ARTICLE TEMPLATE

Statistical Inference for Lindley Random Walks with Correlated Increments

John Grant[a], Jiajie Kong[b], Xin Liu[a], Robert Lund [b], and Jonathan Woody[c]

[a]School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC, USA; [b]Department of Statistics, University of California, Santa Cruz, USA; [c]Department of Mathematics and Statistics, Mississippi State University, Mississippi, MS, USA

ARTICLE HISTORY
Compiled February 14, 2024

ABSTRACT
This paper studies statistical inference for a Lindley random walk model when the increment process driving the walk is strictly stationary. Lindley random walks govern customer waiting times in many queueing models and several natural and business processes, including snow depths, frozen soil depths, inventory quantities, etc. The probabilistic properties of a Lindley walk with time-correlated stationary changes are first reviewed. We provide a streamlined argument that the process has a proper limiting distribution when the mean of the incremental changes is negative, and that the Lindley process is strictly stationary when starting from this stationary distribution. Next, the Markov characteristics of the process are explored when the change process has a Markov structure of first or higher order. A derivation of the model's likelihood is given when the change process is a Gaussian autoregressive time series. An efficient particle filtering method of evaluating and optimizing the likelihood is then devised and studied via simulation.

KEYWORDS
Autoregression, Coupling, Particle Filtering, Storage Model, Statistical Inference.

1. Introduction

A Lindley random walk, also known as a Lindley recursion or a storage model, is a mass balance equation that has classically arisen in inventory and single server queueing setups [1,2]. A Lindley random walk can also be used to describe some environmental process having a "hard boundary at zero" (described further below), including snow depths, streamflows, frozen soil depths, and ice thicknesses. See [3–5], and [6] for Lindley random walk applications to snow depths and other cryospheric quantities.

A Lindley random walk $\{X_n\}_{n=0}^{\infty}$ obeys the recursion

$$X_n = \max(X_{n-1} + C_n, 0), \quad n \geq 1, \tag{1}$$

starting from some initial level X_0. The sequence $\{C_n\}_{n=1}^{\infty}$ is called the change process here.

CONTACT John Grant. Email: jwgrant@clemson.edu

For examples of where a Lindley random walk arises, consider inventory in a warehouse. Let X_n denote the number of items in the warehouse of a particular type on the opening of day n, N_n the number of new items that are delivered to restock the warehouse during day n, and O_n the number of customer orders that are shipped at the end of day n. Then

$$X_{n+1} = \max\{X_n + N_n - O_n, 0\}$$

is the number of items in the warehouse on the morning of day $n+1$, which is (1) with $C_{n+1} = N_n - O_n$. The maximum at zero enters because one cannot deliver more items than what is currently in stock (we have not specified what to do with "lost" orders that cannot be filled due to lack of content). Another mass balance with Lindley structure arises in single server queueing applications. Here, the virtual waiting time, for example, of the nth customer (the time the customer waits until their service begins), denoted by W_n, obeys

$$W_n = \max(W_{n-1} + S_{n-1} - I_n, 0), \tag{2}$$

where I_n is the interarrival time between the $(n-1)$st and nth arriving customers and S_n is the time taken to serve the nth customer. The classical assumptions in queueing theory are that $\{I_n\}_{n=1}^{\infty}$ and $\{S_n\}_{n=0}^{\infty}$ are each independent and identically distributed (IID) non-negative random sequences (and also independent of each other). In this case, $C_n = S_{n-1} - I_n$ is also IID.

The Lindley model in (1) can also be viewed as a censored time series. Censored time series inference has seen significant recent development in the literature [7–12]. Indeed, given a sample $\{X_n\}_{n=1}^{N}$ of the process in 1, one can recover C_n exactly unless $C_n < -X_{n-1}$ (the censoring mechanism is somewhat complex as it depends on past process values).

In environmental Lindley random walk applications, [3] models daily snow depths at day n as the snow depth on the ground yesterday, plus any new snowfall, minus any meltoff or compaction since yesterday. To prevent non-identifiability, one amalgamates any new snow, meltoff, and compaction into a single change variable C_n and the model in (1) arises. Since daily weather is highly correlated, one needs a temporally correlated $\{C_n\}$ for realism. In queueing applications, the server may be inclined to work faster when there are more customers in the queue, inducing correlation in $\{C_n\}$. In the context of the warehouse storage model above, the typical assumption in the literature is that $\{N_n - O_n\}$ is IID. In practice, correlation might arise in various ways. For example, one is more likely to order more stock if the warehouse is empty and less stock if the warehouse is full, thus introducing dependence between $\{O_n\}$ and $\{X_n\}$.

As storage levels cannot be negative, a Lindley random walk has a so-called hard boundary at zero. One could transform Lindley random walk data prior to modeling, analyzing say $\ln(X_n + \epsilon)$ for some small $\epsilon > 0$ as a quantity taking on any real value. While this is not necessarily a bad idea, such a scheme would estimate the probability of an empty store, a key quantity for administrators, as zero under this scheme. As such, it is preferable to model the storage level directly (without transformation).

The majority of research on Lindley random walks has been done in queueing contexts that assume IID $\{C_n\}$. A classic reference here is [2]. In the correlated case, stability of waiting times for stationary $\{S_{n-1}, I_n\}$ were examined in [13,14]. Queues built from a first-order autoregressive processes were studied in [15] and [16]. Dependence can arise in a variety of queueing contexts, including batching and multiple

customer classes; see [17–22].

The purpose of this paper is to examine estimation issues in the Lindley walk when $\{C_n\}$ is strictly stationary, and more specifically, a pth order autoregression. We provide a streamlined proof of stochastic stability that uses stochastic monotonicity and coupling techniques for technical efficiency. When $\{C_n\}$ is a pth order autoregression, the Markov properties (and lack thereof) of the Lindley walk are established. The model's statistical likelihood is then derived and novel particle filtering inference techniques are used to estimate model parameters via maximum likelihood methods.

The rest of this paper proceeds as follows. The next section establishes/reviews ergodicity and stationarity properties of $\{X_n\}$ when $\{C_n\}$ is strictly stationary. Section 3 then establishes several Markov structures for the process, even though $\{X_n\}$ is not a Markov chain itself. The final two sections move to statistical inference issues for the walk. In particular, Section 4 uses the Markov structure established in Section 3 to derive the model's likelihood function when $\{C_n\}$ is a pth order Gaussian autoregression. As the resulting likelihood involves some unwieldy multivariate integrals, a particle filtering approach is devised to evaluate and optimize it in Section 5. A simulation study is given there that demonstrates the accuracy of the approach. The paper concludes with discussion and comments in Section 6.

2. Stability and Stationarity

This section assumes that $\{C_n\}_{n=1}^{\infty}$ is strictly stationary. We want to show that the Lindley random walk is stable whenever $E[C_1] < 0$. We will demonstrate that under this assumption, $\{X_n\}_{n=0}^{\infty}$ will reach a proper limiting distribution, and is thus suitable for performing parameter inference. While some old and recent literature also establish these properties [13,14], we present a simple elementary argument for completeness. A process $\{C_t\}$ is said to be strictly stationary if

$$(C_{t_1}, \ldots, C_{t_n}) \stackrel{\mathcal{D}}{=} (C_{t_1+\tau}, \ldots, C_{t_n+\tau})$$

for all $\tau, t_1, \ldots, t_n \in \mathbb{Z}$ and all $n \in \mathbb{N}^+$.

We first assume that $X_0 = 0$ and define $F_n(x) = P(X_n \leq x)$. To see that X_n is stochastically increasing in n, manipulations with (1) provide

$$\begin{aligned} X_n &= S_n - \min_{0 \leq j \leq n} S_j \\ &= \max_{0 \leq j \leq n} (S_n - S_j) \\ &= \max(0, C_n, C_n + C_{n-1}, \ldots, C_n + \cdots + C_1), \end{aligned}$$

where $S_n = C_1 + \cdots + C_n$.

Using this gives, for $x \geq 0$,

$$\begin{aligned} F_{n+1}(x) &= P(C_1 + C_2 + \cdots + C_{n+1} \leq x; C_2 + \cdots + C_{n+1} \leq x; \ldots; C_{n+1} \leq x) \\ &\leq P(C_2 + \cdots + C_{n+1} \leq x; C_3 + \cdots + C_{n+1} \leq x; \ldots; C_{n+1} \leq x) \\ &= P(C_1 + \cdots + C_n \leq x; C_2 + \cdots + C_n \leq x; \ldots; C_n \leq x) \\ &= F_n(x). \end{aligned}$$

Here, the inequalities follow by dropping terms in the joint probability and the strict

stationary of $\{C_n\}$. This shows that $F_n(x)$ is monotone non-increasing in n for each fixed $x \geq 0$.

From this monotonicity, define $F_\infty(x) = \lim_{n\to\infty} F_n(x)$. It is clear that $F_\infty(x)$ is non-decreasing in x and takes values in [0,1]. To see that $F_\infty(x)$ is also right continuous in x, and hence a distribution function, note that

$$\lim_{h\downarrow 0} F_\infty(x+h) = \lim_{h\downarrow 0} \lim_{n\to\infty} F_n(x+h) = \lim_{n\to\infty} \lim_{h\downarrow 0} F_n(x+h) = \lim_{n\to\infty} F_n(x) = F_\infty(x).$$

The interchange of limit orders is justified by the fact that $F_n(x+h)$ is non-increasing with increasing n and decreasing in h.

To show that $F_\infty(\cdot)$ is a proper cumulative distribution function (not vague), we make the law of large numbers (ergodic) assumption that

$$\frac{C_1 + \cdots + C_n}{n} \longrightarrow \mu_C \tag{3}$$

with probability one as $n \to \infty$ and that $\mu_C < 0$. A negative μ_C is necessary to induce a stable random walk even when $\{C_n\}$ is IID.

Now on any point in the probability space where (3) holds, $C_1 + \cdots + C_n \to -\infty$ as $n \to \infty$. On this path of $\{C_n\}_{n=1}^\infty$, it is relatively easy to argue that

$$X_n = \max(0, C_n, C_n + C_{n-1}, \ldots, C_n + C_{n-1} + \cdots + C_1) \tag{4}$$

is bounded away from positive infinity in n. Hence, $\limsup_{n\to\infty} X_n < \infty$, implying that X_∞, a random variable having the CDF $F_\infty(\cdot)$, is finite with probability one.

When $\mu_C = 0$, a proper limiting distribution may or may not arise. The case where $C_n \equiv 0$ provides an example where a trivial degenerate limiting distribution exits; when $\{C_n\}$ is IID and double exponentially distributed, $\{X_n\}_{n=0}^\infty$ is a null recurrent Markov chain. When $\mu_C > 0$, X_n will converge to infinity almost surely as $n \to \infty$ (assuming that $\{C_n\}$ obeys a law of large numbers) and no proper limiting distribution exists. See [13] for more on these cases, or queueing texts when $\{C_n\}$ is IID.

