Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **540,** 521–531 (2025) https://doi.org/10.1093/mnras/staf764
Advance Access publication 2025 May 13

Novel SIMEX algorithm for autoregressive models to estimate AGN variability

F. Elorrieta ,[1,2,3]★ W. Palma,[1] S. Eyheramendy,[1,4,5] F. E. Bauer[1,6,7] and E. Camacho[6,8]

[1]*Millennium Institute of Astrophysics, Nuncio Monseñor Sótero Sanz 100, Of 104, 7500011, Providencia, Santiago, Chile*
[2]*Departmento de Matematicas, Facultad de Ciencia, Universidad de Santiago de Chile, Av. Libertador Bernardo O'Higgins 3663, 9170200, Estacion Central, Santiago, Chile*
[3]*Center for Interdisciplinary Research in Astrophysics Space Exploration (CIRAS), Universidad de Santiago de Chile, Av. Libertador Bernardo O'Higgins 3363, 9170200, Estacion Central, Santiago, Chile*
[4]*Faculty of Engineering and Sciences, Universidad Adolfo Ibanez, Diagonal Las Torres 2700, 7941169, Peñalolen, Chile*
[5]*Data Observatory (DO) Foundation, ANID Technology Center No. DO210001, Eliodoro Yáñez 2990, 7510277 Provdencia, Santiago, Chile*
[6]*Instituto de Astrofísica and Centro de Astroingeniería, Facultad de Física, Pontificia Universidad Católica de Chile, Campus San Joaquín, Av. Vicuña Mackenna 4860, 7820436 Macul Santiago, Chile*
[7]*Space Science Institute, 4750 Walnut Street, Suite 205, Boulder, Colorado 80301, USA*
[8]*Center for Astrophysics – Harvard and Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*

Accepted 2025 May 7. Received 2025 April 30; in original form 2025 March 21

ABSTRACT

The origin of the variability in accretion discs of active galactic nuclei (AGNs) is still unknown, but its behaviour can be characterized by modelling the time series of optical wavelength fluxes with damped random walk (DRW) being the most popular model for this purpose. The DRW depends on a fluctuation amplitude σ and damping time-scale τ, the latter potentially related to the mass and accretion rate on to the massive black hole. Estimating τ is challenging. Popular methods such as maximum likelihood (ML) and least-square error (LSE) result in biased estimators. This bias typically arises due to: (i) light curve observed with additive noise, (ii) the cadence scheme, and/or (iii) autocorrelation parameter near one – the so-called unit root problem, a known statistical challenge. To improve the estimation procedures for τ, we developed a simulation-extrapolation (SIMEX) methodology in the context of time series analysis. We applied it to both a standard autoregressive process (regular time intervals) and a recently developed irregularly autoregressive process (iAR). Its performance was evaluated under near-unit-root behaviour and additive noise via Monte Carlo simulations. This methodology was also tested in AGN light curves from the Zwicky Transient Facility survey, assessing its accuracy in estimating τ. Simulation results confirm that the SIMEX approach outperforms ML and LSE methods, reducing estimation bias by 30 per cent to 90 per cent. Real-data applications validate the methodology, yielding better fits and lower mean squared errors. A more accurate estimation of τ can contribute to understanding the links between DRW parameters and physical characteristics of AGN.

Key words: methods: data analysis – methods: statistical – galaxies: general – galaxies: nuclei.

1 INTRODUCTION

The origin of variability in accretion discs of active galactic nuclei (AGNs) is still unknown, but its behaviour, in the optical part of the spectrum, can be well described as a stochastic process. To model optical light curves coming from the accretion disc of AGNs, the damped random walk (DRW) process has been a popular choice, playing an instrumental role in statistically quantifying the intrinsic variability of AGN light curves (Kelly, Bechtold & Siemiginowska 2009; MacLeod et al. 2010; Graham et al. 2014; Kasliwal, Vogeley & Richards 2017; Burke et al. 2021; Sánchez-Sáez et al. 2021). This model is characterized by two main parameters: the damping time-scale τ (also called the relaxation time), which quantifies how fast the flux variations lose memory of past values; and the variance σ, which measures the amplitude of the variation. In the power spectral density (PSD) that describes the statistical properties of AGN variability in the frequency domain, τ is associated with the transition between white noise (with a flat slope at low frequencies) and red noise (with a slope $\propto \nu^{-2}$ at high frequencies).

Several authors have found correlations between τ and the physical properties of black holes such as the mass or luminosity (Burke et al. 2021; Arévalo et al. 2024). If these correlations or dependencies are confirmed across several orders of magnitude, i.e. from supermassive black holes (SMBHs) to intermediate-mass black holes (IMBHs) and from faint to bright AGNs, variability could be established as a powerful tool to find and characterize these objects. In that sense, regardless of the stochastic model used, the fit output parameters are tools that can be used to better understand the intrinsic physical mechanism behind the variability in accretion discs. Finally, it

★ E-mail: felipe.elorrieta@usach.cl

© 2025 The Author(s).
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/mnras/article/540/1/521/8129690 by guest on 22 May 2025

is important to point out the empirical challenge in astronomy of understanding and dealing with observational errors in which observing conditions and instrumental effects may play an important role.

MacLeod et al. (2010) have applied the DRW model to quasar light curves, enabling the estimation of characteristic time-scales and providing insights into the underlying physical processes driving quasar variability. The mentioned authors' work, has highlighted that AGN variability often exhibits red-noise behaviour, where fluctuations at longer time-scales dominate, and the DRW model offers a framework to quantify this behaviour.

The accurate modelling of AGN variability remains an ongoing endeavour, with the DRW model serving as a foundational tool but not necessarily the sole solution for all cases e.g. Mushotzy et al. (2008). For example Kozłowski et al. (2010) explored AGN variability by employing structure functions, which offer another perspective on AGN light curve behaviour. Kozłowski (2016, 2017) have also shown that maximum likelihood estimation (MLE) can lead to biased estimates of DRW parameters, particularly when light curves are short or sparsely sampled. Another useful model for estimating the damping time-scale and σ is the *irregular autoregressive model* (iAR) (Eyheramendy, Elorrieta & Palma 2018). This model assumes discrete times as opposed to continuous times (like the DRW model) and can handle irregular gaps in observational times. The errors of the model can follow different distributions. For example Gaussian (as the DRW), Gamma or t-Student, allowing some extra flexibility. The parameters of the iAR are denoted by ϕ corresponding to the autoregressive parameter, and σ, the standard deviation of the model. The parameter ϕ is related to the damping time-scale by the formula,

$$\tau = -\frac{1}{\log(\phi)}. \tag{1}$$

A particular challenge facing these models (DRW and iAR) consists of estimating the relaxation parameter τ as a function of the autocorrelation parameter ϕ of the model, when ϕ is close to 1. This happens when an observed time series remains with significant autocorrelation regardless of large observational gaps. In statistics, this is a known problem in time series analysis that is commonly referred to as the near root estimation problem.

Time series parameter estimation in the near root context is a difficult problem to deal with because in this situation the process is unstable and near to non-stationary conditions. Also, the estimation procedure is conducted close to the boundaries of the parameter space [i.e. (0,1)] generating both theoretical and computational problems (see for example Barreto-Souza & Chan 2024 and references therein).

The near root problem is particularly acute in the context of irregularly observed time series, leading to biased estimators. In order to improve the parameter estimation, we propose to fit the time series with the iAR model, and correct the bias in the estimation of the parameter of the model by an adaptation of the SIMulation-EXtrapolation (SIMEX) algorithm. The SIMEX algorithm is not new in astronomy. It has been implemented in several different contexts (e.g. Freeman et al. 2009; Feigelson & Babu 2012; Juvela et al. 2013; Carstens, Xia & Yadavalli 2017; Rana et al. 2017; Fernández-Hernández, Montiel & Rodríguez-Meza 2019; Escamilla-Rivera, Levi Said & Mifsud 2021).

The SIMEX algorithm was introduced and discussed by Cook & Stefanski (1994), Stefanski & Cook (1995), and Raymond & Stefanski (1996) in the context of linear regression models, to correct for bias in the estimation of the parameters, when the bias was induced by measurement errors on covariates. We adapted the algorithm for the case of autoregressive models, which lack a covariate with measurement error, but where we could assume instead additive noise in the stochastic process. Additive noise arises naturally in astronomical observations with, for example fluctuations in the precision of some instruments. We assessed the performance of the algorithm on simulated series generated from a regular autoregressive model (i.e. AR) and from an irregular autoregressive model (i.e. iAR).

