Arbitrarily-Spaced Time Series Analysis
Preliminary and Incomplete Draft

Andreas Eckner

Current version: February 16, 2010

Abstract

This paper presents methods for analyzing arbitrarily-spaced time series data. The methods do not rely on a transformation to equally-spaced data. The analysis of equally-spaced time series data in this framework coincides with a traditional analysis along the lines of Hamilton (1994), and Brockwell and Davis (2003).

1 Introduction

Arbitrarily-spaced time series data naturally occur in many scientific domains. In economics, macroeconomic data, like the GDP, the unemployment rate, and LIBOR, are reported at vastly different frequencies - quarterly, monthly, and daily, respectively, in this case. In astronomy, radial velocities and spectra of celestial objects are observed at times determined by weather conditions and availability of a time slot at the telescope. In medical studies, a patient's state of health may only be observed at irregular time intervals, and different patients are usually observed at different points in time. There are many more examples in biology and signal processing.

There already exists an extensive body of literature on analyzing equally-spaced time series data, and many basic questions are now well understood. See, for example, Hamilton (1994), Brockwell and Davis (2003), Fan and Yao (2003), and Greene (2007). On the other hand, little literature exists specifically for arbitrarily-spaced time series data. A common approach is to transform such data into equally-spaced observations using some form of interpolation and then to apply existing methods for equally-spaced data. See, for example, Adorf (1995), and Beygelzimer, Erdogan, Ma, and Rish (2005). However, transforming the data in such a way has a couple of significant drawbacks:

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

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

See the Appendix for details. 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, like volatility and autocorrelations, may be subject to a significant and hard to quantify bias.*

Example 1.2 (Causality) *For a 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 value. Hence, while the data generating process may be adapted to a certain filtration $\mathcal{F}$,[1] the linearly interpolated process will in general not be adapted to $\mathcal{F}$ any more. This effect may change the causality relationships in a multivariate time series.*

Example 1.3 (Data Quality) *Converting an arbitrarily-spaced time series to an equally-spaced time series may (i) omit data points if consecutive observations lie close together, and (ii) introduce unnecessary additional data points if consecutive observations lie far part. Hence, the information content of the data set may be reduced, as well as diluted, thus causing statistical inference to be less efficient and potentially biased.*

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

While some authors suggested to use continuous diffusions processes, see Brockwell (2008), exists little theory and few techniques for analyzing arbitrarily-spaced time series data in its unaltered form. The aim of this paper is to provide methods for directly analyzing such data, while maintaining consistency with existing methods for equally-spaced time series observations. Throughout this paper, we mostly view the observed time series as a realization from a continuous-time stochastic process. Nevertheless, most results depend only on weak, if any, assumptions about the data generating process.

2 Basic Definitions

Intuitively, an arbitrarily-spaced time series is a sequence of observation time/value pairs (t_n, X_n), where the observation times are strictly increasing. This notion is made precise by the following

Definition 2.1 *For $n \geq 0$, we call*

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

(ii) $\mathbb{R}^n$ *the observation space,*

(iii) $\mathcal{T}_n = \mathbb{O}_n \times \mathbb{R}^n$ *the space of time series of length n,*

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

(iv) $\mathcal{T} = \cup_{n=0}^{\infty} \mathcal{T}_n$ *the space of (real-valued) time series.*

For notational convenience, we denote by $X^{\oslash}$ the empty 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) $T(X) = \{t_1, \ldots, t_{N(X)}\}$ *the sequence of observation times (of* X*),*

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

When confusion is impossible we will frequently use the less formal notation $((t_n, X_n) : 1 \leq n \leq N(X))$ or $(X_{t_n} : 1 \leq n \leq N(X))$ to denote a time series X with observation times $\{t_1, \ldots, t_{N(X)}\}$ and observations $(X_1, \ldots, X_{N(X)})$. Now that we have defined arbitrarily-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

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

In particular, the most recently available observation time before *the first observation is taken to be the first observation time. While not appropriate for some applications, this convention greatly simplifies the exposition below.*

In general $\mathrm{Prev}^X(t) < \mathrm{Next}^X(t)$ unless $t \in T(X)$ in which case t is both the most recent and next available observation time.

Definition 2.4 (Sampling) *We call* $X[t] = X_{Prev(X,t)}$ *the sampled value of* X *at time* t*. Furthermore, we call* $X[t]_{lin} = (1 - \omega^X(t)) X_{Prev(X,t)} + \omega^X(t) X_{Next^X(t)}$ *with*

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

the linearly interpolated value of X *at time* t*.*

In particular, the sampled value of a time series X *before* the first observation time is equal to the first observation value. While not appropriate for some applications (for example, the earth's temperature was not constant before its first measurement), this convention greatly simplifies some notation and avoids the treatment of a multitude of special cases in the exposition below.

Some basic properties of the sampling and interpolation operator are given by the following

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

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

(ii) $X[t]$ *as a function of* t *with* $t \in [\min(T(X)), \infty)$ *is a right-continuous piecewise-constant function with finite number of discontinuities.*

(iii) $X[t]_{lin}$ *as a function of* t *with* $t \in [\min(T(X)), \infty)$ *is a continuous piecewise-linear function.*

Lemma 2.5 suggests an alternative way of defining arbitrarily-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 arbitrarily-spaced time series as a discretely-observed diffusion process, thereby ruling out a large class of data-generating processes.

Definition 2.6 *For* $X \in \mathcal{T}$, *we call*

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

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

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

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

Note that for equally-spaced time series, the lag operator with $\tau = 1$ is identical to the backshift operator, but this is generally not the case for arbitrarily-spaced data.

Example 2.7 *Let* X *be the time series with observations* $(t_1, X_1) = (0, 1)$, $(t_2, X_2) = (2, -1)$, *and* $(t_3, X_3) = (5, 2.5)$. *Then*

$$X[t] = \begin{cases} NA, & \text{for } t < 0 \\ 1, & \text{for } 0 \leq t < 2 \\ -1, & \text{for } 2 \leq t < 5 \\ 2.5, & \text{for } t \geq 5 \end{cases}$$

and

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

and

$$(B(X))[s] = \begin{cases} NA, & \text{for } s < 2 \\ 1, & \text{for } 2 \leq s < 5 \\ -1, & \text{for } s \geq 5 \end{cases}$$

and

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

The following result elaborates the relationship between the lag operator (L) and the sampling operator ([]).

Lemma 2.8 *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) $L(X,\tau)[t] = X[t-\tau]$ *for* $t \in \mathbb{R}$.

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

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

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 (*iii*). ■

Definition 2.9 *For $X \in \mathcal{T}$ and $a > 0$, the (time-)scaling operator S_a is defined as*

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

Lemma 2.10 *For $X \in \mathcal{T}$ and $a > 0$,*

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

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

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

(iv) $S_a(X)[t] = X[at]$ *for* $t \in \mathbb{R}$.