We now move to cases where the initial condition does not take X_0 as zero, showing that $\{X_n\}$ will still reach the same limiting distribution, regardless of starting point. For a fixed sample path of $\{C_n\}_{n=1}^\infty$, let $\{X_n\}_{n=0}^\infty$ denote the process when $X_0 = 0$ and let $\{X_n^*\}_{n=0}^\infty$ denote the process when the initial condition is X_0^*; that is, the initial level of the process $\{X_n^*\}_{n=0}^\infty$ at time $n = 0$ is X_0^*. Here, X_0^* may or may not be random and both $\{X_n\}_{n=0}^\infty$ and $\{X_n^*\}_{n=0}^\infty$ are driven by the same sample path of $\{C_n\}_{n=1}^\infty$. Define the coupling time $T = \inf\{n \geq 0 : X_n = X_n^*\}$. Since the walk is pathwise ordered, $X_n \leq X_n^*$ for all $n \geq 0$. Thus, once $\{X_n^*\}$ first hits state zero, $X_n = X_n^*$ for all n thereafter; that is,

$$T \leq \inf\{n \geq 0 : X_n^* = 0\}.$$

Applying the classic coupling inequality [23] gives

$$\sup_A \mid P(X_n \in A) - P(X_n^* \in A) \mid \leq P(T_0 > n),$$

where the supremum is taken over all Borel measurable subsets of $[0, \infty)$ and $T_0 = \inf\{n \geq 0 : X_n^* = 0\}$.

An implication of the above is that if T_0 is a proper random variable, the limiting distribution will not depend on the initial state X_0^*. We comment that our results follow from stochastic monotonicity; a Markov structure is not needed.

To see that T_0 is proper, note that

$$\begin{aligned} P(T_0 > n) &= P(X_t^* > 0 \text{ for all } t \text{ in } \{0, 1, \ldots, n\}) \\ &= P(X_0^* + C_1 + \cdots + C_t > 0 \text{ for all } t \text{ in } \{0, 1, \ldots, n\}) \\ &\leq P(X_0^* + C_1 + \cdots + C_n > 0). \end{aligned}$$

However, by (3), $C_1 + \cdots + C_n \to -\infty$ as $n \to \infty$ with probability one. Thus, T_0 is finite with probability one, $\lim_{n\to\infty} P(T_0 > n) = 0$, and the limit distribution does not depend on the initial condition X_0^*.

Next, in order to facilitate parameter inference, we investigate stationarity properties of $\{X_n\}_{n=0}^{\infty}$ when X_0 has its limiting distribution. For this, we need a doubly infinite version of the change process, which we denote by $\{C_n\}_{n=-\infty}^{\infty}$. For stationarity to hold, [13] and (4) show that X_0 must be formed from all past C_ns via

$$X_0 = \sup(0, C_0, C_0 + C_{-1}, \ldots, C_0 + C_{-1} + \cdots + C_{-n}, \ldots). \tag{5}$$

Since X_0 is a function of the past and present C_ts, namely $C_0, C_{-1}, \ldots$, we write the "causal" measurable function in (5) as $X_0 = H_0(C_0, C_{-1}, \ldots)$. Likewise, $X_1 = \max(X_0 + C_1, 0)$ is a causal measurable function of $C_1, C_0, \ldots$; viz., $X_1 = H_1(C_1, C_0, \ldots)$. Generalizing this gives $X_k = H_k(C_k, C_{k-1}, \ldots)$ for a measurable function H_k.

For a general $k \geq 1$ and integer $h > 0$, we have

$$\begin{aligned} (X_0, X_1, \ldots, X_k) &= (H_0(C_0, C_{-1}, \ldots), H_1(C_1, C_0, \ldots), \ldots, H_k(C_k, C_{k-1}, \ldots)) \\ &\stackrel{\mathcal{D}}{=} (H_0(C_h, C_{h-1}, \ldots), H_1(C_{h+1}, C_h, \ldots), \\ &\quad \ldots, H_k(C_{k+h}, C_{k+h-1}, \ldots)) \\ &= (X_h, X_{h+1}, \ldots, X_{h+k}), \end{aligned}$$

where $\stackrel{\mathcal{D}}{=}$ indicates equality in distribution, which follows by shifting the strictly stationary path of $\{C_n\}$ used by h units.

This shows that when X_0 is in its stationary state as generated by the infinite history of the change process, $\{X_n\}_{n=0}^{\infty}$ is a strictly stationary process. A caveat here: one cannot take X_0 to be independent of $\{C_n\}_{n=-\infty}^{0}$ and obtain a strictly stationary $\{X_n\}_{n=0}^{\infty}$.

3. Markov Properties

This section studies the Markov structure of the Lindley random walk process $\{X_n\}$ with correlated changes. Clarifying, a process $\{U_n\}_{n=0}^{\infty}$ is called Markov of order p if

$$P(U_n \leq x \mid U_{n-1}, \ldots, U_{n-p}, \ldots, U_0) = P(U_n \leq x \mid U_{n-1}, \ldots, U_{n-p})$$

for all real x, $n \geq 1$, and $U_0, \ldots, U_{n-1}$.

Before continuing, we introduce a class of strictly stationary time series models for $\{C_n\}$. The most widely used stationary time series model class is the autoregressive moving-average (ARMA) models. ARMA models with autoregressive order $p \geq 0$ and moving-average order $q \geq 0$ are the unique (in mean square) solutions to the difference equation

$$C_n - \mu = \phi_1(C_{n-1} - \mu) + \ldots + \phi_p(C_{n-p} - \mu) + \epsilon_n + \theta_1\epsilon_{n-1} + \ldots + \theta_q\epsilon_{n-q}. \quad (6)$$

Here, $\{\epsilon_n\}$ is a sequence of IID random variables with zero mean and variance $\sigma^2 > 0$, $\phi_1, \ldots, \phi_p$ are the p autoregressive coefficients, and $\theta_1, \ldots, \theta_q$ are the q moving-average coefficients. It is important that $\{\epsilon_n\}$ be IID — more than uncorrelated noise is needed in our Markov structure arguments below. We also assume a causal ARMA model. This stipulation requires all roots of the autoregressive polynomial $1 - \phi_1 z - \cdots - \phi_p z^p$ to lie outside the complex unit circle. For causal models, C_n can be expressed in terms of the current and past ϵ_ns only, namely $\epsilon_n, \epsilon_{n-1}, \ldots$.

For the ARMA(p, q) model in (6), $E[C_n] \equiv \mu$ for all n and $\{C_n\}$ is strictly stationary. For a stable model, we need $\mu < 0$ as shown in the last section (this is henceforth assumed). Computation of the autocovariances $\gamma(h) := \mathrm{Cov}(C_n, C_{n+h})$ proceeds from many classic algorithms. For more on this and other properties of ARMA series, see [24].

Many practitioners nowadays focus on autoregressions (AR) only due to their parsimonious and flexible structure and forecasting ease; that is, $\theta_1 = \cdots = \theta_q = 0$. Indeed, autoregressions are dense in all short memory stationary series. It is easy to see that an AR(p) series is a Markov chain of order p when $\{\epsilon_n\}$ is IID; indeed, (6) explicitly writes X_n as a function of the p past series values and an independent noise that does not depend on past process values.

Perhaps the most commonly used time series model, and one studied below, is the causal first order autoregression (AR(1)). This model obeys

$$C_n - \mu = \phi(C_{n-1} - \mu) + \epsilon_n, \quad (7)$$

where $\mid \phi \mid < 1$ is needed for causality.

A more general class of strictly stationary change processes assumes the causal linear structure

$$C_n = \mu + \sum_{k=0}^{\infty} \psi_k \epsilon_{n-k},$$

where the deterministic weight sequence $\{\psi_k\}_{k=0}^{\infty}$ satisfies $\sum_{k=0}^{\infty} \mid \psi_k \mid < \infty$. Reference [24] shows that any causal ARMA sequence with IID innovations has this representation and discusses the related Wold decomposition for stationary time series.

Recursing (1) yields

$$X_n = \max\left(X_{n-L} + C_{n-L+1} + \cdots + C_n; C_{n-L+2} + \cdots + C_n; \ldots; C_n; 0\right), \quad (8)$$

which shows how X_n depends on X_{n-L} for any $L \geq 1$. Hence, in general, X_n and X_{n+h} will be dependent for all lags h, even when $\{C_n\}$ is m-dependent (say a moving-average of order m). Indeed, (8) implies that X_n and X_{n-L} will be dependent in general for any $L \geq 1$. Note however that this dependence does not necessarily imply the absence of

a Markov structure. For an example of this, consider the causal AR(1) process above. Recursing (7) gives

$$C_n - \mu = \phi^L(C_{n-L} - \mu) + \phi^{L-1}\epsilon_{n-L+1} + \cdots + \phi\epsilon_{n-1} + \epsilon_n.$$

for any $L \geq 1$. Thus, C_n and C_{n-L} are dependent; however, this process is known to be first order Markov [24].

Our first result shows that $\{X_n\}$ is not Markov of any order unless $\{C_n\}$ has additional structure. This corrects a mistaken claim in [12]. In deriving the likelihood for a Lindley process, the authors have assumed that the process is first-order Markov when the change process is AR(1). The proof is given in Appendix A.

Lemma 3.1. *The general Lindley walk $\{X_n\}_{n=0}^{\infty}$ is not Markov of any order.*

Our next result, also proven in Appendix A, establishes the Markov structure of the bivariate process $\{(X_n, C_n)\}$ when $\{C_n\}$ is a pth order Markov chain.

Proposition 3.2. *The process $\{(X_n, C_n)\}$ is a pth order Markov chain when $\{C_n\}$ is a pth order Markov chain (such as the above AR(p) series).*

The proofs for the above two results also establish the following result.

Proposition 3.3. *The $p+1$ dimensional process $\{(X_n, C_n, C_{n-1}, \ldots, C_{n+1-p})\}_{n=p}^{\infty}$ is a first order Markov chain whenever $\{C_n\}$ is a causal AR(p) series.*

This property will be useful for deriving the likelihood of a Lindley process when $\{C_n\}$ is a pth order Gaussian autoregression.

4. Likelihood Structure

The likelihood of a Lindley walk with IID $\{C_n\}$ was studied in [2], see [25] for additional work. Our goal in this section is two-fold. This section clarifies the support set of the distribution of $\mathbf{X} = (X_0, \ldots, X_N)'$ and develops the likelihood function in terms of the distribution of the change process $(C_1, \ldots, C_N)'$ and the initial value X_0. In general, X_n has a point mass at zero and a possible density over $(0, \infty)$ for each fixed n. The complexity of the likelihood obtained motivates a particle filtering approach presented in the next section.

We first study the support set of $\mathbf{X}$. To avoid trite work with discrete cases, assume that $\mathbf{C} = (C_1, \ldots, C_N)'$ has the joint probability density (PDF) $f_{\mathbf{C}}(\mathbf{c})$, where $\mathbf{c} = (c_1, \ldots, c_N)' \in \mathbb{R}^N$. The initial starting level X_0 is non-negative with cumulative distribution $F_{X_0}(x)$; this distribution has a point mass at zero ($F_{X_0}(0) > 0$) and a density on $(0, \infty)$. For a strictly stationary $\{X_n\}$, X_0 must be a function of the past changes $C_0, C_{-1}, \ldots$, as is quantified in (5). The random vector $\mathbf{X}$ has a distribution that is a mixture of densities and mass functions on different domain regions. To quantify these regions, consider a partition of $\mathbb{R}_+^{N+1}$ defined as follows: let $I \subset \mathcal{I} \equiv \{0, 1, 2, \ldots, N\}$, and define a set associated with I via

$$B_I = \{\mathbf{y} \in \mathbb{R}_+^{N+1} : y_n > 0 \ \ \forall n \in I, \ \text{ and } \ y_n = 0 \ \forall n \notin I\}.$$

Here, I contains all indices with positive components. In particular, $B_{\mathcal{I}}$ is the interior of $\mathbb{R}_+^{N+1}$ and $B_\emptyset = \{\mathbf{0}\}$. Note that the $B_I, I \subset \mathcal{I}$, are disjoint and that $\cup_{I \subset \mathcal{I}} B_I = \mathbb{R}_+^{N+1}$;

thus, $\{B_I; I \subset \mathcal{I}\}$ partitions $\mathbb{R}_+^{N+1}$.

