Nonparametric Prediction Intervals and Regions for Random Walks and Renewal Processes

Mulubrhan G. Haile and David J. Olive *
Southern Illinois University

September 20, 2022

Abstract

Nonparametric prediction intervals and regions are given for univariate and vector valued random walks with applications for renewal processes.

KEY WORDS: shorth

1 Introduction

This paper suggests prediction intervals and regions for univariate and vector valued random walks. This section reviews the random walk, renewal processes, nonparametric prediction intervals, and nonparametric prediction regions.

A random walk (with drift) $Y_t = Y_{t-1} + e_t$ where the e_t are independent and identically distributed (iid). Suppose there is a sample $Y_1, ..., Y_n$ and we want a prediction interval (PI) for Y_{n+h}. Then $Y_t = Y_{t-2} + e_{t-1} + e_t = Y_{t-h} + e_{t-h+1} + \cdots + e_t = Y_0 + e_1 + \cdots + e_t$, or $Y_{n+h} = Y_n + e_{n+1} + e_{n+2} + \cdots + e_{n+h} = Y_n + \epsilon_{n,h}$. Let $e_j = Y_j - Y_{j-1}$ for $j = 2, ..., n$. Divide $e_2, ..., e_n$ into blocks of length h and let ϵ_i be the sum of the e_i in each block. Hence $\epsilon_1 = e_2 + \cdots + e_{h+1}$, $\epsilon_2 = e_{h+2} + \cdots + e_{2h+1}$, and $\epsilon_i = e_{(i-1)h+2} + e_{(i-1)h+3} + \cdots + e_{(i-1)h+h+1}$ for $i = 1, ..., m = \lfloor n/h \rfloor$. These ϵ_i are iid from the same distribution as $\epsilon_{n,h}$. The same decomposition can be made for a vector valued random walk, $\boldsymbol{Y}_t = \boldsymbol{Y}_{t-1} + \boldsymbol{e}_t$, where the vectors are $g \times 1$. Thus $\boldsymbol{\epsilon}_i = \boldsymbol{e}_{(i-1)h+2} + \boldsymbol{e}_{(i-1)h+3} + \cdots + \boldsymbol{e}_{(i-1)h+h+1}$ for $i = 1, ..., m$.

The random walk can be written as $Y_t = Y_0 + \sum_{i=1}^t e_i$ where $Y_0 = y_0$ is often a constant. A stochastic process $\{N(t) : t \geq 0\}$ is a counting process if $N(t)$ counts the total number of events that occurred in time interval $(0, t]$. Let e_n be the interarrival time or waiting time between the $(n-1)$th and nth events counted by the process, $n \geq 1$. If the nonnegative e_i are iid with $P(e_i = 0) < 1$, then $\{N(t), t \geq 0\}$ is a *renewal process*. Let $Y_n = \sum_{i=1}^n e_i$ = the time of occurrence of the nth event = waiting time until the

*Mulubrhan G. Haile is Visiting Assistant Professor, Westminster College, Fulton, MO, and David J. Olive is Professor, School of Mathematical & Statistical Sciences, Southern Illinois University, Carbondale, IL 62901-4408 (E-mail: *dolive@siu.edu*).

nth event. Then Y_n is a random walk with $Y_0 = y_0 = 0$. Let $E(e_i) = \mu > 0$. Then $E(Y_n) = n\mu$ and $V(Y_n) = nV(e_i)$ if $V(e_i)$ exists. A Poisson process with rate λ is a renewal process where the e_i are iid EXP(λ) with $E(e_i) = 1/\lambda$. See Ross (2014) for the Poisson process and renewal process. Given $Y_1, ..., Y_n$, then n events have occurred, and the 1-step ahead PI is for the time until the next event, the 2-step ahead PI is for the time until the next 2 events, and the h-step ahead PI is for the time for the next h events.

