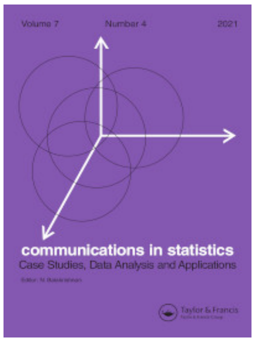

Communications in Statistics: Case Studies, Data Analysis and Applications

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ucas20

Bounded area as a measure of volatility for financial time series

Ferebee Tunno & Latia Carraway

COMMUNICATIONS IN STATISTICS: CASE STUDIES, DATA ANALYSIS AND APPLICATIONS
2022, VOL. 00, NO. 0, 1–13
https://doi.org/10.1080/23737484.2022.2019143

Check for updates

Bounded area as a measure of volatility for financial time series

Ferebee Tunno and Latia Carraway

Department of Mathematics and Statistics, Arkansas State University, Jonesboro, Arkansas, USA

ABSTRACT
This article presents a new way to measure the volatility of financial time series, which is shown to be on a par with arc length for such endeavors. An application involving the clustering of 30 prominent stocks is presented as well.

ARTICLE HISTORY
Received May 2021
Accepted December 2021

KEYWORDS
Bounded area; arc length; volatility; *k*-means++; Rand index

1. Introduction

This article puts forward "bounded area" as a valid measure of volatility for financial time series. The next section defines that term as well as its "arc length" counterpart. It then goes on to give a history of the two concepts in the academic literature. Section 3 outlines the *k*-means++ clustering algorithm as well as the details behind the Rand index. Both are then used in Section 4 to cluster and analyze a collection of 30 prominent stocks via bounded area, arc length, and the standard deviation of returns. Finally, Section 5 puts forward a geometric hypothesis.

2. Bounded Area and Arc Length

Observe the scatterplot of an arbitrary mean-zero process shown in Fig. 1. Typically, one "connects the dots" for better presentation as shown in Fig. 2. The sum of the lengths of these line segments is called the *arc length*. The shaded region between these line segments and the time axis shown in Fig. 3 is called the *bounded area*.

If $\{X_t\}$ is a mean-zero time series observed at times $t = 1, 2, \ldots, n$, then the magnitude of its bounded area is equal to

$$\sum_{t=2}^{n} \left[\int_{t-1}^{t} \left| (X_t - X_{t-1})(u - t) + X_t \right| du \right], \tag{1}$$

CONTACT Ferebee Tunno ftunno@astate.edu Department of Mathematics and Statistics, Arkansas State University, P.O. Box 70, State University, Arkansas 72467, USA

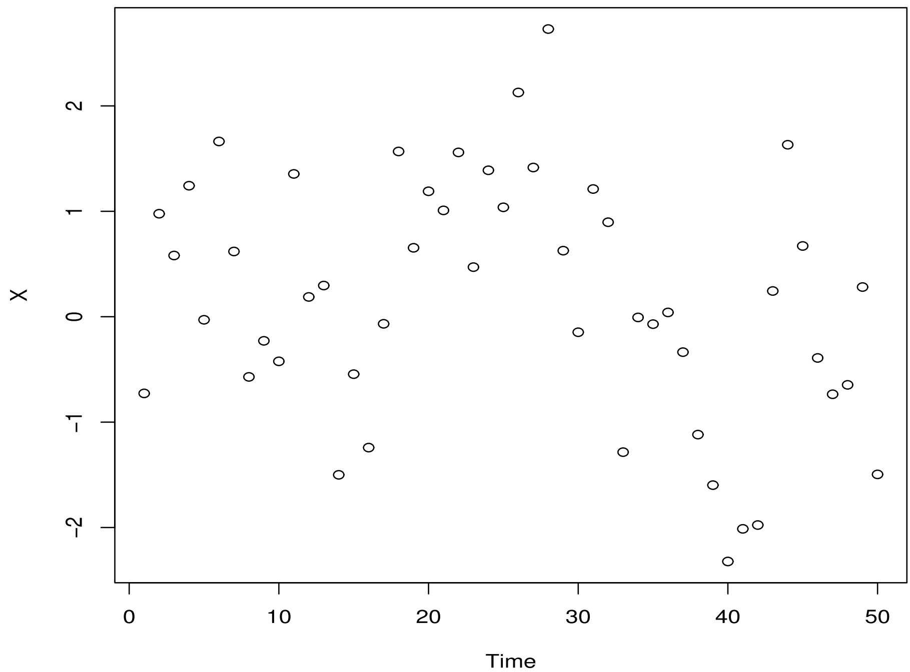

Figure 1. A scatterplot of an arbitrary mean-zero time series.