The joint distribution constructed below is the likelihood of $\mathbf{X}$ as derived in Equation (3) of [26]. Now, consider an $I \subset \mathcal{I}$ that contains a non-empty sequence of indices $i_1, \ldots, i_k$ satisfying $0 \leq i_1 < i_2 < \cdots < i_k \leq N$. For $\mathbf{x} \in B_I$, the cumulative distribution function $F_{\mathbf{X}}(\mathbf{x}) = P[\cap_{i=0}^{N} X_i \leq x_i]$ is differentiable in the variables $x_i, i \in I$, and we write

$$L_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^k}{\partial x_{i_1} \cdots \partial x_{i_k}} F_{\mathbf{X}}(\mathbf{x}) \tag{9}$$

for this density.

For cases where zeroes arise, define $L_{\mathbf{X}}(\mathbf{0}) = F_{\mathbf{X}}(\mathbf{0}) = P[X_0 = 0, \ldots, X_N = 0]$. For each $B \subset B_I$, the positive components in B can be obtained from the map $\mathcal{M}(B) = \{(x_{i_1}, \ldots, x_{i_k}) : \mathbf{x} \in B\}$. We view $L_{\mathbf{X}}(\mathbf{x})$ as a function of the positive components in $\mathbf{x}$ and write $L_{\mathbf{X}}(\mathbf{x}) = L_{\mathbf{X}}(x_{i_1}, \ldots, x_{i_k})$.

Now consider a probability measure μ_I, defined on $\mathbb{R}_+^{N+1}$, such that for $A \in \mathcal{B}(\mathbb{R}_+^{N+1})$,

$$\mu_I(A) = \int_{\mathcal{M}(A \cap B_I)} L_{\mathbf{X}}(\mathbf{x}) \lambda(d\mathbf{x}),$$

where λ is the Lebesgue measure on $\mathcal{M}(B_I)$. If $I = \emptyset$, λ reduces to a discrete Dirac measure. We observe that $\mu_I, I \subset \mathcal{I}$, are mutually singular. The distribution of $\mathbf{X}$ can thus be characterized by

$$\mathcal{L}(\mathbf{x}) = \sum_{I \subset \mathcal{I}} 1_{\{x \in B_I\}} L_{\mathbf{X}}(\mathbf{x});$$

this is our likelihood. Because the data $\mathbf{X}$ are fixed when optimizing over the parameters $\mathbf{\Theta}$ in a likelihood, we also write $\mathcal{L}(\mathbf{\Theta})$ or $\mathcal{L}(\mathbf{\Theta} \mid \mathbf{X})$ for our likelihood. Other variants of notation are used in obvious manners.

We now turn to computing the likelihood of $\mathbf{X}$ from $f_{X_0,\mathbf{C}}(x, \mathbf{c})$, $x \geq 0$, $\mathbf{c} \in \mathbb{R}^N$ — the joint distribution of X_0 and $\mathbf{C}$. When $x > 0$, $f_{X_0,\mathbf{C}}(x, \mathbf{c})$ is a joint density function, and when $x = 0$, $f_{X_0,\mathbf{C}}(0, \mathbf{c}) = \partial^N F_{X_0,\mathbf{C}}(0, \mathbf{c}) / \partial c_1 \cdots \partial c_N)$. We are given the observations $\mathbf{X} = (X_0, X_1, \ldots, X_N)' \in B_I$. Recall that $I = \{i_1, \ldots, i_k\}$ such that $X_m > 0$ for $m \in I$ and $X_n = 0$ for $n \notin I$. Write $I^c \equiv \mathcal{I} \backslash I = \{j_1, \ldots, j_{N-k}\}$. The CDF of $\mathbf{X}$ is

$$F_{\mathbf{X}}(\mathbf{x}) = P(X_0 \leq x_0, X_m \leq x_m, m \in I, X_n = 0, n \in I^c).$$

Without loss of generality, assume that $x_0 > 0$. From (9), the likelihood is

$$\mathcal{L}(\mathbf{\Theta} \mid \mathbf{X} = \mathbf{x}) = \lim_{\substack{h \downarrow 0 \\ m \in I \cup \{0\}}} h^{-\alpha} P(X_m \in (x_m - h/2, x_m + h/2), m \in I \cup \{0\}, X_n = 0, n \in I^c),$$

where $\alpha = |I \cup \{0\}|$, and h is assumed to be smaller than $2x_m$ for all m so that each interval $(x_m - h/2, x_m + h/2)$ is nonempty. From (1), for $m \in I$, $x_m > 0$, implying

that $C_m = X_m - X_{m-1}$; for $n \in I^c$, $x_n = 0$, implying that $C_n \leq -X_{n-1}$. Hence,

$$\mathcal{L}(\boldsymbol{\Theta} \mid \mathbf{X} = \mathbf{x}) =$$
$$\lim_{\substack{h \downarrow 0 \\ m \in I \cup \{0\}}} h^{-\alpha} P\Big(X_m \in (x_m - h/2, x_m + h/2), C_m = X_m - X_{m-1}, m \in I,$$
$$X_n = 0, C_n \leq -X_{n-1}, n \in I^c, X_0 \in (x_0 - h/2, x_0 + h/2) \Big) =$$
$$\lim_{\substack{h \downarrow 0 \\ m \in I \cup \{0\}}} h^{-\alpha} P\Big(X_m \in (x_m - h/2, x_m + h/2), C_m = X_m - X_{m-1}, m \in I,$$
$$C_n \leq -x_{n-1}, n \in I^c, X_0 \in (x_0 - h/2, x_0 + h/2) \Big).$$

Using the change of variables formula for $C_m = X_m - X_{m-1}, m \in I$, we have the likelihood

$$\mathcal{L}(\boldsymbol{\Theta} \mid \mathbf{X}) = \int_{-\infty}^{-X_{j_1 - 1}} \cdots \int_{-\infty}^{-X_{j_{N-k}-1}} f_{X_0, \mathbf{C}}(X_0, X_m - X_{m-1}, m \in I, \mathbf{c}) d\mathbf{c}, \qquad (10)$$

where the integral is over the set $\{\mathbf{c} \in \mathbb{R}^{N-k} : C_n \leq -X_{n-1}, n \in I^c\}$. If X_0 is independent of $\mathbf{C}$, the integrand in (10) becomes $f_{X_0}(X_0) f_{\mathbf{C}}(X_0, X_m - X_{m-1}, m \in I, \mathbf{c})$. The formula in (10) involves high dimensional multiple integrals when $\mathbf{X}$ has many zeros, even after employing the Markov relations in Section 3. If one assumes a Gaussian $\{C_n\}$, the high dimensional integrals induce considerable computational difficulty [27].

Appendix B derives the likelihood when $\{C_n\}$ is a pth order causal autoregression. The expression there, as well as the form in (10), are not computationally convenient because of the high dimensional integrals involved. Because of this, our next section moves to a technique that efficiently simulates this likelihood to a degree where statistical inferences can be accurately made.

5. Particle Filtering Likelihood Evaluation

This section introduces particle filtering methods to approximate and optimize the storage model's likelihood. Since he exact likelihood in (8) is problematic to evaluate, we construct an approximation to it, viewing the problem as a censored time series issue.

5.1. Particle Filtering Methods

As noted in the last section, C_n can be recovered exactly as $X_n - X_{n-1}$ when $X_n > 0$. When $X_n = 0$, we know that $C_n \leq -X_{n-1}$, but we do not know C_n exactly. For such times n, define a censoring indicator $\delta_n = 1$; set $\delta_n = 0$ if C_n can be recovered exactly at time n. For convenience, our initial condition sets $C_1 = X_1$ and we work with the data $X_1, \ldots, X_N$.

For notation, let $\pi_1, \ldots, \pi_r$ denote the ordered times at which C_n is censored and $d_1, \ldots, d_s$ be the ordered times at which C_n can be exactly recovered. Obviously,

$r + s = N$. Define the censored series as

$$C_n^* = \begin{cases} X_n - X_{n-1}, & \text{if } C_n \text{ is recoverable} \\ -X_{n-1}, & \text{if } C_n \text{ is not recoverable} \end{cases}.$$

The likelihood of $(C_1, \ldots, C_N)'$ can be written in terms of the uncensored C_ns as

$$\mathcal{L}(\boldsymbol{\Theta}) = \int_{\{c_{\pi_i} \in (-\infty, c^*_{\pi_i}), i \in \{1,\ldots,r\}\}} \mathcal{N}_{\boldsymbol{\Theta}}\left(c_{1:N}\right) dc_{\pi_1} \ldots dc_{\pi_r}, \tag{11}$$

where we have taken $\{C_n\}$ to be a Gaussian process; this agrees with (10). Here, a joint Gaussian probability density for $(C_1, \ldots, C_N)'$, denoted by $\mathcal{N}_{\boldsymbol{\Theta}}(c_{1:N})$, is assumed. This Gaussian density arises as a consequence of the usual assumption of a Gaussian innovations process in time series. In the method developed below for evaluating (11), the Gaussian setup is crucial, as well as computationally convenient. Other marginal distributions for $\{C_n\}$ are possible. For example, if the quantities being modeled are integers, it would be necessary to develop some form of count time series model likelihood, and the particle filtering methods would need to be altered accordingly.

Literature to evaluate (11) includes [12]. Below, a novel particle filtering method that exploits the autoregressive structure of the series will be devised. We begin with importance sampling, observing that

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Theta}) &= \int_{\{c_{\pi_i} \in (-\infty, c^*_{\pi_i}), i \in \{1,\ldots,r\}\}} \mathcal{N}_{\boldsymbol{\Theta}}\left(c_{1:N}\right) dc_{\pi_1} \ldots dc_{\pi_r} \\ &= \int_{\{c_{\pi_i} \in (-\infty, c^*_{\pi_i}), i \in \{1,\ldots,r\}\}} \frac{\mathcal{N}_{\boldsymbol{\Theta}}\left(c_{1:N}\right)}{q(c_{\pi_1}, \ldots, c_{\pi_r})} q(c_{\pi_1}, \ldots, c_{\pi_r}) dc_{\pi_1} \ldots dc_{\pi_r}, \end{aligned}$$

where $q(\cdot)$ is any probability density function that we call a proposal density function and $c_{1:N} = (c_1, \ldots, c_N)'$. We want $q(\cdot)$ to be easy to sample from and to be supported on the set $\{c_{\pi_i} \in (-\infty, c^*_{\pi_i}], i \in \{1, \ldots, r\}\}$.

Assume that M independent samples are drawn from $q(\cdot)$. Then a law of large numbers approximation of the likelihood in (11) is

$$\mathcal{L}(\boldsymbol{\Theta}) = E_q[W] \approx \frac{1}{M} \sum_{m=1}^{M} \frac{\mathcal{N}_{\boldsymbol{\Theta}}\left(c_{1:N}^{(m)}\right)}{q(c_{\pi_1}^{(m)}, \ldots, c_{\pi_r}^{(m)})},$$

where $W = \mathcal{N}_{\boldsymbol{\Theta}}\left(C_{1:N}\right) / q(C_{\pi_1:\pi_r})$ is viewed as a "weight". The subscript of q on E implies that the expectation is taken relative to the distribution q. In our notation, superscripts of (m) refer to the mth generated sample (particle) of M total.