Proof. The proof is very similar to the one of Lemma 2.8. ■

Since the unit of measurement and starting point of the time scale are usually not of interest, we will focus on (families of) time series operations that are invariant under the lag and time-scaling operator, as will be made precise in the next section.

3 Time Series Operators

This sections discusses operators that take a time series as input and leave as output a transformed time series. In most cases, the observation times of the transformed time series are identical to (or at least a subset of) the observation times of the original time series. We have already encountered a few such operators in the previous section, for example, the backshift, subperiod, and tick spacing operator.

The major difference between time series operators for equally and unequally-spaced observations is, that in the latter case, the observations 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 $O_T : \mathcal{T} \to \cup_{n \geq 0} \mathbb{O}_n$ is the transformation of ticks, $O_V : \mathcal{T} \to \cup_{n \geq 0} \mathbb{R}^n$ the transformation of observations, 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 observations and observation times are the same for the transformed time series. Using this notation, we in particular have $T(O(X)) = O_T(X)$ and $V(O(X)) = O_V(X)$ for any time series $X \in \mathcal{T}$ and time series operator O. This definition of a time series operator is completely general. In practice, most operators will have some special structure.

Definition 3.2 *A time series operator O is adapted (or causal), if for all $X \in \mathcal{T}$*

(i) $t \in T(O(X)) \Leftrightarrow t \in T(O(X\{-\infty, t\}))$ for $t \in \mathbb{R}$,

(ii) $O(X)_t = O(X\{-\infty, t\})_t$ for $t \in T(O(X))$.

The first condition says that each observation time t of the transformed series depends only on information available up to time t (about the original series). Similarly, the second condition says that each observation of the transformed series depends only on information available up to time t (about the original series). Note that Definition 3.2 (ii) implies the apparently stronger property $O(X\{-\infty, t\})_s = O(X)_s$ for $s \leq t$ and $s, t \in T(O(X))$ via

$$O(X\{-\infty, t\})_s = O(X\{-\infty, t\}\{-\infty, s\})_s = O(X\{-\infty, s\})_s = O(X)_s$$

since $\tilde{X} = X\{-\infty, t\}$ is just another time series to which Definition 3.2 (ii) can be applied.

Definition 3.3 *A time series operator O is shift-invariant, if for all $X \in \mathcal{T}$ and $\tau > 0$*

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

or equivalently, if

$$O_T(L(X, \tau)) = O_T(X) + \tau, \text{ and} \tag{3.1}$$

$$O_V(L(X, \tau)) = O_V(X). \tag{3.2}$$

Intuitively, 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 3.4 *A time series operator O is shift-invariant, if and only if, $O\left(X^{\varnothing}\right) = X^{\varnothing}$ and there exist functions $f_m : \mathbb{R}^{2m-1} \to \cup_{n\geq 0}\mathbb{O}_n$ and $g_m : \mathbb{R}^{2m-1} \to \cup_{n\geq 0}\mathbb{R}^n$ for $m > 0$, such that for all $X \in \mathcal{T}$ with $N(X) > 0$,*

$$O_T(X) = t_1 + f_{N(X)}\left((t_j - t_{j-1} : 2 \leq j \leq N(X)), (X_j : 1 \leq j \leq N(X))\right), \tag{3.3}$$

$$O_V(X) = g_{N(X)}\left((t_j - t_{j-1} : 2 \leq j \leq N(X)), (X_j : 1 \leq j \leq N(X))\right), \tag{3.4}$$

$$|O_T(X)| = |O_V(X)|. \tag{3.5}$$

Proof.

- $\Longleftarrow$ The case $m = 0$ is trivial, since $L(X^{\varnothing}, \tau) = X^{\varnothing}$. For $m \geq 1$, it is easy to see that (3.3) implies $O_T(L(X, \tau)) = O_T(X) + \tau$, while (3.4) implies $O_V(L(X, \tau)) = O_V(X)$. (3.5) simply states that the number of observations and observation times have to be the same.

- $\Longrightarrow$ The case $m = 0$ is trivial. For $m \geq 1$, if O is a time series operator without structure (3.3) – (3.5), then there exists a $k \geq 1$ and time series $X^{(1)}, X^{(2)} \in \mathcal{T}_k$ such that either (i) $O_T\left(X^{(1)}\right) \neq O_T\left(X^{(2)}\right) - \tau$ for $\tau = T\left(X^{(2)}\right)_1 - T\left(X^{(1)}\right)_1$, or (ii) $O_V\left(X^{(1)}\right) \neq O_V\left(X^{(2)}\right)$, but with

$$t_j^{(1)} - t_{j-1}^{(1)} = t_j^{(2)} - t_{j-1}^{(2)}, \qquad 2 \leq j \leq k \tag{3.6}$$

and

$$X_j^{(1)} = X_j^{(2)}, \qquad 1 \leq j \leq k. \tag{3.7}$$

However, (3.6) and (3.7) imply that $X^{(2)} = L(X^{(1)}, \tau)$. Hence, if O was shift-invariant, we would have

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

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 (3.3) – (3.5).

■

Frequently, a set of time series operator is naturally indexed by one or more parameters, which gives rise to a family of operators. For example, the family of moving average operators is indexed by the length $\tau > 0$ of the moving average.

Definition 3.5 *A family of time series operators $\{O_i : i \in I\}$ with index set $I = (0, \infty)$ is scale-invariant, if*

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

for all $X \in \mathcal{T}$, $a > 0$, and $i \in I$.

While scale invariance is a property tied to the observation time scale, the following definition is concerned with the scale of observation values.

Definition 3.6 *A time series operator O is homogenous of degree $d > 0$, if $O(aX) = a^d O(X)$ for all $X \in \mathcal{T}$ and $a > 0$. O is a linear operator if it is homogenous of degree $d = 1$.*

For example, the moving average operator is a linear time series operator, while the operator that calculates the (integrated) $p-$variation of a given time series is homogenous of degree $d = p$.

Lemma 3.7 *A time series operator O is homogenous of degree $d > 0$, if and only if, $O\left(X^{\varnothing}\right) = X^{\varnothing}$ and there exist functions $f_m : \mathbb{R}^m \times [-1,1]^m \to \cup_{n\geq 0}\mathbb{O}_n$ and $g_m : \mathbb{R}^m \times [-1,1]^m \to \cup_{n\geq 0}\mathbb{R}^n$ for $m > 0$, such that for all $X \in \mathcal{T}$ with $N(X) > 0$,*

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

where

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

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

Proof. The proof is very similar to the one of Lemma 3.4. ■

4 Convolution Operators

Convolution operators are a class of adapted, shift-invariant (and often homogenous) time series operators, that are particularly tractable. Recall that a finite[3] signed measure on $\left(\mathbb{R}^k, \mathcal{B}_k\right)$ for $k \geq 1$ is a real-valued measurable function μ that satisfies (i) $\mu(\varnothing) = 0$, (ii) there exists a constant $c \in \mathbb{R}$ such that $|\mu(B)| < c$ for all $B \in \mathcal{B}$, and (iii) if $E = +_i E_i$ is a countable disjoint union of sets in $\mathcal{B}_k$, then $\mu(E) = \sum_i \mu(E_i)$.