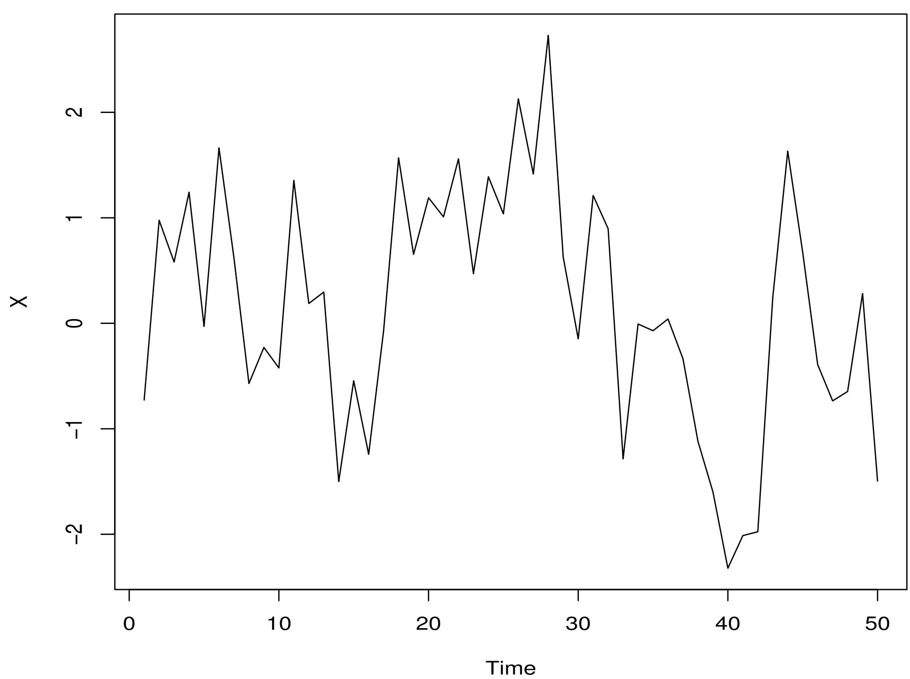

Figure 2. The same scatterplot from Figure 1, but with line segments connecting adjacent points.

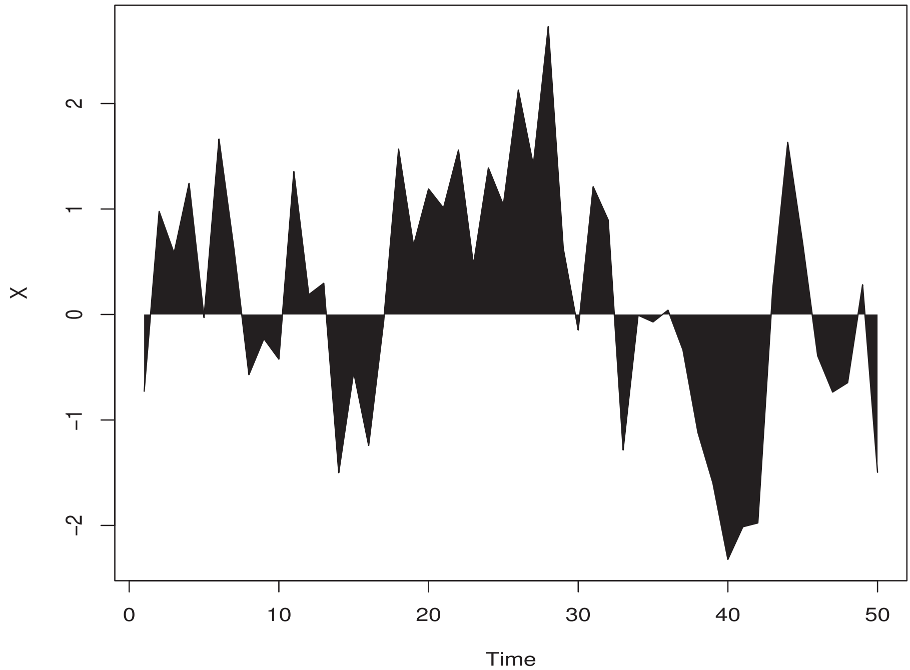

Figure 3. The region between the line segments and the time axis from Figure 2.

while its arc length is equal to

$$\sum_{t=2}^{n} \sqrt{1 + (X_t - X_{t-1})^2}\,. \tag{2}$$

Note that the mean-zero assumption is necessary for (1), but unnecessary for (2). That is, if we define a new process $\{Y_t\}$ such that $Y_t = X_t + k$ with $k \neq 0$, then the arc length calculations for $\{X_t\}$ and $\{Y_t\}$ over $t = 1, 2, \ldots, n$ will be the same, while the bounded area calculations will be different and only correct for $\{X_t\}$.

Tunno (2015) showed that if two independent, stationary, mean-zero ARMA processes with finite second moments are observed over the same period, then a significant difference between their bounded area magnitudes implies a significant difference between their autocovariance structures. Tunno, Gallagher, and Lund (2012) showed that if two independent, stationary ARMA processes with finite fourth moments are observed over the same period, then a significant difference between their arc lengths implies a significant difference between their autocovariance structures.