A "nice" $q(\cdot)$ is sought to facilitate our sampling procedure. For this, we consider an AR(1) scheme to illustrate the ideas; this is easily extendable to AR(p) settings. In the AR(1) case, $\boldsymbol{\Theta}$ contains the three parameters μ, ϕ, and σ^2. The proposal density we use is the conditional probability density of the uncensored data $C_{\pi_1}, \ldots, C_{\pi_r}$ given the censored values $C^*_{\pi_1}, \ldots, C^*_{\pi_r}$:

$$q(c_{\pi_1}, \ldots, c_{\pi_r}) = p(c_{\pi_1}, \ldots, c_{\pi_r} \mid c^*_{\pi_1}, \ldots, c^*_{\pi_r}),$$

where $p(\cdot \mid \cdot)$ is used as notation for a generic conditional probability density function.

Assuming $\pi_1 \neq 1$, the first order Markov property for AR(1) series provides

$$p(c_{\pi_1}, \ldots, c_{\pi_r} \mid c^*_{\pi_1}, \ldots, c^*_{\pi_r}) = \prod_{i=1}^{r} p(c_{\pi_i} \mid c_{\pi_i - 1}, c^*_{\pi_i}). \tag{12}$$

To see (12), note that $c_{\pi_i - 1}$ and the parameters in $\boldsymbol{\Theta}$ determine the normal distribution's mean and standard deviation of c_{π_i}, and $c^*_{\pi_i}$ indicates the upper bound of the truncation. Hence, $p(c_{\pi_i} \mid c_{\pi_i - 1}, c^*_{\pi_i})$ is simply the truncated normal density with support on $(-\infty, c^*_{\pi_i})$, having a mean and variance that are the one-step-ahead prediction of C_{π_i} from $C_{\pi_i - 1}$. To further see this, note that

$$p(c_{\pi_i} \mid c_{\pi_i - 1}, c^*_{\pi_i}) = \frac{\varphi\left(c_{\pi_i} \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i}\right)}{\Phi\left(c^*_{\pi_i} \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i}\right) - \Phi(-\infty)} = \frac{\varphi\left(c_{\pi_i} \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i}\right)}{\Phi\left(c^*_{\pi_i} \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i}\right)}.$$

where $\hat{m}_{\pi_i} = \phi C_{\pi_i - 1}$ and $\hat{r}_{\pi_i} = \sigma^2$ are the one-step-ahead predictions and variances of C_{π_i} from $C_{\pi_{i-1}}$. Here, we have used φ and Φ as notation for the standard normal density and cumulative distribution functions.

A complication here is that $C_{\pi_{i-1}}$ may or may not be censored. If $C_{\pi_{i-1}}$ is not censored, we use the observation $C_{\pi_{i-1}}$; otherwise, we use the generated value of $C_{\pi_{i-1}}$, which is always available from the sampling generation procedure adopted.

Hence, the proposal distribution in (12) is relatively easy to sample from, with weight

$$\begin{aligned} W^{(m)} &= \frac{\mathcal{N}_{\boldsymbol{\Theta}}\left(c^{(m)}_{1:N}\right)}{q\left(c^{(m)}_{\pi_1:\pi_r}\right)} = \frac{\prod_{j=1}^{N} p\left(c^{(m)}_j \mid c^{(m)}_{j-1}\right)}{\prod_{i=1}^{r} p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i - 1}, c_{\pi^*_i}\right)} \\ &= \frac{\prod_{j=1}^{s} p(c_{d_j} \mid c_{d_j - 1}) \prod_{i=1}^{r} p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i - 1}\right)}{\prod_{i=1}^{r} p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i - 1}, c_{\pi^*_i}\right)}. \end{aligned}$$

Equation (12) and $p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i - 1}\right) = \varphi\left(c^{(m)}_{\pi_i} \mid \hat{m}^{(m)}_{\pi_i}, \hat{r}^{(m)}_{\pi_i}\right)$ give

$$\frac{p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i - 1}\right)}{p\left(c^{(m)}_{\pi_i} \mid c^{(m)}_{\pi_i - 1}, c_{\pi^*_i}\right)} = \Phi\left(c^*_{\pi_i} \mid \hat{m}^{(m)}_{\pi_i}, \hat{r}^{(m)}_{\pi_i}\right).$$

Therefore, our form for the weight is

$$W^{(m)} = \prod_{j=1}^{s} \varphi\left(c_{d_j} \mid \hat{m}^{(m)}_{d_j}, \hat{r}^{(m)}_{d_j}\right) \times \prod_{i=1}^{r} \Phi\left(c^*_{\pi_i} \mid \hat{m}^{(m)}_{\pi_i}, \hat{r}^{(m)}_{\pi_i}\right).$$

Summarizing, our algorithm for the AR(1) case is as follows:

1. if $\delta_1 = 1$, C_1 is uncensored and $W_1 = \varphi(C_1 \mid -\mu, \sigma^2/(1-\phi^2))$; if $\delta_i = 0$, C_1 is censored and $W_1 = \Phi(C_1 \mid -\mu, \sigma^2/(1-\phi^2))$, and C_1 is sampled from the truncated normal density $\mathcal{N}\left(\hat{m}_{\pi_i}, \hat{r}_{\pi_i}; -\infty, 0\right)$.

After step 1, repeat steps 2 and 3 until $i = N$:

2. if $\delta_i = 1$, X_i is uncensored and update via

$$W_i = W_{i-1} \times \varphi(c_i \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i});$$

if $\delta_i = 0$, X_i is censored and update via

$$W_i = W_{i-1} \times \Phi(c_i^* \mid \hat{m}_{\pi_i}, \hat{r}_{\pi_i}),$$

3. if $\delta_i = 1$, do nothing; if $\delta_i = 1$, sample C_i from the truncated normal distribution

$$\mathcal{N}\left(\hat{m}_{\pi_i}, \hat{r}_{\pi_i}; -\infty, c_i^*\right)$$

4. Record W_N.

The above process is repeated M times, where M is large enough that law of large number approximations are good. Generally, the larger M is, the better the approximation will be. The final approximated likelihood is

$$\mathcal{L}(\boldsymbol{\Theta}) \approx \frac{1}{M} \sum_{m=1}^{M} W_N^{(m)}.$$

Before closing, we comment on a naive way to simulate the likelihood with a Gaussian $\{C_n\}$. Another way to evaluate (11) decomposes the likelihood as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\Theta}) = \int_{\{c_{\pi_i} \in (-\infty, c_{\pi_i}^*)\}} & \mathcal{N}_{\boldsymbol{\Theta}}\left(c_{\pi_1}, \cdots, c_{\pi_p} \mid c_{d_1}, \cdots, c_{d_p}\right) \\
& \times \mathcal{N}_{\boldsymbol{\Theta}}\left(c_{d_1}, \cdots, c_{d_p}\right) dc_{\pi_1} \cdots dc_{\pi_p} \\
= \mathcal{N}_{\boldsymbol{\Theta}}\left(c_{d_1}, \cdots, c_{d_p}\right) & \int_{\{c_{\pi_i} \in (-\infty, c_{\pi_i}^*)\}} \mathcal{N}_{\boldsymbol{\Theta}}\big(c_{\pi_1}, \cdots, c_{\pi_p} \mid \\
& \qquad c_{d_1}, \cdots, c_{d_p}\big) dc_{\pi_1} \cdots dc_{\pi_p}
\end{aligned}$$

and uses the explicit form of the conditional multivariate normal density in [28] to sample the censored values conditional on the uncensored values. This is more computationally expensive than the proposed particle filtering method because drawing from the conditional normal distribution requires inverting a covariance matrix of dimension equal to the number of censored values. When μ is highly negative and N is large, many data points will be censored and this dimension may be large.

5.2. A Simulation Study

This subsection presents a simulation study that evaluates the performance of our particle filtering estimation methods in the last subsection. The R code and seeds used to generate the data used in our analysis are available from the corresponding author upon request.

We first consider the case of a Gaussian AR(1) $\{C_n\}_{n=1}^N$. The parameters in this setup are μ, ϕ, and σ^2. The mean μ is taken as negative to ensure that $\{X_n\}$ is stable.

The more negative μ is, the more frequently X_n will be zero and censoring occurs. The AR(1) correlation parameter ϕ satisfies $\mid \phi \mid < 1$, which is needed for a causal $\{C_n\}$. We will examine $\phi \in \{-0.5, -0.25, 0, 0.25, 0.5, 0.75\}$, although negative ϕ do not arise in practice as much as positive ϕ. In all simulations, σ^2 is taken as unity.

Each simulated series uses $M = 10,000$ independent particles. The series lengths $N = 100, 250$, and 500 were studied. Usually, $\mathcal{L}(\boldsymbol{\theta})$ obtained via particle filtering is "noisy" due to sampling. This "noisy" likelihood leads to irregular numerical second order derivatives, which complicate getting standard errors of the estimators. A popular fix is called common random number (CRNs). CRN techniques smooth the estimated likelihood by generating a set of random quantities in the particle filtering routines through transformation, keeping them constant across the computations for different sets of parameters. CRN techniques were used to the ensure that the likelihood is relatively smooth with respect to its parameters. This is an essential step with particle filtering methods — see [29] and [30] for discussion and more on CRNs. Finally, the popular quasi-Newton method L-BFGS-B is implemented to optimize the likelihoods. The true model parameters were used as initial guesses in our optimizations. It took, on average, 15s, 45s, and 90s in the coding language R on a Macbook Pro computer to complete an analysis for one simulated series of length $N = 100, 250$, and 500, respectively.

Figure 1 and 2 show boxplots of parameter estimators aggregated from 200 independent series. The sample means of the particle filtering estimators are all close to their true values, with some minor bias present in some cases. This bias decays with increasing sample size n. We remind the reader that likelihood estimation of AR(1) parameters in uncensored settings is also slightly biased (see [31] for a bias quantification). The estimators are compared in the figure with boxplots of a naive estimator that simply fits an AR(1) model to $\{X_n\}_{n=1}^N$ without accounting for the hard boundary at zero. These estimators, shaded in light blue in the figure, are uniformly worse, especially when μ is far below zero and more censoring occurs.

For standard errors of the estimators, Table 1 reports two values: 1) the sample standard deviations of the parameter estimators over the 200 runs (denominator of 199), and 2) the average (over the 200 runs) of standard errors obtained by inverting the Hessian matrix at the maximum likelihood estimate for each run (denominator of 200). These two standard errors are close to one another, providing comfortable agreement. We do not consider standard errors for the poorer naive estimators.

Overall, the performance of the particle filtering estimation for AR(1) series is stellar. One can even get an accurate standard error from one realization of the series by inverting the Hessian matrix at the likelihood estimators. In some particle filtering applications, "particle degeneration" occurs for larger N and results can degrade for these sample sizes. Methods to correct for particle degeneration are discussed in [32–34]; these do not appear needed here.

Some AR(2) cases were also examined. Figures 3 and 4 report results for some selected values of μ, ϕ_1, and ϕ_2 with $\sigma^2 = 1$. Performance is analogous to the AR(1) case, with the naive estimator again performing worse. Table 2 show standard errors for the AR(2) case in an analogous format to those in Table 1. The results are again impressive.

Overall, likelihood inference, the gold standard for statistical estimation, can be conducted for storage models with AR errors.

