Journal of Statistical Modelling: Theory and Applications
Vol. 6, No. 1, 2025, pp. 29-47
Yazd University Press 2025

JSMTA

Research Paper

Time series model with the autoregressive structure on clinical data

Mohammad Ghanemi, Zahra Khodadadi*, Karim Zare
Department of Statistics, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran

Received: March 16, 2025/ Revised: August 28, 2025/ Accepted: September 30, 2025

Abstract: This research explores a time series model with an autoregressive structure applied to clinical data. The motivation for selecting this topic is to develop a probabilistic model for nonlinear time series with interventions in clinical datasets. By distinguishing between the phases before and after the intervention and analyzing the changes during the intervention, the analysis yields a precisely estimated effect of the intervention. The iterative scheme expectation/conditional maximisationeither algorithm is proposed for parameter estimation, and the observed information matrix is derived analytically. A key focus of data analysis is evaluating the robustness of the model's estimates and understanding how minor local disturbances influence the model. The local impact of the model is thoroughly analyzed across three disturbance scenarios. To assess the performance of the proposed methods, simulated datasets are presented, incorporating expectation/conditional maximisationeither estimates to demonstrate the robustness of estimates in the presence of influential outliers. Finally, the proposed method is successfully applied to model new COVID-19 time series cases in the Czech Republic. Appropriate criteria confirm the applicability of the proposed process, alongside the impact of diagnostic analysis.

Keywords: Autoregressive structure; Clinical data; Discontinuous time series model; Expectation/conditional maximisationeither algorithm.
Mathematics Subject Classification (2010): 62M10.

*Corresponding author: zkhodadadi@iau.ac.ir

1 Introduction

In some real-world studies, an unexpected event causes significant changes in a process, providing an opportunity to examine its nature and dynamics.

Intervention models provide an analyzing framework to specify how a certain strategy affects the trend of a response variable and are typically examined within quasi-experimental trials. The overarching approach in intervention studies focuses on two categories of data, pre-intervention and post-intervention, leading to statistical inferences in this context. To model time series data with interrupted behavior due to an intervening factor, the interrupted time series (ITS) analysis method is employed. The validity of ITS methods depends on assumptions regarding the timing of the intervention and the process's response to the intervention (intervention effect). The ITS analysis has attracted broader acceptance among researchers in clinical trials and epidemiology Pooyannik and Khodadadi (2025).

Statistical methods for evaluating the impact of an intervention are increasingly used in clinical economic, and business studies include, difference-in-differences (DID) models, interventional autoregressive (AR) integrated moving average (ARIMA) models, and segmented regression (SR) of the ITS. These three approaches can be used to evaluate the impact of an intervention when data are collected longitudinally and contain pre- and post-intervention factors. Despite some similarities, each model has unique features that may be used to address different types of research questions. Each model also includes strengths and limitations.

The SR analysis of time series, or regression discontinuity, was first introduced by Quandt (1985) and later by Thistlethwaite and Campbell (1960). Since then, the SR analysis of time series has been used in various ways with different parameters in clinical, economic, and educational research.

In April 2005, the government of Ghana established a cost exemption policy and subsequently considered a free maternal healthcare policy. Emmanuel et al. (2014) evaluated the impact of these policies on assisted deliveries from January 2000 to December 2011, using the intervention analysis provided by Box and Tiao (1975). The estimated intervention model showed that removing financial barriers for assisted deliveries significantly increased the number of pregnant women. The free maternal care program significantly increased the monthly number of pregnant women seeking assisted deliveries.

Flammer (2015) studied the reduction of import tariffs as a source of change in a competitive environment by considering control variables in the DID regression model, comparing affected with unaffected sectors over time.

The ARIMA model with interventions can be an effective method for policymakers to make forecasts and evaluate policy interventions or the impact of any significant event affecting donation rates (see Gerlach (2018)). According to Hudson et al. (2019) and Schaffer et al. (2021), the ARIMA model with interventions is highly flexible and easily accounts for seasonal trends, autocorrelation, underlying trends, and various types of intervention effects.

ITS analysis is generally applied in a wide range of fields. For example, it is used in epidemiology (Gasparrini et al., 2009; James Lopez, 2017), econometrics Shadish et al. (2002), and for evaluating the impact of changes in public health policies Kontopantelis et al. (2015). A review of ITS analysis in drug utilization research was provided by

Jandoc et al. (2015). Bernal et al. (2017) examined and designed the ITS model based on the SR and studied specific characteristics of intervention time series. For example, they considered factors such as over-dispersion of time series data, autocorrelation, seasonal trends, and time-varying confounding factors.

Freni-Sterrantino et al. (2019) used a Bayesian hierarchical model in ITS analysis to investigate the impact of the opening of municipal waste incinerators on infant mortality. They conducted a spatial ITS analysis to examine annual risks of infant mortality and sex ratio (female relative to male) within 10 kilometers of eight incinerators in England and Wales during the period from 1996 to 2016. Morales et al. (2020) used seasonal ITS regression analysis to estimate the impact of regulatory changes by the European Union for diclofenac in 2013 among individuals with cardiovascular disease in Denmark, the Netherlands, England, and Scotland.

In medical analysis and clinical studies, the presence of underlying diseases can lead to significantly biased estimates of the effect of an intervention. Ignoring the impact of these comorbidities may result in incorrect conclusions. A potential solution to this issue is the use of controlled time series. Accordingly, Bottomleyet al. (2019) examined the ITS analysis based on control and intervention series in the presence of confounding factors, ensuring that both series exhibit a common trend. They demonstrated that the intervention effect can be estimated by subtracting the control series from the intervention series and analyzing the difference using linear regression or log-linear regression.

Turner et al. (2021) simulated continuous data to compare the performance of a range of statistical methods under various scenarios using the ITS model. Their analysis included different levels and slopes, varying lengths of time series, and lag-1 autocorrelation. They also assessed the performance of the Durbin-Watson test for detecting autocorrelation in the models. The study employed various statistical estimation methods, including ordinary least squares, generalized least squares, Newey-West standard errors, ARIMA, and restricted maximum likelihood.

Ferreira and Arellano-Valle (2018) proposed a linear regression model with errors from the scale mixture of skew-normal (SN) distributions and estimated the parameters using the EM algorithm. They also addressed diagnostic regression issues such as local influence and generalized influential data.

In response to the global COVID-19 pandemic, various counter-strategies have been implemented by different governments to address the crisis. While most developed countries rely on support for healthcare and social systems, developing nations face greater challenges due to lower macroeconomic indicators. Poppe (2020) conducted a country comparison for Colombia, Costa Rica, Peru, Ecuador, Mexico, and Chile. Using multiple ITS analysis with control units, he examined the impact of mandatory public quarantine in Colombia, Peru, and Ecuador, as well as the effect of mask mandates in public settings in Colombia and Chile. The results indicated that the poverty census rate had no significant impact on the national poverty line or urban population mortality rates due to COVID-19. Bordes et al. (2020) investigated the trends in kidney transplantation from both deceased and living donors using density mapping and linear regression through the ITS analysis. They found that kidney transplantation has been significantly impacted in recent months due to the COVID-19 pandemic. Additionally, Suarez-Pierre et al. (2021) analyzed adult organ transplantation data from

1990 to 2019 using the ARIMA model to predict expected transplant rates and waitlist registrations for 2020. Their results indicated a significant correlation between the pandemic and shortage of kidney transplants, as well as waitlist registrations during the COVID-19 period. The findings suggest that the pandemic led to notable disruptions in kidney transplant activities, highlighting the urgent need for strategies to mitigate these effects and ensure that patients awaiting transplantation receive timely care.

Malladi and Lu. (2023) studied the number of living donor transplants performed in the United States, revealing a significant effect of the pandemic on donation rates, with a decrease of 22.6% from 2019 to 2020. They analyzed data on donor transplants obtained from the United Network for Organ Sharing (UNOS) from January 2002 to August 2021, using the ITS model with March 2020 identified as the intervention event. Additionally, a kidney allocation policy implemented in March 2021 was introduced as a second intervention event for kidney donor transplants. The findings underscore the profound disruptions caused by the pandemic on living organ donation, highlighting the need for ongoing monitoring and adaptation of transplant policies to mitigate these impacts and ensure that patients in need of transplants receive timely care.