Theorem 4.1 (Jordan Decomposition) *Let μ be a finite signed measure. There are mutually singular finite measures μ_+ and μ_- so that $\mu = \mu_+ - \mu_-$, and this decomposition is unique.*

See Durrett (2005), Appendix A.8 for details.

Corollary 4.2 *If μ is a finite signed measure on $\left(\mathbb{R}^k, \mathcal{B}_k\right)$ and absolutely continuous with respect to the Lebesgue measure, then there exists a function (density) $f : \mathbb{R}^k \to \mathbb{R}$ such that*

$$\mu(E) = \int_E f(x)\, dx$$

for all $E \in \mathcal{B}_k$.

This corollary is a simple consequence of the Jordan decomposition and the Radon Nikodym theorem, see Durrett (2005), Appendix A.8 for details.

[3]For our purposes, we need to focus on finite measures only, even though the theory of signed measures is more general.

Definition 4.3 *A (univariate time series) kernel* μ *is a finite signed measure of* $(\mathbb{R} \times \mathbb{R}_+, \mathcal{B} \otimes \mathcal{B}_+)$.[4]

Definition 4.4 (Convolution Operator) *For a time series* $X \in \mathcal{T}$ *and kernel* μ*, the convolution* $*^{\mu}(X) = X * \mu$ *is given by*

$$T(X * \mu) = T(X), \tag{4.8}$$

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

where the integration is over the variable s*. If* μ *is absolutely continuous with respect to the Lebesgue measure, then (4.9) can be written as*

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

where f *is the density function of* μ*.*

Lemma 4.5 *If* μ *is a kernel, then the associated convolution operator* $*^{\mu}$ *is shift-invariant.*

Proof. Let $X \in \mathcal{T}$ and $\tau > 0$. Using (4.8) twice, we get

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

showing that $*^{\mu}$ satisfies (3.1). On the other hand, for $t \in T(X) + \tau$,

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

where we used Lemma 2.8 (*iv*) for the second equality. Hence, we also have $(*^{\mu})_V(X) = (*^{\mu})_V(L(X, \tau))$. ■

Definition 4.6 *A convolution operator* $*^{\mu}$*, with associated with a kernel* μ*, is* affine *if*

$$\mu(x, t) \equiv (a + bx)\, \mu_T(t) \tag{4.10}$$

for some constants $a, b \in \mathbb{R}$*, and finite signed measure* μ_T *on* $(\mathbb{R}_+, \mathcal{B}_+)$*.*

Lemma 4.7 *If* μ *is an affine kernel, i.e. of the form (4.10), with* $a = 0$*, then* $*^{\mu}$ *is a linear time series operator in the sense that*

$$(cX + Y) * \mu \equiv c(X * \mu) + Y * \mu$$

for all $X, Y \in \mathcal{T}$ *and* $c \in \mathbb{R}$*. Therefore, such a* μ *will be called a linear kernel.*

[4] $\mathcal{B} \otimes \mathcal{B}_+$ is the Borel $\sigma-$algebra on $\mathbb{R} \times \mathbb{R}_+$, i.e. the $\sigma-$algebra generated by the intervals $[a, b) \times [c, d)$ with $-\infty < a < b < \infty$ and $0 \le c < d < \infty$. See Billingsley (1995), or Durrett (2005).

Proof. First note that

$$\begin{aligned} T\left(\left(cX+Y\right)*\mu\right) &= T\left(cX+Y\right) \\ &= T(cX)\cup T\left(Y\right) \\ &= T\left(c\left(X*\mu\right)\right)\cup T\left(Y*\mu\right), \end{aligned}$$

since a convolution operator leaves all observation times unchanged. Furthermore, for $t \in T\left(\left(cX+Y\right)*\mu\right)$

$$\begin{aligned} \left(\left(cX+Y\right)*\mu\right)_t &= \int_{\mathbb{R}} d\mu\left(\left(cX+Y\right)[t-s],s\right) \\ &= \int_{\mathbb{R}} d\mu\left(cX[t-s]+Y[t-s],s\right) \\ &= \int_{\mathbb{R}} b\left(cX[t-s]+Y[t-s]\right)d\mu_T\left(s\right) \\ &= c\int_{\mathbb{R}} bX[t-s]d\mu_T\left(s\right)+\int_{\mathbb{R}} bY[t-s]d\mu_T\left(s\right) \\ &= c\left(X*\mu\right)_t+\left(Y*\mu\right)_t. \end{aligned}$$

■

Definition 4.8 *For a kernel μ and $p>0$, we call*

$$m\left(\mu,p\right)=\int_{\mathbb{R}}\int_0^\infty |x|^p\,d\mu\left(x,t\right)$$

the $p-$th moment of $*^\mu$. This quantity will prove useful for normalizing various kernels, such as those associated with moving average and return calculations.

For some applications, it is useful to define the convolution (4.9) using linear interpolation.

Definition 4.9 *For a time series $X\in\mathcal{T}$ and kernel μ, the linearly-interpolated convolution $*^\mu_{\text{lin}}\left(X\right)=X*_{\text{lin}}\mu$ is given by*

$$\begin{aligned} &T\left(X*_{\text{lin}}\mu\right)=T\left(X\right), \\ &\left(X*_{\text{lin}}\mu\right)_t=\int_{\mathbb{R}} d\mu\left(X[t-s]_{\text{lin}},s\right), \qquad t\in T\left(X*_{\text{lin}}\mu\right). \end{aligned}$$

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

Theorem 4.10 *Let Y be a Lévy process with $(\mathcal{F}_t : t\geq 0)$ the natural filtration. For a fixed set $T_n\in\mathbb{O}_n$ of observation times with $\min T_n\geq 0$, let $X=\left(T_n,\left(Y_{t_k}:1\leq k\leq n\right)\right)$ denote the corresponding observation time series. Let μ be a linear kernel μ with $\mu_T\left(s\right)=0$ for $s\geq C$ for some constant $C>0$. Then*

$$E\left(\int_0^t d\mu\left(Y_{t-s},s\right)\mid X\right)=\left(X*_{lin}\mu\right)_t \tag{4.11}$$

for all $t\geq C$.

