

Long-Memory and Spurious Breaks in Ecological Experiments

Thomas R. Boucher

Department of Mathematics, Texas A & M University-Commerce, Commerce, USA

Email: thomas.boucher@tamuc.edu

How to cite this paper: Boucher, T.R. (2017) Long-Memory and Spurious Breaks in Ecological Experiments. *Open Journal of Statistics*, 7, 768-779.

<https://doi.org/10.4236/ojs.2017.75054>

Received: November 16, 2015

Accepted: October 8, 2017

Published: October 11, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The impact of long-memory on the Before-After-Control-Impact (BACI) design and a commonly used nonparametric alternative, Randomized Intervention Analysis (RIA), is examined. It is shown the corrections used based on short-memory processes are not adequate. Long-memory series are also known to exhibit spurious structural breaks that can be mistakenly attributed to an intervention. Two examples from the literature are used as illustrations.

Keywords

BACI, Long-Memory, RIA, Short-Memory, Variance Corrections

1. Introduction

Ecological studies often involve data collected over time. Examples are observations of population densities such as the relative abundance of the white sea urchin (*Lytechinus anamesus*) in an area offshore the San Onofre Nuclear Generating Station (Schroeter *et al.* [1]), or of the difference in chlorophyll concentrations between two lakes (Carpenter *et al.* [2]). This data need not satisfy the standard assumption of independent observations but can in fact be autocorrelated. A *long-memory* time series has autocorrelation that decays at a slow hyperbolic rate. Long-memory has been shown to be effective in modeling natural processes such as observations of the yearly minimal water levels of the Nile River and the monthly temperatures for the northern hemisphere (Beran [3]).

Researchers have explored the relationships among long-memory, aggregation, and structural breaks in time series [4] [5] [6]. In [5] the authors show that the number of spurious breaks in a long-memory series approaches infinity as the sample size does, while in [4] [5] and [6] the authors explore the fact that

structural breaks in a time series can create spurious long memory. In [6] the authors propose a test to detect spurious long memory using aggregation tests. Unfortunately, these aggregation tests depend upon very long time series, which is rarely the case in ecological experiments, making undetectable long memory a real danger.

Tests which seek to detect breaks due to an intervention in a series whose true data generating process is long memory are in danger of detecting spurious breaks. Later in this paper we examine two examples taken from the literature. They were chosen because they present instances in the literature where a significant intervention effect was detected when in fact no intervention occurred. A possible explanation for this is the presence of strong correlation in the data, perhaps long-memory, which could have produced the spurious detection of a significant intervention effect. This possibility provided the impetus for this manuscript.

The Before-After-Control-Impact (BACI) design [7] uses two ecological units, one as a control and the other as an impact, *i.e.*, the impact unit has an intervention applied to it. Repeated measurements are taken on each of the units, before and after the intervention. The paired in time differences between impact and control units are the object of statistical analysis. The original BACI analysis uses a 2-sample *t*-test to compare the pre-intervention mean paired difference with the post-intervention mean paired difference. An alternative to the 2-sample *t*-test is Randomized Intervention Analysis (RIA) [2], which uses a permutation test to conduct the comparison. BACI and RIA both assume the pre-intervention differences and the post-intervention differences from the repeated measurements on the control and the impact units form two independent random samples.

In lieu of ignoring the autocorrelation, strategies have been proposed to adjust the BACI and RIA analyses for autocorrelated data. One approach is parametric: estimate the correlation structure using an assumed (short memory) model and use the estimated correlation to adjust the 2-sample *t*-test and confidence interval in the BACI analysis (see Bence [8] for a survey). A second option is to adopt a nonparametric approach, whereby the original data is block resampled, the blocks being groups of observations chosen large enough to take the correlation into account. RIA is essentially block resampling using blocks of size one. This is completely accurate only where the observations are independent. Neither approach is entirely satisfactory when the data is collected from a long-memory process, leaving the experimenter vulnerable to spurious break detection.

This paper is structured as follows: Section 2 contains definitions and some simple derivations. In Section 3 we conduct numerical studies to illustrate the inadequacy of short-memory corrections for long-memory series. Section 4 contains two examples from the literature, and Section 5 concludes the paper with a brief discussion.

