A Framework for the Analysis of Unevenly-Spaced Time Series Data

Andreas Eckner*

First Version: February 16, 2010
Current version: March 28, 2012

Abstract

This paper presents methods for analyzing 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 a data transformation. Care is taken to develop a framework consistent with a traditional analysis for equally-spaced data like 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, convolution operator, trend, momentum

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 such as spectra of celestial objects are taken at times determined by weather conditions, the season, 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, and network traffic analysis.

Research on time series analysis usually specializes along one or more of the following dimensions: univariate vs. multivariate, linear vs. non-linear, parametric vs. non-parametric, evenly vs. unevenly-spaced. There exists an extensive body of literature on analyzing equally-spaced time series 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 at a time when limitations in computing resources favored an analysis of equally-spaced data (and the

---

*Comments are welcome at andreas@eckner.com.

use of linear Gaussian models), since in this case efficient linear algebra routines can be used and many problems have an explicit solution.

As a result, fewer methods exists specifically for analyzing unevenly-spaced time series data. Some authors have suggested an embedding into continuous diffusion processes, see Jones (1981), Jones (1985), and Brockwell (2008). However, the emphasis of this literature has primarily been on modeling univariate ARMA processes, as opposed to developing a full-fledged tool set analogous to the one available for equally-spaced data. In astronomy, a lot of effort has been gone into the task of estimating the spectrum of irregular time series data, see Lomb (1975), Scargle (1982), Bos et al. (2002), and Thiebaut and Roques (2005). Müller (1991), Gilles Zumbach (2001), and Dacorogna et al. (2001) examine unevenly-spaced time series in the context of high-frequency financial data.

Perhaps still the most widely used approach is to transform unevenly-spaced data into equally-spaced observations using some form of interpolation - most often linear - and then to apply existing methods for equally-spaced data. See Adorf (1995), and Beygelzimer et al. (2005). However, transforming data in such a way has a couple of significant drawbacks:

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*

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

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).

Example 1.2 (Causality) *For a given time series, the linearly interpolated observation at a time t (not equal to an 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 in general 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 may reduce and dilute 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*

---

[1] For a precise mathematical definition not offered here, see Karatzas and Shreve (2004) and Protter (2005).

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.*

The aim of this paper is to provide methods for directly analyzing unevenly-spaced time series data, while maintaining consistency with existing methods for equally-spaced time series. In particular, a major goal is to provide a mathematical and conceptual framework for manipulating univariate and multivariate time series with unevenly-spaced observations.

Section 2 introduces unevenly-spaced time series and some elementary operations for them. Section 3 defines time series operators and examines commonly encountered structural features of such operators. Section 4 introduces convolution operators, which are a particularly tractable and widely used class of time series operators. Section 5 provides a large number of examples, such as arithmetic and return operators, for the theory developed in the preceding sections. Section 7 turns to multivariate time series and corresponding vector time series operators. Section 8 provides a systematic treatment of various moving averages, which summarize the average value of a time series over a certain horizon. Section 9 focuses on scale and volatility, while an Appendix summarizes frequently used notation throughout the paper.

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,*

(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 by*

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

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

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

---

[2] See www.nyxdata.com/Data-Products/Historical-Data for a detailed description.

We will frequently use the informal but compact notation $((t_n, X_n) : 1 \leq n \leq N(X))$ and $(X_{t_n} : 1 \leq n \leq N(X))$ 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 introduce methods for extracting basic information from such objects.

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*

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

while the next available observation time is

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

For $\min T(x) \leq t \leq \max T(X)$, $\mathrm{Prev}^X(t) < t < \mathrm{Next}^X(t)$ unless $t \in T(X)$ in which case t is both the most recent and 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, $X[t]_{\mathrm{next}} = X_{\mathrm{Next}^X(t)}$ the next value of X at time t, and $X[t]_{\mathrm{lin}} = \left(1 - \omega^X(t)\right) X_{\mathrm{Prev}(X,t)} + \omega^X(t) X_{\mathrm{Next}^X(t)}$ with*

$$\omega^X(t) = \omega(T(X), t) = \begin{cases} \frac{t - \mathrm{Prev}^X(t)}{\mathrm{Next}^X(t) - \mathrm{Prev}^X(t)}, & \text{if } 0 < \mathrm{Next}^X(t) - \mathrm{Prev}^X(t) < \infty \\ 1, & \text{otherwise} \end{cases}$$

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

Note, 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 not appropriate for some applications, this convention greatly simplifies notation and avoids the treatment of a multitude of special cases in the exposition below.[4]

Remark 2.5 *For a time series $X \in \mathcal{T}$,*

(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 a function of t are right-continuous piecewise-constant functions with finite number of discontinuities.*

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

---

[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 since the result of any calculation involving `NAs` is also `NA`.

These observations suggest an alternative way of defining unevenly-spaced time series, namely as either piecewise-constant or piecewise-linear functions $X : \mathbb{R} \to \mathbb{R}$. However, such a representation 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) *For a time series $X \in \mathcal{T}$, sampling scheme $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$, and observation time sequence $T \in \mathbb{T}$,*

$$X_\sigma \left[T\right] = \left(\left(t_i, X_\sigma \left[t_i\right]\right) : t_i \in T\right)$$

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

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

Definition 2.7 *For a continuous-time stochastic process X^c and fixed observation time sequence $T_X \in \mathbb{T}$, $X = X^c\left[T_X\right]$ is the observation time series of X^c (at observation times T_X).*

By construction, $T\left(X\right) = T_X$ and $X_t = X^c_t$ for $t \in T\left(X\right)$ when sampling from a continuous-time stochastic process.

