algorithms

MDPI

Article

Subsampling Algorithms for Irregularly Spaced Autoregressive Models

Jiaqi Liu †, Ziyang Wang †, HaiYing Wang and Nalini Ravishanker *

Department of Statistics, University of Connecticut, Storrs, CT 06269, USA; jiaqi.3.liu@uconn.edu (J.L.); ziyang.wang@uconn.edu (Z.W.); haiying.wang@uconn.edu (H.W.)

* Correspondence: nalini.ravishanker@uconn.edu

† These authors contributed equally to this work.

Abstract: With the exponential growth of data across diverse fields, applying conventional statistical methods directly to large-scale datasets has become computationally infeasible. To overcome this challenge, subsampling algorithms are widely used to perform statistical analyses on smaller, more manageable subsets of the data. The effectiveness of these methods depends on their ability to identify and select data points that improve the estimation efficiency according to some optimality criteria. While much of the existing research has focused on subsampling techniques for independent data, there is considerable potential for developing methods tailored to dependent data, particularly in time-dependent contexts. In this study, we extend subsampling techniques to irregularly spaced time series data which are modeled by irregularly spaced autoregressive models. We present frameworks for various subsampling approaches, including optimal subsampling under A-optimality, information-based optimal subdata selection, and sequential thinning on streaming data. These methods use A-optimality or D-optimality criteria to assess the usefulness of each data point and prioritize the inclusion of the most informative ones. We then assess the performance of these subsampling methods using numerical simulations, providing insights into their suitability and effectiveness for handling irregularly spaced long time series. Numerical results show that our algorithms have promising performance. Their estimation efficiency can be ten times as high as that of the uniform sampling estimator. They also significantly reduce the computational time and can be up to forty times faster than the full-data estimator.

Keywords: subsampling; irregularly spaced autoregressive model; time series; big data

Citation: Liu, J.; Wang, Z.; Wang, H.; Ravishanker, N. Subsampling Algorithms for Irregularly Spaced Autoregressive Models. *Algorithms* **2024**, *17*, 524. https://doi.org/10.3390/a17110524

Academic Editor: Matthias Templ

Received: 25 September 2024
Revised: 6 November 2024
Accepted: 11 November 2024
Published: 15 November 2024

1. Introduction

Time series analysis is useful in many application domains for understanding patterns over time and for forecasting into the future. Most time series methods have been developed for regularly spaced time series, where the gaps between two consecutive time points are fixed and equal. Regular time series include hourly, daily, monthly, or annually recorded time series. Recently, there has been increasing interest in developing methods for analyzing and forecasting irregularly spaced time series, where the gaps between subsequent observations are not the same. Such irregular time series are observed in diverse fields such as astronomy [1], climatology [2], finance [3], etc. For instance, intra-day transaction-level data in finance consist of prices of a financial asset recorded at each trade within a trading day, resulting in irregular time intervals (in seconds, say) between consecutive trades. An example in real estate consists of the time series of sale prices of houses, where the discrete time gaps (in days or weeks) between subsequent sale dates are typically nonconstant.

Consider an irregularly spaced time series $y_j,\ j = 0, \ldots, m$ consisting of observations at discrete times t_j, where $t_0 < t_1 < \cdots < t_m$. The gaps between consecutive time points, denoted by $d_j = t_j - t_{j-1},\ j = 1, \ldots, m$, are not constant.

To model irregularly spaced time series with discrete gaps, Nagaraja et al. [4] proposed the stationary gap time autoregressive (Gap AR(1)) model:

$$y_0 = \sigma\sqrt{\frac{1}{1-\phi^2}}\varepsilon_0, \quad y_j = \phi^{d_j} y_{j-1} + \sigma\sqrt{\frac{1-\phi^{2d_j}}{1-\phi^2}}\varepsilon_j \text{ for } j = 1, \ldots, m, \tag{1}$$

where ε_j is the error term with the standard normal distribution, i.e., $\varepsilon_j \sim \mathcal{N}(0,1)$, $\phi \in (-1,1)$ is the autoregressive parameter, and $\sigma > 0$ is a scale parameter. They used (1) to model and forecast house prices and then constructed a house price index. Other models have also been constructed in the literature for irregularly spaced time series. Erdogan et al. [5] described a nonstationary irregularly spaced autoregressive (NIS-AR(1)) model and illustrated its use on astronomy data. Anantharaman et al. [6] described Bayesian modeling for an irregular stochastic volatility autoregressive conditional duration (IR-SV-ACD) model and used this for estimating and forecasting inter-transaction gaps and the volatility of financial log-returns.

In this paper, we refer to the Gap AR(1) model in (1) as the irregularly spaced AR(1) or the IS-AR(1) model. Estimation in this model can be handled by the method of maximum likelihood. While a model for irregularly spaced time series enables handling a wider range of temporal scenarios, the statistical analysis involves considerable computational complexity. Specifically, the task of capturing temporal structures by estimating the model parameters becomes more challenging due to the irregular gaps. One potential solution to address these computational challenges is to perform the computations of interest on *subsamples* that are selected from the full data.

Despite extensive recent research on subsampling techniques, subsampling methods tailored for irregularly spaced time series data are sparse. In this article, we fill this gap by proposing novel subsampling methods specifically designed for irregularly spaced time series data analyzed by the IS-AR(1) model. In comparison to other subsampling methods, our proposed algorithms are founded upon the optimality criteria under classical statistical frameworks. It should be noted that the term "subsample" in time series analysis generally refers to a subset of data that are in the form of multiple series, blocks, or sequences, and the main objective of their analysis is to provide estimates of the variance of the full-data estimator; see Carlstein [7], Fukuchi [8], Politis [9]. By contrast, the subsampling methods discussed in this paper are used to provide estimates of model parameters based on subsamples from the full data in situations where obtaining such estimates from the full data is computationally prohibitive.

The remainder of this paper is organized as follows. Section 3 shows full-data estimates for parameters of the IS-AR(1) model. Subsampling methods for the IS-AR(1) model are described in Section 4. These methods include a random subsampling method based on A-optimality (in Section 4.1), information-based subdata selection (in Section 4.2), and a sequential thinning method (in Section 4.3). Section 5.1 illustrates the techniques using simulated data. Lastly, in Section 6, we summarize the key findings of our study.

2. Background

Section 2.1 gives a brief review of irregularly spaced time series, while Section 2.2 reviews subsampling methods from the recent literature.

2.1. Modeling Irregularly Spaced Time Series

Irregularly spaced (or unevenly spaced) time series occur in many domains including astronomy, biomedicine, climatology, ecology, environment, finance, geology, etc. For instance, high-frequency financial transactions typically occur at irregularly spaced time points within a trading day, as each trade is recorded. Within a trading day, the elapsed times (durations) between consecutive trades are not the same for any selected stock. These times also vary between different stocks. In any given time interval (say, one hour), transactions of a stock may occur rapidly, separated by short durations, or occur slowly with

longer durations. Since methods that are used for modeling and forecasting regular time series [10] are not useful for analyzing irregular time series, modified approaches have been developed.

A recent approach for analyzing irregularly spaced time series is the gap time autoregressive (AR) model in (1), which was discussed in [4] for modeling house prices. In this model, the higher the value of d_j, the less useful y_{j-1} becomes for explaining and predicting y_j. Erdogan et al. [5] described AR type models for stationary and nonstationary irregular time series. Similar models were used in [11] and [1] to model and forecast future values of irregularly spaced astronomical light curves from variable stars, for which Elorrieta et al. [12] proposed a bivariate model.

Ghysels and Jasiak [13] proposed the autoregressive conditional duration generalized autoregressive conditionally heteroscedastic (ACD-GARCH) model for analyzing irregularly spaced financial returns, employing the computationally cumbersome Generalized Method of Moments (GMM) approach for estimating parameters. Meddahi et al. [14] proposed a GARCH-type model for irregularly spaced time series by discretizing a continuous time stochastic volatility process, thereby combining the advantages of the ACD-GARCH model [13] and the ACD model [15]. Maller et al. [16] described a continuous version of the GARCH (i.e., the COGARCH) model for irregularly spaced time series, while [17] proposed a multivariate local-level model with score-driven covariance matrices for intra-day log prices, treating the asynchronicity as a missing data problem. Recently, [6] extended the gap time modeling idea of [4] to construct useful time series models to understand volatility patterns in irregularly spaced financial time series, considering the gaps as random variables. An alternate stochastic volatility model treating the gaps as fixed constants was discussed in [18].

Most methods for analyzing long irregularly spaced time series can be computationally challenging. Subsampling methods can help us obtain estimates in a computationally feasible way.

