Title

Consequences of Sampling Frequency on the Estimated Dynamics of AR Processes Using Continuous-Time Models

Authors
Batra, Rohit
Johal, Simran K
Chen, Meng
et al.

Publication Date
2023-07-10

DOI
10.1037/met0000595

Consequences of Sampling Frequency on the Estimated Dynamics of AR Processes using Continuous Time Models

Rohit Batra[1*], Simran K. Johal[1*], Meng Chen[1], and Emilio Ferrer[1]
[1]University of California, Davis

Continuous-time (CT) models are a flexible approach for modeling longitudinal data of psychological constructs. When using CT models, a researcher can assume one underlying continuous function for the phenomenon of interest. In principle, these models overcome some limitations of discrete-time (DT) models and allow researchers to compare findings across measures collected using different time intervals, such as daily, weekly, or monthly intervals. Theoretically, the parameters for equivalent models can be rescaled into a common time interval that allows for comparisons across individuals and studies, irrespective of the time interval used for sampling. In this study, we carry out a Monte Carlo simulation to examine the capability of CT autoregressive (CT-AR) models to recover the true dynamics of a process when the sampling interval is different from the time scale of the true generating process.

We use two generating time intervals (daily or weekly) with varying strengths of the autoregressive parameter and assess its recovery when sampled at different intervals (daily, weekly, or monthly). Our findings indicate that sampling at a faster time interval than the generating dynamics can mostly recover the generating autoregressive effects. Sampling at a slower time interval requires stronger generating autoregressive effects for satisfactory recovery, otherwise the estimation results show high bias and poor coverage. Based on our findings, we recommend researchers use sampling intervals guided by theory about the variable under study, and whenever possible, sample as frequently as possible.

Keywords: Continuous-Time Models, Autoregressive Processes, Sampling Frequency, Dynamic Models

One of the primary goals of psychological research is studying the development of psychological attributes such as cognitive abilities, emotions, and skills of individuals over time. This necessitates the collection of longitudinal data with repeated measures of a process for an individual, such as day-to-day changes in mood of an individual with bipolar disorder, week-to-week changes in stress of a person in a new job, or month-to-month changes in reading ability of a child during grade school. Usually, researchers are interested in modeling the dynamics of the given process, or the systematicity in the pattern of changes over a time period, which depend on the time interval at which a process is sampled (Boker and Nesselroade, 2002; Browne and Nesselroade, 2005).

Most of the commonly used approaches for modeling process dynamics can be categorized as discrete-time (DT) models such as vector autoregressive models, growth curve models, or cross-lagged panel models. DT models examine the relations between consecutive occasions, which are assumed to be equally spaced. Such an assumption provides the benefit of easily interpretable dynamics with respect to the specific time interval. At the same time, it limits the understanding of the true underlying phenomenon, as one cannot interpolate or extrapolate the findings beyond the chosen time interval (Gollob and Reichardt, 1987). It is also likely that this time interval can vary, both within the repeated measures of an individual, or across individuals in a

This paper is accepted for publication in the APA journal *Psychological Methods*.
 The final article will be available, upon publication, via its DOI: 10.1037/met0000595

* These authors contributed equally to this work.

Rohit Batra
Simran K. Johal
Meng Chen
Emilio Ferrer

All code regarding data generation and model estimation is available online at https://github.com/rohitbatra29/Sampling_on_CT-AR.git

Correspondence concerning this article should be addressed to Rohit Batra, University of California, Davis. E-mail: rbat@ucdavis.edu

study, and also across different studies (Voelkle et al., 2012).

For example, an experience sampling method with measurements at multiple times per day may have different intervals for the same individual, such that the interval between two consecutive daytime measurements will be different from the time lag between before and after bedtime measurements. Similarly, different individuals may have longer or shorter sleep periods which can bring differences across individuals in their sampling intervals. Varying time intervals make it difficult to estimate and interpret a DT model and are usually handled by inserting phantom variables at the missing occasions to meet the equal interval assumption (Boker et al., 2018; Voelkle and Oud, 2015).

Continuous-time (CT) models circumvent the issue of varying time intervals by conceptualizing time on a real-number scale[1]. These models assume one underlying continuous function for the process under study that applies to the entire span of time, thereby allowing for unequal time intervals between consecutive measures of individuals (Voelkle and Oud, 2013). Their primary advantage is that one can rescale the dynamic parameters to any discrete time interval, and compare their strengths across different individuals, or studies. Nevertheless, CT models still rely on discrete measurements[2] that serve as snapshots into this underlying trajectory of a process. But what happens when these snapshots do not match the actual dynamics of a process for an individual? For instance, suppose an individual's stress level is a continuous function that is changing mostly on a day-to-day basis as shown in the top panel of Figure 1, but a researcher decides to observe them on a weekly time interval shown in the middle panel of the figure. Theoretically, the stress dynamics for this individual can be estimated using a CT model on the weekly interval data and rescaling the parameters to a daily time interval. However, would we expect these estimated dynamics to be the same had the researcher observed the individual at a daily interval?

The frequency at which a process is sampled can be called the sampling frequency, e.g., weekly sampling frequency for our hypothetical example. In contrast, the frequency at which a researcher is interested in studying the dynamics of a process can be called the data generating frequency[3] or frequency of interest, such as daily frequency of stress for the above example. The discrepancy between the sampling and generating frequencies is an important, yet unaddressed, issue in the CT literature, and there are no established guidelines on the limits of rescaling the CT dynamics. The aim of this paper is to study the consequences of sampling frequency on the estimation of CT models for individual dynamics. In the remaining introduction, we first present DT autoregressive and CT autoregressive models along with their assumptions. Then we provide an overview of recommendations for choosing an appropriate sampling frequency in DT models and consider whether they apply for CT models as well. This forms the motivation for our simulation work.

DT-AR Models

A dynamic process can be defined as a process with systematicity or stability in their movement over time with a feature of self-regulation, where the current state and changes in the process depend on its past state and changes (Boker and Nesselroade, 2002). A commonly used model that formalizes this time-dependence feature using repeated measures from a single individual is an autoregressive (AR) model of first order, given by the equation:

$$y_t = c + \phi \,.\, y_{t-1} + \epsilon_t \tag{1}$$

where the AR parameter ϕ represents the influence of the value of a process at the previous state on the current state at occasion (or discrete time point) t. ϵ_t is the random shock to the system at every new time point t, which are independently and identically distributed with mean zero and variance σ^2_ϵ. c is an additive constant which we assume to be 0 for the rest of the paper. We refer to the above model as the discrete time autoregressive (DT-AR) model because of its use of discrete time points to specify the process.

The sign and strength of the AR parameter ϕ indicates whether the process moves towards or away from an equilibrium point and, if so, at what pace, respectively. In our work, we are concerned with stationary DT-AR models with $\phi \in (0, 1)$, also referred to as *positive* AR systems, in which the lagged parameter is interpreted as bringing the process back to the equilibrium, whereas the random shocks push the system away from it (Browne and Nesselroade, 2005; Ryan and Hamaker, 2021). In this context, $\frac{c}{(1-\phi)}$ serves as the long-term equilibrium point of a dynamic process and can be thought of as the home base or attractor (Oravecz et al., 2011). The DT-AR model makes two important but restrictive assumptions: 1) the process is weakly stationary, that is, the means and variances remain constant over time; and, 2) the time interval (also called, time lag) from any one

[1]Compared to DT models, where time is conceptualized as integers or discrete increments.

[2]As time is conceptualized as a real number, per definition, no matter how densely these snapshots are taken, there always exists an interval of time between the snapshots. Therefore, measurements are in discrete as opposed to continuous time.

[3]An assumption of CT is the continuous existence of the process for all possible frequencies at which a researcher can study such a process. With the use of the term "generating frequency", we do not mean that these processes have a single generating frequency. However, a researcher still needs to study these processes at a particular frequency of interest, and this serves as the data generating frequency. This data generating frequency is helpful as a reference when we make comparisons with a particular sampling frequency in the rest of the paper.

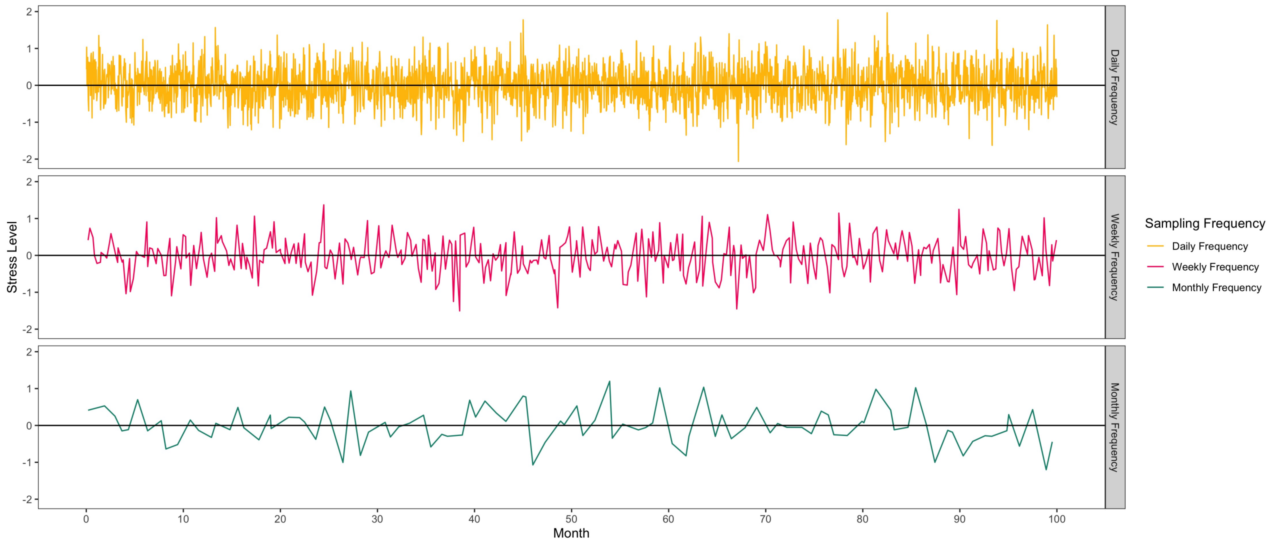

Figure 1

Example of a daily generating process which is sampled at weekly and monthly frequencies.

discrete measure to the next is of the same length (Browne & Nesselroade, 2005; du Toit & Browne, 2007).

The consequence of the second assumption is that the estimated parameters are a function of the time interval used to collect and model the data, which has come to be known as the time lag problem of DT models (Gollob and Reichardt, 1987; Ryan and Hamaker, 2021). Researchers studying the same dynamic process can use different time intervals to collect the data, and this makes it impossible to compare the estimated dynamics across studies. For instance, a researcher might choose to formalize an individual's stress as a DT-AR model with repeated measures collected on a day-to-day interval, whereas another researcher might study the same individual's stress as a DT-AR model but on a week-to-week interval. There is no way to directly compare the parameters estimated from the two because the parameters are not linear functions of time (Voelkle et al., 2012). It is also quite common in empirical research to collect longitudinal data with unequal time intervals between consecutive assessments, and this is usually handled by treating the unobserved time points as missing to meet the assumption of equal time intervals. However, adding missingness can become quite cumbersome if time intervals vary across assessments waves, e.g., for an individual with 101 measures, we can have 100 possible time intervals of varying lengths (Voelkle and Oud, 2015).

CT-AR Models

CT models assume an underlying generating function that works continuously over time for a psychological process, given that this function is smooth and differentiable (Boker, 2002). Unlike DT models, in which the increments in the process are at discrete intervals, the process in CT models moves along the smallest possible increments of time. This shifts the formalization of changes from differences to differential equations. A simple CT model for a single-subject time series is a univariate *Ornstein-Uhlenbeck (O-U) process*, which is a first-order stochastic differential equation given by:

$$dy_t \;=\; \beta\,.\,(y_t - c)\,.\,dt \;+\; \sigma\,.\,dw_t \qquad (2)$$

where, the differential of y_t with respect to time t, $\frac{dy_t}{dt}$, represents the instantaneous rate of change of y_t. This differential includes both a deterministic and a stochastic component, denoted by the two summands respectively. The deterministic part represents the influence of current values of y_t on the instantaneous changes, scaled by the parameter β, termed the auto-effect. Similar to the AR parameter ϕ in a DT-AR model, the sign and value of β determine if these small changes bring the process back to the equilibrium or away from it (Kuiper and Ryan, 2018; Oravecz et al., 2011; Voelkle and Oud, 2013). We assume equilibrium (denoted by c in Equation 2) to be 0 for the discussion here. The stochastic part represents the change in the unidimensional Weiner process w_t or Brownian motion, scaled by the constant σ (Voelkle et al., 2012). This represents the random shocks to the dynamic system that push it away from (or bring it accidentally back to) equilibrium.

In this paper, we only consider O-U processes with $\beta < 0$ that return to equilibrium over time, making it the CT counterpart to the DT-AR model, and we denote it as the continuous time autoregressive (CT-AR) model for the rest of the paper. This model provides the flexibility of accounting for different time intervals across the repeated measures of a process by considering the discrete measures as informative of an underlying continuous trajectory, wherever they fall (Voelkle et al., 2018). Hence, the CT-AR model proves a more flexible alternative to handling unequal time intervals than the DT-AR model. Another benefit of CT-AR models is that one can theoretically rescale the CT parameters to any discrete time interval, regardless of the sampling design. This provides a macroscopic view of the process over different time intervals, and many researchers are interested in capturing the time interval at which the DT parameters are maximum (Deboeck and Preacher, 2016; Dormann and Griffin, 2015).

Next, we briefly explain two important concepts of base time interval and rescaling in CT models, which will be helpful for our further discussion.