Tunno and Perry (2022) showed that if two independent, mean-zero signal-plus-noise processes are observed over the same period, then a significant difference between their bounded area magnitudes implies a significant difference between their underlying structures. They also showed the same to be true for arc length, but revealed that bounded area is a better discriminant between such signals. Tunno (2015) also demonstrated that bounded area distinguishes more accurately than arc length.

Wickramarachchi and Tunno (2015) showed that arc length is a suitable measure of volatility for financial time series and can be used to sort such series into meaningful clusters. In Section 4, the authors will make the case that bounded area is also a credible measure of volatility when clustering financial time series.

3. *k*-means++ algorithm and the Rand index

There are a wide variety of ways to cluster time series data. For a nice survey of all such techniques, see Liao (2005). In this section, the particular clustering algorithm known as *k-means++* is explained.

The original k-means algorithm for partitioning a numerical data set into k disjoint subsets/clusters was first created by MacQueen (1967) and goes as follows:

1. Choose the number of clusters k for your set S.
2. Randomly partition S into k clusters and determine their centers (averages) or directly generate k random points as cluster centers.
3. Assign each member from S to the nearest cluster, using some pre-chosen distance norm.
4. Recompute the new cluster centers.
5. Repeat steps 3 and 4 until things stabilize.

The k-means++ algorithm, proposed independently by Ostrovsky et al. (2012) and Arthur and Vassilvitskii (2007), improves upon the regular k-means algorithm by more carefully selecting the initial centers. k-means++ greatly reduces the possibility of suboptimal clustering by substituting the following algorithm in for the initial random partitioning of data points:

1. Choose one center uniformly at random from among the data points.
2. For each data point x, compute the distance $D(x)$ between x and the nearest center that has already been chosen.
3. Add one new data point at random as a new center, using a weighted probability distribution where point x is chosen with probability proportional to $(D(x))^2$.
4. Repeat steps 2 and 3 until k distinct centers have been chosen.

Now consider a set of elements $S = \{O_1, O_2, \ldots, O_N\}$ with partitions $X = \{x_1, x_2, \ldots, x_k\}$ and $Y = \{y_1, y_2, \ldots, y_k\}$. Next, define the following:

$$a = \text{the number of pairs of elements in } S \text{ that are in the same clusters of } X \text{ and also the same clusters of } Y$$

$$
\begin{aligned}
b &= \text{the number of pairs of elements in } S \text{ that are in different clusters} \\
&\quad\ \text{of } X \text{ and also different clusters of } Y \\
c &= \text{the number of pairs of elements in } S \text{ that are in the same clusters} \\
&\quad\ \text{of } X \text{ but different clusters of } Y \\
d &= \text{the number of pairs of elements in } S \text{ that are in different clusters} \\
&\quad\ \text{of } X \text{ but the same clusters of } Y
\end{aligned}
$$

We then calculate the *Rand index* as follows:

$$\frac{a+b}{a+b+c+d} = \frac{a+b}{\binom{N}{2}} = \frac{2(a+b)}{N(N-1)}.$$

This index is a number between 0 and 1 and gives the proportion of time that X and Y cluster S in the same manner. See Rand (1971) for further details.

Now consider the following contingency table, where n_{ij} stands for the number of elements from S that are contained in both x_i and y_j:

	y_1	y_2	$\cdots$	y_k	
x_1	n_{11}	n_{12}	$\cdots$	n_{1k}	a_1
x_2	n_{21}	n_{22}	$\cdots$	n_{2k}	a_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}	$\cdots$	n_{kk}	a_k
	b_1	b_2	$\cdots$	b_k	N

We then calculate the *adjusted Rand index* as follows:

$$\frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\binom{n_{ij}}{2} - \left[\sum_{i=1}^{k}\binom{a_i}{2}\sum_{j=1}^{k}\binom{b_j}{2}\right]\Big/\binom{N}{2}}{\frac{1}{2}\left[\sum_{i=1}^{k}\binom{a_i}{2}+\sum_{j=1}^{k}\binom{b_j}{2}\right] - \left[\sum_{i=1}^{k}\binom{a_i}{2}\sum_{j=1}^{k}\binom{b_j}{2}\right]\Big/\binom{N}{2}}.$$

This index is a "corrected-for-chance" version of the Rand Index and is also a number bounded above by 1, but has the potential to be negative. See Hubert and Arabie (1985) for further details.

4. Application

Table 1 below lists 30 stocks that the authors believe currently provide a reasonable representation of consequential American market activity. Also included are their bounded area and arc length values using adjusted closing prices from January 2013 to December 2019 ($n = 1{,}762$). In this section, these stocks will be clustered via the k-means++ algorithm[1] using bounded area and arc length as surrogates for volatility.

[1] https://toolslick.com/programming/ml/kmeans

Table 1. Thirty stocks along with their bounded area and arc length values using adjusted closing prices from January 2013 to December 2019.