2.2. Subsampling Methods

Constructing parameter estimates based on subsamples from the full data is a popular technique to speed up computations. While the simplest approach of uniform sampling may not be effective for extracting useful information from a large dataset, optimized subsampling methods do provide a better trade-off between estimation efficiency and computational efficiency. Such methods have attracted significant attention in recent years because they are designed to (a) give higher preference to more informative data points and (b) be subject to less information loss. Typical practices include stochastic subsampling and deterministic subdata selection.

Stochastic subsampling methods are successful because they specify inclusion probabilities that allow more informative data points to have a higher chance of being included in the subsample. In early attempts, Drineas et al. [19] advocated the use of normalized statistical leverage scores as subsampling probabilities in least squares estimation problems, while Yang et al. [20] showed that using the normalized square roots of the statistical leverage scores provides a tighter error bound. Ma et al. [21] examined the statistical properties of estimators resulting from subsampling methods based on statistical leverage scores, they and termed it *algorithmic leverage*. Xie et al. [22] applied the statistical leverage scores for subsampling under a vector autoregressive model.

This approach has been used in several statistical modeling frameworks. Zhu [23] proposed a subsampling method using the gradients of the objective function for linear models. Wang et al. [24] proposed optimal subsampling probabilities under A-optimality and L-optimality criteria for logistic regression. Teng et al. [25] examined the asymptotic properties of subsampling estimators for generalized linear models under unbounded design. The optimal subsampling framework has been extended to other modeling scenarios such as multiclass logistic regression, generalized linear models, and quantile regressions.

While the aforementioned studies are based on sampling with replacement (in which the same data point may be included more than once in the subsample), further research derived optimal subsampling probabilities and a distributed sampling strategy under Poisson sampling (sampling without replacement). Wang et al. [26] further showed that Poisson sampling can be superior to sampling with replacement in terms of estimation efficiency. Poisson sampling for irregular time series analysis is described in Section 4.1.

Deterministic subdata selection uses a particular criterion to determine a subsample without involving additional randomness. Wang et al. [27] introduced a novel method, termed *information-based optimal subsampling* (IBOSS), designed to select data points with extreme values to approximate the D-optimality criterion in designed experiments. Pronzato and Wang [28] later proposed an online selection strategy that leverages directional derivatives to decide whether to include a data point in a subsample, aiming to achieve optimality according to a given criterion. This approach processes only the data points encountered up to the current time, determining whether the current point should be included in the subsample. As a result, it is especially well suited for applications involving streaming data.

Despite the rapid recent developments mentioned above, applications of the stochastic subsampling and deterministic subdata selection methods remain unexplored for irregularly spaced time series.

3. Full-Data Estimation for the IS-AR(1) Model

Given a time series of length m from the IS-AR(1) model in (1), we present the full-data maximum likelihood estimates (MLEs) of the model parameters. Since m is assumed to be large, we can obtain the MLEs of the unknown parameters $\boldsymbol{\theta} = (\phi, \sigma)'$ by maximizing the *conditional* log-likelihood function (which ignores the marginal distribution of the initial data point y_0 at time t_0). Up to a normalizing constant, this has the form

$$\begin{aligned} \mathrm{M}_m(\boldsymbol{\theta}) &= \frac{1}{m}\sum_{j=1}^{m} \ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) \\ &= \frac{1}{m}\sum_{j=1}^{m}\left(-\log\sigma - \frac{1}{2}\log\frac{1-\phi^{2d_j}}{1-\phi^2} - \frac{(y_j - \phi^{d_j} y_{j-1})^2(1-\phi^2)}{2\sigma^2(1-\phi^{2d_j})}\right). \end{aligned} \tag{2}$$

The MLE $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathrm{M}_m(\boldsymbol{\theta})$ is the maximizer of (2) and must be solved numerically since an analytical solution is infeasible. Finding the MLE is challenging due to (a) the domain restriction (i.e., $-1 < \phi < 1$) and (b) the nonconcavity of the objective function. These challenges often cause convergence issues for the the classical Newton–Raphson algorithm.

Since ϕ is a correlation coefficient constrained to lie between -1 and 1, an unconstrained gradient-based optimization method may not work well. To remedy this, we use Fisher's z-transformation to map ϕ from $(-1, 1)$ onto $\mathbb{R}$ so that an unconstrained optimizing algorithm can be used on the transformed parameter. Although the choice of the transformation is not unique and other transformations may be used, Fisher's z-transformation provides a well-understood estimator of the correlation coefficient, especially when the error term ε_j follows a normal distribution [29]. Specifically, let

$$z = \operatorname{atanh}(\phi) = \frac{1}{2}\log\left(\frac{1+\phi}{1-\phi}\right),$$

so that

$$\phi = \tanh(z) = \frac{\exp(2z)-1}{\exp(2z)+1}. \tag{3}$$

We have

$$\frac{\partial\phi}{\partial z} = \operatorname{sech}^2(z) = 1-\phi^2,$$

and

$$\frac{\partial^2 \phi}{\partial z^2} = -2\tanh(z)\mathrm{sech}^2(z) = -2\phi(1-\phi^2),$$

where $\tanh(\cdot)$ and $\mathrm{sech}(\cdot)$ are, respectively, the hyperbolic tangent and secant functions. Below, we show the use of the coordinate ascent algorithm to maximize (2) in terms of z; we then insert the estimate $\hat{z}$ into (3) to obtain the estimate of ϕ.

We denote the gradient vector and Hessian matrix of the per-observation log-likelihood $\ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j)$ in (2) as follows:

$$\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) = \frac{\partial \ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j)}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \dot{\ell}_1(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) \\ \dot{\ell}_2(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) \end{bmatrix}, \tag{4}$$

and

$$\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) = \frac{\partial^2 \ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{bmatrix} \ddot{\ell}_{11}(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) & \ddot{\ell}_{12}(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) \\ \ddot{\ell}_{12}(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) & \ddot{\ell}_{22}(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) \end{bmatrix}. \tag{5}$$

The first and second derivatives of $\ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j)$ with respect to z are, respectively, denoted by

$$\dot{\ell}_{z1}(\boldsymbol{\theta}) = (1-\phi^2)\dot{\ell}_1(\boldsymbol{\theta}) \text{ and } \ddot{\ell}_{z11}(\boldsymbol{\theta}) = (1-\phi^2)^2\ddot{\ell}_{11}(\boldsymbol{\theta}) - 2(1-\phi^2)\phi\dot{\ell}_1(\boldsymbol{\theta}).$$

We describe the coordinate ascent algorithm. Suppose $\phi^{(k)} = \tanh(z^{(k)})$ is obtained in the k-th step of the algorithm. We find the value of σ given $\phi^{(k)}$ by solving $\dot{\ell}_2(\phi^{(k)}, \sigma) = 0$, which gives

$$\sigma^{(k)} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} \frac{(y_j - \phi^{d_j} y_{j-1})^2 (1-{\phi^{(k)}}^2)}{1-{\phi^{(k)}}^{2d_j}}}.$$

Given $\sigma^{(k)}$, we find the value of z in the $(k+1)$-th step by implementing the Newton–Raphson algorithm, i.e., computing

$$z^{(k+1),l+1} = z^{(k+1),l} - \frac{\sum_{j=1}^{m} \dot{\ell}_{z1}(\phi^{(k+1),l}, \sigma^{(k)}; y_j, y_{j-1}, d_j)}{\sum_{j=1}^{m} \ddot{\ell}_{z11}(\phi^{(k+1),l}, \sigma^{(k)}; y_j, y_{j-1}, d_j)}$$

for $l = 0, 1, 2, \ldots$ until convergence, where $\phi^{(k+1),0} = \phi^{(k)}$ and $\phi^{(k+1),l} = \tanh(z^{(k+1),l})$. We alternate this updating of the values of σ and z until the values become stable.

We summarize the aforementioned estimation procedure in Algorithm 1. Here, $\phi^{(0)}$ is an initial value, while ϵ_{NR} and ϵ_{CA} are error tolerances to determine convergence.

