A Framework for the Analysis of Unevenly Spaced Time Series Data

Andreas Eckner*

July 20, 2014

Abstract

This paper presents methods for analyzing and manipulating unevenly spaced time series without a transformation to equally spaced data. Processing and analyzing such data in its unaltered form avoids the biases and information loss caused by resampling. Care is taken to develop a framework consistent with a traditional analysis of equally spaced data, as in Brockwell and Davis (1991), Hamilton (1994) and Box, Jenkins, and Reinsel (2004).

Keywords: time series analysis, unevenly spaced time series, unequally spaced time series, irregularly spaced time series, moving average, time series operator, LTI systems

1 Introduction

Unevenly spaced (also called unequally or irregularly spaced) time series data naturally occurs in many industrial and scientific domains. Natural disasters such as earthquakes, floods, or volcanic eruptions typically occur at irregular time intervals. In observational astronomy, measurements of properties such as the spectra of celestial objects are taken at times determined by seasonal and weather conditions, and availability of observation time slots. In clinical trials (or more generally, longitudinal studies), a patient's state of health may be observed only at irregular time intervals, and different patients are usually observed at different points in time. There are many more examples in climatology, ecology, economics, finance, geology, meteorology, and network traffic analysis. In addition, two special cases of equally spaced time series are naturally treated as unevenly spaced time series: time series with missing observations; and multivariate data sets that consist of time series with different frequencies, even if the observations of the individual time series are reported at regular intervals.

Research on time series analysis is usually specialized along one or more of the following dimensions: univariate vs. multivariate, linear vs. non-linear, parametric vs. non-parametric, and evenly vs. unevenly spaced. There is an extensive body of literature on the analysis of equally spaced data along the first three dimensions; see, for example, Tong (1990), Brockwell and Davis (1991), Hamilton (1994), Brockwell and Davis (2002), Fan and Yao (2003), Box et al. (2004), and Lütkepohl (2010). Much of the basic theory was developed when limitations in computing resources favored the analysis of equally spaced data

*Comments are welcome at andreas@eckner.com.

(and the use of linear Gaussian models); because in such cases, efficient linear algebra routines can be used, and many such problems have explicit solutions.

Fewer methods exists specifically for the analysis of unevenly spaced time series data. Some authors have suggested an embedding into continuous diffusion processes; see Jones (1981), Jones (1985), Belcher et al. (1994) and Brockwell (2008). The emphasis of this literature has primarily been on modeling univariate autoregressive-moving-average (ARMA) processes, as opposed to developing a full-fledged tool set analogous to the one available for equally spaced data. In astronomy, in particular, much effort has been devoted to the task of estimating the spectrum of irregular time series data; see Lomb (1975), Scargle (1982), Bos et al. (2002), Thiebaut and Roques (2005), and Broersen (2008). Müller (1991), Gilles Zumbach (2001), and Dacorogna et al. (2001) examined unevenly spaced time series in the context of high-frequency financial data.

Perhaps the most widely used approach is to transform unevenly spaced data into equally spaced data using some form of interpolation - most often linear - and then to apply existing methods developed for equally spaced data. For examples, see Adorf (1995), and Beygelzimer et al. (2005). Transforming data in such a way, however, has significant drawbacks, examples of which are provided at the end of the introduction.

This paper provides methods for directly analyzing unevenly spaced time series data while maintaining consistency with existing methods developed for equally spaced time series. In particular, a major goal is to provide a unified mathematical and conceptual framework for manipulating univariate and multivariate time series with unevenly spaced observations.

The remainder of the introduction lists some of the drawbacks of transforming unevenly spaced data into equally spaced data. Section 2 introduces unevenly spaced time series and some elementary operations for them. Section 3 defines time series operators and examines their commonly encountered structural features. Section 4 introduces convolution operators, which form a particularly tractable and widely used class of time series operators. Section 5 presents arithmetic, return, and rolling operators as examples for the application of the theory developed in the preceding sections. Section 7 turns to multivariate time series and associated vector time series operators. Moving averages, which summarize the average value of a time series over a certain horizon, are examined in Section 8. Section 9 focuses on the scale and volatility of time series, while an Appendix summarizes frequently used notation.

1.1 Disadvantages of Resampling

Before starting our discussion of unevenly spaced time series, it is helpful to keep in mind a few scenarios where a transformation to equally spaced data has significant disadvantages.

Example 1.1 (Bias) *Let B be standard Brownian motion and $0 \leq a < t < b$. A straightforward calculation shows that the distribution of B_t conditional on B_a and B_b is $N(\mu, \sigma^2)$ with*

$$\mu = \frac{b-t}{b-a} B_a + \frac{t-a}{b-a} B_b,$$

and

$$\sigma^2 = \frac{(t-a)\,(b-t)}{b-a}.$$

Sampling with linear interpolation implicitly reduces this conditional distribution to a single deterministic value, or equivalently, ignores the stochasticity around the conditional mean. Hence, if methods for equally spaced time series analysis are applied to linearly interpolated data, estimates of second moments; such as volatilities, autocorrelations, and covariances; may be subject to a significant and hard-to-quantify bias. See Scholes and Williams (1977), Lundin et al. (1999), Hayashi and Yoshida (2005), and Rehfeld et al. (2011) for examples.

Example 1.2 (Causality) *For a time series, the linearly interpolated observation at time t (not equal to an actual observation time) depends on the value of the previous and subsequent observation. Hence, while the data-generating process may be adapted to a certain filtration $(\mathcal{F}_t)$,[1] the linearly interpolated process will generally not be adapted to $(\mathcal{F}_t)$. This effect may change the causal relationships in a multivariate time series. Furthermore, the estimated predictive ability of a time series model may be severely biased.*

Example 1.3 (Data Loss and Dilution) *Converting an unevenly spaced time series to an equally spaced time series reduces and dilutes the information content of a data set. First, data points are omitted if consecutive observations lie close together, thus causing statistical inference to be less efficient. Second, redundant data points are introduced if consecutive observations lie far part, thus biasing estimates of statistical significance.*

Example 1.4 (Time Information) *In certain applications, the spacing of observations may in itself be of interest and contain relevant information above and beyond the information contained in an interpolated time series. For example, the transaction and quote (TAQ) data from the New York Stock Exchange (NYSE) contains all trades and quotes of listed, and certain non-listed, stocks during any given trading day.[2] The frequency of the arrival of new quotes and transactions is an integral part of the price formation process, determining the level and volatility of security prices.*
In signal processing, the probability of having an observation during a given time interval may depend on the instantaneous amplitude of the studied signal.

2 The Basic Framework

An unevenly spaced time series is a sequence of observation time and value pairs (t_n, X_n) with strictly increasing observation times. This notion is made precise by the following

Definition 2.1 *For $n \geq 1$, we call*

(i) $\mathbb{T}_n = \{(t_1 < t_2 < \ldots < t_n) : t_k \in \mathbb{R}, 1 \leq k \leq n\}$ *the space of strictly increasing time sequences of length n,*

(ii) $\mathbb{T} = \cup_{n=1}^{\infty} \mathbb{T}_n$ *the space of strictly increasing time sequences,*

(iii) $\mathbb{R}^n$ *the space of observation values for n observations,*

(iv) $\mathcal{T}_n = \mathbb{T}_n \times \mathbb{R}^n$ *the space of real-valued, unevenly spaced time series of length n, and*

[1] For a precise mathematical definition not offered here, see Karatzas and Shreve (2004) and Protter (2005).
[2] See www.nyxdata.com/Data-Products/Historical-Data for a detailed description.

(v) $\mathcal{T} = \cup_{n=1}^{\infty} \mathcal{T}_n$ *the space of (real-valued) (unevenly spaced) time series.*

Definition 2.2 *For a time series* $X \in \mathcal{T}$*, we denote*

(i) $N(X)$ *to be the number of observations of* X*, particularly so that* $X \in \mathcal{T}_{N(X)}$*;*

(ii) $T(X) = \left(t_1, \ldots, t_{N(X)}\right)$ *to be the sequence of observation times (of* X*); and*

(iii) $V(X) = \left(X_1, \ldots, X_{N(X)}\right)$ *to be the sequence of observation values (of* X*).*

We will frequently use the informal but compact notation $((t_n, X_n) : 1 \le n \le N(X))$ and $(X_{t_n} : 1 \le n \le N(X))$ to denote a time series $X \in \mathcal{T}$ with observation times $\left(t_1, \ldots, t_{N(X)}\right)$ and observation values $\left(X_1, \ldots, X_{N(X)}\right)$.[3] Now that we have defined unevenly spaced time series, we will introduce methods for extracting basic information from them.

Definition 2.3 *For a time series* $X \in \mathcal{T}$ *and point in time* $t \in \mathbb{R}$ *(not necessarily an observation time), the most recent observation time is*

$$\mathrm{Prev}^X(t) \equiv \mathrm{Prev}(T(X), t) = \begin{cases} \max(s : s \le t, s \in T(X)), & \text{if } t \ge \min(T(X)) \\ \min(T(X)), & \text{otherwise} \end{cases}$$

and the next available observation time is

$$\mathrm{Next}^X(t) \equiv \mathrm{Next}(T(X), t) = \begin{cases} \min(s : s \ge t, s \in T(X)), & \text{if } t \le \max(T(X)) \\ +\infty, & \text{otherwise.} \end{cases}$$

For $\min T(x) \le t \le \max T(X)$, $\mathrm{Prev}^X(t) < t < \mathrm{Next}^X(t)$; unless $t \in T(X)$, in which case t is both the most recent and the next available observation time.

Definition 2.4 (Sampling) *For* $X \in \mathcal{T}$ *and* $t \in \mathbb{R}$*,* $X[t] = X_{\mathrm{Prev}(X,t)}$ *is the sampled value of* X *at time* t*,* $X[t]_{\mathrm{next}} = X_{\mathrm{Next}^X(t)}$ *is the next value of* X *at time* t*, and* $X[t]_{\mathrm{lin}} = \left(1 - \omega^X(t)\right) X_{\mathrm{Prev}(X,t)} + \omega^X(t)\, X_{\mathrm{Next}^X(t)}$ *where*

$$\omega^X(t) = \omega(T(X), t) = \begin{cases} \frac{t - \mathrm{Prev}^X(t)}{\mathrm{Next}^X(t) - \mathrm{Prev}^X(t)}, & \text{if } 0 < \mathrm{Next}^X(t) - \mathrm{Prev}^X(t) < \infty \\ 0, & \text{otherwise} \end{cases}$$

is the linearly interpolated value of X *at time* t*. These sampling schemes are called last-point, next-point, and linear interpolation, respectively.*

Note that the most recently available observation time *before* the first observation is taken to be the first observation time. As a consequence, the sampled value of a time series X *before* the first observation is equal to the first observation value. While potentially inappropriate for some applications, this convention greatly simplifies notation and avoids the treatment of a multitude of special cases in the exposition below.[4]

[3] For equally spaced time series, the reader may be used to using language like "the third observation" of a time series X. For unevenly spaced time series, it is often necessary to distinguish between the third observation *value*, X_{t_3}, and the third observation tuple, or simply the third observation, (t_3, X_3), of a time series.

[4] A software implementation might instead use a special symbol to denote a value that is not available. For example, R (www.r-project.org) uses the constant `NA`, which propagates through all steps of an analysis because the result of any calculation involving `NA`s is also `NA`.

Remark 2.5 *Fix a time series $X \in \mathcal{T}$. Then,*

(i) $X[t] = X[t]_{\mathrm{next}} = X[t]_{\mathrm{lin}} = X_t$ *for* $t \in T(X)$*;*

(ii) $X[t]$ *and* $X[t]_{\mathrm{next}}$*, as functions of* t*, are right-continuous piecewise-constant functions with finite number of discontinuities; and*

(iii) $X[t]_{\mathrm{lin}}$ *as a function of* t *is a continuous piecewise-linear function.*

These observations suggest an alternative way of defining unevenly spaced time series; namely, as either piecewise-constant or piecewise-linear functions $X : \mathbb{R} \to \mathbb{R}$. Such a representation, however, cannot capture the occurrence of identical consecutive observations, and therefore ignores potentially important time series information. Moreover, such a framework does not naturally lend itself to interpreting an unevenly spaced time series as a discretely-observed continuous-time stochastic process, thereby ruling out a large class of data-generating processes.

Definition 2.6 (Simultaneous Sampling) *Fix an observation time sequence* $T_X \in \mathbb{T}$*.*

(i) For a time series $X \in \mathcal{T}$ *and sampling scheme* $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$*, we call*

$$X_\sigma [T_X] = ((t_i, X_\sigma [t_i]) : t_i \in T_X)$$

the sampled time series of X *(using sampling times* T*).*

(ii) For a continuous-time stochastic process X^c*, we call*

$$X^c [T] = \left(\left(t_i, X^c_{t_i}\right) : t_i \in T_X\right)$$

the observation time series of X^c *(at observation times* T_X*).*

In particular, $X[T(X)] = X$ for all time series $X \in \mathcal{T}$.

Definition 2.7 *For* $X \in \mathcal{T}$*, we call*

(i) $\Delta t(X) = ((t_{n+1}, t_{n+1} - t_n) : 1 \leq n \leq N(X) - 1)$ *the time series of tick (or observation time) spacings (of* X*);*

(ii) $X\{s,t\} = ((t_n, X_n) : s \leq t_n \leq t, 1 \leq n \leq N(X))$ *for* $s \leq t$ *the subperiod time series (of* X*) in* $[s,t]$*;*

(iii) $B(X) = ((t_{n+1}, X_n) : 1 \leq n \leq N(X) - 1)$ *the backshifted time series (of* X*), and* B *the backshift operator;*

(iv) $L(X, \tau) = ((t_n + \tau, X_n) : 1 \leq n \leq N(X))$ *the lagged time series (of* X*) with lag* $\tau \in \mathbb{R}$*, and* L *the lag operator; and*

(v) $D(X, \tau) = ((t_n, X[t_n - \tau]) : 1 \leq n \leq N(X)) = L(X[T(X) - \tau], -\tau)$ *the delayed time series (of* X*) with delay* $\tau \in \mathbb{R}$*, and* D *the delay operator.*

A time series X is equally spaced, if the observation values of $\Delta t\,(X)$ are all equal to a constant $c > 0$. For such a time series and for $\tau = c$, the backshift operator is identical to the lag operator (apart from the last observation) and to the delay operator (apart from the first observation). In particular, the backshift, delay, and lag operator are identical for an equally spaced time series with time index $t \in \mathbb{Z}$. These transformations are genuinely different for unevenly spaced data, however, because the backshift operator shifts observation *values*, while the lag operator shifts observation *times*.[5]

Example 2.8 *Let X be a time series of length three with observation times $T\,(X) = (0, 2, 5)$ and observation values $V\,(X) = (1, -1, 2.5)$. Then*

$$X[t] = \begin{cases} 1, & \text{for } t < 2 \\ -1, & \text{for } 2 \le t < 5 \\ 2.5, & \text{for } t \ge 5 \end{cases} \qquad X[t]_{\text{next}} = \begin{cases} 1, & \text{for } t \le 0 \\ -1, & \text{for } 0 < t \le 2 \\ 2.5, & \text{for } t > 2 \end{cases}$$

$$X_{\text{lin}}[t] = \begin{cases} 1, & \text{for } t < 0 \\ 1 - t, & \text{for } 0 \le t < 2 \\ \frac{7}{6}t - \frac{10}{3}, & \text{for } 2 \le t < 5 \\ 2.5, & \text{for } t \ge 5. \end{cases}$$

These three graphs are shown in Figure 1. Moreover,

$$\Delta t\,(X)\,[s] = \begin{cases} 2, & \text{for } s < 5 \\ 3, & \text{for } s \ge 5 \end{cases}$$

because $T\,(\Delta t\,(X)) = (2, 5)$ and $V\,(\Delta t\,(X)) = (2, 3)$, and

$$B\,(X)\,[t] = \begin{cases} 1, & \text{for } t < 5 \\ -1, & \text{for } t \ge 5 \end{cases} \qquad L\,(X, 1)\,[t] = \begin{cases} 1, & \text{for } t < 3 \\ -1, & \text{for } 3 \le t < 6 \\ 2.5, & \text{for } t \ge 6. \end{cases}$$

The following result, which elaborates on the relationship between the lag operator L and the sampling operator, will be heavily used throughout the paper.

Lemma 2.9 *For $X \in \mathcal{T}$ and $\tau \in \mathbb{R}$,*

(i) $T\,(L\,(X, \tau)) = T\,(X) + \tau$,

(ii) $L\,(X, \tau)_{t+\tau} = X_t$ *for* $t \in T\,(X)$,

(iii) $L\,(X, \tau)_t = X_{t-\tau}$ *for* $t \in T\,(L\,(X, \tau))$, *and*

(iv) $L\,(X, \tau)\,[t]_\sigma = X[t - \tau]_\sigma$ *for $t \in \mathbb{R}$ and sampling scheme σ. In other words, depending on the sign of τ, the lag operator shifts the sample path of a time series either backward or forward in time.*

[5]The difference between the lag and delay operator is that the former shifts the information filtration of observation *times and values*, while the latter shifts only the information filtration of observation *values*.