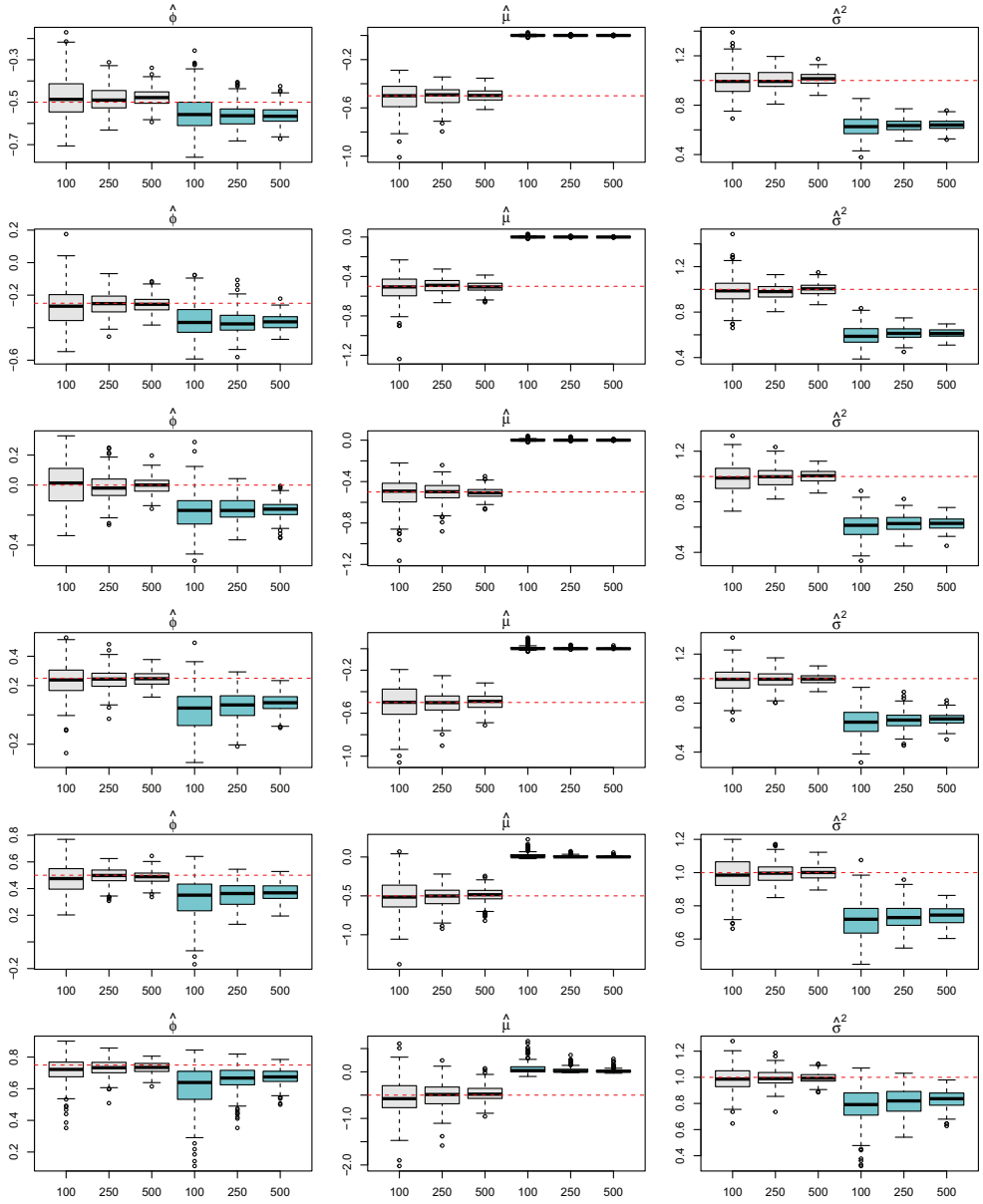

Figure 1. Boxplots of parameter estimators for the Lindley walk with an AR(1) $\{C_n\}$ with $\mu = -0.5$ (Model 1). The dashed lines demarcate true parameter values. All particle filtering estimators appear roughly unbiased, with any bias decaying with increasing sample size. The naive estimators, shaded in light blue, are uniformly poorer.

6. Discussion

This paper investigated Lindley random walks (storage models) in the case where the change process driving the walk is strictly stationary. This essentially extends Lindley process inference to time series settings.

First, the paper established the asymptotic mathematical properties of Lindley processes with correlated changes, providing a streamlined analysis. We then investigated

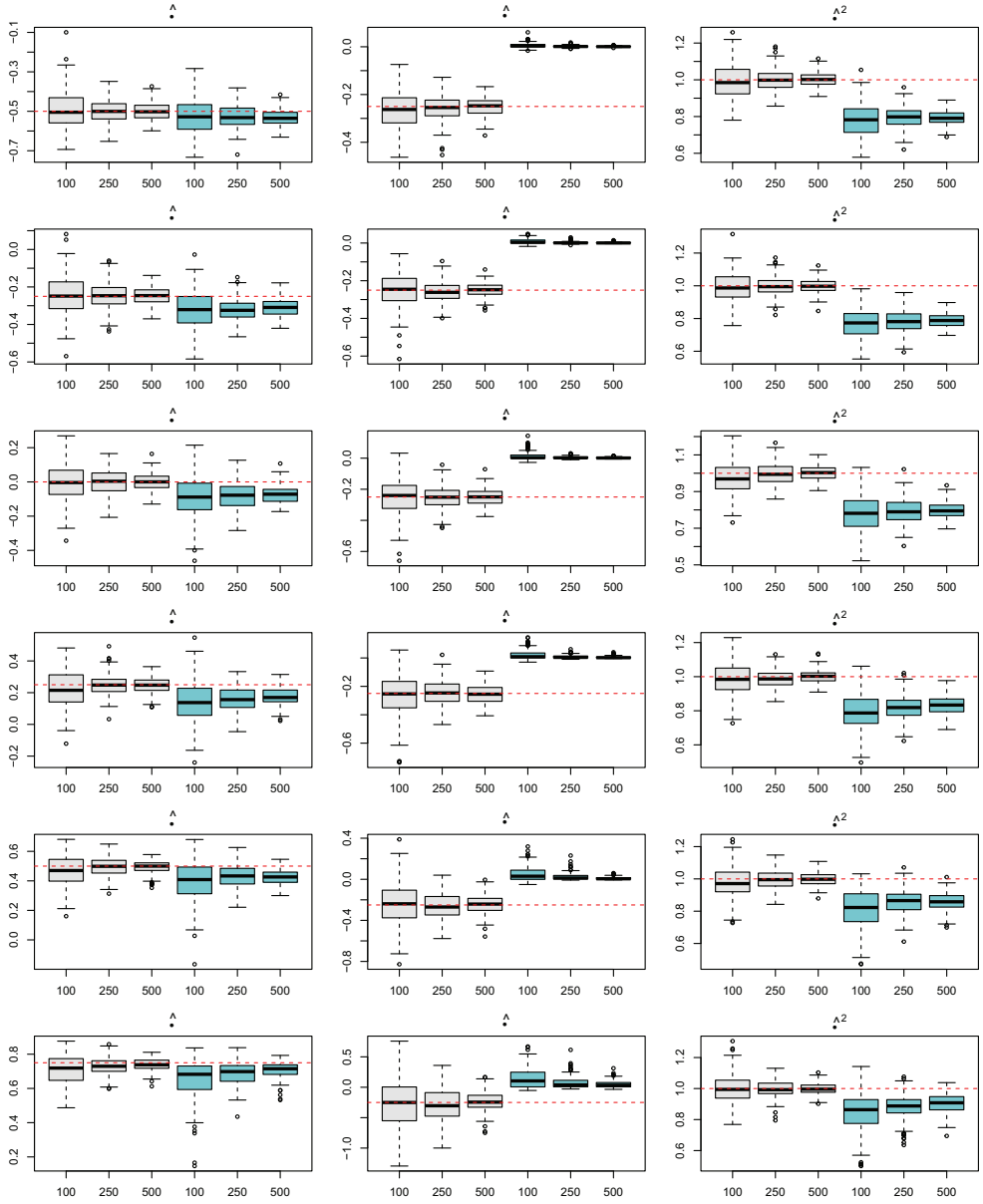

Figure 2. Boxplots of parameter estimators for the Lindley walk with an AR(1) $\{C_n\}$ with $\mu = -0.25$ (Model 2). The dashed lines demarcate true parameter values. All particle filtering estimators appear roughly unbiased, with any bias decaying with increasing sample size. The naive estimators, shaded in light blue, are uniformly poorer.

the Markov (or lack thereof) structure of the Lindley process. The paper then turned to statistical estimation issues, deriving the model's likelihood function in the case of a Gaussian AR(p) change process. Because of the complexity of the resulting expression, a particle filtering method of likelihood approximation was investigated that partitioned the series into segments where the change process was either recoverable or censored. A simulation study showed that the estimation procedure works well; accurate standard errors for the parameters were even achieved.

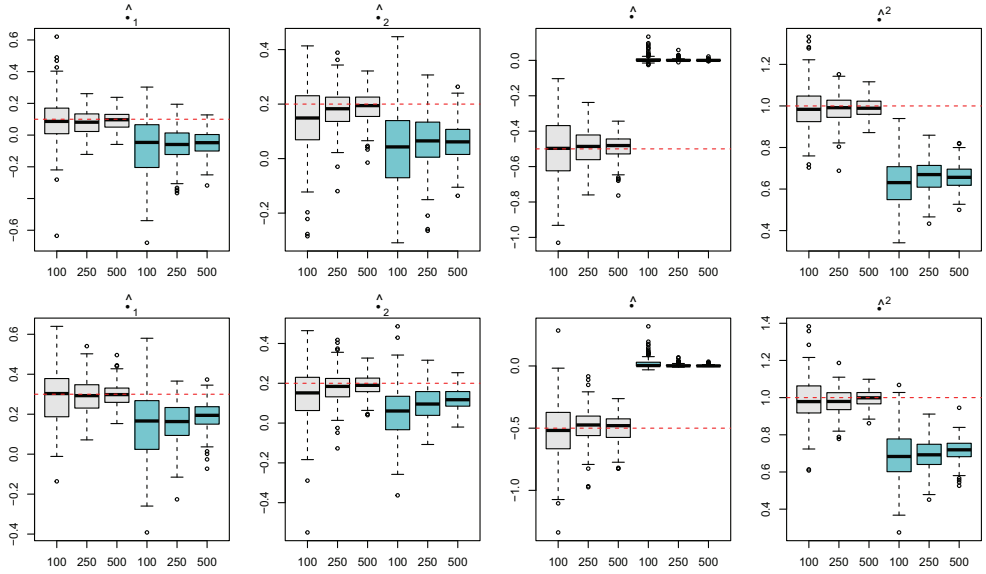

Figure 3. Boxplots of parameter estimators for the Lindley walk with an AR(2) $\{C_n\}$ with $\mu = -0.5$ (Model 3). The dashed lines demarcate true parameter values. All particle filtering estimators appear roughly unbiased, with any bias decaying with increasing sample size. The naive estimators, shaded in light blue, are uniformly poorer.