For completeness, we provide below detailed expressions of the gradient vector and Hessian matrix of the per-observation log-likelihood:

$$\begin{aligned}
\dot{\ell}_1(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) &= \frac{d_j \phi^{2d_j-1}}{1-\phi^{2d_j}} - \frac{\phi}{1-\phi^2} + \frac{\phi e_j^2}{\sigma^2(1-\phi^{2d_j})} \\
&\quad + \frac{1-\phi^2}{\sigma^2(1-\phi^{2d_j})^2} d_j e_j (\phi^{d_j-1} y_{j-1} - \phi^{2d_j-1} y_j) \\
\dot{\ell}_2(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) &= -\frac{1}{\sigma} + \frac{e_j^2(1-\phi^2)}{\sigma^3(1-\phi^{2d_j})}, \\
\ddot{\ell}_{11}(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) &= \frac{2d_j^2 \phi^{2d_j-2}}{(1-\phi^{2d_j})^2} - \frac{d_j \phi^{2d_j-2}}{1-\phi^{2d_j}} - \frac{1+\phi^2}{(1-\phi^2)^2} \\
&\quad + \frac{e_j^2 - 2d_j e_j \phi^{d_j} y_{j-1}}{\sigma^2(1-\phi^{2d_j})} + \frac{2d_j e_j^2 \phi^{2d_j}}{\sigma^2(1-\phi^{2d_j})^2}
\end{aligned}$$

$$+\frac{(1-\phi^2)d_j^2\phi^{2d_j-2}(2y_je_j-y_{j-1}^2+y_j^2)}{\sigma^2(1-\phi^{2d_j})^2}$$
$$+\frac{d_je_j(y_{j-1}-\phi^{d_j}y_j)[(d_j-1)\phi^{d_j-2}-(d_j+1)\phi^{d_j}]}{\sigma^2(1-\phi^{2d_j})^2}$$
$$-\frac{4(1-\phi^2)d_j^2e_j^2\phi^{2d_j-2}}{\sigma^2(1-\phi^{2d_j})^3},$$

$$\ddot{\ell}_{12}(\boldsymbol{\theta};y_j,y_{j-1},d_j)=\frac{2d_je_j\phi^{d_j-1}(\phi^{d_j}y_j-y_{j-1})(1-\phi^2)}{\sigma^3(1-\phi^{2d_j})^2}-\frac{2e_j^2\phi}{\sigma^3(1-\phi^{2d_j})},$$

$$\ddot{\ell}_{22}(\boldsymbol{\theta};y_j,y_{j-1},d_j)=\frac{1}{\sigma^2}-\frac{3e_j^2(1-\phi^2)}{\sigma^4(1-\phi^{2d_j})},$$

where, $e_j=y_j-\phi^{d_j}y_{j-1}$.

Algorithm 1 Numerical optimization algorithm.

1. $z\leftarrow \operatorname{atanh}(\phi^{(0)})$
2. $\sigma\leftarrow\sqrt{\frac{1}{m}\sum_{i=1}^{m}\frac{(y_j-\phi^{d_j}y_{j-1})^2(1-\phi^2)}{1-\phi^{2d_j}}}$
3. $z_o\leftarrow$ Inf
4. **while** $|z-z_o|>\epsilon_{\mathrm{CA}}$
   1. $z_o\leftarrow z$
   2. $\Delta_{\mathrm{NR}}\leftarrow$ Inf
   3. **while** $|\Delta_{\mathrm{NR}}|>\epsilon_{\mathrm{NR}}$
      1. $\Delta_{\mathrm{NR}}\leftarrow\frac{\sum_{j=1}^{m}(1-\phi^2)\dot{\ell}_1(\phi,\sigma;y_j,y_{j-1},d_j)}{\sum_{j=1}^{m}(1-\phi^2)^2\ddot{\ell}_{11}(\phi,\sigma;y_j,y_{j-1},d_j)-2(1-\phi^2)\phi\ddot{\ell}_1(\phi,\sigma;y_j,y_{j-1},d_j)}$
      2. $z\leftarrow z-\Delta_{\mathrm{NR}}$

      end
   4. $\phi\leftarrow\tanh(z)$
   5. $\sigma\leftarrow\sqrt{\frac{1}{m}\sum_{i=1}^{m}\frac{(y_j-\phi^{d_j}y_{j-1})^2(1-\phi^2)}{1-\phi^{2d_j}}}$

   end

4. Subsampling Methods for the IS-AR(1) Model

If the size of the full data m is moderate, we can easily obtain the full-data MLE by maximizing (2) using Algorithm 1. However, it is computationally challenging to implement this algorithm when m is very large. In this case, we may resort to subsampling strategies, using a small fraction of the full data to obtain estimators. We propose three distinct methods: (1) optimal subsampling under A-optimality (**opt**), (2) information-based optimal subdata selection (**iboss**), and (3) sequential thinning (**thin**). We describe these procedures and propose practical algorithms in the following subsections.

4.1. Optimal Subsampling Under A-Optimality

Optimal subsampling is a stochastic subsampling strategy satisfying a certain optimality property. The inclusion of subsampled data points is determined by carefully designed probabilities in order to meet the optimality criterion. In this section, we implement optimal Poisson subsampling, which is suitable for time series and avoids the need to simultaneously access the full data. In comparison to sampling with replacement,

Poisson sampling cannot sample the same data point more than once. Since repeatedly sampled data points do not contribute any new information, Poisson sampling preserves more distinct information from the full data and is therefore more efficient than sampling with replacement. See Wang et al. [26] for a theoretical justification of the superiority of Poisson sampling.

Let η_j be the sampling probability and δ_j be the indicator variable for the inclusion of the j-th data point in the subsample. With u_j randomly sampled from the uniform distribution $U(0,1)$, we set $\delta_j = 1$ when $u_j \leq \eta_j$ and $\delta_j = 0$ otherwise. The actual subsample size from Poisson sampling, which is denoted by $r^* = \sum_{j=1}^{m} \delta_j$, is random. The expected subsample size is $r = \sum_{j=1}^{m} \eta_j$. We obtain the subsample estimator $\breve{\boldsymbol{\theta}}$ by maximizing the following target function:

$$\mathrm{M}_r^{(\mathrm{opt})}(\boldsymbol{\theta}) = \frac{1}{m}\sum_{j=1}^{m} \delta_j \frac{\ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j)}{\eta_j}. \tag{6}$$

The efficacy of $\breve{\boldsymbol{\theta}}$ depends on the subsampling probabilities $\{\eta_j\}_{j=1}^{m}$. We obtain the A-optimal subsampling probabilities which depend on the unknown true parameter and minimize the asymptotic mean squared error of $\breve{\boldsymbol{\theta}}$ [24,26]. In practice, this unknown parameter could be replaced by a pilot estimate $\boldsymbol{\theta}^{(\mathrm{plt})}$ obtained from a pilot subsample of size r_0. The A-optimal subsampling probabilities with the pilot estimate $\boldsymbol{\theta}^{(\mathrm{plt})}$ take the general form

$$\hat{\eta}_j^{(\mathrm{opt})} = \left\| \kappa \left\{ \ddot{\mathbf{M}}_{r_0}^{(\mathrm{plt})}(\boldsymbol{\theta}^{(\mathrm{plt})}) \right\}^{-1} \dot{\ell}(\boldsymbol{\theta}^{(\mathrm{plt})}; y_j, y_{j-1}, d_j) \right\| \wedge 1, \text{ for } j = 1, \ldots, m, \tag{7}$$

where κ is a parameter to ensure that the expected sample size is set around r and prevent it from being too small, the notation $a \wedge b = \min(a, b)$,

$$\ddot{\mathbf{M}}_{r_0}^{(\mathrm{plt})}(\boldsymbol{\theta}) = \frac{1}{r_0}\sum_{j=1}^{m} \delta_j^{(\mathrm{plt})} \ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}; y_j, y_{j-1}, d_j) \tag{8}$$

is the average Hessian matrix with the pilot subsample, and $\delta_j^{(\mathrm{plt})}$ is the indicator variable for inclusion of the j-th data point in the pilot sample. While the exact value of κ requires additional computation, it can be approximated by

$$\kappa_a = \frac{r}{\sum_{j=1}^{m} \left\| \left\{ \ddot{\mathbf{M}}_{r_0}^{(\mathrm{plt})}(\boldsymbol{\theta}^{(\mathrm{plt})}) \right\}^{-1} \dot{\ell}(\boldsymbol{\theta}^{(\mathrm{plt})}; y_j, y_{j-1}, d_j) \right\|}$$

when the sampling rate r/m is small. This is usually the case in practice.

