(X_{t-1} - \overline{X})}{\sum_{t=1}^{n}(X_t - \overline{X})^2}, \quad \hat{\mu}_{Z,YW} = \overline{X}(1 - \hat{\alpha}_{YW}),$$

به دست می‌آید. همان‌طور که قبلا نیز اشاره شد، برآوردهای YW پارامترهای (β, λ, γ) را می‌توان براساس برآورد μ_Z به دست آورد. به منظور برآورد پارامتر θ، امید مرتبه دوم فرایند را در نظر بگیرید

$$\begin{aligned} \mathrm{E}(X_t^2) &= \alpha\theta\mathrm{E}(X_t^2) + \alpha(1+\theta)\mathrm{E}(X_t) + \mathrm{E}(Z_t^2) + 2\alpha\mu_Z\mathrm{E}(X_t) \\ &= \frac{\alpha(1+\theta)}{1-\alpha\theta}\mathrm{E}(X_t) + \frac{\mathrm{E}(Z_t^2)}{1-\alpha\theta} + \frac{2\alpha\mu_Z\mathrm{E}(X_t)}{1-\alpha\theta}, \end{aligned}$$

و با توجه به $\overline{X^2} = \frac{1}{n}\sum_{t=1}^{n} X_t^2$، آن‌گاه

$$\overline{X^2} = \frac{\hat{\alpha}_{YW}(1+\theta)}{1-\hat{\alpha}_{YW}\theta}\overline{X} + \frac{\hat{\mu}_{Z,YW}^2 + \hat{\sigma}_{Z,YW}}{1-\hat{\alpha}_{YW}\theta} + \frac{2\hat{\alpha}_{YW}\hat{\mu}_{Z,YW}\overline{X}}{1-\hat{\alpha}_{YW}\theta}, \tag{8}$$

بنابراین، برآورد θ با استفاده از محاسبه عددی ریشه معادله (۸) به دست می‌آید.

۴ مطالعه شبیه‌سازی

کارایی برآوردگرهای فرایند DEW-INAR(1) براساس شبیه‌سازی مونت‌کارلو تحت مقادیر مختلف حجم نمونه $n = 100, 200, 500, 1000$ با تکرار $h = 100$، مقایسه می‌گردد. هم‌چنین، دو ترکیب مختلف از پارامترهای مدل به صورت $(\alpha, \theta, \lambda, \gamma, \beta) = (0.6, 0.8, 0.3, 1.5, 0.7)$ و $(\alpha, \theta, \lambda, \gamma, \beta) = (0.2, 0.5, 0.7, 2.5, 0.3)$ در نظر گرفته می‌شود. برای ارزیابی عملکرد برآوردگرها، از معیار میانگین توان‌های دوم خطا (MSE)[1] استفاده می‌شود. نتایج در جداول ۲ تا ۴ موید همگرایی برآوردهای CML، MCLS و YW به مقادیر واقعی پارامترها است. علاوه بر این، با افزایش حجم نمونه، مقادیر MSE به تدریج کاهش می‌یابد. طبق جدول ۲، برآورد CML برای مقادیر کوچک پارامترها از عملکرد ضعیف‌تری برخوردار است و با افزایش مقدار α و θ، برآوردهای دقیق‌تری حاصل می‌شود. در میان روش‌های مختلف برآورد نیز، همان‌طور که انتظار می‌رود، روش برآورد پارامتری CML نسبت به روش‌های برآورد ناپارامتری MCLS و YW، دقیق‌ترین نتایج را برحسب مقادیر کمینه MSE، ارائه می‌نماید.

[1] Mean squared error

جدول ۲. برآورد CML پارامترهای مدل DEW-INAR(1) و مقادیر MSE در پرانتز.

n	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\beta}$
	$(\alpha,\theta,\lambda,\gamma,\beta) = (۰/۶, ۰/۸, ۰/۳, ۱/۵, ۰/۷)$				
۱۰۰	۰/۵۸۱(۰/۰۱۲)	۰/۷۹۴(۰/۰۲۲)	۰/۲۶۶(۰/۰۱۵)	۱/۳۹۵(۰/۲۵۹)	۰/۶۸۴(۰/۰۱۱)
۲۰۰	۰/۵۸۱(۰/۰۰۵)	۰/۷۹۳(۰/۰۰۸)	۰/۳۴۱(۰/۰۱۰)	۱/۴۳۳(۰/۲۱۶)	۰/۶۸۸(۰/۰۰۹)
۵۰۰	۰/۵۹۹(۰/۰۰۲)	۰/۸۱۱(۰/۰۰۳)	۰/۳۱۱(۰/۰۰۸)	۱/۴۷۴(۰/۱۸۶)	۰/۶۹۳(۰/۰۰۴)
۱۰۰۰	۰/۵۹۹(۰/۰۰۱)	۰/۸۰۰(۰/۰۰۲)	۰/۳۰۱(۰/۰۰۴)	۱/۴۹۹(۰/۱۰۸)	۰/۷۰۰(۰/۰۰۱)
	$(\alpha,\theta,\lambda,\gamma,\beta) = (۰/۲, ۰/۵, ۰/۷, ۲/۵, ۰/۳)$				
۱۰۰	۰/۲۱۹(۰/۰۱۱)	۰/۵۲۹(۰/۰۲۶)	۰/۶۳۹(۰/۰۱۷)	۲/۲۶۰(۰/۶۳۳)	۰/۲۴۶(۰/۰۱۶)
۲۰۰	۰/۲۰۱(۰/۰۰۷)	۰/۵۱۳(۰/۰۱۲)	۰/۶۸۶(۰/۰۱۵)	۲/۳۵۶(۰/۵۷۹)	۰/۲۵۲(۰/۰۰۹)
۵۰۰	۰/۱۹۹(۰/۰۰۵)	۰/۴۹۸(۰/۰۰۸)	۰/۷۱۲(۰/۰۰۷)	۲/۴۷۶(۰/۳۳۵)	۰/۲۷۴(۰/۰۰۷)
۱۰۰۰	۰/۲۰۱(۰/۰۰۳)	۰/۵۰۱(۰/۰۰۳)	۰/۶۹۵(۰/۰۰۶)	۲/۴۹۸(۰/۱۵۸)	۰/۲۹۶(۰/۰۰۴)