Definition 2.8 *For $X \in \mathcal{T}$, we call*

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

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

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

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

(v) $D\left(X, \tau\right) = \left(\left(t_n, X\left[t_n - \tau\right]\right) : 1 \le n \le N\left(X\right)\right) = L\left(X\left[T\left(X\right) - \tau\right], -\tau\right)$ 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 $\Delta t\left(X\right)$ 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 an infinite number of observations. However, these transformations are genuinely different for unevenly-spaced data, since the backshift operator shifts observation *values*, while the lag operator shifts observation *times*.[5]

---

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

Example 2.9 *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*

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

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

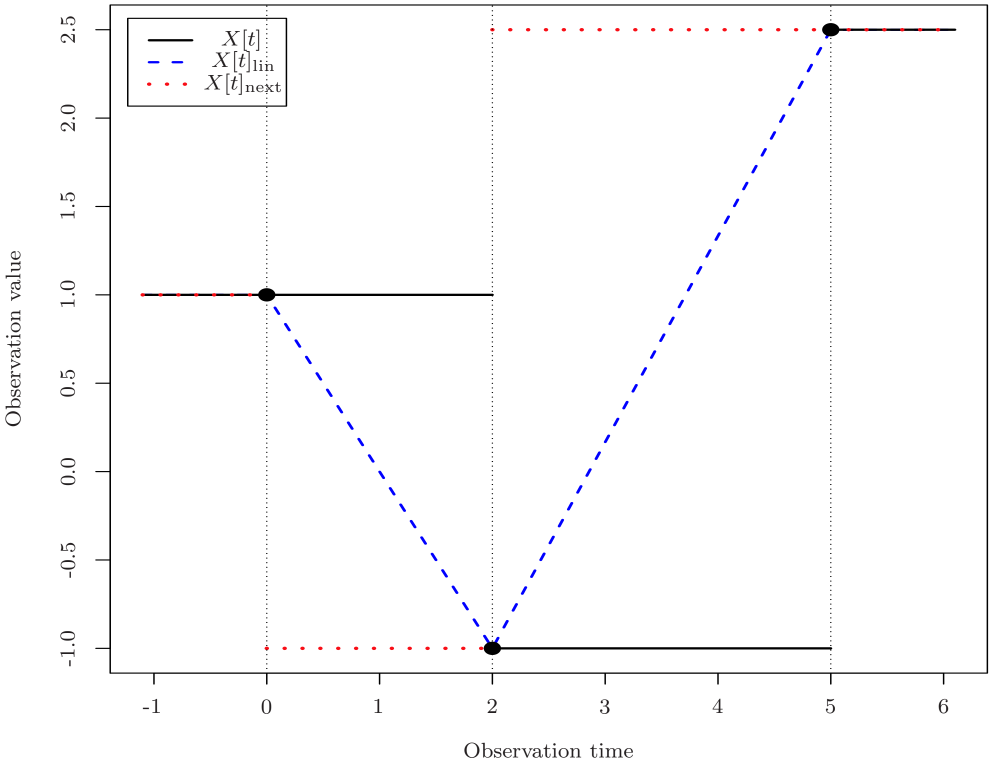

Figure 1: The sampling-scheme dependent graph of the time series X from Example 2.9. The three observation time-value pairs of the time series are represented by black circles.

These three graphs are shown in Figure 1. Moreover,

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

since $T(\Delta t(X)) = (2, 5)$ 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 elaborates the relationship between the lag operator L and the sampling operator.

Lemma 2.10 *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* $t \in T(L(X,\tau))$,

(iv) $L(X,\tau)[t]_\sigma = X[t-\tau]_\sigma$ *for* $t \in \mathbb{R}$ *and sampling scheme* $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$. *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.*

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

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

where the second equality follows from (i). Hence

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

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

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

3 Time Series Operators

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

The key difference between time series operators for equally and unevenly-spaced observations is that in the latter case, the observation values of the transformed series can (and generally do) 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* $O_T : \mathcal{T} \to \cup_{n\geq 1}\mathbb{T}_n$ *is the transformation of observation times,* $O_V : \mathcal{T} \to \cup_{n\geq 1}\mathbb{R}^n$ *the transformation of observation values, and where* $|O_T(X)| = |O_V(X)|$ *for all* $X \in \mathcal{T}$.

The constraint at the end of the definition simply ensures that the number of observation values and observation times for the transformed time series agree. Using this notation, we in particular have $T\left(O\left(X\right)\right)=O_T\left(X\right)$ and $V\left(O\left(X\right)\right)=O_V\left(X\right)$ for any time series $X\in\mathcal{T}$ and time series operator O. The above definition of a time series operator is completely general. In practice, most operators share one or more additional structural features, a list of which follows.

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

The remainder of this section is somewhat technical and may be skimmed over by readers primarily interested in applications. However, the dedicated effort will prove fruitful for the detailed analysis of linear time series operators in Section 6.

3.1 Tick or Time Invariance

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

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

Such operators cover the vast majority of cases 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* $\max T\left(O\left(X\right)\right)\leq\max T\left(X\right)$ *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*

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

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] In many cases, the following convenient characterization is useful.

---

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

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

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

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

Proof.

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

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

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

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

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

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

■

In particular, the lag and subperiod operator are interchangeable for causal, tick-invariant operators.

3.3 Shift Invariance or Stationarity

For many data sets, only the relative position of observation times is of interest, and this property should be reflected in the analysis carried out.

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

$$O(L(X,\tau)) = L(O(X),\tau),$$

or equivalently

$$O_T(L(X,\tau)) = O_T(X) + \tau, \text{ and} \tag{3.4}$$

$$O_V(L(X,\tau)) = O_V(X). \tag{3.5}$$

A shift-invariant operator does not use any special knowledge about the *absolute* values of observation times, but only about their relative position. This intuition is made precise by the following

Lemma 3.8 *A time series operator O is shift invariant, if and only there exist functions $f_m : (0,\infty)^{m-1} \times \mathbb{R}^m \to \cup_{n=1}^{\infty} \mathbb{T}_n$ and $g_m : (0,\infty)^{m-1} \times \mathbb{R}^m \to \cup_{n=1}^{\infty} \mathbb{R}^n$ for $m = 1,2,\ldots,$ such that for all $X \in \mathcal{T}$,*

$$O_T(X) = t_1 + f_{N(X)}(V(\Delta t(X)), V(X)), \tag{3.6}$$

$$O_V(X) = g_{N(X)}(V(\Delta t(X)), V(X)), \tag{3.7}$$

$$|O_T(X)| = |O_V(X)|. \tag{3.8}$$

Proof.