The approximated subsampling probabilities $\hat{\eta}_j^{(\mathrm{opt})}$ in (7) are subject to additional disturbance from the pilot estimation, and small values of $\hat{\eta}_j^{(\mathrm{opt})}$ may inflate the asymptotic variance of the resulting estimator. To address this problem, we mix $\hat{\eta}_j^{(\mathrm{opt})}$ with the uniform subsampling probabilities to prevent the sampling probabilities from getting too small. Specifically, we use

$$(1-\alpha_1)\eta_j^{(\mathrm{opt})} + \alpha_1 \frac{r}{m}, \text{ for } j = 1, \ldots, m,$$

where $\alpha_1 \in (0,1)$ is a tuning parameter. The final subsample estimator $\tilde{\boldsymbol{\theta}}$ is obtained by combining the pilot estimator $\boldsymbol{\theta}^{(\mathrm{plt})}$ and the optimal subsample estimator $\breve{\boldsymbol{\theta}}$:

$$\tilde{\boldsymbol{\theta}} = \left(r_0 \ddot{\mathbf{M}}_{r_0}^{(\mathrm{plt})}(\boldsymbol{\theta}^{(\mathrm{plt})}) + r^* \ddot{\mathbf{M}}_r^{(\mathrm{opt})}(\breve{\boldsymbol{\theta}}) \right)^{-1} \left(r_0 \ddot{\mathbf{M}}_{r_0}^{(\mathrm{plt})}(\boldsymbol{\theta}^{(\mathrm{plt})}) \boldsymbol{\theta}^{(\mathrm{plt})} + r^* \ddot{\mathbf{M}}_r^{(\mathrm{opt})}(\breve{\boldsymbol{\theta}}) \breve{\boldsymbol{\theta}} \right), \tag{9}$$

where

$$\ddot{\mathbf{M}}_r^{(\mathrm{opt})}(\check{\boldsymbol{\theta}}) = \frac{1}{m}\sum_{j=1}^{m}\delta_j \frac{\ddot{\ell}(\check{\boldsymbol{\theta}}; y_j, y_{j-1}, d_j)}{\hat{\eta}_{\alpha_1,j}^{(\mathrm{opt})}}.$$

We summarize the A-optimal subsampling strategy in Algorithm 2.

Algorithm 2 Poisson sampling under A-optimality.

- **Step 1—Pilot Estimates and Preparations:**
  1. Take a simple random sample of size r_0 by sampling without replacement and set indicators $\{\delta_j^0\}_{j=1}^m$.
  2. $\boldsymbol{\theta}^{(\mathrm{plt})} \leftarrow \arg\max_{\boldsymbol{\theta}} \sum_{j=1}^m \delta_j^0 \ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j)$
  3. $\mathbf{M}^{(\mathrm{plt})} \leftarrow r_0^{-1} \sum_{j=1}^m \delta_j^0 \ddot{\ell}(\boldsymbol{\theta}^{(\mathrm{plt})}; y_j, y_{j-1}, d_j)$
  4. Calculate κ_a.
- **Step 2—Subsampling:**

  for i=1 **to** m
  1. $u \leftarrow \mathrm{Unif}(0,1)$
  2. $\eta_j \leftarrow (1-\alpha)\kappa_a \|{\mathbf{M}^{(\mathrm{plt})}}^{-1} \dot{\ell}(\boldsymbol{\theta}^{(\mathrm{plt})}; y_j, y_{j-1}, d_j)\| + \alpha(r/m)$
  3. **if** $u < \eta_j$ **do**
     $\delta_j \leftarrow 1$
     else
     $\delta_j \leftarrow 0$
     end

  end
- **Step 3—Estimation:**
  1. $\check{\boldsymbol{\theta}} \leftarrow \arg\max_{\boldsymbol{\theta}} \sum_{j=1}^m \delta_j \ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j)/(\eta_j \wedge 1)$
  2. Combine $\boldsymbol{\theta}^{(\mathrm{plt})}$ and $\check{\boldsymbol{\theta}}$ using (9) to obtain $\tilde{\boldsymbol{\theta}}$.

4.2. Information-Based Optimal Subdata Selection

The **iboss** method is a deterministic selection approach to obtain a subsample. Its basic motivation is to maximize the determinant of the Fisher information matrix conditioned on the covariates. Under the model specified in (2), the conditional Fisher information for the observed data at time t_j is

$$\begin{aligned}\boldsymbol{I}_j(\boldsymbol{\theta}|y_{j-1}, d_j) &= -\mathbb{E}\left(\begin{bmatrix}\left(\frac{\partial \ell_j}{\partial \phi}\right)^2 & \frac{\partial \ell_j}{\partial \phi}\frac{\partial \ell_j}{\partial \sigma} \\ \frac{\partial \ell_j}{\partial \phi}\frac{\partial \ell_j}{\partial \sigma} & \left(\frac{\partial \ell_j}{\partial \sigma}\right)^2\end{bmatrix}\middle| y_{j-1}, d_j\right) \\ &= \begin{bmatrix} 2c_j(\boldsymbol{\theta})^2 + \frac{d_j^2\phi^{2d_j-2}y_{j-1}^2(1-\phi^2)}{\sigma^2(1-\phi^{2d_j})} & -\frac{2}{\sigma}c_j(\boldsymbol{\theta}) \\ -\frac{2}{\sigma}c_j(\boldsymbol{\theta}) & \frac{2}{\sigma^2}\end{bmatrix},\end{aligned} \tag{10}$$

where

$$c_j(\boldsymbol{\theta}) = \frac{d_j\phi^{2d_j-1}}{1-\phi^{2d_j}} - \frac{\phi}{1-\phi^2}.$$

The information matrix $\boldsymbol{I}_j(\boldsymbol{\theta}|y_{j-1}, d_j)$ has full rank. We select data points with the r largest values of $\det\{\boldsymbol{I}_j(\boldsymbol{\theta}|y_{j-1}, d_j)\}$, $j = 1, \ldots, m$. This can be undertaken efficiently by established methods like some partition-based selection algorithms [30]. We implement the procedure via the following steps:

1. Obtain a pilot estimate $\boldsymbol{\theta}^{(\mathrm{plt})}$ as in Section 4.1.
2. Calculate $\det(\boldsymbol{I}_j(\boldsymbol{\theta}^{(\mathrm{plt})}|y_{t_{j-1}}, d_j))$ for $j = 1, \ldots, m$.
3. Take the r data points with the largest $\det(\boldsymbol{I}_j(\boldsymbol{\theta}^{(\mathrm{plt})}|y_{j-1}, d_j))$ using the Quickselect algorithm. Denote their inclusion indicators as $\delta_j^{(\mathrm{iboss})}$s.
4. Obtain $\breve{\boldsymbol{\theta}}^{(\mathrm{iboss})}$ by maximizing the following target function:

$$\mathrm{M}_r^{(\mathrm{iboss})}(\boldsymbol{\theta}) = \frac{1}{r}\sum_{j=1}^{m} \delta_j^{(\mathrm{iboss})} \ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j).$$

5. Obtain the final subsample estimator $\tilde{\boldsymbol{\theta}}^{(\mathrm{iboss})}$ by

$$\begin{aligned} \tilde{\boldsymbol{\theta}}^{(\mathrm{iboss})} = & \left(r_0 \ddot{\mathbf{M}}_{r_0}^{(\mathrm{plt})}(\breve{\boldsymbol{\theta}}^{(\mathrm{iboss})}) + r \ddot{\mathbf{M}}_r^{(\mathrm{iboss})}(\breve{\boldsymbol{\theta}}^{(\mathrm{iboss})}) \right)^{-1} \\ & \times \left(r_0 \ddot{\mathbf{M}}_{r_0}^{(\mathrm{plt})}(\breve{\boldsymbol{\theta}}^{(\mathrm{iboss})}) \times \boldsymbol{\theta}^{(\mathrm{plt})} + r \ddot{\mathbf{M}}_r^{(\mathrm{iboss})}(\breve{\boldsymbol{\theta}}^{(\mathrm{iboss})}) \times \breve{\boldsymbol{\theta}}^{(\mathrm{iboss})} \right), \end{aligned} \tag{11}$$

where $\ddot{\mathbf{M}}_{r_0}^{(\mathrm{plt})}(\boldsymbol{\theta})$ is defined in (8) and

$$\ddot{\mathbf{M}}_r^{(\mathrm{iboss})}(\breve{\boldsymbol{\theta}}^{(\mathrm{iboss})}) = \frac{1}{r}\sum_{j=1}^{m} \delta_j \ddot{\boldsymbol{\ell}}(\breve{\boldsymbol{\theta}}^{(\mathrm{iboss})}; y_j, y_{j-1}, d_j).$$

We present the **iboss** method in Algorithm 3.

Algorithm 3 Information-based optimal subdata selection.