جدول ۳. برآورد MCLS پارامترهای مدل DEW-INAR(1) و مقادیر MSE در پرانتز.

n	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\beta}$
	$(\alpha,\theta,\lambda,\gamma,\beta) = (۰/۶, ۰/۸, ۰/۳, ۱/۵, ۰/۷)$				
۱۰۰	۰/۵۳۶(۰/۰۱۹)	۰/۶۸۷(۰/۰۳۶)	۰/۴۶۲(۰/۰۶۹)	۲/۰۶۶(۰/۶۲۵)	۰/۶۳۳(۰/۰۳۰)
۲۰۰	۰/۵۵۱(۰/۰۰۹)	۰/۷۰۲(۰/۰۲۹)	۰/۴۲۷(۰/۰۶۶)	۱/۷۱۸(۰/۳۵۳)	۰/۶۴۶(۰/۰۲۸)
۵۰۰	۰/۵۸۶(۰/۰۰۸)	۰/۷۵۹(۰/۰۰۶)	۰/۳۶۲(۰/۰۴۹)	۱/۸۹۶(۰/۲۶۳)	۰/۶۷۷(۰/۰۲۲)
۱۰۰۰	۰/۵۸۸(۰/۰۰۳)	۰/۷۸۶(۰/۰۰۴)	۰/۳۲۹(۰/۰۳۷)	۱/۵۷۹(۰/۱۳۶)	۰/۶۸۷(۰/۰۱۷)
	$(\alpha,\theta,\lambda,\gamma,\beta) = (۰/۲, ۰/۵, ۰/۷, ۲/۵, ۰/۳)$				
۱۰۰	۰/۱۶۳(۰/۰۱۴۷)	۰/۳۶۳(۰/۲۰۱)	۰/۷۴۸(۰/۰۲۲)	۲/۸۴۵(۰/۸۱۶)	۰/۳۳۲(۰/۰۲۸)
۲۰۰	۰/۱۸۶(۰/۰۰۷)	۰/۳۹۷(۰/۱۲۱)	۰/۷۳۷(۰/۰۱۹)	۲/۷۱۹(۰/۷۷۳)	۰/۳۲۵(۰/۰۱۶)
۵۰۰	۰/۱۸۹(۰/۰۰۶)	۰/۴۳۷(۰/۰۹۲)	۰/۷۲۷(۰/۰۱۱)	۲/۵۷۲(۰/۵۴۱)	۰/۳۲۲(۰/۰۰۹)
۱۰۰۰	۰/۱۹۲(۰/۰۰۴)	۰/۵۳۵(۰/۰۵۴)	۰/۷۱۶(۰/۰۰۹)	۲/۵۲۴(۰/۴۴۳)	۰/۳۱۷(۰/۰۰۷)

۵ کاربرد DEW-INAR(1) در برازش داده‌های COVID-19

در این بخش، به بررسی کارایی فرایند DEW-INAR(1) در برازش داده‌های تعداد فوتی ناشی از COVID-19 ارائه شده در سایت سازمان جهانی بهداشت (https://covid19.who.int)، پرداخته می‌شود. سری اول مشاهدات مربوط به تعداد فوت ناشی از COVID-19 در کشور قبرس در بازه زمانی ۱۲ فوریه تا ۲ می سال ۲۰۲۱،

جدول ۴. برآورد YW پارامترهای مدل DEW-INAR(1) و مقادیر MSE در پرانتز.

n	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\beta}$
$(\alpha,\theta,\lambda,\gamma,\beta) = (۰/۶, ۰/۸, ۰/۳, ۱/۵, ۰/۷)$					
۱۰۰	۰/۵۲۹(۰/۰۲۸)	۰/۸۹۶(۰/۱۱۴)	۰/۳۹۳(۰/۰۶۵)	۱/۹۷۵(۰/۴۱۲)	۰/۵۵۲(۰/۰۲۷)
۲۰۰	۰/۵۵۹(۰/۰۱۱)	۰/۸۸۲(۰/۰۸۲)	۰/۳۸۷(۰/۰۶۷)	۱/۷۸۸(۰/۲۷۵)	۰/۵۷۱(۰/۰۲۵)
۵۰۰	۰/۵۸۴(۰/۰۰۸)	۰/۸۱۳(۰/۰۰۹)	۰/۳۵۱(۰/۰۵۳)	۱/۵۹۱(۰/۲۰۱)	۰/۶۴۲(۰/۰۱۷)
۱۰۰۰	۰/۵۹۳(۰/۰۰۷)	۰/۸۰۵(۰/۰۰۸)	۰/۳۳۶(۰/۰۴۹)	۱/۵۳۵(۰/۱۶۱)	۰/۷۳۳(۰/۰۰۹)
$(\alpha,\theta,\lambda,\gamma,\beta) = (۰/۲, ۰/۵, ۰/۷, ۲/۵, ۰/۳)$					
۱۰۰	۰/۲۵۶(۰/۰۱۹)	۰/۴۱۹(۰/۰۹۷)	۰/۶۶۹(۰/۰۲۱)	۲/۱۸۲(۰/۸۹۱)	۰/۴۱۵(۰/۰۲۴)
۲۰۰	۰/۱۷۹(۰/۰۰۸)	۰/۵۷۱(۰/۰۸۴)	۰/۷۳۵(۰/۰۱۸)	۲/۳۳۲(۰/۷۹۱)	۰/۳۶۲(۰/۰۱۷)
۵۰۰	۰/۱۹۳(۰/۰۰۶)	۰/۵۴۱(۰/۰۶۵)	۰/۷۲۲(۰/۰۱۲)	۲/۵۵۸(۰/۵۵۱)	۰/۲۷۴(۰/۰۱۳۷)
۱۰۰۰	۰/۲۰۸(۰/۰۰۴)	۰/۵۲۹(۰/۰۴۷)	۰/۷۱۷(۰/۰۰۹)	۲/۵۲۱(۰/۵۰۴)	۰/۳۲۸(۰/۰۰۹)