Ticker	Stock	Bounded area	Arc length
GOOG	Alphabet, Inc.	420794.1	14341.55
AMZN	Amazon	879981.8	19415.68
AXP	American Express	26650.25	2328.659
AMGN	Amgen, Inc.	50385.62	3478.884
AAPL	Apple	78329.77	3303.17
BIIB	Biogen	72078.95	7925.215
BA	Boeing	168208.1	4690.726
BP	British Petroleum	7968.605	1894.525
CAT	Caterpillar	45777.15	2833.432
CVX	Chevron	19945.89	2534.194
C	Citigroup	16718.25	2158.694
KO	Coca Cola	8546.052	1859.205
DD	DuPont	22072.23	2418.913
XOM	Exxon Mobil	6336.809	2167.457
FB	Facebook	81377.65	3400.563
GE	General Electric	8830.629	1814.721
HD	Home Depot	74493.22	2837.987
HON	Honeywell International	51124.86	2485.025
INTC	Intel	16137.09	1978.727
IBM	International Business Machine	16601.37	2901.64
JNJ	Johnson & Johnson	35178.64	2345.988
JPM	J.P. Morgan Chase	42356.54	2279.517
MCD	McDonald's	65531.2	2541.402
MRK	Merck	18004.95	2048.75
MSFT	Microsoft	51480.39	2286.635
PG	Proctor & Gamble	19546.61	2117.501
UTX	United Technologies	24954.44	2509.084
VZ	Verizon	10056.64	1919.712
WMT	Wal-Mart	24149.18	2164.447
DIS	Walt Disney	29905.43	2458.388

Table 2. Three clusters using bounded area (left) and arc length (right).

Cluster 1	XOM	BP	KO	GE
	VZ	INTC	IBM	C
	MRK	PG	CVX	DD
	WMT	UTX	AXP	DIS
	JNJ			
Cluster 2	JPM	CAT	AMGN	HON
	MSFT	MCD	BIIB	HD
	AAPL	FB		
Cluster 3	BA			

Cluster 1	GE	KO	BP	VZ
	INTC	MRK	PG	C
	WMT	XOM	JPM	MSFT
	AXP	JNJ	DD	DIS
	HON	UTX	CVX	MCD
	CAT	HD	IBM	AAPL
	FB	AMGN		
Cluster 2	BA			
Cluster 3	BIIB			

Although not shown here, initial efforts to use the k-means++ algorithm to cluster the stocks in Table 1 according to their bounded area and arc length values always resulted in Google and Amazon constituting their own (upper) clusters. So as not to let these two "outlier" stocks obscure further partitioning among the other 28 stocks that might be meaningful, they will henceforward not be considered.

Tables 2–5 show the results of using the k-means++ algorithm to cluster the stocks listed in Table 1 according to their bounded area and arc length values, but now with Amazon and Google removed. This time around, Boeing is always its own cluster.

Table 3. Four clusters using bounded area (left) and arc length (right).

Cluster 1	XOM	BP	KO	GE
	VZ	INTC	IBM	C
	MRK	PG	CVX	DD
	WMT	UTX	AXP	DIS
Cluster 2	JNJ	JPM	CAT	AMGN
	HON	MSFT		
Cluster 3	MCD	BIIB	HD	AAPL
	FB			
Cluster 4	BA			

Cluster 1	GE	KO	BP	VZ
	INTC	MRK	PG	C
	WMT	XOM	JPM	MSFT
	AXP	JNJ	DD	DIS
	HON	UTX	CVX	MCD
Cluster 2	CAT	HD	IBM	AAPL
	FB	AMGN		
Cluster 3	BA			
Cluster 4	BIIB			

Table 4. Five clusters using bounded area (left) and arc length (right).

Cluster 1	XOM	BP	KO	GE
	VZ			
Cluster 2	INTC	IBM	C	MRK
	PG	CVX	DD	WMT
	UTX	AXP	DIS	JNJ
Cluster 3	JPM	CAT	AMGN	HON
	MSFT			
Cluster 4	MCD	BIIB	HD	AAPL
	FB			
Cluster 5	BA			

Cluster 1	GE	KO	BP	VZ
	INTC	MRK	PG	C
	WMT	XOM		
Cluster 2	JPM	MSFT	AXP	JNJ
	DD	DIS	HON	UTX
	CVX	MCD		
Cluster 3	CAT	HD	IBM	AAPL
	FB	AMGN		
Cluster 4	BA			
Cluster 5	BIIB			

Table 5. Six clusters using bounded area (left) and arc length (right).