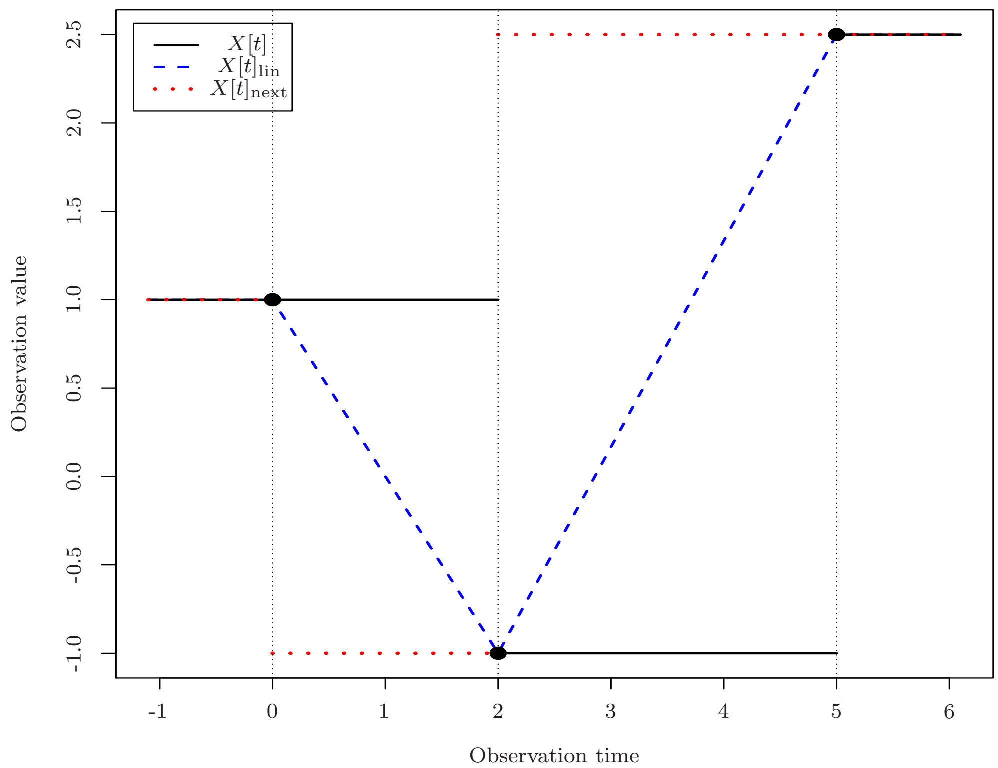

Figure 1: The sampling-scheme-dependent graph of the time series X with $T(X) = (0, 2, 5)$ and $V(X) = (1, -1, 2.5)$. The three observation time-value pairs are denoted by solid black circles.

Proof. Relationships (i) and (ii) follow directly from the definition of the lag operator, while (iii) follows by combining (i) and (ii). For (iv), we note that

$$\text{Prev}^{L(X,\tau)}(t) = \text{Prev}\left(T\left(L\left(X,\tau\right)\right),t\right) = \text{Prev}\left(T\left(X\right)+\tau,t\right) = \text{Prev}\left(T\left(X\right),t-\tau\right)+\tau$$

where the second equality follows from (i). Hence,

$$\begin{aligned} L\left(X,\tau\right)[t] &= L\left(X,\tau\right)_{\text{Prev}^{L(X,\tau)}(t)} \\ &= L\left(X,\tau\right)_{\text{Prev}(T(X),t-\tau)+\tau} \\ &= X_{\text{Prev}(T(X),t-\tau)} \\ &= X\left[t-\tau\right], \end{aligned}$$

where the third equality follows from (ii). The proofs for the other two sampling schemes are similar. ■

At this point, the reader might want to check her understanding by verifying the identity $X = L\left(L\left(X,\tau\right)\left[T\left(X\right)+\tau\right],-\tau\right)$ for all time series $X \in \mathcal{T}$ and $\tau \in \mathbb{R}$.

3 Time Series Operators

Time series operators (also called "systems" in the signal processing literature) take a time series as input and leave as output a transformed time series. We already encountered a few operators; such as the backshift (B), subperiod ($\{\}$), and tick spacing operator (Δ); in the previous section.

The key difference between time series operators for evenly and unevenly spaced observations is that, for the latter, the observation values in the transformed series can depend on the spacing of observation times. This interaction between observation values and observation times calls for a careful analysis and classification of the structure of such operators.

Definition 3.1 *A time series operator is a mapping* $O : \mathcal{T} \to \mathcal{T}$*; or equivalently, a pair of mappings* (O_T, O_V)*; where* $O_T : \mathcal{T} \to \mathbb{T}$ *is the transformation of observation times,* $O_V : \mathcal{T} \to \cup_{n \geq 1} \mathbb{R}^n$ *is the transformation of observation values, and* $|O_T(X)| = |O_V(X)|$ *for all* $X \in \mathcal{T}$*.*

The constraint at the end of the definition ensures that the number of observation values and observation times for the transformed time series agree. In particular, using this notation, we have $T(O(X)) = O_T(X)$ and $V(O(X)) = O_V(X)$ for time series $X \in \mathcal{T}$ and a time series operator O.

Example 3.2 *Fix a sampling scheme* $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$ *and observation time sequence* $T \in \mathbb{T}$*. The mapping that assigns each time series* $X \in \mathcal{T}$ *the sampled time series* $X[T]_\sigma$ *(see Definition 2.6) is a time series operator.*

The above definition of a time series operator is completely general. In practice, most operators share one or more of the structural features listed below.

3.1 Tick Invariance

In many cases, the observation times of the transformed time series are identical to those of the original time series.

Definition 3.3 *A time series operator* O *is said to be tick invariant if* $T(O(X)) = T(X)$ *for all* $X \in \mathcal{T}$*.*

Such operators cover the vast majority of the cases in presented in this paper. Notable exceptions are the lag operator L and various resampling schemes. For some results, a weaker property than tick-invariance is sufficient.

Definition 3.4 *A time series operator* O *is said to be lag free if* $\max T(O(X)) \leq \max T(X)$ *for all* $X \in \mathcal{T}$*.*

In other words, for lag-free operators, when one has finished observing an input time series, the entire output time series is also observable at that point in time.

3.2 Causality

Definition 3.5 *A time series operator* O *is said to be causal (or adapted) if*

$$O(X)\{-\infty, t\} = O(X\{-\infty, t\})\{-\infty, t\} \tag{3.1}$$

for all $X \in \mathcal{T}$ *and* $t \in \mathbb{R}$*.*

In other words, a time series operator is causal if the output up to each time point t depends only on the input up to that time.[6] Equivalently, if two time series are identical up to some time, then a causal transformation produces transformed time series that are likewise identical up to the same time.

In many cases, the following convenient characterization is useful:

Lemma 3.6 *A time series operator O is causal and lag free if and only if the order of O and the subperiod operator is interchangeable; that is*

$$O\left(X\right)\left\{-\infty,t\right\}=O\left(X\left\{-\infty,t\right\}\right) \tag{3.2}$$

for all $X\in\mathcal{T}$ and $t\in\mathbb{R}$.

Proof.

$\Longrightarrow$ If O is lag-free, then

$$\max T\left(O\left(X\left\{-\infty,t\right\}\right)\right)\leq\max T\left(X\left\{-\infty,t\right\}\right)\leq t \tag{3.3}$$

which implies that $O\left(X\left\{-\infty,t\right\}\right)\left\{-\infty,t\right\}=O\left(X\left\{-\infty,t\right\}\right)$ and combined with (3.1) shows (3.2).

$\Longleftarrow$ Applying $\max T\left(.\right)$ to both sides of (3.2) gives

$$\max T\left(O\left(X\left\{-\infty,t\right\}\right)\right)\leq t$$

for all $X\in\mathcal{T}$ and $t\in\mathbb{R}$. Setting $t=\max T\left(X\right)$ shows that O is lag-free. Hence (3.1) follows from (3.2) and (3.3).

■

In particular, the order of a causal, tick-invariant operator and the subperiod operator is interchangeable.

3.3 Shift or Time Invariance

For many data sets, only the relative, but not the absolute, positions of observation times are of interest; and this property should be reflected in the analysis that is carried out.

Definition 3.7 *A time series operator O is said to be shift invariant (or time invariant), if the order of O and the lag operator L is interchangeable. In other words, for all $X\in\mathcal{T}$ and $\tau\in\mathbb{R}$,*

$$O\left(L\left(X,\tau\right)\right)=L\left(O\left(X\right),\tau\right); \tag{3.4}$$

or equivalently,

$$O_T\left(L\left(X,\tau\right)\right)=O_T\left(X\right)+\tau,\ \textit{and} \tag{3.5}$$
$$O_V\left(L\left(X,\tau\right)\right)=O_V\left(X\right). \tag{3.6}$$

[6] If X was a stochastic process instead of a fixed time series, we would call an operator O adapted if $O\left(X\right)$ was adapted to the filtration generated by X.

Lemma 3.8 *A time series operator O is shift invariant if and only if*

$$O\left(X\right)_t = O\left(L\left(X, -t\right)\right)_0 \tag{3.7}$$

for all $X \in \mathcal{T}$ and $t \in T\left(O\left(X\right)\right)$.

Proof.

$\Longrightarrow$ Setting $\tau = -t$ in (3.4) and sampling at time zero gives

$$\begin{aligned} O\left(L\left(X, -t\right)\right)_0 &= L\left(O\left(X\right), -t\right)_0 \\ &= O\left(X\right)_t, \end{aligned}$$

where the second equality follows from Lemma 2.9 (ii).

$\Longleftarrow$ Fix $X \in \mathcal{T}$, $t \in T\left(O\left(L\left(X, \tau\right)\right)\right)$, and $\tau \in \mathbb{R}$. If (3.7) holds, then it particularly holds for $L\left(X, t\right)$ and $L\left(X, t-\tau\right)$, which gives

$$\begin{aligned} O\left(L\left(X, \tau\right)\right)_t &= O\left(L\left(L\left(X, \tau\right), -t\right)\right)_0 \\ &= O\left(L\left(X, \tau - t\right)\right)_0 \\ &= O\left(L\left(X, -\left(t - \tau\right)\right)\right)_0 \\ &= O\left(X\right)_{t-\tau} \\ &= L\left(O\left(X, \tau\right)\right)_t, \end{aligned}$$

where we use Lemma 2.9 (iii) for the last equality.

■

In particular, a (by definition $\mathcal{T}-$valued) time series operator that is shift invariant can be represented by a single real-valued function. Specifically, $O\left(X\right)_t = g\left(X, t\right)$ for $t \in T\left(O\left(X\right)\right)$ and $g : \mathcal{T} \times \mathbb{R} \to \mathbb{R}$ given by $g\left(Y, t\right) = O\left(L\left(Y, -t\right)\right)_0$.

3.4 Time Scale Invariance

The measurement unit of the time scale is generally not of interest in an analysis of time series data. Therefore, we usually focus on (families of) time series operators that are invariant under a rescaling of the time axis.

Definition 3.9 *For $X \in \mathcal{T}$ and $a > 0$, the time-scaling operator S_a is defined as*

$$S_a\left(X\right) = S\left(X, a\right) = \left(\left(a t_n, X_n\right) : 1 \le n \le N(X)\right).$$

Lemma 3.10 *For $X \in \mathcal{T}$ and $a > 0$,*

(i) $T\left(S_a\left(X\right)\right) = aT\left(X\right)$,

(ii) $S_a\left(X\right)_{at} = X_t$ *for* $t \in T\left(X\right)$,

(iii) $S_a\left(X\right)_t = X_{t/a}$ *for* $t \in T\left(S_a\left(X\right)\right)$, *and*

(iv) $S_a(X)[t]_\sigma = X[t/a]_\sigma$ *for* $t \in \mathbb{R}$ *and sampling scheme* σ*. In other words, depending on the sign of* $a-1$*, the time-scaling operator either compresses or stretches the sample path of a time series.*

Proof. The proof is very similar to that of Lemma 2.9. ■

Certain sets of time series operators are naturally indexed by one or more parameters. For example, simple moving averages (see Section 8.1) can be indexed by the length of the moving average time window.

Definition 3.11 *A family of time series operators* $\{O_\tau : \tau \in (0, \infty)\}$ *is said to be timescale invariant if*

$$O_\tau(X) = S_{1/a}(O_{\tau a}(S_a(X)))$$

for all $X \in \mathcal{T}$*,* $a > 0$*, and* $\tau \in (0, \infty)$*.*

The families of rolling return operators (Section 5.2), simple moving averages (Section 8.1), and exponential moving averages (Section 8.2) are timescale invariant.

3.5 Homogeneity

Scaling properties are of interest not only for the time scale, but also for the observation value scale.

Definition 3.12 *A time series operator* O *is said to be homogeneous of degree* $d \geq 0$ *if* $O(aX) = a^d O(X)$ *for all* $X \in \mathcal{T}$ *and* $a > 0$*.*[7]

The tick-spacing operator Δt is homogeneous of degree $d = 0$, moving averages are homogeneous of degree $d = 1$, and the operator that calculates the integrated $p-$variation of a time series is homogeneous of degree $d = p$.

Remark 3.13 *The sampling operator of Example 3.2 is linear (see Section 6), homogeneous of degree* $d = 1$*, and causal iff we use last-point sampling; and is neither tick invariant, shift invariant, nor timescale invariant.*

4 Convolution Operators

Convolution operators are a class of tick-invariant, causal, shift-invariant (and often homogeneous) time series operators that are particularly tractable. To this end, recall that a signed measure on a measurable space (Ω, Σ) is a measurable function μ that satisfies (i) $\mu(\varnothing) = 0$ and if $A = +_i A_i$ is a countable disjoint union of sets in Σ with either $\sum_i \mu(A_i)^- < \infty$ or $\sum_i \mu(A_i)^+ < \infty$, then also (ii) $\mu(A) = \sum_i \mu(A_i)$.

Proposition 4.1 *If* μ *is a signed measure on* (Ω, Σ) *that is absolutely continuous with respect to a* $\sigma-$*finite measure* ν*, then there exists a function (density)* $f : \Omega \to \mathbb{R}$ *such that*

$$\mu(A) = \int_A f(x)\, d\nu(x)$$

for all $A \in \Sigma$*.*

[7] For a time series X and scalar $a \in \mathbb{R}$, aX denotes the time series that results by multiplying each observation value of X by a. See Section 5.1 for a systematic treatment of time series arithmetic.

This result is a consequence of the Jordan decomposition and the Radon Nikodym theorem. See, for example, Section 32 in Billingsley (1995) or Appendix A.8 in Durrett (2005) for details.

Definition 4.2 *A (univariate) time series kernel μ is a signed measure on $(\mathbb{R} \times \mathbb{R}_+, \mathcal{B} \otimes \mathcal{B}_+)$[8] with*

$$\int_0^\infty |d\mu\left(f\left(s\right), s\right)| < \infty$$

for all bounded piecewise linear functions $f : \mathbb{R}_+ \to \mathbb{R}$, where the integration is over the time variable s.

The integrability condition ensures that the value of a convolution is well-defined and finite. In particular, the condition is satisfied if $d\mu\left(x, s\right) = g\left(x\right) d\mu_T\left(s\right)$ for some real function g and finite signed measure μ_T on $\mathbb{R}_+$, which is indeed the case throughout this paper.

Definition 4.3 (Convolution Operator) *For a time series $X \in \mathcal{T}$, kernel μ, and sampling scheme $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$, the convolution $*_\sigma^\mu\left(X\right) = X *_\sigma \mu$ is a time series with*

$$T\left(X *_\sigma \mu\right) = T\left(X\right), \tag{4.8}$$

$$\left(X *_\sigma \mu\right)_t = \int_0^\infty d\mu\left(X[t-s]_\sigma, s\right), \qquad t \in T\left(X *_\sigma \mu\right). \tag{4.9}$$

If μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R} \times \mathbb{R}_+$, then (4.9) can be written as

$$\left(X *_\sigma \mu\right)_t = \int_0^\infty f\left(X[t-s]_\sigma, s\right) ds, \qquad t \in T\left(X *_\sigma \mu\right), \tag{4.10}$$

where f is the density function of μ.

Remark 4.4 (Discrete Time Analog) *The discrete-time analogs to convolution operators are additive (but generally nonlinear), causal, time-invariant filters of the form*

$$Y_n = \sum_{k=0}^{\infty} f_k\left(X_{n-k}\right),$$

where $(X_n : n \in \mathbb{Z})$ is a stationary time series, and $f_0, f_1, \ldots$ are real-valued functions subject to certain integrability conditions. As discussed in Section 6, the representation simplifies further in the case of linearity.

In the remainder of the paper, we primarily focus on last-point sampling and therefore often omit the "σ" from the notation of a convolution operator. All of our results also hold for the other two sampling schemes, but we do not always provide separate derivations.

Proposition 4.5 *The convolution operator $*^\mu$, associated with a (univariate) time series kernel μ, is tick invariant, causal, and shift invariant.*

[8] $\mathcal{B} \otimes \mathcal{B}_+$ is the Borel $\sigma-$algebra on $\mathbb{R} \times \mathbb{R}_+$.