2. Definitions and Derivations

2.1. Time Series

The following facts from time series theory and methodology may be found in standard texts such as [9]. The autocorrelation structure of a stationary time series $\{X_t\}$ is described by its autocorrelation function (ACF) denoted $\rho_X(h)$, where h denotes the time lag between the two random variables:

$$\rho_X(h) := \text{Corr}(X_t, X_{t+h}), \quad h = \pm 1, 2, 3, \dots$$

A *short-memory* time series has an ACF that decays at an exponential rate, *i.e.*, $\rho_X(h)r^h$ approaches a positive constant as $h \rightarrow \infty$ for some $0 < r < 1$. A *long-memory* time series has an ACF that decays at a hyperbolic rate: $\rho_X(h)r^h$ approaches a positive constant as $h \rightarrow \infty$ for some $0 < r < 1$.

Suppose $\{W_t\}$ is a white noise process. We consider the following short-memory process, the autoregressive model of order 1 (AR(1))

$$X_t = \phi X_{t-1} + W_t, \quad |\phi| < 1.$$

The ACF of the AR(1) is given by

$$\rho_X(h) = \phi^h, \quad h = 0, 1, 2, \dots$$

Long-memory processes, as described by fractionally differenced white noise (FD(d)), define the time series $\{X_t\}$ by

$$(1-B)^d X_t = W_t, \quad -0.5 < d < 0.5.$$

where B defined by $BX_t = X_{t-1}$ is the backshift operator and the fractional differencing operator $(1-B)^d$ has the polynomial expansion

$$(1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j, \quad \pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}.$$

The exact form of the ACF of the FD(d) process is known (see [3] pgs. 63ff.):

$$\rho_X(h) = \frac{\Gamma(h+d)\Gamma(1-d)}{\Gamma(h-d+1)\Gamma(d)} = \prod_{0 < i \leq h} \frac{i-1+d}{i-d}, \quad h = 1, 2, \dots \quad (1)$$

The AR(1) and FD(d) are both stationary processes. Fractionally differenced white noise is a classic long-memory time series, having an ACF that decays at a hyperbolic rate: $\rho(h)h^{1-2d}$ approaches a positive constant as $h \rightarrow \infty$. The AR(1) model is short-memory since its ACF converges to zero at an exponential rate as $h \rightarrow \infty$.

2.2. BACI Analysis

Consider the BACI design. Suppose Y_1, \dots, Y_n are observations on the impact site, X_1, \dots, X_n are observations on the control site and D_1, \dots, D_n are the differences between the two:

$$D_t := Y_t - X_t, \quad t = 1, \dots, n.$$

Assume $\{Y_t\}$ and $\{X_t\}$ are jointly stationary, yielding a stationary $\{D_t\}$,

and that $Var(D_i) = \sigma^2$. As is well-known, if D_1, \dots, D_n form a random sample then $Var(\bar{D}) = \sigma^2/n$. However, when D_1, \dots, D_n are realizations from a stationary time series with autocorrelation function $\rho_D(h)$, then

$$Var(\bar{D}) = \frac{\sigma^2}{n} \left[1 + 2 \sum_{h=1}^{n-1} \left(1 - \frac{h}{n} \right) \rho_D(h) \right]. \quad (2)$$

The quantity

$$1 + 2 \sum_{h=1}^{n-1} \left(1 - \frac{h}{n} \right) \rho_D(h) \quad (3)$$

is sometimes called the variance *correction factor* [8] [3].

The estimated correction factor

$$1 + 2 \sum_{h=1}^{n-1} \left(1 - \frac{h}{n} \right) \hat{\rho}(h) \quad (4)$$

is used to adjust the usual estimate of $Var(\bar{D})$, s^2/n , when D_1, \dots, D_n are realizations of a stationary time series. Bence [8] made an extensive investigation of the effect of the estimated correction factor when the autocorrelation is assumed to be that of an AR(1) model and is estimated by $\hat{\rho}(h) = \hat{\phi}^h$, where $\hat{\phi}$ is an estimate of the autoregression coefficient.