Cluster 1	XOM	BP	KO	GE
	VZ			
Cluster 2	INTC	IBM	C	MRK
	PG	CVX	DD	WMT
	UTX	AXP		
Cluster 3	DIS	JNJ	JPM	
Cluster 4	CAT	AMGN	HON	MSFT
Cluster 5	MCD	BIIB	HD	AAPL
	FB			
Cluster 6	BA			

Cluster 1	GE	KO	BP	VZ
	INTC	MRK	PG	C
	WMT	XOM		
Cluster 2	JPM	MSFT	AXP	JNJ
	DD	DIS	HON	UTX
	CVX	MCD		
Cluster 3	CAT	HD	IBM	
Cluster 4	AAPL	FB	AMGN	
Cluster 5	BA			
Cluster 6	BIIB			

For perspective, Table 6 shows the standard deviation of returns for all 30 stocks during the period January 2013 through December 2019. Specifically, if $\{X_t\}$ is a time series observed at times $t = 1, 2, \ldots, n$ with $D_t := X_t - X_{t-1}$ for $t = 2, 3, \ldots, n$, then that standard deviation takes the form

$$\widehat{\sigma} = \sqrt{\frac{1}{n-2}\sum_{t=2}^{n}(D_t - \bar{D})^2}\,,$$

where

$$\bar{D} = \frac{1}{n-1}\sum_{t=2}^{n} D_t\,.$$

This is a common measure of the *volatility* of a stock. For a nice review of volatility in general, see Poon and Granger (2003).

Table 6. Thirty stocks along with the standard deviation of their returns using adjusted closing prices from January 2013 to December 2019.

Ticker	Stock	Std. Dev. of Returns
GOOG	Alphabet, Inc.	12.34592
AMZN	Amazon	18.73062
AXP	American Express	1.026904
AMGN	Amgen, Inc.	2.158978
AAPL	Apple	2.121515
BIIB	Biogen	6.899412
BA	Boeing	3.774907
BP	British Petroleum	0.4198629
CAT	Caterpillar	1.617281
CVX	Chevron	1.217688
C	Citigroup	0.7903046
KO	Coca Cola	0.364998
DD	DuPont	1.1292
XOM	Exxon Mobil	0.802288
FB	Facebook	2.373291
GE	General Electric	0.2585365
HD	Home Depot	1.608202
HON	Honeywell International	1.203946
INTC	Intel	0.587934
IBM	International Business Machine	1.66925
JNJ	Johnson & Johnson	1.087777
JPM	J.P. Morgan Chase	0.9626558
MCD	McDonald's	1.321727
MRK	Merck	0.6708082
MSFT	Microsoft	1.023353
PG	Proctor & Gamble	0.7695518
UTX	United Technologies	1.219441
VZ	Verizon	0.4645543
WMT	Wal-Mart	0.8837024
DIS	Walt Disney	1.218841

Table 7. Three (left) and four (right) clusters using $\widehat{\sigma}$.

Cluster 1	GE	KO	BP	VZ
	INTC	MRK	PG	C
	XOM	WMT		
Cluster 2	JPM	MSFT	AXP	JNJ
	DD	HON	CVX	DIS
	UTX	MCD		
Cluster 3	HD	CAT	IBM	AAPL
	AMGN	FB	BA	BIIB

Cluster 1	GE	KO	BP	VZ
	INTC	MRK	PG	
Cluster 2	C	XOM	WMT	JPM
	MSFT	AXP	JNJ	DD
Cluster 3	HON	CVX	DIS	UTX
	MCD	HD	CAT	
Cluster 4	IBM	AAPL	AMGN	FB
	BA	BIIB		

Tables 7 and 8 show the results of using the k-means++ algorithm to cluster the stocks listed in Table 6 according to $\widehat{\sigma}$, but with Amazon and Google removed. This time around, no single stock constitutes its own cluster.

Table 9 shows both the Rand and adjusted Rand index values for comparing bounded area clusters with arc length clusters, bounded area clusters with $\widehat{\sigma}$ clusters, and arc length clusters with $\widehat{\sigma}$ clusters. Amazon and Google are not present and so $N = 28$.

To the degree that the Rand index is reliable, then all three measures of volatility are cohesive and become more so as the number of clusters increases. To the degree that the adjusted Rand index is reliable, then that cohesion subsides significantly. The authors put more stock in the former (unadjusted)

Table 8. Five (left) and six (right) clusters using $\widehat{\sigma}$.

Cluster 1	GE	KO	BP	VZ
	INTC			
Cluster 2	MRK	PG	C	XOM
	WMT			
Cluster 3	JPM	MSFT	AXP	JNJ
	DD	HON		
Cluster 4	CVX	DIS	UTX	MCD
	HD	CAT	IBM	
Cluster 5	AAPL	AMGN	FB	BA
	BIIB			

Cluster 1	GE	KO	BP	VZ
	INTC			
Cluster 2	MRK	PG	C	XOM
Cluster 3	WMT	JPM	MSFT	AXP
	JNJ			
Cluster 4	DD	HON	CVX	DIS
	UTX			
Cluster 5	MCD	HD	CAT	IBM
Cluster 6	AAPL	AMGN	FB	BA
	BIIB			

Table 9. Rand (left) and adjusted Rand (right) index values for comparing bounded area (BA) clusters with arc length (AL) clusters, bounded area clusters with $\widehat{\sigma}$ clusters, and arc length clusters with $\widehat{\sigma}$ clusters.