Proof. Tick-invariance and causality follow immediately from the definition of a convolution operator. Shift-invariance can be shown either using Lemma 3.8 or directly, which we do here for illustrative purposes. To this end, let $X \in \mathcal{T}$ and $\tau > 0$. Using (4.8) twice we get

$$(*^{\mu})_T (L(X,\tau)) = T(L(X,\tau)) = T(X) + \tau = T(X * \mu) + \tau = (*^{\mu})_T (X) + \tau,$$

showing that $*^{\mu}$ satisfies (3.5). On the other hand, for $t \in T(*^{\mu}(L(X,\tau))) = T(X) + \tau$,

$$\begin{aligned}(X * \mu)_{t-\tau} &= \int_0^{\infty} d\mu\left(X[(t-\tau)-s], s\right) \\ &= \int_0^{\infty} d\mu\left(L(X,\tau)[t-s], s\right) \\ &= (L(X,\tau) * \mu)_t\,,\end{aligned}$$

where we use Lemma 2.9 (*iv*) for the second equality. Hence, we also have $(*^{\mu})_V (X) = (*^{\mu})_V (L(X,\tau))$ and $*^{\mu}$ is therefore shift invariant. ■

Not all time series operators that are tick invariant, causal, and shift invariant can, however, be expressed as convolution operators. For example, the operator that calculates the smallest observation value in a rolling time window (see Section 5.3) is not of this form.

As the following example illustrates, when a time series operator can be expressed as a convolution operator, the associated kernel is not unique.

Example 4.6 *Let O be the operator that sets all observation values of a time series equal to zero, that is, $T(O(X)) = T(X)$ and $O(X)_t = 0$ for all $t \in T(O(X))$ and $X \in \mathcal{T}$. Then $O(X) = *^{\mu}(X)$ for all $X \in \mathcal{T}$ for the following kernels:*

(i) $\mu(x,s) = 0$,

(ii) $\mu(x,s) = \mathbf{1}_{\{s \in N\}}$ *where $N \in \mathcal{B}_+$ is a Lebesgue null set, and*

(iii) $\mu(x,s) = \mathbf{1}_{\{f(s)=x\}}$ *where $f : \mathbb{R}_+ \to \mathbb{R}$ satisfies $\lambda\left(f^{-1}(\{x\})\right) = 0$ for all $x \in \mathbb{R}$.*[9]

When we say that the kernel of a convolution operator has a certain structure, we imply that one of the equivalent kernels is of that structure.

5 Examples

This section gives examples of univariate time series operators that can be expressed either as convolution operators or as simple transformations of them.

5.1 Time Series Arithmetic

It is straightforward to extend the four basic arithmetic operations of mathematics - addition, subtraction, multiplication and division - to unevenly spaced time series.

Definition 5.1 (Arithmetic Operations) *For a time series $X \in \mathcal{T}$ and $c \in \mathbb{R}$, we call*

[9]In other words, f is a function that assumes each value in $\mathbb{R}$ at most on a null set of time points. Examples are $f(s) = \sin(s)$, $f(s) = \exp(-s)$, $f(s) = s^2$.

(i) $c+X$ *(or* $X+c$*) with* $T(c+X) = T(X)$ *and* $V(c+X) = \left(c+X_1, \ldots, c+X_{N(X)}\right)$ *"the sum of* c *and* X*,"*

(ii) cX *(or* Xc*) with* $T(cX) = T(X)$ *and* $V(cX) = \left(cX_1, \ldots, cX_{N(X)}\right)$ *"the product of* c *and* X*," and*

(iii) $1/X$ *with* $T(1/X) = T(X)$ *and* $V(1/X) = \left(1/X_1, \ldots, 1/X_{N(X)}\right)$ *"the inverse of* X*," provided that all observation values of* X *are non-zero.*

Definition 5.2 (Arithmetic Time Series Operations) *For time series* $X, Y \in \mathcal{T}$ *and sampling scheme* $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$*, we call*

(i) $X +_\sigma Y$ *with* $T(X +_\sigma Y) = T(X) \cup T(Y)$ *and* $(X +_\sigma T)_t = X[t]_\sigma + Y[t]_\sigma$ *for* $t \in T(X +_\sigma Y)$ *"the sum of* X *and* Y*," and*

(ii) $X_\sigma Y$ *with* $T(X_\sigma Y) = T(X) \cup T(Y)$ *and* $(X_\sigma Y)_t = X[t]_\sigma Y[t]_\sigma$ *for* $t \in T(X_\sigma Y)$ *"the product of* X *and* Y*,"*

where $T_X \cup T_Y$ *for* $T_X, T_Y \in \mathbb{T}$ *denotes the sorted union of* T_X *and* T_Y*.*

Proposition 5.3 *The arithmetic operators in Definition 5.1 are convolution operators with kernel* $\mu(x,s) = (c+x)\,\delta_0(s)$*,* $\mu(x,s) = cx\delta_0(s)$*, and* $\mu(x,s) = \delta_0(s)/x$*, respectively.*

Proof. For $X \in \mathcal{T}$, $c \in \mathbb{R}$, and $\mu(x,s) = (c+x)\,\delta_0(s)$, by definition, $T(X * \mu) = T(X) = T(c+X)$. For $t \in T(X * \mu)$,

$$
\begin{aligned}
(X * \mu)_t &= \int_0^\infty d\mu\left(X[t-s], s\right) \\
&= \int_0^\infty \left(c + X[t-s]\right) \delta_0(s)\, ds \\
&= c + X_t,
\end{aligned}
$$

and, therefore, also $V(X * \mu) = V(c+X)$. The reasoning for the other two kernels is similar. ■

Proposition 5.4 *The sampling operator is linear, that is,*

$$
(aX +_\sigma bY)[t]_\sigma = aX[t]_\sigma + bY[t]_\sigma
$$

for time series $X, Y \in \mathcal{T}$*, sampling scheme* σ*, and* $a, b \in \mathbb{R}$*.*

Proof. The result follows directly from the definition of the corresponding arithmetic operations. For example, in the case of last-point sampling,

$$
\begin{aligned}
(aX + bY)[t] &= (aX + bY)_{\mathrm{Prev}(T(aX+bY),t)} \\
&= ((aX) + (bY))_{\mathrm{Prev}(T(X) \cup T(Y),t)} \\
&= (aX)\left[\mathrm{Prev}\left(T(X) \cup T(Y), t\right)\right] + (bY)\left[\mathrm{Prev}\left(T(X) \cup T(Y), t\right)\right] \\
&= aX\left[\mathrm{Prev}\left(T(X) \cup T(Y), t\right)\right] + bY\left[\mathrm{Prev}\left(T(X) \cup T(Y), t\right)\right] \\
&= aX_{\mathrm{Prev}^X(t)} + bY_{\mathrm{Prev}^Y(t)} \\
&= aX[t] + bY[t],
\end{aligned}
$$

where we used

$$\text{Prev}^X\left(\text{Prev}\left(T(X) \cup T\left(Y\right), t\right)\right) = \text{Prev}^X\left(t\right)$$

for the second to last equality. ■

Note that in general, $1/X\,[t]_{\text{lin}}$ does not equal $(1/X)\,[t]_{\text{lin}}$ for $X \in \mathcal{T}$ and $t \notin T\left(X\right)$, although equality holds for the other two sampling schemes. Similarly, $\left(X_\sigma Y\right)[t]_{\text{lin}}$ usually does not equal $X\,[t]_{\text{lin}}\,Y\,[t]_{\text{lin}}$ for $X, Y \in \mathcal{T}$ and $t \notin T\left(X\right)$, but equality holds for the other two sampling schemes.

What does "the sum of X and Y" actually mean for two time series X and Y that are not synchronized; or what would we want it to mean? If X and Y are discretely observed realizations of continuous-time stochastic processes X^c and Y^c, respectively, we might want $\left(X+Y\right)[t]$ to be a "good guess" of $(X^c+Y^c)_t$, given all available information. The following example describes two settings where this desirable property holds.

Example 5.5 *Let X^c and Y^c be two independent continuous-time stochastic processes, $T_X, T_Y \in \mathbb{T}$ be fixed observation time sequences, and $X = X^c\,[T_X]$ and $Y = Y^c\,[T_Y]$ be the corresponding observation time series of X^c and Y^c, respectively. Furthermore, let $\mathcal{F}_t = \sigma(X_s : s \leq t)$ and $\mathcal{G}_t = \sigma(Y_s : s \leq t)$ denote the filtration generated by X and Y, respectively, and $\mathcal{H}_t = \mathcal{F}_t \cup \mathcal{G}_t$.*

(i) If X and Y are martingales, then

$$E\left(\left(X^c+Y^c\right)_t \mid \mathcal{H}_t\right) = \left(X+Y\right)[t] \tag{5.11}$$

for all $t \geq \max\left(\min T_X, \min T_Y\right)$, that is, for time points for which both X and Y have at least one past observation.

(ii) If X^c and Y^c are Lévy processes (but not necessarily martingales),[10] *then*

$$E\left(\left(X^c+Y^c\right)_t \mid \mathcal{H}_\infty\right) = X[t]_{\text{lin}} + Y\,[t]_{\text{lin}} = \left(X +_{\text{lin}} Y\right)[t] \tag{5.12}$$

for all $t \in \mathbb{R}$.[11]

Proof. If X^c and Y^c are independent martingales, then

$$\begin{aligned}
E\left(\left(X^c+Y^c\right)_t \mid \mathcal{H}_t\right) &= E\left(X_t^c \mid \mathcal{H}_t\right) + E\left(Y_t^c \mid \mathcal{H}_t\right) \\
&= E\left(X_t^c \mid \mathcal{F}_t\right) + E\left(Y_t^c \mid \mathcal{G}_t\right) \\
&= E\left(X_t^c \mid \mathcal{F}_{\text{Prev}^X(t)}\right) + E\left(Y_t^c \mid \mathcal{G}_{\text{Prev}^Y(t)}\right) \\
&= X^c_{\text{Prev}^X(t)} + Y^c_{\text{Prev}^Y(t)} \\
&= X[t] + Y\,[t] \\
&= \left(X+Y\right)[t],
\end{aligned}$$

[10] A Lévy process has independent and stationary increments, and is continuous in probability. Brownian motion and homogenous Poisson processes are special cases. See Sato (1999) and Protter (2005) for details.

[11] If X^c and Y^c are martingales without stationary increments, then hardly anything can be said about the conditional expectation on the left-hand side of (5.12) without additional information about X^c and Y^c. The LHS could even be larger than the largest observation value of $X +_{\text{lin}} Y$. Details are available from the author upon request.

showing (5.11). For a Lévy process Z and times $s \leq t \leq r$, $Z_r - Z_s$ has an infinitely divisible distribution, and the conditional expectation $E\left(Z_t \middle| Z_s, Z_r\right)$ is therefore the linear interpolation of (s, Z_s) and (r, Z_r) evaluated at time t. Hence, (5.12) follows from similar reasoning as (5.11). ■

Apart from the four basic arithmetic operations, other scalar mathematical transformations can also be extended to unevenly spaced time series. For $X \in \mathcal{T}$ and function $f : \mathbb{R} \to \mathbb{R}$, we call $f(X)$ the time series that results when applying the function f to each observation value of X. For example,

$$\exp(X) = ((t_n, \exp(X_n)) : 1 \leq n \leq N(X)).$$

Elementwise time series operators are convolution operators with kernel $\mu(x, s) = f(x)\,\delta_0(s)$.

5.2 Return Calculations

In many applications, it is of interest to either analyze or report the change of a time series over a fixed time horizon; such as one day, month, or year. This section examines how to calculate such time series returns.

Definition 5.6 *For time series $X, Y \in \mathcal{T}$, we call*

(i) $\Delta^k X = \Delta\left(\Delta^{k-1} X\right)$ *for* $k \in \mathbb{N}$ *with* $\Delta X = ((t_n, X_n - X_{n-1}) : 1 \leq n \leq N(X) - 1)$ *for* $k > 0$ *and* $\Delta^0 X = X$ *the* $k - th$ *order difference time series (of X), and*

(ii) $\mathrm{diff}_\gamma(X, Y)$ *with scale* $\gamma \in \{\text{abs, rel, log}\}$ *the absolute/relative/log difference between X and Y, where*

$$\mathrm{diff}_\gamma(X, Y) = \begin{cases} X - Y, & \text{if } \gamma = \text{abs}, \\ \frac{X}{Y} - 1, & \text{if } \gamma = \text{rel}, \\ \log\left(\frac{X}{Y}\right), & \text{if } \gamma = \text{log}, \end{cases}$$

provided that X and Y are positive-valued time series for $\gamma \in \{\text{rel, log}\}$.[12]

Definition 5.7 (Returns) *For a time series $X \in \mathcal{T}$, time horizon $\tau > 0$, and return scale* $\gamma \in \{\text{abs, rel, log}\}$, *we call*

(i) $\mathrm{ret}_\gamma^{\mathrm{roll}}(X, \tau) = \mathrm{diff}_\gamma(X, D(X, \tau))$ *the rolling absolute/relative/log return time series (of X over horizon τ),*

(ii) $\mathrm{ret}_\gamma^{\mathrm{obs}}(X) = \mathrm{diff}_\gamma(X, B(X))$ *the absolute/relative/log observation (or tick) return time series (of X),*

(iii) $\mathrm{ret}_\gamma^{\mathrm{SMA}}(X, \tau) = \mathrm{diff}_\gamma(X,\ \mathrm{SMA}(X, \tau))$ *the absolute/relative/log simple moving average (SMA) return time series (of X over horizon τ), and*

(iv) $\mathrm{ret}_\gamma^{\mathrm{EMA}}(X, \tau) = \mathrm{diff}_\gamma(X,\ \mathrm{EMA}(X, \tau))$ *the absolute/relative/log exponential moving average (EMA) return time series (of X over horizon τ),*

[12] Again, we use an analogous definition for other sampling schemes. For example, $\mathrm{diff}_{\mathrm{abs,lin}}(X, Y) = X -_{\mathrm{lin}} Y$ denotes the absolute, linearly interpolated difference between X and Y.

provided that X is a positive-valued time series for $\gamma \in \{\mathrm{rel}, \log\}$. See Section 8 for details regarding the moving average operators SMA *and* EMA*.*

The interpretation of the first two return definitions is immediately clear, while Section 8.3 provides a motivation for the last two definitions.

Proposition 5.8 *The return operators* $\mathrm{ret}^{\mathrm{roll}}$, $\mathrm{ret}^{\mathrm{SMA}}$, *and* $\mathrm{ret}^{\mathrm{EMA}}$ *are either convolution operators or convolution operators combined with simple transformations.*

Proof. For a time series $X \in \mathcal{T}$ and time horizon $\tau > 0$, it is easy to see that

$$\mathrm{diff}_\gamma(X, D(X,\tau)) = \begin{cases} X * \mu, & \text{if } \gamma \in \{\mathrm{abs}, \log\}, \\ \exp(X * \mu) - 1, & \text{if } \gamma = \mathrm{rel}, \end{cases}$$

with

$$\mu(x,s) = \begin{cases} x(\delta_0(s) - \delta_\tau(s)), & \text{if } \gamma = \mathrm{abs}, \\ \log(x)(\delta_0(s) - \delta_\tau(s)), & \text{if } \gamma \in \{\mathrm{rel}, \log\}, \end{cases}$$

provided that X is a positive-valued time series for $\gamma \in \{\mathrm{rel}, \log\}$. Using their respective definitions in Section 8, the proofs for the other two return operators are similar. ■

5.3 Rolling Time Series Functions

In many cases, it is desirable to extract a certain piece of local information about a time series. Rolling time series functions allow us to do just that.

Definition 5.9 (Rolling Time Series Functions) *Assume that we are given a function $f : \mathcal{T} \to \mathbb{R}$ that is shift invariant in the sense that $f(X) = f(L(X,\eta))$ for all $\eta \in \mathbb{R}$ and $X \in \mathcal{T}$. For a time series $X \in \mathcal{T}$ and time horizon $\tau \in \mathbb{R}_+ \cup \{\infty\}$, the "rolling function f of X over horizon τ," denoted by* $\mathrm{roll}(X, f, \tau)$, *is the time series with*

$$\begin{aligned} T(\mathrm{roll}(X,f,\tau)) &= T(X), \\ \mathrm{roll}(X,f,\tau)_t &= f(X\{t-\tau, t\}), \qquad t \in T(\mathrm{roll}(X,f,\tau)). \end{aligned}$$

Proposition 5.10 *The class of rolling time series operators is identical to the class of causal, shift-invariant, and tick-invariant operators.*

Proof.

$\Longrightarrow$ By construction, every rolling time series function is causal and tick-invariant. Furthermore, shift-invariance of f implies

$$\begin{aligned} \mathrm{roll}(X,f,\tau)_t &= f(X\{t-\tau,t\}) \\ &= f(L(X,\eta)\{t+\eta-\tau, t+\eta\}) \\ &= \mathrm{roll}(L(X,\eta), f, \tau)_{t+\eta}, \end{aligned}$$

or equivalently, $V(\mathrm{roll}(X,f,\tau)) = V(\mathrm{roll}(L(X,\eta), f, \tau))$ for all $X \in \mathcal{T}$ and $\eta \in \mathbb{R}$. Hence, $\mathrm{roll}(., f, \tau)$ is also shift invariant.