For forecasting, predict the test data $Y_{n+1}, ..., Y_{n+L}$ given the past training data $Y_1, ..., Y_n$. A large sample $100(1-\delta)$% prediction interval for Y_{n+h} is $[L_n, U_n]$ where the coverage $P(L_n \leq Y_{n+h} \leq U_n) = 1 - \delta_n$ is eventually bounded below by $1-\delta$ as $n \to \infty$. We often want $1 - \delta_n \to 1 - \delta$ as $n \to \infty$. A large sample $100(1-\delta)$% PI is asymptotically optimal if it has the shortest asymptotic length: the length of $[L_n, U_n]$ converges to $U_s - L_s$ as $n \to \infty$ where $[L_s, U_s]$ is the population shorth: the shortest interval covering at least $100(1-\delta)$% of the mass.

The shorth estimator of the population shorth will be defined below and used to create large sample PIs that do not require knowing the distribution of the errors e_t. If the data are $Z_1, ..., Z_n$, let $Z_{(1)} \leq \cdots \leq Z_{(n)}$ be the order statistics. Let $\lceil x \rceil$ denote the smallest integer greater than or equal to x (e.g., $\lceil 7.7 \rceil = 8$). Consider intervals that contain c cases $[Z_{(1)}, Z_{(c)}], [Z_{(2)}, Z_{(c+1)}], ..., [Z_{(n-c+1)}, Z_{(n)}]$. Compute $Z_{(c)} - Z_{(1)}, Z_{(c+1)} - Z_{(2)}, ..., Z_{(n)} - Z_{(n-c+1)}$. Then the estimator shorth$(c) = [Z_{(s)}, Z_{(s+c-1)}]$ is the interval with the shortest length.

Suppose the data $Z_1, ..., Z_n$ are iid and a large sample $100(1-\delta)$% PI is desired for a future value Z_f such that $P(Z_f \in [L_n, U_n]) \to 1 - \delta$ as $n \to \infty$. The shorth(c) interval is a large sample $100(1-\delta)$% PI if $c/n \to 1-\delta$ as $n \to \infty$, that often has the asymptotically shortest length. Frey (2013) showed that for large $n\delta$ and iid data, the shorth$(k_n = \lceil n(1-\delta) \rceil)$ prediction interval has maximum undercoverage $\approx 1.12\sqrt{\delta/n}$, and used the large sample $100(1-\delta)$% PI shorth(c) =

$$[Z_{(s)}, Z_{(s+c-1)}] \quad \text{with} \quad \mathrm{c} = \min(\mathrm{n}, \lceil \mathrm{n}[1 - \delta + 1.12\sqrt{\delta/\mathrm{n}}\ ]\ \rceil). \tag{1}$$

To describe the Olive (2013, 2017ab) nonparametric prediction region, Mahalanobis distances will be useful. Let the $g \times 1$ column vector T be a multivariate location estimator, and let the $g \times g$ symmetric positive definite matrix $\boldsymbol{C}$ be a dispersion estimator. Then the ith *squared sample Mahalanobis distance* is the scalar

$$D_i^2 = D_i^2(T, \boldsymbol{C}) = D^2_{\boldsymbol{w}_i}(T, \boldsymbol{C}) = (\boldsymbol{w}_i - T)^T \boldsymbol{C}^{-1} (\boldsymbol{w}_i - T) \tag{2}$$

for each observation $\boldsymbol{w}_i$, where $i = 1, ..., n$. Notice that the Euclidean distance of $\boldsymbol{w}_i$ from the estimate of center T is $D_i(T, \boldsymbol{I}_g)$ where $\boldsymbol{I}_g$ is the $g \times g$ identity matrix. The classical Mahalanobis distance D_i uses $(T, \boldsymbol{C}) = (\overline{\boldsymbol{w}}, \boldsymbol{S})$, the sample mean and sample covariance matrix where

$$\overline{\boldsymbol{w}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{w}_i \quad \text{and} \quad \boldsymbol{S} = \frac{1}{\mathrm{n}-1} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\boldsymbol{w}_\mathrm{i} - \overline{\boldsymbol{w}})(\boldsymbol{w}_\mathrm{i} - \overline{\boldsymbol{w}})^\mathrm{T}. \tag{3}$$

Consider predicting a future test value $\boldsymbol{w}_f$, given past training data $\boldsymbol{w}_1, ..., \boldsymbol{w}_n$ where $\boldsymbol{w}_1, ..., \boldsymbol{w}_n, \boldsymbol{w}_f$ are iid. Prediction intervals are a special case of prediction regions with $g = 1$ so the $\boldsymbol{w}_i$ are random variables.