We also model the flux time series of IMBH candidates. The light curves were obtained from the Forced Photometry service of the Zwicky Transient Facility (ZTF), comprising 5 yr of observations in the g and r bands, with much more sparsely acquired additional i-band observations. This data set presents two challenges, the cadence of 2–3 d, similar to the expected characteristic time-scales of IMBHs, that might introduce bias in the parameter estimation, and the seasonal gaps, due to the fact that the facility is a ground-based observatory plus the observational effect of other astronomical objects like the moon, implying an irregularly observed time series.

The remainder of the paper is organized as follows. In Section 2, the basics of the regular and irregular autoregressive models are presented. The methodologies for parameter estimation are described in Section 3: MLE in Section 3.1; LSE in Section 3.2; the general framework of estimation through the SIMEX algorithm in Section 3.3; and the novel adaptation for autoregressive models in Section 3.4. The finite sample performance of these estimation methods is assessed via Monte Carlo simulations, and the results are shown in Section 4. The proposed estimation methodology is presented in Section 5 on real and assimilated data. We performed a data assimilation experiment, where we generated synthetic light curves of AGN objects, in order to estimate the relaxation parameter in a context where this parameter is known. And also, we show examples of the SIMEX estimation in AGN light curves observed from the ZTF survey. In Section 6, the conclusions and final remarks of this work are presented.

2 MODELS

2.1 Regular autoregressive model

A stochastic process y_t is said to be autoregressive of order one, $AR(1)$, if it is defined by:

$$y_1 = \sigma\,\varepsilon_1, \quad y_t = \phi\,y_{t-1} + \sigma\sqrt{1-\phi^2}\,\varepsilon_t, \tag{2}$$

for $t = 2, \ldots, n$, where ε_t is a white noise sequence with zero mean and unit variance. Moreover, y_t is said to be a weakly stationary process when $|\phi| < 1$ (Brockwell & Davis 1991). The mean and variance of the process is therefore given by

$$E(y_t) = 0 \text{ and } Var(y_t) = \sigma^2 \text{ for all } t, \tag{3}$$

and for any two observational times t, s we can define the autocovariance function as depending only on the absolute time difference,

$$\gamma(|t-s|) = E(y_t\,y_s) = \sigma^2\,\phi^{|t-s|}, \tag{4}$$

as well as the autocorrelation function (ACF),

$$\rho(|t-s|) = \frac{\gamma(|t-s|)}{\gamma(0)} = \phi^{|t-s|}.$$

If the $AR(1)$ process Y_t is observed with an additive error ξ_t with zero mean and known variance σ_ξ^2, the model (2) can be rewritten as,

$$Y_t = y_t + \xi_t \tag{5}$$

where y_t is the $AR(1)$ process without error defined in (2).

Downloaded from https://academic.oup.com/mnras/article/540/1/521/8129690 by guest on 22 May 2025

2.2 Irregular autoregressive model

We denote y_{t_j} as an observation measured at time t_j, and consider an increasing sequence of observational times $\{t_j\}$ for $j = 1, \ldots, n$. As defined by Eyheramendy et al. (2018), the iAR process is:

$$y_{t_1} = \sigma\, \varepsilon_{t_1}, \quad y_{t_j} = \phi^{t_j - t_{j-1}}\, y_{t_{j-1}} + \sigma \sqrt{1 - \phi^{2(t_j - t_{j-1})}}\, \varepsilon_{t_j}, \tag{6}$$

for $j = 2, \ldots, n$, where $\phi \in (0, 1)$ and ε_{t_j} are assumed independent Gaussian random variables with zero mean and unit variance (note that this model allows also for other distributional assumptions Eyheramendy et al. 2018). Note that

$$E(y_{t_j}) = 0 \text{ and } Var(y_{t_j}) = \sigma^2 \text{ for all } y_{t_j}, \tag{7}$$

and the covariance between y_{t_k} and y_{t_j} is $E(y_{t_k}\, y_{t_j}) = \sigma^2\, \phi^{t_k - t_j}$, for $k \geq j$.

Given the results above, the sequence $\{y_{t_j}\}$ corresponds to a second-order or weakly stationary process (Eyheramendy et al. 2018). Note that when the observational time gaps are equal, i.e. $\delta_j = t_j - t_{j-1} = c$, the model reduces to the standard regular autoregressive model AR (Brockwell & Davis 1991). However, astronomical light curves are nearly always observed with irregular times gap. For example in Eyheramendy et al. (2018), the authors discussed that a suitable representation for the observational times of the Vista Variable of the Via Lactea survey (Minniti et al. 2010) is given when the time gaps δ_j are defined by a mixture of two exponential distributions with rates λ_1 and λ_2, and weights w_1 and w_2, such that

$$f(\delta_j|\lambda_1, \lambda_2, \omega_1, \omega_2) = \omega_1 g(\delta_j|\lambda_1) + \omega_2 g(\delta_j|\lambda_2) \quad \forall j = 2, \ldots, n, \tag{8}$$

where $g(.)$ is the exponential distribution. With this mixture model, time-gaps can vary between short and longer periods without observations. Since the time gaps δ_j must be greater than zero, other distributions considered in the literature for a random generation of these time gaps are the $Exponential$, $Gamma$ or $Uniform(a, b)$ with $b > a > 0$.

3 ESTIMATION

To fully specify the model described above, it is required to estimate the two parameters: ϕ and σ. In this study, we focus on the estimation of the parameter ϕ and assume, without loss of generality, that $\sigma = 1$, which can be achieved by standardizing the sequences.

In what follows, we describe three methods for estimating the parameter ϕ. These include the well-known MLE and LSE. The third approach considered is based on the SIMEX algorithm (Stefanski & Cook 1995).

3.1 Maximum likelihood estimation

A well-known method for estimating the parameter ϕ is MLE which consists in optimizing the likelihood function as follows:

$$L(\phi) = \log(\det(\Sigma_\phi)) + Y' \Sigma_\phi^{-1} Y$$
$$\widehat{\phi}_n = \text{argmin } L(\phi),$$

where, $\boldsymbol{Y} = (y_{t_1}, \ldots, y_{t_n})$ represents the vector of random variables that are assumed to follow an autoregressive model and $\boldsymbol{\Sigma}_\phi = \left(\phi^{t_i - t_j}\right)_{i,j=i,\ldots,n}$ corresponds to its covariance matrix.

We observe that for the regular autoregressive model, the variance of the ML estimator is explicitly given by

$$Var(\widehat{\phi}_n) = \frac{1}{n}\left(1 - \phi^2\right).$$

Note also that, this variance approaches 0 as $\phi \to 1$.

3.2 Least-square error estimation

An alternative method for estimating the parameter ϕ is the LSE. In this case, the estimated parameter is obtained as the argument which minimizes the sum of squares errors defined by

$$J(\phi) = \sum_{i=2}^{n} (y_{t_i} - \phi^{t_i - t_{i-1}}\, y_{t_{i-1}})^2$$
$$\tilde{\phi}_n = \text{argmin } J(\phi),$$

where $\tilde{\phi}_n$ is the least-square estimator of ϕ.

3.3 SIMEX method

Before we delve into the specifics for estimating ϕ based on the SIMEX algorithm, we introduce in general how this algorithm works. In Section 3.4, we present the adaptation of this algorithm for the problem in this study.

Cook & Stefanski (1994) introduced the SIMEX method, which has become popular among statisticians and scientists due to its simplicity and versatility (e.g. Li & Lin 2003; Guolo 2008; Wang, Fan & Wang 2010; Amico & Van Keilegom 2018; Li & Huang 2019). This method is effective for models with measurement error in the predictors when the standard errors of measurements are known or can be reasonably estimated. It yields nearly asymptotically unbiased and efficient parameter estimates for various regression models and less-biased non-regression population parameter estimates (e.g. Cook & Stefanski 1994; Stefanski & Cook 1995; Raymond & Stefanski 1996; Carroll, Maca & Ruppert 1999; Piesse & Feng 2008; Küchenhoff, Mwalili & Lesaffre 2006).