Base Time Scale

A base time scale defines the time metric used in a CT model, which is usually set to be equal to the smallest time interval during data collection, or at the discretion of the researcher based on their hypothesis. For a given initial observation y_{t_0}, the solution of the differential equation of the CT-AR model can be found by integrating over Equation 2 and is given by:

$$y_{t_i} \; = \; e^{\beta \,.\, \Delta t_i} \,.\, y_{t_{i-1}} \; + \; \epsilon_{\Delta t_i} \tag{3}$$

where, y_{t_i} represents the i^{th} observation taken at time t, predicted by the previous observation $y_{t_{i-1}}$ scaled by an exponential function and the innovation term $\epsilon_{\Delta t_i}$. Similar to the errors in Equation 1, these innovations are also independent and identically distributed with mean 0 and variance $\sigma^2_{\epsilon} = \sigma^2_y(1 - e^{2*\beta*\Delta_i})$, where σ^2_y is the stationary variance of the process. This innovation variance is a function of β, which is the same auto-effect as before, and $\Delta t_i = t_i - t_{i-1}$, which is the time elapsed between the two consecutive observations. We name the time interval when $\Delta t = 1$ as the base time scale because it serves as the basis relative to which other discrete time intervals are calculated for the purposes of estimating a CT model.

For instance, let us consider again a study of stress in an individual over time. Suppose a researcher decides to collect stress measures from this individual every day, so the base time interval is one day, i.e., $\Delta t = 1$ day. During data collection, however, there might be some days when the data are not measured, or the observation time is different across days, e.g., 9 a.m. one day, but 12 p.m. the next day, which will make the $\Delta t_i = 1.125$ days between these two consecutive measures. The choice of the base time interval can be entirely up to the researcher modeling the process and it can be easier to think of it as the smallest possible sampling interval against which other intervals are scaled.

Rescaling

Rescaling the CT parameter estimates to any discrete time interval is one of the major advantages of CT models, as it makes it easier to compare the DT estimates across different individuals and studies (Voelkle et al., 2012). Using Equations 1 and 3, we can establish this direct relationship between the autoregressive effect from the DT-AR model and the auto-effect from the CT-AR model:

$$\phi_{\Delta t} \; = \; e^{\beta \,.\, \Delta t} \tag{4}$$

where $\phi_{\Delta t}$ represents the DT AR parameter for a particular discrete interval Δt and β is the auto-effect, which does not depend on any discrete interval. For example, the researcher collecting daily data on stress of an individual with a base time interval of $\Delta t = 1$ day can estimate a CT-AR model and use Equation 4 to get the autoregressive effect for a weekly interval with $\Delta t = 7$ days ($\phi_{\Delta t=7} \; = \; e^{\beta \,.\, 7}$). For more information on the derivation of this equivalence, see Voelkle et al. (2012). Unlike DT-AR models, where our inferences about the process are restricted to the time interval used in data collection, we can rescale the inferences from CT-AR models to any discrete interval. Hence, based on this equivalence, one can functionally rescale to any time interval but Voelkle et al. (2012) caution the readers to rescale within 'reasonable limits', e.g., it is unreasonable to rescale the estimated dynamics with a base time interval of minutes ($\Delta t = 1$ minute) to a discrete interval of years ($\Delta t = 1$ year = 524,160 minutes).

Even though the CT model parameters can be rescaled to any time interval, the estimated dynamics still depend on the sampling frequency at which the process is observed by the researcher. In the next section, we discuss the determinants of sampling frequency for a psychological process together with past work that has studied the influence of sampling frequency on the CT dynamics.

Choosing An Appropriate Sampling Frequency

Psychological research relies on observations of discrete measures from longitudinal studies. However, the determination of when or how frequently these measures should be taken in a given time period is usually based on a researcher's hypothesis or knowledge. Such knowledge could be the time lag at which the construct operates, the proposed lag at which the causal effect is highest in a system of variables, or simply the convenience of the research design (Dormann and Griffin, 2015; Dormann and van de Ven, 2014). The use of some of these approaches has drawn criticism in DT modeling. For instance, Collins and Graham (2002) warned against using empirically-derived sampling intervals because they can lead to *chicken and egg situations*. This usually happens when researchers do not have prior knowledge of the shape of the trajectory. If so, they do not know the frequency at which to observe the measure and, without such knowledge, the empirical data collected can be unreliable, ultimately causing issues if someone were to use empirical intervals for future data collection and inferences (Adolph et al., 2008).

Over time, a recommendation has emerged from many methodologists to sample as frequently as possible, the so-called the *microgenetic method* (Adolph et al., 2008). However, such oversampling can lead to high correlations between the measurement error of the closely repeated measures (Boker, 2002), or issues related to the recruitment and attrition of participants in the study (Bolger and Laurenceau, 2013; Janssens et al., 2018). It is possible to minimize some of these biases and estimate the pattern for longer sampling intervals from frequently sampled data but not the other way around – with data collected at longer time lags, there is no information to discretize for shorter time lags and estimate the dynamics of the process using current DT modeling approaches (Adolph et al., 2008). To determine the sampling rate for oscillating functions, there exists rules of thumb like Nyquist-Shannon theorem, where the sampling frequency should be at least twice as frequent as the process frequency of interest (Nyquist, 1928; Shannon, 1948). In this paper, we are not concerned with such processes and thus, we do not use this rule of thumb. Another general advice in the literature has been that the sampling frequency should be *appropriate* for the variable under study. However, researchers do not always have a clear hypothesis of the time interval at which a process operates for an individual and they might use different exploratory intervals to formalize the operationalization.

Here, we ask the question, does one need to consider these recommendations when modeling in CT, or can we sample at any time interval and simply rescale the estimates? For example, instead of collecting daily measures of stress, a researcher might decide to collect weekly measures of stress, which might be too sparse to capture the functional form of how an individual's stress changes on a day-to-day basis, even if theoretically a CT-AR model allows us to rescale the dynamics from weekly to daily time interval. There has been little research exploring the effects of sampling intervals on CT estimation and there are no established guidelines on rescaling from one discrete interval to another. Only recently, work by Adolf et al. (2021) on deriving optimal sampling intervals acknowledges the importance of determining an appropriate sampling frequency for CT models. They propose a formula that uses past estimates of the auto-effect, β, to derive an optimal sampling interval that would result in the most reliable estimation of the CT model from new data. However, their recommended formula is asymptotically derived and its performance has not been examined in the finite number of observations that would be encountered in psychological research. Our work, on the other hand, approaches the idea of data collection and sampling intervals from the perspective of a mismatch between the true underlying process of interest and the interval a researcher uses to collect their data. This is done within the domain of finite numbers of observations. We expand on these aspects in more detail below.

We can rephrase our previous question – *Do we expect the choice of sampling frequency to affect the estimation and inferences about the dynamics from a CT-AR model?* We believe it is possible that the choice of sampling frequency does not affect the dynamics estimated from CT models, due to their flexibility to rescale across discrete intervals. However, it is also possible that an incorrect choice of sampling frequency might entirely miss the dynamics of interest. If so, the ability to rescale the estimates cannot correct such a loss. Our results later show evidence for both these points of view.

Purpose of the Paper

In this paper, we are interested in examining the role of the sampling time interval for recovering the true dynamics of a given process, and the consequences of using a sampling interval different than the interval at which the process operates. This will also help determine some of the limitations of rescaling the estimated CT-AR parameters, for a given sampling frequency. To accomplish this, we use a Monte Carlo simulation study in which we evaluate if it is possible to capture the true dynamics of a CT-AR model under different conditions of generating and sampling frequencies.

In the following section, we introduce the conditions and steps of our simulation procedure along with the criteria used to evaluate the estimated models. Next, we show the results from the simulation, and end with a discussion of our findings and our recommendations to researchers planning to use CT-AR models for their study purposes.

Methods

Simulation Conditions

We are interested in the ability of CT-AR models to recover population dynamics, and whether this ability is affected by various factors, including (a) the data generating frequency (b) the strength of the autoregressive effect, (c) the number of observations, and (d) the equality of the time intervals between measurement occasions. We limited our focus to a single variable with no measurement error in order to reduce the complexity of the simulation, facilitate the interpretation of the results, and better illustrate the ability of CT models to recover population dynamics in the simplest scenario possible.

Data Generating Frequency

The generating process operated at one of the two frequencies: a daily or a weekly time interval. CT models assume that the process can operate at multiple discrete time intervals, and thus there is not necessarily a single time interval at which the process operates. We do not assume that the data generating frequency is the *only* time interval at which the process operates, but the time interval at which the dynamics of the process might be the most representative for that psychological construct. Furthermore, we chose our data generating frequencies to be either daily or weekly, as these frequencies can be representative of processes – such as emotional interactions in close relationship – that are of interest to psychological researchers (Chen et al., 2018; Newell, 1994).

Strength of Autoregressive Parameter

We chose four values for the AR coefficient, ϕ, to represent the dynamics of the generating process: .05, .2, .5, and .8. These AR coefficients represented the stability of the process from day-to-day or week-to-week. These four values were chosen to cover a spectrum of possible processes, with a process whose current observations are weakly dependent on the previous observation represented by weaker AR values on one end, and a process whose current observations strongly depend on the previous observation represented with stronger values at the other end. These coefficient values have been used in previous simulation studies examining autoregressive processes (de Haan-Rietdijk et al., 2017; O'Laughlin et al., 2020).

Number of Observations

We sampled either 100, 300, or 1,000 observations from the true process. We chose 100 as our smallest number of observations based on previous work showing that CT models with fewer than 100 observations on one individual are unlikely to perform optimally (Hecht & Zitzmann, 2020). Therefore, sampling less than 100 observations would make it difficult to separate poor performance of the CT-AR model more generally from poor performance due to sampling frequency. Furthermore, a sample of 100 observations is typical for most Experience Sampling Methods (ESM) designs, and has been used in previous studies investigating the performance of CT-AR models (Chen & Ferrer, 2022; Liu et al., 2021). We also examined the effect of increasing the number of observations, by including samples of 300 and 1,000 observations. A sample of 1,000 observations also allows us to study the asymptotic performance of the CT-AR model, with little influence from sampling error.

Equality of Sampling Intervals

ESM studies are oftentimes designed so that the time interval between measurement occasions is random, and thus not equal between pairs of measurement occasions. This is typically considered an advantage, as prompting participants to respond at random intervals can, for instance, minimize the risk of participants adapting their behavior in anticipation of a prompt, and can let researchers better characterize the process under study (Bolger et al., 2003; Voelkle and Oud, 2013). However, this randomness is ignored when estimating DT models, as these models assume that the time interval between consecutive observations is equal.

Therefore, in our simulation, we generated data with both equal and random time intervals between measurement occasions – the former represents an 'ideal' research scenario that meets the assumptions of DT models, and the latter represents a scenario that better aligns with how ESM research is conducted.

Summary of Simulation Conditions

Our simulation factors were therefore:

- Data Generating Frequency: Daily frequency, Weekly frequency
- Strength of Autoregressive Parameter: .05, .2, .5, .8
- Number of Observations: 100, 300, 1,000
- Equality of Sampling Intervals: Equal intervals, Random intervals

These factors were fully crossed, resulting in a total of 48 conditions. For each condition, we simulated 500 datasets as described in the Simulation Procedure.

Simulation Procedure

An overview of the full simulation procedure is shown in Figure 2. We created our sample datasets by first generating data for the complete trajectory using a time interval of one hour between measurement occasions. Then, we created sample datasets based on three different sampling frequencies, and estimated CT-AR models for each of the samples.

Data Generation

We chose to generate data using an hourly time interval because we needed a discrete realization of the continuous process to sample from that was common to our three sampling frequencies (daily, weekly, and monthly, which will be discussed further below). Furthermore, generating data at an hourly time interval means that it occurs faster than our fastest sampling frequency, allowing us to approximate a continuous process. Finally, hourly data – although uncommon – is possible in psychological research, particularly

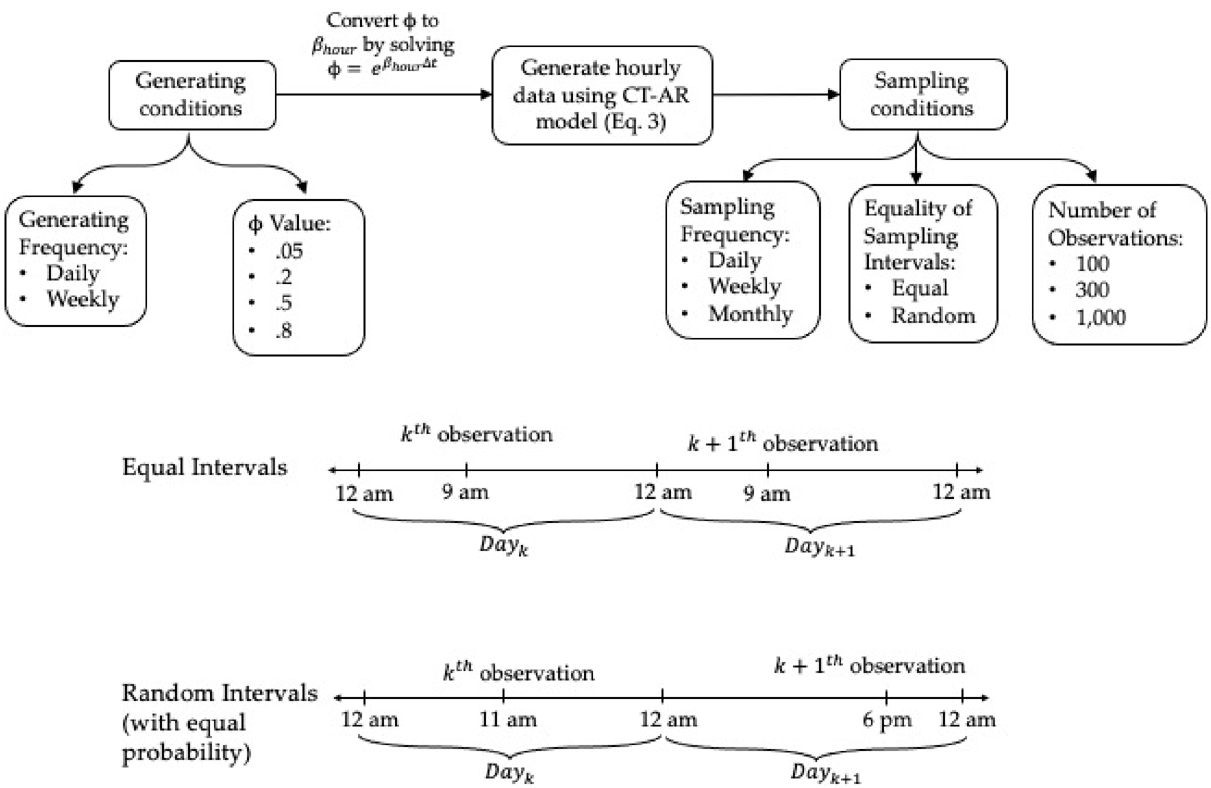

Figure 2

Overview of data generation procedure and simulation conditions

Generating Frequency	AR Parameter (ϕ)	CT hourly Auto-Effect (β_{hour})
Daily	.05	-.125
	.2	-.067
	.5	-.029
	.8	-.009
Weekly	.05	-.018
	.2	-.010
	.5	-.004
	.8	-.001

Table 1

Table of population values
Note. We used the generating AR parameters to get the CT auto-effects at the hourly interval for data generation.

when studying physiological systems (e.g., Jacobson et al., 2019).