$\Longleftarrow$ Note that $V\left(\Delta t(L(X,\tau))\right) = V\left(\Delta t\left(X\right)\right)$ for all $\tau \in \mathbb{R}$. Hence, (3.6) implies $O_T\left(L\left(X,\tau\right)\right) = O_T\left(X\right) + \tau$, while (3.7) implies $O_V\left(L\left(X,\tau\right)\right) = O_V\left(X\right)$. Condition (3.8) ensures that the number of observation values and observation times are the same for the output time series.

$\Longrightarrow$ Assume O is a shift-invariant time series operator without structure (3.6) – (3.8). Then there exists a $k \geq 1$ and time series $X^{(1)}, X^{(2)} \in \mathcal{T}_k$ with

$$V\left(\Delta t\left(X^{(1)}\right)\right) = V\left(\Delta t\left(X^{(2)}\right)\right) \tag{3.9}$$

and

$$V\left(X^{(1)}\right) = V\left(X^{(2)}\right) \tag{3.10}$$

but either (i) $O_T\left(X^{(1)}\right) \neq O_T\left(X^{(2)}\right) - \tau$ for $\tau = \min T\left(X^{(2)}\right) - \min T\left(X^{(1)}\right)$, or (ii) $O_V\left(X^{(1)}\right) \neq O_V\left(X^{(2)}\right)$. However, (3.9) and (3.10) imply that $X^{(2)} = L(X^{(1)},\tau)$. Hence, since O is shift invariant

$$O(X^{(2)}) = O(L(X^{(1)},\tau)) = L\left(O(X^{(1)}),\tau\right)$$

and therefore $O_T\left(X^{(1)}\right) = O_T\left(X^{(2)}\right) - \tau$ and $O_V\left(X^{(1)}\right) = O_V\left(X^{(2)}\right)$, which contradicts the assumption that O does not have the structure (3.6) – (3.8).

■

Corollary 3.9 *A tick-invariant time series operator O is shift invariant, if and only if there exist functions $g_m : (0,\infty)^{m-1} \times \mathbb{R}^m \to \mathbb{R}^m$ for $m = 1, 2, \ldots$, such that*

$$O_V\left(X\right) = g_{N(X)}\left(V\left(\Delta t\left(X\right)\right), V\left(X\right)\right)$$

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

Corollary 3.10 *A causal, tick-invariant time series operator O is shift invariant, if and only if*

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

for all $X \in \mathcal{T}$ and $t \in T\left(X\right)$. In other words, there exists a function $g : \{X \in \mathcal{T} : \max T\left(X\right) = 0\} \to \mathbb{R}$ such that

$$O\left(X\right)_t = g\left(L\left(X\left\{-\infty, t\right\}, -t\right)\right)$$

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

3.4 Time Scale Invariance

The measurement unit of the time scale is generally not of inherent interest in an analysis of time series data. We therefore usually focus on (families of) time series operators that are invariant under the time-scaling operator.

Definition 3.11 *For $X \in \mathcal{T}$ and $a > 0$, the time-scaling operator S_a is defined as*

$$S_a(X) = S(X, a) = ((at_n, X_n) : 1 \leq n \leq N(X)).$$

Lemma 3.12 *For $X \in \mathcal{T}$ and $a > 0$,*

(i) $T(S_a(X)) = aT(X)$,

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

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

(iv) $S_a(X)[t] = X[t/a]$ *for* $t \in \mathbb{R}$ *and sampling scheme* $\sigma \in \{,\mathrm{lin},\mathrm{next}\}$. *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 the one of Lemma 2.10. ■

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

Definition 3.13 *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 3.14 *A time series operator O is said to be homogenous of degree $d \geq 0$, if $O(aX) = a^d O(X)$ for all $X \in \mathcal{T}$ and $a > 0$.*[7]

For example, the tick-spacing operator Δt is homogenous of degree $d = 0$, moving averages are homogenous of degree $d = 1$, while the operator that calculates the (integrated) $p-$variation of a time series is homogenous of degree $d = p$.

---

[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.

Lemma 3.15 *A time series operator O is homogenous of degree $d \geq 0$, if and only if there exist functions $f_m : \mathbb{T}_m \times [-1,1]^m \to \cup_{n=1}^{\infty} \mathbb{T}_n$ and $g_m : \mathbb{T}_m \times [-1,1]^m \to \cup_{n=1}^{\infty} \mathbb{R}^n$ for $m = 1, 2, \ldots$, such that for all $X \in \mathcal{T}$,*

$$\begin{aligned} O_T(X) &= f_{N(X)}\left(T(X), \widetilde{V}(X)\right), \\ O_V(X) &= g_{N(X)}\left(T(X), \widetilde{V}(X)\right)(\max|V(X)|)^d, \\ |O_T(X)| &= |O_V(X)|, \end{aligned}$$

where

$$\widetilde{V}(X) = \begin{cases} V(X) / \max|V(X)|, & \text{if } \max|V(X)| > 0 \\ \mathbf{0}_{|N(X)|}, & \text{otherwise} \end{cases}$$

and where $\mathbf{0}_k$ denotes the $k-$dimensional null vector.

Proof. The proof is very similar to the one of Lemma 3.8. ■

Remark 3.16 *The sampling operator of Example 3.2 is not tick invariant, causal iff we use last-point sampling, not shift invariant, not timescale invariant, homogenous of degree $d = 1$, and linear (see Section 6).*

4 Convolution Operators

Convolution operators are a class of tick-invariant, causal, shift-invariant (and often homogenous) 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 (ii) 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 $\mu(A) = \sum_i \mu(A_i)$.

Lemma 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 $A \in \Sigma$.

This results 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*

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

for all bounded piecewise linear functions f on $\mathbb{R}_+$.