شامل $n = ۱۱۱$ مشاهده و سری دوم مربوط به تعداد فوت در جمهوری چک در بازه زمانی ۳ جولای الی ۱ اکتبر و شامل $n = ۹۱$ مشاهده است. نمودارهای مسیر نمونه‌ای، تابع خودهمبستگی (ACF) و تابع خودهمبستگی جزئی (PACF) داده‌های COVID-19 در شکل ۳ نشان داده شده است. نمودار ACF به طور تدریجی کاهش می‌یابد که نشان‌دهنده ایستایی مشاهدات است. در نمودار PACF هر دو سری داده نیز، تنها تاخیر اول از باند خارج شده و تاخیرهای دیگر در داخل باند قرار دارند. بنابراین، کاهش ناگهانی PACF پس از تاخیر اول، بر معنی‌داری مقدار خودهمبستگی در تاخیر اول دلالت دارد و لذا فرایند خودبازگشتی مرتبه اول توصیه می‌شود.

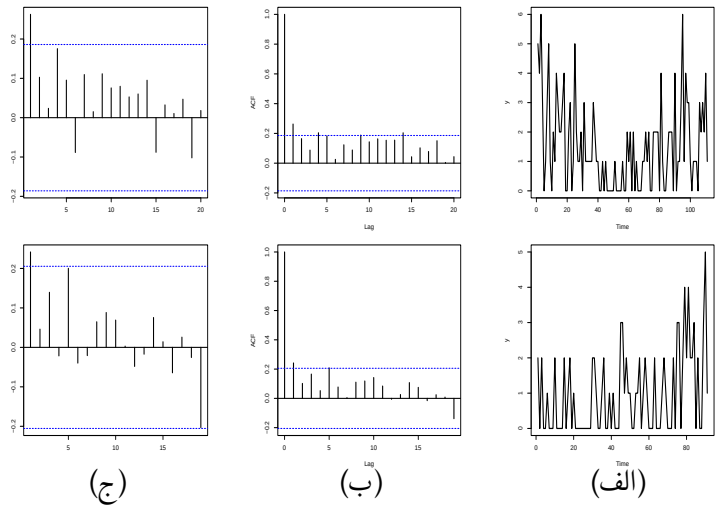
ویژگی‌های آماری داده‌های COVID-19 در جدول ۵ گزارش شده است. همان‌طور که ملاحظه می‌شود دو سری داده بیش‌پراکنده است و ایستایی فرایند با استفاده از آزمون دیکی-فولر افزوده (ADF)[1] و با توجه به p-مقدار آزمون، تایید می‌گردد. فرایند DEW-INAR(1) انواع مختلف پراکندگی را در برمی‌گیرد و لذا انتخاب مناسبی برای داده‌های بیش‌پراکنده است. معیار اطلاع آکائیک (AIC) روش ریاضی برای ارزیابی میزان تناسب یک مدل

جدول ۵. ویژگی‌های آماری و p-مقدار آزمون ADF برای داده‌های COVID-19.

مجموعه داده‌ها	میانگین	واریانس	خودهمبستگی	FDI	p-value
فوتی در قبرس	۱/۴۸۶	۲/۱۶۱	۰/۲۶۳	۰/۰۳۳	۱/۴۵۳
فوتی در چک	۰/۹۴۵	۱/۳۶۳	۰/۲۴۲	۰/۰۴۶	۱/۴۴۲

با داده‌های واقعی است و به منظور مقایسه مدل‌های مختلف و تعیین این‌که کدام یک برای داده‌ها مناسب‌تر است،

[1] Augmented Dickey-Fuller



شکل ۳. الف- مسیر نمونه‌ای، ب- ACF، ج- PACF داده‌های COVID-19 در قبرس (سطر اول) و چک (سطر دوم)

استفاده می‌شود و به صورت $\mathrm{AIC} = -۲\log(L(\hat{\eta})) + ۲k$ محاسبه می‌گردد، که در آن η بردار پارامترهای مدل، $L(\hat{\eta})$ تابع درستنمایی و k تعداد پارامترهای نامعلوم مدل است. هم‌چنین، دیگر آماره‌های نیکویی برازش (GOF)، شامل معیار اطلاع بیزی (BIC)، معیار اطلاع هنان-کوئین (HQIC) و معیار اطلاع آکائیک سازگار (CAIC) نیز به کار برده می‌شود.

در ادامه، فرایند DEW-INAR(1) با برخی فرایندهای INAR(1) کلاسیک، براساس مقادیر آماره‌های نیکویی برازش مقایسه می‌شود. مدل‌های مدنظر شامل PINAR(1) (الاوش و الزید ، ۱۹۸۷)، GINAR(1) (الزید و الاوش ، ۱۹۸۸)، NBIINAR(1) (الاوش و علی ، ۱۹۹۲)، NBRCINAR(1) (ویب ، ۲۰۰۸)، NGINAR(1) (ریستیک و همکاران ، ۲۰۰۹)، NDCINAR(1) (میلتیک ایلیک ، ۲۰۱۶)، NGINAR(1)-ρ (بورگس و همکاران ، ۲۰۱۷)، DCGINAR(1) (ریستیک و همکاران ، ۲۰۱۳)، GADCINAR(1) (نستیک و همکاران ، ۲۰۱۷)، GNBINAR(1) (شما و همکاران ، ۲۰۲۰)، INAR(1)DLi-3P (الیوا و همکاران ، ۲۰۲۰)، INAR(1)DBL (آلتون و همکاران ، ۲۰۲۲) و INAR(1)DPsL (ایرشاد و همکاران ، ۲۰۲۱) است.