The SIMEX method involves adding simulated measurement error with increasing variance to the original data, applying statistical models to these increasingly error-prone data sets, identifying the trend of model parameter estimates against the variance of the added measurement error, and extrapolating this trend back to a point with no measurement error. While it does not provide an analytic solution to the measurement error problem and it is computationally intensive, its strength lies in its broad applicability to both simple and sophisticated measurement error models (Cook & Stefanski 1994). Given that standard errors of measurements are commonly estimated and reported in astronomy, the SIMEX method has potential in a wide range of applications.

3.3.1 The simulation step

The SIMEX procedure begins with a simulation stage where simulated measurement errors are added to the predictor, say X, making it increasingly error-prone. It is assumed that the predictor X is observed with error, i.e. $X = X_{\text{true}} + V$, where V is a random variable representing the measurement error of the predictor. In our

Downloaded from https://academic.oup.com/mnras/article/540/1/521/8129690 by guest on 22 May 2025

implementation, we assume $V \sim N(0, \sigma^2)$. These error-augmented predictors, known as re-measurements of X, are created by first choosing a set of monotonically increasing small numbers, such as $\lambda = \delta, 2\delta, \ldots, l\delta$, where δ is the constant increase in λ in each simulation. For each λ, an artificial error $\sqrt{\lambda}U$ is generated, where U is a random variable, which we also assume $N(0, \sigma^2)$. The re-measurements, $X(\lambda)$, are then $X(\lambda) = X + \sqrt{\lambda}U = X_{\text{true}} + V + \sqrt{\lambda}U$. The variance of the inflated measurement errors of $X(\lambda)$ is $(1 + \lambda)\sigma^2$. σ^2 is considered known, or there is a way of properly estimating it.

Next, the response variable, say Y, is regressed on $X(\lambda)$, i.e. $Y = \beta_0 + \beta_1 X(\lambda) + \epsilon$. As λ increases, $X(\lambda)$ becomes more endogenous, and the regression coefficient estimates become increasingly biased. This relationship forms the basis for the extrapolation. To reduce sampling variation, multiple re-measurements of X are generated for each λ. The naive coefficients for each re-measured X are denoted as $\beta_{k,\text{naive}}(\lambda, B)$, where k indicates intercept ($k = 0$) or slope ($k = 1$), and B is the number of re-measured predictors. The sample mean, $\hat{\beta}_{k,\text{naive}}(\lambda)$, is calculated for extrapolation to reduce sampling error.

The choice of λ is crucial. Stefanski & Cook (1995) suggest restricting λ values to a neighbourhood of zero to minimize extrapolation errors, proposing a range between 0 and 2. Different studies have used various λ values: Carroll et al. (1999) used $\lambda = \frac{1}{2}, \ldots, \frac{4}{2}$ for non-parametric regression, while Piesse & Feng (2008) used $\lambda = \frac{2}{5}, \frac{4}{5}, \ldots, \frac{10}{5}$.

3.3.2 *The extrapolation step*

At this point, there is a series of $\hat{\beta}_{k,\text{naive}}(\lambda)$ corresponding to different values of λ. The objective of the extrapolation stage is to identify the relationship between $\beta_{k,\text{naive}}(\lambda)$ and λ. Recall that the re-measurement of X, $X(\lambda)$, has a measurement error variance equal to $(1 + \lambda)\sigma^2$. Therefore, the $\beta_{k,\text{naive}}(\lambda)$ associated with $\lambda = -1$ is the SIMEX estimate of the true parameter in the error-free model, denoted as $\beta_{k,\text{simex}}$. The SIMEX estimate $\beta_{k,\text{simex}}$ is obtained by extrapolation in the regression model learned for $\beta_{k,\text{naive}}(\lambda)$ and λ.

There are many possible choices for the extrapolation function. The usual candidates include the linear, the quadratic, and the rational extrapolats (Carrol et al. 2006). The linear extrapolat function is

$$\beta_{k,\text{naive}}(\lambda) = \gamma_{k,1} + \gamma_{k,2}\lambda + \epsilon_{k,\text{lin}} \quad (k = 0, 1),$$

and the simple quadratic extrapolant function

$$\beta_{k,\text{naive}}(\lambda) = \psi_{k,1} + \psi_{k,2}\lambda + \psi_{k,3}\lambda^2 + \epsilon_{k,\text{qua}} \quad (k = 0, 1),$$

where the γs and the ψs are the parameters of the extrapolated model to be estimated, and the ϵs are the model residuals. The parameters of the extrapolated models can be estimated with ordinary least squares (OLS) regressions, and the conventional tools of model diagnostics can be applied to assess model fit.

After choosing the appropriate extrapolat model and estimating the model parameters, it is easy to calculate the fitted value for $\lambda = -1$, which is $\hat{\beta}_{k,\text{simex}}$, and the SIMEX process is complete.

3.4 Novel implementation of the SIMEX algorithm for autoregressive models of order one

We consider the observations y_{t_j} for $j = 1, \ldots, n$, obtained at irregular time gaps $t_1, \ldots, t_n$. For the iAR model (6), if the true parameter ϕ is *close* to 1, common statistical methods to estimate it will generate biased estimators. The reason for this problem is that the parameter lies close to the boundary of the parametric space, i.e. $(-1, 1)$, which challenges the computations. We propose to adapt the SIMEX algorithm in order to obtained better estimators.

The original observations $y_{t_1}, \ldots, y_{t_n}$ are transformed by including additive noise to each of them.

$$Y_{t_j} = y_{t_j} + \eta_{t_j}, \tag{9}$$

where η_{t_j} is a zero-mean additive noise with known variance σ_η^2. We implement the SIMEX algorithm to the iAR model on Y_{t_j}. Under the standard SIMEX setting, the predictor with measurement error corresponds to the previous observation $Y_{t_{j-1}}$ for $j = 2, \ldots, n$. Specifically, the two steps develop as follows.

(i) **SIMULATION**: for each $\lambda = \delta, 2\delta, \ldots, l\delta$ simulate:

$$\begin{aligned} Y_{t_j}(\lambda) &= Y_{t_j,\lambda} = Y_{t_j} + \lambda\, \nu_{t_j} \\ &= Y_{t_j,\text{true}} + \eta_{t_j} + \lambda\, \nu_{t_j} \end{aligned}$$

for $j = 1, \ldots, n$ and $\delta \geq 0$, where ν_{t_j} is the white noise sequence with zero mean and variance σ_ν^2 added to the observed time series. Once $Y_{t_1,\lambda}, \ldots, Y_{t_n,\lambda}$ have been simulated, and assumed that they have been generated under the iAR model (6), an estimator $\hat{\phi}_{\lambda,i}$ for $i = 1, \ldots, B$ is obtained. Then to reduce the variance, $\hat{\phi}_\lambda = \frac{1}{B}\sum_{i=1}^{B} \hat{\phi}_{\lambda,i}$ is the estimator of ϕ for a given λ. As λ increases, $\hat{\phi}_\lambda$ becomes increasingly biased.

In our implementation, two scenarios are considered: (1) when the time series has additive error (e.g. in equation 9), and (2) when it does not (e.g. in equation 6). The variance of the inflated measurement errors is $((1 + \lambda)\sigma_\nu^2)$ in the presence of measurement error and $(\lambda\sigma_\nu^2)$ otherwise, since in this case $Y_{t_j} = Y_{t_j,\text{true}}$.

(ii) **EXTRAPOLATION**: after obtaining $\hat{\phi}_\lambda$ for $\lambda = \delta, 2\delta, \ldots, l\delta$, the objective is to find the relationship between λ and $\hat{\phi}_\lambda$. This is done using linear regression analysis. Our implementation offers four alternatives to fit the regression. It can be a polynomial regression of degree from 1 to 4, corresponding to 'linear', 'quadratic', 'cubic', and 'bi-quadratic', respectively. Also, our implementation allows to set different weights to the points generated from the estimates $\widehat{\phi}(\lambda)$ as follows:

(a) **equal**: $w_i = 1/N\ \forall i = 1, \ldots, N$.