$\Longleftarrow$ If O is a causal, shift-invariant, and tick-invariant time series operator, then

$$O(X)_t = O(X\{-\infty, t\})_t.$$

Hence, if we define $f : \mathcal{T} \to \mathbb{R}$ as $f(X) = O(X)_{\max T(X)}$, then we have $O(X)_t = \text{roll}(X, f, -\infty)_t$ for all $X \in \mathcal{T}$ and $t \in T(X)$. We are left to show that f is shift-invariant. Using Lemma 2.9 (*ii*) applied to $O(X)$ and the shift-invariance of O,

$$\begin{aligned} f(X) &= O(X)_{\max T(X)} \\ &= L(O(X), \eta)_{\max T(X)+\eta} \\ &= O(L(X, \eta))_{\max T(X)+\eta} \\ &= O(L(X, \eta))_{\max T(L(X,\eta))} \\ &= f(L(X, \eta)) \end{aligned}$$

for all $\eta \in \mathbb{R}$ and $X \in \mathcal{T}$.

■

In particular, rolling time series functions include convolution operators. Many operators, however, that cannot be expressed as convolution operators are included as well:

Example 5.11 *For a time series $X \in \mathcal{T}$, horizon $\tau \in \mathbb{R}_+ \cup \{\infty\}$, and function $f : \mathcal{T} \to \mathbb{R}$, we call* $\text{roll}(X, f, \tau)$ *with*

(i) $f(Y) = |V(Y)| = N(Y)$ *the rolling number of observations,*

(ii) $f(Y) = \sum_{i=1}^{N(Y)} V(Y)_i$ *the rolling sum,*

(iii) $f(Y) = \max V(Y)$ *the rolling maximum (also denoted by* $\text{rollmax}(X, \tau)$*),*

(iv) $f(Y) = \min V(Y)$ *the rolling minimum (also denoted by* $\text{rollmin}(X, \tau)$*),*

(v) $f(Y) = \max V(Y) - \min V(Y)$ *the rolling range (also denoted by* $\text{range}(X, \tau)$*), and*

(vi) $f(Y) = \frac{1}{N(Y)} \left| \{i : 1 \le i \le N(Y), V(Y)_i \le V(Y)_{N(Y)}\} \right|$ *the rolling quantile,*

of X over horizon τ.

6 Linear Time Series Operators

Linear operators (also called "linear systems" in the signal processing literature) are pervasive when working with vector spaces, and the space of unevenly spaced time series is no exception. To fully appreciate the structure of such operators, we need to take a closer look at the properties of time series sample paths.

6.1 Sample Paths

Definition 6.1 (Sample Paths) *For a time series $X \in \mathcal{T}$ and sampling scheme $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$, the function $\mathrm{SP}_\sigma(X) : \mathbb{R} \to \mathbb{R}$ with $\mathrm{SP}_\sigma(X)(t) = X[t]_\sigma$ for $t \in \mathbb{R}$ is called the sample path of X (for sampling scheme σ). Furthermore,*

$$\mathrm{SP}_\sigma = \{\mathrm{SP}_\sigma(X) : X \in \mathcal{T}\}$$

is the space of sample paths, and $\mathrm{SP}_\sigma(t)$ is the space of sample paths that are constant after time $t \in \mathbb{R}$.

In particular, $\mathrm{SP}_\sigma(X) \in \mathrm{SP}_\sigma(\max T(X))$ for $X \in \mathcal{T}$, because the sampled value of a time series is constant after its last observation.

Lemma 6.2 *The mapping of a time series to its sample path is linear, that is,*

$$\mathrm{SP}_\sigma(aX +_\sigma bY) = a\mathrm{SP}_\sigma(X) + b\mathrm{SP}_\sigma(Y)$$

for all $X, Y \in \mathcal{T}$, each sampling scheme σ, and $a, b \in \mathbb{R}$.

Proof. Using the sample path definition and Proposition 5.4,

$$\begin{aligned}\mathrm{SP}_\sigma(aX +_\sigma bY)(t) &= (aX +_\sigma bY)[t]_\sigma \\ &= aX[t]_\sigma + bY[t]_\sigma \\ &= a\mathrm{SP}_\sigma(X)(t) + b\mathrm{SP}_\sigma(Y)(t)\end{aligned}$$

for all $t \in \mathbb{R}$. ■

Lemma 6.3 *Fix a sampling scheme σ. Two time series $X, Y \in \mathcal{T}$ have the same sample path if and only if the observation values of $X -_\sigma Y$ are identical to zero.*

Proof. The result follows from Lemma 6.2 with $a = 1$ and $b = -1$. ■

The space of sample paths SP_σ (or $\mathrm{SP}_\sigma(t)$ for fixed $t \in \mathbb{R}$) can be turned into a normed vector space, see Kreyszig (1989) or Kolmogorov and Fomin (1999). Specifically, given two elements $x, y \in \mathrm{SP}_\sigma$ and number $a \in \mathbb{R}$, define $x + y$ and ax as the real functions with $(x + y)(t) = x(t) + y(t)$ and $(ax)(t) = ax(t)$, respectively, for $t \in \mathbb{R}$. It is straightforward to verify that

$$\|x\|_{\mathrm{SP}} = \max_{t \in \mathbb{R}} |x(t)|$$

for $x \in \mathrm{SP}_\sigma$ defines a norm on SP_σ and $(\mathrm{SP}_\sigma, \| \, \|_{\mathrm{SP}})$ is therefore a normed vector space.

It is easy to see that

$$\max_{t \in T(X)} |X_t| = \|\mathrm{SP}_\sigma(X)\|_{\mathrm{SP}}$$

for all $X \in \mathcal{T}$ and each sampling scheme σ. Hence,

$$\|X\|_{\mathcal{T}} = \max_{t \in T(X)} |X_t|$$

defines a norm on $\mathcal{T}$, making $(\mathcal{T}, \| \, \|_{\mathcal{T}})$ a normed vector space also.[13]

[13] Strictly speaking, $\mathcal{T}$ is not a vector space because it has no unique zero element. If we consider two time series X and Y to be equivalent if their sample paths are identical, however, then the space of equivalence classes in $\mathcal{T}$ is a well-defined vector space. For our discussion, this distinction is not important, see Lemma 6.11, and we therefore do not distinguish between $\mathcal{T}$ and the space of equivalence classes in $\mathcal{T}$.

Corollary 6.4 *For each sampling scheme σ, the mapping $X \to \mathrm{SP}_\sigma(X)$ is an isometry between $(\mathcal{T}, \| \|_{\mathcal{T}})$ and $(SP_\sigma, \| \|_{SP})$. In other words,*

$$\|X\|_{\mathcal{T}} = \|\mathrm{SP}_\sigma(X)\|_{\mathrm{SP}}$$

for all $X \in \mathcal{T}$.

Lemma 6.5 *The order of the lag operator L and the mapping of a time series to its sample path is interchangeable in the sense that*

$$\mathrm{SP}_\sigma(X)(t-\tau) = \mathrm{SP}_\sigma(L(X,\tau))(t)$$

for all $X \in \mathcal{T}$ and $t, \tau \in \mathbb{R}$, and each sampling scheme σ.
Proof. *The result follows from Lemma 2.9 (iv).* ■

6.2 Bounded Linear Operators

Definition 6.6 *A time series operator O is said to be linear for a sampling scheme σ if $O(aX +_\sigma bY) = aO(X) +_\sigma bO(Y)$ for all $X, Y \in \mathcal{T}$ and $a, b \in \mathbb{R}$.*

In particular, linear time series operators are homogeneous of degree one, although the reverse is generally not true. For example, the operator that calculates the smallest observation value in a rolling time window (see Example 5.11) is homogeneous of degree one but not linear.

Definition 6.7 *A time series operator O is said to be*

(i) bounded if there exists a constant $M < \infty$ such that

$$\|O(X)\|_{\mathcal{T}} \leq M \|X\|_{\mathcal{T}}$$

for all $X \in \mathcal{T}$, and

(ii) continuous (for sampling scheme σ) if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\|X -_\sigma Y\|_{\mathcal{T}} < \delta$$

for $X, Y \in \mathcal{T}$ implies

$$\|O(X) -_\sigma O(Y)\|_{\mathcal{T}} < \varepsilon.$$

Proposition 6.8 *A linear time series operator O is bounded if and only if it is continuous.*

Proof. The equivalence of boundedness and continuity holds for any linear operator between two normed vector spaces; see Kreyszig (1989), Chapter 2.7 or Kolmogorov and Fomin (1999), §29. ■

For the remainder of this paper, we shall exclusively focus on bounded, and therefore continuous, linear operators.

Theorem 6.9 *If a time series kernel μ is of the form*

$$\mu(x, s) = x\mu_T(s), \tag{6.13}$$

where μ_T is a finite signed measure on $(\mathbb{R}_+, \mathcal{B}_+)$, then for each sampling scheme σ the associated convolution operator $*_\sigma^\mu$ is a bounded linear operator.

Proof. First note

$$\begin{aligned}
T\left((aX +_\sigma bY) *_\sigma \mu\right) &= T\left(aX +_\sigma bY\right) \\
&= T(aX) \cup T\left(bY\right) \\
&= T(X) \cup T\left(Y\right) \\
&= T(X *_\sigma \mu) \cup T\left(Y *_\sigma \mu\right) \\
&= T\left(a\left(X *_\sigma \mu\right)\right) \cup T\left(b\left(Y *_\sigma \mu\right)\right) \\
&= T\left(a\left(X *_\sigma \mu\right) +_\sigma b\left(Y *_\sigma \mu\right)\right)
\end{aligned}$$

because convolution and arithmetic operators are tick invariant. For $t \in T\left((aX +_\sigma bY) *_\sigma \mu\right)$,

$$\begin{aligned}
\left((aX +_\sigma bY) *_\sigma \mu\right)_t &= \int_0^\infty d\mu\left((aX +_\sigma bY)\left[t-s\right]_\sigma, s\right) \\
&= \int_0^\infty d\mu\left(aX[t-s]_\sigma + bY[t-s]_\sigma, s\right) \\
&= \int_0^\infty \left(aX[t-s]_\sigma + bY[t-s]_\sigma\right) d\mu_T\left(s\right) \\
&= a\int_0^\infty X[t-s]_\sigma d\mu_T\left(s\right) + b\int_0^\infty Y[t-s]_\sigma d\mu_T\left(s\right) \\
&= a\left(X *_\sigma \mu\right)_t + b\left(Y *_\sigma \mu\right)_t,
\end{aligned}$$

showing that $*_\sigma^\mu$ is indeed linear. Furthermore,

$$\begin{aligned}
\left|(X *_\sigma \mu)_t\right| &= \left|\int_0^\infty X[t-s]_\sigma d\mu_T\left(s\right)\right| \qquad (6.14)\\
&\leq \int_0^\infty \left|X[t-s]_\sigma\right| \left|d\mu_T\left(s\right)\right| \\
&\leq \|X\|_{\mathcal{T}} \int_0^\infty \left|d\mu_T\left(s\right)\right| \\
&= \|X\|_{\mathcal{T}} \|\mu_T\|_{\mathrm{TV}},
\end{aligned}$$

for all $t \in T\left(X\right)$, where $\|\mu_T\|_{\mathrm{TV}}$ is the total variation of the signed measure μ_T, which is finite by assumption. Taking the maximum over all $t \in T\left(X *_\sigma \mu\right)$ on the left-hand side of (6.14) gives

$$\left\|*_\sigma^\mu\left(X\right)\right\|_{\mathcal{T}} \leq \|X\|_{\mathcal{T}} \|\mu_T\|_{TV},$$

which shows the boundedness of $*_\sigma^\mu$. ■

The next subsection shows that, under reasonable conditions, the reverse is also true. In other words, convolution operators with linear kernel of the form (6.13) are the only interesting bounded linear operators. To show this result, we need to take a closer look at how individual time series observations are used by a linear convolution operator of the form (6.13).

Remark 6.10 *Assume we are given a time series $X \in \mathcal{T}$ and a linear convolution operator of form (6.13). Define $t_0 = -\infty$ and $X_{t_0} = X_{t_1}$. For each observation time $t = t_n \in T\left(X\right)$,*

$$(X * \mu)_{t_n} = \mu_T\left(\{0\}\right) X_{t_n} + \sum_{k=1}^{n} \mu_T\left(\left(t_n - t_k, t_n - t_{k-1}\right]\right) X_{t_{k-1}}$$

for last-point sampling,

$$(X *_{\text{next}} \mu)_{t_n} = \sum_{k=1}^{n} \mu_T\left([t_n - t_k, t_n - t_{k-1})\right) X_{t_k}$$

for next-point sampling, and

$$(X *_{\text{lin}} \mu)_{t_n} = \sum_{k=1}^{n} a_{k,n} X_{t_k}$$

for sampling with linear interpolation, where the coefficients $a_{k,n}$ for $1 \leq k \leq n$ depend only on μ_T and the observation time spacings.

6.3 Linear Operators as Convolution Operators

Clearly, a time series contains all of the information about its sample path. On the other hand, the sample path of a time series contains only a subset of the time series information, because the observation times are not uniquely determined by the sample path alone. The following result shows that linear time series operators use only this reduced amount of information about a time series.

Lemma 6.11 *Let O be a linear time series operator for sampling scheme σ. There exists a function $g_\sigma : \mathrm{SP}_\sigma \to \mathrm{SP}_\sigma$ such that $\mathrm{SP}_\sigma(O(X)) = g_\sigma(\mathrm{SP}_\sigma(X))$ for all $X \in \mathcal{T}$. In other words, for linear operators the sample path of the output time series depends only on the sample path of the input time series.*

Proof. For each sample path $x \in \mathrm{SP}_\sigma$, we choose[14] one (of the infinitely many) time series $X \in \mathcal{T}$ with $\mathrm{SP}_\sigma(X) = x$ and define

$$g_\sigma(x) = \mathrm{SP}_\sigma(O(X)).$$

We need to show that g_σ is uniquely defined for each $x \in \mathrm{SP}_\sigma$. Assume there exist two time series $X, Y \in \mathcal{T}$ with $\mathrm{SP}_\sigma(X) = \mathrm{SP}_\sigma(Y)$, but $\mathrm{SP}_\sigma(O(X)) \neq \mathrm{SP}_\sigma(O(Y))$. Because O is linear,

$$aO(X -_\sigma Y) = O(a(X -_\sigma Y)) \tag{6.15}$$

for all $a \in \mathbb{R}$. Lemma 6.3 implies that the observation values of $X -_\sigma Y$ (and therefore also $a(X -_\sigma Y)$) are identical to zero. Hence, (6.15) can be satisfied for all $a \in \mathbb{R}$ only if the observation values of $O(X -_\sigma Y)$ (and therefore also $O(X) -_\sigma O(Y)$) are identical to zero. Applying Lemma 6.3 once more shows that $O(X)$ and $O(Y)$ have the same sample path, in contradiction to our assumption. ∎

A lot more can be said about linear operators that satisfy additional, quite general, properties.

Theorem 6.12 *Let O be a bounded, causal, shift- and tick-invariant time series operator that is linear for sampling scheme σ. Then O is a convolution operator with kernel of the form $\mu(x, s) = x\mu_T(s)$, where μ_T is a finite signed measure on $(\mathbb{R}_+, \mathcal{B}_+)$.*

[14] Given a sample path, a matching time series can be constructed by looking at the kinks (for linear interpolation) or jumps (for the other sampling schemes) of the sample path.

Proof. See Appendix A. ■

Combining Proposition 4.5, and Theorems 6.9 and 6.12 yields the main result of this section:

Theorem 6.13 *The class of bounded, linear, causal, shift- and tick-invariant time series operators coincides with the class of convolution operators with kernels of the form* $\mu(x,s) = x\mu_T(s)$ *where* μ_T *is a finite signed measure on* $(\mathbb{R}_+, \mathcal{B}_+)$.

Remark 6.14 *The class of bounded, linear, shift- and tick-invariant (but not necessarily causal) time series operators coincides with the class of convolution operators with kernel of the form* $\mu(x,s) = x\mu_T(s)$ *where* μ_T *is a finite signed measure on* $(\mathbb{R}, \mathcal{B})$.

Using these results, the discrete-time analogs of convolution operators derived in Remark 4.4 can be further simplified for linear operators.

Remark 6.15 (Discrete Time Analog - continued) *The discrete-time analog of linear convolution operators are causal, linear, shift-invariant filters of the form*

$$Y_n = \sum_{k=0}^{\infty} a_k X_{n-k},$$

with $\sum_{k=0}^{\infty} |a_k| < \infty$, *where* $(X_n : n \in \mathbb{Z})$ *is a stationary time series. The condition that the filter coefficients are absolutely convergent implies that the transformed time series is stationary if the input time series is stationary; see Brockwell and Davis (1991), Proposition 3.2.1. The coefficients of the discrete-time filter are related to the time series kernel (6.13) via the equations given in Remark 6.10.*

6.4 An Application

Definition 6.16 *We say that a time series operator* O *has finite memory if there exists a constant* $C < \infty$ *such that*

$$O(X)_t = O\left(X\left\{\mathrm{Prev}^X(t-C), +\infty\right\}\right)_t$$

for all $X \in \mathcal{T}$ *and* $t \in T(O(X))$.

Corollary 6.17 *A rolling time series function over horizon* $\tau < \infty$ *is a time series operator with finite memory.*

Proof. The result follows immediately from Definition 5.9. ■

The following theorem shows that for linear convolution operators and a quite general class of data-generating processes, the operator applied to the corresponding *observation* time series provides the "best guess" of the operator applied to the unobserved, data-generating process.