A *large sample* $100(1-\delta)\%$ *prediction region* is a set $\mathcal{A}_n$ such that $P(\boldsymbol{w}_f \in \mathcal{A}_n) \geq 1-\delta$ asymptotically. A prediction region is *asymptotically optimal* if its volume converges in probability to the volume of the minimum volume covering region or the highest density region of the distribution of $\boldsymbol{w}_f$.

Like prediction intervals, prediction regions need correction factors. For iid data from a distribution with a $g \times g$ nonsingular covariance matrix, it was found that the simulated maximum undercoverage of prediction region (5) without the correction factor was about 0.05 when $n = 20g$. Hence the correction factor (4) is used to give better coverage for small n. Let $q_n = \min(1-\delta+0.05, 1-\delta+g/n)$ for $\delta > 0.1$ and

$$q_n = \min(1-\delta/2, 1-\delta+10\delta g/n), \quad \text{otherwise.} \tag{4}$$

If $1-\delta < 0.999$ and $q_n < 1-\delta+0.001$, set $q_n = 1-\delta$. Let $D_{(U_n)}$ be the $100q_n$th sample quantile of the D_i where $i = 1, ..., n$.

The large sample $100(1-\delta)\%$ *nonparametric prediction region* for a future value $\boldsymbol{w}_f$ given iid data $\boldsymbol{w}_1, ..., \boldsymbol{w}_n$ is

$$\{\boldsymbol{z} : (\boldsymbol{z}-\overline{\boldsymbol{w}})^T \boldsymbol{S}^{-1}(\boldsymbol{z}-\overline{\boldsymbol{w}}) \leq D^2_{(U_n)}\} = \{\boldsymbol{z} : D^2_{\boldsymbol{z}}(\overline{\boldsymbol{w}}, \boldsymbol{S}) \leq D^2_{(U_n)}\}. \tag{5}$$

The nonparametric prediction region is a large sample prediction region if the iid $\boldsymbol{w}_i$ have a nonsingular covariance matrix, and is asymptotically optimal for a large class of elliptically contoured distribution, including multivariate normal distributions with nonsingular covariance matrices. Regions with smaller asymptotic volumes can exist if the distribution is not elliptically contoured. From Olive (2018), simulated coverage was often near the nominal for $n \geq 20g$, but simulated volumes behaved better for $n \geq 50g$. The shorth PIs do not need the mean or variance of the e_t to exist. Prediction regions that do not need the nonsingular covariance matrix to exist are given by Zhang and Olive (2022).

Section two describes the prediction intervals and regions, while section 3 gives an example and simulations.

2 Prediction Intervals and Regions for the Random Walk

The prediction intervals and regions for the random walks are simple. First consider the random walk $Y_t = Y_{t-1} + e_t$ where the e_t are iid. Find the ϵ_i for $i = 1, ..., m = \lfloor n/h \rfloor$. Assume $n \geq 50h$ and let $[L, U]$ be the shorth(c) PI (1) for a future value of ϵ_f based on $\epsilon_1, ..., \epsilon_m$ with $m \geq 50$. Then the large sample $100(1-\delta)\%$ PI for Y_{n+h} is $[Y_n + L, Y_n + U]$. This PI tends to be asymptotically optimal as along as the e_t are iid. This PI is equivalent to applying the shorth(c) PI on $Y_n + \epsilon_1, ..., Y_n + \epsilon_m$. Note that $\epsilon_h = \epsilon_{n,h} \approx N(h\mu, h\sigma^2)$ for large h by the central limit theorem if $E(e_t) = \mu$ and $V(e_t) = \sigma^2$.