(b) **exponential**: $w_i = \alpha(1 - \alpha)^{i-1}$, where $i = 1, \ldots, N$ and α is a parameter that determines the weight of the first point. Note that for this definition, the weights decrease as the additive error variance increases.

(c) **invexponential**: $w_i = \alpha(1 - \alpha)^{N-i}$, where $i = 1, \ldots, N$ and α is a parameter that determines the weight of the last point. Note that for this definition, the weights increase as the additive error variance increases.

(d) **Gaussian**: $w_i = f_{i\delta}(l\delta/2, \sigma^2)$, where $i = 1, \ldots, l$ and $f_{i\delta}(l\delta/2, \sigma^2)$ is the density of a normal distribution with mean $\mu = l\delta/2$ and variance σ^2 evaluated at the point $i\delta$. Note that for this definition we get higher weights at intermediate points.

Once this model is estimated, the SIMEX estimator of ϕ is then obtained by setting $\lambda = 0$ (when there is no measurement error)

Downloaded from https://academic.oup.com/mnras/article/540/1/521/8129690 by guest on 22 May 2025

or $\lambda = -1$ (when there is measurement error). The Algorithm 1 summarizes the simulation and extrapolation steps:

Algorithm 1: SIMEX algorithm for autoregressive model of order one.

Input: B (number of repetitions); N (number of simulations); δ (gap between points); $type$ (type of fit); $weight$ (weights for the fit).
for $i = 1$ **to** B **do**
  for $j = 1$ **to** N **do**
    (i) Simulate Y_{t_j} as the sum of the observed time series y_{t_j} with a noise ν_{t_j} generated from a normal distribution with variance $\lambda\sigma_\nu^2$ and $\lambda = \delta, 2\delta, \ldots, N\delta$;
    (ii) Estimate the autoregressive parameter $\hat{\phi}_{\lambda,i}$ for the simulated time series Y_{t_j};
  end
  Obtain $\hat{\phi}\lambda$ by averaging over B repetitions, such that: $\hat{\phi}_\lambda = \frac{1}{B}\sum_{i=1}^{B}\hat{\phi}_{\lambda,i}$;
end
Fit a polynomial regression to $\hat{\phi}_\lambda$ using the specified $type$ and $weight$;
Obtain the SIMEX estimate by setting $\lambda = 0$ or $\lambda = -1$ depending on the scenario;

Fig. 1 illustrates the estimation process of the SIMEX algorithm for both scenarios. In panel (a), an exponential weighted quadratic fit over $N = 15$ points with $\delta = 1/30$ is used for a time series without measurement error. In panel (b), an equally weighted quadratic fit with $\delta = 0.5$ and $N = 20$ is used for a time series with measurement error variance $\sigma_\nu^2 = 0.1$. Note that the main difference in the algorithm in the case that the time series has or does not have measurement error is the value of the parameter λ at which the extrapolation is made, since in the case without error it is extrapolated to $\lambda = 0$ and in the case with error it is extrapolated to $\lambda = -1$. In addition, it is important to note that in the case with small measurement error variances, larger delta values are usually required, since the aggregate error variance at each point is $\lambda\sigma_\nu^2$. Finally, in both examples presented in Fig. 1 the SIMEX estimate is closer to the true parameter than the MLE estimate.

4 MONTE CARLO EXPERIMENTS

This section provides a Monte Carlo study that assesses the finite-sample performance of SIMEX, ML, and LSE estimators. We fixed the parameter $\phi = 0.99$ in all the experiments, and we consider two sample sizes $n = 200$ and $n = 500$. Each experiment was repeated $m = 1000$ times. The performance criteria used to evaluate the estimation methods were the standard deviation of the estimated parameters and the difference between the estimated and true parameter, measured by the bias and the root mean square error (RMSE). Tables 1–2 show results for the regular AR model, and Tables 3–4 show the results for the iAR model. Table 1 shows results for the standard AR model. Table 2 shows results for the AR model with additive noise with standard deviation equal to 0.1. In addition, Tables 3–4 show results for the iAR model. Observational times were generated from two distributions, one that guarantees small time gaps such as Uniform(0,1) and another with larger gaps such as Gamma(3,1). The parameter settings for the SIMEX algorithm for all the experiments considered were $N = 15$, $B = 30$, linear fit and points weighted equally. The only parameter that changes between experiments is δ which is defined as $\delta = 1/50$ for the regular spacing case; $\delta = 1/70$ for the gamma gaps case; $\delta = 1/100$ for the uniform time gap case; and $\delta = 0.3$ for the additive noise case.

The Monte Carlo experiments show that the SIMEX estimator generally outperforms the LSE and ML estimators in estimating the autoregressive parameter ϕ. In scenarios with no additive noise and regular observational times (Table 1), SIMEX has the lowest bias and RMSE, especially as the sample size increases from $n = 200$ to $n = 500$, indicating high accuracy and precision. When additive noise ($\sigma_\nu = 0.3$) is introduced (Table 2), SIMEX maintains superior performance, with lower bias and RMSE than LSE and ML across both sample sizes. Under irregular observational times (Tables 3–4), SIMEX continues to demonstrate robustness, particularly with observation times drawn from a Gamma(3,1) distribution, where it shows minimal bias and low RMSE. Even under the more

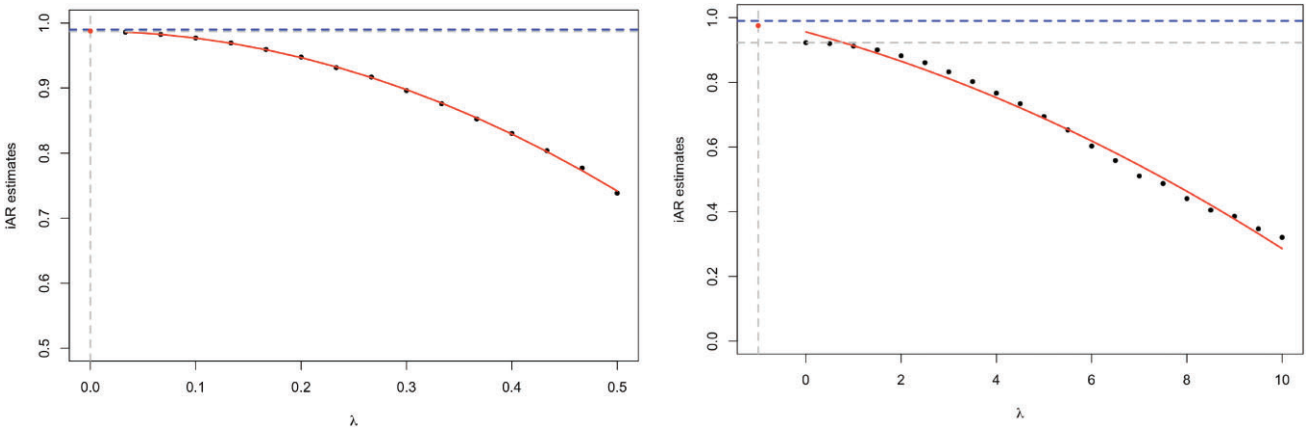


Figure 1. Illustration of the SIMEX estimation algorithm. The black dots are the iAR estimates of the ϕ parameter when the variance measurement error is inflated by λ. The red line represents the model fitted to the black dots. The red dot indicates the SIMEX estimation of the ϕ parameter of the iAR model. The dashed blue line is the true value of the parameter while the horizontal dashed grey line denotes the MLE estimates. Panel (a) shows the case where the time series has no measurement error, and panel (b) illustrates the case with measurement error. For panel (a), the SIMEX algorithm parameters were: $N = 15$, $B = 30$, $\delta = 1/30$, quadratic fit, and exponential weighting of points. For panel (b), the parameters were: $N = 20$, $B = 30$, $\delta = 0.5$, quadratic fit, and equal weighting of points.