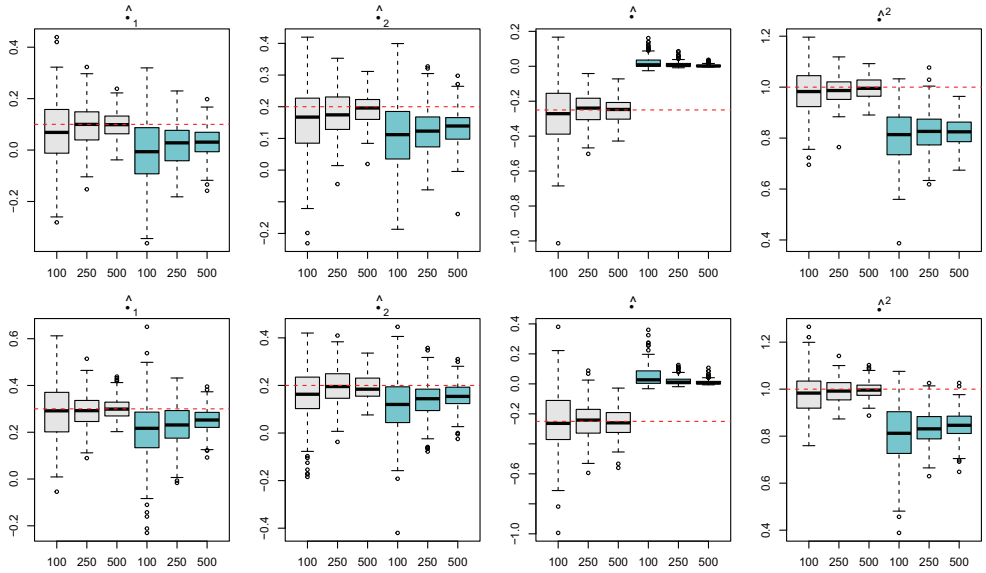

Figure 4. Boxplots of parameter estimators for the Lindley walk with an AR(2) $\{C_n\}$ with $\mu = -0.25$ (Model 4). The dashed lines demarcate true parameter values. All particle filtering estimators appear roughly unbiased, with any bias decaying with increasing sample size. The naive estimators, shaded in light blue, are uniformly poorer.

Several directions for future research are apparent. Queueing applications involving (2) would need to move away from a Gaussian $\{C_n\}$. Here, one wants the $\{I_n\}$ and $\{S_n\}$

processes to be stationary but with exponentially distributed marginal distributions.

A copula way to construct correlated exponential service times $\{S_n\}$ takes a stationary Gaussian process $\{Z_n\}$, standardized so that $E[Z_n] \equiv 0$ and $\mathrm{Var}(Z_n) \equiv 1$, and sets

$$S_n = F^{-1}(\Phi(Z_n)),$$

where $F^{-1}(x) = -\ln(x)/\eta$ is the inverse of the exponential cumulative distribution function with mean $\eta > 0$ (non-exponential distributions can also be made). Another extension involves inventory counts. Here, the process would be count valued, with $\{I_n\}$ and $\{S_n\}$ having a count marginal distribution such as Poisson. Different likelihoods would need to be developed for these non-Gaussian cases.

A detailed application of the methods is being constructed in [35]. This application involves daily frozen soil and lake ice depths and requires periodic versions of the Lindley walk. Here, $\{C_n\}$ is stationary in a periodic sense with a negative overall mean (this said, some day-to-day changes in the height of winter might see mean increases). This structure also arises in [4], where trend components are incorporated in the modeling procedure, but estimation methods are somewhat ad-hoc. In periodic settings, stability of the walk needs to be investigated.

Funding

Jiajie Kong's and Robert Lund's research was supported by National Science Foundation Grant DMS 2113592.

Disclosure Statement

The authors report there are no competing interests to declare.

References

[1] Kendall DG. Some problems in the theory of queues. Journal of the Royal Statistical Society: Series B (Methodological). 1951;13(2):151–173.

[2] Lindley DV. The theory of queues with a single server. In: Mathematical Proceedings of the Cambridge Philosophical Society; Vol. 48; Cambridge University Press; 1952. p. 277–289.

[3] Woody J, Lund R, Grundstein AJ, et al. A storage model approach to the assessment of snow depth trends. Water Resources Research. 2009;45(10):W10426.

[4] Lee J, Lund R, Woody J, et al. Trend assessment for daily snow depths with changepoint considerations. Environmetrics. 2020;31(1):2580–2591.

[5] Woody J, Wang Y, Dyer J. Application of multivariate storage model to quantify trends in seasonally frozen soil. Open Geosciences. 2016;8(1):310–322.

[6] Woody J, Lu Q, Livsey J. Statistical methods for forecasting daily snow depths and assessing trends in inter-annual snow depth dynamics. Environmental and Ecological Statistics. 2020;27(1):609–622.

[7] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological). 1977; 39(1):1–22.

[8] Schmee J, Hahn GJ. A simple method for regression analysis with censored data. Technometrics. 1979;21(4):417–432.
[9] Jones RH. Maximum likelihood fitting of ARMA models to time series with missing observations. Technometrics. 1980;22(3):389–395.
[10] Robinson PM. Estimation and forecasting for time series containing censored or missing observations. Time Series : Proceedings of the International Conference Held at Nottingham University. 1980;.
[11] Hopke PK, Liu C, Rubin DB. Multiple imputation for multivariate data with missing and below-threshold measurements: time-series concentrations of pollutants in the Arctic. Biometrics. 2001;57(1):22–33.
[12] Park JW, Genton MG, Ghosh SK. Censored time series analysis with autoregressive moving average models. Canadian Journal of Statistics. 2007;35(1):151–168.
[13] Loynes RM. The stability of a queue with non-independent inter-arrival and service times. In: Mathematical Proceedings of the Cambridge Philosophical Society; Vol. 58; Cambridge University Press; 1962. p. 497–520.
[14] Lovas A, Rásonyi M. Ergodic theorems for queuing systems with dependent inter-arrival times. Operations Research Letters. 2021;49(5):682–687.
[15] Hwang GU, Sohraby K. On the exact analysis of a discrete-time queueing system with autoregressive inputs. Queueing Systems. 2003;43(1):29–41.
[16] Kamoun F. The discrete-time queue with autoregressive inputs revisited. Queueing Systems. 2006;54(3):185–192.
[17] Fendick KW, Saksena VR, Whitt W. Dependence in packet queues. IEEE Transactions on Communications. 1989;37(11):1173–1183.
[18] Fendick KW, Saksena VR, Whitt W. Investigating dependence in packet queues with the index of dispersion for work. IEEE Transactions on Communications. 1991;39(8):1231–1244.
[19] Heffes H. A class of data traffic processes – covariance function characterization and related queuing results. Bell System Technical Journal. 1980;59(6):897–929.
[20] Heffes H, Lucantoni D. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. IEEE Journal on Selected Areas in Communications. 1986;4(6):856–868.
[21] Kim SH, Whitt W. Are call center and hospital arrivals well modeled by nonhomogeneous Poisson processes? Manufacturing & Service Operations Management. 2014;16(3):464–480.
[22] Sriram K, Whitt W. Characterizing superposition arrival processes in packet multiplexers for voice and data. IEEE Journal on Selected Areas in Communications. 1986;4(6):833–846.
[23] Lindvall T. Lectures on the coupling method. New York (NY): Wiley; 2002.
[24] Brockwell PJ, Davis RA. Time series: theory and methods. 2nd ed. New York (NY): Springer; 1991.
[25] Basawa IV, Bhat UN, Lund R. Maximum likelihood estimation for single server queues from waiting time data. Queueing Systems. 1996;24(1):155–167.
[26] Gonçalves FB, Franklin P. On the definition of likelihood function. arXiv preprint arXiv:190610733. 2019;.
[27] Genz A, Trinh G. Numerical computation of multivariate normal probabilities using bivariate conditioning. In: Monte Carlo and quasi-Monte Carlo methods. New York (NY): Springer; 2016. p. 289–302.
[28] Tong YL. The multivariate normal distribution. New York (NY): Springer; 2012.
[29] Masarotto G, Varin C. Gaussian copula regression in R. Journal of Statistical Software. 2017;77(8):1–26.
[30] Han Z, De Oliveira V. gckrig: An R package for the analysis of geostatistical count data using Gaussian copulas. Journal of Statistical Software. 2018;87(1):1–32.
[31] Tjøstheim D, Paulsen J. Bias of some commonly-used time series estimates. Biometrika. 1983;70(2):389–399.

[32] Djurić PM, Bugallo MF. Particle filtering for high-dimensional systems. In: 2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP); IEEE; 2013. p. 352–355.
[33] Naesseth C, Lindsten F, Schon T. Nested sequential Monte Carlo methods. In: International Conference on Machine Learning; PMLR; 2015. p. 1292–1301.
[34] Rebeschini P, Van Handel R. Can local particle filters beat the curse of dimensionality? The Annals of Applied Probability. 2015;25(5):2809–2866.
[35] Grant J, Kong J, Lund R, et al. Modeling seasonal cryosopheric processes with Lindley random walks. In Preparation. 2024;.
[36] Zeger SL, Brookmeyer R. Regression analysis with censored autocorrelated data. Journal of the American Statistical Association. 1986;81(395):722–729.
[37] Cai Z. Estimating a distribution function for censored time series data. Journal of Multivariate Analysis. 2001;78(2):299–318.
[38] Brandt A, Franken P, Lisek B. Stationary stochastic models. New York (NY): Wiley; 1990.
[39] Vlasiou M. Lindley-type recursions [dissertation]. Eindhoven, the Netherlands: Eindhoven University of Technology; 2006.

Appendix A. Markov Property Proofs

Lemma .1. *The general Lindley walk $\{X_n\}_{n=0}^{\infty}$ is not Markov of any order.*

Proof. To construct a counterexample, let $\{C_n\}$ be a general stationary Gaussian process. We consider the stable case where $\mu_C < 0$ and let $x_0, \ldots, x_{t-p-1}$ be arbitrary strictly positive feasible values for the process. Consider the conditional probability

$$P[X_t \leq y \mid X_{t-1} = 0, \ldots, X_{t-p} = 0; X_{t-p-1} = x_{t-p-1}, \ldots, X_0 = x_0] \tag{1}$$

(this takes $x_{t-p} = \cdots = x_{t-1} = 0$).

We now write this probability strictly in terms of the $\{C_n\}$ process. By the Lindley recursion and the fact that $X_{t-1} = 0$, $\{X_t \leq y\} = \{C_t \leq y\}$ for all $y \geq 0$. Also, the event $[X_{t-1} = 0, \ldots, X_{t-p} = 0; X_{t-p-1} = x_{t-p-1}, \ldots, X_0 = x_0]$ can be written in terms of the C_ts via

$$\left[\cap_{i=1}^{t-p-1} C_i = x_i - x_{i-1} \cap C_{t-p} \leq -x_{t-p-1} \cap_{i=1}^{p-1} C_{t-p+i} \leq 0\right] \tag{2}$$

(more is said about this in the future sections). It follows that the probability in (1) equals

$$P\left[C_t \leq y \mid \cap_{i=1}^{t-p-1} C_i = x_i - x_{i-1} \cap C_{t-p} \leq -x_{t-p-1} \cap_{i=1}^{p-1} C_{t-p+i} \leq 0\right]. \tag{3}$$

For a general Gaussian stationary process, the conditional probability in (3) depends on x_{t-p-1} (and the previous x_ts too). This follows directly from the property that conditional distributions of multivariate normal quantities are again multivariate normal (see Proposition 1.6.6 in [24] for the explicit form). Therefore, $\{X_t\}$ cannot be pth order Markov.

Even in cases where $\{C_t\}$ is first order Markov — say an AR(1) series — the conditional probability in (3) can be shown to depend on x_{t-p-1} when $\phi \neq 0$ due to the fact that we are conditioning on some non-singleton sets in (2) (this takes additional work to see, which we do not provide here). □

Proposition .2. *The process $\{(X_n, C_n)\}$ is a pth order Markov chain when $\{C_n\}$ is a pth order Markov chain (such as the above AR(p) series).*

Proof. For clarity, we first provide the argument for the case where $p = 1$. For this p, the Lindley recursion gives

$$X_{n+1} = \max(X_n + C_{n+1}, 0) := h(X_n, C_{n+1})$$

for the measurable function h defined by $h(x, y) = \max(x + y, 0)$. From this and the first order Markov property of $\{C_n\}$, the joint dynamics of (X_{n+1}, C_{n+1}), given the entire history $C_1, \ldots, C_n$ and $X_0, \ldots, X_n$, are described solely by X_n and C_n — we do not need $C_1, \ldots, C_{n-1}$ or $X_0, X_1, \ldots, X_{n-1}$. The conclusion now follows.