For the vector valued random walk $\boldsymbol{Y}_t = \boldsymbol{Y}_{t-1} + \boldsymbol{e}_t$, find $\boldsymbol{\epsilon}_{1,h}, ..., \boldsymbol{\epsilon}_{m,h}$. The nonparametric $100(1-\delta)\%$ prediction region for a future value $\boldsymbol{\epsilon}_{f,h}$ is

$$\{\boldsymbol{z} : (\boldsymbol{z}-\overline{\boldsymbol{\epsilon}})^T \boldsymbol{S}_h^{-1}(\boldsymbol{z}-\overline{\boldsymbol{\epsilon}}) \leq D^2_{(U_m)}\} = \{\boldsymbol{z} : D^2_{\boldsymbol{z}}(\overline{\boldsymbol{\epsilon}}, \boldsymbol{S}_h) \leq D^2_{(U_m)}\} \tag{6}$$

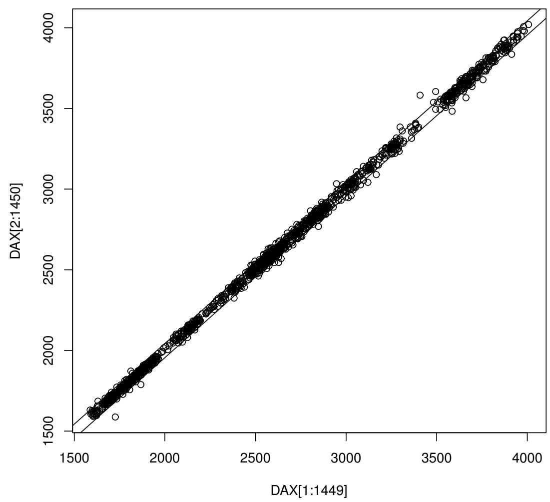

Figure 1: PI Plot of the DAX Data Set

where $\boldsymbol{S}_h$ is the sample covariance matrix of the $\boldsymbol{\epsilon}_{i,h}$ and $D_i^2 = (\boldsymbol{\epsilon}_{i,h} - \overline{\boldsymbol{\epsilon}})^T \boldsymbol{S}_h^{-1} (\boldsymbol{\epsilon}_{i,h} - \overline{\boldsymbol{\epsilon}})$. This prediction region is a hyperellipsoid centered at the sample mean $\overline{\boldsymbol{\epsilon}}$. The following large sample $100(1-\delta)\%$ prediction region for $\boldsymbol{Y}_{n+h}$ shifts the hyperellipsoid (6) to be centered at $\boldsymbol{Y}_n + \overline{\boldsymbol{\epsilon}}$:

$$\{\boldsymbol{z} : [\boldsymbol{z} - (\boldsymbol{Y}_n + \overline{\boldsymbol{\epsilon}})]^T \boldsymbol{S}_h^{-1} [\boldsymbol{z} - (\boldsymbol{Y}_n + \overline{\boldsymbol{\epsilon}})] \leq D^2_{(U_m)}\}. \tag{7}$$

Since $\boldsymbol{Y}_{n+h}$ has the same distribution as $\boldsymbol{Y}_n + \boldsymbol{\epsilon}_{f,h}$, $P(\boldsymbol{Y}_{n+h} \in (7)) = P(\boldsymbol{\epsilon}_{f,h} \in (6)) = 1 - \delta_n$ which is bounded below by $1-\delta$, asymptotically. The prediction region (7) is equivalent to applying the nonparametric prediction region (5) to $\boldsymbol{Y}_n + \boldsymbol{\epsilon}_{1,h}, ..., \boldsymbol{Y}_n + \boldsymbol{\epsilon}_{m,h}$. The prediction region (7) is similar to the Olive (2018) prediction region for the multivariate regression model.