- **Step 1—Pilot Estimates and Preparations:**
  1. Take a simple random sample of size r_0 by sampling without replacement and set indicators $\{\delta_j^0\}_{j=1}^m$.
  2. $\boldsymbol{\theta}^{(\mathrm{plt})} \leftarrow \arg\max_{\boldsymbol{\theta}} \sum_{j=1}^m \delta_j^0 \ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j)$
- **Step 2—Subdata Selection:**
  1. **for** $j = 1$ **to** m
     $I_j \leftarrow \det(\boldsymbol{I}_j(\boldsymbol{\theta}^{(\mathrm{plt})}|y_{j-1}, d_j))$
     end
  2. $I_{(r)} \leftarrow r$-th largest I_j for $j = 1, \ldots, m$
  3. **for** $j = 1$ **to** m
     if $I_j \geq I_{(r)}$ **then** $\delta_j^{(\mathrm{iboss})} \leftarrow 1$ **else** $\delta_j^{(\mathrm{iboss})} \leftarrow 1$
     end
- **Step 3—Estimation:**
  1. $\breve{\boldsymbol{\theta}} \leftarrow \arg\max_{\boldsymbol{\theta}} \sum_{j=1}^m \delta_j^{(\mathrm{iboss})} \ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j)$
  2. Combine $\boldsymbol{\theta}^{(\mathrm{plt})}$ and $\breve{\boldsymbol{\theta}}$ using (11) to obtain $\tilde{\boldsymbol{\theta}}$.

4.3. Sequential Thinning on Streaming Data

The **thin** method proposed in Pronzato and Wang [28] is another deterministic subdata selection approach. Unlike **iboss**, which requires simultaneous access to the full data, **thin** only uses the data points that are already included in the subsample to determine whether the next data point should be included in the subsample or not. This online decision nature of the **thin** algorithm makes it suitable for time series data. We present the main idea of the **thin** method based on D-optimality below.

Let the average information matrix of a subsample indexed by $S = \{\delta_j^S\}_{j=1}^m$ be

$$\mathcal{I}_S(\boldsymbol{\theta}) = \frac{1}{r_S}\sum_{j=1}^{m}\delta_j^S \mathbf{I}_j(\boldsymbol{\theta}|y_{j-1}, d_j), \tag{12}$$

where $r_S = \sum_{j=1}^m \delta_j^S$ and $\mathbf{I}_j(\boldsymbol{\theta}|y_{j-1}, d_j)$ is defined in (10). The contribution of a data point, say $(y_{j'}, y_{j'-1}, d_{j'})$, to the average information matrix, if included in the subsample, can be measured by the directional derivative, which is defined as

$$\zeta_{\mathcal{I}_S(\boldsymbol{\theta}),j'} = \mathrm{tr}\Big(\mathcal{I}_S(\boldsymbol{\theta})^{-1}\boldsymbol{I}_{j'}(\boldsymbol{\theta}|y_{j'-1}, d_{j'})\Big) - 2. \tag{13}$$

The **thin** method aims to include $(y_{j'}, y_{j'-1}, d_{j'})$ in the subsample if $\zeta_{\mathcal{I}_S(\boldsymbol{\theta}),j'}$ is large enough. This is motivated by the key result in optimal experimental design theory that the optimal design consists of design points with the largest directional derivatives in the design space [31,32]. In the context of subsampling, this means that we need to find the subsample with the r largest directional derivatives under the unknown optimal average information matrix. In order to achieve this in an online manner, we need to sequentially estimate the upper r/m quantile for the distribution of the directional derivatives.

Unlike the linear models discussed in Pronzato and Wang [28] for which the information matrix is completely known, the information matrix for our model depends on the unknown parameter $\boldsymbol{\theta}$. We therefore need a pilot step to obtain a pilot estimator. We present the outline of the **thin** method for our problem in the following steps:

1. Reserve the first r_0 data points up to time t_{r_0} as a pilot sample, and use it to obtain a pilot estimate $\boldsymbol{\theta}^{(\mathrm{plt})}$ as in Section 4.1.
2. Calculate $\{\zeta_{\mathcal{I}_{r_0}(\boldsymbol{\theta}^{(\mathrm{plt})}),j}\}_{j=1}^{r_0}$ and obtain its sample upper α-quantile $\hat{C}_{r_0}$, where $\alpha = r/(m-r_0)$ and $\mathcal{I}_{r_0}(\boldsymbol{\theta}^{(\mathrm{plt})})$ is the average information matrix for the pilot sample.
3. For $j = r_0+1, \ldots, m$, let $\mathcal{I}_{j-1}(\boldsymbol{\theta}^{(\mathrm{plt})})$ be the average information matrix and $\hat{C}_{j-1}$ be the estimated quantile from the subsample collected up to time t_{j-1}. If $\zeta_{\mathcal{I}_{j-1}(\boldsymbol{\theta}^{(\mathrm{plt})}),j} > \hat{C}_{j-1}$, include the data point (y_j, y_{j-1}, d_j) in the subsample and calculate the updated $\mathcal{I}_j(\boldsymbol{\theta}^{(\mathrm{plt})})$; otherwise, $\mathcal{I}_j(\boldsymbol{\theta}^{(\mathrm{plt})}) = \mathcal{I}_{j-1}(\boldsymbol{\theta}^{(\mathrm{plt})})$. Calculate the updated $\hat{C}_j$.
4. Let $\delta_j^{(\mathrm{thin})}$ be the indicators for the **thin** subsample collected in Step 3. Obtain the subsample estimator $\breve{\boldsymbol{\theta}}^{(\mathrm{thin})}$ by maximizing

$$\mathrm{M}_r^{(\mathrm{thin})}(\boldsymbol{\theta}) = \frac{1}{r}\sum_{j=r_0+1}^{m}\delta_j^{(\mathrm{thin})}\ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j).$$

5. Obtain the final subsample estimator $\tilde{\boldsymbol{\theta}}^{(\mathrm{thin})}$ by

$$\begin{aligned}\tilde{\boldsymbol{\theta}}^{(\mathrm{thin})} =& \Big(r_0\ddot{\mathbf{M}}_{r_0}^{(\mathrm{plt})}(\breve{\boldsymbol{\theta}}^{(\mathrm{thin})}) + r\ddot{\mathbf{M}}_r^{(\mathrm{thin})}(\breve{\boldsymbol{\theta}}^{(\mathrm{thin})})\Big)^{-1} \\ &\times \Big(r_0\ddot{\mathbf{M}}_{r_0}^{(\mathrm{plt})}(\breve{\boldsymbol{\theta}}^{(\mathrm{thin})})\times\boldsymbol{\theta}^{(\mathrm{plt})} + r\ddot{\mathbf{M}}_r^{(\mathrm{thin})}(\breve{\boldsymbol{\theta}}^{(\mathrm{thin})})\times\breve{\boldsymbol{\theta}}^{(\mathrm{thin})}\Big),\end{aligned} \tag{14}$$

where $\ddot{\mathbf{M}}_{r_0}^{(\mathrm{plt})}(\boldsymbol{\theta})$ is defined in (8) and

$$\ddot{\mathbf{M}}_r^{(\mathrm{thin})}(\breve{\boldsymbol{\theta}}) = \frac{1}{r}\sum_{j=1}^{m}\delta_j^{(\mathrm{thin})}\ddot{\ell}(\breve{\boldsymbol{\theta}}; y_j, y_{j-1}, d_j).$$

A detailed algorithm implementing **thin** is detailed in Algorithm 4.

Algorithm 4 Sequential thinning under D-optimality.