---

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

The integrability condition ensures that the value of a convolution is well defined. In particular, the condition is satisfied if $d\mu(x,s) = g(x)\,\mu_T(s)$ for a 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 $*_\sigma^\mu(X) = X *_\sigma \mu$ is given by*

$$T(X *_\sigma \mu) = T(X), \tag{4.11}$$

$$(X *_\sigma \mu)_t = \int_0^\infty d\mu\left(X[t-s]_\sigma, s\right), \qquad t \in T(X *_\sigma \mu), \tag{4.12}$$

where the integration is over the time variable s. If μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R} \times \mathbb{R}_+$, then (4.12) can be written as

$$(X *_\sigma \mu)_t = \int_0^\infty f\left(X[t-s]_\sigma, s\right) ds, \qquad t \in T(X *_\sigma \mu), \tag{4.13}$$

where f is the density function of μ.

Remark 4.4 (Discrete Time Analog) *The discrete-time analog to convolution operators are additive (and hence 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. In the case of linearity, the representation simplifies further as shown in Section 6.

In the remainder of the paper, we primarily focus on last-point sampling and therefore omit the "σ" from the notation of convolution operators unless necessary. Most subsequent results also hold for the other two sampling schemes, even though we do not always provide a separate proof.

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

Proof. Tick-invariance and causality follow immediately from the definition of a convolution operator. Shift-invariance can be shown either using Corollary 3.10, or directly, which we do here for illustrative purposes. To this end, let $X \in \mathcal{T}$ and $\tau > 0$. Using (4.11) twice we get

$$(*^\mu)_T\left(L(X,\tau)\right) = T\left(L(X,\tau)\right) = T(X) + \tau = T(X * \mu) + \tau = (*^\mu)_T(X) + \tau,$$

showing that $*^\mu$ satisfies (3.4). On the other hand, for $t \in T\left(*^\mu\left(L(X,\tau)\right)\right) = T(X) + \tau$,

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

where we used Lemma 2.10 (*iv*) for the second equality. Hence, we also have $(*^{\mu})_V (X) = (*^{\mu})_V (L(X,\tau))$ and $*^{\mu}$ is therefore shift invariant. ■

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

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

Example 4.6 *Let O be the time series 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 $t \in T(O(X))$ for all $X \in \mathcal{T}$. 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,*

(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]

For the remainder of the paper, when we say that the kernel of a convolution operator has a certain structure, we implicitly mean 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 as convolution operators or a simple transformation of such an operator.

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*

(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."*

Note that in general, $1/X\,[t]_{\mathrm{lin}}$ does not equal $(1/X)\,[t]_{\mathrm{lin}}$ for $X \in \mathcal{T}$ and $t \notin T(X)$, although we have equality for the other two sampling schemes.

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*

---

[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) $X +_\sigma Y$ *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* $(XY)_t = X[t]_\sigma Y[t]_\sigma$ *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*.*

Lemma 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 $t \in T(X * \mu)$

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

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 all time series $X, Y \in \mathcal{T}$*, each sampling scheme* $\sigma \in \{, \text{lin}, \text{next}\}$*, and* $a, b \in \mathbb{R}$*.*

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? For example, if X and Y are discretely observed realizations of continuous time stochastic processes X^c and Y^c, respectively, we might want $(X+Y)[t]$ 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 hold.

Example 5.5 *Let* X^c *and* Y^c *be independent continuous-time stochastic processes,* $T_X, T_Y \in \mathbb{T}$ *fixed observation time sequences, and* $X = X^c[T_X]$ *and* $Y = Y^c[T_Y]$ *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*

$$E\left((X^c + Y^c)_t \mid \mathcal{H}_t\right) = (X+Y)[t]. \tag{5.14}$$

for all $t \geq \max(\min T_X, \min T_Y)$*, 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,[10] *although not necessarily martingales, then*

$$E\left(\left(X^c+Y^c\right)_t \mid \mathcal{H}_\infty\right) = X[t]_{\text{lin}} + Y\left[t\right]_{\text{lin}} = \left(X +_{\text{lin}} Y\right)[t] \tag{5.15}$$

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

Proof. Since X^c and Y^c are independent martingales,

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

showing (5.14). For a Lévy process Z and times $s \le t \le r$, $Z_r - Z_s$ 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, (5.15) follows from similar reasoning as (5.14). ■

Apart from the four basic arithmetic operations, scalar mathematical transformations can also be extended to unevenly-spaced time series.

Definition 5.6 (Elementwise Operations) *For $X \in \mathcal{T}$ and function $f: \mathbb{R} \to \mathbb{R}$, $f(X)$ denotes the time series that results when applying the function f to each observation value of X.*

For example,

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

In particular, 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 fixed time horizons, such as one day, month, or year. This section examines how to calculate such time series returns.

Definition 5.7 *For time series $X, Y \in \mathcal{T}$,*

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

---

[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 (5.15) without additional information. The LHS could even be larger than the largest observation value of $X +_{\text{lin}} Y$. Details are available from the author upon request.

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

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

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

Definition 5.8 (Returns) *For a time series* $X \in \mathcal{T}$*, time horizon* $\tau > 0$*, and return scale* $\gamma \in \{\mathrm{abs},\ \mathrm{rel},\ \log\}$*, 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*),*

(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*),*

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 first two return definitions is immediately clear, while Section 8 provides a motivation for the last two definitions.

Proposition 5.9 *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 a simple transformation.*

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 definition in Section 8, the proof for the other two return operators is similar. ■

---

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

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 to do just that.