3 Example and Simulations

Example 1. Common examples of random walks are stock prices. The EuStockMarkets data set, available from the *R* software, is a multivariate time series with 1860 observations on 4 variables. The observations are the daily closing prices of major European stock indices: Germany DAX, Switzerland SMI, France CAC, and UK FTSE. The data are sampled in business time, i.e., weekends and holidays are omitted. If we consider Y_t =DAX, the plot of the random walk $e_t = Y_t - Y_{t-1}$ is rectangular about the $e = 0$ line for cases 1-1460. Cases 1461-1800 scatter about the $e = 0$ line, but have much more variability (not shown but Figure 9.1 in Haile (2022)). Let cases 1-1450 be the training data, and let cases 1451-1460 be the test data. Figure 1 shows a plot of Y_{t-1} versus Y_t on the vertical axis for $t = 2$ to 1450. The two parallel lines correspond to the one step ahead 95% prediction intervals, which cover slightly more that 95% of the training data.

The remainder of this section gives simulations for the prediction intervals and regions. More simulations and tables are in Haile (2022). With 5000 runs, coverages between 0.94 and 0.96 suggest that there is no reason to believe that the nominal coverage is not 0.95.

Table 1: Random Walk 95% PI

n	dist	h=1	h=2	h=3	h=4
100	N	0.9554	0.9618	0.9432	0.9208
100		4.1675	6.3206	7.2125	7.7710
100	C	0.9594	0.9602	0.9394	0.9204
100		47.2301	570.4022	578.6495	562.9710
100	EXP	0.9602	0.9588	0.9466	0.9238
100		3.6581	6.2683	7.1028	7.6876
100	U	0.9496	0.9610	0.9436	0.9232
100		1.9027	3.2920	3.9957	4.3782
400	N	0.9576	0.9562	0.9590	0.9578
400		4.0667	5.7800	7.2466	8.3284
400	C	0.9562	0.9556	0.9648	0.9580
400		32.8266	72.3425	133.9639	189.4935
400	EXP	0.9632	0.9576	0.9604	0.9578
400		3.3091	5.1449	6.7579	7.9262
400	U	0.9532	0.9480	0.9554	0.9548
400		1.9035	3.1644	4.0582	4.7017
800	N	0.9466	0.9528	0.9532	0.9568
800		4.0192	5.7505	7.0041	8.1543
800	C	0.9536	0.9576	0.9522	0.9536
800		29.7084	65.2802	98.3195	142.1259
800	EXP	0.9592	0.9594	0.9570	0.9540
800		3.1997	5.0468	6.4145	7.6673
800	U	0.9498	0.9506	0.9546	0.9572
800		1.9013	3.1659	3.9642	4.6309

A small random walk simulation was done for the large sample 95% PIs using 5000 runs with $Y_0 = 1$. The errors e_i were iid from four distributions: i) N(1,1), ii) $t_1 \sim Cauchy(1,1)$, iii) EXP(1), and iv) uniform$(0,2)$. Only distribution iii) is not symmetric. We computed the h-step ahead 95% PIs for $h = 1,2,3,4 = L$. We want $n \geq 50L$, but simulations may use smaller n such as $n = 25L$. The asymptotic optimal lengths are i) 3.92, 5.54, 6.79, 7.84, ii) 25.41, 50.82, 76.24, 101.65, iii) 3.00, 4.72, 6.11, 7.22, iv) 1.90, 3.11, 3.87, 4.48.