- **Set Parameters:**
  1. Set $r_0 \geq 2$, $q \in (1/2, 1)$, $\gamma \in (0, q-1/2)$.
  2. $\alpha \leftarrow r/(m-r_0)$
  3. Set $\epsilon_1 > 0$ such that $\epsilon_1 \ll \alpha$
- **Step 1—Pilot Estimates and Preparations:**
  1. $\theta^{(\mathrm{plt})} \leftarrow \arg\max_\theta \sum_{i=1}^{r_0} \ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j)$
  2. $r_* \leftarrow 0$
  3. $\delta_1, \ldots, \delta_{r_0} \leftarrow 0$
  4. **for** $j = 1$ **to** r_0
     $\left| I_j \leftarrow \mathbf{I}_j(\boldsymbol{\theta}^{(\mathrm{plt})} | y_{j-1}, d_j) \right.$
     end
  5. $\boldsymbol{\mathcal{I}}_* \leftarrow r_0^{-1} \sum_{j=1}^{r_0} I_j$
  6. **for** $j = 1$ **to** r_0
     $\left| \zeta_j \leftarrow \mathrm{tr}(\boldsymbol{\mathcal{I}}_*{}^{-1} I_j) - 2 \right.$
     end
  7. $\zeta_{(1)}, \ldots, \zeta_{(r_0)} \leftarrow \mathtt{sort}(\zeta_1, \ldots, \zeta_{r_0})$ where $\zeta_{(1)} \leq \cdots \leq \zeta_{(r_0)}$
  8. $r_0^+ \leftarrow \lceil (1-\alpha/2) r_0 \rceil$, $r_0^- \leftarrow \lfloor (1-3\alpha/2) r_0 \rfloor \vee 1$
  9. $\hat{C}_* \leftarrow \zeta_{(\lceil (1-\alpha) r_0 \rceil)}$
  10. $\beta_0 \leftarrow r_0/(r_0^+ - r_0^-)$
  11. $h \leftarrow \zeta_{(r_0^+)} - \zeta_{(r_0^-)}$
  12. $h_* \leftarrow h/r_0^\gamma$
  13. $\hat{f}_* \leftarrow \left[\sum_{j=1}^{r_0} \mathbf{1}_{|\zeta_j - \hat{C}_*| \leq h_*} \right] / (2 r_0 h_*)$
- **Step 2—Sequential Thinning:**
  for $k = r_0 + 1$ **to** m
  1. **if** $r - r_* > m - k$ **do**
     $\left| \delta_j \leftarrow 1 \right.$ **for** $j = k, \ldots, m$
     $\left| \right.$ **break**
     else if $r_* \geq r$ **do**
     $\left| \delta_j \leftarrow 0 \right.$ **for** $j = k, \ldots, m$
     $\left| \right.$ **break**
     end
  2. $I_* \leftarrow \mathbf{I}_j(\boldsymbol{\theta}^{(\mathrm{plt})} | y_{j-1}, d_j)$
  3. $\zeta_* \leftarrow \mathrm{tr}(\boldsymbol{\mathcal{I}}_*^{-1} I_*)$
  4. **if** $(r_* + r_0)/k \leq \epsilon_1$ **or** $\zeta_* \geq \hat{C}_*$ **do**
     $\left| \delta_k \leftarrow 1 \right.$
     else
     $\left| \delta_k \leftarrow 0 \right.$
     end
  5. $\hat{C}_o \leftarrow \hat{C}_*$
  6. $\hat{C}_* \leftarrow \hat{C}_* + \frac{(\hat{f}_*)^{-1} \bigwedge \beta_0 (k-1)^\gamma}{k^q} (\mathbf{1}_{\zeta_* \geq \hat{C}_*} - \alpha)$

Algorithm 4 *Cont.*

7. $h_* \leftarrow h/k^{\gamma}$
8. $\hat{f}_* \leftarrow \hat{f}_* + k^{-q}[(2h_*)^{-1}\mathbf{1}_{\{|\zeta_* - \hat{C}_o| \le h_*\}} - \hat{f}_*]$
9. $r_* \leftarrow r_* + \delta_k$
10. $\boldsymbol{\mathcal{I}}_* \leftarrow \boldsymbol{\mathcal{I}}_* + \frac{\delta_k}{r_* + r_0}[I_* - \boldsymbol{\mathcal{I}}_*]$

end

- **Step 3—Estimate and Combine:**
  1. $\breve{\boldsymbol{\theta}} \leftarrow \arg\max_{\boldsymbol{\theta}} \sum_{j=r_0+1}^{m} \delta_j \ell(\boldsymbol{\theta}; y_j, y_{j-1}, d_j)$
  2. Combine $\theta^{(\mathrm{plt})}$ and $\breve{\boldsymbol{\theta}}$ by (14)

5. Numerical Results

We perform numerical simulation to evaluate the performance of the subsampling methods. We present the results on estimation efficiency in Section 5.1 and the results on computational efficiency in Section 5.2.

We consider the performances of the uniform subsample estimator (**unif**), the A-optimal subsample estimator (**opt**), the information-based optimal subdata selection estimator (**iboss**), and the sequential thinning estimator (**thin**). We set the full-data sample size as $m = 10^5$ and consider different values of the true parameters; we allow $\phi = 0.1, 0.5, 0.9$ and $\sigma = 1$, 10 and 20. The time gaps d_j are randomly sampled from the set $\{1, 2, 3, 4, 5\}$.

We consider different subsample sizes r = 5000, 10,000, 15,000, 20,000, 25,000, and 30,000, corresponding, respectively, to 5%, 10%, 15%, 20%, 25%, and 30% of the full dataset. We let $r_0 = 1000$. For **opt**, we set the mixing rate $\alpha_1 = 0.1$. To implement the **thin** subsampling procedure, we set the tuning parameters $q = 0.7$, $\gamma = 0.1$, and $\epsilon_1 = 0.001$. We also implement the uniform subsampling method with a subsample size $r + r_0$ as a benchmark for comparisons.

5.1. Estimation Efficiency

We use the empirical mean squared error (MSE) to measure the estimation efficiency of the subsample estimator with respect to the true parameter. We implement the subsampling methods discussed in Section 4 and repeat the simulation 1000 times to calculate the empirical MSE, which is defined as

$$\mathrm{MSE} = \frac{1}{1000}\sum_{s=1}^{1000}\left\|\tilde{\boldsymbol{\theta}}^{(s)} - \boldsymbol{\theta}\right\|^2 = \frac{1}{1000}\sum_{s=1}^{1000}[(\tilde{\phi}^{(s)} - \phi)^2 + (\tilde{\sigma}^{(s)} - \sigma)^2] = \mathrm{MSE}_{\phi} + \mathrm{MSE}_{\sigma}, \quad (15)$$

where $\tilde{\boldsymbol{\theta}}^{(s)} = (\tilde{\phi}^{(s)}, \tilde{\sigma}^{(s)})$ is the subsample estimate in the s-th repetition. Due to the nonconvex nature of the model, some methods failed to converge in a few repetitions. We drop the results in these repetitions when calculating the MSE. This does not affect the accuracy of the MSE because the nonconvergence rate is very low.

The results in terms of the empirical MSE are presented in Figure 1. It is seen that all three proposed subsampling methods outperform the uniform sampling in all the cases. Overall, the **opt** algorithm displays the best performance, especially when σ or ϕ is large. The deterministic methods, **iboss** and **thin**, have higher estimation efficiency when σ is small.

The varying performances in empirical MSE among the proposed subsampling methods are partly due to the different optimality criteria adopted. The A-optimality criterion used by **opt** seeks to minimize the trace of the asymptotic variance matrix, whereas the D-optimality criterion used by both **iboss** and **thin** focuses on minimizing the determinant of the asymptotic variance matrix. As a result, **opt** places greater emphasis on parameter components with larger variances, while **iboss** and **thin** distribute their focus more evenly across all parameters. When σ is small (e.g., $\sigma = 1$), the variance in estimating σ is low

across all subsampling methods, allowing **iboss** and **thin** to express their advantage in reducing the variance of estimating ϕ, thereby outperforming **opt**. However, when $\sigma = 20$, the variance in estimating σ becomes the dominant contributor to the empirical MSE. Since **iboss** and **thin** do not prioritize the estimation of σ, their performance weakens in this case. A similar reasoning applies to the differing performances across various values of ϕ. To better understand this, we further decompose the empirical MSEs for σ and ϕ to validate our explanation.

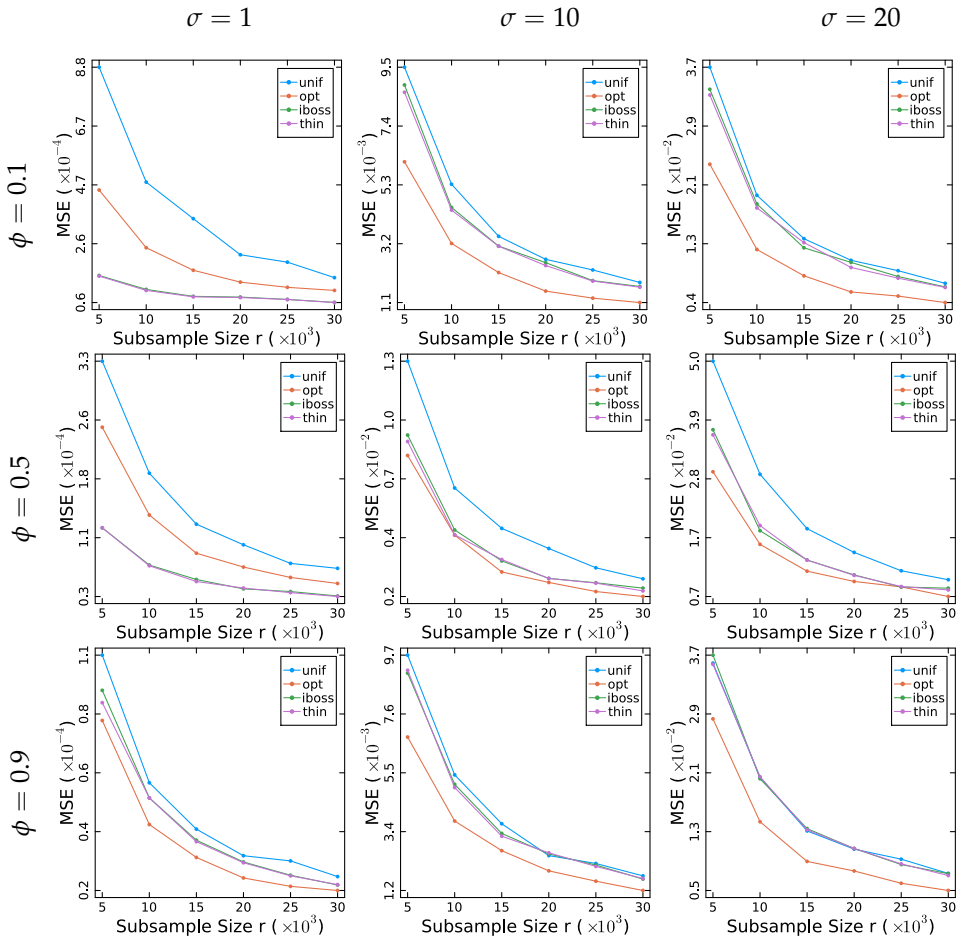

Figure 1. Comparison of MSEs for estimating $\boldsymbol{\theta}$ between different subsampling algorithms across varying subsample sizes r.