Proof. A Lévy process is the sum of three independent stochastic processes; a Brownian motion with drift, a compound Poisson process, and a square integrable pure jump martingale with almost surely finite number of jumps on finite intervals. See Protter (2005), Theorem 42 and Sato (1999). Equation (4.11) is easily shown for each of these three components. ■

5 Examples

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

5.1 Arithmetic Operators

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",

(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".

Definition 5.2 (Arithmetic Time Series Operations) *For time series $X, Y \in \mathcal{T}$, we call*

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

(ii) XY with $T(XY) = T(X) \cup T(Y)$ and $(XT)_t = X[t]\,Y[t]$ for $t \in T(X+Y)$ "the product of X and Y".

If desired, these two definitions can merged by interpreting a constant $c \in \mathbb{R}$ as a time series with a single observation c in the infinite past.

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

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

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

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

The following example examines under what circumstances linear interpolation is appropriate for arithmetic time series operators, as opposed to "last-observation sampling".

Example 5.4 *Let X , Y be two independent martingales. For fixed observation times T_X and T_Y, let $\widehat{X}$ and $\widehat{Y}$ denote the corresponding observation time series of X and Y, respectively. Furthermore, let $\mathcal{F}_t = \sigma(\widehat{X}_s : s \le t)$ and $\mathcal{G}_t = \sigma(\widehat{Y}_s : s \le t)$ denote the filtration generated by X and Y, respectively, and $\mathcal{H}_t = \mathcal{F}_t \cup \mathcal{G}_t$. For $t \in \mathbb{R}$,*

(i) $E\left((X+Y)_t \,|\, \mathcal{H}_t\right) = \widehat{X}[t] + \widehat{Y}[t]$

(ii) $E\left((X+Y)_t \,|\, \mathcal{H}_\infty\right) = \widehat{X}[t]_{lin} + \widehat{Y}[t]_{lin}$.

In particular, for $t \in T(\widehat{X}) \cup T(\widehat{Y})$

$$E\left((X+Y)_t \,|\, \mathcal{H}_t\right) = \left\{ \begin{array}{ll} \widehat{X}_t + \widehat{Y}[t] & \textit{for } t \in T(\widehat{X}) \\ \widehat{X}[t] + \widehat{Y}_t & \textit{for } t \in T(\widehat{Y}) \end{array} \right\}.$$

Hence, for applications that are allowed to depend on past data only, it makes sense to define the sum $\widehat{X} + \widehat{Y}$ of two independent zero-drift time series $\widehat{X}$ and $\widehat{Y}$ as

$$\begin{aligned} T(\widehat{X}+Y) &= T(\widehat{X}) \cup T(\widehat{Y}), \\ (\widehat{X}+\widehat{Y})_t &= \widehat{X}[t] + \widehat{Y}[t], \qquad t \in T(\widehat{X}+\widehat{Y}). \end{aligned} \tag{5.12}$$

In the presence of a drift or for weakly correlated time series, (5.12) can still be used as an approximation.

The following example examines the approximative nature of "last-observation sampling" for arithmetic time series operators.

Example 5.5 *Consider the same setup as in Example 5.4 except that X and Y are martingales of the form*

$$\begin{aligned} X &= \sqrt{1-\rho^2}\,\widetilde{X} + \rho Z, \\ Y &= \sqrt{1-\rho^2}\,\widetilde{Y} + |\rho|\, Z, \end{aligned}$$

where $\widetilde{X}$, $\widetilde{Y}$, and Z are independent martingales, and $\rho \in [-1,1]$ is a correlation parameter. Since $\mathcal{H}_t = \mathcal{H}_{\max(Prev^{\widehat{X}}(t), Prev^{\widehat{Y}}(t))}$ and using the martingale property

$$\begin{aligned} E\left((X+Y)_t \,|\, \mathcal{H}_t\right) &= E\left((X+Y)_t \,|\, \mathcal{H}_{\max(Prev^{\widehat{X}}(t), Prev^{\widehat{Y}}(t))}\right) \\ &= E\left((X+Y)_{\max(Prev^{\widehat{X}}(t), Prev^{\widehat{Y}}(t))} \,|\, \mathcal{H}_{\max(Prev^{\widehat{X}}(t), Prev^{\widehat{Y}}(t))}\right), \end{aligned}$$

so that without loss of generality we need only consider the case $t \in T(\widehat{X}) \cup T(\widehat{Y})$. By symmetry, we only need to consider the case $t \in T(\widehat{X})$ for which $E\left(X_t \,|\, \mathcal{H}_t\right) = X_t$ and

$$\begin{aligned} E\left(Y_t \,|\, \mathcal{H}_t\right) &= Y_{Prev^{\widehat{Y}}(t)} + E\left(Y_t - Y_{Prev^{\widehat{Y}}(t)} \,|\, \mathcal{H}_t\right) \\ &= \widehat{Y}[t] + E\left(E\left(Y_t - Y_{Prev^{\widehat{Y}}(t)} \,|\, \mathcal{H}_t \cup X_{Prev^{\widehat{Y}}(t)}\right) \,|\, \mathcal{H}_t\right) \\ &= \widehat{Y}[t] + E\left(\operatorname{sign}(\rho)\,\rho^2 \left(X_t - X_{Prev^{\widehat{Y}}(t)}\right) \,|\, \mathcal{H}_t\right) \\ &= \widehat{Y}[t] + \operatorname{sign}(\rho)\,\rho^2 \left(X_t - E\left(X_{Prev^{\widehat{Y}}(t)} \,|\, \mathcal{H}_t\right)\right). \end{aligned}$$

Hence, the conditional expectation $E\left((X+Y)_t \,|\, \mathcal{H}_t\right)$ is of the same form as in Example 5.4, plus the product of (i) the correlation parameter ρ^2, and (ii) the $\mathcal{H}_t-$conditional expected change of X (or Y) since the most recent observation time of $\widehat{Y}$ (or $\widehat{X}$).

To summarize, if $\widehat{X}$ and $\widehat{Y}$ are time series that are (i) generated by stochastic processes X and Y that independent (or at least weakly correlated), and (ii) if X and Y are martingales (or at least the spacing of observation times for $\widehat{X}$ and $\widehat{Y}$ is small compared to the absolute value of the drift of X and Y), then $\widehat{X} + \widehat{Y}$ is equal to (or at least a good approximation of) the time series of conditional expectation $E\left((X+Y)_t \,|\, \mathcal{H}_t\right)$ with observation times equal to the union of observation times of $\widehat{X}$ and $\widehat{Y}$.

(Next: $E\left((X+Y)_t \,|\, \mathcal{H}_\infty\right)$?)

5.2 Return Calculations

This section examines calculating time series return over various return horizons and on various scales, and starts by defining a few auxiliary operators.