Theorem 6.18 *Let* $(X_t^c : t \geq 0)$ *be a Lévy process,* $T_X \in \mathbb{T}$ *with* $\min T_X \geq 0$ *be a fixed observation time sequence, and* $X = X^c[T_X]$ *be the corresponding observation time series. Let* O *be a linear convolution operator for sampling scheme* $\sigma =$ "lin". *If* O *has finite memory, then*

(i) the associated kernel $\mu(x,s) = x\mu_T(s)$ *satisfies* $\mu_T \equiv 0$ *on* $(C, +\infty)$ *for some constant* $C < \infty$,

(ii) and

$$E\left(\left(X^c * \mu\right)_t \mid X\right) = \left(\mathrm{SP}_{\mathrm{lin}}\left(X\right) * \mu_T\right)\left(t\right)$$

for $C + \min T\left(X\right) \leq t \leq \max T\left(X\right)$*. In particular*

$$E\left(\left(X^c * \mu\right)_t \mid X\right) = O\left(X\right)_t \tag{6.16}$$

for all $t \in T\left(X\right)$ *with* $t \geq C + \min T\left(X\right)$*.*

Proof.

(i) By the definition of a linear convolution operator,

$$\begin{aligned}\int_0^\infty X[t-s]_{\mathrm{lin}} d\mu_T\left(s\right) &= O\left(X\right)_t \\ &= O\left(X\left\{\mathrm{Prev}^X\left(t-C\right), +\infty\right\}\right)_t \\ &= \int_0^{t-\mathrm{Prev}^X\left(t-C\right)} X[t-s]_{\mathrm{lin}} d\mu_T\left(s\right) \\ &= \int_{[0,C]} X[t-s]_{\mathrm{lin}} d\mu_T\left(s\right) + \int_{(C,t-\mathrm{Prev}^X(t-C)]} X[t-s]_{\mathrm{lin}} d\mu_T\left(s\right)\end{aligned}$$

for all $X \in \mathcal{T}$ and $t \in T\left(X\right)$, where we used the finite memory property for the second equation. Equivalently,

$$\int_{(C,\infty)} X[t-s]_{\mathrm{lin}} d\mu_T\left(s\right) = \int_{(C,t-\mathrm{Prev}^X(t-C)]} X[t-s]_{\mathrm{lin}} d\mu_T\left(s\right) \tag{6.17}$$

for all $X \in \mathcal{T}$ and $t \in T\left(X\right)$. For each time series $X \in \mathcal{T}$ and time point $t \in T\left(X\right)$, there exists a time series Y with identical sample path and $t - C \in T\left(Y\right)$, which implies that $t-\mathrm{Prev}^Y\left(t-C\right) = C$. Hence, the right-hand side of (6.17) is always equal to zero, which can only be the case if μ_T is identical to zero on $\left(C, +\infty\right)$.

(ii) For a Lévy process Z, the conditional expectation $E\left(Z_t \middle| Z_s, Z_r\right)$ for $s \leq t \leq r$ is the linear interpolation of $\left(s, Z_s\right)$ and $\left(r, Z_r\right)$ evaluated at time t. Hence, for $C + \min T\left(X\right) \leq t \leq \max T\left(X\right)$,

$$\begin{aligned}E\left(\left(X^c * \mu\right)_t \mid X\right) &= E\left(\int_0^\infty d\mu\left(X^c_{t-s}, s\right) \mid X\right) \\ &= E\left(\int_0^C X^c_{t-s} d\mu_T\left(s\right) ds \mid X\right) \\ &= \int_0^C E\left(X^c_{t-s} \mid X\right) d\mu_T\left(s\right) \\ &= \int_0^C X\left[t-s\right]_{\mathrm{lin}} d\mu_T\left(s\right) \\ &= \left(\mathrm{SP}_{\mathrm{lin}}\left(X\right) * \mu_T\right)\left(t\right)\end{aligned} \tag{6.18}$$

where Fubini's theorem was used to change the order of integration and the conditional expectation. Additionally, if $t \in T(X)$, then (6.18) simplifies further to

$$\begin{aligned} E\left((X^c * \mu)_t \mid X\right) &= \int_0^C d\mu\left(X\left[t-s\right]_{\text{lin}}, s\right) \\ &= O\left(X\right)_t . \end{aligned}$$

■

As already indicated in Example 1.1, (6.16) generally does not hold for nonlinear time series operators such as volatility or correlation estimators. In these cases, the right-hand side of (6.16) requires correction terms that depend on the exact dynamics of the data-generating process.

7 Multivariate Time Series

Multivariate data sets often consist of time series with different frequencies, and are thus naturally treated as multivariate unevenly spaced time series, even if the observations of the individual time series are reported at regular intervals. For example, macroeconomic data; such as the gross domestic product (GDP), the rate of unemployment, and foreign exchange rates; is released in a nonsynchronous manner and at vastly different frequencies (quarterly, monthly, and essentially continuously, in case of the US). Moreover, the frequency of reporting may change over time.

Many multivariate time series operators are natural extensions of their univariate counterparts, so the concepts of the prior sections require only a few modifications. If one is primarily interested in the analysis of univariate time series, then this section may be skipped on a first reading without detriment to the understanding of the rest of the paper.

Definition 7.1 *For $K \geq 1$, a $K-$dimensional unevenly spaced time series X^K is a $K-$tuple $\left(X_k^K : 1 \leq k \leq K\right)$ of univariate unevenly spaced time series $X_k^K \in \mathcal{T}$ for $1 \leq k \leq K$. $\mathcal{T}^K$ is the space of (real-valued) $K-$dimensional time series.*

Definition 7.2 *For $K, M \geq 1$, a $(K, M)-$dimensional time series operator is a mapping $O : \mathcal{T}^K \to \mathcal{T}^M$, or equivalently, an $M-$tuple of mappings $O^m : \mathcal{T}^K \to \mathcal{T}$ for $1 \leq m \leq M$.*

The following two operators are helpful for extracting basic information from such objects.

Definition 7.3 (Subset Selection) *For a multivariate time series $X^K \in \mathcal{T}^K$ and indices $(i_1, \ldots, i_M)$ with $1 \leq i_j \leq K$ for $j = 1, \ldots, M$, we call*

$$X^K\left(i_1, \ldots, i_M\right) = \begin{cases} \left(X_{i_1}^K, \ldots, X_{i_M}^K\right), & \text{if } M \geq 2, \\ X_{i_1}^K, & \text{if } M = 1, \end{cases}$$

the subset (vector) time series of X^K (for the indices $(i_1, \ldots, i_M)$).

Definition 7.4 (Multivariate Sampling) *For a multivariate time series $X^K \in \mathcal{T}^K$, time vector $t^K \in \mathbb{R}^K$ and sampling scheme $\sigma \in \{, \text{lin}, \text{next}\}$, we call*

$$X^K[t^K]_\sigma = \left(X_k^K[t_k^K]_\sigma : 1 \leq k \leq K\right) \tag{7.19}$$

the sampled value (vector) of X^K at time (vector) t^K.

Unless stated otherwise, we apply univariate time series operators element-wise to a multivariate time series X^K. In other words, we assume that a univariate time series operator O, when applied to a multivariate time series, is replaced by its natural multivariate extension. For example, we interpret $X^K[t]$ for $t \in \mathbb{R}$ as $\left(X_k^K[t] : 1 \leq k \leq K\right)$ and call it the sampled value (vector) of X^K at time t. Similarly, $L(X^K, \tau)$ for $\tau \in \mathbb{R}$ equals $\left(L\left(X_k^K, \tau\right) : 1 \leq k \leq K\right)$, and so on. Of course, whenever there is a risk of confusion, we must explicitly define the extension of a univariate time series operator to the multivariate case.

7.1 Structure

In Section 3 we already took a detailed look at the common structural features among univariate time series operators. With very minor modifications (not shown here), that analysis is still relevant to the multivariate case; however, additional properties are worth considering now.

7.1.1 Dimension Invariance

For many multivariate operators, the dimension of the output time series is either equal to one (typically for data aggregations) or equal to the dimension of the input time series (typically for data transformations).

Definition 7.5 *A $(K, M)-$dimensional time series operator O^K is dimension invariant if $K = M$.*

In particular, if a vector time series operator O^K is the natural multivariate extension of a univariate operator O, then the former is, by construction, dimension invariant.

7.1.2 Permutation Invariance

Many multivariate time series operators have a certain symmetry in the sense that they assume no natural ordering among the input time series.

Definition 7.6 *A $(K, K)-$dimensional time series operator O is permutation invariant if*

$$O\left(p\left(X^K\right)\right) = p\left(O\left(X^K\right)\right)$$

for all permutations $p : \mathcal{T}^K \to \mathcal{T}^K$ and time series $X^K \in \mathcal{T}^K$.

In particular, if a vector time series operator O^K is the natural multivariate extension of a univariate operator O, then the former is, by construction, permutation invariant.

7.1.3 Example

We will now end the discussion of structural features with a brief example.

Definition 7.7 (Merging) *For $X, Y \in \mathcal{T}$, $X \cup Y$ denotes the merged time series of X and Y, where $T(X \cup Y) = T(X) \cup T(Y)$ and*

$$(X \cup Y)_t = \begin{cases} X_t, & \textit{if } t \in T(X) \\ Y_t, & \textit{if } t \notin T(X) \end{cases}$$

for $t \in T(X \cup Y)$.

In particular, if both time series have an observation at the same time point, then the observation value of the first time series takes precedence. Therefore, $X \cup Y$ and $Y \cup X$ are generally not equal unless the observation times of X and Y are disjoint.

The operator that merges two time series is causal, timescale invariant, tick invariant (in the sense that $T(X \cup Y) = T(X) \cup T(Y)$), shift invariant (in the sense that $L(X \cup Y, \tau) = L(X, \tau) \cup L(Y, \tau)$), homogeneous of degree $d = 1$ (in the sense that $(aX) \cup (aY) = a(X \cup Y)$), and neither dimension nor permutation invariant.

7.2 Multivariate Convolution Operators

Definition 7.8 *A $(K, 1)$ −dimensional (or K−dimensional) time series kernel μ is a signed measure on $(\mathbb{R}^K \times \mathbb{R}_+, \mathcal{B}^K \otimes \mathcal{B}_+)$ with*

$$\int_0^\infty |d\mu(f(s), s)| < \infty$$

for all bounded piecewise linear functions $f : \mathbb{R}_+^K \to \mathbb{R}$.[15]

As in the univariate case, the integrability condition ensures that the value of a convolution is well-defined and finite. In particular, the condition is satisfied if $d\mu(x, s) = g(x)\, d\mu_T(s)$ for some real function $g : \mathbb{R}^K \to \mathbb{R}$ and finite signed measure μ_T on $\mathbb{R}_+$.

Definition 7.9 (Multivariate Convolution Operator) *For a multivariate time series $X^K \in \mathcal{T}^K$, K−dimensional kernel μ and sampling scheme σ, the convolution $*_\sigma^\mu(X^K) = X^K *_\sigma \mu$ is a univariate time series with*

$$T\left(X^K *_\sigma \mu\right) = \bigcup_{k=1}^{K} T\left(X_k^K\right), \tag{7.20}$$

$$\left(X^K *_\sigma \mu\right)_t = \int_0^\infty d\mu\left(X^K[t-s]_\sigma, s\right), \qquad t \in T\left(X^K *_\sigma \mu\right). \tag{7.21}$$

If μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^K \times \mathbb{R}_+$, then (7.21) can be written as

$$\left(X^K *_\sigma f\right)_t = \int_0^\infty f\left(X^K[t-s]_\sigma, s\right) ds, \qquad t \in T\left(X^K *_\sigma \mu\right),$$

where f is the density function of μ.

A K−dimensional convolution operator is a mapping of $\mathcal{T}^K \to \mathcal{T}$. More generally, a (K, M)-dimensional convolution operator is an M−tuple of K−dimensional convolution operators $(*_\sigma^{\mu_1}, \ldots, *_\sigma^{\mu_M})$.

Note that a (K, K) −dimensional convolution operator is generally not equivalent to K one-dimensional convolution operators. In the former case, each output time series depends on *all* input time series; whereas in the latter case, each output time series depends only on the corresponding input time series. In particular, the observation times of the output time series of a (K, K) −dimensional convolution operator are the union of observation times of all input time series, see (7.20).

[15] A more general definition is possible, where μ is a signed measure on $(\mathbb{R}^K \times (\mathbb{R}_+)^K, \mathcal{B}^K \otimes (\mathcal{B}_+)^K)$. For our purposes, however, the simpler definition is sufficient.

7.3 Examples

This section gives examples of multivariate time series operators that can be expressed as multivariate convolution operators.

Proposition 7.10 *The arithmetic operators $X+Y$ and XY in Definition 5.2 are multivariate (specifically $(2,1)-$dimensional) convolution operators with kernel $\mu(x,y,s) = (x+y)\delta_0(s)$ and $\mu(x,y,s) = xy\delta_0(s)$, respectively, for the vector time series $(X,Y) \in \mathcal{T}^2$.*

Proof. For $X, Y \in \mathcal{T}$ and $\mu(x,y,s) = (x+y)\delta_0(s)$, we have $T((X,Y) * \mu) = T(X) \cup T(Y)$ by the definition of a multivariate convolution. For $t \in T((X,Y) * \mu)$,

$$\begin{aligned} ((X,Y) * \mu)_t &= \int_0^{\infty} (X[t-s] + Y[t-s])\delta_0(s)\, ds \\ &= X[t] + Y[t], \end{aligned}$$

and also, therefore, $V((X,Y) * \mu) = V(X+Y)$. The reasoning for the multiplication of two time series is similar. ∎

Definition 7.11 (Cross-sectional Operators) *For a function $f : \mathbb{R}^K \to \mathbb{R}$ and sampling scheme σ, the cross-sectional or (contemporaneous) time series operator $\mathrm{C}_\sigma(.,f) : \mathcal{T}^K \to \mathcal{T}$ is given by*

$$\begin{aligned} T\left(\mathrm{C}_\sigma\left(X^K, f\right)\right) &= \bigcup_{k=1}^{K} T\left(X_k^K\right), \\ \left(\mathrm{C}_\sigma\left(X^K, f\right)\right)_t &= f\left(X^K[t]_\sigma\right), \qquad t \in T\left(\mathrm{C}_\sigma\left(X^K, f\right)\right) \end{aligned}$$

for $X^K \in \mathcal{T}^K$.

It is easy to see that a cross-sectional time series operator $C_\sigma(.,f)$ is a $K-$dimensional convolution operator with kernel $\mu^f\left(x^K, s\right) = f\left(x^K\right)\delta_0(s)$.

Example 7.12 *For a multivariate time series $X^K \in \mathcal{T}^K$, we call $\mathrm{C}_\sigma(X^K, f)$ with*

(i) $f(x^K) = \mathrm{sum}(x^K)$ *the cross-sectional sum of X^K (also written $\mathrm{sum}_{\mathrm{C},\sigma}(X^K)$),*

(ii) $f(x^K) = \mathrm{avg}(x^K)$ *the cross-sectional average of X^K (also written $\mathrm{avg}_{\mathrm{C},\sigma}(X^K)$),*

(iii) $f(x^K) = \min(x^K)$ *the cross-sectional minimum of X^K (also written $\min_{\mathrm{C},\sigma}(X^K)$),*

(iv) $f(x^K) = \max(x^K)$ *the cross-sectional maximum of X^K (also written $\max_{\mathrm{C},\sigma}(X^K)$),*

(v) $f(x^K) = \max(x^K) - \min(x^K)$ *the cross-sectional range of X^K (also written $\mathrm{range}_{\mathrm{C},\sigma}(X^K)$), and*

(vi) $f(x^K) = \mathrm{quant}\big(x^K, q\big)$ *the cross-sectional $q-$quantile of X^K (also written $\mathrm{quant}_{\mathrm{C},\sigma}(X^K, q)$).*

It is easy to see that arithmetic and cross-sectional time series operators are consistent with each other. For example, $\mathrm{sum}_{C,\sigma}(X^K)$ equals $X_1^K + \ldots + X_K^K$ for all $K \geq 1$ and $X^K \in \mathcal{T}^K$.

Contemporaneous time series operators, such as the ones in Example 7.12, are useful for summarizing the behavior of high-dimensional time series. For example, a common question among economists is how the *distribution* of household income within a certain country changes over time. If X^K denotes the time series of individual incomes from a panel data set, $\mathrm{quant}_C(X^K, 0.8)$ / $\mathrm{quant}_C(X^K, 0.2)$ is the time series of the inter-quintile income ratio. Similarly, in a medical study, the dispersion of a certain measurement across subjects as a function of time might be of interest.

8 Moving Averages

Moving averages - with exponentially declining weights or otherwise - are used to extract the average value of a time series over a certain time horizon; for example, to smooth out noisy observation values. Moving averages are very similar to kernel smoothing methods, see Wand and Jones (1994) and Hastie et al. (2009), except that they use only past observations and are therefore causal time series operators.

For equally spaced time series data there is only one way of calculating simple moving averages (SMAs) and exponential moving averages (EMAs), and the properties of such linear filters are well understood. For unevenly spaced data, however, there are multiple alternatives (for example, due to the choice of the sampling scheme), all of which may be sensible depending on the underlying data-generating process of a given time series and the desired application.