In order to further examine the performance of the subsampling methods on estimating different types of parameters, we plot MSE_ϕ for estimating ϕ in Figure 2 and MSE_σ for estimating σ in Figure 3. Irrespective of the values of ϕ and σ, **iboss** and **thin** outperform **opt** in estimating ϕ, while they fall short in estimating σ compared to **opt**.

We observe that the proposed subsampling methods provide greater benefit for smaller r. As r increases, the differences between the methods gradually diminish. This is to be expected, since all estimators converge to the full-data estimator when r gets closer to m. Therefore, with larger datasets, the advantage of using optimal subsampling estimators becomes more pronounced.

In the IS-AR(1) model, the predicted value at any given time point is directly influenced by the correlation coefficient ϕ. Therefore, more accurate estimates of ϕ lead to better prediction accuracy. As shown in Figure 2, the optimal subsampling methods provide exceptional performance in estimating ϕ, especially when using the **iboss** and **thin** approaches. Notably, with a subsample size of $r = 5000$, both **iboss** and **thin** outperform the uniform sampling method, which uses a significantly larger subsample size of r = 30,000 for estimating ϕ.

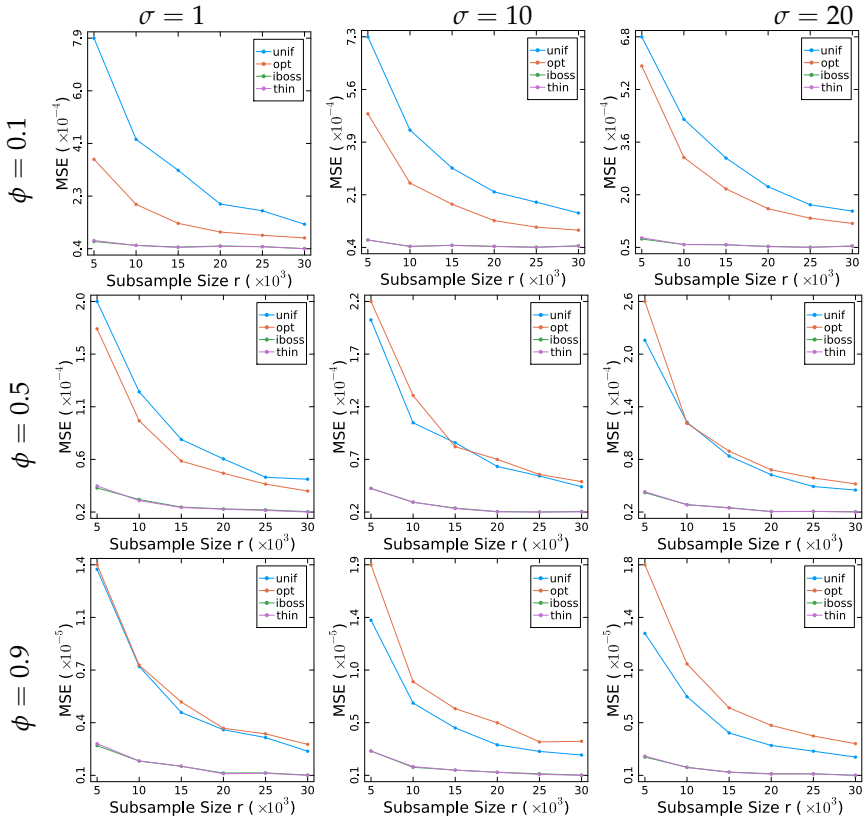

Figure 2. Comparison of MSE_ϕs for estimating ϕ between different subsampling algorithms across varying subsample sizes r.

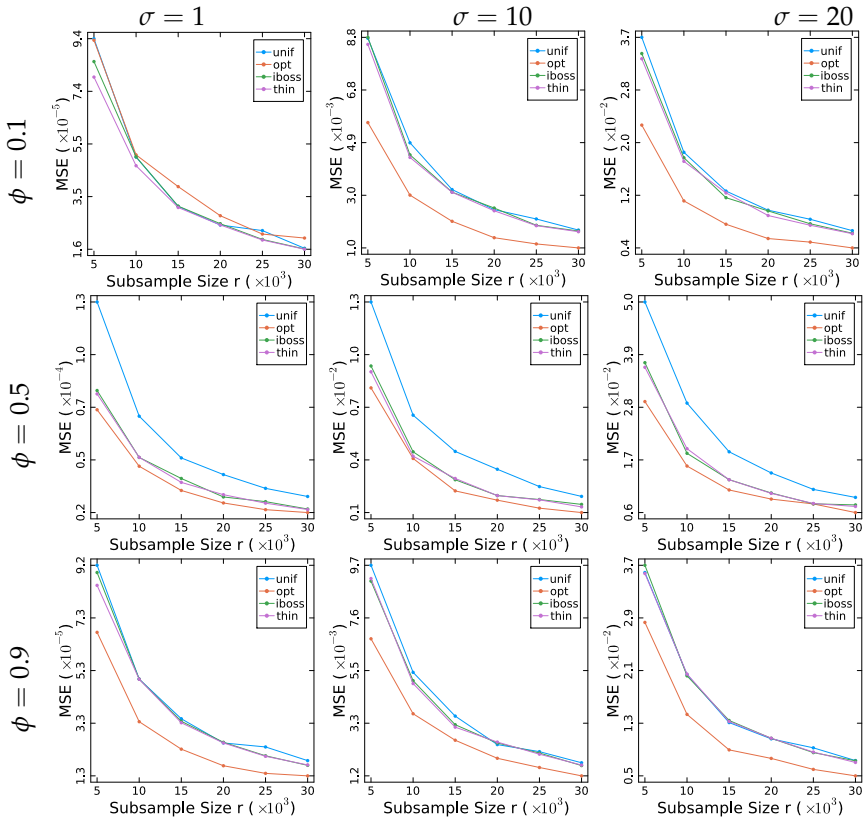

Figure 3. Comparison of MSE_σs for estimating σ between different subsampling algorithms across varying subsample sizes r.

5.2. Time Complexity

To evaluate the computational efficiency of the subsampling methods, we repeat the simulation 30 times and record the average computational times for each method. For comparison, we implement the full-data estimator (**full**) using the algorithm described in Section 3 as a benchmark.

We consider three different full-data sample sizes, $m = 10^4$, 10^5, and 10^6, and six subsample sizes $r = 500, 1000, 1500, 2000, 2500$, and 3000. Table 1 reports the average computational times in milliseconds for the case where $\phi = 0.5$ and $\sigma = 1$. The computation times for other values of ϕ and σ follow a similar pattern. These results are based on simulations using a Julia implementation run on an Apple MacBook Pro with an M1 Pro chip.