	BA and AL	BA and $\widehat{\sigma}$	AL and $\widehat{\sigma}$
3 clusters	0.5714286	0.6375661	0.3835979
4 clusters	0.6825397	0.6137566	0.5820106
5 clusters	0.6904762	0.7195767	0.7962963
6 clusters	0.7142857	0.7619048	0.8095238

	BA and AL	BA and $\widehat{\sigma}$	AL and $\widehat{\sigma}$
3 clusters	0.1681926	0.2634199	0.0295304
4 clusters	0.3773507	0.1140151	0.2012304
5 clusters	0.2076749	0.1749444	0.4267990
6 clusters	0.1843683	0.1480517	0.4078329

index since the k-means++ algorithm should not need the same "correcting-for-chance" as the original k-means algorithm.

5. Triangular information versus rectangular information

(*Note: For this section of the paper, the word "area" is always meant to refer to a positive measure of two-dimensional space. It should also be clear that all time series discussed will be discrete and not continuous.*)

It is a simple matter of geometry that if one knows the lengths of both the hypotenuse and one leg of a right triangle, then one also knows the area of that triangle. It then follows that if one knows the lengths of the line segments connecting adjacent points in $\mathbb{R}^2$ from the set

$$\{(k, y_k), (k+1, y_{k+1}), (k+2, y_{k+2}), \ldots, (k+r, y_{k+r})\},$$

then one also knows the areas of the r right triangles whose hypotenuses are these segments. The specific instance with $r = 7$ is illustrated in Fig. 4.

It is also a geometric fact that if a line segment of known length in $\mathbb{R}^2$ crosses the abscissa, then one can obtain the areas of the two right triangles that appear above and below that abscissa. Specifically, if the endpoints of the line segment have coordinates (α, β) and (δ, ϵ), then it can be shown that those areas are

$$\frac{\beta^2}{2}\left|\frac{\alpha-\delta}{\epsilon-\beta}\right| \qquad \text{and} \qquad \frac{\epsilon^2}{2}\left|\frac{\alpha-\delta}{\epsilon-\beta}\right| .$$

See Fig. 5.

Putting all of these geometric facts together, it then follows that if one knows the arc length of a mean-zero time series $\{X_t\}$ sampled from $t = 1$ to $t = n$,

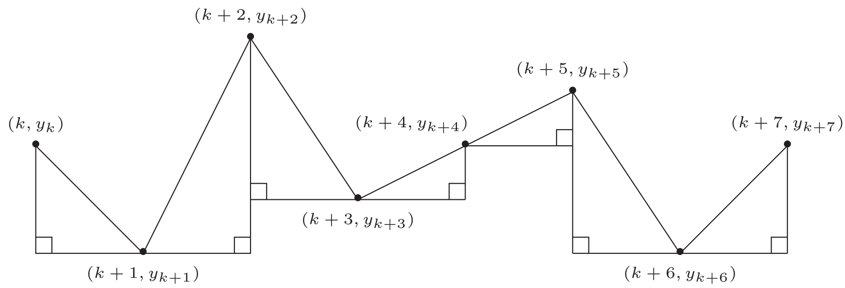

Figure 4. Seven line segments that serve as hypotenuses for seven right triangles. Knowledge of the lengths of these segments implies knowledge of the areas of the triangles.

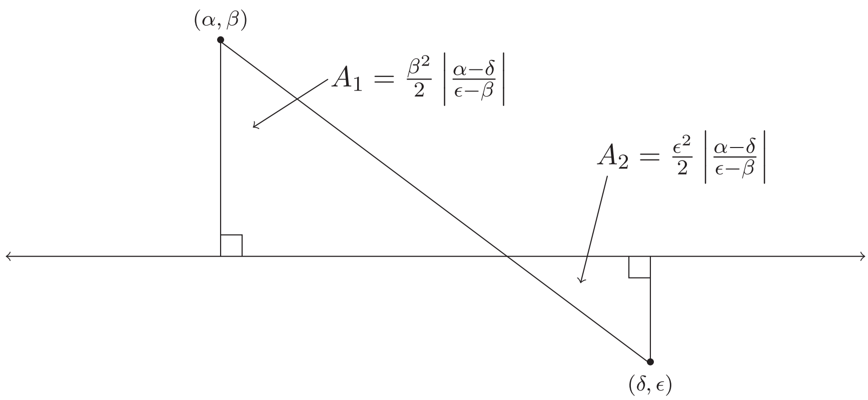

Figure 5. If a line segment of known length in $\mathbb{R}^2$ crosses the abscissa, then one can obtain the areas of the two right triangles that appear above and below that abscissa.

then one also knows the areas of all the right triangles that are created by the $n-1$ individual arc length segments. For arc length segments that lie completely above or below the time axis, let T_i denote the area of the right triangle whose hypotenuse connects points (i, X_i) and $(i+1, X_{i+1})$. For arc length segments that cross the time axis, let $T_{i,u}$ and $T_{i,l}$ denote the areas of the upper and lower right triangles, respectively, that correspond to that segment. In this case, we simply define $T_i = T_{i,u} + T_{i,l}$. See Fig. 6.