Definition 8.1 *A convolution operator* $*_\sigma^\mu$*, associated with a kernel* μ *and sampling scheme* $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$*, is said to be a moving average operator if*

$$\mu(x, s) = x\mu_T(s) = x dF(s), \tag{8.22}$$

where μ_T *is a probability measure on* $(\mathbb{R}_+, \mathcal{B}_+)$ *and* F *its distribution function. We call* μ_T *(and sometimes* μ *itself) a moving average kernel.*

Based on the remarks following Definition 2.4, we know that the first observation value of a moving average is equal to the first observation value of the input time series. Moreover, the moving average of a constant time series is identical to the input time series.

Theorem 8.2 *Fix a sampling scheme* σ *and restrict attention to the set of causal, shift- and tick-invariant time series operators. The class of moving average operators coincides with the class of linear time series operators in the aforementioned set with (i)* $O(X) \geq 0$ *for all* $X \in \mathcal{T}$ *with* $X \geq 0$*, and (ii)* $O(X) = X$ *for all constant time series* $X \in \mathcal{T}$*.*

Proof.

- $\Longrightarrow$ It immediately follows from Definition 8.1 and Proposition 4.5 that a moving average operator satisfies the specified conditions.
- $\Longleftarrow$ By Theorem 6.13, the kernel associated with such a time series operator O is of the form $\mu(x, s) = x\mu_T(s)$ for some finite signed measure μ_T on $(\mathbb{R}_+, \mathcal{B}_+)$. Because $O(X) \geq 0$ for all $X \in \mathcal{T}$ with $X \geq 0$, it follows from a simple measure-theoretic argument that μ_T is a

positive measure. Furthermore, because $O(X) = X$ for all constant time series $X \in \mathcal{T}$, it follows that

$$\int_0^\infty d\mu_T(s) = 1,$$

showing that μ_T is a probability measure on $(\mathbb{R}_+, \mathcal{B}_+)$ and, therefore, that O is a moving average operator.

■

For a moving average kernel μ_T, associated cumulative distribution function F, and $X \in \mathcal{T}$,

$$\begin{aligned} \mathrm{MA}(X, \mu_T) &= \mathrm{MA}(X, F) &&= X * \mu, \\ \mathrm{MA}_{\mathrm{lin}}(X, \mu_T) &= \mathrm{MA}_{\mathrm{lin}}(X, F) &&= X *_{\mathrm{lin}} \mu, \\ \mathrm{MA}_{\mathrm{next}}(X, \mu_T) &= \mathrm{MA}_{\mathrm{next}}(X, F) &&= X *_{\mathrm{next}} \mu \end{aligned} \tag{8.23}$$

with $\mu(x, s) = x\mu_T(s) = xdF(s)$ are three different moving averages of X. If X is non-decreasing, then

$$\mathrm{MA}_{\mathrm{next}}(X, F)_t \geq \mathrm{MA}_{\mathrm{lin}}(X, F)_t \geq \mathrm{MA}(X, F)_t \tag{8.24}$$

for all $t \in T(X)$, because $X[t]_{\mathrm{next}} \geq X[t]_{\mathrm{lin}} \geq X[t]$ for all $t \in \mathbb{R}$ for a non-decreasing time series.

8.1 Simple Moving Averages

Definition 8.3 *For time series $X \in \mathcal{T}$, we define four versions of the simple moving average (SMA) of length $\tau > 0$. For $t \in T(X)$,*

(i) $\mathrm{SMA}(X, \tau)_t = \frac{1}{\tau} \int_0^\tau X[t - s]\, ds$,

(ii) $\mathrm{SMA}_{\mathrm{lin}}(X, \tau)_t = \frac{1}{\tau} \int_0^\tau X[t - s]_{\mathrm{lin}}\, ds$,

(iii) $\mathrm{SMA}_{\mathrm{next}}(X, \tau)_t = \frac{1}{\tau} \int_0^\tau X[t - s]_{\mathrm{next}}\, ds$, *and*

(iv) $\mathrm{SMA}_{\mathrm{eq}}(X, \tau)_t = \mathrm{avg}\{X_s : s \in T(X) \cap (t - \tau, t]\}$

where in all cases the observation times of the input and output time series are identical.

The simple moving averages SMA, $\mathrm{SMA}_{\mathrm{lin}}$ and $\mathrm{SMA}_{\mathrm{next}}$ are moving average operators in the sense of Definition 8.1. Specifically,

$$\begin{aligned} \mathrm{SMA}(X, \tau) &= \mathrm{MA}(X, \mu_T) \\ \mathrm{SMA}_{\mathrm{lin}}(X, \tau) &= \mathrm{MA}_{\mathrm{lin}}(X, \mu_T) \\ \mathrm{SMA}_{\mathrm{next}}(X, \tau) &= \mathrm{MA}_{\mathrm{next}}(X, \mu_T) \end{aligned}$$

with $\mu_T(t) = \frac{1}{\tau}\mathbf{1}_{\{0 \leq t \leq \tau\}}$. However, the simple moving average $\mathrm{SMA}_{\mathrm{eq}}$, where `eq` stands for equally-weighted, cannot be expressed as a convolution operator and is therefore not a moving average operator in the sense of Definition 8.1. Nevertheless, it is useful for demonstrating the difference between the SMAs of equally- and unevenly spaced time series.

The first SMA can be used to analyze discrete observation *values*; for example, to calculate the average FED funds target rate[16] over the past three years. In such a case, it is desirable to weigh each observation value by the amount of time it remained unchanged. The $\mathrm{SMA_{eq}}$ is ideal for analyzing discrete *events*; for example, calculating the average number of casualties per deadly car accident over the past twelve months, or determining the average number of IBM common shares traded on the NYSE per executed order during the past 30 minutes. The $\mathrm{SMA_{lin}}$ can be used to estimate the rolling average value of a discretely-observed continuous-time stochastic processes, with observation times that are independent of the observation values, see Theorem 6.18. Finally, the $\mathrm{SMA_{next}}$ is useful for certain trend and return calculations, see Section 8.3.

That said, the values of all SMAs will generally be quite similar as long as the moving average time horizon τ is considerably larger than the spacing of observation times. The type of moving average used (for example, SMA vs. EMA) and the moving average time horizon will usually have a much greater influence on the outcome of a time series analysis.

Proposition 8.4 *For all $X \in \mathcal{T}$,*

$$\lim_{\tau \searrow 0} \mathrm{SMA_{lin}}(X,\tau)_t = \lim_{\tau \searrow 0} \mathrm{SMA_{next}}(X,\tau)_t = \lim_{\tau \searrow 0} \mathrm{SMA_{eq}}(X,\tau)_t = X_t$$

for all $t \in T(X)$, while

$$\lim_{\tau \searrow 0} \mathrm{SP}(\mathrm{SMA}(X,\tau))(t) = \mathrm{SP}(B(X))(t)$$

for all $t \in \mathbb{R}$.

We will now end our discussion of SMAs by illustrating a connection to the corresponding operator for equally spaced data.

Proposition 8.5 (Equally spaced time series) *If $X \in \mathcal{T}$ is an equally spaced time series with observation time spacings equal to some constant $c > 0$, and if the moving average length τ is an integer multiple of c, then*

$$\mathrm{SMA_{eq}}(X,\tau)_t = \mathrm{SMA_{next}}(X,\tau)_t$$

for all $t \in T(X)$ with $t \geq \min T(X) + \tau - c$. In other words, the simple moving averages $\mathrm{SMA_{eq}}$ and $\mathrm{SMA_{next}}$ are identical after an initial ramp-up period of length $\tau - c$.

Proof. Because $\tau = K\Delta t$ for some integer K, for all $t \in T(X)$ with $t \geq \min T(X) + \tau - c$ we have

$$\begin{aligned}
\mathrm{SMA_{next}}(X,\tau)_t &= \frac{1}{\tau}\int_0^\tau X[t-s]_{\mathrm{next}}\,ds \\
&= \frac{X_t \Delta t(X)_t + X_{t-\Delta t}\Delta t(X)_{t-\Delta t} + \ldots + X_{t-\Delta t(K-1)}\Delta t(X)_{t-\Delta t(K-1)}}{\tau} \\
&= \frac{X_t + X_{t-\Delta t} + \ldots + X_{t-\Delta t(K-1)}}{K} \\
&= \mathrm{avg}\left\{X_s : s \in [t, t-\tau) \cap T(X)\right\} \\
&= \mathrm{SMA_{eq}}(X,\tau)_t\,.
\end{aligned}$$

[16]The FED funds target rate is the desired interest rate (by the Federal Reserve) at which depository institutions (such as a savings bank) lend balances held at the Federal Reserve to other depository institutions overnight. See www.federalreserve.gov/fomc/fundsrate.htm for details.

■

See Eckner (2012) for an efficient $O(N(X))$ implementation in the programming language `C` of simple moving averages and various other time series operators for unevenly spaced data.

8.2 Exponential Moving Averages

This section discusses exponential moving averages, also known as exponentially weighted moving averages. We use the former name in this paper, because the associated weights for unevenly spaced time series are defined only implicitly, via a kernel, as opposed to explicitly, as they are for equally spaced time series.

Definition 8.6 *For a time series $X \in \mathcal{T}$, we define four versions of the exponential moving average (EMA) of length $\tau > 0$. For $t \in \left\{t_1, \ldots, t_{N(X)}\right\}$,*

(i) $\mathrm{EMA}(X,\tau)_t = \frac{1}{\tau}\int_0^\infty X[t-s]\, e^{-s/\tau} ds,$

(ii) $\mathrm{EMA}_{\mathrm{lin}}(X,\tau)_t = \frac{1}{\tau}\int_0^\infty X[t-s]_{\mathrm{lin}}\, e^{-s/\tau} ds,$

(iii) $\mathrm{EMA}_{\mathrm{next}}(X,\tau)_t = \frac{1}{\tau}\int_0^\infty X[t-s]_{\mathrm{next}}\, e^{-s/\tau} ds,$ *and*

(iv) $\mathrm{EMA}_{\mathrm{eq}}(X,\tau)_t = \begin{cases} X_{t_1}, & \text{if } t = t_1 \\ \left(1-e^{-\Delta t_n/\tau}\right) X_{t_n} + e^{-\Delta t_n/\tau}\mathrm{EMA}_{\mathrm{eq}}(X,\tau)_{t_{n-1}}, & \text{if } t = t_n > t_1 \end{cases}$

where in all cases the observation times of the input and output time series are identical.

The exponential moving averages EMA, $\mathrm{EMA}_{\mathrm{lin}}$ and $\mathrm{EMA}_{\mathrm{next}}$ are moving average operators in the sense of Definition 8.1. Specifically,

$$\begin{aligned} \mathrm{EMA}(X,\tau) &= \mathrm{MA}(X,\mu_T) \\ \mathrm{EMA}_{\mathrm{lin}}(X,\tau) &= \mathrm{MA}_{\mathrm{lin}}(X,\mu_T) \\ \mathrm{EMA}_{\mathrm{next}}(X,\tau) &= \mathrm{MA}_{\mathrm{next}}(X,\mu_T) \end{aligned}$$

with $\mu_T(s) = \frac{1}{\tau}e^{-s/\tau}$.

Proposition 8.7 *For all $X \in \mathcal{T}$,*

$$\lim_{\tau\searrow 0}\mathrm{EMA}_{\mathrm{lin}}(X,\tau)_t = \lim_{\tau\searrow 0}\mathrm{EMA}_{\mathrm{next}}(X,\tau)_t = \lim_{\tau\searrow 0}\mathrm{EMA}_{\mathrm{eq}}(X,\tau)_t = X_t$$

for all $t \in T(X)$, while

$$\lim_{\tau\searrow 0}\mathrm{SP}(\mathrm{EMA}(X,\tau))(t) = \mathrm{SP}(B(X))(t)$$

for all $t \in \mathbb{R}$.

The exponential moving average $\mathrm{EMA}_{\mathrm{eq}}$ is motivated by the corresponding definition for equally spaced time series. As the following result shows, it is actually identical to the exponential moving average $\mathrm{EMA}_{\mathrm{next}}$.

Proposition 8.8 *For all $X \in \mathcal{T}$ and time horizons $\tau > 0$,*

$$\mathrm{EMA}_{\mathrm{eq}}(X,\tau) = \mathrm{EMA}_{\mathrm{next}}(X,\tau).$$

In particular, the $\mathrm{EMA}_{\mathrm{eq}}$ *is a moving average in the sense of Definition 8.1.*

Proof. For $n = 1$,

$$\mathrm{EMA}_{\mathrm{eq}}(X,\tau)_{t_1} = X_{t_1} = \mathrm{EMA}_{\mathrm{next}}(X,\tau)_{t_1}.$$

For $1 < n \le N(X)$, by induction,

$$\begin{aligned}
\mathrm{EMA}_{\mathrm{next}}(X,\tau)_{t_n} &= \frac{1}{\tau}\int_0^{\infty} X[t_n - s]_{\mathrm{next}}\, e^{-s/\tau} ds \\
&= \frac{1}{\tau}\int_0^{\Delta t_n} X[t_n - s]_{\mathrm{next}}\, e^{-s/\tau} ds + \frac{1}{\tau}\int_{\Delta t_n}^{\infty} X[t_n - s]_{\mathrm{next}}\, e^{-s/\tau} ds \\
&= X_{t_n}\left(1 - e^{-\Delta t_n/\tau}\right) \\
&+ e^{-\Delta t_n/\tau}\int_{\Delta t_n}^{\infty} X[(t_n - \Delta t_n) - (s - \Delta t_n)]_{\mathrm{next}}\, e^{-(s-\Delta t_n)/\tau} ds \\
&= X_{t_n}\left(1 - e^{-\Delta t_n/\tau}\right) + e^{-\Delta t_n/\tau}\mathrm{EMA}_{\mathrm{next}}(X,\tau)_{t_{n-1}} \\
&= X_{t_n}\left(1 - e^{-\Delta t_n/\tau}\right) + e^{-\Delta t_n/\tau}\mathrm{EMA}_{\mathrm{eq}}(X,\tau)_{t_{n-1}} \\
&= \mathrm{EMA}_{\mathrm{eq}}(X,\tau)_{t_n}.
\end{aligned}$$

■

Hence, the $\mathrm{EMA}_{\mathrm{eq}}$, EMA,[17] and $\mathrm{EMA}_{\mathrm{next}}$ can be calculated recursively. Müller (1991) showed that the same holds true for the $\mathrm{EMA}_{\mathrm{lin}}$:

Proposition 8.9 *For $X \in \mathcal{T}$ and $\tau > 0$,*

$$\begin{aligned}
\mathrm{EMA}_{\mathrm{lin}}(X,\tau)_{t_1} &= X_{t_1}, \\
\mathrm{EMA}_{\mathrm{lin}}(X,\tau)_{t_n} &= e^{-\Delta t_n/\tau}\mathrm{EMA}_{\mathrm{lin}}(X,\tau)_{t_{n-1}} + X_{t_n}(1 - \omega(\tau,\Delta t_n)) \\
&+ X_{t_{n-1}}(\omega(\tau,\Delta t_n) - e^{-\Delta t_n/\tau}),
\end{aligned}$$

for $t_n \in T(X)$ with $n \ge 2$, where

$$\omega(\tau,\Delta t_n) = \frac{\tau}{\Delta t_n}\left(1 - e^{-\Delta t_n/\tau}\right).$$

In particular, $\omega(\tau,\Delta t_n) \approx 0$ for $\tau \ll \Delta t_n$ in which case $\mathrm{EMA}_{\mathrm{lin}}(X,\tau)_{t_n} \approx X_{t_n}$.

See Eckner (2012) for an efficient $O(N(X))$ implementation in the `C` programming language of exponential moving averages and various other time series operators for unevenly spaced data.

[17] The reasoning for this exponential moving average is similar to the proof of Proposition 8.8.

8.3 Continuous-Time Analog

Calculating moving averages in *discrete* time requires us to choose a of the sampling scheme. In contrast, there is no such choice when calculating in *continuous* time, a fact which allows succinct illustration of certain concepts.

Definition 8.10 *Let $\widetilde{X} : \mathbb{R} \to \mathbb{R}$ be a function satisfying suitable integrability conditions. For $\tau > 0$ we call the function $\mathrm{SMA}(\widetilde{X}, \tau) : \mathbb{R} \to \mathbb{R}$ defined by*

$$\mathrm{SMA}(\widetilde{X}, \tau)_t = \frac{1}{\tau} \int_0^\tau \widetilde{X}_{t-s} ds, \qquad t \in \mathbb{R},$$

the simple moving average of $\widetilde{X}$ (with length τ), and $\mathrm{EMA}(\widetilde{X}, \tau) : \mathbb{R} \to \mathbb{R}$ defined by

$$\mathrm{EMA}(\widetilde{X}, \tau)_t = \frac{1}{\tau} \int_0^\infty \widetilde{X}_{t-s} e^{-s/\tau} ds, \qquad t \in \mathbb{R}, \tag{8.25}$$

the exponential moving average of $\widetilde{X}$ (with length τ).

Unless stated otherwise, we apply a time series operator to a real-valued function using its natural analog. For example, $L(\widetilde{X}, \tau)$ for $\tau \in \mathbb{R}$ is the real-valued function with $L(\widetilde{X}, \tau)_t = \widetilde{X}_{t-\tau}$ for all $t \in \mathbb{R}$. Of course, whenever there is a risk of confusion, we must explicitly define this extension.

With the SMA and EMA in continuous time established, we are ready to show several fundamental relationships between the trend measures and returns of a time series.