Table 1. Average estimation times (ms) over 30 repetitions across different estimation methods, with full-data sizes m and subsample sizes r under the setting $\phi = 0.5$ and $\sigma = 1$.

m	r	unif	opt	iboss	thin	full
10^4	500	0.744	1.46	0.999	2.17	17.2
	1000	1.48	2.07	1.34	2.87	17.8
	1500	2.14	2.91	1.67	3.33	17.8
	2000	3.0	3.51	2.01	3.88	17.4
	2500	3.38	4.12	2.51	4.22	17.5
	3000	4.15	4.29	2.96	4.63	17.9
10^5	500	0.781	3.1	4.34	10.6	141
	1000	1.52	3.7	4.3	11.0	147
	1500	2.3	4.45	4.46	11.6	147
	2000	3.03	4.84	4.67	12.0	147
	2500	3.48	5.69	4.67	12.6	147
	3000	4.42	6.71	5.16	13.1	150
10^6	500	1.25	19.5	30.5	79.6	1244
	1000	2.14	19.0	29.8	79.8	1207
	1500	3.15	19.9	30.4	80.6	1224
	2000	4.16	21.3	30.3	81.1	1165
	2500	5.01	21.4	31.7	82.1	1177
	3000	5.65	22.1	29.9	82.4	1274

The entries in the table show a substantial reduction in computation times by using the subsampling methods compared to the time taken for full-data estimation. Unlike the uniform subsampling method, which incurs almost no computational overhead during the subsampling process, optimal subsampling algorithms require additional computations to determine the inclusion probability for each data point. The numerical optimization using subsamples has a time complexity of $\mathcal{O}(r)$, whereas calculating the subsampling probabilities for nonuniform subsampling methods requires a cost of $\mathcal{O}(m)$.

Compared to full-data-based estimation, both **opt** and **iboss** demonstrate exceptional efficiency, reducing the computation time by a factor of $1/40$ on average when the sample size m is sufficiently large. Although not as fast as these two methods, **thin** still achieves significant computational savings, taking less than one-tenth of the time required for the full-data estimation. Additionally, an advantage of using **thin** is that subsample selection can be performed sequentially. Overall, all three proposed optimal subsampling methods for the IS-AR(1) model offer reliable strategies for reducing the computational burden in large-scale data applications.

6. Conclusions

In this paper, we investigated the technique of computationally feasible subsampling in the context of irregularly spaced time series data. We proposed practical algorithms for implementing the **opt**, **iboss**, and **thin** methods for the IS-AR(1) model. The numerical re-

sults demonstrated that the proposed subsampling methods outperform the naive uniform subsampling approach with improved estimation efficiency and show significant benefits in reducing the computation time compared with the full-data estimation.

While our work is currently focused on the IS-AR(1) model, it highlights the potential of optimal subsampling methods for time series data. In future research, these techniques can be extended to more complex models, such as IS-AR(p) for $p > 1$. Typically, as the number of parameters increases, optimal subsampling methods become even more effective in reducing computation times. We hope that our work paves the way for further exploration of subsampling strategies in time series analysis.

Author Contributions: Conceptualization, J.L. and Z.W.; methodology, J.L. and Z.W.; formal analysis, J.L. and Z.W.; investigation, J.L. and Z.W.; writing—original draft preparation, Z.W.; writing—review and editing, N.R. and H.W.; supervision, N.R. and H.W.; project administration, N.R. and H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data used for the paper are computer-generated. Codes for generating the data are available upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Elorrieta, F.; Eyheramendy, S.; Palma, W. Discrete-time autoregressive model for unequally spaced time-series observations. *Astron. Astrophys.* **2019**, *627*, A120. [CrossRef]
2. Mudelsee, M. Trend analysis of climate time series: A review of methods. *Earth-Sci. Rev.* **2019**, *190*, 310–322. [CrossRef]
3. Dutta, C.; Karpman, K.; Basu, S.; Ravishanker, N. Review of statistical approaches for modeling high-frequency trading data. *Sankhya B* **2023**, *85*, 1–48. [CrossRef]
4. Nagaraja, C.H.; Brown, L.D.; Zhao, L.H. An autoregressive approach to house price modeling. *Ann. Appl. Stat.* **2011**, *5*, 124–149. [CrossRef]
5. Erdogan, E.; Ma, S.; Beygelzimer, A.; Rish, I. Statistical models for unequally spaced time series. In Proceedings of the 2005 SIAM International Conference on Data Mining. SIAM, Newport Beach, CA, USA, 21–23 April 2005; pp. 626–630.
6. Anantharaman, S.; Ravishanker, N.; Basu, S. Hierarchical modeling of irregularly spaced financial returns. *Stat* **2024**, *13*, e692. [CrossRef]
7. Carlstein, E. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Stat.* **1986**, *14*, 1171–1179. [CrossRef]
8. Fukuchi, J.I. Subsampling and model selection in time series analysis. *Biometrika* **1999**, *86*, 591–604. [CrossRef]
9. Politis, D.N. Scalable subsampling: Computation, aggregation and inference. *Biometrika* **2023**, 111, 347–354. [CrossRef]
10. Shumway, R. *Time Series Analysis and Its Applications*; Springer: Berlin/Heidelberg, Germany, 2000.
11. Eyheramendy, S.; Elorrieta, F.; Palma, W. An autoregressive model for irregular time series of variable stars. *Proc. Int. Astron. Union* **2016**, *12*, 259–262. [CrossRef]
12. Elorrieta, F.; Eyheramendy, S.; Palma, W.; Ojeda, C. A novel bivariate autoregressive model for predicting and forecasting irregularly observed time series. *Mon. Not. R. Astron. Soc.* **2021**, *505*, 1105–1116. [CrossRef]
13. Ghysels, E.; Jasiak, J. GARCH for irregularly spaced financial data: The ACD-GARCH model. *Stud. Nonlinear Dyn. Econom.* **1998**, *2*, 1–19. [CrossRef]
14. Meddahi, N.; Renault, E.; Werker, B. GARCH and irregularly spaced data. *Econ. Lett.* **2006**, *90*, 200–204. [CrossRef]
15. Engle, R.F.; Russell, J.R. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* **1998**, *66*, 1127–1162. [CrossRef]
16. Maller, R.A.; Müller, G.; Szimayer, A. GARCH modelling in continuous time for irregularly spaced time series data. *Bernoulli* **2008**, *14*, 519–542. [CrossRef]
17. Buccheri, G.; Bormetti, G.; Corsi, F.; Lillo, F. A score-driven conditional correlation model for noisy and asynchronous data: An application to high-frequency covariance dynamics. *J. Bus. Econ. Stat.* **2021**, *39*, 920–936. [CrossRef]
18. Dutta, C. Modeling Multiple Irregularly Spaced High-Frequency Financial Time Series. Ph.D. Thesis, University of Connecticut, Storrs, CT, USA, 2022.
19. Drineas, P.; Mahoney, M.W.; Muthukrishnan, S. Sampling algorithms for l_2 regression and applications. In Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms, Miami, FL, USA, 22–24 January 2006; pp. 1127–1136.
20. Yang, T.; Zhang, L.; Jin, R.; Zhu, S. An explicit sampling dependent spectral error bound for column subset selection. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; Volume 37, pp. 135–143.
21. Ma, P.; Mahoney, M.W.; Yu, B. A statistical perspective on algorithmic leveraging. *J. Mach. Learn. Res.* **2015**, *16*, 861–991.

22. Xie, R.; Wang, Z.; Bai, S.; Ma, P.; Zhong, W. Online decentralized leverage score sampling for streaming multidimensional time series. In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, Okinawa, Japan, 16–18 April 2019; Volume 89, pp. 2301–2311.
23. Zhu, R. Gradient-based sampling: An adaptive importance sampling for least-squares. *Adv. Neural Inf. Process. Syst.* **2018**, *29*, 406–414.
24. Wang, H.; Zhu, R.; Ma, P. Optimal subsampling for large sample logistic regression. *J. Am. Stat. Assoc.* **2018**, *13*, 829–844. [CrossRef]
25. Teng, G.; Tian, B.; Zhang, Y.; Fu, S. Asymptotics of Subsampling for Generalized Linear Regression Models under Unbounded Design. *Entropy* **2022**, *25*, 84. [CrossRef]
26. Wang, J.; Zou, J.; Wang, H. Sampling with replacement vs Poisson sampling: A comparative study in optimal subsampling. *IEEE Trans. Inf. Theory* **2022**, *68*, 6605–6630. [CrossRef]
27. Wang, H.; Yang, M.; Stufken, J. Information-based optimal subdata selection for big data linear regression. *J. Am. Stat. Assoc.* **2019**, *114*, 393–405. [CrossRef]
28. Pronzato, L.; Wang, H. Sequential online subsampling for thinning experimental designs. *J. Stat. Plan. Inference* **2021**, *212*, 169–193. [CrossRef]
29. Casella, G.; Berger, R. *Statistical Inference*; CRC Press: Boca Raton, FL, USA, 2024.
30. Kleinberg, J.; Tardos, E. *Algorithm Design*; Pearson/Addison-Wesley: Hoboken, NJ, USA, 2006.
31. Wynn, H. *Optimum Submeasures with Applications to Finite Population Sampling*; Academic Press: Cambridge, MA, USA, 1982.
32. Fedorov, V.V.; Hackl, P. *Model-Oriented Design of Experiments*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.