Analysis of the ITS is a valuable study design for evaluating the effectiveness of population-level health interventions implemented at a clearly defined point in time. In this research, we aim to introduce a novel modeling approach in the context of the ITS and the SR models, with a specific focus on examining the autocorrelation structure in these models. Among the practical models used to account for autocorrelation effects in modeling, ARIMA models are notable, which we will incorporate into our proposed framework. Another innovative aspect of this study is the exploration of different parameter estimation methods, as well as the estimation of the change point in the time series behavior. Additionally, we will apply our model to real-world data, with a particular emphasis on clinical observations.

2 ITS process based on AR(1) structure with time-varying and autocorrelated SN innovations

In the framework of the ITS analysis, the SR approach is applied to a time series with a linear trend and independently distributed residuals. However, in real-world phenomena, data trends can be highly variable and may be ambiguous or difficult to identify. Due to the challenges associated with modeling autocorrelation structures, SR may not be able to adequately model certain time series. The ITS-ARIMA model serves as an alternative to ITS-SR. Unlike SR, this model regresses recorded outcomes at previous time points and is designed to account for autocorrelation when evaluating the impact of interventions. The ITS-ARIMA model consists of an intervention function and an ARIMA model, which is constructed based on observations prior to the intervention. This approach allows for a more robust analysis of the effects of interventions over time, accommodating the complexities often found in real data sets.
We will examine the asymmetric SN distribution, the ITS-AR process, and transfer functions, and then introduce a new ITS-AR model based on the innovations of the SN distribution and various characteristics of the target process.

2.1 Asymmetric skew-normal distribution

Definition 2.1. *A random variable Z follows the SN distribution with location parameter μ, scale parameter σ^2, and skewness parameter λ, denoted as $Z \sim SN(\mu, \sigma^2, \lambda)$, where the probability density function (PDF) of this distribution is given as follows*

$$f_Z(z) = \frac{2}{\sigma}\phi(\frac{z-\mu}{\sigma})\Phi(\lambda\frac{z-\mu}{\sigma}),$$

such that $\phi(.)$ and Φ represent the PDF and the cumulative distribution function (CDF) of the standard normal distribution, respectively.

The mean and variance of the SN distribution are expressed as follows

$$\begin{aligned} \mathrm{E}(Z) &= \mu + \sigma\sqrt{\frac{2}{\pi}}\frac{\lambda}{\sqrt{1+\lambda^2}}, \\ \mathrm{Var}(Z) &= \sigma^2(1 - \frac{2\lambda^2}{\pi(1+\lambda^2)}). \end{aligned}$$

The stochastic representation of the SN distribution, which will be used for the data generation process, can be based on the convolution of normal and half-normal random variables. Consider the independent random variables $V_0 \sim N(0, \sigma^2)$ and $V_1 \sim N(0, \sigma^2)$; the SN random variable is constructed as follows

$$Z = \mu + \frac{\lambda}{\sqrt{1+\lambda^2}}|V_0| + \frac{1}{\sqrt{1+\lambda^2}}V_1.$$

Considering $|V_0| = W$, the conditional distribution of Z given that $W = \omega$ is represented as

$$Z|W = \omega \sim N\left(\mu + \frac{\lambda}{\sqrt{1+\lambda^2}}\omega, \frac{\sigma^2}{1+\lambda^2}\right), \tag{1}$$

where $W \sim TN(0, \sigma^2)I_0(\omega)$, with TN denoting the truncated normal distribution and $I_A(.)$ representing the indicator function.

Statement 1. If $Z \sim SN(\mu, \sigma^2, \lambda)$, then the conditional distribution is given by

$$W|Z = z \sim TN\left(\frac{\lambda(z-\mu)}{\sqrt{1+\lambda^2}}, \frac{\sigma^2}{1+\lambda^2}\right)I_0(W).$$

2.2 Process ITS-AR

Examining trends (trend lines) in the variables of interest and calculating deviations from what was predicted in the absence of intervention are the main components of the ITS-ARIMA design. We focus on a specific version of the ITS-ARIMA model in which the autoregressive order is set to one and the moving average order is set to zero, where, $p = 1$ and $q = 0$ represent the orders of the AR and moving average processes, respectively. The proposed model is abbreviated as ITS-AR(1).

The first-order AR process with an intervention component, known as ITS-AR(1), is defined as follows

$$Y_t = \alpha Y_{t-1} + \sum_t f(R_t) + \varepsilon_t, \qquad \alpha \in (-1, 1),$$

where $f(R_t) = \beta I_t(\tau)$ is the intervention function at time t, called as step change function.

3 The new asymmetric ITS-AR(1) process with time-varying innovations from the SN distribution

In this section, we introduce a new process that addresses all these issues and can be applied for further statistical testing.

Suppose $\varepsilon_{1,t} \sim SN(\mu_1, \lambda_1, \sigma_1^2)$ and $\varepsilon_{2,t} \sim SN(\mu_2, \lambda_2, \sigma_2^2)$, where $\{\varepsilon_{1,t}\}$ and $\{\varepsilon_{2,t}\}$ are independent of each other. Therefore, we have

$$Y_t = \alpha Y_{t-1} + \beta I_t(\tau) + e_t = \begin{cases} \alpha Y_{t-1} + e_t, & t \leq \tau, \\ \alpha Y_{t-1} + \beta + e_t, & t > \tau, \end{cases} \qquad t \geq 1,$$

where α is the AR coefficient, β is the intervention effect parameter, τ is the time at which the intervention occurred, and autocorrelated innovations are considered as follows

$$e_t = \theta e_{t-1} + \varepsilon_{1,t} I_{1,t} + \varepsilon_{2,t} I_{2,t},$$

in which $I_{1,t} = 1 - I_t(\tau)$ and $I_{2,t} = I_t(\tau)$. The asymmetric dependent ITS-AR(1) process is referred as ITS-AR-SN(1), where the parameter vector is represented by $\Omega = (\alpha, \beta, \theta, \sigma_1, \sigma_2, \mu_1, \mu_2, \lambda_1, \lambda_2)$.

Based on the transformation technique defined as $Y_t - \theta Y_{t-1}$, the ITS-AR-SN(1) process can be formulated as follows

$$Y_t = (\alpha + \theta) Y_{t-1} - \alpha\theta Y_{t-2} + \beta(I_{2,t} - \theta I_{2,t-1}) + \varepsilon_{1,t} I_{1,t} + \varepsilon_{2,t} I_{2,t}. \tag{2}$$

With the transformed equation (2), autocorrelation in e_t and e_{t-1} is eliminated. We constrain the parameters such that $|\alpha|, |\theta| < 1$, to ensure the stationarity of the ITS-AR-SN(1) process.