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

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

(ii) $\operatorname{diff}(X, Y)$ *with* $\sigma \in \{$*"abs", "rel", "log"*$\}$ *the absolute/relative/log difference between X and Y, where*

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

and provided that $\min V(X) > 0$ *and* $\min V(Y) > 0$ *for* $\sigma \in \{$*"rel", "log"*$\}$.

Definition 5.7 (Returns) *For a time series $X \in \mathcal{T}$ and time horizon τ, we call*

(i) $\operatorname{diff}_\sigma(X, L(X, \tau))$ *the rolling absolute/relative/log return (of X over past τ),*

(ii) $\operatorname{diff}_\sigma(X, B(X))$ *the absolute/relative/log observation return (of X),*

(iii) $\operatorname{diff}_\sigma(X, \mathrm{MA}^{II}(X, \tau))$ *the moving average of absolute/relative/log returns (of X over past τ),*

(iv) $\operatorname{diff}_\sigma(X, \mathrm{EWMA}^{II}(X, \tau))$ *the exponentially weighted moving average of absolute/relative/log returns (of X over past τ),*

provided that X is strictly positive for $\sigma \in \{$*"rel", "log"*$\}$.

The first two definitions are immediately clear, while Section 7 provides a theoretical motivation for the last two definitions.

Lemma 5.8 *The return operators* 1, 3 *and* 4 *in Definition 5.7 are either convolution operators or simple transformations of convolution operators.*

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

$$\text{diff}_\sigma(X, L(X,\tau)) = \begin{cases} X * \mu, & \text{if } \sigma = \text{“abs”} \\ \exp(X * \mu) - 1, & \text{if } \sigma = \text{“rel”} \\ X * \mu, & \text{if } \sigma = \text{“log”} \end{cases}$$

with

$$\mu(x,t) = \begin{cases} x(\delta_0(t) - \delta_\tau(t)), & \text{if } \sigma = \text{“abs”} \\ \log(x)(\delta_0(t) - \delta_\tau(t)), & \text{if } \sigma \in \{\text{“rel”}, \text{“log”}\} \end{cases}$$

provided that X is strictly positive for $\sigma \in \{\text{“rel”}, \text{“log”}\}$. The proof for the other two return operators is similar. ■

Definition 5.9 (Rolling Time Series Functions) *Assume given a time series $X \in \mathcal{T}$, a time horizon $\tau > 0$, and a function $f : \mathcal{T}(\tau) \to \mathbb{R}$, where $\mathcal{T}(\tau) = \{X \in \mathcal{T}\colon \max(T(X)) - \min(T(X)) < \tau\}$ denotes the space of time series with observation length less than τ. The “rolling function f of X over horizon τ”, denoted by $\text{roll}(X, f, \tau)$, is the time series with*

$$T(\text{roll}(X,f,\tau)) = T(X),$$
$$\text{roll}(X,f,\tau)_t = f(X\{t-\tau, t\}), \qquad t \in T(\text{roll}(X,f,\tau)).$$

Rolling time series functions cover a wide range of adapted time series operators, for example, convolution operators where the kernel has a bounded support on the time dimension. Many operators that cannot be expressed as convolution operators are included as well.

Example 5.10 *For a time series $X \in \mathcal{T}$, horizon $\tau > 0$, the rolling time series $\text{roll}(X, f, \tau)$ with*

(i) $f(X) = |V(X)|$ *is the rolling number of observations,*

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

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

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

(v) $f(X) = \max V(X) - \min V(X)$ *is the rolling range,* $\text{range}(X, \tau)$,

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

of X over the horizon τ.

6 Multivariate Time Series Operators

Since multivariate operators are a natural extension of the univariate case, this section states definitions without prior motivation and results without proof.

Definition 6.1 *For $K > 1$, a $K-$dimensional arbitrarily-spaced time series X^K is a $K-$tuple of univariate arbitrarily-spaced time series X_k^K with $1 \le k \le K$. $\mathcal{T}^K$ is the space of (real-valued) $K-$dimensional time series.*

Definition 6.2 *For a multivariate time series $X^K \in \mathcal{T}^K$ and time vector $t^K \in \mathbb{R}^K$,*

$$X^K[t^K] = \left(X_k^K[t_k^K] : 1 \le k \le K\right)$$

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

Unless stated otherwise, all time series operators encountered in Section 2-4 are applied element-wise to a multivariate time series X^K, which greatly simplifies some notation. For example, we interpret $X^K[t]$ for $t \in \mathbb{R}$ as $\left(X_k^K[t] : 1 \le k \le K\right)$, the samples values of X^K at time t. Of course, whenever there is a risk of confusion, we must go back to a more precise notation.

Definition 6.3 *A $K-$dimensional time series operator is a mapping $O : \mathcal{T}^K \to \mathcal{T}^M$ with $M \ge 1$, or equivalently, an $M-$tuple of mappings $O^m : \mathcal{T}^K \to \mathcal{T}$ with $1 \le m \le M$. Each mapping O^m in turn is a pair of mappings (O_T^m, O_V^m), where $O_T^m : \mathcal{T}^K \to \cup_{n \ge 0} \mathbb{O}_n$ is the transformation of ticks, and $O_V^m : \mathcal{T}^K \to \cup_{n \ge 0} \mathbb{R}^n$ the transformation of observations, subject to the constraint $|O_T^m(X)| = |O_V^m(X)|$ for all $X^K \in \mathcal{T}^K$.*

Definition 6.4 *For $K \ge 1$, a finite signed measure on $(\mathbb{R}^K \times (\mathbb{R}_+)^K, \mathcal{B}^K)$ is a real-valued measurable function μ that satisfies (i) $\mu(\varnothing) = 0$, (ii) there exists a constant $c \in \mathbb{R}$ such that $|\mu(B)| < c$ for all $B \in \mathcal{B}^K$, and (iii) if $E = +_i E_i$ is a countable disjoint union of sets in $\mathcal{B}^K$, then $\mu(E) = \sum_i \mu(E_i)$.*

Definition 6.5 *For a multivariate time series $X^K \in \mathcal{T}^K$ and finite signed measure μ on $(\mathbb{R}^K \times (\mathbb{R}_+)^K, \mathcal{B}^K)$, the $K-$dimensional convolution $*^\mu\left(X^K\right) = X^K * \mu$ is given by*

$$T\left(X^K * \mu\right) = \cup_{k=1}^K T\left(X_k^K\right), \tag{6.13}$$

$$\left(X^K * \mu\right)_t = \int_{(\mathbb{R}_+)^K} d\mu\left(X^K[t - s^K], s^K\right), \qquad t \in T\left(X^K * \mu\right), \tag{6.14}$$

where $t - s^K$ is the vector $(t - s_k : 1 \le k \le K)$. If μ is absolutely continuous with density f, then we equivalently write (6.14) as

$$\left(X^K * f\right)_t = \int_{(\mathbb{R}_+)^K} f\left(X^K[t - s^K], s^K\right) ds, \qquad t \in T\left(X^K * \mu\right).$$

A $K-$dimensional convolution operator is a mapping $\mathcal{T}^K \to \mathcal{T}$. More generally,

Definition 6.6 *A (K, M)-dimensional convolution operator is an $M-$tuple of $K-$dimensional convolution operators $(*^{\mu_1}, \ldots, *^{\mu_M})$, and therefore a mapping $\mathcal{T}^K \to \mathcal{T}^M$.*

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 for $1 \le k \le K$ depends on all input time series, while in the later case it only depends 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 the input time series, see (6.13).