اعتبارسنجی متقابل[1] Ir(CV) و تعمیم‌های مختلف آن، یکی از پرکاربردترین روش‌ها در انتخاب برآورد بهینه پارامترها در مدل‌های رگرسیونی است (امینی و روزبه (۲۰۱۵) و روزبه (۲۰۱۸) را ببینید). مارکوس و همکاران (۲۰۲۲) از روش اعتبارسنجی متقابل پیش‌بین برای ارزیابی عملکرد پیش‌بینی در مدل NAR(1) استفاده نمودند. بدین منظور، مشاهدات سری زمانی $X_۱, \ldots, X_n$ برای $n^* < n$، به دو دسته آموزش $(X_۱, \ldots, X_{n^*})$ و آزمون

[1] Cross validation

$(X_{n^*+1}, \ldots, X_n)$ تقسیم می‌شوند. برآورد پارامترهای مدل بر اساس نمونه آموزش محاسبه می‌گردد. سپس مشاهدات گروه آزمون، بر اساس پیش‌بینی یک-گام خارج از نمونه [1] (O-MSE) با استفاده از برآوردهای به دست آمده از بخش مشاهدات آموزش، پیش‌بینی می‌شوند. کارایی فرایندهای سری زمانی در پیش‌بینی مشاهدات با استفاده از معیار میانگین قدرمطلق انحرافات $|\hat{X}_t - X_t|$ به دست می‌آید.

به منظور بررسی عملکرد فرایندهای مختلف در پیش‌بینی داده‌های COVID-19، دو معیار میانگین توان‌های دوم خطا (تفاضل مشاهدات واقعی و مقادیر پیش‌بینی متناظر) (MSE) درون نمونه [2] (I-MSE) و MSE خارج از نمونه به کار برده میشود. در محاسبه I-MSE از تمام مشاهدات برای برآورد پارامترها و پیش‌بینی فرایند استفاده می‌شود. در حالی‌که برای محاسبه O-MSE، مشاهدات به دو بخش تقسیم می‌شود. بخش اول برای برآورد پارامترهای فرایند و بخش دوم برای مقایسه پیش‌بینی استفاده می‌گردد. در هر دو داده COVID-19، $n - ۱۰$ مشاهده اول را برای برآورد پارامترهای مدل و ۱۰ مشاهده دیگر برای پیش‌بینی در نظر گرفته می‌شود. در ادامه، مدل‌های مختلف بر اساس مقادیر I-MSE و O-MSE مقایسه می‌شوند. هم‌چنین، روش CV برای مدل DEW-INAR(1) و سایر مدل‌های رقیب ارائه می‌گردد. با استفاده از روش CV پیش‌بینی، ارائه شده توسط مارکوس و همکاران (۲۰۲۲)، پیش‌بینی‌های یک-گام خارج از نمونه به دست می‌آید.

نتایج برآورد ML پارامتر فرایندهای INAR(1) و مقادیر GOF متناظر آن‌ها برای داده‌های فوت ناشی از COVID-19 در جداول ۶ و ۷، دلالت بر برتری فرایند DEW-INAR(1) در برازش داده‌های COVID-19 دارد، زیرا برای هر دو سری داده‌های COVID-19، فرایند DEW-INAR(1) کمترین مقادیر GOF را شامل می‌شود. با توجه به مقادیر I-MSE و O-MSE، می‌توان ادعا نمود که مدل DEW-INAR(1) بهتر از سایر مدل‌ها قادر به پیش‌بینی داده‌های COVID-19 است، زیرا شامل مقادیر I-MSE و O-MSE کوچکتری نسبت به سایر مدل‌ها است. در جدول ۸، میانگین قدرمطلق انحرافات مشاهدات واقعی و مقادیر پیش‌بینی شده خارج از نمونه برای ۱۰ مشاهده گروه آزمون، گزارش شده است. بر اساس نتایج به دست آمده، مدل DEW-INAR(1) بهترین عملکرد را در پیش‌بینی مشاهدات داده‌های COVID-19 ارائه می‌نماید.

۵.۱ تحلیل مانده‌های فرایند DEW-INAR(1)

با جایگذاری مقادیر برآورد پارامترها در رابطه امید ریاضی شرطی و واریانس شرطی معرفی شده، آن‌گاه مانده‌های پیرسن به صورت $e_t = \frac{X_t - \mathrm{E}(X_t|X_{t-1})}{\sqrt{\mathrm{Var}(X_t|X_{t-1})}}$ تعریف می‌شود. نمودار ACF نمونه‌ای مانده‌های پیرسن در شکل ۴ نشان داده شده که بیانگر عدم همبستگی مانده‌ها است.

هم‌چنین با استفاده از آزمون جانگ-باگس و p-مقدارهای به دست آمده به صورت (۰٫۰۳۷, ۰٫۸۸۹)، عدم همبستگی مانده‌ها نیز تائید می‌گردد. نمودار دوره‌نگار تجمعی مانده‌ها نیز در شکل ۵ ارائه شده که نشان‌دهنده توزیع تصادفی مانده‌ها و عدم مشاهده هرگونه روندی در آن‌ها است.

[1] Out of Sample
[2] In of Sample

جدول ۶. برآورد ML فرایندهای INAR(1) و معیارهای GOF متناظر آن‌ها برای داده‌های فوتی‌های ناشی از COVID-19 در قبرس.