Downloaded from https://academic.oup.com/mnras/article/540/1/521/8129690 by guest on 22 May 2025

Table 1. AR estimation of the ϕ parameter.

	n	Parameter	Estimate	Bias	SD	Theoretical SD (∗)	RMSE
SIMEX	200	0.990	0.982	−0.008	0.021	0.010	0.022
LSE	200	0.990	0.965	−0.025	0.024	0.010	0.035
ML	200	0.990	0.965	−0.025	0.022	0.010	0.033
SIMEX	500	0.990	0.996	0.006	0.008	0.006	0.010
LSE	500	0.990	0.981	−0.009	0.011	0.006	0.014
ML	500	0.990	0.981	−0.009	0.010	0.006	0.014

Table 2. AR estimation of the ϕ parameter when the autoregressive model is generated with an additive noise with a standard deviation $\sigma_\nu = 0.3$. (∗) Theoretical SD is under regular times and without additive noise.

	n	Parameter	Estimate	Bias	SD	Theoretical SD (∗)	RMSE
SIMEX	200	0.990	0.977	−0.013	0.033	0.010	0.036
LSE	200	0.990	0.940	−0.050	0.039	0.010	0.063
ML	200	0.990	0.940	−0.050	0.038	0.010	0.062
SIMEX	500	0.990	0.993	0.003	0.014	0.006	0.014
LSE	500	0.990	0.960	−0.030	0.021	0.006	0.036
ML	500	0.990	0.960	−0.030	0.021	0.006	0.036

Table 3. iAR estimation of the ϕ parameter with observational times from Gamma(3,1) distribution. (∗) Theoretical SD is under regular times and without additive noise.

	n	Parameter	Estimate	Bias	SD	Theoretical SD (∗)	RMSE
SIMEX	200	0.990	0.987	−0.003	0.008	0.010	0.009
LSE	200	0.990	0.982	−0.008	0.009	0.010	0.012
ML	200	0.990	0.982	−0.008	0.009	0.010	0.011
SIMEX	500	0.990	0.992	0.002	0.004	0.006	0.005
LSE	500	0.990	0.987	−0.003	0.005	0.006	0.006
ML	500	0.990	0.987	−0.003	0.004	0.006	0.005

Table 4. iAR estimation of the ϕ parameter with observational times from Uniform(0,1) distribution. (∗) Theoretical SD is under regular times and without additive noise.

	n	Parameter	Estimate	Bias	SD	Theoretical SD (∗)	RMSE
SIMEX	200	0.990	0.967	−0.023	0.041	0.010	0.047
LSE	200	0.990	0.942	−0.048	0.043	0.010	0.065
ML	200	0.990	0.942	−0.048	0.039	0.010	0.062
SIMEX	500	0.990	0.991	0.001	0.028	0.006	0.028
LSE	500	0.990	0.971	−0.019	0.020	0.006	0.028
ML	500	0.990	0.971	−0.019	0.019	0.006	0.027

complex scenario with Uniform times, SIMEX consistently provides the most accurate and stable estimates, highlighting its reliability in challenging data conditions. Note that, across scenarios, SIMEX reduces bias by approximately 30 per cent–90 per cent compared to LSE and ML, with the maximum reduction occurring at uniform times for a sample size of $n = 500$, while RMSE is reduced by up to 60 per cent in the experiment with additive noise and sample size $n = 500$.

To enhance the discussion of our Monte Carlo study results, we included in Fig. 2 three density plots that illustrate the distributional behaviour of the three estimation methods (SIMEX, LSE, and ML) for simulated iAR process of size $n = 200$ considering $regular$, $Gamma$, and $uniform$ time gaps, respectively. These plots visually emphasize the comparative performance by showing the concentration and spread of the parameter estimates across simulations. For SIMEX, the density is notably sharper and more centred around the true parameter value, highlighting its reduced bias and higher accuracy. These density plots corroborate the findings from our simulation experiments, further demonstrating SIMEX's robustness and reliability in finite-sample settings.

5 ESTIMATION OF THE DAMPING TIME-SCALE IN AGNS FROM ZTF

In the introduction, we established the relationship between the damping time-scale τ, or relaxation parameter and the autocorrelation parameter ϕ (equation 1). Note that this equation implies the

Downloaded from https://academic.oup.com/mnras/article/540/1/521/8129690 by guest on 22 May 2025

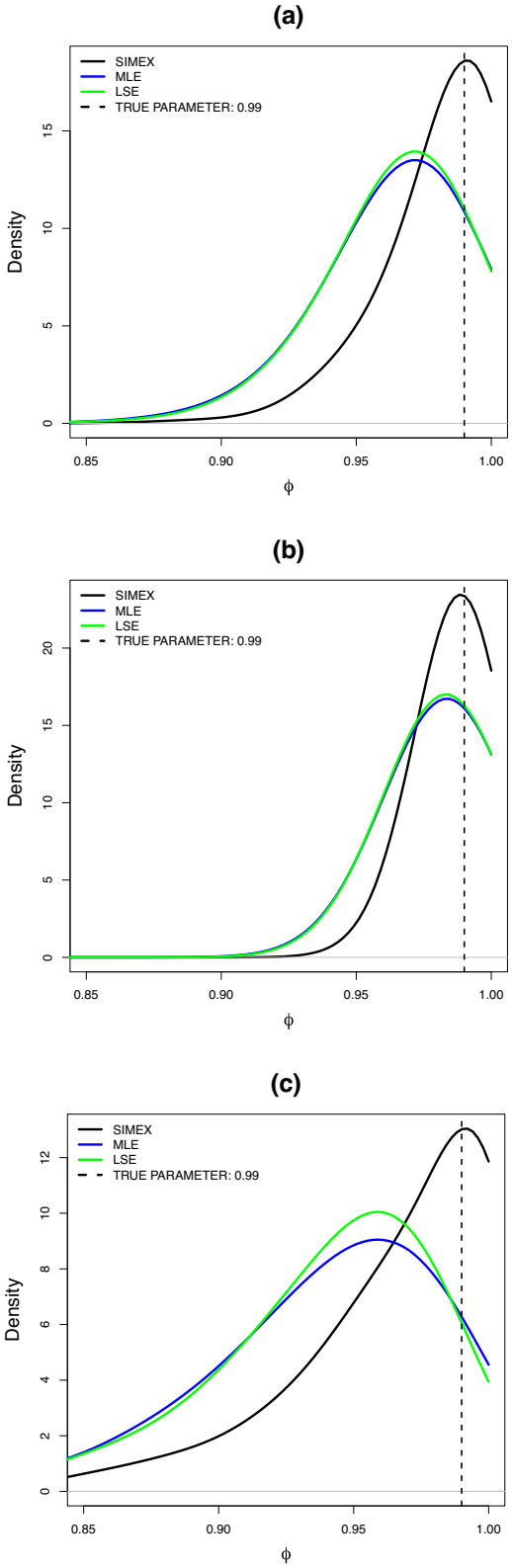


Figure 2. Density of the SIMEX, ML, and LSE estimators for the autoregressive paramater ϕ for a sample size $n = 200$. Figure (a) shows the density of the estimators for the regular AR model of order 1. Figure (b) shows the density of the estimators for the iAR model simulated with a Gamma(3,1) distribution for the observational times. Figure (c) shows the density of the estimators for the iAR model simulated with a Uniform(0,1) distribution for the observational times. The dashed vertical line represents the true value of the parameter ($\phi = 0.99$).

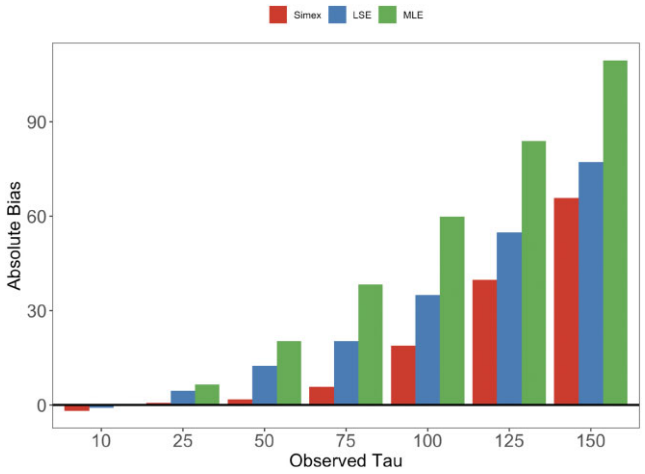


Figure 3. Mean of the absolute bias in the estimation of the parameter τ performed by the MLE, LSE, and SIMEX methods for the iAR model in simulated AGN light curves. The blue, green, and red bar denote the mean estimations obtained by the MLE, LSE, and SIMEX methods, respectively.

following,

$$\frac{\mathrm{d}\tau}{\tau} = \tau \frac{\mathrm{d}\phi}{\phi}. \tag{10}$$

This last equation shows that a small estimation error in ϕ, say 1 per cent, will translate to a huge estimation error in the time relaxation parameter. For example if $\tau = 100$, then the estimation error of this parameter becomes 100 per cent.