The BACI design uses a 2-sample t -test to compare the pre-intervention and post-intervention control-impact mean differences. Let \bar{D}_{Pre} denote the n_1 pre-intervention differences, \bar{D}_{Post} the n_2 post-intervention differences, $n_1 + n_2 = n$, and $\hat{SE}(\bar{D}_{Post} - \bar{D}_{Pre})$ be the estimated standard error of $\bar{D}_{Post} - \bar{D}_{Pre}$ where

$$\hat{SE}(\bar{D}_{Post} - \bar{D}_{Pre}) = \hat{\sigma}_D \sqrt{\frac{1}{n_1} \left[1 + 2 \sum_{h=1}^{n_1-1} \left(1 - \frac{h}{n_1} \right) \hat{\rho}_D(h) \right] + \frac{1}{n_2} \left[1 + 2 \sum_{h=1}^{n_2-1} \left(1 - \frac{h}{n_2} \right) \hat{\rho}_D(h) \right]}.$$

The estimated standard error $\hat{SE}(\bar{D}_{Post} - \bar{D}_{Pre})$ is calculated using (2) and the estimated variance correction (4). The estimates $\hat{\sigma}_D$ and $\hat{\rho}_D$ are obtained from pooling the two sets of differences. Note $\hat{SE}(\bar{D}_{Post} - \bar{D}_{Pre})$ ignores the correlation between the two samples since

$$SE(\bar{D}_{Post} - \bar{D}_{Pre}) = \sqrt{Var(\bar{D}_{Post}) + Var(\bar{D}_{Pre}) - 2Cov(\bar{D}_{Post}, \bar{D}_{Pre})}.$$

This is another reason for the inaccuracy of the method in the presence of long-memory; for a short-memory process the problem will not be as severe.

The assumption that $\{X_t\}$ and $\{Y_t\}$ are jointly stationary allows the use of the 2-sample t -test with equal variances. Combined with the null hypothesis of no intervention effect, this suggests the following approximate test statistic for the 2-sample t -test

$$\frac{\bar{D}_{Post} - \bar{D}_{Pre}}{\hat{SE}(\bar{D}_{Post} - \bar{D}_{Pre})} \sim t_{n_1+n_2-2}.$$

The use of the t -distribution depends on asymptotic theory which requires very large samples when the process is long-memory. For smaller samples it is

not exactly correct, but it is difficult to work out the exact distribution ([3], Section 8.6.3).

2.3. RIA and Permutation Tests

An alternative to using a correction factor for the standard error of the mean is the use of nonparametric methods. The procedure is to resample blocks (the *block bootstrap*), the blocks being chosen large enough to properly capture the autocorrelation. The permutation test used in RIA is essentially block resampling from blocks of size one. This can be effective where the correlation structure is that of a short-memory process since blocks of minimal size are required. However, when long-memory is present the blocks must be large, requiring very large samples.

The problems that correlated data pose for RIA have been studied previously. One examination is in Carpenter *et al.* [2]. The authors simulated data from short-memory AR(1) and MA(1) processes and analyzed these with a permutation test. Recognizing RIA is affected by autocorrelations, the authors recommended a correction to the p-value when dealing with positive autocorrelations, for example, using a declared p-value of 0.01 to get a true p-value of 0.05. However, the autocorrelations used in this study were short-memory and moderate at most in strength. The simulation results in Section 3 suggest such a correction is not adequate for data with long-memory, as in a FD(d) process.

3. Simulations

All simulations were run using the R environment [10]. The R package *fracdiff* [11] calculates the maximum likelihood parameters of a FD(d) model, following Haslett and Raftery [12]. There is a large body of literature concerned with estimating the long-memory parameter d , but this is not the focus of this paper. The package also contains a routine which will simulate observations from the process. In all simulations we assumed the white noise process followed a standard normal distribution.

3.1. Variance Correction Factors

The correction factors for FD(d) and AR(1) processes can be computed from (1) and (3) when the values of d, ϕ are known. **Table 1** below contains the

Table 1. Variance corrections of autoregressive (AR) and fractionally differenced (FD) models, rounded to 2 decimal places. Parameter values are in parentheses.

n	AR(0.7)	AR(0.9)	AR(0.99)	FD(0.3)	FD(0.49)
5	3.08	7.19	8.80	2.40	8.57
10	4.16	12.03	18.12	3.61	17.82
25	5.04	17.56	43.44	6.24	45.10
50	5.36	18.90	78.00	9.46	89.84
100	5.51	19.00	125.79	14.33	178.08

correction factors for the FD(d) process for $d = 0.3, 0.49$ and those of the AR(1) process for values $\phi = 0.7, 0.9, 0.99$. Several sample sizes were used.