Here, we will derive the conditional PDF of the ITS-AR-SN(1) process. Based on the fact that $I_{2,t-1} = 1$ ($I_{1,t} = 1$) leads to $I_{2,t} = 1$ ($I_{1,t-1} = 1$), we consider $P(Y_t < j | i_1 \leq Y_{t-1} < i_1 + d_1,\ i_2 \leq Y_{t-2} < i_2 + d_2),\ d_1, d_2 > 0$ where $d_1, d_2 \to 0$.

$$\begin{aligned} P(Y_t < y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}) &= I_{1,t} P(\varepsilon_{1,t} \leq y_t - (\alpha+\theta) y_{t-1} + \alpha\theta y_{t-2}) \\ &\quad + I_{2,t} P(\varepsilon_{2,t} \leq y_t - (\alpha+\theta) y_{t-1} + \alpha\theta y_{t-2} - \beta(1-\theta)) \\ &\quad + I_{2,t} I_{1,t-1} P(\varepsilon_{2,t} \leq y_t - (\alpha+\theta) y_{t-1} + \alpha\theta y_{t-2} - \beta) \\ &= I_{1,t} F_{\varepsilon_1}(y_t - (\alpha+\theta) y_{t-1} + \alpha\theta y_{t-2}) \\ &\quad + I_{2,t-1} F_{\varepsilon_2}(y_t - (\alpha+\theta) y_{t-1} + \alpha\theta y_{t-2} - \beta(1-\theta)) \\ &\quad + I_{2,t} I_{1,t-1} F_{\varepsilon_2}(y_t - (\alpha+\theta) y_{t-1} + \alpha\theta y_{t-2} - \beta), \end{aligned}$$

where $F_{\varepsilon_2}(.),\ i = 1, 2$ is the CDF of the SN distribution. Therefore, the conditional PDF is given as follows

$$\begin{aligned} f(y_t | y_{t-1}, y_{t-2}) &= I_{1,t} f(y_t - (\alpha+\theta) y_{t-1} + \alpha\theta y_{t-2}) \\ &\quad + I_{2,t-1} f(y_t - (\alpha+\theta) y_{t-1} + \alpha\theta y_{t-2} - (1-\theta)\beta) \end{aligned}$$

$$+I_{2,t}I_{1,t-1}f(y_t-(\alpha+\theta)y_{t-1}+\alpha\theta y_{t-2}-\beta), \tag{3}$$

where $f(.)$ is the CDF of SN distribution. The third statement occurs only for $t=\tau$.
Statement 2. The conditional expectations and variance of the ITS-AR-SN(1) process are obtained as follows

$$\mathrm{E}(Y_t|Y_{t-1},Y_{t-2})=(\alpha+\theta)Y_{t-1}-\alpha\theta Y_{t-2}+\beta(1-\theta^*)+\mu_{\varepsilon,1}I_{1,t}+\mu_{\varepsilon,2}I_{2,t}, \tag{4}$$

where

$$\theta^*=\begin{cases}0, & t=\tau,\\ \theta, & t>\tau,\end{cases}, \qquad \mu_{\varepsilon,i}=\mathrm{E}(\varepsilon_i)=\mu_i+\sigma_i\sqrt{\frac{2}{\pi}}\frac{\lambda_1}{\sqrt{1+\lambda_1^2}}, \quad i=1,2,$$

and

$$\mathrm{Var}(Y_t|Y_{t-1},Y_{t-2})=I_{1,t}\sigma^2_{\varepsilon,1}+I_{2,t}\sigma^2_{\varepsilon,2}, \tag{5}$$

in such a way that

$$\sigma^2_{\varepsilon,i}=\mathrm{Var}(\varepsilon_i)=\sigma_i^2\left(1-\frac{2\lambda^2}{\pi(1+\lambda_i^2)}\right), \quad i=1,2.$$

4 Expectation/conditional maximisationeither algorithm ITS-AR-SN (1) process

We implement the expectation/conditional maximisationeither (ECME) algorithm proposed by Meng and Rubin (1993) to estimate the parameters of the suggested model. The hierarchical form of the SN distribution presented in (1) and (3) for $t=1,\dots,n$, is represented as

$$\begin{aligned}
&\varepsilon_{j,t}|W_t=\omega\sim N\Big(\mu_j+\frac{\lambda_j}{\sqrt{1+\lambda_j^2}}\omega,\frac{\sigma_j^2}{1+\lambda_j^2}\Big), \quad j=1,2,\\
&W_{j,t}\sim TN(0,\sigma_j^2)I_0(W_{j,t}),\\
&\sum_{j=1}^{2}\varepsilon_{j,t}I_{j,t}=Y_t-(\alpha+\theta)Y_{t-1}+\alpha\theta Y_{t-2}-\beta(1-\theta^*).
\end{aligned}$$

Suppose $\boldsymbol{Y}=(Y_1,\dots,Y_n)'$, and $\boldsymbol{W}=\Big((W_{1,1},W_{2,1})',\dots,(W_{1,n},W_{2,n})'\Big)'$, also let the complete data set as $\boldsymbol{C}=(\boldsymbol{Y}',\mathbf{W}')'$, where $\boldsymbol{Y}$ is the observed part and $\boldsymbol{W}$ is the missing part (from a TN distribution), as previously defined. Therefore, the likelihood function based on the complete data is as follow

$$\begin{aligned}
L(\Omega\mid C)=\prod_{t=2}^{n}\sum_{j=1}^{2}\phi\Bigg(\Big(y_t-(\alpha+\theta)y_{t-1}+\alpha\theta y_{t-2}-\beta(I_{2,t}-\theta I_{2,t-1})-\mu_j\\
-\frac{\lambda_j}{\sqrt{1+\lambda_j^2}}W_{j,t}\Big)\frac{\sqrt{1+\lambda_j^2}}{\sigma_j}\Bigg)I_{j,t}\frac{\sqrt{1+\lambda_j^2}}{\sigma_j^2}T\phi\big(\frac{W_{j,t}}{\sigma_j}\big)I_0(W_{j,t})
\end{aligned}$$

$$
\begin{aligned}
&= \prod_{t=2}^{\tau-1} \phi\left(\left(y_t - (\alpha+\theta)y_{t-1} + \alpha\theta y_{t-2} - \mu_1 - \frac{\lambda_1}{\sqrt{1+\lambda_1^2}} W_{1,t}\right)\frac{\sqrt{1+\lambda_1^2}}{\sigma_1}\right) \\
&\times \prod_{t=\tau}^{n} \phi\left(\left(y_t - (\alpha+\theta)y_{t-1} + \alpha\theta y_{t-2} - (1-\theta^*) - \mu_2 - \frac{\lambda_2}{\sqrt{1+\lambda_2^2}} W_{2,t}\right)\frac{\sqrt{1+\lambda_2^2}}{\sigma_2}\right) \\
&\times \prod_{t=2}^{\tau-1} \frac{\sqrt{1+\lambda_1^2}}{\sigma_1^2} T\phi\left(\frac{W_{1,t}}{\sigma_1}\right) \prod_{t=\tau}^{n} \frac{\sqrt{1+\lambda_2^2}}{\sigma_2^2} T\phi\left(\frac{W_{2,t}}{\sigma_2}\right) \prod_{t=2}^{n} \sum_{j=1}^{2} I_0(W_{2,t}).
\end{aligned}
$$

By removing constants, the corresponding conditional log-likelihood function is

$$
\begin{aligned}
\ell(\Omega|C) \propto & -\frac{1+\lambda_1^2}{2\sigma_1^2} \sum_{t=2}^{\tau-1} \left(y_t - (\alpha+\theta)y_{t-1} + \alpha\theta y_{t-2} - \mu_1 - \frac{\lambda_1}{\sqrt{1+\lambda_1^2}} W_{1,t}\right)^2 \\
& -\frac{1+\lambda_2^2}{2\sigma_2^2} \sum_{t=\tau}^{n} \left(y_t - (\alpha+\theta)y_{t-1} + \alpha\theta y_{t-2} - (1-\theta^*)\beta - \mu_2 - \frac{\lambda_2}{\sqrt{1+\lambda_2^2}} W_{2,t}\right)^2 \\
& + \frac{\tau-2}{2}(\log(1+\lambda_1^2) - 2\log(\sigma_1^2)) + \frac{n-\tau+1}{2}(\log(1+\lambda_2^2) - 2\log(\sigma_2^2)) \\
& -\frac{1}{2\sigma_1^2} \sum_{t=2}^{\tau-1} W_{1,t}^2 - \frac{1}{2\sigma_2^2} \sum_{t=\tau}^{n} W_{2,t}^2.
\end{aligned}
$$