To generate the complete trajectory from a process that met the conditions described above, we first transformed the DT AR effect, ϕ, to a CT auto-effect, β, using the "indirect" method (Oud et al., 1993). The "indirect" method solves for the underlying CT auto-effect for a given AR effect and time interval using Equation 4. Note that the use of the "indirect" method is generally not recommended, as it cannot result in a unique solution for the CT auto-effect except under certain conditions (Hamerle et al., 1991; Kuiper and Ryan, 2018; Ryan and Hamaker, 2021; Voelkle et al., 2012). In our simulation, we assumed a negative and real-valued underlying CT auto-effect, satisfying the condition of the "indirect" method and allowing us to transform from ϕ to β_{hour}. We then used β_{hour} in Equation 3 to generate data at the hourly level. For example, suppose the process operated on a daily frequency with an AR parameter of 0.5. We obtained the CT auto-effect on an hourly time interval by assuming that the base time interval was hourly, so that the corresponding time interval for 1 day would be Δt = 24 hours. We then solved the following equation for β_{hour}:

$$0.5 = e^{24 \cdot \beta_{hour}} \tag{5}$$

, resulting in the population hourly auto-effect, β_{hour} = -0.03. All the values of the true DT effects and the corresponding CT auto-effects can be found in Table 1.

To generate the entire trajectory for each replication, we set the first observation and the mean of the process to zero and substituted β_{hour} into Equation 3 to generate the remainder of the process. The innovations are generated for each new time point (i.e., each new hour) using a standard normal distribution with mean zero and variance $\sigma_{\epsilon}^2 = \sigma_y^2(1-e^{2*\beta*\Delta_i})$, where σ_y^2 was set to 0.25 for the rest of the simulation. We decided to keep this variance of the process fixed to a single small value in order to avoid contamination from the data generation process and keep our focus on the examination of simulation conditions mentioned before. Hence, the innovation variance for each condition relies primarily on the auto-effect β and the data generating frequency. The total time period for data generation was chosen such that there were enough data for even the slowest sampling frequency, e.g., monthly, to sample the required number of observations[4].

Sampling

After generating the complete trajectory, we sampled at three different sampling frequencies: daily, where measurements were taken once every 24 hours; weekly, where measurements were taken once every 168 hours; and monthly, where measurements were taken once every 672 hours. Daily and weekly measurements, in particular, are common in psychological studies assessing emotional affect or psychopathological symptoms (Bos, 2021; Castro-Schilo and Ferrer, 2013; Lenderking et al., 2008; Wichers et al., 2021). The exact hour of measurement from a particular period was determined based on whether the time intervals were intended to be equal or random: if measurements were taken with equal time intervals, then the 10^{th} observation of each period was always taken; if measurements were taken with random time intervals, then a random measurement within that period was selected (this process is visualized in Figure 2). Measurements were taken until the required number of observations was reached.

Estimation

We fit a CT-AR model to each sampled dataset using *OpenMx* (Boker et al., 2022; Deboeck and Preacher, 2016; Hunter, 2018; Pritikin et al., 2015). To align with how researchers would estimate a CT-AR model with empirical data, we used the time interval of the sampling frequency as the base time interval for the model. The CT model is estimated in OpenMx using the hybrid Kalman filter and nonlinear optimization (Boker et al., 2022; Chow et al., 2010; Hunter, 2018). The hybrid Kalman filter estimates the latent states of the continuous process to the observed measurement under a given set of parameter values and derives the likelihood associated with this set of parameter values as a byproduct. It uses a series of recursive steps: first is the prediction step, where the value of the process at time t and the covariance matrix are predicted based on the model, given the parameters, and the corrected estimates of the process at the previous time point. Next, the correction step corrects this prediction based on the observed data at time t. These two steps iterate from the first time point to the last and return the likelihood value, which is then used in an optimization algorithm to derive the maximum likelihood estimates for the model parameters.

In addition to the three CT-AR models, we also estimated a DT-AR model for the dataset where the generating frequency and the sampling frequency aligned, e.g., the generating and sampling frequency were both daily or both weekly. We included the DT-AR model in comparison to the CT-AR model because it serves as a useful benchmark for how well our CT model recovers the dynamics when the assumption of equal time intervals is violated. If the time intervals between measurement occasions were equal, then both models should give equivalent results; however, if the measurements are unequally spaced, then the CT-AR model should result in more

[4]Every iteration of our simulation uses 0 as the initial value for data generation which is also the mean of the process, and the process variance is fixed to a small value throughout. Hence, we did not add a burn-in phase to the data generation. If future studies explore these simulations for different values of the process mean or variance, it is recommended to add a burn-in phase for data generation to let the simulated data reflect the generating process.

accurate estimates (de Haan-Rietdijk et al., 2017; Loossens et al., 2021; Voelkle and Oud, 2013).

All code regarding data generation and model estimation is available online at https://github.com/rohitbatra29/Sampling_on_CT-AR.git.

Performance Measures

As detailed in Figure 3, we evaluated a total of four models for each replication: three CT-AR models (one for each sampling frequency), and one DT-AR model when the sampling frequency matched the generating frequency. We assessed how accurately and precisely these models were able to estimate the dynamics of the underlying process.

Relative Bias

Since we were interested in how well the CT-AR model recovers the dynamics of the true process across different sampling frequencies, we used Equation 4 to rescale the estimated auto-effects to the generating time interval. These estimates were then compared to the population AR values as shown in Figure 3. We evaluated bias for the AR effects and not the auto-effects because: (1) most empirical researchers would interpret the results of their model on the original time metric, and not the CT metric, making it important that the AR parameters are unbiased; and, (2) bias can be harder to interpret for the auto-effect, since the auto-effect can range from $-\inf$ to 0, whereas the AR parameter can (in our simulation) only range from 0 to 1 (de Haan-Rietdijk et al., 2017; Kuiper and Ryan, 2018).

The accuracy of the estimated AR parameter was assessed using relative bias, which is calculated as:

$$RB = \frac{\hat{\phi} - \phi}{\phi} \tag{6}$$

where $\hat{\phi}$ represents the estimated AR value, and ϕ the population AR value. We considered a parameter to be substantially biased if the absolute value of the relative bias was greater than 0.1, in line with previous literature (Flora and Curran, 2004). Although we did also consider bias (which is simply the difference between the estimated and population AR values), we chose to focus our discussion on relative bias, due to the availability of cut-off values for deeming an estimate to be substantially biased or not.

Confidence Interval Coverage

In addition to the point estimate, we assessed the coverage of the 95% confidence intervals for the true CT auto-effect, as well as the true AR parameter at the generating frequency. The coverage rate of the auto-effect helps us determine how well the CT-AR model performs under different sampling frequencies. As the data were generated from a CT-AR model, we would expect that the coverage rate of the auto-effect would be satisfactory under optimal sampling conditions.

However, as mentioned before, researchers are more likely to evaluate and interpret the results of the CT-AR model on the original time metric of their study, and not the CT metric. Therefore, it is important that coverage for the true AR value is also satisfactory, as we would hope that rescaling the CT model estimates would still allow us to capture the true AR parameter. Since simply rescaling the limits of the confidence interval for the auto-effect using Equation 4 would provide us the same information as the coverage of the CT auto-effect, we instead used the delta method to transform the 95% confidence intervals of the auto-effect to 95% confidence intervals of the AR parameter (Weisberg, 2014). The delta method, in general, is used to calculate the standard error of a nonlinear transformation of the model parameters based on the model parameter's standard error and the partial derivative of the transformation function. These standard errors can then be used to calculate confidence intervals. In our case, the AR estimate is a nonlinear function of the auto-effect (as shown in Equation 4), and the standard error of the AR parameter is given by:

$$SE(\phi) = \sqrt{var(\beta) \,.\, \Delta t^2 \,.\, e^{2\beta \,.\, \Delta t^2}} \tag{7}$$

Therefore, the 95% confidence intervals for the AR parameter are:

$$e^{\beta \,.\, \Delta t} \pm 1.96 \cdot SE(\phi) \tag{8}$$

An example of using the delta method for the 95% confidence interval of the AR value is shown in Appendix A.

For both types of confidence intervals, the coverage rate was calculated as the proportion of replications within each condition whose confidence intervals contained the true parameter. If this proportion was between 90% and 95%, then coverage was considered optimal; proportions below 90% reflected poor coverage (Collins et al., 2001; Enders and Peugh, 2004).

Results

We analyzed the relative bias and coverage rate of the estimated models across the 48 conditions. For a given data generating frequency of daily or weekly time intervals, we compared: 1) three sampling frequencies – daily, weekly, and monthly; 2) strengths of the AR parameter, = .05, .2, .5, .8; 3) number of observations equal to 100, 300 and 1,000; 4) equal or random time intervals between any two consecutive measurements. For each condition we conducted 500 replications. Within each replication, we evaluated three CT-AR models for the three sampling frequencies and one DT-AR model when the sampling frequency matched the data generating frequency. Convergence was determined when the estimated model was successfully optimized in *OpenMx*. If the model did not converge, a different starting value was used for the

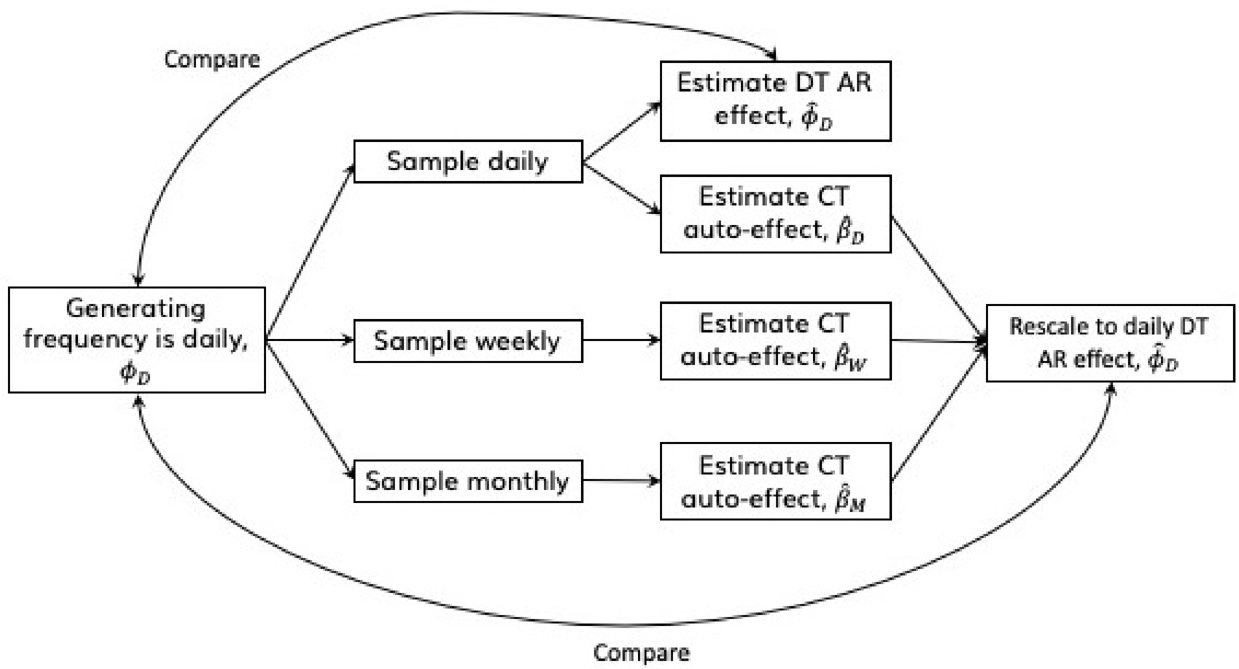

Figure 3

Example of model estimation performance evaluation

optimization of the parameter; if the model failed to converge even after 10 different starting values, it was classified as non-convergent, and the results of such models were not used for further analyses. Figure B1 in Appendix B shows the convergence rate of the estimated CT-AR models across all conditions. Overall, most conditions had perfect convergence, whereas a few showed only satisfactory convergence (but still above 80%). These handful of conditions belonged to the daily generating frequency with 100 observations and random intervals between measurements.

Relative Bias

Using relative bias, we were interested in comparing the rescaled DT parameter estimated from the CT model of the three sampling frequencies to the true AR parameter. Hence, Figures 4 and 5 show the median relative bias of $\hat{\phi}$ with respect to the true generating frequency, daily and weekly, respectively. The value of the rescaled $\hat{\phi}$ estimate is considered substantially biased if the absolute relative bias is above .1 (Flora and Curran, 2004). We also include the results for median bias in the Appendix C, but the conclusions outlined in this section remain fairly similar across the two performance measures.