The most striking difference between the AR(1) and the FD(d) correction factors are the rates at which they increase as the sample size increases. FD(d) processes increase more due to the fact the autocorrelations persist longer. For small sample sizes the AR(1) corrections tend to be equal to or slightly larger than the FD(d) corrections, while for large sample sizes they are too small.

Bence [8] observes that setting ϕ to 0.99 or 0.9999 makes little difference because in either case the confidence bounds will be so broad that little could be claimed on the basis of the estimate. AR(0.99) is competitive for small samples, but for moderate to large samples it underestimates the variance correction factor; setting $\phi = 0.99$ is too conservative for data with very strong long-memory, indicating how serious the situation is. Also, simulated data from $d = 0.49$ has a reasonable probability of returning a non-stationary AR(1) model, particularly for smaller samples, rendering the short-memory correction unusable.

3.2. Size of 2-Sample t Hypothesis Tests with and without AR(1) Variance Corrections

To investigate the size α of 2-sample t -tests when the data are from a long memory process, series of various lengths for several values of d were simulated. Each simulated series was split into two equal halves to be the two series. The case $d = 0$ corresponds to white noise for the errors. The t -test statistic was calculated both with and without the AR(1) variance correction, and the null was rejected if the test statistic exceeded the appropriate critical value. The proportion of rejections was the estimated size of the test. The AR(1) correction used the value of ϕ estimated from the simulated series. Results are in **Table 2**.

For white noise ($d = 0$) processes the uncorrected and AR(1) corrected tests have size approximately equal to the nominal size. As d increases, the sizes of the tests increase. Though the AR(1) performs better than no correction, the performance is very poor for strong long-memory. Also, in the presence of long-memory the size of the test increases from the nominal size as the sample size increases, the performance being worse the stronger the long-memory.

3.3. Randomized Intervention Analysis

Carpenter *et al* appear to have introduced RIA in [2]. The RIA permutation test examines all possible permutations of the observed pre-intervention and post-intervention differences, determines a distribution from this for the absolute value of the difference between the pre-intervention and the post-intervention means and uses this distribution to compute a p-value for the observed data. The case where there is no intervention effect is equivalent to splitting a single series of differences $\{D_t\}$ into two series at the time of the ineffective intervention, and then comparing these two series with a permutation

Table 2. Hypothesis test size Monte Carlo approximations for 10,000 simulated $FD(d)$ with $\alpha = 0.10, 0.05, 0.01$, $d = 0.0, 0.2, 0.4$ and $n_1 + n_2 = 20, 40, 60, 80, 100$ with $n_1 = n_2$. “None” denotes the ordinary 2-sample t -test, “AR” denotes the 2-sample t -test with the AR(1) correction.

α		0.10		0.05		0.01	
d	n	None	AR	None	AR	None	AR
0.0	20	0.107	0.105	0.052	0.044	0.009	0.003
	40	0.099	0.101	0.050	0.047	0.010	0.006
	60	0.100	0.100	0.051	0.049	0.010	0.008
	80	0.099	0.100	0.050	0.048	0.010	0.008
	100	0.102	0.101	0.049	0.047	0.010	0.009
0.2	20	0.221	0.157	0.138	0.075	0.048	0.009
	40	0.275	0.193	0.189	0.110	0.083	0.026
	60	0.309	0.218	0.225	0.137	0.107	0.041
	80	0.332	0.240	0.247	0.154	0.127	0.052
	100	0.352	0.254	0.268	0.169	0.145	0.063
0.4	20	0.361	0.211	0.271	0.114	0.135	0.017
	40	0.472	0.287	0.382	0.181	0.242	0.053
	60	0.529	0.327	0.444	0.228	0.311	0.090
	80	0.571	0.364	0.491	0.269	0.358	0.119
	100	0.603	0.397	0.520	0.293	0.398	0.149

test. The permutation test assumes the differences are independent, an assumption violated by data possessing long-memory.

Computing the exact p-value for a permutation test can be computationally taxing even for moderate sample sizes. The p-value can be approximated via Monte Carlo methods, using random assignments of the data to each of the two samples. The estimate of the p-value is taken to be the ratio of the number of random assignments resulting in an absolute mean difference that meet or exceed the observed difference to the number of random assignments. Since the aim of the simulation is to approximate the distribution of the p-value returned for RIA applied to a $FD(d)$ time series, Monte Carlo methods are again applied to simulate many realizations from a $FD(d)$ process and an approximate p-value is calculated for each.