In a first experiment, we assess the estimation of the parameter τ by the iAR model using the MLE, LSE, and SIMEX methods in simulated AGN light curves generated by assuming a DRW model with a given τ. Synthetic light curves were simulated following a similar approach as in Sánchez-Sáez et al. (2018), i.e. adopting the exact observational times from an example long-duration, high cadence optical light curve of this reference. Specifically, this time sampling considers 175 observed epochs over a period of 1659 d. Using this sampling, we generated $N = 100$ synthetic AGN light curves for values of τ from 10 to 150. As in the Monte Carlo experiments, we use as the parameter set of the SIMEX estimation method, linear and equally weighted fits, $N = 30$ and $\delta = 1/70$. Fig. 3 shows the mean bias in the estimations of the autoregressive parameter for the simulated AGN light curves across different values of τ. Note in this experiment that, although the bias increases for all methods as τ becomes larger, the bias is consistently lower for the SIMEX method compared to the conventional methods (MLE and LSE). Notably, for the SIMEX method, the bias remains below 10 per cent relative to the true value up to a τ of 75. Specifically, at $\tau = 150$, the SIMEX method demonstrates a 40 per cent reduction in bias compared to the MLE method.

To complement the previous analysis, we present an additional experiment to evaluate the accuracy in estimating the damping time-scale τ in terms of the normalized damping time-scale $\rho = \tau/\Delta T$, where ΔT corresponds to the total length of the light curve in years (in this case, $\Delta T = 1659/365 \approx 4.54$). This formulation allows for a direct comparison with those in Kozłowski (2017) and Suberlak, Ivezić & MacLeod (2021), where the recovery of τ becomes inaccurate when ρ exceeds approximately 0.1. Using the same 100 synthetic AGN light curves described in the previous experiment,

Downloaded from https://academic.oup.com/mnras/article/540/1/521/8129690 by guest on 22 May 2025

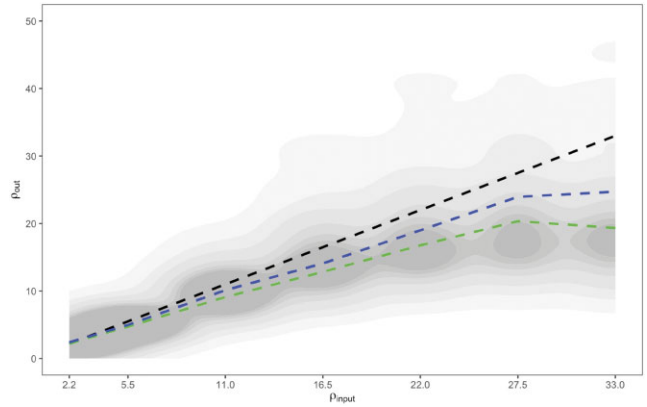


Figure 4. Distribution of the estimated normalized damping time-scale $\rho = \tau/\Delta T$ for simulated AGN light curves, using the SIMEX method applied to the iAR model. For each of the 100 light curves, between 30 per cent and 100 per cent of the epochs were randomly sampled to evaluate the robustness of the estimation. The shaded regions, from darkest to lightest, represent the 50 per cent, 60 per cent, 70 per cent, 80 per cent, 90 per cent, 95 per cent, and 99 per cent confidence intervals. The dashed blue line indicates the mean of the estimations, while the dashed green line corresponds to the median.

we generated new data sets by randomly sampling between 30 per cent and 90 per cent of the observations in each light curve. For each subsampled version, we estimated ρ using the SIMEX method applied to the iAR model. Fig. 4 displays the distribution of estimated ρ values across different subsampling levels and for the full sample. In contrast to the aforementioned studies, our results show that the SIMEX method maintains good accuracy for higher values of rho, with deviations from the true value occurring for rho values above 16.5.

5.1 ZTF light curves example

To assess the performance of the proposed estimation methodology on real data, we analysed a random sample of light curves from stochastic objects, including AGN, quasi-stellar objects (QSO), and Blazars, observed by the ZTF survey (Bellm, et al. 2019). These light curve data were processed by the Automatic Learning for the Rapid Classification of Events (ALeRCE) broker (Förster et al. 2021). The data set consisted of 3116 light curves, each with at least 30 measurements observed in the g band. Among these, 1609 correspond to AGN objects, 220 to Blazars, and 1049 to QSOs.

For these light curves, the damping time-scale parameter τ is unknown. To determine which estimation method provides the most accurate estimate of τ, we evaluate the fit of the iAR model derived from each method estimate. As goodness-of-fit measure of the iAR model we use the mean squared error (MSE) defined as,

$$\text{MSE} = \frac{1}{n}\sum_{j=1}^{n}\left(y_{t_j} - \hat{y_{t_j}}\right)^2.$$

For each light curve, the iAR model parameters were estimated using three methods: ML, LSE, and the SIMEX estimation method proposed in this work. For the SIMEX method, four parameter settings were used, reflecting the scenarios with (Scenario 1) and without measurement error (Scenario 2). In all settings, the number of repetitions for the estimation procedure was fixed at $B = 30$. To summarize:

Table 5. Pearson correlation between the estimations of the ϕ parameter obtained with the SIMEX, MLE, and LSE methods in light curves that are not near-root examples.

Methods	Correlations
SIMEX and MLE	0.988
SIMEX and LSE	0.257
MLE and LSE	0.296

(i) **Setting 1:** under Scenario 1 (no measurement error), and assuming an equally weighted quadratic fit over $N = 15$ points with $\delta = 1/80$.

(ii) **Setting 2:** under Scenario 1, and assuming an equally weighted quadratic fit over $N = 15$ points with $\delta = 1/110$.

(iii) **Setting 3:** under Scenario 2 (with measurement error), and assuming an equally weighted linear fit over $N = 15$ points with $\delta = 0.05$ and measurement error variance $\sigma_\nu = 0.1$.

(iv) **Setting 4:** under Scenario 2, and assuming an equally weighted quadratic fit over $N = 15$ points with $\delta = 0.1$ and measurement error variance $\sigma_\nu = 0.1$.

Among these four parameter settings, the one that minimized the MSE was selected for each light curve. Based on this goodness-of-fit criterion, Setting 3 was the most frequently chosen, accounting for 35.5 per cent (1106) of the examples. Among the remaining light curves, Setting 1 was selected in 24.9 per cent (777) of the cases, Setting 2 in 19.4 per cent (603) of the cases, and Setting 4 in 20.2 per cent (630) of the cases.

Based on the SIMEX estimation setting selected for each light curve, 343 examples were identified with a near-root situation, characterized by ϕ estimated values greater than 0.99, or equivalently, τ estimates exceeding 99.5. In contrast, using the standard MLE method, 264 light curves were identified with near-root estimations.

For light curves where the estimation of ϕ is not close to the unit root, it is expected that the estimation methods should yield similar results. However, this holds true only for the SIMEX and MLE methods. As shown in Table 5, the correlation between these two methods for non-near-root light curves is exceptionally high at 0.988. In contrast, the LSE method shows weak correlations with both SIMEX and MLE. Considering this result and the previously observed limitations of the LSE method in simulation experiments, we focus exclusively on MLE and SIMEX to identify light curves with near-root estimations.

There are 365 light curves with estimates close to the unit root in at least one of the two methods under consideration, either MLE or SIMEX. Focusing on these cases, we observe that the SIMEX method achieves the best fit, as determined by the criterion of the lowest MSE, in 318 of these examples (87.1 per cent). In contrast, the MLE method produces a lower MSE in 47 cases (12.9 per cent). Furthermore, looking at the parameters estimators for ϕ and τ, we notice that the distribution of these estimates tends to concentrate at higher values when the SIMEX method is applied, as can be seen in Fig. 5. For example in terms of τ, the third quartile of the SIMEX estimates is 288.98, substantially higher than the 220.01 for MLE, reinforcing the tendency of SIMEX to produce larger τ estimates compared to the standard maximum likelihood method.