In the $(k+1)$-th iteration of the ECME algorithm, the E-Step requires the calculation of $Q(\Omega|\Omega^{(k)}) = E_{\Omega^{(k)}}(\ell(\Omega|C)|Y)$. Thus, in the E-step of the algorithm, the conditional expectation of the process is given by

$$
\begin{aligned}
Q(\Omega|\Omega^{(k)}) = & -\frac{1+\lambda_1^2}{2\sigma_1^2} \sum_{t=2}^{\tau-1} (y_t - (\alpha+\theta)y_{t-1} + \alpha\theta y_{t-2} - \mu_1)^2 \\
& + \frac{\lambda_1\sqrt{1+\lambda_1^2}}{\sigma_1^2} \sum_{t=2}^{\tau-1} (y_t - (\alpha+\theta)y_{t-1} + \alpha\theta y_{t-2} - \mu_1) E_{1,1}(t) \\
& -\frac{1+\lambda_2^2}{2\sigma_2^2} \sum_{t=\tau}^{n} (y_t - (\alpha+\theta)y_{t-1} + \alpha\theta y_{t-2} - (1-\theta^*)\beta - \mu_2)^2 \\
& + \frac{\lambda_2\sqrt{1+\lambda_2^2}}{\sigma_2^2} \sum_{t=\tau}^{n} (y_t - (\alpha+\theta)y_{t-1} + \alpha\theta y_{t-2} - (1-\theta^*)\beta - \mu_2) E_{2,1}(t) \\
& -\frac{1+\lambda_1^2}{2\sigma_1^2} \sum_{t=2}^{\tau-1} E_{1,2}(t) - \frac{1+\lambda_2^2}{2\sigma_2^2} \sum_{t=\tau}^{n} E_{2,2}(t) + \frac{\tau-2}{2}(\log(1+\lambda_1^2) - 2\log(\sigma_1^2)) \\
& + \frac{n-\tau+1}{2}(\log(1+\lambda_2^2) - 2\log(\sigma_2^2)), \qquad (6)
\end{aligned}
$$

where $E_{j,1}(t) = E(W_{j,t}|Y)$ and $E_{j,2}(t) = E(W_{j,t}^2|Y)$, $j = 1, 2$. These represent the first and second moments of the conditional distribution of the hidden variables $W_{j,t}$ which follow a TN distribution in the interval $(0, \infty)$. The mean and variance of $W_{j,t}$ are represented respectively as

$$
\mu_{W,j} \quad = \quad E(W_{j,t})
$$

$$
\begin{aligned}
&= \frac{\lambda_j}{\sqrt{1+\lambda_j^2}}(y_t-(\alpha+\theta)y_{t-1}+\alpha\theta y_{t-2} \\
&\quad -\beta(I_{2,t}-\theta I_{2,t-1})-\mu_j), \qquad j=1,2 \\
\sigma^2_{W,j} &= \mathrm{Var}(W_{j,t})=\frac{\sigma_j^2}{1+\lambda_j^2}.
\end{aligned}
$$

The expressions $E^{(k)}_{j,1}(t)$ and $E^{(k)}_{j,2}(t)$ are the updated expected values in each iteration of the ECME algorithm with respect to $\mathbf{\Omega}^{(k)}$. Based on the latest estimates $(\hat\alpha^{(k)},\hat\beta^{(k)},\hat\theta^{(k)},\hat\mu_1^{(k)},\hat\mu_2^{(k)},\hat\sigma_1^{2(k)},\hat\sigma_2^{2(k)},\hat\lambda_1^{(k)},\hat\lambda_2^{(k)})$ at iteration k, we optimize the aforementioned Q function in the CM steps of the algorithm as follows

$$
\begin{aligned}
\hat\alpha^{(k+1)} = &\left(\sum_{t=2}^{n}\sum_{j=1}^{2}\frac{1+\hat\lambda_j^{2(k)}}{\hat\sigma_j^{2(k)}}(y_{t-1}-\hat\theta^{(k)}y_{t-2})^2\right)^{-1}\Bigg[\sum_{t=2}^{2}\sum_{j=1}^{2}\frac{1+\hat\lambda_j^{2(k)}}{\hat\sigma_j^{2(k)}}(y_{t-1}-\hat\theta^{(k)}y_{t-2}) \\
&\times\Big(y_t-\hat\theta^{(k)}y_{t-1}-(1-\hat\theta^{*(k)})\hat\beta_j^{(k)}-\hat\mu_j^{(k)}-\frac{\hat\lambda_j^{(k)}}{\sqrt{1+\hat\lambda_j^{2(k)}}}E^{(k)}_{j,1}(t)\Big)I_{j,t}\Bigg].
\end{aligned}
$$

Based on the estimates $(\hat\alpha^{(k+1)},\hat\beta^{(k)},\hat\theta^{(k)},\hat\mu_1^{(k)},\hat\mu_2^{(k)},\hat\sigma_1^{2(k)},\hat\sigma_2^{2(k)},\hat\lambda_1^{(k)},\hat\lambda_2^{(k)})$, the estimator for the parameter β is represented as follows

$$
\hat\beta^{(k+1)}=\frac{\sum_{t=\tau}^{n}\Big(y_t-(\hat\alpha^{(k+1)}+\hat\theta^{(k)})y_{t-1}+\hat\alpha^{(k+1)}\hat\theta^{(k)}y_{t-2}-\hat\mu_2^{(k)}-\frac{\hat\lambda_2^{(k)}}{\sqrt{1+\hat\lambda_2^{2(k)}}}E^{(k)}_{2,1}(t)\Big)}{(n-\tau+1)(1-\hat\theta^{*(k)})}.
$$

Based on the estimates $(\hat\alpha^{(k+1)},\hat\beta^{(k+1)},\hat\theta^{(k)},\hat\mu_1^{(k)},\hat\mu_2^{(k)},\hat\sigma_1^{2(k)},\hat\sigma_2^{2(k)},\hat\lambda_1^{(k)},\hat\lambda_2^{(k)})$, the estimator for the parameter θ is represented as follows

$$
\begin{aligned}
\hat\theta^{(k+1)} = &\Bigg(\frac{1+\hat\lambda_1^{2(k)}}{\hat\sigma_1^{2(k)}}\sum_{t=2}^{\tau-1}(y_{t-1}-\hat\alpha^{(k+1)}y_{t-2})^2+\frac{1+\hat\lambda_2^{2(k)}}{\hat\sigma_2^{2(k)}} \\
&\times\left[(y_{\tau-1}-\hat\alpha^{(k+1)}y_{\tau-2})^2+\sum_{t=\tau+1}^{n}(y_{\tau-1}-\hat\alpha^{(k+1)}y_{\tau-2}-\beta)^2\right]\Bigg)^{-1}\frac{1+\hat\lambda_1^{2(k)}}{\hat\sigma_1^{2(k)}} \\
&\times\sum_{t=2}^{\tau-1}(y_{t-1}-\hat\alpha^{(k+1)}y_{t-2})\left(y_t-\hat\alpha^{(k+1)}y_{t-1}-\hat\mu_1^{(k)}-\frac{\hat\lambda_1^{(k)}}{\sqrt{1+\hat\lambda_1^{2(k)}}}E^{(k)}_{1,1}(t)\right) \\
&+\frac{1+\hat\lambda_2^{2(k)}}{\hat\sigma_2^{2(k)}}(y_{t-1}-\hat\alpha^{(k+1)}y_{t-2}) \\
&\times\left(y_t-\hat\alpha^{(k+1)}y_{t-1}-\beta-\hat\mu_2^{(k)}-\frac{\hat\lambda_2^{(k)}}{\sqrt{1+\hat\lambda_2^{2(k)}}}E^{(k)}_{2,1}(t)\right) \\
&+\frac{1+\hat\lambda_2^{2(k)}}{\hat\sigma_2^{2(k)}}\sum_{t=\tau+1}^{n}(y_{t-1}-\hat\alpha^{(k+1)}y_{t-2}-\beta)
\end{aligned}
$$

$$\times \left(y_t - \hat{\alpha}^{(k+1)} y_{t-1} - \beta - \hat{\mu}_2^{(k)} - \frac{\hat{\lambda}_2^{(k)}}{\sqrt{1+\hat{\lambda}_2^{2(k)}}} E_{2,1}^{(k)}(t) \right).$$