O-MSE	I-MSE	CAIC	HQIC	BIC	AIC	CML	فرایند
۱٫۵۱۶	۲٫۰۷۸	۳۶۸٫۲۳	۳۶۸٫۲۱	۳۷۱٫۴۳	۳۶۶٫۰۱	$\hat{\lambda} = ۱٫۱۶۷۲۵,\ \hat{\alpha} = ۰٫۲۱۸۷۴$	PINAR(1)
۲٫۲۲۱	۳٫۶۷۶	۳۷۲٫۸۳	۳۷۲٫۸۱	۳۷۶٫۰۳	۳۷۰٫۶۱	$\hat{p} = ۰٫۵۷۸۶۶,\ \hat{\alpha} = ۰٫۲۱۱۸۷$	GINAR(1)
۲٫۱۷۹	۲٫۳۸۳	۳۷۶٫۱۹	۳۷۶٫۱۷	۳۷۹٫۳۹	۳۷۳٫۹۷	$\hat{p} = ۱۲٫۲۰۶۵۱,\ \hat{\alpha} = ۱٫۴۸۹۲۷$	NGINAR(1)
۱٫۶۳۷	۱٫۸۹۸	۳۶۲٫۹۵	۳۶۳٫۸۷	۳۶۸٫۷۱	۳۶۰٫۵۷	$\hat{n} = ۳٫۱۸۵۶۲,\ \hat{p} = ۰٫۶۸۰۱۹,\ \hat{\rho} = ۰٫۲۶۲۳۵$	NBRCINAR(1)
۱٫۶۵۳	۱٫۹۰۲	۳۶۲٫۴۲	۳۶۳٫۳۴	۳۶۸٫۱۷	۳۶۰٫۰۴	$\hat{n} = ۲٫۹۵۲۹۴,\ \hat{p} = ۲٫۷۶۲۳۱,\ \hat{\rho} = ۰٫۲۸۶۰۵$	NBIINAR(1)
۱٫۵۰۱	۲٫۹۲۴	۳۸۸٫۱۷	۳۹۱٫۲۵	۳۹۶٫۰۸	۳۸۷٫۹۵	$\hat{\alpha} = ۰٫۲۸۶۲۱,\ \hat{\theta} = ۰٫۹۹۹۸۱,\ \hat{\mu} = ۲٫۸۴۸۲۷$	NDCINAR(1)
۱٫۸۷۸	۲٫۰۷۳	۳۶۰٫۸۲	۳۶۳٫۸۹	۳۶۸٫۷۲	۳۶۰٫۵۹	$\hat{\alpha} = ۰٫۲۴۰۸۱,\ \hat{\rho} = ۱٫۵۰۲۰۷,\ \hat{\mu} = ۱٫۲۰۸۲۴$	NGINAR(1)$\rho-$
۱٫۶۶۷	۱٫۹۰۴	۳۶۹٫۱۷	۳۷۲٫۲۵	۳۷۷٫۰۸	۳۶۸٫۹۵	$\hat{\alpha} = ۰٫۲۹۷۸۲,\ \hat{\theta} = ۰٫۶۴۳۵۴,\ \hat{\mu} = ۱٫۴۴۹۷۵$	DCGINAR(1)
۱٫۶۶۹	۱٫۸۹۶	۳۶۲٫۶۱	۳۶۵٫۶۸	۳۷۰٫۵۱	۳۶۲٫۳۸	$\hat{\alpha} = ۰٫۲۷۲۰۳,\ \hat{\theta} = ۰٫۵۹۹۳۳,\ \hat{\mu} = ۱٫۴۳۸۳۳$	GADCINAR(1)
۱٫۶۷۸	۱٫۹۹۳	۳۶۰٫۷۱	۳۶۳٫۷۸	۳۶۸٫۶۱	۳۶۰٫۴۸	$\hat{\alpha} = ۰٫۴۷۴۰۹,\ \hat{\theta} = ۰٫۷۰۲۸۹,\ \hat{\mu} = ۱٫۴۹۶۲۲$	GNBINAR(1)
۱٫۶۴۸	۱٫۹۰۱	۳۵۶٫۱۶	۳۶۰٫۱۸	۳۶۶٫۶۳	۳۵۵٫۷۹	$\hat{\alpha} = ۰٫۲۹۷۲۶,\ \hat{\beta} = ۰٫۶۲۱۱۵$ $\hat{\lambda} = ۰٫۳۴۹۸۹,\ \hat{p} = ۰٫۲۲۶۵۱$	INAR(1)DLi-3P
۱٫۶۳۹	۱٫۹۰۵	۳۵۷٫۶۳	۳۵۹٫۷۱	۳۶۲٫۹۳	۳۵۷٫۵۱	$\hat{\alpha} = ۰٫۲۰۴۶۹,\ \hat{p} = ۰٫۶۰۲۹۷$	INAR(1)DBL
۱٫۶۴۹	۱٫۹۰۲	۳۵۴٫۰۱	۳۵۷٫۰۸	۳۶۱٫۹۲	۳۵۳٫۷۹	$\hat{\alpha} = ۰٫۲۲۶۵۱,\ \hat{\beta} = ۱٫۵۰۲۵۷,\ \hat{\theta} = ۱٫۰۵۰۱۶$	INAR(1)DPsL
۱٫۲۵۳	۱٫۷۱۱	۳۴۶٫۶۱	۳۵۱٫۴۲	۳۵۹٫۴۸	۳۴۵٫۹۳	$\hat{\alpha} = ۰٫۲۶۷۰۲,\ \hat{\theta} = ۰٫۴۴۶۳۴,\ \hat{\lambda} = ۰٫۱۴۱۳۹$ $\hat{\gamma} = ۱٫۲۷۷۷۱,\ \hat{\beta} = ۰٫۷۴۴۹۸$	DEW-INAR(1)

جدول ۷. برآورد ML فرایندهای INAR(1) و معیارهای GOF متناظر آن‌ها برای داده‌های فوتی‌های ناشی از COVID-19 در چک.