Daily Generating Frequency

When the data generating frequency was daily, we noticed an overall large bias for weekly and monthly sampling frequencies for the smallest AR strength (= .05), as reported in Figure 4(a). This bias was reduced as the number of observations increased, and it was reduced considerably when the data were sampled with random time intervals and 1,000 observations . This implies that processes with little dependence on their previous states and sampling frequencies slower than the true frequency of interest show significant bias when the estimated CT-AR dynamics are rescaled to the true DT interval. For AR strengths greater than .05, monthly sampling frequency still showed bias in most cases, albeit smaller than before. Some exceptions where monthly rescaled estimates were unbiased include the condition when data were sampled at equal intervals with ϕ = .5, or random intervals with ϕ = .5 at sample size of 1,000, or data sampled at random intervals with ϕ = .8 for any number of observations

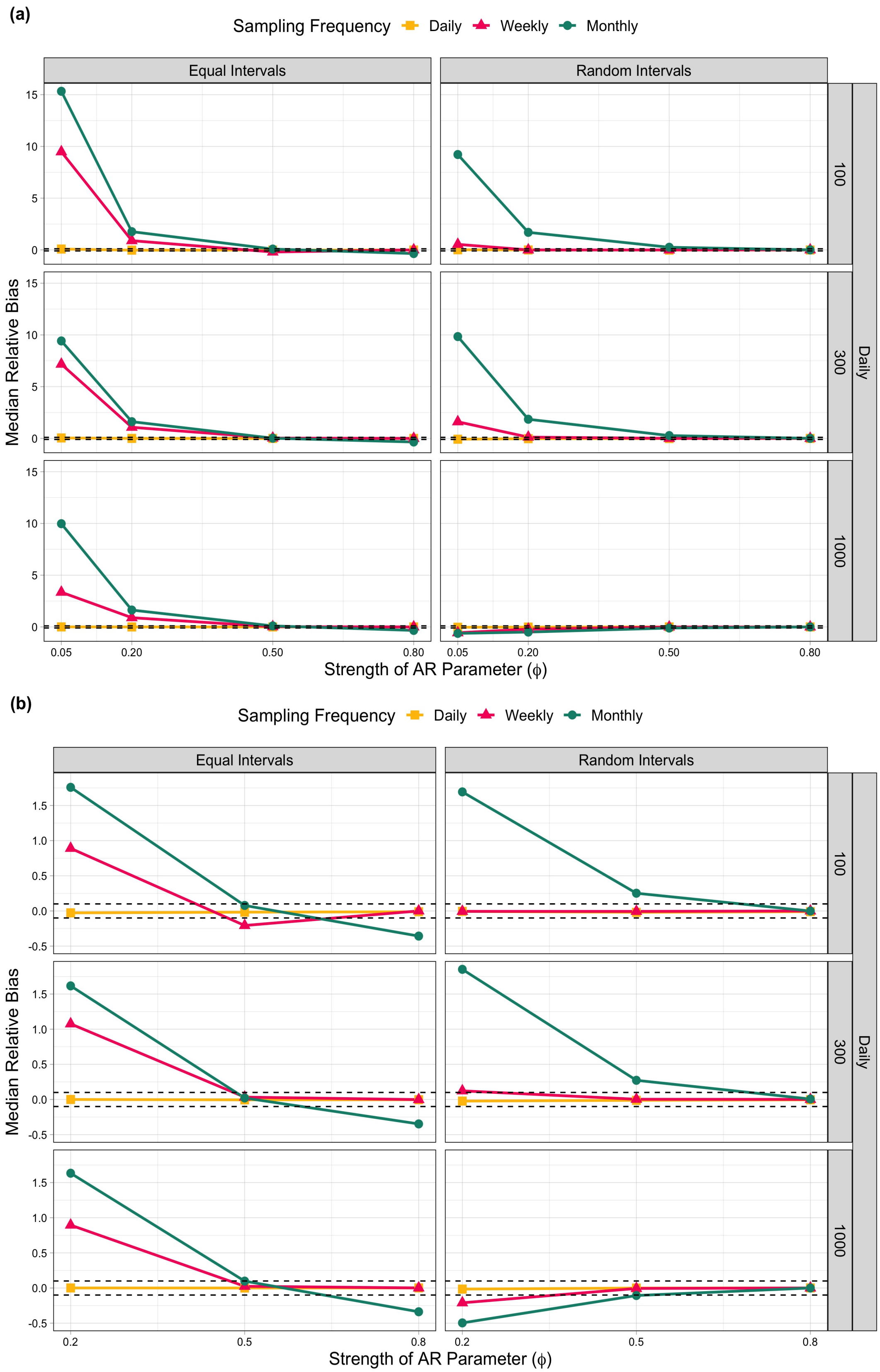

Figure 4

Median relative bias for daily generating frequency.

Note. (a) Relative bias for daily generating frequency for all values of AR parameter; (b) Relative bias for daily generating frequency for all values of AR parameter except when ϕ = .05. Plot (b) provides a zoomed in view of differences in bias between the three sampling frequencies, without distortion from larger values of relative bias.

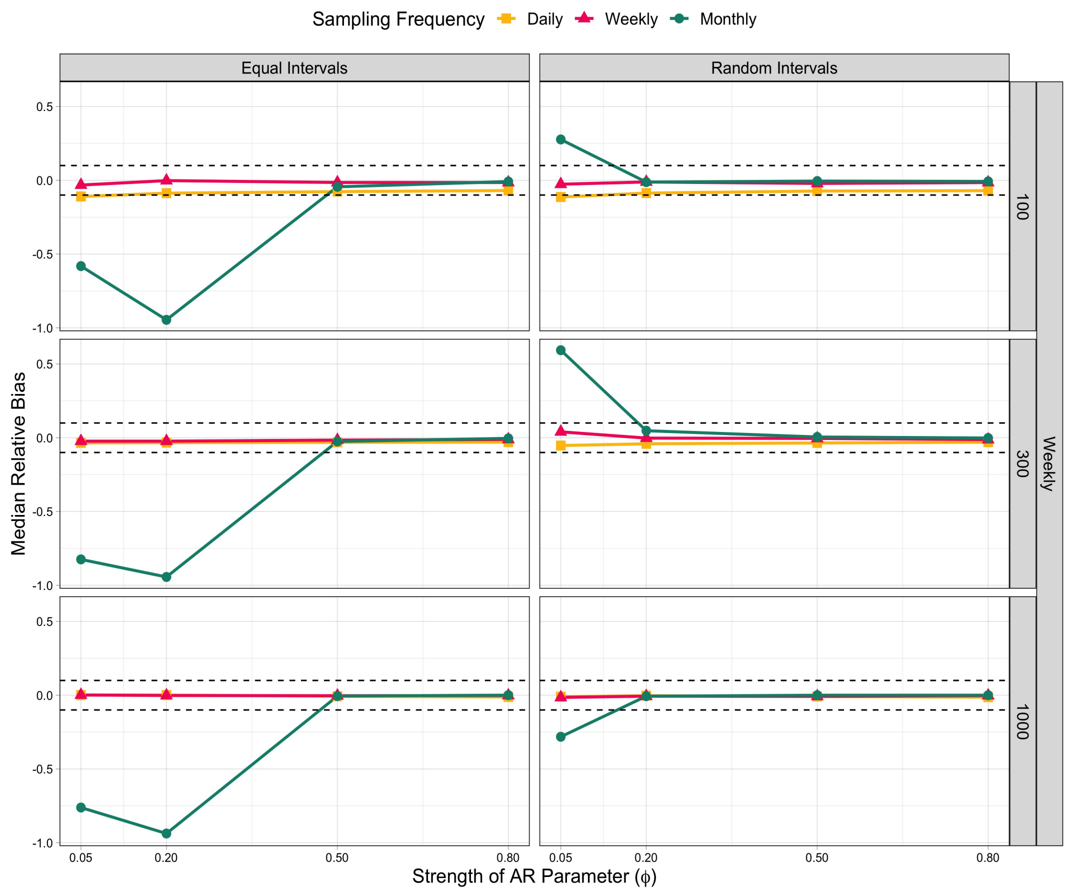

Figure 5

Median relative bias for weekly generating frequency.

(see Figure 4 (b)). On the other hand, rescaled ϕ estimates from data sampled at a weekly frequency were generally unbiased when the data had random intervals. For equal intervals, weekly estimates were unbiased for stronger AR parameters (ϕ = .5, .8) and more observations (T > 100).

The estimates when data were sampled at a daily frequency, i.e., when the generating and sampling frequencies matched, were unbiased across all conditions. That is, across all strengths of the ϕ parameter, all sample sizes and both conditions of equality of sampling intervals, the estimates did not show systematic bias.

Weekly Generating Frequency

When the generating frequency was weekly (Figure 5), there was no bias present for the daily and weekly sampling frequencies, across all the strengths of AR parameters, equality of sampling intervals, and observation sizes. These results signify that the CT-AR model produces unbiased rescaled estimates when sampling frequencies are faster than or matching to the generating frequency. Meanwhile, when the sampling frequency was slower than the data generating frequency, as was the case with monthly sampled data, we saw underestimates of the true parameter for weaker AR values (ϕ = .05, .2) at equal intervals of observation. There was also some bias present for monthly sampling frequency at the smallest AR value of .05 and random intervals, but it was not as prominent as compared to equal intervals. For $\phi \geq .2$, monthly sampled data with random intervals remained unbiased as seen in Figure 5.

Effect of Sampling Intervals and Number of Observations

There were no substantial differences in relative bias when comparing the results across the number of observations. Generally, we noticed that bias in estimates reduced or disappeared as the number of observations increased, but only when $\phi \geq .2$. Furthermore, comparing the two conditions of equality of sampling intervals, data sampled at random intervals tended to be less biased than data sampled at equal intervals. These differences were more salient when the

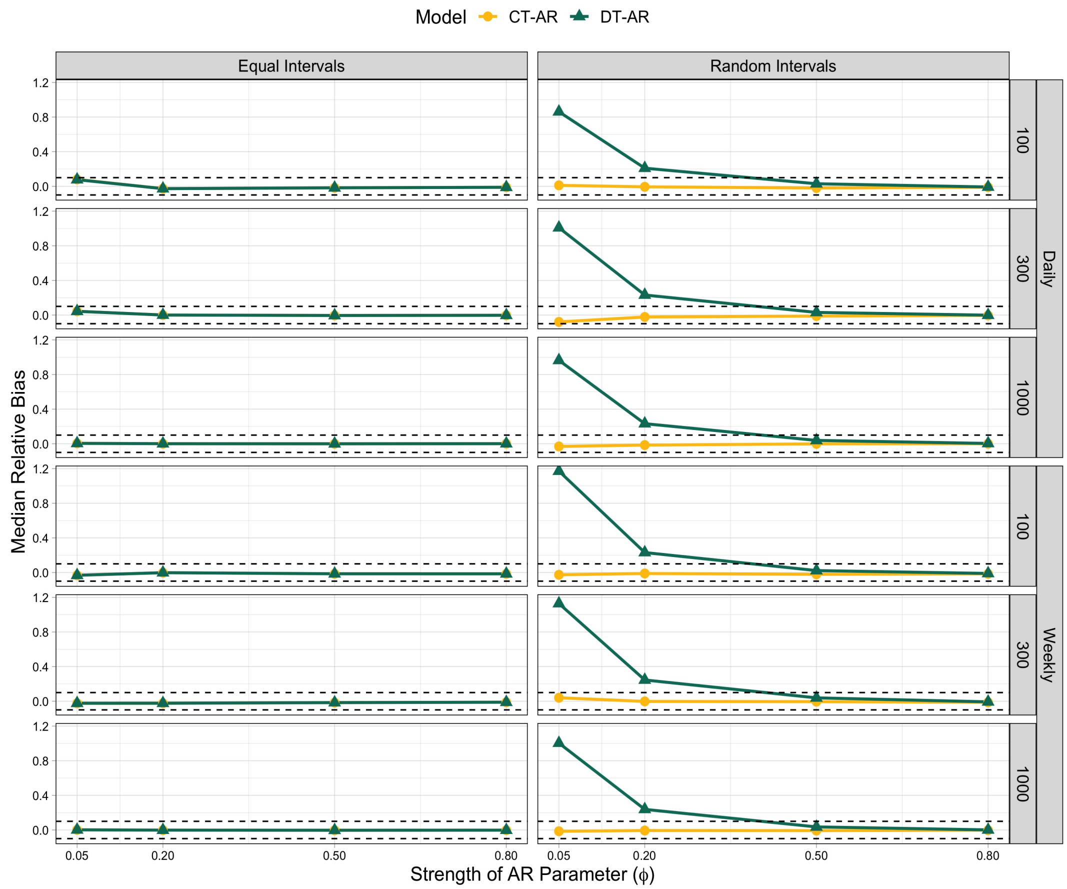

Figure 6

Comparison of relative bias for CT-AR and DT-AR models.

sampling frequency was slower than the true frequency, as for instance weekly and monthly sampling frequency, when the generating frequency was daily.

CT-DT Comparison

In addition to fitting CT-AR models for each of the three sampling frequencies, we also fitted a DT-AR model when the sampling frequency matched the data generating frequency; that is, daily sampling for daily generating dynamics and weekly sampling for weekly generating dynamics. Our aim was to compare the CT-AR and DT-AR models across the type of sampling intervals and the number of observations, and Figure 6 shows this comparison.

When the intervals were equal, the estimates from both CT-AR and DT-AR models were unbiased across all simulation conditions, when the data generating and sampling frequencies are in accordance. This is not surprising, because with invariant intervals between any two consecutive measurements, the data meet the assumption of DT-AR models. In this case, both the DT- and CT- models are correctly specified models. Meanwhile, when the intervals between consecutive measures were random, the CT-AR model was unbiased for all AR strengths, and the DT-AR model was biased for weaker AR values (ϕ = .05, .2), but was unbiased for values of .5 or above. These results are consistent across all sample sizes and the two generating frequencies.