Based on the estimates $(\hat{\alpha}^{(k+1)}, \hat{\beta}^{(k+1)}, \hat{\theta}^{(k+1)}, \hat{\mu}_1^{(k)}, \hat{\mu}_2^{(k)}, \hat{\sigma}_1^{2(k)}, \hat{\sigma}_2^{2(k)}, \hat{\lambda}_1^{(k)}, \hat{\lambda}_2^{(k)})$, the estimators for the parameters μ_1 and μ_2 are represented as follows

$$\begin{aligned}
\hat{\mu}_1^{(k+1)} = & \frac{1}{\tau - 2} \\
& \times \sum_{t=2}^{\tau-1} \left(y_t - (\hat{\alpha}^{(k+1)} + \hat{\theta}^{(k+1)}) y_{t-1} + \hat{\alpha}^{(k+1)} \hat{\theta}^{(k+1)} y_{t-2} - \frac{\hat{\lambda}_1^{(k)}}{\sqrt{1+\hat{\lambda}_1^{2(k)}}} E_{1,1}^{(k)}(t) \right), \\
\hat{\mu}_2^{(k+1)} = & \frac{1}{n-\tau+1} \sum_{t=\tau}^{n} \Big(y_t - (\hat{\alpha}^{(k+1)} + \hat{\theta}^{(k+1)}) y_{t-1} + \hat{\alpha}^{(k+1)} \hat{\theta}^{(k+1)} y_{t-2} \\
& - (1 - \hat{\theta^*}^{(k+1)}) \hat{\beta}^{(k+1)} - \frac{\hat{\lambda}_2^{(k)}}{\sqrt{1+\hat{\lambda}_2^{2(k)}}} E_{2,1}^{(k)}(t) \Big).
\end{aligned}$$

Based on the estimates $(\hat{\alpha}^{(k+1)}, \hat{\beta}^{(k+1)}, \hat{\theta}^{(k+1)}, \hat{\mu}_1^{(k+1)}, \hat{\mu}_2^{(k+1)}, \hat{\sigma}_1^{2(k)}, \hat{\sigma}_2^{2(k)}, \hat{\lambda}_1^{(k)}, \hat{\lambda}_2^{(k)})$, the estimators for the parameters σ_1 and σ_2 are represented as follows

$$\begin{aligned}
\hat{\sigma}_1^{2(k+1)} &= \frac{1+\hat{\lambda}_1^{2(k)}}{2(\tau-2)} \Big[(y_t - (\hat{\alpha}^{(k+1)} + \hat{\theta}^{(k+1)}) y_{t-1} + \hat{\alpha}^{(k+1)} \hat{\theta}^{(k+1)} y_{t-2} - \hat{\mu}_1^{(k+1)})^2 \\
&\quad - E_{1,2}^{(k)}(t) \Big] - \frac{\hat{\lambda}_1^{(k)} \sqrt{1+\hat{\lambda}_1^{2(k)}}}{\tau-2} \\
&\quad \times \sum_{t=2}^{\tau-1} \Big(y_t - (\hat{\alpha}^{(k+1)} + \hat{\theta}^{(k+1)}) y_{t-1} + \hat{\alpha}^{(k+1)} \hat{\theta}^{(k+1)} y_{t-2} - \hat{\mu}_1^{(k+1)} \Big) E_{1,1}^{(k)}(t), \\
\hat{\sigma}_2^{2(k+1)} &= \frac{1+\hat{\lambda}_2^{2(k)}}{2(n-\tau+1)} \sum_{t=\tau}^{n} \Big(y_t - (\hat{\alpha}^{(k+1)} + \hat{\theta}^{(k+1)}) y_{t-1} + \hat{\alpha}^{(k+1)} \hat{\theta}^{(k+1)} y_{t-2} \\
&\quad - (1 - \hat{\theta^*}^{(k+1)}) \hat{\beta}^{(k+1)} - \hat{\mu}_2^{(k+1)} \Big)^2 \frac{1+\hat{\lambda}_2^{2(k)}}{2(n-\tau+1)} \sum_{t=\tau}^{n} E_{2,2}^{(k)}(t) \\
&\quad - \frac{\hat{\lambda}_2^{(k)} \sqrt{1+\hat{\lambda}_2^{2(k)}}}{n-\tau+1} \sum_{t=\tau}^{n} \Big(y_t - (\hat{\alpha}^{(k+1)} + \hat{\theta}^{(k+1)}) y_{t-1} + \hat{\alpha}^{(k+1)} \hat{\theta}^{(k+1)} y_{t-2} \\
&\quad - (1 - \hat{\theta^*}^{(k+1)}) \hat{\beta}^{(k+1)} - \hat{\mu}_2^{(k+1)} \Big) E_{2,1}^{(k)}(t).
\end{aligned}$$

For the ECME algorithm, the CML ECM step is modified as follows:

CML Step: The updates for $\hat{\lambda}_1^{(k+1)}$ and $\hat{\lambda}_2^{(k+1)}$ are obtained by optimizing the log-likelihood functions as follows:

$$\hat{\lambda}_1^{(k+1)} = \arg\max_{\lambda_1} \sum_{t=2}^{\tau-1} \log \left(\frac{2}{\hat{\sigma}_1^{(k)}} \phi\left(\frac{\varepsilon_{1,t} - \hat{\mu}_1^{(k+1)}}{\hat{\sigma}_1^{(k)}} \right) \Phi\left(\lambda_1 \frac{\varepsilon_{1,t} - \hat{\mu}_1^{(k+1)}}{\hat{\sigma}_1^{(k)}} \right) \right),$$

$$\hat{\lambda}_2^{(k+1)} = \arg\max_{\lambda_2} \sum_{t=\tau}^{n} \log\left(\frac{2}{\hat{\sigma}_2^{(k)}}\phi\left(\lambda_2\frac{\varepsilon_{2,t}-\hat{\mu}_2^{(k+1)}}{\hat{\sigma}_2^{(k)}}\right)\Phi\left(\lambda_2\frac{\varepsilon_{2,t}-\hat{\mu}_2^{(k+1)}}{\hat{\sigma}_2^{(k)}}\right)\right),$$

in which $\varepsilon_{1,t} = y_t - (\hat{\alpha}^{(k+1)} + \hat{\theta}^{(k+1)})y_{t-1} + \hat{\alpha}^{(k+1)}\hat{\theta}^{(k+1)}y_{t-2}$, and $\varepsilon_{2,t} = y_t - (\hat{\alpha}^{(k+1)} + \hat{\theta}^{(k+1)})y_{t-1} + \hat{\alpha}^{(k+1)}\hat{\theta}^{(k+1)}y_{t-2} - (1-\hat{\theta^*}^{(k+1)})\hat{\beta}^{(k+1)} - \hat{\mu}_2^{(k+1)}$.

The maximization step in CML requires a one-dimensional search, which can be easily solved using the 'optim' function available in the R software package. As Liu and Rubin (1994) noted, the ECME algorithm exhibits a faster convergence rate compared to the ECM algorithm. The ECME steps are iteratively repeated until a convergence criterion is met.

5 The local impact of the ITS-AR-SN(1) process

According to Cook (1986), the likelihood displacement (LD) is used to assess the impact of the perturbation $\xi \in R^q$, where $\xi = (\xi_1, \ldots, \xi_q)'$, and q represents a suitable number of dimensions. The vectors $\xi_0 = (1, \ldots, 1)$ and $\xi_0 = (0, \ldots, 0)$ correspond to the perturbation and non-perturbation sets of size $q \times 1$, respectively. The LD index is defined as $LD(\xi) = 2(\ell(\hat{\Omega}) - \ell(\hat{\Omega}|\xi))$, such that large values of $LD(\xi)$ indicate that the ML estimate $\hat{\Omega}$ and the estimate $\hat{\Omega}|\xi$ are significantly different, where $\ell(\hat{\Omega})$ and $\ell(\hat{\Omega}|\xi)$ represent the log-likelihood functions of the specified and disturbed models, respectively. According to Garay et al. (2014), for evaluating the impact of disturbances, the log-likelihood function can be replaced with the Q function defined in (6) in the ECME estimation.