O-MSE	I-MSE	CAIC	HQIC	BIC	AIC	CML	فرایند
۱٫۴۱۹	۱٫۷۸۳	۲۵۳٫۰۲	۲۵۲٫۷۷	۲۵۵٫۷۶	۲۵۰٫۷۴	$\hat{\lambda} = ۰٫۷۸۲۱۵,\ \hat{\alpha} = ۰٫۱۷۴۶۱$	PINAR(1)
۱٫۳۴۲	۱٫۶۹۱	۲۴۹٫۱۶	۲۴۸٫۹۱	۲۵۱٫۹۱	۲۴۶٫۸۹	$\hat{p} = ۰٫۴۸۳۴۲,\ \hat{\alpha} = ۰٫۱۲۲۴۳$	GINAR(1)
۱٫۴۵۱	۱٫۸۴۵	۲۵۵٫۷۱	۲۵۵٫۴۵	۲۵۸٫۴۵	۲۵۳٫۴۳	$\hat{p} = ۸٫۳۴۹۴۸,\ \hat{\alpha} = ۱٫۷۹۲۳۴$	NGINAR(1)
۱٫۲۹۱	۱٫۵۸۳	۲۴۹٫۵۴	۲۵۰٫۱۱	۲۵۴٫۶۱	۲۴۷٫۰۷	$\hat{n} = ۱٫۸۹۰۳۷,\ \hat{p} = ۰٫۶۶۷۴۹,\ \hat{\rho} = ۰٫۱۷۳۲۷$	NBRCINAR(1)
۱٫۳۴۲	۱٫۶۰۶	۲۴۸٫۱۹	۲۴۸٫۷۷	۲۵۳٫۲۶	۲۴۵٫۷۳	$\hat{n} = ۱٫۵۹۵۳۸,\ \hat{p} = ۲٫۱۷۹۵۹,\ \hat{\rho} = ۰٫۲۲۸۴۱$	NBIINAR(1)
۱٫۲۰۹	۱٫۵۴۱	۲۴۷٫۱۳	۲۴۹٫۹۱	۲۵۴٫۳۹	۲۴۶٫۸۶	$\hat{\alpha} = ۰٫۱۹۸۹۲,\ \hat{\theta} = ۰٫۹۹۹۹۱,\ \hat{\mu} = ۱٫۲۵۱۰۳$	NDCINAR(1)
۱٫۴۰۷	۱٫۷۲۱	۲۴۱٫۲۲	۲۴۳٫۹۸	۲۴۸٫۴۷	۲۴۰٫۹۴	$\hat{\alpha} = ۰٫۲۲۵۶۹,\ \hat{\rho} = ۱٫۰۸۹۷۱,\ \hat{\mu} = ۲٫۶۱۱۳۷$	NGINAR(1)$\rho-$
۱٫۴۷۶	۲٫۹۳۱	۲۴۸٫۵۲	۲۵۱٫۲۸	۲۵۵٫۷۷	۲۴۸٫۲۴	$\hat{\alpha} = ۰٫۱۷۷۲۱,\ \hat{\theta} = ۰٫۹۹۹۷۲,\ \hat{\mu} = ۰٫۹۶۱۳۱$	DCGINAR(1)
۱٫۴۶۵	۱٫۷۰۵	۲۶۳٫۷۱	۲۶۶٫۴۸	۲۷۰٫۹۷	۲۶۳٫۴۴	$\hat{\alpha} = ۰٫۵۱۸۴۴,\ \hat{\theta} = ۰٫۹۹۹۸۴,\ \hat{\mu} = ۲٫۱۱۸۴۷$	GADCINAR(1)
۱٫۲۸	۱٫۵۱۲	۲۴۲٫۳۱	۲۴۵٫۰۷	۲۴۹٫۵۷	۲۴۲٫۰۳	$\hat{\alpha} = ۰٫۴۱۰۴۶,\ \hat{\theta} = ۰٫۹۲۱۱۲,\ \hat{\mu} = ۰٫۹۵۶۹۱$	GNBINAR(1)
۱٫۳۲۸	۱٫۶۵۳	۲۴۵٫۹۱	۲۴۹٫۴۹	۲۵۵٫۴۸	۲۴۵٫۴۴	$\hat{\alpha} = ۰٫۷۰۱۴۶,\ \hat{\beta} = ۰٫۴۴۳۳۹$ $\hat{\lambda} = ۰٫۳۳۳۹۱,\ \hat{p} = ۰٫۱۳۷۸۹$	INAR(1)DLi-3P
۱٫۲۹۸	۱٫۶۴۳	۲۴۶٫۹۲	۲۴۸٫۸۱	۲۵۱٫۸۱	۲۴۶٫۷۸	$\hat{\alpha} = ۰٫۱۵۳۸۱,\ \hat{p} = ۰٫۵۲۲۸۸$	INAR(1)DBL
۱٫۳۲۸	۱٫۶۵۵	۲۴۳٫۷۲	۲۴۶٫۴۸	۲۵۰٫۹۷	۲۴۳٫۴۴	$\hat{\alpha} = ۰٫۱۳۷۸۹,\ \hat{\beta} = ۲٫۷۳۵۳۳,\ \hat{\theta} = ۱٫۰۹۶۸۷$	INAR(1)DPsL
۰٫۹۳۶	۱٫۲۱۵	۲۲۷٫۷۱	۲۳۱٫۹۴	۲۳۹٫۴۳	۲۲۶٫۸۸	$\hat{\alpha} = ۰٫۲۱۳۴۱,\ \hat{\theta} = ۰٫۳۶۶۹۸,\ \hat{\lambda} = ۰٫۷۲۱۹۱$ $\hat{\gamma} = ۳٫۴۳۱۵۷,\ \hat{\beta} = ۰٫۵۴۸۷۲$	DEW-INAR(1)

جدول ۸. میانگین قدرمطلق انحرافات پیش‌بینی یک-گام خارج از نمونه برای داده‌های COVID-19.