Coverage Rate of 95% Confidence Intervals

We assessed the performance of the three sampling frequencies for the coverage of the population parameter based on two types of 95% confidence intervals. First, we discuss coverage on the CT-AR model auto-effect estimated in our simulation, that is, the confidence intervals around β. This is useful for examining the model's performance on the CT metric. Second, we discuss coverage on the DT metric, where we rescale the CT-AR confidence intervals to DT using the delta transformation and assess the coverage of the ϕ parameters. Since the technique of rescaling and the accuracy of rescaled estimates have been one of our primary points of discussion, this approach represents the scenario where

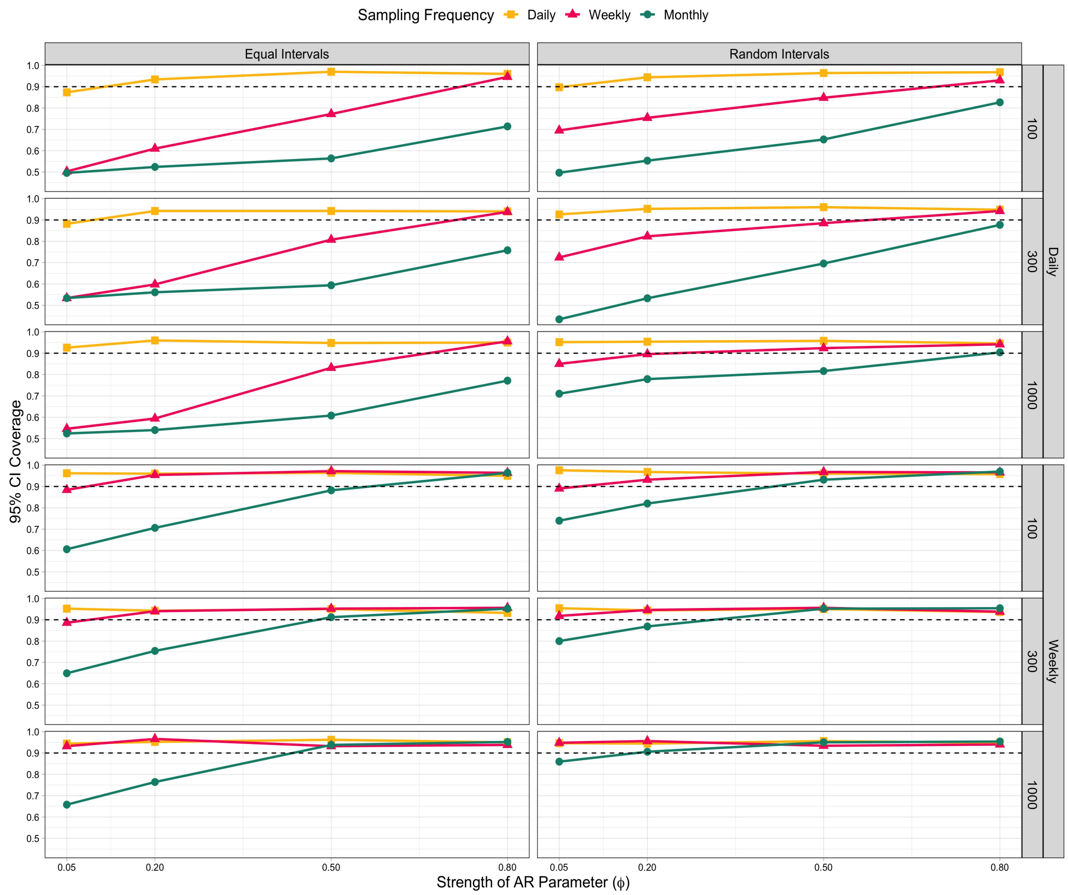

Figure 7

Plot of 95% confidence interval coverage rate of the β parameter in CT metric.

a researcher collects a sample and checks for the sampling properties of their rescaled parameter. We use 90% as the threshold for good coverage (Collins et al., 2001; Enders and Peugh, 2004).

Coverage of β parameter (CT metric)

Figure 7 shows coverage rates for the β parameter. We noticed that as the strength of the AR parameter increased, coverage generally increased for all sampling frequencies. However, this was especially true for sampling frequencies slower than the true frequency, e.g., weekly and monthly sampling frequencies for a daily generating frequency attain good coverage when ϕ is .5 or .8. On the other hand, for the weakest AR value (ϕ = .05), slower sampling frequencies or even sampling frequencies matching the true generating frequency had low coverage. For example, daily sampling frequency showed moderate coverage when the true frequency was also daily in the top three rows of the Figure 7. When the sampling frequency was faster than the generating frequency, coverage was perfect across all conditions, as was the case for data sampled daily when the generating frequency was weekly.

We also observed differences across random and equal sampling intervals. In most cases, with data sampled at random intervals instead of equal intervals, coverage improved notably when the sampling frequency was slower than the generating frequency. For instance, the top 3 rows of Figure 7 show that, when the data were generated at a daily frequency, coverage improved for weekly and monthly sampling frequencies with random intervals, compared to conditions of equal intervals. These results align well with past research showing that random intervals can recover the true dynamics of a given process better than equal time intervals (Voelkle and Oud, 2013). When the sampling frequency was faster than the generating frequency, then the coverage was great across all conditions, regardless of whether the sampling interval was equal or random. This was the case for daily sampled data with weekly generating frequency, as shown in the bottom three rows of Figure 7.

These results are consistent across the three sample sizes

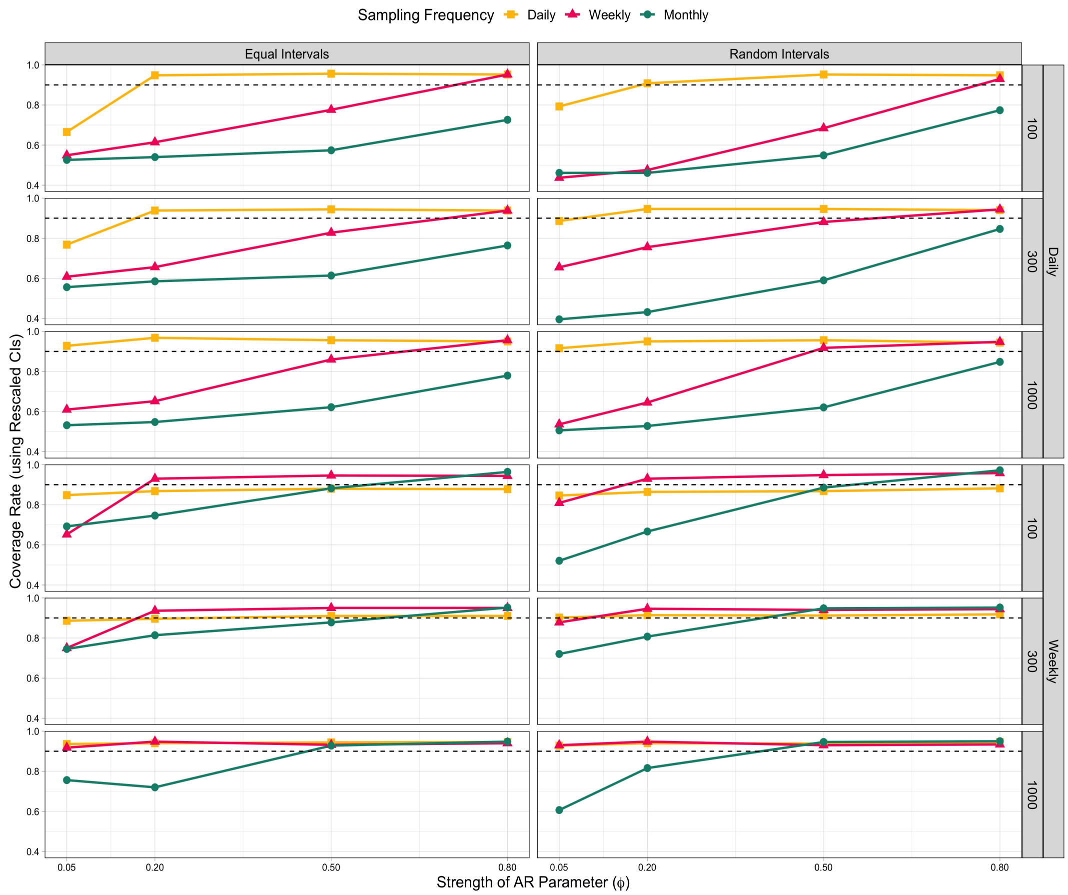

Figure 8

Plot of 95% confidence interval coverage of the ϕ parameter in DT metric.

considered in our simulation. The only difference due to sample size was observed for the monthly sampling frequency, where coverage improved as sample size increased, particularly for weaker AR parameter values. Lastly, the patterns of coverage observed were identical for daily and weekly sampling frequencies when the data had daily and weekly generating frequencies, respectively. This confirms that the model is performs similarly across the two conditions, when the sampling and generating frequency match.

Coverage for ϕ parameter (DT metric)

Figure 8 shows the coverage for the ϕ parameter in the delta-transformed 95% confidence intervals. Results were largely similar between CT and DT metrics, except for the comparison between equal and random intervals. Compared to the results of coverage in the CT metric, the coverage for the DT metric also improved for all sampling frequencies as the value of the AR parameter increased, except when data was sampled monthly at equal intervals under weekly generating frequency for a sample size of 1,000. When the sampling frequency matched the generating frequency, we noticed parameter coverage increased as AR strength increased, with good coverage attained for most AR parameters ($\phi \geq .2$) almost everywhere. This trend was identical for the daily sampling under the daily generating frequency (top three rows in Figure 8) and the weekly sampling under the weekly generating frequency (bottom three rows in Figure 8). When the sampling frequency was faster than the generating frequency, i.e., daily sampling under weekly generating condition, the coverage was great for larger sample sizes (n $\geq$ 300).

When the sampling frequency was slower than the generating frequency, coverage was poor in many conditions. For the daily generating frequency, both weekly and monthly sampling frequency had poor coverage for weaker AR parameters (ϕ = .05, .2). As the AR parameter became stronger (ϕ = .5, .8), weekly showed better coverage at larger sample sizes (n = 1,000), but monthly remained poor throughout. Meanwhile, we found that coverage for the monthly sampling frequency under the weekly generating frequency was poor for weaker AR parameters (ϕ = .05, .2), but improved with stronger AR parameters.

CComparing equal to random sampling intervals, we saw some differences relative to the previous results for the CT coverage. In some cases, random interval sampling showed poor coverage compared to equal interval sampling whereas in other conditions, random sampling showed better coverage than equal. For instance, monthly sampling frequency showed better coverage at equal intervals than random intervals under the weekly generating data. But in many other conditions like weekly sampling under daily generating frequency, random intervals led to better coverage than equal ones.

As for the effect of the number of observations, coverage mostly improved as the number of observations increased from 100 to 300, and to 1,000. Overall, the DT metric showed good coverage of the ϕ parameter, in many cases often similar to the coverage rate of β parameter in the CT metric, but there were some cases, such as with slower sampling frequencies, where coverage was worse. We believe the reason for this is due to the delta transformation, which uses the time interval in its calculation. This makes the DT confidence intervals narrower (in terms of the AR strength they cover) compared to the ones in CT, making it harder for the intervals to cover the parameter of interest.

Discussion

Summary of Results

The purpose of our paper was to determine the extent to which sampling at different frequencies can affect the estimated dynamics of an underlying process. To study this question, we fit a CT-AR model after sampling simulated data at a daily, weekly, or monthly time interval, and evaluated the accuracy of the resulting parameter estimates after rescaling them to the generating time interval. We were particularly interested in investigating whether the ability of CT models to rescale their estimates to any time interval of interest could overcome any mismatch between the generating and sampling frequencies.

The results of our simulation study show that the AR parameter estimates were unbiased and obtained optimal confidence interval coverage rates for two conditions. The first condition was when the sampling frequency was faster than the generating frequency – for example, sampling daily when the generating frequency was weekly. The second condition was when the sampling frequency matched the generating frequency – for example, when the sampling and generating frequencies were both daily.

When the process was sampled at a rate slower than that of the true process and the strength of the AR parameter was low, then the resulting estimates of the AR parameter were biased and confidence interval coverage was poor. However, as the strength of the AR parameter increased, the bias and confidence interval coverage of the slower sampling frequencies improved. These results show that sampling at a frequency that is too slow can result in poor model performance on both the CT and DT metrics, just as it can result in reduced precision for those estimates (Adolf et al., 2021).

We also found an effect of the equality of the sampling intervals, such that sampling at random time intervals generally produced less biased estimates and better coverage rates than equal time intervals, when sampling frequencies were slower than the generating frequency. In fact, the improvement in bias was sometimes to the extent that estimates from samples with random intervals could be unbiased even as the estimates from samples with equal intervals were biased. If the sampling frequency was faster than, or matched, the generating frequency then there was little to no effect of the type of sampling interval, likely because there was already little bias in these situations.

The bias of the AR parameters tended to remain constant even as the number of observations increased. If any differences were present, it tended to be for weaker AR parameters, such that having more observations resulted in less bias. This is in contrast to previous work (Hecht and Zitzmann, 2020), in which 100 observations of a single individual were insufficient for unbiased results, and increasing the number of observations generally improved the performance of the CT-AR model. Our simulation showed that even 100 observations could result in unbiased estimates and satisfactory confidence interval coverage, as long as the sampling frequency matched, or was faster than, the generating frequency. However, we did discover an effect of the number of observations on confidence interval coverage, such that increasing the sample size improved coverage rates.

Finally, when the generating and sampling frequency matched, we also compared the dynamics that were estimated from a CT-AR versus DT-AR model, and our findings echoed previous results in the literature. When the data were sampled at equal time intervals, both models gave the same, unbiased estimates of the dynamics. This is not surprising, since the CT-AR and DT-AR models are equivalent when the time intervals are truly equal (Loossens et al., 2021). When the time intervals were random, however, we found that the CT-AR model resulted in unbiased estimates of the dynamics throughout, while the DT-AR model resulted in biased estimates when the strength of the AR parameter was low. Yet, once the AR parameter was at least .5, the DT-AR model gave unbiased results comparable to that of the CT-AR model. The lack of difference between the DT-AR and CT-AR models at stronger AR values echoes results from Boker et al. (2018), where treating random time intervals as equal gave similar results to a method that corrected for the random time intervals. Our results may be due to how the time intervals were randomly generated – since each measurement occasion was chosen with equal probability from each period, the time intervals were not systematically longer or shorter than in the equal time interval samples, allowing the bias of shorter time intervals to cancel out the bias of longer time intervals.