It is assumed that the specified model $Q(\Omega|\xi)$ is twice continuously differentiable in the neighborhood of $(\hat{\Omega}, \xi_0)$, and we consider the normal curvature $C_l(\Omega) = 2|l'\ddot{F}l| = 2|l'(\Delta'\ddot{Q}\Delta)l|$ at $\xi_0 = \xi$, where

$$\ddot{F} = \frac{\partial^2 Q(\Omega|\xi)}{\partial\xi\partial\xi'}, \quad \Delta = \frac{\partial^2 Q(\Omega|\xi)}{\partial\Omega\partial\xi'}, \quad \ddot{Q} = \frac{\partial^2 Q(\Omega|\xi)}{\partial\Omega\partial\Omega'},$$

such that l is a $q \times 1$ unit vector, and $\ddot{Q}$ is a 9×9 observed information matrix from the specified model. The matrix Δ is a $9 \times q$ disturbance matrix from the disturbed model, evaluated at $\xi_0 = \xi$ and $\Omega = \hat{\Omega}$. We check the local behavior of $LD(\xi)$ and $C_l(\Omega)$ in the direction of a unit vector l, where $\| l \| = 1$.

The local influence analysis is conducted through the maximization of $C_{\max} = \max_{\|l\|=1} C_l$, where the largest absolute eigenvalue $\lambda_{\max}$ and the corresponding eigenvector $l_{\max}$ are derived from the matrix $\ddot{F} = \Delta'\ddot{Q}\Delta$. The t-th observation may have the greatest impact if the absolute value of its corresponding element in $l_{\max}$ is the largest. This method examines the overall local influence $C_t = C_{l_t}(\Omega)$, where l_t is a unit vector of size $q \times 1$ that is zero everywhere except at position t, where it equals one. We consider the stable normal curvature $B_l(\Omega) = \frac{C_l(\Omega)}{2\mathrm{tr}(\Delta'\ddot{Q}^{-1}\Delta)}$, where $0 \le B_l(\Omega) \le 1$, making it suitable for various comparisons. If $M(0)_t = B_{l_t}$ exceeds the criterion $\frac{1}{q} + c^* SM(0)$, where $SM(0)$ is the sample standard error of $M(0)_k$ for $k = 1, \ldots, q$ and c^* is a specific constant, then the t-th observation is considered significantly influential. Having a reference value for $C_{\max}$ and for the elements of $l_{\max}$ aids in evaluating the extent of influence, as discussed by Liu et al. (2015), and Liu et al. (2016).

6 Real data analysis

We are examining the weekly new COVID-19 cases reported in the Czech Republic from March 8, 2020, to December 17, 2023, totaling $n = 198$ cases as reported by the World Health Organization. The first confirmed COVID-19 infections in the Czech Republic were reported on March 1, 2020. Since then, the number of new cases has gradually increased, peaking in February 2022 with over 286,000 confirmed cases. Between 2020 and 2023, the Czech Republic reported approximately 4.73 million confirmed cases.

The vaccination policy in the Czech Republic was implemented on December 27, 2020, with the first booster doses administered on May 26, 2021. The sample trajectory of COVID-19 cases is presented in Figure 1, illustrating two distinct patterns before and after $n = 101$, which corresponds to February 6, 2022. As shown in Figure 1, new cases exhibited a significant downward trend approximately two months after the commencement of vaccinations in February 2022. By February 2022, a total of 17.3 million vaccine doses had been administered, with 6.94 million individuals receiving complete vaccination and 3.95 million receiving booster doses. It can be concluded that the administration of the COVID-19 vaccine booster dose has led to a reduction in new cases and contributed to controlling the COVID-19 pandemic.

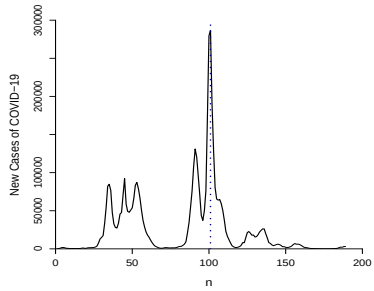

Figure 1: Sample path of COVID-19 dataset cases.

The partial autocorrelation function (PACF) of the COVID-19 data series is illustrated in Figure 2, suggesting a first-order AR model for the data. The fluctuations in the actual dataset are quite substantial. Therefore, we examine the logarithmic transformation of new COVID-19 cases, and some statistical measures of the transformed dataset are summarized in Table 1. Consequently, the COVID-19 data exhibit left skewness (asymmetry) and are platykurtic or flat-tailed (with thin tails). The trajectory of the transformed sample data is displayed in Figure 2.

Table 1: Statistical measurements of the modified COVID-19 dataset.

Kurtosis	Skewness	Standard Deviation	Mean
2.33	-0.24	4.13	8.61

Figure 3 displays the sample path plot of the transformed data set.

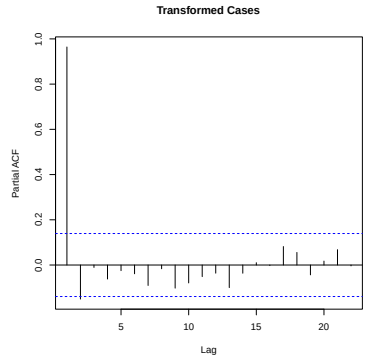

Figure 2: PACF diagram of the transformed dataset.

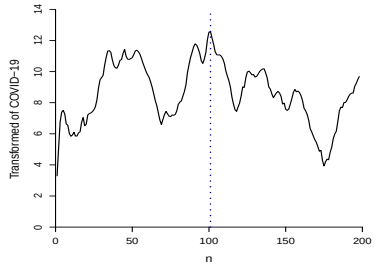

Figure 3: Sample path of the transformed dataset.

6.1 Analysis of residuals

We conducted a residual analysis to assess the adequacy of the model, which indicates the consistency between the data and the fitted model. Considering the standardized Pearson residuals defined as $r(t) = \frac{Y_t - \mathrm{E}(Y_t|Y_{t-1},Y_{t-2})}{\sqrt{\mathrm{Var}(Y_t|Y_{t-1},Y_{t-2})}}$, where $\mathrm{E}(Y_t|Y_{t-1},Y_{t-2})$ and $\mathrm{Var}(Y_t|Y_{t-1},Y_{t-2})$ are presented in (4) and (5).

The analysis of Pearson residuals is presented in Figure 4, indicating an approximately constant variance over time. The histogram reveals that the prediction errors are nearly normally distributed, with a mean of -0.030 (close to 0) and a variance of 0.9784 (close to 1). Based on the autocorrelation function (ACF) plot in Figure 4, no observable patterns or significant autocorrelation were detected in the residuals, supporting the assumption of normal distribution. The p-value for the Ljung-Box test was 0.582, suggesting a white noise series from the Pearson residuals. The cumulative periodogram in Figure 4 clearly shows that the residuals are randomly distributed without any specific trend. These results indicate a good fit for the proposed model.

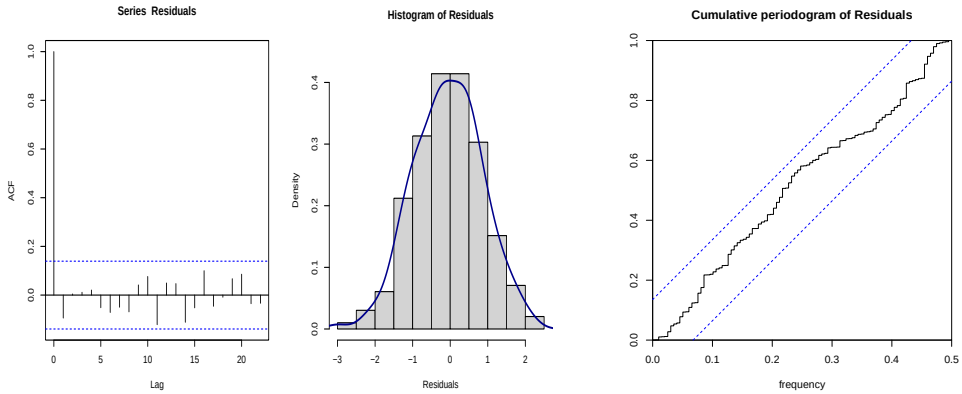

Figure 4: The sample ACF, histogram, and cumulative periodogram of Pearson residuals of the ITS-AR-SN(1) process.