Henceforward, we will use the term *triangular information* to stand for the total area of all the right triangles that the arc length segments for a mean-zero time series sampled from $t = 1$ to $t = n$ create. That is,

$$\text{triangular information} = \sum_{i=1}^{n-1} T_i \, .$$

Thus, the arc length of a sampled mean-zero time series imparts triangular information about that series.

We now compare and contrast triangular information with what will henceforward be referred to as *rectangular information*. For arc length segments that lie completely above or below the time axis, let R_i denote the area of the rectangle

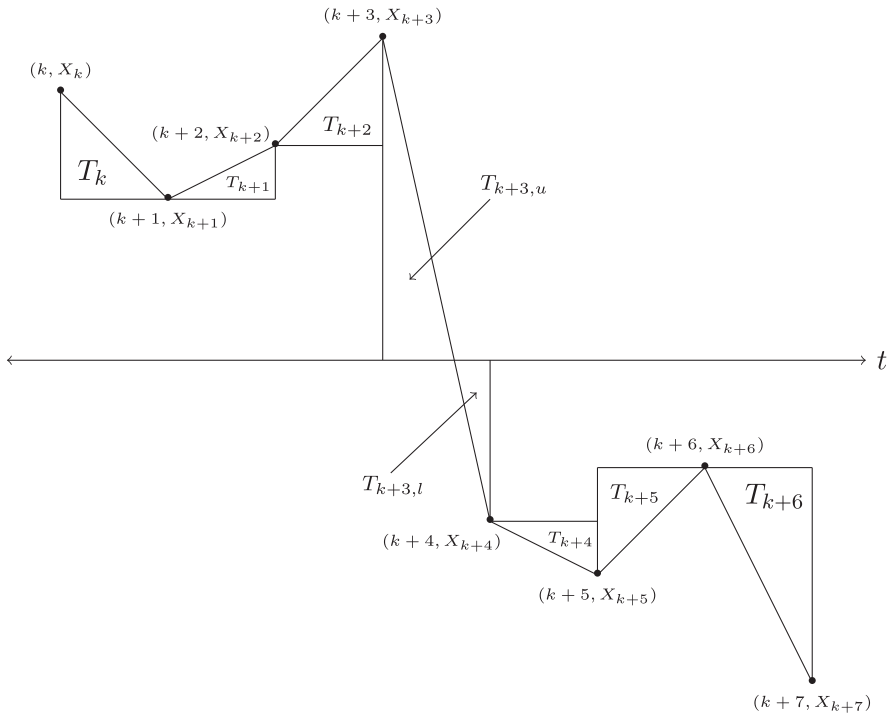

Figure 6. Arc length segments and their corresponding right triangles for a portion of an arbitrary mean-zero time series, where $T_{k+3} = T_{k+3,u} + T_{k+3,l}$.

that lies above or below the triangle whose hypotenuse connects points (i, X_i) and $(i+1, X_{i+1})$. Since there are no rectangles in the vertical band created by arc length segments that cross the time axis, we simply let $R_i = 0$ in this case. Now we put forward a definition analogous to triangular information:

$$\text{rectangular information} = \sum_{i=1}^{n-1} R_i \, .$$

See Fig. 7.

The upshot of this whole discussion is to make the following observation:

$$\text{bounded area} = \text{triangular information} + \text{rectangular information} \, .$$

Specifically, if a mean-zero time series is sampled from $t = 1$ to $t = n$, then

$$\text{bounded area} = \sum_{i=1}^{n-1} (T_i + R_i) \, .$$

Thus, the geometric information imparted by arc length is merely a subset of that imparted by bounded area. This is not to say, however, that arc length information can be recovered from bounded area information.

If future studies reveal either bounded area or arc length to truly be “better” than the other when measuring volatility, then that edge must be directly