Considerations for Rescaling CT Parameters

An advantage of CT models that is often emphasized in the literature is that they allow rescaling of the estimated dynamics to any discrete time interval of interest (Boker et al., 2018; Deboeck and Preacher, 2016; de Haan-Rietdijk et al., 2017; Kuiper and Ryan, 2018; Oud and Delsing, 2010; Ryan and Hamaker, 2021; Voelkle et al., 2012; Voelkle and Oud, 2013). Little mention is given in the literature to how the interplay between the sampling frequency and the true dynamics of the process could affect this rescaling, or the limits to rescaling these estimated dynamics. Our simulation study demonstrated that, *under certain conditions*, researchers can rescale the estimates from CT-AR models and still accurately recover the dynamics of the process under study. Researchers should keep these conditions in mind when designing their study.

Based on our findings, we recommend that researchers sample at a frequency that most closely matches that of the process of interest, or at a faster frequency. Sampling at a frequency slower than that of the process typically resulted in biased estimates and poor confidence interval coverage. This is particularly important when the stability of the process is suspected to be low, as sampling at a frequency slower than that of the true process resulted in the worst performance when the strength of the AR parameter was weak (e.g., .05). As the strength of the AR parameter increased, however, the performance across the different sampling frequencies became comparable. Although the recommendation that researchers sample as frequently as possible is not a new one (e.g., Adolph et al., 2008), our results help quantify this recommendation by showing the decrease in bias and increase in confidence interval coverage that occurs as the sampling scheme better matches that of the generating process. In fact, if the researcher manages to sample at a rate faster than that of the true process, then our results show that performance is almost always optimal in the conditions studied here.

The above recommendations for accurately rescaling CT parameters are summarized in Table 2.

Reconsidering Sampling Frequencies as Sampling Ratios

Throughout our paper, we have discussed the generating and sampling frequencies in metrics of time that are familiar and used empirically, such as daily, weekly, and monthly. Yet we can reframe the discussion of our simulation in terms of the ratio of the generating frequency to the sampling frequency. For example, if the true process operated at a daily time interval and was sampled weekly, we could describe this as sampling at a 1:7 ratio; if the true process operated at a weekly interval and was sampled at a monthly interval, that would be a 1:4 ratio.

Reframing the discussion in this way allows us to generalize our results to metrics of time beyond the ones mentioned in this paper, as well as to other ratios more broadly. A ratio of 1:4 could not only mean sampling every month when the process operates on a weekly basis, but sampling every four days when a process operates daily, or every 2 days when a process operates at a half-day interval.

Regardless of the time metric used, the main takeaways of our paper should still be considered when deciding how frequently to sample. Researchers should strive for a ratio that allows them to be as close as possible to what is likely to be the true frequency, as the ability to recover the dynamic parameter improves the closer the ratio is to 1. For example, even though taking a monthly sample rarely resulted in good performance for either data generating frequency, the performance of the monthly sample was better when the true process operated at a weekly interval (1:4 ratio) than a daily interval (1:28 ratio). Thus, if a researcher designing a study was choosing between two different sampling frequencies, our results suggest that sampling close to the frequency of interest is the best choice in terms of estimation accuracy.

Given our repeated emphasis on sampling at a frequency that is as close as possible to, or faster than, the data generating frequency, researchers may wonder how to determine the generating frequency of their process. In this sense, our results do not escape the "chicken and egg" issue presented earlier: researchers need to first determine the frequency of interest for their process as well as the hypothesized strength of the AR parameter before they can apply the results of our simulation study to decide the appropriate sampling frequency. We believe that this decision would ideally be based on theoretical knowledge about the process of interest and its presumed dynamics.

However, in the absence of such theory, some approaches exist for estimating the frequency that best reflects the process from collected data. As mentioned previously, the method proposed by Adolf et al. (2021) can allow researchers to use past values of the auto-effect to estimate a sampling interval that would result in the highest estimation reliability. Other techniques, such as the variance decomposition approach discussed by Shiyko and Ram (2011), could allow researchers to roughly determine how often the process of interest fluctuates based on collected data. Variance decomposition can provide information, for instance, about whether the process varies mostly between days or mostly within days, and therefore whether multiple measurements a day would be necessary to capture the process.

Researcher's Question	Recommendation
What frequency should I sample my process at?	We recommend researchers to sample as fast as the frequency of interest for a process and if possible, use a faster sampling frequency based on the feasibility of data collection.
How does my hypothesis about the strength of the AR parameter affect my sampling frequency?	A researcher's hypothesis about the strength of the AR parameter greatly determines the flexibility they will have with their sampling frequencies for collecting data. A smaller AR parameter (especially below .5) requires a closer match between the frequency of interest and the sampling frequency.
Should I use equally-spaced or unequally-spaced time intervals during data collection?	Instead of using equidistant time intervals between measurement occasions, researchers should aim to use random intervals for improved accuracy and coverage of their rescaled parameter estimates from a CT-AR model.
How many observations do I need, if I am collecting data on a single individual?	If the stability of the process is hypothesized to be high (higher AR parameter), then researchers can collect as few as 100 observations. However, less stable processes would require more observations for good model performance. Increasing the number of observations only tends to make a difference when the stability of the process is low. Otherwise, even 100 observations is sufficient for unbiased estimates and good coverage rates.
In what scenarios is it safe to use a DT-AR model instead of a CT-AR model?	The CT-AR model is a safer option across the different sampling conditions. However, if a researcher can collect data with an equally-spaced sampling scheme, then the DT-AR model performs as well as the CT-AR model. If a researcher collects data with an unequally-spaced sampling scheme, then the DT-AR model should only be used when the process is hypothesized to be highly stable (stronger AR parameter).

Table 2

Recommendations for researchers on sampling decisions for CT modeling

Methodological Considerations

In the conditions used in our study, the primary difference between our data generating models was the strength of the autoregressive parameter, which when combined with the chosen data generating frequency, translated into differences in the CT auto-effect. However, estimation and inference of both CT-AR and DT-AR models depend not only on the values of the auto-effect or autoregressive coefficients, but also on the innovation variance. In our simulation, the innovation variance was determined by the strength of the autoregressive effect and the process variance which we kept fixed to a small value). Of course, another alternative would have been to manipulate the process variance as another factor in our simulation.

However, we chose to only manipulate the autoregressive effect for a number of reasons. First, we believe that keeping the variance of the process fixed to a small value helps avoid contamination of the effects of interest. Second, the main benefit of manipulating the innovation variances directly - for a fixed strength of the autoregressive effect - would be to change the signal-to-noise ratio of the process. We were able to indirectly include a range of innovation variances in our simulation, as the calculation of the innovation variance depends on the strength of the autoregressive effect and the data generating frequency. Hence, although indirectly, we were able to study the performance of the CT-AR model across a range of signal-to-noise ratios. Furthermore, we believe that increasing the innovation variance for a fixed autoregressive strength would not change the findings and the main recommendations offered here. Instead, we predict that an increased innovation variance would require either more observations or stronger underlying autoregressive effects for good model performance.

Another methodological consideration is the discrepancy in results for confidence interval coverage in the CT versus DT metric. As mentioned in the Results, we believe most of this discrepancy comes from the difference in the standard errors across the two metrics. More specifically, the standard errors used in the DT confidence intervals are smaller than those in the CT confidence intervals. This is illustrated in Figure D1 of the Appendix. We believe this is due to the use of the time interval in the delta method calculation. The smaller standard errors then result in narrower confidence intervals, as can be seen in Figure D2, where the confidence intervals obtained from the delta method are contrasted with those from CT for the same point estimate. Although there are other methods available to rescale confidence intervals to DT (e.g., exponentiating the CT intervals), we believe the delta method is the most theoretically appropriate because it maintains the non-linear relationship between the CT and DT metrics. However, we do not know how this nonlinear transformation of the standard errors from CT to DT metric compares to other approaches for rescaling. This issue should be investigated in future research.

Limitations and Future Directions

As with every simulation study, the results and recommendations we offer are limited to the specific conditions that we examined. In these analyses, we used a univariate autoregressive model of order 1 with manifest variables. We chose to limit our attention to this simple scenario for two reasons. First, even though many researchers are increasingly focused on exploring the dynamics of multiple variables, analyses examining time series of a single manifest variable are still common in psychological research (e.g., Coppersmith et al., 2023; De Haan-Rietdijk et al., 2016; Koval & Kuppens, 2012). Second, we chose to focus on the simplest possible case for our simulation to represent a best-case scenario. This seemingly simple case reduces potential complexity of parameters, facilitates the interpretation of the results, and can better demonstrate the impact of the mismatch between the data generating and sampling frequency. Of course, it is possible to increase the complexity of the generating model in a number of ways – either by extending to a multivariate process with the vector autoregressive model, increasing the order of the model to a lag of 2 or more, or including latent variables to consider a factorial structure and account for measurement error (Chow, 2019; Driver and Voelkle, 2018; Oud and Delsing, 2010).

Although this complexity may better represent the process of interest in psychological research, such complex models are likely to introduce further difficulties into the estimation process, as well as the recommendations we could make to researchers. For example, in the simplest multivariate case involving two variables, each variable could have its own data generating frequency. Evaluating the effect of the mismatch between the sampling and data generating frequency would then be significantly more complicated. For example, the data generating frequency could either match, be too fast, or be too slow for one variable, and then independently match, be too fast, or be too slow for the second variable. Any recommendations regarding the sampling frequency would then need to take all these possible scenarios into account. Future work should investigate how introducing these complexities affects the recommendations offered above and expand those recommendations to situations researchers are likely to encounter.

A second limitation of our model is that we only generated data from a single individual. A future extension of our work could involve incorporating information from multiple individuals using, for example, a multilevel approach (Driver and Voelkle, 2018). Psychological research often involves analyzing the data of multiple individuals, so generating data and estimating models within a multilevel modeling framework represents a common research scenario. With multiple individuals in a single sample, information could be borrowed from each person during the estimation process to potentially improve the estimation of the dynamic parameters. Although having multiple individuals could compensate for having fewer timepoints per individual (Hecht and Zitzmann, 2020), we believe it is unlikely to make up for the consequences of a mismatch between the sampling and generating frequency.

Conclusion

Our simulation study shows that researchers can rescale estimates from the CT-AR model and accurately estimate the dynamics in some circumstances. However, the choice of sampling frequency plays an important role in such rescaling. Our results demonstrated that researchers can generally rescale the estimated auto-effects to the true frequency as long as the sampling frequency is not slower than the data generating frequency, but in other circumstances the accuracy as a result of rescaling depends on a delicate interplay among expected AR strength, differences between the true and sampling frequencies, and the number of observations. This requires researchers to have some knowledge of the frequency at which the process of interest operates. Therefore, even though CT-AR models can theoretically be rescaled to infer about the process studied on a time scale different than the one used in sampling, researchers utilizing CT models cannot escape the need for theory when designing their sampling schedule.

Acknowledgments

We thank Ginette Lafit, Janne K Adolf, and Eva Ceulemans for reading a preprint of the paper and providing us with insightful comments.

References

Adolf, J. K., Loossens, T., Tuerlinckx, F., & Ceulemans, E. (2021). Optimal sampling rates for reliable continuous-time first-order autoregressive and vector autoregressive modeling. [ISBN: 1939-1463 Publisher: American Psychological Association]. *Psychological Methods.*

Adolph, K. E., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological Review*, *115*(3), 527–543. https://doi.org/10.1037/0033-295X.115.3.527

Boker, S. M. (2002). Consequences of Continuity: The Hunt for Intrinsic Properties within Parameters of Dynamics in Psychological Processes. *Multivariate Behavioral Research*, *37*(3), 405–422. https://doi.org/10.1207/S15327906MBR3703_5

Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., Estabrook, R., Bates, T. C., & Mehta, P. (2022). OpenMx User Guide, 195.

Boker, S. M., & Nesselroade, J. R. (2002). A Method for Modeling the Intrinsic Dynamics of Intraindividual Variability: Recovering the Parameters of Simulated Oscillators in Multi-Wave Panel Data. *Multivariate Behavioral Research*, *37*(1), 127–160. https://doi.org/10.1207/S15327906MBR3701_06

Boker, S. M., Tiberio, S. S., & Moulder, R. G. (2018). Robustness of Time Delay Embedding to Sampling Interval Misspecification. In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous Time Modeling in the Behavioral and Related Sciences* (pp. 239–258). Springer International Publishing. https://doi.org/10.1007/978-3-319-77219-6_10

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived [ISBN: 0066-4308 Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA]. *Annual review of psychology*, *54*(1), 579–616.

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford press.

Bos, F. (2021). *Ecological momentary assessment as a clinical tool in psychiatry: Promises, pitfalls, and possibilities* (Doctoral dissertation). University of Groningen.

Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average time series models. [ISBN: 0805846085 Publisher: Lawrence Erlbaum Associates Publishers].

Castro-Schilo, L., & Ferrer, E. (2013). Comparison of nomothetic versus idiographic-oriented methods for making predictions about distal outcomes from time series data [ISBN: 0027-3171 Publisher: Taylor & Francis]. *Multivariate Behavioral Research*, *48*(2), 175–207.

Chen, M., Chow, S.-M., & Hunter, M. D. (2018). Stochastic Differential Equation Models with Time-Varying Parameters. In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous Time Modeling in the Behavioral and Related Sciences* (pp. 205–238). Springer International Publishing. https://doi.org/10.1007/978-3-319-77219-6_9

Chen, M., & Ferrer, E. (2022). Assessing dynamical associations in dyadic interactions across multiple time scales via a bayesian hierarchical vector autoregressive model. *Journal of Social and Personal Relationships*, 02654075221137865.