6.2 Local impact analysis based on real datasets

We conducted a local impact study based on the $M(.)$ index for the COVID-19 dataset. In a preliminary analysis, we used $c^* = 3$ in the reference calculation. Figure 5 displays the $M(.)$ index plots for three disturbance scenarios. As shown in Figure 5, observations 31, 98, and 99 appear to have the highest influence on the ECME estimates under the ITS-AR-SN(1) process for each disturbance scenario. To illustrate the impact of these three observations on the parameters, we re-estimated the models by excluding these potentially influential observations to obtain the ECME parameter estimates.

The robustness of the model can be evaluated by considering the influence of an outlier observation (far from the center) on the ECME parameter estimates. Specifically, we can assess how the ECME estimates of the unknown parameters are affected by small disturbances. To further illustrate that observed values in the samples may contain anomalies, we computed the parameter estimates for the ITS-AR-SN(1) process for the actual data both before and after removing observations 31, 98, and 99, as shown in Table 2. After excluding the anomalous data, the statistical measures of mean, standard deviation, skewness, and kurtosis changed to 8.58, 3.91, -0.25, and 2.41, respectively. The ECME estimates for the process in question are robust against anomalies, as all parameters do not change significantly with or without the anomalies. Therefore, it appears that they have little to no impact under the ITS-AR-SN(1) model.

6.3 Forecasting result

The forecasting of the COVID-19 data set is provided based on the conditional mean, where parameters are substituted by their corresponding ECME estimates. The K-fold cross-validation analysis is one of the most popular methods for evaluating models in regression and classification. However, time series forecasting is frequently substituted by an out-of-sample (OS) evaluation due to the intrinsic serial correlation and non-stationarity of the data.

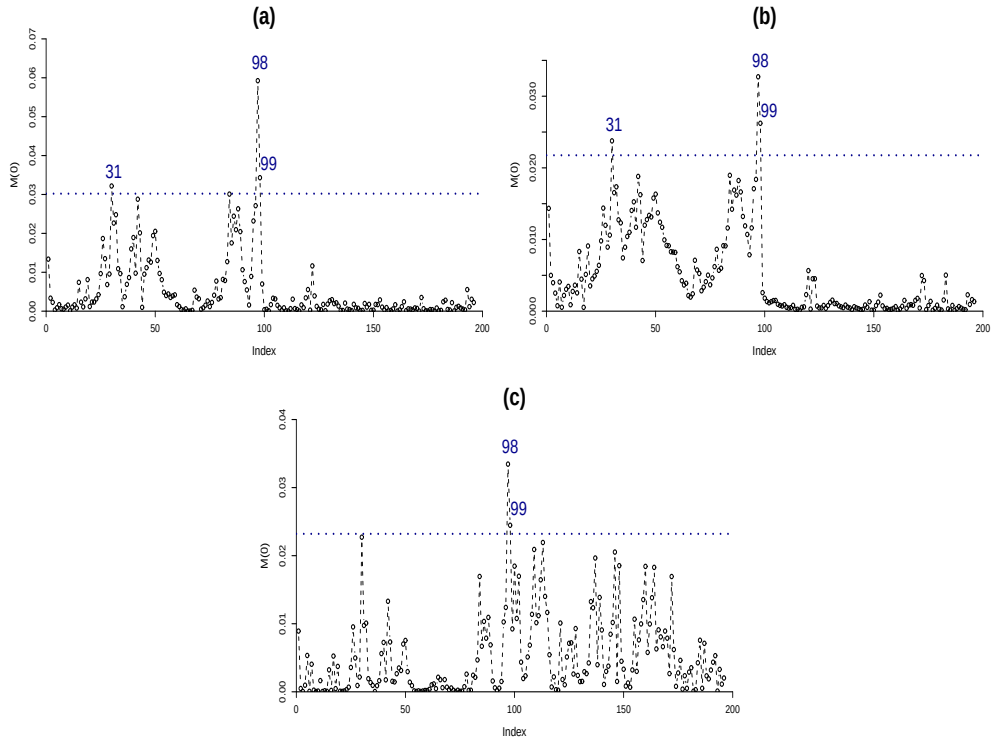

Figure 5: The Index chart of (a) item weight disorder, (b) data, and (c) scale disorders.

Table 2: ECME estimation with or without impact points in the COVID-19 dataset.

Complete data	$\hat{\alpha} = 0.271, \hat{\beta} = 0.418, \hat{\theta} = 0.901, \hat{\mu}_1 = 0.522,$
	$\hat{\mu}_2 = 0.924, \hat{\sigma}_1 = 0.399, \hat{\sigma}_2 = 0.336, \hat{\lambda}_1 = 0.864, \hat{\lambda}_2 = 0.223$
Remove 31-th data	$\hat{\alpha} = 0.272, \hat{\beta} = 0.433, \hat{\theta} = 0.901, \hat{\mu}_1 = 0.513,$
	$\hat{\mu}_2 = 0.924, \hat{\sigma}_1 = 0.393, \hat{\sigma}_2 = 0.336, \hat{\lambda}_1 = 0.822, \hat{\lambda}_2 = 0.219$
Remove 98-th data	$\hat{\alpha} = 0.274, \hat{\beta} = 0.444, \hat{\theta} = 0.903, \hat{\mu}_1 = 0.538,$
	$\hat{\mu}_2 = 0.933, \hat{\sigma}_1 = 0.405, \hat{\sigma}_2 = 0.338, \hat{\lambda}_1 = 0.834, \hat{\lambda}_2 = 0.213$
Remove 99-th data	$\hat{\alpha} = 0.271, \hat{\beta} = 0.433, \hat{\theta} = 0.901, \hat{\mu}_1 = 0.513,$
	$\hat{\mu}_2 = 0.924, \hat{\sigma}_1 = 0.392, \hat{\sigma}_2 = 0.338, \hat{\lambda}_1 = 0.855, \hat{\lambda}_2 = 0.285$
Without all the outliers	$\hat{\alpha} = 0.271, \hat{\beta} = 0.433, \hat{\theta} = 0.901, \hat{\mu}_1 = 0.513,$
	$\hat{\mu}_2 = 0.924, \hat{\sigma}_1 = 0.393, \hat{\sigma}_2 = 0.339, \hat{\lambda}_1 = 0.883, \hat{\lambda}_2 = 0.255$

A portion from the end of the series is excluded for evaluation in order to assess time series issues. Therefore, splitting a dataset into two sets, known as training and test subsamples, is the elementary step in OS evaluation.

The OS analysis is provided to compare the forecasting performances of the ITS-AR-SN(1) model. We consider the full sample sizes of real data series and split the data series into two subsamples, the training subsample with $n - 20$ from the first

data to the 178-th, and the test subsample with 20 last observations. The parameters of the suggested model are calculated using the training sample, and the accuracy of the forecasting is then assessed by obtaining the test observation forecasts using the conditional mean forecasting technique. The ECME estimates of the parameters and the goodness of fit measures values in-sample mean square error (MSE) (IS-MSE) and OS-MSE are given as follows

$$\begin{array}{llll} \hat{\alpha} = 0.271, & \hat{\beta} = 0.418, & \hat{\theta} = 0.901, & \hat{\mu}_1 = 0.522, \quad \hat{\mu}_2 = 0.924, \\ \hat{\sigma}_1 = 0.399, & \hat{\sigma}_2 = 0.336, & \hat{\lambda}_1 = 0.864, & \hat{\lambda}_2 = 0.223, \end{array} .$$

The goodness of fit measures values in-sample mean square error (MSE) (IS-MSE) and OS-MSE are computed as $\mathrm{IS}-\mathrm{MSE} = 0.28,\ \mathrm{OS}-\mathrm{MSE} = 0.34$. Based on the MSE values, we can conclude the ITS-AR-SN(1) model has an adequate fit for COVID-19 data.