6.1 Examples

This section gives a couple of examples of multivariate convolution operators.

Lemma 6.7 *The arithmetic operators $X + Y$ and XY in Definition 5.2 are multivariate convolution operators with kernel $\mu(x, y, t_x, t_y) = \delta_0(t_x)\delta_0(t_y)(x + y)$ and $\mu(x, y, t_x, t_y) = \delta_0(t_x)\delta_0(t_y)\, xy$, respectively, for the time series vector $(X, Y) \in \mathcal{T}^2$.*

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

$$
\begin{aligned}
((X, Y) * \mu)_t &= \int_0^\infty \int_0^\infty \delta_0(s_1)\delta_0(s_2)(X[t - s_1] + Y[t - s_2])ds_1 ds_2 \\
&= X[t] + Y[t],
\end{aligned}
$$

and therefore also $V((X, Y) * \mu) = V(X + Y)$. The reasoning for the other kernel is virtually identical. ■

Definition 6.8 *For a multivariate time series $X^K \in \mathcal{T}^K$, we call*

(i) $\operatorname{avg}_c(X^K) = X^K * \mu$ *with* $\mu(x^K, t^K) = \delta_0(t^K)\operatorname{avg}(x^K)$ *the cross-sectional average of* X^K,

(ii) $\min_c(X^K) = X^K * \mu$ *with* $\mu(x^K, t^K) = \delta_0(t^K)\min(x^K)$ *the cross-sectional minimum of* X^K,

(iii) $\max_c(X^K) = X^K * \mu$ *with* $\mu(x^K, t^K) = \delta_0(t^K)\max(x^K)$ *the cross-sectional maximum of* X^K,

(iv) $\operatorname{range}_c(X^K) = X^K * \mu$ *with* $\mu(x^K, t^K) = \delta_0(t^K)(\max(x^K) - \min(x^K))$ *the cross-sectional range of* X^K,

(v) $\operatorname{quantile}_c(X^K, q) = X^K * \mu$ *with* $\mu(x^K, t^K) = \delta_0(t^K)\operatorname{quantile}(x^K, q)$ *the cross-sectional* $q-$*quantile of* X^K.

Contemporaneous time series operators, such as the ones in Definition 6.8, useful in many applications for summarizing high-dimensional time series. For example, a common question among economic policy makers is how the distribution of household income is changing over time. If X^K denotes the time series of individual incomes, then *quantile*$_c(X^K, 0.8)$ / *quantile*$_c(X^K, 0.2)$ is the time series of the income ratio of the highest and lowest quintile. In biology, the contemporaneous dispersion of body heights of a species as a function of age might be interest, which can be expressed in a similar manner.

7 (Exponentially Weighted) Moving Averages

Moving averages, with exponentially declining weights or otherwise, are useful for summarizing the past behavior of a time series over a certain horizon. For equally-spaced time series data, there is only one way of calculating moving averages (MAs) and exponentially weighted moving averages (EWMAs), and the properties of such linear filters are well understood. For unequally-spaced data, however, there exist a couple of alternatives, all of which may be sensible depending on the data generating process of the given time series and the desired application.

7.1 Moving Averages

Definition 7.1 *For a time series $X \in \mathcal{T}$, we define three versions of the moving average (MA) of length $\tau > 0$. For $t \in T(X)$,*

(i) $\mathrm{MA}^{\mathrm{I}}(X,\tau)_t = \mathrm{avg}\{X_s : s \in [t, t-\tau)\}$,

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

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

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

MA^{I} is most appropriate for analyzing truly discrete *events*, for example, for calculating the average number of casualties per deadly car accident over the past twelve months. Or the average number of IBM common shares traded on the NYSE per executed order during the past 30 minutes. $\mathrm{MA}^{\mathrm{II}}$ seems most appropriate for analyzing truly discrete observation *values*, for example, for calculating the average FED funds target rate[5] over the past three years. In this case, it is desirable to weigh historical target rates by the amount of time such target levels remained unchanged. Finally, $\mathrm{MA}^{\mathrm{III}}$ is most appropriate for estimating the rolling average value of a continuous time stochastic processes with observation times that are independent of the observation values, see Theorem 4.10 for a precise statement.

In most cases, the value of the moving averages $\mathrm{MA}^{\mathrm{II}}$ and $\mathrm{MA}^{\mathrm{III}}$ will be very similar and the choice of the moving average length τ will generally have a much larger influence on the outcome of a time series analysis. If not, then the problem of interest is probably poorly formulated.

Lemma 7.2 *For a time series $X \in \mathcal{T}$, $t \in T(X)$, and moving average length $\tau > 0$:*

$$\left|\mathrm{MA}^{\mathrm{II}}(X,\tau) - \frac{\mathrm{MA}^{\mathrm{I}}(\Delta t(X)\,B(X),\tau)}{\mathrm{MA}^{\mathrm{II}}(\Delta t(X),\tau)}\right| \le \frac{1}{\tau}\max(|V(X)|)\left|Prev^X(t-\tau)-(t-\tau)\right| \tag{7.15}$$

$$\le \frac{1}{\tau}\max|V(X)|\max(\Delta t(X)).$$

[Question: should $\mathrm{MA}^{\mathrm{II}}(\Delta t(X),\tau)$ be replaced by $\mathrm{MA}^{\mathrm{I}}(\Delta t(X),\tau)$?]