Definition 5.10 (Rolling Time Series Functions) *Assume given a time horizon* $\tau \in \overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{\infty\}$ *and a function* $f : \mathcal{T}(\tau) \to \mathbb{R}$*, where* $\mathcal{T}(\tau) = \{X \in \mathcal{T}\text{: } \max(T(X)) - \min(T(X)) \leq \tau\}$ *denotes the space of time series with (temporal) length of at most* τ*. Further assume that* f *is shift invariant, that is,* $f(X) = f(L(X, \eta))$ *for all* $\eta \in \mathbb{R}$ *and* $X \in \mathcal{T}(\tau)$*. For a time series* $X \in \mathcal{T}$*, the "rolling function* f *of* X *over horizon* τ*," denoted by* $\operatorname{roll}(X, f, \tau)$*, is the time series with*

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

Proposition 5.11 *The class of rolling time series operators is identical to the class of causal, shift-invariant and tick-invariant operators.*

In particular, rolling time series functions include convolution operators. However, many operators that cannot be expressed as a convolution operator are included as well:

Example 5.12 *For a time series* $X \in \mathcal{T}$*, horizon* $\tau \in \overline{\mathbb{R}}_+$*, and function* $f : \mathcal{T}(\tau) \to \mathbb{R}$*, we call* $\operatorname{roll}(X, f, \tau)$ *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* $\operatorname{rollmax}(X, \tau)$*),*

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

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

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

of X *over horizon* τ*.*

6 Linear Time Series Operators

Linear operators 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 study some 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 $X \in \mathcal{T}$ since 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 $X, Y \in \mathcal{T}$ *and* $a, b \in \mathbb{R}$.

Proof. By the definition of a sample path 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* $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$. *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 $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$. 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 since is has no unique zero element. However, if we consider two time series X and Y to be equivalent if their sample paths are identical, 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 *The mapping* $X \to \mathrm{SP}_\sigma(X)$ *is an isometry between* $(\mathcal{T}, \| \, \|_{\mathcal{T}})$ *and* $(SP_\sigma, \| \, \|_{SP})$ *for each sampling scheme* $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$*. In other words,*

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

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

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

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

for all $X \in \mathcal{T}$*, and* $t, \tau \in \mathbb{R}$*, and sampling scheme* $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$*.*
Proof. *The result follows from Lemma 2.10 (iv).* ■

6.2 Bounded Linear Operators

Definition 6.6 *A time series operator O is said to be linear for a sampling scheme* $\sigma \in \{, lin, next\}$*, 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 homogenous 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 Section 5.3) is homogenous 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 $X \in \mathcal{T}$*,*

(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*

$$\mu(x, s) = x\mu_T(s), \tag{6.16}$$

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

Proof. First note

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

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

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

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

$$
\begin{aligned}
\left|(X *_\sigma \mu)_t\right| &= \left|\int_0^\infty X[t-s]_\sigma d\mu_T\left(s\right)\right| \qquad (6.17) \\
&\leq \int_0^\infty \left|X[t-s]_\sigma\right| \left|d\mu_T\left(s\right)\right| \\
&\leq \|X\|_{\mathcal{T}} \int_0^\infty \left|d\mu_T\left(s\right)\right| \\
&= \|X\|_{\mathcal{T}} \|\mu_T\|_{\mathrm{TV}},
\end{aligned}
$$

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 (6.17) 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 (6.16) are the only interesting bounded linear operators. To this end, we need to take a closer look at how individual time series observations are used by a linear convolution operator of the form (6.16).

Remark 6.10 *Assume given a linear convolution operator of form* (6.16) *and time series* $X \in \mathcal{T}$. *Define* $t_0 = -\infty$ *and* $X_{t_0} = X_{t_1}$. *For each observation time* $t = t_n \in T\left(X\right)$,

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

for last-point sampling,

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

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 coefficient $a_{k,n}$ *for* $1 \leq k \leq n$ *depend only on* μ_T *and the observation times but not the observation values of* X*.*

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, since 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* $\sigma \in \{, \text{lin}, \text{next}\}$*. There exists a function* $g_\sigma : \text{SP}_\sigma \to \text{SP}_\sigma$ *such that* $\text{SP}_\sigma(O(X)) = g_\sigma(\text{SP}_\sigma(X))$ *for all* $X \in \mathcal{T}$*. In other words, the sample path of the output time series depends only on the sample path of the input time series.*

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

$$g_\sigma(x) = SP_\sigma(O(X)).$$

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

$$aO(X -_\sigma Y) = O(a(X -_\sigma Y)) \tag{6.18}$$

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, (6.18) 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 one more time shows that $O(X)$ and $O(Y)$ have the same sample path, in contradiction to our assumption. ■

A lot more can be said for bounded linear operators that satisfy a couple of 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* $\sigma \in \{, \text{lin}, \text{next}\}$*. Then* O *is a convolution operator.*

Proof. See Appendix A. ■

Combining Theorem 6.9 and 6.12 gives the main result of this section:

---

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

Theorem 6.13 *The class of bounded, linear, causal, shift- and tick-invariant 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* $(\mathbb{R}_+, \mathcal{B}_+)$.

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* $(\mathbb{R}, \mathcal{B})$.

6.4 Application

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

Theorem 6.15 *Let* $(X_t^c : t \geq 0)$ *be a Lévy process,* $T_X \in \mathbb{T}$ *with* $\min T_X \geq 0$ *a fixed observation time sequence, and* $X = X^c[T_X]$ *the corresponding observation time series. Let* $\mu = x\mu_T$ *be a linear kernel with* $\mu_T(s) = 0$ *for* $s \geq C$ *for some constant* $C > 0$. *Hence, the time series operator* $*_{\text{lin}}^{\mu}$ *has a finite memory. Then*

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

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

$$E\left((X^c * \mu)_t \,|\, X\right) = \left(X *_{\text{lin}} \mu\right)_t \tag{6.19}$$

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

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

$$\begin{aligned} E\left((X^c * \mu)_t \,|\, X\right) &= E\left(\int_0^\infty d\mu\left(X_{t-s}^c, s\right) \,|\, X\right) \\ &= E\left(\int_0^C X_{t-s}^c \, d\mu_T(s)\, ds \,|\, X\right) \\ &= \int_0^C E\left(X_{t-s}^c \,|\, X\right) d\mu_T(s) \\ &= \int_0^C X[t-s]_{\text{lin}}\, d\mu_T(s) \\ &= \left(\mathrm{SP}_{\text{lin}}(X) * \mu_T\right)(t) \end{aligned} \tag{6.20}$$

where Fubini's theorem was used to change the order of integration and the conditional expectation. If in addition $t \in T(X)$, then (6.20) simplified further to

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

■

As already indicated by Example 1.1, (6.19) in general does not hold for non-linear time series operators, such as volatility or correlation estimators. In this case, the right-hand side of (6.19) 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, thus naturally leading to a treatment as a multivariate unevenly-spaced time series, even if the observations of individual time series are reported at regular intervals. For example, macroeconomic data like 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, respectively, in the case of the US). Moreover, the frequency of reporting may change over time.