Theorem 8.11 *Let $\widetilde{X} : \mathbb{R} \to \mathbb{R}$ be a differentiable function satisfying suitable integrability conditions. For $\tau > 0$ and $t \in \mathbb{R}$,*

$$\frac{\widetilde{X}_t - \widetilde{X}_{t-\tau}}{\tau} = \mathrm{SMA}\left(\widetilde{X}', \tau\right)_t \tag{8.26}$$

and

$$\frac{1}{\tau} \log\left(\frac{\widetilde{X}_t}{\widetilde{X}_{t-\tau}}\right) = \mathrm{SMA}\left(\log\left(\widetilde{X}\right)', \tau\right)_t, \tag{8.27}$$

if $\widetilde{X} > 0$, where $\widetilde{X}'$ denotes the derivative of $\widetilde{X}$.

Proof. By the fundamental theorem of calculus

$$\widetilde{X}_t - \widetilde{X}_{t-s} = \int_0^s \widetilde{X}'_{t-z} dz$$

$$\log\left(\widetilde{X}_t\right) - \log\left(\widetilde{X}_{t-s}\right) = \int_0^s \log\left(\widetilde{X}\right)'_{t-z} dz$$

for $s \in \mathbb{R}$. Hence, (8.26) and (8.27) follow from the definition of the simple moving average. ■

For an unevenly spaced time series $X \in \mathcal{T}$ (in discrete time),

$$\frac{1}{\tau} \mathrm{ret}_{\mathrm{abs}}^{\mathrm{roll}}(X, \tau)_t = \frac{X_t - X[t-\tau]}{\tau} \approx \mathrm{SMA}_{\mathrm{next}}\left(\frac{\Delta X}{\Delta t(X)}, \tau\right)_t$$

and

$$\frac{1}{\tau}\text{ret}_{\log}^{\text{roll}}\left(X,\tau\right)_t = \frac{1}{\tau}\log\left(\frac{X_t}{X\left[t-\tau\right]}\right) \approx \text{SMA}_{\text{next}}\left(\frac{\Delta\log\left(X\right)}{\Delta t\left(X\right)},\tau\right)$$

for $t \in T\left(X\right)$ and $\tau > 0$ can be used as an approximation of (8.26) and (8.27), respectively. In some cases, the approximation holds exactly. For example, if $t-\tau \in T\left(X\right)$, then $t = t_n$ and $t-\tau = t_{n-k}$ for some $1 \le k < n \le N\left(X\right)$, so that

$$\begin{aligned}
\text{SMA}_{\text{next}}\left(\frac{\Delta X}{\Delta t\left(X\right)},\tau\right)_t &= \frac{1}{\tau}\left(\frac{\Delta X_{t_n}\Delta t_n}{\Delta t_n} + \ldots + \frac{\Delta X_{t_{n-k+1}}\Delta t_{n-k+1}}{\Delta t_{n-k+1}}\right) \\
&= \frac{1}{\tau}\left(\Delta X_{t_n} + \Delta X_{t_{n-1}} + \ldots + \Delta X_{t_{n-k+1}}\right) \\
&= \frac{X_{t_n} - X_{t_n-\tau}}{\tau} \\
&= \frac{X_t - X\left[t-\tau\right]}{\tau}.
\end{aligned} \tag{8.28}$$

A similar result holds for exponential moving averages

Theorem 8.12 *Let $\widetilde{X} : \mathbb{R} \to \mathbb{R}$ be a differentiable function satisfying suitable integrability conditions. For $\tau > 0$ and $t \in \mathbb{R}$,*

$$\frac{\widetilde{X}_t - \text{EMA}(\widetilde{X},\tau)_t}{\tau} = \text{EMA}(\widetilde{X}',\tau)_t \tag{8.29}$$

where $\widetilde{X}'$ denotes the derivative of $\widetilde{X}$.

Proof. The result can again be shown using the fundamental theorem of calculus, but the calculation is tedious. Alternatively, partial integration gives

$$\begin{aligned}
\text{EMA}(\widetilde{X}',\tau)_t &= \frac{1}{\tau}\int_0^\infty \widetilde{X}'_{t-z} e^{-z/\tau} dz \\
&= -\frac{1}{\tau}\widetilde{X}_{t-z}e^{-z/\tau}\Big|_{z=0}^{z=\infty} - \frac{1}{\tau^2}\int_0^\infty \widetilde{X}_{t-z}e^{-z/\tau}dz \\
&= \frac{1}{\tau}\widetilde{X}_t - \frac{1}{\tau}\text{EMA}\left(\widetilde{X},\tau\right)_t.
\end{aligned}$$

■

For an unevenly spaced time series $X \in \mathcal{T}$ (in discrete time),

$$\frac{X_t - \text{EMA}(X,\tau)_t}{\tau} \approx \text{EMA}_{\text{next}}\left(\frac{\Delta X}{\Delta t\left(X\right)},\tau\right)_t$$

can be used as an approximation of (8.29).

9 Scale and Volatility

In this section we predominantly focus on volatility estimation for time series generated by Brownian motion. In particular, we do not examine processes with jumps or time-varying volatility. However, most results can be extended to more general processes; for example, by using a rolling time window to estimate time-varying volatility, or by using the methods in Barndorff-Nielsen and Shephard (2004) to allow for jumps.

First, recall a couple of elementary properties of Brownian motion.

Lemma 9.1 *Let $(B_t : t \geq 0)$ be a standard Brownian motion with $B_0 = 0$. For $t > 0$,*

(i) $E(|B_t|) = \sqrt{2t/\pi}$,

(ii) $E(\max_{0 \leq s \leq t} B_s) = \sqrt{2t/\pi}$, *and*

(iii) $E(\max_{0 \leq r,s \leq t} |B_r - B_s|) = \sqrt{8t/\pi}$.

(iv) Let $(\pi_n : n \geq 1, \pi_n \in \mathbb{T})$ be a refining sequence of partitions of $[0,t]$ with $\lim_{n\to\infty}\text{mesh}(\pi_n) = 0$. Then

$$\lim_{n\to\infty} \pi_n^p B = \lim_{n\to\infty} \sum_{t_i \in \pi_n} \left(B_{t_{i+1}} - B_{t_i}\right)^p = \begin{cases} \infty, & \text{if } 0 \leq p < 2, \\ t, & \text{if } p = 2, \\ 0, & \text{if } p > 2, \end{cases}$$

almost surely.

Proof. B_t is a normal random variable with density $\phi(0,t)$, and therefore the probability density of $|B_t|$ is $2\phi(x,t)\mathbf{1}_{(x\geq 0)}$. Hence,

$$\begin{aligned} E(|B_t|) &= 2\int_0^\infty x\phi(x,t)\,dx \\ &= \frac{2}{\sqrt{2\pi t}}\int_0^\infty xe^{-\frac{x^2}{2t}}\,dx \\ &= -\sqrt{\frac{2t}{\pi}}e^{-\frac{x^2}{2t}}\Big|_0^\infty \\ &= \sqrt{\frac{2t}{\pi}}. \end{aligned}$$

By the reflection principle, $\max_{0\leq s\leq t} B_s$ has the same distribution as $|B_t|$; see Durrett (2005), Example 7.4.3. The third result follows from $\max_{0\leq r,s\leq t}|B_r - B_s| = \max_{0\leq s\leq t} B_s - \min_{0\leq s\leq t} B_s$ and the symmetry of Brownian motion. For the last result see Protter (2005), Chapter 1, Theorem 28. ■

We examine two different types of consistency for the estimation of quadratic variation (and also, therefore, volatility); namely, (i) consistency for a fixed observation time window as the spacing of observation times goes to zero, and (ii) consistency as the length of the observation time window goes to infinity with a constant "density" of observation times.

Definition 9.2 *Assume we are given a stochastic process $(\widetilde{X}_t : t \geq 0)$ and let $\langle \widetilde{X}, \widetilde{X}\rangle$ denote the continuous part of its quadratic variation $[\widetilde{X}, \widetilde{X}]$. A time series operator O is said to be*

(i) a $\pi-$consistent estimator of $\langle \widetilde{X}, \widetilde{X}\rangle$ if, for every $t > 0$ and refining sequence of partitions $(\pi_n : n \geq 1, \pi_n \in \mathbb{T})$ of $[0,t]$ with $\lim_{n\to\infty}\text{mesh}(\pi_n) = 0$,

$$\lim_{n\to\infty} O\left(\widetilde{X}[\pi_n]\right)_t = \langle \widetilde{X}, \widetilde{X}\rangle_t,$$

almost surely.

(ii) *a $T-$consistent estimator of $\langle \widetilde{X}, \widetilde{X} \rangle$, if for every sequence $T = (t_n)_{n\geq 1}$ of strictly increasing observation times with $t_1 = 0$, $\lim_{n\to\infty} t_n = \infty$ and bounded inter-observation times $t_n - t_{n-1}$,*

$$\lim_{n\to\infty} \frac{O\left(\widetilde{X}\left[T \cap [0, t_n]\right]\right)_{t_n}}{\langle \widetilde{X}, \widetilde{X} \rangle_{t_n}} = 1$$

almost surely; and

(iii) *a $T-$consistent volatility estimator if, for every sequence $T = (t_n)_{n\geq 1}$ of strictly increasing observation times with $t_1 = 0$, $\lim_{n\to\infty} t_n = \infty$ and bounded inter-observation times $t_n - t_{n-1}$,*

$$\lim_{n\to\infty} \frac{O\left(\widetilde{X}\left[T \cap [0, t_n]\right]\right)_{t_n}}{\langle \widetilde{X}, \widetilde{X} \rangle_{t_n} / t_n} = 1$$

almost surely.

Note that generally, neither $\pi-$consistent nor $T-$consistent quadratic variation estimators yield unbiased estimates of $\langle \widetilde{X}, \widetilde{X} \rangle$ from the observation time series X, because consistency is required only in the limit.

Lemma 9.3 *For a given a stochastic process $(\widetilde{X}_t : t \geq 0)$, a time series operator O is a $T-$consistent estimator of $\langle \widetilde{X}, \widetilde{X} \rangle$ if and only if*

$$O_\sigma : X \to \frac{1}{\max T(X)} \sqrt{O(X)}$$

is a $T-$consistent volatility estimator.

Definition 9.4 *For a time series $X \in \mathcal{T}$, the realized $p-$variation of X is*

$$\mathrm{Var}(X, p) = \mathrm{Cumsum}\left(|X - B(X)|^p\right),$$

where the time series operator Cumsum *replaces the observation values of a time series with their cumulative sum. In particular, using $p = 1$ gives the total variation, and $p = 2$ gives the quadratic variation $[X, X]$ of X.*

Proposition 9.5 *The realized quadratic variation* $\mathrm{Var}(X, p = 2)$ *is a $\pi-$consistent and $T-$consistent quadratic variation estimator for scaled Brownian motion, $a + \sigma B$, where $a \in \mathbb{R}$ and $\sigma > 0$.*

Proof. Let $\widetilde{X} = a + \sigma B$. For $p = 2$,

$$\mathrm{Var}(X, p)_t = \sum_{t_i \in T(X), t_i < t} \left(X_{t_{i+1}} - X_{t_i}\right)^2$$

for $t \in T(X)$. Lemma 9.1 *(iv)* shows that this operator is a $\pi-$consistent quadratic variation estimator.
For showing $T-$consistency, define $Y_i = \left(X_{t_{i+1}} - X_{t_i}\right)^2$. Then

$$\sum_{i=1}^{\infty} \frac{\mathrm{Var}(Y_i)}{(\sigma^2 t_i)^2} \leq \left(\max_i \mathrm{Var}(Y_i)\right) \sum_{i=1}^{\infty} \frac{1}{(\sigma^2 t_i)^2} < \infty$$

for bounded inter-observation times. Kolmogorov's strong law of large numbers, see Chapter 3 in Shiryaev (1995), implies the second equality in the following set of equations:

$$\begin{aligned}
\lim_{n\to\infty} \frac{\operatorname{Var}(X, p=2)_{t_n}}{\langle \widetilde{X}, \widetilde{X} \rangle_{t_n}} &= \lim_{n\to\infty} \frac{1}{\sigma^2 t_n} \sum_{i=1}^{n-1} Y_i \\
&= \lim_{n\to\infty} \frac{E\left(\operatorname{Var}(X, p=2)_{t_n}\right)}{\sigma^2 t_n} \\
&= \lim_{n\to\infty} \frac{\sigma^2 t_n}{\sigma^2 t_n} \\
&= 1,
\end{aligned}$$

thereby showing the $T-$consistency of the quadratic variation estimator. ■

Definition 9.6 *For a time series $X \in \mathcal{T}$, the average true range (ATR) is given by*

$$\begin{aligned}
\operatorname{ATR}(X, \rho, \tau) &= \operatorname{SMA}(\operatorname{range}(X, \rho), \tau) \\
&= \operatorname{SMA}(\operatorname{rollmax}(X, \rho), \tau) - \operatorname{SMA}(\operatorname{rollmin}(X, \rho), \tau),
\end{aligned}$$

where $\rho > 0$ is the range horizon and $\tau > \rho$ is the smoothing horizon.

The ATR is a popular, robust measure of volatility, see Wilder (1978).

Proposition 9.7 *The mapping*

$$X \longmapsto \sqrt{\frac{\pi}{8\rho}} \operatorname{ATR}(X, \rho, \tau = \max T(X) - \min T(X))$$

for $\rho > 0$ is a $T-$consistent volatility estimator for scaled Brownian motion, $a + \sigma B$, where $a \in \mathbb{R}$ and $\sigma > 0$.

Proof. Let $\widetilde{X} = a + \sigma B$. Lemma 9.1 (iii) shows that for $t > \rho$,

$$\begin{aligned}
E\left(\operatorname{range}\left(\widetilde{X}, \rho\right)_t\right) &= E\left(\max_{t-\rho\le s\le t} \widetilde{X}_s - \min_{t-\rho\le s\le t} \widetilde{X}_s\right) \\
&= \sigma E\left(\max_{t-\rho\le s\le t}\left(\frac{\widetilde{X}_s - \widetilde{X}_{t-\rho}}{\sigma}\right)\right) - \sigma\sqrt{\rho} E\left(\min_{t-\rho\le s\le t}\left(\frac{\widetilde{X}_s - \widetilde{X}_{t-\rho}}{\sigma}\right)\right) \\
&= \sigma\sqrt{\frac{8\rho}{\pi}},
\end{aligned}$$

where we used that

$$W_s \equiv \frac{\widetilde{X}_{t-\rho+s} - \widetilde{X}_{t-\rho}}{\sigma} = B_{t-\rho+s} - B_{t-\rho}, \qquad s \ge 0$$

is a standard Brownian motion. The continuous-time ergodic theorem, see Bergelson et al. (2012), implies

$$\lim_{n\to\infty} \operatorname{ATR}(., \rho, \tau = t_n - t_1) = E\left(\operatorname{range}\left(\widetilde{X}, \rho\right)_\rho\right) = \sigma\sqrt{\frac{8\rho}{\pi}},$$

which gives the desired result. ■

The ATR can also be used to construct a $\pi-$consistent volatility estimator.

Proposition 9.8 *Assume we are given a positive function f on $[0, \infty)$ with*[18]

$$\lim_{x \to 0} f(x) = 0$$

and

$$\lim_{x \searrow 0} \frac{f(x)}{x} = \infty.$$

Fix $t > 0$. For a suitable constant $c(f, t)$, the mapping

$$X \longmapsto c(f, t) \operatorname{ATR}(X, \rho = f(\max \Delta t(X)), \tau = t)$$

is a $\pi-$consistent volatility estimator for scaled Brownian motion, $a + \sigma B$, where $a \in \mathbb{R}$ and $\sigma > 0$.

Proposition 9.9 *The following time series operators are $T-$consistent quadratic variation estimators for scaled Brownian motion.*

Proof. The argument is similar to that of Proposition 9.7. ∎

Lemma 9.10 *(i)* $c_{\mathrm{SMA}}(\rho)\mathrm{SMA}(|X - \mathrm{SMA}(X, \rho)|^2, \tau = \max T(X) - \min T(X))$ *for $\rho > 0$ and suitable constant $c_{\mathrm{SMA}}(\rho)$;*

(ii) $c_{\mathrm{EMA}}(\rho)\mathrm{EMA}(|X - \mathrm{EMA}(X, \rho)|^2, \tau = \max T(X) - \min T(X))$ *for $\rho > 0$ and suitable constant $c_{\mathrm{EMA}}(\rho)$.*

Proof. The argument is again similar to that of Proposition 9.7. ∎

Clearly, there is a trade-off between the robustness and efficiency of volatility estimation. For example, while the realized volatility is the maximum likelihood estimate (MLE) of σ for scaled Brownian motion, it is very sensitive to the presence of measurement noise - much more so than, for example, volatility estimators based on the ATR.

10 Conclusion

This paper presents methods for and analyzing and manipulating unevenly spaced time series in their unaltered form, without a transformation to equally spaced data. Nevertheless, the methods developed are consistent with the existing literature on equally spaced time series analysis. Topics for future research include auto- and cross-correlation analysis, spectral analysis, model specification and estimation, and forecasting methods.

[18] The conditions ensure that, as the mesh size goes to zero, the range horizon ρ goes to zero as well but with the number of time series observations within each range going to infinity at the same time.

Appendices

A Proof of Theorem 6.12

We proceed by breaking down the proof into three separate results.