Proof. For $t \in T(X)$, $t = t_n$ and $\mathrm{Prev}^X(t-\tau) = t_{n-k}$ for some $n, k \in \mathbb{N}$. Using this notation

$$\mathrm{MA}^{\mathrm{I}}(B(X)\,\Delta t(X),\tau) = \frac{X_{t_{n-1}}\Delta t(X)_{t_n} + \ldots + X_{t_{n-k}}\Delta t(X)_{t_{n-k+1}}}{n-k},$$

$$\mathrm{MA}^{\mathrm{II}}(\Delta t(X),\tau) = \frac{\tau}{n-k},$$

[5] 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 http://www.federalreserve.gov/fomc/fundsrate.htm for details.

and

$$\begin{aligned}\mathrm{MA}^{\mathrm{II}}(X) &= \frac{1}{\tau}\int_{t-\tau}^{t} X[s]\,ds \\ &= \frac{X_{t_{n-1}}\Delta t(X)_{t_n} + \ldots + X_{t_{n-k+1}}\Delta t(X)_{t_{n-k+2}} + X_{t_{n-k}}(t_{n-k+1}-(t-\tau))}{\tau} \\ &= \frac{X_{t_{n-1}}\Delta t(X)_{t_n} + \ldots + X_{t_{n-k}}\Delta t(X)_{t_{n-k+1}}}{\tau} \\ &\quad - \frac{1}{\tau}X_{t_{n-k}}\left(\mathrm{Prev}^X(t-\tau)-(t-\tau)\right).\end{aligned}$$

Hence

$$\begin{aligned}& MA^{II}(X,\tau) - \frac{MA^{I}(\Delta t(X)\,B(X),\tau)}{MA^{II}(\Delta t(X),\tau)} \\ &= \frac{1}{\tau}X_{t_{n-k}}\left(\mathrm{Prev}^X(t-\tau)-(t-\tau)\right)\end{aligned}$$

from which (7.15) easily follows. ■

Hence, the value of the moving average $\mathrm{MA}^{\mathrm{II}}(X)$ at a time $t \in T(X)$ is roughly equal to the weighted average observation of X in the time interval $(t, t-\tau]$ with weights equal to $\omega_k = \Delta t(X)_{t_{k+1}}/\tau$ for $t_k \in (t, t-\tau] \cap T(X)$, with equality if $t-\tau \in T(X)$. In particular, for an equally-spaced time series with time scale chosen such that $\Delta t(X) \equiv 1$, the moving averages $MA^{I}(B(X),\tau)$ and $MA^{II}(X,\tau)$ coincide since necessarily $\tau \in \mathbb{N}$.

Lemma 7.3 *Let X be a strictly increasing time series. Then*

$$\mathrm{MA}^{\mathrm{III}}(X,\tau)_t > \mathrm{MA}^{\mathrm{II}}(X,\tau)_t \tag{7.16}$$

for $t \in T(X)$. If in addition X is equally spaced with observation time spacings Δt, and if $\tau = K\Delta t$ for some constant $K \in \mathbb{N}$, then also

$$\mathrm{MA}^{\mathrm{I}}(X,\tau)_t > \mathrm{MA}^{\mathrm{III}}(X,\tau)_t \tag{7.17}$$

for $t \in T(X)$.

Proof. Inequality (7.16) follows directly from the fact that $X[t]_{\mathrm{lin}} > X[t]$ for strictly increasing time series. Inequality (7.17) follows from that fact that for an equally-spaced time series, the moving average $\mathrm{MA}^{\mathrm{II}}$ of X is identical to the moving average MA^{I} applied to the backshifted time series $B(X)$. ■

7.2 Exponentially Weighted Moving Averages

Definition 7.4 *For a time series $X \in \mathcal{T}$, we define three versions of the exponentially weighted moving average (EWMA) of length $\tau > 0$. For $t_n \in T(X)$,*

(i) $\mathrm{EWMA}^{\mathrm{I}}(X,\tau)_{t_n} = \left(1-e^{-\Delta t_n/\tau}\right)X_{t_n} + e^{-\Delta t_n/\tau}\mathrm{EWMA}^{\mathrm{I}}(X,\tau)_{t_{n-1}}$ *for $n \geq 2$, with* $\mathrm{EWMA}^{\mathrm{I}}(X,\tau)_{t_1} = X_{t_1}$ *for $n = 1$,*

(ii) $\mathrm{EWMA}^{\mathrm{II}}(X,\tau)_{t_n} = \frac{1}{\tau}\int_0^\infty X[t_n-s]\,e^{-s/\tau}ds,$

(iii) $\text{EWMA}^{\text{III}}(X,\tau)_{t_n} = \frac{1}{\tau}\int_0^\infty X[t_n - s]_{\text{lin}}\, e^{-s/\tau} ds,$

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

In particular, the first time series value of all three EWMAs is equal to the first observation value of the input time series.

Lemma 7.5 *For time series $X \in \mathcal{T}$,*

$$\text{EWMA}^{\text{II}}(X,\tau)_t = \text{EWMA}^{\text{I}}(B(X),\tau)_t$$

for $t \in T(X) \setminus \{t_1\}$. In other words, the EWMA^{II} *of a time series is identical to the* EWMA^{I} *of the back-shifted time series (apart from the first observation).*

Proof. For $n = 2$, we have

$$EWMA^{II}(X,\tau)_{t_2} = X_{t_1} = B(X)_{t_2} = EWMA^{I}(B(X),\tau)_{t_2},$$

since X_{t_2} is the first observation value of the back-shifted time series $B(X)$. For $n > 2$, by induction

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

■

The following result shows that the EWMA^{III} can be calculated recursively. See also Müller (1991) for calculating EWMAs using various interpolation methods.

Lemma 7.6 *Definition 7.4 of the* EWMA^{III} *is equivalent to*

$$\begin{aligned}
\text{EWMA}^{\text{III}}(X,\tau)_{t_n} = {} & e^{-\Delta t_n/\tau}\text{EWMA}^{\text{III}}(X,\tau)_{t_{n-1}} + X_{t_n}(1 - \omega(\tau,\Delta t_n)) \\
& + X_{t_{n-1}}(\omega(\tau,\Delta t_n) - e^{-\Delta t_n/\tau}),
\end{aligned}$$

for $t_n \in T(X)$ with $n \geq 2$ and initial value $\text{EWMA}^{\text{III}}(X,\tau)_{t_1} = X_{t_1}$*, where*

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

In particular, $\omega(\tau,\Delta t_n) \approx 1$ for $\Delta t_n \ll \tau$ and therefore $\text{EWMA}^{\text{III}}(X,\tau)_{t_n} \approx \text{EWMA}^{\text{III}}(X,\tau)_{t_{n-1}}$*, while $\omega(\tau,\Delta t_n) \approx 0$ for $\Delta t_n \gg \tau$ and therefore* $\text{EWMA}^{\text{III}}(X,\tau)_{t_n} \approx X_{t_n}$*.*

Proof. For $X \in T(X)$ and $t_n \in T(X)$ with $n \geq 2$,

$$\begin{aligned} \mathrm{EWMA}^{\mathrm{III}}(X,\tau)_{t_n} &= \frac{1}{\tau}\int_0^{\infty} X[t_n - s]_{\mathrm{lin}}\, e^{-s/\tau} ds \\ &= \frac{1}{\tau}\int_0^{\Delta t_n} X[t_n - s]_{\mathrm{lin}}\, e^{-s/\tau} ds + \frac{1}{\tau}\int_{\Delta t_n}^{\infty} X[t_n - s]_{\mathrm{lin}}\, e^{-s/\tau} ds \\ &= \frac{1}{\tau}\int_0^{\Delta t_n} \alpha_n^X(s)\, e^{-s/\tau} ds + e^{-\Delta t_n/\tau}\mathrm{EWMA}^{\mathrm{III}}(X,\tau)_{t_{n-1}}, \end{aligned} \tag{7.18}$$

where

$$\alpha_n^X(s) = X_{t_n} - \frac{s}{\Delta t_n}\left(X_{t_n} - X_{t_{n-1}}\right), \qquad 0 \leq s \leq \Delta t_n.$$