Many multivariate time series operators are a natural extension of the univariate counterpart, so that the concepts of the prior sections require only few modifications. If one is primarily interested in the analysis of univariate time series, this section may be skipped on a first reading without impeding 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 basis information about vector time series objects.

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

$$X^K(i_1, \ldots, i_n) = \left(X_{i_1}^K, \ldots, X_{i_M}^K\right)$$

for $M \geq 2$, and $X^K(i_1) = X_{i_1}^K$ for $M = 1$, 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 $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$, we call*

$$X^K[t^K]_\sigma = \left(X_k^K[t_k^K]_\sigma : 1 \leq k \leq K\right) \tag{7.21}$$

is 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 common structural features among univariate time series operators. With very minor modifications (not shown here), the analysis is still relevant to the multivariate case, however, some other properties are also worth considering now.

7.1.1 Dimension Invariance

For many multivariate operators, the dimension of the output time series is either equal to one (typical for data aggregations) or equal to the dimension of the input time series (typical 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 end the discussion of structure 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*

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

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

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

The operator that merges two time series is tick invariant (in the sense that $T(X \cup Y) = T(X) \cup T(Y)$), causal, shift invariant (in the sense that $L(X \cup Y, \tau) = L(X, \tau) \cup L(Y, \tau)$), timescale invariant, homogenous of degree $d = 1$ (in the sense that $(aX) \cup (aY) = a(X \cup Y)$), not dimension invariant, and not permutation invariant.

7.2 Multivariate Convolution Operators

Multivariate convolution operators are a natural extension of the univariate case.

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*

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

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

Definition 7.9 (Multivariate Convolution Operator) *For a multivariate time series* $X^K \in \mathcal{T}^K$, $K-$*dimensional kernel* μ *and sampling scheme* $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$, *the convolution* $*_\sigma^\mu\left(X^K\right) = X^K *_\sigma \mu$ *is given by*

$$T\left(X^K *_\sigma \mu\right) = \bigcup_{k=1}^{K} T\left(X_k^K\right), \tag{7.22}$$

$$\left(X^K *_\sigma \mu\right)_t = \int_0^\infty d\mu\left(X^K[t-s]_\sigma, s\right), \qquad t \in T\left(X^K *_\sigma \mu\right), \tag{7.23}$$

If μ *is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^K \times \mathbb{R}_+$, *then* (7.23) *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 $\mathcal{T}^K \to \mathcal{T}$. More generally, a (K, M)-dimensional convolution operator is an $M-$tuple of $K-$dimensional convolution operators $\left(*_\sigma^{\mu_1}, \ldots, *_\sigma^{\mu_M}\right)$.

Note that a $(K,K)-$dimensional convolution operator is generally not equivalent to K one-dimensional convolution operators. In the former case, the $k-$th output time series depends on *all* input time series, while in the later case it depends only on the $k-$th 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 (7.22).

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* $\mu\left(x,y,s\right) = \delta_0\left(s\right)\left(x+y\right)$ *and* $\mu\left(x,y,s\right) = \delta_0\left(s\right) xy$, *respectively, for the vector time series* $(X,Y) \in \mathcal{T}^2$.

---

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

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

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

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

Definition 7.11 (Cross-sectional Operators) *For a function* $f : \mathbb{R}^K \to \mathbb{R}$, *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 $\mu^f\left(x^K, s\right) = \delta_0(s) f\left(x^K\right)$.

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 average 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* $\mathrm{range}_{\mathrm{C},\sigma}(X^K)$*),*

(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, $\mathrm{sum}_{\mathrm{C},\sigma}(X^K) = X_1^K + \ldots + X_K^K$ for all $K \geq 1$ and $X^K \in \mathcal{T}^K$.

Contemporaneous time series operators, such as the ones in Definition 7.11, 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 is changing over time. If X^K denotes the time series of individual incomes from a panel data set, $\mathrm{quant}_{\mathrm{C}}\left(X^K, 0.8\right) / \mathrm{quant}_{\mathrm{C}}\left(X^K, 0.2\right)$ 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 for summarizing the average value of a time series over a certain time horizon, for example, for the purpose of smoothing noisy observations. Moving averages are therefore closely related to kernel smoothing methods, see Wand and Jones (1994) and Hastie et al. (2009), except that in our case we use only past observations to get a causal time series operator.

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 exist multiple alternatives (for example, due to the choice of the sampling scheme), all of which may be sensible depending on the 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*

$$\mu(x, s) = x\mu_T(s) = x dF(s), \tag{8.24}$$

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

Using the remarks following Definition 2.4, we see 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 that a moving average operator satisfies the mentioned conditions.