Carpenter *et al.* recognized RIA is affected by autocorrelations. They simulated data from short-memory AR(1) and MA(1) processes and ran these through RIA, also checking these for true rejections when a given intervention of size ms occurred, that is, sizes that are multiples m of the standard deviation. As a result they recommend a correction to the p-value when dealing with positive autocorrelations, *i.e.*, using a declared p-value of 0.01 to get a true p-value of 0.05.

Table 3. Quartile summary of Monte Carlo distribution results of estimated RIA p-values for 10,000 random permutations of each of 1000 simulated $FD(d)$ series with the indicated values of d and n .

n	$d = 0.1$			$d = 0.3$			$d = 0.49$		
	Q1	Median	Q3	Q1	Median	Q3	Q1	Median	Q3
10	0.2100	0.4469	0.7062	0.13	0.3648	0.6584	0.09463	0.29745	0.59205
20	0.1779	0.4214	0.6966	0.08017	0.28870	0.61352	0.02045	0.14620	0.46227
50	0.1344	0.4110	0.6842	0.0237	0.1892	0.5373	0.00160	0.04795	0.34013
100	0.1515	0.3740	0.6693	0.0087	0.1196	0.4537	0.0000	0.0121	0.2185

In the simulation a permutation test was applied to each of 1000 simulated long-memory $FD(d)$ series, of the values of d and n indicated. The estimated permutation test p-values were based on 10,000 random permutations of each simulated data set. Note the simulated long-memory series contain no intervention but as mentioned do strongly violate the assumption of independent observations behind the permutation test. Estimated quartiles of the p-value distributions are summarized below in **Table 3**.

For fixed d , as the sample size n increases, the distribution becomes increasingly right-skewed, with the p-values increasingly concentrated near zero. This is also true for fixed n , as the long-memory parameter d increases. The simulation results indicate long-memory data analyzed with a permutation test will result in many false detections of trend or intervention.

4. Data Examples

As mentioned, the R package *fracdiff* [11] calculates the maximum likelihood parameters of a $FD(d)$ model, following Haslett and Raftery [12]. The log-likelihood from fitting an $FD(d)$ model to data is optional output and can be used to test for long-memory vs. the null hypothesis of white noise. Asymptotically, -2 times the log-likelihood follows a Chi-Square distribution under the null hypothesis. However, fearing slow convergence due to long-memory (for long-memory time series the rate of convergence in the Central Limit Theorem occurs at rate $n^{1/2-d}$ rather than $n^{1/2}$. (See Beran [3], Taquq [13])) it is wise to approximate the p-value using Monte Carlo methods. This is accomplished by simulating many white noise series, fitting a $FD(d)$ model to the simulated series and calculating the log likelihood. The estimated p-value is the proportion of simulated series with a value of -2 times the log likelihood which exceeds the observed value of -2 times the log likelihood. Another approximate p-value for a test for the significance of the observed value of d can be computed by calculating the proportion of simulated series with an estimated d which exceeds that observed. The exact hypotheses being tested are $H_0 : d = 0$ (white noise) vs. $H_a : d > 0$ (long-memory in the form of a $FD(d)$ model).

The following two examples were taken from the literature. They were chosen because they present instances in the literature where the BACI analysis with the

short-memory AR(1) variance correction and the RIA analysis utilizing a permutation test returns a significant intervention effect when in fact no intervention occurred. A possible explanation for this is the presence of strong correlation in the data, perhaps long-memory, which could have produced the spurious detection of a significant intervention effect. The observations in the following examples are only approximately equally spaced in time. They were assumed so in order to simplify the analyses.

4.1. Sea Urchin Data

The first example involves data read from figure 4a in Bence [8], concerning the relative abundance of the white sea urchin (*Lytechinus anamesus*) in an area offshore the San Onofre Nuclear Generating Station. The data first appeared in Schroeter *et al.* [1]. The data values are differences in log-transformed density (numbers per square meter) of white sea urchins between an impact site and a control site; a plot appears in Figure 1. It is important to note there was no intervention in the series, despite the apparent and unexplained structural break prior to mid-1981.