When focusing on particular examples, we can observe light curves where the SIMEX method significantly improves the fit to the data. For instance, Figs 6(a)–(d) show the fit of the iAR model using both the SIMEX and MLE estimation methods for the AGN light curve from alert object 'ZTF18acbzncb' observed in the g band by ZTF survey. For this light curve, the τ parameter estimated by

Downloaded from https://academic.oup.com/mnras/article/540/1/521/8129690 by guest on 22 May 2025

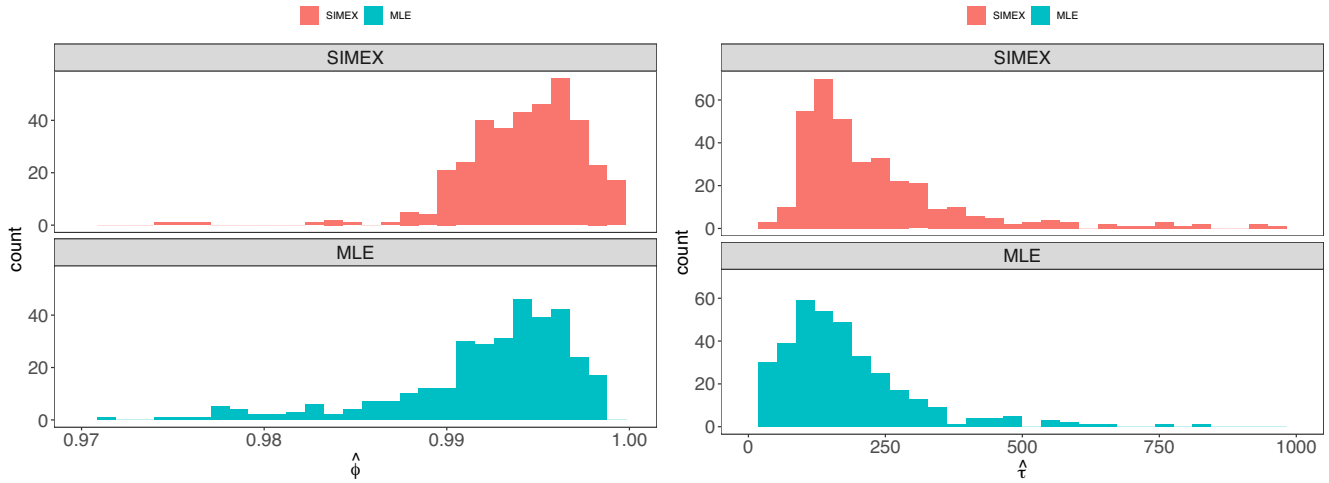


Figure 5. Distribution of the estimated values of the parameters ϕ and τ using both SIMEX and MLE estimation methods in 365 near-root light curves observed in the g band of the ZTF survey.

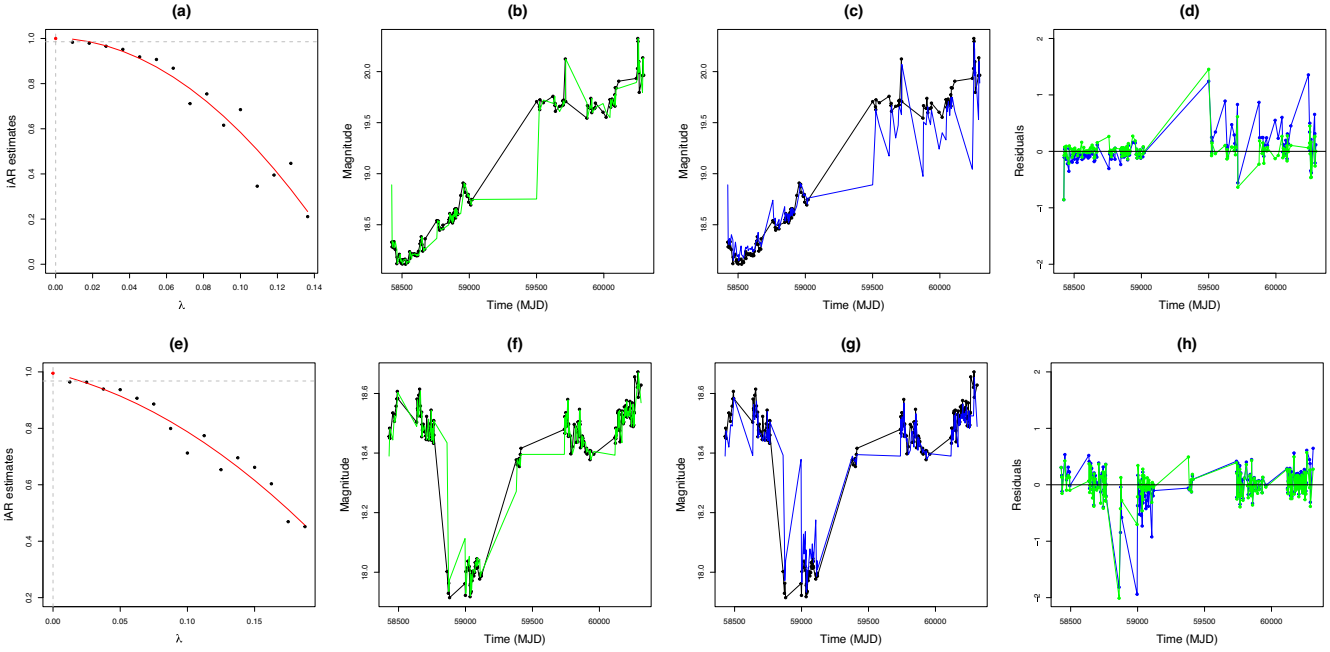


Figure 6. Two examples of AGN light curves observed by ZTF survey fitted by the iAR model using both SIMEX and MLE estimation procedures. Figures (a) and (e) illustrate the SIMEX estimation of the ϕ parameter of the iAR model in both examples. Figures (b) and (f) show the fit of the iAR model using the SIMEX estimate of the autocorrelation parameter and Figures (c) and (g) show the fit of the iAR model using the MLE estimate of this parameter. Finally, Figures (d) and (h) show the residuals of the iAR model fits for both estimation methods. In these figures, black line is the observed light curve, green and blue lines are the fitted (residuals) values obtained by the SIMEX and MLE estimation procedures, respectively.

MLE is 82.83, while using the SIMEX method with parameters according to the Setting 2 mentioned previously yields a τ estimate of 368.24. The difference obtained in the estimation of this parameter implies a reduction in the MSE obtained from the SIMEX estimation (0.033) compared to the MLE estimation (0.056). The improvement in the fit of the SIMEX estimation is most evident in the second half of the light curve, where Fig. 6(c) shows that the fit of the MLE estimation is below the true values. This is corroborated in Fig. 6(d) where the residuals obtained with both estimation methods are compared. We further applied the Box–Ljung test with lag 15 to the standardized residuals obtained by the SIMEX method, obtaining a P value of 0.1037. This result indicates no significant evidence of residual autocorrelation, supporting the goodness of fit of the SIMEX estimation.

A second example is presented in Figs 6(e)–(h). This QSO light curve was also observed in the g band by ZTF survey from alert object 'ZTF18aceivsg'. As in the previous example, the SIMEX estimation provides a higher estimate for the τ parameter compared to the MLE estimation. Specifically, the τ estimate obtained with the SIMEX method was 213.18, while the MLE estimate was 30.50. Additionally, the light curve fit obtained using the SIMEX estimation outperformed that from the MLE estimation, with MSE values of

Downloaded from https://academic.oup.com/mnras/article/540/1/521/8129690 by guest on 22 May 2025

0.059 and 0.094, respectively. As in the previous example, the Box–Ljung test applied to the standardized residuals (lag 15) yielded a high P value (P value $= 0.9134$), suggesting no significant autocorrelation and supporting the adequacy of the SIMEX-based fit.