$\Longleftarrow$ By Theorem 6.13, the kernel associated with O is of the form $\mu(x, s) = x\mu_T(s)$ for some finite signed measure μ_T on $(\mathbb{R}_+, \mathcal{B}_+)$. Since $O(X) \geq 0$ for all $X \in \mathcal{T}$ with $X \geq 0$, it follows that μ_T is a positive measure. Furthermore, since $O(X) = X$ for all constant time series $X \in \mathcal{T}$ it follows

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

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

■

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

$$\begin{aligned} \mathrm{MA}(X, \mu_T) &= \mathrm{MA}(X, F) &&= X * \mu, \\ \mathrm{MA}_{\mathrm{lin}}(X, \mu_T) &= \mathrm{MA}_{\mathrm{lin}}(X, F) &&= X *_{\mathrm{lin}} \mu, \\ \mathrm{MA}_{\mathrm{next}}(X, \mu_T) &= \mathrm{MA}_{\mathrm{next}}(X, F) &&= X *_{\mathrm{next}} \mu \end{aligned} \tag{8.25}$$

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

$$\mathrm{MA}_{\mathrm{next}}(X,F)_t \geq \mathrm{MA}_{\mathrm{lin}}(X,F)_t \geq \mathrm{MA}(X,F)_t \tag{8.26}$$

for all $t \in T(X)$, since $X[t]_{\mathrm{next}} \geq X[t]_{\mathrm{lin}} \geq X[t]$ 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) $\mathrm{SMA}_{\mathrm{next}}(X,\tau)_t = \frac{1}{\tau}\int_0^\tau X[t-s]_{\mathrm{next}}\,ds$,

(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 therefore is not a moving average operator in the sense of Definition 8.1. Nevertheless it is useful for demonstrating the difference between simple moving averages for equally- and unevenly-spaced time series.

The first SMA can be used for analyzing discrete observation *values*, for example, for calculating the average FED funds target rate[16] over the past three years. In this case, it is desirable to weigh observations by the amount of time each value remained unchanged. The $\mathrm{SMA}_{\mathrm{eq}}$ is ideal for analyzing discrete *events*, for example, for calculating the average number of casualties per deadly car accident over the past twelve months, or for determining the average number of IBM common shares traded on the NYSE per executed order during the past 30 minutes. The $\mathrm{SMA}_{\mathrm{lin}}$ can be used to estimate the rolling average value of a continuous time stochastic processes with observation times that are independent of the observation values, see Theorem 6.15. Finally, the $\mathrm{SMA}_{\mathrm{next}}$ is useful for certain trend and return calculations, see Section 8.3.

That said, the value of all SMAs will in general be quite similar as long as the moving average time horizon τ is considerably larger than the spacing of observation times. Rather, 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.

---

[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 at the Federal Reserve to other depository institutions overnight. See www.federalreserve.gov/fomc/fundsrate.htm for details.

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

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

while

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

We end our discussion of SMAs by illustrating the 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*

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

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

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

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

■

See Eckner (2011) 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 prefer the former name in this paper, since weights for unevenly-spaced time series are defined only implicitly, via a kernel, as opposed to explicitly 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 $t \in \{t_1, \ldots, t_{N(X)}\}$,*

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

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

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 $X \in \mathcal{T}$,*

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

while

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

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

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 EMA_{eq}, EMA,[17] and EMA_{next} can be calculated recursively. Müller (1991) has shown that the same holds true for the EMA_{lin}:

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

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

for $t_n \in T\left(X\right)$ *with* $n \geq 2$, *where*

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

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

See Eckner (2011) for an efficient $O\left(N\left(X\right)\right)$ implementation in the programming language `C` of exponential moving averages and various other time series operators for unevenly-spaced data.

8.3 Continuous-Time Analog

Calculating moving averages in *discrete* time requires a choice of the sampling scheme. In contrast, there is no such choice in *continuous* time, a fact that allows to succinctly illustrate certain concepts.

Definition 8.10 *Let* $\widetilde{X} : \mathbb{R} \to \mathbb{R}$ *be a measurable function on* $(\mathbb{R}, \mathcal{B})$. *For* $\tau > 0$ *we call the function* $\text{SMA}(\widetilde{X}, \tau) : \mathbb{R} \to \mathbb{R}$ *defined by*

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

the simple moving average (of $\widetilde{X}$*) of length* $\tau > 0$, *and* $\text{EMA}(\widetilde{X}, \tau) : \mathbb{R} \to \mathbb{R}$ *defined by*

$$\text{EMA}(\widetilde{X}, \tau)_t = \frac{1}{\tau}\int_0^\infty \widetilde{X}_{t-s} e^{-s/\tau} ds, \qquad t \in \mathbb{R}, \tag{8.27}$$

the exponential moving average (of $\widetilde{X}$*) of length* $\tau > 0$.

Unless stated otherwise, we apply a time series operator to real-valued functions using its natural analog. For example, $L(\widetilde{X}, \tau)$ for $\tau \in \mathbb{R}$ is a 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 trend measures and returns of a time series.

---

[17] The reasoning is similar to the proof of Proposition 8.8.

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

$$\frac{\widetilde{X}_t - \widetilde{X}_{t-\tau}}{\tau} = \mathrm{SMA}\left(\widetilde{X}', \tau\right)_t \tag{8.28}$$

and

$$\frac{1}{\tau} \log\left(\frac{\widetilde{X}_t}{\widetilde{X}_{t-\tau}}\right) = \mathrm{SMA}\left(\log\left(\widetilde{X}\right)', \tau\right)_t, \tag{8.29}$$

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

Proof. By the fundamental theorem of calculus

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

for $s \in \mathbb{R}$. Hence, (8.28) and (8.29) follow from the definition of the simple moving average. ■

For time series $X \in \mathcal{T}$ in discrete time,

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

and

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

for $t \in \mathbb{R}$ can be used as an approximation of (8.28) and (8.29), respectively. In some cases, the approximation even holds exactly. For example, if $t - \tau \in T(X)$, then $t = t_n$ and $t - \tau = t_{n-k}$ for some $1 \le k < n \le N(X)$, so that

$$\begin{aligned} \mathrm{SMA}_{\mathrm{next}}\left(\frac{\Delta X}{\Delta t(X)}, \tau\right)_t &= \frac{1}{\tau}\left(\Delta X_{t_n} + \Delta X_{t_{n-1}} + \ldots + \Delta X_{t_{n-k+1}}\right) \\ &= \frac{X_{t_n} - X_{t_n - \tau}}{\tau} \\ &= \frac{X_t - X[t-\tau]}{\tau}. \end{aligned} \tag{8.30}$$

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

$$\frac{\widetilde{X}_t - \mathrm{EMA}(\widetilde{X}, \tau)_t}{\tau} = \mathrm{EMA}(\widetilde{X}', \tau)_t \tag{8.31}$$

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 a bit tedious. Alternatively, partial integration gives

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

■

For time series $X \in \mathcal{T}$ in discrete time,

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

can be used as an approximation of the right-hand side of (8.31).