Integration gives

$$\begin{aligned} \frac{1}{\tau}\int_0^{\Delta t_n} \alpha_n^X(s)\, e^{-s/\tau} ds &= \frac{1}{\tau}\int_0^{\Delta t_n} X_{t_n} e^{-s/\tau} ds - \frac{\left(X_{t_n} - X_{t_{n-1}}\right)}{\Delta t_n}\frac{1}{\tau}\int_0^{\Delta t_n} s e^{-s\tau} ds \\ &= \left(1 - e^{-\Delta t_n/\tau}\right) X_{t_n} - \frac{\left(X_{t_n} - X_{t_{n-1}}\right)}{\Delta t_n}\gamma(\tau, \Delta t_n), \end{aligned} \tag{7.19}$$

where

$$\begin{aligned} \gamma(\tau, \Delta t_n) &= \frac{1}{\tau}\int_0^{\Delta t_n} s e^{-s/\tau} ds \\ &= -s e^{-s/\tau}\Big|_0^{\Delta t_n} + \int_0^{\Delta t_n} e^{-s/\tau} ds \\ &= -\Delta t_n e^{-\Delta t_n/\tau} + \tau\left(1 - e^{-\Delta t_n/\tau}\right). \end{aligned} \tag{7.20}$$

Plugging (7.20) into (7.19) and then in turn into (7.18) gives

$$\begin{aligned} \mathrm{EWMA}^{\mathrm{III}}(X,\tau)_{t_n} &= e^{-\Delta t_n/\tau}\mathrm{EWMA}^{\mathrm{III}}(X,\tau)_{t_{n-1}} + \left(1 - e^{-\Delta t_n/\tau}\right) X_{t_n} \\ &\quad + \left(X_{t_n} - X_{t_{n-1}}\right) e^{-\Delta t_n/\tau} - \frac{\left(X_{t_n} - X_{t_{n-1}}\right)}{\Delta t_n}\tau\left(1 - e^{-\Delta t_n/\tau}\right) \\ &= e^{-\Delta t_n/\tau}\mathrm{EWMA}^{\mathrm{III}}(X,\tau)_{t_{n-1}} + X_{t_n}\left(1 - \tau\frac{1 - e^{-\Delta t_n/\tau}}{\Delta t_n}\right) \\ &\quad + X_{t_{n-1}}\left(\tau\frac{1 - e^{-\Delta t_n/\tau}}{\Delta t_n} - e^{-\Delta t_n/\tau}\right). \end{aligned}$$

■

Remark 7.7 *All three EWMAs can be written a weighted sum of past observations with exponentially declining weights as a function of time. For example, iteratively applying the definition of* EWMA$^{\mathrm{I}}$ *gives*

$$\mathrm{EWMA}^{\mathrm{I}}(X,\tau)_{t_n} = \sum_{k=0}^{n-1} \omega_{n,k}^X X_{t_{n-k}},$$

for $1 \le n \le N(X)$, *with*

$$\omega_{n,k}^X = \begin{cases} \left(1 - e^{-\Delta t_{n-k}/\tau}\right) e^{-(t_n - t_{n-k})/\tau}, & \text{for } 0 \le k < n-1, \\ e^{-(t_n - t_1)/\tau}, & \text{for } k = n-1. \end{cases}$$

Lemma 7.5 and 7.6 can be used to derive corresponding expressions for the other two EWMAs.

8 Conclusion

This paper presents methods for analyzing arbitrarily-spaced time series in its unaltered form, i.e. without a need of first transforming to equally spaced-data. Nevertheless, the developed methods are consistent with existing literature on equally-spaced time series analysis. Since the existing literature on equally-spaced time series analysis is so vast, in this paper we can only scratch the surface of the corresponding methods for arbitrarily-spaced data, and a lot more work remains to be done to extend this theory.

A Proofs

- Calculation for Example 1.1: By definition, the conditional density is given by

$$\begin{aligned} f(B_t \mid B_a, B_b) &= \frac{f(B_a, B_t, B_b)}{f(B_a, B_b)} \\ &= \frac{\phi(B_a, s)\, \phi(B_t - B_a, t - s)\, \phi(B_b - B_t, r - t)}{\phi(B_a, s)\, \phi(B_b - B_a, r - s)} \\ &= \frac{\phi(B_t - B_a, t - s)\, \phi(B_b - B_t, r - t)}{\phi(B_b - B_a, r - s)}, \end{aligned}$$

  where $\phi(c, d)$ is the density function of a normal distribution with mean c and variance $d > 0$, respectively. A straightforward, but tedious calculation yields the desired result.

B Frequently Used Notations and Symbols

$\mathbf{0}_n$	the null vector of length n
$\varnothing$	the empty set
$*^{\mu}$	the convolution operator for signed measure μ, see Definition 4.4
B	the backshift operator, see Definition 2.6
$\mathcal{B}$	the Borel $\sigma-$algebra on $\mathbb{R}$
$\mathcal{B}_+$	the Borel $\sigma-$algebra on $\mathbb{R}_+$
L	the lag operator, see Definition 2.6
$N(X)$	the number of observations of time series X, see Definition 2.2
$\mathbb{O}_n$	the space of strictly increasing time sequences of length n, see Definition 2.1
$\mathbb{R}^n$	$n-$dimensional Euclidean space
$\mathbb{R}_+$	the interval $[0, \infty)$
$\mathcal{T}_n$	the space of time series of length n, see Definition 2.1
$\mathcal{T}$	the space of arbitrarily-spaced time series, see Definition 2.1
$\mathcal{T}(\tau)$	the space of time series in $\mathcal{T}$ with observation length less than τ, see Definition 5.9
$\mathcal{T}^K$	the space of $K-$dimension time series, see Definition 6.1
$T(X)$	the vector of observations times of a time series X, see Definition 2.2
$V(X)$	the vector of observations of a time series X, see Definition 2.2
$X^{\varnothing}$	the empty time series
$X[t]$	the most recent observation of time series X at time t, see Definition 2.4

References

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

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.

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 (2003). *Introduction to Time Series and Forecasting* (Second ed.). Springer.

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

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

Greene, W. H. (2007). *Econometric Analysis* (Sixth ed.). Prentice Hall.

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

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

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.

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