The mean absolute percentage error (MAPE) is a measure of prediction accuracy of a forecasting method. Due to evaluating the prediction methodology, the forecasts of test subsample with $n = 179, \ldots, 198$ of the considered data series are reported in 3, for which we know the actual values. Based on 3, the MAPE values of the COVID-19 data series of forecasting results of the ITS-AR-SN(1) model is obtained as 3.8%, which confirms the accuracy of the prediction results. 0.1cm 6.5pt

Table 3: The forecasting of the COVID-19 data set based on the ITS-AR-SN(1) model.

n	179	180	181	182	183	184	185	186	187	188
Actual values	4.81	5.12	5.68	5.99	6.16	6.79	7.47	7.72	7.68	8.01
Forecasting values	5.01	5.58	5.83	6.41	6.63	6.77	7.46	8.10	8.22	8.12
n	189	190	191	192	193	194	195	196	197	198
Actual values	8.02	8.17	8.39	8.55	8.61	8.62	9.05	9.24	9.49	9.68
Forecasting values	8.51	8.44	8.62	8.84	8.97	9.01	8.99	9.51	9.62	9.86

7 Summary and conclusions

The ITS analysis based on AR structures is frequently employed to assess intervention experiments. The asymmetric SN distribution, characterized by distinct features for process innovations, is appropriate as the innovation parameters can vary with each regime. It is reasonable to expect that, as the regime changes, the distribution parameters will also adjust accordingly in our new process. The current study aims to address these challenges by developing a new ITS-AR(1) process that incorporates regime-dependent innovations, asymmetric SN innovations, and treats the intervention variable as an exogenous variable based on a transition function. The proposed ITS-AR-SN(1) process is estimated using the ECME framework, and a comprehensive extraction of disturbance models is presented based on local impact analysis. In the simulation framework, unknown parameters are estimated using the ECME technique based on the ITS-AR-SN(1) process, which is robust against outliers. The theoretical findings are validated with real data. The local impact index related to the COVID-19 data notably identified observations 31, 98, and 99 as influential points in the disturbance designs. It can be inferred that the number of new cases decreased following the

start of COVID-19 booster dose administration, with the effectiveness of the booster becoming apparent approximately six weeks after the policy was implemented.

Future research can be expanded in various areas. For instance, one topic that requires further investigation is the exploration of the advantages of Bayesian approaches and their application in modeling high-dimensional data, which could provide valuable insights for regression modeling.

Some suggestions for future studies include the following:

- Calculating the E-Bayesian estimators of the distribution parameters for the ITS-AR-SN(1) process under various loss functions and informative prior distributions.
- Robust ITS analysis based on symmetric and asymmetric distributions in various sciences.
- Exploring other statistical models using the ITS-AR-SN(1) distribution family.
- Modeling censored data using the family of asymmetric ITS-AR-SN(1) distributions.
- Developing a discrete version of the ITS-AR-SN(1) distribution using various discretization methods and evaluating the effectiveness of the new discrete distribution in modeling discrete observations.

References

Bernal, J.L., Cummins, S. and Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International Journal of Epidemiology,* **46**(1):348–355.

Bordes, S.J., Montorfano, L., West-Ortiz, W., Valera, R., Cracco, A., Alonso, M., Pinna, A.D. and Ebaid, S. (2020). Trends in US kidney transplantation during the COVID-19 pandemic. *Cureus*, **12**(12): e12075.

Bottomley, C., Scott, J.A.G. and Isham, V. (2019). Analysing interrupted time series with a control. *Epidemiologic Methods*, **8**(1):20180010.

Box, G.E.P and Tiao, G.C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association,* **70**(349):70–79.

Cook, D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society,* Series B, **48**(2):133–169.

Emmanuel, O.O., Adebanji, A.O. and Kakaï, R.G. (2014). Arima-noise model for a segmented intervention analysis of maternal health policy on assisted delivery. *International Journal of Applied Mathematics and Statistics,* **52**:62–71.

Ferreira, C.S. and Arellano-Valle, R.B. (2018). Estimation and diagnostic analysis in skew generalized-normal regression models. *Journal of Statistical Computation and Simulation,* **88**(6):1039–1059.

Flammer, C. (2015). Does product market competition foster corporate social responsibility. *Strategic Management Journal,* **38**:163–183 .

Freni-Sterrantino, A., Ghosh, R.E., Fecht, D., Toledano, M.B., Elliott, P., Hansell, A.L. and Blangiardo, M. (2019). Bayesian spatial modelling for quasi-experimental designs: An interrupted time series study of the opening of Municipal Waste Incinerators in relation to infant mortality and sex ratio. *Environment International,* **128**:109–115.

Garay, A.M., Lachos, V.H., Labra, F.V. and Ortega, E.M.M. (2014). Statistical diagnostics for nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of Statistical Computation and Simulation,* **84**(8):1761–1778.

Gasparrini, A., Gorini, G. and Barchielli, A. (2009). On the relationship between smoking bans and incidence of acute myocardial infarction. *European Journal of Epidemiology,* **24**(10):597–602.

Gerlach, A.L. (2018). *Increasing Organ Donations in Maryland: An Interrupted Time Series Analysis*. Ph.D. Thesis, Walden University, Minneapolis, MN, USA.

Hudson, J., Fielding, S. and Ramsay, C.R. (2019). Methodology and reporting characteristics of studies using interrupted time series design in healthcare. *BMC Medical Research Methodology,* **19**(1):137.

James Lopez, B. (2017). Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International Journal of Epidemiology,* **46**(1):348–355.

Jandoc, R., Burden, A.M., Mamdani, M., Lévesque, L.E. and Cadarette, S.M. (2015). Interrupted time series analysis in drug utilization research is increasing: systematic review and recommendations. *Journal of Clinical Epidemiology,* **68**(8):950–956.

Kontopantelis, E., Doran, T., Springate, D.A., Buchan, I. and Reeves, D. (2015). Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *British Medical Journal,* **350**, h2750.

Liu, C.H. and Rubin, D.B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika,* **81**(4):633–648.

Liu, S., Leiva, V., Ma, T. and Welsh, A.H. (2016). Influence diagnostic analysis in the possibly heteroskedastic linear model with exact restrictions. *Statistical Methods and Applications,* **25**(2):227–249.

Liu, Y., Ji, G. and Liu, S. (2015). Influence diagnostics in a vector autoregressive model. *Journal of Statistical Computation and Simulation,* **85**(13):2632–2655.

Malladi, S. and Lu, Q. (2023). Intervention time series analysis and forecasting of organ donor transplants in the US during the COVID-19 era. *Forecasting,* **5**(1):229–255.

Meng, X.L. and Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika,* **80**(2):267–278.

Morales, D.R., et al. (2020). Impact of EU regulatory label changes for diclofenac in people with cardiovascular disease in four countries: Interrupted time series regression analysis. *British Journal of Clinical Pharmacology,* **87**(3):1129–1140.

Pooyannik, F. abd Khodadadi, Z. (2025). Robust Bayesian Inference of Dynamic Intervention Two-Pieces Normal Autoregressive Process with Local Influence Analysis. *Iranian Journal of Science*, 1–17.

Poppe, A. (2020). Impact of the healthcare system, macro indicator, general mandatory quarantine, and mask obligation on COVID-19 cases and death in six latin American countries: An interrupted time series study. *Frontiers in Public Health*, **8**:607832.

Quandt, R.E. (1985). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, **53**(284):873–880.

Schaffer, A.L., Dobbins, T.A. and Pearson, S.A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. *BMC Medical Research Methodology,* **21**(1):58.

Shadish, W.R., Cook, T.D. and Campbell, D.T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth Cengage Learning.

Suarez-Pierre, A., Choudhury, R., Carroll, A.M., King, R.W., Iguidbashian, J., Cotton, J., Colborn, K.L., Kennealey, P.T., Cleveland, J.C., Pomfret, E. and Fullerton, D.A. (2021). Measuring the effect of the COVID-19 pandemic on solid organ transplantation. *The American Journal of Surgery,* **224**(1):437–442.

Thistlethwaite, D.L. and Campbell, D.T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology,* **51**(6):309.

Turner, S.L., Forbes, A.B., Karahalios, A., Taljaard, M. and McKenzie, J.E. (2021). Evaluation of statistical methods used in the analysis of interrupted time series studies: A simulation study. *Medical Research Methodology,* **21**(1):181.