Let the population forecast error be $e(h)$. For type 1, the asymptotic optimal lengths of the large sample 95% PIs are $3.92\sqrt{h}$ where $e(h) \sim N(h, \sigma^2 = h)$. For type 2, $e(h) \sim C(h, \sigma = h)$: a Cauchy distribution. For type 3, $e(h) \sim G(h,1)$: a Gamma distribution. For type 4, $e(2) \sim$ triangular(0,4). The distribution of the sum of n iid U(0,1) random variables is known as the Irwin-Hall distribution. See Gray and Odell

Table 2: Random Walk 95% Prediction Regions, p=8

n	ψ	type	h=1	h=2	h=3	h=4
400	0	1	0.9426	0.9438	0.9370	0.9214
400	0	2	0.9490	0.9502	0.9444	0.9270
400	0	3	0.9466	0.9530	0.9476	0.9392
400	0	4	0.9416	0.9446	0.9388	0.9216
400	0.354	1	0.9514	0.9446	0.9456	0.9186
400	0.354	2	0.9450	0.9572	0.9460	0.9290
400	0.354	3	0.9556	0.9546	0.9496	0.9314
400	0.354	4	0.9416	0.9412	0.9340	0.9182
400	0.9	1	0.9484	0.9462	0.9424	0.9198
400	0.9	2	0.9524	0.9502	0.9480	0.9310
400	0.9	3	0.9482	0.9576	0.9546	0.9392
400	0.9	4	0.9458	0.9376	0.9346	0.9228
800	0	1	0.9458	0.9450	0.9460	0.9484
800	0	2	0.9516	0.9554	0.9514	0.9506
800	0	3	0.9494	0.9508	0.9480	0.9544
800	0	4	0.9432	0.9408	0.9438	0.9418
800	0.354	1	0.9456	0.9464	0.9478	0.9450
800	0.354	2	0.9474	0.9550	0.9540	0.9488
800	0.354	3	0.9534	0.9516	0.9532	0.9536
800	0.354	4	0.9494	0.9466	0.9480	0.9518
800	0.9	1	0.9436	0.9482	0.9478	0.9450
800	0.9	2	0.9500	0.9494	0.9512	0.9514
800	0.9	3	0.9552	0.9520	0.9514	0.9484
800	0.9	4	0.9474	0.9450	0.9494	0.9464
1600	0	1	0.9506	0.9516	0.9476	0.9464
1600	0	2	0.9522	0.9534	0.9532	0.9514
1600	0	3	0.9496	0.9530	0.9524	0.9522
1600	0	4	0.9418	0.9428	0.9414	0.9430
1600	0.354	1	0.9506	0.9472	0.9504	0.9502
1600	0.354	2	0.9440	0.9520	0.9488	0.9502
1600	0.354	3	0.9506	0.9572	0.9574	0.9570
1600	0.354	4	0.9488	0.9418	0.9444	0.9462
1600	0.9	1	0.9510	0.9496	0.9476	0.9458
1600	0.9	2	0.9492	0.9500	0.9532	0.9474
1600	0.9	3	0.9524	0.9558	0.9548	0.9540
1600	0.9	4	0.9450	0.9508	0.9452	0.9500

(1966), Marengo, Farnsworth, and Stefanic (2017), and Roach (1963).

Results are shown in Table 1. Roughly need $n/h \geq 50$ for good coverage. Thus $n = 100$ was too small for the h-step ahead PIs with $h = 3$ and $h = 4$. The Cauchy distribution needs huge n before the average PI lengths get close to the asymptotically optimal lengths.

A small vector valued random walk simulation was also done for the large sample 95% prediction regions using 5000 runs. We used distributions with nonsingular population covariance matrices. Let $\boldsymbol{u}_t = (u_{t1}, ..., u_{tp})^T$ where the u_{ti} are iid from a type 1) $N(1, 1)$, 2) $1 + t_5$, 3) EXP(1), or 4) U(0,2) distribution. Then $\boldsymbol{e}_t = \boldsymbol{A}\boldsymbol{u}_t$ where the $p \times p$ matrix $\boldsymbol{A} = (a_{ij})$ with the diagonal elements $a_{ii} = 1$, and $a_{ij} = \psi$ for $i \neq j$. The prediction regions are hyperellipsoids, which have volumes (not given), instead of lengths.