When $p > 1$, merely extend the above logic by applying the Lindley recursion p times to get

$$X_{n+1} = h(X_{n-p+1}, C_{n-p+2}, \ldots, C_n, C_{n+1})$$

for a measurable function $h : \mathbb{R}^{p+1} \to \mathbb{R}$ (the form of h is not important, but one may wish to compare to (8) and argue as above). While here, we note that given that past p X_ns and C_ns, X_{n+1} can be written in a form that does not involve $X_{n-p+2}, \ldots, X_n$, but rather only X_{n-p+1} and $C_{n-p+2}, \ldots, C_{n+1}$. □

Appendix B. The Likelihood for AR(p) Changes

We now derive the likelihood when $\{C_n\}$ is an autoregressive process of order p satisfying the AR(p) recursion in (6) — the process need not be Gaussian. The Markov structure identified in Section 3 effectively reduces the integral dimension in (10). Our goal is to explicitly derive the likelihood in terms of its free parameters, which are $\mathbf{\Theta} = (\mu, \sigma^2, \phi_1, \ldots, \phi_p)'$.

Recall that if $X_n > 0$, the value of C_n can be recovered; namely, $C_n = X_n - X_{n-1}$; when $X_n = 0$, we only know that C_n is less than or equal to $-X_{n-1}$, but we do not know its exact value. Previous authors [12,36,37] have viewed this problem as a censored time series issue. The censoring here is not simple; indeed, when $X_n = 0$, the values of C_n are censored depending on $-X_{n-1}$, which is not constant in time and also depends on the past history of the process.

Our derivation partitions the series into segments where C_n is "*recoverable*" or not. We take X_n as observed for all $n \in \{0, \ldots, N\}$. While the derivation is somewhat tedious, this is expected given the difficulties encountered in likelihood evaluation in [12,36] and [37]. Define the first *ariser* time as

$$\kappa_1 := \min_{n>p}\{n : X_n > 0, \ldots, X_{n-p+1} > 0\},$$

which is the first time that p consecutive changes are recoverable. The first *plunger* time is set to

$$\tau_1 := \min_{n>\kappa_1}\{n : X_n = 0\}.$$

For $i \geq 1$, define successive ariser and plunger times as

$$\kappa_{i+1} := \min\{n > \tau_i : X_n > 0, \ldots, X_{n-p+1} > 0\}, \quad \tau_{i+1} := \min\{n > \kappa_{i+1} : X_n = 0\}.$$

Cases where κ_i or τ_i do not occur in $\{1, 2, \ldots, N\}$ are addressed below.

Let $K(N) := \max\{i : \kappa_i \leq N\}$ denote the observed number of plunger times in $\{1, 2, \ldots, N\}$. The i^{th} complete regime of the process contains all times in $R_i := \{\kappa_i + 1, \ldots, \tau_i, \ldots, \kappa_{i+1}\}$ for $i = 1, \ldots, K(N) - 1$. For boundary conditions, set $R_0 = \{1, \ldots, \kappa_1\}$ and $R_{K(N)} = \{\kappa_{K(N)} + 1, \ldots, N\}$ if $k(N) < N$; otherwise, set $R_{K(N)} = \emptyset$. Note that R_0 does not contain X_0. It is not possible to recover C_0 since X_{-1} is unobserved. Likewise, if κ_i does not exist (occur), set $R_0 = \{1, \ldots, N\}$. This blocks the observations into distinct regimes via its ariser times.

For notation, let $A = \{n_1, n_2, \ldots, n_k\}$ for $n_1 < n_2 < \cdots < n_k$ denote k ordered index times in $\{1, \ldots, N\}$. Let $\mathbf{X}_A = (X_{n_1}, \ldots, X_{n_k})'$ denote a $k \times 1$ vector of the ordered process values occurring over $n_i \in A$. The notation $\mathbf{X}_{A-1} = (X_{n_1-1}, \ldots, X_{n_k-1})'$ is used. For convenience, let $\mathbf{X}_n = (X_0, X_1, \ldots, X_n)'$. We use the same notation for the $\{C_n\}$ process: $\mathbf{C}_A = (C_{n_1}, \ldots, C_{n_k})'$ for $A = \{n_1, n_2, \ldots, n_k\}$ and $\mathbf{C}_n = (C_1, \ldots, C_n)'$. For realized values, lowercase notation is used, e.g., $\mathbf{x}_n = (x_0, x_1, \ldots, x_n)'$ and $\mathbf{c}_n = (c_1, \ldots, c_n)'$. For each $n \in \mathbb{N}$, let $n(p) = \{n-p+1, \ldots, n\}$ denote the p consecutive time points ending at time n. For two random variables/vectors Y and Z, the conditional "density" of Y given $Z = z$ is denoted by $f_{Y|Z=z}(\cdot \mid Z = z)$.

At time $n = \kappa_i$, $C_{\kappa_i}, \ldots, C_{\kappa_i - p+1}$ are recoverable with

$$\mathbf{C}_{\kappa_i(p)} = \mathbf{X}_{\kappa_i(p)} - \mathbf{X}_{\kappa_i(p)-1} = (X_{\kappa_i-p+1} - X_{\kappa_i-p}, \ldots, X_{\kappa_i} - X_{\kappa_i-1})',$$

where $\kappa_i(p) = \{k_i - p + 1, \ldots, \kappa_i\}$. Define the ith set of *good times* (recoverable times) G_i, which is a subset of R_i, as

$$G_i = \{n : \kappa_i < n < \tau_i\} = \{k_i + 1, \ldots, \tau_i - 1\},$$

where for each $n \in G_i$, $X_n, \ldots, X_{n-p} > 0$, implying that $C_n, \ldots, C_{n-p}$ are all recoverable. We use the convention $G_i^c = \{n : n \in R_i \cap n \notin G_i\} = \{\tau_i, \ldots, \kappa_{i+1}\}$ for ease of exposition. That is, the complement of the good times of the ith regime only contains times in the ith regime.

Let $\mathcal{L}_{G_i|\kappa_i}(\cdot \mid \mathbf{X}_{\kappa_i})$ denote the conditional distribution of $\mathbf{X}_{G_i}$ given the past $\mathbf{X}_{\kappa_i}$. When $G_i \neq \emptyset$, the change of variables formula and the pth order Markov property of $\{C_n\}$ yield

$$\begin{aligned}
&\mathcal{L}_{G_i|\kappa_i}(\mathbf{X}_{G_i} \mid \mathbf{X}_{\kappa_i}) \qquad (4)\\
&= f_{\mathbf{X}_{G_i}|\mathbf{X}_{\kappa_i}}(\mathbf{X}_{G_i} \mid \mathbf{X}_{\kappa_i})\\
&= f_{\mathbf{X}_{G_i}|\mathbf{C}_{\kappa_i(p)},\mathbf{X}_{\kappa_i-p}}(\mathbf{X}_{G_i} \mid \mathbf{X}_{\kappa_i(p)} - \mathbf{X}_{\kappa_i(p)-1}, \mathbf{X}_{\kappa_i-p})\\
&= f_{\mathbf{C}_{G_i}|\mathbf{C}_{\kappa_i(p)},\mathbf{X}_{\kappa_i-p}}(\mathbf{X}_{G_i} - \mathbf{X}_{G_i-1} \mid \mathbf{X}_{\kappa_i(p)} - \mathbf{X}_{\kappa_i(p)-1}, \mathbf{X}_{\kappa_i-p})\\
&= f_{\mathbf{C}_{G_i}|\mathbf{C}_{\kappa_i(p)}}(\mathbf{X}_{G_i} - \mathbf{X}_{G_i-1} \mid \mathbf{X}_{\kappa_i(p)} - \mathbf{X}_{\kappa_i(p)-1}), \qquad (5)
\end{aligned}$$

where $f_{\mathbf{C}_{G_i}|\mathbf{C}_{\kappa_i(p)}}(\cdot \mid \cdot)$ is the conditional "density" of the changes during the good times G_i in R_i conditional on the p consecutive changes $\mathbf{C}_{\kappa_i(p)}$ ending at the ariser time κ_i prior to the start of R_i. The p^{th} order Markov property of $\{C_n\}$ yields

$$\begin{aligned}
&f_{\mathbf{C}_{G_i}|\mathbf{C}_{\kappa_i(p)}}(\mathbf{X}_{G_i} - \mathbf{X}_{G_i-1} \mid \mathbf{X}_{\kappa_i(p)} - \mathbf{X}_{\kappa_i(p)-1}) =\\
&\prod_{n \in G_i} f_{C_n|\mathbf{C}_{(n-1)(p)}}(X_n - X_{n-1} \mid \mathbf{X}_{(n-1)(p)} - \mathbf{X}_{(n-1)(p)-1}),
\end{aligned}$$

where the notation has $(n-1)(p) := \{n-p, \ldots, n-1\}$. If $G_i = \emptyset$, the convention $f_{\mathbf{X}_{G_i}|\mathbf{X}_{\kappa_i}}(\cdot) = 1$ is assumed.

We will further decompose the times in G_i^c into two sets. The first set collects times where the observations are positive, the other set containing the times where the observations are zero. More precisely, define $\eta_i \subset G_i^c$ as $\eta_i = \{n \in G_i^c : x_n > 0\}$. Then η_i contains the times in R_i where the C_n are recoverable, but the previous p changes are not all recoverable. Likewise, define $\eta_i^c = \{n \in G_i^c : X_n = 0\}$ as those times in R_i where C_n is not recoverable and not all of the previous p changes are recoverable. Our definitions partition each R_i into $R_i = G_i \cup \eta_i \cup \eta_i^c$.

To study the conditional distribution of $\mathbf{X}_{G_i^c}$, we need to identify the zeros in the observations $(X_{\tau_i}, \ldots, X_{\kappa_{i+1}})'$. Let $z_1^i < \ldots < z_{L_i}^i$ denote the indices where those zeros occur. Clearly, $z_1^i = \tau_i$ and $z_{L_i}^i = \kappa_{i+1} - p$. Then $\eta_i = G_i^c / \{z_1^i, \ldots, z_{L_i}^i\}$ and $\eta_i^c = \{z_1^i, \ldots, z_{L_i}^i\}$. Let $\mathcal{L}_{G_i^c|(\tau_i-1)}(\cdot \mid \mathbf{X}_{\tau_i-1})$ denote the distribution of $\mathbf{X}_{G_i^c}$ given $\mathbf{X}_{\tau_i-1}$. Similar to the analysis that produced (10), and using the Markov property of $\{C_n\}$,

$$\begin{aligned}
&\mathcal{L}_{G_i^c|(\tau_i-1)}(\mathbf{X}_{G_i^c} \mid \mathbf{X}_{\tau_i-1}) \\
&= \int_{-\infty}^{-X_{z_1^i-1}} \cdots \int_{-\infty}^{-X_{z_{L_i}^i-1}} f_{\mathbf{C}_{G_i^c}|\mathbf{C}_{(\tau_i-1)(p)}}(\mathbf{X}_{\eta_i} - \mathbf{X}_{\eta_i-1}, \mathbf{c} \mid \\
&\qquad\qquad \mathbf{X}_{(\tau_i-1)(p)} - \mathbf{X}_{(\tau_i-1)(p)-1}) d\mathbf{c}, \qquad (6)
\end{aligned}$$

where the integration domain is $\{\mathbf{c} \in \mathbb{R}^{L_i} : C_n \leq -X_{n-1}, n \in \eta_i^c\}$ and the notation $(\tau_i - 1)(p) = \{\tau_i - p, \ldots, \tau_i - 1\}$ is used. The conditional density in (6) can also be written as a product of the conditional densities $f_{C_n|\mathbf{C}_{(n-1)(p)}}(\cdot)$ for $n \in G_i^c$ (we omit details). If $G_i^c = \emptyset$, then set $\mathcal{L}_{G_i^c|(\tau_i-1)}(\mathbf{X}_{G_i^c} \mid \mathbf{X}_{\tau_i-1})$ to unity (this can only happen in R_0 or $R_{K(N)}$).