6 CONCLUSIONS

In this article, we introduced a novel SIMEX algorithm for estimating the autocorrelation parameter in autoregressive models of order one. We considered regular and irregular autoregressive models with and without additive noise. Through extensive simulations, we demonstrated that standard estimation methods, such as MLE and LSE, can yield biased estimators of the autocorrelation parameter when the true value is close to one. This is a statistical problem known as the unit root problem, which refers to the fact that the parameter is close to the boundary of the parameter space. In this case, the boundary corresponds to a value of one. When $\phi = 1$, the process is non-stationary and unstable. As a consequence, it is not possible to obtain a reliable estimator. When ϕ is close to 1, computer algorithms have difficulty distinguishing the true value of the parameter due to accuracy constraints. This root proximity problem complicates estimation, as the true value may be extremely close to the upper bound of the parameter space (0, 1) and, in many cases, the algorithm fails to produce estimates, making it difficult to assess the variance of the estimator.

Monte Carlo experiments performed in this work show that the SIMEX estimation methodology improves substantially the estimation of the autoregressive model under both regular and irregular sampling, especially under challenging conditions such as near-unit root scenarios or the presence of additive noise. Monte Carlo simulations revealed substantial reductions in bias and improvements in estimation precision compared to standard methods like MLE and LSE. However, above certain thresholds, all method begin to fail by large factors.

We further show that the novel SIMEX algorithm offers an alternative for reliable estimation of the damping time-scale parameter in both simulated and real AGN light curves observed in the ZTF survey. In both empirical applications, the proposed SIMEX method with time dependence methodology works well, validating the practical utility of this approach, yielding better model fits and lower MSE compared to the other estimation techniques. These results highlight the SIMEX method ability to provide more reliable estimates of the damping time-scale parameter (τ), which is essential for understanding AGN physical properties.

The SIMEX method proposed in this work aims to improve the variability estimation of AGN compared to the maximum likelihood estimator. In the context of autoregressive models for irregular time series, there are few works that use the maximum likelihood estimation (for example the iAR model). Generally, the variability estimation of AGN is performed using DRW models based on Markov Chain Monte Carlo (MCMC) methods with nested sampling. It is important to note that the iAR and DRW models are equivalent (under Gaussianity). Therefore, a less biased estimator of the ϕ parameter of the iAR model also allows to have a better estimator of the τ parameter and consequently to improve the quality of fit of the DRW model. MCMC methods can recover the true values of the DRW parameters accurately (Hu, Cai & Wang 2024) but can be computationally expensive. The alternative that we propose in which we apply the SIMEX methodology on the iAR model (and not directly on the DRW model) is computational less expensive. The performance of the MCMC regarding bias is of out the scope of this work.

As mentioned in Section 1, the estimation of the relaxation time from these models (iAR or DRW) assumes that this parameter is constant over time. Exploring this limitation is outside the scope of this work. However, it can be approached from a sequential estimation of the parameter ϕ of the iAR model (see for example Elorrieta et al. 2023). In a future work, an adaptation of the SIMEX estimator for time-varying autoregressive parameter will be considered.

ACKNOWLEDGEMENTS

The authors acknowledge support from the ANID - Millennium Science Initiative, AIM23-0001 awarded to the Millennium Institute of Astrophysics MAS (FE, WP, SE, FEB). FEB acknowledges support from ANID-Chile BASAL CATA FB210003, FONDECYT Regular 1241005.

DATA AVAILABILITY

The ZTF light curves data set is available from the ALeRCE website: http://alerce.science/. The specific light curves used in this work will be shared on reasonable request to the corresponding author. The R code to implement the SIMEX method is available at: https://github.com/felipeelorrieta.

REFERENCES

Amico M., Van Keilegom I., 2018, Annu. Rev. Stat. Appl., 5, 311
Arévalo P., Churazov E., Lira P., Sánchez-Sáez P., Bernal S., Hernández-García L., López-Navas E., Patel P., 2024, A&A, 684, A133
Barreto-Souza W., Chan N. H., 2024, Scand. J, Stat., 51, 402
Bellm Eric C., Kulkarni Shrinivas R., Graham Matthew J., Dekany Richard, Smith Roger M., Riddle Reed, Masci Frank J., Helou George, et al. 2019. Publications of the Astronomical Society of the Pacific, 131, 018002
Brockwell P. J., Davis R. A., 1991, Time series: theory and methods, 2nd edn., Springer Science + Business Media, LLC, New York
Burke C. J. et al., 2021, Science, 373, 789
Carroll R. J., Maca J. D., Ruppert D., 1999, Biometrika, 86, 541
Carroll R. J., Ruppert D., Stefanski L. A., Crainiceanu C. M., 2006, Measurement Error in Nonlinear Models: A Modern Perspective, 2nd edn Chapman and Hall/CRC, Boca Raton, FL
Carstens H., Xia X., Yadavalli S., 2017, Appl. Energy, 188, 563
Cook J. R., Stefanski L. A., 1994, J. Am. Stat. Assoc., 89, 1314
Elorrieta F., Osses L., Cáceres M., Eyheramendy S., Palma W., 2023, in Valenzuela O., Rojas F., Herrera L. J., Pomares H., Rojas I.eds, Theory and Applications of Time Series Analysis. Springer Nature Switzerland, Cham, p. 3
Escamilla-Rivera C., Levi Said J., Mifsud J., 2021, J. Cosmol. Astropart. Phys., 2021, 016
Eyheramendy S., Elorrieta F., Palma W., 2018, MNRAS, 481, 4311
Feigelson E. D., Babu G. J., 2012, Modern statistical methods for astronomy: with R applications. Cambridge Univ. Press, Cambridge
Fernández-Hernández L. M., Montiel A., Rodríguez-Meza M. A., 2019, MNRAS, 488, 5127
Förster F. et al., 2021, AJ, 161, 242
Freeman P. E., Newman J. A., Lee A. B., Richards J. W., Schafer C. M., 2009, MNRAS, 398, 2012
Graham M. J., Djorgovski S. G., Drake A. J., Mahabal A. A., Chang M., Stern D., Donalek C., Glikman E., 2014, MNRAS, 439, 703
Guolo A., 2008, Stat. Methods Med. Res., 17, 555
Hu X.-F., Cai Z.-Y., Wang J.-X., 2024, ApJ, 961, 5
Juvela M., Montillaud J., Ysard N., Lunttila T., 2013, A&A, 556, A63
Kasliwal V. P., Vogeley M. S., Richards G. T., 2017, MNRAS, 470, 3027
Kelly B. C., Bechtold J., Siemiginowska A., 2009, ApJ, 698, 895
Kozłowski S. et al., 2010, ApJ, 708, 927

Downloaded from https://academic.oup.com/mnras/article/540/1/521/8129690 by guest on 22 May 2025

Kozłowski S., 2016, MNRAS, 459, 2787
Kozłowski S., 2017, A&A, 597, A128
Küchenhoff H., Mwalili S. M., Lesaffre E., 2006, Biometrics, 62, 85
Li X., Huang X., 2019, Can. J. Stat., 47, 262
Li Y., Lin X., 2003, J. Am. Stat. Assoc., 98, 191
MacLeod C. L. et al., 2010, ApJ, 721, 1014
Minniti D. et al., 2010, New Astron., 15, 433
Mushotzy R. F., Winter L. M., McIntosh D. H., Tueller J., 2008, ApJ, 684, L65
Piesse A., Feng C. X., 2008, in JSM Proceedings, Survery Research Methods Section, American Statistical Association, Alexandria, VA
Rana A., Jain D., Mahajan S., Mukherjee A., Holanda R., 2017, J. Cosmol. Astropart. Phys., 2017, 010
Raymond J. Carroll, Helmut Küchenhoff F. L., Stefanski L. A., 1996, J. Am. Stat. Assoc., 91, 242
Sánchez-Sáez P. et al., 2021, AJ, 161, 141
Sánchez-Sáez P., Lira P., Mejía-Restrepo J., Ho L. C., Arévalo P., Kim M., Cartier R., Coppi P., 2018, ApJ, 864, 87
Stefanski L. A., Cook J. R., 1995, J. Am. Stat. Assoc., 90, 1247
Suberlak K. L., Ivezić Z., MacLeod C., 2021, ApJ, 907, 96
Wang X.-F., Fan Z., Wang B., 2010, Comput. Stat. Data Anal., 54, 25

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/540/1/521/8129690 by guest on 22 May 2025

© 2025 The Author(s).
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.