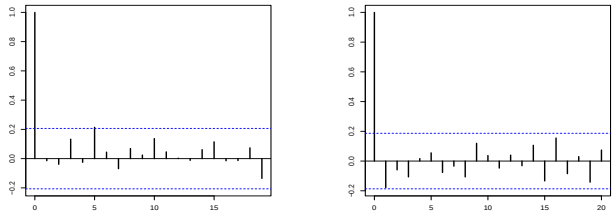
مدل	فوتی‌های قبرس	فوتی‌های چک
PINAR(1)	۱٫۰۹	۰٫۸۸
GINAR(1)	۱٫۴۳	۱٫۰۵
NGINAR(1)	۱٫۵۹	۱٫۱۲
NBRCINAR(1)	۱٫۱۶	۰٫۹۴
NBIINAR(1)	۱٫۱۰	۰٫۹۲
NDCINAR(1)	۰٫۹۷	۰٫۸۶
$\rho-$NGINAR(1)	۱٫۰۹	۰٫۹۳
DCGINAR(1)	۱٫۱۸	۰٫۹۱
GADCINAR(1)	۱٫۱۱	۰٫۹۷
GNBINAR(1)	۱٫۲۴	۱٫۰۶
INAR(1)DLi-3P	۱٫۱۹	۱٫۰۱
INAR(1)DBL	۱٫۱۳	۰٫۹۸
INAR(1)DPsL	۱٫۰۷	۰٫۹۱
DEW-INAR(1)	۰٫۸۳	۰٫۶۹

۵.۲ پیش‌بینی کلاسیک فرایند DEW-INAR(1)

بر اساس (۶)، پیش‌بینی یک گام فرایند، به صورت $\hat{X}_t = \mathrm{E}\big(X_t|X_{t-1}\big) = \alpha X_{t-k} + \mu_Z$ ارائه می‌گردد، که در آن پارامترهای α و μ_Z با مقادیر برآورد ML متناظرشان جایگزین می‌گردند. در شکل ۶، نمودار مسیر نمونه‌ای واقعی (خط مشکی) به همراه مقادیر پیش‌بینی (خط قرمز) تحت فرایند DEW-INAR(1) ارائه شده است. همان‌طور که مشهود است، فرایند DEW-INAR(1) به خوبی مسیر نمونه‌ای هر یک از داده‌های COVID-19 را پیش‌بینی می‌نماید.

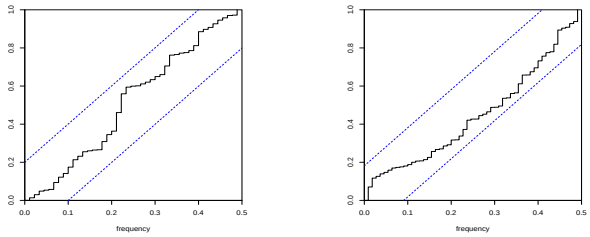
بحث و نتیجه‌گیری

با بررسی ویژگی‌های آماری، انعطاف‌پذیری و جامعیت توزیع نمایی-وایبل گسسته در برازش انواع مختلف داده‌های شمارشی استنتاج می‌شود. هم‌چنین، روش‌های مختلف برآورد پارامترها و مطالعات شبیه‌سازی مونت‌کارلو ارائه شده است. با توجه به نتایج شبیه‌سازی، روش برآورد پارامتری ماکسیمم درستنمایی شرطی، بهترین عملکرد را نسبت به رویکردهای ناپارامتری نشان می‌دهد. با استفاده از داده‌های COVID-19، کارایی فرایند جدید در مقایسه با مدل‌های INAR(1) کلاسیک تایید می‌گردد.



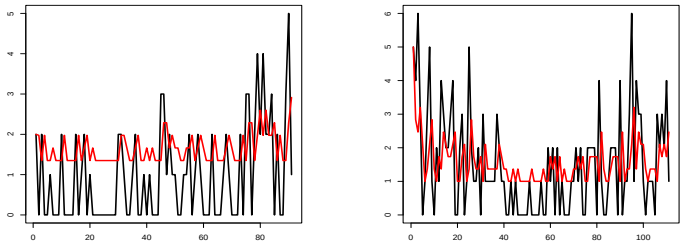
(ب) (الف)

شکل ۴. الف- نمودار ACF نمونه‌ای مانده‌های پیرسن داده‌های COVID-19 در قبرس، ب- نمودار ACF نمونه‌ای مانده‌های پیرسن داده‌های COVID-19 در چک.



(ب) (الف)

شکل ۵. الف- نمودار دوره‌نگار مانده‌های پیرسن داده‌های COVID-19 در قبرس، ب- نمودار دوره‌نگار مانده‌های پیرسن داده‌های COVID-19 در چک.



(ب) (الف)

شکل ۶. الف- مقادیر واقعی (مشکی) و پیش‌بینی (قرمز) داده‌های COVID-19 در قبرس، ب- مقادیر واقعی (مشکی) و پیش‌بینی (قرمز) داده‌های COVID-19 در چک.

تقدیر و تشکر

نویسندگان از سردبیر، داوران و ویراستار محترم مجله که با نظرات ارزشمند خود در بهبود کیفیت مقاله نقش مهمی داشته‌اند، کمال تشکر و سپاسگزاری را دارند.

مراجع

Alizadeh, M., Altun, E., Cordeiro, G.M. and Rasekhi, M. (2017), The Odd Power Cauchy Family of Distributions: Properties, Regression Models and Applications, *Journal of Statistical Computation and Simulation*, **88**(4), 785-807.