It remains to consider the 0^{th} "startup regime". The distribution of X_0, which is the stationary distribution of the process, is absolutely continuous away from zero and has a point mass at zero. While the distribution of X_0 does not have an explicit form, its moment properties are studied in [38] and [39]. Until the first time n such that p consecutive C_ns are observed (which happens at the time $n = \kappa_1$), the joint distribution of $(X_0, X_1, X_2, \ldots, X_{\kappa_1})'$ depends on the joint distribution of $(X_0, C_1, \ldots, C_{\kappa_1})'$ according to (10). More precisely, let $\eta_0 = \{n \in R_0 : X_n > 0\}$ and $\eta_0^c := \{n \in R_0 : X_n = 0\} = \{z_1, \ldots, z_L\}$. Then from (10), the likelihood of $\mathbf{X}_{R_0}$ is

$$\mathcal{L}_{R_0}(\mathbf{x}_{R_0}) = \int_{-\infty}^{-X_{z_1-1}} \cdots \int_{-\infty}^{-X_{z_L-1}} f_{X_0, \mathbf{C}_{\kappa_1}}(X_0, X_m - X_{m-1}, m \in \eta_0, \mathbf{c}) d\mathbf{c}, \qquad (7)$$

where $\mathbf{C}_{\kappa_1} = (C_1, \ldots, C_{\kappa_1})'$ and the integration domain is $\{\mathbf{c} \in \mathbb{R}^L : c_n \leq -X_{n-1}, n \in \eta_0^c\}$.

As a summary of the above, combining (5), (6), and (7) produces our AR(p) likelihood as

$$\mathcal{L}(\boldsymbol{\Theta} \mid \mathbf{X}) = \mathcal{L}_{R_0}(\mathbf{X}_{R_0}) \times \left(\prod_{i=1}^{K(N)} \mathcal{L}_{G_i|\kappa_i}(\mathbf{X}_{G_i} \mid \mathbf{X}_{\kappa_i}) \times \mathcal{L}_{G_i^c|\tau_i-1}(\mathbf{X}_{G_i^c} \mid \mathbf{X}_{\tau_i-1}) \right). \qquad (8)$$

			Model 1 ($\mu = -0.5$)			Model 2 ($\mu = -0.25$)		
ϕ	n		$\hat{\phi}$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\mu}$	$\hat{\sigma}^2$
-0.5	100	mean	-0.4799	-0.5153	0.9894	-0.4936	-0.2644	0.9886
		SD	0.0999	0.1223	0.1154	0.0944	0.0782	0.0912
		$\hat{E}(I'(\theta)^2)$	0.1028	0.1111	0.1166	0.0935	0.0812	0.0912
	250	mean	-0.4852	-0.5043	1.0041	-0.5000	-0.2563	0.9977
		SD	0.0606	0.0774	0.0750	0.0578	0.0526	0.0564
		$\hat{E}(I'(\theta)^2)$	0.0629	0.0685	0.0726	0.0585	0.0505	0.0571
	500	mean	-0.4786	-0.4982	1.0120	-0.4996	-0.2523	1.0023
		SD	0.0411	0.0498	0.0534	0.0419	0.0372	0.0387
		$\hat{E}(I'(\theta)^2)$	0.0433	0.0482	0.0512	0.0410	0.0355	0.0402
-0.25	100	mean	-0.2672	-0.5191	0.9868	-0.2474	-0.2480	0.9926
		SD	0.1243	0.1319	0.1237	0.1080	0.0915	0.0886
		$\hat{E}(I'(\theta)^2)$	0.1203	0.1207	0.1172	0.1071	0.0925	0.0897
	250	mean	-0.2536	-0.4934	0.9801	-0.2455	-0.2621	0.9964
		SD	0.0708	0.0721	0.0666	0.0702	0.0529	0.0584
		$\hat{E}(I'(\theta)^2)$	0.0755	0.0728	0.0715	0.0681	0.0585	0.0571
	500	mean	-0.2559	-0.5078	1.0025	-0.2487	-0.2502	0.9981
		SD	0.0489	0.0523	0.0532	0.0468	0.0372	0.0424
		$\hat{E}(I'(\theta)^2)$	0.0530	0.0525	0.0517	0.0478	0.0408	0.0399
0	100	mean	0.0022	-0.5182	0.9905	-0.0041	-0.2493	0.9706
		SD	0.1374	0.1456	0.1108	0.1072	0.1140	0.0868
		$\hat{E}(I'(\theta)^2)$	0.1282	0.1374	0.1126	0.1133	0.1092	0.0870
	250	mean	-0.0140	-0.5043	0.9944	-0.0004	-0.2543	0.9959
		SD	0.0890	0.0920	0.0762	0.0737	0.0721	0.0566
		$\hat{E}(I'(\theta)^2)$	0.0809	0.0833	0.0703	0.0713	0.0698	0.0557
	500	mean	-0.0012	-0.5090	1.0019	0.0000	-0.2502	1.0017
		SD	0.0532	0.0520	0.0491	0.0494	0.0543	0.0384
		$\hat{E}(I'(\theta)^2)$	0.0574	0.0594	0.0499	0.0503	0.0493	0.0393
0.25	100	mean	0.2405	-0.4988	0.9892	0.2220	-0.2609	0.9860
		SD	0.1194	0.1639	0.1026	0.1113	0.1426	0.0927
		$\hat{E}(I'(\theta)^2)$	0.1239	0.1626	0.1047	0.1113	0.1405	0.0869
	250	mean	0.2416	-0.5071	0.9921	0.2483	-0.2450	0.9875
		SD	0.0774	0.1032	0.0712	0.0637	0.0881	0.0520
		$\hat{E}(I'(\theta)^2)$	0.0787	0.1011	0.0660	0.0691	0.0891	0.0534
	500	mean	0.2454	-0.4962	0.9944	0.2464	-0.2566	1.0010
		SD	0.0512	0.0717	0.0461	0.0512	0.0672	0.0377
		$\hat{E}(I'(\theta)^2)$	0.0549	0.0707	0.0462	0.0489	0.0636	0.0383
0.5	100	mean	0.4702	-0.5056	0.9867	0.4689	-0.2466	0.9793
		SD	0.1064	0.2134	0.1019	0.1020	0.1932	0.0937
		$\hat{E}(I'(\theta)^2)$	0.1099	0.2194	0.0980	0.1001	0.2010	0.0837
	250	mean	0.4941	-0.5148	0.9955	0.4947	-0.2683	0.9942
		SD	0.0639	0.1289	0.0606	0.0631	0.1250	0.0537
		$\hat{E}(I'(\theta)^2)$	0.0679	0.1403	0.0614	0.0616	0.1317	0.0527
	500	mean	0.4863	-0.4908	1.0009	0.4948	-0.2444	0.9968
		SD	0.0468	0.0977	0.0443	0.0440	0.0906	0.0380
		$\hat{E}(I'(\theta)^2)$	0.0461	0.0962	0.0428	0.0429	0.0920	0.0366
0.75	100	mean	0.7119	-0.5585	0.9867	0.7053	-0.2611	0.9992
		SD	0.0885	0.3957	0.0992	0.0897	0.4161	0.0940
		$\hat{E}(I'(\theta)^2)$	0.0840	0.3987	0.0935	0.0786	0.3804	0.0857
	250	mean	0.7319	-0.5093	0.9966	0.7292	-0.2988	0.9981
		SD	0.0535	0.2795	0.0592	0.0462	0.2706	0.0540
		$\hat{E}(I'(\theta)^2)$	0.0497	0.2564	0.0567	0.0474	0.2445	0.0519
	500	mean	0.7332	-0.4629	0.9951	0.7384	-0.2398	0.9997
		SD	0.0371	0.1673	0.0404	0.0337	0.1548	0.0373
		$\hat{E}(I'(\theta)^2)$	0.0333	0.1728	0.0386	0.0319	0.1751	0.0354

Table 1. Standard errors for the Lindley walk with an AR(1) $\{C_n\}$. The results report the sample standard deviation (SD) of the particle filtering parameter estimators from the 200 independently generated series, and the average of the 200 standard errors obtained by inverting the Hessian matrix ($\hat{E}(I'(\theta)^2)$) at the maximum likelihood estimate over these same runs. Both standard errors roughly agree.

			Model 3 ($\mu = -0.5$)				Model 4 ($\mu = -0.25$)			
ϕ_1	n		$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\mu}$	$\hat{\sigma}^2$
0.1	100	mean	0.0887	0.1436	-0.5035	0.9854	0.0704	0.1551	-0.2722	0.9827
		SD	0.1436	0.1247	0.1704	0.1084	0.1242	0.1168	0.1647	0.0900
		$\hat{E}(I'(\theta)^2)$	0.1252	0.1295	0.1686	0.1077	0.1133	0.1150	0.1489	0.0895
	250	mean	0.0807	0.1825	-0.4988	0.9848	0.0955	0.1756	-0.2450	0.9876
		SD	0.0767	0.0727	0.1025	0.0698	0.0784	0.0712	0.0880	0.0528
		$\hat{E}(I'(\theta)^2)$	0.0769	0.0793	0.1037	0.0664	0.0691	0.0701	0.0932	0.0540
	500	mean	0.0915	0.1900	-0.4886	0.9913	0.0992	0.1930	-0.2520	0.9953
		SD	0.0563	0.0553	0.0710	0.0454	0.0498	0.0489	0.0696	0.0404
		$\hat{E}(I'(\theta)^2)$	0.0534	0.0548	0.0732	0.0464	0.0485	0.0492	0.0672	0.0382
0.3	100	mean	0.2867	0.1428	-0.5187	0.9873	0.2871	0.1613	-0.2611	0.9789
		SD	0.1348	0.1334	0.2383	0.1115	0.1196	0.1107	0.1950	0.0882
		$\hat{E}(I'(\theta)^2)$	0.1281	0.1272	0.2249	0.1036	0.1138	0.1144	0.2017	0.0863
	250	mean	0.2912	0.1806	-0.4832	0.9804	0.2917	0.1968	-0.2453	0.9899
		SD	0.0829	0.0850	0.1350	0.0651	0.0700	0.0771	0.1262	0.0521
		$\hat{E}(I'(\theta)^2)$	0.0763	0.0773	0.1345	0.0613	0.0690	0.0696	0.1307	0.0527
	500	mean	0.2970	0.1897	-0.5006	0.9964	0.2997	0.1900	-0.2562	0.9967
		SD	0.0536	0.0499	0.1064	0.0442	0.0469	0.0509	0.0933	0.0349
		$\hat{E}(I'(\theta)^2)$	0.0529	0.0534	0.0968	0.0439	0.0485	0.0488	0.0917	0.0373

Table 2. Standard errors for the Lindley walk with an AR(2) $\{C_n\}$. The results report the sample standard deviation (SD) of the particle filtering parameter estimators from the 200 independently generated series, and the average of the 200 standard errors obtained by inverting the Hessian matrix ($\hat{E}(I'(\theta)^2)$) at the maximum likelihood estimate over these same runs. Both standard errors roughly agree. Both Model 3 and Model 4 fix $\phi_2 = 0.2$.