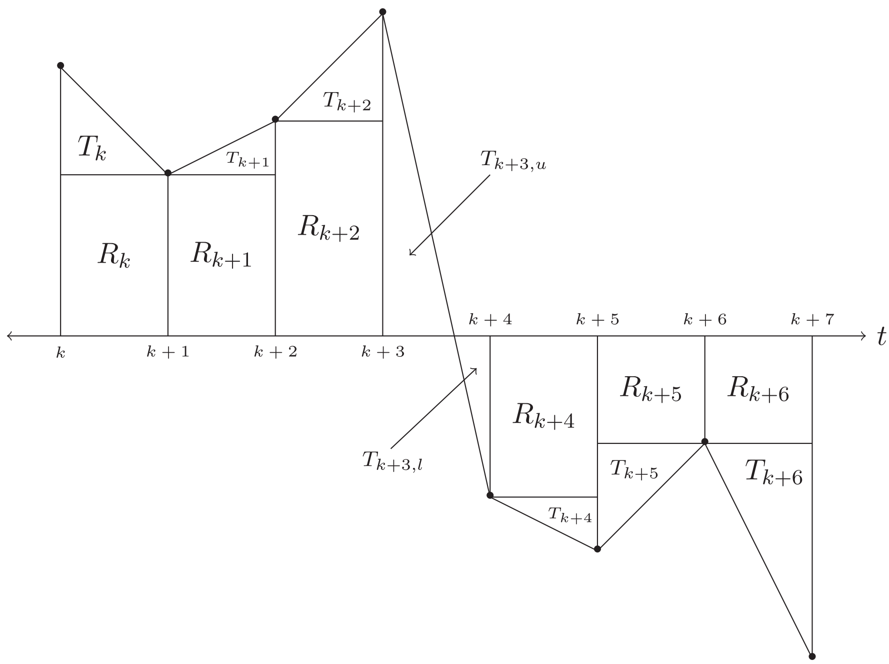

Figure 7. Triangular and rectangular information for a portion of an arbitrary mean-zero time series, where $T_{k+3} = T_{k+3,u} + T_{k+3,l}$ and $R_{k+3} = 0$.

connected to the presence or absence of rectangular information. Which of the two it might be is a subject for further consideration.

Another future pursuit will be to see if bounded area and arc length are meaningful surrogates for the conditional volatility associated with a GARCH process. To get a feel for how this task will be executed, recall that if $\{\epsilon_t\}$ is a GARCH process and if $\mathcal{E}_t$ is an information set based on events up to time t, then $\mathrm{Var}(\epsilon_t \mid \mathcal{E}_{t-1}) = \sigma_t^2$ is the conditional volatility associated with ϵ_t. It can also be shown that $\mathrm{Cov}(\epsilon_t^2, \epsilon_{t+h}^2) = \mathrm{Cov}(\sigma_t^2, \sigma_{t+h}^2)$.

Now let $\{\epsilon_{t,A}\}$ and $\{\epsilon_{t,B}\}$ be independent, stationary GARCH processes with conditional volatility processes $\{\sigma_{t,A}^2\}$ and $\{\sigma_{t,B}^2\}$, respectively. If we define $X_t := \epsilon_{t,A}^2$ and $Y_t := \epsilon_{t,B}^2$, then comparing the dynamics between $\{\sigma_{t,A}^2\}$ and $\{\sigma_{t,B}^2\}$ is equivalent to testing

$$\begin{aligned} H_0 : \gamma_X(h) &= \gamma_Y(h) \quad \text{for all } h \qquad \text{vs.} \\ H_1 : \gamma_X(h) &\neq \gamma_Y(h) \quad \text{for at least one } h\,, \end{aligned}$$

where either bounded area or arc length will be the test statistic pivot.

Acknowledgments

The authors wish to give special thanks to Dr. Gauri Guha for his assistance with this project.

References

Arthur, D., and S. Vassilvitskii. 2007. "k-means++: The Advantages of Careful Seeding." *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027–1035.

Hubert, L. J., and P. Arabie. 1985. "Comparing Partitions." *Journal of Classification* 2:193–218.

Liao, T. W. 2005. "Clustering of Time Series Data – A Survey." *Pattern Recognition* 38:1857–1874.

MacQueen, J. 1967. "Some Methods for Classification and Analysis of Multivariate Observations." *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* 1:281–297.

Ostrovsky, R., Y. Rabani, L. J. Schulman, and C. Swamy. 2012. "The Effectiveness of Lloyd-Type Methods for the k-means Problem." *Journal of the ACM* 59:28.

Poon, S., and C. W. J. Granger. 2003. "Forecasting Volatility in Financial Markets: A Review." *Journal of Economic Literature* 41:478–539.

Rand, W. M. 1971. "Objective Criteria for the Evaluation of Clustering Methods." *Journal of the American Statistical Association* 66:846–850.

Tunno, F. 2015. "Bounded Area Tests for Comparing the Dynamics Between ARMA Processes." *Communications in Statistics: Theory and Methods* 44:3921–3941.

Tunno, F., C. Gallagher, and R. Lund. 2012. "Arc Length Tests for Equivalent Autocovariances." *Journal of Statistical Computation and Simulation* 82:1799–1812.

Tunno, F., and M. Perry. 2022. "Signal Discrimination Without Denoising." *Communications in Statistics: Simulation and Computation* 51:626–646.

Wickramarachchi, T., and F. Tunno. 2015. "Using Arc Length to Cluster Financial Time Series According to Risk." *Communications in Statistics: Case Studies, Data Analysis, and Applications* 1:217–225.