Al-Osh, M.A. and Aly, E.E.A.A. (1992), First Order Autoregressive Time Series with Negative Binomial and Geometric Marginals, *Communications in Statistics-Theory and Methods*, **21**, 2483-2492.

Al-Osh, M.A. and Alzaid, A.A. (1987), First-order Integer-valued Autoregressive (INAR(1)) Process, *Journal of Time Series Analysis*, **8**, 261-275.

Altun, E., El-Morshedy, E. and Eliwa, E.S. (2022), A Study on Discrete Bilal Distribution with Properties and Applications on Integer-valued Autoregressive Process, *Revstat-Statistical Journal*, **20**, 501-528.

Alzaid, A.A. and Al-Osh, M.A. (1988), First-order Integer-valued Autoregressive (INAR(1)) Process: Distributional and Regression Properties, *Statistica Neerlandica*, **42**, 53-61.

Amini, M. and Roozbeh, M. (2015), Optimal Partial Ridge Estimation in Restricted Semiparametric Regression Models, *Journal of Multivariate Analysis*, **136**, 26-40.

Borges, P., Bourguignon, M. and Molinares, F.F. (2017), A Generalized NGINAR(1) Process with Inflated Parameter Geometric Counting Series, *Australian & New Zealand Journal of Statistics*, **59**, 137-150.

Eliwa, M.S., Altun, E., El-Dawoody, M. and El-Morshedy, M. (2020), A New Three-parameter Discrete Distribution with Associated INAR(1) Process and Applications, *IEEE Access*, **8**, 91150-91162.

Eliwa, M.S. and El-Morshedy, M. (2021), A One-parameter Discrete Distribution for Over-dispersed Data: Statistical and Reliability Properties with Applications, *Journal of Applied Statistics*, **49**, 2467-2487.

El-Morshedy, M., Eliwa, M.S. and Altun, E. (2020), Discrete Burr-Hatke Distribution with Properties, Estimation Methods and Regression Model, *IEEE Access*, **8**, 74359-74370.

Gómez-Déniz, E. and Calderín-Ojeda, E. (2011), The Discrete Lindley Distribution: Properties and Applications, *Journal of Statistical Computation and Simulation*, **81**(11), 1405-1416.

Huang, J. and Zhu, F. (2021), A New First-order Integer-valued Autoregressive Model with Bell Innovations, *Entropy*, **323**, 713.

Irshad, M.R., Chesneau, C., D'cruz, V. and Maya, R. (2021), Discrete Pseudo Lindley Distribution: Properties, Estimation and Application on INAR(1) Process, *Mathematical and Computational Applications*, **26**(4), 76.

Karlsen, H. and Tjøstheim, D. (1988), Consistent Estimates for the NEAR(2) and NLAR(2) Time Series Models, *Journal of the Royal Statistical Society. Series B (Methodological)*, **50**(2), 313-320.

Krishna, H. and Pundir, P.S. (2009), Discrete Burr and Discrete Pareto Distributions, *Statistical Methodology*, **6**, 177-188.

Liu, Z. and Zhu, F. (2021), A New Extension of Thinning-based Integer-valued Autoregressive Models for Count Data, *Entropy*, **23**(1), 62.

Lívio, T., Mamode Khan, N., Bourguignon, M. and Bakouch, S.H. (2018), An INAR(1) Model with Poisson-Lindley Innovations, *Economics Bulletin*, **38**(3), 1505-1513.

Marques, P.C.F., Graziadei, H. and Lopes, H.F. (2022), Bayesian Generalizations of the Integer-valued Autoregressive Model, *Journal of Applied Statistics*, **49**(2), 336-356.

McKenzie, E. (1986), Autoregressive Moving-average Processes with Negative Binomial and Geometric Distributions, *Advances in Applied Probability*, **18**, 679-705.

Miletić Ilić, A.V. (2016), A Geometric Time Series Model with a New Dependent Bernoulli Counting Series, *Communications in Statistics-Theory and Methods*, **45**, 6400-6415.

Mohammadpour, M. and Shirozhan M. (2020), An INAR(1) Model Based on Negative Binomial Thinning Operator with Serially Dependent Noise, *Journal of Statistical Sciences*, **14**(1), 215-232.

Nakagawa, T. and Osaki, S. (1975), Discrete Weibull Distribution, *IEEE Transaction on Reliability*, **24**, 300-301.

Nastić, A.S., Ristić, M.M. and Miletić Ilić, A.V. (2017), A Geometric Time Series Model with an Alternative Dependent Bernoulli Counting Series,*Communications in Statistics-Theory and Methods*, **46**(2), 770-785.

Ristić, M.M., Bakouch, H.S. and Nastić, A.S. (2009), A New Geometric First-order Integer-valued Autoregressive (NGINAR(1)) Process, *Journal of Statistical Planning and Inference*, **139**, 2218-2226.

Ristić, M.M., Nastić, S.A. and Miletić Ilić, V.A. (2013), A Geometric Time Series Model with Dependent Bernoulli Counting Series, *Journal of Time Series Analysis*, **34**(4), 423-516.

Roozbeh, M. (2018), Optimal QR-based Estimation in Partially Linear Regression Models with Correlated Errors Using GCV Criterion, *Computational Statistics and Data Analysis*, **117**, 45-61.

Shamma, N., Mohammadpour, M. and Shirozhan, M. (2020), A Time Series Model Based on Dependent Zero Inflated Counting Series, *Computational statistics*, **35**, 1737-1757.

Weiß, C.H. (2008), Thinning Operations for Modeling Time Series of Count-a Survey, *AStA Advances in Statistical Analysis*, **92**, 319-341.

Weiß, C.H. (2018), *An Introduction to Discrete-Valued Time Series*, John Wiley & Sons Ltd.