Table 2 shows some results when $p = 8$, giving the coverages. Roughly need $n/h \geq 20p$ to get good coverages. Thus $n = 400$ was too small for $p = 8$ with $h = 3$ or $h = 4$, although undercoverage was small for $h = 3$.

4 Discussion

Other prediction intervals and regions can by used since the $\boldsymbol{\epsilon}_i$ are iid. For $g = 1$ with $\boldsymbol{\epsilon}_i = \epsilon_i$, the asymptotically optimal shorth PI was used. Convergence of the shorth PI to the population shorth can be slow. See Grübel (1988). The Zhang and Olive (2022) prediction regions can also be used for the vector valued random walk.

Data sets where the future data does not behave like the past data are common, and then the prediction intervals and regions tend to perform poorly. In Example 1, cases 1-1460 appear to follow one random walk, while cases 1461-1800 follow another random walk with more variability.

Some prediction intervals for stochastic processes include Pan and Politis (2016), Vidoni (2004), and Vit (1973). Mykland (2003) described how to convert prediction regions into investment strategies. Pankratz (1983, p. 106) notes that the random walk model has been found to be a good model for many stock price time series.

Plots and simulations were done in *R*. See R Core Team (2020). Programs are in the collection of functions *tspack.txt*. See (http://parker.ad.siu.edu/Olive/tspack.txt). Tables 1 and 2 used the functions `rwpisim` and `rwprsim` for the random walk simulations.

REFERENCES

Frey, J. (2013), "Data-Driven Nonparametric Prediction Intervals," *Journal of Statistical Planning and Inference*, 143, 1039-1048.

Gray, H.L., and Odell P.L., (1966), "On Sums and Products of Rectangular Variates," *Biometrika*, 53, 615-617.

Grübel, R. (1988), "The Length of the Shorth," *The Annals of Statistics,* 16, 619-628.

Haile, M.G. (2022), "Inference for Time Series after Variable Selection," Ph.D. Thesis, Southern Illinois University. See (http://parker.ad.siu.edu/Olive/shaile.pdf).

Marengo, J.E., Farnsworth, D.L., and Stefanic, L. (2017), "A Geometric Derivation of the Irwin-Hall Distribution," *International Journal of Mathematics and Mathematical Sciences*, 2017, online.

Mykland, P.A. (2003), "Financial Options and Statistical Prediction Intervals," *The Annals of Statistics*, 31, 1413-1438.

Olive, D.J. (2013), "Asymptotically Optimal Regression Prediction Intervals and Prediction Regions for Multivariate Data," *International Journal of Statistics and Probability*, 2, 90-100.

Olive, D.J. (2017a), *Robust Multivariate Analysis*, Springer, New York, NY.

Olive, D.J. (2017b), *Linear Regression*, Springer, New York, NY.

Olive, D.J. (2018), "Applications of Hyperellipsoidal Prediction Regions," *Statistical Papers*, 59, 913-931.

Pan, L., and Politis, D.N. (2016), "Bootstrap Prediction Intervals for Markov Processes," *Computational Statistics* & *Data Analysis*, 100, 467-494.

Pankratz, A. (1983), *Forecasting with Univariate Box-Jenkins Models*, Wiley, New York, NY.

R Core Team (2020), "R: a Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

Roach, S.A. (1963), "The Frequency Distribution of the Sample Mean Where Each Member of the Sample is Drawn from a Different Rectangular Distribution," *Biometrika*, 50, 508-513.

Ross, S.M. (2014), *Introduction to Probability Models*, 11th ed., Academic Press, San Diego, CA.

Vidoni, P. (2004), "Improved Prediction Intervals for Stochastic Process Models," *Journal of Time Series Analysis*, 25, 137-154.

Vit, P. (1973), "Interval Prediction for a Poisson Process," *Biometrika*, 60, 667-668.

Zhang, L., and Olive, D.J. (2022), "A Data Splitting Prediction Region," preprint at (http://parker.ad.siu.edu/Olive/ppdatsplittinf.pdf).