Chow, S.-M. (2019). Practical tools and guidelines for exploring and fitting linear and nonlinear dynamical systems models [ISBN: 0027-3171 Publisher: Taylor & Francis]. *Multivariate behavioral research*, *54*(5), 690–718.

Chow, S.-M., Ho, M.-h. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalence and Differences Between Structural Equation Modeling and State-Space Modeling Techniques. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(2), 303–332. https://doi.org/10.1080/10705511003661553

Collins, L. M., & Graham, J. W. (2002). The effect of the timing and spacing of observations in longitudinal studies of tobacco and other drug use: Temporal design considerations [ISBN: 0376-8716 Publisher: Elsevier]. *Drug and Alcohol Dependence*, *68*, 85–96.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures [Place: US Publisher: American Psychological Association]. *Psychological Methods*, *6*(4), 330–351. https://doi.org/10.1037/1082-989X.6.4.330

Coppersmith, D. D., Ryan, O., Fortgang, R. G., Millner, A. J., Kleiman, E. M., & Nock, M. K. (2023). Mapping the timescale of suicidal thinking. *Proceedings of the National Academy of Sciences*, *120*(17), e2215434120.

De Haan-Rietdijk, S., Gottman, J. M., Bergeman, C. S., & Hamaker, E. L. (2016). Get over it! a multilevel threshold autoregressive model for state-dependent affect regulation. *Psychometrika*, *81*, 217–241.

Deboeck, P. R., & Preacher, K. J. (2016). No Need to be Discrete: A Method for Continuous Time Mediation Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 61–75. https://doi.org/10.1080/10705511.2014.973960

de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete- vs. Continuous-Time Modeling of Unequally Spaced Experience Sampling Method Data. *Frontiers in Psychology*, *8*, 19.

Dormann, C., & Griffin, M. A. (2015). Optimal time lags in panel studies. [ISBN: 1939-1463 Publisher: American Psychological Association]. *Psychological methods*, *20*(4), 489.

Dormann, C., & van de Ven, B. (2014). Timing in Methods for Studying Psychosocial Factors at Work. In M. F. Dollard, A. Shimazu, R. Bin Nordin, P. Brough, &

M. R. Tuckey (Eds.), *Psychosocial Factors at Work in the Asia Pacific* (pp. 89–116). Springer Netherlands. https://doi.org/10.1007/978-94-017-8975-2_4

Driver, C. C., & Voelkle, M. C. (2018). Hierarchical Bayesian continuous time dynamic modeling. [ISBN: 1939-1463 Publisher: American Psychological Association]. *Psychological Methods*, *23*(4), 774.

du Toit, S. H. C., & Browne, M. W. (2007). Structural Equation Modeling of Multivariate Time Series. *Multivariate Behavioral Research*, *42*(1), 67–101. https://doi.org/10.1080/00273170701340953

Enders, C. K., & Peugh, J. L. (2004). Using an EM Covariance Matrix to Estimate Structural Equation Models With Missing Data: Choosing an Adjusted Sample Size to Improve the Accuracy of Inferences [Publisher: Routledge _eprint: https://doi.org/10.1207/S15328007SEM1101_1]. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(1), 1–19. https://doi.org/10.1207/S15328007SEM1101_1

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. [ISBN: 1939-1463 Publisher: American Psychological Association]. *Psychological methods*, *9*(4), 466.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models [ISBN: 0009-3920 Publisher: JSTOR]. *Child development*, 80–92.

Hamerle, A., Nagl, W., & Singer, H. (1991). Problems with the estimation of stochastic differential equations using structural equations models [ISBN: 0022-250X Publisher: Taylor & Francis]. *Journal of Mathematical Sociology*, *16*(3), 201–220.

Hecht, M., & Zitzmann, S. (2020). Sample size recommendations for continuous-time models: Compensating shorter time series with larger numbers of persons and vice versa [ISBN: 1070-5511 Publisher: Taylor & Francis]. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–8.

Hunter, M. D. (2018). State Space Modeling in an Open Source, Modular, Structural Equation Modeling Environment, 19.

Jacobson, N. C., Chow, S.-M., & Newman, M. G. (2019). The Differential Time-Varying Effect Model (DTVEM): A tool for diagnosing and modeling time lags in intensive longitudinal data. *Behavior Research Methods*, *51*(1), 295–315. https://doi.org/10.3758/s13428-018-1101-0

Janssens, K. A., Bos, E. H., Rosmalen, J. G., Wichers, M. C., & Riese, H. (2018). A qualitative approach to guide choices for designing a diary study [ISBN: 1471-2288 Publisher: Springer]. *BMC medical research methodology*, *18*(1), 1–12.

Koval, P., & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion*, *12*(2), 256.

Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval [ISBN: 1070-5511 Publisher: Taylor & Francis]. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(5), 809–823.

Lenderking, W. R., Hu, M., Tennen, H., Cappelleri, J. C., Petrie, C. D., & Rush, A. J. (2008). Daily process methodology for measuring earlier antidepressant response [ISBN: 1551-7144 Publisher: Elsevier]. *Contemporary Clinical Trials*, *29*(6), 867–877.

Liu, S., Ou, L., & Ferrer, E. (2021). Dynamic mixture modeling with dynr. *Multivariate Behavioral Research*, *56*(6), 941–955.

Loossens, T., Tuerlinckx, F., & Verdonck, S. (2021). A comparison of continuous and discrete time modeling of affective processes in terms of predictive accuracy. *Scientific Reports*, *11*(1), 6218. https://doi.org/10.1038/s41598-021-85320-4

Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.

Nyquist, H. (1928). Certain topics in telegraph transmission theory [ISBN: 0096-3860 Publisher: IEEE]. *Transactions of the American Institute of Electrical Engineers*, *47*(2), 617–644.

O'Laughlin, K. D., Liu, S., & Ferrer, E. (2020). Use of composites in analysis of individual time series: Implications for person-specific dynamic parameters [ISBN: 0027-3171 Publisher: Taylor & Francis]. *Multivariate behavioral research*, 1–18.

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods*, *16*(4), 468–490. https://doi.org/10.1037/a0024375

Oud, J. H. L., van Leeuwe, J., & Jansen, R. (1993). Kalman Filtering in discrete and continuous time based on longitudinal LISREL models. *Advances in longitudinal and multivariate analysis in the behavioral sciences, ITS, Nijmegen, Netherlands*, 3–26.

Oud, J. H., & Delsing, M. J. (2010). Continuous time modeling of panel data by means of SEM. *Longitudinal research with latent variables* (pp. 201–244). Springer.

Pritikin, J. N., Hunter, M. D., & Boker, S. M. (2015). Modular open-source software for item factor analysis [ISBN: 0013-1644 Publisher: Sage Publications Sage CA: Los Angeles, CA]. *Educational and psychological measurement*, *75*(3), 458–474.

Ryan, O., & Hamaker, E. L. (2021). Time to intervene: A continuous-time approach to network analysis and centrality [ISBN: 1860-0980 Publisher: Springer]. *Psychometrika*, 1–39.

Shannon, C. E. (1948). A mathematical theory of communication [ISBN: 0005-8580 Publisher: Nokia Bell Labs]. *The Bell system technical journal*, *27*(3), 379–423.

Shiyko, M. P., & Ram, N. (2011). Conceptualizing and estimating process speed in studies employing ecological momentary assessment designs: A multilevel variance decomposition approach. *Multivariate Behavioral Research*, *46*(6), 875–899.

Voelkle, M. C., Gische, C., Driver, C. C., & Lindenberger, U. (2018). The Role of Time in the Quest for Understanding Psychological Mechanisms. *Multivariate Behavioral Research*, *53*(6), 782–805. https://doi.org/10.1080/00273171.2018.1496813

Voelkle, M. C., & Oud, J. H. L. (2015). Relating Latent Change Score and Continuous Time Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(3), 366–381. https://doi.org/10.1080/10705511.2014.935918

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*(2), 176–192. https://doi.org/10.1037/a0027543

Voelkle, M. C., & Oud, J. H. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes [ISBN: 0007-1102 Publisher: Wiley Online Library]. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 103–126.

Wichers, M., Riese, H., Hodges, T. M., Snippe, E., & Bos, F. (2021). A narrative review of network studies in depression: What different methodological approaches tell us about depression [ISBN: 1664-0640 Publisher: Frontiers]. *Frontiers in psychiatry*, 1865.

Appendix A
Example of Delta Method

In general, the delta method is used to calculate the standard error for a nonlinear transformation of a parameter or set of parameters. If the transformation, $f(\theta)$, involves only a single parameter theta, then:

$$var(f(\theta) \ = \ var(\theta)\ .\ (\frac{\partial f}{\partial \theta})^2$$

Here, we have defined in Equation 4:

$$f(\beta) \ = \ e^{\beta\ .\ \Delta t}$$

Thus, we can define the variance of the ϕ parameter as:

$$var(\phi_{\Delta t}) \ = \ var(\beta)\ .\ (\Delta t\ .\ e^{\beta\ .\ \Delta t})^2$$

For example, suppose the true generating frequency and sampling frequency were both daily, the estimated CT auto-effect was -2.37, and its standard error was 1.06. Then we could calculate the 95% confidence interval for the phi parameter as:

$$e^{-2.37} \ \ 1.96\ .\ (1.06 \ \cdot \ 1 \ \cdot \ e^{-2.37})$$

Appendix B
Plot of Convergence

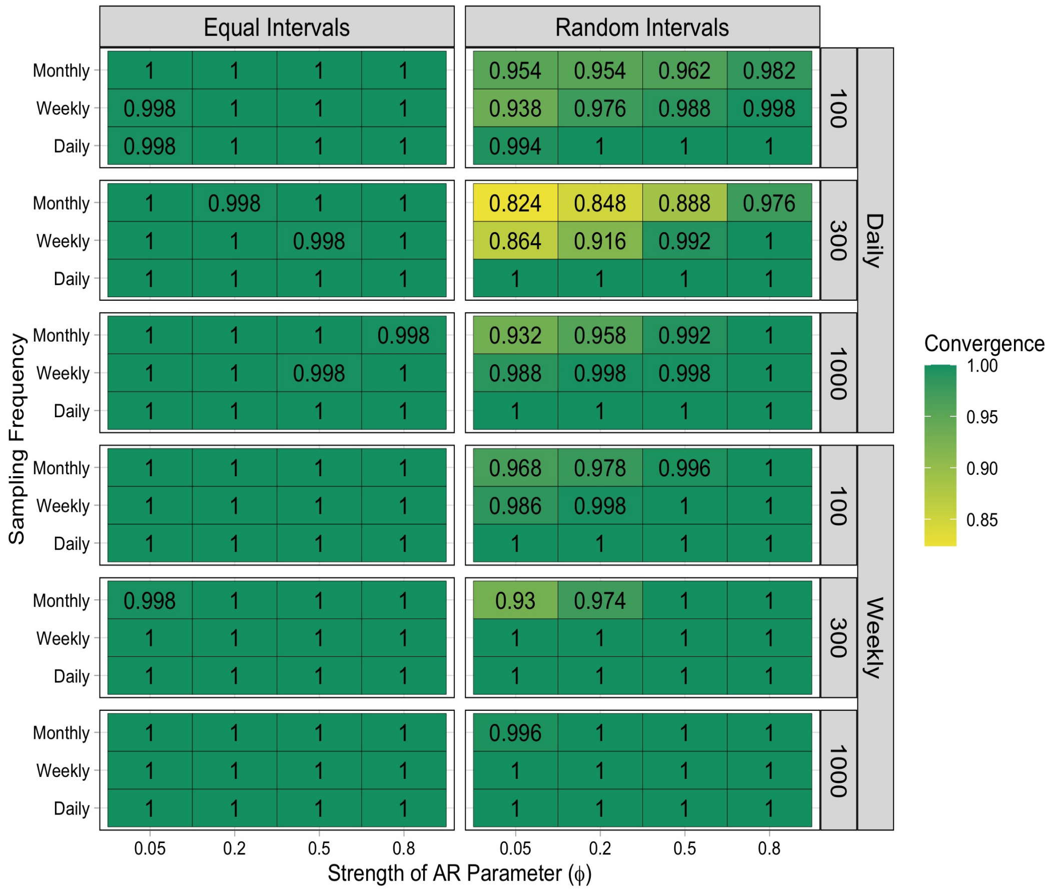

Figure B1

Convergence rate for the CT-AR models (out of 500 replications for each condition)