The analysis by Bence estimated the mean difference with a t confidence interval. He assumed an AR(1) correlation structure after the Durbin-Watson test detected significant autocorrelation. However, estimation returned a non-stationary model, ruling out the AR(1) and another indication the data may possess long-memory.

Fitting a long-memory model to the sea urchin data yielded the estimate $\hat{d}=0.3$. The value of the approximate chi-square test statistic equaled 34.66, with a Monte Carlo approximate p-value of 0.00161. The test for significance of

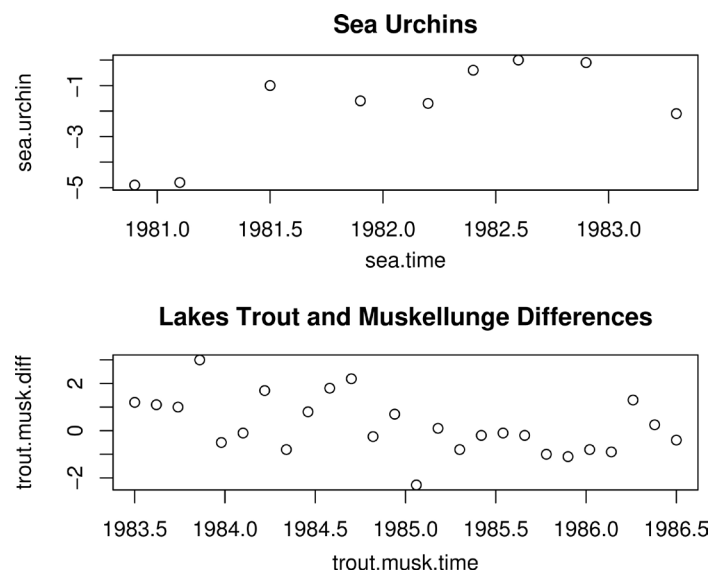


Figure 1. Approximate differences in log-transformed density (numbers per square meter) of white sea urchins between an impact site and a control site (see [1] [8]); the approximate interlake differences in chlorophyll concentrations between Big Muskellunge and Trout lakes, 1984-1986 (see Carpenter *et al.* [2]).

the long-memory parameter yielded a Monte Carlo approximate p-value of 0.01222. The long-memory corrected 95% confidence interval (using (3) with the estimated $FD(d)$ ACF using \hat{d}) for the mean difference is -1.84 ± 5.13 , indicating the mean difference is equal to 0. This compares with the (from Bence [8]) OLS estimate (-1.83 ± 1.36), conventional 2-stage estimate (-2.40 ± 3.25) and bias-corrected 2-stage estimate (-3.41 ± 26.26). Moderate long-memory in the data is a possible explanation for the non-stationary AR(1) fitted model and the unexpected structural break.

4.2. Interlake Differences

Carpenter *et al.* ([2], in figure 5) report an example using the difference in chlorophyll concentrations between two lakes (Big Muskellunge Lake and Trout Lake in the Northern Highlands Lake District of Wisconsin). This data was read from the figure and a plot appears in figure 3. RIA was significant even though no effect was evident from the mid-1985 intervention. The plot of the data reveals a possible trend; no explanation was given for this.

Durbin-Watson does not detect a statistically significant autocorrelation at lag 1 (p-value = 0.346), ruling out the AR(1). The fitted $FD(d)$ model yielded the estimate $\hat{d} = 0.15$. The value of the approximate chi-square test statistic equaled 81.21, with a Monte Carlo approximate p-value of 0.09075. The test for the significance of the long-memory parameter yielded a Monte Carlo approximate p-value of 0.05785. Weak to moderate long-memory in the data is a possible explanation of the significance of RIA, creating a false trend which was detected by the permutation test as a spurious break due to the intervention.

5. Discussion

Murtaugh [14] [15] and Stewart-Oaten [16] debated the effectiveness of the BACI and RIA designs. However, their points concerned incorrect specification of the process mean structure and not the process autocorrelation structure. An examination of RIA for correlated data is in Carpenter *et al.* [2]. However, the autocorrelations studied were short-memory and moderate at most in strength. Bence [8], recognized short-memory variance corrections may not always be adequate for ecological data, that the actual correlation structure of the process may be more elaborate than that of a short-memory process. The simulation results in Section 3 indicate short-memory variance corrections in the 2-sample t -test used in the BACI analysis are not adequate for data with long-memory.