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

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

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

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

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

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,

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

The reflection principle implies that $\max_{0\leq s\leq t} B_s$ 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 therefore also 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 constant "density" of observation times.

Definition 9.2 *Assume 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* $\lim_{n\to\infty}\mathrm{mesh}(\pi_n) = 0$,

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

almost surely.

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

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

almost surely.

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

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

almost surely.

Note that in general, neither $\pi-$consistent nor $T-$consistent quadratic variation estimators can be used to obtain an unbiased estimate of $\langle\widetilde{X},\widetilde{X}\rangle$ from the observation time series X, since consistency is required only in the limit.

Lemma 9.3 *A time series operator O is $T-$consistent estimator of $\langle \widetilde{X}, \widetilde{X} \rangle$, if and only if*

$$O_\sigma : X \to \sqrt{O\left(X\right)} \max T\left(X\right)$$

is a $T-$consistent volatility estimator.

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

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

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 $p = 2$ the quadratic variation $[X, X]$ of X.*

Proposition 9.5 *The realized quadratic variation* $\mathrm{Var}(X, p = 2) = [X, X]$ *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. For $p = 2$,

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

for $t \in T(X)$. Lemma 9.1 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

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

due to the bounded mesh size of T. Kolmogorov's strong law of large numbers, see Chapter 3 in Shiryaev (1995), implies the second equality in the following set of equations:

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

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*

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

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

Proof. Lemma 9.1 (iii) shows that for $t > \rho$,

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

where we used that

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

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

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{\sigma}\sqrt{\frac{\pi}{8\rho}}\operatorname{ATR}\left(.,\rho,\tau=t_n-t_1\right) &= \frac{1}{\sigma}\sqrt{\frac{\pi}{8\rho}}E\left(\operatorname{range}\left(\widetilde{X},\rho\right)_\rho\right)\\
&= 1,
\end{aligned}$$

or in other words, the ATR is a $T-$consistent volatility estimator for scaled Brownian motion.
■

Note that the ATR is not $\pi-$consistent since it makes only limited use of the information contained in a time series.

Proposition 9.8 *The following time series operators are $T-$consistent quadratic variation estimator for scaled Brownian motion:*

Lemma 9.9 *(i)* $c_p(\rho)\operatorname{SMA}(\left|X-\operatorname{SMA}(X,\rho)\right|^2, \tau = \max T(X) - \min T(X))$ *for* $\rho > 0$ *and suitable constant* $c_p(\rho)$*,*

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

Proof. The argument is similar to the one for the realized quadratic variation and the ATR.
■

Clearly, there is a trade-off between 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 not a consistent volatility estimator in the presence of either jumps or measurement noise.

10 Conclusion

This paper presented methods for and analyzing and manipulating unevenly-spaced time series in their unaltered form, without a transformation to equally-spaced data. I aimed to make the developed methods 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.

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 $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$. 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*

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

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 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

$$\mathrm{SP}_\sigma \left(O\left(X\right)\right)\left(t\right) = \mathrm{SP}_\sigma \left(O\left(L\left(X, -t\right)\right)\right)\left(0\right) \tag{A.33}$$

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

$$\mathrm{SP}_\sigma \left(O\left(L\left(X, -t\right)\right)\right) = g_\sigma \left(\mathrm{SP}_\sigma \left(L\left(X, -t\right)\right)\right). \tag{A.34}$$

Combining (*A.33*) and (*A.34*) 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 in addition 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 $\sigma \in \{, \mathrm{lin}, \mathrm{next}\}$. There exists a finite signed measure μ_T on $(\mathbb{R}, \mathcal{B})$ such that*

$$\mathrm{SP}_\sigma \left(O\left(X\right)\right)\left(t\right) = \int_{-\infty}^{+\infty} \mathrm{SP}_\sigma \left(X\right)\left(t-s\right) d\mu_T\left(s\right) \tag{A.35}$$

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*.35) for $t=0$, or equivalently, that there exists a finite signed measure μ_T on $(\mathbb{R},\mathcal{B})$ such that

$$h_\sigma\left(\mathrm{SP}_\sigma\left(X\right)\right) = \int_{-\infty}^{+\infty} \mathrm{SP}_\sigma\left(X\right)(-s)\, d\mu_T\left(s\right) \tag{A.36}$$

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

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

and

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

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

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

Since 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 ($A.36$). To this end, note that the sample path of each time series $X \in \mathcal{T}$ can be written as a weighted sum of 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,

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

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. Hence, we have shown the desired result for last-point sampling, and the finals steps of the proof are completely analogous for next-point sampling.

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 an integral 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 $t \in T(X) = T(O(X))$. Since 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 $s \leq t$, 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
$*^\mu$	the convolution operator associated with a signed measure μ, see Definition 4.3
B	the backshift operator, see Definition 2.8
$\mathcal{B}$	the Borel $\sigma-$algebra on $\mathbb{R}$
$\mathcal{B}_+$	the Borel $\sigma-$algebra on $\mathbb{R}_+$
L	the lag operator, see Definition 2.8
D	the delay operator, see Definition 2.8
$N(X)$	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 $[0, \infty)$
$\overline{\mathbb{R}}_+$	$[0, \infty) \cup \{\infty\}$
$\mathcal{T}_n$	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
$\mathcal{T}(\tau)$	the space of time series in $\mathcal{T}$ with temporal length less than τ, see Definition 5.10
$\mathcal{T}^K$	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
$X[t]$	the most recent observation value of time series X at time t, see Definition 2.4
$X^c[T_X]$	the observation time series of a continuous-time stochastic process X^c, see Definition 2.7

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.

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.

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. (2011). 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.