Appendix C
Plots for Bias

Appendix D

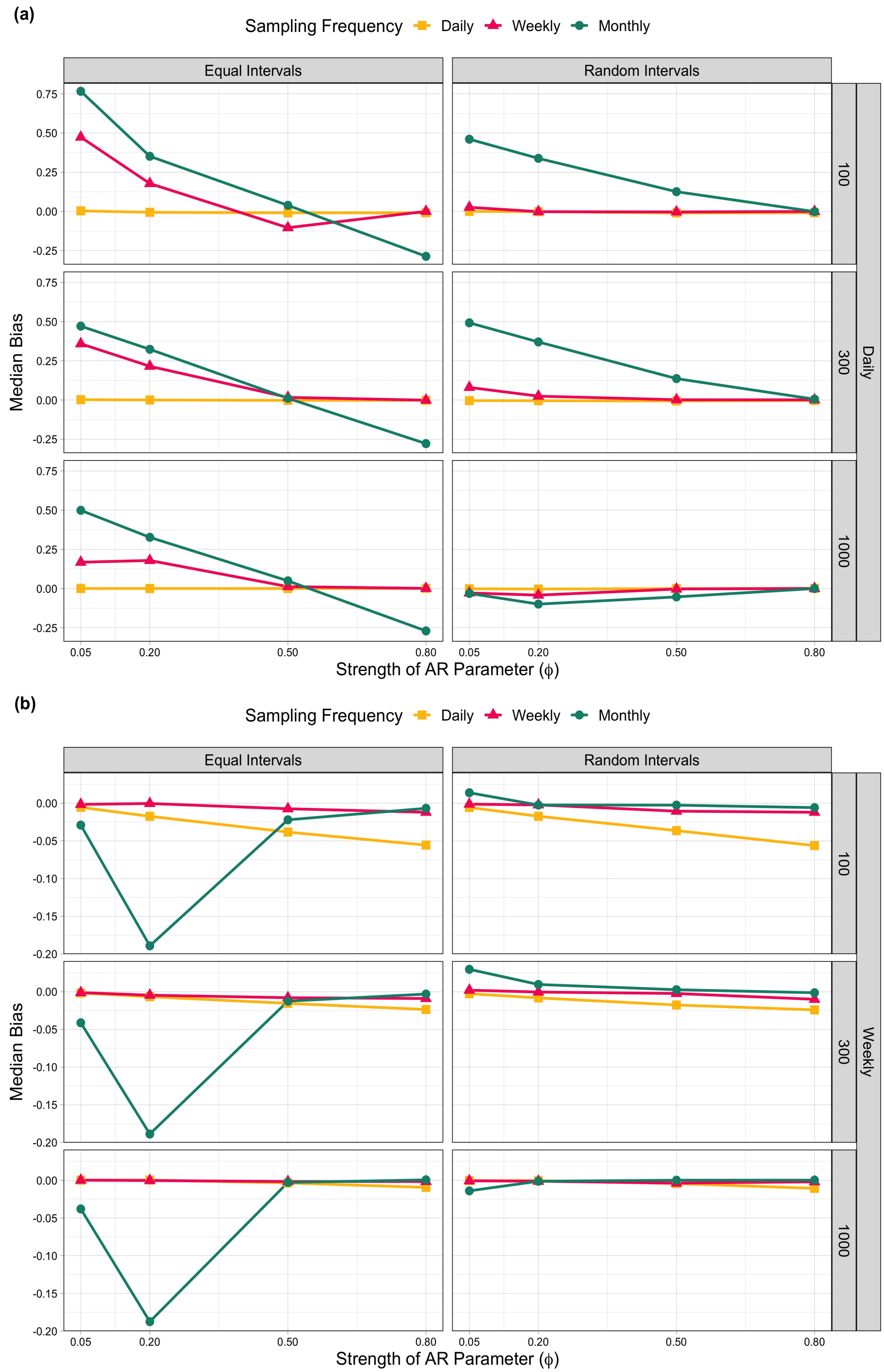

Figure C1

Median Bias plots for daily and weekly generating frequencies.

Note. ((a) Plot of bias for the daily generating frequency; (b) Plot of bias for the weekly generating frequency.

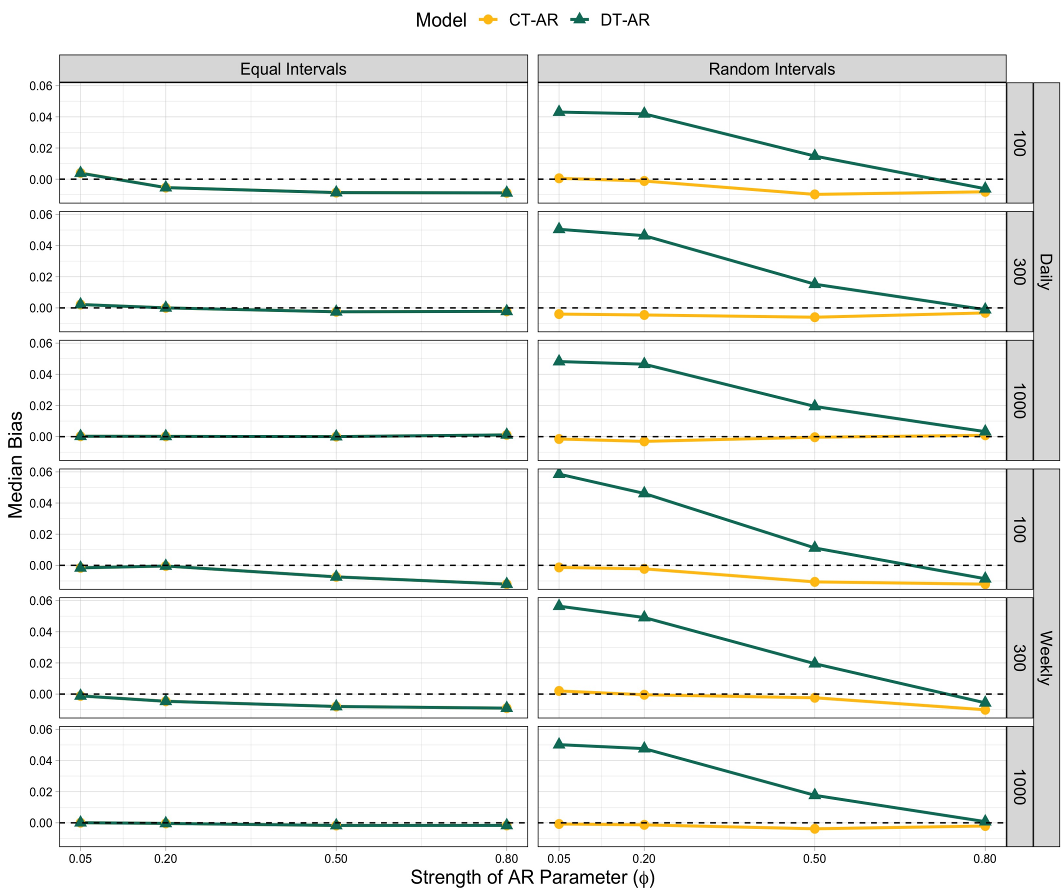

Figure C2

Plot of median bias for comparison of CT-AR and DT-AR models.

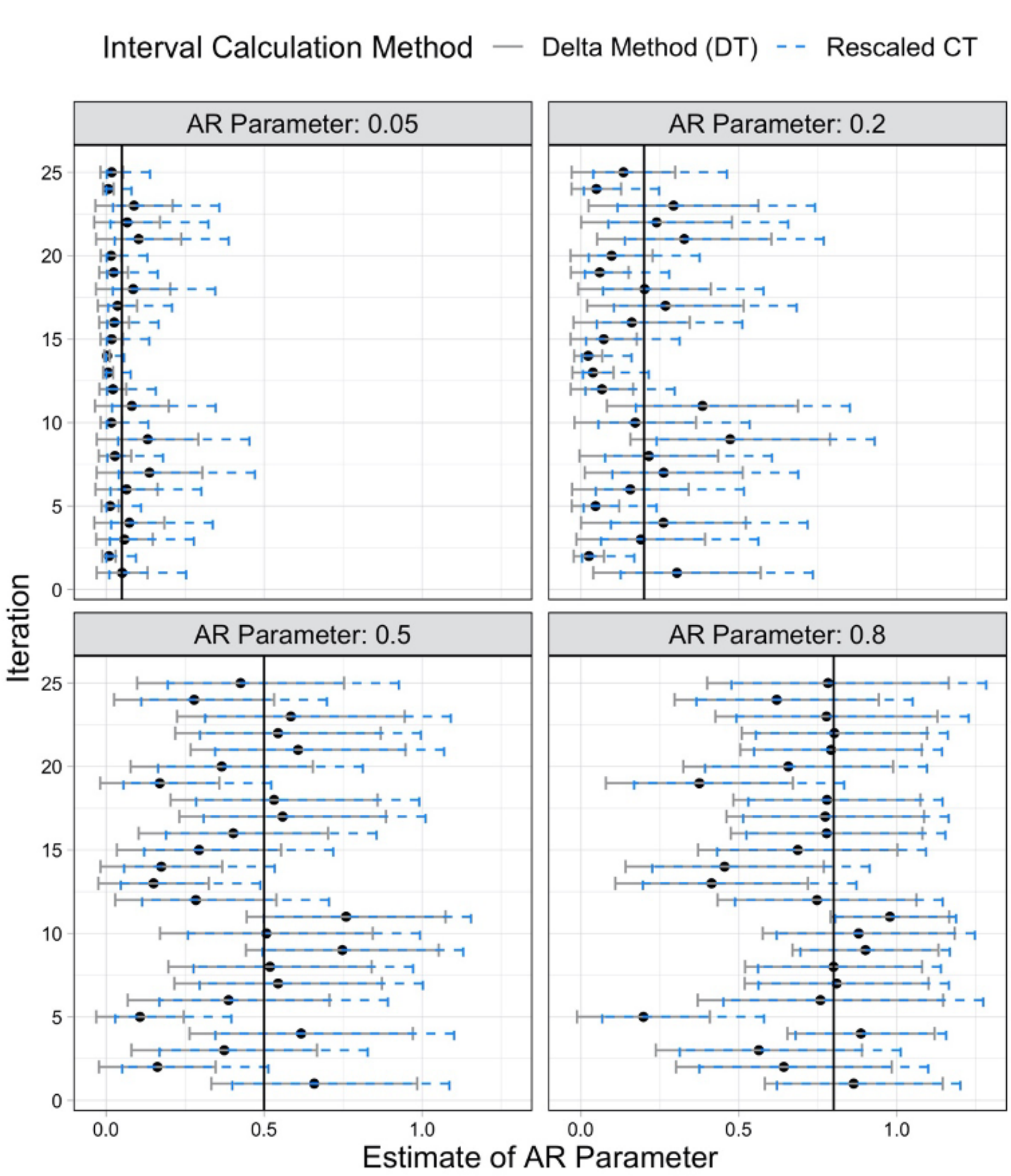

Figure D1

Illustration of the confidence intervals obtained from the delta method (DT metric) versus CT

Note. All estimates are on the DT metric, to make comparisons easier. To transform the confidence intervals from the CT metric to the DT metric, we either applied the delta method transformation, or directly applied Equation 4 to the bounds of the CT confidence intervals. The graph shown here includes 25 random iterations from a single condition where the data generating frequency was weekly, the process was sampled daily, and there were 100 observations. These findings are overall consistent across other conditions.

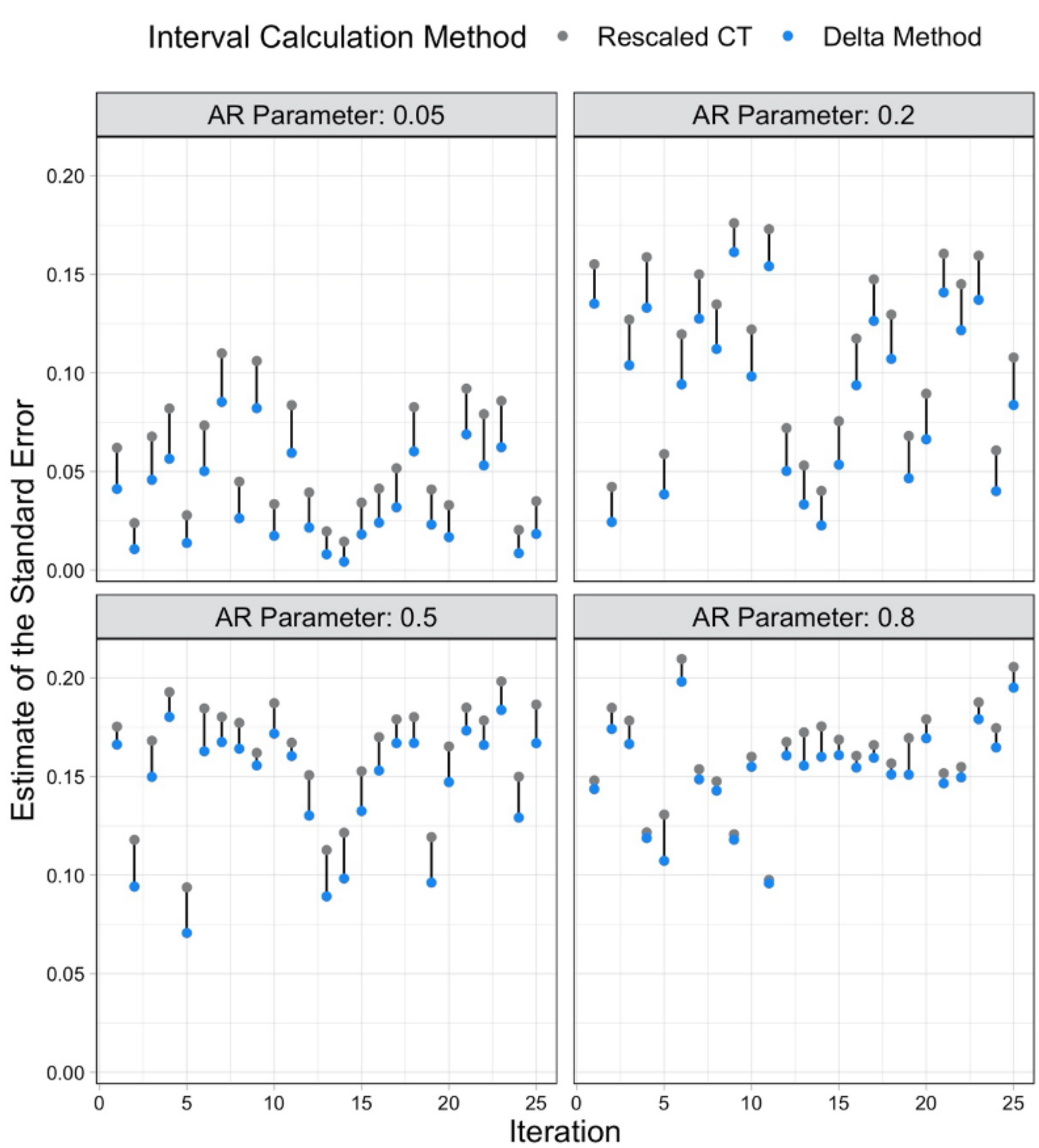

Figure D2

Difference in standard error estimates between the delta method (DT metric) and CT

Note. To ensure that comparisons were done on the same metric, all SEs are on the DT metric. This was done either by applying the delta method, or by the more "direct" approximation of dividing the length of the Rescaled CT confidence intervals (see Figure D1) by 2 * 1.96. The graph shown here includes 25 random iterations from a single condition where the data generating frequency was weekly, the process was sampled daily, and there were 100 observations. These findings are overall consistent across other conditions.