However, the BACI design and analysis will work better than RIA in these situations because it is amenable to a simple long-memory variance correction which will improve its performance. It is also known [3] that the sample mean is nearly optimal when estimating the location parameter of a Gaussian long-memory series, in the sense that the sample mean is unbiased and its efficiency when compared to the best linear unbiased estimator is very high.

Researchers have examined the relationships among long-memory, aggregation

and structural breaks in time series [4] [5] [6]. Tests such as RIA which seek to detect breaks in a series whose true data generating process is a long memory process may result in spurious break detection. In the other direction, structural breaks in a time series may cause the manifestation of long-memory behavior. RIA would not be sensitive to structural breaks in a series since RIA does not detect the time at which a change occurred.

One solution is to detect and account for the breaks in a series, correct for them and then analyze the corrected time series. However, aggregation tests [6] for long memory depend upon very long time series, which is rarely the case in ecological experiments. Researchers should consider using long-memory corrections when short-memory corrections return non-stationary models, data exhibits persistent autocorrelation, an intervention where none occurred, or a trend with no apparent explanation.

References

- [1] Schroeter, S.C., Dixon, J.D., Kastendiek, J., Smith, R.O. and Bence, J.R. (1993) Detecting the Ecological Effects of Environmental Impacts: A Case Study of Kelp Forest Invertebrates. *Ecological Applications*, **3**, 331-350.
<https://doi.org/10.2307/1941836>
- [2] Carpenter, S.R., Frost, T.M., Heisey, D. and Kratz, T.K. (1989) Randomized Intervention Analysis and the Interpretation of Whole-Ecosystem Experiments. *Ecology*, **70**, 1142-1152. <https://doi.org/10.2307/1941382>
- [3] Beran, J. (1994) *Statistics for Long-Memory Processes*. Chapman and Hall.
- [4] Chan, K.S. and Tsai, H. (2005) Temporal Aggregation of Stationary and Nonstationary Discrete-Time Processes. *Journal of Time Series Analysis*, **26**, 613-624.
<https://doi.org/10.1111/j.1467-9892.2005.00430.x>
- [5] Granger, C.W.J. and Hyung, N. (2013) Occasional Structural Breaks and Long Memory. *Annals of Economics and Finance*, **14-2**, 721-746.
- [6] Ohanissian, A., Russell, J. and Tsay, R. (2008) True or Spurious Long Memory? A New Test. *Journal of Business & Economic Statistics*, **26**, 161-175.
<https://doi.org/10.1198/073500107000000340>
- [7] Murdoch, W.W. and Stewart-Oaten, A. (1986) Environmental Impact Assessment: "Pseudoreplication" in Time? *Ecology*, **67**, 929-940.
<https://doi.org/10.2307/1939815>
- [8] Bence, J.R. (1995) Analysis of Short Time Series: Correcting for Autocorrelation. *Ecology*, **76**, 628-639. <https://doi.org/10.2307/1941218>
- [9] Brockwell, P.J. and Davis, R.A. (1998) *Time Series: Theory and Methods*. Springer, NY.
- [10] R Development Core Team (2005) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
<http://www.R-project.org>
- [11] Fraley, C. and Leisch, F. (2012) *fracdiff: Fractionally Differenced ARIMA (p,d,q) Models*. R Package Version 1.4-2.
- [12] Haslett, J. and Raftery, A.E. (1989) Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource (with Discussion). *Applied Statistics*, **38**, 1-50. <https://doi.org/10.2307/2347679>

- [13] Fox, R. and Taqqu, M.S. (1986) Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series. *The Annals of Statistics*, **14**, 517-532. <https://doi.org/10.1214/aos/1176349936>
- [14] Murtaugh, P.A. (2002) On Rejection Rates of Paired Intervention Analysis. *Ecology*, **83**, 1752-1761. [https://doi.org/10.1890/0012-9658\(2002\)083\[1752:ORROPI\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[1752:ORROPI]2.0.CO;2)
- [15] Murtaugh, P.A. (2003) On Rejection Rates of Paired Intervention Analysis: *Reply Ecology*, **84**, 2799-2802. <https://doi.org/10.1890/03-3022>
- [16] Stewart-Oaten, A. (2003) On Rejection Rates of Paired Intervention Analysis: Comment. *Ecology*, **84**, 2795-2799. <https://doi.org/10.1890/02-3115>