Lemma A.1 *Let O be a shift-invariant time series operator that is linear for sampling scheme σ. There exists a function $h_\sigma : \mathrm{SP}_\sigma \to \mathbb{R}$ (as opposed to $\mathrm{SP}_\sigma \to \mathrm{SP}_\sigma$ in Lemma 6.11) such that*

$$\mathrm{SP}_\sigma \left(O\left(X\right)\right)\left(t\right) = h_\sigma \left(\mathrm{SP}_\sigma \left(L\left(X, -t\right)\right)\right) \tag{A.30}$$

for all time series $X \in \mathcal{T}$ and all $t \in \mathbb{R}$. In other words, the sample path of the output time series depends only on the sample path of the input time series, and this dependence is shift-invariant.

Proof. By Lemma 6.5 and the shift-invariance of O,

$$\begin{aligned}\mathrm{SP}_\sigma \left(O\left(X\right)\right)\left(t-\tau\right) &= \mathrm{SP}_\sigma \left(L\left(O\left(X\right), \tau\right)\right)\left(t\right) \\ &= \mathrm{SP}_\sigma \left(O\left(L\left(X, \tau\right)\right)\right)\left(t\right)\end{aligned}$$

for all $X \in \mathcal{T}$ and $t, \tau \in \mathbb{R}$. Setting $t = 0$ and $\tau = -t$ gives

$$\mathrm{SP}_\sigma \left(O\left(X\right)\right)\left(t\right) = \mathrm{SP}_\sigma \left(O\left(L\left(X, -t\right)\right)\right)\left(0\right) \tag{A.31}$$

for all $t \in \mathbb{R}$. According to Lemma 6.11 there exists a function $g_\sigma : \mathrm{SP}_\sigma \to \mathrm{SP}_\sigma$ such that

$$\mathrm{SP}_\sigma \left(O\left(L\left(X, -t\right)\right)\right) = g_\sigma \left(\mathrm{SP}_\sigma \left(L\left(X, -t\right)\right)\right). \tag{A.32}$$

Combining (*A.*31) and (*A.*32) yields

$$\mathrm{SP}_\sigma \left(O\left(X\right)\right)\left(t\right) = \mathrm{SP}_\sigma \left(O\left(L\left(X, -t\right)\right)\right)\left(0\right) = g_\sigma \left(\mathrm{SP}_\sigma \left(L\left(X, -t\right)\right)\right)\left(0\right).$$

Hence, the desired function h_σ is given by $h_\sigma\left(x\right) = g_\sigma\left(x\right)\left(0\right)$ for $x \in \mathrm{SP}_\sigma$. ■

For operators that are bounded, even more can be said.

Theorem A.2 *Let O be a bounded and shift-invariant time series operator that is linear for sampling scheme σ. Then there exists a finite signed measure μ_T on $(\mathbb{R}, \mathcal{B})$ such that*

$$\mathrm{SP}_\sigma \left(O\left(X\right)\right)\left(t\right) = \int_{-\infty}^{+\infty} \mathrm{SP}_\sigma \left(X\right)\left(t-s\right) d\mu_T\left(s\right) \tag{A.33}$$

for all $t \in \mathbb{R}$, or equivalently

$$\mathrm{SP}_\sigma \left(O\left(X\right)\right) = \mathrm{SP}_\sigma \left(X\right) * \mu_T,$$

where "$*$" denotes the convolution of a real function with a Borel measure.

Proof. By Lemma A.1 there exists a function $h_\sigma : \mathrm{SP}_\sigma \to \mathbb{R}$ such that

$$\begin{aligned}\mathrm{SP}_\sigma\left(O\left(X\right)\right)(t) &= h_\sigma\left(\mathrm{SP}_\sigma\left(L\left(X,-t\right)\right)\right)\\ &= h_\sigma\left(\mathrm{SP}_\sigma\left(\widetilde{X}\right)\right)\\ &= \mathrm{SP}_\sigma\left(O\left(\widetilde{X}\right)\right)(0),\end{aligned}$$

where $\widetilde{X} = L\left(X,-t\right)$. Furthermore, by Lemma 6.5 (setting $t=-s$ and $\tau=-t$)

$$\mathrm{SP}_\sigma\left(X\right)(t-s) = \mathrm{SP}_\sigma\left(\widetilde{X}\right)(-s).$$

Hence, it suffices to show (*A*.33) for $t=0$, or equivalently, that there exists a finite signed measure μ_T on $(\mathbb{R},\mathcal{B})$ such that

$$h_\sigma\left(\mathrm{SP}_\sigma\left(X\right)\right) = \int_{-\infty}^{+\infty} \mathrm{SP}_\sigma\left(X\right)(-s)\, d\mu_T\left(s\right) \tag{A.34}$$

for all $X \in \mathcal{T}$.

Let us for a moment consider only the last-point and next-point sampling scheme. Define the set of indicator time series $I_{a,b} \in \mathcal{T}$ for $-\infty \le a \le b < \infty$ by

$$T\left(I_{a,b}\right) = \begin{cases} \left(a-1,a,b\right), & \text{if } a<b,\\ \left(a\right), & \text{if } a=b,\end{cases}$$

and

$$V\left(I_{a,b}\right) = \begin{cases} \left(0,1,0\right), & \text{if } a<b,\\ \left(0\right), & \text{if } a=b.\end{cases}$$

The sample path of an indicator time series is given $\mathrm{SP}_\sigma\left(I_{a,b}\right)(t) = \mathbf{1}_{[a,b)}(t)$ for $t \in \mathbb{R}$, which gives these time series their name. For disjoint intervals $[a_i,b_i)$ with $a_i \le b_i$ for $i \in \mathbb{N}$, we define

$$\mu_T\left(\bigcup_i [a_i,b_i)\right) = h_\sigma\left(\mathrm{SP}_\sigma\left(\sum_i I_{(-b_i,-a_i)}\right)\right).$$

Because O and therefore h_σ are bounded,

$$\begin{aligned}\mu_T\left(\bigcup_i [a_i,b_i)\right) &= h_\sigma\left(\mathrm{SP}_\sigma\left(\sum_i I_{(-b_i,-a_i)}\right)\right)\\ &\le M\left\|\mathrm{SP}_\sigma\left(\sum_i I_{(-b_i,-a_i)}\right)\right\|_{\mathrm{SP}}\\ &= M < \infty.\end{aligned}$$

Combined with the linearity of O we have that μ_T is a finite countably additive set function on $(\mathbb{R},\mathcal{A})$ where

$$\mathcal{A} = \left\{\bigcup_i [a_i,b_i) : -\infty \le a_i \le b_i < \infty \text{ for } i \in \mathbb{N}\right\}.$$

By Carathéodory's extension theorem, there exists a unique extension of μ_T (for simplicity, also called μ_T) to $(\mathbb{R}, \mathcal{B})$.

We are left to show that μ_T satisfies ($A.34$). To this end, note that the sample path of each time series $X \in \mathcal{T}$ can be written as a weighted sum of the sample paths of indicator time series. For example, in the case of last-point sampling,

$$\mathrm{SP}\left(X\right) = \sum_{i=0}^{N(X)} X_{t_i} \mathrm{SP}_\sigma\left(I_{t_i,t_{i+1}}\right)$$

where we define $t_0 = -\infty$, $X_{t_0} = X_{t_1}$, and $t_{N(X)+1} = +\infty$. Using the linearity of O, and therefore h_σ,

$$\begin{aligned}
h_\sigma\left(\mathrm{SP}\left(X\right)\right) &= h_\sigma\left(\sum_{i=0}^{N(X)} X_{t_i} \mathrm{SP}\left(I_{t_i,t_{i+1}}\right)\right) \\
&= \sum_{i=0}^{N(X)} X_{t_i} h_\sigma\left(\mathrm{SP}\left(I_{t_i,t_{i+1}}\right)\right) \\
&= \sum_{i=0}^{N(X)} X_{t_i} \mu_T\left([-t_{i+1}, -t_i\right) \\
&= \sum_{i=0}^{N(X)} X_{t_i} \mu_T\left([0 - t_{i+1}, 0 - t_i\right).
\end{aligned}$$

The last expression equals

$$\int_{-\infty}^{+\infty} X\left[0 - s\right] d\mu_T\left(s\right) = \left(\mathrm{SP}\left(X\right) * \mu_T\right)\left(0\right),$$

see Remark 6.10. Thus, we have shown the desired result for last-point sampling, and the final steps of the proof for next-point sampling are completely analogous.

For sampling with linear interpolation, the proof is complicated by the fact that the sample path $\mathrm{SP}_{\mathrm{lin}}\left(I_{a,b}\right)$ of an indicator function is not an indicator function but rather a triangular function. However, a given step function can be approximated arbitrarily close by a sequence of trapezoid functions (which are elements of $\mathrm{SP}_{\mathrm{lin}}$), and the measure μ_T can be defined as the limiting value of h_σ when applied to this sequence of trapezoid functions. The rest of the proof then proceeds as above, but is notationally more involved. Alternatively, one can invoke a version of the Riesz representation theorem. Specifically, the sample path $\mathrm{SP}_{\mathrm{lin}}\left(X\right)$ of each time series $X \in \mathcal{T}$ is constant outside an interval of finite length and the space $\mathrm{SP}_{\mathrm{lin}}$ can therefore be embedded in the space of continuous real functions that vanish at infinity. A version of the Riesz representation theorem shows that bounded, linear functionals on the latter space can be written as integrals with respect to a finite, signed measure; see Arveson (1996), Theorem 5.2. ■

Proof of Theorem 6.12. By Theorem A.2 and tick-invariance, there exists a finite signed measure μ_T on $(\mathbb{R}, \mathcal{B})$ such that

$$\begin{aligned} O(X)_t &= \mathrm{SP}_\sigma(O(X))(t) \\ &= \int_{-\infty}^{+\infty} \mathrm{SP}_\sigma(X)(t-s)\, d\mu_T(s) \\ &= \int_{-\infty}^{+\infty} X[t-s]_\sigma\, d\mu_T(s), \end{aligned}$$

for all $t \in T(X) = T(O(X))$. Because O is causal,

$$O(X)_t = O(X+Y)_t$$

for all time series $Y \in \mathcal{T}$ with $\mathrm{SP}_\sigma(Y)(s) = 0$ for $s \le t$, which implies

$$\begin{aligned} 0 &= \int_{-\infty}^{+\infty} Y[t-s]_\sigma\, d\mu_T(s) \\ &= 0 + \int_{-\infty}^{0} Y[t-s]_\sigma\, d\mu_T(s) \end{aligned}$$

for all such time series. Hence, μ_T is identical to the zero measure on $(-\infty, 0)$, and μ_T can be restricted to $(\mathbb{R}_+, \mathcal{B}_+)$, which makes O a convolution operator. ■

B Frequently Used Notation

$\mathbf{0}_n$	the null vector of length n
$*_\sigma^\mu$	the convolution operator associated with signed measure μ and sampling scheme σ
B	the backshift operator, see Definition 2.7
$\mathcal{B}$	the Borel $\sigma-$algebra on $\mathbb{R}$
$\mathcal{B}_+$	the Borel $\sigma-$algebra on $\mathbb{R}_+$
L	the lag operator, see Definition 2.7
D	the delay operator, see Definition 2.7
$N(X)$	the number of observations of time series X, see Definition 2.2
$\mathbb{T}_n$	the space of strictly increasing time sequences of length n, see Definition 2.1
$\mathbb{T}$	the space of strictly increasing time sequences, see Definition 2.1
$\mathbb{R}^n$	$n-$dimensional Euclidean space
$\mathbb{R}_+$	the interval $[0, \infty)$
σ	one of three sampling schemes, see Definition 2.4 and 2.6
$\mathrm{SP}_\sigma(X)$	the sample path of a time series X with sampling scheme σ, see Definition 6.1
SP_σ	the space of time series sample paths for sampling scheme σ, see Definition 6.1
$\mathcal{T}_n$	the space of time series of length n, see Definition 2.1
$\mathcal{T}$	the space of unevenly spaced time series, see Definition 2.1
$\mathcal{T}^K$	the space of $K-$dimension time series, see Definition 7.1
$T(X)$	the vector of observation times of a time series X, see Definition 2.2
$V(X)$	the vector of observation values of a time series X, see Definition 2.2
$X[t]_\sigma$	the sampled value of time series X at time t with sampling scheme σ, see Def. 2.4
$X^c[T_X]$	the observation time series of a continuous-time stochastic process X^c, see Def. 2.6

References

Adorf, H. M. (1995). Interpolation of irregularly sampled data series - a survey. In *Astronomical Data Analysis Software and Systems IV*, pp. 460–463.

Arveson, W. (1996). Notes on measure and integration in locally compact spaces. Lecture Notes, Department of Mathematics, University of California Berkeley.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 2*, 1–37.

Belcher, J., J. S. Hampton, and G. Tunnicliffe-Wilson (1994). Parameterization of continuous time autoregressive models for irregularly sampled time series data. *Journal of the Royal Statistical Society. Series B 56*, 141–155.

Bergelson, V., A. Leibman, and C. G. Moreira (2012). From discrete- to continuous-time ergodic theorems. *Ergodic Theory and Dynamical Systems 32*, 383–426.

Beygelzimer, A., E. Erdogan, S. Ma, and I. Rish (2005). Statictical models for unequally spaced time series. In *SDM*.

Billingsley, P. (1995). *Probability and Measure* (Third ed.). New York: John Wiley & Sons.

Bos, R., S. de Waele, and P. M. T. Broersen (2002). Autoregressive spectral estimation by application of the burg algorithm to irregularly sampled data. *IEEE Transactions on Instrumentation and Measurement 51*(6), 1289–1294.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (2004). *Time Series Analysis: Forecasting and Control* (4th ed.). Wiley Series in Probability and Statistics.

Brockwell, P. J. (2008). Lévy-driven continuous-time arma processes. Working Paper, Department of Statistics, Colorado State University.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (Second ed.). Springer.

Brockwell, P. J. and R. A. Davis (2002). *Introduction to Time Series and Forecasting* (Second ed.). Springer.

Broersen, P. M. T. (2008). Time series models for spectral analysis of irregular data far beyond the mean data rate. *Measurement Science and Technology 19*, 1–14.

Dacorogna, M. M., R. Gençay, U. A. Müller, R. B. Olsen, and O. V. Pictet (2001). *An Introduction to High-Frequency Finance*. Academic Press.

Durrett, R. (2005). *Probability: Theory and Examples* (Third ed.). Duxbury Advanced Series.

Eckner, A. (2012). Algorithms for unevenly-spaced time series: Moving averages and other rolling operators. Working Paper.

Fan, J. and Q. Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Series in Statistics.

Gilles Zumbach, U. A. M. (2001). Operators on inhomogeneous time series. *International Journal of Theoretical and Applied Finance 4*, 147–178.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning* (Second ed.). Springer. http://www-stat.stanford.edu/~tibs/ElemStatLearn.

Hayashi, T. and N. Yoshida (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli 11*, 359–379.

Jones, R. H. (1981). Fitting a continuous time autoregression to discrete data. In D. F. Findley (Ed.), *Applied Time Series Analysis II*, pp. 651–682. Academic Press, New York.

Jones, R. H. (1985). Time series analysis with unequally spaced data. In E. J. Hannan, P. R. Krishnaiah, and M. M. Rao (Eds.), *Time Series in the Time Domain, Handbook of Statistics*, Volume 5, pp. 157–178. North Holland, Amsterdam.

Karatzas, I. and S. E. Shreve (2004). *Brownian Motion and Stochastic Calculus* (Second ed.). Springer-Verlag.

Kolmogorov, A. N. and S. V. Fomin (1999). *Elements of the Theory of Functions and Functional Analysis*. Dover Publications.

Kreyszig, E. (1989). *Introductory Functional Analysis with Applications*. Wiley.

Lomb, N. R. (1975). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science 39*, 447–462.

Lundin, M. C., M. M. Dacorogna, and U. A. Müller (1999). Correlation of high frequency financial time series. In P. Lequex (Ed.), *The Financial Markets Tick by Tick*, Chapter 4, pp. 91–126. New York: John Wiley & Sons.

Lütkepohl, H. (2010). *New Introduction to Multiple Time Series Analysis*. Springer.

Müller, U. A. (1991). Specially weighted moving averages with repeated application of the ema operator. Working Paper, Olsen and Associates, Zurich, Switzerland.

Protter, P. (2005). *Stochastic Integration and Differential Equations* (Second ed.). New York: Springer-Verlag.

Rehfeld, K., N. Marwan, J. Heitzig, and J. Kurths (2011). Comparison of correlation analysis techniques for irregularly sampled time series. *Nonlinear Processes in Geophysics 18*, 389–404.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.

Scargle, J. D. (1982). Studies in astronomical time series analysis. ii. statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal 263*, 835–853.

Scholes, M. and J. Williams (1977). Estimating betas from nonsynchronous data. *Journal of Financial Economics 5*, 309–327.

Shiryaev, A. N. (1995). *Probability* (Second ed.). Springer.

Thiebaut, C. and S. Roques (2005). Time-scale and time-frequency analyses of irregularly sampled astronomical time series. *EURASIP Journal on Applied Signal Processing 15*, 2486–2499.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press.

Wand, M. and M. Jones (1994). *Kernel Smoothing*. Chapman and Hall/CRC.

Wilder, W. (1978). *New Concepts in Technical Trading Systems*. Trend Research.